Disparités spatiales et mobilité résidentielle

Laurent Gobillon

Thèse d'habilitation à diriger des recherches

Soutenue publiquement le 25 mars 2010

Université d'Aix-Marseille II

Jury :

Didier Blanchet (Rapporteur)

Pierre-Philippe Combes (Directeur)

Miren Lafourcade (Rapporteur)

Thierry Magnac

Thierry Mayer (Rapporteur)

Alain Trannoy

Version du 11/01/2010

Table des matières

Remerciements

Avec le temps, il devient difficile de remercier toutes les personnes qui m'ont soutenu ou aidé dans mes activités de recherche. Je tiens cependant à exprimer ma gratitude à Pierre-Philippe Combes qui suit ma progression depuis mon master et sur qui j'ai toujours pu compter aux différentes étapes de ma carrière. Je remercie également les membres du laboratoire du CREST pour les échanges scientifiques stimulants durant ma thèse. J'ai notamment bénéficié du soutien de Francis Kramarz qui m'a permis d'effectuer un post-doctorat à *University College London*. Je suis aussi reconnaissant à Gilles Duranton pour avoir rendu possible mon séjour en post-doctorat à la *London School of Economics* et m'avoir beaucoup apporté sur les thèmes d'économie géographique. Je remercie par ailleurs les membres de l'INED pour leur accueil et l'ambiance particulièrement chaleureuse régnant à l'institut : je peux ainsi travailler au quotidien avec une certaine paix intérieure. J'ai beaucoup apprécié les moments passés avec mes co-auteurs à réfléchir sur des sujets d'intérêt commun. Je suis particulièrement reconnaissant à Thierry Magnac pour m'avoir montré une façon plus intuitive d'aborder l'économétrie. Enfin, je remercie ma famille pour m'avoir toujours soutenu dans ma vie professionnelle.

Introduction

Depuis la Seconde Guerre Mondiale, l'aménagement du territoire est devenu une préoccupation importante des décideurs publics. Des politiques ont été mises en place pour influencer l'organisation spatiale des activités à l'échelle nationale. A l'origine, les mesures adoptées avaient un objectif d'équité et visaient à lutter contre les disparités grandissantes entre villes et campagnes. Plus récemment, elles ont changé d'objectif et ont favorisé la concentration de certaines activités. Elles ont par exemple encouragé la création de pôles de compétitivité dans une optique d'efficacité (Duranton et al., 2008), ou tenté de regrouper les moyens hospitaliers en un nombre limité de grosses structures. L'importance quantitative des effets de la concentration spatiale sur lesquels sont basées ces politiques est pourtant restée longtemps mal connue. En particulier, il n'y a pas encore de réponse définitive aux questions suivantes : le regroupement des activités a-t-elle un effet important sur la productivité? Vaut-il mieux une agglomération indifférenciée des activités ou une spécialisation sectorielle?

Parallèlement, des politiques de développement urbain ont été implémentées pour revitaliser les quartiers défavorisés dans une perspective d'équité spatiale. Ces politiques partent du présupposé que les habitants des quartiers défavorisés ont une localisation résidentielle menant à une situation défavorable sur le marché du travail. Des incitations fiscales aux entreprises ont été mises en place dans des zones franches situées dans certains quartiers pauvres. L'objectif de ces incitations est de relancer localement l'activité et de créer des emplois. Toutefois, dans quelle mesure les disparités spatiales de chômage résultent-elles effectivement d'effets locaux et non simplement d'un tri spatial des individus selon certaines caractéristiques influençant leur employabilité (comme par exemple le diplôme) ? L'implantation locale d'entreprises va-t-elle être suffisante pour résorber le chômage local ? Les politiques publiques localisées ont rarement fait l'objet d'évaluations économétriques poussées jusqu'à récemment et seuls des rapports relativement descriptifs ont essayé de mesurer leur impact.

L'objet de mes travaux est de quantifier les disparités spatiales, d'identifier leurs déterminants et d'évaluer l'effet des politiques publiques localisées. Les disparités spatiales constituent un thème à la croisée de plusieurs domaines de l'économie. Mes travaux s'inscrivent donc non seulement en économie géographique et urbaine, mais aussi en économie du travail et en économie de la santé.

Une partie de mes travaux évalue si les disparités spatiales sur le marché du travail et en santé résultent principalement d'effets de composition ou s'il existe des effets locaux qui sont quantitativement importants. Elle précise le rôle du tri spatial des individus, des entreprises ou des hôpitaux selon leurs caractéristiques observables ou inobservables. Une implication de mes travaux est aussi de pouvoir déterminer à quel point les résultats des études sur les effets locaux sont biaisés par le fait que les effets de composition ne sont pas correctement pris en compte.

Un deuxième pan de mes recherches tente d'identifier, à l'aide d'une approche structurelle, les principaux mécanismes menant à une plus grande productivité des entreprises dans les grandes agglomérations. L'approche repose sur une comparaison des distributions de productivités entre grandes villes et petites villes. La méthode d'analyse proposée peut aussi être appliquée dans d'autres contextes. Je montre en particulier comment elle peut être utilisée pour évaluer les problèmes d'accès à l'emploi des femmes en comparant les distributions de salaires des deux sexes.

Dans le cas où un regroupement d'individus ou d'entreprises dans l'espace a une influence sur les marchés locaux du travail ou des biens, par exemple en créant des interactions augmentant la productivité et les salaires, il importe de mieux comprendre comment les agents économiques choisissent leur localisation. Le tri spatial des individus provient principalement de la mobilité résidentielle et des choix de localisation sur le territoire. Une dernière partie de mes travaux s'intéresse à la mobilité des ménages. Je me focalise surtout sur le rôle de la mobilité comme moyen d'ajuster son logement au cours du cycle de vie. J'étudie aussi la mobilité liée à des raisons professionnelles sans pour autant faire de lien avec les disparités spatiales.

Plusieurs pistes de recherche peuvent être envisagées dans la continuité de mes travaux. Tout d'abord, il conviendrait d'analyser simultanément les disparités géographiques et les migrations entre bassins locaux de l'emploi. En outre, le modèle structurel sur la productivité des entreprises dans les agglomérations pourrait être complété en prenant en compte les effets de sélection spatiale des entrepreneurs selon leurs caractéristiques (observables et inobservables). Par ailleurs, les disparités spatiales en santé peuvent en partie résulter des politiques de remboursement aux hôpitaux pour les soins prodigués. Il est possible d'évaluer l'importance du mode de remboursement puisque des changements de réglementation ont eu lieu en France à partir de 2004. Enfin, les différences d'accès à l'emploi pourraient être précisées en analysant le processus de promotions à partir du fichier de paie qu'une grande entreprise accepte de mettre à disposition pour des activités de recherche.

I. Les disparités spatiales entre marchés locaux

Les économistes du travail prennent depuis longtemps en compte les différences entre individus et entreprises dans leurs études. Après avoir intégré l'hétérogénéité observée dans leurs analyses, de nombreuses techniques ont été développées pour prendre en compte les caractéristiques inobservables. Parallèlement, la littérature en économie géographique s'est développée en insistant sur les mécanismes d'agglomération et de dispersion selon les coûts de transport. Les modèles d'économie urbaine ont la plupart du temps étudié l'organisation spatiale des villes pour une main-d'œuvre homogène ou stratifiée dans une dimension (comme le revenu). Ce n'est que depuis les années 2000 que le rôle de l'hétérogénéité des agents a vraiment commencé à être précisé dans les modèles et reste un thème de recherche prisé. Mes travaux empiriques s'inscrivent dans cette perspective et revisitent des questions géographiques en mettant en évidence les effets spécifiques à l'hétérogénéité des travailleurs et des entreprises.

I.1. Les problèmes de tri spatial dans la littérature

Les travaux récents sur les disparités spatiales font généralement le lien entre un modèle micro-économique estimé sur données individuelles et des disparités géographiques mesurées à un niveau plus agrégé. Dans cette perspective, un indicateur de résultat individuel à une date donnée que l'on note ici Y_{it} est considéré. Cet indicateur peut être le salaire, une indicatrice de chômage, ou une indicatrice de décès. Le niveau de résultat d'une zone est calculé comme la moyenne de l'indicateur de résultat dans la zone. Il est ensuite possible de quantifier les disparités spatiales de résultats entre zones.

Le résultat d'un individu i donné peut dépendre de ses caractéristiques observées X_{it} et de ses caractéristiques inobservables u_i. Certains chocs individuels dépendant du temps ε_{it} sont inobservables.

Notons $a(i,t)$ la zone où est localisé l'individu à un instant donné. Le résultat de l'individu peut aussi dépendre d'indicateurs locaux $Z_{a(i,t)t}$ captant des effets d'agglomération ou de ségrégation, ainsi que des effets d'infrastructure ou de réglementation locale. Certains effets locaux $\eta_{a(i,t)t}$ sont cependant inobservables.

On peut considérer que le résultat individuel suit un modèle linéaire simple pour illustrer l'approche généralement adoptée par les articles. Ce modèle s'écrit :

$$Y_{it} = X_{it}\gamma + Z_{a(i,t)t}\delta + u_i + \eta_{a(i,t)t} + \varepsilon_{it} \qquad (1)$$

Cette équation peut être estimée à l'aide d'économétrie de panel.

Les déterminants des disparités spatiales peuvent alors être étudiés en moyennant le modèle (1) pour chaque zone. Pour une variable individuelle V_i ou V_{it}, on note $\overline{V}_{at}$ sa moyenne au niveau de la zone a à date donnée. Une fois moyenné, le modèle se réécrit :

$$\overline{Y}_{at} = \overline{X}_{at}\gamma + Z_{at}\delta + \overline{u}_{at} + \overline{\eta}_{at} + \overline{\varepsilon}_{at} \qquad (2)$$

On peut alors effectuer une analyse de la variance de l'équation (2) à la façon d'Abowd, Kramarz et Margolis (1999). Cette analyse permet d'évaluer l'importance respective des différents facteurs influençant le résultat d'une zone. Toutefois, il n'est pas possible de déterminer de façon exacte le pouvoir explicatif des différents facteurs car certaines variables sont corrélées entre elles.

A ce stade, l'analyse peut éventuellement permettre de conclure si ce sont les effets de composition $\overline{X}_{at}\gamma + \overline{u}_{at}$ ou les effets locaux $Z_{at}\delta$ qui ont le pouvoir explicatif le plus important. Il est à noter que les variances des effets inobservables $\overline{\eta}_{at}$ et $\overline{\varepsilon}_{at}$ peuvent généralement être estimées. On peut donc en déduire si les facteurs résiduels locaux et individuels ont eux aussi un rôle important à jouer.

Il est par ailleurs possible d'évaluer l'existence de tri spatial des individus selon leurs caractéristiques observables en calculant la corrélation entre $X_{it}\beta$ et $Z_{a(i,t)t}\delta$. De la même façon, il est possible d'évaluer l'existence de tri spatial des individus selon leurs caractéristiques inobservables en calculant la corrélation entre u_i et $Z_{a(i,t)t}\delta$.

Jusqu'au début des années quatre-vingt dix, les chercheurs disposaient rarement de données individuelles pour estimer des modèles du type (1). Ils étaient donc contraints de se focaliser sur des spécifications du type (2) avec toutes les limites qu'entraîne cette restriction. En particulier, les effets individuels inobservables ne peuvent être pris en compte qu'à l'aide de

données individuelles en panel. Les régressions agrégées ne peuvent pas prendre en compte ces effets et les résultats sont biaisés si les moyennes locales des effets individuels inobservables sont corrélées avec des variables locales.

Par ailleurs, se limiter à des régressions agrégées peut entraîner des problèmes d'identification. En effet, considérons par exemple que les variables individuelles observées correspondent à des indicatrices de diplôme. Leurs moyennes au niveau de chaque zone sont les parts locales de diplômes. Supposons maintenant que l'on soit intéressé par les effets d'interactions locales mesurés justement par les parts locales de diplômes. La spécification (2) ne permet pas d'identifier séparément les effets de composition et les effets d'interactions car ces deux types d'effets sont capturés simultanément par les mêmes parts locales de diplômes. En revanche, une estimation des effets au niveau individuel avec une spécification de type (1) permet de les identifier séparément. L'identification est assurée par la variation de diplôme entre individus au sein d'une zone donnée.

Dans les sous-sections suivantes, je présente mes études menées sur données individuelles qui ont permis une avancée sur la compréhension des disparités spatiales. Leur apport vis-à-vis de la littérature internationale consiste à mieux prendre en compte l'hétérogénéité inobservée soit au niveau des individus, soit au niveau des zones.

I.2. Les disparités spatiales sur le marché du travail

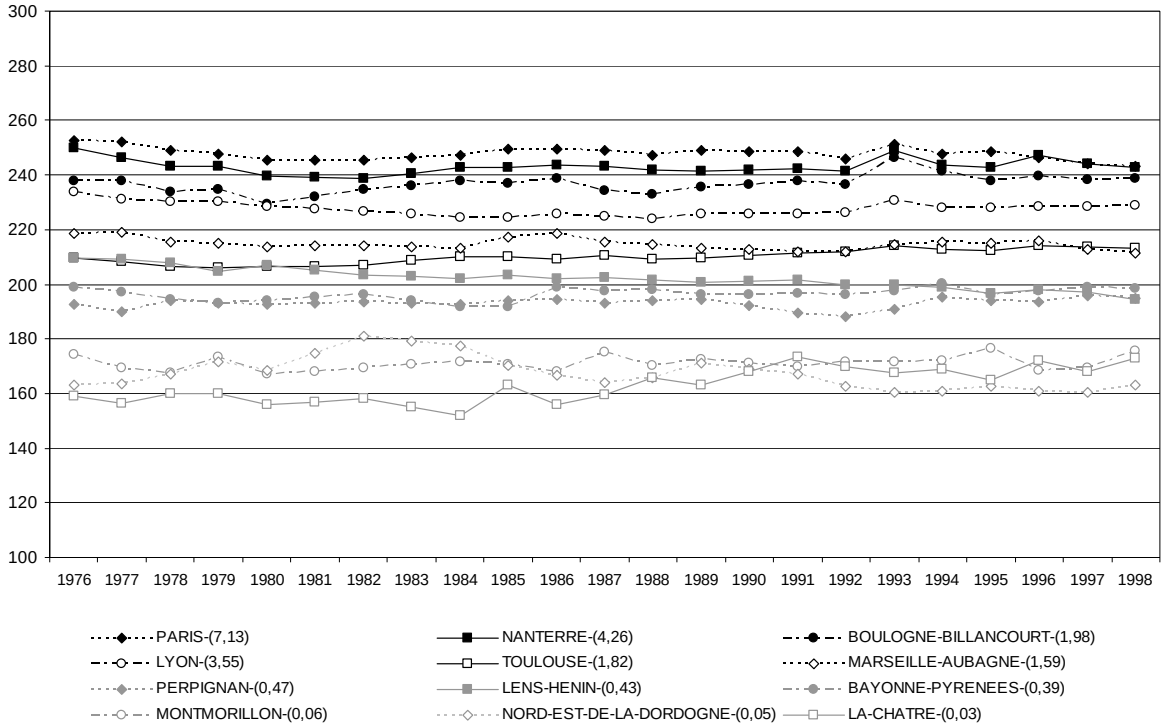
Les autorités publiques s'interrogent depuis longtemps sur les disparités territoriales de développement. Les migrations des populations des campagnes déjà peu peuplées vers les grandes villes en plein essor ont conduit le gouvernement à s'inquiéter de l'existence d'un désert français après la Seconde Guerre Mondiale. Les disparités de densité de population vont de pair avec des différences de rémunérations, les travailleurs étant généralement mieux rémunérés dans les grandes villes. On peut se demander si ces différences sont dues simplement à des différences spatiales de composition de la main d'œuvre ou a de véritables effets d'agglomération. L'essor des grandes villes n'a pas empêché l'émergence de disparités intra-urbaines avec la coexistence de banlieues résidentielles riches et de banlieues défavorisées. Là encore, on peut se demander si ces disparités sont dues à de simples effets de composition des populations ou si elles s'accompagnent de réels effets de ségrégation et des

différences d'infrastructures locales. Mes travaux essaient de préciser le rôle des effets de composition dans l'explication des disparités spatiales sur le marché du travail.

I.2.1. L'impact des effets d'agglomération sur les salaires locaux

Les disparités de coûts salariaux entre zones d'emploi sont importantes en France. La Figure 1 représente le coût journalier moyen pour un sous-échantillon de zones d'emploi sur la période 1976-1998. Le coût est environ 60% plus élevé dans certaines zones d'emploi de la région parisienne que dans d'autres plus rurales comme Montmorillon (en Poitou-Charentes), le Nord-est de la Dordogne (en Aquitaine) ou La Châtre (dans le Centre). L'ampleur des disparités est relativement persistante dans le temps. L'existence d'écarts de coût moyen entre zones d'emploi est confirmée par des indicateurs calculés sur l'ensemble des zones d'emploi françaises.

Figure 1 : Evolution du coût journalier moyen (*détrendé*)
pour un sous-échantillon de zones d'emploi



Note : Le chiffre entre parenthèses après le nom de la zone d'emploi correspond à la part de la zone d'emploi dans l'emploi total français.
Source : cf. Combes, Duranton et Gobillon (2008b).

La littérature considère habituellement que les disparités spatiales de coût correspondent à des différences spatiales de productivité. Elle propose trois types d'explications à leur existence (cf. Combes, Duranton et Gobillon, 2008a et 2008b).

Le premier type d'explication considère que les disparités spatiales de productivité sont générées par des différences de dotations physiques entre zones. En effet, certains lieux pourraient bénéficier de ressources naturelles spécifiques. La localisation sur une côte ou une voie fluviale améliorerait l'accès au marché. Le climat lui-même peut avoir un effet sur la productivité des travailleurs. La dotation physique d'une zone peut aussi prendre la forme de capital public comme des infrastructures de transport, et inclure les institutions locales. Ce type d'explication est souvent avancé par les économistes de la croissance (Temple, 1999). Pour la France, il existe assez peu de différences institutionnelles sur le territoire. Le pays se caractérise en revanche par une inégale répartition spatiale des infrastructures publiques comme les aéroports internationaux, les gares de trains à grande vitesse et les réseaux autoroutiers. De plus, le relief et les conditions climatiques sont très variés.

Le second type d'explication est l'existence d'interactions locales entre travailleurs et entreprises situés en un même lieu qui seraient à l'origine de gains de productivité conduisant à des salaires plus élevés. Ces interactions peuvent être de nature diverse.

Si l'on considère que les entreprises ne peuvent embaucher des individus qu'à proximité de leur localisation, le marché du travail peut être considéré dans une large mesure comme local. Plus la densité locale de population est forte, plus la division locale du travail par spécialisation des tâches peut être poussée et engendrer des gains de productivité. Alternativement, Helsley et Strange (1990) considèrent que l'appariement entre travailleurs et entreprises s'améliorerait avec la taille locale du marché. En effet, les chances pour une entreprise de trouver un travailleur adéquat pour un emploi donné augmentent avec le nombre de candidats. Réciproquement, les chances pour un travailleur de trouver un emploi adapté à ses compétences augmentent avec le nombre d'entreprises.

La proximité des producteurs de biens finaux et des producteurs de biens intermédiaires favorise leurs interactions en réduisant les coûts de transport, ce qui permet d'augmenter leur productivité (cf. Krugman et Venables, 1995).

Enfin, il peut exister des externalités technologiques dues à des échanges d'informations et de connaissances qui sont plus importantes dans les zones denses. Ces interactions améliorent la productivité et peuvent plus généralement être source de progrès technique favorisant la

croissance (Lucas, 1988). La littérature distingue les externalités technologiques selon qu'elles ont lieu au sein d'un même secteur (*externalités de localisation*) ou sont plutôt intersectorielles (*externalités d'urbanisation*).

A partir de données américaines, Henderson, Kuncoro et Turner (1995) montrent qu'il existe un effet des externalités de localisation sur la croissance de l'emploi. Ces résultats contrastent avec Glaeser, Kallal, Schleifer et Scheinkman (1992) qui trouvent que la spécialisation aurait plutôt un effet néfaste et que ce serait la diversité industrielle locale qui aurait un impact positif sur la croissance de l'emploi. Ciccone et Hall (1996) mettent quant à eux en évidence que les externalités d'urbanisation mesurées par la densité d'emploi ont un effet positif sur les salaires locaux. Jusqu'à ce que j'entreprenne mes travaux, il existait peu d'études sur la France. Combes (2000) montre que la spécialisation ne favorise que rarement la croissance de l'emploi local mais que la diversité peut avoir un effet positif pour certains secteurs technologiques et de services. Combes, Magnac et Robin (2004) étudient conjointement le nombre d'établissements et leur taille moyenne dans les bassins d'emploi. Ils montrent que les externalités locales captées par ces deux variables sont de nature statique plutôt que dynamique. Alors que les externalités locales ont un effet significatif sur le nombre local d'établissements, leur influence est bien moindre sur la taille moyenne des établissements.

Les interactions peuvent avoir lieu non seulement au sein d'une zone géographique, mais aussi entre zones. Typiquement, la littérature en commerce international s'interroge sur l'importance de la localisation d'une région ou d'un pays pour ses échanges commerciaux avec d'autres régions ou pays (Head et Mayer, 2004). Il convient donc d'évaluer l'effet de l'accès au marché sur la productivité locale.

Le troisième type d'explication est directement lié aux variations spatiales de la composition locale de la main-d'œuvre. En effet, lorsque les travailleurs qualifiés se concentrent en certains lieux, la productivité et les salaires y sont plus élevés. Une explication à ce tri spatial des travailleurs peut être que les différents secteurs ne nécessitent généralement pas les mêmes proportions de qualifications et ne sont pas localisés aux mêmes endroits. Les effets de composition si souvent évoqués en économie du travail ont été ignorés jusqu'aux années 2000. Les premiers travaux notables sont Glaeser et Maré (2001) pour les États-Unis et Duranton et Monastiriotis (2002) pour le Royaume-Uni. Ces deux études montrent que les variations spatiales de composition de la main-d'œuvre sont un déterminant important des disparités spatiales de salaire. Comme nous l'avons déjà souligné, ignorer les effets de composition lorsqu'on évalue l'effet des externalités locales peut conduire à des biais dans les

estimations. C'est le cas si le niveau local de qualification est corrélé avec la densité de population ou la concentration sectorielle locale. Ce type de problème a été ignoré jusqu'à une période récente.

J'ai été amené à évaluer l'importance des trois types de facteurs (dotations, interactions, et compositions de la main-d'œuvre) permettant d'expliquer les différences de salaires entre zones d'emploi. Les données utilisées sont le panel des Déclarations Annuelles des Salaires pour la période 1976-1998. Ce panel contient de l'information sur tous les épisodes d'emploi des travailleurs nés en octobre d'une année paire. L'analyse est limitée aux salariés du privé à temps plein. Les données contiennent des informations sur l'âge des individus X_{it}, leur salaire journalier Y_{it}, l'industrie (NAF114) dans laquelle ils travaillent. Elles peuvent par ailleurs être utilisées pour construire des indicateurs agrégés au niveau de la zone d'emploi. Ces indicateurs peuvent être communs à tous les secteurs comme la densité d'emploi au mètre carré ou l'accès au marché (défini comme une moyenne de l'emploi des autres zones pondéré par l'inverse de la distance à ces zones) Z_{at}. Ils peuvent aussi dépendre du secteur comme la spécialisation locale dans une industrie Z_{ast} définie comme la part de cette industrie dans l'emploi local divisée par la part de cette industrie dans l'économie.
La spécification retenue peut se décomposer en deux équations :

$$Y_{it} = X_{it}\gamma + Z_{a(i,t)s(i,t)t}\theta + \beta_{a(i,t)t} + u_i + \varepsilon_{it} \tag{3}$$

$$\beta_{at} = Z_{at}\delta + \eta_{at} \tag{4}$$

La première explique le salaire journalier avec les caractéristiques des individus (observables ou inobservables), la spécialisation au niveau zone-secteur, et des effets fixes locaux β_{at}. Les effets fixes locaux sont ensuite expliqués par la densité d'emploi et l'accès au marché, un résidu η_{at} captant les effets locaux inobservables dont les effets des aménités locales (celles-ci étant difficiles à mesurer en pratique).

L'estimation du modèle se fait en deux étapes. La première équation est estimée à l'aide des techniques de panel usuelles (estimation *within*). La seconde équation est estimée à l'aide de techniques linéaires (moindres carrés ordinaires, moindres carrés quasi-généralisés) tout en prenant en compte l'erreur d'échantillonnage sur la variable dépendante qui provient de

l'estimation de première étape. Cette approche en deux étapes a l'avantage de faciliter la prise en compte des effets locaux inobservables. En effet, une approche plus directe en une seule étape oblige à les ignorer car la structure de la matrice de covariance écrite dans la dimension intra-individuelle est très compliquée du fait des mouvements de travailleurs entre zones d'une année sur l'autre. Moulton (1990) montre que lorsqu'on ignore les effets locaux inobservables, les écarts-type des variables agrégées sont très biaisés. Ce résultat a été vérifié dans notre contexte en comparant les résultats des deux approches (estimations en une ou deux étapes).

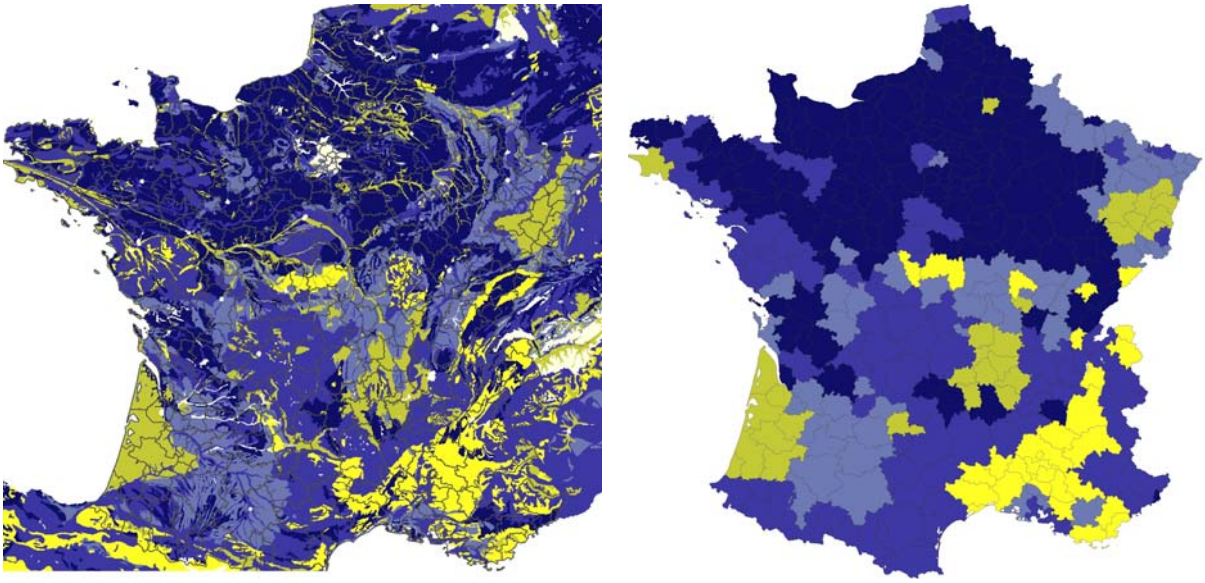
Mes travaux (Combes, Duranton et Gobillon, 2008a) montrent l'importance des effets de composition de la main-d'œuvre (en termes de caractéristiques observables et inobservables) qui expliqueraient environ 40 à 50% des disparités spatiales de salaires. La corrélation entre les effets de composition et les effets fixes locaux est significativement positive de l'ordre de 0.3, ce qui correspondrait à un tri spatial de la main d'œuvre la plus productive dans les zones les plus denses. La spécialisation locale dans certains secteurs a l'effet positif attendu sur les salaires, mais son pouvoir explicatif est faible et presque négligeable.
Lorsque les effets fixes locaux sont régressés sur le logarithme de la densité locale d'emploi, on trouve que l'effet de la densité d'emploi est positif et significatif, de l'ordre de 3.7%. Cet ordre de grandeur est bien plus faible que celui de l'article de référence de Ciccone et Hall (1996), pour qui l'effet était plutôt de l'ordre de 5%. Il est possible de vérifier que la différence d'effet provient de la prise en compte de la composition locale de la main d'œuvre dans notre cas. En effet, si les effets de composition locale n'étaient pas pris en compte dans notre régression de première étape, l'effet de la densité serait plutôt de l'ordre de 6%. Enfin, on peut noter que l'accès au marché a un effet positif sur les salaires. Les différences spatiales d'accès au marché expliquent une partie non négligeable des disparités spatiales de salaire, mais leur pouvoir explicatif est plus faible que celui des différences spatiales de densité d'emploi. Il en est de même pour les effets locaux résiduels. Il est donc probable que les infrastructures aient un rôle plus faible à jouer sur les salaires que les externalités d'agglomération.

La régression de seconde étape peut être sujette à des problèmes de biais d'endogénéité. En effet, il est possible que la productivité locale ait un effet sur la localisation des emplois et donc sur la densité d'emplois au mètre carré. Les instruments généralement utilisés pour la densité sont des variables historiques sur lesquelles le niveau de productivité a peu d'effet. Ici,

les variables utilisées sont les densités de population de la fin du 19e siècle et du début du 20e siècle qui permettent d'instrumenter par le niveau et la croissance passés de la densité. Ces instruments ont peu de chances d'être endogènes car les techniques productives ont fortement évolué en un siècle et une guerre sépare les dates des instruments et la période étudiée. Nous trouvons que l'instrumentation affecte peu nos résultats puisque l'effet de la densité reste de l'ordre de 3%.

Dans un travail plus récent (Combes, Duranton, Gobillon et Roux, 2009), d'autres instruments liés à la nature du sol sont utilisés. En effet, les caractéristiques du sol ont pu influencer les productivités agraire ou industrielle passées mais sont déconnectées de la production française d'aujourd'hui. Les instruments de sol ou sous-sol sont construits à partir des données de la European Soil Database (ESDB). Ces données reportent pour des cellules de terrain de un kilomètre sur un kilomètre : la nature du minerai qui constitue le sol, la capacité de rétention d'eau, l'érodabilité, la profondeur de la roche, le contenu en carbone, et les caractéristiques hydrologiques. Les cellules géographiques sont réagrégées au niveau de la zone d'emploi pour pouvoir servir d'instrument. A titre d'exemple, la Figure 2 représente la capacité en eau du sous-sol avant et après agrégation au niveau des bassins d'emploi :

Figure 2 : Capacité en eau du sous-sol avant et après agrégation par zone d'emploi



Panel (a). Données d'origine

Panel (a). Données transformées

Note : dans les deux panels, le noir correspond à « très élevé » (plus de 190mm), le gris foncé à « élevé » (entre 140 et 190 mm), le gris à « moyen » (entre 100 et 140 mm), le gris clair à « faible » (entre 5 et 100 mm), et le gris très clair à « très faible » (entre 0 et 5 mm). Les valeurs manquantes pour la zone autour de Paris sont représentées en blanc.
Source : cf. Combes, Duranton, Gobillon et Roux (2009).

Lorsque la densité d'emploi n'est instrumentée que par les variables de sol, le coefficient obtenu pour la densité est de l'ordre de 4-5%, donc supérieur à celui obtenu lorsque les instruments retenus sont les variables de population passée. Toutefois, les instruments sans être faibles, ne sont pas d'aussi bon prédicteurs de la densité que les variables de population passée. Lorsque les deux jeux de variables sont utilisés comme instruments, on trouve à nouveau un coefficient de la densité de l'ordre de 3%. Comme on dispose de deux jeux d'instruments différents, il est possible de faire des tests de sur-identification pour tester leur validité. Les tests montrent de façon générale que nos instruments sont valides.

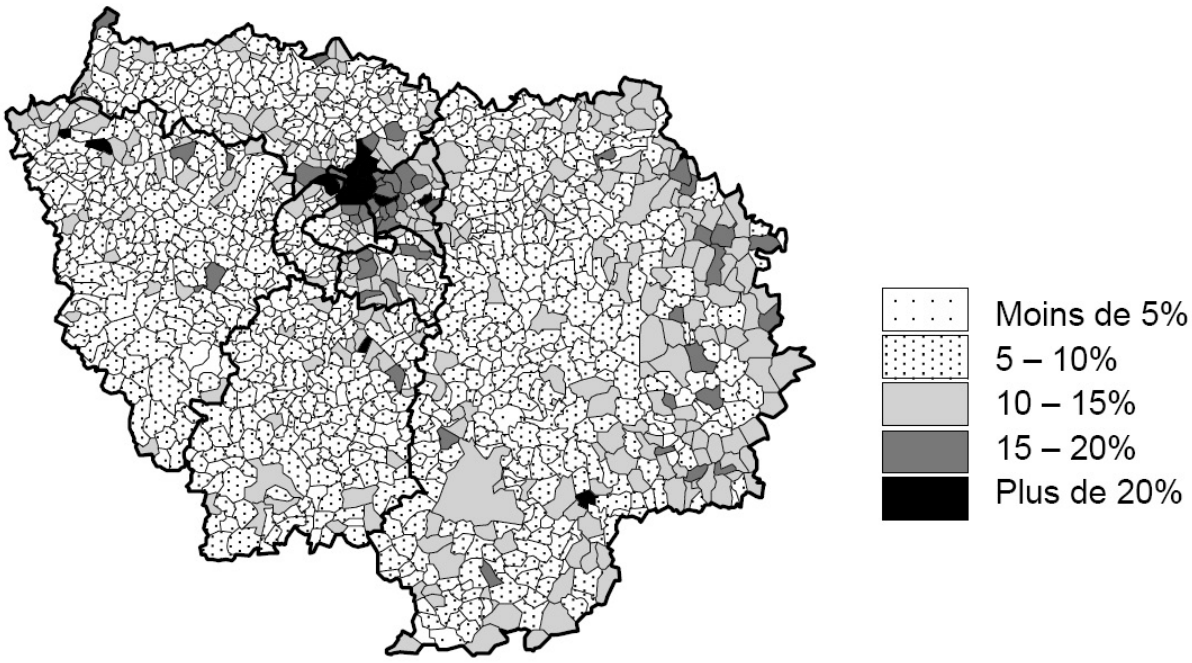
En résumé, les disparités spatiales de salaire sont principalement le résultat d'effets de composition. Alors que les externalités d'agglomération et l'accès au marché ont aussi un rôle important à jouer, les effets de spécialisation sont moindres tout comme les effets d'aménités. Des politiques publiques souhaitant rééquilibrer les salaires par soucis d'équité devraient tenter d'affecter la localisation des individus et des emplois. Une façon d'influencer la localisation de la main-d'œuvre qualifiée est de développer des centres d'enseignement supérieur en certains lieux. Toutefois, on peut se demander quels sont les effets que peut entraîner ce type de politique publique en termes d'efficacité. En effet, la densité d'emploi qui caractérise les grandes villes est généralement le résultat d'un processus d'agglomération dépendant de la présence de main d'œuvre qualifiée. Si les centres d'enseignement supérieur sont transférés des grandes villes à des villes plus petites, il peut s'ensuivre une baisse de la densité locale dans certains endroits et une baisse des effets d'agglomération.

I.2.2. Les effets locaux dans les différences spatiales de chômage

Les disparités spatiales sur le marché du travail peuvent non seulement concerner les salaires mais aussi le chômage. Elles peuvent se manifester non seulement entre bassins d'emploi ou entre agglomérations, mais aussi entre communes au sein d'une même agglomération.
La France a connu des émeutes en novembre 2005 avec plus de 9000 voitures brûlées en trois semaines et plus de 2900 personnes interpellées pendant la même période. Ces émeutes ont débuté à Clichy-sous-Bois. Des émeutes ont aussi eu lieu en 2007 à Villiers-le-Bel. Il est possible de vérifier que les émeutes concernent tout d'abord des communes des banlieues défavorisées de Seine Saint-Denis où le taux de chômage est élevé.

La Figure 3 représente le taux de chômage communal en région Ile-de-France. Elle met en évidence les fortes disparités qui existent entre le Nord-Est de Paris / Seine Saint-Denis où le taux de chômage communal est supérieur à 15% voire 20%, et l'ouest parisien / Yvelines où le taux de chômage communal est souvent inférieur à 5%. Cette différence correspond à une séparation spatiale entre les banlieues pauvres du Nord-Est et les banlieues résidentielles riches de l'Ouest.

Figure 3 : Taux de chômage communaux en Ile-de-France, calculés à partir du Recensement de 1999

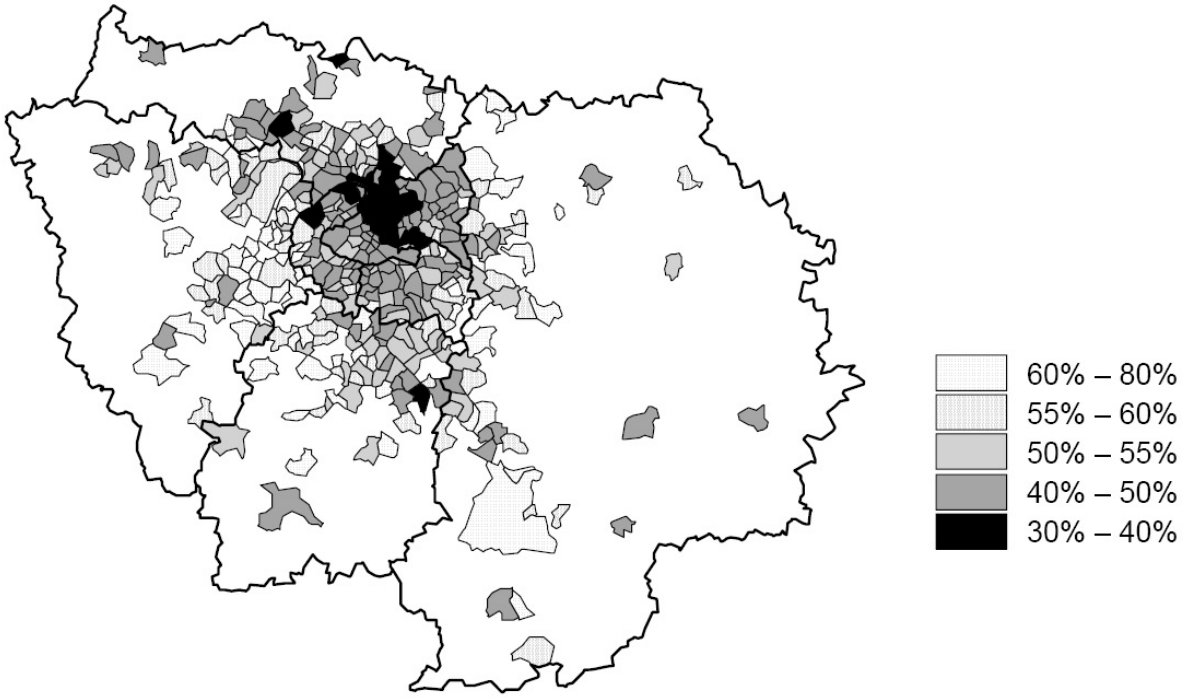


Source : cf.Gobillon, Magnac et Selod (2009a).

Une importante littérature s'interroge sur les explications du chômage local et met en avant les effets de ségrégation et de problèmes physiques d'accès à l'emploi. Alors que les études empiriques sur ce thème datent d'environ cinquante ans, les travaux théoriques sont plus récents et se sont développés depuis les deux dernières décennies. Les analyses reposent pour la plupart sur des mécanismes dynamiques comme les mécanismes de recherche d'emploi (Pissarides, 2000) qui trouveraient une contrepartie empirique dans les épisodes de chômage et les chances de retour à l'emploi. La Figure 4 représente donc la probabilité d'avoir retrouvé un emploi au bout de 24 mois selon la commune de la région parisienne (elle est calculée à partir d'un estimateur de Kaplan-Meier prenant en compte la censure). Le tableau d'ensemble reste le même que dans le cas du taux de chômage : les chômeurs habitant les communes du

Nord-Est ont plus de difficultés à retrouver un emploi que les chômeurs habitant les banlieues résidentielles de l'Ouest.

Figure 4 : Probabilités communales brutes d'avoir trouvé un emploi dans les 24 premiers mois pour les chômeurs inscrits à l'ANPE



Note : La probabilité de retrouver un emploi n'est calculée que pour les communes de plus 5000 habitants pour assurer la fiabilité statistique des résultats.
Source : cf.Gobillon, Magnac et Selod (2009a).

Il est possible de distinguer les mécanismes pouvant expliquer ces disparités spatiales selon l'étape du processus de recherche d'emploi auquel ils ont lieu : la prise de contact avec un employeur potentiel, la proposition d'emploi, l'acceptation ou le refus de l'offre par le chômeur. Si en moyenne, les chômeurs ont localement moins de contacts, moins de propositions d'emploi ou sont amenés à refuser plus souvent les offres, les retours à l'emploi seront localement moins fréquents.

Les caractéristiques des individus peuvent directement avoir un effet sur le retour à l'emploi en influençant les trois étapes du processus de recherche d'emploi. Par exemple, un travailleur qualifié pourra avoir plus de contacts avec des employeurs potentiels s'il est plus efficace à recueillir de l'information sur les emplois vacants ou si la demande de travail est plus importante pour les travailleurs qualifiés que pour les travailleurs peu qualifiés. Il est aussi susceptible de recevoir plus d'offres d'emplois en écrivant de meilleures lettres de motivation

ou en étant plus convaincant durant les entretiens. Cependant, il est possible qu'un travailleur qualifié rejette une offre plus facilement car il est susceptible d'avoir ou d'anticiper plus d'offres alternatives. D'autres caractéristiques comme le sexe, l'âge, la situation familiale ou la nationalité peuvent aussi jouer sur le processus de recherche d'emploi. Ainsi, les effets de composition constituent un déterminant potentiel des disparités spatiales de sortie du chômage.

Les problèmes d'accès physique à l'emploi et de ségrégation peuvent aussi intervenir aux trois étapes de la recherche d'emploi.

Les problèmes d'accès physique à l'emploi peuvent diminuer le nombre de contacts des chômeurs avec les employeurs car les zones éloignées des emplois proposent généralement peu d'emplois vacants. Les chômeurs peuvent évidemment chercher à distance de leur lieu de résidence. Toutefois, il est probable que leur recherche soit moins efficace du fait d'un manque d'information sur les offres d'emploi distantes. Turner (1997) montre par exemple qu'aux Etats-Unis l'information sur les offres est souvent locale puisque les entreprises postent surtout leurs offres sur les vitrines ou dans les journaux locaux. Les chômeurs vivant à distance des offres doivent donc se déplacer pour pouvoir obtenir l'information sur les offres. Un manque d'information peut aussi amener les chômeurs à chercher dans de mauvais endroits proposant des emplois inadaptés à leurs compétences (Ihlandfeldt, 1997).

En outre, rechercher un emploi à distance du lieu de résidence est coûteux en temps et en argent, et peut pousser certains chômeurs à restreindre leur périmètre de recherche à proximité de leur domicile. Dans ce contexte, l'accès à une voiture ou à des transports en commun peut limiter les coûts de recherche et pousser les chômeurs à augmenter leur périmètre de recherche (Stoll, 1999).

Les incitations à chercher un emploi peuvent aussi dépendre de la distance aux centres d'emploi. En effet, si on omet les banlieues résidentielles riches, le coût du logement décroit généralement avec la distance aux emplois. Les chômeurs habitant loin des centres d'emplois doivent généralement s'acquitter d'un loyer moins élevé et ont donc moins d'incitations à chercher un emploi de façon active (Smith et Zenou, 2003).

La distance aux emplois peut aussi réduire l'occurrence des offres. C'est le cas lorsque les employeurs sont moins enclins à faire une offre à un travailleur habitant loin de l'entreprise parce que ses migrations alternantes pourraient rendre le travailleur moins productif. En effet,

le travailleur aurait plus de chances d'arriver en retard ou d'être fatigué du fait des trajets dans les transports (Zenou, 2002).
Enfin, la distance à une offre d'emploi peut réduire la probabilité d'acceptation. En effet, certains chômeurs pourraient refuser une offre s'ils anticipent des migrations alternantes trop coûteuses par rapport au salaire proposé (Zax et Kain, 1996). En d'autres termes, ils refuseraient l'offre si le salaire proposé net des coûts de transport est en-dessous de leur salaire de réserve.
es mécanismes à l'œuvre sont présentés plus en détail dans Gobillon, Selod et Zenou (2007) qui proposent une synthèse théorique de la littérature.

La ségrégation résidentielle peut avoir un effet négatif sur la prise de contact avec des employeurs potentiels car le réseau social est généralement assez local. Quand la part de travailleurs déconnectés du marché du travail (comme les chômeurs) est importante dans une zone, les travailleurs ont moins de chance de connaître des voisins qui travaillent et pourraient les renseigner sur les emplois vacants (Calvo-Armengol et Jackson, 2004).
Même si un individu d'une zone ségréguée réussit à prendre contact avec un employeur, ses chances de recevoir une offre peuvent être plus faibles (Wilson, 1996). En effet, il est possible qu'il soit discriminé sur la base du lieu où il réside. Cette pratique est connue sous le nom de discrimination territoriale ou délit de sale adresse. Ce type de discrimination peut être lié aux goûts de l'employeur, par exemple s'il n'aime pas la manière d'être de la population de certains quartiers (à cause de leur façon de parler, de s'habiller, etc.). Il peut aussi être lié aux goûts des clients de l'employeur qui n'ont pas envie d'interagir avec les habitants de certains quartiers. Enfin, l'employeur peut faire de la discrimination statistique envers l'individu par manque d'information sur ses caractéristiques. Dans ce cas, il attribue à l'individu les caractéristiques moyennes des habitants de son quartier. L'employeur peut alors décider de ne pas faire d'offre d'emploi à l'individu si les travailleurs localisés dans son quartier de résidence sont en moyenne moins productifs ou plus souvent criminels.

Dans Gobillon, Magnac et Selod (2009a), nous tentons d'expliquer les disparités spatiales de retour à l'emploi par trois types de facteurs : les effets de composition, les problèmes d'accès physique à l'emploi et les effets de ségrégation.
Les données utilisées sont celles du Fichier Historique de l'ANPE (1993-2003) qui regroupent l'information sur tous les épisodes de chômage pour les chômeurs inscrits. Comme il est nécessaire d'être inscrit à l'ANPE pour toucher les indemnités chômage, environ 90% des

chômeurs sont présents dans le fichier. Les jeunes n'ayant jamais travaillé sont sous-représentés. Nous faisons de l'échantillonnage en flux et ne conservons que les chômeurs dont l'épisode de chômage débute le premier semestre de l'année 1996. Ces chômeurs sont cependant suivis sur l'ensemble de la période. Les épisodes de chômage peuvent se terminer par un retour à l'emploi, la passage en inactivité, ou une censure correspondant principalement à une absence aux contrôles ou à la fin de la période d'étude.

Les données permettent de prendre en compte les effets de composition grâce à des variables socio-économiques comme le diplôme, l'âge, la situation familiale (vie en couple, nombre d'enfants), l'existence d'un handicap, la nationalité. Elles contiennent aussi la commune de résidence qui permet d'identifier la localisation des chômeurs en Ile-de-France. Les données sont appariées avec des variables communales construites à partir du recensement de 1999 et permettant de capter des effets de ségrégation : les proportions communales des différents groupes de nationalités et de diplômes. Elles sont aussi appariées avec des indicateurs d'accès physique à l'emploi en transport en commun et en véhicule privé. Pour une commune donnée, l'accès à l'emploi est mesuré par le rapport entre le nombre d'emplois et la population dans la zone géographique accessible depuis la commune en moins de quarante-cinq minutes aux heures de pointe le matin par le type de transport considéré. La limite de quarante-cinq minutes a été choisie pour être supérieure à la durée moyenne de migrations alternantes en Ile-de-France qui est de trente-quatre minutes.

Afin de déterminer l'importance des différents déterminants des disparités spatiales de retour à l'emploi, nous considérons un modèle de durée à hasard proportionnel où le hasard de base est spécifique à chaque commune. La sortie du chômage considérée est le retour à l'emploi, les sorties vers l'inactivité et les absences aux contrôles étant assimilées à des censures. Pour un individu i de caractéristiques X_i habitant la commune $j(i)$, la fonction de hasard s'écrit :

$$\lambda\left(t \middle| X_i, j(i)\right) = \lambda_{j(i)}(t)\exp(X_i\beta) \tag{5}$$

où $\lambda_j(t)$ est le hasard de base pour la commune j. Cette équation permet de séparer les effets de composition et les effets propres à chaque commune. Les coefficients des variables individuelles sont estimés en maximisant la vraisemblance partielle stratifiée par commune

(Ridder et Tunali, 1999). La fonction de hasard intégré de chaque commune est ensuite estimée en utilisant l'estimateur de Breslow (1974).

Dans une seconde étape, il est possible de résumer le hasard de chaque commune par un hasard de base $\lambda(t)$ et un effet fixe local α_j :

$$\lambda_j(t) = \alpha_j \lambda(t) \tag{6}$$

L'intuition de la méthode utilisée pour estimer les effets fixes communaux est de remplacer le membre de gauche par un estimateur de première étape, passer en logarithme et utiliser des techniques d'estimation pour modèles linéaires (en réalité, la méthode d'estimation s'avère un peu plus compliquée et est présentée en détails dans l'article).

Il est finalement possible dans une troisième étape d'expliquer les effets fixes locaux à l'aide des indicateurs de ségrégation et d'accès à l'emploi Z_j et des facteurs locaux inobservables η_j :

$$\ln \alpha_j = Z_j \gamma + \eta_j \tag{7}$$

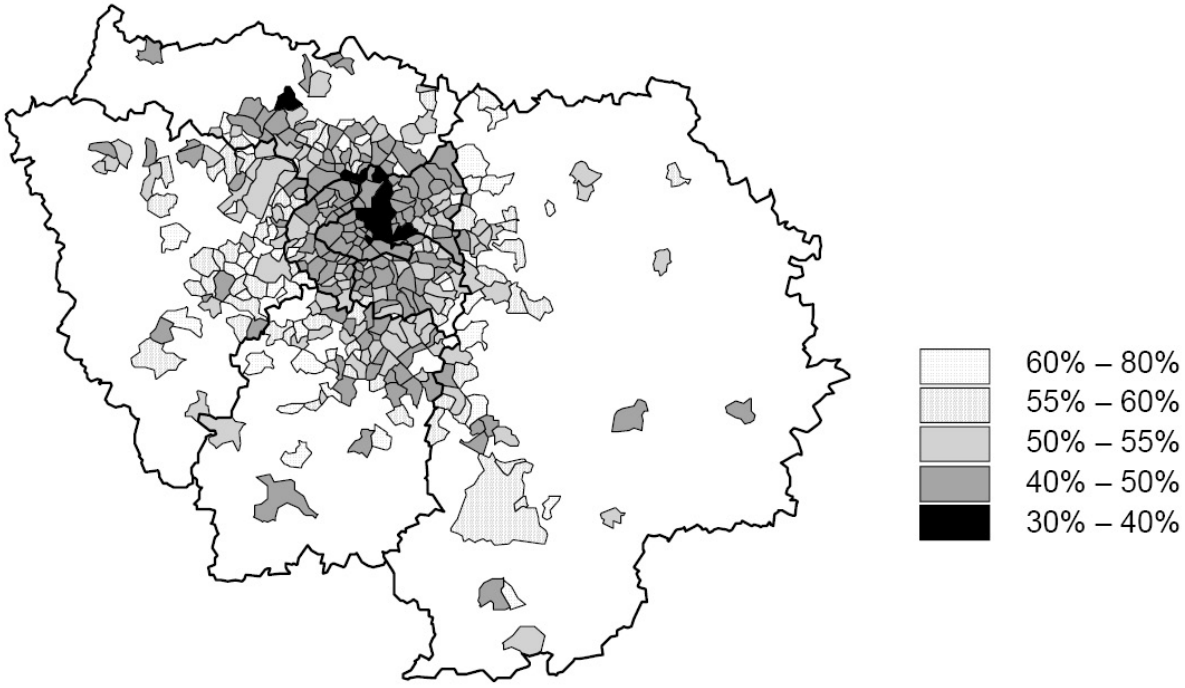
Cette équation est estimée à l'aide de techniques linéaires.

L'estimation de l'ensemble des équations permet d'évaluer séparément les effets de composition et les effets locaux, eux-mêmes décomposés en effets de ségrégation et d'accès à l'emploi. On peut ensuite agréger le modèle au niveau communal et effectuer une analyse spatiale de la variance pour évaluer dans quelle mesure les différents facteurs expliquent les disparités spatiales de retour à l'emploi.

Les variables explicatives ont l'effet attendu sur le retour à l'emploi : les femmes retrouvent moins souvent un emploi, tout comme les personnes célibataires et les peu diplômés. Il est à noter qu'il existe un fort effet négatif pour les handicapés et les Africains, qu'ils soient maghrébins ou viennent d'un pays subsaharien. La Figure 5 représente les probabilités communales de trouver un emploi dans les vingt-quatre premiers mois nettes des effets de composition. Une comparaison avec la Figure 4 montre que les disparités spatiales nettes sont plus faibles que les disparités spatiales brutes. Toutefois, il existe toujours un contraste entre les communes du Nord-Est où le retour à l'emploi est difficile et les communes de l'Ouest où il est plus facile. En fait, on peut montrer que seulement environ 30% des disparités spatiales

de retour à l'emploi peuvent être expliquées par des effets de composition. Les individus trouvant plus facilement un emploi du fait de leurs caractéristiques observables habitent très souvent des communes avantageuses puisque la corrélation entre effets individuels et effets locaux est de l'ordre de 0.4 – 0.5.

Figure 5 : Probabilités communales nettes d'avoir trouvé un emploi dans les 24 premiers mois pour les chômeurs inscrits à l'ANPE



Note : La probabilité de retrouver un emploi n'est calculée que pour les communes de plus de 5000 habitants pour assurer la fiabilité statistique des résultats.
Source : cf.Gobillon, Magnac et Selod (2009a).

Concernant les 70% de disparités spatiales restantes, environ 60% peuvent être expliquées par des indices de ségrégation. En particulier, la part communale des Africains (maghrébins ou subsahariens) a un effet négatif fort sur le retour à l'emploi. Il est cependant difficile de dire si cet effet provient de réels effets de ségrégation ou d'un tri spatial des chômeurs selon leurs caractéristiques inobservables tel que ceux qui ont moins de chances de retrouver un emploi habitent dans des communes où la part d'Africains est plus élevée. En effet, le modèle de Cox permet difficilement de prendre en compte l'hétérogénéité individuelle inobservable. Cependant notre spécification est déjà très flexible avec un hasard de base non spécifié. Baker et Melino (2000) montrent qu'avoir une spécification non-paramétrique tant du hasard que de l'hétérogénéité individuelle inobservée (sous forme de points de masse) rendrait de toute

façon instable l'estimation du modèle et les résultats obtenus ne seraient pas fiables. Enfin, le pouvoir explicatif des variables d'accès physique à l'emploi est relativement faible et les coefficients des variables n'ont pas forcément le signe attendu. La densité d'emploi en véhicule privé a un effet négatif sur le retour à l'emploi, ce qui est en désaccord avec les théories sur les problèmes d'accès physique à l'emploi. Une explication possible est qu'il existe des banlieues résidentielles aisées dans l'ouest parisien qui sont situées à distance des emplois mais dont les habitants chômeurs ont des caractéristiques particulières leur permettant de retrouver facilement un emploi.

Il est possible de conduire des estimations spécifiques pour différents groupes de nationalité (Français, Maghrébins, Africains Sub-sahariens) afin d'évaluer si l'espace joue différemment pour chaque groupe (Gobillon, Magnac et Selod, 2009b). Les résultats montrent que les disparités spatiales sont importantes pour tous les groupes de nationalités. Il existe par ailleurs un tri spatial des Africains désavantagés par leurs caractéristiques observables dans les communes désavantagées. Il est aussi possible de calculer la corrélation entre les effets communaux des Français et des Africains. Cette corrélation est élevée et de l'ordre de 0.6. Les communes défavorables au retour à l'emploi pour les français seraient donc souvent les mêmes que celles qui sont défavorables pour les Africains.
Enfin, l'effet négatif de la part locale d'Africains Sub-sahariens sur les effets fixes communaux est bien plus faible pour les Africains Sub-sahariens (et même non significatif) que pour les français. Cette différence suggère des effets de réseaux pour les Africains Sub-sahariens qui bénéficieraient de la présence de populations de même nationalité.

De façon générale, mes travaux suggèrent l'importance de la dimension spatiale dans le retour à l'emploi. Les pouvoirs publics peuvent alors s'interroger sur les politiques permettant de réduire ces disparités. Comme les logements publics sont fortement concentrés dans les banlieues défavorisées, une possibilité pourrait être de diversifier leur localisation en encourageant leur implantation dans des communes plus aisées. Une partie des populations défavorisées pourrait alors se retrouver dans des zones plus favorables à l'obtention d'un emploi. La loi relative à la solidarité et au renouvellement urbains (SRU) mise en place en 2000 avait pour objectif d'imposer un pourcentage de 20% de logements sociaux dans toutes les grandes communes sous peine d'une amende annuelle. Cette mesure s'est révélée inefficace, nombre de communes riches préférant augmenter leurs taxes locales afin de payer l'amende plutôt que d'augmenter leur nombre de logements sociaux.

I.3. L'effet des politiques locales sur l'emploi

L'amplitude des disparités spatiales sur le marché du travail est influencée non seulement par les effets de composition, d'infrastructures, d'agglomération ou d'interactions, mais aussi par les politiques localisées.
En particulier, les sites d'un territoire peuvent présenter un niveau d'attractivité pour les entreprises très différent selon le niveau local des taxes. Par ailleurs, même si le taux de taxation est uniforme sur l'ensemble du territoire pour une taxe donnée, il peut exister des crédits locaux dans certaines zones qui les rendent plus attractives.

En France, les Zones Franches Urbaines (ZFU) ont été mises en place en 1997, 2004 et 2006, pour favoriser l'implantation des entreprises sur certains sites et la croissance locale de l'emploi. Les mesures en faveur des entreprises consistent en l'exonération d'une partie de la taxe foncière, de l'impôt sur les bénéfices et des charges sociales. Toutefois ces mesures ne garantissent pas forcément une résorption du chômage local car les entreprises peuvent souhaiter employer des travailleurs d'autres quartiers. En particulier, le niveau de qualification des travailleurs habitant les ZFU peut ne pas être adapté aux besoins des nouvelles entreprises. Bien qu'en 1997, des closes de recrutement local aient été imposées comme conditions pour bénéficier d'exonérations fiscales, elles ont rarement été contrôlées en pratique durant les premières années suivant la mise en place des Zones Franches Urbaines. Des travaux ont montré que les ZFU auraient un effet positif sur la création locale d'établissements et d'emplois (Rathelot et Sillard, 2009), mais que leur effet sur le retour à l'emploi des chômeurs habitant les communes traitées resterait assez mitigé (Gobillon, Magnac et Selod, 2009c).

En Grande-Bretagne, il existe une taxe foncière dépendant de la valeur des bâtiments utilisés par les entreprises. Jusqu'au début des années quatre-vingt-dix, les autorités des juridictions anglaises fixaient le taux de taxation au niveau local. Je me suis intéressé dans Duranton, Gobillon et Overman (2009) à l'effet de la taxation locale des bâtiments sur l'implantation locale des entreprises et la croissance locale de l'emploi.
La taxation locale a fait l'objet d'une littérature théorique assez importante sur la concurrence entre juridictions par la fixation du niveau local de taxe. En particulier, l'analyse a porté sur la capacité de la concurrence à entraîner un niveau de taxe permettant de financer un niveau efficace de biens publics (cf. Epple and Nechyba, 2004). Les économistes urbains se sont

aussi interrogés sur la capitalisation des taxes locales et le niveau optimal de taxation (cf. Fujita, 1989).

La littérature empirique date de cinquante ans et a évolué au cours du temps. A l'origine, les études s'intéressaient à des unités spatiales de taille importante comme les états américains pour lesquels des données étaient disponibles (Bartik, 1991). Les résultats obtenus suggèrent une relation positive entre le niveau local de taxation et l'entrée des entreprises. Des papiers plus récents (Guimaraes, Figueiredo et Woodward, 2004) se sont intéressés à des unités spatiales plus petites et trouvent des résultats similaires. Cependant, plusieurs problèmes méthodologiques sont ignorés dans la littérature. Tout d'abord, les entreprises souhaitant choisir un lieu d'implantation ont le choix entre un très grand nombre de sites dont la plupart des caractéristiques sont inobservables et peuvent être corrélées avec des caractéristiques des entreprises et le niveau local de taxation. Lorsque les caractéristiques inobservables des sites ne sont pas intégrées à l'analyse, les résultats peuvent être biaisés. Par ailleurs, les entreprises elles-mêmes sont hétérogènes. Il peut y avoir un tri spatial des entreprises selon leurs caractéristiques inobservables comme l'efficacité du manager. Lorsque cette hétérogénéité n'est pas prise en compte, les résultats peuvent là encore être biaisés. Enfin, certains aspects du système de taxation peuvent être endogènes aux décisions des entreprises, et les estimations peuvent souffrir d'un biais de causalité inverse.

Dans notre article, nous proposons d'étudier pour la Grande-Bretagne sur la période 1984-1989, l'effet de la taxation locale sur l'implantation des entreprises et la croissance locale de l'emploi. Nous utilisons une méthodologie permettant de prendre en compte les biais causés par les caractéristiques inobservables des lieux et des entreprises ainsi que par les problèmes de causalité inverse. Plus précisément, nous nous intéressons à une taxe sur les propriétés non résidentielles et examinons son effet sur l'entrée et la croissance des établissements des secteurs manufacturiers.

Un modèle simple est développé pour expliquer comment la taxation locale peut influencer les décisions des établissements. On considère que les établissements d'une juridiction donnée produisent un bien final nécessitant des quantités de travail et de bâtiments complémentaires. Le coût de location des bâtiments inclut une taxe payée par les occupants (et non les propriétaires). En accord avec le contexte de la Grande-Bretagne, cette taxe est le produit de la valeur historique des bâtiments et d'un taux variant avec la juridiction. Les établissements existent durant deux périodes. Ils s'implantent en première période et subissent un choc de demande en seconde période. Ce choc est tiré dans une distribution donnée et varie d'une entreprise à l'autre. Les établissements réagissent différemment selon l'ampleur du choc.

Dans tous les cas, nous considérons qu'ils ne peuvent pas renégocier le loyer, ce qui correspond au type de contrats généralement signés puisque les loyers ne sont le plus souvent révisés que tous les cinq ans. Même au moment des révisions des loyers, la latitude pour l'ajustement est limitée. Ainsi, les établissements signent en première période un contrat pour un loyer qui est le même aux deux périodes.

Les établissements subissant un choc négatif souhaitent sortir du marché ou diminuer leur taille. En fait, ils ne sortent du marché que si leur choc est inférieur à la borne conduisant à un profit nul sur le marché en seconde période. Si le choc est supérieur à cette borne, l'établissement sous-loue une partie de son bâtiment. Les établissements subissant un choc positif ont quand à eux trois options : garder leur taille, augmenter leur taille sur le site occupé, ou augmenter leur taille après s'être relocalisés sur un site où les loyers sont plus faibles. Les arbitrages d'un établissement sont effectués en tenant en compte du fait qu'une augmentation de la taille des bâtiments loués entraîne une réévaluation des bâtiments et donc une augmentation de la taxe qu'il doit payer tant sur la partie des bâtiments déjà occupée que sur la partie supplémentaire. L'augmentation de la taille d'un établissement n'a donc lieu que si le choc est assez important pour assurer un profit supérieur à celui obtenu par l'établissement s'il reste de même taille. La relocalisation n'a lieu que s'il existe des opportunités intéressantes en termes de loyer sur un autre site pour compenser les coûts de relocalisation.

Dans ce contexte, une augmentation de la taxe locale entre les deux périodes entraîne plus de sorties du marché. En effet, une plus grande proportion des établissements faisant face à un choc de demande négatif est incapable de survivre sur le marché. Elle entraîne aussi une diminution du nombre d'entreprises augmentant leur taille dans la juridiction car une réévaluation devient plus coûteuse. Enfin, une augmentation de la taxe locale entraîne une augmentation du nombre d'entreprises souhaitant se relocaliser car plus d'entreprises ayant intérêt à croître souhaitent changer de site pour bénéficier d'un loyer plus avantageux.

Ainsi, l'augmentation de la taxe locale dans une zone a un effet ambigu sur la croissance de l'emploi des établissements survivants de cette zone. L'effet n'est négatif que si la sortie du marché des « mauvais » établissements (en termes de croissance) est dominée par la relocalisation des « bons établissements » et le ralentissement des établissements restant dans la zone. Notre travail empirique permet d'estimer la résultante des effets, mais pas d'identifier séparément les effets de sélection et de ralentissement. De plus, comme l'augmentation de la taxe entraîne plus de sorties et de relocalisations libérant de l'espace, il devrait y avoir plus d'entrées de nouvelles entreprises.

Dans notre application empirique, l'emploi d'un établissement e_{it} est régressé sur le taux local de taxation $r_{a(i,t)t}$ où $a(i,t)$ est la localisation de l'établissement, tout en contrôlant pour les caractéristiques observables X_{it} et inobservables u_i (effets fixes) des établissements, ainsi que les caractéristiques inobservables locales à un niveau très fin $\theta_{z(i,t)t}$ (effets fixes) où $z(i,t)$ est la localisation à une échelle très fine :

$$e_{it} = \gamma r_{a(i,t)t} + X_{it}\beta + u_i + \theta_{z(i,t)t} + \varepsilon_{it} \tag{8}$$

Cette équation peut être estimée avec des méthodes linéaires après avoir fait disparaître les deux sources d'hétérogénéité inobservée par différenciation. En effet, il est possible de faire disparaître l'hétérogénéité spatiale inobservée en construisant des paires d'établissements très proches (à moins d'un kilomètre) et en considérant la différence de leur emploi. Si on considère que le terme d'hétérogénéité spatiale ne varie pas de façon brusque dans l'espace, il disparaîtra lorsqu'on fait de la différenciation spatiale. L'équation obtenue contient encore la différence spatiale des effets fixes des établissements. Il est possible de faire disparaître cette source d'hétérogénéité par différenciation temporelle en utilisant l'estimateur intra pour les paires d'établissements. Ainsi, le modèle est estimé en différence de différence. Il est à noter que cette procédure ne fait pas disparaître les problèmes de causalité inverse qui existent si la croissance de l'emploi a un effet sur l'évolution de la taxation locale. Pour prendre en compte ce type de problèmes, la double différence du taux local de taxation est instrumentée par la double différence de variables concernant les partis politiques locaux. Ces variables sont les parts des politiciens locaux affiliés aux trois partis principaux (les conservateurs, les travaillistes et les démocrates).

L'estimation est effectuée sur le panel d'établissements anglais *Annual Respondent Database* (ARD) sur la période 1994-1989. A partir de 1990, le taux de taxation devient en effet uniforme sur le territoire. Une particularité importante de la base de données est qu'elle contient le code postal des établissements. En Grande-Bretagne, ce code permet d'identifier avec une grande précision la localisation des établissements (à cent mètres près).

Lorsque nous estimons l'équation d'emploi (8) en niveau, nous trouvons que le taux local de taxation a un effet positif significatif sur l'emploi. L'effet reste positif quand l'estimation est menée sur l'équation d'emploi écrite en différence spatiale. Il reste aussi positif mais devient

faible et non significatif quand le modèle est estimé en double différence. Enfin, l'effet devient négatif et significatif quand le taux de taxe local est instrumenté par les variables de partis politiques locaux. Ces résultats suggèrent que l'effet mesuré pour le taux local de taxation est sensible à la prise en compte de le tri spatial des établissements selon leurs caractéristiques inobservables et aux problèmes de causalité inverse.

L'entrée locale des établissements a aussi été étudiée avec une méthodologie proche de la différenciation spatiale. Nous avons considéré que le profit d'un établissement envisageant un lieu d'entrée de petite taille z situé dans la juridiction a noté Π_{izt} dépend du taux local de taxation, des caractéristiques observables de l'établissement Z_{it} et inobservables v_i (effets fixes), d'un effet fixe très local φ_{zt} et d'un effet fixe juridictionnel κ_a tel que :

$$\Pi_{izt} = \lambda r_{at} + Z_{it}\xi + v_i + \varphi_{zt} + \kappa_a + \varepsilon_{izt} \tag{9}$$

où ε_{izt} est un choc pouvant dépendre de la localisation et suivant une loi de valeurs extrêmes. Là encore, les deux sources d'hétérogénéité inobservée peuvent biaiser les résultats si les effets fixes établissements ou les effets fixes locaux sont corrélés avec les variables explicatives. Notre stratégie d'estimation consiste à faire disparaître l'ensemble des effets fixes en une seule étape. En effet, considérons deux petits sites z_1 et z_2 se trouvant proches de la frontière entre deux juridictions et localisés des deux côtés de cette frontière. La probabilité pour un établissement de choisir le site z_1 conditionnellement à se localiser sur l'un des deux sites s'écrit :

$$P(i \in z_1 | i \in \{z_1, z_2\}) = \frac{1}{1 + \exp\left[\lambda\left(r_{a_2 t} - r_{a_1 t}\right) + \kappa_{a_2} - \kappa_{a_1}\right]} \tag{10}$$

L'écriture de la probabilité conditionnelle fait simultanément disparaître tous les effets d'établissement (observés et inobservés) ainsi que les effets très locaux car ces effets sont les mêmes pour l'établissement des deux côtés de la frontière.

L'estimation du modèle se fait en maximisant la vraisemblance conditionnelle. Que le modèle contienne des effets fixes pour les juridictions ou non, l'effet du taux local de taxation est

positif mais non significatif. Ce résultat semble en accord avec le modèle mais l'estimation du coefficient de taxe locale manque de précision.

I.4. Les disparités spatiales de santé

Bien que la littérature sur les disparités spatiales se soit principalement développée en économie géographique pour analyser des aspects du marché du travail (cf. Combes et Overman, 2004, pour une synthèse de la littérature sur l'Europe), les méthodes d'analyse utilisées peuvent être appliquées à d'autres domaines comme la santé. On peut par exemple vouloir étudier les disparités spatiales de performances hospitalières lors du traitement des patients. Les performances d'un hôpital sont généralement mesurées par la propension des patients à décéder lorsqu'ils sont atteints d'une pathologie grave comme un infarctus aigu du myocarde (IAM).

Il peut être intéressant de décomposer les disparités spatiales de mortalité dans les hôpitaux en fonction des effets de composition des patients, des effets de composition hospitalière, des effets liés aux variations spatiales dans l'utilisation des procédures de soins, et des effets locaux comme l'impact de la concurrence locale entre hôpitaux. Ce type d'information est important pour pouvoir évaluer dans quelle mesure les politiques publiques sont susceptibles d'influencer la mortalité locale. En effet, le financement des hôpitaux publics français au début des années 2000 est organisé autour d'un système de dotations par enveloppe. Un budget global est ainsi réparti entre les autorités régionales qui distribuent à leur tour leur dotation entre les hôpitaux de la région. Si les disparités spatiales de mortalité sont dues à des différences interrégionales d'utilisation de procédures innovantes, les autorités peuvent changer l'allocation de budget entre régions. Si les disparités spatiales de mortalité sont plutôt causées par des différences dans le niveau de concentration des moyens hospitaliers au niveau local, une région donnée peut pousser à la concentration locale lorsqu'elle alloue son budget par négociations bilatérales avec les hôpitaux.

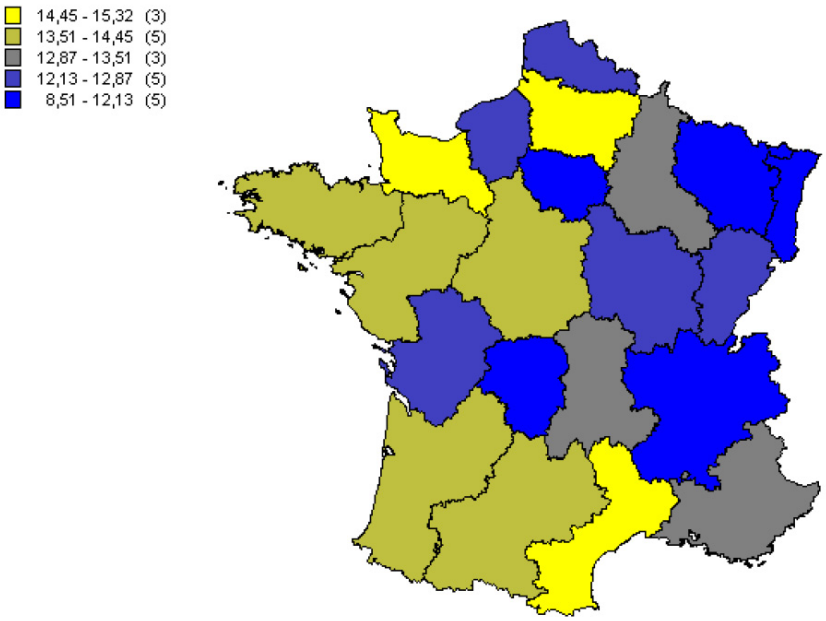
Les articles en économie de la santé s'intéressant à la dimension spatiale à l'intérieur d'un pays ont principalement étudié l'effet de la concurrence locale sur la mortalité dans les hôpitaux (Gaynor, 2006) sans pour autant tenter de caractériser les disparités spatiales.

Dans Gobillon et Milcent (2009), nous essayons d'identifier les principaux déterminants des disparités spatiales de mortalité par crise cardiaque en France. L'infarctus est la pathologie la plus étudiée dans la littérature car le taux de mortalité est élevé et les chances de survie

dépendent beaucoup des technologies utilisées pour traiter les patients. En France, les maladies ischémiques ont longtemps été la première cause de mortalité avec d'être dépassées depuis peu par le cancer.

L'unité spatiale retenue est la région car la pratique courante est de traiter les patients dans leur région de résidence (c'est le cas de plus de 90% des patients). La Figure 6 représente les probabilités régionales de décéder dans les quinze premiers jours pour les patients admis dans un hôpital pour un IAM. Il existe un contraste entre d'un côté, Paris et l'Est de France où la mortalité est faible, et de l'autre, l'Ouest et le Sud où la mortalité est élevée. La différence entre les extrêmes (l'Alsace et le Languedoc-Roussillon) est d'environ 80%.

Figure 6 : Probabilités régionales pour les patients atteints d'un IAM de décéder dans les quinze premiers jours



Note : La probabilité de décéder est calculée comme un moins la survie moyenne dans les hôpitaux de la région. Les sorties autres que le décès sont traitées comme des censures.
Source : cf. Gobillon et Milcent (2009).

Un déterminant de ces disparités spatiales est la différence de composition des patients hospitalisés entre régions. Les régions du sud-est par exemple concentrent une population plus âgée. L'alcoolisme, le diabète et l'obésité causés par des habitudes de vie spécifiques à certaines régions peuvent aussi mener à une mortalité plus élevée. Enfin, il est possible que les régions diffèrent par les proportions de patients ayant un historique de problèmes cardiaques ou de circulation qui constituent eux aussi des facteurs à risque.

Un second déterminant des disparités spatiales de mortalité est la composition régionale hospitalière. On distingue généralement les hôpitaux selon qu'ils sont privés à but lucratif, privés à but non lucratif (PSHP) ou publics. Il existe une littérature américaine montrant que la qualité des soins aux personnes âgées est meilleure dans les hôpitaux privés à but lucratif que dans les hôpitaux privés à but non lucratif (McClellan et Staiger, 2000). En France, la propension à décéder d'un IAM est plus faible dans les hôpitaux privés à but lucratif que dans les hôpitaux publics (Milcent, 2005). Une particularité des hôpitaux privés est qu'ils utilisent plus de procédures innovantes pour traiter les infarctus. Ainsi, si la proportion d'hôpitaux privés varie sur le territoire, il en sera de même pour les traitements pouvant affecter la mortalité.

Enfin, un troisième déterminant est l'organisation locale du système de soins. Les régions peuvent différer selon leurs infrastructures en général ou suivant l'efficacité de l'allocation des patients aux différents hôpitaux. Elles peuvent aussi différer selon le degré de concurrence hospitalière dans les agglomérations. Kessler et McClellan (2000) montrent qu'aux Etats-Unis, la concurrence locale (mesurée pour une zone donnée comme l'inverse d'un indice d'Herfindahl calculé à partir du nombre de patients des hôpitaux de cette zone) aurait un effet positif sur l'efficacité hospitalière.[1] Le système de soins américain reste cependant très différent du système français. En France, les hôpitaux publics constituent une grande part de l'offre de soins et sont à but non lucratifs. Ils ne sont donc pas vraiment incités par la concurrence à améliorer leur qualité des soins. Cependant les médecins français du public peuvent souhaiter intégrer un hôpital privé où ils seront mieux rémunérés. Ils essaieront alors de faire preuve d'efficacité pour être embauchés. Il existerait donc tout de même une forme de concurrence en qualité qui pourrait affecter positivement la qualité locale des hôpitaux. Concentrer les moyens hospitaliers dans un petit nombre d'hôpitaux permet par ailleurs de faire des économies d'échelle en partageant du matériel technologique coûteux. Un indice d'Herfindahl captant simultanément la concurrence en qualité et les économies d'échelle a donc théoriquement un effet de signe ambigu sur la mortalité des patients.

---

[1] L'indice d'Herfindahl pour une zone a donnée s'écrit comme $H_a = \sum_{j \in a} \left(\frac{n_j}{n^a} \right)^2$ où j indice l'hôpital, n_j est le nombre de patients dans l'hôpital j, et n^a est le nombre total de patients dans la zone a. H_a varie de $1/N^a$ à 1, où N^a est le nombre d'hôpitaux dans la zone. Quand $H_a = 1/N^a$, les patients de la zone sont équi-distribués dans les N^a hôpitaux. Quand $H_a = 1$, tous les patients de la zone sont traités dans un seul hôpital.

Nous avons essayé de déterminer l'importance relative des trois types de déterminants pour expliquer les différences spatiales de mortalité des patients hospitalisés pour un IAM. Nous utilisons les données du Programme de Médicalisation des Systèmes d'Information (PMSI) pour la période 1998-2003. Les données contiennent des informations sur les séjours des patients et sont exhaustives sur la période considérée. Il est ainsi possible de connaître la durée en jours des séjours, la commune d'implantation de l'hôpital traitant, l'origine du patient (son domicile, un autre service ou hôpital), son type de sortie (un décès, un retour au domicile ou un transfert). L'âge, le sexe et les diagnostics secondaires des patients sont aussi précisés. Comme les patients ne sont pas suivis lors d'un transfert, nous limitons l'analyse aux patients provenant de leur domicile. Ces données sont appariées avec les données hospitalières des Statistiques Annuelles des Etablissements de santé (SAE) qui fournissent des informations sur les hôpitaux comme leur statut, leur nombre de lits en général ou en chirurgie. Enfin, les données sont complétées par des variables communales captant le niveau de richesse de la commune.

Nous modélisons la durée de séjour des patients hospitalisés pour un IAM en utilisant un modèle de Cox stratifié par hôpital. Ainsi, chaque hôpital a un hasard de base spécifique. La sortie analysée est le décès, les autres sorties étant considérées comme censurées. La spécification est donc la même que dans Gobillon, Magnac et Selod (2009a). Les hasards hospitaliers sont aussi décomposés multiplicativement dans une seconde étape entre un hasard de base commun à tous les hôpitaux et un effet fixe hospitalier. Finalement, les effets fixes hospitaliers sont régressés dans une troisième étape sur des variables hospitalières et géographiques. Une fois que tous les coefficients du modèle sont estimés, le modèle est moyenné au niveau régional et une analyse de la variance est effectuée. Cette analyse doit permettre de déterminer l'importance respective des différents déterminants des disparités régionales de mortalité des patients atteints d'un IAM.

Les résultats sur les variables explicatives individuelles montrent, conformément à l'intuition, que les personnes plus âgées décèdent plus fréquemment. Il en est de même pour les femmes car elles ont un système de circulation différent de celui des hommes et les procédures de soins sont souvent moins adaptées à leur cas (Milcent et al., 2007). Les diagnostics secondaires ont des effets positifs ou négatifs. L'effet négatif contre-intuitif obtenu pour certains d'entre eux pourrait s'expliquer par un meilleur suivi et des meilleurs soins pour des

patients jugés à risque. Les traitements innovants ont un fort effet négatif sur la mortalité. Lorsqu'on calcule la probabilité régionale de décéder dans les quinze premiers jours nette des effets des variables explicatives, les régions extrêmes restent les mêmes mais la différence de probabilité de décéder diminue à 47%.

Lorsqu'on tente de préciser les facteurs hospitaliers expliquant la mortalité, la part de patients traités pour un IAM dans l'hôpital a un effet négatif significatif. Ce résultat suggère que les hôpitaux qui se sont spécialisés dans le traitement des IAM sont plus efficaces, peut-être à cause d'un phénomène d'apprentissage ou d'un équipement plus performant. La part de lits en chirurgie a aussi un effet négatif significatif peut-être parce que les hôpitaux effectuant plus d'opérations chirurgicales ont un équipement de meilleure qualité. On peut aussi remarquer que la différence de mortalité entre les hôpitaux privés à but lucratif et les hôpitaux publics est nulle. Les hôpitaux privés à but lucratif ne seraient donc pas plus efficaces que les hôpitaux publics à traitement égal des patients. Cependant, ils proposent des traitements plus innovants qui entrainent une mortalité plus faible. Enfin, on peut noter que parmi les facteurs géographiques, la concentration des patients dans un petit nombre d'hôpitaux au niveau de l'agglomération urbaine a un effet négatif sur la mortalité, ce qui est compatible avec l'existence d'économies d'échelle.

L'analyse spatiale de la variance montre que même si les variables démographiques expliquent en partie les disparités régionales de mortalité tout comme la concentration hospitalière locale, le facteur déterminant reste la part régionale de traitements par des actes innovants. Ainsi, les autorités publiques peuvent souhaiter diminuer les disparités régionales de mortalité parce qu'il y a des inégalités de traitement. Une façon de réduire les disparités pourrait être de changer les allocations budgétaires aux régions, en augmentant les budgets dans les régions où les traitements innovants sont moins fréquents et en les diminuant là où ils sont plus fréquents.

II. La productivité des entreprises dans les villes : une approche distributionnelle

Mes études sur les disparités spatiales reposent toutes sur une analyse spatiale de la variance d'un indicateur de résultat. En général, la moyenne locale de cet indicateur est expliquée par des variables locales et des effets de composition locale. Une limite importante de mes études, récurrente en économie géographique et urbaine, est qu'il est difficile d'identifier séparément les effets des différents mécanismes sous-jacents par manque de données plus précises sur ces mécanismes. Une partie de mes travaux a essayé de préciser les mécanismes à l'œuvre en utilisant la forme des distributions locales d'indicateurs de résultat plutôt que de simples moyennes. La méthode d'analyse peut être utilisée dans des contextes très variés et a aussi été appliquée pour étudier les problèmes d'accès à l'emploi des femmes à partir d'une comparaison de leur distribution de salaires avec celle des hommes.

II.1. Identifier les effets d'agglomération et de concurrence locale

Les entreprises et les travailleurs sont plus productifs dans les grandes agglomérations. En accord avec ce fait stylisé mis en avant depuis plus d'un siècle, nous avons montré que les salaires augmentent avec la densité d'emploi.

Pendant longtemps, l'augmentation de la productivité avec la taille des villes a été attribuée à des économies d'agglomération dont les origines ont été détaillées dans la section précédente. Tous les mécanismes que nous avons évoqués reposent sur le fait que la concentration des entreprises et des travailleurs dans l'espace les rend plus productifs.

Plus récemment, une explication alternative basée sur des mécanismes de sélection a été avancée. Cette explication développée dans la littérature en commerce international (Melitz, 2003) stipule que plus les marchés sont grands, plus ils attirent d'entreprises et se caractérisent par une concurrence plus importante. Ainsi, une plus grande fraction des entreprises les moins productives sort du marché. Une productivité moyenne des entreprises plus élevée dans les grandes villes pourrait donc être le résultat d'un processus Darwinien de sélection où seuls les plus productifs survivent.

Dans Combes, Duranton, Gobillon, Puga et Roux (2009), nous avons essayé d'évaluer empiriquement l'importance des mécanismes d'agglomération et de sélection menant à une productivité plus importante dans les grandes villes. L'idée générale est de proposer un

modèle théorique incorporant les deux types de mécanismes puis d'estimer leur contribution respective à partir de la forme de distributions locales de productivité. Nous partons du modèle de Melitz et Ottaviano (2008) qui incorpore le mécanisme de sélection, en prenant soin d'éliminer les hypothèses paramétriques sur la forme des distributions locales de productivité. Nous incorporons dans le modèle un mécanisme d'agglomération assez général dans l'esprit de Lucas et Rossi-Hansberg (2002). Les prédictions théoriques montrent que même si les effets de sélection et d'agglomération permettent d'obtenir une productivité moyenne plus importante dans les grandes villes, ils ont des prédictions différentes sur la forme de la distribution observée des productivités. Nous utilisons ensuite une relation structurelle entre les distributions de productivité des petites et des grandes villes pour estimer les paramètres relatifs aux effets d'agglomération et aux effets de sélection.

Le modèle théorique part d'une situation où un individu habitant un lieu donné consomme un bien numéraire et des biens différenciés. Il maximise son utilité sous contrainte budgétaire et consomme une certaine quantité de chaque bien.

Les biens différentiés sont produits dans un contexte de concurrence monopolistique. Une entreprise souhaitant produire un bien doit payer un coût d'entrée et la production d'une unité de bien nécessite une certaine quantité de travailleurs qui correspond au coût marginal en main d'œuvre. La quantité de travailleurs requise diffère selon les entreprises, et pour une entreprise donnée, elle est tirée aléatoirement dans une distribution donnée. Nous ne faisons aucune hypothèse sur cette distribution contrairement à la littérature sur le thème. Les entreprises donc le coût marginal est supérieur au prix auquel elles peuvent vendre leur bien sortent du marché car elles ne sont pas rentables. Ainsi, seule une fraction des variétés de bien est produite à l'équilibre.

Le bien numéraire est produit avec une unité de travail sous l'hypothèse de rendements constants. Il peut être échangé librement quand on considère qu'il y a plusieurs localisations. Ainsi, le coût de la main d'œuvre pour produire le bien est toujours l'unité. Le bien numéraire est consommé à hauteur des revenus qui ne sont pas utilisés pour l'achat de biens différentiés.

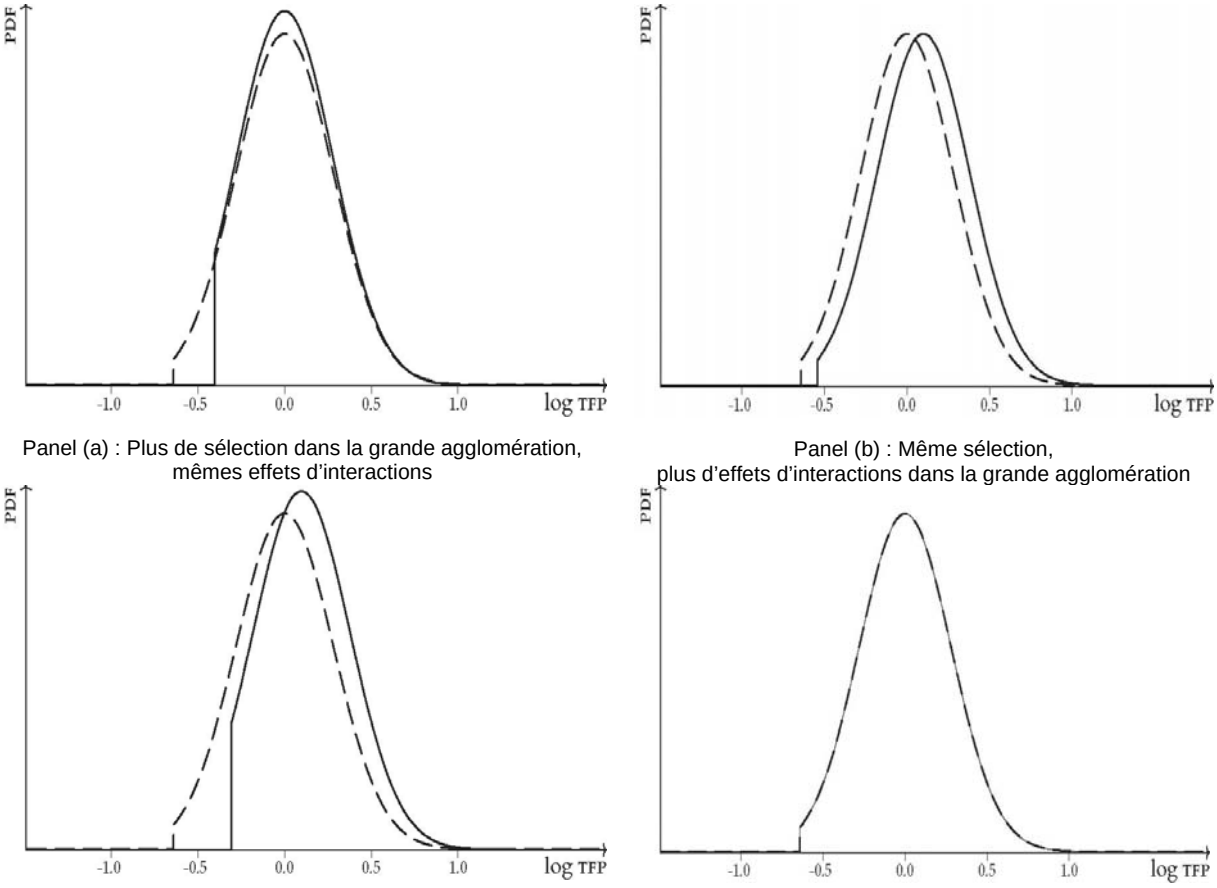
L'entrée sur le marché des entreprises produisant des biens différentiés a lieu tant que le profit espéré est supérieur à zéro. La condition de profit nul donne le nombre d'entreprises entrant sur le marché.

Les effets d'agglomération sont pris en compte en considérant qu'un travailleur est rendu plus productif par ses interactions avec les autres travailleurs. Ces interactions correspondent à un échange d'idées entre travailleurs. Ainsi, la productivité d'un travailleur augmente

proportionnellement avec le nombre de travailleurs sur le marché. La quantité de travailleurs nécessaire pour produire une unité de bien décroit donc avec la taille du marché.
On peut exprimer les prédictions du modèle en termes de productivité des entreprises, égale à la quantité produite divisée par la quantité de travail utilisée. Une entreprise ne décide de produire que lorsque sa productivité est supérieure à un certain seuil. Sa décision est influencée par les effets d'interaction qui peuvent augmenter sa productivité.

A titre d'illustration, on peut considérer une économie avec deux agglomérations de taille différente. On fait l'hypothèse restrictive que la distribution des coûts dans laquelle les entreprises de chaque agglomération font un tirage sont les mêmes. Autrement dit, il n'y a pas de tri spatial des entreprises *ex ante* selon leur coût marginal. Deux cas polaires pour la concurrence et les interactions sont envisagés. Pour ce qui est de la concurrence, on peut considérer qu'à un extrême, les entreprises ne vendent qu'aux consommateurs de leur agglomération et ne sont en concurrence qu'avec les entreprises locales (*concurrence locale)*. A l'autre extrême, les entreprises vendent aux consommateurs des deux agglomérations avec la même facilité et sont donc en concurrence avec les entreprises des deux agglomérations (*concurrence globale)*. En ce concerne les interactions, on peut considérer qu'à un extrême, les travailleurs interagissent exclusivement avec les travailleurs de la même agglomération (*interactions locales)*. A l'autre extrême, les travailleurs interagissent avec les travailleurs des deux agglomérations avec la même facilité (*interactions globales*).
La combinaison des différentes possibilités donne quatre cas pour lesquels il est possible de comparer les distributions d'équilibre des log-productivités entre les deux agglomérations. Ces distributions sont représentées sur la Figure 7.

Figure 7 : Distributions de log-productivités dans une grande et une petite agglomération, selon le type d'effet de sélection et d'interactions



Panel (a) : Plus de sélection dans la grande agglomération, mêmes effets d'interactions

Panel (b) : Même sélection, plus d'effets d'interactions dans la grande agglomération

Panel (c) : Plus de sélection et plus d'effets d'interactions dans la grande agglomération

Panel (d) : Même sélection, mêmes effets d'agglomération

Note : la ligne continue fait référence à la grande agglomération et la ligne en pointillé fait référence à la petite agglomération.
Source : cf. Combes, Duranton, Gobillon, Puga et Roux (2009).

Cas 1 *(concurrence locale et interactions globales)****.*** Le Panel (a) représente la distribution des log-productivités pour une grande agglomération (ligne continue) et pour une petite agglomération (ligne en pointillé) dans le cas où les entreprises ne vendent que localement et où les travailleurs profitent d'interactions avec les travailleurs des deux agglomérations. Comparée à la distribution de la petite agglomération, celle de la grande agglomération est tronquée à gauche du fait d'effets de sélection plus importants causés par une concurrence locale plus importante. La troncation à gauche implique que le pic de la distribution de la grande agglomération est plus haut que celui de la petite agglomération, mais les deux pics ont lieu pour le même niveau de productivité.

Cas 2 *(concurrence globale et interactions locales)***.** Le Panel (b) représente les distributions de log-productivités dans le cas où toutes les entreprises sont en concurrence et où les travailleurs n'interagissent qu'avec les travailleurs de leur agglomération. Comparée à la distribution de la petite agglomération, celle de la grande agglomération est translatée à droite du fait d'interactions plus importantes qui bénéficient localement à toutes les entreprises de la grande agglomération.

Cas 3 *(concurrence locale et interactions locales)***.** Le Panel (c) représente les distributions de log-productivités dans le cas d'une concurrence locale et d'interactions locales. Comparée à la distribution de la petite agglomération, celle de la grande agglomération est tronquée à gauche et translatée à droite. En fait, on obtient une combinaison des deux cas précédents sans déformation supplémentaire parce que le modèle est spécifié tel que les effets de concurrence et d'interactions soient « séparables ».

Cas 4 *(concurrence globale et interactions globales)***.** Le Panel (d) représente les distributions de log-productivités dans le cas d'une concurrence globale et d'interactions globales. Dans ce cas, les deux distributions de log-productivités sont confondues. Il convient de noter que l'absence de différenciation des deux distributions ne signifie pas l'absence de mécanismes de concurrence et d'interactions, mais simplement que ces deux types d'effets jouent de la même façon pour les deux agglomérations.

Ces exemples suggèrent une façon de tester empiriquement l'existence d'effets de concurrence et d'interactions en accord avec le modèle. Considérons la transformation de la distribution de log-productivités d'une petite agglomération de référence j en celle d'une grande agglomération i. On peut résumer la troncation à gauche par un paramètre S correspondant au seuil de troncation, et la translation par un paramètre A. Les deux paramètres sont estimés en s'appuyant sur une comparaison des fonctions quantiles des distributions de log-productivités des deux agglomérations notées $\lambda_i(\bullet)$ et $\lambda_j(\bullet)$. Il est possible de montrer que les quantiles vérifient la spécification :

$$\lambda_i(r_S(u)) = \lambda_j(S + (1-S)r_S(u)) + A \text{ avec } u \in [0,1] \tag{11}$$

où $r_S(u) = \max\left(0, \frac{-S}{1-S}\right) + \left[1 - \max\left(0, \frac{-S}{1-S}\right)\right]u$.

Cette spécification peut être expliquée à partir de deux cas polaires. Lorsqu'il n'y a pas de sélection ($S = 0$), la spécification se résume à dire que le paramètre de translation est égal à la différence des quantiles des deux agglomérations à chaque rang de leur distribution $\lambda_i(u) - \lambda_j(u)$. Lorsqu'il n'y a pas de translation ($A = 0$) et que la sélection joue dans le sens où la distribution de l'agglomération i est une version tronquée à gauche de celle de l'agglomération j ($S > 0$), la spécification se réécrit simplement $\lambda_i(u) = \lambda_j(S + (1-S)u)$: la log-productivité à un rang donné dans la distribution de l'agglomération i se trouve à un rang supérieur dans la distribution de l'agglomération j. Plus la troncation de la distribution de l'agglomération i est importante, plus le rang qu'il faut considérer dans la distribution de l'agglomération j est élevé.

Les paramètres de troncation et de translation peuvent être estimés en utilisant la continuité d'égalités entre quantiles (11). En effet, on peut remplacer chaque quantile par sa contrepartie empirique et minimiser la distance entre les quantiles calculés aux rangs appropriés. La minimisation est similaire à celle que l'on effectue pour une continuité de moments (Carrasco et Florens, 2000) et est exposée en détails dans Gobillon et Roux (2009).

Pour estimer le modèle, nous apparions trois bases de données administratives françaises sur la période 1994-2002. La première contient les données des Bénéfices Réels Normaux et du Régime Simplifié d'Imposition (BRN-RSI) qui relatent des informations sur la valeur ajoutée des entreprises ainsi que sur la valeur des actifs productifs et financiers. Nous utilisons la mesure du capital construite par Boutin et Quantin (2006) comme la somme des valeurs des biens évalués à leur coût historique. Les données contiennent aussi le secteur d'activité de l'entreprise en NES36.

La seconde base est le Système d'Identification du Répertoire des ENtreprises (SIREN). Elle contient l'information annuelle sur tous les établissements du secteur privé français, en excluant les secteurs de la finance et de l'assurance. Les données contiennent en particulier l'identifiant de la commune d'implantation à partir de laquelle on peut déterminer l'agglomération d'implantation.

Enfin, la troisième base contient les Déclarations Annuelles des Données Sociales (DADS), une base de données appariées employeurs-employés qui est exhaustive sur la période considérée. Elle inclut des informations sur le nombre d'heures travaillées de chaque employé dans chaque établissement. Elle renseigne aussi sur la catégorie socio-professionnelle à deux chiffres des employés et permet donc de prendre en compte la qualification de la main-

d'œuvre. Nous utilisons les catégories proposées par Burnaud et Chenu (2001) pour agréger la main-d'œuvre en trois catégories de qualification : basse, moyenne et haute.

Les données nous permettent d'estimer des équations de TFP par secteur avec différentes méthodes (moindres carrés ordinaires ; Olley et Pakes, 1995 ; Levinsohn et Petrin, 2003 ; parts des coûts dans la production sous hypothèse de rendements constants). Nous régressons ainsi le logarithme de la valeur ajoutée sur le logarithme du capital, le logarithme de la quantité de main d'œuvre (mesurée par le nombre d'heures travaillées), et les parts des différentes qualifications dans la main d'œuvre. Nous assimilons ensuite le résidu de l'équation de TFP à la log-productivité des entreprises. Ce résidu est estimé comme la différence entre la valeur ajoutée observée et la valeur ajoutée prédite par l'équation de TFP. Nous pouvons ensuite reconstruire pour chaque secteur ou pour l'ensemble des secteurs mélangés, des distributions de log-productivités pour la catégorie des grandes agglomérations (plus de 200 000 habitants) et la catégorie des petites agglomérations (moins de 200 000 habitants).

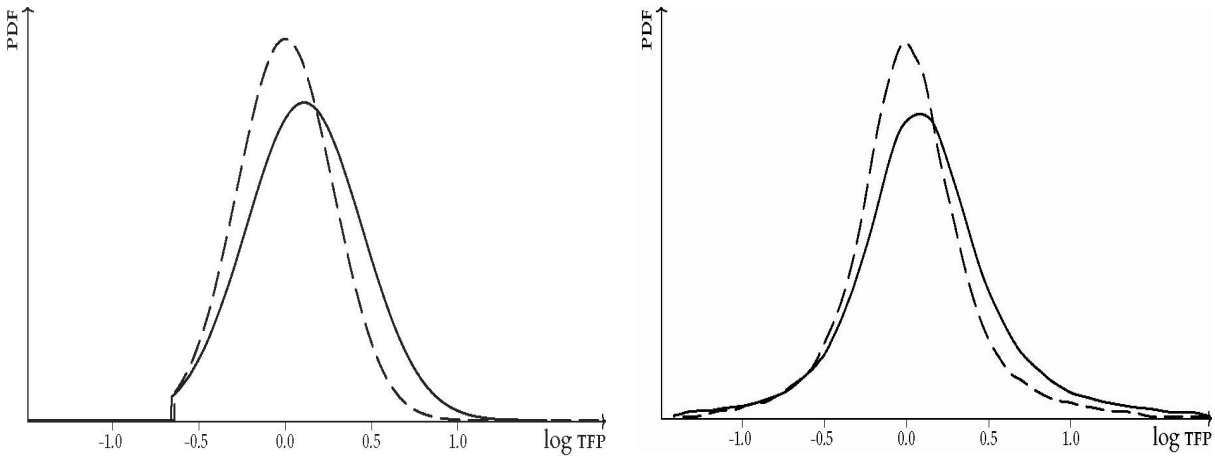
Les résultats de l'estimation du paramètre de troncation et du paramètre de translation (obtenus lorsque l'équation de TFP est estimée par moindres carrés ordinaires) peuvent être commentés pour les seize secteurs de l'étude. Nous trouvons que le paramètre de translation est toujours positif. Il est significatif pour chacun des secteurs à l'exception d'un seul. Quand le paramètre de translation est estimé pour l'ensemble des secteurs, sa valeur implique que les effets d'interactions dans une grande agglomération entraînent une productivité de 12% plus élevée que dans une petite agglomération. D'après le modèle théorique, les résultats suggèrent que les interactions sont locales, ce qui est en accord avec la littérature empirique (cf. Rosenthal et Strange, 2004). Pour 11 secteurs sur 16, le paramètre de troncation n'est pas significatif. Il est négatif et significatif pour 4 secteurs et pour les secteurs pris dans leur ensemble. Il n'est positif et significatif que pour un secteur. Dans tous les cas, les valeurs du paramètre sont très petites. Les résultats suggèrent qu'il y a peu de différence dans les effets de sélection entre les petites et les grandes agglomérations. Ils n'impliquent pas qu'il n'existe pas d'effets de sélection mais plutôt qu'ils ont la même importance dans les grandes et les petites agglomérations. D'après le modèle, les résultats sont compatibles avec une situation où les entreprises sont en concurrence au niveau national et non sur des marchés locaux séparés.

Il est possible de vérifier que les effets d'interactions sont des meilleurs prédicteurs que les effets de concurrence pour expliquer les différences de distributions de log-productivités entre

grandes et petites agglomérations. Les résultats suggèrent toutefois qu'il reste une part inexpliquée dans ces différences.

Il est possible d'étendre le modèle en considérant que les effets d'interactions bénéficient plus aux entreprises très productives qu'aux entreprises moins productives. Cet effet croissant des interactions avec la productivité est modélisé en considérant maintenant que les interactions conduisent à une transformation linéaire de la log-productivité. Les interactions ont donc simultanément un effet de dilatation et un effet de translation sur la log-productivité, au lieu d'un simple effet de translation. Le Panel (a) de la Figure 8 représente la distribution de log-productivités dans une grande agglomération (ligne continue) et dans une petite agglomération (ligne en pointillé) quand la concurrence est globale mais que les interactions sont locales. La distribution de la grande agglomération est plus aplatie que cette de la petite agglomération. Les distributions empiriques des log-productivités pour les grandes agglomérations et les petites agglomérations représentées sur le Panel (b) ont des profils très similaires à ceux du Panel (a).

Figure 8 : Distributions des log-productivités dans une grande agglomération et une petite agglomération, les interactions ayant un effet linéaire



Panel (a). Distributions théoriques

Panel (b). Distributions observées, Population > 200 000 vs. Population < 200 000

Source : cf. Combes, Duranton, Gobillon, Puga et Roux (2009).

Il est possible d'évaluer empiriquement l'importance de la dilatation causée par l'effet croissant des interactions avec la productivité. Dans 8 secteurs, le coefficient de dilatation est statistiquement différent de l'unité. Pour 7 secteurs (et les secteurs pris dans leur ensemble), le

coefficient de dilatation est supérieur à l'unité. Dans un seul secteur, le coefficient de dilatation est inférieur à l'unité. Il y aurait donc une tendance pour la distribution des log-productivités dans les grandes agglomérations à être plus dilatée que celle dans les petites agglomérations. Les entreprises les plus productives profiteraient donc plus des interactions que les entreprises les moins productives. Pour les secteurs pris dans leur ensemble, l'avantage productif des entreprises des grandes villes provenant des interactions est en moyenne de 9%, mais il monte à 14% pour les entreprises du premier quartile et n'est que de 5% pour celles du dernier quartile.

En résumé, une méthode a été proposée pour évaluer l'importance respective des effets de concurrence et d'interactions en comparant les formes des distributions de log-productivités dans les grandes agglomérations et les petites agglomérations. Les résultats suggèrent qu'il existerait des interactions locales bénéficiant aux entreprises des grandes villes avec un effet supérieur pour les plus productives d'entre elles. Ils suggèrent aussi que la concurrence serait plutôt globale et de même intensité dans les grandes et les petites agglomérations.

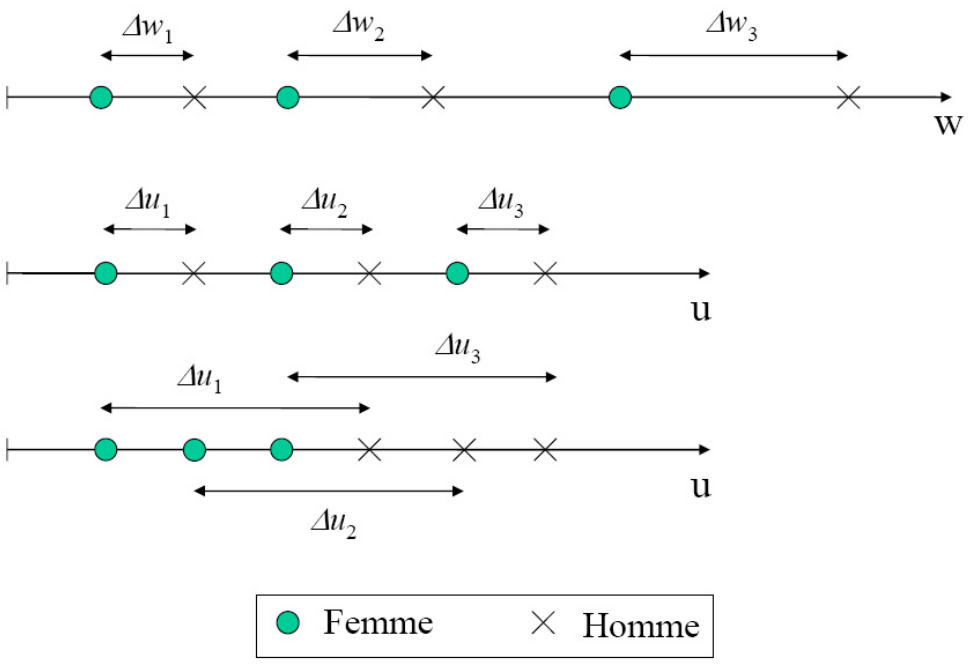
II.2. Tester l'existence d'un plafond de verre pour les femmes

Mes travaux méthodologiques sur la comparaison entre distributions m'ont aussi amené à m'intéresser à un autre thème éloigné de l'économie géographique : les différences de salaire hommes-femmes. En effet, une littérature qui se développe actuellement montre que le salaire moyen moins élevé des femmes serait dû à une sous-représentation des femmes dans les emplois les mieux rémunérés. Ce phénomène est appelé « plafond de verre » (*glass ceiling*) pour évoquer l'idée que les femmes ne peuvent pas dépasser une certaine limite dans la position des emplois qu'elles peuvent occuper. Depuis Albrecht, Bjorklund et Vroman (2003), les articles empiriques qui essaient de tester l'existence d'un plafond de verre s'appuient sur une comparaison des quantiles des distributions de salaires des hommes et des femmes à un rang donné. Ils considèrent qu'il existe un plafond de verre si la différence de quantiles est croissante avec le rang dans la distribution.

Dans Gobillon, Meurs et Roux (2009) nous mettons en avant le fait que cette approche est discutable parce qu'elle mélange deux dimensions : la position d'emploi proprement dite et le salaire qui y est associé. Les interprétations faites à partir des différences de quantiles peuvent alors être inexactes comme le suggère la Figure 9. Supposons une échelle de positions

d'emploi classique où le salaire augmente plus que proportionnellement avec la position de l'emploi. Les emplois sont occupés alternativement par des hommes et des femmes (axe 1). Dans ce cas, la différence de quantiles entre hommes et femmes est croissante le long de l'échelle de positions. D'après la littérature, on devrait alors conclure qu'il existe un plafond de verre alors qu'en moyenne les hommes et les femmes ont autant de chances d'occuper chaque emploi. Il est possible d'éliminer les espacements entre les salaires de positions consécutives en considérant la différence de rangs dans l'échelle des positions plutôt que la différence de quantiles (axe 2). La différente de rangs est constante le long de l'échelle, ce qui suggère un égal accès aux emplois pour les hommes et les femmes. Cette conclusion semble être une bonne réponse. Toutefois le raisonnement reste erroné et il est possible de construire une situation où la différence de rangs est constante mais où il existe un plafond de verre. C'est le cas lorsque les trois positions d'emploi inférieures sont occupées par des femmes et les trois positions d'emploi supérieures sont occupées par des hommes (axe 3).

Figure 9 : Différence de quantiles et de rangs entre hommes et femmes dans différents contextes



Source : cf. Gobillon, Meurs et Roux (2009).

La confusion provient du fait que les rangs considérés sont ceux dans les distributions de salaires respectives des hommes et des femmes, et ne sont pas directement liés au rang dans

l'échelle des positions d'emploi. L'analyse devrait plutôt considérer l'échelle de positions d'emploi et examiner comment la différence d'accès aux emplois entre hommes et femmes varie le long de cette échelle. On peut alors considérer qu'il existe un plafond de verre si les femmes n'ont pas accès aux emplois aux rangs les plus élevés dans l'échelle de positions d'emploi. Plus généralement, on peut considérer que les femmes rencontrent des barrières aux emplois les mieux rémunérés si leur accès relatif aux emplois par rapport aux hommes diminue le long de l'échelle des positions d'emploi.

Nous proposons donc un modèle d'assignement d'emplois (cf. Sattinger, 1993) où l'accès aux emplois relatif des femmes par rapport aux hommes influence le processus de recrutement le long de l'échelle des positions d'emplois qui sont indexées par le salaire offert. Les travailleurs souhaitent obtenir l'emploi le mieux rémunéré et sont en concurrence jusqu'à ce que l'un d'eux soit embauché. Les travailleurs restant sont alors en concurrence pour le second emploi le mieux rémunéré. Le processus de concurrence prend fin lorsque tous les emplois sont occupés.

Le modèle permet un désavantage des femmes lorsqu'elles postulent à une position d'emploi. Ainsi, si une femme et un homme se présentent pour occuper une position d'emploi donnée, la femme pourra avoir une probabilité d'être recrutée plus faible que l'homme. En outre, les chances relatives des femmes d'être recrutées par rapport aux hommes peuvent dépendre de la position de l'emploi le long de l'échelle des salaires. Les barrières d'accès à l'emploi que rencontrent les femmes sont en particulier susceptibles de résulter de deux types de discrimination. Les employeurs peuvent faire de la discrimination envers les femmes simplement par goût (Becker, 1971). Ils peuvent aussi faire de la discrimination statistique en attribuant aux candidats femmes la productivité moyenne qu'ils attribuent à la population féminine (Arrow, 1971 ; Phelps, 1972).

La différence d'accès aux emplois entre femmes et hommes le long de l'échelle des salaires est résumée par une fonction d'accès. A chaque rang de la distribution des salaires, cette fonction prend la valeur du rapport entre les probabilités d'une femme et d'un homme d'obtenir l'emploi correspond au rang s'ils faisaient tous deux acte de candidature. Il est possible de définir formellement trois cas particuliers de différence d'accès. On dira qu'il existe une discrimination uniforme envers les femmes si leur fonction d'accès prend une valeur constante inférieure à l'unité à tous les rangs de la distribution de salaires. On dira qu'il existe des barrières aux emplois les mieux rémunérés pour les femmes si la fonction d'accès est plus faible aux rangs les plus élevés. Cette définition englobe la situation où les femmes

n'ont aucun accès aux emplois les mieux rémunérés, auquel cas on dira qu'il existe un phénomène de plafond de verre (*glass ceiling*). Enfin, on dira qu'il existe un « plancher collant » (*sticky floor*) si la fonction d'accès est supérieure à un aux rangs les plus faibles.
Il est possible d'établir à partir du modèle que la fonction d'accès à l'emploi est de la forme :

$$h(u) = \frac{n'_f(u) / n_f(u)}{n'_m(u) / n_m(u)} \tag{12}$$

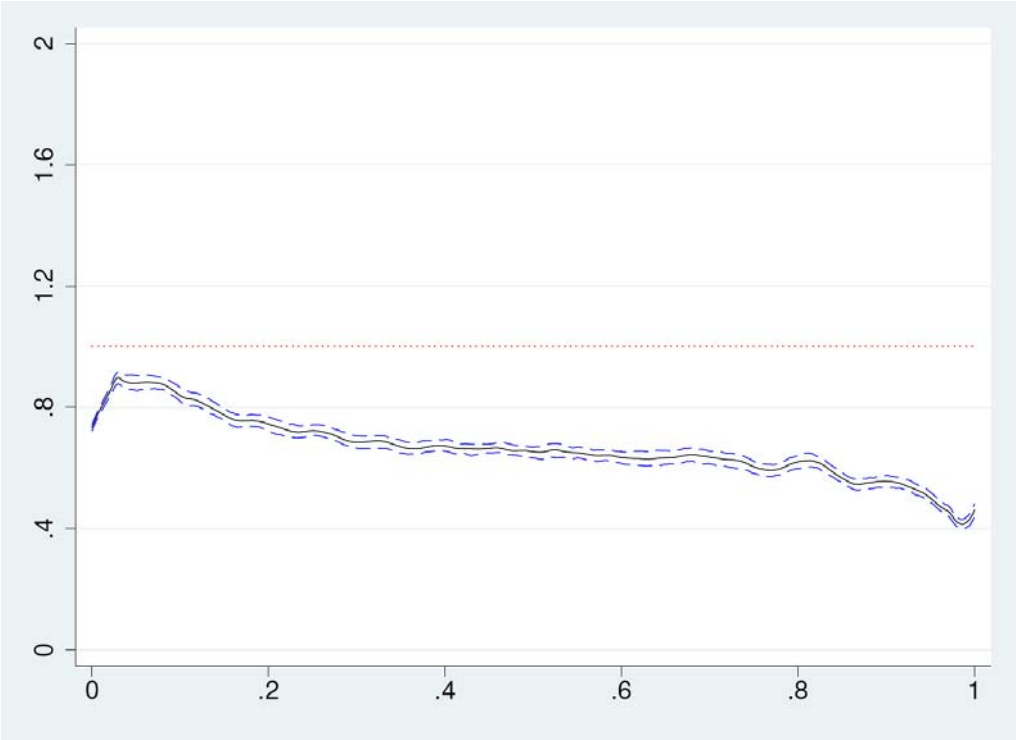
où $n_j(u)$ est la proportion de travailleurs de sexe j encore à la recherche d'un emploi au rang u dans la population globale. Autrement dit, la fonction d'accès correspond au rapport des taux de retour à l'emploi des femmes et des hommes à chaque rang. Cette relation structurelle peut être utilisée pour estimer non paramétriquement la fonction d'accès car les termes de droite peuvent facilement être estimés à partir des données.

Dans notre application empirique, nous utilisons la version exhaustive des Déclarations Annuelles des Salaires (DADS) pour l'année 2003. Les données contiennent des informations sur tous les travailleurs du privé. En particulier, il est possible de connaître le secteur dans lequel ils sont employés, leur contrat (temps plein / temps partiel), leur salaire journalier, leur catégorie socio-professionnelle, leur âge et s'ils sont nés en France. Une limite des données est qu'elles ne contiennent aucune information sur le diplôme.
Comme nous souhaitons estimer notre modèle, nous restreignons notre attention à un sous-échantillon pour lequel ses hypothèses ont le plus de chances d'être vérifiées. Pour éviter le problème de salaire minimum qui tasse tous les emplois dans le bas de la distribution des salaires, nous n'étudions que les cadres. Le modèle suppose l'existence d'un marché homogène où les hommes et les femmes sont en concurrence pour les mêmes positions d'emploi. Ainsi, nous limitons notre analyse aux travailleurs à temps plein âgés de 40 à 45 ans. A cet âge, les femmes encore sur le marché sont motivées par leur carrière professionnelle et sont en concurrence avec les hommes pour les mêmes emplois.

La fonction d'accès relatif aux emplois des femmes et des hommes que nous estimons sur ce sous-échantillon est toujours en-dessous de un et décroissante avec le rang comme le montre la Figure 10. Ainsi, les femmes ont un moins bon accès aux emplois quel que soit leur rang

dans la distribution des salaires et leur accès tend à décroitre avec le rang (de 0.9 à 0.4). Par exemple, la probabilité d'une femme d'obtenir un emploi au rang 0.05 est 12% plus basse que celle d'un homme. Au rang 0.95, la différence de probabilité d'obtention d'un emploi est de 50%.

Figure 10 : Fonction d'accès à l'emploi estimée pour l'ensemble des secteurs



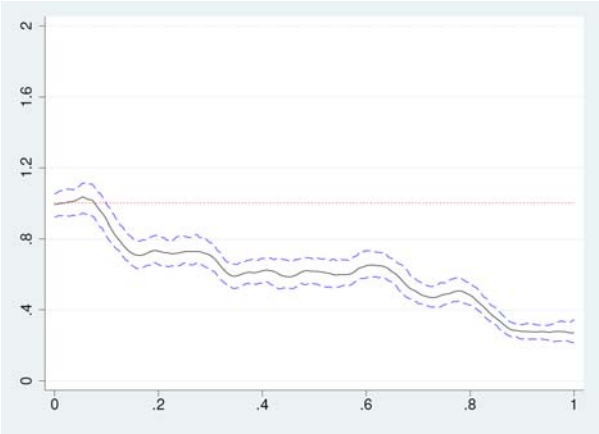
Source : cf. Gobillon, Meurs et Roux (2009).

Nous étudions aussi plus spécifiquement les secteurs de la banque et de l'assurance. En effet, ces deux secteurs présentent des similarités puisqu'ils sont tous deux intensifs en main-d'œuvre et se caractérisent par une proportion importante de cadres. De plus, la proportion de femmes est au-dessus de la moyenne, ce qui est assez courant dans les secteurs des services. Le système de carrières est cependant très différent dans les deux secteurs. En effet, les banques se basent sur une classification des positions d'emploi et un système de promotion assez rigides hérités de la période où elles appartenaient au secteur public. Les compagnies d'assurance donnent quant à elle une plus grande place à l'individualisation des carrières (Dejonghes et Gasnier, 1990).

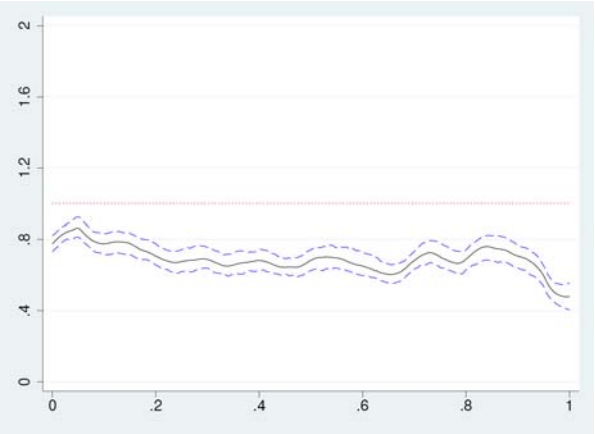
La Figure 11 montre qu'il existe de grandes différences entre les fonctions d'accès des secteurs de l'assurance et de la banque. Pour l'assurance (Panel a), la fonction d'accès décroit

fortement de 1 à 0.3, ce qui suggère d'importantes barrières aux emplois les mieux rémunérés pour les femmes. Pour les banques (Panel b), la fonction d'accès décroit lentement de 0.8 à 0.6, et son profil est proche d'être constant, suggérant une discrimination presque uniforme.

Figure 11 : Fonction d'accès à l'emploi estimée
pour les secteurs de la banque et de l'assurance



Panel (a). Assurance



Panel (b). Banque

Source : cf. Gobillon, Meurs et Roux (2009).

En fait, il est possible d'évaluer l'importance respective des barrières aux emplois les mieux rémunérés pour les femmes dans les secteurs de l'assurance et de la banque à partir de la pente de la fonction d'accès. En effet, plus la pente est importante, plus il existe de telles barrières aux emplois. Nous souhaitons donc approximer la fonction d'accès avec une fonction linéaire, évaluer les coefficients de pente pour les deux secteurs, et les comparer. La procédure d'estimation est très similaire à celle de Combes, Duranton, Gobillon, Puga et Roux (2009). En effet, considérons la cumulative $F_j(\bullet)$ de la variable aléatoire correspondant au rang d'un individu de sexe j dans la distribution de salaire des positions d'emploi. Notons aussi $u_j(\bullet)$ son inverse qui est la fonction quantile. Par définition, on a :

$$v = F_f\left[u_f(v)\right] = F_m\left[u_m(v)\right] \text{ pour } v \in [0,1] \tag{13}$$

Pour une valeur donnée des paramètres de la fonction linéaire d'accès, il est possible d'évaluer numériquement la fonction quantile de chaque sexe provenant du modèle. On remplace par ailleurs les fonctions cumulatives par leur contrepartie empirique. On peut alors

minimiser la distance quadratique entre les deux cumulatives empiriques évaluées aux quantiles appropriés déduits du modèle théorique pour estimer les paramètres de la fonction linéaire d'accès. Il est à noter que le critère de minimisation quadratique suit asymptotiquement une statistique de Cramer Van-Mises sous l'hypothèse que la fonction d'accès est bien linéaire. Il est donc possible de tester la validité de la représentation linéaire de la fonction d'accès.

L'estimation des paramètres de la fonction d'accès conduit pour l'ensemble des secteurs à une pente de -0.28 : lorsqu'on considère un emploi un décile plus haut dans la distribution des salaires, l'accès à l'emploi relatif des femmes par rapport aux hommes diminue de 2.8%. La décroissance est beaucoup plus importante dans le secteur de l'assurance avec une baisse de 6.0% par décile, tandis qu'elle est plus faible dans le secteur bancaire avec une baisse de seulement 0.7% par décile. Il est à noter que la spécification linéaire de la fonction d'accès n'est rejetée dans aucun des cas. Comme la pente de la fonction d'accès pour le secteur bancaire est proche de zéro, nous avons aussi essayé d'estimer une spécification où la fonction d'accès prend une valeur constante à tous les rangs dans ce secteur. La constante est estimée à 0.672 et la spécification n'est pas rejetée par notre test. Ainsi, la fonction d'accès serait approximativement constante pour le secteur bancaire.

Nous montrons aussi comment notre modèle peut être étendu pour prendre en compte l'hétérogénéité individuelle observée chez les travailleurs. En effet, cette hétérogénéité (ici l'âge à l'année près et le fait d'être né français) peut être incorporée dans la probabilité d'accès aux emplois. Les travailleurs sont en revanche tous en concurrence quelles que soient leurs caractéristiques. Une application empirique permet d'évaluer une fonction d'accès nette des effets de composition pris en compte par les quelques variables explicatives individuelles qui peuvent être construites à partir de nos données. Nous trouvons que le profil de la fonction d'accès nette est très proche de celui de la fonction d'accès brute qui avait déjà été calculée.

Enfin, il est possible d'étendre le modèle en considérant que la concurrence n'a pas lieu sur le marché national mais plutôt à l'intérieur de chaque entreprise. On peut alors estimer la fonction d'accès moyenne des entreprises. En particulier, la fonction d'accès estimée est nette de l'hétérogénéité des entreprises dans le niveau des salaires. Une application pour les grandes entreprises montre que le profil de cette nouvelle fonction d'accès est là encore proche du profil de la fonction d'accès brute. Il est cependant plus lisse du fait de la prise en compte de l'hétérogénéité des entreprises.

III. La mobilité résidentielle au cours du cycle de vie

Une critique récurrente des travaux sur les disparités spatiales sur le marché du travail est que la localisation des individus pourrait être endogène. En effet, les individus sont susceptibles de prendre en compte le niveau de salaire qu'ils peuvent obtenir dans les différentes zones et de choisir la zone où ils reçoivent l'offre d'emploi leur permettant d'atteindre le niveau de rémunération le plus élevé. Il conviendrait donc d'étudier le lien entre les disparités spatiales de salaires et les migrations.

La mobilité résidentielle ne se résume pourtant pas à un ajustement de la localisation lié au marché du travail. Elle intervient tout au long du cycle de vie non seulement pour se rapprocher d'un emploi mais aussi pour ajuster la consommation d'aménités locales et le logement à de nouveaux besoins et contraintes. On peut donc essayer de penser la mobilité dans ses relations non seulement avec le marché du travail mais aussi avec la distribution spatiale des aménités et le marché du logement. Je précise maintenant le processus de mobilité résidentielle et les mécanismes à l'œuvre dans les choix des ménages.

III.1. Les ajustements en logement

Les choix de logement peuvent être pensés dans une perspective de cycle de vie où des jeunes habitant chez leurs parents vont peu à peu gravir l'échelle menant à la propriété (Ortalo-Magné et Rady, 2006), effectuer des ajustements spécifiques à la cessation d'activité et enfin adapter leur résidence à la fin de vie.

Alors que les jeunes effectuent leurs études puis occupent un premier emploi, ils ne disposent généralement pas de ressources suffisantes pour acheter un logement. Il est possible d'emprunter pour accéder à la propriété mais il est nécessaire de disposer d'un apport personnel suffisant pour pouvoir obtenir un crédit immobilier auprès d'une banque. Les jeunes vont donc vivre en location tout en accumulant du patrimoine grâce à leurs revenus et des épargnes successives. Des ajustements sont effectués vers un logement plus grand lors de la mise en couple et de la naissance des enfants. Les jeunes ménages arrivent finalement à accéder à la propriété lorsqu'ils disposent d'un apport personnel suffisant suite à l'épargne ou à un choc comme un héritage. Toutefois, leurs moyens ne sont pas forcément assez importants pour acheter tout de suite un logement assez grand ou d'assez bonne qualité. Il est donc possible qu'ils déménagent à nouveau plus tard vers un logement plus adapté si leur

patrimoine le leur permet. Les enfants quittent ensuite le domicile parental puis a lieu la cessation d'activité. Ces nouveaux évènements peuvent suffisamment affecter les ménages pour qu'ils ajustent à nouveau leur logement. Alors que s'approche la fin de vie, l'un des conjoints peut décèder. Le conjoint survivant cherche alors parfois un logement plus adapté à la vie seule et une localisation plus proche de services à la personne et de ses enfants qui sont susceptibles de lui apporter de l'aide. Lorsqu'aux grands âges, les problèmes de santé deviennent trop importants, un déménagement dans une maison de retraite ou de soins permet souvent d'obtenir un soutien public ou privé nécessaire jusqu'au décès.
J'ai étudié plusieurs phases de ce processus un peu schématique. Tout d'abord, j'ai travaillé sur les politiques du logement favorisant l'accès à la propriété. Je me suis aussi intéressé à la mobilité des seniors qui est peu étudiée dans la littérature, et plus particulièrement aux choix de logement au moment du passage à la retraite et au moment du veuvage.

III.1.1. L'accession à la propriété

Dans beaucoup de pays développés, l'accès à la propriété est encouragée soit directement par des prêts subventionnés qui complémentent les prêts privés, soit par des avantages fiscaux qui rendent la propriété plus attractive que la location (cf. Rosen, 1985). Enfin, d'autres formes de subventions d'Etat contribuent à l'apport personnel des ménages et leur permettent de remplir les conditions d'emprunt sur le marché du crédit privé. En particulier, ce type de subventions est utile pour les jeunes qui n'ont pas eu le temps d'accumuler du patrimoine.

Dans Gobillon et Le Blanc (2008), je me suis intéressé à la mise en place du Prêt à Taux Zéro (PTZ) en décembre 1995 qui était destiné à favoriser l'accès à la propriété. Le PTZ consiste en un prêt sans intérêt de montant limité auquel à l'origine, les primo-accédants sont les seuls éligibles. Il est généralement remboursé après les autres prêts. Ainsi, le PTZ peut être vu comme une subvention à l'apport personnel des ménages.
Pour estimer l'effet du PTZ, un modèle de mobilité et de choix de statut incorporant les contraintes d'emprunt sur le marché du crédit est d'abord estimé sur la période 1992-1996 durant laquelle très peu de PTZ sont attribués. L'introduction du PTZ est ensuite simulée et son impact sur la mobilité, l'accès à la propriété et la valeur du logement, est évalué. Un apport à la littérature internationale est d'avoir étudié l'effet d'une politique publique simultanément sur la mobilité résidentielle et le choix de statut, ce qui n'est fait que dans très

peu d'études (Zorn, 1989 ; Ioannides et Kan, 1996). Par ailleurs, nous évaluons aussi l'effet du PTZ sur la valeur des logements achetés dans un cadre quasi-structurel cohérent.

L'accès à la propriété est une préoccupation importante des pouvoirs publics depuis la Seconde Guerre Mondiale. Les aides, d'abord apportées à la construction, ont peu à peu été réorientées vers la demande de logement dans les années soixante-dix. A l'origine, les aides à la personne ont pris la forme de prêts subventionnés comme le Prêt d'Accession à la Propriété (PAP) et le Prêt Conventionné (PC). Ces prêts ont eu du succès jusqu'au milieu des années quatre-vingt car le taux d'intérêt réel était très bas à cause de l'inflation. Cependant, la baisse de l'inflation les a ensuite rendus moins attractifs. De plus, l'insolvabilité des emprunteurs est devenue fréquente, ce qui a poussé les banques à resserrer les contraintes de crédit (Lacroix, 1995). Le PTZ a donc été introduit fin 1995 pour remplacer le PAP, les autres aides restant inchangées.
Seuls les primo-accédants ayant de faibles ou de moyens revenus sont éligibles au PTZ. De plus, un seul PTZ peut être attribué par ménage. Le montant du prêt dépend de la taille de la famille et décroit avec le revenu. Il dépend aussi de la zone géographique où le logement est localisé (Ile-de-France ou province). Le PTZ ne peut pas excéder 20% de la valeur du logement et 50% du montant total des prêts contractés pour financer l'achat du logement. Les ménages les plus modestes peuvent repousser le remboursement du PTZ de quinze à dix-neuf ans après la date d'achat, donc après que les autres prêts aient été remboursés. Les ménages peuvent bénéficier d'un PTZ pour acheter un logement ancien, mais uniquement si les réparations dépassent un seuil rarement atteint. Ainsi, les ménages éligibles aux PTZ sont principalement ceux qui souhaitent acheter un logement neuf. Environ 120000 ménages ont bénéficié du PTZ en moyenne chaque année entre 1996 et 1999 (Thomas et Grillon, 2001). Le PTZ représente en moyenne 35% de la valeur des logements achetés. Le montant moyen du PTZ est de 15000 euros.

Contrairement à de nombreuses politiques du logement dans d'autres pays, la cible du PTZ n'est pas les ménages les plus pauvres. Tout d'abord, il ne peut être attribué qu'aux ménages ayant accès au marché du crédit. Ces ménages sont en général plus riches que la plupart des locataires. De plus le critère de revenu ne rend pas seulement les ménages à bas revenus éligibles, mais aussi les familles à revenus moyens, voire même des familles relativement aisées. Il est possible de vérifier à l'aide de l'enquête Revenus Fiscaux qu'en 1998, le critère de revenus rendait 94% des locataires vivant en dehors de l'Ile-de-France éligibles au PTZ. Il

est cependant à noter que ce manque de ciblage n'est pas une exception en France, et que plusieurs politiques du logement partagent cette caractéristique. Le coût du PTZ pour le gouvernement est élevé puisqu'il est compris entre 800 et 900 millions d'euros par an. Ce coût représente entre 7000 et 8000 euros par bénéficiaire, c'est-à-dire environ la moitié du montant des PTZ reçus par les ménages (cf. Gobillon et Le Blanc, 2005). A titre de comparaison, le coût moyen des allocations logement est d'environ 1000 euros par an et bénéficiaire.

Les données disponibles en France ne permettent qu'une étude à court terme des effets du Prêt à Taux Zéro puisqu'il n'existe pas de données longitudinales sur le logement. Nous utilisons deux enquêtes en coupe de l'INSEE qui sont représentatives de la population française. La première est l'Enquête Logement de 1996 dans laquelle les caractéristiques des logements et des ménages sont décrites en détail. Cette enquête contient des informations rétrospectives sur la période de quatre ans précédant la date d'enquête. Elle apporte ainsi des renseignements sur la mobilité entre 1992 et 1996 ainsi que sur le statut par rapport au logement aux deux dates. Une limite des données est qu'elles ne contiennent pas d'information sur le patrimoine des ménages.

Nous utilisons donc une seconde enquête, l'Enquête Patrimoine de 1997, qui donne des informations sur le montant et les composantes du patrimoine des ménages. Les données contiennent aussi les mêmes caractéristiques socio-démographiques que l'Enquête Logement, dont le revenu du ménage.

Nous restreignons notre attention aux locataires du secteur privé. En effet, les propriétaires ne sont pas éligibles au PTZ dont nous souhaitons simuler l'introduction. Par ailleurs, les locataires du public paient des loyers dont le niveau est bien inférieur à celui sur le marché privé. Leur mobilité est donc plus faible que celle des ménages louant dans le secteur privé (Gobillon, 2001). Du fait des particularités du secteur public et de la faible mobilité qui y règne, nous avons choisi d'exclure ce secteur de l'analyse par mesure de simplicité.

Le modèle théorique sur lequel est basée notre spécification économétrique est une version à deux périodes du modèle dynamique de Ioannides et Kan (1996). Le modèle permet d'étudier l'effet des contraintes d'emprunt sur la mobilité résidentielle, le choix de statut et la valeur d'achat des logements à l'aide des données dont nous disposons. Il est possible de le résoudre analytiquement contrairement à sa version dynamique. Le modèle capture les principaux mécanismes intervenant au cours du cycle de vie.

Au début de la première période, un ménage locataire détient un patrimoine et loue un logement d'une certaine taille. Le ménage a une utilité inter-temporelle dépendant de la consommation d'un bien non-durable, du flux de services dérivé de l'occupation de son logement, et de son patrimoine de seconde période. Ce contexte correspond à celui d'un ménage myope qui résume ses possibilités futures par son patrimoine futur. Ce patrimoine est son épargne si le ménage est locataire en seconde période, et c'est la somme de son épargne et de la valeur de son logement s'il est propriétaire. Le flux de services dérivé du logement est supposé proportionnel à la quantité de logement, cette quantité étant un résumé uni-dimensionnel du logement. Ainsi, la quantité de logement entre directement dans la fonction d'utilité.

En début de première période, le ménage reçoit un revenu. Il a le choix entre trois options : rester dans son logement, accéder à la propriété, et déménager en louant son nouveau logement.

Dans le cas où le ménage ne déménage pas, sa quantité de logement reste inchangée, et il ne décide que de sa consommation courante et de son transfert de patrimoine à la période suivante sous forme d'épargne. En cas de déménagement, le ménage peut aussi choisir sa quantité de logement. Cependant, la mobilité résidentielle est supposée entraîner un coût monétaire fixe de déménagement.

Dans le cas de l'accès à la propriété, le ménage a un accès limité au marché du crédit. Plus spécifiquement, le ménage doit faire face à deux contraintes d'emprunt imposées par les banques. La première est une contrainte de revenu. Le ménage ne peut pas effectuer un emprunt tel que la somme à rembourser chaque année soit supérieure à 30% de son revenu courant. La seconde est une contrainte d'apport minimum. Le ménage doit faire un apport d'au moins 20% de la valeur du logement au moment de l'achat. Ces deux contraintes (qui correspondent à celles des prêts les plus courants en France), permettent de déterminer la valeur maximale que le ménage peut financer pour acheter un logement. Sous l'hypothèse que le taux d'intérêt du prêt est le même que celui de l'épargne, la contrainte budgétaire du ménage stipule que son patrimoine en seconde période est égal à la somme de son patrimoine courant et de son revenu à laquelle sont soustraits sa consommation courante, le coût de son logement et les coûts de déménagement. Le coût du logement est le coût d'opportunité d'utiliser l'argent en logement plutôt qu'en placement sous forme d'épargne. Ce coût est négatif si le logement prend suffisamment de valeur. Nous supposons que ce cas ne se présente pas par mesure de simplicité.

Dans le cas de la location du nouveau logement, le contrainte budgétaire est la même sauf que le coût du logement est le loyer. Dans le cas où le ménage reste dans son logement, la contrainte budgétaire est similaire à celle de la location, sauf que le ménage ne paie pas de coûts de déménagement. Le ménage est en outre soumis à une contrainte supplémentaire : il ne peut ajuster sa quantité de logement.

Le ménage maximise son utilité sous contraintes. Le programme de maximisation peut être décomposé en deux étapes. Le ménage détermine tout d'abord son utilité optimale pour chacune des trois options : rester dans son logement, accéder à la propriété et déménager en étant locataire de son nouveau logement. Il choisit ensuite l'option lui garantissant le niveau d'utilité le plus élevé.

Gobillon et Le Blanc (2004) établissent des prédictions du modèle sous certaines conditions. Il est possible de montrer que le choix entre location et propriété en cas de mobilité dépend de la position respective des coûts du logement associés aux deux options. La force des contraintes d'emprunt mesurée par le rapport entre la valeur maximale que le ménage peut financer et sa valeur optimale en l'absence de contraintes a un effet négatif sur la mobilité résidentielle et l'accès à la propriété. Enfin, la décision de mobilité suit une règle (S,s) : le ménage ne déménage que si la différence entre sa quantité optimale de logement et sa quantité courante est assez importante par rapport aux coûts de mobilité. Autrement dit, il existe un intervalle d'inaction autour de la quantité courante tel que le ménage décide de rester dans son logement lorsque la quantité optimale est comprise dans cet intervalle. Il ne déménage que si la quantité optimale est en-dessous de la borne inférieure de l'intervalle ou au-dessus de la borne supérieure.

Le modèle est estimé simultanément sur les données de l'Enquête Logement de 1996 et les données de l'Enquête Patrimoine de 1997. L'Enquête Logement apporte de l'information sur les décisions de mobilité et de choix de statut, les revenus et, pour certains sous-groupes de la population, le loyer optimal et le loyer initial. L'Enquête Patrimoine apporte quant à elle de l'information sur les revenus et la valeur maximale.

Une spécification économétrique est dérivée du modèle sous l'hypothèse que la fonction d'utilité est log-Cobb-Douglas (et sous quelques hypothèses techniques supplémentaires). Cette spécification contient six équations.

La première a trait à la décision de mobilité résidentielle et correspond à la différence d'utilité entre les deux options : rester dans son logement et déménager en étant locataire de son nouveau logement. Elle est de la forme :

$$U_r - U_l = X_1\gamma_1 - \theta_1\left(\ln L_l - \ln L_{t-1}\right)^2 + \varepsilon_1 \tag{14}$$

où les variables explicatives X_1 et les inobservables ε_1 captent les coûts de déménagement de l'individu considéré, et la différence quadratique entre le loyer optimal en cas de déménagement en louant le nouveau logement L_l et le loyer à la période courante L_{t-1} correspond au déséquilibre entre quantité optimale et quantité courante de logement. Ainsi, les composantes de l'équation de mobilité sont ceux de la règle (S,s).

La deuxième équation a trait au choix de statut en cas de mobilité et correspond à la différence d'utilité entre être locataire et être propriétaire de son nouveau logement lors d'un déménagement. Elle est de la forme :

$$U_l - U_p = X_2\gamma_2 - \ln\left(\frac{\rho_t}{p_t}\right) + \theta_2 1_{\{V_{uc}^p > V_{\max}\}}\left(\ln V_{uc}^p - \ln V_{\max}\right) + \varepsilon_2 \tag{15}$$

où ρ_t / p_t désigne le rapport entre les niveaux courants du loyer unitaire et du prix unitaire (définis comme loyer et le prix par unité de logement), X_2 et ε_2 sont respectivement les variables explicatives et les inobservables pouvant négativement influencer les anticipations de prix unitaire par les ménages et encourager la location. Ici, V_{uc}^p est la valeur optimale du logement en l'absence de contraintes d'emprunt et $V_{\max}$ la valeur maximale qu'un ménage peut financer. La différence $\ln V_{uc}^p - \ln V_{\max}$ appliquée uniquement aux ménages contraints capte alors la force des contraintes d'emprunt. Ainsi, l'accès à la propriété est moins attractif lorsque le ménage anticipe une évolution des prix à la baisse et qu'il est plus fortement contraint sur le marché du crédit.

Le loyer optimal n'est observé que pour les ménages mobiles locataires de leur nouveau logement. Une spécification du loyer optimal est donc introduite. Elle constitue la troisième équation du modèle et s'écrit :

$$\ln L_r = X_3\gamma_3 + \phi_1 \ln Y_t + \varepsilon_3 \tag{16}$$

où X_3 et ε_3 sont respectivement les variables explicatives et les inobservables permettant de prendre en compte l'hétérogénéité de goût et Y_t est le niveau courant des revenus.

Il est possible que les revenus du ménage introduits dans l'équation de loyer soient endogènes du fait de la présence de caractéristiques individuelles inobservables. Ce problème d'endogénéité est pris en compte grâce à la quatrième équation qui est une spécification de revenus :

$$\ln Y_t = X_4\gamma_4 + \varepsilon_4 \tag{17}$$

où X_4 et ε_4 sont des caractéristiques observables et inobservables des ménages.

La valeur maximale ne peut être calculée que pour les ménages de l'Enquête Patrimoine. Une spécification de la valeur maximale est donc introduite. Elle constitue la cinquième équation du modèle et s'écrit :

$$\ln V_{\max} = X_5\gamma_5 + \phi_2 \ln Y_t + \varepsilon_5 \tag{18}$$

où X_5 et ε_5 sont des caractéristiques observables et inobservables du ménage pouvant affecter son patrimoine et donc son apport personnel. Le revenu courant joue non seulement directement sur la contrainte de revenus mais sert aussi d'approximation au revenu permanent qui influence le patrimoine.

Enfin dans les données de l'Enquête Logement, le loyer à la date initiale n'est observé que pour les ménages ne déménageant pas. La sixième équation permet d'imputer ce loyer aux autres ménages :

$$\ln L_{t-1} = X_6\gamma_6 + \varepsilon_6 \tag{19}$$

où X_6 et ε_6 sont des caractéristiques observables et inobservables du logement servant à l'imputation.

Comme les revenus des ménages sont communs aux deux bases, il est possible d'autoriser une corrélation entre le résidu de l'équation de revenus et les résidus de toutes les autres équations. Les résidus de toutes les équations sont alors corrélés (certes de façon restrictive).

Du fait du nombre élevé d'équations et des non linéarités de certaines d'entre elles, le modèle est estimé par maximum de vraisemblance simulée.

Les estimations montrent qu'en accord avec la théorie, plus le loyer optimal est éloigné du loyer à la date initiale, plus les ménages ont tendance à déménager. De plus, la force des contraintes diminue l'accès à la propriété.
Le modèle prédit par ailleurs que le taux de ménages contraints sur le marché du crédit est de 53%. Ce taux est inférieur à celui obtenu par Zorn (1989) pour l'ensemble de la population américaine, qui est de 61%. Nous simulons aussi un desserrement des contraintes d'emprunt se traduisant par une hausse de 10% de la valeur maximale pour l'ensemble des ménages. La proportion de ménages contraints diminue alors de 3.5% et le flux de ménages choisissant la propriété augmente de 6%. Alors qu'un quart des ménages additionnels accédant à la propriété aurait choisi une mobilité avec location du nouveau logement en l'absence de contrainte, trois quarts d'entre eux n'auraient pas déménagé. Ainsi, les transitions d'une situation d'immobilité à l'accès à la propriété dominent les transitions de la location à la propriété comme aux USA (Zorn, 1989).

Nous simulons ensuite l'introduction du PTZ qui est grossièrement équivalent à une subvention à l'apport calculée en soustrayant au montant initial du PTZ (qui est possible d'évaluer à partir des règles d'éligibilité), la valeur actualisée des remboursements futurs. La subvention est simplement ajoutée à la valeur maximale que le ménage peut financer. Nous trouvons que l'introduction du PTZ aurait bénéficié à 530000 ménages en quatre ans. Comme nous n'avons pas pris en compte la restriction du PTZ aux nouveaux logements et que certains propriétaires achètent des anciens logements, ce chiffre constitue une borne supérieure. D'après le Ministère du Logement, le chiffre réel calculé pour la période 1996-1999 est de 420000. Nos simulations montrent que le PTZ serait à l'origine de l'accès à la propriété de 75000 ménages supplémentaires. En l'absence de PTZ, 70% de ces ménages additionnels seraient restés dans leur logement et 30% d'entre eux auraient déménagé en louant leur nouveau logement.
Au niveau de l'efficacité, le PTZ souffre d'un effet d'aubaine de 85% : c'est-à-dire que 85% des ménages bénéficiant d'un PTZ auraient choisi d'accéder à la propriété même en l'absence de ce prêt. Ce chiffre est cohérent avec ceux proposés lors d'autres évaluations : par exemple, un rapport des Ponts et Chaussées considère que l'effet d'aubaine est compris entre 75% et 90% (Welhoff, 2004).

Un autre objectif du PTZ est de stimuler la construction de logements de meilleure qualité en les rendant accessibles aux ménages malgré leur prix plus élevé. En fait, la réponse de la valeur moyenne d'achat au desserrement des contraintes d'emprunt est la somme de deux effets. D'une part, les ménages qui auraient décidé d'accéder à la propriété même en l'absence de PTZ vont acheter des logements plus coûteux. D'autre part, les ménages marginaux qui accèdent à la propriété à cause du PTZ sont moins riches que les ménages supra-marginaux et vont acheter des logements moins coûteux. L'effet du PTZ est donc en théorie ambigu. Nous trouvons grâce à des simulations que l'introduction du PTZ aurait pour effet de diminuer la valeur moyenne du logement de 3000 euros. Ainsi, l'effet de sélection dominerait. Le PTZ pourrait donc encourager la construction de logements de moins bonne qualité ou localisés en périphérie des villes où le prix de la terre est généralement plus faible. Il est possible de vérifier que la distribution spatiale des primo-accédants sur la période 1998-2002 est bien différente pour les ménages ayant bénéficié du PTZ et ceux qui n'en ont pas bénéficié. Parmi les non bénéficiaires, 30% habitent en milieu rural et 34% vivent dans une aire urbaine de plus de 100000 habitants. Parmi les bénéficiaires, ces chiffres sont respectivement de 48% et 19% (Daubresse, 2003). Le PTZ pourrait alors être à l'origine de problèmes d'accès physique aux emplois.

On peut calculer le rapport entre la somme d'argent supplémentaire utilisée par les ménages en achat de logement suite à l'introduction du PTZ, et le coût public de la mesure. Ce rapport, appelé « effet multiplicateur », prend une valeur aux alentours de 1.2. Cette valeur est relativement faible, les politiques publiques implémentées dans d'autres pays ayant un effet multiplicateur qui est souvent de l'ordre de 2.

En résumé, le PTZ augmente significativement l'accès à la propriété, en particulier celui de ménages qui seraient restés dans leur logement en son absence. Toutefois, il créé un effet d'aubaine important et son impact quantitatif est relativement faible par rapport à son coût. Il est aussi susceptible d'encourager la construction de logements de qualité assez faible ou localisés à distance des emplois. Par ailleurs, il est à noter que l'effet du PTZ que nous avons calculé ne constitue certainement qu'une bonne supérieure puisque la restriction du PTZ aux logements neufs et l'augmentation des prix due au PTZ n'ont pas été prises en compte.

III.1.2. Les choix de logement des séniors

De nombreux pays connaissent un vieillissement de la population sans précédent. Les gouvernements se demandent comment ce changement démographique va remettre en question le système des pensions de retraite et le système de soins. Les ajustements en logement des seniors et leur impact sur le marché de l'immobilier sont moins souvent au centre des débats. Pourtant, certains évènements du cycle de vie, comme la retraite ou le veuvage, sont spécifiques aux seniors et peuvent affecter les besoins en logement de façon particulière.

Les études sur les séniors sont généralement très générales. Elles concernent une population de plus de 55 ans et analysent conjointement l'effet de tous les évènements économiques et démographiques ayant lieu aux âges élevés (cf. Venti et Wise, 1990, pour les USA ; Ermisch et Jenkins, 1999, pour la Grande-Bretagne). Dans mes travaux, j'ai tenté de mettre en évidence les particularités de la décision de mobilité et des choix de logements lors de deux évènements spécifiques : le passage à la retraite et le veuvage. Les études que j'ai menées montrent que les comportements diffèrent fortement selon la transition étudiée.

La mobilité au moment de la retraite

Les arbitrages des ménages sont profondément modifiés au moment de la retraite. Alors que la localisation des emplois occupés joue le rôle d'aimant dans le choix d'un site de résidence durant la vie active, ce pôle d'attraction disparaît complètement au moment de la retraite. Parallèlement, les retraités connaissent une forte augmentation de leur temps libre et une diminution de leurs revenus comme ils ne bénéficient plus que de pensions dont le montant s'élève à environ 80% du niveau de leurs salaires passés.

Gobillon et Wolff (2009) étudient la mobilité résidentielle et les choix de logement des seniors au moment du passage à la retraite. Les arbitrages des ménages peuvent être schématisés en considérant un seul individu devenant retraité. Comme les revenus de cet individu baissent, il est poussé à diminuer ses coûts de logement et donc à réduire la taille ou la qualité de son logement. En effet, un logement plus petit ou de moins bonne qualité est généralement associé à un loyer plus faible pour un locataire, et des coûts d'entretien plus faibles pour un propriétaire. Cependant, un nouveau retraité pourrait vouloir utiliser son temps de loisir supplémentaire à son domicile et valoriserait alors plus son logement. Dans ce cas, il

peut vouloir augmenter la taille de son logement. Ces deux mécanismes suffisent pour affirmer qu'un nouveau retraité pourrait vouloir ajuster la taille et la qualité de son logement à la hausse ou à la baisse. Effectuer de tels ajustements entraîne une mobilité qui a des coûts monétaires directs comme les coûts de transport du mobilier ou les frais de transaction pour les propriétaires. Un déménagement entraîne aussi des coûts non monétaires puisque le capital local sur le lieu de résidence, comme la connaissance des lieux ou les interactions sociales avec les voisins et commerçants, est généralement perdu lors d'un changement de logement. L'individu ne déménagera donc que si les bénéfices qu'il tire d'un ajustement de sa localisation et de son logement sont supérieurs aux coûts de mobilité.

Bien sûr, il existe plusieurs variations à ce scénario. En particulier, la décision de mobilité d'un couple va dépendre du statut d'activité des deux membres du couple. Lorsque l'un des membres prend sa retraite et que l'autre continue à travailler, le couple pourra décider de rester dans son logement et d'attendre que le second membre du couple prenne sa retraite pour déménager. La localisation du lieu de travail joue alors le rôle d'ancre géographique.

En outre, un ménage mobile a le choix entre louer et acheter son nouveau logement. La littérature considère généralement que ce choix revient à un arbitrage basé sur les coûts des deux options. Pour un locataire, le coût est son loyer. Pour un propriétaire, son coût inclut les coûts de maintenance et un coût d'opportunité de placement. En effet, le rendement du logement dépend de la croissance des prix immobiliers. Le ménage supporte un coût d'opportunité quand ce rendement est inférieur à celui d'actifs alternatifs. Lorsqu'ils sont jeunes, les ménages doivent généralement emprunter de l'argent pour pouvoir acheter un logement. C'est moins souvent le cas pour les ménages prenant leur retraite car ils ont eu le temps d'accumuler du patrimoine. En particulier, un grand nombre d'entre eux sont déjà propriétaires et peuvent revendre leur logement pour en acheter un nouveau. Toutefois, ces opérations immobilières entrainent des dépenses notariales assez élevées. Comme la période de vie durant laquelle les nouveaux retraités peuvent amortir ces dépenses est plus courte que pour les jeunes, les dépenses notariales peuvent être dissuasives.

Le comportement des ménages dépend aussi de leurs préférences pour le présent qui ont pu influencer leur comportement d'épargne. En effet, un individu ayant une forte préférence pour le présent tend à dépenser ses revenus pendant sa vie active et à occuper un grand logement à un coût élevé. Lors de sa cessation d'activité ses revenus baissent et comme il n'a pas d'épargne, il est amené à déménager pour occuper un logement plus petit. Considérons maintenant un individu qui a une faible préférence pour le présent et préfère accumuler du patrimoine qu'il pourra utiliser après sa cessation d'activité lorsqu'il a plus de temps libre. Cet

individu consomme peu durant sa période d'activité et occupe un petit logement. Après sa retraite, il possède du patrimoine qu'il peut utiliser pour financer un logement plus grand et plus coûteux dont il peut bénéficier durant son temps libre.

Les arbitrages des ménages dépendent aussi de leurs préférences pour vivre à l'intérieur plutôt qu'à l'extérieur de leur logement. Les ménages ayant une forte préférence pour vivre en dehors de leur logement seront enclins à investir dans un jardin plutôt que d'augmenter significativement la taille de leur logement au moment de la retraite. La localisation a aussi plus d'importance pour eux et ils pourraient préférer habiter un lieu proposant beaucoup d'aménités de consommation comme un climat agréable, un accès à la côte ou de beaux paysages. Il est considéré au moins depuis Roback (1982) que le niveau des loyers et des salaires compense la présence d'aménités de consommation. En effet, les sites présentant un niveau d'aménités de consommation élevé se caractérisent aussi par des salaires plus faibles et des loyers plus élevés. Comme les retraités ne sont plus sur le marché du travail, ils auront tendance à se localiser sur le site où les aménités de consommation sont principalement compensées par des salaires faibles (Knapp et Graves, 1988). De tels sites incluent de larges espaces au climat agréable où l'occupation du sol n'est pas très contrainte, comme par exemple des localisations rurales du Sud et du Sud-Ouest de la France (dans des départements comme l'Ardèche et le Lot). En revanche, les localisations agréables sur la côte auront tendance à se caractériser par des loyers élevés car l'occupation du sol y est plus contrainte, et ils n'attireront probablement que les retraités les plus riches. Empiriquement, il a été montré que les retraités préfèrent habiter les lieux plus ensoleillés (Chen et Rosenthal, 2008). Les retraités pourraient aussi vouloir passer plus de temps avec leurs enfants et petits enfants, et essayer d'habiter plus près de leur famille (De Coulon et Wolff, 2006).

Le choix de localisation aura tendance à affecter les ajustements en logement. Il est probable que les ménages veuillent diminuer la taille de leur logement si les coûts d'habitation sur leur site de destination sont élevés, et si le climat très favorable permet de passer des moments agréables à l'extérieur. En revanche, ils auront tendance à augmenter la taille de leur logement si les coûts d'habitation sur le site de destination sont faibles et que le climat est défavorable.

Les choix de localisation et les ajustements en logement des ménages au moment du passage à la retraite sont étudiés empiriquement dans Gobillon et Wolff (2009). Alors que la plupart des prédictions théoriques sont ambigües, les tendances prédominantes au moment de la cessation d'activité peuvent ainsi être précisées. Deux bases de données sont utilisées : l'enquête Trois Générations de 1992 et le Panel Européen des Ménages sur la période 1994-2001.

Les données de l'enquête Trois Générations contiennent de l'information sur les grands-parents des personnes enquêtées qui sont âgés de 68 à 92 ans. Des questions rétrospectives identifient les déménagements au moment de leur retraite ainsi que les raisons (parfois multiples) de leur mobilité résidentielle. L'enquête montre que les déménagements au moment de la retraite sont un phénomène important puisque 31.5% des grands-parents affirment avoir déménagé. Parmi les nouveaux retraités mobiles, 44.1% se sont relocalisés dans une autre région. Ils sont 16.5% à avoir déménagé pour vivre plus près de leur famille et 15.5% pour ajuster leur logement. Les raisons liées à la géographie sont plus rarement évoquées (9.2%) et sont de nature multiple : alors que certains nouveaux retraités mobiles souhaitent se rapprocher de leur lieu de naissance, d'autres migrent vers un climat plus favorable dans le Sud et le Sud-Ouest de la France. Enfin, une large fraction des nouveaux retraités mobiles (29.1%) a déménagé pour des raisons professionnelles. Il y a plusieurs explications à ce résultat : l'existence de nouvelles contraintes financières suite à la cessation d'activité, la perte d'un logement de fonction, la perte d'un financement professionnel du logement.

Les raisons avancées dépendent aussi des caractéristiques des grands-parents répondants. Les femmes déclarent plus souvent des motivations familiales. Les raisons géographiques sont plus fréquentes chez les plus éduqués et les plus riches. Une explication à ce résultat est l'attrait des villes situées sur des côtes ensoleillées (comme Biarritz, Cannes ou Nice) où le logement est généralement coûteux et abordable uniquement pour les plus riches. Les ajustements en logement dépendent aussi du niveau des revenus. 14% des individus dans le premier quartile de revenus reportent qu'ils ont déménagé pour diminuer la taille de leur logement, contre seulement 2.9% des individus dans le dernier quartile. Réciproquement, 4.2% seulement des individus du premier quartile reportent qu'ils ont déménagé pour augmenter la taille de leur logement, contre 12.5% des individus dans le dernier quartile.

Nous avons analysé plus précisément les mécanismes sous-jacents à la mobilité et aux choix de logements avec le Panel Européen des Ménages. Le panel suit les individus sur toute la période 1994-2001 même lorsqu'ils déménagent. Il est possible de vérifier que pour les seniors ayant cessé leur activité durant la période d'observation, il existe un pic de mobilité l'année même de la retraite : 43% des individus mobiles déménagent exactement l'année de leur cessation d'activité. La proportion est aussi relativement importante l'année suivant la retraite et s'élève à 26%. En comparaison, seulement 17% des déménagements ont lieu durant les deux années précédant le passage à la retraite. La mobilité est plus fréquente chez les

nouveaux retraités possédant une résidence secondaire. Ce résultat suggère que les résidences secondaires pourraient être des logements privilégiés lors de la cessation d'activité.

Il est possible d'analyser la décision de mobilité avec un modèle probit pour les seniors âgés de 55 à 70 ans, l'unité d'observation étant l'individu-année. Nous trouvons les résultats classiques qu'avoir un conjoint qui travaille, être propriétaire, vivre dans une maison, et être resté longtemps dans son logement décroit la probabilité de déménager. Le nombre de pièces en excès, défini comme le nombre de pièces moins le nombre de personnes du ménage, n'a pas d'effet significatif. Ainsi, l'ajustement de la taille du logement ne serait pas une des raisons principales à un déménagement. Enfin la retraite elle-même a un effet positif sur la mobilité quand elle a eu lieu l'année considérée. Elle a un effet positif mais non significatif quand elle a lieu l'année précédente. L'effet de la retraite est plus fort pour les locataires que pour les propriétaires, peut-être parce qu'ils ont pris l'habitude de déménager, qu'ils n'ont plus de quoi payer leur loyer au moment de la retraite ou que leurs coûts de déménagement sont plus faibles.

Il est aussi possible d'analyser les ajustements en logement entre la situation l'année précédant un déménagement et celle qui le suit. Les ménages mobiles tendent à corriger un déséquilibre entre le nombre de pièces et le nombre d'occupants lorsqu'ils déménagent. Toutefois, il ne faut pas perdre de vu que même s'il y a correction du déséquilibre, ce n'est pas le déséquilibre lui-même qui cause la mobilité. Il est intéressant de constater que les ménages mobiles tendent à améliorer la qualité de leur logement. Ainsi, des ménages qui n'avaient pas de baignoire, de douche, d'eau chaude, de système de chauffage performant, de terrasse ou de jardin tendent à être mieux équipés après un déménagement. De plus, la qualité du voisinage est généralement meilleure : moins de ménages occupent un quartier jugé dangereux ou concentrant des problèmes de vandalisme.

Par ailleurs, les propriétaires mobiles tendent à rester propriétaires. L'inertie est moindre pour les locataires, peut-être parce que certains d'entre eux n'avaient choisi la location qu'à cause d'une mobilité importante liée à leur activité professionnelle durant leur vie active. Ces ménages peuvent en revanche choisir un logement définitif en propriété au moment de la retraite. Parallèlement, les ménages mobiles occupant une maison tendent aussi à occuper une maison après leur déménagement. L'inertie est plus faible pour les ménages mobiles occupant un appartement. Alors que certains arrivent à accéder à une maison, il est probable que nombre d'entre eux n'ont pas les ressources financières pour y arriver, d'autant plus que les maisons sont plus fréquemment en propriété que les appartements. Les ménages mobiles

choisissent souvent une localisation avec un climat agréable dans le Sud ou le Sud-Ouest lors de leur déménagement.

En résumé, nos résultats empiriques montrent que la mobilité résidentielle est substantielle au moment de la retraite. Les ménages tendent à ajuster la taille de leur logement pour corriger un déséquilibre entre le nombre de pièces et le nombre d'habitants. L'ajustement de la taille du logement n'est pourtant pas un facteur déterminant de la mobilité. D'autres facteurs souvent liés à la qualité du logement jouent un rôle plus important dans l'occurrence des déménagements. En fait, les ménages mobiles tendent en moyenne à améliorer la qualité de leur logement et de leur environnement suite à leur déménagement.

La mobilité au moment du veuvage

Les déterminants des comportements de mobilité et d'ajustement en logement sont bien différents lorsque l'un des membres d'un couple décède et que son conjoint reste seul(e). Il existe certes une diminution des revenus comme au moment de la retraite, le conjoint survivant ne touchant environ que 60% de la pension de retraite du conjoint décédé sous forme de pension de réversion. Mais les changements de préférences ne sont plus liés à l'augmentation du temps libre et concernent plutôt le mode d'occupation du logement et l'aide à la personne. Les décisions de mobilité résidentielle et de choix de logement au moment du veuvage ont été analysées dans Bonnet, Gobillon et Laferrère (2009).

Une analyse descriptive montre que la proportion d'individus veufs dans la population augmente naturellement avec l'âge et est plus importante chez les femmes que chez les hommes. Par exemple, le taux de veuvage à 80 ans est de 60% pour les femmes nées en 1920 contre seulement 17% pour les hommes. Pour faciliter l'exposition de mes travaux, j'utiliserai à partir de maintenant le terme de « veuve » pour parler du conjoint survivant sans faire de distinction par rapport au sexe. Il est aussi possible de noter qu'à âge donné, la proportion de veuves diminue avec la cohorte. Cet effet est dû à l'augmentation de l'espérance de vie et à l'occurrence du veuvage plus tard dans le cycle de vie. La plupart des veuves âgées de 60 à 85 ans vivent seules. La co-résidence avec les enfants est rare, et devient même moins fréquente chez les générations plus jeunes (Flipo, Le Blanc et Laferrère, 1999).

La législation peut affecter les choix de logement des nouvelles veuves. En effet, d'après la législation sur les mariages appliquée jusqu'en 2001, les biens acquis durant le mariage sont

mis en commun, et la moitié d'entre eux appartient à chaque conjoint. Lors d'un décès, la moitié des biens continue d'appartenir au conjoint survivant, et l'autre moitié revient aux héritiers. Cette seconde moitié, dont éventuellement un logement en propriété, est divisée entre le conjoint survivant et les enfants. Dans la plupart des cas, le transfert de propriété aux enfants ne change pas les modalités d'occupation du logement pour la veuve. Si le seul bien du ménage est le logement, les enfants peuvent en revanche pousser la veuve à déménager pour vendre le logement et diviser le produit de la vente entre les héritiers. Ce type de situation peut cependant être évité si le conjoint décédé a fait un testament garantissant l'usufruit du logement à la veuve pour le temps qui lui reste à vivre. D'après ces règles, on peut s'attendre à ce que plus une veuve a d'enfants, plus il est probable qu'elle soit forcée de déménager.

D'autres mécanismes basés sur les revenus et les préférences peuvent induire des ajustements en logement. En particulier, la pension de réversion n'est généralement pas suffisante pour compenser les coûts de logement et la veuve peut vouloir déménager pour diminuer sa consommation de logement. De plus, le logement est un bien public partiel pour lequel il peut y avoir d'importantes économies d'échelle (Nelson, 1988). Lorsqu'un des deux membres du couple décède, ces économies d'échelle disparaissent. Cependant, il en est de même pour les effets de congestion et certaines des pièces peuvent par ailleurs être utiles pour loger des aides de soins, des membres de la famille ou des visiteurs permettant de surmonter la solitude. De façon générale, nous pensons que pour une nouvelle veuve, les bénéfices associés à l'occupation d'un grand logement sont relativement faibles comparés aux coûts d'occupation, surtout lorsque la pension de réversion est relativement faible. Il est donc probable qu'une nouvelle veuve souhaite diminuer sa consommation de logement. C'est d'autant plus vrai dans le cas où elle fait face à une contrainte de liquidité et est forcée de déménager.

Si les choix de logement du couple ont été effectués en anticipant le veuvage de l'un des membres du couple, la consommation de logement sera plus proche de l'optimum de la veuve et un déménagement est moins probable.

Une diminution de la consommation de logement peut être effectuée en diminuant la qualité du logement, ou en diminuant le nombre de pièces. L'ajustement de la taille du logement par les seniors a été étudié pour plusieurs pays. Venti et Wise (2001) montrent qu'aux USA, les personnes âgées ne réduisent pas la valeur de leur logement si ce n'est lorsqu'elles doivent faire face à un choc comme le veuvage. Ermisch et Jenkins (1999) pour la Grande-Bretagne et Angelini et Laferrère (2008) pour les pays européens, trouvent que la mobilité résidentielle est

plus faible pour les personnes âgées et mène souvent à une diminution de la taille du logement, en particulier aux âges élevés.

Du fait de l'existence de coûts de déménagement, il n'est pas toujours rentable pour une nouvelle veuve de déménager. En fait, une nouvelle veuve ne déménagera que si sa consommation optimale de logement est suffisamment éloignée de sa consommation courante pour compenser les coûts de mobilité. La faible mobilité des personnes âgées peut être expliquée par l'existence de coûts de mobilité non monétaires importants. En effet, les personnes âgées sont généralement en moins bonne santé, ont pris des habitudes de vie dans leur logement et leur quartier au cours du temps, et ont une connaissance de leur voisinage qui est perdue si elles déménagent. De plus, les nouvelles veuves sont en général assez âgées et disposent donc de moins de temps pour rentabiliser un coût de déménagement. Par ailleurs, il est probable que les nouvelles veuves propriétaires soient moins mobiles que celles qui sont locataires car leurs coûts de mobilité sont généralement plus importants. Ensuite, comme les coûts de maintenance pour les propriétaires sont généralement assez élevés, on s'attend à ce que les nouvelles veuves propriétaires deviennent plus souvent locataires que les autres ménages propriétaires lors d'un déménagement, du fait de la baisse de leurs revenus.

Le décès d'un conjoint peut aussi changer les préférences de localisation du ménage. En effet, les deux conjoints peuvent avoir des préférences différentes qui ont mené à un compromis sur leurs choix de logement durant leur vie en couple. Le veuvage peut permettre au conjoint survivant d'exprimer ses préférences personnelles et de choisir un autre logement. Le décès d'un conjoint peut aussi changer les préférences du conjoint survivant car la veuve peut avoir besoin d'une aide à la personne en particulier si elle a des problèmes de santé. Une nouvelle veuve pourrait souhaiter se relocaliser près de ses enfants ou à un endroit où les aménités de consommation peuvent lui permettre de vivre indépendamment. En particulier, les services à la personne sont généralement plus courants dans les villes. Une nouvelle veuve propriétaire de son logement en milieu rural peut vouloir migrer vers une ville pour bénéficier d'un meilleur accès aux services à la personne. Comme la proportion d'appartements en location est plus importante dans les villes qu'à la campagne et que la proportion de maisons en propriété y est plus faible, la nouvelle veuve mobile pourra être amenée à changer de type de logement en devenant locataire d'un appartement.

Une analyse empirique de la mobilité et des choix de logement est menée à partir des Enquêtes Logement de 1996 et 2002. Des questions rétrospectives permettent d'identifier s'il y a eu un déménagement durant les quatre années précédant chaque enquête. Les données

apportent aussi de l'information sur plusieurs caractéristiques du logement et des ménages quatre ans avant chaque enquête. L'enquête de 2002 contient par ailleurs une question sur l'existence d'enfants en-dehors du ménage.

Nous étudions la mobilité résidentielle durant les quatre ans précédant les enquêtes. L'échantillon retenu est celui des ménages dont la personne de référence est retraitée ou inactive, et dont l'âge est compris entre 60 et 85 ans quatre ans avant la date d'enquête. Nous limitons ainsi les choix liés à l'activité professionnelle et les départs en maison de retraite. Nous considérons le veuvage pour les couples vivant seuls et définissons un veuvage comme une diminution de la taille du ménage de deux membres à un membre, le membre restant ayant un statut de veuve à la date d'enquête.

De façon guère surprenante, le taux de veuvage est plus important aux âges élevés. Durant la période 1998-2002 par exemple, environ 14% des couples dont la personne de référence est âgée de 65 à 69 ans connaissent un décès. Le taux correspond est de 30% pour les couples dont la personne de référence est âgée de 80 à 84 ans. Nous souhaitons comparer la mobilité des nouvelles veuves avec celle des couples stables (ie. deux individus en couple à la date d'enquête et quatre ans avant) et celle des veuves stables (ie. un individu veuf vivant seul à la date d'enquête et quatre ans avant). Durant la période 1998-2002, le taux de mobilité des nouvelles veuves, qui s'élève à 13.3%, est plus de deux fois supérieur à celui des couples stables. Il est aussi à noter que le taux de mobilité des veuves stables est bien plus faible que celui des nouvelles veuves et s'élève à 7.9%.

Nous étudions la mobilité au moment du veuvage à l'aide d'un modèle probit. Les résultats confirment les statistiques descriptives : toutes choses égales par ailleurs, les nouvelles veuves sont plus mobiles que les couples stables et les veuves stables. Il est donc probable que si un veuvage induit une mobilité, celle-ci ait lieu moins de quatre ans après le décès du conjoint et non plus tard. La mobilité décroit jusqu'à 80 ans et ensuite augmente. Les ménages ayant des enfants en-dehors de leur domicile sont plus mobiles que ceux qui n'en ont pas. Ce résultat est compatible avec une relocalisation des parents plus près de leurs enfants soit pour recevoir leur aide, soit pour s'occuper des petits enfants. Les propriétaires sont moins mobiles que les locataires du public, eux-mêmes moins mobiles que les locataires du privé, ce qui a déjà été observé pour la population en général (Gobillon, 2001). Le nombre de pièces en excès n'a pas d'effet significatif sur la mobilité comme dans l'étude sur les déménagements au moment du passage à la retraite menée par Gobillon et Wolff (2009).

Nous avons aussi effectué des régressions séparément pour les nouvelles veuves, les couples stables et les veuves stables. L'âge et l'existence d'enfants en-dehors du domicile ont un effet particulier pour les nouvelles veuves. La mobilité des nouvelles veuves ne décroit pas avec l'âge et augmente plus au-delà de 80 ans que pour les autres groupes. La mobilité plus importante des nouvelles veuves aux âges avancés peut être due à des problèmes de santé. Alors que dans un couple, une personne avec des problèmes de santé peut s'appuyer sur l'aide de son conjoint, une veuve âgée doit souvent déménager pour pouvoir bénéficier d'une aide à la personne. Comme les individus en maison de retraite sont exclus de notre échantillon, la mobilité résidentielle importante entre logements privés au-dessus de 80 ans est cohérente avec le fait que les nouvelles cohortes de personnes âgées essaient de vivre indépendamment aussi longtemps que possible. L'existence d'enfants en-dehors du ménage augmente la propension à déménager des nouvelles veuves (significativement à 10%), mais n'a pas d'effet significatif pour les couples stables et les veuves stables. Peut-être les nouvelles veuves déménagent-elles lorsqu'elles ont des enfants pour bénéficier de leur aide. Elles ont peut-être aussi subi des pressions pour déménager au moment de l'héritage. Parmi les personnes dont le conjoint décède, les femmes ont une propension à déménager plus importante que les hommes. Une explication possible est l'existence de problèmes de santé plus importants aux grands âges (Cambois, Désesquelles et Ravaud, 2003). Enfin, le nombre de pièces en excès a un effet positif sur la mobilité pour les nouvelles veuves et non pour les couples stables. Ainsi il est plus probable que leur mobilité soit causée par un ajustement de leur consommation de logement, en particulier à cause de contraintes financières plus importantes.

Nous avons aussi étudié les ajustements en logement des ménages suite à un déménagement durant les périodes 1992-1996 et 1998-2002 à l'aide de modèles logit multinomiaux. Tout d'abord, les changements de taille de logement ont été examinés en distinguant pour le nombre de pièces : une hausse, une baisse, et l'absence de changement. La baisse de la taille du logement est plus courante après 75 ans. Par ailleurs, le nombre de pièces en excès a un effet positif sur la propension à diminuer la taille du logement et un effet négatif sur la propension à l'augmenter. Ainsi, les déménagements auraient tendance à corriger le déséquilibre entre la taille des logements et le nombre d'occupants comme dans le cas des déménagements liés au passage à la retraite. Les nouvelles veuves diminuent significativement plus la taille de leur logement que les couples stables. De plus, pour les ménages mobiles habitant initialement une maison, les transitions vers un appartement sont plus courantes chez les nouvelles veuves que chez les couples stables. Une explication à cette différence pourrait être que les maisons sont généralement localisées dans des quartiers

résidentiels et situés assez loin des services d'aide à la personne. Déménager d'une maison à un appartement pourrait permettre aux nouvelles veuves d'améliorer leur accès aux services en ce rapprochant d'un centre-ville. Par ailleurs, parmi les propriétaires mobiles, les transitions vers la location sont plus courantes pour les nouvelles veuves que pour les couples stables. Ce résultat est compatible avec le besoin plus important des veuves de diminuer les coûts d'occupation du logement et de se rapprocher d'un centre-ville. Il est à noter qu'environ un-tiers des nouvelles veuves propriétaires qui choisissent de louer leur nouveau logement le font dans le secteur public qui propose des logements spécifiques adaptés aux personnes âgées. Enfin, les nouvelles veuves mobiles déménagent plus souvent que les couples stables vers une plus grande commune. Une explication possible est la présence plus importante dans les grandes villes de services à la personne qui sont probablement plus prisés par les veuves.

Nous avons aussi examiné les raisons avancées par les ménages à leur déménagement. La principale raison avancée par les nouvelles veuves (26% d'entre elles) est leur souhait de vivre plus près de leur famille ou de leur lieu de naissance. Cette raison n'est avancée que par 15% des veuves stables et 12% des couples stables. La seconde raison donnée par les nouvelles veuves est la diminution de la taille du logement (18%), les proportions correspondantes étant seulement de 12% pour les veuves stables et 5% pour les couples stables. La troisième raison avancée par les veuves est liée à la qualité du quartier et à la localisation (13%). Ce type de raisons est plus souvent avancé par les couples stables (21%) et les veuves stables (16%). Il est à noter qu'environ un cinquième des nouvelles veuves mobiles déclarent déménager pour une autre raison. Laferrère (2005) observe que cette déclaration devient plus fréquente avec l'âge et suggère qu'elle pourrait être liée à des problèmes de santé.

Nous avons finalement comparé la distance aux enfants pour différents sous-groupes de la population à l'aide de l'Enquête Logement 2002. Toutes choses égales par ailleurs, les nouvelles veuves mobiles habitent plus près de leurs enfants que les nouvelles veuves non mobiles et les couples stables (mobiles ou non). Habiter plus près des enfants pourrait permettre aux nouvelles veuves d'obtenir une aide plus régulière.

Des simulations évaluent l'effet de l'augmentation du nombre de veuves sur la demande de logements durant la période s'étendant jusqu'en 2030. Le nombre de veuves augmentera en moyenne d'environ 19 000 individus par an. Ces nouvelles veuves représenteront une demande supplémentaire de logements d'environ 9%. Cette proportion varie selon le type de

logement considéré. Elle est bien supérieure pour les petits logements d'une ou deux pièces pour lesquels elle varie dans la fourchette 13%-19% selon le scénario considéré.

En résumé, nos résultats empiriques suggèrent que les nouvelles veuves sont plus mobiles que les couples et diminuent leur consommation de logement. En moyenne, elles tendent à déménager vers des appartements en location dans de plus grandes municipalités où les services à la personne sont plus accessibles. L'augmentation de la demande de logements due aux veuves additionnelles sera significative dans les vingt années qui viennent (plus particulièrement, celle de petits logements) mais reste peu importante.

III.2. La migration comme moyen de bénéficier d'opportunités d'emploi

Nous avons vu que durant la vie active, la mobilité résidentielle peut être causée par des ajustements en logement et plus particulièrement par l'accès à la propriété. Les analyses qui ont été menées ne prennent pas en compte l'espace. Pourtant, tant le coût du logement que les revenus peuvent dépendre de la localisation. Une littérature sur les migrations s'est développée parallèlement à la littérature sur les choix de logement. Elle s'intéresse principalement à l'effet d'opportunités d'emploi localisées sur les changements de lieu de résidence. Les coûts locaux du logement tiennent jusqu'à maintenant une place secondaire dans les travaux sur les migrations, probablement parce que des données fiables couvrant le territoire à une échelle fine sont rarement disponibles (c'est en particulier vrai pour la France). Il existe un enjeu particulier à intégrer les choix de localisation dans les analyses des disparités spatiales sur le marché du travail. En effet, lorsque le lieu de résidence est choisi sur la base d'une opportunité d'emploi, une mesure simple de l'effet de la localisation sur le salaire ou le chômage individuel peut souffrir d'un biais d'endogénéité. Considérons par exemple que certains individus aient la chance de recevoir une offre d'emploi avantageuse sur un site donné et décident de migrer vers ce site. L'effet sur le salaire d'être localisé sur le site captera alors non seulement l'effet de facteurs locaux exogènes, mais aussi l'effet de sélection (i.e. le bonus de salaire des travailleurs ayant eu la chance de pouvoir bénéficier d'une opportunité d'emploi sur le site).

Traditionnellement, les études sur le lien entre migrations et salaires utilisent des données en coupe pour estimer un modèle de Roy à trois équations (cf. Robinson et Tomes, 1982). Le modèle contient une équation de salaire en cas de migration, une équation de salaire en l'absence de migration, et une spécification binaire de la décision de migrer dépendant de la différence de salaires en cas de migration et en l'absence de migration (ainsi que d'autres facteurs non liés aux salaires). Il est important de noter que le salaire de migration n'est observé que pour les migrants, et le salaire en l'absence de migration n'est observé que pour les non migrants. L'enjeu est généralement d'évaluer si le rendement de variables, tel que le diplôme ou l'expérience sur le marché du travail, est plus important en cas de migration qu'en l'absence de migration.
Ce type d'études a deux limites principales. La première est de ne pas vraiment modéliser le choix de localisation. En effet, les individus choisissent de migrer ou non, mais ne choisissent pas de destination en cas de migration. Par ailleurs, les estimations sont généralement en coupe et ne permettent pas d'examiner les rendements des caractéristiques inobservables dont l'estimation de l'effet nécessite des données longitudinales.

Une généralisation du modèle de Roy à une équation de salaire par localisation est proposée par Dahl (2002). L'objet d'intérêt est la variation des rendements des diplômes entre Etats américains. Une équation de salaire par Etat est donc spécifiée ainsi qu'une équation de choix de localisation. Il est considéré que les travailleurs choisissent l'Etat leur permettant d'atteindre le niveau d'utilité le plus élevé, l'utilité associée à un Etat dépendant du salaire. Le modèle peut être estimé en deux étapes avec une méthode de correction du biais de sélection pour les équations de salaire proche de la méthode d'Heckman. En effet, il est montré que la probabilité pour un travailleur de choisir l'Etat où il habite effectivement est une statistique généralement suffisante pour calculer le terme corrigeant l'effet de sélection dans les équations de salaire. L'auteur propose de calculer non paramétriquement la probabilité de choisir chaque Etat, d'évaluer ensuite le terme correctif de chaque équation de salaire, et d'estimer enfin chaque équation de salaire. Les résultats obtenus montrent que les rendements du diplôme sont assez peu biaisés par les problèmes de sélection.

Dans Gobillon et Le Blanc (2003), j'ai plutôt essayé de prendre en compte la dimension panel dans le modèle classique de migration. L'objectif était de préciser le rôle des caractéristiques inobservables dans la décision de migration et les salaires.

La spécification du modèle est donnée par :

$$M_{it} = 1_{\{M_{it}^* > 0\}} \text{ où } M_{it}^* = \ln Y_{it}^m - \ln Y_{it}^r - C_{it} \tag{20}$$

$$\ln Y_{it}^m = X_{it}\beta + u_i^m + \varepsilon_{it}^m \tag{21}$$

$$\ln Y_{it}^r = X_{it}(\beta + \gamma) + u_i^r + \varepsilon_{it}^r \tag{22}$$

$$C_{it} = Z_{it}\delta + u_i + \varepsilon_{it} \tag{23}$$

L'indicatrice de migration M_{it} suit un modèle à variable latente M_{it}^* dont la valeur dépend de la différence entre le revenu de migration Y_{it}^m et le revenu en l'absence de migration Y_{it}^r, ainsi que des coûts migratoires C_{it}. Les deux types de revenus ainsi que les coûts sont expliqués par des variables explicatives (X_{it} ou Z_{it}) et des résidus. Dans chaque équation, les résidus sont décomposés en l'effet des caractéristiques individuelles inobservables (u_i^m, u_i^r ou u_i) et des chocs qui peuvent correspondre à des opportunités d'emploi ou des coûts inattendus (ε_{it}^m, ε_{it}^r ou ε_{it}). Les corrélations entre caractéristiques individuelles inobservables sont permises. En revanche, les chocs sont supposés indépendants et identiquement distribués. La variance du choc de coût est fixée à un pour que le modèle soit identifié. Tous les chocs sont supposés être normaux.

Cette spécification permet d'évaluer si les caractéristiques observables ont des effets différents en cas de migration et en l'absence de migration (en testant si $\gamma = 0$). Elle permet aussi de tester si les caractéristiques inobservables ont le même rendement (en évaluant les variances et la corrélation de leurs effets). Enfin, elle permet de décomposer le terme d'Heckman prenant en compte la sélection des individus dans les équations de salaire en une partie liée aux caractéristiques individuelles inobservables et une partie liée aux chocs. Le modèle est identifié grâce en particulier aux migrations répétées de certains individus. Il est estimé en maximisant la vraisemblance qui est en partie simulée.

Les données utilisées pour les estimations sont celles du Panel Européen des Ménages pour la période 1994-2000. Ce panel a la particularité de suivre les individus lorsqu'ils déménagent. Sur la période étudiée, nous avons pu avoir accès au code commune du lieu de résidence des individus. Il nous a permis de déterminer si les individus avaient effectué des migrations

inter-commmunales. Nous restreignons notre étude aux chefs de ménages actifs de plus de vingt-cinq ans.

Une question rétrospective permet de connaître les motivations à déménager des ménages mobiles. Alors que 32% d'entre eux déclarent avoir déménagé pour une raison liée à l'emploi, ils sont 40% à déclarer une raison liée au logement et 25% à déclarer une autre raison (liée à l'environnement ou à la famille en particulier). Les déterminants des migrations inter-communales ne sont donc pas seulement liés à l'emploi. Il serait certainement souhaitable de ne prendre en compte que les migrations à moyenne ou longue distance pour lesquelles la motivation à migrer est plus souvent liée à l'emploi, mais la taille de l'échantillon ne permet pas de s'y limiter sous peine d'avoir trop peu de migrations répétées pour identifier le modèle. Cependant, les raisons autres que celles liées à l'emploi sont indirectement prises en compte dans les variables explicatives des coûts de déménagement. Enfin, nous utilisons les revenus des ménages et non les salaires pour l'estimation du modèle car ils font plus sens lorsque les ménages sont composés de plus d'un membre qui travaille.

Les résultats montrent que les variables explicatives ont des rendements assez différents en cas de migration et en l'absence de migration. Par exemple, les revenus des plus âgés sont plus faibles en cas de migration ainsi que pour les chefs de ménages nés à l'étranger, ceux ayant des enfants, ceux avec un diplôme technique peu élevé, ceux étant au chômage et les femmes vivant seules. En revanche, nous trouvons que les effets individuels inobservables des équations de revenu ont une corrélation très proche de un et ont des variances qui ne sont pas significativement différentes. Ainsi, les caractéristiques individuelles inobservables auraient dans notre cas des rendements similaires.

Par ailleurs, même si la variance des résidus dans les équations de salaire est principalement due aux caractéristiques individuelles inobservables (à hauteur d'environ 70%), le biais de sélection provient principalement des chocs. Une explication de ce phénomène est que la décision de migration est expliquée par la différence de revenus en cas de migrations et en l'absence de migration. La différence des effets individuels inobservables est proche de zéro alors que celle de chocs ne l'est pas.

En résumé, même si les caractéristiques individuelles inobservables ont un impact sur le niveau des salaires, elles ne sont pas à l'origine des effets de sélection dans notre contexte. Ce sont plutôt les chocs comme les opportunités d'emploi à distance du lieu de résidence ou même un licenciement qui créent les effets de sélection les plus importants.

Une limite de notre travail est que le modèle n'est pas dynamique alors qu'on peut s'attendre à ce que les migrants anticipent les retombées d'une migration sur le moyen ou long terme. Par ailleurs, le choix d'une localisation n'est pas pris en compte. Gould (2007) propose une approche apportant une solution au moins partielle à ces problèmes. Il considère un modèle dynamique dans lequel les individus sont classés à chaque date dans des catégories définies en croisant le lieu de résidence (ville ou campagne) et le type d'activité (étudiant, inactif, travailleur peu qualifié, travailleur qualifié). L'hétérogénéité individuelle inobservée est prise en compte en considérant que les individus peuvent être de trois types inobservables et en modélisant les probabilités d'appartenir à ces trois types. Les estimations montrent que les travailleurs qualifiés gagnent plus dans les villes que dans les campagnes, mais que ce n'est pas le cas des travailleurs peu qualifiés. En outre, pour les travailleurs qualifiés habitant à la campagne, l'expérience professionnelle acquise en ville est plus valorisée que l'expérience acquise en milieu rural. Ce n'est pas le cas pour les travailleurs peu qualifiés habitant à la campagne.

Il serait intéressant de généraliser ce type d'approche à un contexte où chaque site constitue une localisation pour pouvoir étudier les disparités spatiales sur le marché du travail sans être trop restrictif sur les effets spatiaux.

IV. Pistes de recherche

IV.I. Améliorer l'analyse des disparités spatiales de salaire

L'un des challenges les plus importants de la littérature sur les disparités spatiales est certainement de réussir à prendre en compte l'endogénéité du choix de localisation des individus et des établissements. L'idéal serait de disposer d'une expérience contrôlée dans laquelle un sous-échantillon aléatoire d'individus serait alloué aléatoirement dans l'espace. Ce type d'expérience est généralement impossible à implémenter en pratique. Plusieurs pistes alternatives de recherche sont envisageables.

Une première piste consiste à trouver une sous-population spécifique dans un contexte particulier pour laquelle le choix de localisation est exogène. On peut par exemple s'intéresser à la localisation des femmes de militaires assignés par l'armée sur des sites géographique spécifiques. Aslund, Osth et Zenou (2009) étudient quant à eux les réfugiés politiques en Suède car ils sont alloués par le gouvernement en différents lieu du territoire pour des raisons qui ne sont pas liées à l'emploi. Une limite de ce type d'approche est que les résultats obtenus, mêmes s'ils évitent les biais d'endogénéité liés aux choix de localisation, ne concernent souvent qu'une petite sous-population et peuvent difficilement être généralisés.

Une autre piste pour étudier les disparités spatiales de salaires est de s'inspirer de l'approche de Dahl (2002) pour corriger le biais de sélection dans un contexte de panel. Ce type d'approche est cependant assez lourde quand des effets fixes individuels sont inclus dans l'équation de salaire (où les équations de salaire, si une équation est spécifiée par localisation). Il est possible de réduire la dimension du problème en considérant que les individus ne peuvent appartenir qu'à un nombre limité de catégories d'après leurs caractéristiques inobservées. En outre, il est nécessaire de disposer de conditions d'exclusion pour que le modèle soit identifié, une variable au moins devant se trouver dans la spécification du choix de localisation et non dans la ou les équations de salaire. La variable exclue est souvent le lieu de naissance, mais ce type de condition reste discutable. En effet, un individu peut être à même de bénéficier plus facilement d'opportunités d'emploi sur son lieu de naissance à cause de son réseau social ou parce qu'il dispose de plus d'informations. Ces opportunités d'emploi peuvent conduire à un salaire plus élevé.

Une autre limite des travaux sur les disparités spatiales de salaire est que les effets individuels inobservés sont supposés ne pas dépendre du temps. Il est cependant possible que la productivité des travailleurs évolue différemment dans les villes et à la campagne. En particulier, il est probable que les travailleurs des secteurs nécessitant des connaissances de pointe s'améliorent plus rapidement dans les grandes villes au contact des personnes disposant de ces connaissances. Dans les équations de salaire utilisées pour étudier les disparités spatiales de productivité, on peut donc vouloir préciser les effets d'apprentissage en supposant que les effets fixes individuels peuvent varier avec la durée passée dans les différents types de localisation.

IV.2. Compléter le modèle structurel sur la productivité des entreprises

Par ailleurs, des effets de sélection supplémentaires peuvent apparaître dans le cadre du modèle structurel comparant les distributions de productivité des grandes agglomérations et des petites agglomérations. En effet, la sélection des entreprises sur le marché ne se limite pas à la disparition des entreprises les moins productives. Il peut aussi exister *ex ante* un tri spatial des entrepreneurs entre grandes agglomérations et petites agglomérations. Il devient dès lors problématique de considérer que les distributions de productivités dans les deux catégories d'agglomérations sont *ex ante* mêmes, même si cette hypothèse restrictive est souvent à la base des modèles théoriques incluant une hétérogénéité des entreprises (une exception récente étant Behrens, Duranton et Robert-Nicoud, 2009). Elle est aussi à la base du modèle structurel que nous avons estimé. Il conviendrait de trouver un moyen de justifier empiriquement cette hypothèse ou de la relâcher. En particulier, une analyse de la localisation des entrepreneurs par rapport à leur lieu de naissance devrait pouvoir donner une première idée du tri spatial des entrepreneurs.

Une autre limite du modèle structurel est que l'hétérogénéité de la main-d'œuvre en terme de productivité est imparfaitement prise en compte. En effet, les régressions permettant d'estimer la productivité des entreprises ne prennent en compte les variations de caractéristiques de la main-d'œuvre que par l'entremise des proportions des travailleurs en trois classes de qualification. Une approche alternative, appliquée dans des travaux en cours, consiste à calculer la productivité des entreprises à partir d'une équation de salaire estimée au niveau des

travailleurs. L'hétérogénéité individuelle est prise en compte en introduisant des effets fixes individuels dans l'équation de salaire. Une mesure de la productivité des entreprises nette de l'effet des facteurs individuels peut alors être construite.

Par ailleurs le modèle structurel est statique. Il ne conduit pas à des prédictions claires sur les mouvements d'entrée-sortie des entreprises que l'on observe au cours du temps dans la réalité. Le modèle pourrait être complété en précisant la dynamique locale du marché du travail dans les agglomérations. Une telle approche devrait permettre de mieux comprendre comment la distribution locale des productivités des entreprises évolue au cours du temps en lien avec la démographie d'entreprises. Empiriquement cette évolution peut être précisée en comparant les positions des entreprises entrantes et sortantes dans les distributions locales de productivités.

IV.3. Evaluer l'effet du mode de financement des hôpitaux sur les soins

L'étude des disparités régionales de mortalité par infarctus aigu du myocarde a montré qu'une grande partie de ces disparités est due à des variations spatiales de traitement par des actes coûteux. En fait, ces actes sont plus souvent effectués par des hôpitaux privés à but lucratif que par des hôpitaux publics. Ce contraste peut s'expliquer par une différence de mode de financement. Jusqu'en 2003, les hôpitaux publics sont financés par enveloppe budgétaire. Ils doivent donc faire des arbitrages entre les procédures utilisées et limiter dans une certaine mesure le recours aux procédures coûteuses. Les hôpitaux privés à but lucratif bénéficient quant à eux d'un remboursement à l'acte auquel participent les patients. Comme les patients des hôpitaux privés à but lucratifs sont généralement assez aisés, ces hôpitaux peuvent être enclins à utiliser les procédures de pointe plus souvent que les hôpitaux publics.

Des changements dans les politiques de financement ont été implémentés dès 2004. En effet, après cette date, les hôpitaux publics sont peu à peu remboursés sur la base de groupes homogènes de séjours (GHS). Les séjours des patients sont regroupés dans les GHS selon leurs similitudes médicales en termes de prise en charge. Il existe des règles supplémentaires qui stipulent que certaines procédures particulières peuvent être remboursées à 100%. Ces règles ont un intérêt particulier dans le cas de l'infarctus aigu du myocarde, puisque le stent[2]

[2] Le stent est un consommable médical en forme de ressort que l'on introduit dans l'artère ou la veine à l'endroit où il y a eu le blocage menant à l'infarctus. Il permet de garder l'artère ou la veine dilatée.

entre dans la catégorie des procédures pleinement remboursées. Ainsi, les hôpitaux publics devraient avec un meilleur accès à des traitements de pointe dès 2004. Les changements de mode de financement dans le privé n'interviennent qu'en 2005. On peut donc s'attendre à ce que les différences de traitements se resserrent entre les hôpitaux du public et les hôpitaux du privé à but lucratif en 2004.

Il paraît donc intéressant d'étudier l'évolution de la différence de traitements entre hôpitaux publics et hôpitaux privés à but lucratif durant la période 2003-2004. Pour chaque hôpital, un indicateur de contrainte d'utilisation des stents peut être construit comme la différence de proportions de stents administrés dans les six derniers mois de l'année et les six premiers mois de l'année. Si un hôpital est contraint, on s'attend à ce que ses médecins administrent plus de stents en début d'année quand des consommables sont disponibles, qu'en fin d'année quand il ne reste plus de consommables en stock. Il est alors possible d'examiner dans quelle mesure les hôpitaux publics sont moins contraints en 2004 qu'en 2003, et comparer l'évolution de leurs contraintes avec celle des hôpitaux privés à but lucratif. On peut aussi examiner s'il en a résulté une diminution des disparités régionales de mortalité par crise cardiaque.

IV.4. Approfondir l'étude des difficultés d'accès à l'emploi des femmes

Les travaux qui ont été entrepris pour mesurer les problèmes d'accès des femmes aux emplois les mieux rémunérés reposent sur l'utilisation du salaire comme index de la position d'un poste dans la hiérarchie des emplois. Cette mesure est imparfaite car le salaire peut être influencé par la productivité individuelle. Une approche alternative consisterait à utiliser le système de grades de la fonction publique ou d'une grande entreprise pour caractériser la position réelle d'un poste dans la hiérarchie des emplois. Les données contenant ce type de grade sont difficiles à obtenir.

Une grande entreprise du CAC 40 a cependant accepté de donner accès à son fichier de paie en panel sur la période 2007-2009 dans le cadre d'une convention. Une extension des travaux déjà entrepris aura pour objectif de quantifier les différences d'accès à l'emploi entre hommes et femmes au sein d'un grade donné, et entre grades. En effet, pour un grade donné, il existe plusieurs types d'emploi qui sont rémunérés différemment. Les femmes peuvent ne pas avoir accès à certains de ces emplois. De plus, les femmes peuvent avoir un moins bon accès à certains grades, en particulier ceux correspondant aux emplois les mieux rémunérés.

Par ailleurs, une limite du modèle d'assignement d'emplois utilisé pour mesurer les différences d'accès entre hommes et femmes est sa nature statique. Il serait intéressant de préciser la dynamique de promotions grâce à la dimension panel des données de la grande entreprise du CAC 40. On peut ainsi examiner les différences de promotions des hommes et des femmes au sein d'un grade donné, et notamment les différences de promotions aux emplois servant de tremplin pour accéder au grade suivant. On peut aussi examiner les différences de promotions de grades.

Bibliographie

Abowd J., Kramarz F. et D. Margolis (1999), “High wage workers and high wage firms”, *Econometrica*, 67(2), pp. 251–333.

Albrecht J., Björklund A. et S. Vroman (2003), “Is There a Glass Ceiling in Sweden?”, *Journal of Labour Economics*, 21(1), pp. 145-177.

Angelini, V. et A. Laferrère (2008), “Home, houses and residential mobility”, in Börsch-Supan A. et al. (eds.), Health, Ageing and Retirement in Europe (2004-2007). Starting the Longitudinal Dimension. Mannheim Research Institute for the Economics of Aging, pp. 99-107.

Arrow J.K. (1971), “The theory of discrimination”, in O. Aschenfelter & A. Rees (eds), 1974, Discrimination in Labor Markets, NJ: Princeton UP.

Aslunf O., Osth J. et Y. Zenou (2009), “How important is access to jobs? Old question – improved answer”, *Journal of Economic Geography*, à paraître.

Baker M. et A. Melino (2000), “Duration Dependence and Non-Parametric Heterogeneity: A Monte-Carlo Study”, *Journal of Econometrics*, 96, pp. 357-393.

Bartik T. (1991), *Who benefits from state and local economic development policies?*, Kalamazoo, Michigan: W. E. Upjohn Institute for Employment Research.

Becker G. (1971), *The Economics of Discrimination*, Second Edition, The University of Chicago Press, 178p.

Bonnet C., Gobillon L. et A. Laferrère (2009), “The effect of Being Widowed on Housing and Location Choices”, CREST Working Paper 2008-12, revise-and-resubmit au *Journal of Housing Economics*.

Behrens K., Duranton G. et F. Robert-Nicoud (2009), “Sorting, Selection and Agglomeration”, Document de Travail.

Breslow N. (1974), “Covariance Analysis of Censored Survival Data”, *Biometrics*, 30, pp. 89-99.

Burnod G. et A. Chenu (2001), “Employés qualifiés et non-qualifiés: Une proposition d’aménagement de la nomenclature des catégories socioprofessionnelles”, *Travail et Emploi*, 86, pp. 87–105.

Boutin X. et S. Quantin (2006), “Une méthodologie d’évaluation comptable du coût du capital des entreprises françaises : 1994-2002”, Working Paper G2006-09, INSEE DESE.

Calvo-Armengol A. et M. Jackson (2004), “Social networks in determining unemployment: Patterns, dynamics and inequality”, *American Economic Review*, 94, pp. 191-206.

Cambois E., Désesquelles A. et J.F. Ravaud (2003), “The gender disability gap”, *Population and Societies*, 386.

Carrasco M. et J.P. Florens (2000), “Generalization of GMM to a continuum of moment conditions”, *Econometric Theory*, 16(6), pp. 797–834.

Chen Y. et S. Rosenthal (2008), “Local amenities and life-cycle migration: Do people move for jobs or for fun?”, *Journal of Urban Economics*, 64, pp. 519-537.

Ciccone A. et R. Hall (1996), “Productivity and the density of economic activity”, *American Economic Review*, 86(1), pp. 54–70.

Combes P.Ph. (2000), “Economic structure and local growth: France, 1984–1993”, *Journal of Urban Economics*, 47(3), pp. 329–355.

Combes P.Ph., Duranton G. et L. Gobillon (2008a), “Spatial Wage Disparities: Sorting Matters!”, *Journal of Urban Economics*, 63(2), pp. 723-742.

Combes P.Ph., Duranton G. et L. Gobillon (2008b), “Le rôle des marchés locaux du travail dans la concentration spatiale des activités économiques”, *Revue de l'OFCE*, 104, pp. 141-177.

Combes P.Ph., Duranton G., Gobillon L. et S. Roux (2009), “Estimating agglomeration economies with history, geology and worker effects”, CEPR Working Paper 6728, forthcoming in *The Economics of Agglomeration*, Edward Glaeser (ed.,), NBER, Cambridge, MA.

Combes P.Ph., Duranton G., Gobillon L., Puga D. et S. Roux (2009), “The productivity advantages of large cities: Distinguishing agglomeration from firm selection”, CEPR Working Paper 7191, revise-and-resubmit à *Econometrica*.

Combes P.Ph., Magnac T. et J.M. Robin (2004), “The dynamics of local employment in France”, *Journal of Urban Economics*, 56, pp. 217-243.

Combes P.Ph. et H. Overman (2004), “The Spatial Distribution of Economic Activities in the European Union”, in Vernon Henderson and Jacques-François Thisse (eds.), Handbook of Regional and Urban Economics, volume 4. Amsterdam: North-Holland, pp. 2845-2910.

Dahl G. (2002), “Mobility and the return to education: Testing a Roy model with multiple markets”, *Econometrica*, 70(6), pp. 2367–2420.

De Coulon A. et F.C. Wolff (2006), “The location of immigrants at retirement: Stay/return or ‘va-et-vient’ ?”, IZA Discussion Paper, 2224.

Dejonghe V. et C. Gasnier (1990), “Pratiques salariales et gestion du personnel dans les banques et les assurances”, *Travail et Emploi*, 45(3), pp. 40-46.

Duranton G., Gobillon L. et H. Overman (2009), “Assessing the Effects of Local Taxation using Microgeographic Data”, CEPR Working Paper 5856, resoumis à *The Economic Journal*.

Duranton G., Martin P., Mayer T. et F. Mayneris (2008), “Les pôles de compétitivité : que peut-on en attendre ?”, CEPREMAP, Editions Rue d’Ulm.

Duranton G. et V. Monastiriotis (2002), “Mind the gaps: The evolution of regional earnings inequalities in the UK 1982–1997”, *Journal of Regional Science*, 42(2), pp. 219–256.

Epple D. et T. Nechyba (2004), “Fiscal decentralization”, in Vernon Henderson and Jacques-François Thisse (eds.), Handbook of Regional and Urban Economics, volume 4. Amsterdam: North-Holland, pp. 2423-2480.

Ermisch J. et S. Jenkins (1999), “Retirement and housing adjustment in later life: evidence from the British Household Panel Survey”, *Labour Economics*, 6, pp. 311-333.

Flipo A., Le Blanc D. et A. Laferrère (1999), “De l'histoire individuelle à la structure des ménages”, *Insee Première*, 649.

Fujita M. (1989), *Urban Economic Theory: Land Use and City Size*, Cambridge, Cambridge University Press.

Gaynor M. (2006), “What do we know about competition and quality in health care markets?”, NBER Working Paper 12301.

Glaeser E., Kallal H., Scheinkman J. et A. Schleifer (1992), “Growth in cities”, *Journal of Political Economy*, 100(6), pp. 1126–1152.

Glaeser E. et D. Maré (2001), “Cities and skills”, *Journal of Labor Economics*, 19(2), pp. 316–342.

Gobillon L. (2001), “Emploi, Logement et Mobilité Résidentielle”, *Economie et Statistique*, 349-350, pp. 77-98.

Gobillon L. et D. Le Blanc (2003), “Migrations, Income and Skills”, Document de travail CREST 2003-4.

Gobillon L. et D. Le Blanc (2004), “L’impact des contraintes d’emprunt sur la mobilité résidentielle et les choix de statut d’occupation des ménages : un modèle simple de demande”, *Annales d’Economie et de Statistiques*, 74, pp. 14-46.

Gobillon L. et D. Le Blanc (2008), “Economic Effects of Upfront Subsidies to Ownership: the case of the Prêt à Taux Zéro in France”, *Journal of Housing Economics*, 17(1), pp. 1-33.

Gobillon L., Magnac T. et H. Selod (2009a), “The effect of location on finding a job in the Paris region”, CEPR Working Paper 6199, *Journal of Applied Econometrics*, à paraître.

Gobillon L., Magnac T. et H. Selod (2009b), “Spatial Disparities in the Unemployment Duration of Disadvantaged Groups”, Document de Travail.

Gobillon L., Magnac T. et H. Selod (2009c), “Do Unemployed Workers Benefit from Enterprise Zones: the French Experience”, Document de Travail.

Gobillon L., Meurs D. et S. Roux (2009), “Estimating gender differences in access to jobs: Females trapped at the bottom of the ladder”, Document de travail PSE 2009-36.

Gobillon L. et C. Milcent (2008), “Regional Disparities in Mortality by Heart Attack: evidence from France”, PSE Working Papers 2008-12.

Gobillon L. et S. Roux (2009), “Quantile-based inference of parametric transformations between two distributions”, Document de Travail.

Gobillon L., Selod H. et Y. Zenou (2007), “The mechanisms of spatial Mismatch”, *Urban Studies*, 44(12), pp. 2401-2427.

Gobillon L. et F.C. Wolff (2009), “Housing and location choices of retiring households: Theory and evidence”, Document de Travail INED, 162.

Gould E. (2007), “Cities, Workers and Wages: A structural Analysis of the Urban Wage Premium”, *Review of Economic Studies*, 74(2), pp. 477-506.

Graves P. et T. Knapp (1988), “Mobility behavior of the elderly”, *Journal of Urban Economics*, 24, pp. 1-8.

Guimaraes P., Figueiredo O. et D. Woodward (2004), “Industrial location modeling: Extending the random utility framework”, *Journal of Regional Science*, 44(1), pp. 1-20.

Helsley R. et W. Strange (1990), “Matching and Agglomeration Economies in a System of Cities”, *Regional Science and Urban Economics*, 20, pp. 189-212.

Head K. et T. Mayer (2004), “The Empirics of Agglomeration”, in Vernon Henderson and Jacques-François Thisse (eds.), Handbook of Regional and Urban Economics, volume 4. Amsterdam: North-Holland, pp. 2609-2669.

Henderson J., Kuncoro A. et M. Turner (1995), “Industrial Development in Cities”, *Journal of Political Economy*, 103, pp. 1067-1090.

Ihlanfeldt K. (1997), “Information on the spatial distribution of job opportunities within Metropolitan Areas”, *Journal of Urban Economics*, 41, pp. 218-242.

Ioannides Y et K. Kan (1996), “Structural Estimation of Residential Mobility and Housing Tenure Choice”, *Journal of Regional Science*, 36(3), pp. 335-363.

Kessler D. et M. McClellan (2000), “Is hospital competition socially wasteful?”, *Quaterly Journal of Economics*, 115(2), pp. 577-615.

Krugman P. et A. Venables (1995), “Globalization and The Inequality of Nations”, *Quarterly Journal of Economics*, 110, pp. 857-880.

Lacroix T. (1995), “Le recul de l'accession sociale”, *Economie et Statistique*, 288-289, pp. 11-41.

Laferrère A. (2005), “Old age and housing: dissaving, adjusting consumption, and the role of children”, Working Paper.

Levinsohn J. et A. Petrin (2003), “Estimating production functions using inputs to control for unobservables”, *Review of Economic Studies*, 70(2), pp. 317–342.

Lucas R. (1988), “On the mechanics of economic development”, *Journal of Monetary Economics*, 22(1), pp. 3–42.

Lucas R. et E. Rossi-Hansberg (2002), “On the internal structure of cities”, *Econometrica*, 70(4), pp. 1445–1476.

McClellan M. et D.O. Staiger (2000), “Comparing Hospital Quality at For-Profit and Not-for-Profit Hospitals”, in D.M. Cutler (ed.), The Changing Hospital Industry: Comparing Not-for-Profit and For-Profit Institutions, University of Chicago Press.

Melitz M. (2003), “The impact of trade on intraindustry reallocations and aggregate industry productivity”, *Econometrica*, 71(6), pp. 1695–1725.

Melitz M. et G. Ottaviano (2008), “Market size, trade and productivity”. *Review of Economic Studies*, 75(1), pp. 295–316.

Milcent C. (2005), “Hospital Ownership, Reimbursement Systems and Mortality rates”, *Health Economics*, 14, pp. 1151.1168.

Milcent C., Dormont B., Durand-Zaleski I. et P. Steg (2007), “Gender Differences in Hospital Mortality and Use of Percutaneous Coronary Intervention in Acute Myocardial Infarction”, Circulation, 115(7), pp. 823-826.

Moulton B. (1990), “An illustration of a pitfall in estimating the effects of aggregate variables on micro units”, *Review of Economics and Statistics*, 72(2), pp. 334–338.

Olley G. et A. Pakes (1996), “The dynamics of productivity in the telecommunication equipment industry”, *Econometrica*, 64(6), pp. 1263–1297.

Ortalo-Magné F. et S. Rady (2006), “Housing Market Dynamics: On the Contribution of Income Shocks and Credit Constraints”, *Review of Economic Studies*, 73(2), pp. 459-485.

Phelps E. (1972), “The statistical theory of racism and sexism”, *American Economic Review*, 62, pp. 639-651.

Pissarides C. (2000), *Equilibrium Unemployment Thoery*, MIT Press, 226p.

Rathelot R. et P. Sillard (2009), “Zones Franches Urbaines : quels effets sur l’emploi salarié et les créations d’établissements ?”, *Economie et Statistique*, 415-416, pp. 81-96.

Ridder G. et I. Tunali (1999), “Stratified partial likelihood estimation”, *Journal of Econometrics*, 92(2), pp. 193-232.

Roback J. (1982), “Wages, rents and the quality of life”, *Journal of Political Economy*, 90(6), pp. 1257– 1278.

Robison C. et N. Tomes (1982), “Self-Selection and Interprovincial Migration in Canada”, *The Canadian Journal of Economics*, 15(3), pp. 474-502.

Rosen H. (1985), “Housing Subsidies”, in Auerbach A, Feldstein M (Eds), Handbook of Public Economics, volume 1. North Holland: Amsterdam, pp. 375-420.

Rosenthal S. et W. Strange (2004), “Evidence on the nature and sources of agglomeration economies”, in Vernon Henderson and Jacques-François Thisse (eds.), Handbook of Regional and Urban Economics, volume 4. Amsterdam: North-Holland, pp. 2119–2171.

Sattinger M. (1993), “Models of the Distribution of Earnings”, *Journal of Economic Literature*, 31(2), pp. 831-880.

Smith T. et Y. Zenou (2003), “Spatial mismatch, search effort and urban spatial structure”, *Journal of Urban Economics*, 54, pp. 129-156.

Stoll M. (1999), “Spatial job search, spatial mismatch, and the employment and wages of racial and ethnic groups in Los Angeles”, *Journal of Urban Economics*, 46, pp. 129-55.

Temple J. (1999), “The new growth evidence”, *Journal of Economic Literature*, 37(1), pp. 112–156.

Thomas M. et J. Grillon (2001), “Prêt à 0% : un bilan après cinq ans d'existence”, ANIL.

Turner S. (1997), “Barriers to a better break: employer discrimination and spatial mismatch in metropolitan Detroit”, *Journal of Urban Affairs*, 19, pp. 123–141.

Venti S.F. et D.A. Wise (1990), “But they don't want to reduce housing equity”, in D.A. Wise (ed.), Issues in the Economics of Aging. University of Chicago Press, Chicago, pp. 13-29.

Venti S.F. et D.A. Wise (2001), *Aging and Housing Equity: Another Look*, NBER Working Paper, 8608.

Wilson J. (1996), *When Work Disappears: The World of the New Urban Poor*, New York: Alfred A. Knopf.

Zax J. et J. Kain (1996), “Moving to the suburbs: Do relocating companies leave their Black employees behind?”, *Journal of Labor Economics*, 14, pp. 472-504.

Zenou Y. (2002), “How do firms redline workers?”, *Journal of Public Economics*, 52(3), pp. 391-408.

Zorn P. (1989), “Mobility-Tenure Decisions and Financial Credit: Do Mortgage Qualification Requirements Constrain Homeownership?”, *AREUA Journal*, 17(1), pp. 1-16.

Annexe A. Curriculum vitae

Laurent GOBILLON

Né le 01/01/75 à Colmar (35 ans)
Célibataire
Institut National d'Etudes Démographiques (INED)
133 Boulevard Davout
75980 Paris Cedex 20
☏ : + 33 (0) 1 56 06 20 16, Fax : + 33 (0) 56 06 21 99
Email : laurent.gobillon@ined.fr
Page web : http://laurent.gobillon.free.fr/

FORMATION

2003-2004 : **London School of Economics** (LSE), études post-doctorales.

2003-2004 : **University College of London** (UCL), études post-doctorales.

1999-2002 : **Ecole des Hautes Etudes en Sciences Sociales** (EHESS), Doctorat en « Analyse et Politique Economiques », sous la Direction de Jean-Marc Robin.
Titre : « Mobilité Résidentielle et Marchés Locaux de l'Emploi ».
Laboratoire de rattachement : Laboratoire de Microéconométrie du **CREST**.
Soutenance le 20 décembre 2002 à Paris.
Mention très honorable avec félicitations du jury.

Proposition pour prix de thèse et publication.

Jury : Brigitte Dormont (rapporteur), Lionel Fontagné, Marc Ivaldi, Philippe Martin (rapporteur) et Jean-Marc Robin.

1996-1999 : **ENSAE** (Ecole Nationale de la Statistique et de l'Administration Economique), diplômé en 1999.

1998-1999 : **DEA d'Analyse et Politique Economiques** (cohabilité par l'Ecole Polytechnique, l'EHESS, l'ENS et l'ENSAE), diplômé, mention bien.

1993-1996 : Math sup, math spé M' au Lycée Schweitzer de Mulhouse (Haut-Rhin).

1993 : Baccalauréat section C, diplômé, mention bien.

STATUT

2010- : **Institut National d'Etudes Démographiques**, co-responsable de l'unité 09, « Démographie Economique ».

2004- : **Institut National d'Etudes Démographiques**, chargé de recherche (2ème classe jusqu'en 2008, puis 1ère classe).

2009- : **Centre for Economic Policy Research** (CEPR), membre (*research affiliate*).

2008- : **Centre de Recherches en Economie et Statistiques** (CREST), chercheur associé.

2007- : **Laboratoire d'Economie Appliquée** (INRA-LEA), chercheur associé.

2003-2004 : **London School of Economics**, étudiant post-doctoral. Financement du *Economic and Social Research Council* (ESRC). Mi-temps d'Octobre 2003 à Juillet 2004.

2003-2004 : **University College of London**, étudiant post-doctoral. Bourse *Research Training Mobility* (RTN) de la Commission Européenne. Temps plein de Janvier à Septembre 2003. Mi-temps d'Octobre 2003 à Juillet 2004.

1999-2002 : **CREST**, élève chercheur. Allocation de Recherche CREST.

CENTRES D'INTERET

Economie Géographique (migrations, marchés locaux de l'emploi)
Economie Urbaine (*Spatial Mismatch*, ségrégation)
Economie du Logement (mobilité résidentielle, choix de statut, contraintes d'emprunt)
Economie de la Santé (disparités spatiales du système de soins)
Economie du Travail (inégalités hommes-femmes)
Econométrie (modèles multi-niveaux)

PUBLICATIONS DANS DES JOURNAUX INTERNATIONAUX A COMITE DE LECTURE

Gobillon L., Magnac T. et H. Selod (2010), « The effect of location on finding a job in the Paris region », à paraître dans le *Journal of Applied Econometrics*.

Combes P.Ph., Duranton G. et L. Gobillon (2008), « Spatial Wage Disparities: Sorting Matters! », *The Journal of Urban Economics*, 63(2), p. 723-742.

Gobillon L. et D. Le Blanc (2008), « Economic Effects of Upfront Subsidies to Ownership: the case of the Prêt à Taux Zéro in France », *Journal of Housing Economics*, 17(1), p. 1-33.

Gobillon L., Selod H. et Y. Zenou (2007), « The mechanisms of Spatial Mismatch », *Urban Studies*, 44(12), pp. 2401-2427.

CHAPITRES DE LIVRE EN ANGLAIS

Combes P.Ph, Duranton G., Gobillon L. et S. Roux (2009), « Estimating agglomeration economies with history, geology and worker effects », à paraître dans *The Economics of Agglomeration*, Edward Glaeser (ed.), University of Chicago Press.

DOCUMENTS DE TRAVAIL EN ANGLAIS

Combes P.Ph., Duranton G., Gobillon L., Puga D. et S. Roux (2009), « The productivity advantages of large cities: Distinguishing agglomeration from firm selection », CEPR Working Paper 7191, revise-and-resubmit à *Econometrica*.

Duranton G., Gobillon L. et H. Overman (2009), « Assessing the Effects of Local Taxation using Microgeographic Data », CEPR Working Paper 5856, revise-and-resubmit à l'*Economic Journal*.

Bonnet C., Gobillon L. et A. Laferrère (2009), « The effect of Being Widowed on Housing and Location Choices », Document de Travail CREST 2008-12, revise-and-resubmit au *Journal of Housing Economics*.

Gobillon L. et F.C. Wolff (2009), « Housing and location choices of retiring households: Theory and evidence », Document de Travail INED 162, revise-and-resubmit à *Urban Studies*.

Gobillon L., Meurs D. et S. Roux (2009), « Estimating Gender Differences in Access to Jobs: Females Trapped at the Bottom of the Ladder », CEPR Working Paper 7475.

Gobillon L. et C. Milcent (2009), « Spatial Disparities in Hospital Performances », Document de Travail PSE 2008-12 (version sous un autre titre).

Gobillon L. et H. Selod (2007), « The effect of segregation and spatial mismatch on unemployment: evidence from France », CEPR Working Paper 6198.

Gobillon L. (2004), « The Estimation of Cluster Effects in Linear Panel Models », Document de travail.

Gobillon L. et D. Le Blanc (2003), « Migrations, Income and Skills », Document de travail CREST 2003-47.

PUBLICATIONS DANS DES JOURNAUX FRANÇAIS A COMITE DE LECTURE

Combes PPh., Duranton G. et L. Gobillon (2008), « Le rôle des marchés locaux du travail dans la concentration spatiale des activités économiques », *Revue de l'OFCE*, 104, p. 141-177.

Gobillon L. et H. Selod (2007), « Les déterminants locaux du chômage en région parisienne », *Economie et Prévision*, 180-181, p 19-38.

Gobillon L. et A. Laferrère (2006), « Les choix de logement des personnes âgées : consommation et épargne », *Revue Française d'Economie*, 20(3), pp. 115-161.

Gobillon L. et D. Le Blanc (2005), « Quelques effets économiques du prêt à taux zéro », *Economie et Statistique*, 381-382, pp. 63-89.

Gobillon L. et D. Le Blanc (2004) « L'impact des contraintes d'emprunt sur la mobilité résidentielle et les choix de statut d'occupation des ménages : un modèle simple de demande », *Annales d'Economie et de Statistiques*, 74, pp. 14-46.

Gobillon L. (2001), « Emploi, Logement et Mobilité Résidentielle », *Economie et Statistique*, 349-350, pp. 77-98.

DOCUMENTS DE TRAVAIL EN FRANCAIS

Gobillon L. (2008), « Une Synthèse de la Littérature sur la Consommation de Logement des Ménages », Document de travail.

ARTICLES DE VALORISATION DE LA RECHERCHE

Bonnet C., Gobillon L. et A. Laferrère (2007), « Un changement de logement suite au décès du conjoint ? », *Gérontologie et Société*, 121, pp. 195-210.

Bonnet C. et L. Gobillon (2007), « Choix de logement et mobilité résidentielle suite au décès du conjoint », Chapitre d'ouvrage in *Vieillissement de la Population et Logement*, sous la direction de Catherine Bonvalet, Férial Drosso, Francine Benguigui et Phuong Mai Huynh, 508p, La Documentation Française.

Gobillon L. et A. Laferrère (2007), « Les choix de logement des personnes âgées : consommation et épargne », Chapitre d'ouvrage in *Vieillissement de la Population et Logement*, sous la direction de Catherine Bonvalet, Férial Drosso, Francine Benguigui et Phuong Mai Huynh, 508p, La Documentation Française.

Gobillon L. et H. Selod (2005), « Chômage et Ségrégation Urbaine », *Lettre du CREST n°62*.

Gobillon L. (2005), recension de l'ouvrage Kazepov Y. (ed), *Cities of Europe*, Blackwell Publishing, 2005, *Population*, 60(1-2), pp. 189-191.

Combes P.Ph., Duranton G. et L. Gobillon (2003), « Externalités d'agglomération et inégalités salariales entre zones d'emploi », complément C du rapport du *Conseil d'Analyse Economique (CAE)* intitulé « Compétitivité », pp. 163-183.

Gobillon L. et H. Selod (2003), « Les déterminants spatiaux du chômage en Ile-de-France », complément B du rapport du *Conseil d'Analyse Economique (CAE)* intitulé « Ségrégation Urbaine et Intégration Sociale », pp. 171-187.

Gobillon L. et H. Selod (2000), « Ségrégation urbaine et accès à l'emploi », *Lettre du* CREST *n°35 et Variance n°18*.

ARTICLES DANS LA PRESSE

Gobillon L., Magnac T. et H. Selod (2009), « Discrimination ethnique et discrimination ubaine », *Libération*, 6 mai 2009.

PRESENTATIONS A DES SEMINAIRES ET COLLOQUES

(Articles présentés à l'origine sous différentes versions et avec différents titres)

« Do unemployed workers benefit from enterprise zones? The French Experience » (avec Harris Selod and Thierry Magnac) :

Septembre 2009 :	Conférence de la European Association of Labour Economists, Tallin (Estonie)
Août 2009 :	64th Meeting of the Econometric Society, Barcelone (Espagne)
Novembre 2008 :	55th Annual North American Meetings of the Regional Science Association, New-York (USA)

« Spatial Disparities in the Unemployment Duration of Disadvantaged Groups » (avec Harris Selod and Thierry Magnac) :

Janvier 2009 :	Deuxième conference du réseau RTN « Transnationality of Migrants », Louvain-la-Neuve (Belgique)

« Ségrégation résidentielle, accès à l'emploi et chômage » :

Janvier 2009 :	Communication à la demi-journée Ecole d'Economie de Paris – MEDDAT sur « Les mécanismes de la ségrégation urbain, résidentielle et scolaire », 19 Janvier 2009, La Défense (Paris)

« Estimating gender differences in access to jobs: Females trapped at the bottom of the ladder » :

Janvier 2009 :	INED, Unité 09 : « Démographie Economique », Paris

« Regional Disparities in the mortality rate by heart attack: evidence from France » :

Octobre 2009 :	XXVIe congrès international de la population, Marrakech (Maroc)
Mai 2008 :	XXIVe Journées de Microéconomie Appliquée, Saint-Denis La Réunion (France)
Avril 2008 :	INED, Unité 06 : « Mobilité, logement et entourage », Paris
Novembre 2007 :	54rd Annual North American Meetings of the Regional Science Association International, Savannah (USA)
Septembre 2007 :	16th European Workshop Econometrics and Health Economics, Bergen (Norvège)
Juin 2007 :	INRA, LEA : « Laboratoire d'Economie Appliquée », Paris
Mai 2007 :	INED, Unité 05 : « Mortalité, santé et épidémiologie de l'INED », Paris

« The effect of location on finding a job in the Paris region » :

Avril 2007 :	Séminaire TEMA de l'Université de Paris I
Novembre 2006 :	53th Annual North American Meetings of the RSAI, Toronto (Canada)
Novembre 2006 :	International Symposium on Statistical Analysis of Spatio-Temporal Data, Tokyo (Japon)
Novembre 2006 :	Urban Economics Workshop at the University of Tokyo (Japon)
Mai 2006 :	INED, Unité 09 : « Démographie Economique », Paris
Mai 2006 :	Séminaire d'économie spatiale du CORE, Louvain-la-Neuve (Belgique)

« Assessing the Effects of Local Taxation using Micro-geographic Data » :

Mars 2007 :	Séminaire d'Econometrie de l'Université de Toulouse
Juillet 2006 :	Conference: « Empirical Method for the Study of Economic Agglomeration », Kyoto (Japon)
Décembre 2004 :	Spatial Statistics Workshop, Université de Paris I

« Les choix de logement des personnes âgées : consommation et épargne » :

Juin 2005 :	Séminaire du Laboratoire d'Economie de l'Université de Nantes
Mai 2005 :	Lundis de l'INED démodynamique, Paris
Novembre 2004 :	Séminaire « Patrimoine Immobilier et Retraites » au PUCA, Paris
Septembre 2004 :	INED, Unité 09 : « Démographie économique », Paris
Septembre 2004 :	INED, Unité 06 : « Mobilité, territoire, habitat et sociabilité », Paris

« Spatial Wage Disparities: Sorting Matters! » :

Janvier 2004 :	Groupe de travail sur le travail et les inégalités, Paris-Jourdan
Janvier 2004 :	Journées Doctorales de l'ADRES, Paris
Octobre 2003 :	Seminar of Applied Microeconomics, University College London
Août 2003 :	58th Econometric Society European Meeting, Stockholm (Suède)
Avril 2003 :	Séminaire TEAM-EUREQua à l'Université de Paris I
Janvier 2003 :	CEPR Workshop : *The Economic Geography of Europe: Measurement, Testing and Policy Simulations*, Villars (Suisse)
Octobre 2002 :	Conseil d'Analyse Economique, pour le Groupe de Travail « Compétitivité », Paris

« Migrations, Income and Skills » :

Août 2003 :	18th European Economic Association Annual Congress, Stockholm (Suède)

« Economic Effects of Upfront Subsidies to Ownership: the case of the Prêt à Taux Zéro in France » :

Décembre 2003 :	Séminaire de recrutement au GREQAM, Marseille
Avril 2003 :	Séminaire de recrutement de l'Université de Cergy-Pontoise
Mars 2003 :	Lunch Metrics, IFS, Londres
Octobre 2002 :	Network of Excellence : *The Evaluation of Labor Market Policies*, Amsterdam (Pays-Bas)
Juin 2002 :	International Economics Field Seminar, London School of Economics
Mai 2002 :	7ème Colloque des Jeunes Economètres, Mont Saint-Odile
Mai 2002 :	Séminaire d'Economie Appliquée de l'INRA Toulouse
Février 2002 :	Séminaire de l'EPEE, Université d'Evry Val d'Essonne
Août 2001 :	16th European Economic Association Annual Congress, Lausanne (Suisse)
Juin 2001 :	18èmes Journées de Micro-Economie Appliquée, Nancy

« Le chômage des banlieues : problème d'accès à l'emploi ou effet de ségrégation ? » :

Avril 2002 :	Conseil d'Analyse Economique, pour le Groupe de Travail « Obstacles à l'Intégration Sociale et Moyens d'y Remédier », Paris

« Emploi, Logement et Mobilité Résidentielle » :

Avril 2001 :	Première Ecole de Printemps en Economie Géographique, Pau
Novembre 2000 :	Journées Doctorales de Paris
Avril 2000 :	Jamboree, Doctorat Européen, Bonn (Allemagne)

Discussions d'articles

Mai 2008 :	« How Gender Discrimination Can Explain Fertility Behaviours and Family-Firendly Policies ? » (Magali Recoules), Journées de Microéconomie Appliquées, Saint-Denis – Ile de La Réunion (France).
Novembre 2007 :	« Endettement et Activité Féminine : le Cas de la France » (Cédric Houdré), séminaire de la direction des statistiques démographiques et sociales (DSDS – INSEE).
Novembre 2007 :	« Moving to High Quality of Life » (Jordan Rappaport), 54rd Annual North American Meetings of the Regional Science Association International, Savannah (USA)
Juillet 2007 :	« Zones Franches Urbaines : quels effets sur l'emploi et les créations d'établissements » (Roland Rathelot et Patrick Sillard), séminaire du Département des Etudes Economiques d'Ensemble (DESE – INSEE).
Mars 2007 :	« The Social Multiplier and Labour Market Participation of Mothers » (Eric Maurin et Julie Moschion), journée de l'unité de démographie économique de l'INED : « Travail et Parentalité font-il bon ménage? »

ENSEIGNEMENT

2001-2002 : Université d'Evry, chargé de travaux dirigés en microéconomie.

1999-2001 : ENSAE, chargé de travaux dirigés en microéconomie.

1999-2000 : Paris IX Dauphine, chargé de cours de microéconomie.

AUTRES ACTIVITES

Rapporteur pour : *Review of Economic Studies, Journal of Labour Economics, Journal of Urban Economics, Regional Science and Urban Economics, Journal of Economic Geography, Urban Studies, Journal of Population Economics, Economica, Journal of Regional Science, Natural Resources Forum, Annales d'Economie et de Statistique, Population, La Revue Economique, Economie et Statistique, Economie et Prévision, La Revue Française d'Economie.*

2009-2012 : Membre de l'équipe CEPR dans le projet : « European Firms In a Global Economy: internal policies for external competitiveness » (EFIGE) financé par la Communauté Européenne dans le cadre du programme FP7.

2007-2010 : Responsable de l'équipe INED pour le réseau européen RTN : « Transnationality of Migrants » (TOM).

2008-2009 : Tuteur INED du Doctorant Matthieu Solignac (directeur de thèse : David Margolis) dans le cadre du Programme Doctoral Européen de Démographie.

2007. Participation au Comité de sélection des projets répondant à l'appel à projet « Mobilité Professionnelle » lancé par la DARES. Rapporteur pour deux des propositions soumises.

2005 : Organisation de la journée de l'unité 09 (Démographie Economique), INED, sur le thème : « Disparités Spatiales et Migrations ».

CONTRATS

2006-2008 : Participation à un contrat DREES-MIRE sur le thème : « Production et traitement des discriminations : nouvelles approches ». Nom du projet : *Discriminations et Territoires*, avec : Harris Selod, Thierry Magnac et Claudia Senik. Durée : 3 ans.

2003-2006 : Participation à un contrat d'Aide Concertée d'Activité (ACI) du Ministère de la Recherche sur le thème : « Développement Urbain Durable », avec Madior Fall, Thierry Magnac, et Harris Selod, durée : 36 mois.

2003-2006 : Participation à un contrat de la DREES afin d'étudier les disparités spatiales du système hospitalier à l'aide des données du *Programme de Médicalisation des Système d'Information (PMSI), 1997-2003* , avec Carine Milcent, durée : 36 mois.

REFERENCES

Pierre-Philippe Combes (CNRS, GREQAM). Vieille Charité, 2 rue de la Charité, 13236 Marseille Cedex 2. ☏ +33 (0) 4 91 14 07 33, Fax : +33 (0) 4 91 90 02 77, Email : ppcombes@univmed.fr

Gilles Duranton (University of Toronto). Department of Economics, 150 St George Street, Toronto, Ontario, M5S 3G7, Canada. ☏ +44 (0) 1 416 978 2678, Fax : +33 (0) 1 416 978 6713, Email : gilles.duranton@utoronto.ca

Jean-Marc Robin (Professeur à l'Université de Paris I). Maison des Sciences Economiques, EUREQua, 106-112 Boulevard de l'Hôpital, 75647 Paris Cedex 13.
☏ +33 (0) 1 44 07 82 08, Fax : +33 (0) 1 44 07 82 31, Email : Jean-Marc.Robin@ens.fr

Thierry Magnac (Professeur à l'Université de Toulouse 1 Sciences Sociales- IDEI). Manufacture des Tabacs, 21 allée de Brienne, 31000 Toulouse.
☏ +33 (0) 5 61 12 88 71, Fax : +33 (0) 5 61 12 86 37, Email : magnac@cict.fr

Annexe B. Articles

Les articles sont présentés dans l'ordre d'apparition du texte.

B.1. Estimating agglomeration economies with history, geology and worker effects

Combes PPh., Duranton G., Gobillon L. et S. Roux (2009), "Estimating agglomeration economies with history, geology and worker effects", CEPR Working Paper 6728, à paraître dans *The Economics of Agglomeration*, Edward Glaeser (ed.), NBER, Cambridge, MA.

45 pages

Estimating agglomeration economies with history, geology, and worker effects

Pierre-Philippe Combes[*†]
University of Aix-Marseille
Gilles Duranton[*‡]
University of Toronto
Laurent Gobillon[*§]
Institut National d'Etudes Démographiques
Sébastien Roux[*‖]
Centre de Recherche en Économie et Statistique

13 January 2009

Abstract: Does productivity increase with density? We revisit the issue using French wage and TFP data. To deal with the 'endogenous quantity of labour' bias (i.e., urban agglomeration is consequence of high local productivity rather than a cause), we take an instrumental variable approach and introduce a new set of geological instruments in addition to standard historical instruments. To deal with the 'endogenous quality of labour' bias (i.e., cities attract skilled workers so that the effects of skills and urban agglomeration are confounded), we take a worker fixed-effect approach with wage data. We find modest evidence about the endogenous quantity of labour bias and both sets of instruments give a similar answer. We find that the endogenous quality of labour bias is quantitatively more important.

Key words: agglomeration economies, instrumental variables, wages, TFP
JEL classification: R12, R23

[*]We are grateful to Alejandra Castrodad for superb research assistance with the GIS work and to France Guérin-Pace for providing us with historical data for French municipalities. Thanks also to Jason Faberman, Megan MacGarvie, Stuart Rosenthal, Will Strange, AEA, NARSA and NBER conference participants, and more particularly to the editor, Ed Glaeser. Financial support from the Centre National de la Recherche Scientifique (Combes) and from the Canadian Social Science and Humanities Research Council (Duranton) is gratefully acknowledged.

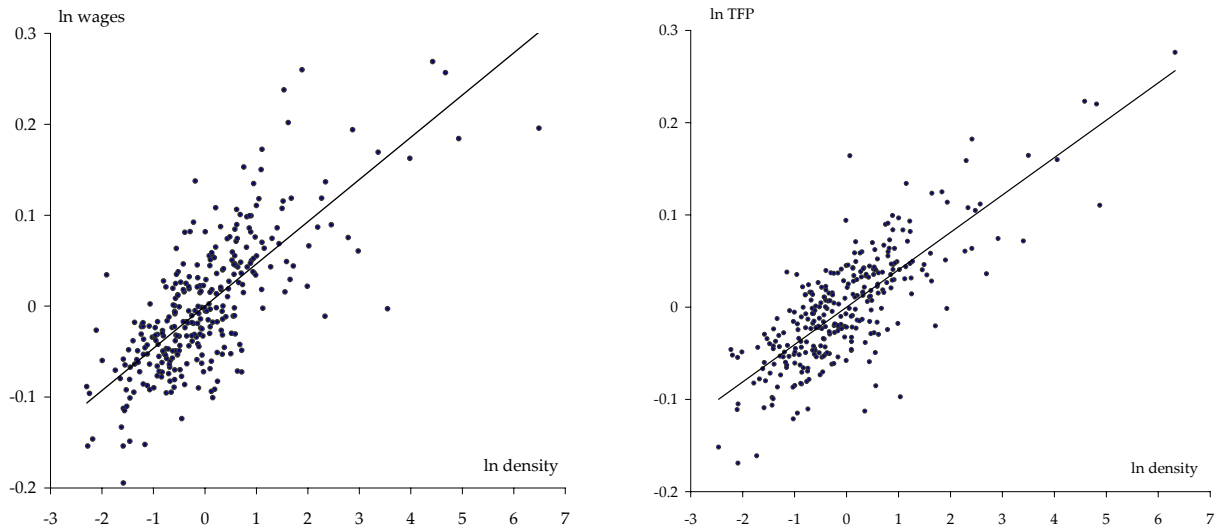
[†]GREQAM, 2 Rue de la Charité, 13236 Marseille cedex 02, France (e-mail: ppcombes@univmed.fr; website: http://www.vcharite.univ-mrs.fr/pp/combes/). Also affiliated with the Centre for Economic Policy Research.

[‡]Department of Economics, University of Toronto, 150 Saint George Street, Toronto, Ontario M5S 3G7, Canada (e-mail: gilles.duranton@utoronto.ca; website: http://individual.utoronto.ca/gilles/default.html). Also affiliated with the Centre for Economic Policy Research.

[§]Institut National d'Etudes Démographiques, 133 Boulevard Davout, 75980 Paris cedex 20, France (e-mail: laurent.gobillon@ined.fr; website: http://laurent.gobillon.free.fr/).

[‖]Centre de Recherche en Économie et Statistique (CREST-INSEE), 15 Boulevard Gabriel Péri, 92245 Malakoff cedex, France (e-mail: roux@ensae.fr).

Figure 1: Productivity and employment density in France



(a) Wages and employment density (306 employment areas, 1976-1996 average)

(b) TFP (Olley-Pakes) and employment density (306 employment areas, 1994-2002 average)

Source: DADS, BRN, RSI, SIREN and authors' calculations. All variables are centred around their mean. The R-squared is 56% in panel (a) and 61% in panel (b). See the rest of the paper for the details of the calculations.

1. Introduction

Productivity and wages are higher in larger cities and denser areas. This fact was first noted by Adam Smith (1776) and Alfred Marshall (1890) and has been confirmed by the modern empirical literature on this topic (see Rosenthal and Strange, 2004 for a review). The measured elasticity of local productivity with respect to employment density is typically between 0.04 and 0.10. We confirm this on French data. Figure 1 (a) plots mean log wages against employment density over 1976-1996 for 306 French employment areas. The measured density elasticity of wages is 0.05. Figure 1 (b) conducts a similar exercise using log TFP for the same 306 employment areas over 1994-2002. The measured density elasticity of TFP is 0.04.

To draw inference from figure 1, two fundamental identification problems must be dealt with. First, density and measures of productivity (wage or TFP) may be simultaneously determined. This could happen because more productive places tend to attract more workers and, as a result, become denser. An alternative explanation, albeit equivalent from an econometric perspective, is that there may be a missing local variable that is correlated with both density and productivity. We refer to this issue as the 'endogenous quantity of labour' problem. Since Ciccone and Hall (1996), a standard way to tackle this problem is to use instrumental variables (IV).

The second major identification problem is that more productive workers may sort into denser areas. This may occur for a variety of reasons. For instance, skilled workers may have a stronger preference for high density, perhaps because density leads to better cultural amenities. Alternatively, the productivity benefits of high density may be stronger for skilled workers. These explanations suggest that it is not only density that we expect to be simultaneously determined with productivity but also the characteristics of the local workforce. To make matters worse, we expect characteristics that are not usually observed by the statistician such as ambition or work discipline to matter and be spatially unevenly distributed. For instance, French university professors may have similar observable characteristics everywhere but a disproportionate fraction of the better ones are working in or around Paris. We refer to this problem as the 'endogenous quality of labour' problem. Since Glaeser and Maré (2001), a standard way to tackle this problem is to use the longitudinal dimension of the data.

One may also be concerned that density affects productivity in a myriad of ways, direct and indirect (see Duranton and Puga, 2004 for a review). Denser markets allow for a more efficient *sharing* of indivisible facilities (e.g., local infrastructure), risks, and the gains from variety and specialisation. Next, denser markets also allow for a better *matching* between employers and employees, buyers and suppliers, partners in joint-projects, or entrepreneurs and financiers. This can occur through both a higher probability of finding a match and a better quality of matches when they occur. Finally, denser markets can facilitate *learning* about new technologies, market evolutions, or new forms of organisation. Some of these mechanisms (e.g., matching) may have instantaneous effects while others (e.g., learning) may take time to materialise.[1]

Our paper addresses the issues of endogenous quantity and endogenous quality of labour. We do not attempt to distinguish between the different channels through which density could affect productivity and only aim at estimating a total net effect of density on wages. To deal with the endogenous quantity of labour problem we take an IV approach using both history and geology as sources of exogenous variation for population. To deal with the endogenous quantity of labour problem, we proceed as in Combes, Duranton, and Gobillon (2008*a*) and use the longitudinal dimension of extremely rich wage data. We impose individual fixed effects and local time-varying fixed effects in a wage regression. This allows us to separate local from individual effects and reconstruct some local wages net of individual observed and unobserved effects. Note that both approaches are necessary to identify the effect of density on productivity. Neither approach on its own would be sufficient.

Our main results are the following. The raw elasticity of mean wages to density is

[1]Even if the overall effect is positive, they may also be many negative effects of density on productivity due to crowding or congestion.

slightly below 0.05. Controlling only for the endogenous quantity of labour bias lowers this estimate to around 0.04. Historical and geological instruments lead to roughly the same answer. Controlling only for the endogenous quality of labour bias yields an even lower density elasticity of 0.033. Controlling for both source of biases leads to a coefficient of 0.027. When we also control for the fact that agglomeration economies spill over the spatial units boundaries, our preferred estimate for the elasticity of wages to local density stands at 0.02. These results are broadly confirmed when we use an alternative measure of productivity, TFP, rather than wages.

We draw a number of conclusions from this work. First, even though we control for two major sources of bias we still find evidence of small but significant agglomeration effects. Second, the sorting of workers across places is a quantitatively more important issue than their indiscriminate agglomeration in highly productive locations. Third, the importance of unobserved labour quality implies that wages should be favoured over TFP and other productivity measures since wage data are our main hope to deal with unobserved worker characteristics.

The rest of this chapter is as follows. Section 2 provides a simple model of productivity and wages in cities and discusses the two main estimation issues. Section 3 presents the wage data and our approach to the endogenous quality of labour bias. Section 4 presents our instruments and discusses the details of our instrumentation strategy. Our results for wages are presented in section 5 while those for productivity follow in section 6. Finally, section 7 concludes.

2. Identification issues when estimating agglomeration effects

We consider a simple theoretical model of the relationship between local characteristics and wages or productivity. Take a competitive firm i operating under constant returns to scale. Its output y_i depends on the amounts of capital k_i and labour l_i it uses and its total factor productivity A_i:

$$y_i = A_i k_i^{\alpha} l_i^{1-\alpha} \,. \tag{1}$$

If all firms face the same interest rate r, the first-order conditions for profit maximisation imply that the wage *rate* is given by:

$$w_i = (1-\alpha) \left(\frac{\alpha}{r}\right)^{\alpha/(1-\alpha)} A_i^{1/(1-\alpha)} \,. \tag{2}$$

Taking logs directly leads to:

$$\ln w_i = \text{Constant} + \frac{1}{1-\alpha} \ln A_i \,. \tag{3}$$

The whole focus of the agglomeration literature is then on how the local characteristics of area a where firm i is located determine productivity. We assume that TFP depends on a vector of local characteristics X_a and (observed and unobserved) firm characteristics μ_i:

$$\ln A_i = X_{a(i)}\varphi + \mu_i \,. \tag{4}$$

Inserting into (3) implies:

$$\ln w_i = \text{Constant} + \frac{1}{1-\alpha}\left(X_{a(i)}\varphi + \mu_i\right). \tag{5}$$

This equation can in principle be estimated using wage data and local characteristics. An alternative strategy is to insert (4) into (1), takes logs, and estimate:

$$\ln y_i = \alpha \ln k_i + (1-\alpha)\ln l_i + X_{a(i)}\varphi + \mu_i. \tag{6}$$

Hence both wage and firm level (TFP) data can be used to estimate the coefficients of interest in the vector φ or $\varphi/(1-\alpha)$.[2] The first identification problem when estimating (5) or (6) is that the effect of local characteristics, $X_{a(i)}$, on wages and productivity may not be causal (endogenous quantity of labour bias). In other words, unobserved local determinants of firm productivity that are part of the error term μ_i may well be correlated with $X_{a(i)}$. Second, local characteristics of workers that are not observed, and therefore not included in $X_{a(i)}$, may not be comparable across areas (endogenous quality of labour bias). Again this creates some correlation between μ_i and $X_{a(i)}$.[3]

To provide further justification for equations (5) and (6) and clarify some issues regarding endogenous quantity of labour bias, we note that the literature on the microfoundations of agglomeration (e.g. Duranton and Puga, 2004) typically leads to equilibria where the wage in area a, w_a, depends on a local productivity shifter, B_a, and the local workforce, N_a:

$$w_a = B_a N_a^{\delta}, \tag{7}$$

This equation is consistent with (5) when agglomeration effects are such that $A_i^{1/(1-\alpha)} = B_a N_a^{\delta} e^{\mu_i}$. B_a can be viewed as a short-hand for all variables other than N_a in X_a. With $\delta > 0$, individual wages increase with N_a. If N_a is exogenously determined, it can be part of the vector of local characteristics X_a and δ be appropriately estimated with OLS. Following Roback (1982) and subsequent literature, we may assume instead that workers choose their city of residence. This choice is determined, through utility maximisation, by

[2]Combes, Mayer, and Thisse (2008*b*, chapter 11) provide a more complete model of local productivity and a precise discussion of a number of issues including those that relate to the prices of factors, intermediates and final output.

[3]In addition, when estimating (6), factors might be endogenous as well. This issue is discussed in section 6.

the difference between the wage and the local cost of living:

$$U_a = w_a - C_a N_a^{\lambda}, \tag{8}$$

In any city, the cost of living increases with the city workforce and depends on other characteristics such as amenities which have utility costs (and benefits). A spatial equilibrium equalises utility across cities.

Assuming $\lambda > \delta$ and normalising equilibrium utility to zero, the above yields:

$$N_a = \left(\frac{B_a}{C_a}\right)^{\frac{1}{\lambda-\delta}} \qquad \text{and} \qquad w_a = B_a^{\frac{\lambda}{\lambda-\delta}} C_a^{\frac{-\delta}{\lambda-\delta}}. \tag{9}$$

At one extreme if there are no differences in productivity across cities other than due to population difference (i.e., $B_a = B$) and only costs vary, the OLS coefficient on log workforce when regressed against log wage will be (appropriately) δ. In the opposite case where costs are the same everywhere ($C_a = C$) and only productivity vary, the regression will estimate instead λ. In the general case where both costs and productivity vary, the estimated coefficient on log workforce will be between δ and λ.[4] The intuition for that result is that if the variation in local workforce comes solely from local costs, it is 'exogenous' and the proper coefficient is estimated. If instead the workforce is determined by the variation in productivity, wages in equilibrium only reflect the extent to which local costs increase with the size of the workforce. We need to keep this point in mind for our estimation strategy below.

This problem actually goes deeper than that. Our model considers only two factors of production, labour and physical capital, which is mobile and whose price can reasonably be taken to be constant everywhere. Then the term associated with its price r enters the constant and raises no further problem. However, land may also enter as a factor of production. Unlike for capital, the price of land varies across areas. Following again Roback (1982), we expect better consumption amenities to draw in more population and imply higher land prices. Firms will thus use less land. In turn, this lowers the marginal product of labour when land and labour are imperfect substitutes in the production function (as might be expected). Put differently, non-production variables may affect both population patterns and be capitalised into wages. To deal with this problem, we could attempt to control for local variables that directly affect consumer utility and thus land prices. However, our range of controls is limited and we are reluctant to use a broad range of local amenities since many of them are likely to be simultaneously determined with wages.

[4]Using the results from (9), it is easy to show that the estimated coefficient for log workforce will be: $\frac{\lambda Var(\ln B_a)+\delta Var(\ln C_a)-(\delta+\lambda)Cov(\ln B_a, \ln C_a)}{Var(\ln B_a)+Var(\ln C_a)-Cov(\ln B_a, \ln C_a)}$. With zero covariance between B_a and C_a and equal variance, this reduces to $\frac{\delta+\lambda}{2}$.

Faced with reverse causality and missing variables that potentially affect both wages and the density of employment, our strategy is to rely on instrumental variables.[5] Hence, we are asking to our instrument to deal with both the reverse-causality problem and the missing variable issue highlighted here.[6]

Turning to the endogenous quality of labour bias, note that the quantity derived in equation (2) and used throughout the model is a wage rate per efficiency unit of labour. Even if we are willing to set aside the issue that different types of labour should be viewed as different factors of production, not all workers supply the same number of efficiency units of labour per day. However, the data for individual workers is about their daily earnings, that is their wage *rate* times the efficiency of their labour. For worker j employed by firm i it is convenient to think of their earnings as being $W_j = w_{i(j)} \times s_j$ where their level of skills s_j is assumed to map directly into the efficiency of their labour. Hence, individual skills must be conditioned out from the regression to estimate (5) properly. Otherwise, any correlation between local characteristics and the skills of the local workforce will lead to biased estimates for agglomeration effects. Put differently, the quality of workforce in an area is likely to be endogenous. Previous work on French data (Combes *et al.*, 2008*a*) leads us to be believe that this is a first-order issue.

To deal with this problem of endogenous labour quality, a number of approaches can be envisioned. The first would be to weigh the workforce by a measure of labour quality at the area level and try to instrument for labour quality just like we instrument for labour quantity. Instruments for labour quality are very scarce. The only reasonable attempt is by Moretti (2004) who uses land-grant colleges in US cities to instrument for the local share of workers with higher education. In any case, this is unlikely to be enough because we also expect unobservables such as ambition or work discipline to matter and be spatially unevenly distributed (Bacolod, Blum, and Strange, 2009).

[5]Alternative approaches may include focusing on groups of workers for which there is an element of exogeneity in their location decision. One could think for instance of spouses of military personnel. However such groups are likely to be very specific. Another alternative may be to look at 'natural experiments' that led to large scale population and employment changes. Such experiments are very interesting to explore a number of issues. For instance, Davis and Weinstein (2008) estimate the effects of the US bombing of Japanese cities during World War II on their specialisation to provide some evidence about multiple equilibria. Redding and Sturm (2008) use the division of Germany after World War II to look at the effects of market potential. However such natural experiments are not of much relevance to study productivity since the source of any such large scale perturbation (e.g., the bombing of Japanese cities) is also likely to affect productivity directly and there is no natural exclusion restriction.

[6]The issue with instrumenting is that the number of possible instruments is small while there are potentially dozens of (endogenous) variables that can describe a local economy. In view of this problem, our strategy is to consider parsimonious specifications with no more than one or two potentially endogenous variables. The drawback is that the exclusion restriction for the instruments (i.e., lack of correlation between the instruments and the error) is more difficult to satisfy than with a greater number of controls. Despite this, we think that a more demanding exclusion restriction is preferable to the addition of inappropriate, and possibly endogenous, controls.

To tackle sorting heads on, previous literature has attempted to use area characteristics at a different level of spatial aggregation. For instance, Evans, Oates, and Schwab (1992) use metropolitan characteristics to instrument for school choice while Bayer, Ross, and Topa (2008) use location at the block level and assume an absence of sorting conditional on neighbourhood choice.[7] In our data, although we know location at the municipal level, we are loathe to make any strong spatial identifying assumption of that sort. A more satisfactory alternative would be to estimate a full system of equations, modelling explicitly location choice. Unfortunately, due to both the difficulty of finding meaningful exclusion restrictions and the complications introduced by the discrete nature of the choice among many locations, this is a difficult exercise. Dahl (2002) proposes a new approach to this problem but this can be applied to cross-section data only.

The last existing approach is to use the longitudinal dimension of the data as in Glaeser and Maré (2001), Moretti (2004) and Combes *et al.* (2008*a*). This is the approach we follow. The details of our methodology are described in the next section.

3. Sorting and wage data

Choice of spatial zoning, sectoral aggregation, and explanatory variables

The choice of geographical units could in principle be of fundamental importance. With the same data, there is no reason why a partial correlation that is observed for one set of spatial units should also be observed for an alternative zoning. In particular, the shape of the chosen units may matter. However, Briant, Combes, and Lafourcade (2007) compare the results of several standard exercises in spatial economics using both official French units, which were defined for administrative or economic purposes, and arbitrarily defined ones of the same average size (i.e., squares on a map). Their main finding is that to estimate agglomeration effects, the localisation of industries, and the distance decay of trade flows across areas, the *shape* of units makes no difference.

With respect to our choice of units, we opt for French employment areas ('zones d'emploi'). Continental France is fully covered by 341 employment areas, whose boundaries are defined on the basis of daily commuting patterns. Employment areas are meant to capture local labour markets and most of them correspond to a city and its catchment area or to a metropolitan area. This choice of relatively small areas (on average 1,500 km^2) is consistent with previous findings in the agglomeration literature that agglomeration effects are in part very local (Rosenthal and Strange, 2004). Nevertheless, we are aware

[7]Opposite spatial identifying assumptions are made. In Evans *et al.* (1992), the choice of the more aggregate area is assumed to be exogenous while location choice at a lower spatial level is not. Bayer *et al.* (2008) assume instead that randomness prevails at the lower level of aggregation and not at the higher level of aggregation.

that *different spatial scales* may matter with respect to agglomeration effects (see Briant *et al.*, 2007, and previous literature). We need to keep this important issue in mind when deciding on a specification.

Turning to the level of sectoral aggregation, a key question regards whether the benefits from agglomeration stem from the size of the overall local market (*urbanisation economies*) or from geographic concentration at the sector level (*localisation economies*). Although we want to focus on overall scale effects, sector effects cannot be discarded. Previous results for France suggest that they matter although they are economically far less important than overall scale effects (Combes *et al.*, 2008*a*). In the following, we work at the level of 114 three-digit sectors.[8]

The main explanatory variable we are interested in is employment density. It is is our favourite measure of local scale. Since Ciccone and Hall (1996), density-based measures have often been used to assess overall scale effects. Their main advantage compared to alternatives measures of size such as total employment or total population is that density-based measures are more robust to the zoning. In particular, Greater Paris is divided into a number of employment areas. The true economic scale of these Parisian employment areas is much better captured by their density than any absolute measure of employment.

To repeat, French employment areas are relatively small and determined by commuting patterns. On the other hand, input-output linkages may not be limited by commuting distances. Hence we expect some agglomeration effects to take place at a scale larger than employment areas. There is by now a lot of evidence that the market potential of an area matters (Head and Mayer, 2004). In some regressions, we thus also consider the market potential of an area that we define as the sum of the density of the other areas weighted by the inverse distance to these areas.[9] Experimenting with other measures leads to very similar results.

Main wage data

We use an extract from the Déclarations Annuelles des Données Sociales (DADS) or Annual Social Data Declarations database from the French statistical institute (INSEE). The DADS are collected for pension, benefits and tax purposes. Establishments must fill a report for each of their employees every year. An observation thus corresponds to an

[8]We view this level of aggregation as a reasonable compromise. On the one hand, we need finely defined sectors in wage regressions and for TFP estimation. On the other hand, localisation economies are expected to be driven by similarities in customers, suppliers, workers, and technology and thus take place at a fairly broad level of sectoral aggregation.

[9]We retain a simple specification for market potential and do not aim to derive it from a 'New Economic Geography' model (Head and Mayer, 2004). Alternative specifications for market potential are highly correlated with the one we use. See Head and Mayer (2006) for further evidence and discussion of this fact.

employee-establishment-year combination. The extract we use covers all employees in manufacturing and services working in France and born in October of even-numbered years.

For each observation, we know the age, gender, and occupation at the two-digit level. Except for a small sub-sample, education is missing. We also know the number of days worked but not hours for all years so that we restrict ourselves to full-time employees for whom hours are set by law. For earnings, we focus on total labour costs deflated by the French consumer price index. We refer to the real 1980 total labour cost per full working day as the wage. The data also contains basic establishment level information such as location and three-digit sector.

The raw data contains 19,675,740 observations between 1976 and 1996 (1981, 1983, and 1990 are missing). The details of the cleaning of the data is described in Combes *et al.* (2008*a*). After selecting only full-time workers in the private sector, excluding outliers, dumping a number of industries with reporting problems, and deleting observations with coding problems, we end up with 8,826,422 observations. For reasons of computational tractability, we keep only six points in time (every four years: 1976, 1980, 1984, 1988, 1992, and 1996), leaving us with 2,664,474 observations.

Using the above data, we can construct a number of variables for each year. Our main explanatory variable, employment density can be readily calculated from the data.[10] So can market potential. For each area and sector, we also compute the number of establishments, the share of workers in professional occupations, and the share of the sector in local employment. As controls we also use three amenities variables. These amenities variables are the share of population located on a sea shore, mountains, and lakes and waterways. These variables come from the French inventory of municipalities. We aggregate them at the level of employment areas, weighting each municipality by its population.[11] Table 1 below reports a number of descriptive statistics for French employment areas.

Three wages

The simplest way to implement equation (5) is to compute the mean wage for each area and year, and take its log:

$$W_{at}^{1} \equiv \ln \overline{w}_{at} \equiv \ln \left(\frac{1}{N_{at}} \sum_{j \in (a,t)} w_{jt} \right). \tag{10}$$

where w_{jt} is the wage of worker j and year t and N_{at} the number of workers in area a and year t.

[10] We keep in mind that the years are not the same for the wage and TFP regressions. For each set of regressions, the explanatory variables are constructed from the corresponding data sources.

[11] Each employment area contains on average more than 100 municipalities.

We can then use W^1_{at} as dependent variable to be explained by local employment density and other local characteristics in equation (14). Using a simple log mean like W^1_{at} throws a number of problems. First, when using mean wages we do nothing regarding the endogenous quality of labour bias. Second, we do not condition out sector effects.[12]

To deal with these two problems, a first solution is to use all the available observables about workers and proceed as follows. We first compute a mean wage per employment area, sector, and year:

$$\overline{w}_{ast} \equiv \frac{1}{N_{ast}} \sum_{j \in (a,s,t)} w_{jt}. \tag{11}$$

This wage can then be regressed on a number of (mean) characteristics of the workers and the local sector. More specially we can estimate the following first step regression:

$$\ln \overline{w}_{ast} = W^2_{at} + \gamma_s + X_{ast}\varphi + \epsilon_{ast}. \tag{12}$$

In this equation, γ_s is a sector dummy, and X_{ast} is a set of characteristics for sector s in area a and year t and the workers employed therein. To capture sector effects we use in X_{ast} the (log) share of local employment in sector s and the (log) number of local establishments in this sector. Also in X_{ast}, the mean individual characteristics are the age, its square, and the shares of employment in each of 6 skill groups.[13] In equation (12), the coefficient of interest is W^2_{at}, a fixed effect for each employment area and year. When estimating (12), all local sector and mean individual characteristics are centred and the observations are weighted by the number of workers in each cell to avoid heteroscedasticity.

The coefficients W^2_{at} can, in a second step, be regressed on local employment density and other local characteristics as stipulated by equation (5). While further details and justifications about the estimation of (12) are given in Combes *et al.* (2008*a*), three important issues need to be briefly discussed. First, the approach described here first estimates local fixed effects before using them as dependent variable in a second step. We prefer this two-step approach to its one-step counterpart for reasons made clear below.

Next, estimating (12) with OLS may condition out sectoral effects but it does not take care of the possible simultaneity between mean sector wages and local sector characteristics. A high level of specialisation in a certain sector may induce high wages in this sector. Alternatively high local wages may simply be a reflection of strong local advantage also

[12]One further (minor) issue need to be mentioned. We take the log of mean wages rather than the mean of log (individual) wages. When viewing local wages as an aggregate of individual wages, the log of mean wages is not the proper aggregate to consider. Mean log wages should be used instead. However, the former is easier to implement than the latter, especially for those who do not have access to micro-data. In any case, this issue is empirically unimportant since the correlation between log mean wages and mean log wages is 0.99.

[13]The shares of each skill in local sector employment capture the effects of both individual characteristics at the worker level and the interactions between workers. The two cannot be separately identified with aggregate data.

leading to a high level of specialisation. We acknowledge this concern at the sector level but we do not deal with it. The main reason is that the coefficients for local specialisation and the number of establishments, although significant, only explain a very small part of the variation in (12) (Combes *et al.*, 2008*a*).

Finally, controlling for observable labour market characteristics including one-digit occupational categories (for lack of control for education) attenuates concerns about the endogenous quality of labour bias. However, they do not eradicate them entirely.

A more powerful way to deal with the endogenous quality of labour bias is to estimate:

$$\ln w_{jt} = W^3_{a(jt)t} + \gamma_{s(jt)} + X^1_{a(jt)s(jt)t}\varphi^1_{s(jt)} + X^2_{jt}\varphi^2 + \theta_j + \epsilon_{jt}. \tag{13}$$

This equation is estimated at the level of individual workers and contains a worker fixed effect θ_j which controls for all fixed individual characteristics.[14] The use of individual data also allows us to control for individual characteristics X^2_{jt} (age and its square) separately from (centred) local industry characteristics X^1_{ast}. The latter contain the share of local employment of the sector, the local number of firms in the sector, and the local share of professional workers. The coefficient of interest in equation (13) is the wage index W^3_{at} for each area and year after conditioning out sector effects, observable time-varying individual characteristics, and all fixed individual characteristics. If we ignore again the possible endogeneity of local sector characteristics, the main issue when estimating (13) regards the endogeneity of location or sector choices. However, because we have sector effects and time-varying local effects, W^3_{at}, problems only arise when we have spatial or sector sorting based on the worker-specific errors. In particular, there is no bias when sorting is based on the explanatory variables, *including individual, area-year, and industry fixed effects*. More concretely, there is a bias when the location decision is driven by the exact wage that the worker can get at locations in a given year but there is no bias when workers base their location decision on the average wage of other workers in an area and their own characteristics, i.e., when they make their location decision on the basis of their expected wages. See Combes *et al.* (2008*a*) for further discussion.

Note that we prefer this two-step approach, which first estimates (12) or (13) before regressing W^2_{at} or W^3_{at} on local characteristics, to its corresponding one-step counterpart for three reasons. First, we can properly take into account correlations between area-sector variables and error terms at the area level. Second, a two-step approach allows us to account for area-specific error terms when computing the standard errors for the coefficients we estimate. Doing so is important because Moulton (1990) shows that standard errors can be seriously biased otherwise. Accounting for area-specific errors with a one-step

[14] Equation (13) is identified from both the movers (to identify the difference between W^3_{at} and $W^3_{a't+1}$) and the stayers (to identify the difference between $W^3_{\mathbf{a}t}$ and W^3_{at+1}).

Table 1: Summary statistics for our main variables (averages across 306 employment areas).

	Mean	Std. dev.
Mean wage (1976–1996, in 1980 French Francs, per day)	207.9	15.8
W^1	5.3	0.074
W^2	5.2	0.070
W^3	-0.04	0.049
Employment density (workers per sq. km)	64.4	543.0
ln employment density	2.4	1.2
Market potential (workers km per sq. km)	108.1	139.9
ln market potential	4.4	0.7
1831 Urban population density (inh. per sq. km)	38.2	419.8
1881 Urban population density (inh. per sq. km)	106.8	1232.3
Sea (average % municipalities on a coast line)	8.8	21.1
Lake (average % municipalities on a lake)	17.2	12.9
Mountain (average % municipalities on a mountain)	9.8	19.7

Source: DADS for the first eight lines, historical censuses for the next two, and 1988 municipal inventory for the last three. For sea, lake and mountain, we have for each employment area the percentage of municipalities on a coast, with a lake, or on a mountain. We average this quantity across employment areas.

approach is not possible given that workers can move across areas. Third, we can conduct a variance decomposition for the second stage.[15]

Finally, to avoid identifying out of the temporal variation, we average the three wage variables and all the explanatory variables across the six years of data we use.[16] Before turning to our results, it is interesting to note that these three local wage variables are strongly correlated with one another. The correlation between W^1 and W^2 is 0.87, the correlation between between W^1 and W^3 is 0.81, while the correlation between W^2 and W^3 is 0.91. Table 1 reports a number of descriptive statistics for French employment area.

4. Instruments

That the estimation of agglomeration economies could be plagued by simultaneity was first articulated by Moomaw (1981). To preview of our IV approach, we note first that using historical variables such as long lags of population density to instrument for the size

[15]It is also true that using as dependent variable a coefficient estimated in a previous step introduces some measurement error. The procedure used in Combes *et al.* (2008*a*) to control for this problem shows that it makes no difference because the coefficients are precisely estimated at the first step.

[16]These averages are weighted by by the number of workers in the area for each year to obtain a wage index for the average worker in the area over time. By contrast, our final regressions for the cross-section of employment areas assess whether denser areas make their average firm more productive. There is no longer any reason to weigh the observations (by the number of workers) in these regressions.

or density of local population is standard since Ciccone and Hall's (1996) pioneering work. To the extent that (i) there is some persistence in the spatial distribution of population and (ii) the local drivers of high productivity today differ from those of a long gone past, this approach is defensible. An alternative strategy is to use the nature of soils since geology is also expected to be an important determinant of settlement patterns. Some soils are more stable than others and can thus support a greater density of people. More fertile lands may have also attracted people in greater number, etc. To the extent that geology affects the distribution of population (i.e., labour supply) and does not otherwise cause productivity (i.e., labour demand) because fertile lands are no longer a relevant driver of local wealth, it can provide reasonable instruments to explain the distribution of employment. Except by Rosenthal and Strange (2008) in a slightly different context, geology has not been used to instrument for the distribution of population.

Description of the instruments

Our first set of instruments is composed of historical populations from early French censuses. For 26 French censuses prior to our earliest year of data (1976) we know the 'urban' population for each municipality. Among available censuses we choose the earliest one from 1831 and another from 1881, 50 years later.[17] We also experimented with other years. Unfortunately, urban population in historical censuses is only reported above a threshold of 5,000. For 1831, there are 35 employment areas for which no municipality had an urban population above 5,000. A small majority of them are rural areas while the others are densely populated employment areas with strong municipal fragmentation. We think of this as being measurement error. To minimise weak instrument problems, we drop these 35 employment areas.

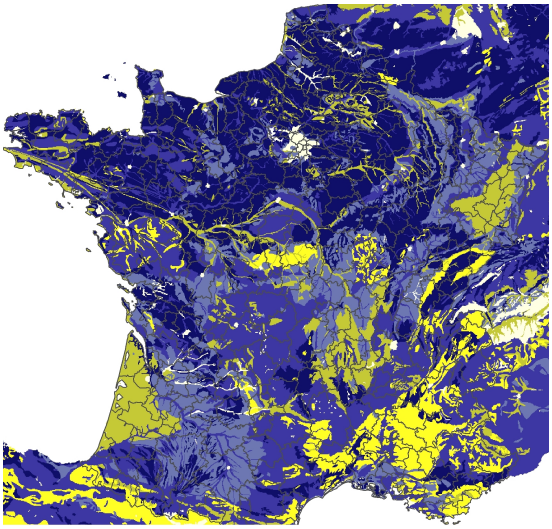
Our second group of instruments is composed of geological variables from the European Soil Database (ESDB) compiled by the European Soil Data Centre. The data originally come as a raster data file with cells of 1 km per 1 km. We aggregated it at the level of each employment area.[18] Given that soil characteristics are usually discrete, we use the value that appears most often in each area. To take an illustrative example, the initial and transformed data for the water capacity of the subsoil are represented in figure 2. For a small number of densely populated employment areas in Greater Paris, the most important category is sometimes "missing". When this is the case, we turn to the

[17]Because they are in log, using these two variables together allows us to instrument for both past 1831 level and past growth between 1831 and 1881.

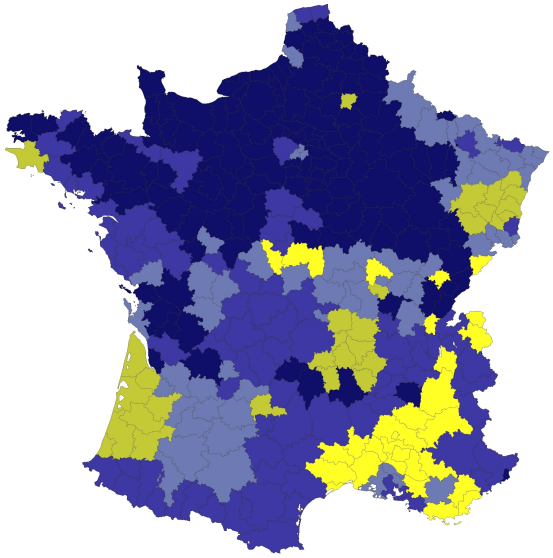
[18]To aggregate the information from 1 km by 1 km pixels to employment areas, the zonal statistics tool from ArcGIS 9 was used. The tool uses the zones defined in the zone dataset (in our case French employment areas), and internally converts the vectors into a zone raster, which it aligns with the value raster dataset for soils.

Figure 2: Geological characteristics: Water capacity of the subsoil

Panel A. Original data



Panel B. Transformed data



Source: European Soil Database. Panel A represents the initial raster data. Panel B represents the transformed version of the same data after imputation of the missing values for 7 employment areas in Greater Paris. In both panels, the darkest shade of grey corresponds to 'very high' (i.e., above 190 mm), the second darkest shade corresponds to 'high' (between 140 and 190 mm) followed by 'medium' (100 – 140 mm), 'low' (5 – 100 mm), and 'very low' (0 – 5 mm). Missing values in panel A (around Paris) are in white.

second most important category. In the rare instances where the information is missing from all the pixels in an employment area, we impute the value of a neighbouring area (chosen because it takes similar values for other soil characteristics). For instance, the water capacity of the subsoil in Central Paris is missing. We impute the value of that of its close neighbour Boulogne-Billancourt.

In total, we generate 12 variables from the ESDB.[19] The first four describe the nature of the soils according to the mineralogy of their subsoil (3 categories) and topsoil (4 categories) and the nature of the dominant parent material at a broad level of aggregation (6 categories) and at a finer level (with 20 categories). More precisely, the mineralogy variables describe the presence of various minerals in the topsoil (the first layer of soil, usually 5 to 15 cm deep) and the subsoil (the intermediate layer between the topsoil and the bedrock). The dominant parent material of the soil is a description of the underlying

[19]The ESDB (v2 Raster Archive) contains many more characteristics. For France, some of them like the soil code according to the standard FAO classification are poorly reported. A large number of characteristics also contain categories that refer to land use (e.g., 'urban' or 'agriculture') and are thus not appropriate here. More generally, characteristics *a priori* endogenous to human activity were discarded. Finally, some characteristics such as the secondary dominant parent material stroke us as anecdotal and unlikely to yield relevant instruments.

geological material (the bedrock). Soils usually get a great deal of structure and minerals from their parent material. The more aggregate dominant parent material variable (in 6 categories) contains entries such as igneous rocks, glacial deposits, or sedimentary rocks. Among the latter, the detailed version of the same variable (with 20 categories) distinguishes between calcareous rocks, limestone, marl, and chalk.

The next seven geological characteristics document various characteristics of the soil including the water capacity of the subsoil (5 categories) and topsoil (3 categories), depth to rock (4 categories), differentiation (3 categories), erodibility (5 categories), carbon content (4 categories), and hydrogeological class (5 categories). Except for the hydrogeological class which describes the circulation and retention of underground water, the meaning of these variables is relatively straightforward. Finally, we create a measure of local terrain ruggedness by taking the mean of maximum altitudes across all pixels in an employment area minus the mean of minimum altitudes. This variable thus captures variations of altitude at a fine geographical scale.

Relevance of the instruments

Following equations (5) and (6), the specifications we want to estimate are:

$$\ln W_a = \text{Constant} + X_a \varphi^W + \mu_a^W \tag{14}$$

and

$$\ln TFP_a = \text{Constant} + X_a \varphi^{TFP} + \mu_a^{TFP}, \tag{15}$$

where μ_a^W and μ_a^{TFP} are the errors terms for the wage and TFP equations. The vector of dependent variables X_a contains the three amenity variables discussed above, (log) employment density, and sometimes market potential. These last two variables are suspected of being simultaneously determined with wages and TFP.

Estimating the effect of employment density and market potential on local wages and productivity using instrumental variables can yield unbiased estimates provided that the instruments satisfy two conditions, relevance and exogeneity. Formally, these conditions are

$$Cov(Density_a, Z_a|.) \neq 0, \qquad Cov(MarketPotential_a, Z_a|.) \neq 0, \tag{16}$$

and

$$Cov(\mu_a^X, Z_a) = 0 \qquad \text{for } X = W \text{ and } X = TFP, \text{ respectively}, \tag{17}$$

where Z denotes the set of instruments. We begin by discussing the ability of our instruments to predict contemporaneous employment density and market potential conditionally to the other controls.

The stability of population patterns across cities over time is a well documented fact (see Duranton, 2007, for a recent discussion). This stability is particularly strong in France

Table 2: R-squareds of univariate regressions and pairwise correlations : historical vs. density and market potential (1976-1996)

	ln(employment density)		ln(market potential)	
ln(1831 density)	0.57	(0.75)	0.05	(0.24)
ln(1881 density)	0.78	(0.88)	0.10	(0.33)
ln(1831 market potential)	0.21	(0.46)	0.96	(0.98)
ln(1881 market potential)	0.22	(0.47)	0.99	(0.99)

306 observations.
Adjusted R-squared in plain text and pairwise correlations between parentheses.

(Eaton and Eckstein, 1997). The raw data confirm this. Table 2 presents pairwise correlations between our four historical instruments and current employment density and market potential.[20] For the sake of comparison with geology variables below, we also report the R-squareds of the corresponding univariate regressions. We can see that the log urban population densities of 1831 and 1881 are good predictors of current employment density. Past market potentials computed from 1831 and 1881 urban populations also predict current market potential extremely well.

Turning to geological characteristics, we expect the nature of soils and their characteristics to be fundamental drivers of population settlements. Soil characteristics arguably determine their fertility. Since each soil characteristic is described by several discrete variables, it is not meaningful to run pairwise correlations as with historical variables. Instead, table 3 reports the R-squared when regressing employment density and market potential against various sets of dummies for soil characteristics. The results show that some geological characteristics like the dominant parent material or the depth to rock have good explanatory power. Other soil characteristics such as their mineralogy or their carbon content are less powerful predictors of current population patterns. Note also that soil characteristics tend to be better at explaining the variations of market potential than employment density. This is not surprising since most soil characteristics vary relatively smoothly over fairly large spatial scales while variations in density are more abrupt and take place at smaller spatial scales.

While the correlations and R-squareds reported in tables 2 and 3 are interesting, equation (16) makes clear that the relevance of an instrument depends on the *partial* correlation of the instrumental variables and the endogenous regressor. To assess these partial correlations, table 4 presents the results of OLS regressions of log density on our instrumental

[20]We use the measures of density used for our wage regressions (1976-1996). Our measures of density for the TFP regressions differ slightly since they are calculated from a slightly different source and cover different years.

Table 3: R-squareds when regressing density and market potential on soil characteristics

	ln(emp. density)	ln(market pot.)
Subsoil mineralogy (2 dummies)	0.02	0.06
Topsoil mineralogy (3 dummies)	0.02	0.06
Dominant parent material (5 dummies)	0.11	0.31
Dominant parent material (19 dummies)	0.13	0.48
Topsoil water capacity (2 dummies)	0.03	0.23
Subsoil water capacity (3 dummies)	0.01	0.32
Depth to rock (3 dummies)	0.10	0.35
Soil differentiation (2 dummies)	0.07	0.19
Erodibility (4 dummies)	0.04	0.19
Carbon content (3 dummies)	0.04	0.04
Hydrogeological class (4 dummies)	0.01	0.04
Hydrogeological class (4 dummies)	0.05	0.10

Adjusted R-squareds. 306 observations.

variables and controls. Table 5 reports results for a similar exercise with market potential.

Column 1 of table 4 examines the partial correlation between employment density and 1831 population density while conditioning out amenities (sea, lake, and mountain). Column 2 performs a similar regression using 1881 instead of 1831 population density. In both columns, the coefficient on past density is highly significant and close to unity. In columns 3 to 9, we regress contemporaneous employment density on a series of soil dummies concerning their mineralogy, dominant parent material, water capacity, carbon content, depth to rock, and soil differentiation. For lack of space, we do not report all the coefficients but it must be noted that at least one dummy is significant at 5% in each regression.

The comparison of R-squareds in columns 1-2 versus 3-9 shows immediately that long lags of population density explain a greater share of the variations in contemporaneous employment density than soil characteristics. To make a more formal assessment of the relevance of our instruments we turn to the weak instrument tests developed by Stock and Yogo (2005).[21] Table 4 presents the relevant F-statistics. The two lagged density instruments in columns 1 and 2 have F-statistics close to 400 and 1000, respectively. This makes them very strong in light of the critical values reported by Stock and Yogo (2005) in their tables 1-4. The soil instruments are weaker by comparison and fall below the critical

[21]Stock and Yogo (2005) provide two tests for weak instruments. They are both based on a single F-statistic of the instrumental variables but use different thresholds. The first one tests the hypothesis that two-stage least square (TSLS) small sample bias is small relative to the OLS endogeneity bias ('bias test'). Second, an instrument is considered strong if, from the perspective of the Wald test, its size is 'close' to its level for all possible configurations of the IV regression ('size test'). Note that instruments may be weak in one sense but not another, and instruments may be weak in the context of TSLS but not when using limited information maximum likelihood (LIML).

Table 4: First stage: Density

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
ln(1831 density)	0.906								
	(0.046)[a]								
ln(1881 density)		0.924							
		(0.030)[a]							
Ruggedness			- 0.710						
			(0.224)[a]						
Subsoil mineralogy	N	N	N	Y	N	N	N	N	N
Dominant parent material (6 categories)	N	N	N	N	Y	N	N	N	N
Subsoil water capacity	N	N	N	N	N	Y	N	N	N
Soil carbon content	N	N	N	N	N	N	Y	N	N
Depth to rock	N	N	N	N	N	N	N	Y	N
Soil differentiation	N	N	N	N	N	N	N	N	Y
R-squared	0.58	0.78	0.07	0.07	0.17	0.06	0.10	0.15	0.11
F-test (H_0 – All instruments zero)	395.7	1018.8	10.0	5.5	9.1	1.7	6.5	12.6	12.3
Partial R-squared	0.57	0.77	0.03	0.04	0.13	0.02	0.06	0.11	0.08

Dependent variable: ln(employment density). 306 observations for each regression.
All regressions include a constant and three amenity variables (sea, lake, and mountain).
Standard errors in parentheses. *a*, *b*, *c*: corresponding coefficient significant at 1, 5, 10%.

Table 5: First stage: Market potential

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
ln(1831 market pot.)	1.026								
	(0.012)[a]								
ln(1881 market pot.)		0.970							
		(0.007)[a]							
Ruggedness			- 0.339						
			(0.111)[a]						
Subsoil mineralogy	N	N	N	Y	N	N	N	N	N
Dominant parent material (6 cat.)	N	N	N	N	Y	N	N	N	N
Subsoil water capacity	N	N	N	N	N	Y	N	N	N
Soil carbon content	N	N	N	N	N	N	Y	N	N
Depth to rock	N	N	N	N	N	N	N	Y	N
Soil differentiation	N	N	N	N	N	N	N	N	Y
R-squared	0.97	0.99	0.23	0.24	0.43	0.41	0.28	0.44	0.31
F-test (H_0 – All instruments zero)	7106.47	21503.0	9.4	7.3	23.1	26.3	11.3	41.2	24.0
Partial R-squared	0.96	0.99	0.03	0.05	0.28	0.26	0.10	0.29	0.14

Dependent variable: ln(market potential). 306 observations for each regression.
All regressions include a constant and three amenity variables (sea, lake, and mountain).
Standard errors in parentheses. *a*, *b*, *c*: corresponding coefficient significant at 1, 5, 10%.

values of Stock and Yogo (2005) with TSLS. To avoid the pitfalls of weak instruments, a number of possible strategies can be envisioned. First, it would be possible to increase the strength of some soil instruments by considering only the more relevant dummies and dropping insignificant ones. In absence of a well articulated theory of how soils affects economic development, we acknowledge an element of 'data mining' in our use of soil characteristics. We are nonetheless reluctant to push it to such extremes. Second, we experiment below with estimation strategies that are less sensitive to weak instruments such as limited information maximum likelihood (LIML) as advocated by Andrews and Stock (2007). Third, we repeat the same regressions with different sets of soil instruments and see how this affects the coefficient(s) of interest. Obtaining the same answer over and over again would be reassuring.

In table 5, we repeat the same exercise with market potential using lagged values of that variable and the same set of soil instruments as in table 4. Both historical and soil variables are much stronger instruments for market potential than for employment density. For historical variables, the reason is that market potential is computed as a weighted mean of employment density. As a result this washes out much idiosyncratic variation and naturally yields higher R-squareds. Put differently, soil variables are better replicating the smooth evolution of market potential than the spikes of employment density. The facts that in column 1 the coefficient on 1831 market potential is essentially one and that the partial R-squared is 95% also indicate that we should not expect much difference between OLS and TSLS below.

Because both market potential and soil characteristics vary smoothly over space, one may worry that the good explanatory power of soil characteristics may be spurious. This will be the case if some large areas with particular soil characteristics spuriously overlap with areas of particularly high or low market potential. However, a detailed reading of the coefficients on soil dummies (not reported in table 5) indicates that this is not the case. For instance, areas for which the dominant parent material is conditionally associated with the lowest market potential are eolian sands, molasse (sand stone), and ferruginous residual clay. Sands, which drain very fast, and ferruginous clay, a heavy soil which does not drain at all, do not lead to very fertile soils. On the other hand, the parent materials associated with a high market potential are loess, a notably fertile type of soil, and chalk, a stable and porous soil which can be very fertile provided it is deep enough. Similarly, a high water capacity of the subsoil is associated with a higher market potential as could be expected.

Instrument exogeneity

Equation (17) gives the second condition that must be satisfied by a valid instrument: orthogonality to the error term. Intuitively, the difficulty in inferring the effect of density

and market potential on wages and TFP arises because of the possibility that a missing local characteristic or some local shocks might be driving both population location and economic outcomes. To overcome this problem, we require instruments which affect wages and TFP only through the spatial distribution of population. That is, as made clear above, we need our instruments to affect the supply of labour, but not directly productivity. We now discuss the *a priori* arguments why our instruments may (or may not) satisfy this exogeneity condition.

We begin with historical variables dating back to 1831. Long-lagged values of the same variable obviously remove any simultaneity bias caused by 'contemporaneous' local shocks. For such simultaneity to remain, we would need these shocks to have been expected in 1831 and have determined population location at the time. This is extremely unlikely. However, endogeneity might also arise because of some missing permanent characteristic that drives both past population location and contemporaneous productivity. A number of first-nature geographic characteristics such as a coastal location may indeed explain both past population location and current economic outcomes. In our regressions we directly control for a number of such first-nature characteristics (coast, mountain, lakes and waterways).

Hence, the validity of long population lags rests on the hypothesis that the drivers of population agglomeration in the past are not related to modern determinants of local productivity after controlling for first-nature characteristics of places. The case for this relies on the fact that the French economy in the late 20th century is very different from what it was in 1831. First, the structure of the French economy in the late 20th century differs a lot from that of 1831. In 1831, France was only starting its industrialisation process, whereas it is de-industrialising now. Manufacturing employment was around 3 million in 1830 against more than 8 million at its peak in 1970 and less than 6 today (Marchand and Thélot, 1997). Then, agriculture employed 63% of the French workforce against less than 5% today. Since 1831, the workforce has also doubled. Second, the production techniques in agriculture, manufacturing and much of the service industries are radically different today from what they were more than 150 years ago. With technological change, the location requirements of production have also changed considerably. For instance, the dependance of manufacturing on sources of coal and iron has disappeared. Third, the costs of shipping goods and transporting people from one location to another have declined considerably. Actually, 1831 coincides with the construction of the first French railroads. Subsequently, cars, trucks and airplanes have further revolutionised transport. At a greater level of aggregation, trade has also become much easier because of European integration over the last 50 years. Fourth, other drivers of population location not directly related to production have changed as well. With much higher standards of living, households are arguably more willing to trade greater efficiency against good

amenities (Rappaport, 2007). Some previously inhospitable parts of the French territory such as its Languedocian coast in the South have been made hospitable and are now developed, etc. Finally, since 1831, France has been ruled by, successively, a king, an emperor, and presidents and prime ministers from 5 different republics. The country also experienced a revolution in 1848, a major war with Germany in 1870, and two world wars during the 20th century.

With so much change, a good case can indeed be made that past determinants of population location are not major drivers of current productivity. As a result, historical variables are the instrument of choice for current population patterns since Ciccone and Hall (1996). They have been widely used by the subsequent literature.

Although the *a priori* case for historical instruments is powerful, nothing guarantees that it is entirely fool-proof. The fact that long lags of the population variables usually pass over-identification tests and other *ex post* diagnostics may not constitute such a strong argument in favour of their validity. Population variables are often strongly correlated with one another so that any permanent characteristics that affects both measures of past population location and contemporaneous productivity may go un-noticed due to the weak power of over-identification tests when the instruments are very similar and thus highly correlated.

We now consider geological characteristics. The *a priori* case for thinking that geological characteristics are good instruments hinges first on the fact that they have been decided mostly by nature and do not result from human activity. This argument applies very strongly to a number of soil characteristics we use. For instance, soil mineralogy and their dominant parent material were determined millions of years ago. Other soil characteristics might seem more suspect in this respect. For instance, a soil's depth to rock or its carbon content might be an outcome of human activity. In the very long-run, there is no doubt that human activity plays a role regarding these two characteristics. Whether recent (in geological terms) economic activity can play an important role is more doubtful (e.g., Guo and Gifford, 2002). A second caveat relates to the measurement of some soil characteristics. In particular, it is hard to distinguish between a soil's intrinsic propensity to erodibility from its actual erosion (see Seybold, Herrick, and Brejda, 1999). In relation to these two worries, our wealth of soil characteristics implies that we can meaningfully compare the answers given by different soil characteristics as instruments in different regressions. We can also use over-identification tests to assess this issue more formally.

Nonetheless, that soils predate patterns of human settlement does not ensure that any soil characteristics will automatically satisfy condition (17) and be valid instruments. More specifically, we expect soil characteristics to have been a major determinant of local labour demand in the past. The main argument for the validity of geological in-

struments is then that soil quality is no longer expected to be relevant in an economy where agriculture represents less than 5% of employment. We also exclude agricultural activities from our data. Put differently, the case for geological characteristics relies on the fact that this important, though partial, determinant of past population location is now largely irrelevant. Hence, like with historical instruments, the *a priori* case for geological instruments is strong but there is no way to be entirely sure.

It is important to note that the cases for the validity of historical and geological variables as instruments differ. Historical variables are 'broad' determinants of current population location. Soil characteristics are narrower but more 'fundamental' determinants of current population location. Put differently, although we expect soils to have determined history, they were not the sole determinants of population patterns in 1831. Geological characteristics also explain current patterns of employment density over and above past employment density. If one group of instruments fails, it is unlikely that the second will do so in the same way. Finally, it is also important to keep in mind that these two sets of instruments can only hope to control for the endogenous quantity of labour bias. That a higher density can lead to the sorting of better workers in these areas is not taken care of by these instruments. Put differently, we expect the endogenous quality of labour bias to remain. Moving from crude measures of wage such W^1 to more sophisticated ones (W^2 and most of all W^3) is designed to tackle this second issue.

5. Main wage results

Table 6 presents the results of three simple regressions for our three wages: W^1, the mean local wage as computed in (10), W^2, the wage index after conditioning out sector effects and observable individual characteristics as estimated in (12), and W^3, the wage index from (13) which also conditions out individual fixed effects. In columns 1, 2, and 3, these three wages are regressed on log employment density controlling for three amenity variables (coast, lakes and waterways, mountain) using OLS. The measured density elasticity of mean wages is at 0.048. This is very close to previous results in the literature (Ciccone and Hall, 1996, Ciccone, 2002). Controlling for sector effects in column 2 implies a marginally higher estimate of 0.051 for the density elasticity and significantly improves the explanatory power of employment density. This suggests that, although the local characteristics of the sector of employment matter, conditioning out sector effects does not affect our estimates of the density elasticity. Controlling also for unobserved individual characteristics yields a significantly lower elasticity of 0.033. This suggests that a good share of measured agglomeration effects are in fact attributable to the unobserved

Table 6: Local wages as a function of density: OLS and historical instruments

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
	W^1 OLS	W^2 OLS	W^3 OLS	W^1 TSLS	W^2 TSLS	W^3 TSLS	W^1 TSLS	W^2 TSLS	W^3 TSLS
ln(density)	0.048	0.051	0.033	0.040	0.042	0.026	0.040	0.044	0.027
	(0.002)^a	(0.002)^a	(0.001)^a	(0.003)^a	(0.002)^a	(0.002)^a	(0.003)^a	(0.002)^a	(0.002)^a
Instruments used:									
ln(1831 density)	-	-	-	Y	Y	Y	Y	Y	Y
ln(1881 density)	-	-	-	-	-	-	Y	Y	Y
First stage statistics	-	-	-	395.7	395.7	395.7	518.7	518.7	518.7
Over-id test p-value	-	-	-	-	-	-	0.99	0.19	0.21
R-squared	0.56	0.72	0.65	-	-	-	-	-	-

306 observations for each regression.
All regressions include a constant and three amenity variables (sea, lake, and mountain).
Standard errors in parentheses. a, b, c: corresponding coefficient significant at 1, 5, 10%.

characteristics of the workforce. More specifically, workers who command a higher wage on labour market sort in denser areas.

In columns 4, 5, and 6, we perform the same regressions as in columns 1, 2, 3 but we instrument employment density with 1831 urban population density. Compared to their corresponding OLS coefficients, the TSLS coefficients for employment density are marginally lower. The instrument is very strong with a first-stage F (or Cragg-Donald) statistic close to 400. In columns 6, 7, and 8, we add 1881 population density as a second instrument for employment density. The results are virtually undistinguishable from those of columns 4, 5, and 6. With two instruments, it is also possible to run Sargan tests of over-identification. They are passed in all three cases with p-values above 10%. However, we can put only a limited weight on this test because the correlation between 1881 and 1831 density is high at 0.75.

If we think of table 6 as our baseline, a number of findings are worth highlighting. The density elasticity of mean wages is 0.048 (column 1). Controlling for the endogenous quality of labour bias through a fixed-effect estimation reduces the size of the coefficient by about a third to 0.033 (column 3). Controlling for the endogenous quantity of labour bias using long historical lags as instruments reduces it by another fifth to 0.027 (column 9). Hence this table provides evidence about both the quality and quantity of labour being simultaneously determined with productivity. It also suggests that the endogenous quality of labour bias is more important than the quantity bias.

Next, table 7 reports results for number of regressions which all use geological characteristics as instruments for employment density. Following the results of table 4, we expect geological instruments to be on the weak side. Furthermore, table 5 also makes clear that geological characteristics appear to explain market potential better than employment

Table 7: Local wages as a function of density: geological instruments

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
	W^1	W^2	W^3	W^3	W^3	W^3	W^3	W^3
	TSLS	TSLS	TSLS	LIML	LIML	LIML	LIML	LIML
ln(density)	0.042	0.047	0.038	0.038	0.048	0.050	0.048	0.047
	(0.010)^a	(0.008)^a	(0.006)^a	(0.006)^a	(0.005)^a	(0.005)^a	(0.005)^a	(0.005)^a
Instruments used:								
Subsoil mineralogy	Y	Y	Y	Y	Y	N	N	N
Ruggedness	Y	Y	Y	Y	N	N	N	N
Depth to rock	N	N	N	N	Y	Y	Y	N
Soil carbon content	N	N	N	N	N	Y	N	N
Topsoil water capacity	N	N	N	N	N	N	Y	Y
Dominant parent material (6 cat.)	N	N	N	N	N	N	N	Y
First stage statistics	6.2	6.2	6.2	6.2	8.3	8.2	8.1	6.8
Over-id test p-value	0.99	0.90	0.67	0.67	0.45	0.12	0.34	0.15

306 observations for each regression.
All regressions include a constant and three amenity variables (sea, lake, and mountain).
Standard errors in parentheses. a, b, c: corresponding coefficient significant at 1, 5, 10%.

density. Hence, we need to keep in mind that our geological instruments are correlated with a variable, market potential, that is (for the time being) missing from the regression and suspected to have an independent effect on wages. As a consequence, IV estimations that rely solely on geological characteristics may not perform very well and should be interpreted with caution.

In each of the regressions in table 7, we use two different soil characteristics. Since, except for ruggedness, each soil characteristic is documented with a series of dummy variables, we could technically run over-identification tests while instrumenting for only one characteristic. However such tests may not be economically meaningful since we would end up testing for over-identification using the particular categorisation of the ESDB. We experimented extensively with soil characteristics. The results we report in the table are representative of what is obtained using any combination of the soil characteristics listed in the table. With them, over-identification tests are usually passed. This is not the case with the other soil characteristics.

More precisely, in column 1 of table 7, we regress mean wages on density and other controls using subsoil mineralogy and ruggedness as instruments for employment density. We obtain a density elasticity of 0.042, which is consistent with what we find in table 6 when we use historical variables. We repeat the same regression in columns 2 and 3 using W^2 and W^3 as dependent variables. In columns 3, the coefficient is slightly above its OLS counterpart rather than slightly below when using historical instruments. The difference is nonetheless not significant. Before going any further, note that the low first-stage statistics in columns 1-3 raises some questions about the strength of these geological instruments. With weak instruments a number of authors (e.g., Stock and Yogo,

Table 8: Local wages as a function of density: historical and geological instruments

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
	W^1	W^2	W^3	W^3	W^3	W^3	W^3	W^3
	TSLS	TSLS	TSLS	GMM	TSLS	TSLS	TSLS	TSLS
ln(density)	0.040	0.042	0.027	0.027	0.027	0.027	0.027	0.027
	(0.003)^a	(0.002)^a	(0.002)^a	(0.002)^a	(0.002)^a	(0.002)^a	(0.002)^a	(0.002)^a
Instruments used:								
ln(1831 density)	Y	Y	Y	Y	Y	Y	Y	Y
Subsoil mineralogy	Y	Y	Y	Y	N	Y	N	N
Ruggedness	N	N	N	N	Y	Y	Y	N
Hydrogeological class	N	N	N	N	N	N	Y	N
Topsoil water capacity	N	N	N	N	N	N	N	Y
First stage statistics	138.7	138.7	138.7	116.2	208.7	108.1	69.8	76.2
Over-id test p-value	0.98	0.83	0.15	0.13	0.31	0.21	0.53	0.02

306 observations for each regression.
All regressions include a constant and three amenity variables (sea, lake, and mountain).
Standard errors in parentheses. a, b, c: corresponding coefficient significant at 1, 5, 10%.

2005) now argue for the superiority of the LIML estimator to the TSLS estimator. Column 4 of table 7 reports the LIML estimate for a specification similar to column 3. The TSLS and LIML results are the same.[22]

In columns 5 to 8, we report LIML results regarding our preferred measure of wages, W^3, for further combinations of instruments. The coefficient on employment density is positive and highly significant in all cases. However, it is above its OLS counterpart rather than below, even more so than in column 4. This discrepancy between the IV results using history in table 6 and those using geology in table 7 is due to the fact that soil variables are not only correlated with the employment density, but also with the market potential, which is missing. As a result, the density elasticities in table 7 may be biased upwards. To see this, note that in column 4 the correlation between the predicted values of employment density obtained from the instrumental regression and actual density is 0.29. The correlation between predicted density and actual market potential (omitted from the regression) is nearly as high at 0.27. In column 5, the problem is even worse since the correlation between predicted and actual density is 0.37 while the correlation between predicted density and market potential is 0.48.[23]

To explore this problem further, we now consider historical and geological instruments at the same time. Table 8 reports the results for a number of regressions using both 1831

[22]In the other regressions, the differences in the point estimates and standard errors between TSLS and LIML remain small. The differences with respect to the over-identification tests are sometimes more important. This is due to the greater power of the Anderson-Rubin test under LIML relative to the Sargan test used with TSLS.

[23]This is consistent with the fact that over-identification tests are passed only for the small set of regressions reported in the table.

Table 9: Local wages as a function of density and (exogenous) market potential: historical and geological instruments

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
	W^1	W^2	W^3	W^3	W^3	W^3	W^3	W^3	W^3
	OLS	OLS	OLS	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS
ln(density)	0.042	0.048	0.026	0.020	0.020	0.020	0.020	0.020	0.020
	(0.003)^a	(0.002)^a	(0.001)^a	(0.002)^a	(0.002)^a	(0.002)^a	(0.002)^a	(0.002)^a	(0.002)^a
ln(market pot.)	0.024	0.012	0.027	0.034	0.034	0.034	0.034	0.034	0.034
	(0.006)^a	(0.004)^a	(0.003)^a	(0.003)^a	(0.003)^a	(0.003)^a	(0.003)^a	(0.003)^a	(0.003)^a
Instruments used:									
ln(1831 density)	-	-	-	Y	Y	Y	Y	Y	Y
Subsoil mineralogy	-	-	-	Y	N	N	N	N	N
Ruggedness	-	-	-	N	Y	N	N	N	N
Subsoil water capacity	-	-	-	N	N	Y	N	N	N
Depth to rock	-	-	-	N	N	N	Y	N	N
Erodibility	-	-	-	N	N	N	N	Y	N
Soil differentiation	-	-	-	N	N	N	N	N	Y
First stage statistics	-	-	-	128.5	191.6	80.9	96.9	77.9	130.0
Over-id test p-value	-	-	-	0.72	0.82	0.42	0.54	0.37	0.10
R-squared	0.59	0.73	0.73	-	-	-	-	-	-

306 observations for each regression.
All regressions include a constant and three amenity variables (sea, lake, and mountain).
Standard errors in parentheses. a, b, c: corresponding coefficient significant at 1, 5, 10%.

density and some soil characteristics. In all cases, the instruments are strong because of the presence of 1831 density. Subsoil mineralogy (along with 1831 density) is used in columns 1 to 3 to instrument for density and explain W^1, W^2, and W^3. The results are the same as those of columns 4-6 of table 6 which use only 1831 density while they differ more with those of columns 1-3 of table 7 which use subsoil mineralogy (together with ruggedness) but not 1831 density. This is unsurprising given that 1831 density is a much stronger instrument. Using a GMM-IV estimation rather than TSLS in column 4 does not change anything. Using ruggedness or hydrogeological class in columns 5-7 also implies a similar coefficient on density. With these three soils characteristics (and 1831 density) the over-identification test is passed. For the other soil characteristics however, this test is failed. An example is given in column 8 with topsoil water capacity. This is in line with the results of the previous table that a majority of soil characteristics do not give the same answer as 1831 density when used as instruments to estimate the density elasticity of wages.

To confirm that this problem is due to the strong correlation between soil characteristics and market potential, table 9 reports results for a number of regressions in which market potential is added as a control. In columns 1-3, our measures of wages W^1, W^2, and W^3 are regressed on density and market potential using OLS. The measured elasticity of wages with respect to market potential is between 0.01 and 0.03. It is also interesting to note

Table 10: Local wages as a function of density and (endogenous) market potential: historical and geological instruments

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
	W^1	W^2	W^3	W^3	W^3	W^3	W^3	W^3	W^3
	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS
ln(density)	0.033	0.040	0.020	0.018	0.019	0.020	0.020	0.020	0.020
	(0.003)^a	(0.003)^a	(0.002)^a	(0.002)^a	(0.002)^a	(0.002)^a	(0.002)^a	(0.003)^a	(0.002)^a
ln(market pot.)	0.034	0.020	0.034	0.048	0.039	0.036	0.036	0.033	0.034
	(0.006)^a	(0.005)^a	(0.003)^a	(0.007)^a	(0.005)^a	(0.005)^a	(0.006)^a	(0.010)^a	(0.007)^a
Instruments used:									
ln(1831 density)	Y	Y	Y	Y	Y	Y	Y	Y	Y
ln(1881 density)	Y	Y	Y	N	N	N	N	N	N
ln(1831 m. pot.)	Y	Y	Y	N	N	N	N	N	N
Erodibility	N	N	N	Y	N	N	N	N	Y
Soil carbon content	N	N	N	Y	Y	N	N	N	N
Subsoil water capacity	N	N	N	N	Y	Y	N	N	N
Depth to rock	N	N	N	N	N	Y	Y	N	N
Ruggedness	N	N	N	N	N	N	Y	Y	N
Soil differentiation	N	N	N	N	N	N	N	Y	Y
First stage statistics	298.0	298.0	298.0	8.3	17.0	23.0	19.8	8.3	10.5
Over-id test p-value	0.57	0.36	0.67	0.62	0.19	0.36	0.54	0.11	0.14

306 observations for each regression.
All regressions include a constant and three amenity variables (sea, lake, and mountain).
Standard errors in parentheses. a, b, c: corresponding coefficient significant at 1, 5, 10%.

that the density elasticity is slightly lower than in column 1-3 of table 6 where market potential is omitted. In columns 4 to 9, we instrument employment density with 1831 density and a range of soil characteristics. The density elasticity is very stable at 0.02 while the market potential elasticity is also very stable at 0.034. Importantly, the over-identification tests are passed (whereas they fail without market potential as a control). More generally, the over-identification test is passed for most combinations of geological instruments and 1831 density. The main systematic failure occurs when the dominant parent material dummies are used. It should be noted that 1831 density is a much stronger instrument and as a result it 'does most of the work' in generating the predicted density at the first stage. This greater strength of past density may explain the stability of the coefficients. Nonetheless, in each of the IV regressions of table 9, at least one soil dummy (and usually more) is significant (and usually highly so). This implies that we can run meaningful over-identification tests. The fact that their p-values is usually well above 10% is strongly suggestive that 1831 density and a broad range of soil characteristics all support this 0.02 estimate for the density elasticity of wages.

Finally in table 10 we consider that market potential could also be endogenous. In columns 1-3, we use only historical instruments: 1831 and 1881 density as well as 1831 market potential. The results for W^3 in column 3 are similar to the IV results in table 8. In columns 4-9, we use 1831 employment density with, in each regression, two different

soils characteristics among erodibility, carbon content, subsoil water capacity, depth to rock, ruggedness, and soil differentiation. The over-identification test is always passed in the table. Although the results are not reported here, this test is also passed for all the other pairwise combinations of these characteristics (except the combination soil differentiation and carbon content for which the test marginally fails). For our preferred concept of wage, W^3, the coefficients on density and market potential are very stable and confirm the estimates of column 3 with historical instruments and those of the previous table where market potential is taken to be exogenous. This stability across columns 3-9 is interesting because geological variables and past density in columns 4-9 are not as strong sets of instrument as the combination of past density and past market potential. Our preferred estimate for the elasticity of wages with respect to employment density is 0.02. With respect to market potential, our preferred estimate is at 0.034.

While regressing mean wages on employment density leads to a measured elasticity of 0.05, adding further controls and correcting for the endogenous quality and quantity of labour biases bring this number down to about 0.02.

6. TFP

Firm and establishment data

To construct our establishment-level data, we proceed as follows. We first put together two firm-level data sets: the BRN ('Bénéfices Réels Normaux') and the RSI ('Régime Simplifié d'Imposition'). The BRN contains the balance sheet of all firms in the traded sectors with a turnover above 730,000 euros. The RSI is the counterpart of the BRN for firms with a turnover below 730,000 euros. Although the details of the reporting differs, for our purpose these two data sets contain essentially the same information. Their union covers nearly all French firms.

For each firm we have a firm identifier and detailed annual information about its output and its consumption of intermediate goods and materials. This allows us to construct a reliable measure of value added. To estimate TFP (see below), we use a measure of capital stock based on the sum of the reported book values of productive and financial assets.[24] We also experimented with TFP estimations using the cost of capital rather than assets values following the detailed methodology developed by Boutin and Quantin (2006).

[24]In this respect, we proceed like Syverson (2004). Nevertheless, valuing assets at their historical costs is not without problems. We minimise them by estimating TFP at the three-digit level with 114 sectors. Indeed, the capital stocks of firms within the same sector are likely to be of the same vintage when sectors are more narrowly defined. We also use year dummies. An alternative would be to deflate assets using economic criteria. However, our panel is rather short which makes it difficult to trace the original investments. Our procedure also differs from that of Olley and Pakes (1996) who use a permanent inventory method.

Since firms can have many establishments at many locations, we also use the SIREN data ('Système d'Identification du Répertoire des ENtreprises'). It is an exhaustive registry of all establishments in the traded sectors. For each establishment and year, SIREN reports both a firm and an establishment identifier, a municipality code, and total employment. Note finally that BRN, RSI, and SIREN only report total employment and not hours worked.

To obtain information about hours, we return to the DADS which report them after 1993. Hence, for 1994-2002, we use another, this time exhaustive, DADS dataset.[25] Using the individual information about hours and two-digit occupation that this source contains, we can aggregate it at the establishment level to obtain the hours for all employees and by skill group. We emphasise this because of the suspected importance of labour quality. To avoid estimating too many coefficients for different types of labour, we aggregate two-digit occupational categories into 3 groups: high-, intermediate- and low-skill workers following the classification of Burnod and Chenu (2001).

To merge these four data sets, we extend the procedure of Aubert and Crépon (2003). At the establishment level, we first match SIREN with DADS using the establishment identifier present in both datasets. This establishment-level data set (sector and hours by skill group) is needed below to create a number of local characteristics. Next, we aggregate this establishment dataset at the firm level using the firm identifier. Finally, we merge this firm data with RSI and BRN to recover firm-level information. For each firm between 1994 and 2002, we end up with its value added, the value of its assets, and total hours worked by establishment and skill group. The total number of observations for 1994 is 942,506. This number rises slowly over the period.

Finally, and to avoid dealing with the complications of TFP estimation for multi-establishment firms for which capital and output are known only at the firm level, we restrict our attention to single-establishment firms to estimate TFP.[26] Because the information about very small firms tends to be noisy, we only retain firms with more than 5 employees.

Constructing area-year measures of TFP

We now turn to TFP and start by constructing productivity measures for each employment area and year from TFP regressions. We estimate TFP for 114 sectors separately. For simplicity, we ignore sector subscripts for the coefficients. For firm i in a given sector,

[25] Unfortunately this data cannot be used for our wage regressions because the different years have not been linked up.

[26] With multi-establishment firms, we need to impute the same residual estimated from a firm-level production function to all establishments of the same firm. This is a strong assumption that we would rather not make. In results not reported here, we nonetheless experimented with TFP estimated from multi-establishment firms.

its value-added va_{it} is specified as:

$$\ln va_{it} = \alpha \ln k_{it} + \beta \ln l_{it} + \sum_m \beta_m^S q_{imt} + \phi_t + \varepsilon_{it}, \tag{18}$$

where k_{it} is the capital of firm i, l_{it} its labour (in hours), q_{imt} the share of labour hours in skill group m, ϕ_t a year fixed effect, and ε_{it} an error term measuring firm TFP. The way we introduce skill shares is justified in Hellerstein, Neumark, and Troske (1999).

Three important issues are worth highlighting at this stage. First, we face the same problem as with wages regarding input quality and more particularly labour. Unfortunately, workers characteristics are typically scarce in firm- or establishment-level data. We use the strategy used in (18) based on occupational categories to control for labour quality.[27] This is obviously a less powerful set of controls than the individual fixed effects used in the wage regressions above.

Second, we can hope to control for the two main factors of production, capital and labour, but not for other factors, land in particular.[28] As argued above, the price of land is expected to affect the consumption of land and thus production while, at the same time, be correlated with other local characteristics. Again, instrumenting for these local characteristics is the solution we consider here. Furthermore, output prices are unobserved and are likely to be correlated with local characteristics as well. To the extent that we think of our work as looking into the determinants of local value added rather than pure productivity, this need not bother us much here.[29]

The third issue about TFP estimation is related to the fact that input choices are expected to be endogenous. This issue has received a lot of attention in the literature (see Ackerberg, Caves, and Frazer, 2006, for a recent contribution). For our purpose, this endogeneity bias matters only when it differs across areas. Our main TFP results were estimated using Olley and Pakes (1996). See Appendix A for details about the OP approach. This approach allows us to recover r_{it}, an estimator of ε_{it}. We then average it within sectors, areas, and years:

$$r_{ast} \equiv \frac{1}{L_{ast}} \sum_{i \in (a,s,t)} l_{i,t} r_{i,t}, \tag{19}$$

where $L_{ast} \equiv \sum_{i \in (a,s,t)} l_{i,t}$ is the total number of hours worked in area a, sector s, and year t. A first measure of the local productivity of the average firm in area a and year t denoted

[27] An obvious way to deal with the unobserved quality of the workforce is to use fixed-effects but unfortunately their use is often problematic with firm-level data because of the sluggish adjustment of capital. See Fox and Smeets (2007) for a more thorough attempt to take (observable) input quality into account when estimating TFP. Like us they find that measures of labour quality are highly significant but taking labour quality into account does not reduce the large dispersion of TFP across firms.

[28] We also expect the functional form to matter although we limit ourselves to simple specifications here.

[29] In a different context where one is interested in distinguishing between price and productivity effects, such benign neglect may not be warranted. See for instance Combes, Duranton, Gobillon, Roux, and Puga (2007). Note that this issue also applies to wages.

TFP^1_{at} is obtained by averaging equation (19) across sectors within areas and years with weights equal to the number of firms:

$$\text{TFP}^1_{at} \equiv \frac{1}{n_{at}} \sum_{s \in (a,t)} n_{ast} r_{ast}, \tag{20}$$

where n_{ast} and n_{at} are the total numbers of firms for area a, sector s, and year t and for area a and year t, respectively.

This measure of TFP does not control for the local sector structure. To control for the fact that high productivity sectors may have a propensity to locate in particular areas, we regress r_{ast} on a full set of sector fixed effect, γ_s:

$$r_{ast} = \gamma_s + \iota_{ast}. \tag{21}$$

This equation is estimated with WLS where the weights are the numbers of establishments associated with each observation.[30] To estimate a productivity index TFP^2_{at}, we average the estimated residuals of (21) for each area and year:

$$\text{TFP}^2_{at} \equiv \frac{1}{n_{at}} \sum_{s \in (a,t)} n_{ast} \widehat{\iota}_{ast}. \tag{22}$$

TFP^2_{at} can thus be interpreted as a productivity index net of sector effects.

We finally compute a third local productivity index TFP^3_{at} controlling for variables at the area and sector level, X_{ast}. For that purpose, we estimate the equation:

$$r_{ast} = \text{TFP}^3_{at} + \gamma_s + X_{ast}\varphi + \epsilon_{ast}. \tag{23}$$

This equation is estimated with WLS where weights are once again the numbers of establishments associated with each observation. It mimics equation (12) for wages and uses the same (centred) local characteristics (same sector specialisation, number of firms, share of professionals, average age, and average squared-age). The main difference however is that these characteristics are constructed using the TFP data and not the wage data.

For comparison, we also estimate equation (18) with OLS. Denote $\widehat{\varepsilon}_{it}$ the estimated residual for firm i. We then define:

$$r^{\text{OLS}}_{ast} \equiv \frac{1}{L_{ast}} \sum_{i \in (a,s,t)} l_{it} \widehat{\varepsilon}_{it}, \tag{24}$$

the OLS counterpart to (19). It is possible to recompute our three measures of local productivity TFP^1_{at}, TFP^2_{at}, and TFP^3_{at} using (24) rather than (19). Below we compare the

[30] These weights give more importance to sectors and areas for which a larger number of $r_{i,t}$ are considered when constructing $r_{s,a,t}$. For these area-sector-years, the sampling error on $r_{s,a,t}$ is usually smaller. Weighing should thus reduce the impact of the sampling error on the dependent variable that comes from the first-stage estimation.

coefficients in our main regressions using local productivity indices computed from OP and OLS.

One aspect of the simultaneity bias at the area level is that establishments may produce more and grow larger in areas where the local productivity is higher. It is possible to control for that by introducing area and year fixed effects g_{at} in equation (18):

$$\ln va_{it} = \alpha \ln k_{it} + \beta \ln l_{it} + \sum_m \beta_m^S q_{imt} + \phi_t + g_{at} + \varepsilon_{it}. \tag{25}$$

This equation is estimated with OLS. Since this equation is estimated for each sector, the area-year fixed effects depend on the sector and can be rewritten with a sector subindex, g_{ast}. We can then define $r_{ast}^{\text{FE}} \equiv g_{ast}$, the fixed effect counterpart to (19) and (24), and construct once again our three measures of local productivity.[31]

Finally, we average our estimates across years as we did for wages to avoid identifying out of the temporal variation.[32] Before going to our results, note that our local productivity variables are strongly correlated with one another. Using OP estimates, the correlation between TFP^1 and TFP^2 is 0.93, the correlation between TFP^1 and TFP^3 is also 0.93, and the correlation between TFP^2 and TFP^3 is 0.98. For TFP^3, the correlation between OP and OLS estimates is 0.96, the correlation between OP and fixed effects estimates is 0.91, and the correlation between OLS and fixed effects is also 0.91. Finally the correlation between TFP^3 estimated with OP in and mean wages (W^1) is 0.77.[33] This correlation raises to 0.88 after correcting wages of sector effects (W^2) or to 0.87 after correcting wages of sector and worker effects (W^3).

Results

Table 11 presents the results of three regressions for our three measures of local OP productivity: TFP^1, the mean productivity computed in (20), TFP^2, the local productivity controlling for sector fixed-effects as estimated in (21), and TFP^3, the local productivity estimated in (23) which conditions out a broader set of sector effects. This table mirrors the 'wage' table 6 for productivity. In columns 1, 2, and 3, these three measures of local productivity are regressed on log employment density controlling for amenities using OLS. The mean elasticity of TFP with respect to density is at 0.04 for mean productivity, 0.041 when taking out sector effects, and 0.047 when also controlling for the local sector structure. In columns 4, 5, and 6, we instrument employment density with 1831 urban

[31] We also experimented with a number of alternative TFP approaches such as GMM, cost shares, IV cost shares, Levinsohn and Petrin (2003), etc.

[32] Like with wages, these averages are now unweighted.

[33] Recall that the years over which TFP and wages are computed are not the same.

Table 11: Local TFP (OLLEY-PAKES) as a function of density: OLS and historical instruments

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
	TFP^1 OLS	TFP^2 OLS	TFP^3 OLS	TFP^1 TSLS	TFP^2 TSLS	TFP^3 TSLS	TFP^1 TSLS	TFP^2 TSLS	TFP^3 TSLS
ln(density)	0.040	0.041	0.047	0.031	0.034	0.038	0.033	0.035	0.039
	(0.002)^a	(0.002)^a	(0.002)^a	(0.003)^a	(0.002)^a	(0.002)^a	(0.002)^a	(0.002)^a	(0.002)^a
Instruments used:									
ln(1831 density)	-	-	-	Y	Y	Y	Y	Y	Y
ln(1881 density)	-	-	-	-	-	-	Y	Y	Y
First stage statistics	-	-	-	371.4	371.4	371.4	429.1	429.1	429.1
Over-id test p-value	-	-	-	-	-	-	0.07	0.18	0.37
R-squared	0.63	0.70	0.75	-	-	-	-	-	-

306 observations for each regression.
All regressions include a constant and three amenity variables (sea, lake, and mountain).
Standard errors in parentheses. a, b, c: corresponding coefficient significant at 1, 5, 10%.

population density. The TSLS coefficients for employment density are marginally lower than in columns 1, 2, and 3. In columns 7, 8, and 9, we add 1881 population density to instrument for contemporaneous employment density. Although the Sargan test of over-identification marginally fails in column 7 with a p-value of 7%, the results are very close to those of columns 4, 5, and 6.

Comparing these results to those of table 6 for wages, we note the following. First, instrumenting for contemporaneous employment density with deep historical lags lowers the coefficients in roughly the same proportion in both cases. This confirms our finding of a mild simultaneity bias regarding the quantity of labour. Second, controlling for sector effects in TFP^3 compared to TFP^1 raises the coefficient on employment density just like it does when considering W^2 instead of W^1 (although the increase is slightly more important here).[34] A stronger effect of density after conditioning out sector effects is consistent with the notion that sectors located in less dense areas may gain less from overall density, and perhaps more from same sector specialisation or another sector characteristics that is conditioned out in TFP^3.[35] Third, it is also interesting to note that, when a direct comparison is possible, the density elasticities for wages tend to be above those for TFP. From the theoretical framework developed above (and particularly equations 5 and 6), we actually expect the coefficients on employment density to be higher for wages by a factor equal to the inverse of the labour share ($\frac{1}{1-\alpha}$). With labour coefficients typically between 0.5 and 0.75, the difference between the two sets of estimates is of the right magnitude, although a bit smaller than expected.

[34]While TFP^1 may be taken to be the counterpart of W^1, TFP^3 corresponds to W^2. Because we cannot control for input quality well, there is no TFP concept that corresponds to W^3.

[35]This higher coefficient on density with TFP^3 is also consistent with possible correlations between unobserved input quality and the local structure of production.

Table 12: Local TFP (OLLEY-PAKES) as a function of density: geological instruments

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
	TFP^1 TSLS	TFP^2 TSLS	TFP^3 TSLS	TFP^3 LIML	TFP^3 LIML	TFP^3 LIML	TFP^3 LIML	TFP^3 LIML
ln(density)	0.054	0.041	0.045	0.045	0.048	0.047	0.045	0.046
	(0.009)^a	(0.007)^a	(0.007)^a	(0.007)^a	(0.005)^a	(0.005)^a	(0.007)^a	(0.005)^a
Instruments used:								
Subsoil mineralogy	Y	Y	Y	Y	Y	N	N	N
Ruggedness	Y	Y	Y	Y	N	N	N	N
Depth to rock	N	N	N	N	N	Y	N	N
Soil carbon content	N	N	N	N	N	Y	Y	N
Topsoil water capacity	N	N	N	N	N	N	Y	Y
Dominant parent material (6 cat.)	N	N	N	N	Y	N	N	Y
First stage statistics	5.5	5.5	5.5	5.5	5.9	7.4	5.2	6.0
Over-id test p-value	0.58	0.60	0.38	0.38	0.26	0.19	0.15	0.22

306 observations for each regression.
All regressions include a constant and three amenity variables (sea, lake, and mountain).
Standard errors in parentheses. a, b, c: corresponding coefficient significant at 1, 5, 10%.

To assess the sensitivity of our results to the approach used to estimate TFP, we reproduce in table 14 of Appendix B some of the regressions of table 11 using alternative local productivity indices. These measures of local TFP are constructed from the OLS estimates of (18) and from (25) which computes local productivity fixed effects. When TFP is estimated with OLS instead of OP, the coefficients on density are close, though not exactly the same.[36] When TFP is estimated with local fixed effects instead of OP, we find lower coefficients on density. At this stage, our best estimate of the density elasticity of TFP is at 0.04.[37]

Turning to geological instruments, table 12 mirrors for TFP what table 7 does for wages.[38] Columns 1-3 use subsoil mineralogy and ruggedness to instrument for employment density using our three measures of TFP as dependent variables. The coefficients on density are higher than with historical instruments in table 11. Such a difference between

[36]When TFP is estimated with OP we must drop the first year of data and firms with no investment. Estimating TFP with OLS on the same sample of firms as with OP makes no difference with respect to OLS estimates of local productivity.

[37]Comparing these results to the main study about agglomeration effects using TFP data in the literature, Henderson (2003), is not easy. First, Henderson (2003) uses very different US data for which value added cannot be measured directly and focuses on five industries only. Second, he focuses on sector effects and uses as key independent variable the number of plants in the local industry. We focus instead on total local employment conditioning out local industry shares (among others) in some TFP measures. Third, he estimates TFP and the effects of local characteristics in one stage using a different specification for productivity, which includes firm fixed effects. Finally, he tackles endogeneity problems using a GMM approach. Despite these differences, his findings of strong heterogeneity across industries and modest to high scale effects at the industry level are consistent with ours.

[38]That is, aside from the difference in dependent variables, the regressions are exactly the same. The values taken by employment density differ very slightly because of the differences in years between the wage and TFP data and the difference in data source.

geological and historical instruments is also observed with wages.[39] To repeat, it reflects the fact that geological instruments have a larger correlation with market potential than local density. The results of columns 3 are confirmed in column 4 when LIML rather than TSLS is used and in column 5-8 when different sets of instruments are used. It is interesting to note that the over-identification tests are passed for the same specifications as with wages (and they also fail for the same unreported regressions as well).

To mirror again the analysis performed with wages, table 15 of Appendix B performs the regressions of table 8 with TFP rather than wages using historical and geological instruments at the same time. The results are again extremely consistent with the wage results. The coefficients on employment density in table 15 with both sets of instruments are the same as those that use historical instruments only in table 11. This near-equality also holds with wages. Furthermore, over-identification tests appear to be passed or failed with the same combinations of instruments again. An exception is column 8 with dominant parent material and topsoil water capacity. The test is passed with wages with a p-value of 15%, while it is failed with TFP (p-value of 5%).

In table 16 of Appendix B we add market potential as explanatory variable as we do with wages in table 9. We again use the exact same specifications as with wages. Adding market potential to the OLS specifications lowers the coefficient on employment density for TFP. The elasticity of TFP with respect to market potential is about half the density elasticity. These two results closely mirror what happens in our wage regressions when we add market potential as an explanatory variable. In the second part of table 16 we instrument employment density with 1831 density and a range of soil characteristics. The coefficient on density declines by about 0.01 point to 0.033 while that on market potential increases by about the same amount to 0.027. This again is very close to what happens in the wage regressions. Interestingly, the same combinations of instruments pass the over-identification tests with both wages and TFP. The failure of the Sargan test in the last column of table 16 is an exception.

Finally, in table 13, market potential is also assumed to be endogenous. As with wages in table 10, we instrument density and market potential with historical and soil variables. The main result is that instrumenting for market potential leaves its coefficient unchanged. The IV coefficient on employment density is also unchanged. This is the same outcome as with wages. In tables 17 and 18 of Appendix B, we repeat the same exercise but use TFP indices estimated with OLS and with local fixed effects as in equation (25). As in previous comparisons, the results for OLS and OP TFP are very close. With (local) fixed-effect TFP, the density and market potential elasticities are lower than with OLS and OP TFP. However

[39] As previously, the coefficients on density are also slightly above those obtained with wages for similar regressions.

Table 13: Local TFP (OLLEY-PAKES) as a function of density and (endogenous) market potential: historical and geological instruments

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
	TFP[1] TSLS	TFP[2] TSLS	TFP[3] TSLS	TFP[3] TSLS	TFP[3] TSLS	TFP[3] TSLS	TFP[3] TSLS	TFP[3] TSLS	TFP[3] TSLS
ln(density)	0.028	0.030	0.035	0.034	0.034	0.034	0.034	0.034	0.035
	(0.003)[a]	(0.002)[a]	(0.002)[a]	(0.003)[a]	(0.003)[a]	(0.003)[a]	(0.003)[a]	(0.004)[a]	(0.003)[a]
ln(market pot.)	0.025	0.027	0.026	0.021	0.023	0.021	0.022	0.024	0.018
	(0.005)[a]	(0.004)[a]	(0.004)[a]	(0.009)[b]	(0.007)[a]	(0.006)[a]	(0.008)[a]	(0.013)[c]	(0.009)[b]
Instruments used:									
ln(1831 density)	Y	Y	Y	Y	Y	Y	Y	Y	Y
ln(1881 density)	Y	Y	Y	N	N	N	N	N	N
ln(1831 m. pot.)	Y	Y	Y	N	N	N	N	N	N
Erodibility	N	N	N	Y	N	N	N	N	Y
Soil carbon content	N	N	N	Y	Y	N	N	N	N
Subsoil water capacity	N	N	N	N	Y	Y	N	N	N
Depth to rock	N	N	N	N	N	Y	Y	N	N
Ruggedness	N	N	N	N	N	N	Y	Y	N
Soil differentiation	N	N	N	N	N	N	N	Y	Y
First stage statistics	230.6	230.6	230.6	8.2	16.4	22.8	20.2	8.0	10.5
Over-id test p-value	0.16	0.43	0.17	0.68	0.90	0.60	0.29	0.04	0.16

306 observations for each regression.
All regressions include a constant and three amenity variables (sea, lake, and mountain).
Standard errors in parentheses. a, b, c: corresponding coefficient significant at 1, 5, 10%.

we observe the same stability in the coefficients across regressions. This suggests that the method used to estimate TFP matters with respect to the point estimates for the density and market potential elasticities (though by only 0.01). However, the choice of TFP estimation does not matter otherwise.

7. Conclusions

We revisit the estimation of local scale effects using large scale French wage and TFP data. To deal with the 'endogenous quantity of labour' bias (i.e., urban agglomeration is consequence of high local productivity rather than a cause), we take an instrumental variable approach and introduce a new set of geological instruments in addition to standard historical instruments. To deal with the 'endogenous quality of labour' bias (i.e., cities attract skilled workers so that the effects of skills and urban agglomeration are confounded), we take a worker fixed effect approach.

Our first series of findings relates to the endogenous quantity of labour bias. (i) Long lags of our endogenous explanatory variables make for strong instruments. (ii) Geological characteristics are more complicated instruments to play with. (iii) Nevertheless, geological and historical instruments lead to similar answers once the regression is properly

specified: The simultaneity problem between employment density and local wages / productivity is relatively small. It reduces the impact of density by around a fifth.

Our second finding relates to the endogenous quality of labour bias. (iv) Better workers are located in more productive areas. This sorting of workers by skills (observed and unobserved) is quantitatively more important than the endogenous quantity of labour bias. (v) In our regressions, we address sorting using the panel dimension of our wage data. The density elasticity is divided by almost 2. Applying this type approach to TFP is problematic. We thus put more weight on our wage results than we do on our TFP results. Nonetheless, the high degree of consistency between wage and TFP results is reassuring.

We believe the priority for future work should be to develop more sophisticated approaches to deal with the sorting of workers across places. Awaiting progress on this issue, our preferred estimates for the elasticity of wages to density is at 0.02 and around 0.035 for the density elasticity of TFP. For market potential, we find elasticities around 0.035 for wages and 0.025 for TFP. Finally, our result about the relative importance of the two biases raises an interesting question. To which extend does it reflect particular features of the French housing and labour market institutions? One may imagine that in a country like the US with greater labour mobility and a much flatter housing supply curve (in at least part of the country), the endogenous quantity of labour bias might dominate. Further research should inform this question.

Appendix A. Implementation of Olley and Pakes (1996)

The error term in (18) is rewritten as $\varepsilon_{it} \equiv v_{it} + \xi_{it}$ where v_{it} is the part of the error term that influences the decision of the firm regarding its factors and ξ_{it} is an independent noise. The crucial assumption is that capital investment, I_{it}, can be written as a function of the error term, v_{it}, and current capital: $I_{it} \equiv f_t(k_{it}, v_{it})$ with $\partial f_t / \partial v_{it} > 0$. The investment function can be inverted to yield: $v_{it} = f_t^{-1}(k_{it}, I_{it})$. Equation (18) can then be rewritten as:

$$\ln va_{it} = \alpha \ln k_{it} + \beta \ln l_{it} + \sum_m \beta_m^S q_{imt} + \phi_t + f_t^{-1}(k_{it}, I_{it}) + \xi_{it}. \tag{A1}$$

This equation can be estimated in two stages. Denote $\Phi_t(k_{it}, I_{it}) \equiv f_t^{-1}(k_{it}, I_{it}) + \alpha \ln k_{it} + \phi_t$. Equation (A1) becomes:

$$\ln va_{it} = \beta \ln l_{it} + \sum_m \beta_m^S q_{imt} + \Phi_t(k_{it}, I_{it}) + \xi_{it}. \tag{A2}$$

This equation can be estimated with OLS after approximating $\Phi_t(k_{it}, I_{it})$ with a third-order polynomial, crossing k_{it}, I_{it}, and year dummies. Its estimation allows us to recover some estimators for the labour and skill share coefficients ($\widehat{\beta}$ and $\widehat{\beta}_m^S$). It is then possible to construct $z_{it} \equiv \ln va_{it} - \widehat{\beta} \ln l_{it} - \sum_m \widehat{\beta}_m^S q_{imt}$. Furthermore, the error v_{it} is rewritten as the projection on its lag and an innovation: $v_{it} \equiv h(v_{it-1}) + \zeta_{it-1}$. Using $v_{it-1} = f_{t-1}^{-1}(k_{it-1}, I_{it-1}) = \Phi_{t-1}(k_{it-1}, I_{it-1}) - \alpha \ln k_{j,t-1} - \phi_{t-1}$, the value-added equation then becomes:

$$z_{it} = \alpha \ln k_{it} + \phi_t + h\left(\widehat{\Phi}(k_{it-1}, I_{it-1}) - \alpha \ln k_{jt-1} - \phi_{t-1}\right) + \psi_{it}, \tag{A3}$$

where ψ_{it} is a random error. The function $h(.)$ is approximated by a third-order polynomial and equation (A3) is estimated with non-linear least squares. It allows us to recover some estimators of the capital coefficient $\widehat{\alpha}$ and the year dummies $\widehat{\phi}_t$. Firm TFP is then defined as $r_{it} \equiv z_{it} - \widehat{\alpha} \ln k_{it} - \widehat{\phi}_t$. It is an estimator of ε_{it}. For further details about the implementation procedure in stata used in our paper, see Arnold (2005).

Although the OP method allows us to control for simultaneity, it has some drawbacks. In particular, we need to construct investment from the data: $I_{it} = k_{it} - k_{it-1}$. As a consequence it can be computed only for firms that are present in two consecutive years. Other observations must be dropped. Furthermore, the investment equation $I_{it} = f_t(k_{it}, v_{it})$ can be inverted only if $I_{it} > 0$. Hence, we can keep only observations for which $I_{it} > 0$. This double selection may introduce a bias, for instance, if (i) there is greater 'churning' (i.e. entry and exits) in denser areas, and (ii) age and investment affect productivity positively. Then, more establishments with a low productivity may be dropped in high density areas. In turn, this may increase the measured difference in local productivity between areas of low and high density. Re-estimating OLS TFP on the same sample of firms used for OP shows that this is, fortunately, not the case on French data.

Appendix B. Further results

Table 14: Local TFP (OLS and FIXED-EFFECTS) as a function of density: OLS and historical instruments

Variable	TFP estimated with OLS [1]	[2]	[3]	[4]	TFP with fixed effects [5]	[6]	[7]	[8]
	TFP^1 OLS	TFP^3 OLS	TFP^1 TSLS	TFP^3 TSLS	TFP^1 OLS	TFP^3 OLS	TFP^1 TSLS	TFP^3 TSLS
ln(density)	0.035	0.049	0.029	0.042	0.027	0.040	0.018	0.033
	(0.002)^a	(0.002)^a	(0.002)^a	(0.002)^a	(0.002)^a	(0.002)^a	(0.002)^a	(0.002)^a
Instruments used:								
ln(1831 density)	-	-	Y	Y	-	-	Y	Y
ln(1881 density)	-	-	Y	Y	-	-	Y	Y
First stage statistics	-	-	432.3	432.3	-	-	432.3	432.3
Over-id test p-value	-	-	0.21	0.10	-	-	0.85	0.75
R-squared	0.63	0.75	-	-	0.45	0.67	-	-

306 observations for each regression.
All regressions include a constant and three amenity variables (sea, lake, and mountain).
Standard errors in parentheses. a, b, c: corresponding coefficient significant at 1, 5, 10%.

Table 15: Local TFP (OLLEY-PAKES) as a function of density: historical and geological instruments

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
	TFP^1 TSLS	TFP^2 TSLS	TFP^3 TSLS	TFP^3 TSLS	TFP^3 GMM	TFP^3 TSLS	TFP^3 TSLS	TFP^3 TSLS
ln(density)	0.031	0.034	0.038	0.038	0.039	0.039	0.039	0.038
	(0.003)^a	(0.002)^a	(0.002)^a	(0.002)^a	(0.002)^a	(0.002)^a	(0.002)^a	(0.002)^a
Instruments used:								
ln(1831 density)	Y	Y	Y	Y	Y	Y	Y	Y
Subsoil mineralogy	Y	Y	Y	Y	N	Y	N	N
Ruggedness	N	N	N	N	Y	Y	Y	N
Hydrogeological class	N	N	N	N	N	N	Y	N
Topsoil water capacity	N	N	N	N	N	N	N	Y
First stage statistics	129.7	129.7	129.7	103.3	194.1	100.3	64.5	125.7
Over-id test p-value	0.14	0.56	0.78	0.66	0.14	0.46	0.33	0.05

306 observations for each regression.
All regressions include a constant and three amenity variables (sea, lake, and mountain).
Standard errors in parentheses. a, b, c: corresponding coefficient significant at 1, 5, 10%.

Table 16: Local TFP (OLLEY-PAKES) as a function of density and (exogenous) market potential: historical and geological instruments

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
	TFP[1] OLS	TFP[2] OLS	TFP[3] OLS	TFP[3] TSLS	TFP[3] TSLS	TFP[3] TSLS	TFP[3] TSLS	TFP[3] TSLS	TFP[3] TSLS
ln(density)	0.036	0.037	0.043	0.033	0.033	0.034	0.033	0.033	0.033
	(0.002)[a]	(0.002)[a]	(0.002)[a]	(0.003)[a]	(0.003)[a]	(0.003)[a]	(0.003)[a]	(0.003)[a]	(0.003)[a]
ln(market pot.)	0.017	0.019	0.017	0.027	0.027	0.027	0.028	0.027	0.028
	(0.004)[a]	(0.004)[a]	(0.004)[a]	(0.004)[a]	(0.004)[a]	(0.004)[a]	(0.004)[a]	(0.004)[a]	(0.004)[a]
Instruments used:									
ln(1831 density)	-	-	-	Y	Y	Y	Y	Y	Y
Subsoil mineralogy	-	-	-	Y	N	N	N	N	N
Ruggedness	-	-	-	N	Y	N	N	N	N
Subsoil water capacity	-	-	-	N	N	Y	N	N	N
Depth to rock	-	-	-	N	N	N	Y	N	N
Erodibility	-	-	-	N	N	N	N	Y	N
Soil differentiation	-	-	-	N	N	N	N	N	Y
First stage statistics	-	-	-	115.0	170.6	71.4	86.6	70.7	116.1
Over-id test p-value	-	-	-	0.27	0.50	0.69	0.32	0.42	0.05
R-squared	0.64	0.72	0.76	-	-	-	-	-	-

306 observations for each regression.
All regressions include a constant and three amenity variables (sea, lake, and mountain).
Standard errors in parentheses. a, b, c: corresponding coefficient significant at 1, 5, 10%.

Table 17: Local TFP (OLS) as a function of density and (endogenous) market potential: historical and geological instruments

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
	TFP[1] TSLS	TFP[2] TSLS	TFP[3] TSLS	TFP[3] TSLS	TFP[3] TSLS	TFP[3] TSLS	TFP[3] TSLS	TFP[3] TSLS	TFP[3] TSLS
ln(density)	0.023	0.022	0.037	0.035	0.035	0.036	0.037	0.036	0.035
	(0.002)[a]	(0.002)[a]	(0.002)[a]	(0.003)[a]	(0.003)[a]	(0.003)[a]	(0.003)[a]	(0.004)[a]	(0.003)[a]
ln(market pot.)	0.030	0.030	0.024	0.022	0.026	0.019	0.012	0.020	0.020
	(0.004)[a]	(0.004)[a]	(0.004)[a]	(0.010)[b]	(0.007)[a]	(0.007)[a]	(0.008)	(0.013)	(0.009)[b]
Instruments used:									
ln(1831 density)	Y	Y	Y	Y	Y	Y	Y	Y	Y
ln(1881 density)	Y	Y	Y	N	N	N	N	N	N
ln(1831 m. pot.)	Y	Y	Y	N	N	N	N	N	N
Erodibility	N	N	N	Y	N	N	N	N	Y
Soil carbon content	N	N	N	Y	Y	N	N	N	N
Subsoil water capacity	N	N	N	N	Y	Y	N	N	N
Depth to rock	N	N	N	N	N	Y	Y	N	N
Ruggedness	N	N	N	N	N	N	Y	Y	N
Soil differentiation	N	N	N	N	N	N	N	Y	Y
First stage statistics	232.2	232.2	232.2	8.2	16.4	22.9	20.2	8.0	10.5
Over-id test p-value	0.54	0.56	0.04	0.73	0.29	0.05	0.14	0.38	0.87

306 observations for each regression.
All regressions include a constant and three amenity variables (sea, lake, and mountain).
Standard errors in parentheses. a, b, c: corresponding coefficient significant at 1, 5, 10%.

Table 18: Local TFP (FIXED EFFECTS) as a function of density and (endogenous) market potential: historical and geological instruments

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
	TFP[1] TSLS	TFP[2] TSLS	TFP[3] TSLS	TFP[3] TSLS	TFP[3] TSLS	TFP[3] TSLS	TFP[3] TSLS	TFP[3] TSLS	TFP[3] TSLS
ln(density)	0.011	0.013	0.028	0.030	0.029	0.030	0.030	0.030	0.030
	(0.002)[a]	(0.002)[a]	(0.002)[a]	(0.003)[a]	(0.003)[a]	(0.003)[a]	(0.003)[a]	(0.004)[a]	(0.003)[a]
ln(market pot.)	0.032	0.027	0.022	0.013	0.015	0.011	0.008	0.012	0.010
	(0.004)[a]	(0.004)[a]	(0.004)[a]	(0.009)	(0.007)[b]	(0.007)	(0.008)	(0.013)	(0.009)
Instruments used:									
ln(1831 density)	Y	Y	Y	Y	Y	Y	Y	Y	Y
ln(1881 density)	Y	Y	Y	N	N	N	N	N	N
ln(1831 m. pot.)	Y	Y	Y	N	N	N	N	N	N
Erodibility	N	N	N	Y	N	N	N	N	Y
Soil carbon content	N	N	N	Y	Y	N	N	N	N
Subsoil water capacity	N	N	N	N	Y	Y	N	N	N
Depth to rock	N	N	N	N	N	Y	Y	N	N
Ruggedness	N	N	N	N	N	N	Y	Y	N
Soil differentiation	N	N	N	N	N	N	N	Y	Y
First stage statistics	232.1	232.1	232.1	8.2	16.4	22.9	20.2	8.0	10.5
Over-id test p-value	0.63	0.15	0.89	0.68	0.79	0.56	0.38	0.33	0.57

306 observations for each regression.
All regressions include a constant and three amenity variables (sea, lake, and mountain).
Standard errors in parentheses. a, b, c: corresponding coefficient significant at 1, 5, 10%.

References

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2006. Structural identification of production functions. Processed, UCLA.

Andrews, Donald W.K. and James H. Stock. 2007. Inference with weak instruments. In Richard Blundell, Whitney K. Newey, and Torsten Persson (eds.) *Advances in Econometrics: Proceedings of the Ninth World Congress of the Econometric Society, volume 3*. Cambridge: Cambridge University Press.

Arnold, Jens Matthias. 2005. Productivity estimation at the plant level: A practical guide. Processed, Bocconi University.

Aubert, Patrick and Bruno Crépon. 2003. La productivité des salariés âgés : une tentative d'estimation. *Economie et Statistique* 0(368):65–94.

Bacolod, Marigee, Bernardo S. Blum, and William C. Strange. 2009. Skills in the city. *Journal of Urban Economics* 65:forthcoming.

Bayer, Patrick, Stephen L. Ross, and Giorgio Topa. 2008. Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of Political Economy* 116(6):1150–1196.

Boutin, Xavier and Simon Quantin. 2006. Une méthodologie d'évaluation comptable du coût du capital des entreprises françaises : 1984–2002. Working paper, INSEE-DESE G2006-09.

Briant, Anthony, Pierre-Philippe Combes, and Miren Lafourcade. 2007. Does the size and shape of geographical units jeopardize economic geography estimations? Processed, Paris-jourdan Sciences Economiques.

Burnod, Guillaume and Alain Chenu. 2001. Employés qualifiés et non-qualifiés: Une proposition d'aménagement de la nomenclature des catégories socioprofessionnelles. *Travail et Emploi* 0(86):87–105.

Ciccone, Antonio. 2002. Agglomeration effects in Europe. *European Economic Review* 46(2):213–227.

Ciccone, Antonio and Robert E. Hall. 1996. Productivity and the density of economic activity. *American Economic Review* 86(1):54–70.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2008*a*. Spatial wage disparities: Sorting matters! *Journal of Urban Economics* 63(2):723–742.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Sébastien Roux, and Diego Puga. 2007. The productivity advantages of large markets: Distinguishing agglomeration from firm selection. Memeographed, University of Toronto.

Combes, Pierre-Philippe, Thierry Mayer, and Jacques Thisse. 2008*b*. *Economic Geography*. Princeton (NJ): Princeton University Press.

Dahl, Gordon B. 2002. Mobility and the return to education: Testing a Roy model with multiple markets. *Econometrica* 70(6):2367–2420.

Davis, Donald R. and David E. Weinstein. 2008. A search for multiple equilibria in urban industrial structure. *Journal of Regional Science* 48(1):29–65.

Duranton, Gilles. 2007. Urban evolutions: The fast, the slow, and the still. *American Economic Review* 97(1):197–221.

Duranton, Gilles and Diego Puga. 2004. Micro-foundations of urban agglomeration economies. In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland, 2063–2117.

Eaton, Jonathan and Zvi Eckstein. 1997. Cities and growth: Theory and evidence from France and Japan. *Regional Science and Urban Economics* 27(4–5):443–474.

Evans, William N., Wallace E. Oates, and Robert M. Schwab. 1992. Measuring peer group effects: A study of teenage behavior. *Journal of Political Economy* 100(5):966–991.

Fox, Jeremy T. and Valérie Smeets. 2007. Do input quality and structural productivity estimates drive measured differences in firm productivity? Processed, University of Chicago.

Glaeser, Edward L. and David C. Maré. 2001. Cities and skills. *Journal of Labor Economics* 19(2):316–342.

Guo, L.B. and R.M. Gifford. 2002. Soil carbon stocks and land use change: A meta analysis. *Global Change Biology* 8(4):345–360.

Head, Keith and Thierry Mayer. 2004. The empirics of agglomeration and trade. In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland, 2609–2669.

Head, Keith and Thierry Mayer. 2006. Regional wage and employment responses to market potential in the EU. *Regional Science and Urban Economics* 36(5):573–594.

Hellerstein, Judith K., David Neumark, and Kenneth R. Troske. 1999. Wages, productivity, and worker characteristics: Evidence from plant-level production functions and wage equations. *Journal of Labour Economics* 17(3):409–446.

Henderson, J. Vernon. 2003. Marshall's economies. *Journal of Urban Economics* 53(1):1–28.

Levinsohn, James and Amil Petrin. 2003. Estimating production functions using inputs to control for unobservables. *Review of Economic Studies* 70(2):317–342.

Marchand, Olivier and Claude Thélot. 1997. *Le travail en France (1800-2000)*. Paris: Nathan.

Marshall, Alfred. 1890. *Principles of Economics*. London: Macmillan.

Moomaw, Ronald L. 1981. Productivity and city size? A critique of the evidence. *Quarterly Journal of Economics* 96(4):675–688.

Moretti, Enrico. 2004. Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics* 121(1):175–212.

Moulton, Brent R. 1990. An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *Review of Economics and Statistics* 72(2):334–338.

Olley, G. Steven and Ariel Pakes. 1996. The dynamics of productivity in the telecommunication equipment industry. *Econometrica* 64(6):1263–1297.

Rappaport, Jordan. 2007. Moving to nice weather. *Regional Science and Urban Economics* 37(3):375–398.

Redding, Stephen J. and Daniel M. Sturm. 2008. The costs of remoteness: Evidence from German division and reunification. *American Economic Review* 98(5):1766–1797.

Roback, Jennifer. 1982. Wages, rents and the quality of life. *Journal of Political Economy* 90(6):1257–1278.

Rosenthal, Stuart S. and William C. Strange. 2004. Evidence on the nature and sources of agglomeration economies. In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland, 2119–2171.

Rosenthal, Stuart S. and William C. Strange. 2008. The attenuation of human capital spillovers. *Journal of Urban Economics* 64(2):373–389.

Seybold, C.A., J.E. Herrick, and J.J. Brejda. 1999. Soil resilience: A fundamental component of soil quality. *Soil Science* 164(4):224–234.

Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: Printed for W. Strahan, and T. Cadell.

Stock, James H. and Motohiro Yogo. 2005. Testing for weak instruments in linear IV regression. In Donald W.K. Andrews and James H. Stock (eds.) *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge: Cambridge University Press, 80–108.

Syverson, Chad. 2004. Market structure and productivity: A concrete example. *Journal of Political Economy* 112(6):1181–1222.

B.2. The effect of location on finding a job in the Paris region

Gobillon L., Magnac T. et H. Selod (2009), “The effect of location on finding a job in the Paris region”, CEPR Working Paper 6199, *Journal of Applied Econometrics*, à paraître.

52 pages

The effect of location on finding a job in the Paris region*

L. Gobillon†, T. Magnac‡, H. Selod§

August 7, 2009

*We are grateful to a co-editor and two anonymous referees for their comments and we thank as well the participants at seminars at INRA, CREST, University of Tokyo, Louvain-la-Neuve and the North American Meetings of the Regional Science Association International 2006. We thank the French National Employment Agency (ANPE) and the Ministry of Transport's Regional Directorate (DREIF) for providing us with the data. All remaining errors are ours. The findings, interpretations, and conclusions expressed in this paper are ours and do not represent the view of our employers, including the World Bank, its executive directors, or the countries they represent.

†Institut National d'Etudes Démographiques (INED), PSE-INRA and CREST. Address: INED, 133 boulevard Davout, 75980 Paris Cedex 20, France. E-mail: laurent.gobillon@ined.fr.

‡ Toulouse School of Economics (GREMAQ & IDEI) Address: Manufacture des Tabacs, 21 allée de Brienne, 31000 Toulouse, France. E-mail: magnac@cict.fr.

§The World Bank and CREST. Address: 1818 H Street NW, Washington DC 20433, United States of America: hselod@worldbank.org.

Abstract

An important but empirically debated issue in spatial economics is whether spatial differences in unemployment reflect residential sorting on invididual characteristics or a true effect of location. We investigate this issue in the 1,300 French muncipalities that constitute the Paris region and across which there is overwhelming evidence of spatial disparities in unemployment durations. We resort to a methodology that enables us to disentangle individual and unspecified local effects. In order to control for individual determinants, we estimate a proportional hazard model stratified by municipality using an exhaustive dataset of all unemployment spells starting in the first semester of 1996. This model allows us to recover a survival function for each municipality that is purged of individual observed heterogeneity. We show that only around 30% of the spatial disparities in the propensity to find a job are explained by individual characteristics. Nearly 70% of the remaining disparities are captured by local indicators which we show to be mainly correlated with local measures of residential segregation. We are also able to show that local and individual characteristics reinforce one another in their contribution to spatial disparities in unemployment duration.

Keywords: Unemployment, Duration models, Economic geography, Urban economics, Stratified partial likelihood.

JEL Codes: C41, J64, R23

1 Introduction

The determinants of urban unemployment have raised the interest of economists for decades. In the US, two major trends of literature have tried to explain how location could have an adverse impact on employment, involving a variety of mechanisms. The first set of works is the so-called *spatial mismatch* literature which investigates how the physical disconnection from jobs can exacerbate unemployment among low-skilled minority workers (see Ihlanfeldt and Sjoquist, 1998, for an empirical survey, and Gobillon, Selod and Zenou, 2007, for a theoretical one). The second set of works investigates the impact of *residential segregation* on the poor labor-market outcomes of ghetto residents (see e.g. Wilson, 1996, Cutler and Glaeser, 1997). In both literatures, papers usually resort to cross-section methods and try to explain individual unemployment probabilities or local unemployment rates (see e.g. Ihlanfeldt, 1993, Conley and Topa, 2002, Weinberg, 2000 and 2004).

In this paper, we focus on the local determinants of unemployment *duration*. Only a few papers, mainly on the US, have studied unemployment dynamics at the individual level with a spatial perspective (Holzer, Ihlanfeldt and Sjoquist, 1994, Rogers, 1997, Dawkins, Shen and Sanchez, 2005, Johnson, 2006). In these works, authors usually investigate the impact of local indicators proxying for spatial mismatch or residential segregation in an unemployment duration model. They typically estimate a proportional hazard model with a single baseline hazard common to all locations, a set of individual variables and local indicators. We adopt a much broader approach that consists in estimating a baseline hazard function for each location while controlling for individual characteristics in a proportional hazard model. This key methodological innovation, known as the Stratified Partial Likelihood Estimator (SPLE), was first proposed by Ridder and Tunali (1999) in another context. We apply it in this paper to a large administrative dataset containing records of unemployment spells and adapt it to include some new econometric features.

Compared to the previous literature, the advantages of our empirical strategy are threefold. First, we do not need to choose a specific function for the local hazard functions. We can thus measure the overall effects of location without only focusing on a few arbitrarily selected mechanisms, proxied by criticizable local indicators. Second, we allow the effect of location to vary depending on the time spent unemployed. We can thus assess the effect of location on the short

run (say, after 6 months) and on the long run (say, after two years). Third, the model is sufficiently versatile to allow us to further restrict local hazard rates while controling for the generality of these restrictions.

The estimation procedure has three steps. The first step consists in estimating a proportional hazard model with an unspecified municipality-specific hazard baseline hazard. In the second step of the estimation procedure, we impose that the municipality effects are multiplicative in the hazard rate. This multiplicative component model summarizes the local effects through a single indicator. In the third stage, we assess how this local indicator may capture local determinants reflecting the different mechanisms put forward by the literature. This is done by regressing municipality effects on these variables and computing their explanatory power. We do not interpret this last stage as a causal regression because omitted variable or reverse causality concerns cannot be dismissed. Our procedure however ensures that the previous stages are immune to these endogeneity issues so that we can separate the robust estimation of local effects from less robust results obtained in the third stage.

Yet, we cannot easily deal with individual unobserved heterogeneity in our proportional hazard specification. This is in line with Baker and Melino (2000)'s finding that identification of both flexible hazard rates and the distribution of unobserved heterogeneity is fragile in empirical studies. Moreover, individual unobserved heterogeneity like, for instance, omitted variables related to preferences, is partially captured since we model the hazard rate function at the level of the municipality. Furthermore, we apply goodness of fit test procedures under the form of a Kolmogorov statistic as developed in Andrews (1997) and show that the model fits the data very well at the level of each municipality.

Our approach requires a very large dataset comprising enough unemployment spells in each location. We use a unique exhaustive administrative dataset available for the Paris region from which we extract unemployment spells that started in the first semester of 1996. Unemployment spells can end in three different ways: finding a job, dropping out of the labor force, and right-censorship (including exits for unknown reasons). We model the first two exits in an independent competing risk framework.

Our main empirical results are as follows. We find that controling for individual characteristics explains around 30% of the spatial disparities across municipalities in unemployment durations

until finding a job. Among the main individual determinants of unemployment duration as education, it should be stressed that some nationalities, Africans in particular, experience significantly much larger durations until finding a job. Furthermore, the association between average individual characteristics at the municipality level and baseline hazards is positive. Presumably because of sorting effects, individual and local effects reinforce each other in their contribution to spatial disparities of unemployment duration. In other words, durations are not only larger because of an adverse individual characteristic but also because the average of this adverse characteristic is larger at the local level. Finally, nearly 70% of the remaining local disparities are captured by local indicators, mainly segregation indices.

The rest of the paper is structured as follows. In section 2, we provide a short survey of the litterature on how segregation and bad physical accessibility to jobs can increase unemployment duration. Section 3 presents the data and a selection of descriptive statistics to measure spatial disparities. Section 4 details the SPLE method. Section 5 discusses the results. Finally, section 6 concludes.

2 Why should location influence unemployment duration?

The duration of unemployment depends on many factors. To discuss this issue in an orderly manner, it is useful to adopt a job-search perspective considering that exit from unemployment can occur at the end of a three-stage process. In the first stage, workers must wait some time before coming into *contact* with a job opportunity. In the second stage, an *offer* from an employer may materialize. Finally, workers may *accept or reject* the offer depending on whether the offered wage is greater or smaller than their reservation wage. With this framework in mind, job seekers who, on average, wait long before experiencing contacts with employers and who have few chances to transform their contacts into offers and matches should experience long unemployment spells. For instance, educated workers could be advantaged in the first stage if they are more efficient in obtaining information about jobs and in contacting firms, or if labor demand is biased in their favor. They may also have an advantage in the second stage if they write better application letters and resumes, and fare better during interviews. However, educated workers may be more likely to reject an offer when they face or anticipate many well-paid outside offers. Other individual

and family characteristics such as gender, race/ethnicity, age, experience, marital status or the number and age of children and dependants should also be expected to affect unemployment duration through one or several stages of the job-acquisition process.

This section describes how *location*, i.e. the *disconnection from job opportunities* (in cases where job opportunities are unevenly distributed within a metropolitan area) and/or *residential segregation* (in terms of education, race/ethnicity/nationality or employment status), can also influence the duration of unemployment. We decompose the effects on each stage of the job-acquisition process.

Disconnection from job opportunities may directly affect the time spent searching for a job in the first stage of the process. Indeed, job-seekers residing in areas with few local job vacancies or in areas located far away from employment centers are exposed only to a small pool of vacancies. Residing in loose local labor markets, they should spend more time searching before getting into contact with a potential employer. Of course, job-seekers also have the possibility to search for jobs in other areas. But having to search away from one's area of residence penalizes job seekers. At least three reasons may come into view. Firstly, because of informational frictions, *job-seekers may not search efficiently far away from their residences.* For instance, workers residing far away from job opportunities may not hear about job offers when firms resort to recruiting methods that favor the local labor force (i.e. by posting 'wanted' signs in retail shops, or by choosing not to publicize job offers beyond a certain distance). Alternatively, job-seekers may obtain only partial information on the location of distant jobs or may have only a vague idea about the types of jobs offered in parts of the metropolitan area they are not familiar with. They may end up searching in the wrong places (Ihlanfeldt, 1997, Stoll and Raphael, 2000). Secondly, because search is costly, *workers may restrict their search horizon at the vicinity of their neighborhood.* They may search less often in order to reduce the number of job-search trips or may not search at all for jobs located in distant places. In this context, access to public transport or car ownership can reduce job-search costs and expand the job-search horizon (Stoll, 1999). Thirdly, the individual search effort may depend on the local cost of living so that *workers residing in areas disconnected from job opportunities may not search intensively.* It has been argued that workers residing in such areas usually incur low housing costs and thus may feel relatively little less pressure to actively

search for a job in order to pay their rent (Smith and Zenou, 2003, Pattachini and Zenou, 2006).

Disconnection from job opportunities may also reduce the frequency of job proposals in the second stage. *Employers may then be reluctant to propose jobs to distant workers* because commuting long distances would make these workers less productive (they would show up late or be tired due to excessive commuting, see Zenou, 2002).

Distance to job opportunities may also reduce the probability of a job acceptance in the third stage. Indeed, *workers may reject a job offer that would involve commutes that are too long* if commuting to that job would be too costly in view of the proposed wage (Zax and Kain, 1996). In other words, distance is likely to make the offered wage net of commuting costs drop below a worker's given reservation wage.

The effect of *residential segregation* on the first stage of the job-acquisition process is also likely to be harmful to the extent that job contacts often occur through friends and relatives (Mortensen and Vishwanath, 1994). Because *social networks are at least partly localized*, when the unemployment rate is high in a given area, workers are less likely to know employed neighbors that can let them know about existing vacancies (Calvó and Jackson, 2004, Selod and Zenou, 2006).

Residential segregation is also likely to reduce the probability for a worker residing in a segregated area of receiving a job offer. This is because employers may discriminate against residentially segregated workers, a practice known as *redlining* (see Wilson, 1996, for stories of firms not hiring workers located in 'bad' neighborhoods). For employers, the motivation can hinge upon the stigma or prejudice associated with the residential location of candidates (*sheer discrimination*), or because they consider that, on average, workers from stigmatized areas have bad work habits or are more likely to be criminal (*statistical discrimination*). In industries and jobs in which workers are in contact with customers, employers may discriminate against residentially-segregated workers in order to satisfy the perceived prejudices of their clients, a practice known as *customer discrimination* (Holzer and Ihlanfeldt, 1998). In France, the issue of *redlining* is increasingly being put forward in the public debate to account for the unemployment of the young adults that reside in distressed areas. To our knowledge, however, the issue has not yet been studied empirically.

All these economic mechanisms suggest that the rate at which workers leave unemployment, and thus the duration of unemployment, depends on both individual characteristics and local

features. In the present paper, we propose a methodology to disentangle individual and unspecified local effects. We explore the nature of local effects by regressing them on indices of segregation and distance to job opportunities. We assess the overall impact of these indices on finding a job, but we do not try to identify through which specific mechanisms they percolate.

3 Description of the Data

3.1 The area of study

The paper focuses on the Ile-de-France region (the Paris region hereafter), an administrative unit of 10.9 million inhabitants distributed over 1,280 municipalities centered around the city of Paris and the 20 administrative subdistricts of Paris (which will be treated as municipalities in the analysis). These 1,300 spatial units may have very different population sizes which range from 225,000 in the most populous Parisian subdistrict to small villages located some 80 km away from the center of Paris. They correspond more or less to the Paris Metropolitan Area as can be seen from Graph 1 which represents the population density in each municipality.

[*Insert Graph* 1]

Graph 2 provides evidence that the studied area exhibits large spatial disparities in the local unemployment rates across municipalities. In particular, the unemployment rates in municipalities located to the North-East of Paris are more than four times higher than in most municipalities located to the West.

[*Insert Graph* 2]

3.2 The ANPE historical file

We use the historical file of job applicants to the National Agency for Employment (*Agence Nationale pour l'Emploi* or *ANPE* hereafter) for the Paris region to study spatial disparities in unemployment durations. This database provides a quasi-exhaustive list of unemployment spells in the region as it has been estimated that 90% of job seekers in France are indeed registered with the ANPE (Chardon and Goux, 2003). The reason is that registering with the ANPE is

a prerequisite for unemployment workers to be able to claim their unemployment benefits. A significant share of those not eligible for unemployment benefits—as for instance first time job seekers—also register with the ANPE to assist them in their job search.

The ANPE is organized in hundreds of local agencies and unemployed workers register in the agency closest to their residence.[1] The exhaustive dataset that we have for the Paris region contains information on the exact date of an application (the very day), the unemployment duration (in days), and the reason for which the application came to an end. Along with the municipality where the individual lives and registers, a set of socio-economic characteristics were reported upon registration with the employment agency: age, gender, nationality, diploma, marital status, number of children and disabilities. To build our working sample, we select individuals who applied to the employment agency between January 1 and June 30, 1996 and who lived in the Paris region at that time. As we have information on unemployment spells until 2003, starting as early as 1996 enables us to follow unemployed workers over a long period of time and to minimize the number of incomplete spells due to the end of the observation period (which only concerns .6% of the exits in our sample. After deleting the very few observations for which socio-economic characteristics are missing, we end up with $430,695$ observations on individual unemployment durations. More details on the construction and the contents of the dataset are given in Appendix A.

We group the reasons given for the termination of the application with the agency into three types: (1) finding a job, (2) exiting to non-employment, and (3) right censoring, which groups together unknown destinations and incomplete spells.[2] In the following, we assume that right-censoring is independent of the durations until exiting to a job or non-employment and that these two exits are independent, conditional on all observed characteristics including the municipality of residence. A large proportion of exits are right-censored (55.3%), of which 29.5% correspond to an

[1] Except in very specific occupations (artists, ...).

[2] An exit to non-employment covers the following situations: a training period, an illness, a pregnancy, a job accident (as some unemployed workers can in fact work for a very small number of hours), an exemption from the rule imposing to actively search for a job, retirement, or military service. Unknown destinations can result from mobility between four subregions (see text below), an absence at a control, an expulsion for some misbehavior, an absence after a notification, a training or job refusal, a fake statement, the lack of a positive action to search for a job, and other unspecified cases.

absence at a control.[3] The remaining unemployment spells mostly end up with a job (28%) even if exit to non-employment is far from negligible (16.7%). The average unemployment duration for individuals finding a job is 269 days whereas it stands at a higher level of 368 days for individuals who exit to non-employment. The higher unemployment duration for exits to non-employment could possibly reflect the discouragement of workers that could not find a job after a long time.

A crucial issue in this data arises because residential mobility across municipalities could blur our estimation of spatial disparities. In particular, because mobility can change the way unemployment spells are recorded, it could give rise to (*i*) right and left censoring, to (*ii*) measurement errors of local effects, and (*iii*) to departures from the independence assumption because of right censoring. To see how these issues may emerge and why we believe, however, that they may be relatively minor, consider the following lines of reasoning:

First, the French Employment Agency is organized geographically in large subregions called ASSEDIC and when residential mobility brings about a change in ASSEDIC, the unemployment spell is right-censored and a (mistakenly fresh) spell is started in the new ASSEDIC. The unemployment spell is thus cut into two halves, the first spell being right censored, and the second spell being left-censored. Fortunately, there are only 4 ASSEDIC located in the Paris region (West, East, South-East and Paris) and in our date only 4.83% of unemployed workers change ASSEDIC as stated in the reasons for exits. Moreover, double counting is also mitigated by the fact that we consider a flow-sampling window of only six months starting at the beginning of 1996.

Second, even if mobility takes place within the same ASSEDIC, it might bring a change in municipality and local agency. Although the spell is registered as uninterrupted, the stated place of residence may either correspond to that recorded at the origin or at the destination agency.[4] Yet, measurement errors of local effects, if anything, would likely attenuate our measures of spatial disparities.

[3] There is no evidence that these absences would mainly concern unemployed workers that neglected to report they found a job. Indeed, a 2005 follow-up survey on a small random sample of unemployed workers having left the ANPE showed that only approximately half of absentees at controls did find a job, which is not in contradiction with the assumption of independence between right censoring and finding a job.

[4] The administrative treatment by ANPE of spells which ended in a local agency different from the one where it started is very obscure. The two spells are registered and one of them is deleted apparently without following any precise written rule.

Finally, in many aspects, the mobility decision could be partly independent of exiting to a job or to non-employment. For instance, the prospect of saving money by living with one's parents might be an exogenous factor conditional on the individual and local effects. On the other hand, mobility is clearly endogenous if a job seeker moves with the prospect of waiting until finding a job near the new location. Given that mobility is low in France with respect to the US (Baccaini, Courgeau and Desplanques, 1993), we believe that these issues have second order effects compared to the main substantive effects of infrastructure, redlining or discrimination to which we now turn.

3.3 Spatial disparities

The Paris region exhibits stark socio-economic disparities which can broadly be depicted as follows. In the North-East, the population is usually little educated, poor, and composed of blue collar workers. Recent migrant minorities are over-represented. In the West, the population is very educated, rich, and comprises mostly white collars. Minorities of recent immigration waves are under-represented.

To further characterize disparities across municipalities and differences in municipality environments, we compute municipality-specific segregation and job-accessibility variables using several sources.

3.3.1 Census measures of segregation and job accessibility

Segregation is accounted for by the municipality proportion of education and nationality groups computed from the 1999 Population Census. Job accessibility is measured by the job density around each municipality. More precisely, for each municipality we are able to identify all the other municipalities than can be reached within 45 minutes for a given transport mode (private vehicles or public transport). The 45-minute cut-off has been chosen just above the average commuting time of 34 minutes in the Paris region (DREIF-INSEE, 1997). This defines a group of accessible municipalities for which we can calculate the overall job density (the ratio of the number of jobs located in the area to the number of occupied and unoccupied workers residing in the same area).[5] Data on the location of jobs and workers are from the 1999 census. Travel times between

[5] For a discussion of alternative indicators see Gobillon and Selod (2007).

municipalities are estimated at morning peak hours by the French Department of Transportation for 2000 using a transport survey on the Paris region (*Enquête Générale des Transports*).

We compute indices of spatial disparities on these local segregation and job-accessibility variables. The indices we compute are the inter-decile ratio, the inter-decile range, the Gini index and the coefficient of variation, and results are reported in Table 1.[6] We find that spatial disparities across muncipalities are very large for the percentage of African nationalities as the inter-decile ratio is over 9 for the percentages of citizens from North Africa and sub-Saharan Africa. This means that the frequency of African citizens is 9 times larger in municipalities at the ninth decile than in municipalities at the first decile. Spatial disparities are also large for segregation in terms of education and stands near 4 for the percentage of individuals with a university degree and around 2.5 for the percentage of individuals with a technical degree. Measures of job accessibility also exhibit significant spatial disparities. The inter-decile ratio for job densities by public transport is 3.

[*Insert Table* 1]

3.3.2 Spatial Disparities for the Unemployed

Spatial disparities in the characteristics of unemployed workers also exhibit a similar pattern over the Paris region. Table 2 reports similar indices of spatial disparities across municipalities for several variables of the ANPE historical file. We measure the spatial disparities in the occurrence of exit types, the unemployment duration conditionally on the type of exit, and the individual variables that we use in our empirical analysis below. As with the census data, the indices we compute are the inter-decile ratio, the inter-decile range, the Gini index and the coefficient of variation.

We first comment the spatial disparities in the proportions of individuals who respectively experience an exit to a job, an exit to non-employment, and right-censoring. For simplicity, we restrict our comments to the inter-decile ratio but other indicators give qualitatively similar results.

[6] To compute the spatial inter-decile index of a variable, we construct the empirical distribution function of the local average of the variable. Observations are weighted by the population in each municipality. We smooth the empirical distribution by a Gaussian kernel with a Silverman's rule of thumb bandwidth and deciles are retrieved using a very fine grid ($1,000,000$ points).

The inter-decile ratio is fairly large for the probability that unemployment finishes with an exit to a job as it reaches 1.73. This means that, if we order municipalities with respect to the proportion of unemployment spells ending with an exit to a job, an unemployment spell has 73 percent more chances to end with an exit to a job in the municipality at the ninth decile than in the municipality at the first decile. The inter-decile ratio is smaller for the probability of an exit to non-employment (1.37) and for right-censoring (1.32). These variations in spatial disparities across exit types calls for a careful conditioning on local effects.

If we now look at unemployment durations conditionally on the type of exit, the inter-decile ratio for unemployment spells ending with an exit to a job reaches 1.37. This means that an unemployment spell ending with an exit to a job lasts 37 percent longer in the municipality at the ninth decile than in the municipality at the first decile. For unemployment spells ending with an exit to non-employment, the inter-decile ratio is even greater and stands at 1.43.

[*Insert Table* 2]

As the above data is right-censored because of exits to other states, these statistics are difficult to interpret. This is why we also assess disparities between municipalities with the help of duration models. For each type of exit and for each municipality, we compute the Kaplan-Meier estimator of the survival function (which takes into account right-censorship). Disparities by exit type can then be assessed by comparing the survival function across municipalities for any chosen duration. As survival functions are well estimated only when the number of unemployed workers is large enough, we restrict our attention to municipalities with a population greater than 5,000 inhabitants in 1999. Graph 3 represents the probability of finding a job before 24 months for each municipality of the Paris region. Disparities are large: the probability of finding a job before 24 months is below 40% in many municipalities of the North-East, whereas it is above 55% in many municipalities of the West. Graph 4 represents the probability of exiting to non-employment before 24 months for each municipality. Contrary to the graph for exit to a job, no specific pattern emerges. This contrast suggests that while job search outcomes strongly depend on location, this is less the case for labor-market participation decisions.

[*Insert Graph* 3 *and* 4]

There are also noticeable spatial disparities in some of the socio-demographic characteristics of

unemployed workers. Whereas the spatial disparities in age, sex or marital status are small (see Table 2), there are much larger disparities for some categories of nationality, education, and family size, as well as for disability. The inter-decile ratio for instance is greater than 5 for the proportion of Africans. In other words, the proportion of Africans among unemployed workers in municipalities at the ninth decile is 5 times greater than in municipalities at the first decile. The inter-decile ratio is above 5 for unemployed workers having three children or more, around 4 for unemployed workers with no diploma, and above 2.5 for disabled unemployed workers.

In conclusion, spatial disparities of individual characteristics are large. Results in Tables 1 and 2 are hard to compare, data sources being so different. Likely origins of differences between them could be reporting errors and composition effects although we did not investigate the point further. It very much depends on the spatial disparities in entries into unemployment which we cannot assess with the data that are available to us.

4 The econometric model

We analyze the empirical associations between unemployment durations and the local context (segregation and job accessibility) using a three-stage procedure. First, we specify a proportional hazard model (PH model hereafter) with individual covariates and a municipality-specific baseline hazard. Parameters related to individual variables are estimated using the stratified partial likelihood estimator (SPLE hereafter) as proposed by Ridder and Tunali (1999). Municipality-specific integrated baseline hazards are then recovered using the Breslow estimator. Second, municipal baseline hazards are restricted to be a multiplicative function of an aggregate baseline hazard function and of municipality effects which are both estimated using the first-stage outputs. A third and final descriptive stage consists in regressing the municipality effects on local indicators of segregation (municipality composition) and job accessibility.

Our approach can be justified as follows. The first two stages allow to estimate municipality effects. For computational reasons, this would be unfeasible by maximum likelihood estimation in one stage only since the number of municipalities $(1,300)$ is too large. The final stage consists in regressing those municipality effects on aggregate variables. It enables us to analyse the correlation between spatial effects and segregation or job accessibility indices, although we do not claim to estimate

causal effects in the last stage. Our procedure in three steps guards us against specification errors at each stage. First stage estimates are robust to misspecification of the multiplicative model and of the descriptive regression of municipality effects. Second stage estimates are robust to errors at the descriptive regression stage. Furthermore, all steps contribute to the empirical analysis of spatial disparities in unemployment durations.

4.1 Model Specification

Consider an individual i who enters unemployment (i.e. who enters the ANPE file). His unemployment spell lasts until he finds a job (exit labeled e) or drops out of the labor force (exit labeled ne). The unemployment spell is right-censored if the individual disappears from the records during the observation period or has not experienced an exit before the last day of observation in the panel. A latent duration T_k is associated with each exit $k \in \{e, ne\}$. For an individual i, we denote $\lambda_k(\bullet \,|X_i, j(i))$ the conditional hazard rate for exit k where X_i is a set of individual explanatory variables – that are fixed over time in our application – and $j(i)$, where $j(i) \in \{1, ..., J\}$, is the municipality where the individual is located. We adopt the proportional hazard assumption separating the effect of individual characteristics and the effect of local clusters by writing:

$$\lambda_k(t\,|X_i, j(i)) = \theta_k^{j(i)}(t) \exp(X_i \beta_k) \text{ for } k \in \{e, ne\}, \tag{1}$$

where $\theta_k^j(t)$ is the baseline hazard rate function for municipality j and exit k. Observe that the effect of local variables is not of the proportional hazard type at this stage since the municipality-specific baseline hazard rate is fully flexible. Additionally, the two latent durations and right-censorship are assumed to be independent so that our framework is an independent competing risk model where observations are clustered.

Observe also that the above specification features come at the expense of overlooking unobserved individual heterogeneity whose presence can bias the estimation of the hazard rates and parameters. Latent durations associated with different types of exit might also be dependent if the effect of individual unobserved heterogeneity influencing the different types of exit are correlated. Lancaster (1990) proposes to introduce individual unobserved heterogeneity in a partial likelihood model by modeling it as a gamma distribution and to estimate parameters using an EM algorithm. Yet, the procedure is burdensome and unfeasible in samples where the number of

observations is as large as in ours. An alternative way to proceed would be to difference out individual unobserved terms using multiple spells. In theory, this could be done by redefining clusters as couples (municipality, individual) but the number of applicants appearing twice or more is very small (about 8%) and the issue about biases caused by residential mobility could then become serious. Also note that Baker and Melino (2000) argue that it is difficult to empirically identify the unobserved heterogeneity distribution and flexible hazard rates, and that in the current application, the hazards are fully flexible at the municipality levels. For all these reasons, we decided not to incorporate individual unobserved heterogeneity in our econometric specification (1). We nevertheless discuss below the consequences of its presence on our empirical results. Specifically, we will pay attention to the effects of the sorting of individuals across municipalities according to their unobserved characteristics.

4.2 Stratified Partial Likelihood Estimation (SPLE)

Our estimation follows Ridder and Tunali (1999). Start with the estimation of the effects of individual explanatory variables using the SPLE. Denote $\Omega^j(t)$ the set of individuals at risk of exiting unemployment in municipality j at time t. The probability of individual i experiencing a type-k exit at time t conditionally on someone in the same municipality experiencing a type-k exit is:[7]

$$P_i(t,k) = \frac{\exp(X_i\beta_k)}{\sum_{n\in\Omega^{j(i)}(t)} \exp(X_n\beta_k)} \tag{2}$$

Observe that conditioning on the municipality population at risk (instead of the whole population at risk) makes all municipality-specific baseline hazards cancel out so that we do not need to specify its functional form. The stratified partial likelihood function (calculated on all unemployed workers who experience an exit to a job or to non-employment) is:

$$L = \prod_i P_i(t_i,k_i) = \prod_k L_k(\beta_k) \tag{3}$$

[7] This formula is exact only when time is continuous. In our data where time is expressed in days, several individuals may exit the same day and it is impossible to order them depending on their time of exit. Nevertheless, following Breslow (1974), we consider (2) as an approximation of the conditional probability of exit. In practice, when an individual exits a given day, the risk set includes all the other individuals who exit the same day.

where t_i is the time of exit of individual i, k_i is the type of exit of individual i, and $L_k(\beta_k) = \prod_{i|k_i=k} P_i(t_i, k_i)$ is constructed from all unemployment spells that end with a type-k exit. $L_k(\beta_k)$ is the partial likelihood obtained in the hypothetical context where there is only one possible exit k and where unemployment spells are censored if they end up with the other exit. Notice that each set of parameters β_k can be separately estimated by maximizing the corresponding term L_k under the independent competing risks assumption. Denote $\widehat{\beta}_k$ the estimator.

We can now turn to the estimation of the municipality baseline hazard function. For exit k, the Breslow estimator of the integrated baseline hazard of municipality j, $\Theta_k^j(t)$, is defined as:

$$\widehat{\Theta}_k^j(t) = \int_0^t \frac{I(C^j(s) > 0)}{\sum_{i \in \Omega^j(s)} \exp(X_i \widehat{\beta}_k)} dN_k^j(s), \tag{4}$$

where $I(\bullet)$ is the indicator function, $C^j(s) = card\ \Omega^j(s)$, and $dN_k^j(s)$ is a dummy that equals one if someone in municipality j experiences a type k-exit in an arbitrarily short period of time before date s (and zero otherwise). For each t, the variance of $\widehat{\Theta}_k^j(t)$ can be recovered from Ridder and Tunali's formulas.[8]

4.3 Estimation of Spatial Effects

In the second stage, for each type of exit, we estimate municipality effects that summarize the municipality-specific baseline hazard rates by a single quantity. This is a restriction of the general model. Since the estimation procedure can be applied to each type of exit separately, we restrict our attention to a given exit k and drop subscript k for readability. It should be kept in mind that all parameters analyzed below are exit-specific.

We assume that the municipality-specific baseline hazard rates take a multiplicative form:

$$\theta^j(t) = \alpha^j \theta(t) \tag{5}$$

where α^j is a municipality fixed effect and $\theta(t)$ is a general baseline hazard function. Here, we depart from Ridder and Tunali who adopt an additive form. Indeed, we find it more natural to use

[8] This can be done using their equations $A25$, $A27$ and $A29$ and setting $K = 1$, $t_0 = 0$ and $t_1 = T$ in their equation (22).

a multiplicative specification since, when combining (1) and (5), we obtain a proportional hazard model where spatial effects enter multiplicatively. This is presumably quite restrictive though worth investigating.

Instead of directly implementing the functional estimation of (5), we divide the period $[0, \infty)$ into M intervals whose lower (resp. upper) bound is t_{m-1} (resp. t_m), for $m = 1, ..., M$ (where $t_0 = 0$ and $t_M = \infty$). If we denote $\theta_m = \int_{t_{m-1}}^{t_m} \theta(s)\, ds$ the increment of the aggregate integrated baseline hazard over the interval m, the increment of the integrated baseline hazard rate over a time interval m in a municipality j is given by

$$y_m^j = \Theta^j(t_m) - \Theta^j(t_{m-1}) = \alpha^j \theta_m.$$

An estimate of the average hazard rate y_m^j can be constructed from equation (4) and is:[9]

$$\hat{y}_m^j = \widehat{\Theta}^j(t_m) - \widehat{\Theta}^j(t_{m-1}).$$

Using equation (5), we can now set up the estimated model as a minimum distance problem (or *asymptotic least squares*, see Gouriéroux *et al.*, 1985) by writing that:

$$\ln\left(\hat{y}_m^j\right) = \ln\left(\alpha^j\right) + \ln(\theta_m) + \varepsilon_m^j \tag{6}$$

where $\varepsilon_m^j = \ln(\hat{y}_m^j) - \ln(y_m^j)$ is the residual due to the sampling variability of estimated hazard rates (see Appendix B.2.2 for the computation of the covariance matrix).

There are two statistical issues of importance. First, note that (6) is ill-defined when $\widehat{y}_m^j$ takes the value zero. This happens when there is no exit of type k in municipality j in the time interval $[t_{m-1}, t_m]$. Corresponding observations are ignored in the estimation. It is a small sample issue that can be safely ignored if the number of observations is large as in most municipalities. In practice, there is a trade-off between small sample biases and precision when choosing the intervals. Trading off optimally bias and precision by constructing optimal data-driven intervals is out of the scope of this paper.

Second, equation (6) is a two-component model that can be estimated using weighted least squares where the weights are given by the square root of the inverse of the covariance matrix of residuals ε_m^j. However, this minimum distance estimator is known to perform badly in small samples as

[9] Dealing with the last interval is specific and detailed in the Appendix B.2.1.

shown by Altonji and Segal (1996).[10] We chose to use a slight modification of their equally weighted estimator which is simpler and better behaved. We simply weigh the estimation by the number of unemployed workers at risk at the beginning of the intervals in the municipalities (see Appendix B.2.3). Indeed, the average hazard rate computed for any given time interval (the dependent variable in (6)) is usually computed with more accuracy when the number of unemployed workers at risk is large. We also used other weighting schemes that yielded minor changes to the results.

The final descriptive stage consists in regressing municipality effects on aggregate explanatory variables at the municipality level. We specify:

$$\ln\left(\alpha^j\right) = Z^j\gamma + \eta^j \tag{7}$$

where Z^j are municipality variables and η^j are random terms. As municipality fixed effects are estimated in the previous stages, their exact value is not observed. Introducing these estimators in equation (7), we obtain:

$$\widehat{\ln\left(\alpha^j\right)} = Z^j\gamma + \eta^j + \xi^j \tag{8}$$

where $\xi^j = \widehat{\ln\left(\alpha^j\right)} - \ln\left(\alpha^j\right)$ is a sampling error. Equation (8) is estimated using weighted least squares where the weight is the initial number of unemployed workers in the municipality (see Appendix B.3). This weighting has two justifications. As above, the sampling error decreases with the number of unemployed workers. Second, weighting by the number of unemployed workers can be justified if we assume that municipalities can be decomposed into smaller areas of fixed population in which exit from unemployment is subject to an idiosyncratic shock with variance σ^2 (but affected in the same way by municipality variables). In this context, the aggregate random term η^j at the municipality level in equation (7) is an average of the smaller areas' idiosyncratic shocks. We thus assume that the terms η^j have a variance of the form $\sigma^2/C^j(0)$ where $C^j(0)$ is the initial number of unemployed workers in municipality j.

[10] More specifically, they argue that the optimal weights in minimum distance could be no better than arbitrary and simple weights. The issue of which weights are better is standing since none of these weights are optimal in small samples. Additionally, correcting small sample biases by bootstrap or jackknife does not perform better (Horowitz, 1998).

5 Results

We now comment the results of the empirical analysis whose estimation stages were described in the previous section. We first examine the estimated coefficients of the individual explanatory variables obtained using the stratified partial likelihood estimator (stage 1). We then describe the spatial disparities in municipality survival functions obtained from the model. We finally turn to the results concerning the municipality fixed effects (stage 2) and regress them on local variables measuring residential segregation and job accessibility (stage 3).

5.1 Individual Determinants of Unemployment Durations

Table 3 reports the coefficients estimated using SPLE for each type of exit (job and non-employment). Remember that the effects of individual variables should be interpreted as affecting multiplicatively the hazard rates (through the term $\exp(X_i\beta_k)$ in (1)).

[*Insert Table* 3]

Results are as expected although the magnitude of the effects of some variables is surprisingly large. First, for both exits, younger people have shorter unemployment spells. Although negative and significant, the effect of age is marginally decreasing (in absolute value) as evidenced by the square term. Note that it is never positive in any reasonable age range. Second, women exit significantly more slowly to a job than men (around -18%) while their exit rate to non-employment is much larger (around $+35\%$). Similarly, having children (whatever their number) decreases the exit rate to a job and increases the exit rate to non-employment. Being in a couple significantly increases exit rates both to a job and to non-employment.

The strongest effects are for nationality. Africans and other non-European citizens have an exit rate to a job that is between 45% and 66% lower than the French. In contrast, the effect of nationality variables on the hazard rate for exit to non-employment is significant only for North Africans and the magnitude of the coefficient is much lower than for exit to a job.

Education variables also have a strong effect. Overall, education affects the exit rate to a job more than the exit rate to non-employment. For instance, compared to a university degree, a basic degree lowers the exit rate to a job by 59% while it decreases the exit rate to non-employment

by "only" 42%. The shadow wage (i.e. the opportunity cost of time in non participation) is less affected by education than market wages.

We perform two specification checks. First, it is interesting to compare our results with the results of the estimation of a standard Cox model where the baseline hazard function is restricted to be constant across municipalities. We use the same individual variables X_i as covariates to which we add the third-stage municipality variables $Z^{j(i)}$ (segregation indices and job accessibility measures). Under these assumptions, our three-stage procedure collapses into one stage only and results are not robust to misspecification in the second and third stage of our procedure. The estimated coefficients of individual variables (and their standard errors) are very close to those obtained using our estimation method (SPLE) in stage 1.[11] The estimated coefficients of geographic variables are also quite close to what we will obtain in stage 3 (see below). Nevertheless, the standard errors of the estimated coefficients of aggregate variables obtained with a standard Cox model are at least one-third smaller than those obtained in our third-stage estimation. There are two explanations for this difference. First, the standard Cox model does not account for aggregate unobserved effects whereas our estimation method does. It is widely known that this can lead to very biased standard errors (see Moulton, 1990). Second, there can be an efficiency loss when using our three stage approach since we are not estimating all equations at the same time.

As a second specification check, we compare the Kaplan-Meier estimates of the survival function until an exit to a job with the estimates derived in our model for each municipality.[12] We use a Kolmogorov statistic as developed in Andrews (1997). The difficulty in the procedure – as detailed in Appendix B.1 – is the estimation of the variance of the test statistic evaluated by bootstrap. The results show that our model provides a very good fit at the municipality level. In Graph 5 we report the empirical frequency across municipalities of the p-values associated with the null hypothesis of a good fit. The empirical frequency of municipalities in which we reject a good fit is indeed lower than the level of the test. It shows that we are able to describe unemployment durations at the municipality level in a very satisfactory way. This justifies our specification and,

[11] Complete results are available upon request.

[12] We are thankful to a referee for this suggestion.

in particular, our choice of not modelling unobserved heterogeneity.

[*Insert Graph* 5]

5.2 Describing spatial disparities in unemployment duration

We now assess the relative importance of individual characteristics and spatial effects in explaining the spatial disparities in unemployment durations until finding a job or leaving to non-employment.

5.2.1 Comparing the Explanatory Power of Individual and Spatial Effects

The respective explanatory power of individual and spatial determinants is evaluated using two complementary approaches.

The first approach consists in comparing indices of spatial disparities obtained from the Kaplan-Meier estimators and from our model. While Kaplan-Meier estimators represent the raw data and do not control for observed individual determinants of durations, the survival functions obtained from the model (as computed from the integrated baseline hazard functions in equation (4)) do control for individual determinants. In Table 4, we report various disparity indices (inter-decile range and ratio, Gini index and coefficient of variation) of the survival functions after 6 and 24 months both for Kaplan-Meier and for the model. For exits to a job, we find that individual variables explain only around 24% of spatial disparities at 6 months and around 15% at 24 months. To see how these are calculated, consider for instance that the inter-decile ratio at 24 months is 1.503 for the Kaplan-Meier estimate, and 1.426 for the survival function from the model. A coefficient of determination could thus be defined as $(.503 - .426)/.503$, which is equal to 15.3%. This shows that even after controlling for the characteristics of local unemployed workers, spatial disparities in finding a job remain large. This is a common theme in the literature (see Maurin, 2004).

[*Insert Table* 4]

Note that the comparisons, which rely on the usual estimators of the survival functions, are only heuristic. Indeed, they are not based on an analytical relationship between the Kaplan-Meier estimators, and the effect of individual variables and the municipality survival functions of the model.

Our second approach to estimate the relative importance of the contributions of individual and local characteristics to spatial disparities in unemployment durations does not make use of the Kaplan-Meier estimator and has firmer analytical grounds. It decomposes the average integrated hazard at the municipality level. To see why such an analysis is interesting, write for a given individual i, the logarithm of the integrated hazard as the sum of the effect of individual observed characteristics $X_i\beta$ and the logarithm of the municipality integrated baseline hazard $\Theta^{j(i)}(T)$:

$$\ln \Lambda \left(T \left| X_i, j\left(i\right)\right.\right) = X_i\beta + \ln \Theta^{j(i)}\left(T\right) \tag{9}$$

because of the proportional hazard assumption. When we take the expectation of this equation in the population of unemployed workers in a given municipality j exiting at time t, we obtain the following decomposition of the log-integrated hazard of that municipality:

$$E(X_i\beta \mid i \in \Omega^j\left(0\right), T = t) + E(\ln \Theta^{j(i)}\left(T\right) \mid i \in \Omega^j\left(0\right), T = t) = X^j(t)\beta + \ln \Theta^j\left(t\right) \tag{10}$$

where $\Omega^j\left(0\right)$ is the population of unemployed workers in the municipality and where $X^j(t)$ is the expectation of individual characteristics in the sub-population $i \in \Omega^j\left(0\right), T = t$. In practice, the two right-hand side terms can be consistently estimated via the first stage estimations and their sum yields an estimate of the total effect. We use this decomposition between the effect of explanatory variables and the municipality integrated baseline hazard function to analyze the spatial variance across municipalities of the hazard.

We can even go further in the multiplicative model defined by equation (5) since it implies that:

$$\ln \Theta^j\left(t\right) = \ln \Theta\left(t\right) + \ln \alpha^j \tag{11}$$

and the spatial variance of the expression in equation (10) does not depend any longer on the hazard.

In Table 5, we report the results of the analysis of the spatial variance across municipalities using equation (10) for durations of various lengths: short (6 months), intermediate (12 months), long (24 months) and the multiplicative model corresponding to equation (11). Averages of individual observable characteristics explain around 30% of spatial differences in exits to jobs. This confers to individual variables slightly more explanatory power than what we obtained in Table 4, although it remains quite low. For instance at 12 months, the spatial variance of the log-integrated baseline hazard is equal to .0560 while the spatial variance of the average log-integrated

hazard in the municipality (LHS of (10)) is .0786. A pseudo-coefficient of determination is thus $(.0786 - .0560)/.0786$, which is equal to 29%.

[*Insert Table* 5]

5.2.2 Spatial Sorting and Spatial Effects

To understand what the remaining spatial disparities capture, it is useful to rewrite equation (10) by decomposing the log municipality integrated baseline hazard function into linear components at time t:

$$\ln \Theta^j(t) = \ln \Theta(t) + Z^j \gamma(t) + \eta^j(t) + \varepsilon^j(t) \tag{12}$$

where $\Theta(t)$ is the aggregate integrated baseline hazard, Z^j are observed municipality characteristics whose coefficients vary with time and $\eta^j(t)$ are unobserved municipality effects due to geographic features such as infrastructures. The last term $\varepsilon^j(t)$ stands for spatial sorting of individuals into different municipalities. After controlling for individual observed characteristics, (12) shows that the remaining spatial disparities can be due not only to local characteristics (observed or not) but also to variations in the local average of individual unobserved characteristics. The lack of identification of these different effects is one form of the so-called reflection problem of Manski (1993).

5.2.3 The spatial correlation between integrated baseline hazards and composition effects

Returning to the analysis of equation (10), it is also meaningful to calculate the correlations between the municipality composition effects $(X^j(t)\beta)$ and the logarithm of the municipality integrated baseline hazard $(\ln \Theta^j(t))$ at 6, 12 and 24 months. This correlation can be interpreted in three ways. It can reflect some sorting on observable municipality effects $(Z^j\gamma(t))$, some sorting on unobservable municipality characteristics $(\eta^j(t))$, or a correlation between the local average of individual observed variables and the local average of individual unobserved variables $(\varepsilon^j(t))$.

Correlations are shown in Table 5. For exits to jobs, the correlation between the municipality composition effects and the municipality integrated baseline hazard is positive (for instance .17 at 12 months), whereas, for exits to non-employment, it is negative ($-.04$ at 12 months). In

order to assess the robustness of these findings, Graph 6 plots, for exits to jobs, the locally aggregated predictor $X^j\beta(t)$ as a function of the logarithm of the municipality integrated baseline hazard $\ln \Theta^j(t)$ at 12 months for municipalities with more than 10,000 inhabitants. The positive association between these variables appears from these plots. It means that individual and local variables reinforce each other when affecting unemployment exits.

[*Insert Graph* 6]

Using equation (12), this correlation can be interpreted in three ways related to sorting. First, unemployed workers who are less likely to find a job because of their observable characteristics could sort in municipalities with bad observable neighborhood attributes (for instance where there are many foreigners, as these neighborhoods could be redlined by xenophobic employers). Second, they could also sort themselves in municipalities with bad unobservable attributes (for instance in municipalities which have a bad reputation among employers for some unobserved reason). Third, municipality aggregate of observed and unobserved individual characterics could be positively correlated (for instance workers with no diploma may be less efficient in job-search). In our opinion, one of the main result of this paper is thus to show that disparities in individual characteristics are reinforced by disparities in local characteristics due to residential sorting.

Finally, we investigate whether places that enhance job finding do slow down the exit to non-employment. To do that, we compute the correlations between the municipality integrated baseline hazards for finding a job and for exit to non-employment at 6, 12 and 24 months (weighing by the number of unemployed workers at risk). We find that for short and medium horizons (6 and 12 months), there is little correlation between the two types of local effects (resp. $-.028$ and .033). However, in the long run (24 months), the correlation is positive and stands at .176. In municipalities where job exits are more likely to occur, exits to non-employment are also more likely to take place at least in the long run. This result can be understood by comparing reservation wages, shadow wages and job offers. In this framework, unemployed workers exit to non employment when their reservation wage falls below their shadow wage. Our result suggests that this is more likely to happen in the long run in municipalities where unemployed workers are more likely to exit to a job. This could happen if municipalities where residents can easily find a job are also those where holding a job is more likely. Spouses could thus more likely become non participant because their opportunity cost of time increases with their spouse's income.

5.3 Municipality Effects and Spatial Characteristics

5.3.1 Multiplicative Component Model

As explained above, we further restrict the hazard function and consider a multiplicative municipality hazard specified as the product of a municipality effect and an aggregate baseline hazard (see Equation (5)). To implement this approach, we divide the time line into $M = 9$ intervals, with the first eight intervals lasting 90 days and the remaining one lasting the rest of the period. To assess whether the multiplicative specification is too restrictive, we compare the value of disparity indices obtained with the unspecified municipality hazard with those obtained with the multiplicative hazard (see Table 4 and Table 5). We find that the multiplicative hazard reproduces well spatial disparities for finding a job although it performs poorly for exit to non-employment at 6 months (though not for the Gini). This good fit justifies the use of municipality effects as an adequate summary to study the determinants of spatial disparities for finding a job.[13]

In line with the theories presented in Section 2, we investigate how municipality fixed effects can be explained by segregation and job accessibility. Segregation is measured here by the composition of the municipality population by education and by nationality. Job accessibility is measured by local job density (as defined in Section 3.2).

5.3.2 Partial Correlations with Spatial Caracteristics

Table 6 reports various regressions of municipality effects on those spatial characteristics. We computed a pseudo-R^2 to assess the explanatory power of the model taking into account the sampling error (see Appendix B.3). When using only segregation indices as explanatory variables (column 1), we are able to explain 72.4% of the variance of municipality fixed effects. Job accessibility indices (column 2) have a much lower explanatory power since the pseudo R^2 is only 25.9%. This suggests that spatial disparities in finding a job are more strongly associated with differences in the local level of segregation than with variations in job accessibility. When using both segregation and job accessibility indices (column 3), the pseudo R^2 reaches 73.0%.

We now comment on the coefficient of the latter regression (column 3). Large municipality

[13] Results in Table 5 show more discrepancies between the multiplicative model and the unrestricted model in terms of spatial variance and correlations.

effects in finding a job are associated with a large proportion of unskilled workers and of non-French citizens (especially citizens from sub-Saharan Africa). This is consistent with the existence of redlining (according to nationality and skill) as well as with a social network effect. Municipality effects in finding a job are also correlated with local job accessibility, especially by private transport, but the coefficients of both private and public job accessibility measures are negative in contradiction with the spatial mismatch theory.

[*Insert Table* 6]

Of course, there are some other interpretations of the results which are based on possible omitted local variables, reverse causality or sorting on individual unobservables. There can be omitted local variables correlated with segregation or job-accessibility measures. Our surprising result for job accessibility could be explained if the job density indices captured the low quality and high congestion of transports for instance.

Reverse causality can occur if local unemployment acts as an attraction or a repulsion force on population and jobs. This could affect the job accessibility measure and the segregation indices (provided that the population categories are differentially attracted or repulsed). To take the example of segregation for instance, French people may flee municipalities where it is more difficult for unemployed workers to find a job. This would increase the local proportion of foreigners, especially Africans and could explain the negative coefficient of the municipality proportion of Africans on finding a job.

Municipality explanatory variables can capture the local average of individual unobserved variables if there is a correlation between $Z^j\gamma$ and the omitted term ε^j as defined in equation (12) above. This is the case for instance when individuals with a given unobserved attribute (such as motivation to search for a job) choose their location depending on observable municipality variables (attractive residential neighborhoods where jobs are not easily accessible). This may explain the negative effect of the job accessibility index by private transport.

6 Conclusion

In this paper, we study the spatial disparities in exits from unemployment across municipalities in the Paris region. We use a unique and exhaustive administrative dataset which contains all registered unemployment spells over the 1996-2003 period. This dataset contains some individual characteristics of unemployed workers as well as their residential location. It is merged with spatial indices of segregation and job accessibility computed from the census and a transport survey.

Our methodology is based on the estimation of independent competing risk duration models with two exits (finding a job and dropping out of the labor force). We constructed measures of raw spatial disparities across municipalities from the local survival functions after 24 months. We find that there are very large disparities. The local composition of workers' characteristics can explain around 30% of the disparities in finding a job. Our local indices (especially residential segregation measures) capture nearly 70% of the remaining differences. Furthermore, we showed that disparities in individual characteristics are reinforced by disparities in local characteristics due to residential sorting. The latter finding lends credit to the idea that spatial factors exacerbate non spatial factors in the determination of unemployment, or that the most fragile unemployed workers tend to cumulate local and invididual disadvantages.

Our work nevertheless considered broad local effects related to segregation and job accessibility without trying to investigate and disentangle the specific mechanisms at work, which could be the basis for furture research. Another extension of this work could be to compute municipality survival functions by nationality group or class of diploma. This would enable us to assess the extent to which the effect of local factors may differ for these groups. It would also be interesting to study spatial disparities at a much finer scale were the data available. Indeed, our accessibility measures are only at the municipality level whereas accessibility can differ even between two small neighborhoods (e.g. when they are separated by a railroad). Working at a finer geographic scale may also allow for an investigation of other important issues such as the role of spatialized social networks, which are likely to occur within a limited geographic area (see Bayer, Ross and Topa, 2008; Gobillon and Selod, 2007).

A Data Appendix

Over the 1993-2003 period, the panel contains 10,290,225 unemployment spells. We selected unemployment spells beginning between January 1 and June 30, 1996 which form a subsample of 451,191 unemployment spells. By keeping observations corresponding to unemployed workers between 16 and 54 years old only, we ended up with a dataset comprising 433,802 unemployment spells. After deleting observations with missing values and coding problems, the final sample is composed of 430,695 observations. Descriptive statistics about variables in our final dataset are given in Table A1.

B Computational details

B.1 First-stage estimation

We want to test for each municipality that the empirical survival function as estimated by Kaplan-Meier estimation is equal to the survival function predicted by the model.

B.1.1 Construction of the Kolmogorov test statistic

Let $S(t,k)$ be the survival function of the data for exit k and let $S(t,k \mid x,\delta)$ be the conditional survival function of the model. Here, δ denotes parameters β_k and the municipality baseline hazard. The null hypothesis writes:

$$H_0 : S(t,k) = \int S(t,k \mid x,\delta) dF(x)$$

where $dF(x)$ is the probability measure of covariates. This is an adaptation of Andrews (1997) with some differences since the latter paper considers null hypotheses of the form:

$$H_0 : S(t,k\,|x)F(x) = S(t,k\,|x,\delta)F(x) \text{ a.s. } F(x).$$

where $S(t,k\,|x)$ is the conditional survival function of the data. The adaptation of the proofs of Andrews (1997) are out of the scope of this paper.

Our sample consists in individuals, $i = 1,...,N$, of characteristics X_i for whom we observe unemploment duration t_i and type of exit. We restrict our attention to exits to job and we drop

index k from the survival functions. Computing the test statistic for exits to non-employment follows the same principles.

Let $\hat{H}_N^j(t)$ be the Kaplan-Meier estimator of the survival function at duration t in municipality j. Let $\hat{S}_N^j(t;\delta)$ be the survival function until an exit to a job in municipality j, as predicted by the model i.e.

$$\hat{S}_N^j(t) = \frac{1}{N_j} \sum_{i=1, i\in j}^{N} S_i(t \mid X_i, \hat{\delta}),$$

where N_j denotes the number of unemployed workers in municipality j, and where $\hat{\delta}$ denotes the SPLE estimator and the functional Breslow estimator of the baseline hazard rate. The conditional Kolmogorov statistic for municipality j is:

$$CK_N^j = \sqrt{N_j} \max_{i\in j} \left| \hat{H}_N^j(t_i) - \hat{S}_N^j(t_i) \right|$$

In practice, we trim out durations in the last percentile when computing our test statistic. This is because the survival functions are not estimated with accuracy at that percentile and the test statistic takes artificially large values at finite distance.

Alternatively, we could also consider another statistic which is the mean square difference of the two survival functions, in analogy with a Cramer-von Mises statistic:

$$QM_N^j = \frac{1}{\sqrt{N_j}} \sum_{i\in j} \left(\hat{H}_N^j(t_i) - \hat{S}_N^j(t_i) \right)^2$$

For the two test statistics, we need to compute the distribution under the null hypothesis. We proceed by bootstrap as proposed by Andrews (1997).

B.1.2 Computation of the distribution of the test statistic.

We now explain how to compute the distribution of the test statistic for a given municipality. We drop index j for simplicity. For an individual i having a censored unemployment spell, the duration before censorship, $t_{ic} = t_i$, is an exogenous characteristic of the individual. If the unemployment spell is not censored, censorship is not relevant and its duration is not taken into account. We denote $\widetilde{X}_i$ the exogenous information i.e. $\widetilde{X}_i = (X_i, t_{ic})$ for censored individuals and $\widetilde{X}_i = (X_i, \cdot)$ for uncensored ones.

The asymptotic distribution of Andrews' test statistic is computed by (semi)-parametric bootstrap for B replications, $b = 1, ..., B$. For each replication, we generate a duration for each individual using the proportional hazard model. The data are generated conditionally on the exogenous characteristics $\widetilde{X}_i$, the estimated parameters for exit to job $\widehat{\beta}_e$ and the Breslow estimator. The procedure to simulate durations is the following.

For an individual i, the integrated hazard for exit to job at duration t writes:

$$\Lambda\left(t\left|X_i\right.\right) = \Lambda\left(t\right)\exp\left(X_i\beta_e\right)$$

where $\Lambda\left(t\right)$ is the integrated baseline hazard. As the survival function is $S\left(t\left|X_i\right.\right) = \exp\left[-\Lambda\left(t\left|X_i\right.\right)\right]$, we have:

$$\Lambda\left(t\right) = \frac{-1}{\exp\left(X_i\beta_e\right)}\ln S\left(t\left|X_i\right.\right)$$

To obtain simulated durations, we draw the value of the survival function in a uniform distribution $[0, 1]$, $\widetilde{S}_i^b$, and replace unknown parameters by their estimates:

$$\widehat{\Lambda}_i^b = \frac{-1}{\exp\left(X_i\widehat{\beta}_e\right)}\ln\widetilde{S}_i^b$$

There remains to invert function $\widehat{\Lambda}$ at point $\widehat{\Lambda}_i^b$ to recover duration $\widetilde{t}_i^b$ for individual i. In practice, $\widehat{\Lambda}_i^b$ increases piecewise and there is a duration at which the function $\widehat{\Lambda}$ makes a jump from $\underline{\Lambda}_i^b$ to $\overline{\Lambda}_i^b$ such that $\underline{\Lambda}_i^b \leqslant \widehat{\Lambda}_i^b < \overline{\Lambda}_i^b$. We define $\widetilde{t}_i^b$ as the duration at which function $\widehat{\Lambda}$ makes this jump.

A practical issue is that $\widehat{\Lambda}$ cannot be computed above the upper bound $\widehat{\Lambda}\left(t^{\max}\right)$, where $t^{\max}$ is the largest duration in the municipality sample. If a simulated value is such that $\widehat{\Lambda}_i^b > \widehat{\Lambda}\left(t^{\max}\right)$, we set duration to $\widetilde{t}_i^b = t^{\max}$. This small sample bias should disappear asymptotically when the number of individuals in each municipality tends to infinity.

For individuals who were not censored, the generated duration is $t_i^b = \widetilde{t}_i^b$. For individuals who were right-censored, the generated duration is $t_i^b = \min\left(\widetilde{t}_i^b, t_{ic}\right)$. The generated exit is finding a job if $\widetilde{t}_i^b < t_{ic}$, and it is censorship if $\widetilde{t}_i^b > t_{ic}$. We then construct the Kaplan-Meier's estimator denoted $\hat{H}_N^{j,b}\left(t\right)$ where the index j of the municipality has been reintroduced. For individual i, we compute the value of the Kaplan-Meier's estimator of his municipality at the generated duration as: $\hat{H}_i^b = \hat{H}_N^{j,b}\left(t_i^b\right)$. In the same way, we use the Breslow's estimator of any municipality j to construct a survival function which is denoted $\hat{S}_N^{j,b}\left(t\right)$. For individual i, we compute the value of the survival function of his municipality at the generated duration as: $\hat{S}_i^b = \hat{S}_N^{j,b}\left(t_i^b\right)$.

The test statistic of municipality j computed for the bth replication can be written as:

$$CK_N^{j,b} = \sqrt{N_j} \sup_{i \in j} \left| \hat{H}_i^b - \hat{S}_i^b \right|$$

As previously, we trim out durations in the last percentile when computing our test statistic. B bootstrap samples of size N_j are simulated. The p-value of the test statistic is defined as:

$$p_{jN}^B = \frac{1}{B} \sum_{b=1}^{B} \mathbf{1}\{CK_N^{j,b} > CK_N^j\}.$$

B.2 Second-stage estimation

B.2.1 Finite-sample issues

We first explain how we take into account finite sample issues when establishing equation (6). For that purpose, we refined appropriatly the quantities involved in (6). We divide the period into M intervals $[t_{m-1}, t_m]$, $m = 1, ..., M$. We denote $\theta_m = \frac{1}{t_m - t_{m-1}} \int_{t_{m-1}}^{t_m} \theta(s)\, ds$ the average baseline hazard over the interval m and $d_m^j = \int_{t_{m-1}}^{t_m} I(C^j(s) > 0)\, ds$ the length of time within interval m when some individuals in municipality j are at risk. In particular, $d_m^j < t_m - t_{m-1}$ in the last time interval in which there are some unemployed workers at risk in municipality j. The average hazard rate over a time interval m in municipality j where some people are at risk ($d_m^j > 0$) is given by $y_m^j = \frac{1}{d_m^j} [\Theta^j(t_m) - \Theta^j(t_{m-1})]$. An estimator of this average hazard rate can be constructed from equation (4) and writes: $\hat{y}_m^j = \frac{1}{d_m^j} [\widehat{\Theta}^j(t_m) - \widehat{\Theta}^j(t_{m-1})]$. We can then re-establish formula (6) where the quantities have been redefined.

B.2.2 Covariance matrix of the sampling errors

We now give the formulas to compute the covariance matrix of $(\varepsilon_m^j)_{j,m}$,which are the sampling errors in equation (6), using Ridder and Tunali's appendix (RT hereafter). We first introduce the following notations that will be used below:

$$\begin{aligned} S_j^0(\beta, s) &= \sum_{i \in \Omega^j(s)} \exp(X_i \beta) \\ S_j^1(\beta, s) &= \sum_{i \in \Omega^j(s)} X_i \exp(X_i \beta) \end{aligned}$$

where $\Omega^j(s)$ is the set of unemployed workers still at risk in municipality j at time s. Note that whereas $S_j^0(\beta, s)$ is a 1×1 matrix, $S_j^1(\beta, s)$ is a $1 \times K$ matrix, where K is the number of explanatory variables in the first stage. We also denote $C^j(s) = card\ \Omega^j(s)$ the number of unemployed workers still at risk in municipality j at time s. According to RT (A28), we have:

$$\exp \varepsilon_m^j = \eta_m^j + \frac{1}{\sqrt{N}} c_{jm}' \xi \tag{13}$$

where $N = \sum_j C^j(0)$ is the number of unemployed workers in the Paris region and:

$$\eta_m^j = \tfrac{1}{d_m^j} \int_{t_{m-1}}^{t_m} I\left(C^j(s) > 0\right) \left[\tfrac{1}{S_j^0(\beta,s)} dN^j(s) - \theta^j(s)\, ds\right] \quad \text{(RT A22)}$$

$$c_{jm} = -\tfrac{1}{d_m^j} \int_{t_{m-1}}^{t_m} I\left(C^j(s) > 0\right) \tfrac{S_j^1(\beta^*,s)}{\left[S_j^0(\beta^*,s)\right]^2} dN^j(s) \quad \text{(RT A27)}$$

$$\xi = \sqrt{N}\left(\widehat{\beta} - \beta\right)$$

where β^* is a value between β and $\widehat{\beta}$ (coming from a Taylor expansion not detailed here), $dN^j(s)$ is a dummy that equals one if someone in municipality j experiences an exit in an arbitrarily short period of time before date s (and zero otherwise), and $d_m^j = \int_{t_{m-1}}^{t_m} I\left(C^j(s) > 0\right) ds$. Here, ξ is uncorrelated with η_m^j. From equation (13), it is possible to get:

$$V\left(\exp \varepsilon_m^j\right) = V\left(\eta_m^j\right) + c_{jm}' V c_{jm} \quad \text{(RT A29)}$$

$$cov\left(\exp \varepsilon_m^j, \exp \varepsilon_n^k\right) = c_{jm}' V c_{kn} \text{ for } j \neq k \text{ or } m \neq n \quad \text{(RT A30)}$$

where $V = V\left(\widehat{\beta}\right)$. These covariance-matrix terms of $\left(\exp \varepsilon_m^j\right)_{j,m}$ can be estimated computing estimators of all terms on the right-hand sides. An estimator of V is obtained from the Fisher information matrix of SPLE. In practice, there is no need to have the theoretical formula to get this estimator as it is directly recovered from the estimation software. Some estimators of $V\left(\eta_m^j\right)$ and c_{jm} are:

$$\widehat{V}\left(\eta_m^j\right) = \tfrac{1}{\left(d_m^j\right)^2} \int_{t_{m-1}}^{t_m} I\left(C^j(s) > 0\right) \tfrac{1}{\left[S_j^0\left(\widehat{\beta},s\right)\right]^2} dN^j(s) \quad \text{(from RT A25)}$$

$$\widehat{c}_{jm} = -\tfrac{1}{d_m^j} \int_{t_{m-1}}^{t_m} I\left(C^j(s) > 0\right) \tfrac{S_j^1\left(\widehat{\beta},s\right)}{\left[S_j^0\left(\widehat{\beta},s\right)\right]^2} dN^j(s) \quad \text{(from RT A27)}$$

These estimators have to be programmed to be computed. From the covariance matrix of $(\exp \varepsilon_m^j)_{j,m}$, we get the covariance matrix of $(\varepsilon_m^j)_{j,m}$ using the delta method.

B.2.3 Estimation

Formulas

We first give some notations we use in this section. Denote J the number of municipalities and M the number of time intervals. For any $JM \times 1$ matrix X, X_j refers to the $M \times 1$ matrix defined by $X_{(j-1)M+[1:M],1}$. For any given $JM \times JM$ matrix X, $X_{j,k}$ refers to the $M \times M$ submatrix defined by $X_{(j-1)M+[1:M],(k-1)M+[1:M]}$ where $[1:M]$ is the vector of integers from 1 to M.
The equation to estimate is (6) where we fix $\theta_1 = 1$ to secure identification. We stack the observations of (6) and obtain:

$$Y = A\alpha + G\theta + \varepsilon \tag{14}$$

where A is a $JM \times J$ matrix such that $A_{(j-1)M+m,k} = 1$ if $j = k$ and $A_{(j-1)M+m,k} = 0$ otherwise, G is a $JM \times (M-1)$ matrix such that $G_{(j-1)M+m,l} = 1$ if $m = l$ and $A_{(j-1)M+m,l} = 0$ otherwise, $Y = \left(\ln y_1^1, ..., \ln y_M^J\right)'$ and $\varepsilon = \left(\varepsilon_1^1, ..., \varepsilon_M^J\right)'$ are some $JM \times 1$ vectors, $\alpha = (\ln \alpha_1, ..., \ln \alpha_J)'$ is a $J \times 1$ vector and $\theta = (\ln \theta_2, ..., \ln \theta_M)'$ is a $(M-1) \times 1$ vector.

Denote $\Delta = diag\,(N_{11}, ..., N_{JM})$ the $JM \times JM$ diagonal matrix where N_{jm} is the number of unemployed workers in municipality j still at risk at the beginning of interval m. After weighting equation (14) with $\Delta^{1/2}$, it becomes:

$$\Delta^{1/2}Y = \Delta^{1/2}A\alpha + \Delta^{1/2}G\theta + \Delta^{1/2}\varepsilon$$

Denote W the projector in the dimension orthogonal to $\Delta^{1/2}A$. Using the first stage of Frisch-Waugh theorem, we obtain the WLS estimator of θ:

$$\begin{aligned} \widehat{\theta} &= (G'\Delta^{1/2}W\Delta^{1/2}G)^{-1}G'\Delta^{1/2}W\Delta^{1/2}Y \\ &= \theta + (G'\Delta^{1/2}W\Delta^{1/2}G)^{-1}G'\Delta^{1/2}W\Delta^{1/2}\varepsilon \end{aligned} \tag{15}$$

The second stage of the Frisch-Waugh theorem gives the WLS estimator of α:

$$\begin{aligned}
\widehat{\alpha} &= (A'\Delta A)^{-1}A'\Delta\left[Y - G\widehat{\theta}\right] \\
&= (A'\Delta A)^{-1}A'\Delta\left[Y - G\theta - G\left(G'\Delta^{1/2}W\Delta^{1/2}G\right)^{-1}G'\Delta^{1/2}W\Delta^{1/2}\varepsilon\right] \\
&= \alpha + (A'\Delta A)^{-1}A'\Delta\left[\varepsilon - G\left(G'\Delta^{1/2}W\Delta^{1/2}G\right)^{-1}G'\Delta^{1/2}W\Delta^{1/2}\varepsilon\right]
\end{aligned}$$

Denote $\Gamma = A'\Delta A$, $\Phi = G'\Delta^{1/2}W\Delta^{1/2}G$ and $\Psi = G'\Delta^{1/2}W\Delta^{1/2}V\Delta^{1/2}W\Delta^{1/2}G$, where $V = V(\varepsilon)$. We have:

$$V\left(\widehat{\theta}\right) = \Phi^{-1}\Psi\Phi^{-1} \tag{16}$$

Also, we get:

$$\begin{aligned}
V(\widehat{\alpha}) &= \Gamma^{-1}A'\Delta V\Delta A\Gamma^{-1} \\
&\quad +\Gamma^{-1}A'\Delta GV\left(\widehat{\theta}\right)G'\Delta A\Gamma^{-1} \\
&\quad -\Gamma^{-1}A'\Delta\left(V\Delta^{1/2}W\Delta^{1/2}G\Phi^{-1}G' + G\Phi^{-1}G'\Delta^{1/2}W\Delta^{1/2}V\right)\Delta A\Gamma^{-1}
\end{aligned} \tag{17}$$

Computation

We have:

$$\begin{aligned}
\Phi &= \sum_{j=1}^{J}\left(W\Delta^{1/2}G\right)_j'\left(W\Delta^{1/2}G\right)_j = \sum_{j=1}^{J}\overline{G}_j'\Delta_{j,j}\overline{G}_j \\
\Psi &= \sum_{j,k=1}^{J}\left(W\Delta^{1/2}G\right)_j'\Delta_{j,j}^{1/2}V_{j,k}\Delta_{k,k}^{1/2}\left(W\Delta G\right)_k = \sum_{j,k=1}^{J}\overline{G}_j'\Delta_{j,j}V_{j,k}\Delta_{k,k}\overline{G}_k \\
\Gamma &= diag\left[tr\left(\Delta_{1,1}\right),...,tr\left(\Delta_{J,J}\right)\right]
\end{aligned}$$

where for any given variable Z_j of dimension $M\times 1$, $\overline{Z}_j$ is its counterpart centered with its weighted average: $\overline{Z}_j = Z_j - \frac{1}{tr(\Delta_{j,j})}tr\left(\Delta_{j,j}Z_j\right)$.

We also have:

$$\begin{aligned}
\left(A'\Delta V\Delta A\right)_{j,k} &= N_j'V_{j,k}N_k \\
\left(A'\Delta GV\left(\widehat{\theta}\right)G'\Delta A\right)_{j,k} &= N_{j-}'V\left(\widehat{\theta}\right)N_{k-}
\end{aligned}$$

where $N_j = \left(N_{j,1},...,N_{j,M}\right)'$ and $N_{j-} = \left(N_{j,2},...,N_{j,M}\right)'$.

Moreover, $V\Delta^{1/2}W\Delta^{1/2} = \overline{V}\Delta$ where $\overline{V}$ is defined such that any of its given submatrix $\overline{V}_{j,k}$ writes: $\overline{V}_{j,k} = V_{j,k} - \frac{1}{tr(\Delta_{j,j})}\left(1_M \otimes N_j'\right) V_{j,k}$ with $\otimes$ the Kronecker product and 1_M a $M \times 1$ matrix filled with the value 1. Hence, we have:

$$\left(A'\Delta V\Delta^{1/2}W\Delta^{1/2}G\Phi^{-1}G'\Delta A\right)_{j,k} = N_j'\left(\overline{V}'\Delta G\Phi^{-1}G'\right)_{j,k} N_k$$

Moreover, $G\Phi^{-1}G' = J.J_J \otimes \begin{pmatrix} 0 & 0 \\ 0 & \Phi^{-1} \end{pmatrix}$ with J_J the $J \times J$ matrix filled with the value $1/J$.

Hence, $\left(\overline{V}\Delta G\Phi^{-1}G'\right)_{j,k} = \sum_l \left(\overline{V}\Delta\right)_{j,l} \begin{pmatrix} 0 & 0 \\ 0 & \Phi^{-1} \end{pmatrix} = \left(\sum_l \overline{V}_{j,l}\Delta_{l,l}\right) \begin{pmatrix} 0 & 0 \\ 0 & \Phi^{-1} \end{pmatrix}$.

B.3 Third-stage estimation

The third-stage equation to estimate is given by (8). When we stack the observations, we obtain:

$$\widehat{\alpha} = Z\gamma + \eta + \xi \tag{18}$$

where $\widehat{\alpha} = \left(\widehat{\ln \alpha_1}, ..., \widehat{\ln \alpha_J}\right)'$, $\eta = (\eta_1, ..., \eta_J)'$ and $\xi = (\xi_1, ..., \xi_J)'$ are some $J \times 1$ vectors, and $Z = (Z_1', ..., Z_J')'$ is a $J \times K$ matrix. We suppose that $\left(\eta_j\right)_{1,...,J}$ have a covariance matrix $\upsilon^2 Q^{-1}$ where $Q = diag(N_{11}, ..., N_{J1})$. Equation (18) is estimated with weighted least squares where the weights are the square-roots of the numbers of unemployed workers at the initial date ($Q^{1/2}$). The estimated coefficients write:

$$\widehat{\gamma} = \left(Z'QZ\right)^{-1} Z'Q\widehat{\alpha}$$

and their covariance matrix is:

$$\begin{aligned} V\left(\widehat{\gamma}\right) &= \left(Z'QZ\right)^{-1} Z'Q\left[V\left(\xi\right) + \upsilon^2 Q^{-1}\right] QZ\left(Z'QZ\right)^{-1} \\ &= \left(Z'QZ\right)^{-1} Z'QV\left(\xi\right) QZ\left(Z'QZ\right)^{-1} + \upsilon^2\left(Z'QZ\right)^{-1} \end{aligned}$$

It is possible to construct a consistent estimator of υ^2 using the residuals $\widehat{\eta + \xi} = Q^{1/2}\widehat{\alpha} - Q^{1/2}Z\widehat{\gamma}$. This estimator is found from the following calculation sequence:

$$\widehat{\eta + \xi}'\widehat{\eta + \xi} = (\eta + \xi)'\left[I - Q^{1/2}Z\left(Z'QZ\right)^{-1} Z'\right]' Q\left[I - Z\left(Z'QZ\right)^{-1} Z'Q^{1/2}\right](\eta + \xi)$$

where we made the approximation (for N large enough) that:

$$\widehat{\eta+\xi}'\widehat{\eta+\xi} \approx (\eta+\xi)' Q (\eta+\xi)$$

We thus have:

$$E\left[\widehat{\eta+\xi}'\widehat{\eta+\xi}\right] \approx \upsilon^2 J + tr\left[QV(\xi)\right]$$

when $V(\xi)$ has been computed from the first-stage estimation. An estimator of υ^2 can then be defined as:

$$\widehat{\upsilon}^2 = \left[\widehat{\eta+\xi}'\widehat{\eta+\xi} - tr\left[QV(\xi)\right]\right]/J$$

We introduce an error rate coming from sampling error as:

$$err = \frac{tr\left[QV(\xi)\right]}{\widehat{\eta+\xi}'\widehat{\eta+\xi}}$$

We also construct a pseudo-R^2 defined as:

$$R_p^2 = \frac{V_Q^e(Z\widehat{\gamma})}{V_Q^e(Z\widehat{\gamma}) + \widehat{\upsilon}^2 J}$$

where $V_Q^e(\cdot) = \left(Z\widehat{\gamma} - \overline{Z}\widehat{\gamma}\right)' Q \left(Z\widehat{\gamma} - \overline{Z}\widehat{\gamma}\right)/tr(Q)$ is the empirical variance obtained when weighting observations with weights Q (where $\overline{Z} = tr(QZ)/trQ$). Note that when there is no sampling error, this pseudo-R^2 is equal to the usual R^2.

References

[1] Altonji J.G. and L.M. Segal (1996), "Small Sample Bias in GMM estimation of Covariance Structures", Journal of Business & Economic Statistics, 14(3), 353-65.

[2] Andrews D. (1997), "A Conditional Kolmogorov test", Econometrica, 65(5), 1097-1128

[3] Baccaini B., Courgeau D. and G. Desplanques (1993), "Les migrations internes en France de 1982 à 1990. Comparaison avec les périodes antérieures", 48(6), 1771-89.

[4] Baker M. and A. Melino (2000), "Duration Dependence and Non-Parametric Heterogeneity: A Monte-Carlo Study", Journal of Econometrics, 96, 357-93.

[5] Bayer P., Ross S. and G. Topa (2008), "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes", Journal of Political Economy, 116(6), 1150-96.

[6] Breslow N.E. (1974), "Covariance Analysis of Censored Survival Data", Biometrics, 30, pp. 89-99.

[7] Calvo-Armengol A. and M. Jackson (2004), "Social networks in determining unemployment: Patterns, dynamics and inequality", American Economic Review, 94, 191-206.

[8] Chardon O. and D. Goux (2003), "The new european definition of BIT unemployment" (in French), Economie et Statistiques, 362, 67-83.

[9] Conley T. and G. Topa (2002), "Socio-economic distance and spatial patterns in unemployment", Journal of Applied Econometrics, 17(4), 303-327.

[10] Cutler D. and E. Glaeser (1997), "Are Ghettos Good or Bad?", Quarterly Journal of Economics, 112, 827-872.

[11] Dawkins C., Shen Q. and T. Sanchez (2005), "Race, Space and Unemployment Duration", Journal of Urban Economics, 58, 91-113.

[12] DREIF (Direction Régionale de l'Equipement Ile-de-France) / INSEE Ile-de-France (1997), "Daily moves: Ile-de-France population more mobile" (in French), 2 pages.

[13] Gobillon L. and H. Selod (2007), "The effect of segregation and spatial mismatch: evidence from France", CEPR Working Paper 6198.

[14] Gobillon L., Selod H. and Y. Zenou (2007), "The mechanisms of spatial mismatch", Urban Studies, 44 (12), 2401-27 .

[15] Gouriéroux C., A. Monfort, E. Renault and A. Trognon (1985), "Asymptotic Least Squares" (in French), Annales de l'INSEE, 58, 91-122.

[16] Holzer H. and K. Ihlanfeldt (1998) "Customer discrimination and employment outcomes for minority workers", Quarterly Journal of Economics, 113, 835-867.

[17] Holzer H., Ihlanfeldt K. and D. Sjoquist (1994), "Work, search and travel among white and black youth", Journal of Urban Economics, 35, 320-345.

[18] Horowitz J., (1998), "Bootstrap Methods for Covariance Structures", Journal of Human Resources, 33, 39-61.

[19] Ihlanfeldt K. (1993), "Intra-urban job accessibility and Hispanic youth employment rate", Journal of Urban Economics, 33, 254-271.

[20] Ihlanfeldt K. (1997), "Information on the spatial distribution of job opportunities within Metropolitan Areas", Journal of Urban Economics, 41, 218-242.

[21] Ihlanfeldt K. and D. Sjoquist (1998), "The spatial mismatch hypothesis: a review of recent studies and their implications for welfare reform", Housing Policy Debate, 9, 849-892.

[22] Johnson R. (2006) "Landing a job in urban space: the extent and effects of spatial mismatch", Regional Science and Urban Economics, 36, 331-72.

[23] Lancaster T. (1990), The Econometric Analysis of Transition Data, Cambridge University Press: Cambridge.

[24] Manski C. F. (1993), "Identification of Endogenous Social Effects: The Reflection Problem", Review of Economic Studies, 60(3), 531-42.

[25] Maurin E. (2004), The French Ghetto (in French), ed. Seuil, 95p.

[26] Moulton B.R. (1990), "An Illustration of Pitfalls in estimating the Effects of Aggregate variables on Micro Units", Review of Economics and Statistics, 72, 334-8.

[27] Mortensen D. et T. Vishwanath (1994), "Personal contacts and earnings. It is who you know!", Labour Economics, 1, 187-201.

[28] Patacchini E. Y. Zenou (2006), "Search activities, cost of living, and local labor markets", Regional Science and Urban Economics, 36(2), 227-48.

[29] Ridder G. and I. Tunali (1999), "Stratified partial likelihood estimation", Journal of Econometrics, 92(2), 193-232.

[30] Rogers C. (1997), "Job Search and Unemployment Duration: Implications for the Spatial Mismatch Hypothesis", Journal of Urban Economics, 42, 109-32.

[31] Selod H. and Y. Zenou (2006), "City structure, job search, and labor discrimination. Theory and policy implications", Economic Journal, 116, 1057-87.

[32] Smith T. and Y. Zenou (2003), "Spatial mismatch, search effort and urban spatial structure", Journal of Urban Economics, 54, 129-156.

[33] Stoll M. (1999), "Spatial job search, spatial mismatch, and the employment and wages of racial and ethnic groups in Los Angeles", Journal of Urban Economics, 46, 129-55.

[34] Stoll M. and S. Raphael (2000), "Racial differences in spatial job search patterns: Exploring the causes and consequences", Economic Geography, 201-223.

[35] Weinberg B. (2000), "Black residential centralization and the spatial mismatch hypothesis", Journal of Urban Economics, 48, 110-134.

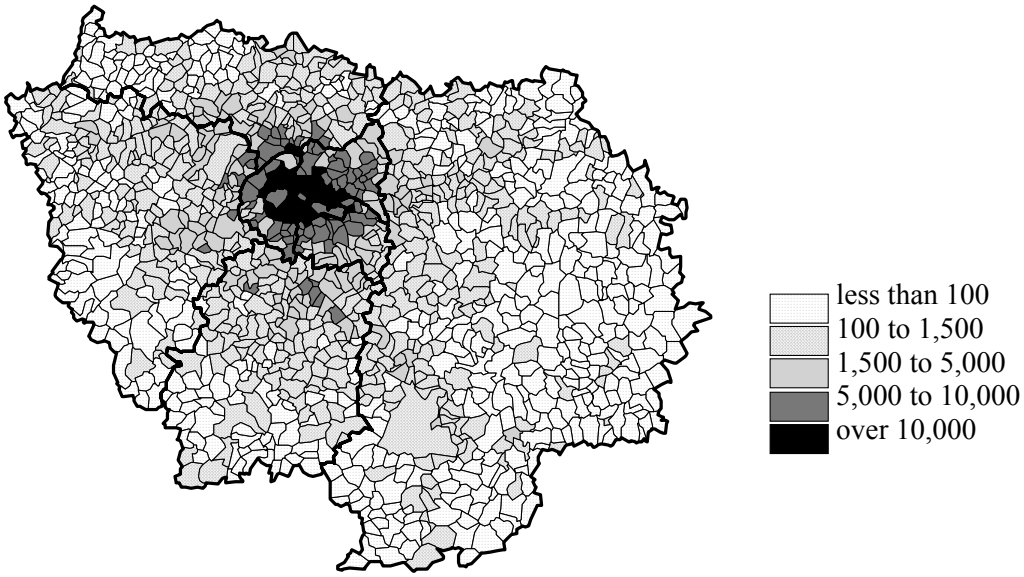
[36] Weinberg B. (2004), "Testing the spatial mismatch hypothesis using inter-city variations in industrial composition", Regional Science and Urban Economics, 34, 505-32.

[37] Wilson,J. (1996), When Work Disappears: The World of the New Urban Poor, New York: Alfred A. Knopf.

[38] Zax J. and J. Kain (1996), "Moving to the suburbs: Do relocating companies leave their Black employees behind?", Journal of Labor Economics, 14, 472-504.

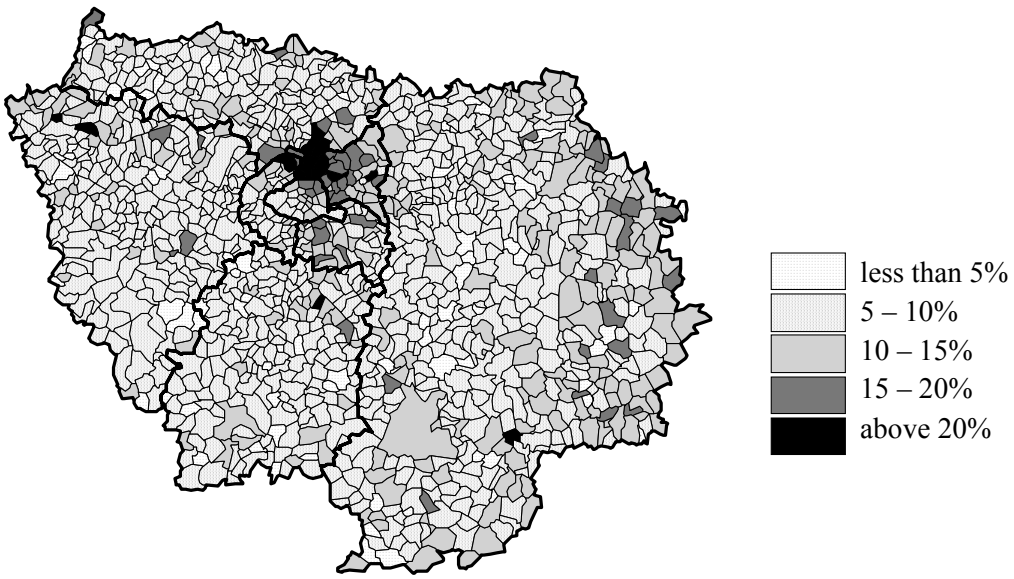
[39] Zenou Y. (2002), "How do firms redline workers?", Journal of Public Economics, 52(3), pages 391-408.

Graph 1: Population density (per sq km) in the Paris region in 1999



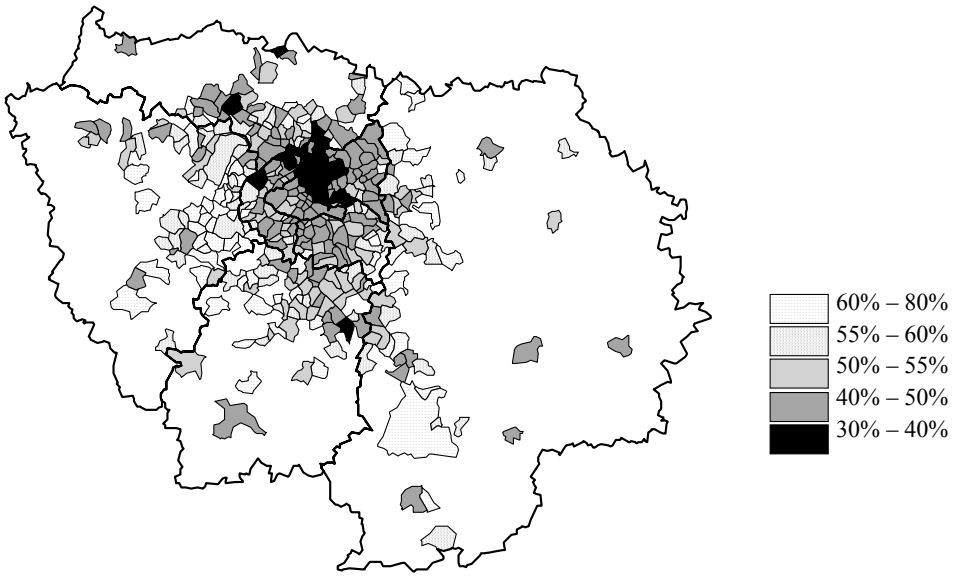
Source: constructed from the 1999 Population Census, INSEE. The geographical unit is the subdistrict for the city of Paris or the municipality for the rest of the region. Bold lines represent the boundaries of the city of Paris (the turtle-shaped area in the middle of the map) and of the seven surrounding subregional administrative districts (*départements*).

Graph 2: Unemployment rates in the Paris region in 1999



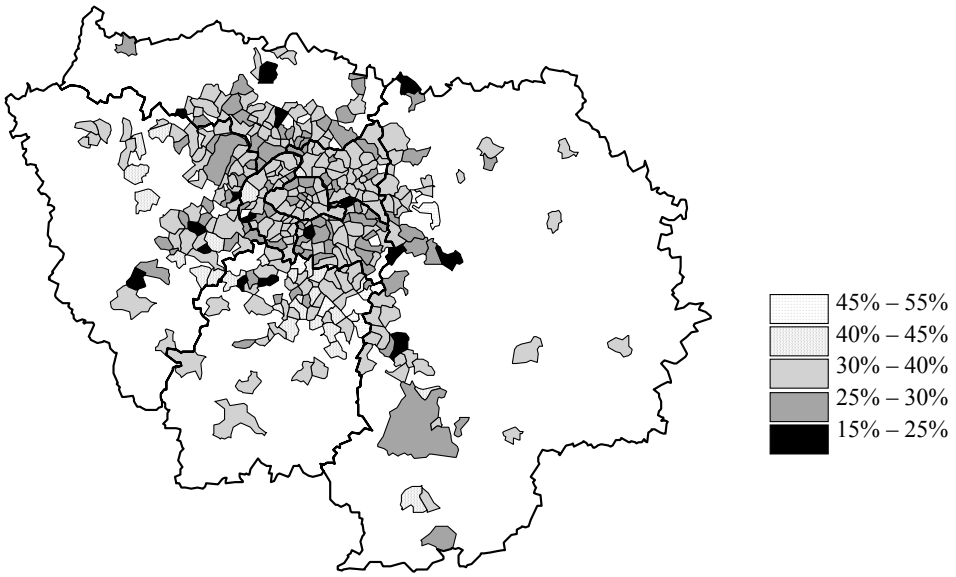
Source: constructed from the 1999 Population Census, INSEE.

Graph 3: probability of finding a job before 24 months (Kaplan-Meier) for municipalities with at least 5,000 inhabitants



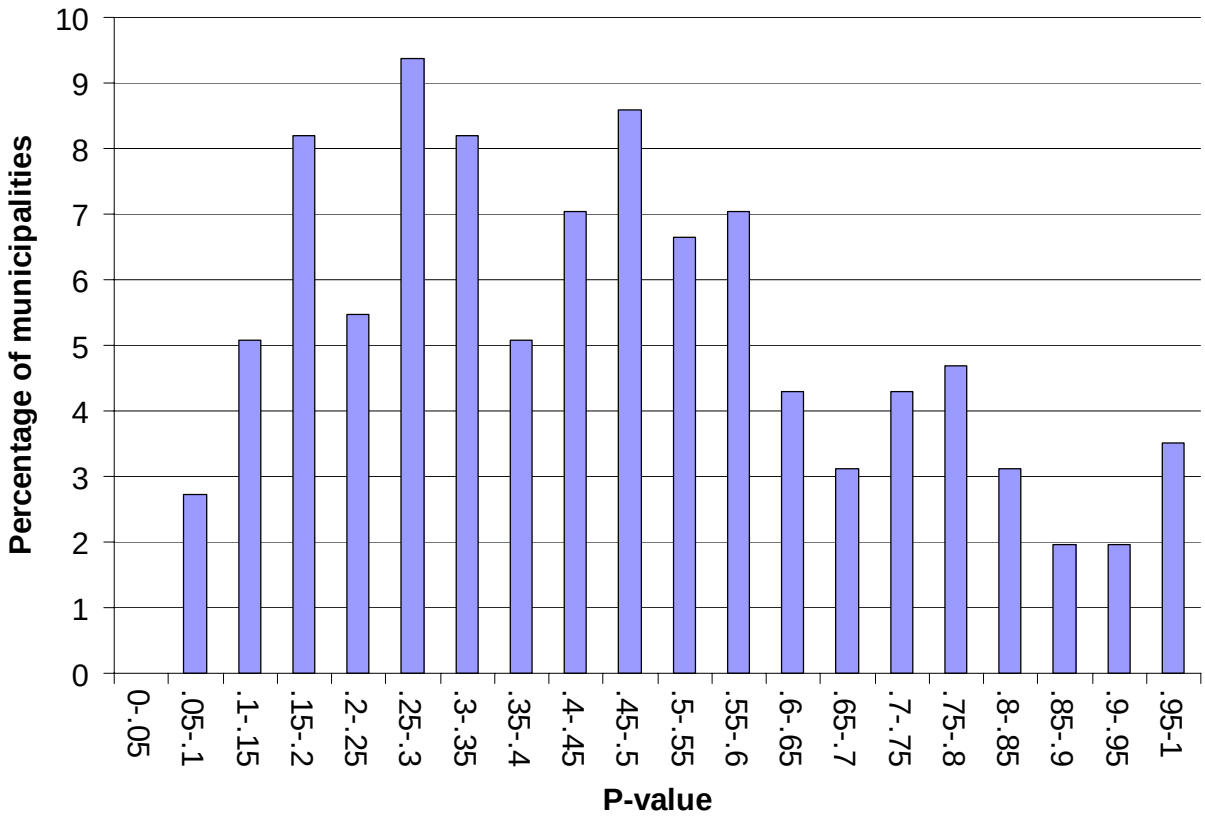
Source: constructed from the ANPE file.

Graph 4: probability of leaving for non-employment before 24 months (Kaplan-Meier) for municipalities with at least 5,000 inhabitants



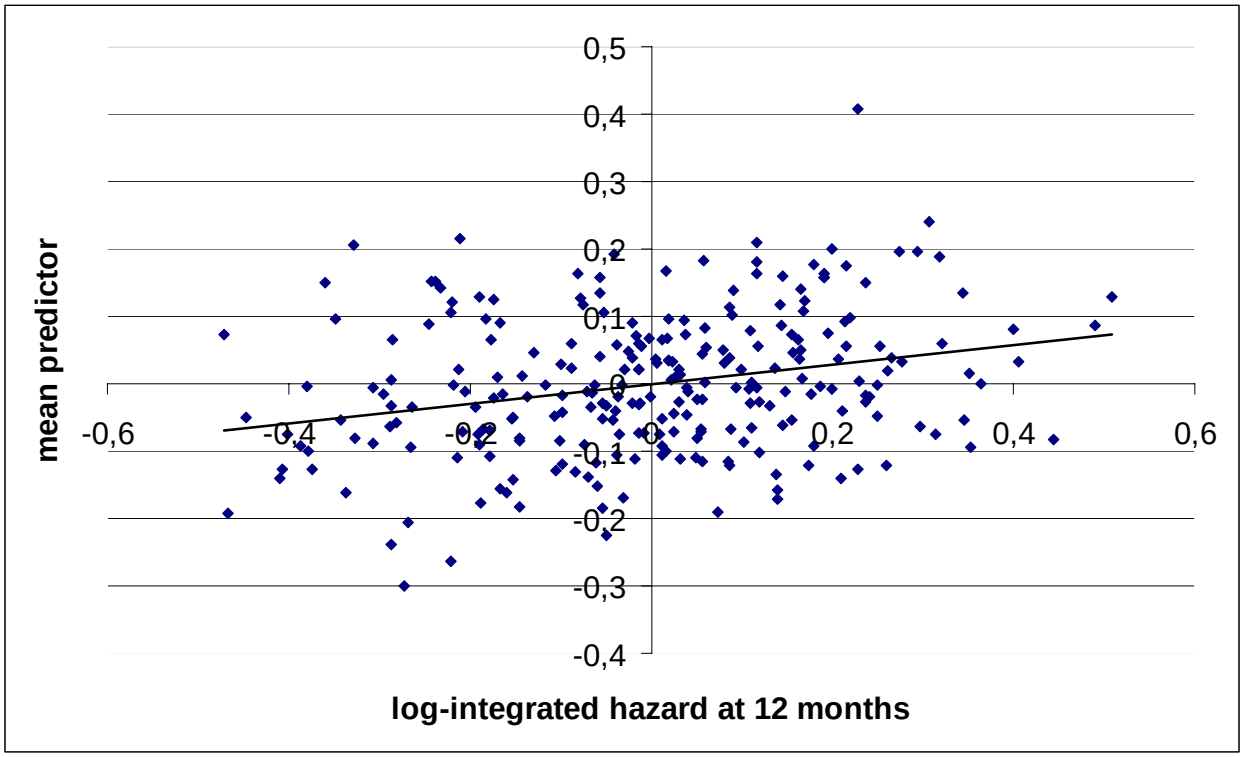
Source: constructed from the ANPE file.

Graph 5: The spatial distribution of p-values of the Kolmogorov test statistic



Source: constructed from the ANPE file.

Graph 6: Municipality average of individual effects $X^j\beta(12)$
as a function of log-integrated hazard at 12 months for exit to job



Source: constructed from the ANPE file.
Municipalities with more than 10,000 inhabitants in 1999. Variables are centered. Correlation between variables: .27.

Table 1: Spatial inequality indices for measures of segregation and job accessibility calculated using census data

Variables	Mean	q90/q10	q90-q10	Gini	Coeff. of variation	Number of obs.
Segregation variables						
Unemployment rate	.116	2.408	.100	.196	.355	1300
% French	.813	1.277	.198	.054	.095	1300
% European (other)	.070	2.080	.050	.160	.294	1300
% North African	.054	9.464	.097	.396	.730	1300
% Sub-Saharan African	.026	9.371	.046	.389	.700	1300
% Other Nationality	.037	6.522	.057	.361	.688	1300
% Secondary School Diploma	.399	1.895	.248	.130	.227	1300
% Technical diploma	.192	2.526	.156	.175	.311	1300
% High school Diploma	.148	1.453	.054	.076	.137	1300
% College Diploma	.261	3.883	.333	.270	.479	1300
Job-accessibility variables						
45mn job density by public transport	1.062	2.995	1.009	.211	.459	1300
45mn job density by car	.856	1.615	.400	.104	.181	1300

Source: constructed from the 1999 Population Census and the 2000 General Transport Survey (*Enquête Globale de Transport*). The unemployment rate is weighted by the labor force. Nationality rates are weighted by the population. Diploma rates are computed for the population over 15 and are weighted by the population over 15. Job-accessibility variables are weighted by the labor force.

Table 2: Spatial inequality indices calculated using the ANPE file

Variables	Mean	q90/q10	q90-q10	Gini	Coeff. of variation	Number of obs.
Exit types and unemployment spells						
% Exit to job	.280	1.734	.152	.121	.224	1289
% Exit to non-employment	.167	1.370	.052	.070	.148	1289
% Right-censoring	.553	1.322	.152	.218	.426	1289
Duration if exit to job	276	1.374	87	.070	.138	1254
Duration if exit to non-employment	369	1.433	131	.083	.179	1156
Duration if right-censoring	334	1.753	179	.130	.281	849
Characteristics of unemployed workers						
% Age	32.610	1.080	2.499	.017	.032	1289
% Male	.518	1.164	.078	.033	.068	1289
% Female	.482	1.177	.078	.035	.073	1289
% Single	.606	1.327	.174	.062	.113	1289
% Couple	.394	1.599	.175	.095	.174	1289
% 0 child	.613	1.352	.185	.065	.116	1289
% 1 child	.163	1.458	.061	.085	.174	1289
% 2 children	.124	1.875	.074	.135	.264	1289
% 3 children	.057	2.814	.056	.212	.404	1289
% 4 children	.023	5.281	.032	.306	.569	1289
% 5 children and more	.019	6.938	.032	.378	.703	1289
% French	.782	1.315	.214	.060	.107	1289
% European (other)	.064	2.636	.061	.209	.402	1289
% North African	.077	5.444	.110	.305	.541	1289
% Sub-Saharan African	.045	5.665	.063	.303	.531	1289
% Other Nationality	.032	1.671	.051	.371	.695	1289
% College diploma	.239	3.942	.315	.285	.513	1289
% High School (excluding final year)	.165	1.636	.079	.106	.205	1289
% High school (final year and diploma) and technical diploma	.327	2.199	.231	.152	.272	1289
% Secondary school	.269	2.455	.226	.179	.314	1289
% Disabled	.033	2.631	.030	.195	.399	1289

Source: constructed from the ANPE file, sample of workers whose unemployment spell started between January 1996 and June 1996. All indices are weighted by the number of unemployed workers.

Table 3: Estimation results of the first-stage equation (SPLE)

Variables	Job	Non-employment
Age /100	-2.9289***	-9.0729***
	(.2801)	(.3253)
(Age/100) squared	1.210***	11.330***
	(.387)	(.442)
Male	*<ref>*	*<ref>*
Female	-.1819***	.3486***
	(.0060)	(.0079)
Single	*<ref>*	*<ref>*
Couple	.1089***	.0710***
	(.0077)	(.0094)
No child	*<ref>*	*<ref>*
1 child	-.0815***	.0834***
	(.0093)	(.0110)
2 children	-.0266**	.0375***
	(.0106)	(.0130)
3 children	-.1312***	.0352**
	(.0149)	(.0174)
4 children	-.1823***	.0428*
	(.0245)	(.0260)
5 children and more	-.2425***	.0852***
	(.0299)	(.0281)
French	*<ref>*	*<ref>*
European (other)	-.0510***	-.1732***
	(.0124)	(.0168)
North African	-.4455***	-.0810***
	(.0143)	(.0154)
Sub-Saharan African	-.6638***	-.0244
	(.0209)	(.0198)
Other Nationality	-.5629***	.0248
	(.0231)	(.0224)
College diploma	*<ref>*	*<ref>*
High School (first grade)	-.2296***	-.0970***
	(.0089)	(.0118)
High school (other grade) and technical diploma	-.3349***	-.2176***
	(.0078)	(.0107)
Secondary school	-.5872***	-.4252***
	(.0095)	(.0119)
Not disabled	*<ref>*	*<ref>*
Disabled	-03837***	.4653***
	(.0197)	(.0168)
Number of observations	430,695	

Source: constructed from the ANPE file.
***: significant at 1% level; **: significant at 5% level; *: significant at 10% level.
Monthly dummy variables were also included to control for seasonality but are not reported in the table.

Table 4: Disparity indices at the municipality level

Statistics on durations	Mean	P90 / P10	P90 - P10	Gini	Coeff. of variation
Until exit to job					
Survival at 6 months					
Kaplan-Meier	.801	1.173	.127	.035	.065
Model	.811	1.132	.100	.028	.053
Multiplicative model	.811	1.135	.103	.027	.055
Survival at 24 months					
Kaplan-Meier	.533	1.503	.213	.088	.164
Model	.537	1.426	.188	.076	.143
Multiplicative model	.535	1.416	.185	.076	.142
Until exit to non employment					
Survival at 6 months					
Kaplan-Meier	.893	1.059	.051	.012	.027
Model	.896	1.049	.043	.011	.024
Multiplicative model	.894	1.068	.059	.011	.037
Survival at 24 months					
Kaplan-Meier	.681	1.200	.123	.039	.093
Model	.686	1.175	.109	.036	.075
Multiplicative model	.683	1.175	.110	.036	.087

Source: constructed from the ANPE file.
Fixed effects in the multiplicative model are computed using 8 intervals of 90 days and one interval covering the remaining days. Municipalities are weighted by the number of unemployed workers.

Table 5: Analysis of spatial variance (at the municipality level)

	Exit to Job			Exit to non-employment		
	Variance	Correlation With $X^j\beta(t)$	Number of observations	Variance	Correlation With $X^j\beta(t)$	Number of observations
$X^j\beta(6)$	.0086	1	24,069	.0063	1	15,155
$\ln H_k(6)$	.0811	.3512	24,069	.0424	-.0898	15,155
$\ln H_m(6)$	.0614	.1890	24,069	.0411	-.2083	15,155
$\ln H_m(6)+ X^j\beta(6)$	.0787	.4983	24,069	.0407	.1839	15,155
$\ln H_{mm}(6)$	.0566	.1464	24,069	.0379	-.1287	15,155
$\ln H_{mm}(6)+ X^j\beta(6)$	.0717	.4772	24,069	.0402	.2708	15,155
$X^j\beta(12)$	.0133	1	10,780	.0081	1	6,584
$\ln H_k(12)$	.0766	.3217	10,780	.0370	.0478	6,584
$\ln H_m(12)$	.0560	.1706	10,780	.0337	-.0436	6,584
$\ln H_m(12)+ X^j\beta(12)$	.0786	.5553	10,780	.0404	.4092	6,584
$\ln H_{mm}(12)$	.0552	.1050	10,780	.0438	.0261	6,584
$\ln H_{mm}(12)+ X^j\beta(12)$	.0741	.5140	10,780	.0529	.4158	6,584
$X^j\beta(24)$	.0264	1	3,210	.0151	1	2,206
$\ln H_k(24)$	.0647	.1812	3,210	.0399	-.0015	2,206
$\ln H_m(24)$	.0490	.0292	3,210	.0357	-.1151	2,206
$\ln H_m(24)+ X^j\beta(24)$	.0775	.6069	3,210	.0454	.4749	2,206
$\ln H_{mm}(24)$	.0509	.0567	3,210	.0495	-.0262	2,206
$\ln H_{mm}(24)+ X^j\beta(24)$	.0814	.6141	3,210	.0632	.4659	2,206

Source: constructed from the ANPE file.

Fixed effects in the multiplicative model are computed using 8 intervals of 90 days and one interval covering the remaining days.

$X^j\beta(t)$: average effect of explanatory variables for individuals exiting to a job around time t at the municipality level (windows of 2 months before and after are used). $\ln H_k(t)$: log of integrated hazard at t days using the Kaplan-Meyer estimator. $\ln H_m(t)$: log of integrated hazard at t days for the model. $\ln H_{mm}(t)$: log of integrated hazard at t days for the model under the multiplicative assumption. Statistics are computed weighting municipalities by their number of exiting unemployed workers.

Table 6: Regressions of town fixed effects (for exit to job) on municipality variables

	(1)	(2)	(3)
Constant	-6.717***	-6.147***	-6.644***
	(.120)	(.035)	(.121)
Proportion of technical diplomas	1.861***		1.966***
	(.337)		(.338)
Proportion of high school diplomas	-.078		-.260
	(.384)		(.386)
Proportion of college diplomas	.099		.354*
	(.178)		(.192)
Proportion of European (other)	-1.394***		-1.402***
	(.246)		(.245)
Proportion of North Africans	-1.756***		-1.344***
	(.220)		(.250)
Proportion of Sub-Saharan Africans	-3.775***		-3.872***
	(.513)		(.512)
Proportion of other nationalities	-.458*		-.491**
	(.246)		(.245)
Job density within 45mins by public transport		-.092***	.000
		(.016)	(.012)
Job density within 45mins by private transport		-.517***	-.176***
		(.047)	(.051)
Number of observation	1254	1254	1254
Weighted number of observations	430602	430602	430602
Error rate	.468	.246	.473
Pseudo-R^2	.724	.259	.730

Source: constructed from the ANPE file.
***: significant at 1% level; **: significant at 5% level; *: significant at 10% level.
Estimates are computed using 8 intervals of 90 days and one interval covering the remaining days. Municipalities are weighted by the number of unemployed workers.

Table A1: Descriptive statistics on variables used in the study

Variable	Number of obs.	Mean	Standard deviation	Minimum	Maximum
Exit types and unemployment spells					
Exit to job	430,695	.280	.449	.000	1.000
Exit to non-employment	430,695	.167	.373	.000	1.000
Right-censoring	430,695	.672	.470	.000	1.000
Duration if exit to job	120,502	273	337	1	2818
Duration if exit to non-employment	71,807	368	452	1	2813
Duration if right-censoring	238,386	287	378	1	2829
Characteristics of unemployed workers					
Age	430,695	32.610	9.222	16.000	54.000
Male	430,695	.518	.500	.000	1.000
Female	430,695	.482	.500	.000	1.000
Single	430,695	.606	.489	.000	1.000
Couple	430,695	.394	.489	.000	1.000
0 child	430,695	.613	.487	.000	1.000
1 child	430,695	.163	.369	.000	1.000
2 children	430,695	.124	.330	.000	1.000
3 children	430,695	.057	.233	.000	1.000
4 children	430,695	.023	.150	.000	1.000
5 children and more	430,695	.019	.137	.000	1.000
French	430,695	.782	.413	.000	1.000
European (other)	430,695	.064	.245	.000	1.000
North African	430,695	.077	.267	.000	1.000
Sub-Saharan African	430,695	.045	.207	.000	1.000
Other Nationality	430,695	.032	.175	.000	1.000
College diploma	430,695	.239	.427	.000	1.000
High School (excluding final year)	430,695	.165	.371	.000	1.000
High school (final year and diploma) and technical diploma	430,695	.327	.469	.000	1.000
Secondary school	430,695	.269	.443	.000	1.000
Disabled	430,695	.033	.178	.000	1.000
Segregation variables					
Unemployment rate	430,695	.127	.043	.000	.246
% French	430,695	.794	.078	.569	1.000
% European (other)	430,695	.071	.020	.000	.265
% North African	430,695	.064	.041	.000	.218
% Sub-Saharan African	430,695	.030	.019	.000	.086
% Other Nationality	430,695	.041	.027	.000	.230
% No diploma	430,695	.412	.091	.202	.663
% Technical diploma	430,695	.191	.057	.040	.402
% High school	430,695	.145	.020	.000	.266
% University	430,695	.252	.123	.030	.571
Job-accessibility variables					
45mn job density by public transport	430,695	1.085	.436	.076	19.920
45mn job density by car	430,695	.860	.152	.152	1.200

Source: constructed from the ANPE file.

B.3. Assessing the Effects of Local Taxation using MicroGeographic Data

Duranton G., Gobillon L. et H. Overman (2009), "Assessing the Effects of Local Taxation using MicroGeographic Data", CEPR Working Paper 5856, revise-and-resubmit à l'*Economic Journal*, version révisée.

43 pages

Assessing the effects of local taxation using microgeographic data[§]

Gilles Duranton[‡]
University of Toronto
Laurent Gobillon[§]
Institut National d'Etudes Démographiques
Henry G. Overman[¶]
London School of Economics

Revised version: 7 August 2009

ABSTRACT: We study the impact of local taxation on the location and growth of firms. Our empirical methodology pairs establishments across jurisdictional boundaries to estimate the impact of taxation. Our approach improves on existing work as it corrects for unobserved establishment heterogeneity, for unobserved time-varying site-specific effects, and for the endogeneity of local taxation. Applied to data for English manufacturing establishments, we find that local taxation has a negative impact on employment growth, but no effect on entry.

Key words: Local taxation, spatial differencing, borders, regression discontinuity.
JEL classification: H22, H71, R38.

[§]Thanks to the editor, Antonio Ciccone, two anonymous referees, Tom Davidoff, Fritz Foley, Thomas Klier, Thierry Mayer, John Quigley, Helen Simpson and to seminars and conferences participants in Berkeley, Boston, Ghent, Glasgow, Hamburg, Kansas City, Kiel, Kyoto, London, Oslo, Paris, Seattle, Vancouver, and Venice. We also owe special gratitude to Tim Besley and Tom Holmes. Financial support from the Economic and Social Research Council (ESRC Grant R000239878) and the Leverhulme Trust is gratefully acknowledged. This work contains statistical data from ONS which is Crown copyright and reproduced with the permission of the controller of HMSO and Queen's Printer for Scotland. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

[‡]Department of Economics, University of Toronto, 150 Saint George Street, Toronto, Ontario M5S 3G7, Canada (e-mail: gilles.duranton@utoronto.ca; website: http://individual.utoronto.ca/gilles/default.html). Also affiliated with the Center for Economic Policy Research, and the Center for Economic Performance at the London School of Economics.

[§]Institut National d'Etudes Démographiques, 133 bd Davout, 75980 Paris Cédex 20, France (e-mail: laurent.gobillon@ined.fr; website: http://laurent.gobillon.free.fr/). Also affiliated with the Center for Economic Policy Research, PSE-INRA, and CREST.

[¶]Department of Geography and Environment, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom (e-mail: h.g.overman@lse.ac.uk; website: http://personal.lse.ac.uk/OVERMAN/). Also affiliated with the Centre for Economic Policy Research, and the Centre for Economic Performance at the London School of Economics.

1. Introduction

This paper develops new empirical methodologies to identify the effects of local taxation on the location and growth of firms. This issue has been the focus of an extensive theoretical literature and our paper is not the first paper to consider these issues empirically.[1] Bartik (1991) summarises the earlier literature. Evidence from the 1960s and 1970s suggested there was no effect of taxes on firm location decisions. Bartik's own work focusing on the 1980s suggested a negative relationship and a number of subsequent papers have confirmed that finding. Much of this work used fairly large spatial units (mostly US states). More recent work (e.g., Guimaraes, Figueiredo, and Woodward, 2004) has moved towards smaller spatial units such as US counties with similar results.

The existing literature, however, has failed to resolve three main problems when assessing this impact. First, firms are faced with the choice of a large number of heterogenous locations. Many of these site's characteristics are unobserved and likely to be correlated with other explanatory variables such as plant characteristics and local taxation, thus biasing the results. Second, firms themselves are heterogenous. Again much of this heterogeneity is unobservable so that the sorting of firms according to these characteristics provides another source of bias. Third, aspects of the tax system may be endogenous to firm decisions, which may lead to a reverse causality bias.

This paper deals with all of these problems. Our approach is to use spatial differencing between neighbouring firms across borders to control for the first source of bias. Incorporating panel and instrumental variable techniques then deals with the other two biases. When applying the methodology to data for English manufacturing establishments we find that local taxation of non-residential property has a negative impact on employment growth, but no effect on entry. We show that methodologies that do not address these

[1] One can identify at least three strands of theoretical literature. The issue of tax competition has received considerable attention. See Wilson (1999) for a review. A similarly lengthy debate has centred around local public good provision and the Tiebout (1956) hypothesis that inter-jurisdictional competition helps achieve the efficient provision of local public goods. See Epple and Nechyba (2004) for a recent review. Finally, capitalisation of local taxes and the efficient taxation of land are key concerns for urban economists. This has been a subject of discussion since George (1884). See Fujita (1989) for a presentation of the arguments and Arnott (2004) for a recent discussion of its applicability.

three problems give substantively different results. That a tax on non-residential property should affect *employment* might seem surprising. Our model shows how revaluations of properties in case of expansion act as a break on the growth of establishments. Well documented rigidities in rents also imply that tax increases may not fully capitalise and can lead establishments to relocate or exit altogether.

The rest of the paper is structured as follows. In section 2 we outline our methodology and relate it to the existing literature. Section 3 outlines our data. Section 4 presents our findings on the impact of local taxation on employment while section 5 presents results for entry. Section 6 concludes.

2. Methodology

A. *Model*

To motivate our approach we develop a simple model of the impact of local taxation on establishment location and employment.[2] Because we use our model to interpret our findings we embed several aspects of our institutional setting as we proceed.[3]

Establishments use labour and building space to produce a final good. To keep things simple assume that building space and labour are perfect complements.[4] By choice of units, one unit of building and one unit of labour are needed to produce one unit of the final good. The unit rental cost of buildings in jurisdiction a comprises two elements: the building rent, b_a, and the local tax, referred to in the UK as business rates, r_a. Building rents are endogenously determined in a way described below. Consistent with our empirical setting we assume that the tax is paid by occupiers, not landlords. It is the product of a rate, which varies across jurisdictions, times the historical value of the building.

[2]We use the term establishments to refer to our basic units of observation. In the data that we use, these are sometimes part of larger multi-establishment firms. We ignore the complications this introduces.

[3]The UK rates system is further described below. See also Hale and Travers (1994).

[4]As long as labour and buildings remain gross complements, qualitatively similar results are obtained. Gross complementarity between building and labour is a natural assumption for most sectors.

We take labour as numeraire. The price of the final good is exogenous, common across all jurisdictions, and equal to $1 + p$. Establishments live for up to two periods and new establishments can enter every period. During an establishment's first period, demand for their good is normalised to unity. Period 1 profit of an establishment i initially located in jurisdiction a at time t is given by

$$\pi^1_{at}(i) = (p - b_{at} - r_{at}) \tag{1}$$

In their second period, firms face a shock to the demand for their final good and a tax shock.[5] For the tax, we assume $\tilde{r}_{at+1}$ is drawn at the beginning of period $t + 1$ from a distribution $f(.)$ over $[\underline{r}, \overline{r}]$. This is consistent with our empirical setting where the tax is fully decentralised and set freely by each jurisdiction every year.[6] We assume that $\underline{r} \geq 0$ and that $\overline{r}$ is not too high in a sense made clear below. Demand in the second period is given by $1 + \tilde{\rho}_{it+1}$ where $\tilde{\rho}_{it+1}$ is also drawn at the beginning of period $t + 1$ from a continuous distribution $g(.)$ over the support $[-1, \overline{\rho}]$ with $\overline{\rho} > 0$. Firms respond differently to negative and positive shocks and we consider each in turn.

We rule out the possibility that establishments can renegotiate rents downwards in response to either a tax or a demand shock. This helps limit the number of possible firm responses that we need to consider. It is also consistent with the institutional setting during our period of analysis when most industrial establishments sign long term rental contracts of 20 years or more with a review of the rent only every five years or so. Even at the time of rent reviews, scope for adjustment is limited because virtually all commercial contracts in the UK include an upward only clause (Crosby, Lizieri, Murdoch, and Ward, 1998). In light of this, we assume that new establishments sign a lease in t which sets building rent for both t and $t + 1$. For simplicity and without affecting the qualitative nature of our results, we assume $b_{at} = b_{at+1}$. We will return to this issue below in discussing when rents respond to changes in local taxes.

[5] To keep our model transparent we use shocks affecting demand for final goods. We could instead use productivity shocks with a fully specified demand system to derive similar results.

[6] While we consider an exogenous tax shock, our empirical analysis worries about endogeneity.

Given that rent cannot be reduced, after a negative shock, $\rho_i < 0$, an establishment wants to either exit or downsize.[7] Exit leads to second period profit $\pi^2_{t+1}(i) = \pi_E(i) = 0$ where the subscript E stands for exit. Downsizing occurs through subletting part of the original building unit. In the UK, commercial contracts give tenants "the right to sell or sublet the unexpired term of the lease, with landlords being unable to withhold their consent unreasonably" (Crosby *et al.*, 1998, p. 3). We assume that subletting part of the initial unit of building yields a unit rent $\underline{b}$ where $\underline{b} < b_{at} + r_{at+1}$ in equilibrium. That is, firms cannot make profits by subletting. Downsizing to satisfy the new level of demand and renting unused building space implies second period profit $\pi^2_{t+1}(i) = \pi_D(i) = (1 + \rho_i)p - \rho_i\underline{b} - b_{at} - r_{at+1}$.

Following a positive demand shock, $\rho_i \geq 0$, an establishment has three options. First, it can grow by expanding employment and building space on its current site.[8] The unit rent for extra building space is b_{at+1}, the same as for the original unit of building. The main issue with increasing the amount of building space on a given site is that, in our setting, it typically led to a revaluation of the building (Hale and Travers, 1994). This change in the tax base implies that the unit rate increases for the expansion as well as the original unit. We assume that re-valuation leads the value of the tax base to be multiplied by $\delta > 1$. Profit in this case is $\pi^2_{t+1}(i) = \pi_G(i) = (1 + \rho_i)(p - b_{at} - \delta r_{at+1})$.

Alternatively, an establishment can decide to forego this expansion opportunity and continue to produce only one unit of final good. Profit when staying the same is $\pi^2_{t+1}(i) = \pi_S(i) = p - b_{at} - r_{at+1}$. Finally an establishment can incur cost c and relocate to another jurisdiction where the unit rental cost of buildings is equal to $\overline{b}$ such that in equilibrium $\underline{b} < \overline{b} < b_{at} + \delta r_{at+1}$. Profit in case of relocation is $\pi^2_{t+1}(i) = \pi_R(i) = (1 + \rho_i)(p - \overline{b}) - c$.

Hence, depending on its demand shock, ρ_{it+1}, and the level of tax in period $t + 1$, r_{at+1}, an establishment faces five possible choices: exit (leading to second period profit π_E),

[7] Another alternative would be to relocate and leave the existing lease. To keep the number of options manageable, we assume that moving cost, c, would make this alternative prohibitively costly. More precisely, we assume $c > b_{at} + \overline{r} - \underline{b}$ where $\underline{b}$ is the unit rental costs of buildings when subletting (as defined below).

[8] It could also occur by renting adjacent sites. We ignore this possibility here.

downsize (leading to π_D), stay put (π_S), grow (π_G), or relocate (π_R).

We now define four thresholds, $\rho_{ED}(r_{at+1})$, $\rho_{DS}(r_{at+1})$, $\rho_{SG}(r_{at+1})$, and $\rho_{GR}(r_{at+1})$. They correspond to particular realisations of ρ such that an establishment is indifferent between exit and downsize, between downsize and stay at its original size, between stay at its original size and grow locally, or between grow locally and relocate, respectively. Provided the tax, r_{at+1}, is not too high and relocation costs, c, are large enough, the above expressions for π_E, π_D, π_S, π_G, and π_R imply $-1 \leq \rho_{ED}(r_{at+1}) < \rho_{DS}(r_{at+1}) < \rho_{SG}(r_{at+1}) < \rho_{GR}(r_{at+1})$. Establishments that face very negative demand shocks prefer to exit. For a less negative shock, they remain in business but downsize. For a small positive shock, establishments retain their original size. For an intermediate positive shock, they grow. For a large positive shock, they choose to relocate. We briefly explain the ranking of these each of these thresholds in turn.

Because establishments that remain in business are stuck with their lease and because they can sublet below the rental cost they face, establishments which experience a large negative demand shock exit rather than downsize to a very small size. Interestingly, $\partial\rho_{ED}(r_{at+1})/\partial r_{at+1} > 0$. That is, a higher tax in $t+1$ induces more exits. More exits also imply that surviving establishments are those that have experienced a less negative demand shock.

Next, it is easy to show that $\rho_{DS}(r_{at+1}) = 0$. Establishments that face a small negative shock are left with unnecessary building space. Without the possibility to leave their lease, they prefer to sublet it rather than leave it empty. Establishments that face a small positive shock would like to expand. However, adding building space implies a revaluation of the tax base and a higher tax. Whenever the demand shock is positive, but not large enough to offset this increase in the tax, establishments prefer to keep their original size instead of expanding. Hence, $\rho_{SG}(r_{at+1}) > 0$. It is also easy to see that $\partial\rho_{SG}(r_{at+1})/\partial r_{at+1} > 0$. A higher tax in $t+1$ makes the cost of a revaluation higher and thus can only be justified for larger demand shocks. That is, for establishments that remain in jurisdiction a, a higher rate of taxation leads to lower employment growth on average.

Finally, establishments that face a large enough positive demand shock prefer to relocate. The fixed cost of relocation can only be justified for those establishments that need to expand a lot. Because a higher tax implies a greater gain from relocation, we have $\partial\rho_{GR}(r_{at+1})/\partial r_{at+1} < 0$.

To close the model, we assume a competitive land market for new establishments. Free entry means that establishments enter until they make zero expected profit $E(\Pi_{at}) = \pi^1_{at} + E(\pi^2_{t+1}) = 0$ and b_{at} adjust to ensure this holds across all jurisdictions.[9] Put differently, when new entrants sign contracts current and expected future taxes are fully capitalised in b_{at}. However, the realisation of r_{at+1} is not capitalised for continuing establishments. Loosely speaking, taxes are capitalised in the long run but not in the short run. Furthermore, clearing on the land market implies that more exits and more relocations in a jurisdiction are matched by more entries when the supply of sites is constant (a reasonable assumption for the UK).

The main result of our model is thus that the tax rate affects the use of building space by establishments. In turn, the complementarity between factors implies that taxation affects the employment decision of establishments. There are two separate channels. Higher taxes lead to both a *growth slow-down* and a *selection* effect. More precisely, higher taxes reduce employment growth by inducing more establishments to keep their size constant instead of growing. The growth slow-down effect is driven entirely by building revaluations that lead to higher taxes and make small expansions unprofitable. In terms of selection, higher taxes in $t+1$ imply more exits for establishments with a negative shock and more relocations for establishments with a positive shock. The selection effect is thus ambiguous. The selection effect is driven by imperfect capitalisation of higher taxes into rents for existing establishments. This affects the local profitability of establishments and,

[9] Developing this expression is not very enlightening. When r_{at+1} and ρ_{t+1} are independent, expected profit is well defined and equal to $E(\Pi_{at}) = \pi^1_{at} + \int_{\underline{r}}^{\overline{r}} \int_{-1}^{\overline{\rho}} \pi^{2*}_{t+1}(r_{at+1}, \rho_{t+1}) f(r_{at+1}) g(\rho_{t+1}) \mathrm{d}r_{at+1} \mathrm{d}\rho_{t+1}$ where $\pi^{2*}_{t+1}(r_{at+1}, \rho_{t+1})$ is optimal second-period profit.

in turn, their location choice.[10]

Overall, the effect of higher taxes on employment in the establishments that stay is thus ambiguous. Higher taxes lead to less employment when selection through exits is dominated by selection through relocation and the growth slow-down effect.[11] In our empirical analysis, we can estimate an overall effect but cannot identify the selection and growth slow down effects separately.

Our model also offers some predictions about entry. Through the selection effect, higher taxes imply more exits and more relocations. In turn, this should lead to more entries. In the second part of the paper, we assess the effect of taxes on entries.

We could extend our framework to incorporate other factors of production. The tax could cause establishments to change their use of these other inputs. In turn, this could impact on employment. Because of these omitted inputs, we need to be cautious about the interpretation of our results. Specifically, we are only able to estimate the overall effect of taxation on employment even though other cross-factor effects may be at work.[12]

Because higher taxes affect the employment of establishments, a naive reading of our model would lead us to estimate

$$e_{it} = \alpha r_{at} + \epsilon_{it} \tag{2}$$

where e_{it} is the log employment of establishment i at time t and ϵ_{it} an error term that captures the demand shock ρ_{it} in reduced form. The main parameter of interest is α. It captures the (net) effect on employment of the (log) local tax, r_{at}.

[10] In the model, there is imperfect capitalisation because of the combination of rigid rents, uncertainty about future taxes, and costly relocation. Without rent rigidity, renegotiation in a competitive land market implies full capitalisation of r_{at+1} into b_{at+1}. Without uncertainty, r_{at+1} would be capitalised into the initial level of rent b_{at} just like r_{at} since new establishments face a competitive land market. Finally, without costly adjustment, establishments could leave their lease and negotiate a new one competitively.

[11] We note that these effects can be quantitatively large. If in a jurisdiction there are lots of establishments below ρ_{GR}, a further unexpected increase in taxation can lead to many relocations and a much lower growth rate in employment for remaining establishments.

[12] The effects of taxation may also vary across places. See, for instance, Baldwin, Forslid, Martin, Ottaviano, and Robert-Nicoud (2004). In this paper we only estimate an average effect.

B. ***Heterogenous establishments and heterogenous locations***

Unlike in the model, we expect production establishments to be very heterogenous ex-ante and much of this heterogeneity to be unobservable. This is a first likely source of bias if establishments with different unobserved characteristics sort across jurisdictions with different tax rates. We can enrich the specification (2) and add establishment characteristics:

$$e_{it} = \alpha r_{at} + X_{it}\beta + \mu_i + \epsilon_{it} \tag{3}$$

where β captures the effect of time-varying establishment-specific observable variables, X_{it}. An establishment fixed effect, μ_i, captures the impact of unobservable time invariant establishment characteristics.

The second issue to consider is that not all sites are the same. There are a large number of heterogenous sites and these sites come at very different rental prices. For instance, Thompson and Tsolacos (2001) document a sixfold difference in the rental price between industrial sites close to Heathrow airport and those in the suburbs of Leeds. It is unlikely that these differences in rents only reflect differences in local taxation.

Traditional empirical approaches have worried about the fact that the costs of factors of production differ across geographical regions. We are also concerned with heterogeneity at a much finer geographical scale. A wide variety of factors affect both the attractiveness of sites and the success of establishments once they choose their site. For example, the attractiveness of a site may depend on access to the road network while improvements to that network may affect the performance of establishments at that site. Similarly, changes in congestion can have different implications for establishments located close, but not very close to one another as can the entry or exit of big buyers or suppliers. Overall, there are many reasons to expect considerable site heterogeneity, possibly at a very fine spatial scale, and varying over time.

If this is the case, assessing the impact of taxation requires us to control for both fixed and time-varying site characteristics since they are likely to be correlated with taxes or the demand shock for establishments. Some of these site characteristics may be observable and thus can be controlled for directly. However many are likely to be unobservable. This

implies that we are interested in a specification like

$$e_{it} = \alpha r_{at} + X_{it}\beta + \mu_i + \gamma_a + \theta_{zt} + \epsilon_{it} \tag{4}$$

where γ_a is a time-invariant effect for jurisdiction a and θ_{zt} is a time-varying effect for location z, possibly at a finer spatial scale than a. Note that the establishment fixed effects also control for unobserved time-invariant site-specific effects (if establishments do not move) leaving θ_{zt} to control for unobserved time-*varying* site-specific effects.[13]

C. *Time and spatial differencing*

Estimating (4) by OLS ignoring the unobservable effects gives a consistent estimate of α and β only if $\text{Cov}\left([r_{at}, X_{it}], \mu_i + \gamma_a + \theta_{zt} + \epsilon_{it}\right) = 0$. This condition is unlikely to hold, if only because observable establishment characteristics X_{it} are likely to be correlated with unobservable establishment characteristics μ_i. More interestingly, in a spatial context, the site-specific effect θ_{zt} is likely to be correlated across neighbouring sites. This raises the possibility that, within a jurisdiction, there could be omitted variables driving both the average site-specific effect and the tax rate. That is, r_{at} is likely to be correlated with θ_{zt}.

To control for establishment and jurisdictional fixed effects we can use the panel dimension of our data to calculate the *within* estimator (alternatively, we could calculate the first-difference estimator). The *within* transformation is obtained, as usual, by centring all observations around their mean. For any variable y, for observation i, let $\overline{y}_i$ denote the time average and define $\tilde{y}_{it} \equiv y_{it} - \overline{y}_i$. We can then rewrite equation (4) as:

$$\tilde{e}_{it} = \alpha \tilde{r}_{at} + \tilde{X}_{it}\beta + \tilde{\theta}_{zt} + \tilde{\epsilon}_{it} \tag{5}$$

So far our approach for dealing with heterogenous establishments and time-invariant spatial heterogeneity is standard. Because we have a panel of establishment data, we are able to control for observed time-varying characteristics of establishments, as well as condition out unobserved time-invariant characteristics of both establishments and jurisdictions through the inclusion of establishment and jurisdiction fixed effects.

[13] Empirically, we cannot separately identify the time-invariant establishment-specific and jurisdiction-specific effects because we condition both out using establishment fixed effects.

This specification will give consistent estimates of α and β if $\text{Cov}\left(\left[\tilde{r}_{at}, \tilde{X}_{it}\right], \tilde{\theta}_{zt} + \tilde{\epsilon}_{it}\right) = 0$. This condition, although weaker than that necessary for consistency of OLS, is still unlikely to hold because changes in the unobserved site-specific variables $\tilde{\theta}_{zt}$ may well be correlated with changes in the tax rate as discussed above.

The standard solution to this problem would be to find a suitable instrument for the tax rate. In our context, one possibility is to use local political variables (denoted s_{at}) to instrument for tax rates. Changes in these variables are certainly likely to cause changes in local tax rates (i.e., they satisfy the relevance condition for a suitable instrument). The issue is whether they satisfy the exogeneity condition $\text{Cov}(\tilde{s}_{at}, \tilde{\theta}_{zt} + \tilde{\epsilon}_{it}) = 0$. In this respect note that changes in θ_{zt} are likely to be correlated across sites within jurisdictions while the 'average' θ_{zt} in a jurisdiction may be correlated with voting behaviour. This obviously raises the possibility that the political variables, $\tilde{s}_{at}$, are correlated with the unobserved local effects, $\tilde{\theta}_{zt}$.

To address this, we propose an alternative to instrumenting. Return to equation (5) and proceed as follows. Define Δ_d as the spatial difference operator which takes the difference between each establishment and any other establishment located at distance less than d from that establishment. Applying this spatial difference operator to (5) gives:

$$\Delta_d \tilde{e}_{it} = \alpha \Delta_d \tilde{r}_{at} + \Delta_d \tilde{X}_{it} \beta + \Delta_d \tilde{\theta}_{zt} + \Delta_d \tilde{\epsilon}_{it} \tag{6}$$

Now, we impose the crucial identifying assumption that site-specific effects change smoothly across space. That is, for d sufficiently small $\Delta_d \tilde{\theta}_{zt} \approx 0$. Noting also that taxes will be the same for establishments within the same jurisdiction, this gives us:

$$\Delta_d \tilde{e}_{it} = \Delta_d \tilde{X}_{it} \beta + \Delta_d \tilde{\epsilon}_{it} \tag{7}$$

for establishments in the same jurisdiction and:

$$\Delta_d \tilde{e}_{it} = \alpha \Delta_d \tilde{r}_{at} + \beta \Delta_d \tilde{X}_{it} + \Delta_d \tilde{\epsilon}_{it} \tag{8}$$

for establishments across jurisdictional boundaries. This shows that we can use neighbouring establishments located across jurisdictional boundaries to identify the effects of

taxation. We can also use neighbouring establishments within the same jurisdiction to improve our estimates of the effect of establishment-specific variables.

The easiest way to understand the methodology is to consider two neighbouring jurisdictions (A, B) where all sites are located on a straight line running through the two jurisdictions. With distance from the left hand end of jurisdiction A represented on the horizontal axis, figure 1 shows the change in the site-specific effect at each location. Assume that the increase in the site-specific effect is highest for sites at the left hand end of jurisdiction A and lowest at the right hand end of B. The line $\Delta\theta_z$ shows the effect on establishment employment at each location everything else equal. Everything else is not equal, however. Assume that jurisdictions base tax increases on the *average* change in site characteristics in the jurisdiction. Jurisdiction A sees larger increases in taxes than jurisdiction B. The impact on employment (again ceteris paribus) is shown as $\alpha\Delta r_a$. The overall employment effect is then $\Delta\theta_z + \alpha\Delta r_a$. Note that we assume that the impact of site-specific effects tends to outweigh the increase in taxation, so that *on average* employment growth in A is higher than in B. Thus, even after conditioning out establishment-specific effects, it appears that employment is positively related to taxes. Put differently, to estimate the effect of taxation on employment one needs to condition out site-specific effects.

To do this, consider two sites on the border between A and B. By assumption, the establishments experience approximately the same site-specific effect. All else equal, their employment would grow by roughly the same amount. However, the establishment in A sees a higher tax increase and this offsets the impact of the improved site. The overall effect on employment is given by $\alpha\Delta r_a + \Delta\theta_z$ which is larger for the establishment in B than in A. The difference between the two is $\alpha(\Delta r_a - \Delta r_b)$. Since our procedure compares the growth in employment in the establishment on the boundary in A to the growth in employment in the establishment on the boundary in B, we now correctly identify the negative relationship between increased taxation and employment. Of course, this example is just one possibility, but it does serve to emphasise exactly how our methodology

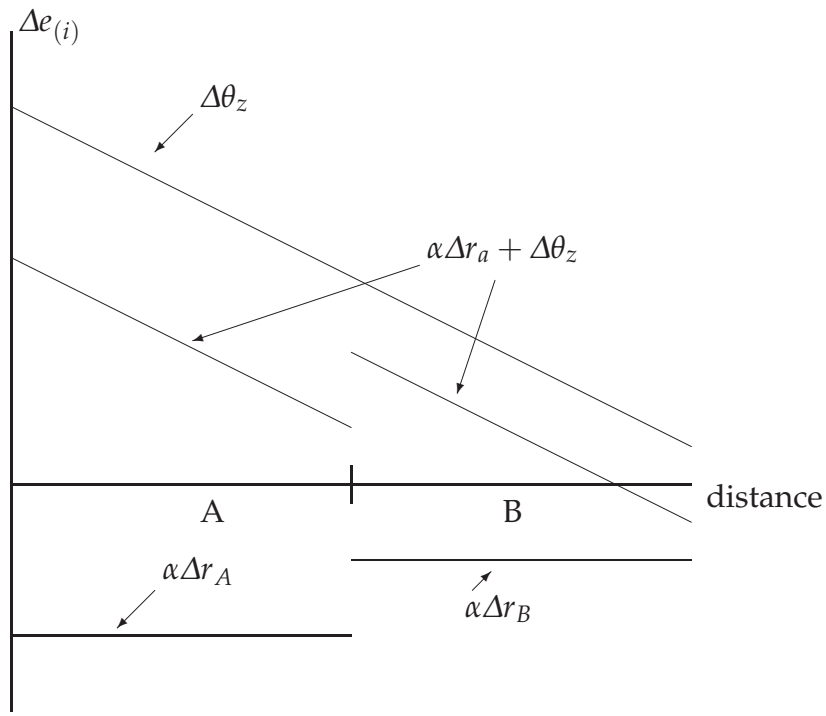


Figure 1. Boundary discontinuities

allows us to correctly identify the impact of taxes on employment.

Formally, estimating equation (6) will give consistent estimates of α and β if $\text{Cov}\left(\left[\Delta_d \tilde{r}_{at}, \Delta_d \tilde{X}_{it}\right], \Delta_d \tilde{\epsilon}_{it}\right) = 0$. Because of the absence of θ_{zt}, this condition is weaker than the one required for consistency of the non-spatially differenced IV specification.

D. *The endogeneity of taxation*

However, instrumenting still has a role to play. First if $\text{Cov}\left(\left[\Delta_d \tilde{r}_{at}, \Delta_d \tilde{X}_{it}\right], \Delta_d \tilde{\epsilon}_{it}\right) \neq 0$ instrumenting using appropriately transformed political variables $(\Delta_d \tilde{s}_{at})$ gives consistent estimates provided $\text{Cov}\left(\Delta_d \tilde{s}_{at}, \Delta_d \tilde{\epsilon}_{it}\right) = 0$. Again, this condition is weaker than before because of the absence of θ_{zt}. There is a second, more subtle, reason why instrumenting after spatial-differencing may still be necessary. For spatial differencing to remove θ_{zt}, we have assumed that $\Delta_d \tilde{\theta}_{zt} \approx 0$. This condition may fail because of spatial discontinuities (e.g. borders are not randomly located) or because the distance d we take may be too large in practice. In this case, instrumenting using appropriately transformed political variables

$(\Delta_d \tilde{s}_{at})$ gives consistent estimates provided $\text{Cov}(\Delta_d \tilde{s}_{at}, \Delta_d \tilde{\theta}_{zt} + \Delta_d \tilde{\epsilon}_{it}) = 0$. Again, this condition is more likely to be satisfied because Δ_d removes any component that varies smoothly across space and drives both political shares and site-specific effects.

Finally, before turning to the implementation, note that space differencing and the *within* transformation have implications for the error structure. As usual, directly implementing (8) yields consistent estimates for the coefficients but does not give the correct standard errors. Appendix A shows how to correct the standard errors.[14]

To summarise, we use the panel dimension of our data to remove establishment and jurisdiction fixed effects. We then identify establishments with similar site-specific effects but that face different tax rates. These are establishments that are located close to one another but in different jurisdictions. We use these establishments to identify the impact of local taxation on employment. Later in the paper, we use similar ideas to consider the impact of local taxation on the entry of new establishments. As there are some differences in the implementation, we leave the details to section 5.

The novelty in our approach is to exploit the fact that, although unobserved site characteristics vary across space, they are likely to be highly spatially correlated. Like Holmes (1998) we identify establishments that are close to one another, but on different sides of jurisdictional boundaries.[15] Because of the spatial correlation in site characteristics, these establishments will have very similar unobserved site characteristics, but will face different tax rates. We look at the extent to which differences in employment between these pairs of establishments can be explained by different local tax rates. This spatial differencing allows us to control for unobserved site characteristics, be they fixed or time-varying. Hence, compared to standard approaches, ours has the added advantage that it can control for variations in site characteristics that occur within jurisdictions as

[14] As we discussed above we ignore the complication introduced by multi-plant firms. We also do not correct for the fact that taxes vary by jurisdiction rather than establishment (Moulton, 1990).

[15] See also Black (1999). The methodology proposed by Holmes (1998) and Black (1999) has been repeated elsewhere, but most applications only use cross-sectional data and do not address endogeneity. The three exceptions are Gibbons and Machin (2003) who consider endogeneity and Kahn (2004) and Chirinko and Wilson (2008) who use some longitudinal information. Our analysis improves on these existing methodologies by incorporating instrumented panel data techniques into the spatial discontinuity approach.

well as the fact that these unobserved effects may vary over time. We also note that spatial differencing also allow us to relax the exclusion restriction when instrumenting.

3. Data

To implement our methodology, needs data satisfying a number of requirements. First, we need to have a panel of individual establishment level data. Cross-sectional data do not allow us to use the *within* transformation to remove establishment and jurisdiction-specific effects. We then need to be able to precisely locate these establishments so that we can identify which pairs of establishments are neighbours. Finally, we need to identify a local tax which is time-varying. We would prefer this tax to be economically significant to increase the chances of detecting any impact on location and employment decisions. Data satisfying all of these requirements is available for England for the six year period from 1984 to 1989. We first describe the establishment level data set we use before turning to details of the particular tax that we consider.

A. *Establishment data*

Establishment level data for years 1984 to 1989 come from the Annual Respondent Database (ARD) which underlies the Annual Census of Production in the UK. Collected by the Office for National Statistics (ONS), the ARD is a rich data set providing information on all UK establishments from 1973 onwards. We face two restrictions on time period. Changes to the tax system restrict our focus to years before 1990, while changes to the ARD restrict us to years after 1983. We also restrict ourselves to English manufacturing establishments.[16] For every establishment, we know postcode, five-digit industrial classification, and number of employees. In our analysis of employment we use only the information coming from a sub-sample of 'selected' establishments required to make a detailed return in any year. These establishments are generally larger and their employ-

[16]We ignore Scotland because it operated a different local tax system, Wales because it is not covered by the data set that provides our instruments, and Northern Ireland because special permission is required to access ARD data for establishments located there.

ment information is of better quality than for non selected establishments.[17] The precise sampling frame for selected establishments and further description of this data can be found in Griffith (1999). In our analysis of entry we use the exhaustive data (that includes 'selected' and 'non-selected' establishments).

The postcode reported in the ARD is very useful for locating establishments. In the UK, postcodes typically refer to one property or a very small group of dwellings. The Ordnance Survey (OS) CODE-POINT data set gives spatial coordinates for all UK postcodes. By merging this data with the ARD we generate very detailed information about the location of all English manufacturing establishments. For all but a tiny percentage of matched establishments the OS acknowledges a potential location error below 100 metres. For the remaining observations, the maximum error is a few kilometres. Overall, we expect a very high level of precision for our location data (see Duranton and Overman, 2005, for further discussion).

As the data requirements for our spatial differencing methodology are already somewhat restrictive, we only consider establishment specific variables that can be calculated for all establishments. For the ARD, this means only controlling for establishment age. For establishments already in the panel in 1976 (the earliest year for which we have information), we are unable to assess exact age. With regressions running over 1984-1989, this truncation censors age for establishments older than 8 years in 1984. For consistency across years, we thus construct a dummy for establishments that are 8 years or older in 1984. We also use age and age squared interacted with a dummy for being an establishment younger than 8 years old.

There are 25,579 establishments in the ARD that are located in England and that report employment at least once within our study period. We make several sample restrictions. We deleted 122 establishments where the Local Authority code was missing since we need this to append the tax rate. We then dropped 374 establishments where we could

[17] The better quality of data is not the only reason to focus on selected establishments. Implementation of the algorithm to identify neighbours and calculate data for pairs of neighbouring establishments is infeasible with the large samples that include non-selected establishments.

not identify coordinates.[18] A further 723 establishments change sector during the period and 2,547 move site (i.e., change coordinates). In both cases these changes frequently reflect coding errors rather than genuine changes in activity or location, so we drop these establishments. This leaves us with a sample of 21,813 establishments and a total of 61,785 observations.

B. *Local taxation and political variables*

As already outlined above, the tax that we consider is a property tax on non-residential property known as the UK business rate, set and collected by jurisdictions called Local Authorities (hereafter LAs). Before 1990, tax rates varied over time and jurisdictions. We are fortunate because this tax consisted only of a flat rate that applied to all occupied non-residential properties. This simplicity implies that the only source of variation is the tax rate itself.

The UK business rates were subject to a major reform in 1990 which essentially eliminated jurisdictional variation. This reform provides a large amount of exogenous variation in the tax rate. Unfortunately we cannot exploit it because all properties were also revalued in 1990. Since we do not know by how much each property was revalued, we cannot compute the change in taxation faced by each establishment. Hence, as mentioned above, while the production data restricts our study period to start in 1984, the reforms to the tax system mean it must end in 1989.[19]

UK business rates represented a considerable tax on business. In 1990 the average annual rates bill per square foot was over £4, compared with an average rent bill per square foot of just over £13 (Bond, Denny, Hall, and McCluskey, 1996). Hence rates bills

[18] We directly identified coordinates for around 90% of establishments. The main problem for the remaining 10% results from the creation of new postcodes. To increase matching rates, we checked our data against a data set of postcode updates. For observations with postcodes still unmatched, we imputed co-ordinates as follows. In the UK, a postcode is either six or seven digits. We first drop the last digit and assign establishments the mean coordinates of all postcodes sharing the same truncated postcode. If this failed to produce a match, we dropped two, then three digits from the postcode and again matched on mean coordinates where possible. This left us with 1.4% of establishments that could not be given a grid reference.

[19] For a comprehensive discussion of the reforms see Hale and Travers (1994).

make up roughly 25 per cent of the total occupancy costs of rented commercial property. In 1992 business paid £13bn in local rates, almost equal to the £15bn that they paid in corporation tax for the same year.[20]

As a first approximation, none of this money was used to finance local services for businesses. During the 1980s, the responsibilities of LAs include social and health services for their residents, community safety, some aspects of education, housing, arts, culture, and environment. Very little is directly provided to businesses.[21] Even basic services such as refuse collection need to be separately organised by businesses. The only LA responsibility that can affect businesses directly is planning (and some trading regulations that affect mostly retail). Despite a national set of planning regulations, LAs differ in their speed of processing applications. It is hard to believe, however, that there were substantial within jurisdictional changes in the efficiency of local planning offices over the time period we consider. Instead, differences with respect to planning efficiency and restrictiveness, if they matter at all, are expected to be part of the fixed effect of these jurisdictions. Of course, if changes to taxes are used to fund services that directly or indirectly benefit businesses, then it is still of interest to understand the net effect.

During our period of study, tax rates were set locally by the 366 LAs in England, which entirely covered its land area. Large cities comprise several LAs (e.g., 33 for London). The tax paid depended on the value of the buildings occupied. The tax rates were known as 'rate poundages' and the value of the buildings as the 'rateable value'. We use the rate poundages as explanatory variable in our regressions. They were changed yearly. Rateable values of buildings were fixed in 1973 and did not change until 1990. The average rate for 1987 was 229.5 pence per pound of 1970 rateable value with a standard deviation of 32.2 pence. The ratio between the highest business rate (Oldham in Manchester) and the lowest (Kensington and Chelsea in London) was about three. These two LAs are far from

[20] Additional evidence that the tax is significant is provided by Bennett (1986) using a cost of capital approach.

[21] Clearly, there may be some indirect effects through, for instance, crime or education. However we expect these effects to spill over across jurisdictions so that establishments on both sides of any border face similar conditions.

each other and this difference will not be taken into account when we spatial difference. The largest ratio between two *neighbouring* LAs is nearly two. There was also significant time-series variation. There was for instance an average increase in the business rates of nearly 10 per cent between 1988 and 1989. Over the entire period ($1984 - 89$) the average increase was 45 per cent with a standard deviation of 0.15. Again looking at neighbouring LAs, we observe a 25 per cent decrease in Kensington and Chelsea against a 31 percent increase in neighbouring Hammersmith over the period. Overall, there is more than sufficient variation in the data to perform our estimations.

Finally, the data on shares of political parties that we use for instrumenting, come from the British Local Election Database which is available through the UK Data Archive. See Rallings and Thrasher (2004).

4. Results

To show the impact of spatial differencing, we also estimate using standard techniques. Results for the non-spatial specifications (i.e., using standard techniques and ignoring the micro-spatial nature of the data) are reported in table 1. We use log employment (e_{it}) as the dependent variable. As explanatory variables we include log tax rate (r_{at}), the three variables described above to capture the impact of age (X_{it}) and a full set of industry-year dummies (two-digit industries) for which we do not report coefficients.[22] Table 2 reports results for comparable specifications that use spatial differencing for establishments matched by industry and years. For the sake of comparison, we restrict the sample to be the same for all specifications by only including those establishments that we use to estimate our preferred specification: the instrumented, fixed effect spatial-differencing specification discussed below (the final specification in table 2). This restriction requires

[22]Industry-year dummies allow for industries to have different μ, γ and θ. In the spatial difference specifications we allow for this by only pairing establishments if they are part of the same industry. This restriction in defining pairs limits the degrees of freedom sufficiently that we cannot allow α and β to differ across industries. Thus, we impose the same assumption here to facilitate comparison. We could drop this assumption if we paired establishments across industries. But this would impose an identical distribution of θ across industries. As we are interested in the econometric issues arising because of the presence of θ we prefer to impose the restriction on α and β.

	OLS	WITHIN	WITHIN IV
(log) tax rate	0.210[a]	0.071[b]	0.363[a]
	(0.060)	(0.030)	(0.104)
age censored dummy	0.569[a]	0.214[a]	0.223[a]
	(0.091)	(0.033)	(0.034)
age	0.023	0.036[a]	0.036[a]
	(0.047)	(0.011)	(0.011)
age squared	0.005	-0.001	-0.001
	(0.005)	(0.001)	(0.001)
Adjusted R-squared	0.12		
Number of observations	13490	13490	13490
Number of establishments	4414	4414	4414

Notes: Regression of (log) employment on (log) local tax rates and age variables. First column reports results from OLS, second column (WITHIN) allows for establishment specific fixed effects, third column (WITHIN IV) further instruments local tax using local political variables. Standard errors under coefficients. [a], [b] and [c] denote significance at the 1%, 5%, and 10% level respectively.

Table 1. Non-spatial regression results

establishments to be in a pair in which both establishments simultaneously report employment in at least two years.[23]

Starting with 21,813 establishments, 7,938 of them are dropped because they only report employment once. Of the remaining 13,875 establishments, 8,792 have a neighbour within 1 kilometre (the distance threshold that we use for our reported results). However, only 4,414 of these establishments are in at least one pair where both establishments simultaneously report employment in at least two years. These establishments form the restricted sample for which we report results in the text. To show that this sample is representative, Appendix B reports results for the same specifications but without imposing this sample restriction.

The first regression in table 1 reports results from estimating equation (4) using pooled OLS for 1984-1989. The results for the age variables show that, as expected, older establishments have higher employment. Our main focus, however, is on the role of taxation.

[23]Consider an establishment that reports employment for (say) only 1987 and 1988. It is in the sample if it has at least one neighbour and that neighbour also reports employment for these two years. We impose this restriction because our implementation of spatial-differencing uses fixed effects for pairs. An establishment can be part of more than one pair satisfying this condition. However, we only use each establishment once in estimating the non-spatial specifications.

In the cross-section, higher tax rates are associated with higher employment.

As we noted above, one possible explanation of this positive correlation is that some establishments are larger than others for unobserved reasons and larger establishments happen to be located in higher tax jurisdictions. The second specification in table 1 allows for this possibility by introducing an establishment-specific fixed effect and calculating the *within* estimator. The coefficient on tax rate is divided by three suggesting that much of the positive correlation between employment and tax rate is indeed due to unobserved characteristics of establishments. Note, however, that the effect remains positive and significant.

The remaining problem that we need to tackle is that the tax rate may be correlated with the error in equation (5). There are two possible sources for this correlation. First, there may be a feedback from employment to tax rate. This feedback will be positive when local politicians tax local business more when it is doing well. Alternatively, and working in the opposite direction, it could be that LAs can afford to keep taxes low during 'good times'. In the UK context, this alternative may arise because good times imply a lesser need for social expenditure. This being said, we expect the first effect to dominate and taxation to go up when local business is doing well. Second, there may be other *time-varying* characteristics that are positively correlated with both employment and tax rates and that we do not control for through the use of establishment-specific fixed effects.

To solve these problems we use local political variables to instrument for tax rates. The full set of instruments includes the share of local politicians affiliated with the three main political parties (Conservative, Labour and Liberal Democrat), a set of dummies indicating whether the LA is controlled by one of the three main parties and a set of interactions giving the share of the three main parties if they control the LA. There are many smaller parties that play a role in local politics and we aggregate these in to 'other' and treat them as the omitted category. The R-squared of the first stage regression is 69%.

For the interested reader Appendix C provides further details.[24]

The third column in table 1 shows what happens when we use these instruments for the level of taxation in the fixed effects specification that we reported in column 2. Surprisingly, perhaps, the coefficient on taxation increases after instrumenting. It would be tempting to conclude that, contrary to most priors, LAs taxed businesses according to jurisdictions (social) needs rather than local business' capacity to pay. That is, there is a negative correlation between tax rates and the omitted variables so that instrumenting leads to a higher coefficient on taxation. Nonetheless, a positive effect with an elasticity close to 0.4 strikes us as implausible. As we will see below it appears that this increase in the coefficient on taxation following instrumentation occurs because the unobserved time-varying site-specific effect is correlated with both employment and local political variables.

Table 2 presents two sets of results (with and without corrected standard errors) for three different spatial specifications that parallel those presented in table 1. In the first column, we spatial difference equation (4) and estimate using OLS. In the second, we spatial difference (5) and estimate using a fixed effect for each pair of establishments. Finally in the third column, we instrument the tax rate in the spatially differenced *within* specification using spatially differenced political variables as in the non-spatial specification.

The results use a distance threshold of 1 kilometre to identify neighbours. In our choice of threshold, we face a tradeoff between sample size and the extent to which the spatially varying site-specific effects are equal across neighbouring sites. We chose the minimum threshold which gives sufficient observations to identify the effect of local taxes (remembering that identification comes from cross jurisdiction border pairs). We use neighbouring establishments within the same jurisdiction to improve our estimates of the

[24]One might be tempted to test for the validity of instruments given that we have more instruments than endogenous regressors. We would argue, however, that such a test is invalid because all our instruments are based on the same underlying assumption (that local politics is independent of changes to establishments' employment).

effect of establishment-specific variables.[25]

Uncorrected standard errors			
spatial difference of	OLS	WITHIN	WITHIN IV
(log) tax rate	0.846[a]	0.111	-1.024[a]
	(0.225)	(0.119)	(0.314)
age censored dummy	0.738[a]	0.134[a]	0.132[a]
	(0.076)	(0.028)	(0.028)
age	0.068[c]	0.042[a]	0.041[a]
	(0.039)	(0.009)	(0.009)
age squared	-0.003	-0.003[a]	-0.003[a]
	(0.004)	(0.001)	(0.001)
Corrected standard errors			
spatial difference of	OLS	WITHIN	WITHIN IV
(log) tax rate	0.846[b]	0.111	-1.024[b]
	(0.379)	(0.167)	(0.420)
age censored dummy	0.738[a]	0.134[a]	0.132[b]
	(0.138)	(0.051)	(0.049)
age	0.068	0.042[a]	0.041[a]
	(0.071)	(0.016)	(0.015)
age squared	-0.003	-0.003[c]	-0.003[a]
	(0.008)	(0.002)	(0.002)
Adjusted R-squared	0.04		
Number of observations	18370	18370	18370
Number of establishments	4414	4414	4414
Number of pairs	6087	6087	6087

Notes: Regression of spatial difference of (log) employment on spatial difference of (log) local tax rates and age variables. First column reports results from OLS, second column (WITHIN) allows for establishment specific fixed effects, third column (WITHIN IV) further instruments local tax using local political variables. Standard errors under coefficients. First block of results report uncorrected standard errors. Second block of results report standard errors corrected according to Appendix A. [a], [b] and [c] denote significance at the 1%, 5% and 10% level respectively.

Table 2. Spatial differencing regression results

Note that, although the overall sample of establishments is restricted to be identical for both the spatial and non-spatial specifications, the number of observations is higher for the spatially differenced specifications (18,370 compared to 13,490). This is because each establishment can be involved in more than one pair. Specifically, we have 6,087 unique

[25]We get the same results if we restrict attention only to the 164 establishments that are part of cross-jurisdiction border pairs. The spatially differenced within IV specification gives a coefficient of -1.072 as opposed to the -1.024 reported in the text. Both coefficients are significant at the 1% level.

pairs as compared to 4,414 unique establishments suggesting that each establishment has on average three neighbours.[26]

Before turning to the individual coefficients, comparing the two blocks of results (with and without corrected standard errors), we see the corrections outlined in Appendix A generally increase standard errors by around 50%. In our context, this results in minor changes in significance, but does not change overall findings. In other contexts it could, suggesting that the correction should usually be implemented. Turning to the coefficients, we see that, apart from changes in significance, the results on the age variables are essentially unchanged. As before, we focus on the tax rate and note that after spatial differencing we get a higher correlation between (spatially differenced) tax rate and (spatially differenced) employment than previously, with a coefficient of 0.846 compared to 0.210 with OLS. These coefficients have the same probability limit when there are no unobserved establishment or site-specific effects, or when these effects are uncorrelated with tax rates or other included explanatory variables. A possible reason for the higher correlation between employment and taxation after controlling for site-specific effects is that areas with poor sites had higher tax rates thus biasing downward the coefficient on taxes in the non-spatial OLS estimation. A correlation between having 'poor sites' and higher taxes is certainly believable given that de-industrialising LAs tended to vote for very left-wing councils that then greatly increased taxation.[27]

However, this interpretation assumes that establishment fixed effects are either absent or uncorrelated with local tax rates. To account for possible correlation we use, as before, the panel dimension of our data to control for unobserved establishment heterogeneity. Column 2 of table 2 reports results when we both spatially difference and allow for fixed effects for pairs of establishments. These pair fixed effects not only control for time-invariant unobserved establishment heterogeneity but also for other time-invariant

[26]Working against this, is the fact that both establishments in the pair must simultaneously report employment data in at least two years. For this distance threshold the first effect dominates. That need not be the case for other distance thresholds.

[27]Sheffield under the leadership of the (then) leftwing firebrand David Blunkett had the highest business rate in the country in 1990 and Liverpool led by the notorious Derek Hatton ranked 12th.

local effects such as the propensity of some jurisdictions to provide better services and thus have consistently better performing establishments.

Comparing results across the non-spatial and spatially differenced specifications we see that after controlling for establishment fixed effects, the coefficient on the tax rate is again higher after spatial differencing. However because of higher standard errors, it is hard to provide a definitive interpretation for this comparison. More significantly, comparing across the spatially differenced specifications, we see that allowing for pair fixed effects reduces our estimate of the positive correlation between taxation and employment relative to the spatially differenced OLS results. The coefficient even becomes insignificant. This confirms our finding from the non-spatial specifications that establishment fixed effects appear to be positively correlated with tax rates: LAs with 'good' establishments charge higher taxes.

As with the non-spatial fixed effects specifications, we still want to control for the fact that tax rates may be endogenous. To do this, we instrument using the spatial difference of the same set of political variables that we use for the non-spatial specification. Results are shown in column 3 of table 2. Taxation now has a negative effect on employment. As argued above, the difference between the spatial and non-spatial results suggests that unobserved time-varying site-specific effects are correlated with both employment and local political variables. Spatial differencing removes these site-specific effects, ensuring that our instruments are valid and thus allowing us to identify the negative effect of taxes on establishment employment.

Pulling the results together, spatial differencing offers two improvements over existing methodologies. First, comparing the non-instrumented spatial and non-spatial regressions allows us to identify the nature of the relationship between site-specific effects and local taxation. Second, and more importantly, because spatial differencing removes unobserved time-varying site-specific effects it makes it far easier to find valid instruments that allows us to identify the negative relationship between local taxation and employment. Quantitatively, spatial differencing greatly affects the results. In what is arguably the

best estimation using standard techniques (that does not control for time-varying site-specific effects), we find a positive elasticity around 0.4. Our preferred specification with spatial differencing reverses the sign of the coefficient and leads to an elasticity around -1 (although not very precisely estimated). In a nutshell, instead of a positive effect of taxation on employment spatial differencing shows a significant negative effect of taxation on employment. From our model we know that his occurs because of some combination of the growth-slowing and selection effects of higher taxes. Focusing on entries allows us to focus specifically on the second of these effects and it is to this issue that we now turn.

5. Entries

A. *Methodology and data*

We now turn to the effects of local taxation on entry.[28] This is an important issue for three reasons. First, the rate at which new establishments enter LAs is an important outcome that deserves attention. Second, our employment estimates confound selection and growth slow-down effects as made clear by our model. Assuming the supply of sites is fixed and that the land market clears so that exits are matched with entries, an analysis of entries offers an opportunity to focus specifically on selection effects. Third spatial differencing provides solutions to the same three problems (establishment heterogeneity, site heterogeneity, and endogeneity) that we addressed when looking at employment growth. However, there are subtle differences which are worth highlighting.

Consider establishment i that wishes to enter in year t. It can choose between all available sites, indexed by z. As before, the jurisdiction that sets the tax for the establishment depends on the site occupied, and is indexed by a. Profit maximisation can be performed in two stages. First, establishment i computes the highest profit it can achieve, Π_{iz}, at each

[28] Rathelot and Sillard (2008) develop an analysis of the effect of a local tax on capital on entries using a similar approach. Details of the implementation differ.

site z. It then selects the site offering the highest profit.[29] We assume that the highest profit for establishment i entering in year t at site z can be written as:

$$\Pi_{izt} = \lambda r_{at} + Z_{it}\zeta + \nu_i + \kappa_a + \varphi_{zt} + \epsilon_{izt} \tag{9}$$

where ν_i is an establishment fixed effect, κ_a is a jurisdiction fixed effect, Z_{it} are explanatory variables at the establishment level, φ_{zt} is a site-specific effect, and ϵ_{izt} is an establishment site-specific shock.[30] Establishment i will choose the site z that gives the highest expected profit. When the shocks ϵ_{izt} follow an appropriate *iid* extreme value distribution, the probability of choosing site z, P_{izt}, is logistic and is given by

$$P_{izt} = \frac{\exp E(\Pi_{izt})}{\sum_{z=1}^{Z} \exp E(\Pi_{izt})} \tag{10}$$

where $E(\cdot)$ is the expectation operator and the summation is across all possible sites Z.

The standard approach to estimating the coefficients λ and ζ is to ignore the site-specific effect, φ_{zt}, and estimate a conditional logit model. To do this, one creates a set of establishment-jurisdiction observations and defines $c_{ia} = 1$ if establishment i locates in jurisdiction a and $c_{ia} = 0$ otherwise. The coefficients can then be estimated by maximising the log likelihood of the conditional logit model:

$$\log L_{cl} = \sum_{i=1}^{I} \sum_{a=1}^{A} c_{ia} \log P_{ia} \tag{11}$$

where for simplicity we drop the time subscripts as establishments only enter once.

As is well recognised, application of the conditional logit model can be problematic when the set of possible jurisdictions is large. One solution is to take a random sub-sample of jurisdictions, although this has implications for the efficiency of the estimator and the small sample properties are unknown. Another possibility, recently proposed by

[29] We ignore any possible interaction between the location decisions of entrants. This seems reasonable in established manufacturing industries where existing establishments drive local wages, determine product market competition etc. As discussed in the text, the fact that the effect on profits of these factors will be highly correlated across neighbouring sites then justifies our approach.

[30] The establishment fixed effect, ν_i, mirrors μ_i in (4). Similarly the jurisdiction effect, κ_a, and the site-specific effect, φ_{zt}, are the counterparts of γ_a and θ_{zt}. Finally both (9) and (4) contain coefficients for the effect of local taxation (λ and α) and establishment-level variables (ζ and β). Note that our approach for entries is also consistent with a more general setting where the specification includes variables that are individual-site-specific, ϕ_{izt}, in addition to individual effects, ν_i, and site effects, φ_{zt}.

Guimaraes, Figueiredo, and Woodward (2003) is to use the fact that, under certain conditions, the log likelihood of the Poisson model is identical to that of the conditional logit. Estimating a Poisson regression is computationally much easier, though the equivalence between the likelihoods only holds in the absence of establishment specific variables (i.e., Z_{it}). In any case both solutions ignore the site-specific effects. As we now show, spatial differencing provides an alternative which controls for site-specific effects and which, in other contexts, would allow for the inclusion of establishment specific variables.

Our approach is as follows. Consider two neighbouring sites, z_1 and z_2, close to the border between two jurisdictions a_1 and a_2. z_1 is located in jurisdiction a_1 and z_2 in a_2. Since the two locations are close, we assume $\varphi_{z_1 t} \approx \varphi_{z_2 t}$. This is the same identification assumption made in section 2 to derive our employment specification (8). To repeat, this assumption is justified by the fact that site-specific effects (labour market conditions, access to markets and major facilities, etc) vary smoothly across space. The probability of choosing z_1 conditional on locating in one of these two neighbouring sites is:

$$P\left(i \in z_1 \,|i \in \{z_1, z_2\}\right) = \frac{P_{iz_1}}{P_{iz_1} + P_{iz_2}} \tag{12}$$

When the shocks ϵ_{izt} follow an appropriate *iid* extreme value distribution, the probability of choosing one of the sites is logistic and is given by:

$$P\left(i \in z_1 \,|i \in \{z_1, z_2\}\right) = \frac{1}{1 + \exp(\lambda\left(r_{a_2 t} - r_{a_1 t}\right) + \kappa_{a_2} - \kappa_{a_1})} \tag{13}$$

Note that unlike equation (10) above, this specification conditions out both establishment- and site-specific effects because these effects are the same at locations z_1 and z_2.[31] Recall that in standard conditional logit models observed site-specific factors are computationally hard to deal with. By contrast our approach directly conditions out both observed and unobserved site-specific factors in a way that is easy to implement. In addition, we do not need to rely on the assumption of the Independence of Irrelevant Alternatives which

[31] We assumed in (9) that the effect of establishment characteristics did not depend on location. However our spatial differentiation approach is more powerful than this since any interaction between establishment and site characteristics is conditioned out by spatial differencing provided the characteristics of the local environment vary smoothly over space.

underlies the conditional logit model. This is a distinct advantage as this assumption is unlikely to hold in spatial settings (see, for example, Head and Mayer, 2004).

Equation (13) only involves jurisdictional level variables so we can estimate λ directly from an aggregate logit model where the observation units are the (border-side, time) pairs.[32] In the estimation, the observations are weighted by the number of entrants for consistency with equation (13).

In a nutshell, we select entrants located close to jurisdictional boundaries and examine their decision to choose to locate on one particular side of the border. The main conceptual difference with our employment regression (8) is that we consider the location decision of a new establishment choosing between neighbouring *jurisdictions* rather than comparing the employment outcomes for (existing) neighbouring establishments. For entry, this is the appropriate way to control for both time invariant establishment specific effects and any unobserved site-specific effects common to both sides of a border.[33] As local tax rates do not vary smoothly across space at jurisdictional boundaries, we can use entrants on either side of these boundaries to identify the effect of taxes after conditioning out time-varying site-specific factors and time-invariant establishment-specific effects. Interestingly one step is enough to eliminate both establishment and site effects for entry, whereas two steps are necessary when studying employment. This reflects the fact that instead of comparing each establishment with a matched establishment on the other side of the border we compare each establishment with itself on the two sides of the border making the *within* transformation redundant. In passing, we note that the same idea cannot be used for exits.[34]

[32]We need to estimate 366 jurisdiction dummies. These dummies are identified from the time variation in entry rates at the borders.

[33]To see this, note that when choosing between two neighbouring sites, the entrant compares profits between them. Time-varying site-specific factors, which vary smoothly across space, will affect profits at both sites in the same way. Hence, these factors do not enter into the location decision. A similar argument applies to unobserved time-invariant establishment characteristics.

[34]Indeed, we can observe the exit of an establishment only on the side of the border where it entered. Hence, unlike with entry, we cannot compare an establishment with itself on both sides of a border. Alternatively, one might want to apply the same methodology as for employment by matching establishments with their closest neighbour(s). However, since establishments exit only once, we would not be able to control for establishment unobserved heterogeneity.

As with the employment specifications, endogeneity must also be addressed since the local tax rate may be simultaneously determined with the rate of entry. To control for this, we estimate a two-stage IV *logit* model instrumenting the difference in tax rate, $r_{a_2t} - r_{a_1t}$, in the logit specification, with the predicted difference in tax rate from a first stage regression using spatially-differenced local political variables. The simplicity of this method comes at a price because, as is well known, correcting the errors to allow for possible correlation between the residual of the instrumentation equation and that of the entry equation is far from straightforward.[35]

B. *Results*

To construct the entry data, we need to detect all entrants in the ARD. Because of a change in 1984 in the way the registry of establishments was constructed, there is a large amount of artificial entry in 1984 and 1985 (i.e., establishments enter the data set for the first time during those two years even though they already existed prior to 1984). See Griffith (1999) for further discussion. As a result, we ignore entries for these two years and focus instead on 81,042 newly reporting establishments between 1986 and 1989.

For consistency with the employment regressions we would like to identify all entrants within 1 kilometre of jurisdiction boundaries. The easiest way to do this would be to draw 1 kilometre buffers around boundaries. Lacking a set of digital boundaries for this time period, we instead proceed indirectly and identify the set of border entrants from the ARD itself. To do this, for each entrant we searched for the closest establishment located in each of the neighbouring LAs and retained only those entrants that had such a neighbour within one kilometre. Since this detection procedure is only meant to compute distances to the LA border for each entrant (rather than find a match for a pair), we considered all possible establishments in all sectors and all years as potential neighbours. We expect this procedure to catch nearly all entrants located within one kilometre of a border.

[35] As an alternative, we used an IV *probit* model that estimates the entry and instrumenting equation simultaneously. This approach corrects standard errors but at the cost that it is not fully consistent with the theoretical specification. The results are the same as with the two-stage *logit* model.

	CL	CL IV	LOGIT	LOGIT IV
(log) tax rate	0.397[a]	0.521	0.108	0.809
	(0.079)	(0.883)	(0.177)	(0.921)
Number of entrants	19,337	19,337	19,337	19,337

Notes: Number of entrants as a function of local tax rates. First column (CL) reports results for conditional logit, second column (CL IV) instruments local taxes using political variables, third column (LOGIT) reports results from a logit model for spatially differenced variables; fourth column (LOGIT IV) instruments local taxes. Standard errors under coefficients. [a], denotes significance at the 1% level. Estimates in the first two columns are from a Poisson regression using the equivalence result from Guimaraes *et al.* (2003).

Table 3. Results for entries

As for employment, we implement both the standard methodology (i.e., conditional logit) and our spatial differencing approach. Results for two non-spatial specifications, are given in columns 1 and 2 of table 3 while columns 3 and 4 report results for two comparable spatial specifications. Again, for the sake of comparison, we have restricted the sample to be the same for all specifications. This restriction requires establishments to locate within 1 kilometre of a boundary between two English LAs between 1986 and 1989. Imposing this requirement leaves us with a sample of 19,337 establishments. To show that this sample is representative, Appendix B once again reports results for the same specifications but without imposing this restriction.

Starting with the conditional logit we see, from the results reported in column 1, that there appears to be a positive effect of tax rates on entry. Column 2 shows what happens when we correct for endogeneity by replacing the actual tax rates with the predicted tax rates from a first stage regression of tax rates on political variables. Once instrumenting, we find that the coefficient increases but it also becomes insignificant because of a much higher standard error. Correcting the standard errors to allow for the fact that we are instrumenting would only reinforce this finding.[36] Columns 3 and 4 show that we reach the same conclusions for entry using our spatially differenced approach.

A positive effect of local taxation on entries (although insignificant) might seem surprising in light of its negative effect on employment. Nonetheless, having a positive effect

[36]Hence our decision only to implement the theoretically consistent two stage conditional logit procedure rather than an instrumented probit specification.

of taxation on entries is consistent with our model. Recall that higher taxes lead to more exits and relocations. This creates vacant sites. In turn, these vacant sites are occupied by new entrants. Overall, our findings about entries are suggestive that the selection effect highlighted by the model plays a role. This selection effect might be more complicated than in our model. In particular, it could be that exiting and relocating establishments in high tax jurisdictions are large and capital intensive. They might be replaced by smaller and less capital intensive firms which are less sensitive to high local taxes. We also note that our model assumes that the supply of sites is fixed. The development of new sites might be negatively affected by high local taxes (a force countering our selection effect).[37]

6. Conclusion

We propose a new approach to assess the effects of local taxation. Our results show the importance of controlling for both unobserved establishment-specific and unobserved site-specific characteristics and possible simultaneity. Simple OLS results suggest a positive relationship between employment and taxes. Allowing for unobserved establishment-specific effects and instrumenting for local taxation, we still find a positive relationship between employment and taxes. Allowing for unobserved location-specific effects and instrumenting for local taxation, we find a negative significant relationship between employment and taxes. By contrast we find that local taxation has no effect on the entry of new establishments.

Beyond our methodological contribution, this analysis also suggests that the study of local taxation and, more broadly, that of decentralised public intervention faces serious endogeneity problems whereby local public decisions depend strongly on very local conditions, which are extremely difficult to control for. As shown here, properly controlling for such local conditions is a necessary condition to obtain reliable estimates.

[37] We could also imagine that jurisdictions where more new sites can be developed raise taxes to maximise tax revenue. This could lead to a positive correlation between new developments and taxes. This should normally be corrected by our instrumentation strategy.

The second broad lesson is that even taxes that are seemingly close to an 'ideal' tax that would be free of distortion can in practice generate significant distortions. Despite the fact that the UK business rates were close to George's (1884) 'pure' land tax, revaluations in case of expansions and frictions in the rental market implied that increases in local taxation had an adverse effect on employment.

References

Arnott, Richard. 2004. Does the Henry George Theorem provide a practical guide to optimal city size? *American Journal of Economics and Sociology* 63(5):1057–1090.

Baldwin, Richard E., Rikard Forslid, Philippe Martin, Gianmarco I. P. Ottaviano, and Frederic Robert-Nicoud. 2004. *Economic Geography and Public Policy*. New Jersey: Princeton University Press.

Bartik, Timothy. 1991. *Who benefits from state and local economic development policies?* Kalamazoo, Michigan: W. E. Upjohn Institute for Employment Research.

Bennett, Robert. 1986. The impact of non-domestic rates on profitability and investment. *Fiscal Studies* 7(1):34–50.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics* 119(1):249–275.

Black, Sandra E. 1999. Do better schools matter? Parental valuation of elementary education. *Quarterly Journal of Economics* 114(2):587–599.

Bond, Stephen, Kevin Denny, John Hall, and William McCluskey. 1996. Who pays business rates? *Fiscal Studies* 17(1):19–35.

Chirinko, Robert S. and Daniel J. Wilson. 2008. State investment tax incentives: A zero-sum game? *Journal of Public Economics* 92(12):2362–2384.

Crosby, Neil, Colin Lizieri, Sandi Murdoch, and Charles Ward. 1998. Implications of changing lease structures on the pricing of lease contracts. The cutting edge, Royal Institution of Chartered Surveyors.

Duranton, Gilles and Henry G. Overman. 2005. Testing for localization using micro-geographic data. *Review of Economic Studies* 72(4):1077–1106.

Epple, Dennis and Thomas Nechyba. 2004. Fiscal decentralization. In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland.

Fujita, Masahisa. 1989. *Urban Economic Theory: Land Use and City Size*. Cambridge: Cambridge University Press.

George, Henry. 1884. *Progress and Poverty: An Inquiry into Causes of Industrial Depressions, and of Increase of Want with Increase of Wealth. The Remedy*. London: W. Reeves.

Gibbons, Steve and Stephen Machin. 2003. Valuing english primary schools. *Journal of Urban Economics* 53(2):197–219.

Griffith, Rachel. 1999. Using the ARD establishment level data: an application to estimating production functions. *Economic Journal* 109(456):F416–F442.

Guimaraes, Paulo, Octavio Figueiredo, and Douglas Woodward. 2003. A tractable approach to the firm location decision problem. *Review of Economics and Statistics* 85(1):201–204.

Guimaraes, Paulo, Octavio Figueiredo, and Douglas Woodward. 2004. Industrial location modeling: Extending the random utility framework. *Journal of Regional Science* 44(1):1–20.

Hale, Rita and Tony Travers. 1994. *A National or Local Tax: A Study of the Non Domestic Rate*. London: The Chartered Institute of Public Finance and Accountancy.

Head, Keith and Thierry Mayer. 2004. Market potential and the location of Japanese investment in Europe. *Review of Economics and Statistics* 86(4):959–972.

Holmes, Thomas J. 1998. The effect of state policies on the location of manufacturing: Evidence from state borders. *Journal of Political Economy* 106(4):667–705.

Kahn, Matthew E. 2004. Domestic pollution havens: Evidence from cancer deaths in border counties. *Journal of Urban Economics* 56(1):51–69.

Moulton, Brent R. 1990. An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *Review of Economics and Statistics* 72(2):334–338.

Rallings, C. and M. Thrasher (eds.) . 2004. *Local Elections in Britain: a statistical digest*. Plymouth: Local Government Chronicle Elections Centre.

Rathelot, Roland and Patrick Sillard. 2008. The importance of local corporate taxes in business location decisions: Evidence from French micro data. *Economic Journal* 118(527):499–514.

Thompson, Robert and Sotiris Tsolacos. 2001. Industrial land values – a guide to future markets? *Journal of Real Estate Research* 21(1/2):55–76.

Tiebout, Charles M. 1956. A pure theory or local expenditures. *Journal of Political Economy* 64(5):416–424.

Wilson, John D. 1999. Theories of tax competition. *National Tax Journal* 52(2):269–304.

Appendix A. Correction of the standard errors

When spatially differencing, an establishment i that has n neighbours will be in n pairs. This induces correlation in the error for all n of these pairs. The correlation arises because ϵ_{it} (the error of establishment i) enters the error of each pair. This imposes a particular structure to the covariance matrix which we use to correct the standard error. This appendix gives the details of that correction. Note that Bertrand, Duflo, and Mullainathan

(2004) consider similar issues when proposing their correction for the standard errors of difference-in-difference estimators to measure treatment effects. There are several key differences between our correction and the one proposed there. First, their correction is difficult to apply to specifications with a large number of establishments and locations because they estimate the treatment effect directly without rewriting the model in difference-in-difference. Second, their correction requires the covariance matrix to be block diagonal. This means that it is not applicable to situations, like the one here, where there is no obvious way to construct closed sets of neighbours (because establishment A may be a neighbour to establishment B, who may be neighbour to C etc). In short we deal with an error structure which is considerably more complex. This comes at a cost: we ignore issues arising from serial correlation of the errors that are the key concern of that paper.

After spatial-differentiation and within-pair projection, the model can be re-written as follows:

$$W\Delta e = \alpha W\Delta r + W\Delta X\beta + W\Delta\epsilon \tag{A 1}$$

$$= Z\gamma + W\Delta\epsilon \tag{A 2}$$

where for any variable v_{it}, observations have been stacked in *pair* and time order, $\gamma' = \left(\alpha,\beta'\right)'$ and $Z = (W\Delta r, W\Delta X)$. The OLS estimator is then:

$$\widehat{\gamma} = \left(Z'Z\right)^{-1} Z'\Delta e = \gamma + \left(Z'Z\right)^{-1} Z'\Delta\epsilon \tag{A 3}$$

We suppose that the residuals ϵ_{it} are *iid* with variance σ^2. The variance of the OLS estimator is:

$$V\left(\widehat{\gamma}\right) = \sigma^2 ABA \tag{A 4}$$

where $A = \left(Z'Z\right)^{-1}$ and $B = Z'\Delta\Delta'Z$. Matrix A is easy to compute since Z can be obtained after spatial-differenciation and a projection *within*-pair of the explanatory variables. Matrix B can also be computed after using an algorithm to obtain $\Delta'Z$. This algorithm relies on the fact that Δ has a simple structure. Indeed, denote $p \in \{1,...,P\}$ where P is the

number of pairs and N_p the number of years that pair p appears in the data. We can decompose Δ in blocks such that $\Delta = (\Delta'_1,..., \Delta'_P)'$ where, for instance, block Δ_p writes:

$$\begin{pmatrix} ... & 0 & 1 & 0 & ... & 0 & -1 & 0 & ... & ... \\ ... & 0 & 0 & 1 & 0 & ... & 0 & -1 & 0 & ... \\ ... & 0 & 0 & 0 & 1 & 0 & ... & 0 & -1 & ... \end{pmatrix} \tag{A 5}$$

(supposing $N_p = 3$). The first line corresponds to the first year that the pair is in the data, the second line to the second year, etc... Each column of Δ can contain the values 1 and -1 several times depending on the number of times an establishment has been matched with neighbours in the corresponding year. For a column i of Δ, denote j_i the identifier of the corresponding establishment and t_i the corresponding date. If *establishment* is the vector containing all the establishment identifiers and *year* is the vector containing the years, j_i and t_i can be retrieved from the first element in column i that takes a value different from 0. For example, in the GAUSS language, column i of Δ is then of the form $(establishment\ .eq\ establishment\,[j_i]).*(year\ .eq\ year[t_i]) - (establishment_n\ .eq\ establishment\,[j_i]).*(year_n\ .eq\ year[t_i])$ where $establishment_n$ contains the identifier of the neighbouring establishments and $year_n$ their years. Element (n,k) of $\Delta' Z$ can be computed using column n of Δ and column k of Z. The whole matrix $\Delta' Z$ is obtained from a loop over n and k.

We now propose an estimator of σ^2. Denote $\widehat{W\Delta\epsilon}$ the vector of residuals from the OLS estimation. We have:

$$\widehat{W\Delta\epsilon} = W\Delta e - Z\widehat{\gamma} = M_Z W\Delta\epsilon \tag{A 6}$$

where M_Z is the projector in the dimension orthogonal to Z. We then get:

$$\widehat{W\Delta\epsilon}'\widehat{W\Delta\epsilon} = \epsilon'\Delta' W M_Z W\Delta\epsilon \tag{A 7}$$

From this formula, we obtain:

$$\begin{aligned} E\left(\widehat{W\Delta\epsilon}'\widehat{W\Delta\epsilon}\right) &= Etr\left(\widehat{W\Delta\epsilon}'\widehat{W\Delta\epsilon}\right) && \text{(A 8)} \\ &= \sigma^2 tr\left(\Delta' W\Delta\right) - \sigma^2 tr\left[Z\left(Z'Z\right)^{-1} Z'\Delta\Delta' W\right] && \text{(A 9)} \\ &= \sigma^2 tr\left(W\Delta\Delta'\right) - \sigma^2 tr\left(AB\right) && \text{(A 10)} \end{aligned}$$

We can recover $tr\,(AB)$ very easily from A and B. It is also possible to simplify the expression: $tr\,(W\Delta\Delta')$. We can write $\Delta\Delta'$ in blocks corresponding to pairs. Indeed, the (p,q)-block writes: $\Delta_p\Delta_q'$. W is block diagonal. Thus, the (p,q)-block of $W\Delta\Delta'$ writes $W_p\Delta_p\Delta_q'$ where W_p is the (p,p)-block of W. Hence, we get: $tr\,(W\Delta\Delta') = \sum_p tr\left(W_p\Delta_p\Delta_p'\right)$. As we have $\Delta_p\Delta_p' = 2I_{T_p}$ and $tr\left(W_p\right) = T_p - 1$ (where T_p is the number of years that pair p appears in the data), we finally get: $tr\,(W\Delta\Delta') = 2\,(N-P)$. An unbiased *(and consistent)* estimator of σ^2 is then:

$$\widehat{\sigma}^2 = \frac{1}{2\,(N-P) - tr\,(AB)}\widehat{W\Delta\epsilon}'\widehat{W\Delta\epsilon} \tag{A 11}$$

We can finally deduce an estimator of the variance of $\widehat{\gamma}$:

$$\widehat{V}\left(\widehat{\gamma}\right) = \widehat{\sigma}^2 ABA \tag{A 12}$$

We now compute the standard errors when instrumenting. The model is:

$$W\Delta e = \alpha W\Delta r + W\Delta X\beta + W\Delta\epsilon \tag{A 13}$$

$$W\Delta r = Y\delta + \xi \tag{A 14}$$

with $Y = W\Delta P$ where P are some political variables, and $cov\,(Y,\xi) = cov\,(Y,W\Delta\epsilon) = 0$ by assumption. Denote $\widehat{\delta}$ the OLS estimator of δ obtained from equation (A 14) and $V = \widehat{V}\left(\widehat{\delta}\,|X\right)$ an estimator of its covariance. This covariance estimator may simply be the usual OLS estimator. It may also take into account clusters at the jurisdiction level. Equation (A 13) rewrites:

$$W\Delta e = \alpha Y\widehat{\delta} + W\Delta X\beta + W\Delta\epsilon + \phi \tag{A 15}$$

$$= \widetilde{Z}\gamma + W\Delta\epsilon + \phi \tag{A 16}$$

with $\widetilde{Z} = \left(Y\widehat{\delta},W\Delta X\right)$ and $\phi = \alpha\left(Y\delta - Y\widehat{\delta}\right)$. ϕ is such that $E\,(\phi\,|X,Y) = 0$ and $V\,(\phi\,|X,Y) = \alpha^2 YVY'$ with $V = V\left(\widehat{\delta}\,|Y\right)$. The IV estimator is:

$$\widehat{\gamma}_{IV} = \left(\widetilde{Z}'\widetilde{Z}\right)^{-1}\widetilde{Z}'W\Delta e = \gamma + \left(\widetilde{Z}'\widetilde{Z}\right)^{-1}\widetilde{Z}'\left(W\Delta\epsilon + \phi\right) \tag{A 17}$$

Assuming that $\widehat{\delta}$ is known, the variance of the IV estimator can be approximated by:

$$V\left(\widehat{\gamma}_{IV}\right) \approx \left(\widetilde{Z}'\widetilde{Z}\right)^{-1}\widetilde{Z}'\left(\sigma^2 W\Delta\Delta' W + \alpha^2 YVY'\right)\widetilde{Z}\left(\widetilde{Z}'\widetilde{Z}\right)^{-1} \tag{A 18}$$

$$\approx \widetilde{A}\left(\sigma^2\widetilde{B} + \alpha^2\widetilde{C}\right)\widetilde{A} \tag{A 19}$$

with $\widetilde{A} = \left(\widetilde{Z}'\widetilde{Z}\right)^{-1}$, $\widetilde{B} = \widetilde{Z}'W\Delta\Delta'W\widetilde{Z}$ and $\widetilde{C} = \widetilde{Z}'YVY'\widetilde{Z}$. $\widetilde{A}$ and $\widetilde{B}$ are easy to compute (see the OLS case). $\widetilde{C}$ is also easy to compute since we can first compute $Y'\widetilde{Z}$, which has a small dimension. We now propose an estimator of σ^2. Denote $\widehat{W\Delta\epsilon + \phi}$ the vector of residuals from the IV second-stage estimation. We have:

$$\widehat{W\Delta\epsilon + \phi}'\widehat{W\Delta\epsilon + \phi} = \left(W\Delta\epsilon + \phi\right)' M_{\widetilde{Z}}\left(W\Delta\epsilon + \phi\right) \tag{A 20}$$

Consider for a while that Z is non random (i.e., $\widehat{\delta}$ is non random). We have:

$$E(\widehat{W\Delta\epsilon + \phi}'\widehat{W\Delta\epsilon + \phi}) = trE\left(\epsilon'\Delta'WM_{\widetilde{Z}}W\Delta\epsilon\right) + trE\left(\phi' M_{\widetilde{Z}}\phi\right) \tag{A 21}$$

$$= \sigma^2\left[2(N-P) - tr(\widetilde{A}\widetilde{B})\right] + \alpha^2\left[tr(VY'Y) - tr(\widetilde{A}\widetilde{C})\right] \tag{A 22}$$

A *(consistent)* estimator of σ^2 is then:

$$\widehat{\sigma}^2_{IV} = \frac{\widehat{W\Delta\epsilon + \phi}'\widehat{W\Delta\epsilon + \phi} - \widehat{\alpha}^2_{IV}\left[tr\left(\widehat{V}Y'Y\right) - tr\left(\widetilde{A}\widehat{\widetilde{C}}\right)\right]}{2\left(N-P\right) - tr\left(\widetilde{A}\widetilde{B}\right)} \tag{A 23}$$

with $\widehat{V}$ an estimator of V obtained from the first-stage equation, $\widehat{\widetilde{C}} = \widetilde{Z}'Y\widehat{V}Y'\widetilde{Z}$. Note that when the residuals ξ_{it} are *iid* with variance θ^2, we have $tr\left(\widehat{V}Y'Y\right) = N\widehat{\theta}^2$ and $tr\left(\widetilde{A}\widehat{\widetilde{C}}\right) = \widetilde{\theta}^2 tr\left(P_{\widetilde{Z}}P_Y\right)$. The estimator of σ^2 becomes:

$$\widehat{\sigma}^2_{IV} = \frac{\widehat{W\Delta\epsilon + \phi}'\widehat{W\Delta\epsilon + \phi} - \widehat{\alpha}^2_{IV}\widehat{\theta}^2\left[N - tr\left(P_{\widetilde{Z}}P_Y\right)\right]}{2\left(N-P\right) - tr\left(\widetilde{A}\widetilde{B}\right)} \tag{A 24}$$

Finally, the variance of the IV estimator can be approximated by:

$$\widehat{V}\left(\widehat{\gamma}_{IV}\right) = \widetilde{A}\left(\widehat{\sigma}^2_{IV}\widetilde{B} + \widehat{\alpha}^2_{IV}\widehat{\widetilde{C}}\right)\widetilde{A} \tag{A 25}$$

Appendix B. Non-spatial results for different samples of establishments

Tables 4 and 5 present results for the non-spatial employment specifications using less restrictive samples than those used in the text. These results should be compared to those

in table 1, where establishments must be (i) in a pair in which (ii) both establishments simultaneously report employment in at least two years (the sample used for our preferred specification). Table 4 uses the largest possible sample for each specification imposing neither restriction (i) or (ii). Table 5 uses the largest possible sample of establishments that are part of a pair. That is, we impose restriction (i) but not restriction (ii). As can be seen from the comparison of tables 1, 4 and 5, all point estimates are of the same sign and do not significantly differ from each other.

Table 6 present results for the non-spatial entry specifications using all entrants. These results should be compared to those in table 3, where establishments must enter within 1 kilometre of an LA boundary to be part of the sample. All point estimates are of the same sign and do not significantly differ from each other.

	OLS	WITHIN	WITHIN IV
(log) tax rate	0.222[a]	0.131[a]	0.481[a]
	(0.029)	(0.016)	(0.062)
age censored dummy	0.652[a]	0.290[a]	0.294[a]
	(0.033)	(0.016)	(0.015)
age	-0.015	0.053[a]	0.055[a]
	(0.018)	(0.005)	(0.005)
age squared	0.011[a]	-0.002[a]	-0.003[a]
	(0.002)	(0.001)	(0.001)
Adjusted R-squared	0.13		
Number of observations	61785	53684	53684
Number of establishments	21813	13875	13875

Notes: Standard errors under coefficients. [a], [b] and [c] denote significance at the 1%, 5% and 10% level respectively. Largest possible sample. First column (OLS) only restricted by data availability. Second and third column require at least two observations per establishment. Compare to table 1 where establishments must be (i) in a pair in which (ii) both establishments simultaneously report employment in at least two years.

Table 4. Non-spatial regression results for largest possible samples

	OLS	WITHIN	WITHIN IV
(log) tax rate	0.156[a]	0.105[a]	0.472[a]
	(0.042)	(0.023)	(0.080)
age censored dummy	0.662[a]	0.208[a]	0.211[a]
	(0.054)	(0.025)	(0.025)
age	0.004	0.046[a]	0.047[a]
	(0.030)	(0.008)	(0.008)
age squared	0.008[b]	-0.002[a]	-0.003[a]
	(0.003)	(0.001)	(0.001)
Adjusted R-squared	0.13		
Number of observations	25579	22803	22803
Number of establishments	5564	5852	5852

Notes: Standard errors under coefficients. [a], [b] and [c] denote significance at the 1%, 5% and 10% level respectively. Sample restricted to establishments that are part of a pair. First column (OLS) only restricted by data availability. Second and third column require at least two observations per establishment. Compare to table 1 where establishments must be (i) in a pair in which (ii) both establishments simultaneously report employment in at least two years.

Table 5. Non-spatial regression results for largest possible samples of establishments in pairs

	CL	CL IV
(log) tax rate	0.633[a]	0.267
	(0.070)	(0.244)
Number of establishments	81,042	81,042

Notes: Number of entrants as a function of local tax rates. First column (CL) reports results for conditional logit, second column (CL IV) instruments local taxes using political variables. Largest possible sample. Standard errors under coefficients. [a], denotes significance at the 1% level. Estimates are from a Poisson regression using the equivalence result from Guimaraes *et al.* (2003).

Table 6. Non-spatial regression results for largest possible sample

Appendix C. First stage regression

Table 7 presents the results from the first stage regression of spatially difference (log) tax rates on the exogenous variables and spatially differenced instruments. We present the *within* version that is used for instrumenting the spatially differenced specification in the text and report both corrected and uncorrected standard errors.

spatial difference of	(1)	(2)
age censored dummy	-0.0029	-0.0029
	(0.0036)	(0.0036)
age	-0.0006	-0.0006
	(0.0011)	(0.0011)
age squared	0.0001	0.0001
	(0.0001)	(0.0001)
share Conservative	0.1279[a]	0.1279[a]
	(0.0436)	(0.0482)
share Labour	0.2466[a]	0.2466[a]
	(0.0429)	(0.0470)
share Liberals	0.1236[a]	0.1236[a]
	(0.0394)	(0.0426)
Conservative controlled	0.0813[a]	0.0813[a]
	(0.0178)	(0.0226)
Labour controlled	-0.1715[a]	-0.1715[a]
	(0.0154)	(0.0173)
Liberal controlled	-0.1376[a]	-0.1376[a]
	(0.0244)	(0.0281)
share Conservative (if control)	-0.2278[a]	-0.2278[a]
	(0.0325)	(0.0431)
share Labour (if control)	0.2645[a]	0.2645[a]
	(0.0317)	(0.0355)
share Liberal (if control)	0.2659[a]	0.2659[a]
	(0.0428)	(0.0497)
Number of observations	18370	18370
Number of establishments	6087	6087

Notes: Standard errors under coefficients. [a], [b] and [c] denote significance at the 1%, 5% and 10% level respectively. Results from first stage regression of spatial difference of tax rates on establishment fixed effects, exogenous variables and political instruments. Column (1) presents results with uncorrected standard errors. Column (2) presents results with standard errors corrected according to Appendix A.

Table 7. First stage regression results

B.4. Spatial disparities in hospital performances

Gobillon L. et C. Milcent (2009), “Spatial disparities in hospital performances”, PSE Working Paper 2008-12 (sous un autre titre), version revisée.

57 pages

Spatial disparities in hospital performances*

Laurent Gobillon† Carine Milcent‡

November 25, 2009

*We are grateful to Chong-en Bai, Marteen Lindeboom, Thierry Magnac, Francesco Moscone and Jordan Rappoport for very useful discussions, as well as the participants of the 16th EU Health Econometrics Workshop in Bergen and the 54th NARSC in Savannah. This project was carried out with the financial support of the French Direction of Research, Studies, Evaluation and Statistics (DREES). We are responsible for all remaining errors.

†INED, PSE-INRA and CREST. INED, 133 boulevard Davout, 75980 Paris Cedex. Tel: 0033(1)56062016. Email: laurent.gobillon@ined.fr. Webpage: http://laurent.gobillon.free.fr.

‡PSE (CNRS-EHESS-ENPC-ENS), 48 boulevard Jourdan, 75014 Paris. email: milcent@pse.ens.fr. Webpage: http://www.pse.ens.fr/milcent/index.html.

Abstract

Spatial disparities in mortality can result from spatial differences in patient characteristics, treatments, hospital characteristics, and local healthcare market structure. To distinguish between these explanatory factors, we estimate a flexible duration model on stays in hospital for a heart attack using a French exhaustive dataset. Over the 1998-2003 period, the raw disparities in the propensity to die within 15 days between the extreme regions reaches 80 %. It decreases to 47 % after taking into account the patients' characteristics and their treatments. We conduct a variance analysis to explain regional disparities in mortality. Whereas spatial variations in the use of innovative treatments play the most important role, spatial differences in the local concentration of patients also play a significant role.

Keywords: spatial health disparities, economic geography, stratified duration model

JEL code: I11, C41

1 Introduction

In many countries, spatial disparities between local markets are large and raise some major policy concerns. Whereas the focus of the attention is often the labour market (Duranton and Monastiriotis, 2002; Combes and Overman, 2004; Mion and Naticchioni, 2009), disparities also occur on other markets such as housing or health. This paper develops a new approach to explaining the spatial disparities in healthcare quality.

In the health literature, some studies quantify the international variations in healthcare reimbursment and utilization (Wagstaff and van Doorslaer, 2000) and the interregional variations in health care delivery (Sutton and Lock, 2000). Other papers are interested in the determinants of quality within a given country and exploit the spatial dimension to construct some control variables or instruments. Geweke, Gowrisankaran and Town (2003) study the effect of hospital on mortality and instrument the hospital choice with the distance between the place of residence and hospitals. A growing strand of the literature is interested in the effect of the local healthcare market structure on health outcome. Most authors try to estimate the marginal effect of local competition on health quality (see Gaynor, 2006 for a survey). However, they do not assess how spatial variations in competition can explain spatial disparities in quality.

In fact, evaluating the marginal effect of some factors on mortality and assessing how some spatial variations in these factors can explain spatial disparities in mortality are two related but different exercices. For instance, it is usually found that sex significantly affects mortality. If there was no variation in the share of females across the territory though, the differences in the local sex composition would not explain the disparities in mortality across locations. The same arguments

apply when considering local determinants such as local competition indices. It may happen that local competition has a significant marginal effect on mortality but only small spatial variations. In that case, it does not explain the spatial disparities in mortality on the territory.

In this paper, we conduct a variance analysis of regional disparities in mortality by acute myocardial infarction (AMI) in France. We quantify these disparities and assess the importance of the factors which may explain them. We can distinguish three types of factors according to the literature on health. First, the spatial disparities in mortality can be explained by some differences in the local composition of patients (case-mix) if there is some spatial sorting according to individual attributes related to the propensity to die (such as age or sex). Second, they can be caused by hospital attributes such as ownership status which is usually found to affect hospital performances. McClellan and Staiger (2000) show that within specific markets in the US, the quality of care to the elderly would be better in for-profit hospitals than in not-for-profit hospitals. Milcent (2005) finds that in France, patients in for-profit hospitals have a lower probability of death when having a heart-attack than patients in public hospitals.[1] Hospitals also exhibit some variations in equipment, innovative treatments, physician skills and activities that can be related to differences in health outcome (Tay, 2003). Third, spatial disparities in mortality can come from differences between local healthcare markets. In particular, the local competition measured with a Herfindahl index is often found to have a significant negative impact on mortality (Kessler and McClellan, 2000).

[1] Other references include Hansman (1996), Newhouse (1970), Cutler and Horwitz (1998), Gowrisankaran and Town (1999), Silverman and Skinner (2001), Sloan et al. (2001), Kessler and McClellan (2002), Shortell and Highes (1998), Ho and Hamilton (2000).

We estimate a model at the individual level where the propensity of patients to die during their stay in hospital is specified as a function of the three types of explanatory factors. We then average the model at the regional level and conduct a variance analysis in the spirit of the literature in labour economics (Abowd, Kramarz and Margolis, 1999) and economic geography (Combes, Duranton and Gobillon, 2008). Estimations are conducted on a unique matched patient-hospital dataset from some exhaustive French administrative records over the 1998-2003 period. This original dataset contains some information on the demographic characteristics of patients, their diagnoses and their treatments. It also provides some details on the hospitals where the patients are treated, like the location, the ownership status and the capacity.

More specifically, we use a very flexible econometric specification building on Ridder and Tunali (1999) and Gobillon, Magnac and Selod (2010). We first estimate a Cox duration model stratified by hospital (i.e. each hospital has a specific baseline hazard) using the stratified partial likelihood estimator (SPLE). The individual variables included in the model are the patient characteristics (demographic shifters and secondary diagnoses) and treatments (as they are patient-specific). Their effects are estimated while taking properly into account the hospital unobserved heterogeneity. Our approach also allows to recover some hospital hazard functions without specifying them parametrically. We then go further and specify the hospital hazards as the product of some hospital fixed effects and a baseline hazard. We show how to estimate the hospital fixed effects using some moment conditions. The estimated hospital fixed effects are regressed on some hospital and local variables. We finally average the model at the regional level and make a spatial variance analysis.

We show that regional disparities in mortality are quite large. In particular, the raw difference in the propensity to die within 15 days between the extreme regions reaches 80%. After taking into account the individual variables, this difference drops to 47%. A variance analysis at the regional level shows that regional differences in innovative treatments play a major role in explaining the regional disparities in mortality. A local Herfindahl index computed from the number of patients in each hospital also plays a significant role. Results suggest that spatial differences in the local concentration of patients partly explain spatial differences in mortality.

In a first section, we present the different factors which may explain the spatial differences in healthcare quality in France and review the corresponding literature. A second section describes our dataset. We then present in a third section some descriptive statistics on the regional disparities in mortality, demand factors and supply factors. The fourth section details the econometric methodology used to identify the causes of the regional disparities in mortality. The fifth section summarizes the results of the model.

2 Heart attack in the French context

2.1 The French context

The aim of the paper is to quantify the regional disparities in the mortality of patients hospitalized in France for a heart attack and identify their key determinants. As mentioned earlier, three potential explanations of these disparities are the spatial differences in the composition of patients (case-mix), the composition of hospitals, and the local healthcare market structure (in particular

the intensity of competition between hospitals). Whereas the local composition of patients can be viewed as the local demand for healthcare, the local composition of hospitals and competition are related to the local supply.

On the demand side, spatial differences in demographic shifters and secondary diagnoses possibly related to specific regional behaviours may explain regional disparities in mortality.

2.1.1 Hospitals' characteristics

On the supply side, the local composition of hospitals can affect the local propensity of AMI patients to die if hospitals differ in their efficiency to treat patients and are distributed over space according to their efficiency. We now briefly describe the French system to highlight how hospitals can differ in their efficiency.

The public sector is under a global budget system as well as part of the private sector. Private hospitals which benefit from this budget are Not-For-Profit hospitals (NFP). Every year, the government determines the global budget and chooses how to divide it between regions. The regional budget is shared between NFP and public hospitals according to the budget of the previous year and through bilateral bargaining between the regional regulator and the hospital managers. NFP and public hospitals have to grant access to hospital care to every patient and cannot make any profit. Also, public and NFP hospitals provide similar high-tech care. By contrast, the other hospitals in the private sector (namely For-Profit hospitals) are paid by fee-for-services and can select patients. The selection is usually done to maximize profit, taking into account the health status of patients. FP hospitals have no constraint on profits. Overall, hospitals thus have

different incentives to provide health care to patients depending on their status (public or private) and reimbursement rule (fee-for-service or global budget).

Importantly, there is no segmentation of the healthcare market by insurance status interfering with the effect of the hospital status as in the US. Indeed, people with a managed care insurance in the US must choose an establishment in a given subset of hospitals which may have a specific status. As a consequence, the effect of the hospital status is intertwined with the effect of the insurance status. In France, hospital expenditures are fully reimbursed by a unique public compulsory insurer which funds come from taxes. Patients can freely choose their hospital and there is no segmentation by insurance status. Hence, the true effect of the hospital status can be identified more easily as in the case of Taiwan (Lien, Chou and Liu, 2008).

Milcent (2005) finds for France that ownership significantly affects the mortality rate. Her results suggest that patients in FP hospitals have a lower probability of death but face a greater uncertainty on the quality of care. FP, NFP and public hospitals are distributed unevenly on the French territory, in particular for demographic and historical reasons. Hence, there are some regional disparities in the local composition of hospitals which can yield some regional disparities in mortality of AMI patients treated in hospitals.

Because of the reimbursement rules, hospitals do not have the same incentives to treat patients with innovative procedures. Indeed, FP hospitals are financed via a fee-for-service system. Innovative supplies involving expensive devices that can be used only once such as angioplasty or stent[2] for heart attack are reimbursed ex-post in addition to the fee-for-service payment. By contrast,

[2] See below for a definition of the angioplasty and stent.

the reimbursement of public and NFP hospitals does not depend on the number of procedures which are performed. Therefore, FP hospitals have more incentives to perform innovative procedures than public and NFP hospitals. Milcent (2005) finds for France that innovative treatments decrease the mortality of AMI patients. Spatial differences in the use of innovative treatments (which are related to the spatial sorting of FP hospitals) can cause regional differences in mortality of AMI patients.

The efficiency of hospitals may also be affected by the intensity of hospital activities because of some learning by doing. As a consequence, we will investigate the role played by the spatial disparities in the occupancy rate and the proportion of patients treated for an AMI, in explaining the spatial disparities in mortality. Finally, larger hospitals can bear larger fixed costs related to equipment as these costs can be shared between more patients. Consequently, we will also evaluate whether spatial disparities in the hospital size proxied by the number of beds (in total and in surgery) are related to spatial differences in mortality.

2.1.2 Competition between hospitals

There is a growing body of US literature on the effect of local competition between hospitals on their efficiency (see Gaynor, 2006, for a survey). Whereas some papers investigate situations where prices are set by hospitals, most studies focus on cases where prices are fixed and hospitals can only choose quality of care. In France, prices are regulated in both the public and private sectors. Hence, the competition between hospitals would only be based on quality. Moreover, public and NFP hospitals which are paid under the global budget system are not allowed to make any profit.

As a consequence, these hospitals do not have incentives to compete with each other even by providing a better quality of care contrary to their US counterparts. By contrast, FP hospitals can make some profit and thus have some incentives to attract patients. This can be done by providing some services of better quality than in other FP hospitals, as well as in public and NFP hospitals. The higher the number of private hospitals, the more important is competition based on quality. On the other hand, when patients in an area are scattered across many small hospitals rather than a few large ones, there is not much learning-by-doing in each hospital and the average quality of care in the area could then be lower.

When a patient chooses a hospital where to be treated, he takes into account the accomodation and catering which differ in private and public hospitals. More importantly, he is attracted by the physicians who are the most efficient and can provide the best care. As FP hospitals want to attract patients to make profit, they will try to get the most efficient physicians. This is a specific form of competition based on quality. In fact, the best physicians have some incentives to work in FP hospitals because of the specific payment rule which differs from the one applied to public and NFP hospitals.

In the public sector, the staff (including physicians and nurses) consists of salaried civil servants. Their wages do not depend on their performance. One day a week, though, they can work outside their hospital, in particular in a FP hospital. Physicians working in NFP hospitals are also salaried but their wages are far higher than in public hospitals. In FP hospitals, some physicians are salaried and the others are self-employed. The working time of physicians as well as their wages usually depend on the number of patients. Moreover, physicians receive additional fees when performing

innovative procedures. Overall, their income is far larger than in public hospitals. Consequently, physicians usually compete to get a job in the private sector and only the best of them are selected.[3] Interestingly, this competition has an effect on medical practices in public hospitals. As physicians want to get employed by private hospitals, they perform some innovative procedures to increase their reputation and skills with learning by doing. Dormont and Milcent (2006) show that in public hospitals, the proportion of patients treated with innovative procedures is significant.

Hospitals which ischemic service has grown large are usually those succeeding in attracting patients because of a better reputation. The best physicians have gathered there, can still improve with learning-by-doing, and perform better than in other hospitals which have become smaller. The concentration of patients in a few large hospitals rather than many small ones in an area could then be associated with a better average local quality.

In the US, the local competition among hospitals affects the patients' propensity to die (Gaynor, 2006). Was it the case in France, spatial disparities in the local competition among hospitals could partly explain the regional differences in mortality. In our study, we account for the intensity of competition between hospitals through a Herfindahl index which measures the concentration of patients in a few large hospitals rather than many small ones and is expected to have a positive effect on mortality (Kessler and McClellan, 2000; Town and Vistnes, 2001; Gowrisankaran and Town, 2003). Note however that for France, the local concentration of patients could also reflect some learning-by-doing or the gathering of efficient physicians in the same

[3] There is a large literature on the incentives for physicians depending on the payment's rule (Hart and Holmstrom, 1987; Pauly, 1990; Blomqvist, 1991; Milgrom and Roberts, 1992; Newhouse, 1996; McGuire, 2000).

place. In that case, the effect of the Herfindahl index would rather be negative. Overall, the sign of its effect is theoretically ambiguous and remains an empirical issue.

Our empirical approach will allow us to assess the respective importance of the determinants in explaining the regional disparities in the mortality of AMI patients.

2.2 Treatments of heart attack

In this paper, we focus on one single disease. Indeed, evidence shows that the effect of characteristics on mortality is disease-specific (Wray et al., 1997). We select the Acute Myocardial Infarction (heart attack) for four reasons. First, it belongs to the ischemic-disease group that has been the primary cause of mortality in France, before getting second recently after cancer. Second, mortality from AMI has been widely studied in the literature to assess the quality of hospital care in the US and the UK. This literature can be used for comparison (see Goworisakaran and Town, 2003, for the US, and Propper, Burgess and Green, 2004, for the UK). Third, AMI is a well-defined pathology with only a few re-admissions due to its clinical definition. Fourth, mortality from AMI is an event frequent enough to yield some reliable statistical results.

Heart attacks occur when arteries or veins which irrigate the heart are clogged. In hospitals, patients can benefit from various treatments and procedures to improve the blood flow in clogged arteries. These include bypass surgery, cardiac catheters, percutaneous transluminal coronary angioplasty (PTCA) and stent. A catheter is a thin flexible pipe which is installed in a vein. It may also be used for cleaning arteries in order to improve the blood flow. A bypass surgery reroute, or "bypass", is a vein or artery collected from the patient's body and set up to derive

blood from coronary arteries. In some cases, the stent and the angioplasty are some alternative procedures to the bypass which yield a better quality of life after home return. An angioplasty consists in inflating a balloon in a blockage to create a channel. This procedure is costly as it induces for one stay an increase in costs which ranges from 30% to 60% (Dormont and Milcent, 2002). The stent is a spring-shaped prosthesis which is used as a complement to angioplasty. The use of stent with an angioplasty significantly improves the results. Angioplasties and stents are some innovative treatments over the 1998-2003 period. We will study how the spatial variations in treatments can explain regional differences in mortality.

In this article, the term *stent* refers to an angioplasty together with one or more stents, the term *angioplasty* refers to an angioplasty without stent, and the term *catheterism* refers to a catheterism without angioplasty and stent.

2.3 Spatial features

We now propose a spatial overview of heart attack. First note that AMI patients who want to be treated in a NFP or a public hospital have to go to a hospital within their region of residence. However, some patients are sometimes transferred to a neighbouring hospital in another region. Also, a patient who gets sick in another region may be cured there. Over the 1998-2003 period, the proportion of AMI patients being treated within their region of residence is very high at 92.9%. This proportion is slightly lower for FP hospitals (91.4%) than for public hospitals (93.1%) and NFP (95.8%). These statistics support the fact that regions can be viewed as local healthcare markets for heart attack.

Depending on their residential location, patients do not face the same supply of healthcare, as the local composition of hospitals by status and mode of reimbursment varies widely across space. In 1999, the proportion of beds in public hospitals is large in the west and in Franche Comté (in the east) where it reaches 80%, whereas it is only 46% in the PACA region (the southern *French Riviera*). The proportion of NFP hospitals is the highest in some eastern regions at the German border (Alsace and Lorraine) for historical reasons. Conversely, the proportion of beds in FP hospitals is larger in the south-east (around the *French Riviera* region) where the population is older and richer.

The local proportions of patients treated for an AMI in the different types of hospitals mimic the distribution of bed capacities. For instance, Graph 1 shows that around Paris and in southern regions, the proportion of patients treated in a FP hospital is higher. These regions are often characterized by a substantial use of innovative procedures like stents, as shown in Graph 2. In fact, the rank correlation between the proportion of stents and the proportion of patients in FP hospitals is .61. When considering NFP hospitals instead of FP hospitals, the correlation is still quite high at .44.

[*Insert Graphs* 1 *and* 2]

We also computed the probability of death within 15 days (see Graph 3).[4] This probability is quite low in the Paris region, the east and south-east. It is larger in the west and south-west. There is no obvious relationship between the probability of dying and the proportions of stents or FP hospitals (rank correlations: $-.09$ and .14 respectively). However, south eastern regions which have a large

[4] See below for more details on how this probability is computed.

proportion of FP hospitals performing innovative treatments also concentrate older people who are more likely to die. Hence, it is necessary to perform an econometric analysis to disentangle the effect of age and more generally of individual attributes (demographic characteristics and secondary diagnosis) from that of innovative procedures, hospital characteristics, and local healthcare market structure.

[*Insert Graph* 3]

3 The dataset

3.1 Data sources on patients, hospitals and areas

We use the PMSI dataset (*Programme de Médicalisation des Systèmes d'Information*) which provides the records of all patients discharged from any French acute-care hospital over the 1998-2003 period. It is compulsory for hospitals to provide these records on a yearly basis.[5] Three nice features of this dataset are that it provides some information at the patient level, it keeps track of hospitals across time, and it is exhaustive both for the public and private sectors.[6] A limit of the data is that patients cannot be followed across time if they come back later to the same hospital or if they change hospital.

The dataset contains some information on the demographic characteristics of patients (age and sex), as well as some very detailed information on the diagnoses and treatments. In our

[5] An exception is local hospitals for which it is not compulsory. This does not affect our study since these hospitals do not take care of AMI patients.

[6] It should be mentioned however that only 90% of the private sector was covered in 1998 and 95% in 1999.

analysis, we can thus take into account all secondary coronary diagnoses as well as all techniques used to cure patients. One may argue that some comorbidity factors are not recorded. However, McClellan and Staiger (1999) show that much more detailed medical data on disease severity and comorbidity do not add much when taking into account the heterogeneity among patients. The dataset also provides us with the type of entry (whether the patients come from their place of residence, another care service in the same hospital or another hospital) as well as the type of exit (death, home return, transfer to another hospital or transfer to another care service).

We only keep patients whose pathology was coded as an acute myocardial infarction in the tenth international code of disease (ICD-10-CM). Before 35, heart attacks are often related to a heart disfunction. As a consequence, we restrain our attention to the patients more than 35 following the OMS definition, which leaves us with $421,185$ stays. As we cannot keep track of patients when they are transferred to another hospital, we restrict our sample to patients who come from their place of residence. After deleting observations with missing values for the variables used in our study (that are only very few), we end up with $341,861$ stays for patients distributed across $1,105$ hospitals.

We match our dataset with the hospital records from the SAE survey (*Statistiques Annuelles des Etablissements de santé*) that was conducted every year over the 1998-2003 period. The SAE survey contains some information on the municipality where each hospital is located, the number of beds (in surgery and in total) and the number of days that beds are occupied (in surgery and in total). The matching rate is very good and reaches 97% of the patients.

The municipality code in the SAE survey also allowed us to match our dataset with some

wealth variables at the municipality level coming from other sources. These variables will be used in our estimations to take into account the spatial differences in the funding of public and NFP hospitals. Indeed, local authorities sometimes take into account the local level of poverty when dispatching the budget across hospitals. Our municipality variables include the municipal unemployment rate computed from the 1999 population census, the median household income from the 2000 Income Tax dataset and the existence of a poor area in the municipality (poor areas being defined by a 1997 law under the label *zones urbaines sensibles*). Also, thanks to the municipality code, we could identify the urban area in which hospitals are located.[7] We computed a local index of competition between hospitals within each urban area. This index is a Herfindahl index at the urban area level using the number of patients in hospitals within each urban area.[8] In our analysis, we will also take into account the size of the healthcare market surrounding each patient's hospital as it may affect their efficiency. This size is measured by the number of beds in the urban area, the patient's hospital being excluded. When constructing urban area variables, we were confronted with a few hospitals in municipalities which do not belong to any areas or to

[7] An urban area (*aire urbaine*) is defined as an urban center (which includes more than 5,000 jobs) and the municipalities in its catchment area. There are 359 urban areas in mainland France and they do not cover the whole territory (as some municipalities are excluded and remain rural).

[8] The Herfindahl index for an urban area u is $H_u = \sum_{j \in u} \left(\frac{p_j}{p^u}\right)^2$ where j indices the hospitals, p_j is the number of patients in the hospital j, and $p^u = \sum_{j \in u} p_j$ is the total number of patients within the urban area u. H_u increases from $\frac{1}{n_u}$ to 1 as the concentration of patients increases, where n_u is the number of hospitals in the urban area u. When $H_u = \frac{1}{n_u}$, the patients are equi-distributed between the n_u hospitals. When $H_u = 1$, they are all treated within one hospital.

several of them. We thus introduced some dummies for these two cases as controls. As we will use hospital variables which should be time-invariant in our analysis (see Section 5 and 6), all hospital and geographic variables are averaged across years.

3.2 Preliminary statistics

For each hospital, we computed a gross survival function for exit to death using the Kaplan-Meier estimator. This estimator treats other exits (home return and transfers) as censored. As we are mostly interested in disparities across regions, we computed the average survival function by region.[9] Observations were weighted by the number of patients still at risk in the hospitals.[10] We selected the region with the highest survival function (Alsace), the region with the lowest survival function (Languedoc-Roussillon), and the Paris region (Ile-de-France) that is the most densely populated. Graph 4 represents the survival functions of these three regions as well as their confidence intervals (Graph A1 in appendix represents the survival functions for all the regions and Table A1 ranks the regions according to survival after 15 days). It shows that the two extreme

[9] We could have directly computed a survival function for each region. However, we believe that the relevant unit at which the treatment of patients takes place is the hospital. Also, our approach at the hospital level parallels the model presented in Section 5.

[10] When the length of stay increases, the number of patients in a given hospital decreases. Above a given length of stay, there is no patient at risk anymore and the hospital is not taken into account in the computation of the survival function. Hence, there is a selection of hospitals as the length of stay increases. We limited our analysis to lengths of stays below fifteen days to minimize the effect of our assumption.

average survival functions are significantly different.

[*Insert Graph* 4]

Table 1 reports some disparity indices between regions in the probability of death within 1, 5, 10 and 15 days (defined as one minus the Kaplan-Meier). These indices are the max/min ratio, the Gini index and the coefficient of variation. The Gini indices and coefficients of variation are computed in two stages. First, we compute the average of a given individual variable (for instance, a death dummy) by region. Then, we compute the regional disparity indices for the resulting variable (in our example, the regional proportion of deaths), weighting the observations by the number of patients in the region. Global indices like the Gini index (.07) and the coefficient of variation (.218) remain quite small and suggest that disparities are not systematic. The max/min ratio shows that regional disparities are significant. Indeed, the difference in the probability of death within 15 days between the Maximum (Languedoc-Roussillon) and the Minimum (Alsace) is 80%. Interestingly, disparities are a bit larger for the probability of death within 1 day (Max/Min ratio of 94%). This may be due to different behaviours across regions in transfers and home returns in the early days of AMI stays.

As mentioned earlier, the regional disparities in mortality may be explained with some regional disparities in demand factors (demographic shifters and secondary diagnosis) or in supply factors, whether they are related to hospitals (treatments and establishment characteristics) or to the local healthcare market structure. To disentangle these three types of effects, we present Gini indices which are some global measures of disparities and are not sensitive to the level of magnitude as

the max/min ratio.[11] We consider in the sequel that disparities are small when the index is inferior to .1, they are moderate for an index from .1 to .2, they are large for an index from .2 to .3, and they are very large for an index above .3.

We first consider variables related to patients which were averaged at the regional level. There are significant disparities across regions for some demographic variables: the Gini index is moderate for females aged $35-55$ (.12) and males who are more than 85 (.11). For diagnoses, the Gini index reaches .23 for surgical French DRGs (.23), .15 for the severity index[12] and .13 for a history of vascular diseases and for stroke. Note that the Gini index is most often moderate for diagnoses related to specific behaviours before the heart attack such as obesity (.17), excessive smoking (.16), and alcohol problems (.14). Regional disparities in the use of procedures are important. The Gini index goes up to .53 for dilatations other than PTCA and .37 for the cabbage or coronary artery bypass surgery. More widespread procedures like angioplasty and stent still have a large Gini index which takes the value .28 and .21, respectively.

Overall, the Gini indices show that potential explanations of disparities in the propensity to die can be related to demographic characteristics, diagnoses and procedures. One should keep in mind though that the explanatory power of a given variable when studying the regional disparities in mortality is closely related to its variance and its effect on death. Considering the variance, the

[11] Alternatively, we could also comment the results obtained with the coefficients of variation which are similar.

[12] We use Deyo's adaptation of the Charlson co-morbidity index to measure the severity of co-morbidities (Deyo, 1992; Ghali, 1996). The Charlson index, which is expressed as a six-level variable, is constructed for each stay. This index is greater than 0 when a surgical procedure has been carried out on the patient. Validation exercises have shown that this index predicts well mortality in longitudinal data (Hamilton and Hamilton, 1997).

different types of patient-specific variables still look like good candidates (even if their respective importance is different).

[*Insert Table* 1]

We then computed regional disparity indices for the hospital and geographic variables used in our regressions. Whereas hospital variables measure capacities and status (public, NFP and FP), geographic variables are mostly meant to capture the effects related to the structure of the local healthcare market (like competition). For a given variable, we constructed its regional average, weighting the observations by the number of patients in the hospitals. The resulting regional average is then used to compute disparity indices at the regional level. Results are reported in Table 2. As previously, we only comment Gini indices.

There are large disparities across regions in the size of hospitals measured by the total number of patients (.23) or the number of AMI patients (.27). Disparities are even larger for the number of beds (.49) and the number of beds in surgery (.47). These disparities point out some sorting of hospitals according to their size. Finally, disparities are smaller but still large for the hospital status and more specifically for being a FP hospital (.24). The regional disparities in hospital characteristics may thus play a role when trying to explain the regional disparities in mortality.

Concerning geographic variables, we observe some very large disparities in the number of beds in the urban area (Gini index .66) which is not surprising as there is a lot of variation in the population of regions. Disparities are also significant for the Herfindhal index computed at the urban area level (.20). Indeed, hospitals are unevenly distributed in the territory, for historical reasons, public policy and consequences of competition. This creates some regional differences in

average Herfindahl Index. Regional disparities in municipality variables capturing some geographic heterogeneity in wealth are at best moderate, the Gini index reaching .17 for the presence of a poor area in the municipality.

[*Insert Table* 2]

In summary, demand and supply factors are all some potential candidates to explain the regional disparities in mortality. We now present our empirical methodology to assess their respective explanatory power.

4 Econometric method

We first give a brief description of the econometric model explaining the propensity to die before turning to a more formal presentation of our approach. We build our specification around hospital units to properly take into account their heterogeneity and use a Cox duration model at the patient level stratified by hospital. Hence, each hospital has its own hazard function which captures its specific behaviour. Ridder and Tunali (1999) explain how to estimate this model using the stratified partial likelihood estimator (SPLE) and establish the theoretical properties of the estimators. Their methodology has been used in some studies related to education and unemployment, but not in health economics. Lindeboom and Kerkhofs (2000) apply their methodology to quantify the effect of school on the job sickness of teachers and Gobillon, Magnac and Selod (2010) use it to analyze the effect of location on finding a job in the Paris region.

The model contains some patient-specific explanatory variables (demographic shifters, diagnoses and treatments), as well as a specific survival function for each hospital which is left unspecified.

The flexible modelling of the hospital heterogeneity allows us to recover some robust estimators of the coefficients of patient-specific explanatory variables. These coefficients are then used in the estimation of the hospital survival functions, which are in turn averaged at the regional level to study the regional disparities in mortality net of the effect of patient-specific variables.

We then link the remaining regional disparities to some local differences in hospital and geographic characteristics. For that purpose, we make the additional assumption that the hospital hazards write multiplicatively as the product of a hospital fixed effect and a baseline hazard. We show how to estimate the hospital fixed effects using moment conditions. We explain them with hospital and geographic variables and finally average the model at the regional level to perform a regional variance analysis.[13]

We now present our approach more formally. For each patient, we observe the length of stay in the hospital and the type of exit (death, home return or transfer). In the sequel, we only study exit to death. All other exits are treated as censored. We specify the hazard function of a patient i in a hospital $j(i)$ as:

$$\lambda\left(t\,|X_i, j\left(i\right)\right) = \theta_{j(i)}\left(t\right)\exp\left(X_i\beta\right) \tag{1}$$

where $\theta_j(\cdot)$ is the instantaneous hazard function for hospital j, X_i are the patient-specific explanatory variables and β are their effect on death. The model is estimated maximizing the stratified partial likelihood. The contribution to likelihood of a patient i who dies after a duration t_i is his

[13] A tempting alternative approach is to estimate all the coefficients in one stage introducing all the patient, hospital and geographic variables in a simple Cox model. However, such an approach does not take into account the hospital unobserved heterogeneity. Consequently, standard errors of the coefficients may be highly biased (see Moulton, 1990). Our approach properly takes into account the hospital unobserved heterogeneity.

probability of dying conditionally on someone at risk in his hospital dying after this duration. It writes:

$$P_i = \frac{\exp\left(X_i\beta\right)}{\sum\limits_{i\in\Omega_{j(i)}(t_i)} \exp\left(X_i\beta\right)} \tag{2}$$

where $\Omega_j\left(t\right)$ is the set of patients at risk at day t in hospital j, i.e. the set of patients that are still in hospital j after staying there for t days. The partial likelihood to be maximized then writes: $L = \prod\limits_i P_i$. Denote $\widehat{\beta}$ the estimated coefficients of patient-specific explanatory variables. It is possible to compute the integrated hazard function $\Theta_j\left(t\right)$ of any hospital j using the estimator proposed by Breslow (1974). It writes:

$$\widehat{\Theta}_j\left(t\right) = \int\limits_0^t \frac{I\left(N_j\left(s\right) > 0\right)}{\sum\limits_{i\in\Omega_j(s)} \exp\left(X_i\widehat{\beta}\right)} dN_j\left(s\right) \tag{3}$$

where $I\left(\cdot\right)$ is the indicator function, $N_j\left(s\right) = card\ \Omega_j\left(s\right)$, and $dN_j\left(s\right)$ is the number of patients exiting from hospital j between the days s and $s+1$. From the Breslow's estimator, we compute a survival function for each hospital j as $\exp(-\widehat{\Theta}_j\left(t\right))$ (an estimator of its standard error is recovered using the delta method). The hospital survival functions will be averaged at the regional level to study regional disparities in mortality after any number of days. As the hospital hazards are left completely unspecified, the study of regional disparities in death using regional averages remains very general.

We then study the determinants of hospital disparities by specifying the hospital hazard rates in a multiplicative way:

$$\theta_j\left(t\right) = \alpha_j\theta\left(t\right) \tag{4}$$

where α_j is a hospital fixed effect and $\theta\left(t\right)$ is a baseline hazard common to all hospitals. We show

in appendix how to estimate the parameters using empirical moments derived from (4).[14] Note that we need an identifying restriction since α_j and $\theta(t)$ can be identified separately only up to a multiplicative constant. We impose for convenience that: $\frac{1}{N}\sum_t N_t\theta(t) = 1$ where N_t is the number of patients still at risk at the beginning of day t and $N = \sum_t N_t$. After some calculations (see appendix), we get:

$$\theta(t) = \left(\frac{1}{N^2}\sum_{j,t} N^j N_t \theta_j(t)\right)^{-1}\left(\frac{1}{N}\sum_j N^j \theta_j(t)\right) \tag{5}$$

$$\alpha_j = \left(\frac{1}{N^j}\sum_t N_{jt}\theta(t)\right)^{-1}\left(\frac{1}{N^j}\sum_t N_{jt}\theta_j(t)\right) \tag{6}$$

where N_{jt} is the number of patients at risk at time t in hospital j, $N^j = \sum_t N_{jt}$, and the sum on t, $\sum_t$, goes from $t = 1$ to $t = T$ (here, we fixed $T = 30$ for convenience). An estimator of $\theta(t)$

[14] In doing so, we depart from the log-linear estimation method proposed by Gobillon, Magnac and Selod (2010). Our approach is more adequate when exits are scarce as in our case. Indeed, Gobillon, Magnac and Selod split the timeline into K intervals denoted $[t_{k-1}, t_k]$. Introduce $\theta_k = \int_{t_{k-1}}^{t_k} \theta(t)\,dt/(t_k - t_{k-1})$ and $y_{jk} = [\Theta_j(t_k) - \Theta_j(t_{k-1})]/(t_k - t_{k-1})$. Integrating (4) over each interval and taking the log, they get: $\ln y_{jk} = \ln\alpha_j + \ln\theta_k$. y_{jk} is not observed but can be replaced by a consistent estimator: $\widehat{y}_{jk} = \left[\widehat{\Theta}_j(t_k) - \widehat{\Theta}_j(t_{k-1})\right]/d_k^j$ where d_k^j is the amount of time in interval $[t_{k-1}, t_k]$ where at least one patient is at risk. The equation to estimate is then: $\ln\widehat{y}_{jk} = \ln\alpha_j + \ln\theta_k + \psi_{jk}$ where $\psi_{jk} = \ln\widehat{y}_{jk} - \ln y_{jk}$ is the sampling error. This equation can be estimated with standard linear panel methods. The authors use weighted least square where the weights are the number of individuals at risk at the beginning of the interval. A limit of this method is that $\ln y_{jk}$ can be replaced by its estimator $\ln\widehat{y}_{jk}$ only if $\widehat{y}_{jk} \neq 0$. When it is not the case, observations should be discarded from the sample. When implementing this approach in our case, this could be an issue as exits are scarce and a significant number of observations should be discarded when the time spent in the hospitals gets long. In practice however, the results obtained with the two approaches are quite similar.

denoted $\widehat{\theta}(t)$ can be obtained, replacing $\theta_j(t)$ by the estimator $\widehat{\theta}_j(t) = \widehat{\Theta}_j(t) - \widehat{\Theta}_j(t-1)$ on the right-hand side of equation (5). An estimator of α_j denoted $\widehat{\alpha}_j$ can then be derived, replacing $\theta_j(t)$ and $\theta(t)$, respectively by $\widehat{\theta}_j(t)$and $\widehat{\theta}(t)$, on the right-hand side of equation (6). We show in appendix how to compute the covariance matrices of $\widehat{\theta} = \left(\widehat{\theta}(1), ..., \widehat{\theta}(T)\right)'$ and $\widehat{\alpha} = (\widehat{\alpha}_1, ..., \widehat{\alpha}_J)'$. We then explain the hospital fixed effects with some hospital and geographic variables denoted Z_j. We specify: $\alpha_j = \exp\left(Z_j\gamma + \eta_j\right)$ where γ are the effects of hospital and geographic variables on death, and η_j includes some unobserved hospital and geographic effects. For a given hospital j, taking the log and replacing the hospital fixed effect with its estimator, we get:

$$\ln \widehat{\alpha}_j = Z_j\gamma + \eta_j + \phi_j \tag{7}$$

where $\phi_j = \ln \widehat{\alpha}_j - \ln \alpha_j$ is the sampling error on the hospital fixed effect. Equation (7) can be estimated using weighted least squares where the weight is the number of patients in the hospital. The standard errors and R-square (adjusted to take into account the sampling error), are computed as proposed by Gobillon, Magnac and Selod (2010). Note that for a given hospital, equation (7) is well defined only when there is at least one patient dying in the hospital over the $1998 - 2003$ period (otherwise the quantity $\widehat{\alpha}_j$ from which we take the log would be zero). This condition may not be verified for hospitals that have only a few patients. In fact, these hospitals have a negligible weight and they are discarded from our sample. We finally average equation (7) at the regional level and conduct a variance analysis for the resulting equation.

5 Results

Table 3 reports the estimation results of the first-stage equation (1). The demographic characteristics have the usual effect on the propensity to die. Females are more likely to die than males. This is consistent with care being more protective for males than for females possibly because of biological differences like the smaller target vessel size and the more important vessel tortuosity of females (Milcent et *al.*, 2007). Also, the propensity to die increases with age.

Among variables related to the diagnosis, the severity index is found to have a positive effect on the propensity to die. Intuitively, one also expects secondary diagnoses to have a positive effect as they deteriorate health. This is true empirically for renal failure, stroke, heart failure, conduction disease, alcohol. Other secondary diagnoses have a more surprising negative effect: diabetes, obesity, excessive smoking, vascular disease, peripheral arterial disease, previous coronary artery disease, and hypertension. These results may be explained by preventive health care. Indeed, these secondary diagnoses may point at patients who are monitored more carefully before and after having a heart attack (Milcent, 2005).

All treatments have the expected negative effect on the propensity to die: CABG, catheterism, PTCA, other dilatation and stent. The stent, which is the most innovative procedure, has the strongest negative effect. After taking into account these treatments, the DRG index capturing the heaviness of surgical procedures has a positive effect on the propensity to die. This can reflect the increased chances of dying because of more cumbersome and risky surgery.

[*Insert Table* 3]

From the estimated coefficients $\widehat{\beta}$, we constructed an integrated hazard for each hospital using Breslow's estimator and averaged the corresponding hospital survival functions by region (weighting by the number of patients at risk in the hospitals).[15] Regions at extremes are the same as when studying the raw data: Alsace (at the German border) usually exhibits the highest survival function and Languedoc-Roussillon (in the South-East) the lowest. Graph 5 represents the survival functions (as well as their confidence intervals) for these two extremes and for Ile-de-France (The Paris region).[16] The difference between the extreme regions is smaller but still significant.

[*Insert Graph* 5]

We quantify the regional disparities computing the same disparity indices as for raw data for the probability of death within 1, 5, 10 and 15 days (defined as one minus the survival function of the model). Results reported in Table 4 show that the difference in the probability of death within

[15] This kind of aggregation is quite common in the labour literature. For instance, Abowd, Kramarz and Margolis (1999) estimate a wage equation that includes some firm fixed effects. They then compute some industry fixed effects as the averages of the estimated firm fixed effects for firms belonging to each industry (weighting the observations by the number of workers in the firms).

[16] Graph A3 in appendix represents the survival functions for all the regions and Table A2 ranks the regions according to survival after 15 days. Curves obtained with the model are not strictly comparable with those obtained from raw data with the Kaplan-Meier estimator as instantaneous hospital hazards were normalized with an *ad-hoc* rule. To get close to comparability, we multiplied instantaneous hospital hazards by a constant which was chosen in such a way that in absence of hospital heterogeneity (i.e. $\theta_j(t) = \theta(t)$ for all t), the expected integrated hazard at day 1 is equal to the expected integrated hazard obtained from the raw data (defined as minus the logarithm of the Kaplan-Meier estimator). This normalization allows to obtain an average survival function of the same magnitude as the one obtained from raw data with the Kaplan-Meier estimator.

15 days between the extreme regions has decreased from 80% to 47% (this corresponds to a 41% decrease). More systematic disparity indices like the coefficient of variation and the Gini index also decrease, but to a lesser extent (by 19% and 17%, respectively). In a variance-analysis spirit, we defined a pseudo-R^2 as one minus the ratio between the variance in the probability of death within a given number of days computed from the model and the variance computed from raw data. At 1 and 5 days, the pseudo-R^2 is nearly 60%. Hence, patients' characteristics and treatments would explain more than half of the regional disparities in early death. However, it is lower at 10 days (48%), and decreases even more to reach 40% at 15 days. These results suggest that part of the early regional disparities may be due to different timings of death events across regions. Also, there may be some specific regional behaviour for transfers and home returns which would affect the local composition of patients and hence would have an impact on the difference between the hospital survival functions obtained from the model and from the Kaplan-Meier estimators. Interestingly, the ranking of regions obtained for death within 15 days is not that different from the one obtained from the raw data (unweighted rank correlation: .70). This means that the ranking of regions does not change much after taking into account individual variables.

[*Insert Table* 4]

We then supposed that each instantaneous hospital hazard can be written multiplicatively as the product of a hospital fixed effect and a baseline hazard. The parameters of the multiplicative model are estimated using empirical moments as explained in the previous section. Graph 6 displays the baseline hazard and the confidence interval at each day. Remember that the weighted average of the instantaneous baseline hazards is normalized to zero. We obtain that the baseline hazard

decreases sharply in the first two days and then more smoothly until the eighth day. It remains constant afterward. The sharp decrease just after entry in the hospital can be explained by violent deaths that are quite common in early days of heart attacks.

[*Insert Graph* 6]

We then regress the hospital fixed effects on a set of hospital and geographic variables. Results are reported in Table 5 (estimated regional dummies corresponding to the specification of Column 3 are reported in Appendix A3). When we only introduce hospital variables (Column 1), the adjusted-R^2 is quite low at .13.[17] It is larger at .23 when only geographic variables enter the specification (Column 2). Interestingly, when introducing both groups of variables (Column 3), the R^2 at .28 is below the sum of R^2 of the two separate regressions (.36), which suggests that variables are quite correlated. Also, it is higher than the R^2 of each separate regression, which suggests that each of the two groups has some explanatory power of its own.

We now comment on the sign of the estimated coefficients for the full specification (Column 3).

As regards the effect of hospital characteristics, we find that the propensity to die is nearly the same in FP hospitals and public hospitals. This result may look surprising but it comes from the fact that we take into account innovative treatments (mainly angioplasty and stent). If we drop the variables related to innovative treatments from the first-stage specification, the propensity to die in public hospitals becomes higher than in FP hospitals (see Table A3 in appendix). Hence, the higher efficiency of FP hospitals would come from a wider use of innovative treatments. We also find that the propensity to die in a NFP hospital is lower than in a public or an FP hospital.

[17] The adjusted-R^2 takes into account the sampling error.

The proportion of patients in the hospital treated for an AMI has a negative and significant effect. It is possible that hospitals concentrating AMI patients have specialized in heart-related pathologies and thus have a higher efficiency. The number of beds as well as their occupation rate has no effect on mortality. The propensity to die is lower in hospitals with a higher proportion of beds in surgery (whether taking into account innovative treatments or not). In fact, hospitals with a high concentration of beds in surgery could have specialized in serious diseases and have a higher-quality staff. The propensity to die also decreases with the occupation rate of beds in surgery (significantly at 10% only). It is possible that hospitals with a higher occupation rate are more efficient and more likely to attract patients.

The Herfindahl index which measures the local concentration of patients has a significant negative effect. This result suggests that when patients in an area are distributed across a few large hospitals rather than many small ones, the mortality in that area tends to be lower. The number of beds in the urban area has a positive significant effect which turns out to be negative but not significant when innovative treatments are not taken into account. An interpretation can be that larger markets propose more innovative treatments but would also lead to some inefficiency in healing patients. These effects would compensate but after taking into account the innovative treatments, only the net inefficiency effect would remain. The municipality variables do not have much effect. The presence of a poor area in the municipality has a positive effect on mortality, but it is significant only at the 10% level.

At last, regional dummies always have a negative effect compared to the reference (Languedoc-Roussillon) and their effect is most often significant. Differences may be explained by unobserved

regional factors such as the regional differences in hospital budgets and in the propensity to transfer patients when they are likely to die. Note that standard errors are quite large and two regions need to be far enough in the distribution of regional effects for the difference between their effects to be significant. The ranking of regional effects is only weakly correlated with the probability of death within 15 days obtained from raw data (unweighted rank correlation: .20) and with that obtained from the model (unweighted rank correlation: $-.11$).

[*Insert Table* 5]

We now study the variations in mortality at the regional level. Taking the logarithm of equation (1) with the multiplicative assumption (4), and computing the average for any region r gives:

$$\frac{1}{N^r} \sum_{i|j(i)\in r} \ln \lambda \left(t \left| X_i, j\left(i\right)\right.\right) = \overline{X}^r \beta + \overline{\ln \alpha}^r + \theta\left(t\right) \tag{8}$$

where N^r is the number of patients in region r, $\overline{X}^r$ is the regional average of individual explanatory variables and $\overline{\ln \alpha}^r$ is the regional average of hospital fixed effects weighted by the number of patients in the hospitals. This equation states how at the regional level, the average hazard at t days for patients entering an hospital for an AMI relates to their average characteristics, the average hospital effects, and the baseline hazard at t days. We qualitatively assess the relative explanatory power of right-hand side terms in (8) computing their variance and their correlation with the left-hand side term (in a way similar to Abowd, Kramarz and Margolis, 1999). In fact, the larger the variance and the correlation, the higher the explanatory power. In practice, as β and $\overline{\ln \alpha}^r$ are not observed, we use their estimators $\widehat{\beta}$ and $\widehat{\overline{\ln \alpha}}^r$ (the latter being defined as the regional weighted average of $\widehat{\ln \alpha_j}$) to compute the right-hand side terms. An estimator of the left-hand side

term is obtained from the sum of right-hand side terms. Using the same approach, we also assess the explanatory power of $\overline{X}_s^r\widehat{\beta}$ for some sub-groups $\overline{X}_s^r$ of explanatory variables. Importantly, note that this procedure measures the explanatory power *ex ante* before any filtering process of patients through transfers or home returns. We can further assess the explanatory power of hospital and geographic variables. Taking the log of the expression of hospital fixed effects and averaging at the regional level, we get:

$$\overline{\ln \alpha}^r = \overline{Z}^r \gamma + \overline{\eta}^r$$

where $\overline{Z}^r$ and $\overline{\eta}^r$ are the regional averages of explanatory variables and unobserved terms, respectively. We can assess the explanatory power of $\overline{Z}^r\gamma$ and $\overline{Z}_s^r\gamma$, for some sub-groups $\overline{Z}_s^r$ of explanatory variables, in the same way as for individual variables (replacing γ by its estimator).

We find that individual variables have a far larger power than hospital effects in explaining regional disparities in mortality (see Table 6a). Indeed, their variance is five to six times larger. Interestingly, among the individual variables, it is the innovative treatments which have the largest explanatory power. This means that regional disparities in innovative treatments are a key factor in explaining regional disparities in mortality. This has some important consequences for the regional funding of innovative equipment. Of course, the regional composition in age and sex also plays a role. Note that the sum of variances for groups of individual variables is far smaller than their sum. This comes from fairly large correlations between groups. In particular, regions where patients are aged and mostly females are also those in which more innovative treatments are performed (correlation between the demographic effects and the effect of innovative treatments:

.57).

[*Insert Table* 6*a*]

The hospital and geographic effects have a larger variance than the demographic composition effects, which suggests that their role in explaining regional disparities is significant. Concerning regional disparities in hospital fixed effects, the local composition by ownership status does not have a noticeable explanatory power (Table 6b). As regards geographic variables, the local size of the surrounding market (measured by the local number of beds except those in the patient's hospital) and the Herfindahl index play a significant role.[18] At last, residual local effects captured by regional dummies have a large variance. This means that some unobserved regional factors have a large effect on regional disparities in mortality.

[*Insert Table* 6*b*]

6 Conclusion

In this paper, we studied the regional disparities in mortality for patients admitted in hospitals for a heart attack. This was done using a unique matched patients-hospitals dataset over the 1998-2003 period constructed from exhaustive administrative records. For patients, this dataset contains some information on demographic characteristics (sex and age), diagnoses and treatments. For

[18] Note that the local size of the surrounding market and the local concentration of patients have an effect that is positively correlated with hospital fixed effects. However, their correlation with the overall integrated hazard (last column in Table 6b) is negative. This is because these effects are more than compensated by regional fixed effects and the effects of innovative treatments.

hospitals, it gives some details on the location, status, rules of reimbursement and capacity.

We showed that regional disparities are fairly large. The difference in mortality rate between the extreme regions reaches 80%. We analyzed the causes of these disparities using a Cox duration model stratified by hospital. The model contains some patient-specific explanatory variables (demographic shifters, diagnoses and treatments), as well as a specific survival function for each hospital which is left unspecified. The flexible modelling of the hospital heterogeneity allows us to recover some robust estimators of the coefficients of patient-specific explanatory variables. These coefficients are then used in the estimation of the hospital survival functions which capture the differences in hospital behaviours when treating patients. Hospital survival functions are in turn averaged at the regional level to study the regional disparities in mortality net of the effect of patient-specific variables. Regional disparities are then lower but remain significant: the difference in mortality rate between the extreme regions is still 47%. Interestingly, the extent to which patients are treated with innovative procedures at the regional level plays a major role in the decrease of the disparities.

We then assessed to what extent the remaining regional disparities could be explained with spatial differences in hospitals' characteristics and local healthcare market structure. This was done regressing hospital survival functions on hospital and geographic variables, and averaging the model at the regional level. We found that once treatments have been taken into account, the status of hospitals does not play much. By contrast, the local concentration of patients plays a significant role. When patients in an area are distributed across a few large hospitals rather than many small ones, the mortality in that area tends to be lower. After hospital and geographic variables have

been taken into account, some significant regional disparities still remain.

A limit of our analysis is that patients were not tracked in the data when they were transferred to another hospital. For patients who were transferred, we had to consider that the length of stay was censored. An interesting extension of our work would be to study how hospitals interact through transfers and to what extent the transfer of patients to another hospital affects their propensity to survive. Space may play a major role in transfers as some hospitals are isolated and others are close to an establishment specialized in heart surgery.

7 Appendix: second-stage estimation

In this appendix, we explain how to construct some estimators of the baseline hazard and hospital fixed effects. We first average equation (4) across time, weighting the observations by the number of patients at risk at each date. We obtain:

$$\frac{1}{N}\sum_{t} N_t \theta_j(t) = \alpha_j \frac{1}{N}\sum_{t} N_t \theta(t)$$

where N_t is the number of patients at risk at the beginning of period t, $N = \sum_t N_t$ with $\sum_t$ the sum from 1 to T days (with $T = 30$ in the application). A natural identifying restriction is that the average of instantaneous hazards equals one: $\frac{1}{N}\sum_t N_t \theta(t) = 1$. We obtain:

$$\alpha_j = \frac{1}{N}\sum_{t} N_t \theta_j(t) \tag{9}$$

It could be possible to construct an estimator of hospital fixed effects from this formula, but weights (namely: N_t) are not hospital-specific and thus do not reflect hospital specificities. Hence,

we propose another estimator of hospital fixed effects in the sequel which we believe better capture hospital specificities.

We also average equation (4) across hospitals, weighting by the number of patients at risk (summed across all dates) in each hospital. We get:

$$\frac{1}{N}\sum_{j} N^{j}\theta_{j}(t) = \frac{1}{N}\left(\sum_{j} N^{j}\alpha_{j}\right)\theta(t)$$

where $N^{j} = \sum_{t} N_{jt}$ with N_{jt} the number of patients at risk in hospital j at the beginning of date t (such that $N = \sum_{j} N^{j}$). Replacing α_j with its expression (9), we obtain: $\theta(t) = \left(\frac{1}{N^2}\sum_{j,t} N^{j}N_{t}\theta_{j}(t)\right)^{-1}\left(\frac{1}{N}\sum_{j} N^{j}\theta_{j}(t)\right)$. An estimator of the hazard rate at date t in hospital j can be constructed from Breslow's estimator such that $\widehat{\theta}_{j}(t) = \widehat{\Theta}_{j}(t) - \widehat{\Theta}_{j}(t-1)$. A natural estimator of the baseline hazard is then:

$$\widehat{\theta}(t) = \left(\frac{1}{N^2}\sum_{j,t} N^{j}N_{t}\widehat{\theta}_{j}(t)\right)^{-1}\left(\frac{1}{N}\sum_{j} N^{j}\widehat{\theta}_{j}(t)\right)$$

We then construct an estimator of a given hospital fixed effect α_j averaging equation (4) across time for this hospital and weighting by the number of patients at risk at the beginning of each day in this hospital. We obtain:

$$\frac{1}{N^{j}}\sum_{t} N_{jt}\theta_{j}(t) = \alpha_{j}\frac{1}{N^{j}}\sum_{t} N_{jt}\theta(t)$$

An estimator of the hospital fixed effect is then:

$$\widehat{\alpha}_{j} = \left(\frac{1}{N^{j}}\sum_{t} N_{jt}\widehat{\theta}(t)\right)^{-1}\left(\frac{1}{N^{j}}\sum_{t} N_{jt}\widehat{\theta}_{j}(t)\right) \tag{10}$$

We also computed the asymptotic variances of $\widehat{\theta} = \left(\widehat{\theta}(1), ..., \widehat{\theta}(T)\right)'$ and $\widehat{\alpha} = (\widehat{\alpha}_1, ..., \widehat{\alpha}_J)'$, denoted V_θ et V_α, with the delta method. Indeed, the covariance matrix of $\widehat{\theta}_J = \left(\widehat{\theta}_1(1), ..., \widehat{\theta}_J(T)\right)'$ can

be estimated from Ridder et Tunali (1999). Its estimator is noted $\widehat{V}_{\theta_J}$. We can then compute the estimators: $\widehat{V}_\theta = \left(\frac{\partial\widehat{\theta}}{\partial\widehat{\theta}'_J}\right)\widehat{V}_{\theta_J}\left(\frac{\partial\widehat{\theta}'}{\partial\widehat{\theta}_J}\right)$ and $\widehat{V}_\alpha = \left(\frac{\partial\widehat{\alpha}}{\partial\widehat{\theta}'_J}\right)\widehat{V}_{\theta_J}\left(\frac{\partial\widehat{\alpha}'}{\partial\widehat{\theta}_J}\right)$. The vectors $\frac{\partial\widehat{\theta}}{\partial\widehat{\theta}_J}$ and $\frac{\partial\widehat{\alpha}}{\partial\widehat{\theta}_J}$ are given by:

$$\frac{\partial\widehat{\theta}(t)}{\partial\widehat{\theta}_k(\tau)} = \frac{NN^k}{\sum\limits_{j,t} N^j N_t \widehat{\theta}_j(t)} 1_{\{t=\tau\}} - \frac{NN^k N_\tau}{\left[\sum\limits_{j,t} N^j N_t \widehat{\theta}_j(t)\right]^2} \sum_j N^j \widehat{\theta}(t) \tag{11}$$

$$\frac{\partial\widehat{\alpha}_j}{\partial\widehat{\theta}_k(\tau)} = \frac{N_{k\tau}}{\sum\limits_t N_{j,t}\widehat{\theta}(t)} 1_{\{k=j\}} - \widehat{\alpha}_j \frac{\sum\limits_t N_{j,t}\frac{\partial\widehat{\theta}(t)}{\partial\widehat{\theta}_k(\tau)}}{\sum\limits_t N_{j,t}\widehat{\theta}(t)} \tag{12}$$

In practice, to simplify the computations, we neglected the second term on the right-hand side of (12). This is only a slight approximation that does not have much impact on the estimated variance of $\widehat{\alpha}_j$. It amounts to neglect in (10) the variations of $\frac{1}{N^j}\sum\limits_t N_{jt}\widehat{\theta}(t)$ with respect to the terms $\widehat{\theta}_j(t)$ compared to the variations of $\frac{1}{N^j}\sum\limits_t N_{jt}\widehat{\theta}_j(t)$. Put differently, $\widehat{\theta}(t)$ is supposed to be non-random in (10).

References

[1] Abowd M., Kramarz F. and D. N. Margolis (1999), "High wage workers and high wage firms", *Econometrica*, 67(2), pp. 251–333.

[2] Blomqvist A. (1991), "The doctor as double agent: Information asymmetry, health insurance, and medical care", *Journal of Health Economics*, 10(4), pp. 411-432.

[3] Breslow N.E. (1974), "Covariance Analysis of Censored Survival Data", *Biometrics*, 30, pp. 89-99.

[4] Combes PPh., Duranton G. and L. Gobillon (2008), "Spatial Wage Disparities: Sorting Matters", *Journal of Urban Economics*, 63(2), pp. 723-742.

[5] Combes PPh. and H. Overman (2004), "The Spatial Distribution of Economic Activities in the European Union", in *Handbook of Urban and Regional Economics, vol. 4*, pp. 2845-2910, Henderson V. and J.F. Thisse (eds.), Elsevier-North Holland, Amsterdam.

[6] Cutler D.M. and J.R. Horwitz (1998), "Converting Hospitals from Not-for-Profit to For-Profit States: Why and What Effects?", NBER Working Paper 6672.

[7] Deyo R., Cherkin D. and M. Ciol (1992), "Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases", *Journal of Clinical Epidemiology*, 45, pp. 613-619.

[8] Dormont B. and C. Milcent (2002), "Quelle régulation pour les hôpitaux publics français?", *Revue d'Economie Politique*, 17(2), pp. 117-142.

[9] Dormont B. and C. Milcent (2006), "Innovation diffusion under budget constraint", *Annales d'Economie et de Statistique*, Forthcoming.

[10] Duranton G. and V. Monastiriotis (2002),"Mind the Gaps: The Evolution of Regional Inqualities in the U.K. 1982-1987", *Journal of Regional Science*, 42(2), pp. 219-265.

[11] Gaynor M. (2006), "What do we know about competition and quality in health care markets?", NBER Working Paper 12301.

[12] Geweke J., Gowrisankaran G. and R. Town (2003), "Bayesian Inference For Hospital Quality in a Selection Model", *Econometrica*, 71, pp. 1215-1238.

[13] Ghali W., Hall R., Rosen A., Ash, A. and M. Moskowitz (1996), "Searching for an improved clinical comorbidity index for use with ICD-9-CM administrative data", *Journal of Clinical Epidemiology*, 49, pp. 273-278.

[14] Gobillon L., Magnac T. and H. Selod (2010), "The effect of location on finding a job in the Paris region", *Journal of Applied Econometrics*, forthcoming.

[15] Gowrisankaran G. and R. Town (1999), "Estimating quality of Care in Hospitals Using Instrumental Variables", *Journal of Health Economics*, 18, pp. 747-67.

[16] Gowrisankaran G. and R. Town (2003), "Competition, Payers, and Hospital Quality", *Health Services*, 38, pp. 1403-1422.

[17] Hamilton B. and V. Hamilton (1997), "Estimating surgical volume-outcome relationships applying survival models: accounting for frailty and hospital fixed effects", *Health Economics*, 6, pp. 383-395.

[18] Hansman H.B. (1996), *The ownership of Enterprise*, Cambridge, Havard University Press.

[19] Hart O.D. and B. Holmstrom (1987), "The theory of contracts", in *Advances in Economic Theory*, Bewley T. (ed.), Cambridge University Press, Cambridge.

[20] Ho V. and B. Hamilton (2000), "Hospital mergers and acquisitions: Does market consolidation harm patients?", *Journal of Health Economics*, 19(5), pp. 767-791.

[21] Johansson S.R. (1991), "The health transition: the cultural inflation of morbidity during the decline of mortality", *Health transition review*, 1, pp. 39-68.

[22] Jollis J., Peterson E., DeLong E., Mark D., Collins S., Muhlbaier L. and D. Prior (1994), "The Relation between the Volume of Coronary Angioplasty Procedures at Hospitals Treating Medicare Beneficiaries and Short-Term Mortality", *The New England Journal of Medicine*, 24(331), pp. 1625-1629.

[23] Kessler D. and M. McClellan (2000), "Is hospital competition socially wasteful?", *Quaterly Journal of Economics*, 115(2), pp. 577-615.

[24] Kessler D. and M. McClellan (2002), "The effects of hospital ownership on medical productivity", *RAND Journal of Economics*, 33(3), pp. 488-506.

[25] Lancaster T. (1990), *The Econometric Analysis of Transition Data*, Cambridge University Press: Cambridge.

[26] Lien H., Chou S. and J. Liu (2008), "Hospital ownership and performance: Evidence from stroke and

[27] cardiac treatment in Taiwan", *Journal of Health Economics*, 27, pp. 1208–1223.

[28] Lindeboom M. and M. Kerkhofs (2000), "Multistate Models for Clustered Duration Data - an Application to Workplace Effects on Individual Sickness Absenteeism", *The Review of Economics and Statistics*, 82(4), pp. 668-684.

[29] Lindeboom, M. and E. van Doorslaer (2004), "Cut-point Shift and Index Shift in Self-reported Health", *Journal of Health Economics*, 23, pp. 1083-1099.

[30] McClellan M. and D. Staiger (1999) "The Quality of Health Care Providers", NBER Working Paper 7327.

[31] McClellan M. and D.O. Staiger (2000), "Comparing Hospital Quality at For-Profit and Not-for-Profit Hospitals", in *The Changing Hospital Industry: Comparing Not-for-Profit and For-Profit Institutions*, D.M. Cutler ed., University of Chicago Press.

[32] McGuire T.G. (2000), "Physician agency", in *Handbook of Health Economics*, pp. 461-536, Culyer A.J. and J.P. Newhouse (eds.), Elsevier, Amsterdam.

[33] Milcent C. (2005), "Hospital Ownership, Reimbursement Systems and Mortality rates", *Health Economics*, 14, pp. 1151–1168.

[34] Milcent C., Dormont B., Durand-Zaleski I. and P.G. Steg (2007), "Gender Differences in Hospital Mortality and Use of Percutaneous Coronary Intervention in Acute Myocardial Infarction", *Circulation*, 115(7), pp. 823-826.

[35] Milgrom P. and J. Roberts (1992), *Economics, Organization and Management*, Prentice Hall, Englewood Cliffs.

[36] Mion G. and P. Naticchioni (2009), "The Spatial Sorting and Matching of Skills and Firms", *Canadian Journal of Economics*, 42(1), pp. 28-55.

[37] Moulton B. (1990), "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units", *The Review of Economics and Statistics*, 72(2), pp. 334-338.

[38] Newhouse J. (1970), "Toward a Theory of Nonprofit Institutions", *American Economic Review*, 60(1), pp. 64-74.

[39] Newhouse J.P. (1996), "Reimbursing health plans and health providers: efficiency in production versus selection", *Journal of Economic Literature*, 34(3), pp. 1236-1263.

[40] Pauly M.V. and M. Redish (1973), "The Not-for-Profit Hospital as a Physicians' cooperative", *American Economic Review*, 63, pp. 87-99.

[41] Pauly M.V. (1990), "The rational nonpurchase of long-term-care insurance", *Journal of Political Economy*, 98(1), pp. 153-168.

[42] Picone G., Shin-Yi C. and F. Sloan (2002), "Are for-profit hospital conversions harmful to patients and to Medicare ?", *RAND Journal of Economics*, 33(3), pp. 507-523.

[43] Propper C., Burgess S. and K. Green (2004), "Does competition between hospitals improve the quality of care? Hospital death rate and the NHS internal market", *Journal of Public Economics*, 88(7-8), pp. 1247-1272.

[44] Ridder G. and I. Tunali (1999), "Stratified partial likelihood estimation", *Journal of Econometrics*, 92(2), pp. 193-232.

[45] Shortell S.M. and E.F.X. Highes (1998), "The effects of Regulation, Competition, and Ownership on Mortality Rates Among Hospital Inpatients", *New England Journal of Medecine*, 318, pp. 1100-1107.

[46] Silverman E.M.and J.S. Skinner (2001), "Are For-Profit Hospitals Really different? Medicare Upcoding and Market Structure", NBER working paper 8133.

[47] Sloan F., Picone, G., Taylor D. and S. Chou (2001), "Hospital Ownership and Cost and Quality of Care: Is There a Dime's Worth of Difference?", *Journal of Health Economics*, 20(1), pp. 1-21.

[48] Sutton M. and P. Lock (2000), "Regional Differences in Health Care Delivery: implications for a national resource allocation formula", *Health Economics*, 9, pp. 547-559.

[49] Tay A. (2003), "Assessing competition in hospital care markets: The importance of accounting for quality differentiation", *RAND Journal of Economics*, 34(4), pp. 786-814.

[50] Town R. and G. Vistnes (2001), "Hospital Competition in HMO networks", *Journal of Health Economics*, 20, pp. 733-753.

[51] Wagstaff A. and E. van Doorslaer (2000), "Equity in Health Care Finance and Delivery", in *Handbook of Health Economics*, pp. 1723-1760, Culyer A.J. et J.P. Newhouse (eds.), Elsevier.

[52] Wray N., Hollinghsworth J., Petersen N., and C. Ashton (1997), "Case-Mix Adjustment Using Administrative Databases: A Paradigm to Guide Future Research", *Medical Care Research and Review*, 54, pp. 326-356.

[53] Wu S. (2001), "Adapting to Heart Conditions: A Test of the Hedonic Treadmill", *Journal of Health Economics*, 20, pp. 495-508.

Table 1: disparity indices computed from
the regional averages of individual variables

	Mean	Min	Max	Max/Min	Std. Dev.	Coeff. of variation	Gini
Number of AMI patients	21448	6335	44393	7.008	11534	.538	.295
Death	.080	.059	.098	1.675	.0098	.123	.070
Female, 35-55 years old	.024	.015	.032	2.145	.0050	.207	.117
Female, 55-65 years old	.026	.021	.034	1.609	.0031	.117	.066
Female, 65-75 years old	.072	.060	.089	1.475	.0074	.103	.056
Female, 75-85 years old	.109	.093	.134	1.435	.0100	.092	.050
Female, over 85 years old	.087	.059	.110	1.852	.0126	.144	.081
Male, 35-55 years old	.187	.135	.239	1.771	.0291	.155	.088
Male, 55-65 years old	.139	.116	.158	1.372	.0137	.099	.057
Male, 65-75 years old	.175	.145	.195	1.343	.0139	.079	.042
Male, 75-85 years old	.134	.105	.159	1.510	.0172	.129	.074
Male, more than 85 year old	.046	.027	.062	2.259	.0090	.196	.108
Excessive smoking	.120	.062	.196	3.160	.0350	.293	.164
Alcohol problems	.012	.004	.017	4.148	.0029	.248	.137
Obesity	.063	.018	.111	6.273	.0196	.313	.170
Diabetes mellitus	.155	.092	.208	2.254	.0240	.156	.077
Hypertension	.299	.203	.373	1.833	.0369	.123	.067
Renal failure	.050	.028	.078	2.760	.0085	.171	.088
Conduction disease	.197	.134	.247	1.843	.0218	.111	.060
Peripheral arterial disease	.063	.036	.109	3.019	.0145	.231	.113
Vascular disease	.044	.025	.078	3.109	.0108	.248	.128
History of coronary artery disease	.040	.017	.070	4.000	.0100	.250	.134
Stroke	.032	.020	.048	2.448	.0055	.173	.092
Heart failure	.158	.128	.204	1.598	.0184	.116	.064
Severity index	.283	.143	.438	3.054	.0737	.261	.147
Cabbage or Coronary Bypass surgery	.009	.001	.036	36.312	.0068	.740	.372
Cardiac catheterization	.190	.130	.271	2.081	.0347	.182	.100
Percutaneous transluminal coronary Angioplasty (PTCA)	.054	.010	.106	1.914	.0270	.497	.277
Other dilatation than PTCA	.002	.000	.005	\	.0016	.994	.534
Percutaneous revascularization using coronary stents (PCI – stenting)	.245	.107	.411	3.836	.0909	.372	.206
Surgical French DRGs	.037	.016	.077	4.650	.0154	.418	.232

Source: computed from the PMSI dataset (1998-2003). Observations used to construct the disparity indices are weighted by the number of AMI patients.

Table 2: disparity indices computed from
the regional averages of hospital and geographic variables

	Mean	Min	Max	Max/Min	Std. Dev.	Coeff. of variation	Gini
Proba. of death within 1 day (KM)	.018	.012	.023	1.940	.003	.158	.089
Proba. of death within 5 days (KM)	.055	.038	.066	1.721	.008	.139	.078
Proba. of death within 10 days (KM)	.088	.061	.107	1.749	.011	.122	.067
Proba. of death within 15 days (KM)	.127	.085	.153	1.800	.016	.128	.070
Number of patients	4466	2363	9974	4.221	2235	.501	.231
Number of AMI patients	386	173	968	5.585	233	.604	.270
Proportion of AMI patients	.089	.061	.157	2.561	.0276	.310	.136
Public	.754	.590	.935	1.584	.1002	.133	.076
Not-for-profit	.045	.000	.261	\	.0525	1.157	.559
For-profit	.201	.060	.367	6.129	.0855	.426	.242
Unemployment rate	.158	.126	.225	1.789	.0282	.178	.098
Poor area in the municipality	.638	.363	.947	2.612	.1944	.305	.173
Municipality median income	13927	11552	17455	1.511	1601	.115	.060
Proportion of beds in surgery	.392	.323	.451	1.395	.030	.076	.043
Number of beds in surgery	753	243	3172	13.062	939	1.246	.469
Proportion of occupied surgery beds	.855	.781	.901	1.153	.0318	.037	.021
Number of beds	1933	595	8488	14.253	2538	1.313	.487
Proportion of occupied beds	.819	.774	.865	1.118	.0247	.030	.017
Number of beds in the urban area	8251	1107	47033	42.475	15016	1.820	.664
Herfindahl index for hospitals in the urban area	.592	.130	.893	6.874	.226	.382	.206

Source: computed from the PMSI, the SAE, and the municipality datasets (1998-2003). Observations used to construct the disparity indices are weighted by the number of AMI patients.

Table 3: estimated coefficients for the individual variables

Variable	Estimate
Female, 55-65 years old	.5426*** (.1113)
Female, 65-75 years old	1.0444*** (.0965)
Female, 75-85 years old	1.3935*** (.0943)
Female, over 85 years old	1.7681*** (.0941)
Male, 35-55 years old	-.3561*** (.1015)
Male, 55-65 years old	.2313** (.0986)
Male, 65-75 years old	.8173*** (.0948)
Male, 75-85 years old	1.2867*** (.0941)
Male, over 85 years old	1.6713*** (.0948)
Excessive smoking	-.4751*** (.0410)
Alcohol problems	.3311*** (.0654)
Obesity	-.2424*** (.0413)
Diabetes mellitus	-.0595*** (.0180)
Hypertension	-.5795*** (.0155)
Renal failure	.3636*** (.0183)
Conduction disease	.8490*** (.0126)
Peripheral arterial disease	-.033 (.0242)
Vascular disease	-.4508*** (.0282)
History of coronary artery disease	-.2408*** (.0290)
Stroke	.2967*** (.0237)
Heart failure	.0569*** (.0134)
Severity index	.1105*** (.0148)
Cabbage or Coronary Bypass surgery	-.7477*** (.0853)
Cardiac catheterization	-1.2587*** (.0299)
Percutaneous Transluminal Coronary Angioplasty	-.6760*** (.0385)
Other dilatation than PTCA	-.874*** (.2181)
Percutaneous revascularization using coronary stents (PCI – stenting)	-1.0207*** (.0261)
Surgical French DRGs	.2852*** (.0358)

Source: computed from the PMSI dataset (1998-2003). Note: ***: significant at 1%; **: significant at 5%; *: significant at 10%. Number of observations: 341,861; mean log-likelihood: -.449369.

Table 4: disparity indices computed from
the regional probability of death obtained from the model

	Mean	Min	Max	Max/Min	Std. Dev.	Coeff. of variation	Gini
Probability of death within 1 day	.019	.015	.024	1.549	.002	.098	.053
Probability of death within 5 days	.057	.050	.073	1.458	.005	.084	.043
Probability of death within 10 days	.086	.074	.108	1.449	.008	.089	.047
Probability of death within 15 days	.116	.098	.144	1.471	.013	.108	.058

Source: computed from the PMSI dataset (1998-2003). Note: for a given region, the probability of death is defined as one minus the regional average of the model survival functions of all hospitals located within the region.

Table 5: regression of hospital fixed effects on hospital and geographic variables

Variable	Regression (1)	Regression (2)	Regression (3)
Constant	-5.917***	-6.195***	-7.003***
	(.216)	(1.445)	(1.517)
For-profit hospital	.286***		.014
	(.041)		(.056)
Not-for-profit hospital	.016		-.166**
	(.071)		(.075)
Proportion of AMI patients in the hospital	-1.049***		-.468**
	(.175)		(.234)
Number of beds (in log)	.107***		.006
	(.016)		(.024)
Occupation rate of beds	.144		.197
	(.223)		(.224)
Proportion of beds in surgery	-.142		-.303***
	(.090)		(.091)
Occupation rate of beds in surgery	-.263		-.267*
	(.161)		(.157)
Median municipality income		.074	.177
		(.148)	(.155)
Presence of a poor area in the municipality		.079**	.065**
		(.031)	(.031)
Municipality unemployment rate		-.321	.046
		(.565)	(.583)
Number of beds in the urban area		.064***	.061**
		(.022)	(.027)
Herfindahl index for the healthcare structure		-.250***	-.283***
		(.089)	(.094)
Regional dummies	Non	Oui	Oui
Number of hospitals	789	834	789
Corresponding number of patients	332,827	333,810	332,827
Adjusted-R^2	.201	.230	.282

Source: computed from the PMSI, the SAE, and the municipality datasets (1998-2003). Note: ***: significant at 1%; **: significant at 5%; *: significant at 10%. We introduced a dummy for the municipality not to be in an urban area (*dummy for rural area*), and a dummy for the municipality to be related to several urban areas (*dummy for multi-polarized municipality*).

Table 6a: variance analysis at the regional level (first stage)

Group of variables from which we consider the effect	Variance	Correlation with the integrated hazard
Integrated hazard	.036856	1.000
Individual variables (averaged at the regional level)	.028623	.898
Innovative treatments	.010155	.740
Non-innovative treatments	.000039	-.136
Diagnoses	.002114	.396
Demographic variables (age x sex)	.005198	.833
Log- hospital fixed effects (averaged at the regional level)	.007159	.475

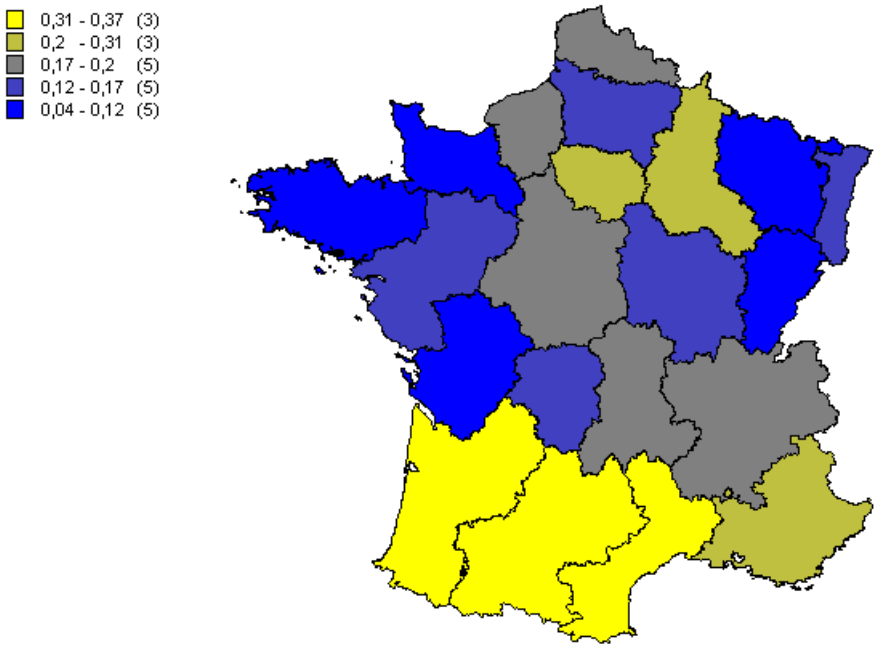
Source: computed from the PMSI, the SAE, and the municipality datasets (1998-2003).

Table 6b: variance analysis at the regional level (third stage)

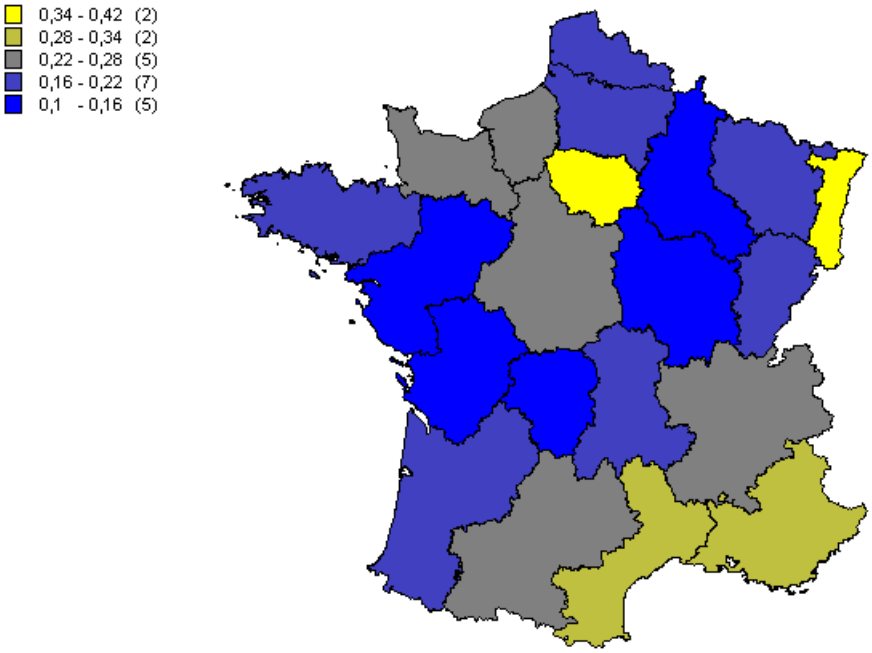
Group of variables from which we consider the effect	Variance	Corr. with log-hosp. fixed effects	Correlation with the integrated hazard
Log- hospital fixed effects (averaged at the regional level)	.007159	1.000	.475
Hospital and geographic variables (averaged at the regional level)	.006012	.987	.403
Hospital variables	.000400	.106	.738
Status and mode of reimbursement	.000078	.319	.624
Proportion of AMI patients	.000110	.083	.387
Beds (capacity and occupation rate)	.000136	-.012	.444
Geographic Variables	.006319	.939	.210
Municipality variables	.010411	.193	-.448
Income-related variables	.000096	-.512	-.290
Dummies for the municipality to be rural or multi-polarized	.000137	.086	-.410
Number of beds in the urban area	.002314	.236	-.284
Herfindahl index for healthcare structure	.002333	.353	-.357
Regional dummies	.010347	.541	.614

Source: computed from the PMSI, the SAE, and the municipality datasets (1998-2003).

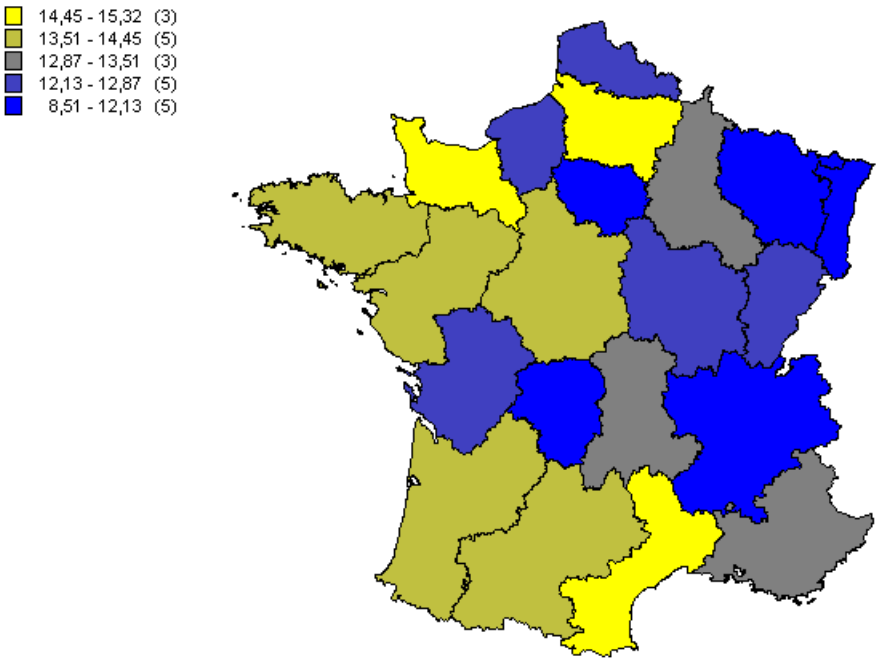
Graph 1: Regional proportions of patients treated by for-profit hospitals



Graph 2: Regional proportions of patients treated with stents

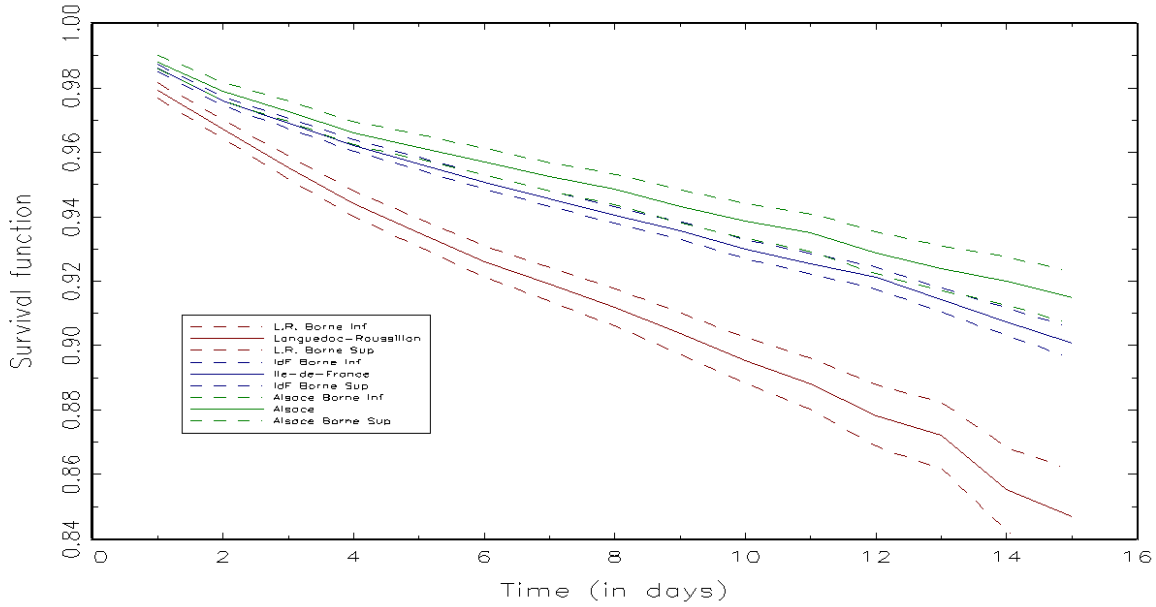


Graph 3: Regional probability of death within fifteen days (in %)



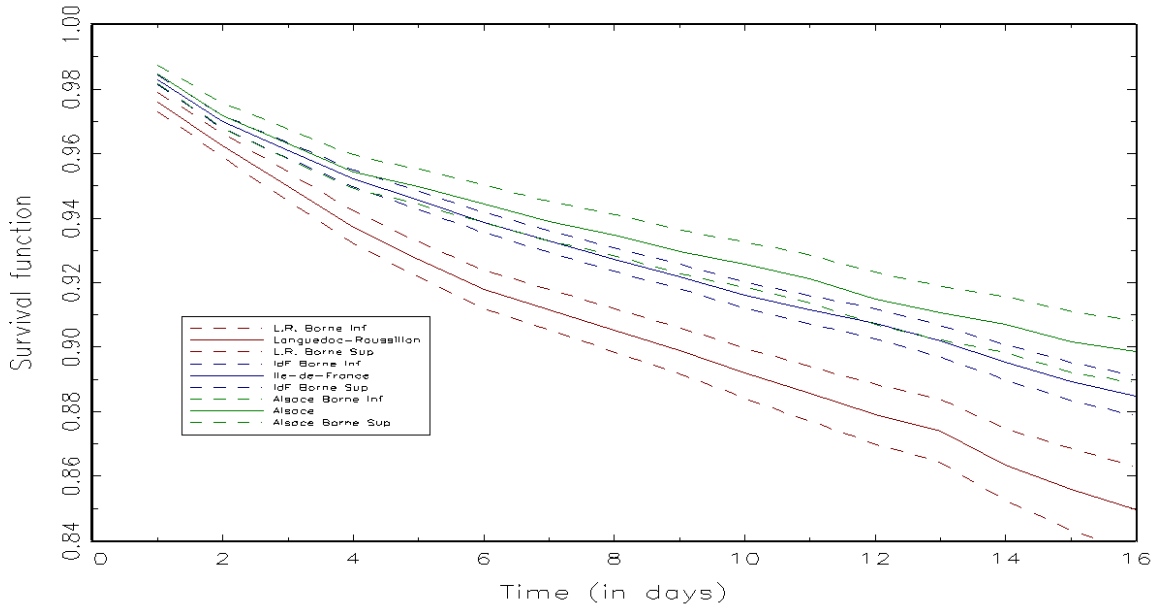
Note: for a given region, the probability of death is defined as one minus the regional average of the Kaplan-Meier survival functions of all hospitals located within the region.

Graph 4: Sample of regional survival functions (Kaplan-Meier)



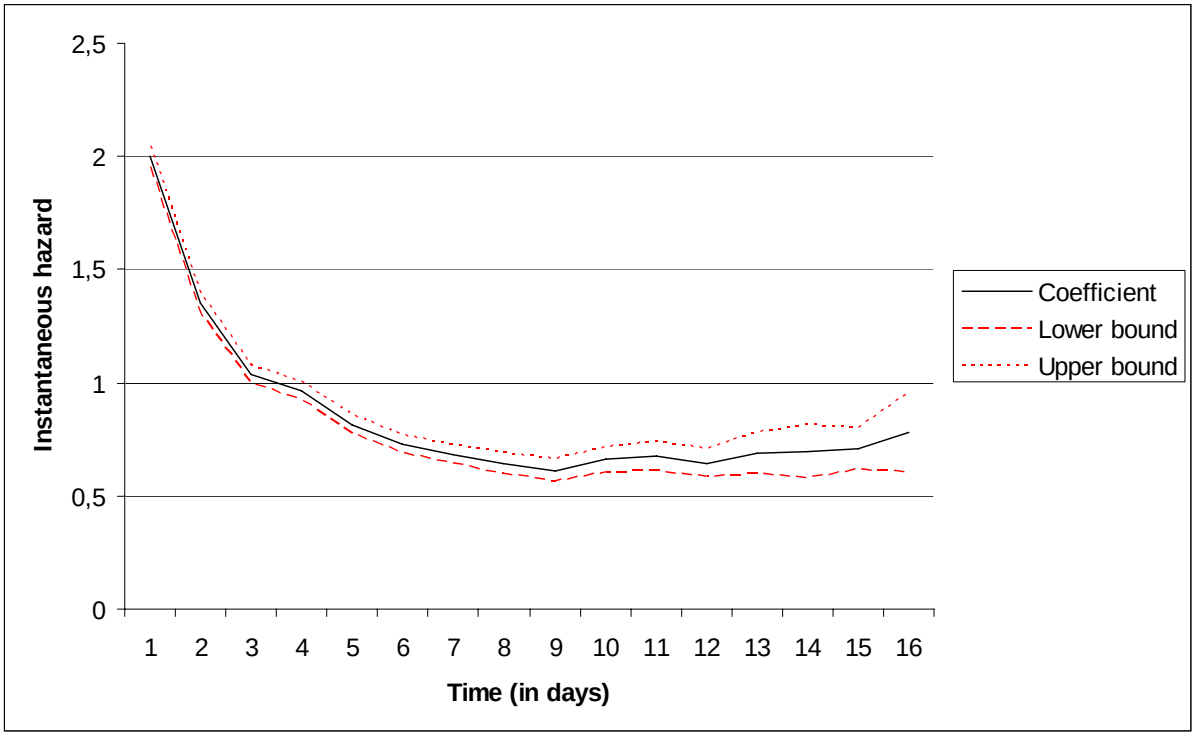
Source: computed from the PMSI dataset (1998-2003). Note: for a given region, the survival function is defined as the regional average of the Kaplan-Meier survival functions of all hospitals located within the region.

Graph 5: Sample of regional survival functions (model)



Source: computed from the PMSI dataset (1998-2003). Note: for a given region, the survival function is defined as the regional average of the model survival functions of all hospitals located within the region.

Graph 6: Baseline instantaneous hazard for exit to death



Source: computed from the PMSI dataset (1998-2003).

Appendix

Table A1: regional probability of death within 15 days (Kaplan-Meier)

Region code	Name	Probability of death
91	Languedoc-Roussillon	15.31%
22	Picardie	15.28%
25	Basse-Normandie	14.48%
53	Bretagne	14.45%
73	Midi-Pyrénées	13.97%
52	Pays de la Loire	13.67%
72	Aquitaine	13.64%
24	Centre	13.51%
83	Auvergne	13.46%
21	Champagne-Ardenne	12.96%
93	Provence – Alpes Côte d'Azur	12.93%
54	Poitou–Charentes	12.87%
31	Nord-Pas-de-Calais	12.73%
26	Bourgogne	12.71%
23	Haute-Normandie	12.60%
43	Franche-Comté	12.13%
82	Rhône-Alpes	12.03%
74	Limousin	11.88%
41	Lorraine	11.67%
11	Ile-de-France	9.93%
42	Alsace	8.51%

Source: computed from the PMSI dataset (1998-2003). Note: for a given region, the probability of death is defined as one minus the regional average of the Kaplan-Meier survival functions of all hospitals located within the region.

Table A2: regional probability of death within 15 days (model)

Numéro de région	Nom	Probability of death	
91	Languedoc-Roussillon	14.40%	(1)
72	Aquitaine	13.95%	(7)
93	Provence – Alpes Côte d'Azur	11.06%	(11)
83	Auvergne	12.85%	(9)
22	Picardie	12.00%	(2)
31	Nord-Pas-de-Calais	11.92%	(13)
24	Centre	11.68%	(8)
73	Midi-Pyrénées	11.68%	(5)
53	Bretagne	11.25%	(4)
25	Basse-Normandie	11.13%	(3)
82	Rhône-Alpes	11.11%	(17)
11	Ile-de-France	11.06%	(20)
41	Lorraine	11.03%	(19)
43	Franche-Comté	10.81%	(16)
52	Pays de la Loire	10.60%	(6)
21	Champagne-Ardenne	10.59%	(10)
23	Haute-Normandie	10.47%	(15)
74	Limousin	9.92%	(18)
54	Poitou–Charentes	9.87%	(12)
26	Bourgogne	9.85%	(14)
42	Alsace	9.84%	(21)

Source: computed from the PMSI dataset (1998-2003). Note: for a given region, the probability of death is defined as one minus the regional average of the Kaplan-Meier survival functions of all hospitals located within the region. In the last column, the ranking of the regions obtained from raw data is reported in brackets.

Table A3: regression of hospital fixed effects on aggregated variables, innovative treatments are not taken into account

Variable	Regression (1)	Regression (2)	Regression (3)
Constant	-4.926***	-5.518***	-5.838***
	(.200)	(1.400)	(1.439)
For-profit hospital	-.211***		-.220***
	(.038)		(.053)
Not-for-profit hospital	-.257***		-.232***
	(.066)		(.071)
Proportion of AMI patients in the hospital	-.604***		-.491**
	(.162)		(.222)
Number of beds (in log)	-.012		-.009
	(.015)		(.023)
Occupation rate of beds	-.049		-.115
	(.207)		(.212)
Proportion of beds in surgery	-.296***		-.284***
	(.084)		(.086)
Occupation rate of beds in surgery	-.101		-.127
	(.150)		(.150)
Median municipality income		.071	.172
		(.143)	(.147)
Presence of a poor area in the municipality		.010	.001
		(.030)	(.030)
Municipality unemployment rate		.242	.437
		(.547)	(.551)
Number of beds in the urban area		-.043**	-.052**
		(.021)	(.026)
Herfindahl index for the healthcare structure		-.087	-.275***
		(.086)	(.088)
Regional dummies	Non	Oui	Oui
Number of hospitals	789	834	789
Corresponding number of patients	332,827	333,810	332,827
Adjusted-R^2	.159	.155	.270

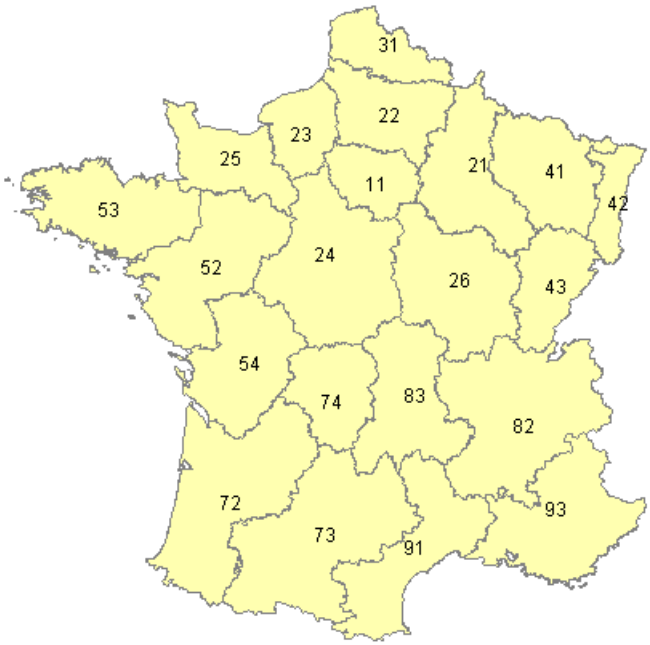
Source: computed from the PMSI, the SAE, and the municipality datasets (1998-2003). Note: ***: significant at 1%; **: significant at 5%; *: significant at 10%. We introduced a dummy for the municipality not to be in an urban area (*dummy for rural area*), and a dummy for the municipality to be related to several urban areas (*dummy for multi-polarized municipality*).

Table A4: regional dummies obtained from the hospital fixed-effect regression

Region code	Name	Coefficient	
91	Languedoc-Roussillon	< Reference >	(1)
25	Basse-Normandie	-.170*	(3)
		(.090)	
41	Lorraine	-.174**	(19)
		(.088)	
22	Picardie	-.181**	(2)
		(.086)	
53	Bretagne	-.181**	(4)
		(.080)	
72	Aquitaine	-.189**	(7)
		(.075)	
21	Champagne-Ardenne	-.227**	(10)
		(.090)	
93	Provence – Alpes Côte d'Azur	-.228***	(11)
		(.071)	
26	Bourgogne	-.228**	(4)
		(.088)	
24	Centre	-.229***	(8)
		(.083)	
74	Limousin	-.233**	(18)
		(.102)	
83	Auvergne	-.233***	(9)
		(.089)	
43	Franche-Comté	-.242**	(16)
		(.101)	
54	Poitou–Charentes	-.242***	(12)
		(.087)	
52	Pays de la Loire	-.261***	(6)
		(.079)	
82	Rhône-Alpes	-.264***	(17)
		(.074)	
73	Midi-Pyrénées	-.267***	(5)
		(.077)	
31	Nord-Pas-de-Calais	-.284***	(13)
		(.070)	
42	Alsace	-.294***	(21)
		(.098)	
23	Haute-Normandie	-.316***	(15)
		(.087)	
11	Ile-de-France	-.583***	(20)
		(.108)	

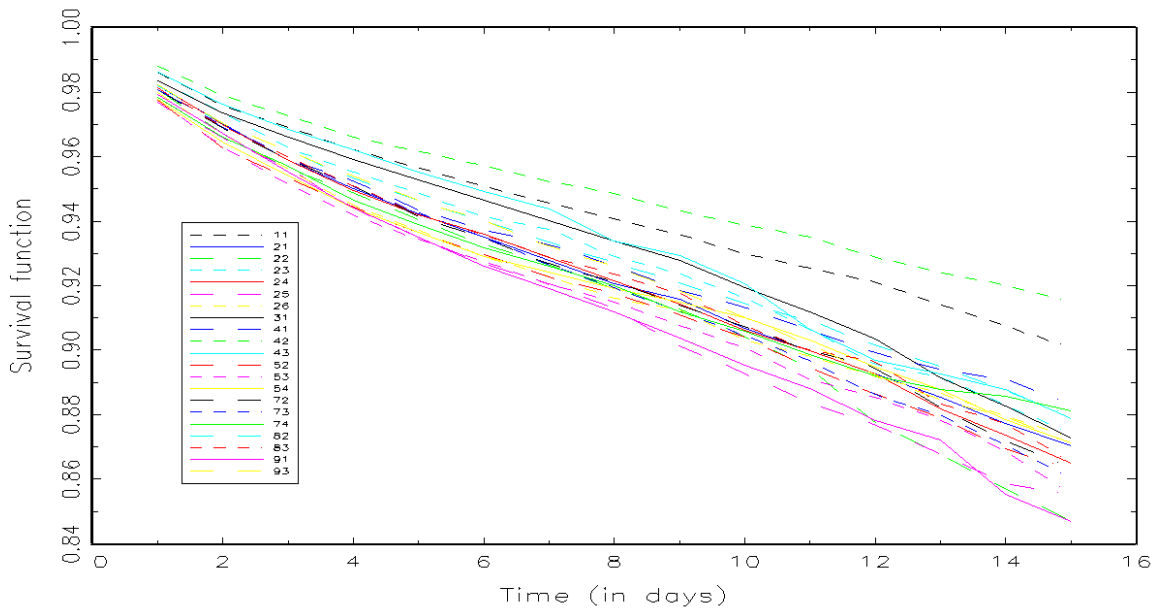
Source: computed from the PMSI, the SAE, and the municipality datasets (1998-2003). Note: in the last column, the ranking of the regions obtained from raw data is reported in brackets.

Graph A1: Map of the French Regions



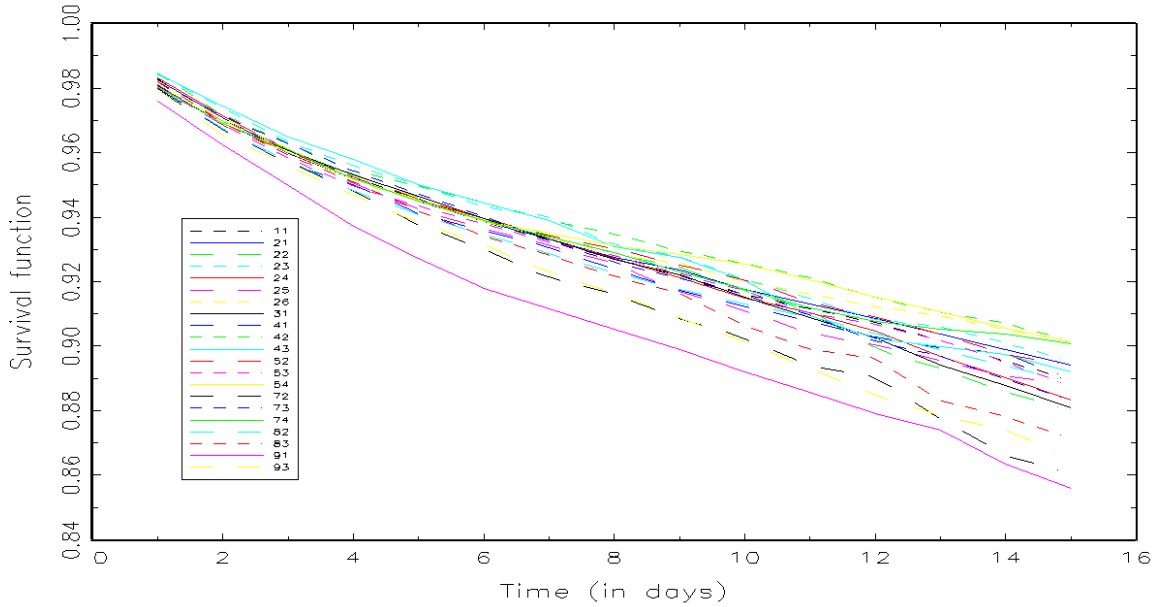
Regions. 11: Ile-de-France; 21: Champagne-Ardenne; 22: Picardie; 23: Haute-Normandie; 24: Centre; 25: Basse-Normandie; 26: Bourgogne; 31: Nord Pas-de-Calais; 41: Lorraine; 42: Alsace; 43: Franche-Comté; 52: Pays de la Loire; 53: Bretagne; 54: Poitou-Charentes; 72: Aquitaine; 73: Midi-Pyrénées; 74: Limousin; 82: Rhônes-Alpes; 83: Auvergne; 91: Languedoc-Roussillon; 93: Provence - Alpes Côtes d'Azur.

Graph A2: Regional survival functions (Kaplan-Meier)



Source: computed from the PMSI dataset (1998-2003). Note: for a given region, the survival function is defined as the regional average of the Kaplan-Meier survival functions of all hospitals located within the region.

Graph A3: Regional survival functions (model)



Source: computed from the PMSI dataset (1998-2003). Note: for a given region, the survival function is defined as the regional average of the model survival functions of all hospitals located within the region.

B.5. The productivity advantages of large cities: distingushing agglomeration from firm selection

Combes PPh., Duranton G., Gobillon L., Puga D. et S. Roux (2009), “The productivity advantages of large cities: distingushing agglomeration from firm selection”, CEPR Working Paper 7191, revise-and-resubmit à *Econometrica*.

45 pages

The productivity advantages of large cities: Distinguishing agglomeration from firm selection

Pierre-Philippe Combes[*†]
University of Aix-Marseille and *CEPR*

Gilles Duranton[*‡]
University of Toronto and *CEPR*

Laurent Gobillon[*§]
Institut National d'Etudes Démographiques, *PSE-INRA*, and *CREST*

Diego Puga[*§]
IMDEA, *Universidad Carlos III* and *CEPR*

Sébastien Roux[*‖]
CREST - INSEE

February 2009

Abstract: Firms are more productive on average in larger cities. Two explanations have been offered: agglomeration economies (larger cities promote interactions that increase productivity) and firm selection (larger cities toughen competition allowing only the most productive to survive). To distinguish between them, we nest a generalised version of a seminal firm selection model and a standard model of agglomeration. Stronger selection in larger cities left-truncates the productivity distribution whereas stronger agglomeration right-shifts and dilates the distribution. We assess the relative importance of agglomeration and firm selection using French establishment-level data and a new quantile approach. Spatial productivity differences in France are mostly explained by agglomeration.

Key words: agglomeration, firm selection, productivity, cities

JEL classification: C52, R12, D24

[*]We thank Kristian Behrens, Stéphane Grégoir, Marc Melitz, Peter Neary, Gianmarco Ottaviano, Giovanni Peri, Stephen Redding, John Sutton, Dan Trefler, and conference and seminar participants for comments and discussions. We gratefully acknowledge funding from the Agence Nationale de la Recherche (grant COMPNASTA), the Canadian Social Science and Humanities Research Council, the Centre National de la Recherche Scientifique, the Comunidad de Madrid (grant PROCIUDAD-CM), the European Commission's Seventh Research Framework Programme (contract number 225551, collaborative project European Firms in a Global Economy — EFIGE), and the Fundación Ramón Areces.

[†]GREQAM, 2 Rue de la Charité, 13236 Marseille cedex 02, France (e-mail: ppcombes@univmed.fr; website: http://www.vcharite.univ-mrs.fr/pp/combes/).

[‡]Department of Economics, University of Toronto, 150 Saint George Street, Toronto, Ontario M5S 3G7, Canada (e-mail: gilles.duranton@utoronto.ca; website: http://individual.utoronto.ca/gilles/default.html).

[§]Institut National d'Etudes Démographiques, 133 Boulevard Davout, 75980 Paris cedex 20, France (e-mail: laurent.gobillon@ined.fr; website: http://laurent.gobillon.free.fr/).

[§]Madrid Institute for Advanced Studies (IMDEA) Social Sciences, Antiguo pabellón central del Hospital de Cantoblanco, Carretera de Colmenar Viejo km. 14, 28049 Madrid, Spain (e-mail: diego.puga@imdea.org; website: http://diegopuga.org).

[‖]Centre de Recherche en Économie et Statistique (CREST), 15 Boulevard Gabriel Péri, 92245 Malakoff cedex, France (e-mail: sebastien.roux@ensae.fr).

1. Introduction

Firms and workers are, on average, more productive in larger cities. This fact — already discussed by Adam Smith (1776) and Alfred Marshall (1890) — is now firmly established empirically (see Rosenthal and Strange, 2004, and Melo, Graham, and Noland, 2009, for reviews and summaries of existing findings). Estimates of the magnitude of this effect range between a 2 and 7 percent productivity increase from a doubling of city size for a large range of city sizes, depending on the sector and details of the estimation procedure.

For a long time, the higher average productivity of firms and workers in larger cities has been attributed to 'agglomeration economies'. These agglomeration economies are thought to arise from a variety of mechanisms such as the possibility for similar firms to share suppliers, the existence of thick labour markets ironing out firm-level shocks or facilitating matching, or the possibility to learn from the experiences and innovations of others (see Duranton and Puga, 2004, for a review). All these mechanisms share a common prediction: the concentration of firms and workers in space makes them more productive.

More recently, an alternative explanation has been offered based on 'firm selection'. The argument builds on work by Melitz (2003), who introduces product differentiation and international or inter-regional trade in the framework of industry dynamics of Hopenhayn (1992). Melitz and Ottaviano (2008) incorporate endogenous price-cost mark-ups in this framework and show that larger markets attract more firms, which makes competition tougher.[1] In turn, this leads less productive firms to exit. This suggests that the higher average productivity of firms and workers in larger cities could instead result from a stronger Darwinian selection of firms.

Our main objective in this paper is to distinguish between agglomeration and firm selection in explaining why average productivity is higher in larger cities. To do so, our first step is to free the framework of Melitz and Ottaviano (2008) from distributional assumptions and generalise it to many cities. We then combine this model with a fairly general model of agglomeration in the spirit of Fujita and Ogawa (1982) and Lucas and Rossi-Hansberg (2002). This nested model allows us to parameterise the relative importance of agglomeration and selection. The main prediction of our model is that, while selection and agglomeration effects both make average firm log productivity higher in larger cities, they have different predictions for how the shape of the log productivity distribution varies with city size. In particular, stronger selection effects in larger cities should lead to a greater left truncation of the distribution of firm log productivities in larger cities, as the least productive firms exit. Stronger agglomeration effects in larger cities should lead instead to a greater rightwards shift of the distribution of firm log productivities in larger cities, as agglomeration effects make all firms more productive. To the extent that more productive firms are better able to reap the benefits of agglomeration, agglomeration should also lead to an increased dilation of the distribution of firm log productivities in larger cities.

We then use this predictions to assess the relative importance of agglomeration and firm selection for different sectors using data for all French firms. Our structural estimation is in two steps.

[1] Bernard, Eaton, Jensen, and Kortum (2003) also develop a model with heterogenous firm productivity levels and endogenous mark-ups but, unlike in Melitz and Ottaviano (2008), these mark-ups are not affected by market size. In Nocke (2006), more able entrepreneurs sort into larger markets because competition there is more intense.

We first estimate total factor productivity at the establishment level. Next, we develop a new quantile approach to compare the distribution of establishment log productivities for each sector across metropolitan areas of different sizes. As stipulated by the model, we estimate the extent to which the log productivity distribution in large cities is left-truncated (evidence of differences in selection effects) or dilated and right-shifted (evidence of differences in agglomeration effects) compared to the log productivity distribution in small cities.

This empirical approach offers a number of benefits. First, it allows both agglomeration economies and firm selection to play a role, instead of focusing on just one or the other. Second, while firmly grounded in a nested model, our approach identifies selection and agglomeration from features that are common to a much broader class of models. Basically, it relies on fiercer competition eliminating the weakest firms and on agglomeration economies raising everyone's productivity — possibly to different extents. Third, we do not rely on particular distributional assumptions of firms' productivity nor on a particular moment of the data. Fourth, our approach does not attempt to identify selection by looking for cutoffs in the lower tail of the log productivity distribution, which may be obscured by measurement error, nor by looking for greater log productivity dispersion in larger cities, which is not a necessary consequence of selection. Instead, it estimates differences in truncation across areas from their entire distributions using the fact that greater truncation raises the density distribution proportionately everywhere to the right of the cutoff.

Our main finding is that productivity differences between French metropolitan areas are explained mostly by agglomeration. On the other hand, we find no systematic evidence of stronger selection in larger cities. We begin with the simplest characterisation of agglomeration economies: a common upwards shift in log productivity. This shift alone is able to explain most of the differences in the log productivity distribution between cities of different sizes, and corresponds to a productivity gain across the board of about 12 percent for establishments in metropolitan areas with population above 200,000 relative to establishments located elsewhere. Even with this simple characterisation of agglomeration economies, there are just no sizeable differences in left truncation across cities of different sizes. We then also allow for the possibility that more productive establishments are better able to reap the benefits from agglomeration, which dilates upwards the log productivity distribution. This additional consequence of agglomeration is also supported by the data. While the average productivity gain is now about 9 percent , establishments at the bottom quartile of the log productivity distribution are only 5 percent more productive in metropolitan areas with population above 200,000 than elsewhere whereas establishments at the top quartile are about 14 percent more productive in larger cities.

While our results about agglomeration and selection apply broadly to manufacturing and business services as well as to most particular sectors, we can find some exceptions within sufficiently detailed sectoral classifications. A few sectors do not seem to benefit from stronger agglomeration economies in large metropolitan areas. A few sectors also exhibit stronger selection in large metropolitan areas. However, such exceptions play almost no role in explaining differences across cities in the aggregate log productivity distribution. Finally, none of our results appears sensitive to our choice of estimation technique for productivity nor to the sample of establishments.

Our paper is related to the large agglomeration literature building on Henderson (1974) and Sveikauskas (1975), and surveyed in Duranton and Puga (2004), Rosenthal and Strange (2004) and Head and Mayer (2004). We extend it by considering an entirely different reason for the higher average productivity in larger cities. It is also related to the pioneering work of Syverson (2004) who examines the effect of market size on firm selection in the ready-made concrete sector and the emerging literature that follows (Del Gatto, Mion, and Ottaviano, 2006, Del Gatto, Ottaviano, and Pagnini, 2008). A first difference with Syverson's work is that we build our empirical approach on a nested model of selection and agglomeration rather than on a model incorporating selection alone. Considering agglomeration and selection simultaneously allows us to identify robust differences in predictions between the two types of mechanisms. A second difference is that, instead of examining differences in summary statistics across locations, we develop a quantile approach that traces differences throughout the log productivity distribution. A third difference is that we consider firms not only in the ready-made concrete sector but in the entire economy. Our paper is finally related to Carrasco and Florens (2000), since our quantile approach adapts their results for an infinite set of moments to deal with an infinite set of quantile equalities.[2]

The rest of this paper is organised as follows. The next section proposes a generalisation of Melitz and Ottaviano (2008) and combines it with an agglomeration model. Section 3 describes our econometric approach. Section 4 discusses the data and the details of our empirical implementation. The baseline results are then presented in Section 5. Section 6 introduces a more general version of our theoretical model and econometric approach, and Section 7 presents our corresponding main results. Finally, Section 8 discusses some additional issues, and Section 9 concludes.

2. A nested model of selection and agglomeration

To build the theoretical foundations of our empirical approach, we nest a generalised version of the firm selection model of Melitz and Ottaviano (2008) and a model of agglomeration economies in the spirit of Fujita and Ogawa (1982) and Lucas and Rossi-Hansberg (2002).

An individual consumer's utility is given by

$$U = q^0 + \alpha \int_{i\in\Omega} q^i \mathrm{d}i - \frac{1}{2}\gamma \int_{i\in\Omega} (q^i)^2 \mathrm{d}i - \frac{1}{2}\eta \left(\int_{i\in\Omega} q^i \mathrm{d}i \right)^2 , \tag{1}$$

[2]There is also a large literature in international trade that explores whether good firms self-select into exporting or learn from it. Early studies (Clerides, Lach, and Tybout, 1998, Bernard and Jensen, 1999) conclude at the predominance of self-selection by observing that exporting firms have better pre-determined characteristics. More recent work by Lileeva and Trefler (2007) shows that lower US tariffs provided less productive Canadian firms with an opportunity to invest and improve their productivity to export to the US. A similar type of question can be raised regarding the higher productivity of firms in import-competing sectors. Pavcnik (2002) uses trade liberalisation in Chile to provide evidence about both selection (exit of the least productive firms and factor reallocation towards the more productive firms) and increases in productivity when firms have to compete with importers. Both strands of literature usually identifies selection from changes over time either in trade policy or along the firm life-cycle. With city size changing only slowly over time, we need to use instead a cross-sectional approach. The other difference with the trade literature is that we implement a structural model rather than run reduced-form regressions. We postpone further discussion of how our results fit with the implications from this trade literature to the concluding section.

where q^0 denotes the individual's consumption of a homogenous numéraire good, and q^i her consumption of variety i of a set Ω of differentiated varieties. The three positive demand parameters α, γ, and η are such that a higher α and a lower η increase demand for differentiated varieties relative to the numéraire, while a higher γ reflects more product differentiation between varieties.[3]

Utility maximisation subject to the budget constraint yields the following inverse demand for differentiated variety i by an individual consumer:

$$p^i = \alpha - \gamma q^i - \eta \int_{j \in \Omega} q^j \mathrm{d}j \,, \tag{2}$$

where p^i denotes the price of variety i. It follows from (2) that varieties with too high a price are not consumed. This is because, by (1), the marginal utility for any particular differentiated variety is bounded. Let $\bar{\Omega}$ denote the set of varieties with positive consumption levels in equilibrium, ω the measure of $\bar{\Omega}$, and $P \equiv \frac{1}{\omega} \int_{j \in \bar{\Omega}} p^j \mathrm{d}j$ the average price of varieties with positive consumption. Integrating equation (2) over all varieties in $\bar{\Omega}$, solving for $\int_{j \in \Omega} q^j \mathrm{d}j$, and substituting this back into equation (2), we can solve for an individual consumer's demand for variety i as:

$$q^i = \begin{cases} \frac{1}{\gamma + \eta\omega}(\alpha + \frac{\eta}{\gamma}\omega P) - \frac{1}{\gamma} p^i & \text{if } p^i \leqslant \bar{h} \equiv P + \frac{\gamma(\alpha - P)}{\gamma + \eta\omega} \,, \\ 0 & \text{if } p^i > \bar{h} \,. \end{cases} \tag{3}$$

The price threshold, $\bar{h}$, in equation (3) follows immediately from the restriction $q^i \geqslant 0$. By the definition of P and equation (2), $P < \alpha$ so that $\bar{h} > P$.

The numéraire good is produced under constant returns to scale using one unit of labour per unit of output. It can be freely traded when we consider more than one location. This implies that the cost to firms of hiring one unit of labour is always unity.[4]

Differentiated varieties are produced under monopolistic competition. By incurring a sunk entry cost s, a firm is able to develop a new variety that can be produced using h units of labour per unit of output. Given that the cost of each unit of labour equals one unit of the numéraire, h is also the marginal cost. The unit labour requirement h differs across firms and for each of them it is randomly drawn from a distribution with known probability density function $g(h)$ and cumulative $G(h)$. Melitz and Ottaviano (2008) derive most of their results under the assumption that $g(h)$ is a Pareto distribution. By contrast, we do not adopt any particular distribution for $g(h)$. To simplify the derivation of the results, we only require $G(.)$ to be continuous and differentiable. Firms with a marginal cost higher than the price at which consumer demand becomes zero are unable to cover their marginal cost and exit. The set of goods varieties that end up being produced in equilibrium is therefore $\bar{\Omega} = \{i \in \Omega \mid h \leqslant \bar{h}\}$.

Since all varieties enter symmetrically into utility, we can index firms by their unit labour requirement h instead of the specific variety i they produce. Re-writing the individual consumer demand of (3) in terms of $\bar{h}$ and multiplying this by the mass of consumers C yields the following

[3]The specification in (1) is often referred to as the quadratic utility model of horizontal product differentiation. It has been used in industrial organisation by, for instance, Dixit (1979) and Vives (1990) and has become popular in location modelling following Ottaviano, Tabuchi, and Thisse (2002).

[4]The unit cost for labour holds provided there is some production of the numéraire good everywhere. Given the quasi-linear preferences, this requires that income is high enough, which is easy to ensure.

expression for the total sales of an individual firm:

$$Q(h) = Cq(h) = \begin{cases} \frac{C}{\gamma}[\bar{h} - p(h)] & \text{if } p(h) \leqslant \bar{h} \,, \\ 0 & \text{if } p(h) > \bar{h} \,. \end{cases} \tag{4}$$

Given that the entry cost is sunk when firms draw their value of h, active firms set prices to maximise operational profits given by

$$\pi(h) = [p(h) - h]Q(h) \,. \tag{5}$$

Maximising $\pi(h)$ in (5) subject to (4) yields the optimal pricing rule

$$p(h) = \frac{1}{2}(h + \bar{h}) \,. \tag{6}$$

Substituting (4) and (6) into (5) we obtain equilibrium operational profits:

$$\pi(h) = \frac{C}{4\gamma}(\bar{h} - h)^2 \,. \tag{7}$$

Entry into the monopolistically competitive industry takes place until *ex-ante* expected profits are driven to zero. The operational profits expected prior to entry must therefore be exactly offset by the sunk entry cost:

$$\frac{C}{4\gamma}\int_0^{\bar{h}} (\bar{h} - h)^2 g(h)\mathrm{d}h = s \,. \tag{8}$$

The free-entry condition (8) implicitly defines the marginal cost cutoff $\bar{h}$ as a function of the distribution $g(h)$, the sunk entry cost s, the mass of consumers C, and the degree of product differentiation parameter γ. We note that while we rely on the framework developed by Melitz and Ottaviano (2008), the existence of a marginal cost cutoff should be a general property of a whole class of market selection models.

We now turn to the agglomeration components of the model. Workers are endowed with a single unit of working time each that they supply inelastically. Each worker is made more productive by interactions with other workers. More specifically, when interacting with W other workers, the effective units of labour supplied by an individual worker during their unit working time is $a(W)$, where $a(0) = 1$, $a' > 0$ and $a'' < 0$. We can think of such interactions as exchanges of ideas between workers, where being exposed to a greater diversity of ideas makes each worker more productive. This motivation for agglomeration economies based on interactions between workers can be found in, amongst others, Fujita and Ogawa (1982) and Lucas and Rossi-Hansberg (2002) — we introduce a discrete version of their spatial decay for interactions below. Duranton and Puga (2004) review micro-foundations of numerous alternative agglomeration mechanisms, which also result in a reduced form like $a(W)$. We assume that such interactions benefit workers across occupations, i.e., regardless of whether they produce any particular variety of the differentiated

good or the numéraire good (a simplifying assumption that we relax below).[5] This, given the unit payment per effective unit of labour supplied, implies that the total labour income of each worker in any occupation is $a(W)$.

A firm with unit labour requirement h hires $l(h) = Q(h)h/a(W)$ workers at a total cost of $a(W)l(h) = Q(h)h$. The natural logarithm of the firm's productivity is then given by

$$\phi = \ln\left(\frac{Q}{l}\right) = \ln\left[a(W)\right] - \ln(h) \ . \tag{9}$$

The probability density function of firms' log productivities is then

$$f(\phi) = \begin{cases} 0 & \text{for } \phi < A - \ln(\bar{h}) \ , \\ \frac{e^{A-\phi}\, g\left(e^{A-\phi}\right)}{G(\bar{h})} & \text{for } \phi \geqslant A - \ln(\bar{h}) \ , \end{cases} \tag{10}$$

where

$$A \equiv \ln\left[a(W)\right] \ . \tag{11}$$

The numerator of $f(\phi)$, $e^{A-\phi}\, g\left(e^{A-\phi}\right)$ follows from using equation (9) and the change of variables theorem, while the denominator $G(\bar{h})$ takes care of the fact that firms with a unit labour requirement above $\bar{h}$ exit. The model can now be solved sequentially by first using the free entry condition of equation (8) to solve for the equilibrium cut-off unit labour requirement $\bar{h}$. We can then substitute $\bar{h}$ into (10) to obtain the equilibrium distribution of firm productivities. Finally, equation (6) gives prices and the definition of $\bar{h}$ in (3) allows us to compute the mass of varieties produced, ω.

To understand how selection and agglomeration forces contribute to determining the distribution of firms' log productivities we must consider what is the relevant mass C of consumers that each firm sells to and what is the relevant mass W of people that each worker interacts with. Before deriving our formal results in a general setting below, let us first develop a simpler example that illustrates the key properties of our framework with particular clarity.

An illustrative example

Consider two polar possibilities for both demand and interactions in an economy with two cities. In terms of demand, at one extreme we can think of firms selling only to consumers in their city and thus competing with other local firms only (*local product-market competition*). At the other extreme, firms can sell with equal ease to consumers anywhere and thus compete with firms everywhere (*global product-market competition*). In terms of interactions, at one extreme we can think of workers interacting exclusively with other workers living in the same city (*local interactions*). At the other

[5] More realistically, we would expect the benefits of agglomeration to depend not just on the interactions between workers but also on the characteristics of firms. In particular, more productive firms (i.e., those with a lower h) may be more able to reap the benefits of agglomeration. We explicitly allow for this possibility in Section 6, where we extend both our model and empirical approach. Regarding the assumption that the benefits of agglomeration economies percolate to the numéraire sector, this ensures that such benefits get reflected in individual worker earnings so that these increase with city size, consistent with the empirical evidence (see, e.g., Glaeser and Maré, 2001, Wheaton and Lewis, 2002, Combes, Duranton, Gobillon, and Roux, 2009). If we re-wrote the model so that, counterfactually, earnings per worker not only did not increase but decreased with city size, then selection forces could be much weakened or even reversed.

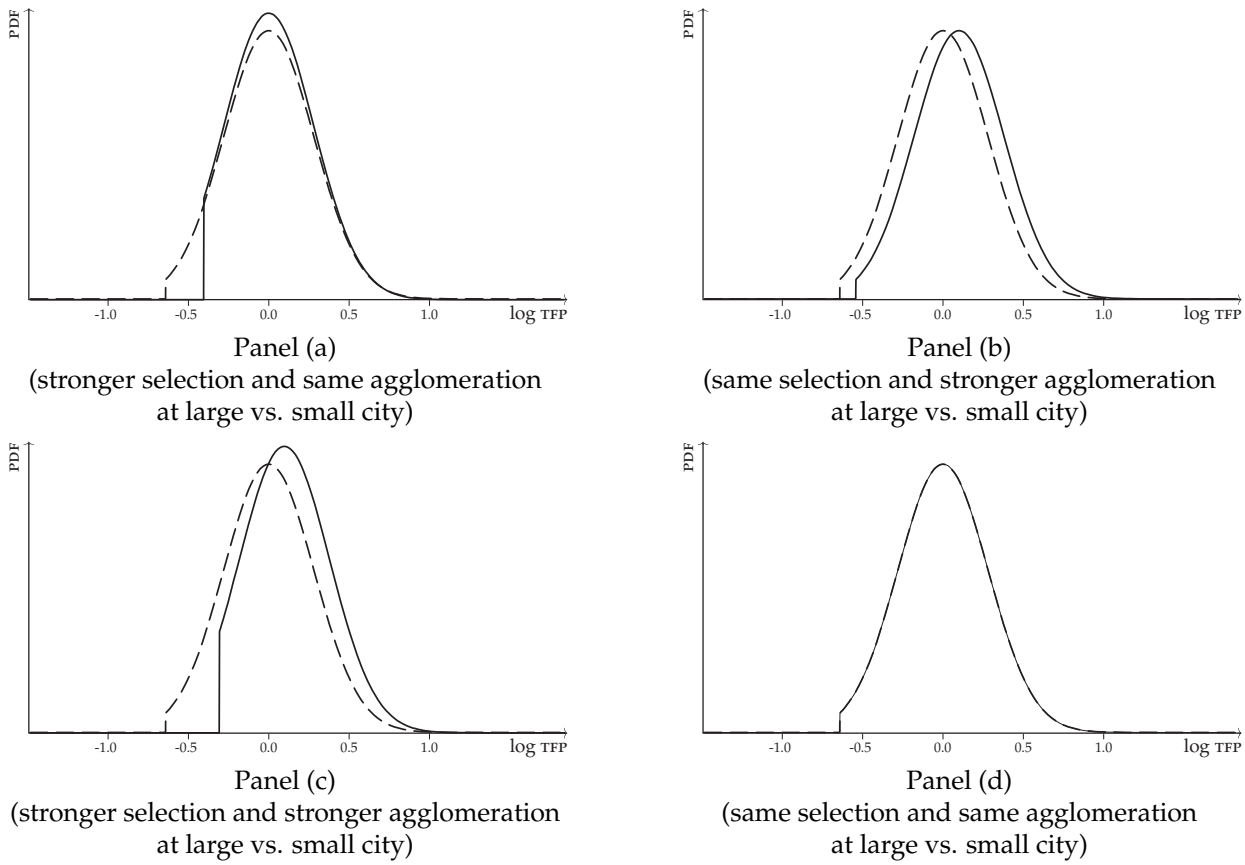


Figure 1: Log productivity distributions in large (solid) and small cities (dashed)

extreme, workers can interact with equal ease with workers living anywhere (*global interactions*). The combination of these possibilities gives us four cases.[6] We now compare in each of the four cases the distribution of firms' log productivities across two cities of different population size (a large city indexed 1 and a small city indexed 2).

Case 1 (*local product-market competition and global interactions*)**.** Panel (a) in Figure 1 plots the distribution of firms' log productivities in a city with a large population (continuous line) and in a city with a small population (dashed line) in the case where firms only sell in their local city and workers enjoy interactions with the same intensity with workers from everywhere (i.e., $C_1 > C_2$ and $W_1 = W_2$). Compared with the distribution in the small city, the large-city distribution is left-truncated as a consequence of firm selection. This left truncation implies that the peak of the large city distribution is higher than that of the small city distribution and that the two peaks occur at the same level of productivity. To understand this greater truncation in the large city, note that if the number of active firms in the large city was the same as in the small city, every large-city firm would sell proportionately more. Formally, using equations (4) and (6), total sales for an individual firm can be expressed as $\frac{C}{2\gamma}(\bar{h} - h)$. Hence, for a given measure of firms ω, that is, for a given $\bar{h}$,

[6]We later derive our formal results in a general setting with multiple cities where firms can sell to consumers everywhere, but there are potentially some additional costs incurred when they sell outside their city, and workers interact with other workers everywhere, but interactions with workers in other cities are potentially weaker.

sales increase proportionately with the population of consumers C. However, the larger individual firm sales associated with a larger C make further entry profitable and, by equation (8), they must be offset by a lower $\bar{h}$ to restore zero *ex-ante* expected profits.[7]

To understand how firms in different ranges of the productivity distribution are affected by city size and $\bar{h}$, note that, from (4) and (6), the price elasticity of demand faced at equilibrium by a firm with unit labour requirement h can be written as follows:

$$\epsilon(h) \equiv -\frac{p(h)}{Q(h)}\frac{\mathrm{d}Q(h)}{\mathrm{d}p(h)} = \frac{\bar{h}+h}{\bar{h}-h}\,. \tag{12}$$

Demand becomes more price-elastic as h increases or as $\bar{h}$ decreases. Thus, each firm in the large city (where $\bar{h}$ is lower) faces a more elastic demand, and hence charges a lower markup, $p - h = (\bar{h} - h)/2$, than a firm with the same h in the small city. The combination of more consumers, further entry, and the ensuing lower markups in the large city affects firms' sales differently depending on their h.[8] In the large city, firms with high productivity, and hence high markups, enjoy smaller profit margins but larger sales than their small-city counterparts. Low productivity firms, however, have both smaller profit margins and smaller sales in the large city than in the small city. In short, product market competition is tougher in a large city than in a small city, and this affects firms with low productivity and hence low price-cost margins the most. Some low-productivity firms that would have been able to survive in a small city cannot lower their prices any further and must exit in the large city. It is this exit at the low-productivity end that leads to the large-city log productivity distribution being a left-truncated version of the small-city distribution (see Lemma 1 below for a formal proof).

Case 2 (*global product-market competition and local interactions*)**.** Panel (b) in Figure 1 plots the distribution of firms' log productivities in a city with a large population (continuous line) and in a city with a small population (dashed line) in the case where every firm competes with the same intensity with firms from everywhere and workers only interact with workers in their city (i.e., $C_1 = C_2$ and $W_1 > W_2$).[9] Compared with the distribution in the small city, the large-city distribution is right-shifted. Since interactions are local, workers in the larger city benefit from

[7]The reason why a larger measure of firms leads to a lower $\bar{h}$ is the following. By (3), even if firms were to keep their prices constant following entry (leaving P unchanged), the business stealing effect of entry (larger ω) is enough to make the sales of more expensive varieties drop to zero. In turn, by (6), this lower $\bar{h}$ induces firms to lower their prices which, by (3), further reduces $\bar{h}$.

[8]This is best seen by considering the effect on a firm's sales $Q(h)$ of a small increase in C. From (4) and (6), $\frac{\mathrm{d}Q(h)}{\mathrm{d}C} = \frac{1}{2\gamma}\left[\bar{h} - h + C\frac{\mathrm{d}\bar{h}}{\mathrm{d}C}\right]$. From the free entry condition of (8), $\frac{\mathrm{d}\bar{h}}{\mathrm{d}C} = -2\gamma s/[C^2\int_0^{\bar{h}}(\bar{h}-h)g(h)\mathrm{d}h]$. It follows that $\frac{\mathrm{d}Q(h)}{\mathrm{d}C} > 0$ if and only if $(\bar{h}-h)\frac{C}{2\gamma}\int_0^{\bar{h}}(\bar{h}-h)g(h)\mathrm{d}h > s$. The expression on the left-hand side of this inequality is twice the firm's markup times *ex-ante* expected sales. Since there are zero expected *ex-ante* profits, for firms near the top of the productivity distribution (those with the lowest values of h and thus highest markup), twice their *ex-post* markup times *ex-ante* expected sales must be higher than the sunk entry cost, and this inequality holds, so their sales increase as city size increases. For firms near the bottom of the productivity distribution (those with a value of h close to the cutoff $\bar{h}$), the inequality fails to hold, so their sales fall as city size increases.

[9]To facilitate visual comparisons, we re-scale the combined size of the large and small cities from panel to panel to keep the sets C and W for the small city constant across cases, thus making the distribution of firms' log productivities in the small city identical in all four panels. This is done for the purpose of plotting the graph only, and does not change the qualitative comparison between the small-city and large-city distributions. These graphs are drawn using a log-normal distribution, but recall that our analytical results are distribution-independent. We use a log-normal distribution for the graphs because it matches well the empirically-observed distributions presented later in the paper.

being exposed to a wider range of ideas than workers in the small city and this makes them more productive. As a result, all large-city firms achieve higher log productivity than their small-city counterparts (i.e., log productivity ϕ is higher in the large city for every h). Since product-market competition is global, all firms can sell to consumers everywhere and this eliminates differences between cities in the strength of the firm selection mechanism. Hence, the log productivity cut-off $\ln[a(W)] - \ln(\bar{h})$ simply moves rightwards to the same extent as the rest of the log productivity distribution. Consequently the large-city log productivity distribution is simply a right-shifted version of the small-city distribution (again, see Lemma 1 below for a formal proof). Thus, agglomeration acts like the tide that lifts all boats.

Case 3 (*local product-market competition and local interactions*)**.** Panel (c) in Figure 1 plots the distribution of firms' log productivities in a city with a large population (continuous line) and in a city with a small population (dashed line) in the case where firms only sell in their local market and workers only interact with workers in their city (i.e., $C_1 > C_2$ and $W_1 > W_2$). Compared with the distribution in the small city, the large-city distribution is both left-truncated and right-shifted (again, see Lemma 1 below for a formal proof). With local product-market competition, large-city markups are lower and this left-truncates the distribution of firms' log productivities to exactly the same extent as under case 1. With local interactions, large-city workers are more productive and this right-shifts the distribution of firms' log productivities (truncated by firm selection) to exactly the same extent as under case 2.[10]

Case 4 (*global product-market competition and global interactions*)**.** When every firm competes with the same intensity with firms from everywhere and every worker enjoys interactions with the same intensity with workers from everywhere (i.e., $C_1 = C_2$ and $W_1 = W_2$), the distribution of firms' log productivities in a city with a large population is exactly the same as in a city with a small population. Panel (d) in Figure 1 plots the probability density function of the distribution of firms' log productivities, $f(\phi)$, in this final case where it is independent of city size. Note that the fact that the distribution of firms' log productivities does not depend on city size in this case does not imply that there are no selection or agglomeration effects. It simply implies that selection and agglomeration effects are equally strong everywhere. If there were no selection or agglomeration effects, the distribution of firm productivities would be given by $f(\phi) = e^{-\phi}\, g\left(e^{-\phi}\right)$. Relative to this underlying distribution, the actual distribution of firms' log productivities in both cities is both left-truncated and right-shifted.

[10]The absence of interactions between selection and agglomeration mechanisms is a consequence of having kept the assumption of quasi-linear preferences of Melitz and Ottaviano (2008), which eliminates income effects in the market for differentiated varieties. The introduction of income effects would create an interaction between agglomeration and firm selection that would result in further left truncation of the large-city log productivity distribution. This is because, with income effects, the log productivity advantages of agglomeration would translate into a larger market for differentiated varieties in the large city. This would reinforce the increase in local product-market competition caused by the larger population, and strengthen firm selection. Thus, with income effects, agglomeration would appear as a right shift in the log productivity distribution, while selection as well as interactions between selection and agglomeration would appear as a left truncation. More complicated interactions between selection and agglomeration mechanisms would also appear if the benefits from agglomeration for each worker varied depending on which firm they were working for. See Section 6.

Parameterising the strength of selection and agglomeration

For expositional clarity, we have so far focused on the polar possibilities of either local or global product-market competition and either local or global interactions. We now generalise our analysis to also consider intermediate cases for both. We do so by parameterising both the spatial decay of product market competition, creating differences in firm selection across cities, and the spatial decay of interactions, creating differences in agglomeration economies across cities. As before, we compare the distribution of log productivities across cities of different size. Suppose we have I cities. Let us denote the population of city i by N_i and order cities from largest to smallest in terms of population: $N_1 > N_2 > \cdots > N_{I-1} > N_I$.

In the case of product-market competition, we can introduce additional costs for selling differentiated varieties in a different city. Suppose that markets are segmented and that selling outside the city where a firm is located involves iceberg trade costs so that τ (> 1) units need to be shipped for one unit to arrive at destination. Since firms now potentially sell in all cities, the free entry condition of equation (8) becomes

$$\frac{N_i}{4\gamma}\int_0^{\bar{h}_i}(\bar{h}_i - h)^2 g(h)\mathrm{d}h + \sum_{j\neq i}\frac{N_j}{4\gamma}\int_0^{\bar{h}_j/\tau}(\bar{h}_j - \tau h)^2 g(h)\mathrm{d}h = s\ , \tag{13}$$

for city i. The first term on the left-hand side captures operational profits from local sales and the second-term summation the operational profits from out-of-city sales. Note that only city i firms with marginal costs $h < \bar{h}_j/\tau$ sell in city j, where $\bar{h}_j$ is the cutoff for local firms in j, since city i firms must be able to cover not just production but also trade costs. Note also that the cases of purely local or purely global product-market competition discussed above can still be captured as particular cases. The case of local product-market competition corresponds to $\tau = \infty$, which turns equation (13) into equation (8) with $C = N_i$. The case of global product-market competition corresponds to $\tau = 1$, which turns equation (13) into equation (8) with $C = \sum_{j=1}^{I} N_j$. In addition, we can now also consider intermediate cases where $1 < \tau < \infty$.

Regarding interactions, we can think of these as being subject to some spatial decay. Specifically, let us redefine the relevant argument for the interactions function $a(\cdot)$ as the sum of local population and outside population, with the latter adjusted by some decay factor as in Fujita and Ogawa (1982) and Lucas and Rossi-Hansberg (2002). This implies that the effective labour supplied by an individual worker in city i is $a(N_i + \delta\sum_{j\neq i} N_j)$, where the decay parameter δ measures the strength of across-city relative to within-city interactions ($0 < \delta < 1$). From equation (9), the log productivity of a firm with marginal cost h in city i is given by $\phi = \ln\left[a(N_i + \delta\sum_{j\neq i} N_j)\right] - \ln(h)$. Thus, the gain in log productivity due to interactions in city i (a local measure of the strength of agglomeration) of equation (11) can be redefined as

$$A_i \equiv \ln\left[a(N_i + \delta\sum_{j\neq i} N_j)\right]\ . \tag{14}$$

The case of local interactions discussed above corresponds to $\delta = 0$, which implies $A_i = \ln[a(N_i)]$. The case of global interactions discussed above corresponds to $\delta = 1$, which implies that $A_i = \ln[a(\sum_{j=1}^{I} N_j)]$. In addition, we can now also consider intermediate cases where $0 < \delta < 1$.

The distribution of firms' log productivities still has its probability density function given by equation (10), which, using subindex i to specify the city, becomes

$$f_i(\phi) = \begin{cases} 0 & \text{for } \phi < A_i - \ln(\bar{h}_i) \ , \\ \frac{e^{A_i - \phi} g\left(e^{A_i - \phi}\right)}{G(\bar{h}_i)} & \text{for } \phi \geqslant A_i - \ln(\bar{h}_i) \ . \end{cases} \tag{15}$$

In anticipation of the econometric approach developed in the next section, it is also useful write the corresponding cumulative density function, $F_i(\phi)$. To do that compactly, we need to introduce some additional notations. Let

$$S_i \equiv 1 - G(\bar{h}_i) \tag{16}$$

denote the proportion of firms that fail to survive product-market competition in city i (a local measure of the strength of selection). To further simplify notation, let us define

$$\tilde{F}(\phi) \equiv 1 - G\left(e^{-\phi}\right) \tag{17}$$

as the underlying cumulative density function of log productivities we would observe in all cities in the absence of any selection ($\bar{h}_i = \infty$, $\forall i$) and in the absence of any agglomeration ($A_i = 0$, $\forall i$). Without selection ($\bar{h}_i = \infty$, $\forall i$) all entrants survive regardless of their draw of h. Without agglomeration ($A_i = 0$, $\forall i$), $\phi = -\ln(h)$. Equivalently, $h = e^{-\phi}$. Using the change of variables theorem then yields (17) above. We can then write the cumulative density function of the distribution of log productivities for active firms in city i as

$$F_i(\phi) = \max\left\{0, \frac{\tilde{F}(\phi - A_i) - S_i}{1 - S_i}\right\} \ . \tag{18}$$

Relative to the underlying distribution given by (17), agglomeration shifts the distribution rightwards by A_i while selection eliminates a share S_i of entrants (those with lower productivity values). The next section develops an econometric approach to estimate the relative magnitude across cities of agglomeration, as measured by A_i, and selection, as measured by S_i. The following proposition contains our main theoretical result, with predictions for how these expressions vary across cities of different sizes.

Proposition 1. Suppose there are I cities ranked from largest to smallest in terms of population: $N_1 > N_2 > \cdots > N_{I-1} > N_I$, that workers are equally productive in any local firm, that interactions across cities decay by a factor δ, where $0 \leqslant \delta \leqslant 1$, and that selling in a different city raises variable costs by a factor τ, where $1 \leqslant \tau \leqslant \infty$.

i. Agglomeration leads to the distribution of firms' log productivities being right-shifted by A_i, and if $\delta < 1$ this right shift is greater the larger a city's population: $A_1 > A_2 > \ldots > A_{I-1} > A_I$.

ii. Firm selection left-truncates a share S_i of the distribution of firms' log productivities, and if $\tau > 1$ this truncation is greater the larger a city's population: $S_1 > S_2 > \ldots > S_{I-1} > S_I$.

iii. If there is no decay in interactions across cities, so that $\delta = 1$, then there are no differences in shift across cities: $A_i = A_j$, $\forall i, j$. If there is no additional cost incurred when selling in a different city, so that $\tau = 1$, then there are no differences in truncation across cities: $S_i = S_j$, $\forall i, j$.

Proof Consider any two areas i and j such that $i < j$ (and thus $N_i > N_j$). The extent of the right shift is A_i in city i and A_j in city j. Using equation (14), $0 \leqslant \delta < 1$ directly implies $A_i > A_j$, and $\delta = 1$ implies $A_i = A_j$. Turning to selection, the proportion of truncated values of $\tilde{F}$ is S_i in city i and S_j in city j. The free entry condition (13) for cities i and j can be rewritten:

$$\frac{N_i}{4\gamma}\int_0^{\bar{h}_i}(\bar{h}_i - h)^2 g(h)\mathrm{d}h + \frac{N_j}{4\gamma}\int_0^{\bar{h}_j/\tau}(\bar{h}_j - \tau h)^2 g(h)\mathrm{d}h + \sum_{k\neq i, k\neq j}\frac{N_j}{4\gamma}\int_0^{\bar{h}_j/\tau}(\bar{h}_j - \tau h)^2 g(h)\mathrm{d}h = s\ , \tag{19}$$

$$\frac{N_j}{4\gamma}\int_0^{\bar{h}_j}(\bar{h}_j - h)^2 g(h)\mathrm{d}h + \frac{N_i}{4\gamma}\int_0^{\bar{h}_i/\tau}(\bar{h}_i - \tau h)^2 g(h)\mathrm{d}h + \sum_{k\neq i, k\neq j}\frac{N_j}{4\gamma}\int_0^{\bar{h}_j/\tau}(\bar{h}_j - \tau h)^2 g(h)\mathrm{d}h = s\ . \tag{20}$$

Subtracting equation (20) from (19) and simplifying yields:

$$N_i\, \nu(\bar{h}_i, \tau) = N_j\, \nu(\bar{h}_j, \tau)\ . \tag{21}$$

where

$$\nu(z, \tau) \equiv \int_0^z (z - h)^2 g(h)\mathrm{d}h - \int_0^{z/\tau}(z - \tau h)^2 g(h)\mathrm{d}h\ . \tag{22}$$

It follows from (21) and $N_i > N_j$ that

$$\nu(\bar{h}_i, \tau) < \nu(\bar{h}_j, \tau)\ . \tag{23}$$

Differentiating (22) with respect to z yields:

$$\begin{aligned}\frac{\partial \nu(z,\tau)}{\partial z} &= 2\left[\int_0^z (z-h)g(h)\mathrm{d}h - \int_0^{z/\tau}(z - \tau h)g(h)\mathrm{d}h\right] \\ &= 2\left[(\tau - 1)\int_0^{z/\tau} h g(h)\mathrm{d}h + \int_{z/\tau}^z (z-h)g(h)\mathrm{d}h\right]\ .\end{aligned} \tag{24}$$

If $1 < \tau \leqslant \infty$, then $\partial\nu(z,\tau)/\partial z > 0$, and thus, by equation (23), $\bar{h}_i < \bar{h}_j$. Hence, by equation (16), $S_i > S_j$. If $\tau = 1$, then by equation (24), $\partial\nu(z,\tau)/\partial z = 0$, and thus $\bar{h}_i = \bar{h}_j$ and $S_i = S_j$. □

3. Econometric approach

We now develop an econometric approach to estimate the parameters that quantify the importance of selection and agglomeration in the theoretical model for cities of different sizes. The observable information is the cumulative distribution of log productivities in each city. Ideally, we would like to use this information to estimate parameters A_i and S_i from equation (18) for each city. However, this is not possible because the baseline cumulative of log productivities $\tilde{F}$ is not observed. Nevertheless, the following lemma shows that we can get around this issue by comparing the distribution of log productivities across two cities of different sizes i and j to difference out $\tilde{F}$ from equation (18).

Lemma 1. Consider two distributions with cumulative density functions F_i and F_j. Suppose F_i can be obtained by shifting rightwards by A_i some underlying distribution with cumulative density function $\tilde{F}$ and left-truncating a share $S_i \in [0,1)$ of its values:

$$F_i(\phi) = \max\left\{0, \frac{\tilde{F}(\phi - A_i) - S_i}{1 - S_i}\right\} . \tag{25}$$

Suppose F_j can be obtained by shifting rightwards by a different value $A_j \neq A_i$ the same underlying distribution $\tilde{F}$ and left-truncating a different share $S_j \neq S_i$ of its values:

$$F_j(\phi) = \max\left\{0, \frac{\tilde{F}(\phi - A_j) - S_j}{1 - S_j}\right\} . \tag{26}$$

Let

$$A \equiv A_i - A_j , \tag{27}$$

$$S \equiv \frac{S_i - S_j}{1 - S_j} . \tag{28}$$

If $S_i > S_j$, then F_i can also be obtained by shifting F_j by A and left-truncating a share S of its values:

$$F_i(\phi) = \max\left\{0, \frac{F_j(\phi - A) - S}{1 - S}\right\} . \tag{29}$$

If $S_i < S_j$, then F_j can also be obtained by shifting F_i rightwards by $-A$ and left-truncating a share $\frac{-S}{1-S}$ of its values:

$$F_j(\phi) = \max\left\{0, \frac{F_i(\phi + A) - \frac{-S}{1-S}}{1 - \frac{-S}{1-S}}\right\} . \tag{30}$$

Proof See the Appendix. □

We are going to use (29) and (30) to get an econometric specification that can be estimated from the data. An advantage of our approach is that we do not need to specify an ad-hoc underlying distribution of log productivities $\tilde{F}$, which one cannot observe empirically. A limitation is that we are not able to separately identify A_i, A_j, S_i and S_j from the data, but only $A = A_i - A_j$ and $S = (S_i - S_j)/(1 - S_j)$. In other words, we are able to make statements about the relative strength of agglomeration economies in large cities compared to small cities and about the relative strength of firm selection in large cities compared to small cities, but not about the absolute strength of agglomeration economies or firm selection. Parameter A measures how much stronger is the right shift (induced by agglomeration economies in the theoretical model) in city i relative to the smaller city j. In particular, it corresponds to the difference between cities i and j in the strength of agglomeration-induced productivity gains. Note that our empirical approach also allows for the possibility that $A < 0$, in which case there would be less rather than more right shift in larger cities. Parameter S measures how much stronger is the left truncation (induced by firm selection in the theoretical model) in city i relative to the smaller city j. In particular, it corresponds to the difference between cities i and j in the share of entrants eliminated by selection, relative to share of surviving entrants in city j. Note that our empirical approach also allows for the possibility that $S < 0$, in which case there would be less rather than more left truncation in larger cities.

A quantile specification

To obtain the key relationship to be estimated, we rewrite the two equations (29) and (30) in quantiles and combine them into a single expression. Assuming that $\tilde{F}$ is invertible, F_i and F_j are also invertible. We can then introduce $\lambda_i(u) \equiv F_i^{-1}(u)$ to denote the uth quantile of F_i and $\lambda_j(u) \equiv F_j^{-1}(u)$ to denote the uth quantile of F_j. If $S > 0$, equation (29) applies and can be rewritten as

$$\lambda_i(u) = \lambda_j\left(S + (1-S)u\right) + A \ , \quad \text{for } u \in [0,1] \ . \tag{31}$$

If $S < 0$, equation (30) applies and can be rewritten as

$$\lambda_j(u) = \lambda_i\left(\frac{u-S}{1-S}\right) - A \ , \quad \text{for } u \in [0,1] \ . \tag{32}$$

Making the change of variable $u \to S + (1-S)\,u$ in (32), this becomes

$$\lambda_j\left(S + (1-S)\,u\right) = \lambda_i\left(u\right) - A \ , \quad \text{for } u \in \left[\frac{-S}{1-S}, 1\right] \ . \tag{33}$$

We can then write the following equation that combines (31) and (33):

$$\lambda_i(u) = \lambda_j\left(S + (1-S)u\right) + A \ , \quad \text{for } u \in \left[\max\left(0, \frac{-S}{1-S}\right), 1\right] \ . \tag{34}$$

Equation (34) cannot be directly used for the estimation because the set of ranks $\left[\max\left(0, \frac{-S}{1-S}\right), 1\right]$ depends on the true value of S, which is not known. We thus make a final change of variable $u \to r_S(u)$, where $r_S(u) = \max\left(0, \frac{-S}{1-S}\right) + \left[1 - \max\left(0, \frac{-S}{1-S}\right)\right] u$, which transforms (34) into

$$\lambda_i\left(r_S(u)\right) = \lambda_j\left(S + (1-S)r_S(u)\right) + A \ , \quad \text{for } u \in [0,1] \ . \tag{35}$$

Equation (35) provides the key relationship that we wish to fit to the data. It states how the quantiles of the log productivity distribution in a large city i are related to the quantiles of the log productivity distribution in a small city j via the relative agglomeration/shift parameter A and the relative selection/truncation parameter S.

A suitable class of estimators

To estimate A and S, we use the infinite set of equalities given by (35) which can be rewritten in more general terms as $m_\theta(u) = 0$ for $u \in [0,1]$, where $\theta = (A, S)$ and

$$m_\theta(u) = \lambda_i\left(r_S(u)\right) - \lambda_j\left(S + (1-S)r_S(u)\right) - A \ . \tag{36}$$

We turn to a class of estimators studied by Gobillon and Roux (2008) who adapt to an infinite set of equalities the results derived by Carrasco and Florens (2000) for an infinite set of moments. Let $\hat{m}_\theta(u)$ denote the empirical counterpart of $m_\theta(u)$, where the true quantiles λ_i and λ_j have been replaced by some estimators $\hat{\lambda}_i$ and $\hat{\lambda}_j$ (see the appendix for details on how these estimators are constructed). We can then introduce an error minimization criterium based on a quadratic norm of functions, following Carrasco and Florens (2000). Let $\mathcal{L}^2$ denote the set of $[0,1]^2$ integrable

functions, $\langle \cdot,\cdot \rangle$ denote the inner product such that for any functions y and z in $\mathcal{L}^2$, we have: $\langle y,z \rangle = \int_0^1 \int_0^1 y(u)z(v)\mathrm{d}u\mathrm{d}v$, and $\| \cdot \|$ denote the corresponding norm. Consider a linear bounded operator B on $\mathcal{L}^2$. Let B^* denote its self-adjoint, such that we have: $\langle By,z \rangle = \langle y,B^*z \rangle$. Then, B^*B can be defined through a weighting function $\ell(\cdot,\cdot)$ such that: $(B^*By)(v) = \int_0^1 y(u)\ell(v,u)\mathrm{d}u$ and thus $\|By\|^2 = \int_0^1 \int_0^1 y(u)\ell(v,u)y(v)\mathrm{d}u\mathrm{d}v$. Let $n = (n_i,n_j)'$, where n_i and n_j denote respectively the number of observations of the distributions F_i and F_j. The vector of parameters θ can then be estimated as

$$\hat{\theta} = \arg\min_{\theta} \|B_n \hat{m}_\theta\| \ , \tag{37}$$

where B_n is a sequence of bounded linear operators.[11] Gobillon and Roux (2008) show that if $\hat{\lambda}_i$ and $\hat{\lambda}_j$ are some appropriate estimators that are consistent, asymptotically normal and continuous twice-differentiable, and if the sequence B_n is such that $\|B_n \hat{m}_\theta\| \to \|Bm_\theta\|$ as $\min(n_i,n_j) \to +\infty$, where B is a bounded linear operator verifying $Bm_\theta = 0 \Longrightarrow m_\theta = 0$, then the estimator $\hat{\theta}$ defined by (37) is consistent and asymptotically normal. They also show how to determine the weights $\ell(v,u)$ leading to the optimal estimator.

Implementation

While it is possible to compute the weights leading to the optimal estimator, they cannot be used in practice because they depend on the true value of the parameters θ. Alternatively, we could rely on a simple weighting scheme such that $\ell(v,u) = 0$ for $u \neq v$ and $\ell(v,v) = \delta_d$ where δ_d is a Dirac mass. With this weighting scheme, the estimator would simplify to:

$$\hat{\theta} = \arg\min_{\theta} \left(\int_0^1 [\hat{m}_\theta(u)]^2 \mathrm{d}u \right) \ . \tag{38}$$

This estimator is the mean-square error on m_θ. However, it has the undesirable feature that it treats the quantiles of the two distributions asymmetrically. In particular, it compares the quantiles of the actual city i log productivity distribution to the quantiles of a left-truncated and right-shifted city j distribution, when it would also be possible to compare the quantiles of the actual city j distribution to the quantiles of a modified city i distribution. We thus implement a more robust estimation procedure that treats the quantiles of the two distributions symmetrically. As a first step, we derive an alternative set of equations to (35) for this reverse comparison. Making the change of variable $u \to \frac{u-S}{1-S}$ in (31), this becomes

$$\lambda_j(u) = \lambda_i \left(\frac{u-S}{1-S} \right) - A \ , \quad \text{for } u \in [S,1] \ . \tag{39}$$

We can then write the following alternative equation to (34) that combines (32) and (39):

$$\lambda_j(u) = \lambda_i \left(\frac{u-S}{1-S} \right) - A \ , \quad \text{for } u \in [\max(0,S),1] \ . \tag{40}$$

[11] The following mild assumption is made to ensure that the model described by $m_\theta(u) = 0$ for $u \in [0,1]$ is identified: there exist K ranks (as many as parameters we wish to estimate) $u_i, \ldots, u_K$ such that the system $m_\theta(u_i) = 0$ for $i = 1, \ldots, K$ admits a unique solution in θ.

Let $\tilde{r}_S(u) = \max(0, S) + [1 - \max(0, S)]\, u$. With a final change of variable $u \to \tilde{r}_S(u)$ on (40), this provides a new set of equalities $\tilde{m}_\theta(u) = 0$, for $u \in [0, 1]$, where

$$\tilde{m}_\theta(u) = \lambda_j\left(\tilde{r}_S(u)\right) - \lambda_i\left(\frac{\tilde{r}_S(u) - S}{1 - S}\right) + A\ . \tag{41}$$

Let $\hat{\tilde{m}}_\theta(u)$ denote the empirical counterpart of $\tilde{m}_\theta(u)$, where the true quantiles λ_i and λ_j have been replaced by some estimators $\hat{\lambda}_i$ and $\hat{\lambda}_j$ (see the appendix for details). The estimator we actually use is then

$$\hat{\theta} = \arg\min_\theta M(\theta)\ , \quad \text{where} \quad M(\theta) = \int_0^1 \left[\hat{m}_\theta(u)\right]^2 \mathrm{d}u + \int_0^1 \left[\hat{\tilde{m}}_\theta(u)\right]^2 \mathrm{d}u\ . \tag{42}$$

In the results below, we report $\hat{\theta} = (\hat{A}, \hat{S})$, as well as a measure of goodness of fit $R^2 = 1 - \frac{M(\hat{A},\hat{S})}{M(0,0)}$. Standard errors of the estimated parameters are bootstrapped drawing some observations out of the log productivity distribution with replacement. For each bootstrap iteration, we first re-estimate TFP for each observation employed in the iteration, and we then re-estimate θ. Finally, we use the distribution of estimates of θ that results from all bootstrap iterations to compute the standard errors.

4. Data and TFP estimations

Data

To construct our data for 1994–2002, we merge together three large-scale, French, administrative data sets from the French national statistical institute (INSEE).

The first is BRN-RSI ('Bénéfices Réels Normaux' and 'Régime simplifié d'imposition') which contains annual information on the balance sheet of all French firms, declared for tax purposes. We extract information about each firm's output and use of intermediate goods and materials to compute a reliable measure of value added for each firm and year. We also retain information about the value of productive and financial assets to compute a measure of capital. This is done using the sum of the reported book values at historical costs. The sector of activity at the three-digit level is also available and a unique identifier for each firm serves to match these data with the other two data sets.

The second data set is SIREN ('Système d'Identification du Répertoire des ENtreprises') which contains annual information on all French private sector establishments, excluding finance and insurance. From this data set, we retain the establishment identifier, the identifier of the firm to which the establishment belongs (for matching with BRN-RSI), and the municipality where the establishment is located. We allocate each municipality to its metropolitan area ('Aire Urbaine') when it is part of one.

The third data set is DADS ('Déclarations Annuelles de Données Sociales'), a matched employer-employee data set, which is exhaustive during the study period. This includes the number of paid hours for each employee in each establishment and her two digit occupational category, which allows us to take labour quality into account. The procedure of Burnod and Chenu (2001) is

then used to aggregate total hours worked at each establishment by workers in each of three skill groups: high, intermediate and low skills.

To sum up, for each firm and each year between 1994 and 2002, we know the firm's value added, the value of its capital, and its sector of activity. For each establishment within each firm, we know its location, and the number of hours worked by its employees by skill level.[12] We retain information on all establishments from all firms with 6 employees or more in all manufacturing sectors and in business services, with the noted exception of finance and insurance. We end up with data on 153,130 firms and 203,300 establishments observed at least once during the study period.

TFP *estimation*

For simplicity of exposition, we have set up the model of Section 2 so that labour is the only input. However, all results extend trivially to a model with capital and workers with multiple skill levels, provided technology is homothetic, capital costs are equal at all locations, and from the point of view of an individual firm multiple types of workers are perfect substitutes (up to a scaling factor to capture the impact of skills on efficiency units).

For the purpose of estimation, we assume more specifically that the technology to generate value added at the firm level (V_t) is Cobb-Douglas in the firm's capital (k_t) and labour (l_t), and use t to index time (years). We also allow for three skilled levels, and use $l_{s,t}$ to denote the share of the firm's workers with skilled level s:

$$V_t = (k_t)^{\beta_1} \left(l_t \sum_{s=1}^{3} \varsigma_s l_{s,t} \right)^{\beta_2} e^{\beta_{3,t}+\phi_t} , \tag{43}$$

where β_1, β_2 and the three ς_s are common to all firms within a sector, $\beta_{3,t}$ varies by detailed subsector of that sector, and ϕ_t is firm-specific. Taking logs yields

$$\ln(V_t) = \beta_1 \ln(k_t) + \beta_2 \ln(l_t) + \beta_2 \ln \left(\sum_{s=1}^{3} \varsigma_s l_{s,t} \right) + \beta_{3,t} + \phi_t . \tag{44}$$

To linearise (44), we use the approximation in Hellerstein, Neumark, and Troske (1999). If the share of labour with each skill does not vary much over time or across firms within each sector, so that $l_{s,t} \approx \zeta_s$, then

$$\beta_2 \ln \left(\sum_{s=1}^{3} \varsigma_s l_{s,t} \right) \approx \beta_2 \left[\ln \left(\sum_{s=1}^{3} \varsigma_s \zeta_s \right) - 1 \right] + \sum_{s=1}^{3} \sigma_s l_{s,t} , \tag{45}$$

where $\sigma_s \equiv \beta_2 \varsigma_s / (\sum_{s=1}^{3} \varsigma_s \zeta_s)$. Substituting equation (45) into (44) yields:

$$\ln(V_t) = \beta_{0,t} + \beta_1 \ln(k_t) + \beta_2 \ln(l_t) + \sum_{s=1}^{3} \sigma_s l_{s,t} + \phi_t , \tag{46}$$

where $\beta_{0,t} \equiv \beta_{3,t} + \beta_2 [\ln(\sum_{s=1}^{3} \varsigma_s \zeta_s) - 1]$.

We obtain log TFP by estimating equation (46) separately for each sector in level 2 of the Nomenclature Economique de Synthèse (NES) sectoral classification, which leaves us with 16

[12] The merged data set contains much more information than is usually available. For instance, US-based research relies either on sectoral surveys or on five-yearly censuses for which value added is difficult to compute. We instead have exhaustive annual data. We also have information on the number of hours worked by skill level instead of total employment as is often the case.

manufacturing sectors and business services.[13] We let $\beta_{0,t}$ be the sum of a year-specific component and a sector-specific component at level 3 of the NES classification (which contains 63 subsectors for our base 16 sectors). Denote by $\hat{\beta}_{0,t}$, $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\sigma}_s$ the estimates of $\beta_{0,t}$, β_1, β_2 and σ_s, respectively. Let $\hat{\phi}_t = \ln(V_t) - \hat{\beta}_{0,t} - \hat{\beta}_1 \ln(k_t) - \hat{\beta}_2 \ln(l_t) - \sum_{s=1}^{3} \hat{\sigma}_s l_{s,t}$. We then measure log TFP for each firm by the firm-level average of $\hat{\phi}_t$ over the period 1994–2002,

$$\hat{\phi} = \frac{1}{T} \sum_{t=1}^{T} \hat{\phi}_t \ , \tag{47}$$

where T denotes the number of years the firm is observed in 1994–2002.

For our baseline results, we estimate equation (46) using ordinary least squares (OLS). Later, we report as robustness checks the results obtained with the methods proposed by Olley and Pakes (1996) and Levinsohn and Petrin (2003) to account for the potential endogeneity of capital and labour, as well as simple cost share estimates of TFP. Details on how TFP estimates are constructed in our context using these methods are relegated to the Appendix.

Since data for value added and capital is only available at the firm level, in the baseline results we restrict the sample to firms with a single establishment (which account for 85 percent of firms, 68 percent of establishments, and 46 percent of employment). Later, we take advantage of establishment-level data on hours worked by skill and report as robustness checks results for all firms, including those with establishments in multiple locations. We do so by estimating the following relationship between each firm's log TFP and the set of cities where it has establishments, separately for each sector:

$$\hat{\phi} = \sum_{i=1}^{I} \nu_i l_i + \epsilon \ , \tag{48}$$

where i indices metropolitan areas (there are 364 in France), and l_i denotes the share of a firm's labour (in hours worked) in area i, averaged over the period 1994–2002. Parameter ν_i is common to all firms and establishments in area i. Let $\hat{\nu}_i$ be the OLS estimate of ν_i and $\hat{\epsilon} = \hat{\phi} - \sum_{i=1}^{I} \hat{\nu}_i l_i$. Establishment-level log TFP is then computed as $\hat{\nu}_i + \hat{\epsilon}$. Note that for firms with a single establishment, $\hat{\nu}_i + \hat{\epsilon} = \hat{\phi}$ as before.

5. Baseline results

The quantile approach described in Section 3 estimates the amount of left truncation and right shift that, when applied to one distribution of firms' log productivities, best approximate another distribution of firms' log productivities. To implement the approach, after estimating TFP for mono-establishment firms as described in Section 4 using OLS, we must choose which two distributions to compare. For our baseline estimates, we lump urban areas together based on their population size. In particular, we compare the distribution of firms' log productivities in urban

[13] Unfortunately, we cannot include banking and insurance in our estimation because the location of establishments is not available for these sectors, which have distinct reporting rules. We also exclude distribution and consumer services from our main estimations. The assignment of a specific location to distribution (which involves moving goods across locations) is difficult and the estimation of a production function in consumer services is more problematic (but see the bottom-right panel of Figure 2 for an illustration from consumer services).

Table 1: Baseline estimation results, cities with pop.> 200,000 vs. pop.< 200,000

Sector	OLS, mono-establishments							
	A	S	R^2	A	R^2	S	R^2	obs.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Food, beverages, tobacco	0.06	0.01	0.90	0.07	0.82	0.03	0.40	22,049
	(0.01) *	(0.00) *		(0.01) *		(0.01) *		
Apparel, leather	0.09	-0.04	0.42	0.05	0.18	-0.01	0.06	5,804
	(0.04) *	(0.06) *		(0.01) *		(0.00) *		
Publishing, printing, recorded media	0.20	-0.03	0.85	0.17	0.73	0.00	0.00	9,236
	(0.01) *	(0.01) *		(0.01) *		(0.02)		
Pharmaceuticals, perfumes, soap	0.09	-0.04	0.76	0.02	0.01	-0.02	0.58	1,069
	(0.05)	(0.04)		(0.04)		(0.01)		
Domestic appliances, furniture	0.14	0.00	0.82	0.14	0.81	0.03	0.09	6,362
	(0.01) *	(0.01)		(0.01) *		(0.02) *		
Motor vehicles	0.10	-0.01	0.54	0.08	0.41	0.00	0.02	1,442
	(0.02) *	(0.02)		(0.02) *		(0.01)		
Ships, aircraft, railroad equipment	0.11	-0.02	0.75	0.08	0.45	-0.01	0.08	1,016
	(0.03) *	(0.01)		(0.03) *		(0.02)		
Machinery	0.08	-0.01	0.97	0.08	0.93	0.01	0.03	14,736
	(0.01) *	(0.00) *		(0.01) *		(0.01)		
Electric and electronic equipment	0.08	-0.01	0.97	0.08	0.94	0.02	0.03	5,749
	(0.01) *	(0.00)		(0.01) *		(0.02)		
Building materials, glass products	0.07	0.00	0.83	0.06	0.81	0.00	0.04	3,196
	(0.02) *	(0.01)		(0.01) *		(0.01)		
Textiles	0.06	0.00	0.61	0.06	0.57	0.00	0.00	3,365
	(0.02) *	(0.01)		(0.02) *		(0.01)		
Wood, paper	0.10	-0.01	0.91	0.09	0.85	0.01	0.02	5,872
	(0.01) *	(0.01)		(0.01) *		(0.01)		
Chemicals, rubber, plastics	0.09	0.00	0.95	0.09	0.95	0.01	0.14	5,337
	(0.01) *	(0.01)		(0.01) *		(0.01)		
Basic metals, metal products	0.08	0.00	0.97	0.08	0.96	0.01	0.07	14,305
	(0.01) *	(0.00)		(0.01) *		(0.01) *		
Electric and electronic components	0.08	0.00	0.94	0.08	0.93	0.01	0.15	2,579
	(0.02) *	(0.01)		(0.02) *		(0.01)		
Consultancy, advertising, business services	0.19	-0.02	0.95	0.17	0.88	0.09	0.06	37,041
	(0.01) *	(0.01) *		(0.01) *		(0.03) *		
All sectors	0.11	-0.02	0.71	0.09	0.57	0.00	0.00	139,143
	(0.00) *	(0.00) *		(0.00) *		(0.00) *		

areas with over 200,000 people with the corresponding distribution in urban areas with less than 200,000 people and rural areas. Later, we report as robustness checks results comparing urban areas using alternative population size classes, results comparing individual urban areas, results using employment areas instead or urban areas, split by employment density instead of population size, and results using alternative methods to estimate TFP.[14]

Columns (1) and (2) in Table 1 show our baselines estimates of A and S for two-digit manufacturing and business service sectors. Recall that $S \equiv \frac{S_i - S_j}{1 - S_j}$, where i corresponds to large cities and j corresponds to small cities. If $S > 0$, then the strength of selection increases with city size. If $S = 0$, then the strength of selection does not vary with city size. Recall also that $A \equiv A_i - A_j$. If $A > 0$, then the strength of agglomeration increases with city size. If $A = 0$, then the strength of agglomeration does not vary with city size. Column (3) in Table 1 reports a pseudo-R^2 as defined in Section 3.

In column (1), A is always positive. Statistical significance at the 5 percent level is marked with an asterisk next to the bootstrapped standard errors reported in parenthesis. A is significantly different from zero in all cases but one. For all sectors it takes a value $A = 0.11$, which implies a 12 percent productivity increase. These results suggest that agglomeration economies are stronger in large cities than in small cities. Our model shows that the extent to which agglomeration economies vary across cities of different size is closely related to the extent to which interactions are local or global (national in this case). Our results are consistent with a situation where interactions are quite local. This matches the empirical literature looking at the spatial decay of different types of agglomeration economies (see Rosenthal and Strange, 2004). We postpone the discussion of the economic significance of our findings until later.

For 11 sectors out of 16, S is not statistically different from zero. It is negative and significant in four sectors and for all sectors pooled together.[15] It is positive and significant in one sector only. In all cases, however, S remains small. This suggests that there is not much difference between large and small cities in the strength of selection. Note that this does not imply that selection is not important. It simply suggests that its importance is similar in cities of different sizes. Our model shows that the extent to which selection varies across cities of different sizes is closely related to the extent to which product market competition is local or global (national in this case). Our results are consistent with a situation where French firms compete with similar intensity on national markets regardless of their location.

Column (4) in Table 1 reports our estimates of A when we impose the restriction $S = 0$ (no difference in the strength of selection between large and small cities). Not surprisingly given how close to zero the estimates of S in column (2) are, the estimates of A in column (4) are very close

[14]Whenever one estimates firm-level TFP, measurement errors are likely to result in a few extreme outliers. To minimise the impact of such outliers in our estimates of truncation and shift, we exclude the 1 percent of observations with the highest TFP values and the 1 percent of observations with the lowest TFP values in each city size class. It is important to trim extreme values in both city size classes to avoid biasing the estimate of S. While our estimates lose some precision when we trim fewer outliers, results are qualitatively unchanged.

[15]Having negative estimates of S, even if small in absolute value, may seem strange in light of the model. We argue in Section 7 that it is an artifact of the simple way in which we capture agglomeration economies in these baseline results. As we shall see, once we generalise our approach, the estimates of S become zero.

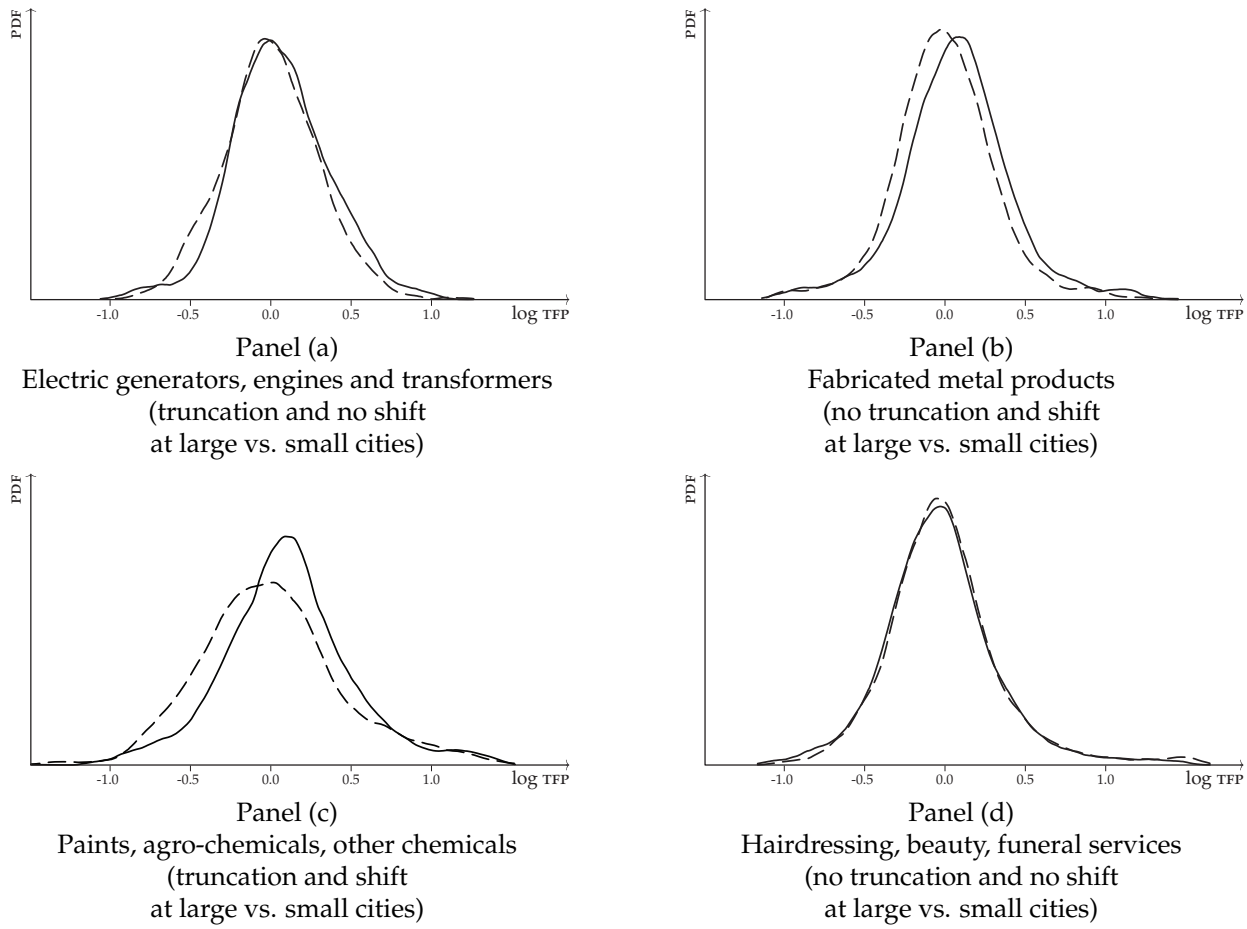
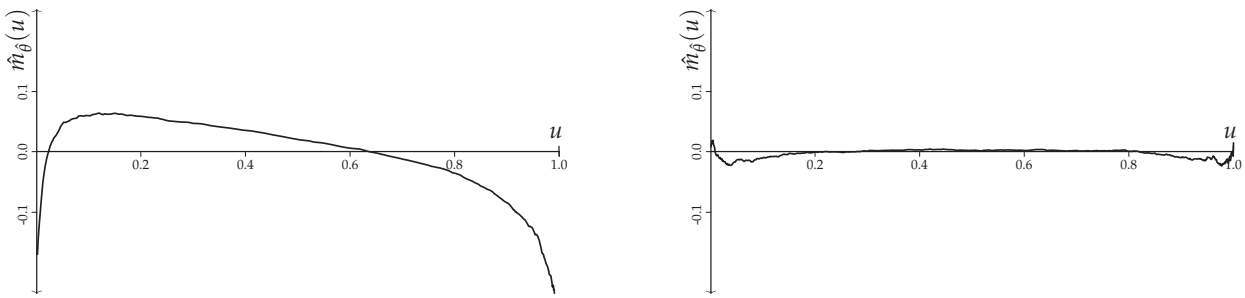


Figure 2: Empirical log productivity distributions in large (solid) and small cities (dashed)

to those in column (1). Column (5) reports the corresponding pseudo-R^2 and shows that for most sectors the fit does not deteriorate too much relative to column (3).

Column (6) in Table 1 reports our estimates of S when we impose the restriction $A = 0$ (no difference in the strength of agglomeration between large and small cities). In each and every case the estimate for S in column (6) is larger than or equal to its corresponding estimate in column (2). This suggests that if we do not allow agglomeration to vary across cities of different sizes, we pick up part of the agglomeration effects as variation in selection. Column (7) reports the pseudo-R^2 under the restriction $A = 0$. A comparison with column (3) shows that fit deteriorates very substantially in all sectors but one. Overall, the results of columns (4)-(7) reinforce those of columns (1)-(3) by underscoring the robustness of our finding that agglomeration economies are stronger in large cities than in small cities and the absence of significant differences in selection effects.[16]

[16]Other estimation methods for TFP, such as the procedures developed by Olley and Pakes (1996) and Levinsohn and Petrin (2003) and cost shares lead to similar conclusions. Considering all establishments as opposed to only mono-establishment firms also yields very similar results. Finally, our results are not affected by trimming 2 percent instead of 1 percent of observations at both extremes of the log productivity distribution. We do not report these results for our baseline estimation. They are available upon request. Below we report results using these alternative estimation methods and alternative samples of establishments to assess the robustness of our main results.



Panel (a)
Baseline results (A and S)

Panel (b)
Main results (A, D and S)

Figure 3: Estimation errors by quantile

We also find broadly similar results for more finely defined sectors, i.e., sectors as defined by the level 3 of the NES classification. There are two differences however. First, because these more disaggregated sectors are sometimes small, differences between small and large cities are less frequently statistically significant than with two-digit sectors. Second, although the general trend in the data is stronger agglomeration in larger cities and no difference in selection, looking at more finely defined sectors shows clearly that this pattern is not universal. Figure 2 represents the distributions of log productivities in small and large cities for four three-digit sectors. These closely match the four theoretical patterns of Figure 1. Electric generators, engines and transformers (Panel a) is a rare example of a sector exhibiting stronger selection effects in large cities but no difference in agglomeration effects. Fabricated metal products (Panel b) illustrates the opposite and more common pattern of no difference in selection and stronger agglomeration in large cities. Paints, agro-chemicals, and other chemicals (Panel c) is a case of both stronger agglomeration and stronger selection effects in large cities. Finally, hairdressing, beauty, and funeral services (Panel d) provides, as might have been expected, an example of a sector where the distribution of firms' log productivities is almost exactly the same in small and large cities.

While the pseudo-R^2 of column (3) is high (including above 0.90 in a majority of sectors), looking at the fit of the estimation in more detail suggests that our baseline analysis does not fully explain the differences between the distributions of log productivities in large and small cities. Panel (a) in Figure 3 (ignore Panel b for now) provides some insight into what is missing from our model and empirical approach as we have presented it so far. The graph plots, for all sectors combined (bottom row of Table 1), the values of $\hat{m}_{\hat{\theta}}(u)$. That is, the Figure plots for each quantile (given by a point on the horizontal axis) the difference between its value in the large city distribution and the value that results from shifting and truncating the small-city log-productivity distribution using the estimated values of A and S. There is very marked pattern, where errors tend to be positive for the lower quantiles and negative for the higher quantiles. This indicates that, by forcing all establishments to have the same productivity boost from locating in a large city, we are giving establishments at the lower end of the productivity distribution too large a boost and establishments at the upper end of the productivity distribution too small a boost. In other words, the figure indicates that, contrary to what we have assumed so far, more productive establishments

benefit more from agglomeration. We now extend our model and empirical approach to allow for this.[17]

6. When more productive establishments benefit more from agglomeration

So far we have taken the simple view that agglomeration economies raise the log productivity of all establishments in larger cities by the same amount. We now generalise our theoretical and empirical frameworks to allow the magnitude of agglomeration economies to be systematically related not just to city size but also to individual productivity. In particular, we conjecture that, while agglomeration economies raise the productivity of all firms in larger cities, they raise the productivity of the most productive firms to a greater extent.

Extending the model

Let us thus relax the assumption that workers are equally productive regardless of the firm they work for. Suppose instead that workers are more productive when they work for a more efficient firm (i.e., one with a lower h) and that this effect is enhanced by interactions with other workers. In particular suppose that the effective units of labour supplied by an individual worker in their unit working time are $a(N_i + \delta \sum_{k \neq i} N_k) h^{-(D_i - 1)}$, where

$$D_i \equiv \ln \left[d(N_i + \delta \sum_{k \neq i} N_k) \right] , \tag{49}$$

$d(0) = 1$, $d' > 0$ and $d'' < 0$ (the model seen up until this point was equivalent to assuming $D_i = 1$). In this case, the natural logarithm of the productivity of a firm with unit cost h in city i is given by

$$\phi = \ln \left(\frac{Q}{l} \right) = A_i - D_i \ln(h) . \tag{50}$$

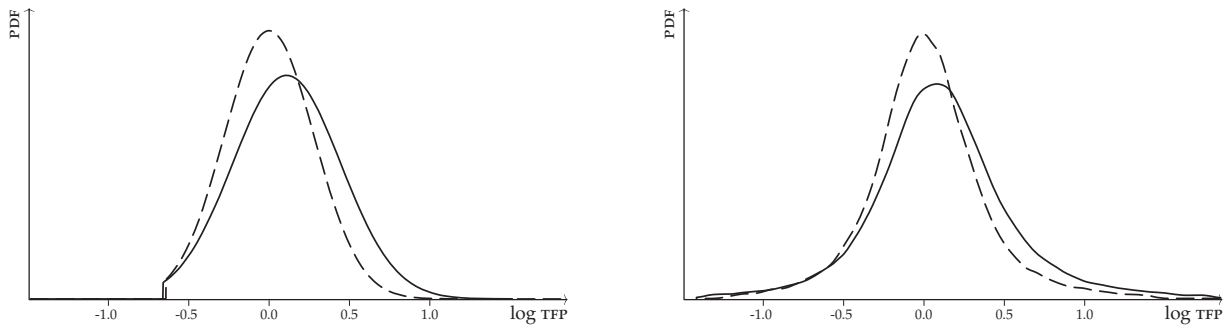
We can then write the cumulative density function of the distribution of log productivities for active firms in city i as

$$F_i(\phi) = \max \left\{ 0, \frac{\tilde{F}\left(\frac{\phi - A_i}{D_i}\right) - S_i}{1 - S_i} \right\} . \tag{51}$$

Relative to the underlying log productivity distribution $\tilde{F}$, agglomeration both dilates the distribution by a factor D_i and shifts the distribution rightwards by A_i, while selection eliminates a share S_i of entrants (those with lower productivity values). Proposition 1 can then be rewritten as follows.

Proposition 1′. Suppose there are I cities ranked from largest to smallest in terms of population: $N_1 > N_2 > \cdots > N_{I-1} > N_I$, that workers are more productive when they work for a more

[17] It is also worth noting that, regardless of the general downward-sloping error pattern, errors for the first few quantiles are negative before quickly becoming positive above the first 2 percent of quantiles. This is a sign that the small negative value estimated for S ($S = -0.02$) leads to a very bad fit at the very bottom of the distribution even if it helps improve the overall fit. We return to this issue below, where we argue that not allowing more productive establishments to get an additional productivity boost in large cities results in a downward bias for S as well as in an upward bias for A.



Panel (a)
Theory: same selection and stronger agglomeration at large vs. small city, full model

Panel (b)
Data: OLS TFP estimates for all sectors Pop.> 200,000 vs. Pop.< 200,000

Figure 4: Log-productivity distributions in large (solid) and small cities (dashed)

efficient (lower h) firm and that this effect is enhanced by interactions, that interactions across cities decay by a factor δ, where $0 \leqslant \delta < 1$, and that selling in a different city raises variable costs by a factor τ, where $1 \leqslant \tau \leqslant \infty$.

i. Agglomeration leads to the distribution of log productivities being dilated by a factor D_i and right-shifted by A_i, and if $\delta < 1$ this dilation and right shift are both greater the larger a city's population: $D_1 > D_2 > \ldots > D_{I-1} > D_I$ and $A_1 > A_2 > \ldots > A_{I-1} > A_I$.

ii. Firm selection left-truncates a share S_i of the distribution of log productivities, and if $\tau > 1$ this truncation is greater the larger a city's population: $S_1 > S_2 > \ldots > S_{I-1} > S_I$.

iii. If there is no decay in interactions across cities, so that $\delta = 1$, then there are no differences in dilation nor in shift across cities: $D_i = D_j$ and $A_i = A_j$, $\forall i, j$. If there is no additional cost incurred when selling in a different city, so that $\tau = 1$, then there are no differences in truncation across cities: $S_i = S_j$, $\forall i, j$.

Proof Consider any two areas i and j such that $i < j$ (and thus $N_i > N_j$). The dilation factor is D_i in cities i and D_j in city j while the extent of the right shift is A_i in city i and A_j in city j. If $0 \leqslant \delta < 1$, by equation (49), $D_i > D_j$ and, by equation (14), $A_i > A_j$. If instead $\delta = 1$, by the same two equations, $D_i = D_j$ and $A_i = A_j$. The proportion of truncated values of $\tilde{F}$ is S_i in city i and S_j in city j. The free entry conditions of equations (19) and (20) still apply. If $1 < \tau \leqslant \infty$, by equation (24), $\bar{h}_i < \bar{h}_j$ and thus, by equation (16), $S_i > S_j$. If $\tau = 1$, by the same two equations, $\bar{h}_i = \bar{h}_j$ and thus $S_i = S_j$. □

Panel (a) of Figure 4 re-draws Panel (b) of Figure 1 for the full model. It plots the distribution of firms' log productivities in a city with a large population (solid line) and in a city with a small population (dashed line) with global product-market competition and local interactions (i.e., when there are stronger agglomeration economies in large cities than in small cities but the same strength of selection in both). The difference with respect to Panel (b) of Figure 1 is that more productive firms now benefit even more from locating in large cities, so that stronger agglomeration economies get reflected in both a right shift and a dilation of the log-productivity distribution. This can be

seen graphically in the peak for the large city being lower and in the gap between the distributions getting larger as we move towards the right. Panel (b) of Figure 4 plots the actual distribution of log productivities in large cities (urban areas with over 200,000 people, solid line) and small cities (urban areas with less than 200,000 people and rural areas, dashed line) for all sectors together using OLS TFP estimates. We can see it looks remarkably similar to the theoretical benchmark in Panel (a) of Figure 4. To show more formally that the extended model is better at capturing differences in productivity between cities of different sizes, we now extend our econometric approach.

Extending the econometric approach

We now show how to incorporate dilation into our econometric approach. The following generalization of Lemma 1 shows that, by comparing once again the distribution of log productivities across two cities of different sizes i and j, we can difference out $\tilde{F}$ from equation (51).

Lemma 1′. Consider two distributions with cumulative density functions F_i and F_j. Suppose F_i can be obtained by dilating by a factor D_i and shifting rightwards by A_i some underlying distribution with cumulative density function $\tilde{F}$ and also left-truncating a share $S_i \in [0,1)$ of its values:

$$F_i(\phi) = \max\left\{0, \frac{\tilde{F}\left(\frac{\phi - A_i}{D_i}\right) - S_i}{1 - S_i}\right\} . \tag{52}$$

Suppose F_j can be obtained by dilating by a different factor $D_j \neq D_i$ and shifting rightwards by a different value $A_j \neq A_i$ the same underlying distribution $\tilde{F}$ and also left-truncating a different share $S_j \neq S_i$ of its values:

$$F_j(\phi) = \max\left\{0, \frac{\tilde{F}\left(\frac{\phi - A_j}{D_j}\right) - S_j}{1 - S_j}\right\} . \tag{53}$$

Let

$$D \equiv \frac{D_i}{D_j} , \tag{54}$$

$$A \equiv A_i - DA_j , \tag{55}$$

$$S \equiv \frac{S_i - S_j}{1 - S_j} . \tag{56}$$

If $S_i > S_j$, then F_i can also be obtained by dilating F_j by D, shifting it by A, and left-truncating a share S of its values:

$$F_i(\phi) = \max\left\{0, \frac{F_j\left(\frac{\phi - A}{D}\right) - S}{1 - S}\right\} . \tag{57}$$

If $S_i < S_j$, then F_j can also be obtained by dilating F_j by $\frac{1}{D}$, shifting it rightwards by $-\frac{A}{D}$ and left-truncating a share $\frac{-S}{1-S}$ of its values:

$$F_j(\phi) = \max\left\{0, \frac{F_i\left(D\phi + A\right) - \frac{-S}{1-S}}{1 - \frac{-S}{1-S}}\right\} . \tag{58}$$

Proof See the Appendix. □

To estimate the set of parameters $\theta = (A, D, S)$, we first rewrite (57) and (58) in quantiles. If $S > 0$, equation (57) applies and can be rewritten as

$$\lambda_i(u) = D\lambda_j\left(S + (1-S)u\right) + A\ , \quad \text{for } u \in [0,1]\ . \tag{59}$$

If $S < 0$, equation (58) applies and can be rewritten as

$$\lambda_j(u) = \frac{1}{D}\lambda_i\left(\frac{u-S}{1-S}\right) - \frac{A}{D}\ , \quad \text{for } u \in [0,1]\ . \tag{60}$$

Performing on (57) and (58) exactly the same steps we performed on (29) and (30) to obtain (36) and (41) yields:

$$m_\theta(u) = \lambda_i\left(r_S(u)\right) - D\lambda_j\left(S + (1-S)r_S(u)\right) - A\ , \tag{61}$$

$$\tilde{m}_\theta(u) = \lambda_j\left(\tilde{r}_S(u)\right) - \frac{1}{D}\lambda_i\left(\frac{\tilde{r}_S(u) - S}{1-S}\right) + \frac{A}{D}\ . \tag{62}$$

The estimator we use is still given by (42), where $\hat{m}_\theta(u)$ and $\hat{\tilde{m}}_\theta(u)$ still denote the empirical counterparts of $m_\theta(u)$ and $\tilde{m}_\theta(u)$, now redefined in (61) and (62).

7. Main empirical results

For our main results, we rely again on TFP estimated as described in Section 4. Relative to our baseline results, we now estimate how the distribution of firms' log productivities in large cities is best approximated by shifting, dilating and truncating the distribution of firms' log productivities in small cities. We now estimate a shift parameter, A, a dilation parameter, D, and a truncation parameter, S, comparing large cities (urban areas with over 200,000 people) and small cities (urban areas with less than 200,000 people and rural areas) for 16 manufacturing and business service sectors and all sectors together.

Columns (1), (2), and (3) of Table 2 present our main results for A, D, and S. Recall that, in the extended model that serves as a basis for these main results, greater agglomeration economies in large cities result in the distribution of log productivities in large cities being both right shifted and dilated relative to the distribution in small cities, i.e., in $A > 0$ and $D > 1$. The value of A corresponds to the average increase in log productivity that would arise in large cities relative to small cities absent any selection.[18] When $A > 0$, values of D above unity are evidence that agglomeration economies in large cities benefit more the more productive firms, whereas values of D below unity would indicate that agglomeration economies benefit less the more productive firms. As in the baseline, positive values of S correspond to the distribution of firms' log productivities in large cities being more truncated than in small cities, whereas negative values correspond to more truncation in small cities.

[18]Note that we normalise our log-TFP estimates so that our estimates of A in Table 2 are directly comparable with those of Table 1. This involves choosing units of value added so that average log-TFP in small cities is zero, which affects neither D nor S.

Table 2: Main estimation results, cities with pop.> 200,000 vs. pop.< 200,000

Sector	OLS, mono-establishments				
	A	D	S	R^2	obs.
	(1)	(2)	(3)	(4)	(5)
Food, beverages, tobacco	0.07	0.94	0.00	0.95	22,049
	(0.00) *	(0.02) *	(0.00)		
Apparel, leather	0.04	1.37	0.01	0.98	5,804
	(0.01) *	(0.05) *	(0.01)		
Publishing, printing, recorded media	0.17	1.26	0.00	0.98	9,236
	(0.01) *	(0.04) *	(0.00)		
Pharmaceuticals, perfumes, soap	0.05	1.19	-0.01	0.91	1,069
	(0.05)	(0.11)	(0.04)		
Domestic appliances, furniture	0.13	1.22	0.01	0.98	6,362
	(0.01) *	(0.04) *	(0.01) *		
Motor vehicles	0.08	1.29	0.01	0.80	1,442
	(0.03) *	(0.14) *	(0.03)		
Ships, aircraft, railroad equipment	0.10	1.09	-0.01	0.81	1,016
	(0.03) *	(0.18)	(0.03)		
Machinery	0.08	1.04	0.00	0.98	14,736
	(0.01) *	(0.03)	(0.00)		
Electric and electronic equipment	0.08	1.02	0.00	0.97	5,749
	(0.01) *	(0.05)	(0.01)		
Building materials, glass products	0.06	1.06	0.00	0.89	3,196
	(0.02) *	(0.06)	(0.02)		
Textiles	0.05	1.14	0.00	0.92	3,365
	(0.01) *	(0.08) *	(0.01)		
Wood, paper	0.09	1.11	0.00	0.99	5,872
	(0.01) *	(0.05) *	(0.01)		
Chemicals, rubber, plastics	0.08	1.04	0.00	0.96	5,337
	(0.01) *	(0.04)	(0.01)		
Basic metals, metal products	0.08	1.06	0.00	1.00	14,305
	(0.01) *	(0.02) *	(0.00)		
Electric and electronic components	0.08	0.99	0.00	0.94	2,579
	(0.02) *	(0.07)	(0.03)		
Consultancy, advertising, business services	0.19	1.05	-0.01	0.96	37,041
	(0.02) *	(0.04)	(0.03) *		
All sectors	0.09	1.22	0.00	1.00	139,143
	(0.00) *	(0.01) *	(0.00)		

In column (1) of Table 2, A is always positive and, like in Table 1, it is significant at 5 percent in all cases but one. The only difference with Table 1 is that the estimates for A are now slightly lower. For instance, when considering all sectors, we find $A = 0.09$ in Table 2 whereas $A = 0.11$ for the baseline in Table 1. Column (2) in Table 2 reports our estimates of D. In eight sectors D is statistically different from unity. For seven of these sectors (and for all sectors together), D is above unity. In only one of these sectors, D is below unity. In the other sectors, D is not statistically different from unity although the point estimates are usually above one. There is thus a tendency for the distribution of firms' log productivities to be more dilated in larger cities for about half the sectors and for all sectors combined. With $A > 0$, the finding that D is often above unity is indicative that it is the most productive firms that benefit the most from agglomeration. For all sectors, $A = 0.09$ and $D = 1.22$ imply that firms are on average 9 percent more productive in large cities but that this productivity advantage is 14 percent for firms at the first quartile and only 5 percent for firms at the bottom quartile.[19]

Turning to S in column (2), there is only one case of a sector, domestic appliances and furniture, with a positive and significant value for S, although this value is small at 0.01. There is also only one case,consultancy, advertising, and business services, with a negative and significant value for S, again small at -0.01. In all other cases, the estimated value of S is not significantly different from zero. This lack of significance is not due to imprecise estimates. On the contrary, the standard errors for S are small, like the standard errors for A. Adding to this, we note that in 11 cases out of 17 (including all sectors combined), the estimated value for S is precisely 0.00. These results provide even stronger evidence than our baseline results that there are no differences between small and large cities in the truncation of the distribution of firms' log productivities. Market selection appears to have a similar intensity across cities in France irrespective of their size.

To summarise, firms are more productive in large cities. However, this is not because tougher competition makes it more difficult for the least productive firms to survive. The productivity advantages of large cities arise because agglomeration economies boost the productivity of all firms, and in about half of the sectors this increase in productivity is strongest for the most productive firms.

More generally, a comparison between Tables 1 and 2 suggests that it is important to allow for more productive firms to benefit more from agglomeration. When one fails to do so, as in our baseline results of Table 1, estimates of A and S become biased as they attempt to approximate a dilation. In particular, when we do not allow for $D > 1$, we tend to overestimate A and underestimate S (the latter even becoming negative in several cases). It is also clear from the comparison of Tables 1 and 2 that the fit is better when considering A, D, and S instead of only A and S. Unsurprisingly, the improvement in the fit is strongest for those sectors with strong dilation. For instance, in Apparel and leather, the pseudo-R^2 goes from 0.42 to 0.98 when adding D to the estimation. Overall, the fit in column (4) of Table 2 is very good. The pseudo-R^2 is always above

[19]By shifting, dilating and truncating the small city distribution using the estimated values of A, D, and S, we obtain a predicted productivity advantage of 4.6 percent for firms at the bottom quartile and 14.1 percent for firms at the top quartile. In the empirical distribution for large cities, we find that these advantages are 4.7 percent and 13.9 percent, respectively. For the bottom and top deciles of the large city distribution and relative to the small city distribution, we find productivity advantages of 0.9 percent and 21.1 percent, respectively.

Table 3: Robustness, cities with pop.> 200,000 vs. pop.< 200,000, alternative estimation methods and alternative samples

Method	A	D	S	R^2	obs.
	(1)	(2)	(3)	(4)	(5)
	all sectors, mono-establishments				
Ordinary Least Squares	0.09 (0.00) *	1.22 (0.01) *	0.00 (0.00)	1.00	139,143
Olley-Pakes	0.08 (0.01) *	1.10 (0.04) *	0.00 (0.00) *	0.98	56,920
Levinsohn-Petrin	0.09 (0.00) *	1.10 (0.01) *	0.00 (0.00)	1.00	101,714
Cost shares	0.07 (0.00) *	1.20 (0.01) *	0.00 (0.00)	0.98	139,143
	all sectors, all establishments				
Ordinary Least Squares	0.09 (0.00) *	1.12 (0.00) *	0.00 (0.00) *	1.00	199,235
Olley-Pakes	0.10 (0.00) *	1.10 (0.00) *	0.01 (0.00) *	0.99	97,246
Levinsohn-Petrin	0.11 (0.00) *	1.04 (0.00) *	0.00 (0.00) *	0.93	155,106
Cost shares	0.08 (0.00) *	1.10 (0.00) *	0.00 (0.00) *	1.00	199,235

0.80 and it is even above 0.95 in a majority of sectors and very close to 1.00 for all sectors combined.

The importance of accounting for the greater benefit of agglomeration economies for large firms is perhaps most clearly seen when one compares the plots of estimation errors by quantile in the two panels of Figure 3. Panel (a) corresponds to our baseline results, and it was the clear downward-sloping pattern in this panel that lead us to add parameter D to the estimation. Panel (b) corresponds to our main results and plots the difference between the value of each quantile in the large city distribution and the value of the same quantile in the distribution that results from shifting, dilating, and truncating the small-city log-productivity distribution using the estimated values of A, D, and S (the estimated value of S being in fact zero, thus leading to no truncation). Estimation errors are greatly reduced relative to those of Panel (a) and the clear downward-sloping pattern apparent in Panel (a) is gone. In fact errors in Panel (b) are almost uniformly zero except for a little wiggle at the both extremes, where productivity values are more scattered and the fit between the distributions inevitably loses precision.

Robustness

To assess the strength of the findings of Table 2, we now turn to a series of robustness checks. First, one might question our TFP estimates. While OLS is arguably the most transparent method to estimate TFP, it does not account for the possible simultaneous determination of productivity

and factor usage. In the top half of Table 3, we report results for all sectors combined using four alternative methods to estimate TFP. The first line of results in Table 3 reports the same OLS results as in the last line of Table 2 to ease comparisons. The next line reports results for the same estimation of A, D, and S using the approach proposed by Olley and Pakes (1996) instead of OLS. The Olley-Pakes estimate of A is very similar to its corresponding OLS value, 0.08 instead of 0.09. The estimates of S are also very close tot 0.00 in both cases. The only difference when using Olley-Pakes is that the tiny amount of truncation (the estimated parameter is $S = 0.004$) is statistically significant whereas it is insignificant with OLS. Finally, there is a small difference for the dilation parameter D. It is equal to 1.10 when using Olley-Pakes against 1.22 with OLS. Overall the differences between OLS and Olley-Pakes TFP are small and could be due to the substantially smaller sample size with Olley-Pakes. The number of establishments used for the estimation drops from 139,143 with OLS to 56,920 with Olley-Pakes. This is due to the need to observe establishments over time in the latter to compute investment. Estimating TFP using the method proposed by Levinsohn and Petrin (2003) in the third line of results in Table 3 yields estimates that are very similar to those of Olley-Pakes TFP. The fourth line reproduces the estimation of A, D, and S when the underlying TFP is estimated using a simple cost-share approach.[20] The results are again very similar. While we do not report detailed sectoral results for these alternative TFP estimations, we note that they are close to the results reported in Table 2.

The bottom half of Table 3 replicates the same four estimations of A, D and S as the first panel but this time considering all establishments, affiliated and unaffiliated, instead of only mono-establishment firms. For OLS TFP, the results are very close to those with mono-establishment firms, except for less dilation in big cities. The next three lines report results for the alternative approaches to TFP estimations as described above. The results are very close to their corresponding results for mono-establishments firms in the first panel of the same table. They are also close to those obtained with OLS TFP. Overall we conclude that neither the sample of establishment we use nor the specific method we implement to estimate TFP have much bearing on our results.

Table 4 reports additional results using first a different zoning and then different urban groupings. The first line reproduces again our main results comparing urban areas with over 200,000 people and urban areas with less than 200,000 people and rural areas for all sectors together. The second line of results in that table repeats the same estimation of A, D, and S (OLS, all sectors, mono-establishments), but compares employment areas with above and below median employment density instead of urban areas with above and below 200,000 people. Employment area boundaries are drawn to capture local labour markets on the basis of commuting patterns whereas urban area boundaries are drawn to capture cities. While the total number of areas is roughly similar (341 contiguous employment areas instead of 364 urban areas and the rural areas that surround them), differences are substantial. For instance, Greater Paris is classified as a single urban area but is made up of 16 separate employment areas. Nevertheless, the estimated coefficients for A, D, and S are very similar. The fit is also very good. Table 5 in Appendix D reports

[20] We do not use the method proposed by Syverson (2004) using instrumented cost shares. This approach, which uses local demand shocks as instruments, is valid only for industries with very localised markets. It is not suitable for a broad cross-section of sectors nor when pulling all sectors together.

Table 4: Robustness, alternative comparisons

Comparison	OLS, all sectors, mono-establishments				
	A	D	S	R^2	obs.
	(1)	(2)	(3)	(4)	(5)
Cities with pop. > 200,000 vs. pop.< 200,000	0.09 (0.00) *	1.22 (0.01) *	0.00 (0.00)	1.00	139,143
Employment areas above vs. below median density	0.09 (0.00) *	1.21 (0.01) *	0.00 (0.00)	1.00	139,143
Paris vs. cities with pop. 1–2 million	0.12 (0.00) *	1.13 (0.02) *	0.00 (0.00)	0.98	47,480
Cities with pop. 1–2 million vs. pop. 200,000–1 million	0.04 (0.00) *	1.05 (0.02) *	0.00 (0.00)	0.95	37,443
Cities with pop. 200,000–1 million vs. pop. < 200,000	0.01 (0.00) *	1.09 (0.01) *	0.00 (0.00)	0.97	91,666
Paris vs. Lyon (pop. 10,381,376 vs. 1,529,824)	0.09 (0.01) *	1.13 (0.03) *	0.00 (0.01)	0.92	41,738
Lyon vs. Grenoble (pop. 1,529,824 vs. 486,022)	0.03 (0.01) *	1.05 (0.05)	0.00 (0.01)	0.62	6,925
Grenoble vs. Troyes (pop. 486,022 vs. 168,605)	0.05 (0.02) *	1.11 (0.09)	0.01 (0.02)	0.86	1,943

detailed, sector by sector, results for French employment areas to compare with those of Table 2. The results are very close. We conclude that using two groups of cities according to population size or two groups of employment areas according to employment density yields very similar results.

The next three lines of Table 4 return to the urban zoning but divide French cities and towns into four groups instead of just two. They are Paris, other cities with population above one million (Lyon, Marseille, and Lille), cities with population between 200,000 and one million, and the rest of the country. Starting with the estimate of A, average TFP for establishments located in Paris is about 12 percent higher than for establishments in other cities with a population above one million (the population of Paris is between six and ten times larger). There is also a sizeable gap between these three other large cities with population above one million and cities with population between 200,000 and one million (the estimate for A is about one third as large but then so is roughly the gap in terms of average population). Finally the gap between the last two groups is small but nonetheless statistically significant. Turning to the dilation parameter D, this remains significantly above unity when considering this more detailed grouping of cities. The distribution of firms' log productivities appears more dilated in larger cities than in smaller cities when considering any two consecutive groups of cities, capturing once again the larger boost for the most productive firms. Finally, regarding the truncation parameter S, our previous results are confirmed with no difference in market selection between any two consecutive groups of cities.

Grouping cities, as we have done so far, is useful because it ensures that we have enough observations to estimate parameters accurately and reduces the impact of idiosyncrasies associated with any particular city. Nevertheless, we now report the in last three lines of Table 4 results for

pairwise comparisons of individual cities that are illustrative of our general results. The four cities used in these comparisons are Paris (the largest, with a population above 10 million), Lyon (the second largest, with a population around 1.5 million), Grenoble (about half a million), Troyes (a smaller city, with a population below 200,000). Although the number of observations becomes small for the comparison between Grenobles and Troyes, the estimate of A remains significant. A trebling of population between Troyes and Grenoble or between Grenoble and Lyon is associated with a 3 to 5 percent increase in average TFP. The productivity gap reflected in the estimate of A for the comparison between Paris and Lyon is of the same magnitude, once we account for the fact that Paris is larger than Lyon by a factor of nearly seven. As would be expected in light of previous results, there are no differences in the strength of selection. Note also that the fit deteriorates in this last part of Table 4, since we must work with a much smaller number of observations for the last two rows comparing individual cities.

8. Discussion

The consequences of unobserved prices

As is often the case in the estimation of production functions, we do not observe prices and estimate TFP by studying how much value (instead of physical output) an establishment can produce with given inputs. Even if prices were observed, it would be unclear whether higher prices reflect higher price markups or higher quality products, that is the ability of firms to produce more value out of the same inputs.[21]

While we cannot solve this missing price problem (and the related quality issue), we now show that, to the extent it may affect our estimates, it cannot be driving our key results. In fact, if anything, it would work against us by leading us to underestimate the effects of agglomeration (greater shift and dilation in larger cities). The inability to observe prices implies that our estimated log productivities do not capture the physical quantity of output relative to inputs, but instead the value of output relative to inputs. Expressed in terms of our model, we do not measure ϕ, as given by equation (50), but instead

$$\psi = \ln\left(\frac{pQ}{l}\right) = \ln(p) + A_i - D_i \ln(h) = \ln(p) + \phi \ . \tag{63}$$

Thus, by not taking prices out, we are shifting log productivities rightwards by the value of log prices, $\ln(p)$. The problem is that log prices are systematically related both to city size (through $\bar{h}$) and to individual productivity (through h). Recall that, by equation (6), prices are given by $p = \frac{1}{2}(h + \bar{h})$.

[21] To solve these two problems, a first possibility is to focus on homogeneous goods for which quantities are directly observed (Syverson, 2004, Collard-Wexler, 2007, Foster, Haltiwanger, and Syverson, 2008). An alternative is to focus on specific industries with localised markets for which direct measures of quality are available, like newspapers and restaurants (Berry and Waldfogel, 2006). These two solutions can only be applied to a small number of industries. The last alternative in the literature is to consider detailed product-level information, including prices, to recover the price markup of firms and back up their 'true' productivity (De Loecker, 2007). With this approach, to disentangle whether higher prices reflect larger markups or superior quality, one still has to make specific assumptions about how quality is produced and about the functional form of demand (e.g., firms selling to consumers with a constant elasticity of substitution across products).

In terms of the relationship with city size,

$$\frac{\partial \ln(p)}{\partial \bar{h}} = \frac{1}{h + \bar{h}} > 0 \,. \tag{64}$$

If $\bar{h}$ differs across cities, then by looking at ψ instead of ϕ we are overestimating log productivities for every for h, but we are doing so by more in smaller cities (where $\bar{h}$ is then larger). Hence one consequence of not observing prices is that we may underestimate A, the parameter capturing the common shift in the log productivity distribution of large cities relative to small cities.

In terms of the relationship with individual productivity,

$$\frac{\partial^2 \ln(p)}{\partial \bar{h}\, \partial h} = -\frac{1}{(h + \bar{h})^2} < 0 \,. \tag{65}$$

Thus, if $\bar{h}$ differs across cities, the problem of underestimating log productivities in large versus small cities is greater for the most productive firms. Hence, another consequence of not observing prices is that we may underestimate D, the parameter capturing to what extent more productive firms get an extra productivity boost from locating in large cities.

To sum up, if $\bar{h}$ differs across cities (which necessarily implies $S \neq 0$), then by not observing prices we will underestimate the consequences of agglomeration (the shift captured by A and the dilation captured by D). However, recall that our results show that $S = 0$ for all sectors combined and for nearly all individual sectors. Note also that not observing prices does not affect the estimation of S (the value of log TFP at which each distribution might be left-truncated changes, but the share S of establishments in the small city distribution that are truncated out of the large city distribution does not). Thus, our finding that $S = 0$ also implies that estimating TFP through value added does not bias our estimates of A and D.

A comparison with approaches based on summary statistics

A key innovation of our approach is that we simultaneously consider agglomeration economies and market selection as possible causes of the productivity advantages of large cities using the entire distribution of firm log productivities. While this makes a comparison of our findings with those in the literature difficult, we now briefly relate our results to previous contributions studying either agglomeration economies or market selection alone on the basis of summary statistics.

Starting with Sveikauskas (1975), the empirical literature on agglomeration typically estimates the elasticity of some measure of average productivity, like average TFP, with respect to some measure of local scale, such as total population or employment density. More recent studies have paid particular attention to addressing two potential problems for identifying agglomeration economies with that approach. First, it is possible that the higher productivity of individual establishments is a cause rather than a consequence of larger or denser cities. This could happen because more productive cities tend to attract more workers and firms and, as a result, become larger and denser. It is also possible that size and productivity are simultaneously determined. This could happen if there is some local characteristic that is correlated with both size and productivity. Starting with Ciccone and Hall (1996), the standard way to tackle this potential problem is to use instrumental

variables when regressing average productivity on local size or density. The main finding is that reverse causality or simultaneity is only a minor issue in practice. A second identification issue is that more productive workers may sort into denser areas. This could happen if skilled workers have a stronger preference for the amenities typically found in larger cities or if larger cities provide greater productive benefits for more skilled workers. The standard way to deal with this issue is to use detailed data on worker characteristics or even to exploit a panel to incorporate individual fixed effects in a regression of individual wages on city size or density (Glaeser and Maré, 2001, Combes, Duranton, and Gobillon, 2008). This issue of endogenous labour quality turns out to be substantially more important in practice than endogenous labour quantity. In light of this, we take advantage of having information on the hours worked by each employee in each establishment and their detailed occupational code, to incorporate detailed labour quality into our TFP estimation.

We can compare the magnitude of agglomeration economies that we find with that of earlier studies looking at agglomeration alone. We do this by turning our estimate of A (the common shift in productivity of large city establishments relative to their small city counterparts) into an elasticity of average TFP with respect to density. An average employee in a French city with population above 200,000 works in a municipality 10.5 times as dense as an average person in the rest of the country. This ratio of 10.5 implies that our estimate of $A = 0.09$ for all sectors combined in Table 2 is equivalent to an elasticity of TFP with respect to employment density of 0.038. Using the same French data as we do but quantifying agglomeration economies by the usual approach of regressing mean TFP for French employment areas on employment density in those areas, Combes *et al.* (2009) find an elasticity of 0.035. More broadly this estimate of 0.038 is within the usual range of between 0.02 and 0.10. The main conceptual difference of our approach with respect to this literature on agglomeration is that we take into account that the higher average productivity in large cities could be caused by stronger selection as well as by stronger agglomeration economies.

Turning to market selection, existing approaches are harder to compare to ours. Our model shares with other models relating selection to market size, such as Syverson (2004) and Melitz and Ottaviano (2008), the prediction that tougher competition leads to a left truncation of the distribution of productivity in large cities relative to small cities. Unfortunately, detecting left truncation on the basis of summary statistics such as the mean or variance of firms' productivity is not straightforward. Greater left truncation increases average productivity, but so does any model of agglomeration. Both selection and agglomeration can also explain an increase in the median or the bottom decile of local productivity. In the model of Syverson (2004), left truncation also implies a decrease in the variance of productivities. We note that this result depends crucially on distributional assumptions.[22] Furthermore, it is possible that the strength of both selection and agglomeration increase with city size in certain sectors. Even if the shape of the distribution was such that truncation reduced dispersion, agglomeration could simultaneously increase dispersion through a dilation of the distribution, and thus make the separation of selection and agglomeration

[22]The result that the variance of productivity increases with left truncation holds in Syverson's model and, more generally, for productivity distributions with log-concave density. However, this result would be reversed if one considered instead a productivity distribution with log-convex density, such as the Pareto distribution commonly used in this literature (on the relationship between the variance of a left truncated distribution and log-concavity and log-convexity, see Heckman and Honore, 1990).

based on dispersion measures alone very difficult. A key difference of our approach is that we consider simultaneously selection and agglomeration and look at all the moments of the productivity distributions, so that we do not rely on particular distributional restrictions. Finally, Syverson focuses on one sector, ready-made concrete, chosen because of very particular characteristics. We look instead at a broad cross-section of sectors.[23]

Given these differences with existing approaches, a detailed comparison of results would not be informative. Instead we now ask how large selection effects would need to be in our data to generate the differences in average productivity that we observe in the absence of any agglomeration economies. To conduct this exercise we solve for S so as to equalise mean productivity in small and large cities with $A = 0$ and $D = 1$. We find that to explain a difference in mean log TFP of 0.09 between cities with population above and below 200,000 (as we have here for all sectors), S should be equal to 0.14. When doing the same calculation sector by sector we find that selection effects of similar magnitude would be needed to explain observed differences in mean productivity. Put differently, for selection effects to be the main force at play behind existing differences in average productivity across cities, they would need to be a full order of magnitude larger than our current estimates.

9. Concluding comments

To assess the relative importance of local agglomeration effects and market selection to explain the higher productivity of firms in larger cities, we nest a standard model of agglomeration with a generalised version of the firm selection model of Melitz and Ottaviano (2008). The main prediction of our model is that stronger selection in larger cities left-truncates the firms' productivity distribution while stronger agglomeration right-shifts and dilates it. A similar prediction would emerge from a much broader class of models nesting agglomeration and selection provided the underlying distribution of the firms' productivities is the same everywhere and selection effects can be separated from agglomeration effects. An important benefit of our structural approach is that it allows for a tight parametrisation of the strength of agglomeration effects relative to selection.

To implement this model on exhaustive French establishment-level data, we develop a new quantile approach that allows us to estimate a relative change in left truncation, shift, and dilation between two distributions. When applied to distributions of firms' log productivities, this quantile approach is fully consistent with our theoretical framework.

Our main finding is that productivity differences across urban areas in France are mostly explained by agglomeration. The distribution of firms' log productivities in large French cities is remarkably well described by taking the distribution of firms' productivities in small French cities, dilating it, and shifting it to the right. This holds for the productivity distributions of firms across all sectors as well as most two-digit sectors when considered individually. It also holds when comparing cities with population above and below 200,000, more finely subgroups of cities,

[23] The approach developed in Del Gatto *et al.* (2006) and Del Gatto *et al.* (2008) also differs significantly from ours. They make distributional assumptions about productivity and assess whether more open sectors exhibit a smaller dispersion of productivity.

and even individual cities. This finding is also robust to the choice of zoning since taking French employment areas instead of French urban yields similar results. Our bottom line is that, relative to the rest of the country, the distribution of firms log productivities in cities with population above 200,000 is shifted to the right by of 0.09 and dilated by a factor of 1.22. Firms in large cities are thus on average about 9 percent more productive than in small cities. Because of dilation, this productivity advantage is only of 5 percent for firms at the bottom quartile and 14 percent for firms at the top quartile. On the other hand we find no difference between small and large cities in terms of left truncation of the log productivity distribution.

These findings are interesting and raise a number of questions regarding future research. Most models of agglomeration economies can easily replicate a shift but far fewer imply a dilation (Duranton and Puga, 2004). In our model, dilation arises from a simple technological complementarity between the productivity of firms and that of workers. Such complementarity could arguably be generated from more subtle interactions between firms and workers (assuming for instance some heterogeneity among workers as well). Furthermore this type of complementarity might also have some interesting implications with respect to location choices for both firms and workers as well as implications regarding the dynamics of firm productivity and workers' career paths.

That there are no differences in market selection might seem surprising to some. The emphasis however should be on the word difference. The fact that firms' log productivity distributions all exhibit a positive skew would be consistent with some selection if the underlying distribution of productivity were symmetric (or negatively skewed). However such selection appears to take place everywhere in France with the same intensity. As shown by our model, this is consistent with the French market being highly integrated. Different findings could certainly emerge when comparing different countries. Furthermore, our finding of no difference in selection across places is consistent with the usual finding in the trade literature that trade liberalisation raises productivity mostly through selection. Poorly integrated markets might show big differences in the intensity of market selection whereas highly integrated markets might have very little. Any transition between these two states involves changes in selection. For instance, when a country liberalises its imports, many low productivity firms may be eliminated by stronger competition from foreign competitors. However, as trade liberalisation proceeds further, the toughness of competition and thus the strength of market selection will converge between the home and foreign countries. This end result of no large spatial differences in the strength of selection is what we find when comparing cities across France. At a very different spatial scale, we also suspect that for many consumer services selection could be very potent at a very fine level of aggregation such as the neighbourhood. A new hairdresser on a stretch of street is likely to affect other hairdressers along that stretch more than a new machine-tool producer will affect other machine-tool producers in the same city.

Appendix A. Proof of Lemma 1 and Lemma 1'

We now prove Lemma 1'. The proof of Lemma 1 is a particular case of this for $D_i = D_j = 1$ and hence $D = 1$. Consider first the case $S_i > S_j$. We apply the change of variables $\phi \to \frac{\phi - A}{D}$, which

turns equation (53) into

$$F_j\left(\frac{\phi - A}{D}\right) = \max\left\{0, \frac{\tilde{F}\left(\frac{\phi - A_i}{D_i}\right) - S_j}{1 - S_j}\right\} . \tag{A1}$$

Dividing by $1 - S$ and adding $\frac{-S}{1-S}$ to all terms in this equation yields

$$\frac{F_j\left(\frac{\phi - A}{D}\right) - S}{1 - S} = \max\left\{\frac{-S}{1-S}, \frac{\tilde{F}\left(\frac{\phi - A_i}{D_i}\right) - S_i}{1 - S_i}\right\} . \tag{A2}$$

Since, with $S_i > S_j$, $S > 0$, we have $\frac{-S}{1-S} < 0$, and we finally obtain

$$\max\left\{0, \frac{F_j\left(\frac{\phi - A}{D}\right) - S}{1 - S}\right\} = \max\left\{0, \frac{\tilde{F}\left(\frac{\phi - A_i}{D_i}\right) - S_i}{1 - S_i}\right\} = F_i(\phi) . \tag{A3}$$

Consider now the case $S_i < S_j$. We apply the change of variables $\phi \to D\phi + A$, which turns equation (52) into

$$F_i(D\phi + A) = \max\left\{0, \frac{\tilde{F}\left(\frac{\phi - A_j}{D_j}\right) - S_i}{1 - S_i}\right\} . \tag{A4}$$

Dividing by $1 - \frac{-S}{1-S}$ and adding S to all terms in this equation yields

$$\frac{F_i(D\phi + A) - \frac{-S}{1-S}}{1 - \frac{-S}{1-S}} = \max\left\{S, \frac{\tilde{F}\left(\frac{\phi - A_j}{D_j}\right) - S_j}{1 - S_j}\right\} . \tag{A5}$$

Since, with $S_i < S_j$, $S < 0$, we finally obtain

$$\max\left\{0, \frac{F_i(D\phi + A) - \frac{-S}{1-S}}{1 - \frac{-S}{1-S}}\right\} = \max\left\{0, \frac{\tilde{F}\left(\frac{\phi - A_j}{D_j}\right) - S_j}{1 - S_j}\right\} = F_j(\phi) . \tag{A6}$$

□

Appendix B. Implementation

In this appendix, we explain how we compute the minimisation criterium of equation (42), used to estimate the values of the parameters. We do so for our main results, including the dilation parameter D introduced in Section 6. For our baseline results, the implementation is the same with $D = 1$.

First note that the data consist of a set of log productivities in large cities (indexed by i) and in small cities (indexed by j), ranked in ascending order and denoted Φ_i and Φ_j respectively. From these data, for any θ, we need to be able to evaluate $\hat{m}_\theta(u)$ and $\hat{\tilde{m}}_\theta(u)$ at any ranks $u \in [0,1]$ to compute $M(\theta) = \int_0^1 [\hat{m}_\theta(u)]^2 \mathrm{d}u + \int_0^1 \left[\hat{\tilde{m}}_\theta(u)\right]^2 \mathrm{d}u$. For that purpose, we construct some estimators

$\hat{\lambda}_i$ and $\hat{\lambda}_j$ of the quantiles $\lambda_i(u)$ and $\lambda_i(u)$. Focusing on large cities (replace i with j for small cities), we start from the set of log productivities

$$\Phi_i = [\phi_i(0), \ldots, \phi_i(E_i - 1)]' \ , \tag{A7}$$

where E_i is the number of establishments in i and $\phi_i(0) < \ldots < \phi_i(E_i - 1)$. We can construct the sample quantiles at the observed ranks as $\hat{\lambda}_i \left(\frac{k}{E_i}\right) = \phi_i(k)$ for $k \in \{0, \ldots, E_i - 1\}$. For any other rank $u \in]0,1[$, the estimators of the quantiles are recovered by linear interpolation:

$$\hat{\lambda}_i\,(u) = (k_i^* + 1 - uE_i)\,\hat{\lambda}_i\left(\frac{k_i^*}{E_i}\right) + (uE_i - k_i^*)\,\hat{\lambda}_i\left(\frac{k_i^* + 1}{E_i}\right) \ , \tag{A8}$$

where $k_i^* = \lfloor uE_i \rfloor$ and $\lfloor . \rfloor$ denotes the integer part. From equation (A8) and the corresponding expression for j, we can use the empirical counterparts of equations (61) and (62),

$$\hat{m}_\theta(u) = \hat{\lambda}_i\,(r_S(u)) - D\hat{\lambda}_j\,(S + (1 - S)r_S(u)) - A \ , \tag{A9}$$

$$\hat{\tilde{m}}_\theta(u) = \hat{\lambda}_j\,(\tilde{r}_S(u)) - \frac{1}{D}\hat{\lambda}_i\left(\frac{\tilde{r}_S(u) - S}{1 - S}\right) + \frac{A}{D} \ , \tag{A10}$$

to compute $\hat{m}_\theta(u)$ and $\hat{\tilde{m}}_\theta(u)$ at any rank u and for any θ. We then consider $K = 1001$ ranks evenly distributed over the interval $[0,1]$. These ranks are denoted u_k, $k \in \{0, \ldots, K\}$, with $u_0 = 0$ and $u_K = 1$. We approximate the two subcriteria using the formulas:

$$\int_0^1 [\hat{m}_\theta(u)]^2 \mathrm{d}u \approx \frac{1}{2}\sum_{k=1}^{K}\left\{[\hat{m}_\theta(u_k)]^2 + [\hat{m}_\theta(u_{k-1})]^2\right\}(u_k - u_{k-1}) \ , \tag{A11}$$

$$\int_0^1 \left[\hat{\tilde{m}}_\theta(u)\right]^2 \mathrm{d}u \approx \frac{1}{2}\sum_{k=1}^{K}\left\{\left[\hat{\tilde{m}}_\theta(u_k)\right]^2 + \left[\hat{\tilde{m}}_\theta(u_{k-1})\right]^2\right\}(u_k - u_{k-1}) \ . \tag{A12}$$

The estimated parameters $\hat{\theta}$ are those which minimise the sum of these two quantities.

Appendix C. Implementation of alternative approaches to productivity

Olley-Pakes

In this appendix, we present three alternative approaches to TFP estimation. The first is the methodology proposed by Olley and Pakes (1996) to account for the endogeneity of production factors when estimating the parameters of equation (46). These authors consider that the residual ϕ_t can be decomposed into an unobserved factor φ_t which is potentially correlated with labour and capital, and an uncorrelated error term η_t such that: $\phi_t = \varphi_t + \eta_t$. They suppose that the unobserved factor φ_t can be rewritten as its projection on its lag and an innovation: $\varphi_t = \kappa\,(\varphi_{t-1}) + \xi_t$. They also make the crucial assumption that capital investment at time t depends on the capital stock and the unobserved factor φ_t: $I_t = i_t\,(k_t, \varphi_t)$. The function i_t is supposed to be strictly increasing in the unobserved factor. It can be inverted such that: $\varphi_t = f_t\,(k_t, I_t)$. Equation (46) can then be rewritten as:

$$\ln(V_t) = \beta_2 \ln(l_t) + \textstyle\sum_{s=1}^{3} \sigma_s l_{s,t} + \Psi_t\,(k_t, I_t) + \eta_t \ , \tag{A13}$$

where the auxiliary function Ψ_t is defined as

$$\Psi_t\left(k_t,I_t\right) = \beta_{0,t} + \beta_1 \ln(k_t) + f_t\left(k_t,I_t\right) \ . \tag{A14}$$

Equation (A13) can be estimated with OLS after $\Psi_t\left(k_t,I_t\right)$ has been replaced with a third-order polynomial crossing k_t, I_t and year dummies. This allows to recover some estimators of the labour and skill share coefficients ($\hat{\beta}_2$ and $\hat{\sigma}_s$), as well as the auxiliary function ($\hat{\Psi}_t$). It is then possible to construct the variable

$$v_t = \ln(V_t) - \hat{\beta}_2 \ln(l_t) - \sum_{s=1}^{3} \hat{\sigma}_s l_{s,t} \ . \tag{A15}$$

From equation (A14), the lagged value of the unobserved factor φ_{t-1} can be approximated by $\hat{\Psi}_{t-1}\left(k_{t-1},I_{t-1}\right) - \beta_{0,t-1} - \beta_1 \ln(k_{t-1})$. Using equations (A13), (A14), (A15), and the projection of the unobserved factor on its lag, the value-added equation then becomes:

$$v_t = \beta_{0,t} + \beta_1 \ln(k_t) + \kappa\left(\hat{\Psi}_{t-1}\left(k_{t-1},I_{t-1}\right) - \beta_{0,t-1} - \beta_1 \ln(k_{t-1})\right) + \vartheta_t \ , \tag{A16}$$

where ϑ_t is a random error. The function $\kappa(.)$ is approximated by a third-order polynomial and equation (A16) is estimated with non-linear least squares. We thus recover some estimators of the year dummies ($\hat{\beta}_{0,t}$) and the capital coefficients ($\hat{\beta}_1$). An estimator of ϕ_t is then given by $\hat{\phi}_t = v_t - \hat{\beta}_{0,t} - \hat{\beta}_1 \ln(k_t)$.

Although the Olley-Pakes method allows us to control for simultaneity, it has some drawbacks. In particular, we need to construct investment from the data: $I_t = k_t - k_{t-1}$. Since investment enters lagged into equation (A16), we must observe firms for at least three consecutive years to compute their TFP with this method. Other observations must be dropped. Furthermore, the investment equation $I_t = i_t(k_t,\varphi_t)$ can be inverted only if $I_t > 0$. Hence, we can keep only observations for which $I_t > 0$. This double selection may introduce a bias, for instance, if (i) there is greater 'churning' (i.e. entry and exits) in denser areas, and (ii) age and investment affect productivity positively. Then, more establishments with a low productivity may be dropped in high density areas. In turn, this may increase the measured difference in local productivity between areas of low and high density. Re-estimating OLS TFP on the same sample of firms used for Olley-Pakes shows that this is, fortunately, not the case on French data.

Levinsohn-Petrin

We also implement the approach proposed by Levinsohn and Petrin (2003). Its main difference with Olley and Pakes (1996) is that the quantity of inputs is used to account for the unobservables instead of investment. The unobserved factor is then rewritten as $\varphi_t = f_t\left(k_t,Ic_t\right)$ where Ic_t is the consumption of inputs. Otherwise, the estimation procedure remains the same. However, we lose fewer observations since the use of materials instead of investment means we need to observe firms for two consecutive years instead of three.

Cost shares

Alternatively, a TFP measure can be constructed using cost shares as estimates of the labour and capital coefficients in equation (46). The costs of labour and capital were evaluated by Boutin

and Quantin (2006) for each cell defined by the 3-digit industry, the year, and the number of firm employees (less than 5, 5–20, 20–50, 100, more than 100). The share of capital (resp. labour) in these costs is denoted $\hat{\beta}^s_{1,t}$ (resp. $\hat{\beta}^s_{2,t}$). Implicitly, we assume that returns to scale equal one as we have: $\hat{\beta}^s_{1,t} + \hat{\beta}^s_{2,t} = 1$. The predicted value-added based on capital and labour is $\ln V^p_t = \hat{\beta}^s_{1,t} \ln(k_t) + \hat{\beta}^s_{2,t} \ln(l_t)$. The following specification can then be estimated with OLS:

$$\ln(V_t) - lnV^p_t = \beta_{0,t} + \sum_{s=1}^{3} \sigma_s l_{s,t} + \widetilde{\phi}_t \tag{A17}$$

Denoting $\hat{\beta}^s_{0,t}$ and $\hat{\sigma}^s_s$ the estimated coefficients, the TFP measure is given by:

$$\hat{\widetilde{\phi}}_t = \ln(V_t) - lnV^p_t - \hat{\beta}^s_{0,t} - \sum_{s=1}^{3} \hat{\sigma}^s_s l_{s,t} \tag{A18}$$

For all methods, the TFP of a firm is the firm-level average of yearly TFP over the period 1994–2002. The TFP estimates we recover with these four approaches are highly correlated. The correlation between OLS TFP and Olley-Pakes TFP is 0.73. The correlation between OLS TFP and Levinsohn-Petrin TFP is 0.85. The correlation OLS TFP and cost-shares TFP is 0.93. Unsurprisingly, these alternative methods to estimate TFP give results which are qualitatively similar for A, D, and S at the sector level. These results are available upon request.

Appendix D. Estimations for employment areas

Table 5: Employment areas above vs. below median density

Sector	OLS, mono-establishments				
	A	D	S	R^2	obs.
	(1)	(2)	(3)	(4)	(5)
Food, beverages, tobacco	0.07	0.94	0.01	0.97	22,049
	(0.00) *	(0.02) *	(0.00) *		
Apparel, leather	0.04	1.35	0.01	0.99	5,804
	(0.01) *	(0.05) *	(0.01)		
Publishing, printing, recorded media	0.17	1.23	0.00	0.98	9,236
	(0.01) *	(0.05) *	(0.00)		
Pharmaceuticals, perfumes, soap	0.06	1.15	-0.01	0.86	1,069
	(0.06)	(0.13)	(0.06)		
Domestic appliances, furniture	0.13	1.20	0.01	0.98	6,362
	(0.01) *	(0.04) *	(0.01) *		
Motor vehicles	0.09	1.19	0.01	0.85	1,442
	(0.02) *	(0.13) *	(0.02)		
Ships, aircraft, railroad equipment	0.09	1.10	0.00	0.85	1,016
	(0.03) *	(0.17)	(0.02)		
Machinery	0.09	1.06	0.00	0.98	14,738
	(0.01) *	(0.02) *	(0.00)		
Electric and electronic equipment	0.08	0.98	0.00	0.96	5,749
	(0.01) *	(0.04)	(0.00)		
Building materials, glass products	0.07	1.10	0.00	0.94	3,196
	(0.01) *	(0.05)	(0.01)		
Textiles	0.06	1.11	0.01	0.94	3,363
	(0.01) *	(0.07) *	(0.01)		
Wood, paper	0.09	1.13	0.00	0.98	5,872
	(0.01) *	(0.05) *	(0.01)		
Chemicals, rubber, plastics	0.08	1.11	0.01	0.96	5,335
	(0.01) *	(0.04) *	(0.01)		
Basic metals, metal products	0.07	1.06	0.00	1.00	14,305
	(0.01) *	(0.02) *	(0.00)		
Electric and electronic components	0.08	0.99	-0.01	0.94	2,579
	(0.01) *	(0.06)	(0.01)		
Consultancy, advertising, business services	0.19	1.06	-0.01	0.96	37,041
	(0.02) *	(0.03)	(0.03) *		
All sectors	0.09	1.21	0.00	1.00	139,143
	(0.00) *	(0.01) *	(0.00)		

References

Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum. 2003. Plants and productivity in international trade. *American Economic Review* 93(4):1268–1290.

Bernard, Andrew B. and J. Bradford Jensen. 1999. Exceptional exporter performance: Cause, effect, or both? *Journal of International Economics* 47(1):1–25.

Berry, Steven and Joel Waldfogel. 2006. Product quality and market size. Processed, Yale University.

Boutin, Xavier and Simon Quantin. 2006. Une méthodologie d'évaluation comptable du coût du capital des entreprises françaises: 1984–2002. Working Paper G2006–09, INSEE-DESE.

Burnod, Guillaume and Alain Chenu. 2001. Employés qualifiés et non-qualifiés: Une proposition d'aménagement de la nomenclature des catégories socioprofessionnelles. *Travail et Emploi* 0(86):87–105.

Carrasco, Marine and Jean-Pierre Florens. 2000. Generalization of GMM to a continuum of moment conditions. *Econometric Theory* 16(6):797–834.

Ciccone, Antonio and Robert E. Hall. 1996. Productivity and the density of economic activity. *American Economic Review* 86(1):54–70.

Clerides, Sofronis, Saul Lach, and James R. Tybout. 1998. Is learning by exporting important? Micro-dynamic evidence from Colombia, Mexico, and Morocco. *Quarterly Journal of Economics* 113(3):903–947.

Collard-Wexler, Allan. 2007. Productivity dispersion and plant selection in the ready-mix concrete industry. Processed, New York University.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2008. Spatial wage disparities: Sorting matters! *Journal of Urban Economics* 63(2):723–742.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux. 2009. Estimating agglomeration effects with history, geology, and worker fixed-effects. In Edward L. Glaeser (ed.) *The Economics of Agglomeration*. Cambridge, MA: National Bureau of Economic Research, forthcoming.

De Loecker, Jan. 2007. Product differentiation, multi-product firms and estimating the impact of trade liberalization on productivity. Processed, Princeton University.

Del Gatto, Massimo, Giordano Mion, and Gianmarco I.P. Ottaviano. 2006. Trade integration, firm selection and thecosts of non-Europe. Discussion Paper 5730, Centre for Economic Policy Research.

Del Gatto, Massimo, Gianmarco I.P. Ottaviano, and Marcello Pagnini. 2008. Openness to trade and industry cost dispersion: Evidence from a panel of Italian firms. *Journal of Regional Science* 48(1):97–129.

Dixit, Avinash K. 1979. A model of duopoly suggesting a theory of entry barriers. *Bell Journal of Economics* 10(1):20–32.

Duranton, Gilles and Diego Puga. 2004. Micro-foundations of urban agglomeration economies. In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland, 2063–2117.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review* 98(1):394–425.

Fujita, Masahisa and Hideaki Ogawa. 1982. Multiple equilibria and structural transition of non-monocentric urban configurations. *Regional Science and Urban Economics* 12(2):161–196.

Glaeser, Edward L. and David C. Maré. 2001. Cities and skills. *Journal of Labor Economics* 19(2):316–342.

Gobillon, Laurent and Sébastien Roux. 2008. Quantile-based inference of parametric transformations between two distributions. Processed, CREST-INSEE.

Head, Keith and Thierry Mayer. 2004. The empirics of agglomeration and trade. In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland, 2609–2669.

Heckman, James J. and Bo E. Honore. 1990. The empirical content of the Roy model. *Econometrica* 58(5):1121–1149.

Hellerstein, Judith K., David Neumark, and Kenneth R. Troske. 1999. Wages, productivity, and worker characteristics: Evidence from plant-level production functions and wage equations. *Journal of Labour Economics* 17(3):409–446.

Henderson, J. Vernon. 1974. The sizes and types of cities. *American Economic Review* 64(4):640–656.

Hopenhayn, Hugo. 1992. Entry, exit, and firm dynamics in long run equilibrium. *Econometrica* 60(5):1127–1150.

Levinsohn, James and Amil Petrin. 2003. Estimating production functions using inputs to control for unobservables. *Review of Economic Studies* 70(2):317–342.

Lileeva, Alla and Daniel Trefler. 2007. Improved market access to foreign markets raises plant-level productivity... for some plants. Working Paper 13297, National Bureau of Economic Research.

Lucas, Robert E., Jr. and Esteban Rossi-Hansberg. 2002. On the internal structure of cities. *Econometrica* 70(4):1445–1476.

Marshall, Alfred. 1890. *Principles of Economics*. London: Macmillan.

Melitz, Marc and Gianmarco I. P. Ottaviano. 2008. Market size, trade and productivity. *Review of Economic Studies* 75(1):295–316.

Melitz, Marc J. 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica* 71(6):1695–1725.

Melo, Patricia C., Daniel J. Graham, and Robert B. Noland. 2009. A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics* 39:forthcoming.

Nocke, Volker. 2006. A gap for me: Entrepreneurs and entry. *Journal of the European Economic Association* 4(5):929–956.

Olley, G. Steven and Ariel Pakes. 1996. The dynamics of productivity in the telecommunication equipment industry. *Econometrica* 64(6):1263–1297.

Ottaviano, Gianmarco I. P., Takatoshi Tabuchi, and Jacques-Franccois Thisse. 2002. Agglomeration and trade revisited. *International Economic Review* 43(2):409–436.

Pavcnik, Nina. 2002. Trade liberalization, exit, and productivity improvements: Evidence from Chilean plants. *Review of Economic Studies* 69(1):245–276.

Rosenthal, Stuart S. and William Strange. 2004. Evidence on the nature and sources of agglomeration economies. In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland, 2119–2171.

Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: Printed for W. Strahan, and T. Cadell.

Sveikauskas, Leo. 1975. Productivity of cities. *Quarterly Journal of Economics* 89(3):393–413.

Syverson, Chad. 2004. Market structure and productivity: A concrete example. *Journal of Political Economy* 112(6):1181–1222.

Vives, Xavier. 1990. Trade association disclosure rules, incentives to share information, and welfare. *Rand Journal of Economics* 21(3):409–430.

Wheaton, William C. and Mark J. Lewis. 2002. Urban wages and labor market agglomeration. *Journal of Urban Economics* 51(3):542–562.

B.6. Estimating gender differences in access to jobs: Females trapped at the bottom of the ladder

Gobillon L., Meurs D. et S. Roux (2009), "Estimating gender differences in access to jobs: Females trapped at the bottom of the ladder", CEPR Working Paper 7475.

55 pages

Estimating Gender Differences in Access to Jobs: Females Trapped at the Bottom of the Ladder*

Laurent Gobillon† Dominique Meurs‡ Sébastien Roux§

September 16, 2009

Abstract

In this paper, we propose a job assignment model allowing for a gender difference in access to jobs. Males and females compete for the same job positions. They are primarily interested in the best-paid jobs. A structural relationship of the model can be used to empirically recover the probability ratio of females and males getting a given job position. As this ratio is allowed to vary with the rank of jobs in the wage distribution of positions, barriers in females' access to high-paid jobs can be detected and quantified. We estimate the gender relative probability of getting any given job position for full-time executives aged $40-45$ in the private sector. This is done using an exhaustive French administrative dataset on wage bills. Our results show that the access to any job position is lower for females than for males. Also, females' access decreases with the rank of job positions in the wage distribution, which is consistent with females being faced with more barriers to high-paid jobs than to low-paid jobs. At the bottom of the wage distribution, the probability of females getting a job is 12% lower than the probability of males. The difference in probability is far larger at the top of the wage distribution and climbs to 50%.

Keywords: gender, discrimination, wages, quantiles, job assignment model, glass ceiling.

JEL Classification: J16, J31, J71

*We are grateful to the participants of seminars and conferences at the University of Hanover, INED (Paris), CREST (Paris), ESEM (Barcelona), EALE (Tallin) for their comments, and especially to Francis Kramarz and Ronald Oaxaca for interesting and useful discussions.

†INED, PSE-INRA, CREST and CEPR. Address: Institut National d'Etudes Démographiques (INED), 133 Boulevard Davout, 75980 Paris Cedex 20, France. Email: laurent.gobillon@ined.fr.

‡University of Paris 10 (EconomiX) and INED. Address: Institut National d'Etudes Démographiques (INED), 133 Boulevard Davout, 75980 Paris Cedex 20, France. Email: dominique.meurs@ined.fr.

§CREST-INSEE and PSE-INRA. Centre de Recherche en Economie et Statistique (CREST), 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France. Email: sebastien.roux@ensae.fr.

1 Introduction

A growing body of literature shows that the gender wage gap is mostly due to the under-representation of females in well-paid occupations. This phenomenon has been called "a glass ceiling effect" to evoke the idea that there is an unspoken rationale which impedes females from holding the highest positions in firms. Following the strand of research initiated by Albrecht, Bjorklund and Vroman (2003), empirical papers use quantile regressions to study the gender difference in access to jobs. They consider that there is a glass ceiling when the gap between the highest centiles of males and females's wage distribution is larger than the gap between lower centiles.

We argue that this approach confuses two dimensions, the job position and the associated wage, possibly leading to inaccurate interpretations. Figure 1 proposes a simple scheme illustrating this point. Suppose a classic job ladder where the wage increases more than proportionally with the rank. Positions are occupied alternately by a female and a male (axis 1). The gender quantile difference for high-paid jobs is larger than for low-paid jobs, which means that the gender wage gap widens along the job ladder. It is tempting to conclude that there is a glass ceiling but this interpretation is arguable as the odds of a female (or a male) to occupy a position are roughly constant along the job ladder. It is possible to control for the unequal spacing between the wages of consecutive positions considering the difference between the ranks of the gender wage distributions instead of the quantiles. We obtain what seems to be a right answer as the gender rank difference is constant along the job ladder (axis 2). However, this is misleading as a setting where there is an obvious glass ceiling can also generate a constant gender rank difference. This is the case when the females occupy the three lowest positions on the job ladder and the males occupy the three highest positions (axis 4).

[*Insert Figure* 1]

The confusion arises because the analysis is based on the ranks in the two gender wage distributions and these ranks are not directly related to the position of jobs on a common job ladder. A sound analysis should rather consider a hierarchy of job positions and investigate how the gender difference in access to jobs may depend on the rank along this ladder. The simplest way to order jobs is probably to consider their rank in the wage distribution of positions. A glass ceiling effect occurs when females have no access to the jobs with the highest ranks in the wage distribution of positions. More generally, females are faced with barriers to high-paid positions when their relative access to jobs compared to males decreases with the rank of jobs.

In this paper, we propose a job assignment model which shows how the relative access to jobs of males and females influences their position along the job ladder. Workers rank jobs according to the wage. For each position, competition occurs among workers who were not selected for a better job, and the employer may favour males over females. We introduce an access function which measures the gender difference in access to jobs depending on their rank in the wage distribution of positions. This function is defined as

the probability ratio of females and males getting a job of a given rank. In an empirical section, we use a structural relationship of the model to assess the importance of the barriers to high-paid jobs that females are faced with. Estimations are conducted for full-time executives aged 40-45 working in French private and public firms.

Our work builds on the literature on job assignment models which posits the existence of heterogenous job positions (see Sattinger, 1993; Teulings, 1995; Fortin and Lemieux, 2002; Costrell and Loury, 2004). In our model, each position is characterized by a specific wage offer to applicants. Male and female workers apply for the best-paid job. The match between each worker and the position is characterized by a quality which affects the profit of the firm. The manager of the best-paid job selects the applicant who is the most valuable. The manager of the second best-paid job hires an individual among the remaining workers, and so on.

We assume that managers take into account the gender of applicants in their hiring process. Employers may expect males to have an average productivity which is higher than the one of females, in line with some statistical discrimination (Arrow, 1971; Phelps, 1972; Coate and Loury, 1993). They may also prefer to hire males rather than females simply because of their tastes (Becker, 1971). Employers choose an applicant on the basis of their utility which depends on the expected profit of the firm and their tastes. As the gender may affect the employers' utility through the two types of discrimination, females may have a lower access to jobs than males. Barriers in the access to jobs are allowed to vary depending on the rank of the job in the wage distribution of positions.

A simple way to characterize the gender relative access is to consider one female worker and one male worker applying for the same job position. Their relative access to the job can then be defined as their relative chances of getting the job. Accordingly, we define an access function $h(u)$ as the probability ratio of a female and a male getting a job of rank u. We formally define three particular cases: some uniform discrimination against females in the access to jobs ($h(u) = \gamma < 1$ at all ranks), some barriers to high-paid jobs ($h(.)$ decreasing with the rank) and a sticky floor ($h(u) > 1$ at lower ranks). For a given access function and a given share of females in the population of workers, the model predicts the numbers of males and females competing for a job at each rank in the wage distribution of positions. It also predicts the gender quantile difference for a given wage distribution of job positions. In a simulation exercise, we consider a constant access function and allocate males and females into job positions with our model. We are able to exhibit an empirical wage distribution[1] for which the model predicts a gender quantile difference increasing with the rank. Whereas the literature would conclude to the existence of a glass ceiling, there is none. Our illustrative example thus confirms that the usual interpretation of the gender quantile difference can be misleading.

In the empirical part of the paper, we use a structural relationship derived from our model to estimate the access function non parametrically from the ranks of males and females in the wage distribution of positions. The estimations are conducted on some French data collected from the employers for tax purposes in 2003,

[1] This wage distribution is computed for full-time executives aged 40-45 in the banking industry.

the *Déclarations Annuelles des Salaires* (DADS). These data are exhaustive for the private sector.

Our analysis is related to a few empirical works which directly investigate the gender difference in positions along the job ladder. Pekkarinen and Vartiainen (2006) show on Finish data that among blue-collar workers, females have to reach a higher productivity threshold to get promoted than males. Winter-Ebner and Zweimüller (1997) find on Austrian data that the gender difference in detailed occupations remains mostly unexplained after controlling for the differences in endowments and discontinuities in labor market experience. However, this kind of studies is usually limited by the lack of detailed information on the individual positions along the job ladder. Here, we consider that the wage is a reasonable proxy for the position in the job hierarchy: a higher wage corresponds to a better position. Killinsworth and Reimers (1983) argue that neither the type nor the rank of a position is perfectly indexed by the wage. This is particularly true for blue collars for whom wages increase significantly with job tenure. Also, some blue collars occupy jobs which are paid at the minimum wage but do not correspond to the same hierarchical position. Hence, we restrict our attention to executives whose wage reflects more closely the rank along the job ladder. We only keep full-time workers aged 40-45 for whom job positions can be considered to be on a single market in line with our model.

Our results show that females have a lower access to jobs than males at all ranks in the wage distribution. Also, their access decreases with the rank, which is consistent with more barriers to high-paid jobs than to low-paid jobs. At the bottom of the wage distribution (5th percentile), the probability of females getting a job is 12% lower than the probability of males. The difference in probability is far larger at the top of the wage distribution (95th percentile) and climbs to 50%. We also restrict our analysis to specific industries as they constitute more homogenous labour markets. We consider more specifically banking and insurance as they are labour intensive with a large share of females, and have different wage policies in France. Banks rely on a rigid job classification inherited from the early eighties when they belonged to the public sector. By contrast, insurance companies propose some careers which are much more individualized. Regarding females, there are far more barriers to high-paid jobs than to low-paid jobs in the insurance industry. Differences in barriers are smaller in the banking industry. In particular, when approximating the access function with a linear specification, we find that the slope of the access function is more than eight times steeper in the insurance industry than in the banking industry. Also, at high ranks (95th percentile), the relative access to jobs of females compared to males is nearly two times smaller in the insurance industry (27%) than in the banking industry (60%).

We then extend our model to take into account the individual observed heterogeneity in the access to jobs. We find that when controlling for age and being born in a foreign country, results remain unchanged. This is in line with our use of an homogeneous population. We also make an alternative assumption on the extent of the labour market, supposing that the competition of workers for jobs occurs within each firm rather than on the national market. We estimate the average access function across large firms employing more than 150 full-time executives aged $40 - 45$. When pooling all industries, results are quite similar to those obtained

when competition is supposed to occur on the national market. For the specific insurance industry, results are a bit different as for females, we find less barriers to high-paid jobs than to low-paid jobs. This change is generated by some heterogeneity in the level of wages among firms.

The rest of the paper is organized as follows. In section 2, we present our baseline model. Our econometric strategy to estimate the access function is detailed in section 3. We then describe our dataset and report some stylized facts in section 4. We comment our estimation results in section 5. Finally, the model is extended to take into account the individual observed heterogeneity and segmented markets in section 6. Concluding remarks are given in the last section.

2 The model

2.1 Setting

We first present a simple model where gender differences in access to jobs yield a specific assignment of male and female workers into jobs and some gender differences in wages. Consider a countable number of workers applying for a countable number of job positions. There is a proportion n_m of males in the whole population of workers which we rather refer to as the *measure* of males for clarity hereafter, and a measure $n_f = 1 - n_m$ of females. The workers do not differ otherwise. We now introduce some mechanisms which determine how males and females are assigned to job positions.

The utility of a worker only depends on his daily wage. Hence, a worker is primarily interested in the job yielding the highest wage. Job positions are heterogenous such that each job position is associated to a specific fixed wage through a contract. This corresponds to a setting of imperfect information where employers do not observe *ex ante* the match between the applicants and the job position when they post their job offer (see Cahuc and Zylberberg, 2004, chapter 6 for a discussion). The wage associated to a contract is not allowed to depend on the gender of the applicant. We suppose that two job positions cannot be associated with the same wage offer so that each job can be uniquely identified by its *rank* in the wage distribution.[2] Workers apply for the best ranked job as it offers the highest wage. Those who are not selected apply for the second best ranked job, and so on.

For any job position of given rank u, the manager screens all the applicants (that is to say, all the workers not hired for jobs of higher rank). The match between the manager and any given worker i is characterized by a quality $\varepsilon_i(u)$ which determines the expected profit associated to the job through the expression:

$$\Pi_u(i) = \theta_j(u) \exp[\varepsilon_i(u)] \tag{1}$$

The multiplicative term $\theta_j(u)$ captures the expected productivity for each gender. There is some statistical discrimination against females where the manager expects a lower average productivity for females than for

[2] The wage distribution is supposed to be exogenous. We could introduce some mechanisms on the labour market to endogenize this distribution but it is beyond the scope of this paper as the wage setting is of no use in our empirical approach.

males. The manager observes the match quality so that he can evaluate how much profit he can make from the job if hiring the applicant. However, the manager does not only take into account the profit when choosing a worker but also his tastes for the gender of the worker. He thus rather considers his utility which is given by:

$$V_u(i) = \ln \mu^*_{j(i)}(u) + \ln \Pi_u(i) \tag{2}$$

where $j(i)$ is the gender of individual i and $\mu^*_j(u)$ captures the taste of the manager for gender j. Taste discrimination is taken into account by a lower taste parameter for females than for males. The utility of the manager can be rewritten in reduced form as:

$$V_u(i) = \ln \mu_{j(i)}(u) + \varepsilon_i(u)$$

where $\ln \mu_j(u) = \ln \mu^*_j(u) + \ln \theta_j(u)$ captures all the gender-specific effects (which cannot be identified separately in our application) and reflects the overall value of a gender for a job position at a given rank. According to this specification, females' access to jobs is allowed to vary with the position as the gender-specific term varies with the rank of the position in the wage distribution: females may have a lower access to better ranked jobs.

The manager chooses the applicant who grants him the highest level of utility. The maximization program of the manager is then:

$$\max_{i \in \Omega(u)} V_u(i) \tag{3}$$

where $\Omega(u)$ is the set of workers available for the job ($\Omega(1)$ being the whole population of workers). This set contains all the workers who were not selected for jobs of rank above u, i.e. who did not draw a match quality high enough to get selected for those jobs. The set of workers available for the job of rank u can thus be defined recursively as:

$$\Omega(u) = \left\{ i \,\middle|\, \text{for all } \widetilde{u} > u,\ V_{\widetilde{u}}(i) < \max_{k \in \Omega(\widetilde{u})} V_{\widetilde{u}}(k) \right\} \tag{4}$$

The resulting allocation of workers is a Nash equilibrium. Workers have no incentive to move from their position. This is because the worker occupying the best position has no incentive to move to a less-paid job. The worker occupying the second best position cannot move to the best position as it is already occupied. Hence, he has no incentive to move, and so on. Also, managers have no incentive to fire an employee as they cannot find a better worker on the market. We assume that at the equilibrium, there is a bijection between workers and job positions so that any job position is filled and any worker is employed.[3]

It is possible to determine for a given job, a closed formula for the probability that the selected worker is of gender j under some additional assumptions. The maximization program of the manager given by (3) and (4) is a multinomial model with two specificities. First, the choice set consists in all workers still available after better ranked job positions have been filled. There would be a selection process based on match qualities

[3]In particular, this rules out the existence of workers not being hired and dropping out of the labour force, and job positions being not filled possibly because the offered wage is below the reservation wage of available individuals.

if the match quality of the workers available for the job was correlated with their match quality for better ranked jobs. We suppose that the match qualities are drawn independently across jobs to avoid this kind of selection mechanism. Second, the choice set contains an infinite but countable number of workers. We adapt the standard theory of multinomial choice models to this setting following Dagsvik (1994). For any job of given rank u, the share of available workers being of gender j is given by $\frac{n_j(u)}{n_f(u)+n_m(u)}$ where $n_j(u)$ is the measure of gender-j workers available for a job of rank u (such that we have: $n_j(1) = n_j$). We suppose that the points of the sequence $\{j(i), \varepsilon_i(u)\}$, $i \in \Omega(u)$ are the points of a Poisson process with intensity measure $\frac{n_j(u)}{n_f(u)+n_m(u)} \exp(-\varepsilon) d\varepsilon$. In particular, this assumption ensures that for any given job, the probability of preferring a worker in any given finite subgroup of available workers follows a logit model. Under this assumption, the following formula is verified by the probability that the worker chosen for the job of rank u is of gender j:

$$P(j(u) = j) = n_j(u)\,\phi_j(u) \tag{5}$$

with

$$\phi_j(u) = \frac{\mu_j(u)}{n_f(u)\,\mu_f(u) + n_m(u)\,\mu_m(u)} \tag{6}$$

where $\phi_j(u)$ is the unit probability of a gender-j worker getting the job. This probability depends on the measures of available workers of each gender, as well as the specific value attributed by the manager to each gender.

2.2 Characterization of the equilibrium

We can then determine for each gender j a differential equation which should be verified by the measure of available workers at each rank. Consider an arbitrarily small interval du in the unit interval. The proportion of jobs in this small interval is du since ranks are equally spaced (and dense) in the unit interval. The measure of jobs occupied by workers of a given gender j is then $n_j(u)\,\phi_j(u)\,du$. For this gender, the measure of workers available for a job of rank $u - du$ can be deduced from the measure of workers available for a job of rank u substracting the workers who get the jobs of ranks between $u - du$ and u :

$$n_j(u - du) = n_j(u) - n_j(u)\,\phi_j(u)\,du \tag{7}$$

From this equation, we obtain when $du \to 0$:

$$n_j'(u) = \phi_j(u)\,n_j(u) \tag{8}$$

For each gender, the decrease in the measure of available workers as the rank decreases can be expressed as the product of the measure of available workers and their unit probability of getting a job. Replacing the unit probability by its expression given by (6), we end up with two equations to determine, for the two genders, the measures of available workers at each rank in the wage distribution of job positions. We have the following existence theorem which proof is relegated in Appendix A:

Theorem 1 *Suppose that $\mu_m(\cdot)$ and $\mu_f(\cdot)$ are C^1 on $(0,1]$ and there is a constant $c > 0$ such that $\mu_m(u) > c$ and $\mu_f(u) > c$ for all $u \in (0,1]$, then there is a unique two-uplet $\{n_f(\cdot), n_m(\cdot)\}$ verifying (8) where $\phi_j(\cdot)$ is given by (6).*

We assume in our theorem that the gender-value functions must take their value above a strictly positive threshold, such that males and females can access all jobs. This assumption is made for the unit probabilities to be always well-defined as the denominator in their formula then cannot be zero. In some specific cases, we can extend the model to the case where the access of a gender to some jobs is completely denied and show that the model still has a solution. Consider for instance the case where females cannot access the best-paid jobs of ranks above a given threshold $\widetilde{u}$ because of a glass ceiling effect but have access to all jobs of ranks below this threshold. In that case, all the jobs of ranks above the threshold are occupied by males. For jobs of rank below the threshold, there is then a measure n_f of available females competing with a measure $n_m - (1 - \widetilde{u})$ of available males (provided that not all males have been hired for the best-paid jobs). It is possible to apply our existence theorem on the subset of ranks below the threshold and get a global solution on the whole set of ranks using a continuity argument.

Also note that the theorem can be extended to the case where the gender-value functions are not continuous, but rather discontinuous at a finite number of ranks. First consider the case where there is only one point of discontinuity. It is possible to apply the existence theorem separately for the subset of ranks below that point, and the subset of ranks above that point. The solution on the whole set of ranks can be recovered from the solutions on the two subsets of ranks using again a continuity argument. This procedure can easily be extended to the case where there are more points of discontinuity.

2.3 Gender differences in access to job

We now characterize the gender difference in access to jobs under the conditions of our existence theorem. We first consider the function which measures the relative preferences of managers for females compared to males:

$$h(\cdot) \equiv \frac{\mu_f(\cdot)}{\mu_m(\cdot)} \tag{9}$$

This function can be re-interpreted as a measure of the gender relative access to jobs and we label it the "access function". Indeed, consider one male worker and one female worker applying for a job position of given rank u. These two workers have different chances of getting the job as they are not of the same gender. The access function evaluated at rank u is the probability ratio of the female and the male being hired for the job position as we have from equations (6) and (9):

$$h(u) = \frac{\phi_f(u)}{\phi_m(u)} \tag{10}$$

When the access function takes the value one at all ranks, males and females have the same chances of getting each job position. When the access function takes a value lower than one for a job position of given rank,

females have less chances than males of getting the job. This situation may correspond to the case where there is some discrimination against females in the access to the job.

It is then possible to formally define some uniform discrimination against females in the access to jobs considering that the chances of females getting a job are uniformly lower than the chances of males at all ranks in the wage distribution of job positions:

Definition 1 *There is some* ***uniform access discrimination*** *if for any* u, $h(u) = \gamma < 1$.

By contrast, we can consider that there are more barriers for females to high-paid jobs than to low-paid jobs when they have a lower access to jobs at higher ranks:

Definition 2 *Females are faced with* ***more barriers to high-paid jobs*** *than to low-paid jobs if there are some ranks* u_0 *and* u_1 *such that for any* $u \in]u_0, u_1[$ *and* $v > u_1$, *we have* $h(u) > h(v)$ *and* $h(v) < 1$.

Females are faced with more barriers to high-paid jobs than to low-paid jobs when the access function is continuous, strictly decreasing and takes some values lower than one at the highest ranks. It is also case when the access function is a two-step function with the second step at a value lower than one. In particular, when the second step takes a zero value there is a glass ceiling: females have no access to the best-paid jobs.[4]

Finally, we can give a definition of the *sticky floor* which would correspond to females being preferred for low-paid jobs:

Definition 3 *There is a* ***sticky floor*** *if there are some ranks* u_0 *and* u_1 *such that for any* $u < u_0$ *and for any* $v \in]u_0, u_1[$, *we have:* $h(u) > h(v)$ *and* $h(u) > 1$.

Note that it is possible to have for females a sticky floor and barriers to high-paid jobs at the same time.

We now consider an example of access function verifying each definition (uniform access discrimination, more barriers to high-paid jobs and sticky floor) to shed some light on the mechanisms at stake in the model. For each access function, we determine numerically for each gender the measure of available workers at each rank at the equilibrium.[5] For that purpose, we need to set the proportion of females n_f to a given value which is chosen to be 22.4%.[6] For a job of rank u in the wage distribution of positions, denote by $v_j(u) = \frac{n_j(u)}{n_j}$ its

[4]Very often in the literature, the glass ceiling is more loosely defined. It is considered that there is a glass ceiling effect when the females' access to jobs is particularly low for top positions.

[5]For females, we use the algorithm proposed by Bulirsch and Stoer (for the implementation, see Press et al., 1992, p. 724-732) to solve the differential equation giving $n_f(\cdot)$. Plugging (6) into (8) for females, and using (10), we get: $n'_f(u) = \frac{n_f(u)h(u)}{n_m(u)+n_f(u)h(u)}$. Summing (8) for the two genders and integrating between 0 and u, we also get: $n_f(u) + n_m(u) = u$. From the two equations, we obtain the differential equation for females: $n'_f(u) = \frac{n_f(u)h(u)}{u-n_f(u)+n_f(u)h(u)}$. This differential equation is solved backward from the highest to the lowest rank using the initial condition $n_f(1) = n_f$. After the differential equation for females has been solved, we deduce the solution for males using the relationship $n_m(u) = u - n_f(u)$.

[6]This value corresponds to the proportion of females among workers aged $40 - 45$ occupying full-time executive jobs in the private sector (see next section for some details on the data).

rank in the wage distribution of gender j. We plot $v_j(u) - u$ which has the following interpretation: when $v_j(u) > u$ (resp. $v_j(u) < u$), a gender-j worker holding a job of rank u in the wage distribution of positions is ranked better (resp. worse) in the wage distribution of his gender. This means that the proportion of workers holding a job of rank above u is lower (resp. higher) for gender-j workers than for the whole population.

We first consider the case where the access function is uniform and takes the value $\gamma = .8$ at all ranks. We plot on Figure 2 for each gender, the difference between the rank in the wage distribution of that gender and the rank in the wage distribution of job positions. We obtain for males a curve which is below zero and $U-$shaped, and for females a curve which is above zero and bell shaped with a maximum .064 at the rank $u_0 = .35$. The intuitions behind the curves are the following (explanations on how mechanisms affect the curves are given for females only for brevity). Males have a better access than females to jobs with a high rank in the wage distribution of job positions and are more often hired. The proportion of males getting high-paid jobs is thus larger than the proportion of females. When the rank decreases (but is higher than u_0), some more females are rejected to low-paid jobs. This makes the difference between the rank in the wage distribution of females and the rank in the wage distribution of job positions increase. However, the stock of males looking for a job decreases faster than the stock of females. This makes the number of males finding a job decrease faster than the number of females and get very small. At ranks lower than u_0, the number of females finding a job is high enough to counterbalance their lower access to jobs and the rank in the wage distribution of females thus gets closer to the rank in the wage distribution of job positions. As males still have a better access to jobs of rank below u_0, the proportion of males getting a job is still higher than the proportion of females as the rank decreases. Hence, effects related to the difference in stock between males and females get larger as the rank decreases and females finally catch up with males when the rank gets to zero.

[*Insert Figure* 2]

We then consider the case where females are faced with more barriers to high-paid jobs than to low-paid jobs, and the access function is of the form: $h(u) = .8 - .3u$. The curve of females represented on Figure 3 remains bell shaped although the differences between the rank in the wage distribution of females and the rank in the wage distribution of job positions are usually larger than in the case of a uniform access discrimination. For instance, the maximum of the curve is now at .140 instead of .064. This is because the females' access to high-paid jobs is lower than in the previous case due to more barriers to high-paid jobs. More females are thus available for less-paid jobs. Note however that the maximum of the curve is reached at a higher rank than in the case of a uniform access discrimination (.42 instead of .35). Indeed, the access to jobs of females increases as the rank decreases, and the difference between the rank in the wage distribution of females and the rank in the wage distribution of job positions thus stabilizes more quickly.

[*Insert Figure* 3]

We finally study a situation where there is at the same time more barriers to high-paid jobs and a sticky

floor, the access function being $h(u) = 1.2 - .4u$. Curves represented on Figure 4 exhibit an intricate profile. For females, the curve has the same profile as in the case of barriers to high-paid jobs for ranks above the threshold $u_1 = .2$. However, for ranks below u_1, the difference between the rank in the wage distribution of females and the rank in the wage distribution of job positions becomes negative and the profile is U-shaped. This occurs because below the threshold u_1, males have a lower access to jobs than females and their access to jobs decreases as the rank decreases. Hence, curves are reversed compared to the profile associated to the case where females are faced with more barriers to high-paid jobs.

[*Insert Figure* 4]

2.4 Gender quantile differences

The recent empirical literature on discrimination against females has focused on the difference between the quantiles of the wage distributions of males and females. Typically, when this difference is increasing with the rank, it is usually said that there is a glass ceiling (see Albrecht, Björklund and Vroman, 2003). However, this intepretation does not rest on any straightforward rationale and has two caveats. First, it does not control for the spacing between wages and thus mixes the rank of positions on the job ladder with earnings. Second, the rank at which quantiles are computed has a different meaning for the two genders. For males, it corresponds to the rank in the wage distribution of males. For females, it corresponds to the rank in the wage distribution of females. In this subsection, we show that it is possible to generate a gender quantile difference which is increasing with the rank even if there is no glass ceiling and the difference in access to jobs between males and females is the same at all ranks.

We first solve the model when the access function is constant with $h(u) = .672$ at all ranks and the proportion of females is the one in banking (28.7%).[7] The numerical solution allows to compute $v_j(u) = \frac{n_j(u)}{n_j}$ as well as $u_j = v_j^{-1}$ which gives for a job of given rank in the wage distribution of gender j, its rank in the wage distribution of job positions. We can then relate the quantile function of gender j denoted $\lambda_j(\cdot)$ to the quantile function of job positions $\lambda(\cdot)$ through the relationship: $\lambda_j(v) = \lambda[u_j(v)]$. The gender quantile difference is given by:

$$(\lambda_m - \lambda_f)(v) = \lambda[u_m(v)] - \lambda[u_f(v)] \tag{11}$$

We can compute the gender quantile difference using the solution $u_j(\cdot)$ of the model and the wage distribution of job positions in banking for $\lambda(\cdot)$. The gender quantile difference represented on Figure 5 is an increasing function above rank .6. Whereas the increase is small just above that rank, the curve becomes very steep above rank .9. The literature would conclude to a glass ceiling whereas there is none.

Also note that the profile of the gender quantile difference is very sensitive to the wage distribution of job positions. Indeed, consider alternatively a wage distribution of job positions which is uniform on the

[7] These choices are made clear in the empirical section. Indeed, we will show that the difference in access to jobs between males and females is nearly uniform in the banking industry and that the access function takes values close to .672 at all ranks.

interval $[\alpha, \alpha+\theta]$ where α and θ are some positive parameters, so that we have $\lambda(u) = \alpha + \theta u$. The gender quantile difference then corresponds to the gender rank difference up to a scale parameter.[8] Figure 5 shows that the gender quantile difference now has a bell shaped profile which is very different from the increasing profile found earlier. This sensitivity of the gender quantile difference to the shape of the wage distribution of job positions is another argument toward the unreliability of interpretations based on the profile of gender quantile differences.

[*Insert Figure* 5]

Economic interpretations should rather rely on the primitive function of a model which is the access function in our case. We now propose an econometric approach to estimate the access function non parametrically from the data.

3 Estimation strategy

3.1 Estimating the access function

We now show how the access function can be estimated from a cross-section dataset containing for each worker some information on his gender and his wage. First recall that the access function can be reinterpreted as the unit probability ratio of females and males getting a given job. From equation (8), each unit probability can be rewritten as:

$$\phi_j(u) = \frac{n'_j(u)}{n_j(u)} \tag{12}$$

We introduce for gender-j workers, the random variable corresponding to their rank in the wage distribution of job positions, U_j. The cumulative (resp. density) of this variable is denoted F_{U_j} (resp. f_{U_j}). The cumulative verifies the relationship: $F_{U_j}(u) = n_j(u)/n_j$. Hence, each unit probability can be rewritten as:

$$\phi_j(u) = \frac{f_{U_j}(u)}{F_{U_j}(u)} \tag{13}$$

The numerator and denominator of the gender-j unit probability only depend on the distribution of ranks of gender-j workers in the wage distribution of job positions.

For a given gender, the numerator and denominator of the unit probability only depend on the distribution of ranks of workers of that gender in the wage distribution of job positions. This means that in practice, the ranks of workers of each gender in the wage distribution of job positions are enough to estimate the unit probabilities, and thus the access function. These ranks can be computed very easily from the data.

For each gender, we construct some estimators of the numerator and denominator of the unit probability of getting a job. The Rosenblatt-Parzen Kernel estimator of the density $f_{U_j}(\cdot)$ is given by:

$$\widehat{f}_{U_j}(u) = \frac{1}{\omega_{jN} N_j} \sum_{i|j(i)=j} K\left(\frac{u-u_i}{\omega_{jN}}\right)$$

[8] The value of the parameter θ is needed in our simulations and is fixed such that the variance of the uniform wage distribution is the same as the variance of the wage distribution of job positions in the banking industry.

where $K(\cdot)$ is a Kernel, ω_{jN} is the bandwidth, $j(i)$ is the gender of individual i and u_i is his rank in the wage distribution of job positions. In our application, the Kernel is chosen to be Epanechnikov and the bandwidth takes the value given by the rule of thumb (Silverman, 1986). A standard estimator of the cumulative $F_{U_j}(\cdot)$ is given by:

$$\begin{aligned}\widehat{F}_{U_j}(u) &= \int_{-\infty}^{u} \widehat{f}_{U_j}(u)\,du \\ &= \frac{1}{N_j}\sum_{i|j(i)=j} L\left(\frac{u-u_i}{\omega_{jN}}\right)\end{aligned}$$

wherer $L(u) = \int_{-\infty}^{u} K(v)\,dv$. For gender j, an estimator of the unit probability of getting a job is then $\widehat{\phi}_j(u) = \widehat{f}_{U_j}(u)/\widehat{F}_{U_j}(u)$. We finally obtain an estimator of the access function:

$$\widehat{h}(u) = \frac{\widehat{\phi}_f(u)}{\widehat{\phi}_m(u)} \tag{14}$$

This estimator is computed for a grid of 1000 ranks in $[0,1]$ which are equally spaced. The confidence interval of the access function at each rank is computed by bootstrap with replacement (100 replications).

3.2 Discussion

It is possible to reinterpret our estimator of the access function drawing a parallel between our specification and duration models. Indeed, we implicitely assumed the existence of a timeline in our model, which runs in the direction opposite to ranks. This is because workers prefer being hired for high-paid jobs, and only those who are not selected turn to low-paid jobs. The unit probability of getting a job in a small rank interval $[u-du, u]$ for a worker available for jobs below rank u is similar to the instantaneous hazard of getting a job in a small *duration* interval $[t, t+dt]$ for a worker still looking for a job after a *duration* t. For the two frameworks to match, we just need the analogical duration to verify: $T_j = 1 - U_j$.

The unit probability of getting a job can then be rewritten as the instantaneous hazard of the analogical duration denoted $\lambda_j(\cdot)$.[9] Indeed, we have: $F_{U_j}(1-t) = S_{T_j}(t)$ and $f_{U_j}(1-t) = f_{T_j}(t)$ where S_{T_j} (resp. f_{T_j}) is the survival (resp. density) function of the analogical duration. Hence, we obtain from (13):

$$\phi_j(1-t) = \frac{f_{T_j}(t)}{S_{T_j}(t)} = \lambda_j(t)$$

Our empirical strategy thus amounts to estimate for each gender the density and survival functions of the analogical duration to construct an estimator of the instantaneous hazard function. Our estimator of the access function is then the ratio of the two gender instantaneous hazards.

[9] This approach is quite similar to Donald, Green and Paarsch (2000) who consider that wages are some non-negative quantities such as time spells, and approximate their distribution parametrically using duration modelling. Whereas their approach is descriptive, we are rather interested in recovering the key function of our theoretical model. Also, the variable that we assimilate to a duration is the analogical duration (one minus the rank) rather than the wage.

An alternative approach could be to express for each gender, the instantaneous hazard function as the derivative of the survival function. An estimator of the survival function is given by the Kaplan-Meier estimator (see for instance Lancaster, 1990). The logarithm of a smoothed version of this estimator can then be derived to recover the instantaneous hazard. Once again, the ratio of the two estimated gender instantaneous hazards gives an estimator of the access function. We did not follow this path as the estimator we used was more straightforward. However, the parallel with duration models will prove to be very useful when we will extend our model to take into account some individual observed hetereogeneity.

4 Descriptive statistics

4.1 The data

The wage distributions of job positions, males and females are constructed from the *Déclarations Annuelles de Données Sociales* (DADS) or Annual Social Data Declarations database. These data are collected by the French Institute of Statistics (INSEE) from the employers for tax purposes every year since 1994. They are *exhaustive* for all private and public firms in the private sector. For each job, the data contain some information on the industry, contract type (full-time/part-time), daily wage, socio-professional category, age, sex and country of birth (France/foreign country[10]) of the employee. A limitation is that the education level of employees is not reported.

As our model is static, we consider the single year 2003. For that year, there are $20,599,456$ jobs in $1,599,865$ firms. We want to restrict our attention to a subpopulation of workers for which the assumptions of our model are more likely to be verified. Because of the mimimum wage, some blue collars and clerks may be paid the same wage although they are ranked differently along the job ladder. Also, the job tenure has an important effect on the wage of blue collars even if they do not move to another job position. We discard low-skilled workers from our analysis to avoid these issues and rather focus on workers with an executive job position (business managers, top executives, engineers and marketing staff). There are $2,173,975$ executive job positions in $318,852$ firms.

We want to study a homogenous market where males and females compete for the same positions. For that purpose, we restrict our sample to executives working full time and aged $40-45$. Executive females still on the market at those ages usually have not experienced career interruptions, are more career-oriented and compete for jobs with males. Having a range of only six years for age limits the cohort effects. Table 1 shows that for the $40-45$ age bracket, there are $354,968$ executive job positions in $86,989$ firms. 22.4% of these executives are females. The wage distribution is skewed to the right and the mean daily wage (139 euros) is higher than the median daily wage (109 euros). The dispersion is very large and the standard error of wages stands at 602 euros. There is a large gender gap in wages as the gender difference in median wage is as large as 17 euros.

[10]For workers born in a foreign country, our data do not allow us to distinguish which country it is.

We will have a more careful look at two industries: banking and insurance. These industries share some similarities as they are both labor-intensive and employ both a high proportion of executives. Also, the proportion of females is above the average, which is quite usual in service industries: 28.7% (resp. 36.9%) of executives are females in banking (resp. insurance). This ensures that there is a large pool of female executives competing for promotion with their male counterparts. The common organizational features of the two industries contrast with the differences in their wage structure. There is a far larger gender gap in median wage for insurance (21 euros) than for banking (13 euros). Also, the wage dispersion is far larger in the banking industry than in the insurance industry. It is of particular interest to study the females' access to high-paid jobs in the two industries as the economic performances in these industries heavily rely on the quality of the management of human resources (Bartel, 2004). A discrimination in access to jobs against females may result in a less efficient matching between workers and job positions with large economic consequences.

[*Insert Table* 1]

4.2 Gender wage distributions

In line with the literature (Albrecht, Björklund and Vroman, 2003), we compare the wage distributions of male and female full-time executives aged $40 - 45$ working in a private or public firm. Figure 6 represents for each gender, the wage distribution as a function of the rank.[11] Males have a higher wage than females at every rank and the gap widens as the rank increases. Figure 7 shows that the wage difference is 15% at the bottom of the distribution (5th percentile) and that it goes up to 26% at the top of the distribution (95th percentile). This increase is usually interpreted as a glass ceiling effect.

[*Insert Figures* 6 *and* 7]

However, when computing the wage difference between males and females at a given rank, this rank does not have the same definition for each gender. For males, it is the rank in the wage distribution of males. For females, it is the rank in the wage distribution of females. There is no straightforward intuition on how the difference between these two ranks is taken into account in the glass ceiling interpretation. It is possible to link these two ranks in a descriptive way though, relating them to the rank in the wage distribution of job positions.

Figure 8 represents the rank in the wage distribution of each gender as a function of the rank in the wage distribution of job positions. If males and females had the same access to jobs (in particular, through the same chances of being promoted), the two curves would be confounded with the bisector. This is not the case for our sample. Consider for instance the rank .5 in the wage distribution of job positions. 50% of workers (males or females) are paid more than the wage corresponding to this rank (which is the median). The rank in the wage distribution of males (resp. females) corresponding to the median is .46 (resp. .63). Hence,

[11] Confidence intervals are not reported because they are nearly confounded with the curves as we have wealth of data.

whereas 54% of males get a wage higher than the median, this proportion is only 37% for females. The larger the gap between these proportions at a given rank in the wage distribution of job positions, the less females have access to jobs above this rank compared to males. In the next section, we rely on our model to evaluate the difference in access to jobs between males and females at any given rank in the wage distribution of job positions.

[*Insert Figure* 8]

5 Results

Figure 9 represents the estimator of the access function $\widehat{h}$ and the confidence interval at each rank of the wage distribution of job positions. Recall that $\widehat{h}(u)$ can be interpreted as the gender probability ratio of getting a job at rank u. When $\widehat{h}(u) > 1$, females have a better access to the job than males. When $\widehat{h}(u) < 1$, males have a better access to the job than females. As the access function takes values which are always lower than one, the probability of getting a job at any rank is lower for females than for males. However, the values are close to one for the first ranks, indicating that females and males are treated almost the same way for the less-paid jobs. For instance, the probability of females getting a job at rank .05 is only 12% lower than the probability of males as shown in Table 2. Between the ranks .2 and .8, the access to job slightly decreases for females compared to males. After rank .8, the access function decreases more sharply pointing at the difficulty females have getting hired. The probability of females getting a job at rank .95 is 50% lower than the probability of males.

[*Insert Figure* 9]

[*Insert Table* 2]

We now look at the banking and insurance industries which are closely related as shown by the recent take-over across these two industries. These industries have different wage policies. Banks rely on a job classification and a regulation which are quite rigid as they are inherited from the period when banks belonged to the public sector. Insurance companies give more weight to the individualization of careers (Dejonghes and Gasnier, 1990). We find that there is a sharp constrast in the access function between the two industries. For insurance, the access function decreases sharply from rank 1 to rank .3 pointing at more barriers for females to high-paid jobs than to low-paid jobs (Figure 10). For banking, it decreases very slowly from rank .8 up to the highest rank and the pattern is closer to some uniform discrimination (Figure 11).

We can assess more accurately to what extent there are more barriers to high-paid jobs than to low-paid jobs from the slope of the access function. Indeed, the larger the slope, the larger the difference between the barriers to high-paid jobs and low-paid jobs. We thus estimate a linear specification of the access function, $h(u) = a - b.u$, and compare the value of the slope parameter b for all the pooled industries, banking and insurance (see Appendix B for the details on the procedure). We obtain that for the pooled industries, an increase of one decile in the wage distribution of job positions ($\Delta u = .1$) yields a decrease in the access to

jobs of females relative to males of 2.8% ($b = .28$) as shown in Table 3. Whereas the decrease is smaller in the banking industry at .7%, it is more than two times larger in the insurance industry at 6.0%. Interestingly, a statistical test shows that the linearity of the access function is not rejected at the five percent level for the pooled industries, as well as for banking and insurance. As the slope of the access function is small in the banking industry, we tried to approximate the access function of that sector with a constant specification: $h(u) = \gamma$. The constant is estimated to be .672 and the specification is not rejected at the five percent level. Hence, the access function in the banking sector is nearly constant.

[*Insert Figures* 10 *and* 11]

[*Insert Table* 3]

The example of these two industries confirms how difficult it is to interpret the gender quantile difference expressed as a function of the rank in the gender wage distribution. As shown on Figures 12 and 13, the gender wage difference exhibits a huge increase at the highest ranks. According to the literature, this would suggest more barriers for females to high-paid jobs than to low-paid jobs in the two industries. Whereas for insurance, this interpretation is consistent with the results of our model, this is much more arguable for banking.

[*Insert Figures* 12 *and* 13]

6 Individual and market heterogeneity

So far, we have considered that workers are heterogenous only in the gender dimension in the sense that all workers of a given gender have *ex-ante* the same chances of getting a job of a given rank. However, in our data, workers can differ in age and country of birth, and there is no reason why their access to jobs cannot be influenced by these factors. As a consequence, we propose an extension of the model that takes into account the individual observed characteristics.

Also, we implicitely assumed that all workers compete on the national market. This is arguable as some individuals make their whole career in a large firm which can be considered as an internal market. We show how to rewrite the model and redefine the access function under the alternative assumption that each firm is a separate market and workers within each firm compete with each other but not with outsiders.

We provide some estimations of the access functions for each of these extensions in a last subsection.

6.1 Individual observed characteristics

Males may get the best jobs because they have some specific characteristics which make them more valuable for the manager. We now show how the individual observed heterogeneity can be included in our model and controlled for when estimating the access function. We suppose that an individual i of gender j can be characterized by a vector X_i of observable attributes (different from his gender). We consider for simplicity

that each attribute only takes discrete values. The individual characteristics may directly influence the productivity of the worker and thus the profit of the manager which is respecified as:

$$\Pi_u(i) = \theta_{j(i)}(u|X_i)\exp[\varepsilon_i(u)]$$

where $\theta_j(u|X_i)$ not only captures gender differences in expected profit but also differences related to the worker's characteristics. The taste of the manager for workers may not only depend on their gender but also on their characteristics (possibly in interaction with their gender) so that the utility of the manager is rewritten as:

$$V_u(i) = \ln \mu^*_{j(i)}(u|X_i) + \ln \Pi_u(i)$$

where $\mu^*_j(u|X_i)$ is a taste parameter which can depend on the worker's characteristics. The utility of the manager in reduced form is given by:

$$V_u(i) = \ln \mu_{j(i)}(u|X_i) + \varepsilon_i(u)$$

where $\ln \mu_j(u|X_i) = \ln \mu^*_j(u|X_i) + \ln \theta_j(u|X_i)$ captures all the effects related to the worker's characteristics (including his gender).

For a given job of rank u, $\Omega(u)$ is the set of all available workers whatever their characteristics. An individual applying for the job competes with all the other workers in this set. We assume that the points of the sequence $\{j(i), X_i, \varepsilon_i(u)\}$, $i \in \Omega(u)$ are the points of a Poisson process with intensity measure $\frac{n_j(u|X)}{n_f(u|X)+n_m(u|X)}P(X)\exp(-\varepsilon)d\varepsilon$ where $P(X)$ is the probability of a worker having the characteristics X and $n_j(u|X)$ is the measure of gender-j workers with characteristics X available for a job of rank u. The probability that the worker chosen for the job is of gender j then verifies the formula (5) except that the unit probability is now:

$$\phi_j(u|X_i) = \psi^{-1}(u)\,\mu_j(u|X_i) \tag{15}$$

where $\psi(u)$ is a competition term verifying:

$$\psi(u) = n_f(u)\,E_{X_k}\left[\mu_f(u|X_k)|k \in \Omega_f(u)\right] + n_m(u)\,E_{X_k}\left[\mu_m(u|X_k)|k \in \Omega_m(u)\right] \tag{16}$$

with $\Omega_j(u)$ the set of gender-j workers available for a job of rank u whatever their characteristics such that $\Omega(u) = \Omega_m(u) \cup \Omega_f(u)$, and $n_j(u)$ the measure of workers in this set. The workers included in this set have some characteristics leading to a quality of matches with job positions which is on average lower than the one of workers occupying jobs of higher rank. This is the result of a filtering process where the workers with characteristics yielding better matches with job positions have succeeded more often in getting a job which is better paid. In (16), $E_{X_k}\left[\mu_j(u|X_k)|k \in \Omega_j(u)\right]$ is the average (exponentiated) effect of individual characteristics (including gender) on the utility of the manager for workers available for a job of rank u. When $\mu_j(u|X_i) = \mu_j(u)$, the formula (15) collapses into (6) which corresponds to the case where there is no individual observed heterogeneity.

For gender j, we now determine the dynamics of the measure of individuals with characteristics X available for a job of rank u. In fact, the measure of individuals available for a job of rank $u - du$ can be deduced from the measure of individuals available for a job of rank u substracting those who found a job of rank between $u - du$ and u:

$$n_j\left(u - du\,|X\right) = n_j\left(u\,|X\right) - n_j\left(u\,|X\right)\phi_j\left(u\,|X\right)du$$

Having $du \to 0$, we get:

$$n_j'\left(u\,|X\right) = \phi_j\left(u\,|X\right)n_j\left(u\,|X\right) \tag{17}$$

This formula is similar to (8) for a homogenous population of workers except that the unit probability now depends on the measures of workers in competition for the job with characteristics other than X. It is possible to show the following existence theorem which proof is relegated in Appendix A:

Theorem 2 *Suppose that X can only take a finite number of values X^p, $p = 1, ..., P$; $\mu_m\left(\cdot\,|X^p\right)$ and $\mu_f\left(\cdot\,|X^p\right)$ are C^1 on $(0, 1]$ for each p; and there is a constant $c > 0$ such that $\mu_m\left(u\,|X^p\right) > c$ and $\mu_f\left(u\,|X^p\right) > c$ for all $u \in (0, 1]$ and all p. Then there is a unique $2P$-uplet $\left\{n_f\left(\cdot\,|X^p\right), n_m\left(\cdot\,|X^p\right)\right\}_{p=1,...,P}$ verifying (17) where $\phi_j\left(\cdot\right)$ is given by (15).*

We can introduce an access function for each subgroup of the population characterized by a set of characteristics X:

$$h\left(u\,|X\right) \equiv \frac{\mu_f\left(u\,|X\right)}{\mu_m\left(u\,|X\right)} \tag{18}$$

Using equations (15) and (18), the access function can be rewritten as:

$$h\left(u\,|X\right) = \frac{\phi_f\left(u\,|X\right)}{\phi_m\left(u\,|X\right)} \tag{19}$$

This formula is similar to the one obtained for a homogenous population. Interestingly, even if the workers with characteristics X compete with some workers having other characteristics, $h\left(u\,|X\right)$ can be rewritten as the unit probability ratio of females and males with characteristics X getting the job of rank u. This is because females and males compete with exactly the same pool of individuals and the competition terms in the unit probabilities of the two genders are the same.

Two different empirical exercices can be conducted in this setting. If the population subgroups for every set of characteristics are large enough, it is possible to estimate an access function for each subgroup. The access functions can then be compared across groups to assess whether, for females, the barriers to high-paid jobs vary with characteristics. Another more general exercise consists in estimating an access function for the whole population which is net of the effect of individual characteristics. Such an access function first need to be defined. We make the additional assumption that the (exponentiated) effect of individual characteristics (including gender) on the utility of the manager takes the following semi-parametric multiplicative form:

$$\mu_j\left(u\,|X\right) = \widetilde{\mu}_j\left(u\right)\exp\left(X\delta_j\right) \tag{20}$$

Under this assumption, the probability ratio of getting a job of rank u for a female and a male with the characteristics of the reference category (i.e. such that $X = 0$) is: $\widetilde{h}(u) = \widetilde{\mu}_f(u)/\widetilde{\mu}_m(u)$. We call $\widetilde{h}(\cdot)$ the *net access function* and show how it is related to the access function of the whole population which was defined in section 2 (re-labelled the *gross access function*). In fact, the unit probability of getting a job of rank u for available gender-j workers verifies:

$$\begin{aligned} \phi_j(u) &= E_{X_k}\left[\phi_j(u|X_k)|k \in \Omega_j(u)\right] \\ &= \psi^{-1}(u)\,\widetilde{\mu}_j(u)\,E_{X_k}\left[\exp(X_k\delta_j)|k \in \Omega_j(u)\right] \end{aligned} \tag{21}$$

From (10) and (21), we get the following relationship:

$$h(u) = r(u)\,\widetilde{h}(u) \text{ with } r(u) = \frac{E_{X_k}\left[\exp(X_k\delta_f)|k \in \Omega_f(u)\right]}{E_{X_k}\left[\exp(X_k\delta_m)|k \in \Omega_m(u)\right]} \tag{22}$$

The gross access function $h(\cdot)$ can thus be decomposed multiplicatively into the net access function $\widetilde{h}(\cdot)$ and a corrective term corresponding to the gender ratio of the average (exponentiated) individual effects $r(\cdot)$. There are two reasons for this ratio to differ from one: available male and female workers can have different characteristics, and the return of the characteristics can differ across genders. The ratio varies across ranks as the result of a filtering process. Among the workers of gender j, those with the highest expected value for the manager (ie. those for which the effect of individual characteristics $X_i\delta_j$ is the highest) are usually going to find a job first. A worker finding a job of a given rank is not used to compute the ratio at lower ranks.

We can construct an estimator of the net access function using (22). We have: $\widetilde{h}(u) = h(u)/r(u)$. An estimator of the gross access function is given by (14). We need an estimator of the gender ratio of the average (exponentiated) individual effects. We first explain how to estimate the coefficients of the individual variables for each gender. As we have seen in section 3, the model can be seen formally as a duration model where the time line is the axis of ranks running from $u = 1$ to $u = 0$. For each gender j, the unit probability of getting a job of rank u is:

$$\phi_j(u|X) = \psi^{-1}(u)\,\widetilde{\mu}_j(u)\exp(X\delta_j) \tag{23}$$

which can be re-interpreted as an instantaneous hazard corresponding to a Cox model. It is possible to estimate the coefficients of the individual variables from the partial likelihood computed for each of the two gender subsamples. Denote by $P_{ij}(u|X_i)$ the probability of a gender-j worker i with characteristics X_i of getting a job of rank in the interval $[u - du, u]$ conditionally on someone in the set of available workers $\Omega_j(u)$ getting a job of rank in that interval. This probability can be written as:

$$P_{ij}(u|X_i) = \frac{\phi_j(u|X_i)}{\sum\limits_{k\in\Omega_j(u)}\phi_j(u|X_k)} = \frac{\exp(X_i\delta_j)}{\sum\limits_{k\in\Omega_j(u)}\exp(X_k\delta_j)} \tag{24}$$

The coefficients δ_j can be estimated maximizing the partial likelihood $\frac{1}{N_j}\sum\limits_{i|j(i)=j}\ln P_{ij}(u|X_i)$. We denote

by $\widehat{\delta}_j$ the corresponding estimator.[12] We can then recover an estimator of the gender ratio of the average (exponentiated) individual effects. Indeed, for gender j, an estimator of $E_{X_k}\left[\exp\left(X_k\delta_j\right)|k\in\Omega_j\left(u\right)\right]$ at any observed rank $u_i\in\left\{\frac{1}{N_j},\frac{2}{N_j},...,1\right\}$ is given by:

$$\widehat{E}_{ij}=\frac{1}{N_j\left(u_i\right)}\sum_{k\in\Omega_j(u_i)}\exp\left(X_k\widehat{\delta}_j\right) \tag{25}$$

where $N_j\left(u_i\right)$ is the number of gender-j workers in the sample available for the job of rank u_i. It is possible to construct a smooth estimator at any rank u using a kernel:

$$\widehat{E}_j\left(u\right)=\sum_{i|j(i)=j}p_{ij}\widehat{E}_{ij}\text{ with }p_{ij}=\frac{K\left(\frac{u-u_i}{h_{jN}}\right)}{\sum\limits_{i|j(i)=j}K\left(\frac{u-u_i}{h_{jN}}\right)} \tag{26}$$

where $K\left(\cdot\right)$ is an Epanechnikov Kernel and h_{jN} is the bandwidth chosen to take the value given by the rule of thumb (Silverman, 1986). An estimator of the gender ratio of the average (exponentiated) individual effects is then:

$$\widehat{r}\left(u\right)=\frac{\widehat{E}_f\left(u\right)}{\widehat{E}_m\left(u\right)} \tag{27}$$

We finally get an estimator of the net access function: $\widehat{\widetilde{h}}\left(u\right)=\widehat{h}\left(u\right)/\widehat{r}\left(u\right)$.

6.2 Segmented markets

We have supposed so far that all the workers compete for jobs on the national market. We now consider the alternative situation where there are Z firms in the economy and each firm consists in a submarket of several jobs. Workers compete for job positions on each submarket, but there is no competition across submarkets. The assignment of workers to jobs within each firm is of the same type as the assignment on the national market which has been described in the previous subsection. For a given firm z, the access function for a subgroup of the population with characteristics X in the firm is defined as:

$$h^z\left(u\,|X\right)\equiv\frac{\mu_f^z\left(u\,|X\right)}{\mu_m^z\left(u\,|X\right)}$$

where u corresponds to the rank in the wage distribution of jobs positions *within the firm*, and $\mu_j^z\left(u\,|X\right)$ is the taste parameter corresponding to gender j and characteristics X which enters the utility of the manager written in reduced form.

We want to recover an access function for the whole population which is net of the effect of individual characteristics. We first make the additional assumption that the (exponentiated) effect of individual characteristics (including gender) on the utility of the manager takes the following semi-parametric multiplicative

[12] If the sets of coefficients obtained for the two genders are very similar, one may want to impose the restriction: $\delta_j=\delta$. The coefficients can then be estimated maximizing the partial likelihood stratified by gender on the sample of all workers (see Ridder and Tunali, 1999, for details).

form for each firm and job:[13]

$$\mu_j^z\left(u\,|X\right)=\widetilde{\mu}_j^z\left(u\right)\exp\left(X\delta_j\right) \tag{28}$$

Under this assumption, the probability ratio of getting a job of rank u in firm z for a female and a male with the characteristics of the reference category (i.e. such that $X=0$) is: $\widetilde{h}^z\left(u\right)=\widetilde{\mu}_f^z\left(u\right)/\widetilde{\mu}_m^z\left(u\right)$, the net access function of the firm computed at rank u. As we are interested in recovering an average net access function for the whole population of workers, we focus on a weighted average of the net access functions of firms where the weight is the proportion of workers in each firm (denoted p^z):

$$\widetilde{h}\left(u\right)=E_z\left[p^z\widetilde{h}^z\left(u\right)\right] \tag{29}$$

In order to estimate the average net access function, we need to construct some estimators of the proportion of workers and the net access function of each firm. An estimator of the proportion of workers in firm z is given by $\widehat{p}^z=\frac{N^z}{N}$ where N^z is the number of workers in the firm. We can also construct an estimator of the net access function of the firm from its relationship with the gross access function of the firm in the same way as when workers compete on the national market. The relationship is given by:

$$h^z\left(u\right)=r^z\left(u\right)\widetilde{h}^z\left(u\right)\ \text{ with } r^z\left(u\right)=\frac{E_{X_k}\left[\exp\left(X_k\delta_f\right)\Big|k\in\Omega_f^z\left(u\right)\right]}{E_{X_k}\left[\exp\left(X_k\delta_m\right)|k\in\Omega_m^z\left(u\right)\right]} \tag{30}$$

The estimator of the net access function of the firm is derived from some estimators of the gross access function and the corrective term accounting for the individual observed heterogeneity. The gross access function of the firm can be estimated using the approach of Section 3, and the estimator is denoted by $\widehat{h}^z\left(u\right)$. The corrected term can be estimated in two stages. First, the coefficients of individual variables are computed maximizing the partial likelihood stratified by firm on the subsample of the gender (Ridder and Tunali, 1999). Denote by $P_{ij}^z\left(u\,|X_i\right)$ the probability of a gender-j worker i in firm z with characteristics X_i of getting a job of rank in the interval $[u-du,u]$ conditionally on someone in the risk set $\Omega_j^z\left(u\right)$ getting a job of rank in that interval. This probability can be written as:

$$P_{ij}^z\left(u\,|X_i\right)=\frac{\phi_j^z\left(u\,|X_i\right)}{\sum\limits_{k\in\Omega_j^z(u)}\phi_j^z\left(u\,|X_k\right)}=\frac{\exp\left(X_i\delta_j\right)}{\sum\limits_{k\in\Omega_j^z(u)}\exp\left(X_k\delta_j\right)} \tag{31}$$

The coefficients δ_j are then estimated maximizing the partial likelihood $\frac{1}{N_j}\sum\limits_{i|j(i)=j}\ln P_{ij}^{z(i)}\left(u\,|X_i\right)$ and we denote by $\widehat{\delta}_j$ the corresponding estimator. For each firm z, we then apply the strategy explained in subsection 6.1 to recover an estimator of the gender ratio of the average (exponentiated) individual effects denoted by $\widehat{r}^z\left(u\right)$. At a given rank u, an estimator of the net access function of a given firm is then $\widehat{\widetilde{h}}^z\left(u\right)=\widehat{h}^z\left(u\right)/\widehat{r}^z\left(u\right)$, and an estimator of the average net access function is given by:

$$\widehat{\widetilde{h}}\left(u\right)=\sum_z\widehat{p}^z\widehat{\widetilde{h}}^z\left(u\right) \tag{32}$$

[13] In particular, the coefficients of the explanatory variables are supposed to be the same across firms. This assumption was necessary to avoid some estimation problems due to the lack of observations in some firms to identify the coefficients of individual variables.

For the sake of comparison, we will also compute an average gross access function in our application which is obtained by replacing the estimated net access function of each firm in equations (32) by their estimated gross access function.

6.3 Results

We now present the results for the two extensions of the model. We first comment the estimated net access function obtained when workers compete on the national market. The individual explanatory variables included in the specification are some dummies for each age between 41 and 45 (the reference category being 40), and a dummy for being born in a foreign country.[14] Figure 14 shows that the net access function is just above the gross access function. However, the two curves are very close, which is consistent with the average effect of individual characteristics being similar for males and females available for a job at each rank. The specific industries of banking and insurance also exhibit a pattern where the gross and net access functions are nearly confounded (see Graphs A.1 and A.2 in appendix). Overall, the individual observed heterogeneity captured by the variables in our data does not explain much of the gross access function.

[*Insert Figure* 14]

We then turn to the estimation of the access function when competition occurs within each firm.[15] We limit our sample to large firms employing 150 full-time executives aged $40-45$ or more. Indeed, many workers getting their first job in a large firm make their whole career in that firm. In our sample, only .5% of firms are large, but they employ 33% of workers. The median wage in large firms reaches 114 euros, which is a bit larger than for the whole sample (109 euros). By contrast, the wages are far less dispersed with a standard deviation of 132 euros compared to 602 euros for the whole sample. Figure 15 shows that the average access function when workers compete on each submarket has a profile quite similar to the access function when all workers of large firms compete on a common national market,[16] although it is smoother probably because the firm heterogeneity in the level of wages is conditioned out in the estimation.[17] The similarity between the two curves is confirmed when evaluating some linear specifications of the access function in the two cases.[18] The estimated specifications are respectively $h(u) = .74 - .09u$ and $h(u) = .69 - .05u$ which are very close. Interestingly, our linear specification test is rejected only when competition occurs on the national market and not when it occurs on segmented submarkets. This difference arises because we conditioned out the firm

[14] The estimated coefficients of individual variables are reported in Tables A.1 and A.2.

[15] The estimated coefficients of individual variables are reported in Tables A.3 and A.4.

[16] As the gross and net access functions are usually close, the gross access functions are the ones used when comparing the results obtained when the market is national and when the market is segmented.

[17] We did not represent the confidence intervals of the three curves on Figure 5 otherwise the figure would be too difficult to read. The values of the curves at a given rank are usually not significantly different.

[18] The technical details are relegated in Appendix B.

heterogeneity only when workers compete on segmented submarkets.

[*Insert Figure* 15]

We performed the same exercise for the insurance and banking industries (see Figures 16 and 17). For banking, the average access function when competition occurs on each segmented submarket has a profile similar to the access function when competition occurs on the national market. Curves seem to differ for insurance. We estimated a linear specification of the two access functions to ease the comparison. We obtained for insurance respectively $h(u) = .93 - .66u$ did competition occur on the national market and $h(u) = .74 - .41u$ when it occurs on each segmented submarket. Hence, the access function would begin at a lower level when competition occurs on each segmented submarket but its slope would be less steep, suggesting less barriers for females in the access to high-paid jobs. The difference between the two access functions can be explained by some heterogeneity in the level of wages among firms. This heterogeneity is wiped out only when competition is supposed to occur within firms (this is because we conduct some within-firm estimations in the spirit of what is done for linear panel data models). In any case, the differences in barriers to high-paid jobs and low-paid jobs are more important in the insurance industry than in the banking industry and for pooled industries. Our results are thus qualitatively robust to the assumption on the extent of the market where workers compete for jobs.

[*Insert Figures* 16 *and* 17]

7 Conclusion

In this paper, we proposed a job assignment model where there is a gender difference in access to jobs. Males and females compete for some heterogenous job positions characterized by different levels of wages. Workers want to get hired for the best-paid jobs. There are barriers which make females less likely to get some of the job positions than males. Our model predicts how these barriers yield differences in the wage distributions of the two genders. Simulations show that even if the gender relative access is constant across jobs, the model can generate a gender quantile difference increasing with the rank. The literature would conclude to a glass ceiling whereas there is none. This questions the validity of the usually glass ceiling interpretation.

We then used a structural relationship of the model to estimate the gender difference in access to jobs at each rank of the wage distribution of positions. Our model was estimated on the 2003 *Déclarations Annuelles des Salaires* (DADS) which is *exhaustive* for all public and private firms. We found that at the bottom of the wage distribution of positions, the probability of females getting a given job is 12% lower than the probability of males. The difference between these probabilities is far larger at the top of the wage distribution of positions and climbs to 50%. These results are in line with a lower access to high-paid jobs for females. They are robust to the inclusion of individual observed heterogeneity in the analysis and to different assumptions on the extent of the market on which workers are in competition for the job positions.

Our model was initially designed to study the consequences of gender differences in access to jobs on the ranks of males and females in the wage distribution of job positions. Alternatively, it could be applied to other subgroups of the population such as the French and the immigrants. Also, it could be interesting to extend our model to a dynamic setting to study the changes in the ranks of males and females in the wage distribution of job positions through job changes, promotions and lay-offs.

References

[1] Albrecht J., Björklund A. and S. Vroman (2003), "Is There a Glass Ceiling in Sweden?", *Journal of Labour Economics*, 21(1), pp. 145-177.

[2] Arrow J.K. (1971), *The theory of discrimination*, Working paper n°30A, Princeton University, in O. Aschenfelter & A. Rees Eds, 1974, Discrimination in Labor Markets, NJ: Princeton UP.

[3] Bartel A. (2004), "Human Resource Management and Organizational Performance: Evidence from Retail Banking", *Industrial and Labor Relations Review*, 57(2), 181-203.

[4] Becker G. (1971), *The Economics of Discrimination*, Second Edition, The University of Chicago Press, 178p.

[5] Cahuc P. and A. Zylverberg (2004), *Labour Economics*, 872p, MIT Press.

[6] Coate S. and G. Loury (1993), "Will affirmative action policy eliminate negative stereotypes?", *American Economic Review*, 83(5), pp. 1220-40

[7] Combes P., Duranton G., Gobillon L., Puga D. and S. Roux (2009), "The productivity advantages of large cities: distinguishing agglomeration from firm selection", *CEPR Working paper 7191.*

[8] Costrell R. and G. Loury (2004), "Distribution of Ability and Earnings in a Hierarchical Job Assignment Model", *Journal of Political Economy*, 112(6), pp. 1322-1363.

[9] Dagsvik J. (1994), "Discrete and Continuous Choice, Max-Stable Processes, and Independence from Irrelevant Attributes", *Econometrica*, 62(5), pp. 1179-1205.

[10] Dejonghe V. and C. Gasnier (1990), "Pratiques salariales et gestion du personnel dans les banques et les assurances", *Travail et Emploi*, 45(3), pp. 40-46.

[11] Donald S., Green D. and H. Paarsch (2000), "Differences in Wage Distributions between Canada and the United States: An Application of a Flexible Estimator Distribution Functions in the Presence of Covariates", *The Review of Economic Studies*, pp. 609-633.

[12] Fortin N. and T. Lemieux (2002), "Are Women's Wage Gains Men's Losses? A Distributional Test", *American Economic Review*, 90(2), pp. 456-460.

[13] Killingsworth M. and C. Reimers (1983), "Race, Ranking, Promotions, and Pay at a Federal Facility: A Logit Analysis", *Industrial and Labor Relations Review*, 36, pp. 92-107.

[14] Knott M. (1974), "The distribution of the Cramér-von Mises statistic for small sample sizes", *Journal of the Royal Statistical Society B*, 36(3), pp. 430–438.

[15] Lancaster T. (1990), *The Econometric Analysis of Transition Data*, Cambridge University Press: Cambridge.

[16] Lazear E. and S. Rosen. (1990), "Male-Female Wage Differentials in Job Ladders", *Journal of Labor Economics*, 8, pp. 106-123.

[17] Pekkarinen T. and J. Vartiainen, (2006), "Gender differences in promotion on a job ladder: Evidence from Finnish metalworkers", *Industrial and Labor Relations Review*, 59(2), pp. 285-301.

[18] Phelps E. (1972), "The statistical theory of racism and sexism", *American Economic Review*, 62, pp. 639-651.

[19] Press W., Teukolsky S., Vettering W. and B. Flannery (1992), *Numerical Recipes in C*, Second Edition, Cambridge University Press, 994p.

[20] Ridder G. and I. Tunali (1999), "Stratified partial likelihood estimation", *Journal of Econometrics*, 92(2), 193-232.

[21] Sattinger M. (1993), "Models of the Distribution of Earnings", *Journal of Economic Literature*, 31(2), pp. 831-880.

[22] Silverman B. (1986), *Density Estimation for Statistics and Data Analysis*, Chapmal & Hall, 175p.

[23] Teulings C. (1995), "The Wage Distribution in a Model of the Assignment of Skills to Jobs", *Journal of Political Economy*, 103(2), pp. 280-315.

[24] Van der Vaart A.W. (1998), *Asymptotic Statistics*, 443p, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

[25] Winter-Ebmer R. and J. Zweimüller (1997), "Unequal assignment and unequal promotion in job ladders", *Journal of Labor Economics*, 15(1), pp. 43-71.

Appendix A : Proof of Theorems 1 and 2

First note that Theorem 1 is a special case of Theorem 2 where P is fixed to one. Hence, we only presents the proof of Theorem 2. The proof revolves around the application of the Cauchy-Lipschitz theorem. Plugging (16) into (15), and plugging the resulting expression into (17), we get for any j and p:

$$n_j'\left(u\left|X^p\right.\right) = \frac{n_j\left(u\left|X^p\right.\right)\mu_j\left(u\left|X^p\right.\right)}{\sum\limits_{j=f,m;\ q=1,\ldots,P} n_j\left(u\left|X^q\right.\right)\mu_j\left(u\left|X^q\right.\right)} \tag{33}$$

Introduce the vectors

$$\begin{aligned}\overline{\mu}(u) &= \left[\mu_f\left(u\left|X^1\right.\right),..,\mu_f\left(u\left|X^P\right.\right),\mu_m\left(u\left|X^1\right.\right),...,\mu_m\left(u\left|X^P\right.\right)\right]' \\ \overline{n}(u) &= \left[n_f\left(u\left|X^1\right.\right),..,n_f\left(u\left|X^P\right.\right),n_m\left(u\left|X^1\right.\right),...,n_m\left(u\left|X^P\right.\right)\right]'\end{aligned}$$

A stacked version of (33) is given by:

$$\overline{n}'(u) = g\left(u,\overline{n}(u)\right) \tag{34}$$

with:

$$g\left(u,\overline{n}(u)\right) = \frac{\overline{n}(u).*.\overline{\mu}(u)}{\left\langle\overline{n}(u),\overline{\mu}(u)\right\rangle}$$

where $\langle\cdot,\cdot\rangle$ denotes the Euclydian scalar product and for any two vectors V_1 and V_2 of same dimension, $V_1.*.V_2$ is the vector where any element i is the product of the elements i of V_1 and V_2.

The equation (34) is a first-order differential equation. The denominators of all elements of $g(\cdot,\cdot)$ are strictly positive on $\widetilde{\Phi} = (0,1]\times\left[0,n_f\left(X^1\right)\right]\times...\times\left[0,n_f\left(X^P\right)\right]\times\left[0,n_m\left(X^1\right)\right]\times...\times\left[0,n_m\left(X^P\right)\right]$ where $n_j\left(X^p\right)$ is the measure of gender-j workers with characteristics X^p. This is because there is a constant $c>0$ such that $\mu_m\left(u\left|X^p\right.\right)>c$ and $\mu_f\left(u\left|X^p\right.\right)>c$ for all p and all $u\in(0,1]$. As $\mu_m\left(\cdot\left|X^p\right.\right)$ and $\mu_f\left(\cdot\left|X^p\right.\right)$ are C^1 on $(0,1]$ for all p, it is then straightforward to show that $g(\cdot,\cdot)$ is C^1 on $\widetilde{\Phi}$. This yields that on any compact set $[\varepsilon,1]\times\left[0,n_f\left(X^1\right)\right]\times...\times\left[0,n_f\left(X^P\right)\right]\times\left[0,n_m\left(X^1\right)\right]\times...\times\left[0,n_m\left(X^P\right)\right]$, $g(\cdot,\cdot)$ is Lipshitzienne and (34) has a unique solution for $\overline{n}(\cdot)$ on $[\varepsilon,1]$. As this is true for ε arbitrarily close to zero, (34) has a unique solution for $\overline{n}(\cdot)$ on $(0,1]$.

Appendix B : Linear access function

In this appendix, we explain how to approximate the gross access function by a linear function of ranks. We estimate a specification of the form: $h(u) = a - b.u$ and test whether this specification fits the data. This is done in the case of a national job market and some separate firm job submarkets.

B.1. National market

The random variable corresponding to the rank of a gender-j worker in the wage distribution of job positions is denoted U_j. Its cumulative function is given by: $F_{U_j}(u) = \frac{n_j(u)}{n_j}$. Its quantile function is $u_j(\cdot)$. We have by definition:

$$v = F_{U_j}[u_j(v)]$$

From this equation, we get:

$$F_{U_f}[u_f(v)] = F_{U_m}[u_m(v)] \tag{35}$$

We use $\frac{N_j}{N}$ where $N = N_f + N_m$ as an estimator of n_j. For a given linear specification of $h(u)$, we can solve the model for $n_f(u)$ and deduce $n_m(u)$ summing (8) for the two genders and integrating between 0 and u, as we get the equality $n_m(u) = u - n_f(u)$. We can then deduce the quantile function of U_j as it writes: $u_j(v) = {n_j}^{-1}(n_j.v)$. The parameters a and b are estimated minimizing the distance between the left and right-hand sides of (35) after replacing F_{U_j}, $j \in \{m, f\}$ by their empirical counterparts. Denoting $\theta = (a, b)$, the minimization program is:

$$\min_{\theta} C(\theta) \text{ with } C(\theta) = \int_0^1 \left[\widehat{F}_{U_f}[u_f(v)] - \widehat{F}_{U_m}[u_m(v)]\right]^2 du \tag{36}$$

Details on how to evaluate the minimization criterium are given in Combes et al. (2009).

It is possible to use the minimization criterium to conduct a specification test. We have $\frac{N_j}{N} \xrightarrow{P} p_j$ (the proportion of gender-j workers in the population) where $N = N_f + N_m$ and $p_f + p_m = 1$. Using Donsker's theorem and Slutsky's lemma (see Van der Vaart, 1998, example 20.11), we get:

$$N^{1/2}\begin{pmatrix} \widehat{F}_{U_f}[u_f(v)] - v \\ \widehat{F}_{U_m}[u_m(v)] - v \end{pmatrix} \Longrightarrow \begin{pmatrix} \frac{1}{\sqrt{p_f}} B_f(v) \\ \frac{1}{\sqrt{p_m}} B_m(v) \end{pmatrix} \tag{37}$$

where $B_f(\cdot)$ and $B_m(\cdot)$ are some independent Brownian bridges.

Applying the continuous fonction $\Psi(x_1, x_2) = (x_1 - x_2)^2$ to (37), we get:

$$N\left[\widehat{F}_{U_f}[u_f(v)] - \widehat{F}_{U_m}[u_m(v)]\right]^2 \Longrightarrow \left(\frac{1}{p_f} + \frac{1}{p_m}\right) B(v)^2 \tag{38}$$

where $B(v) = \left(\frac{1}{p_f} + \frac{1}{p_m}\right)^{-1/2} \left[\frac{1}{\sqrt{p_f}} B_f(v) - \frac{1}{\sqrt{p_m}} B_m(v)\right]$. It is easy to show that $B(\cdot)$ is a Brownian bridge. Indeed, $B(\cdot)$ is Gaussian by construction and we have:

$$\begin{aligned} cov\left[B(u), B(v)\right] &= \left(\frac{1}{p_f} + \frac{1}{p_m}\right)^{-1} \left[\frac{1}{p_f} cov\left[B_f(u), B_f(v)\right] + \frac{1}{p_m} cov\left[B_m(u), B_m(v)\right]\right] \\ &= \left(\frac{1}{p_f} + \frac{1}{p_m}\right)^{-1} \left[\frac{1}{p_f}(u \wedge v - uv) + \frac{1}{p_m}(u \wedge v - uv)\right] \\ &= u \wedge v - uv \end{aligned}$$

Integrating (38) over the $[0,1]$ interval, we obtain:

$$N\left(\frac{1}{p_f} + \frac{1}{p_m}\right)^{-1} C(\theta) \Longrightarrow \int_0^1 B(v)^2 \, dv$$

and the right-hand side follows a Cramer Van-Mises statistic which threshold at the 5% level is .46136 (see Knott, 1974). We can approximate the left-hand side replacing p_j by $\frac{N_j}{N}$ and θ by $\widehat{\theta}$, and then conduct a specification test where the hypothesis we test is the equality (35).

B.2. Separated market for each firm

For any firm z, we consider that the gross access function takes the linear form $h^z(u) = a - b.u$. We then have:

$$d^z(v) = 0 \text{ with } d^z(v) = F^z_{U^z_f}\left[u^z_f(v)\right] - F_{U^z_m}\left[u^z_m(v)\right] \tag{39}$$

where $F^z_{U^z_j}(\cdot)$ is the cumulative function of the random variable U^z_j corresponding to the rank of a gender-j worker in the wage distribution of job positions in firm z, and $u^z_j(\cdot)$ is the corresponding quantile function. Denote:

$$\widehat{d^z}(u) = \widehat{F^z}_{U^z_f}\left[u^z_f(v)\right] - \widehat{F^z}_{U^z_m}\left[u^z_m(v)\right]$$

where $\widehat{F^z}_{U^z_j}(\cdot)$ is the empirical counterpart of $F^z_{U^z_j}(\cdot)$. We can recover an estimator of the parameters θ using the minimization program:

$$\min_\theta C^Z(\theta) \text{ with } C^Z(\theta) = \sum_z P^z \int_0^1 \widehat{d^z}(u)^2 \, du$$

where P^z is a weight (in practice, it is the proportion of workers in firm z). The minimization criterium can be computed in a way similar to the one in (36).

Once again, it is possible to use the minimization criterium to conduct a specification test. We have $\frac{N^z_j}{N} \xrightarrow{P} p^z_j$ (the proportion of gender-j workers in firm z) where $N = \sum_{z=1}^{Z} N^z_f + \sum_{z=1}^{Z} N^z_m$ and $\sum_{z=1}^{Z} p^z_f + \sum_{z=1}^{Z} p^z_m = 1$.

Using again Donsker's theorem and Slutsky's Lemma, we get:

$$N^{1/2}\begin{pmatrix} \widehat{F^1}_{U_f^1}\left[u_f^1(v)\right]-v \\ \widehat{F^1}_{U_m^1}\left[u_m^1(v)\right]-v \\ ... \\ \widehat{F^Z}_{U_f^Z}\left[u_f^Z(v)\right]-v \\ \widehat{F^Z}_{U_m^Z}\left[u_m^Z(v)\right]-v \end{pmatrix} \Longrightarrow \begin{pmatrix} \frac{1}{\sqrt{p_f^1}}B_f^1(v) \\ \frac{1}{\sqrt{p_m^1}}B_m^1(v) \\ ... \\ \frac{1}{\sqrt{p_f^Z}}B_f^Z(v) \\ \frac{1}{\sqrt{p_m^Z}}B_m^Z(v) \end{pmatrix} \tag{40}$$

where $B_j^z(\cdot)$, with $j \in \{f, m\}$ and $z \in \{1, ..., Z\}$ are some independent Brownian Bridges.

Applying the continuous fonction $\Psi(x_1, ..., x_{2Z}) = \sum_{z=1}^{Z} P^z.(x_{2z-1} - x_{2z})^2$ to (40), we get:

$$N\sum_{z=1}^{Z} P^z.\widehat{d^z}(v)^2 \Longrightarrow \sum_{z=1}^{Z} P^z.\left(\frac{1}{p_f^z}+\frac{1}{p_m^z}\right)B^z(v)^2$$

where $B^z(\cdot)$, $z \in \{1, ..., Z\}$ are some independent Brownian bridges. Integrating this equation over the $[0, 1]$ interval, we obtain:

$$N.\left[\sum_{z=1}^{Z} P^z.\left(\frac{1}{p_f^z}+\frac{1}{p_m^z}\right)\right]^{-1} C^Z(\theta) \Longrightarrow \int_0^1 B(v)^2\,dv$$

where the right-hand side follows a Cramer Van-Mises statistic. We can approximate the left-hand side replacing p_j^z by $\frac{N_j^z}{N}$ and θ by $\widehat{\theta}$, and then conduct a specification test where the hypothesis we test is the set of equalities (39).

Figure 1: Gender quantile and rank differences in two different contexts

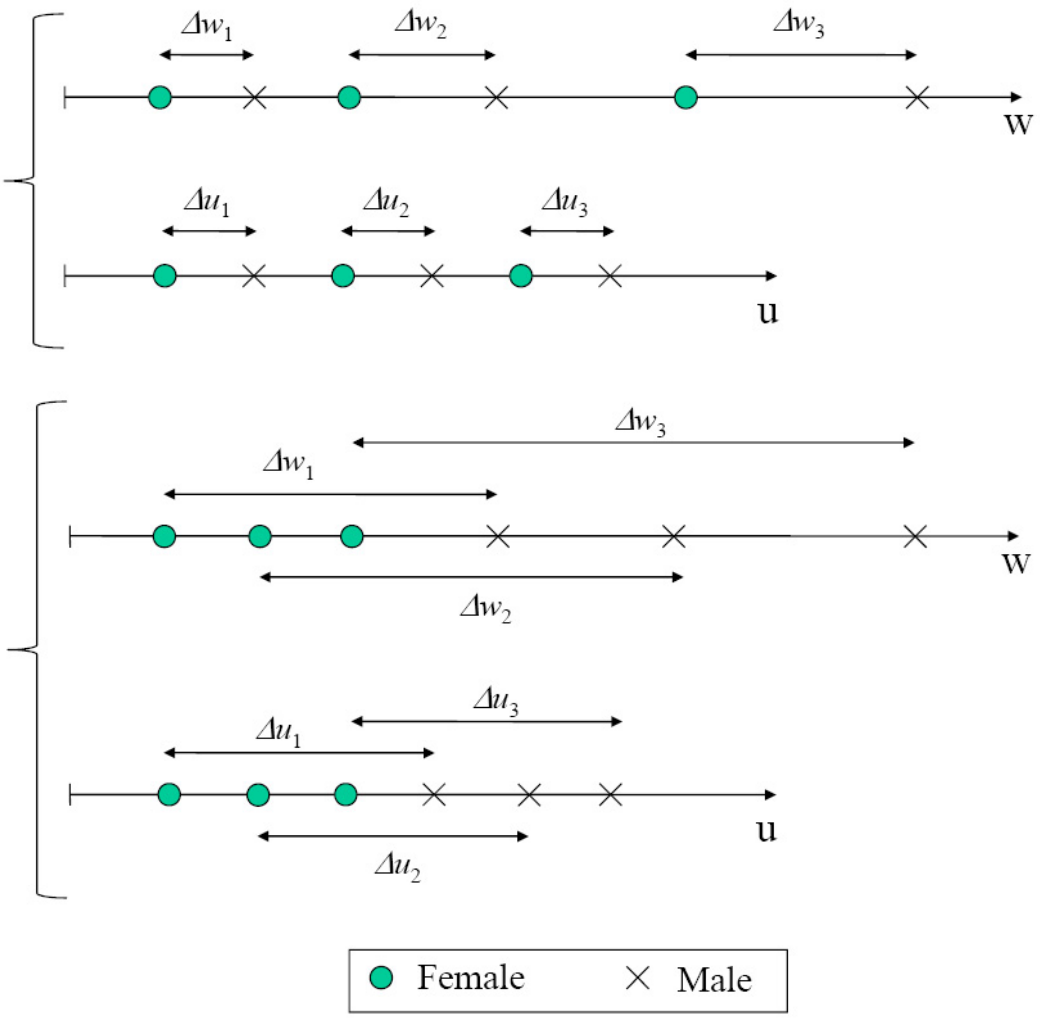
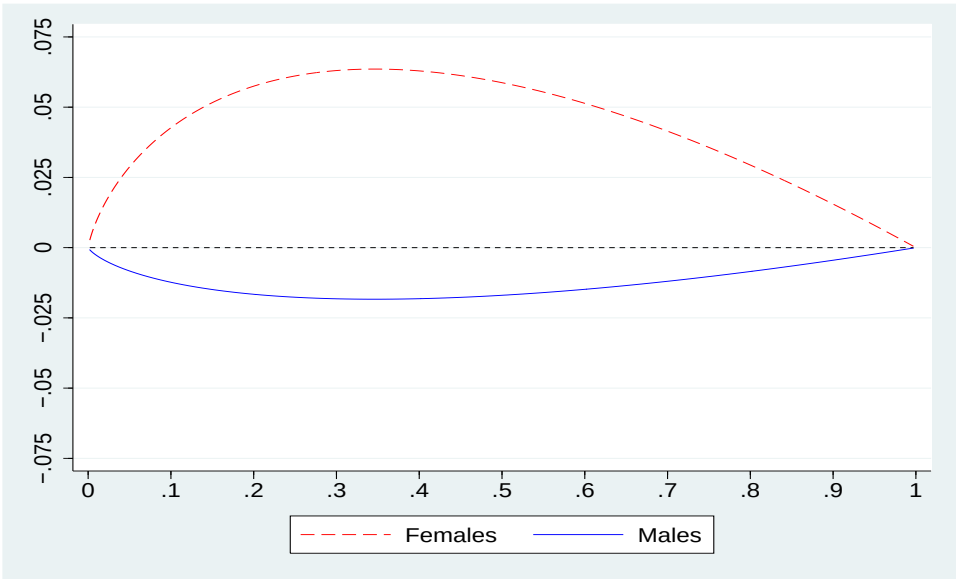
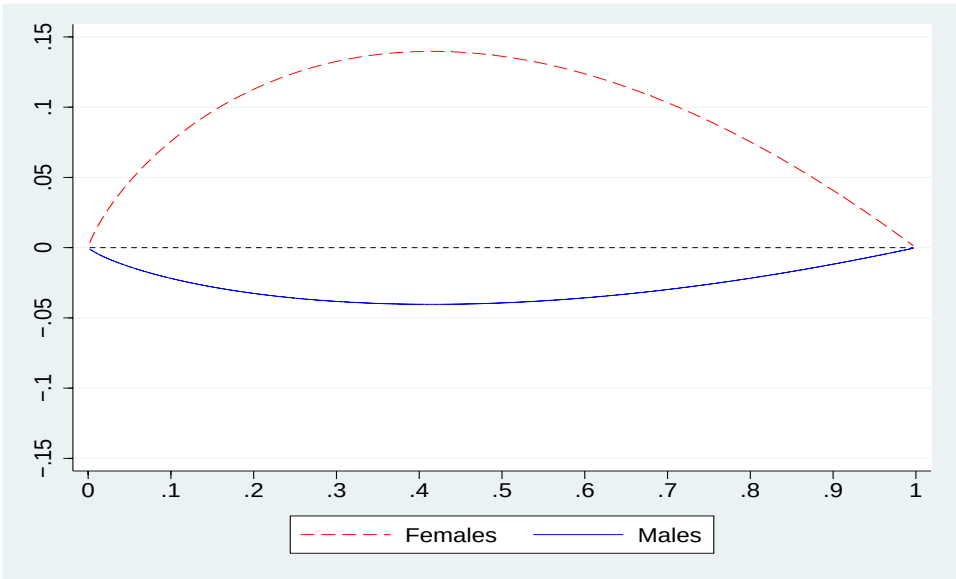


Figure 2: Difference between gender rank and job position rank: $v_j(u) - u$, $h(u) = .8$



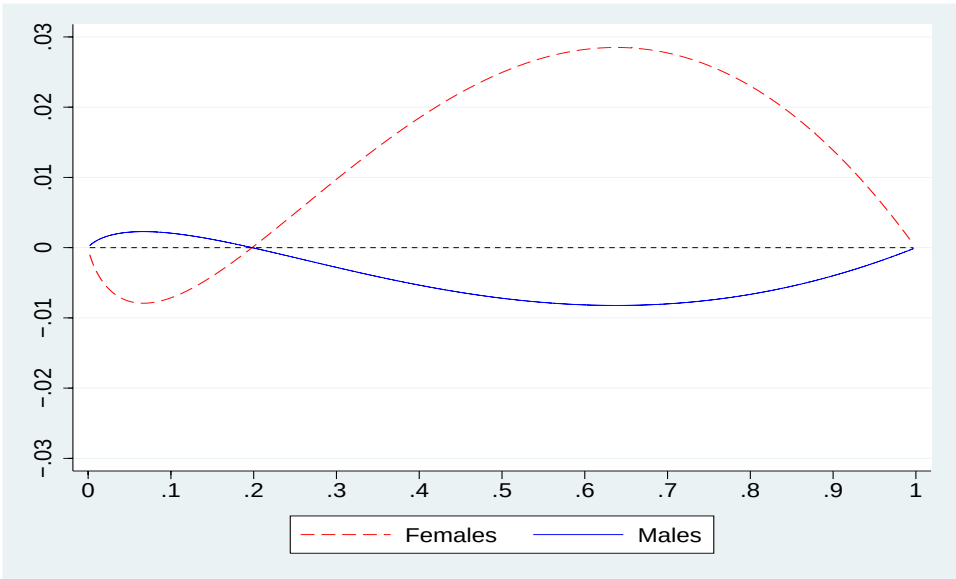
Note: for a job of rank u in the wage distribution of positions, $v_j(u) = n_j(u)/n_j$ denotes the rank in the wage distribution of gender-j workers. $v_j(u)$ is computed as the result of a differential equation as explained in Section 2.3. The computation involves the use of the initial condition: $n_f = n_f(1) = .224$.

Figure 3: Difference between gender rank and job position rank: $v_j(u) - u$, $h(u) = .8 - .3u$



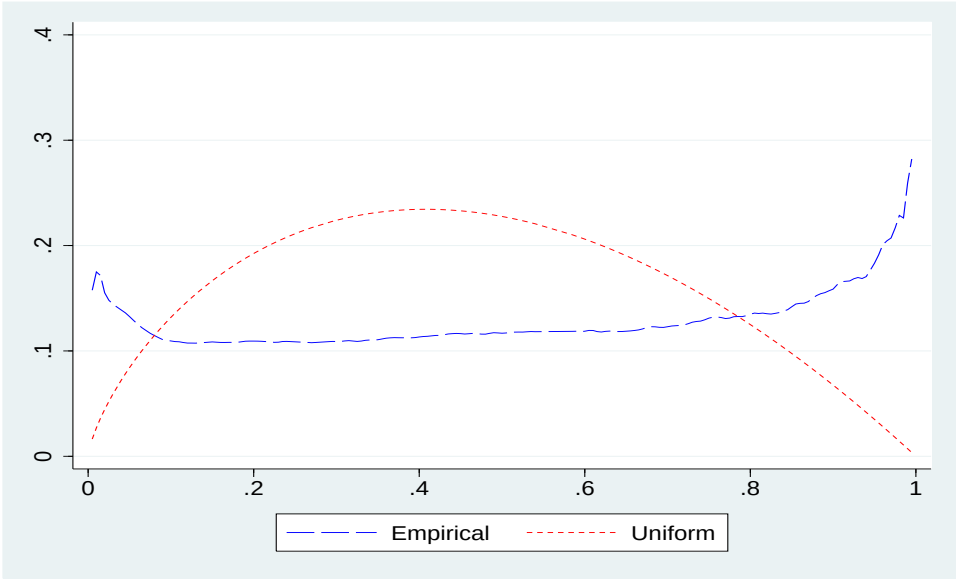
Note: for a job of rank u in the wage distribution of positions, $v_j(u) = n_j(u)/n_j$ denotes the rank in the wage distribution of gender-j workers. $v_j(u)$ is computed as the result of a differential equation as explained in Section 2.3. The computation involves the use of the initial condition: $n_f = n_f(1) = .224$.

Figure 4: Difference between gender rank and job position rank: $v_j(u) - u$, $h(u) = 1.2 - .4u$



Note: for a job of rank u in the wage distribution of positions, $v_j(u) = n_j(u)/n_j$ denotes the rank in the wage distribution of gender-j workers. $v_j(u)$ is computed as the result of a differential equation as explained in Section 2.3. The computation involves the use of the initial condition: $n_f = n_f(1) = .224$.

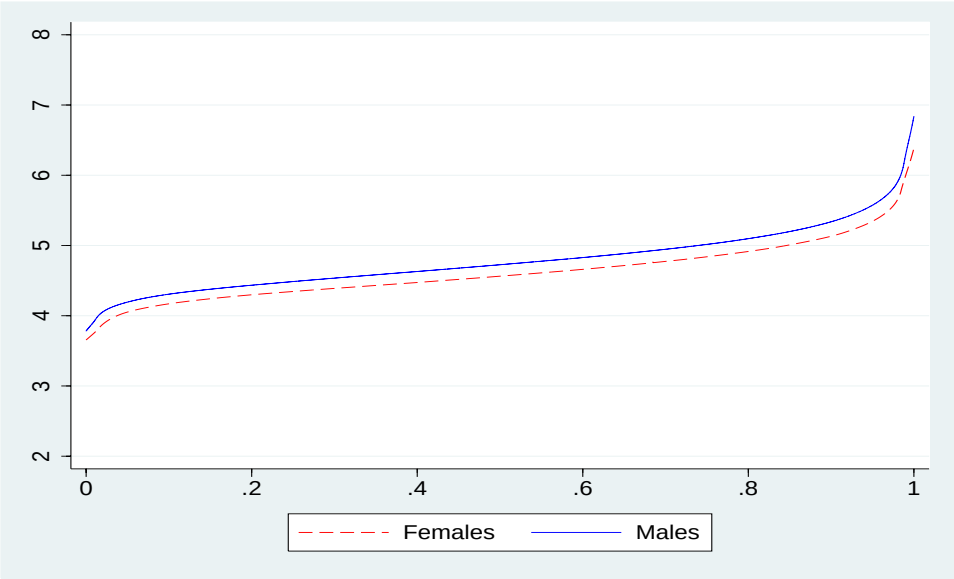
Figure 5: Gender quantile difference (M-F), $h(u) = .672$



Source: DADS, 2003, full-time executives of the banking industry aged 40-45.

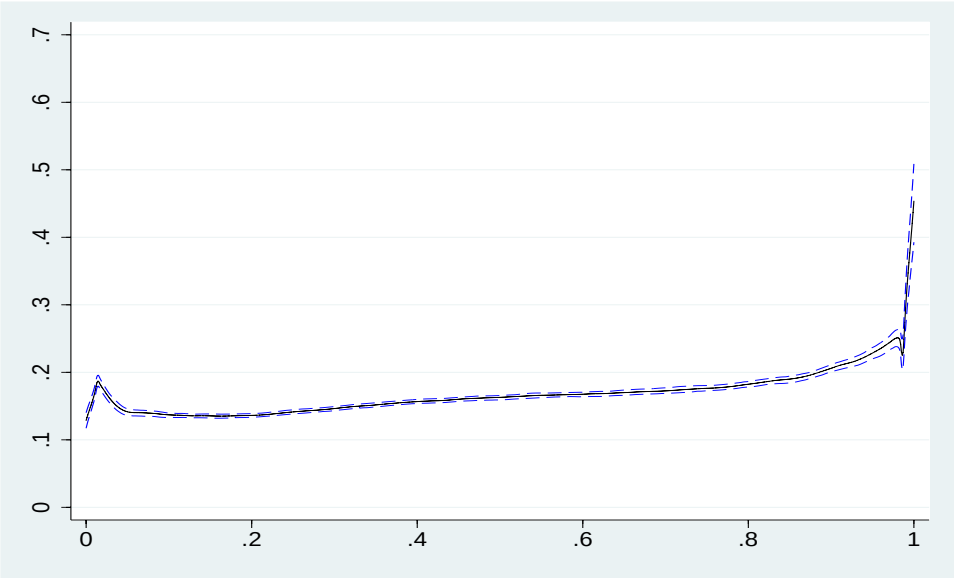
Note: The curve labelled *Empirical* represents the gender quantile difference when the wage distribution of job positions is supposed to be the empirical wage distribution in the banking sector (and the proportion of females is fixed to the one in that industry: 28.7%). The curve labelled *Uniform* represents the gender quantile difference when the wage distribution of job positions is supposed to be uniform over the interval [0,1.64] (the upper bound of the interval ensuring that the gender quantile difference is of the same magnitude as for the curve *Empirical*). The gender quantile difference is the difference between the quantile of males and the quantile of females.

Figure 6: Gender log-wage as a function of gender rank, pooled industries



Source: DADS, 2003, full-time executives of the Private Sector aged 40-45.

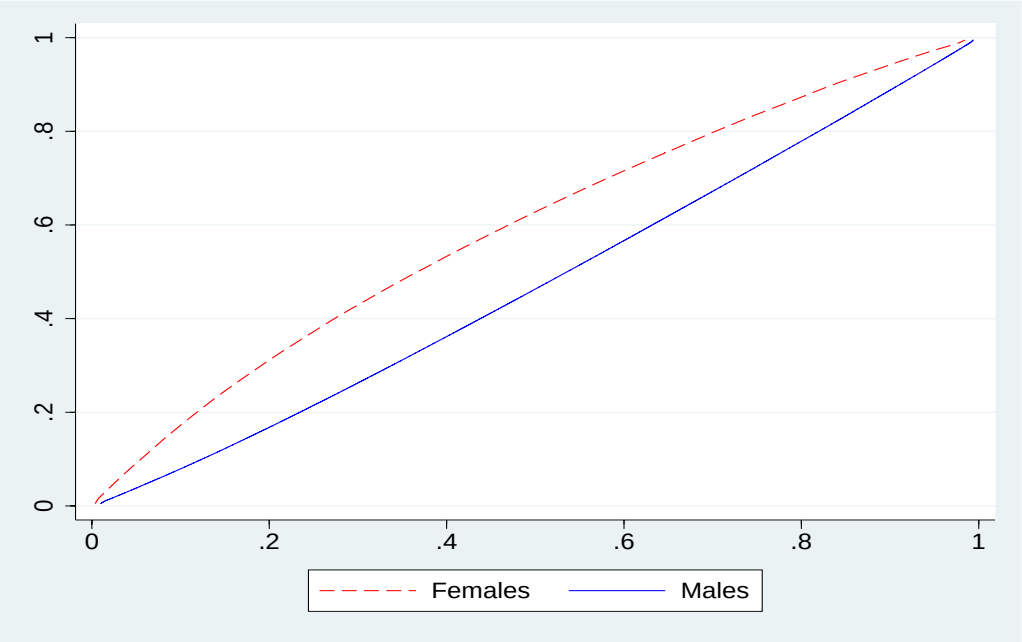
Figure 7: Difference in log-wage (M-F) as a function of gender rank, pooled industries



Source: DADS, 2003, full-time executives of the Private Sector aged 40-45.

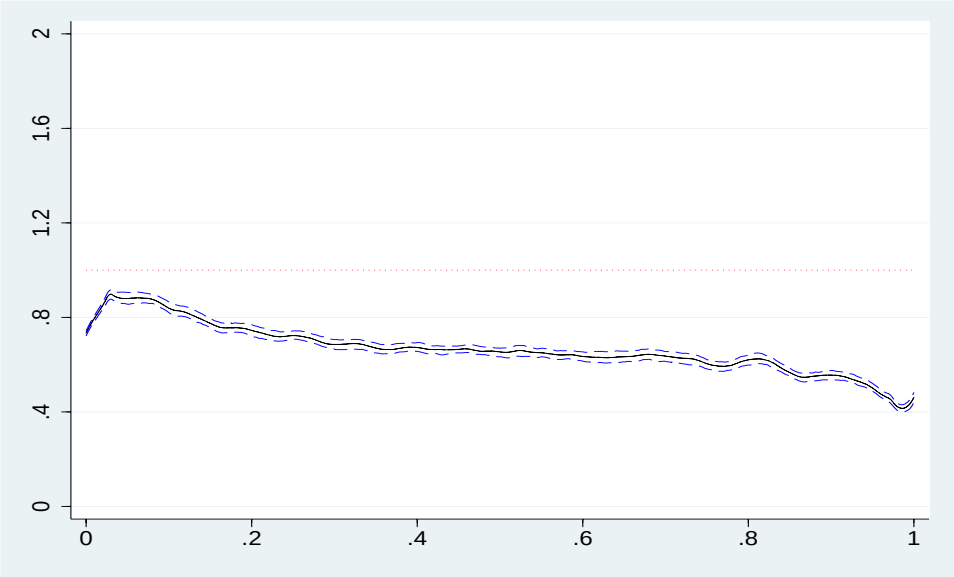
Note: Bounds of the confidence interval estimated by bootstrap (100 replications) are represented in dashed lines.

Figure 8: Gender rank as a function of job rank, pooled industries



Source: DADS, 2003, full-time executives of the Private Sector aged 40-45.

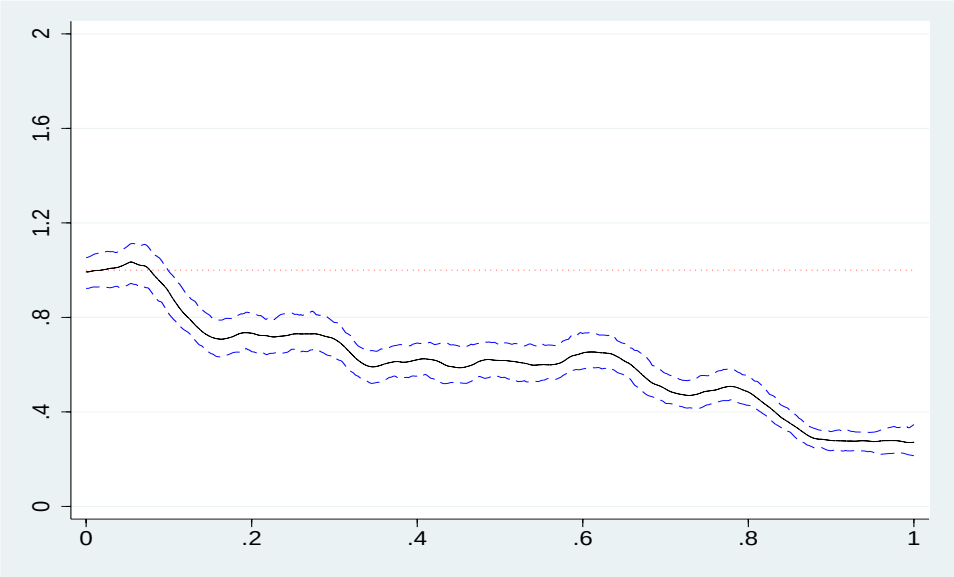
Figure 9: Access function (F/M) as a function of job rank, pooled industries



Source: DADS, 2003, full-time executives of the Private Sector aged 40-45.

Note: See Section 3 for details on the estimation method. Bounds of the confidence interval estimated by bootstrap (100 replications) are represented in dashed lines.

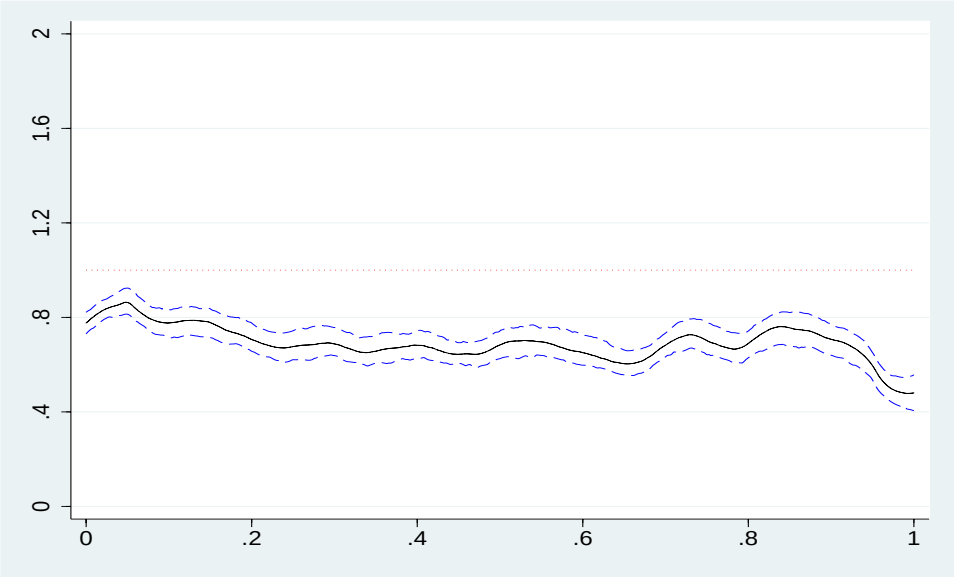
Figure 10: Access function (F/M) as a function of job rank, insurance industry



Source: DADS, 2003, full-time executives of the insurance industry aged 40-45.

Note: See Section 3 for details on the estimation method. Bounds of the confidence interval estimated by bootstrap (100 replications) are represented in dashed lines.

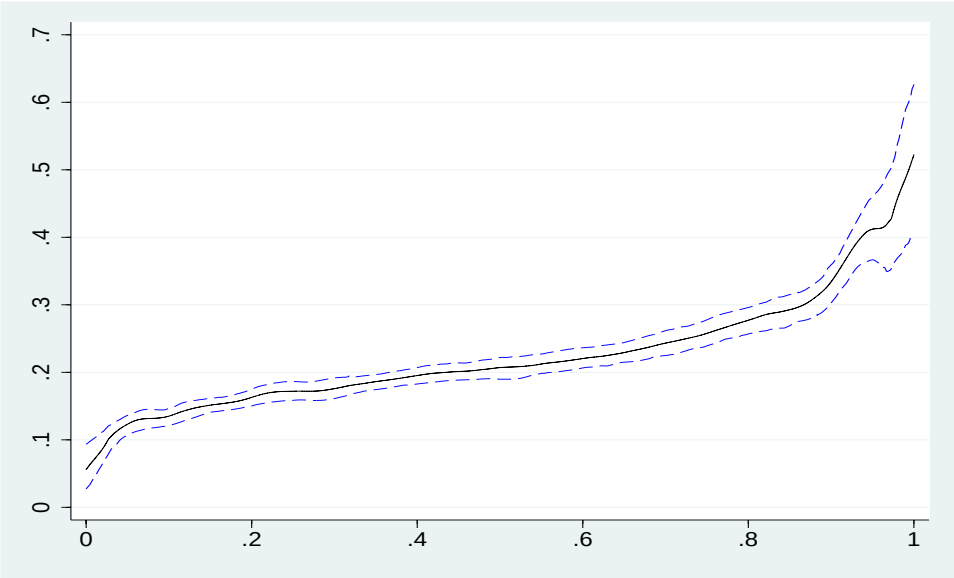
Figure 11: Access function (F/M) as a function of job rank, banking industry



Source: DADS, 2003, full-time executives of the banking industry aged 40-45.

Note: See Section 3 for details on the estimation method. Bounds of the confidence interval estimated by bootstrap (100 replications) are represented in dashed lines.

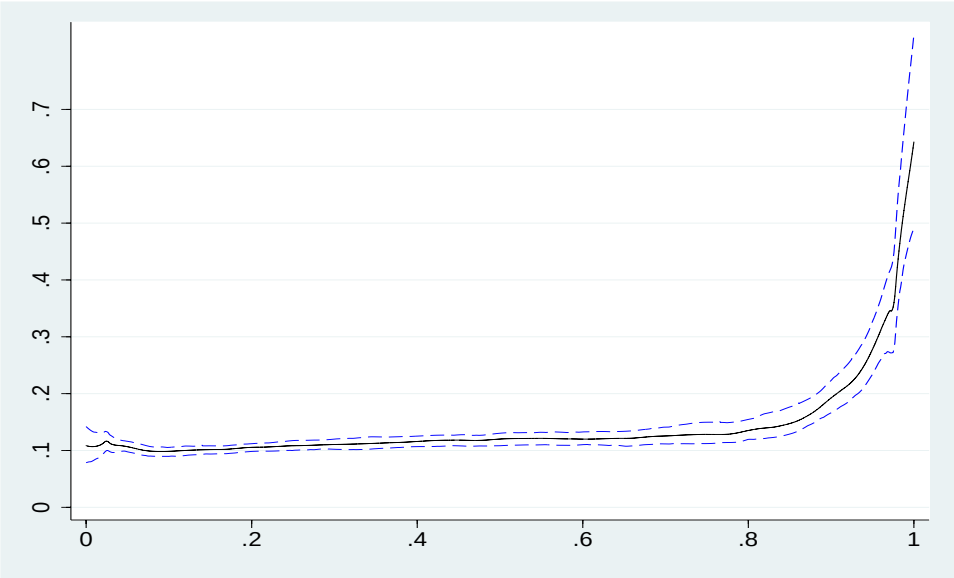
Figure 12: Difference in log-wage (M-F) as a function of gender rank, insurance industry



Source: DADS, 2003, full-time executives of the insurance industry aged 40-45.

Note: Bounds of the confidence interval estimated by bootstrap (100 replications) are represented in dashed lines.

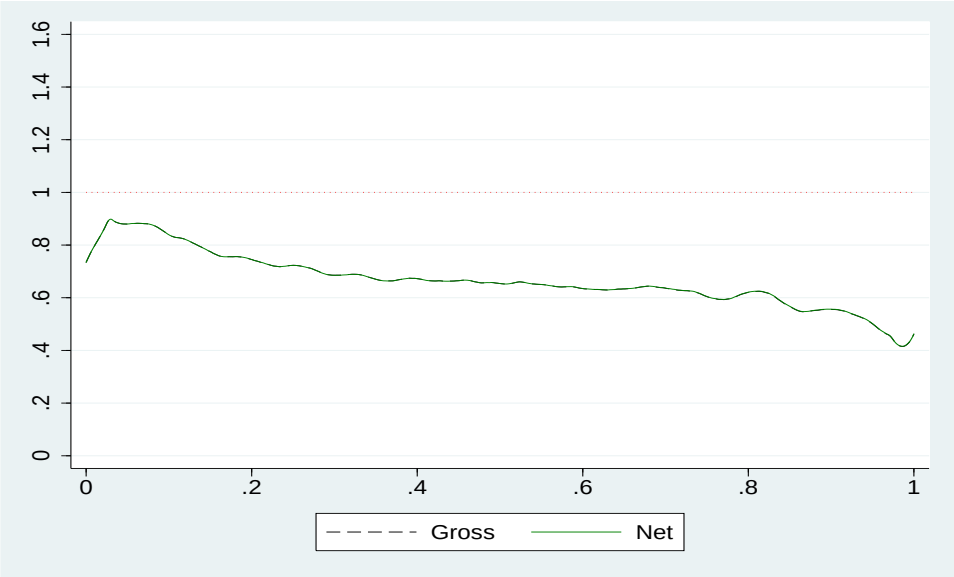
Figure 13: Difference in log-wage (M-F) as a function of gender rank, banking industry



Source: DADS, 2003, full-time executives of the banking industry aged 40-45.

Note: Bounds of the confidence interval estimated by bootstrap (100 replications) are represented in dashed lines.

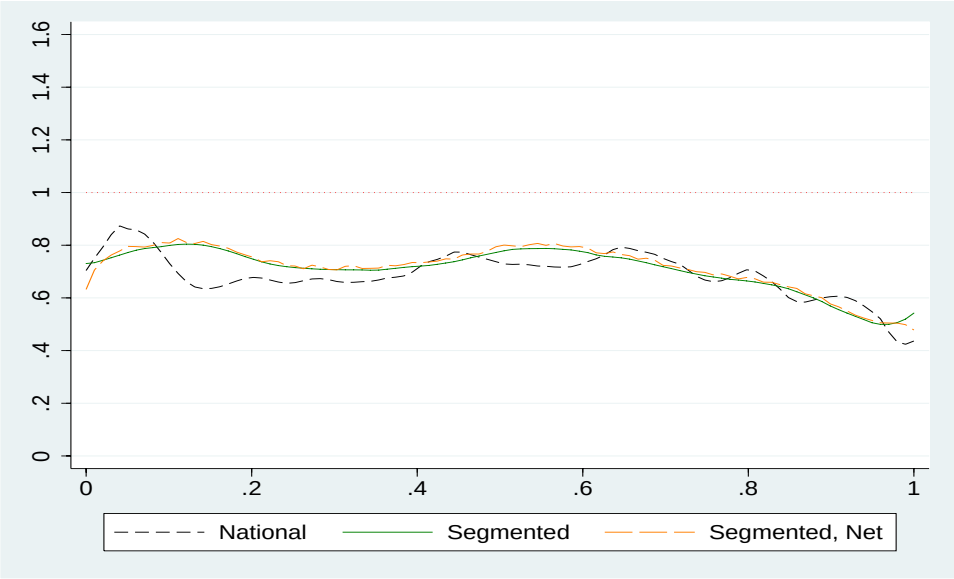
Figure 14: Access function (F/M) as a function of job rank, pooled industries, individual observed heterogeneity taken into account



Source: DADS, 2003, full-time executives of the Private Sector aged 40-45.

Note: The curve labelled *Gross* (in black dashed line) represents the access function computed without taking into account the individual observed heterogeneity. The curve labelled *Net* (in green solid line) represents the access function obtained when taking into account the individual observed heterogeneity (see Section 6.1 for details on the estimation method).

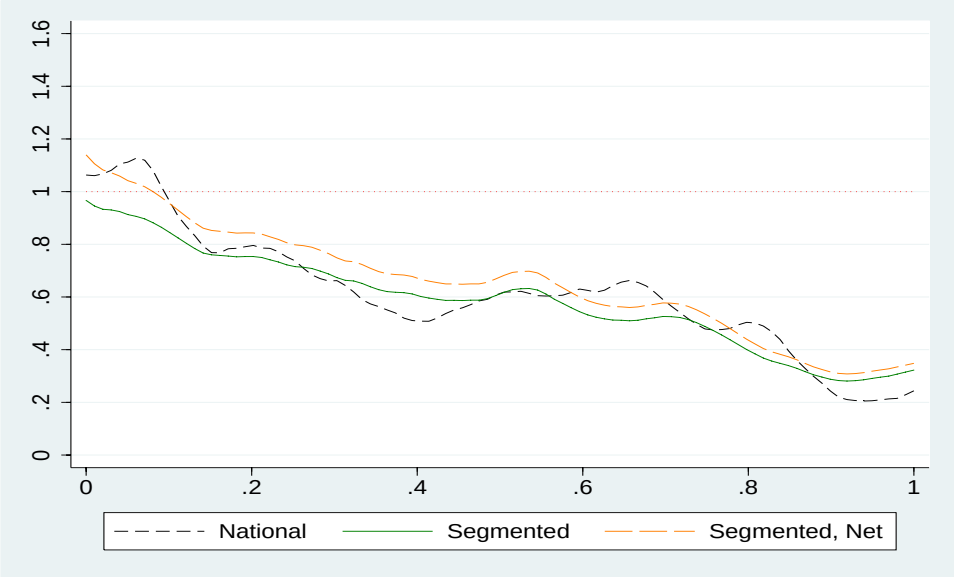
Figure 15: Average access function (F/M) as a function of job rank, large firms, segmented markets



Source: DADS, 2003, full-time executives aged 40-45 in firms employing more than 150 such executives.

Note: The curve labelled *National* (in black dashed line) represents the access function computed for the national market without taking into account the individual observed heterogeneity. The curve labelled *Segmented* (in green solid line) represents the average access function computed across segmented submarkets (each submarket being a large firm) without taking into account the individual observed heterogeneity. The curve labelled *Segmented, Net* (in orange long-dashed line) represents the average access function computed across segmented submarkets when taking into account individual observed heterogeneity (See Section 6.2 for details on the estimation method).

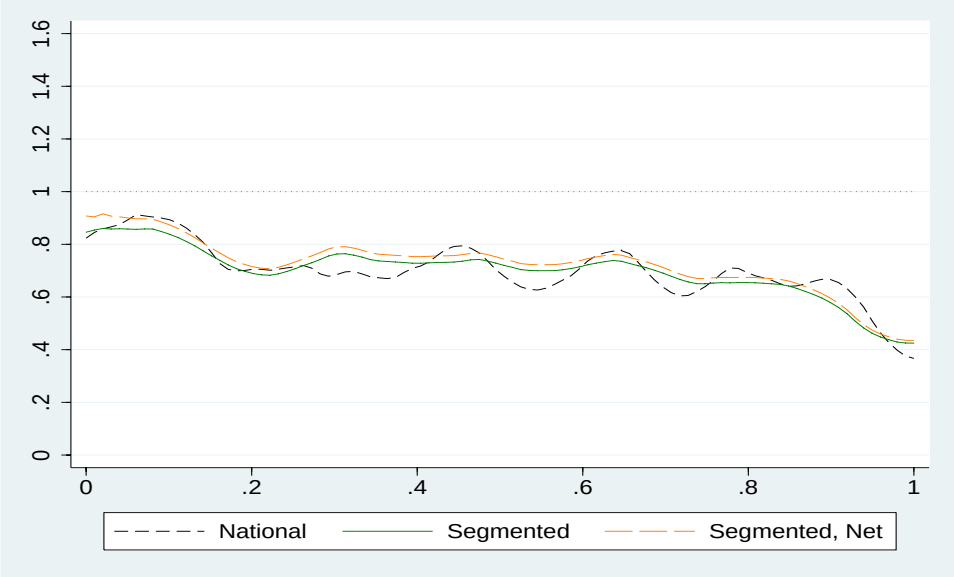
Figure 16: Average access function (F/M) as a function of job rank, large firms, insurance industry, segmented markets



Source: DADS, 2003, full-time executives aged 40-45 in firms of the insurance industry employing more than 150 such executives.

Note: The curve labelled *National* (in black dashed line) represents the access function computed for the national market without taking into account the individual observed heterogeneity. The curve labelled *Segmented* (in green solid line) represents the average access function computed across segmented submarkets (each submarket being a large firm) without taking into account the individual observed heterogeneity. The curve labelled *Segmented, Net* (in orange long-dashed line) represents the average access function computed across segmented submarkets when taking into account individual observed heterogeneity (See Section 6.2 for details on the estimation method).

Figure 17: Average access function (F/M) as a function of job rank, large firms, banking industry, segmented markets



Source: DADS, 2003, full-time executives aged 40-45 in firms of the banking industry employing more than 150 such executives.

Note: The curve labelled *National* (in black dashed line) represents the access function computed for the national market without taking into account the individual observed heterogeneity. The curve labelled *Segmented* (in green solid line) represents the average access function computed across segmented submarkets (each submarket being a large firm) without taking into account the individual observed heterogeneity. The curve labelled *Segmented, Net* (in orange long-dashed line) represents the average access function computed across segmented submarkets when taking into account individual observed heterogeneity (See Section 6.2 for details on the estimation method).

Table 1: Descriptive statistics by subgroup of firms

Sector	Nb. firms	Nb. jobs	% females	Wages, all		
				Median	Mean	Std
All firms	86,989	354,968	22.4	109	139	602
Large firms	429	115,531	22.3	114	134	132
Banking	545	18,628	28.7	104	142	449
Banking, large firms	38	11,197	30.7	110	149	273
Insurance	507	9,360	36.9	107	125	74
Insurance, large firms	20	5,491	37.2	107	120	62

Source: DADS, 2003, full-time executives of the Private Sector aged 40-45.

Table 1: Descriptive statistics by subgroup of firms (cont.)

Sector	Wages, Females			Wages, Males		
	Median	Mean	Std	Median	Mean	Std
All firms	96	119	434	113	145	642
Large firms	103	118	101	118	139	139
Banking	95	120	211	108	150	514
Banking, large firms	110	149	273	114	160	317
Insurance	94	105	49	115	136	84
Insurance, large firms	106	102	40	115	131	70

Source: DADS, 2003, full-time executives of the Private Sector aged 40-45.

Table 2: Values of the access function at different ranks

	p1	p5	p10	p25	p50	p75	p90	p95	p99
All firms	.79	.88	.84	.72	.65	.60	.56	.50	.42
	[.78,.80]	[.86,.91]	[.82,.87]	[.71,.74]	[.63,.67]	[.58,.62]	[.54,.58]	[.49,.52]	[.40,.44]
Large firms	.75	.86	.74	.66	.73	.67	.61	.55	.42
	[.73,.77]	[.83,.89]	[.71,.78]	[.63,.68]	[.70,.77]	[.63,.70]	[.57,.64]	[.52,.57]	[.39,.46]
Banking	.80	.86	.78	.68	.68	.70	.71	.60	.48
	[.76,.85]	[.81,.92]	[.72,.83]	[.62,.74]	[.63,.74]	[.64,.78]	[.64,.77]	[.54,.65]	[.41,.55]
Banking, large firms	.84	.89	.93	.69	.71	.67	.64	.47	.35
	[.79,.90]	[.83,.97]	[.84,1.04]	[.61,.79]	[.62,.80]	[.59,.76]	[.56,.71]	[.41,.54]	[.28,.44]
Insurance	1.00	1.03	.91	.73	.62	.49	.28	.27	.27
	[.93,1.07]	[.93,1.10]	[.82,1.00]	[.67,.82]	[.55,.68]	[.43,.55]	[.24,.32]	[.23,.31]	[.22,.34]
Insurance, large firms	1.06	1.12	.99	.72	.61	.47	.22	.21	.24
	[.96,1.17]	[.99,1.24]	[.85,1.16]	[.62,.88]	[.51,.71]	[.39,.53]	[.17,.28]	[.16,.26]	[.18,.32]

Source: DADS, 2003, full-time executives of the Private Sector aged 40-45.

Note: See Section 3 for the estimation method. Bounds of the confidence intervals reported in brackets are estimated by bootstrap (100 replications).

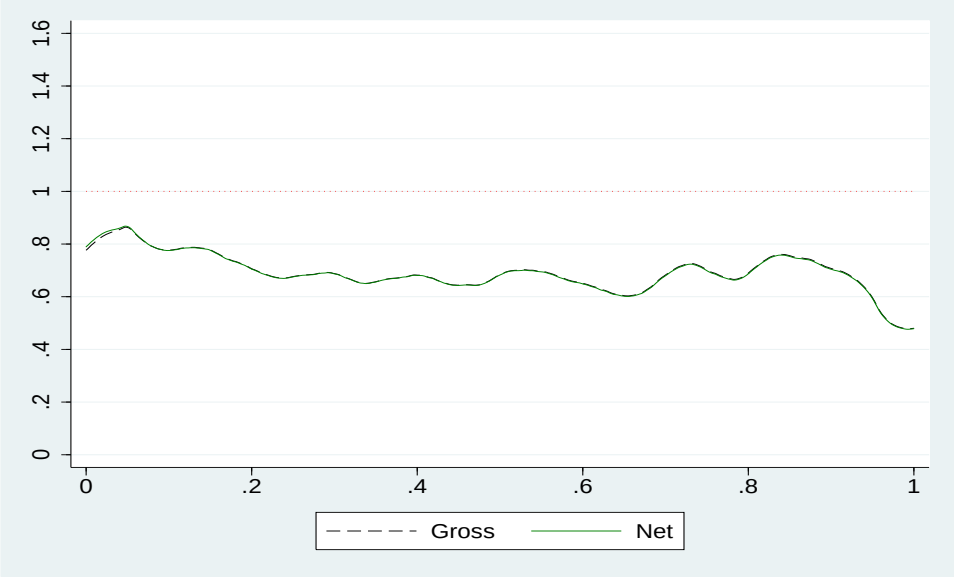
Table 3: Linear specification of the access function

Sector	National market			Segmented markets		
	Const	Slope	Stat	Const	Slope	Stat
All firms	.80	.28	.453			
	[.79,.81]	[.26,.29]				
Large firms	.74	.09	.828	.69	.05	.434
	[.72,.76]	[.07,.14]		[.64,.70]	[-.02,.07]	
Banking	.71	.07	.066			
	[.67,.75]	[-.01,.15]				
Banking, large firms	.83	.26	.077	.77	.25	.300
	[.77,.89]	[.16,.39]		[.67,.82]	[.11,.33]	
Insurance	.90	.60	.078			
	[.85,.95]	[.54,.69]				
Insurance, large firms	.93	.66	.056	.74	.41	.339
	[.82,1.01]	[.53,.79]		[.60,.79]	[.23,.51]	

Source: DADS, 2003, full-time executives of the Private Sector aged 40-45.

Note: We report the estimated coefficients of a linear specification of the access function, $h(u) = a - b.u$. Bounds of the confidence intervals are estimated by bootstrap (100 replications) and are given in brackets. We also report the statistic of a specification test for which the threshold at the 5% level is .461. The method used to estimate the coefficients and to compute the test statistic is detailed in Appendix B.

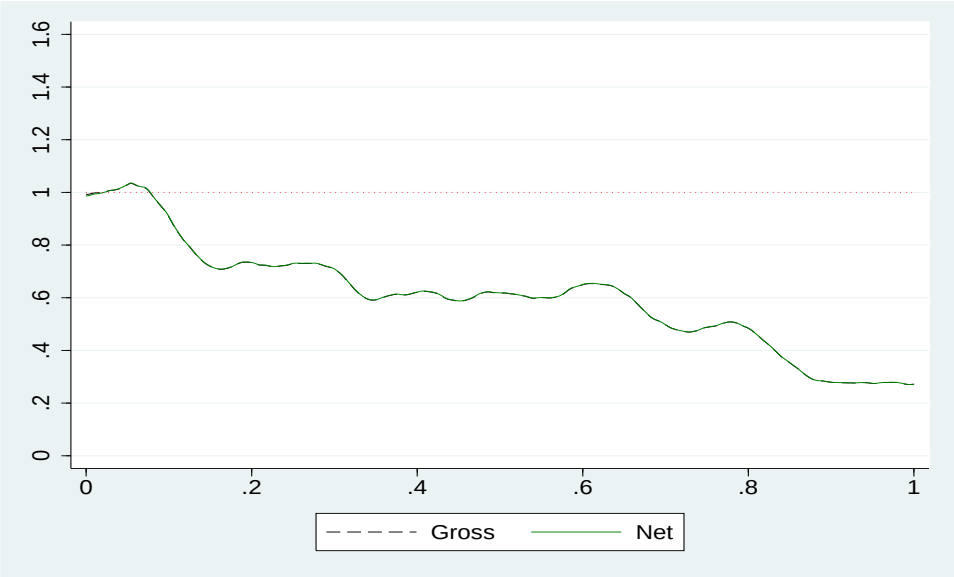
Figure A.1: Average access function (F/M) as a function of job rank, insurance industry, individual observed heterogeneity taken into account



Source: DADS, 2003, full-time executives of the insurance industry aged 40-45.

Note: The curve labelled *Gross* (in black dashed line) represents the access function computed without taking into account the individual observed heterogeneity. The curve labelled *Net* (in green solid line) represents the access function obtained when taking into account the individual observed heterogeneity (see Section 6.1 for details on the estimation method).

Figure A.2: Average access function (F/M) as a function of job rank, banking industry, individual observed heterogeneity taken into account



Source: DADS, 2003, full-time executives of the banking industry aged 40-45.

Note: The curve labelled *Gross* (in black dashed line) represents the access function computed without taking into account the individual observed heterogeneity. The curve labelled *Net* (in green solid line) represents the access function obtained when taking into account the individual observed heterogeneity (see Section 6.1 for details on the estimation method).

Table A.1: Coefficients of individual variables for males

Sector	age41	age42	age43	age44	age45	foreigner
All firms	.033***	.057***	.071***	.072***	.088***	.101***
	(.006)	(.006)	(.006)	(.007)	(.007)	(.006)
Large firms	.021	.031**	.028**	.047***	.056***	.148***
	(.013)	(.013)	(.013)	(.013)	(.013)	(.012)
Banking	-.021	-.015	.038	-.014	-.016	.317***
	(.030)	(.030)	(.030)	(.030)	(.030)	(.031)
Banking, large firms	.005	.012	-.020	-.032	-.019	.281***
	(.044)	(.045)	(.045)	(.045)	(.046)	(.045)
Insurance	.032	.064	.020	-.001	.054	.126***
	(.045)	(.044)	(.044)	(.045)	(.045)	(.048)
Insurance, large firms	-.020	-.010	-.089	-.117	-.059	.227***
	(.072)	(.072)	(.069)	(.072)	(.072)	(.079)

Source: DADS, 2003, full-time male executives of the Private Sector aged 40-45.

Note: The coefficients are estimated by maximizing the partial likelihood on the subsample of males (cf. Section 7.1). Standard errors are given in parentheses. Level of significance: ***: 1%, **: 5%, *: 10%.

Table A.2: Coefficients of individual variables for females

Sector	age41	age42	age43	age44	age45	foreigner
All firms	-.023***	-.008	-.000	-.005	.015	.145***
	(.012)	(.012)	(.012)	(.012)	(.012)	(.011)
Large firms	-.020	-.031	.014	.030	.059**	.190***
	(.024)	(.024)	(.024)	(.024)	(.025)	(.023)
Banking	-.061	-.023	.050	.010	.060	.359***
	(.045)	(.045)	(.046)	(.047)	(.048)	(.045)
Banking, large firms	-.127**	-.107*	-.021	-.039	.005	.297***
	(.064)	(.063)	(.064)	(.066)	(.069)	(.064)
Insurance	.015	-.070	.008	.049	.059	.148**
	(.059)	(.057)	(.058)	(.057)	(.058)	(.063)
Insurance, large firms	-.182**	-.317***	.028	.023	-.006	.160
	(.088)	(.089)	(.091)	(.093)	(.092)	(.107)

Source: DADS, 2003, full-time female executives of the Private Sector aged 40-45.

Note: The coefficients are estimated by maximizing the partial likelihood on the subsample of females (cf. Section 7.1). Standard errors are given in parentheses. Level of significance: ***: 1%, **: 5%, *: 10%.

Table A.3: Coefficients of individual variables for males, segmented markets

Sector	age41	age42	age43	age44	age45	foreigner
Large firms	.031**	.059***	.082***	.126***	.155***	.074***
	(.013)	(.013)	(.013)	(.013)	(.013)	(.013)
Banking, large firms	.026	.034	.006	.006	.014	.195***
	(.045)	(.045)	(.046)	(.046)	(.047)	(.046)
Insurance, large firms	.017	.052	-.034	-.022	.030	.152*
	(.073)	(.073)	(.070)	(.073)	(.073)	(.081)

Source: DADS, 2003, full-time executives aged 40-45 in firms employing more than 150 such executives.

Note: The coefficients are estimated by maximizing the partial likelihood stratified by firm on the subsample of males (cf. Section 7.2). Standard errors are given in parentheses. Level of significance: ***: 1%, **: 5%, *: 10%.

Table A.4: Coefficients of individual variables for females, segmented markets

Sector	age41	age42	age43	age44	age45	foreigner
Large firms	-.013	-.025	.042*	.066***	.104***	.124***
	(.024)	(.024)	(.025)	(.025)	(.026)	(.024)
Banking, large firms	-.106	-.075	.302	-.055	.065	.257***
	(.066)	(.064)	(.065)	(.067)	(.070)	(.065)
Insurance, large firms	-.207**	-.309***	.021	.052	-.008	.172
	(.090)	(.091)	(.092)	(.094)	(.093)	(.108)

Source: DADS, 2003, full-time executives aged 40-45 in firms employing more than 150 such executives.

Note: The coefficients are estimated by maximizing the partial likelihood stratified by firm on the subsample of females (cf. Section 7.2). Standard errors are given in parentheses. Level of significance: ***: 1%, **: 5%, *: 10%.

B.7. Housing and location choices of retiring households: Theory and evidence

Gobillon L. et F.C. Wolff (2009), “Housing and location choices of retiring households: Theory and evidence”, Document de Travail INED, 162, revise-and-resubmit à *Urban Studies*, version révisée.

29 pages

Housing and location choices of retiring households: Evidence from France [#]

Laurent Gobillon [*] François-Charles Wolff [**]

November 2009

Abstract: In this paper, we study the mobility and housing choices of the elderly when retiring using household data collected in France. From a theoretical viewpoint, individuals are likely to decrease their housing quantity because of an income loss when retiring, but they may also increase it to benefit from more housing comfort for leisure. Using the 1992 Trois Générations survey, we first show that housing mobility at retirement is substantial in France, with a variety of self-reported motives. Then, using the 1994-2001 French Europanel survey, we find evidence of both upsizing and downsizing for mobile recent retirees. In many cases, housing adjustments lead to a correction of the initial disequilibrium between the number of rooms and the number of occupants. However, a significant proportion of mobile recent retirees improve the quality of their dwelling.

Keywords: retirement, housing, mobility, panel data, France

JEL Classification: J26, R21

[#] We are indebted to Harris Selod and to three anonymous referees for their insightful comments on a previous draft. We also would like to thank Tim Halliday, Andy Mason, Sang Hyop Lee, and seminar participants at Ined (Paris), University of Hawaï (Manoa), the 23èmes Journées de Microéconomie Appliquée (Nantes), Cergy, Nancy, Marseille (GREQAM), Strasbourg (BETA) and CEE (Paris). Any remaining errors are ours.

[*] Corresponding author. INED, PSE, CREST and CEPR. INED, 133 Boulevard Davout, 75980 Paris Cédex 20, France. Tel: 00 33 (0) 1 56 06 20 16. Email: laurent.gobillon@ined.fr http://laurent.gobillon.free.fr/

[**] LEMNA, Université de Nantes, France. CNAV and INED, Paris, France.
Email: francois.wolff@univ-nantes.fr http://www.sc-eco.univ-nantes.fr/~fcwolff

1. Introduction

In many European countries, housing markets are getting tighter over the years, the main market indicators like housing prices, transactions or mortgages showing strong growth except over the very recent period. It is well acknowledged that both macroeconomic and demographic factors have a deep influence on the housing markets (Hardman and Ioannides, 1995). In particular, many countries are beginning to experience a great demographic transition because of the baby-boom. The high fertility rates after the Second World War have resulted in a high concentration of people born between 1946 and 1964. This is currently leading to an increase in the number of retiring individuals, a phenomenon whose impact has so far not been fully appreciated. This surge should have both economic and social implications on housing markets. It therefore matters to understand the housing decisions of people who retired.

Although the housing choices of the population as a whole have been widely analyzed, the literature on housing in later life is less developed. Some papers have studied the living arrangements and the informal care from children to parents when sharing a home (Börsch-Supan et alii, 1992; Hoerger et alii, 1996; Pezzin and Schone, 1999). The intergenerational co-residence was regarded as a mode of assistance to the elderly. A few papers have rather looked at the dynamics of housing adjustments in later life, especially at retirement.

As retirement goes along with a loss of income and a rise in leisure time, significant housing and location changes are expected. Feinstein and McFadden (1989) and Venti and Wise (1990) find that mobility is high at retirement in the United States. Ermisch and Jenkins (1999) investigate the determinants of residential mobility for the elderly in Britain as well as the subsequent housing adjustments made by movers using the first five waves of the British Household Panel Survey. They consider the whole population aged 55 and over as moves are scarce in later life. Subsequent more recent studies on European countries have used the same kind of approach. Laferrère (2005) analyzes the mobility and housing adjustments of the French elderly using some cross-sectional data with retrospective questions, while Tatsiramos (2006) estimates competing risk hazard models to study the determinants of residential mobility of older households using the European Community Household Panel data.

The purpose of our contribution is to investigate in detail the housing and location choices of retiring people using longitudinal household data collected in France. Our analysis departs from most existing studies which have proposed some general analyses of housing adjustments in

later life and have not really focused on retirement issues only. To motivate our empirical analysis, we begin with a theoretical discussion and show that individuals may want to either increase or decrease both their dwelling size and housing quality when retiring, if they decide to move. These inconclusive predictions occur because of the trade-off between the loss of income and additional leisure time which follow retirement.

To shed light on housing adjustments, we rely on the following two French datasets. First, we use the 1992 *Trois Générations* cross-section survey to compute some descriptive statistics on residential mobility when retiring and on self-reported motives for moving. Various reasons are given by retiring individuals to motivate their move, including living closer to the family, finding a place with a better climate, and changing the size of the dwelling. We then study more closely the mobility process and housing adjustments using the 1994-2001 French Europanel survey.[1] We show that about 60% of movers decide to move at the exact retirement date or one year later. There is evidence of both upsizing and downsizing for mobile recent retirees. Housing adjustments mostly lead to a correction of the initial disequilibrium between the number of rooms and the number of occupants. However, a significant proportion of mobile recent retirees improve the quality of their dwelling.

The rest of this paper is organized as follows. In Section 2, we discuss the housing choices when retiring and the related literature. In Section 3, we present our results on the self-reported motives of respondents for moving at retirement. In Section 4, we investigate more closely the timing of residential mobility and consider the various housing adjustments made by movers using panel data. Finally, Section 5 concludes.

2. Theoretical background

The purpose of this section is to present the mechanisms explaining the mobility and housing choices when retiring. Let us first consider the schematic case of an individual holding a job whose earnings are used for the consumption of some goods and the payment of housing costs. Retirement has two main consequences on the different constraints faced by the agent. On the one hand, the individual income switches from his wages to a lower retirement pension. This negative income effect makes him likely to decrease his consumption of goods and creates some pressure to reduce the housing cost by downsizing. On the other hand, the new retiree benefits

[1] The French Europanel survey is the French version of the European Community Household Panel survey.

from more leisure time. Hence, he values more his dwelling as he can spend more time at home.

Given these changes in financial and time constraints, the retiree can choose between staying in his dwelling and moving. If he stays, he cannot adjust his dwelling to his new retirement needs. If he moves, he can do such adjustments, but mobility involves a moving cost. There are some monetary costs like transportation fees and transaction costs when selling a house. There are also some utility losses as the social capital of movers is likely to be lost (DiPasquale and Glaeser, 1999). Private benefits stemming from a localized network of friends clearly decrease when moving to another location.

In this setting, the individual moves only if the benefits derived from a housing adjustment are worth incurring the moving costs. This implies that the optimal dwelling size when retiring has to be sufficiently below or above the optimal dwelling size when working to motivate the individual to move (Gobillon and Le Blanc, 2004; Flavin and Nakagawa, 2008).[2] If the two dwelling sizes are close, the individual prefers to stay in his dwelling. Note that size is not the only dimension that the individual can adjust when moving. There are also many qualitative aspects to be taken into account, like bath, toilets, central heating, garden, etc. These extra-dimensions will be considered in our empirical application.

Of course, there are many variations to this schematic example. For instance, the mobility of a couple is typically a joint decision which depends on the labour status of the two partners. When one of them retires and the other one keeps working (in particular because it is mandatory in France to work until a given age which is often 60 to get a full pension), the two spouses may decide to stay in their dwelling and wait for the two partners to be retired before moving. The location of the job position then plays as an anchor.

Also, a household can choose between renting and owning when moving. The literature usually considers that this choice depends on the relative user cost of the two alternatives.[3] For a renter, the user cost is the rent. For an owner, it includes the maintenance costs and a possible opportunity cost if the return of financial assets (in which the household could have invested the value of the purchased property) is larger than the return of the housing property (that depends on the increase in housing prices). Households often need to borrow on the credit market to purchase

[2] It is important to stress that in these studies, the dwelling size captures a wide definition of housing consumption, which includes the multiple dimensions of quality as highlighted both in our theoretical and empirical analysis.

[3] The tenure choice has been widely studied in the literature (see Henderson and Ioannides, 1983; Linneman and Wachter, 1989; Zorn, 1989; Gobillon and Le Blanc, 2008).

a dwelling. However, those reaching the age of retirement have had time to accumulate wealth. Many are already owners and can sell their house and use the money as down payment when purchasing a new dwelling. Note however that there are some non-negligible notary expenses when buying a house. As the life expectancy is shorter for retirees than for instance a young couple who tries to access ownership, these expenses may deter some retirees from buying a new house.

Housing choices also depend on the preferences of households for consumption and leisure at the present, which may have influenced their saving behaviour when they were on the labour market. We can illustrate the effect of these preferences on the mobility behaviour when retiring with two polar cases. First consider an individual who wants to consume and spend a significant amount of time in leisure. He tends to consume a lot and may rent a large house when working. He does not save much and may be short of money when retiring. He is then expected to downsize. Now consider an individual who wants to accumulate some wealth that he may enjoy later and chooses to occupy a time-consuming job. He does not consume much and lives in a small dwelling for which he pays a low rent while working. When retiring, he has some wealth which can be used for his leisure time and is then expected to upsize his dwelling.

Also, some households may prefer to spend time outside their dwelling rather than inside. In that case, they are less likely to upsize their dwelling, but may invest more in a garden. Location matters and movers may choose places that propose consumption amenities such as nice weather, an access to the coast or an enjoyable countryside. It has been considered since Roback (1982) that the levels of rents and wages compensate for the presence of consumption amenities. Indeed, places where there are some amenities should be characterized at the spatial equilibrium by some lower wages or some higher rents. As retirees are not on the labour market anymore, they are more attracted by places where consumption amenities are compensated by low wages (Knapp and Graves, 1988). Such places include large spaces where the weather is nice and the occupation of land is not much constrained, as for instance locations in more rural areas in the South or South West (in *départements* like Ardèche, Lot, etc). By contrast, nice spots on the coast are likely to exhibit high rents as the occupation of land is constrained and they may thus attract only wealthy retirees. Some papers have shown that retired people have a strong preference for living in sunny areas (Hogan and Steinnes, 1998; King et alii, 1998; Chen and Rosenthal, 2008). Retirees also have more time to spend with their children and grandchildren, and they may

choose to live closer to their families (De Coulon and Wolff, 2006).[4]

The location choice is expected to affect the housing adjustments of households. On the one hand, downsizing is likely if the housing costs are very high on the site of destination or if the climate is very nice there and the individuals intend to spend a lot of time outside. On the other hand, upsizing is likely if housing costs are very low on the site of destination or if the climate is bad.

To summarize, our discussion shows that the theoretical predictions on housing adjustments when retiring are not clear-cut and the housing choices of households thus clearly remain an empirical question. Using two French datasets, we investigate in the next two sections the pattern of self-reported motives for the mobility decision as well as the adjustments to housing size and quality of mobile households.

3. The pattern of self-reported motives for mobility

As highlighted in the previous section, several reasons may explain mobility when retiring. We shed some light on these reasons using self-reported motives given by respondents in the survey *Trois Générations* (3G hereafter). This survey was conducted in 1992 by the French old-age pension fund and its aim was to provide a better understanding of the dynamics of family relationships.

The data set is based on a sample of randomly selected families including at least three adult members of consecutive generations living in France (see Jellal and Wolff, 2002). The sample was constructed first by selecting some adults born between 1939 and 1943 who have at least one parent and one adult child alive. From the information on location that those adults provided, one of the parents and one of the adult children were then interviewed, using the same questionnaire. The sample comprises 1,958 middle-aged adults, 1,217 grandparents and 1,493 young adults.

We restrict our attention to the grandparents aged between 68 and 92 in 1992 who have thus retired.[5] The respondents provide detailed information on their characteristics (biography, income, health, among others), their current dwelling and their relationships with other

[4] While retirees certainly enjoy spending time with their children and caring for their grandchildren, providing informal care to disabled elderly parents may be a very cumbersome and unpleasant task. There may be then some incentives for young retirees to move far away from their parents (Konrad et alii, 2002).

generations. The survey also includes a retrospective question about residential mobility at the date of retirement: "*Did you move when you retired (or when your spouse retired) or during the two or three years after retirement?*". When a household moves, we have some information on the new location, i.e. whether the respondent lives in the same region or in a different region.[6]

A first finding is that mobility at retirement is a significant phenomenon in France: 31.5% of grandparents claim that they moved when retiring. Among them, 55.9% stayed in the same region, while 44.1% relocated in another region. The survey also includes a retrospective question on the motive for moving. Respondents can choose one motive only in a list of several reasons that can be divided into five broad categories: family, geography, work, housing and other motives. A drawback of this self-reported information is that moving may have multiple causes. Also, more emotional reasons like going back to the place of birth or family motives may be selected more often than economic factors.

In Table 1, we report descriptive statistics on the main reason given by retired respondents for their move. First, 16.5% of them claim that they wanted to live closer to other family members. Most often, retirees intended to find a dwelling near their children. Two reasons can explain this choice. Parents may enjoy visiting their children and providing them with grandchild care (Cardia and Ng, 2003, Dimova and Wolff, 2009). Parents may also want to live near their children to benefit from upstream resources at old age.[7] A change in the housing size is quoted by 15.5% of movers. This item can be related to the sub-optimality of housing size or quality when retiring and a need for housing adjustments to reach the new optimum. Interestingly, the proportion of respondents who moved to a larger dwelling is nearly the same as the proportion of respondents who moved to a smaller dwelling.

Insert Table 1 here

Reasons related to geography are reported by 9.2% of respondents. Among them, some mobile retirees are mainly motivated by going back to their place of birth, while others are more interested in migrating to a sunny place like the South or the South-West of France. A substantial

---

[5] As most of them were born between 1910 and 1920, this generation has experienced little unemployment, but did not benefit from the spread of education observed in France after World War II (Attias-Donfut and Wolff, 2005).

[6] A shortcoming here is that we do not have any detailed information on the previous and current location. The reference to the region may have different interpretations depending on the respondents' feelings on their place of residence before and after the move.

[7] It should be kept in mind that the focus of the survey is on tri-generational families, for which there are private transfers both in the upward and downward directions. The family motive is thus likely to be overrepresented in our sample.

fraction of mobile retirees (29.1%) claims that they have moved for a professional reason. This can be explained in several ways. First, people suffer from a loss of income when retiring. The most financially constrained households may not have enough money to stay in their current dwelling, especially if they are renters. Also, some respondents may have access to a free furnished apartment while holding a job, or they may benefit from partial funding for their housing from their employer. All these benefits are lost when retiring.[8] Finally, the residual "other motives" category (for which we do not have any additional information) is quoted by many respondents (29.8%).

In Table 2, we investigate whether individual characteristics are correlated with the self-reported motives for moving. As retirees are interviewed several years after experiencing mobility, we focus on more permanent characteristics (gender, educational attainment) as well as on retrospective questions (occupational status when working, living as a couple when retiring). We also use the current level of household income as a proxy for resources after retirement.

Insert Table 2 here

According to the 3G survey, women and people living alone when retiring put forward some family motives much more frequently. This is not surprising as respondents can visit their children more easily when living closer to their place of residence. Motives related to geography are mainly reported by mobile retirees who are highly educated, executives, or with a high level of income. Living in a sunny place on the coast (in cities like Biarritz, Cannes or Nice) is usually costly and could deter non-wealthy households from moving there.

When moving at retirement, farmers and self-employed people (who account for respectively 17.2% and 13.8% of the sample) have to give up their previous dwelling more often than workers with another occupation. This should be due to the sale of the workplace that often includes the living place. Employees and blue collars report a change in the size of the dwelling more often than self-employed and highly-educated respondents. Housing adjustments exhibit different patterns depending on income. A downsizing motive is reported by 14.6% of respondents in the first quartile of income, while the proportion is only 2.9% in the fourth quartile. Conversely, an upsizing motive is quoted by only 4.2% of respondents in the first quartile compared to 12.5% in the fourth quartile.

[8] Farmers and self-employed for instance are expected to sell their workplace when retiring, and their dwelling is often part of the workplace.

4. Panel evidence on mobility and housing adjustments

4.1. The French Europanel survey

We study housing adjustments using the Europanel survey conducted in France over the 1994-2001 period. It is a longitudinal survey on households carried out by the French Institute of Statistics, whose aim is to provide some comparable information for several European countries. It starts with the interview of 7,344 households in 1994. These households are then tracked across the different waves even if they move. The sample size significantly decreases over time due to attrition, but the decrease has a magnitude similar to the one observed for other national longitudinal surveys (Perracchi, 2002).[9]

At each wave, the survey provides for every household the date of arrival in the current dwelling. It also includes the location of that dwelling, its size (area in square meters), its number of rooms, its equipment, whether it is a detached house or an apartment, and whether the household is a homeowner or renter. Very detailed information is available on the quality and the surroundings of the dwelling: whether the dwelling is too noisy, too dark, in poor condition or in need of repairs, whether there is some pollution or vandalism in the surrounding area, etc.

We construct two different samples to study the housing choices when retiring. The first one includes all the observations related to the individuals who retire during the panel period.[10] We call it the retiring sample. We assume that age at retirement cannot be below 50 and the few observations violating this criterion were deleted. This leaves us with a sample of exactly 627 individuals (3,892 person-year observations).[11] The focus of this sample is on at-risk individuals, since the selected respondents retired during the panel period. Nevertheless, some uncertainty remains on the measure of mobility at retirement even when using this restricted sample. For instance, we do not have any information on moves just after retirement for individuals retiring at the end of the panel. Also, some people in the panel claim that they are out of the labour force at age 55 and are still so eight years later. Although these people may certainly be regarded as retired after age 60 (which is in many cases the mandatory date for retirement in France), they are

[9] For instance, the number of households is 6,722 in 1995 and 5,343 in 2001.

[10] To select these observations, we construct at each date a variable indicating the occupational status. Then, we keep only the observations of individuals who report they were occupied in the first interview and possibly later, but were retired at the time of their last interview or before.

not included in our retiring sample.

Our second sample, called the elderly sample, is based on a simple age criterion. We keep observations only for people aged between 55 and 70 when first interviewed. This allows us to make a comparison between our results and those obtained for Britain by Ermisch and Jenkins (1999) who selected individuals aged 55 and older (although we choose to put an upper bound on the age interval). This second sample includes 3,029 individuals corresponding to 15,491 person-year observations. The number of observations is far larger than in the retiring sample, which can be helpful as mobility of the elderly remains rather scarce. However, retiring people are not as well targeted as with the retiring sample.

4.2. Factors influencing residential mobility

We first describe the magnitude of residential mobility. Using the retiring sample, we find that the average annual mobility rate is 3.4% (see Table 3). The annual mobility rate is lower for the elderly sample, as it reaches only 2.5%. Ermisch and Jenkins (1999) obtain a slightly higher rate (3.3%) using their sample of people aged 55 and over extracted from the British Household Panel Survey (BHPS). Our mobility rate is also much lower than the one observed in the United States (Feinstein and McFadden, 1989). It is well acknowledged that housing mobility is less important in France than in other countries. Finally, simple calculations show that the proportion of people moving at least once during the panel period is equal to 18.8% when using the retiring sample (and 11.7% when using the elderly sample).[12]

Insert Table 3 here

It is not surprising to find a lower propensity to move when using the elderly sample as all the individuals are not necessarily at risk. Indeed, some individuals in this sample are out of labour force during the entire 1994-2001 period. In particular, this is the case for the individuals who are already retired at the beginning of the panel and may have made a move related to retirement before the first survey date. In both the elderly and retiring samples, mobile respondents usually make one move only during the 1994-2001 period (Table 3). Using the

---

[11] In our empirical analysis, we follow Ermisch and Jenkins (1999) and Tatsiramos (2006) and do not account for the possibility of selective attrition, i.e mobile retirees may have a higher probability of leaving the panel than non-movers. For further details on attrition in the European Community Household Panel survey, see Behr et alii (2005).

[12] The proportion of movers is lower in the Europanel survey than in the 3G survey. This can occur because the period of reference is not the same, as we use a retrospective question on retired people in the 3G survey. Also, in the 3G survey, a move at retirement is reported for an individual whether he or his partner retired. Finally, attrition may

retiring sample, we find that 88.1% of mobile respondents move once, 10.2% of them move twice and only 1.7% of them move three times. Multiples moves may be caused by the dissatisfaction with a new dwelling, or the unexpected inheritance of some housing assets.

We now describe the timing of residential mobility for movers. We only consider movers of the retiring sample who are in the panel at least two years before retirement and are still observed two years after retirement. We examine whether these movers leave their current dwelling just after retirement or whether they anticipate by moving before retirement. In Figure 1, we represent the distribution of these movers over the five-year period that begins two years before retirement. We find a peak of residential mobility in the year of retirement: 43% of mobile individuals move exactly the same year as they retire. The proportion of moves made one year after retirement is also quite high, as it stands at 26%. Conversely, only 17% of moves occur during the two years before retirement. The lower tail of the distribution exhibits an increasing profile. These results support the fact that the residential mobility observed in our sample is related to retirement.[13]

Insert Figure 1 here

As owning additional housing assets is likely to favour mobility, we study the role of secondary home ownership on the mobility decision. We report in Table 4 for the retiring subsamples of movers and non-movers, the proportion of households owning a secondary home in t, depending on whether or not they owned a secondary home in t-1. Among movers, only 50% of households owning a secondary home before their move still own a secondary home after their move. The corresponding proportion for non-movers is about 30 percentage points higher (79.7%). This finding suggests that at retirement, mobile people owning some additional housing assets before their move would often live in their secondary home after their move.

Insert Table 4 here

We also estimate Probit models to highlight the variables which influence the decision to move between dates *t-1* and *t*. We use the elderly sample, and observations are defined on an

---

reduce the proportion of movers in the Europanel survey.

[13] Additional results (not reported here) show that it is difficult to explain the timing of residential mobility. For instance, gender, marital status, income, homeownership and size of the previous dwelling do not influence the timing of the mobility decision. However, individuals who were farmers or self-employed during their working life move more often at the exact date of retirement. When retiring, farmers and self-employed stop renting their job-related assets if they are renters, or sell their workplace if they are owner. Also, owners who sell their equipment tools have some liquid assets that can be used to purchase a new dwelling.

individual-year basis.[14] According to Table 5, the probability of moving is higher for "younger" respondents and for those being married in *t-1*, while gender, the number of dependent children and the limitations of daily activities for health reasons do not have any significant effect. Mobility is higher for highly-educated respondents. This could be due to a permanent wealth effect which is imperfectly captured by income in *t-1*. Mobility is clearly related to the retirement decision. The probability of moving is much higher when retiring between *t-1* and *t*, but mobility is not significantly affected by retirement between *t-2* and *t-1*.

Insert Table 5 here

Having a spouse working in *t-1* strongly reduces the probability of moving. This is not surprising as costs incurred when moving are probably much higher.[15] Housing conditions also play an important role in the mobility process of the elderly. Living in a detached house in *t-1* reduces the probability of moving. Interestingly, we do not find any significant effect of the number of excess rooms (which is defined as the number of rooms minus the number of persons living in the household). Thus, mobility would not be mainly driven by adjustments to the dwelling size, and other factors such as location or housing quality may matter more. We find that the duration of residence and ownership have a large negative effect on mobility. Mobility costs are higher for owners as they have to incur some significant transaction costs and tend to invest more in local amenities and social capital (DiPasquale and Glaeser, 1999).

Table 5 also reports the estimates of Probit regressions for each housing status in *t-1*. Results show that there are some differences between renters and owners. Indeed, the influence of retirement on the probability of moving is much higher for renters than for owners. It may be that renters are the households who moved the most during their career and that they waited until retirement to choose a stable location. Alternatively, renters may be unable to pay their rent after retirement as their income decreases. We also find that the duration of residence has a significant negative effect only for renters, which may be due to rent control. After a few years, the rent paid by a renter is likely to be far below the market rent.

---

[14] The standard errors of the Probit estimates are adjusted for clustering at the individual level.

[15] In our regression, we follow Ermisch and Jenkins (1999) and include a dummy variable for the spouse being employed. However, we do not study the possibility that the retirement of a spouse initiates a residential move. Sédillet and Walraet (2002) have shown that in France women are much more likely than men to take into account the occupational situation of their spouse when making their retirement decision. It could then be that the

4.3. Housing adjustments

We now study the housing adjustments of movers. Specifically, we assess whether the households increase or decrease the size and quality of their dwelling as the theory on this issue leads to inconclusive predictions. Since individuals are tracked over time in the Europanel survey, we are able to compare their housing situation in *t* with the one in *t-1*.

In Table 6, we compare the size of the dwelling before and after moving. For a given housing status and number of excess rooms in *t-1*, we report the change in the number of excess rooms for movers. We obtain very similar results for owners and renters. When the number of excess rooms is zero or negative, individuals tend to increase the size of their dwelling when moving. When the number of excess rooms is at least three, households choose to reduce the size of their dwelling.[16]

Insert Table 6 here

We then estimate Probit models on the subsample of movers to evaluate the effect of the number of rooms in *t-1* on the change in the number of excess rooms, all else being equal. We run two separate regressions. The first one explains the decrease in the number of excess rooms, while the second one explains the increase in the number of excess rooms. Results reported in Table 7 show that there is a "downsizing" effect for movers previously living in a dwelling which was large compared to the number of people living in it, while there is an "upsizing" effect for movers previously living in a dwelling which was small.

Hence, our results suggest that in France, elderly people tend to correct their initial disequilibrium between the number of rooms and the number of occupants when retiring. This finding is in accordance with the results described in Ermisch and Jenkins (1999) for Britain. Nevertheless, it should be kept in mind that adjusting the size of the dwelling was not a key factor when explaining mobility (see Table 5). It is only when retirees have decided to move for one among various reasons that they take the opportunity to choose a more appropriate number of rooms given their situation.

Insert Table 7 here

We also study the changes in housing quality using two sets of variables. The first one

---

anticipation of mobility at the retirement of the household head triggers the retirement of his spouse. The retirement decision of the spouse is thus likely to be endogenous and we chose not to include it in the regressions.

[16] Interestingly, homeowners with two excess rooms before their move keep this number of excess rooms after their move. As in France, the average number of children is two, it may be that movers keep a large dwelling to

includes all types of housing equipment both in *t-1* and in *t*. If there is on average an improvement of the equipment when moving, there should be more households in the cell "not in *t-1*, present in *t*" than in the cell "present in *t-1*, not in *t*". Results reported in Table 8 suggest that mobile respondents tend to move to dwellings that are better equipped. First, the few households who did not have very basic facilities (bath, shower, warm water, etc) move to dwellings equipped with them. Secondly, households tend to access better heating systems after moving, in particular central heating or electric fires. Third, mobile households have more often access either to a terrace or to a non-vegetable garden after their move.[17]

Insert Table 8 here

The second set of variables characterizes the problems faced by households when living in their home. There may be some problems related to the dwelling itself such as the dwelling being too dark, too noisy, poorly heated, or with material in a bad condition (non-isolated windows for instance). As shown by Table 8, all these problems occur much less frequently after moving. For instance, 13% of respondents report that their dwelling was too dark in *t-1* but is not in *t*, while only 3% of respondents have a dwelling too dark in *t* but not in *t-1*. There may also be some insecurity or pollution problems in the neighbourhood. Again, movers tend to escape these negative externalities. For instance, 20.6% of movers claim that there were insecurity and vandalism problems in their neighbourhood in *t-1*, but do not report this kind of troubles in *t*. Conversely, only 5% of movers live in an insecure place after their move whereas it was not the case before their move.

We then examine whether movers are renters or owners in *t* depending on their housing tenure in *t-1* (Table 9). We find that the housing tenure before moving plays a major role: 65% of mobile households who were renters in *t-1* are still renters in *t* and 35% have accessed ownership. Conversely, 86% of mobile owners are also owners of their new dwelling in *t*. The inertia in the owner status may result from intrinsic preferences for homeownership. Also, homeowners who sell a dwelling have some liquid assets that can be used to purchase another dwelling. The lower inertia in the renter status may be explained by some specific mobility behaviours of renters during the working period. People who move frequently because of their professional occupation

---

accomodate their children when they are visiting.

[17] The proportion of individuals having an independent kitchen is 90% before the move. After a move, we note that fewer respondents have access to an independent kitchen. However, the interpretation of this result is intricate.

sometimes choose to rent a dwelling as mobility is much easier in that case. These renters may wait for retirement to purchase a dwelling.

Insert Table 9 here

We can also assess from Table 9 whether movers choose to live in an apartment or a detached house in *t* depending on their type of dwelling in *t-1*. We find that transitions from an apartment to a detached house are rather frequent and account for 45.5% of respondents living in an apartment in *t-1*. Conversely, most of the respondents occupying a detached house in *t-1* choose to live in a detached house again in *t*. This may be due to a wealth effect as detached houses are mostly in ownership. Indeed, homeowners can use the money they get from selling their previous detached house to purchase another detached house. It is clearly more interesting to choose a detached house rather than an apartment as there is more privacy and less trouble from neighbours when living in a detached house. Moreover, there is often a garden going along with a detached house, which makes it more valuable.

We also find that retiring individuals are likely to move to more pleasant areas when changing their location. A significant proportion of movers choose to leave the Paris region when retiring. While about 18% of mobile households resided in this region before moving, the proportion decreases to 12.8% after moving. In Paris, housing is much more expensive than anywhere else in France and even if the city is characterized by unique cultural features, it is somewhat polluted by fumes and loud noises from various sources. According to the data, the most attractive areas (i.e. those where retirees are more likely to live after moving) are mainly located in the South-West and in the South of France.[18] Finally, mobile respondents tend to move to smaller municipalities after retiring. For instance, the proportion of mobile households living in rural areas (less than 2,000 inhabitants) is 27% before moves and 31% after moves.

5. Conclusion

In this paper, we have analyzed the mobility and housing choices of retiring people in France. These issues are especially relevant for that country given the ongoing and forthcoming retirement of the large cohorts of baby-boomers born in the 1950s and 1960s. From a theoretical viewpoint, individuals may want to either increase or decrease their dwelling size and quality

---

Indeed, in many modern dwellings, the kitchen is part of the living room, so that having an independent kitchen does not mean living in a higher quality dwelling.

[18] On average, these areas are sunny and located near the sea.

when retiring. This is because retirees suffer from a loss of income due to retirement and moving costs, but have more leisure time and thus value housing more.

Using two French data sets, we have shown that residential mobility is substantial when retiring and that various motives like housing adjustments, climate and family location are put forward by mobile households for their move. The use of panel data provides additional results on residential mobility at retirement. Our main conclusions are that a significant proportion of people choose to move the year they retire or one year later, and that housing conditions (including location) are often better after moving. Furthermore, housing adjustments most often lead to a correction of the initial disequilibrium between the number of rooms and the number of occupants. "Downsizing" is mainly observed for movers previously living in a dwelling with some excess rooms, while "upsizing" is more frequent for movers whose living space is constrained. Interestingly, all these results are very similar to those found for the UK by Ermisch and Jenkins (1999).

Several extensions to our contribution may come to mind. Firstly, it would be of interest to compare the residential mobility and motives for moving of workers and retirees. Young adult workers are expected to upsize more often because of births and wealth accumulation. Also, it is unlikely that young workers can afford to move to sunny places on the coast as housing there is usually costly.

Secondly, one could gather and use detailed information on consumption amenities and housing prices to analyze their effect on the location decision. Indeed, retirees should be attracted by places with more consumption amenities and lower prices. Note that studying the effect of housing prices on the location decision without taking into account the consumption amenities may lead to a downward bias as housing prices are usually higher in places where there are more consumption amenities.

Thirdly, it could be worth assessing the direct impact of the residential mobility of retirees on the local housing prices. Indeed, prices are likely to increase in places with consumption amenities where retirees tend to move (such as places characterized by sunny weather). This in turn could prevent younger households from locating there. The change in the demographic composition of cities may then have an impact on the level of local taxes which could in turn affect mobility. A general equilibrium framework is clearly needed to better understand this kind of mechanisms and more generally all the consequences of a demographic shock like the baby-boom on housing

choices and local development.

Finally, comparative evidence on residential mobility in European countries could help to better understand migration inflows of retirees from Northern to Southern European countries. In some areas of Mediterranean Europe, international migrations at retirement clearly have some large effects on the local economy (King et alii, 1998). We leave all these issues for future research.

References

Attias-Donfut C., Wolff F.C. (2005), "Generational memory and family relations", in M. Johnson, V. Bengtson, P. Coleman, T. Kirkwood (eds), *The Cambridge Handbook of Age and Ageing*, Cambridge University Press.

Behr A., Bellgardt E., Rendtel U. (2005), "Extent and determinants of panel attrition in the European Community Household Panel", *European Sociological Review*, vol. 21, pp. 489-512.

Börsch-Supan A., Hajivassiliou V., Kotlikoff L., Morris J. (1992), "Health, children, and elderly living arrangements", in D.A. Wise (ed.), *Topics in the Economics of Aging*, University of Chicago Press, Chicago, pp. 79-103.

Cardia E., Ng S. (2003), "Intergenerational time transfers and childcare", *Review of Economic Dynamics*, vol. 6, pp. 431-454.

Chen Y., Rosenthal S. (2008), "Local amenities and life-cycle migration: Do people move for jobs or for fun?", *Journal of Urban Economics*, vol. 64, pp. 519-537.

De Coulon A., Wolff F.C. (2006), "The location of immigrants at retirement: Stay/return or 'va-et-vient' ?", *IZA Discussion Paper*, n° 2224.

Dimova R., Wolff F.C. (2009), "Does grandchild care enhance maternal labor supply? Evidence from around Europe", *Journal of Population Economics*, forthcoming.

DiPasquale D., Glaeser E.L. (1999), "Incentives and social capital: Are homeowners better citizens ?", *Journal of Urban Economics*, vol. 45, pp. 354-384.

Ermisch J., Jenkins S. (1999), "Retirement and housing adjustment in later life: evidence from the British Household Panel Survey", *Labour Economics*, vol. 6, pp. 311-333.

Feinstein J., McFadden D. (1989), "The dynamics of housing demand by the elderly: wealth, cash flow, and demographic effects", in D.A. Wise, (ed.), *Economics of Aging,* University of Chicago Press, Chicago, pp. 55-91.

Flavin M., Nakagawa S. (2008), "A model of housing in the presence of adjustment costs: A structural interpretation of habit persistence", *American Economic Review*, vol. 98, pp. 474-495.

Gobillon L., Le Blanc D. (2004), "L'impact des contraintes d'emprunt sur la mobilité résidentielle et les choix de statut d'occupation des ménages : un modèle simple de demande", *Annales d'Economie et de Statistique*, vol. 74, pp. 15-46.

Gobillon L., Le Blanc D. (2008), "Economic effects of upfront subsidies to ownership: the case of the Prêt à Taux Zéro in France", *Journal of Housing Economics*, vol. 17, pp. 1-33.

Graves P., Knapp T. (1988), "Mobility behavior of the elderly", *Journal of Urban Economics*, vol. 24, pp. 1-8.

Hardman A.M., Ioannides Y.M. (1995), "Moving behaviour and the housing market", *Regional Science and Urban Economics*, vol. 25, pp. 21-39.

Henderson V., Ioannides Y. (1983), "A model of housing tenure choice", *American*

Economic Review, vol. 73, pp. 98-113.

Hoerger T., Picone G., Sloan F. (1996), "Public subsidies, private provision of care and living arrangements of the elderly", *Review of Economics and Statistics*, vol. 78, pp. 428-440.

Hogan T., Steinnes, D. (1998), "A logistic model of the seasonal migration decision for elderly households in Arizona and Minnesota", *Gerontologist*, vol. 38, pp. 152-158.

Jellal M., Wolff F.C. (2002), "Cultural evolutionary altruism : Theory and evidence", *European Journal of Political Economy*, vol. 18, pp. 241-262.

King R., Warnes A.M., Williams A.M. (1998), "International retirement migration in Europe", *International Journal of Population Geography*, vol. 4, pp. 91-111.

Konrad K.A., Kunemund H., Lommerud K.E., Robledo J.R. (2002), "Geography of the family", *American Economic Review*, vol. 92, pp. 981-998.

Laferrère A. (2005), "Old age and housing: dissaving, adjusting consumption, and the role of children", *mimeographed*, INSEE.

Linneman P., Wachter S.M. (1989), "The impacts of borrowing constraints on homeownership", *AREUEA Journal*, vol. 17, pp. 389-402.

Perrachi F., (2002), "The European Community Household Panel: A review", *Empirical Economics*, vol. 27, pp. 63-90.

Pezzin L.E., Schone B.S. (1999), "Intergenerational household formation, female labor supply and informal caregiving. A bargaining approach", *Journal of Human Resources*, vol. 34, pp. 475-503.

Roback J. (1982), "Wages, Rents and the Quality of Life", *Journal of Political Economy*, vol. 90, pp. 1257-1278.

Sédillot B., Walraet E. (2002), "La cessation d'activité au sein des couples : y a-t-il interdépendance des choix ?", n° 357-358, pp. 79-98.

Tatsiramos K. (2006), "Residential mobility and housing adjustment of older households in Europe", *IZA Discussion Paper*, n° 2435.

Venti S.F., Wise D.A. (1990), "But they don't want to reduce housing equity", in D.A. Wise (ed.), Issues in the *Economics of Aging*, University of Chicago Press, Chicago, pp. 13-29.

Zorn P.M. (1989), "Mobility-tenure decisions and financial credit: Do mortgage qualification requirements constrain homeownership?", *AREUEA Journal*, vol. 17, pp. 1-16.

Table 1. Self-reported motives for moving

Self-reported motive	%
Family motive	16.5
To live closer to parents	1.8
To live closer to children	14.7
Geographic motive	9.2
To live closer to native country	6.3
To live in a sunny area	2.9
Professional motive	29.1
Housing motive	15.5
To live in a larger dwelling	7.9
To live in a smaller dwelling	7.6
Other motive	29.8
Number of observations	383

Source: *1992 Trois Générations Survey*.

Table 2. Self-reported motives for moving, by respondent's characteristics

Variable	Motive for mobility					Chi² (prob.)	
	Family	Geographic	Professional	Housing	Other		
Gender							
Male	10.3	11.2	30.8	15.0	32.7	4.69 (.321)	(N=107)
Female	18.8	8.3	28.3	15.6	29.0		(N=276)
Living as a couple							
No	21.9	5.2	19.8	14.6	38.5	11.34 (.023)	(N=96)
Yes	14.6	10.5	32.1	15.7	27.2		(N=287)
Education							
No education	14.9	5.2	25.9	20.1	33.9	26.02 (.011)	(N=174)
Primary	19.7	10.6	34.5	12.7	22.5		(N=142)
Above primary	12.1	16.7	25.8	9.1	36.4		(N=66)
Occupation							
Farmer	9.1	3.0	40.9	10.6	36.4	57.45 (.000)	(N=66)
Self-employed	9.7	7.6	56.6	7.6	18.9		(N=53)
Executive	13.0	17.4	23.9	13.0	32.6		(N=46)
Employee	22.9	5.2	20.8	21.9	29.2		(N=96)
Blue collar	15.9	12.2	17.1	22.0	32.9		(N=82)
Out of labour force	27.5	15.0	22.5	7.5	27.5		(N=40)
Income							
Quartile 1	10.4	6.3	30.2	18.8	34.4	12.27 (.424)	(N=96)
Quartile 2	21.4	8.3	29.8	11.9	28.6		(N=84
Quartile 3	21.4	7.1	28.6	15.3	27.6		(N=98)
Quartile 4	13.3	14.3	27.6	15.2	29.5		(N=105)
Total	16.5	9.2	29.1	15.5	29.8	-	(N=382)

Source: *1992 Trois Générations Survey*.

Table 3. The magnitude of residential mobility

Variables	Individuals retiring during the panel	Individuals between 55 and 70 when first interviewed
Annual movement rate	3.4%	2.5%
	(N=3,892)	(N=15,491)
Proportion of movers	18.8%	11.7%
	(N=627)	(N=3,029)
Number of moves for movers		
1 move	88.1%	90.9%
2 moves	10.2%	8.5%
3 moves	1.7%	.6%
	(N=118)	(N=353)

Source: *1994-2001 Europanel Survey*.

Table 4. Distribution of individuals, by mobility decision and secondary home ownership in *t-1* and *t* (row percentage)

Moving in *t* (N=118)			
Owns a secondary home in *t-1*	Owns a secondary home in *t*		
	No	Yes	
No	86.4%	13.6%	(N=66)
Yes	50.0%	50.0%	(N=52)
Not moving in *t* (N=509)			
Owns a secondary home in *t-1*	Owns a secondary home in *t*		
	No	Yes	
No	94.2%	5.8%	(N=396)
Yes	20.3%	79.7%	(N=113)

Source: *1994-2001 Europanel Survey*.

Table 5. Determinants of residential mobility between years *t-1* and *t*

Variables	All	Renters in *t-1*	Owners in *t-1*
Constant	-.228	.561	-.672
	(.49)	(.75)	(1.07)
Female	.008	-.107	.063
	(.15)	(1.26)	(.92)
Age in *t-1*	-.019**	-.023**	-.023**
	(3.46)	(2.68)	(2.98)
Married in *t-1*	.182**	.054	.329**
	(2.89)	(.65)	(3.31)
Number of dependent children in *t-1*	-.035	-.072	-.014
	(.95)	(1.31)	(.25)
Health condition limiting daily activities in *t-1*	.041	.007	.072
	(.81)	(.09)	(1.05)
Education (ref: no diploma)			
Primary	.114	.252**	-.075
	(1.64)	(2.65)	(.78)
Vocational	.110	.371**	-.102
	(1.23)	(2.88)	(.81)
High school	.221*	.603**	-.086
	(2.06)	(3.59)	(.59)
Undergraduate	.230*	.211	.062
	(2.30)	(1.17)	(.50)
Graduate, postgraduate	.265**	.619**	-.171
	(2.67)	(4.20)	(1.20)
Retiring between *t-1* and *t*	.441**	.578**	.368**
	(4.52)	(3.69)	(2.75)
Retiring between *t-2* and *t-1*	.183	.271	.139
	(1.43)	(1.35)	(.78)
Spouse working in *t-1*	-.322**	-.463**	-.326*
	(3.23)	(2.95)	(2.39)
Household income in *t-1* (log)	.016	-.013	.029
	(.87)	(.35)	(1.55)
Detached house in *t-1*	-.336**	.155	-.763**
	(3.97)	(1.38)	(6.96)
Number of excess rooms in *t-1*	-.017	-.081	.034
	(.77)	(1.94)	(1.20)
Duration in residence in *t-1* (10e-2)	-.489**	-1.548**	.183
	(2.62)	(4.85)	(.83)
Home ownership in *t-1*	-.496**		
	(7.54)		
Number of observations	13,590	3,124	10,466
Number of moves	313	161	152
Pseudo R^2	.081	.092	.079
Log likelihood	-1369.4	-575.7	-731.3

Source: *1994-2001 Europanel Survey*.

Note: Probit models with robust standard errors, adjusted for clustering at the individual level. The different regressions also include 8 regional dummy variables and 5 dummy variables related to the size of the area. Absolute value of robust t-statistics are in parentheses, and significances levels are respectively 1% (**) and 5% (*).

Table 6. Change in excess rooms among movers, by number of excess rooms in t-1 (row percentages)

Renter in *t-1* (N=157)				
Number of excess rooms in *t-1*	Adjustment in the number of excess rooms between *t-1* and *t*			
	Decrease	Same	Increase	
≤ 0	.0	44.1	55.9	(N=34)
1	22.8	43.9	33.3	(N=57)
2	34.0	38.3	27.7	(N=47)
≥ 3	68.4	15.8	15.8	(N=19)
Owner in *t* (N=133)				
Number of excess rooms in *t-1*	Adjustment in the number of excess rooms between *t-1* and *t*			
	Decrease	Same	Increase	
≤ 0	.0	35.3	64.7	(N=17)
1	18.5	40.7	40.7	(N=27)
2	19.4	64.6	16.1	(N=31)
≥ 3	72.4	17.2	10.3	(N=58)

Source: *1994-2001 Europanel Survey*.

Table 7. Determinants of decrease/increase in number of excess rooms

Variables	Decrease	Increase
Constant	-3.725*	1.751
	(2.09)	(1.06)
Female	-.105	.240
	(.47)	(1.13)
Age in *t-1*	.039	-.043*
	(1.72)	(2.10)
Married in *t-1*	-.711*	.935**
	(2.40)	(3.54)
Health condition limiting daily activities in *t-1*	.343	-.150
	(1.62)	(.76)
Education (ref: no diploma)		
Primary	.109	.067
	(.42)	(.24)
Vocational	-.248	.077
	(.68)	(.24)
High school	-.043	.604
	(.11)	(1.63)
Undergraduate	-.547	.655
	(1.34)	(1.66)
Graduate, postgraduate	-.644	.481
	(1.64)	(1.21)
Household income in *t-1* (log)	-.147	.104
	(1.80)	(1.12)
Number of rooms in *t-1*	.825**	-.577**
	(7.07)	(5.12)
Home ownership in *t-1*	-.236	-.206
	(1.01)	(.91)
Detached house in *t-1*	.198	.330
	(.78)	(1.37)
Duration in residence in *t-1* (x 10e-2)	.000	.014
	(.07)	(1.93)
Number of observations	266	266
Number of decrease/increase	89	78
Pseudo R^2	.375	.296
Log likelihood	-106.0	-113.2

Source: *1994-2001 Europanel Survey*.

Note: Probit models with robust standard errors. The different regressions also include 8 regional dummy variables. Absolute value of robust t-statistics are in parentheses, and significances levels are respectively 1% (**) and 5% (*).

Table 8. Changes in housing conditions for movers (rows percentage)

Changes	Not in *t-1*, not in *t*	Not in *t-1*, present in *t*	Present in *t-1*, not in *t*	Present in *t-1*, present in *t*
Housing equipment (N=289)				
Independent kitchen	2.4	4.8	11.0	81.7
Rooms for professional use	92.7	1.4	5.9	.0
Bath. Shower	.0	6.6	.0	93.5
WC in the housing	.0	4.8	.3	94.8
Warm running water	.3	4.8	.0	94.8
Central heating	10.4	18.7	11.5	59.4
Electric fires	60.0	16.2	11.7	12.1
Other means of heating	36.2	16.2	24.8	22.8
Non-vegetable garden. terrace	18.9	22.1	14.8	44.5
Problems related to housing (N=323)				
Housing too small	80.6	6.2	12.3	.9
Noisy neighbours	84.6	2.8	10.5	2.2
Noisy surroundings	67.3	9.6	17.9	5.3
Dwelling too dark	84.0	3.1	13.0	.0
Inefficient heating	84.3	4.4	9.4	1.9
Leak in the roof	89.2	2.8	8.1	.0
Damp on the walls. on the floor	76.1	4.7	17.4	1.9
Windows or floor in bad condition	83.6	2.5	13.0	.9
Pollution. heavy traffic	74.3	7.4	13.9	4.3
Insecurity. Vandalism	70.1	5.0	20.6	4.4

Source: *1994-2001 Europanel Survey*.

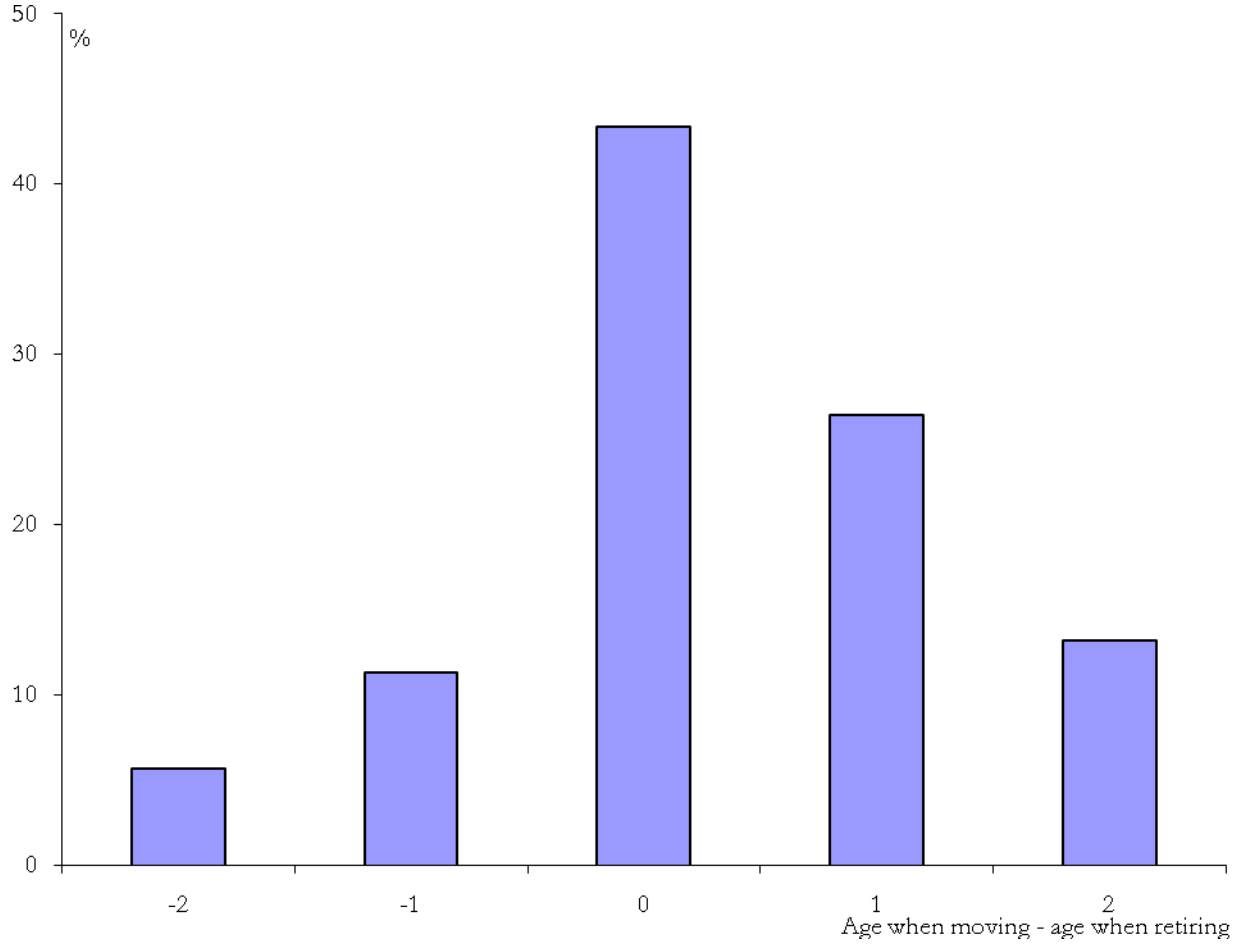
Note: the difference in the number of households for housing equipment and problems faced by the housing is due to some missing values.

Table 9. Tenure and housing-type transition for movers (row percentage)

Tenure transition (N=323)			
	Tenure in *t*		
Tenure in *t-1*	Renter	Owner	
Renter	65.0	35.0	(N=162)
Owner	13.7	86.3	(N=161)
Housing-type transition (N=291)			
	Type in *t*		
Type in *t-1*	Apartment	Detached house	
Apartment	54.5	45.5	(N=123)
Detached house	26.0	74.0	(N=168)

Source: *1994-2001 Europanel Survey*.

Figure 1. Difference between the age when moving and the age when retiring (for movers)



Source: *1994-2001 Europanel Survey*.

B.8. The effect of Being Widowed on Housing and Location Choices

Bonnet C., Gobillon L. et A. Laferrère (2009), “The effect of Being Widowed on Housing and Location Choices”, CREST Working Paper 2008-12, revise-and-resubmit au *Journal of Housing Economics*, version révisée.

37 pages

The effect of widowhood on housing and location choices*

Carole Bonnet (INED)† Laurent Gobillon (INED, PSE-INRA and CREST)‡
Anne Laferrère (INSEE and CREST)§

January 8, 2010

Abstract

The number of elderly persons living alone is increasing and their influence on the housing market is getting larger. This paper investigates the effect of the loss of a spouse on housing and location choices. A partner's death induces a decrease in income which may lead to downsizing. Widowhood may also reveal new preferences, such as the need to be close to care givers and health services. We estimate the effect of a transition to widowhood on housing consumption and location choices using the French Housing Surveys. Widowhood significantly increases residential mobility, especially at older ages and for those who have children. Mobile widows tend to live closer to their relatives but do not move to co-reside with a child. Housing and location adjustments are consistent with new widows moving to dwellings that are smaller, more often apartments and in the rental sector, and on average located in larger municipalities where services are more accessible. The housing demand of widows will be significant in the next twenty years, especially the demand for small dwellings.

*We are grateful to Catherine Bonvalet, Anne Solaz, Joëlle Gaymu, François-Charles Wolff, as well as to two anonymous referees, for their useful comments.

†INED, 133 Boulevard Davout, 75980 Paris Cédex 20. Tel: 00 33 (0) 1 56 06 22 36. email: carole.bonnet@ined.fr.

‡INED. Tel: 00 33 (0) 1 56 06 20 16. email: laurent.gobillon@ined.fr.

§INSEE, 18 Boulevard Adolphe Pinard, 75675 Paris Cedex 14. Tel: 00 33 (0) 1 41 17 55 74. email: anne.laferrere@insee.fr.

1 Introduction

Population ageing will change many societies in unprecedented ways. Governments are debating to what extent demographic changes are a threat to the financial sustainability of pension and healthcare systems. How housing adjustments at older ages may have an impact on the housing market is less often investigated. In this context, widowhood is becoming more important, with the arrival of large baby-boom cohorts at the age of the loss of a spouse or partner (Kalogirou and Murphy, 2006). According to the official household projections of the French Institute of Statistics, the number of couples aged 60 and over will increase by 28 percent between 2010 and 2030 while the number of single-person households aged 60 and over will be 60 percent higher (Jacquot, 2007). Among the latter, many will be widows.

Widowhood affects welfare in many ways. It affects income and living standards as the survivor's pension is smaller than the partner's income. The budget share of housing is large and housing consumption presents economies of scale that are lost when the partner dies. For these reasons, a surviving spouse may want to downsize. Widowhood also affects living arrangements. It is well documented that a fair amount of care to the disabled elderly is provided by a spouse, most often by the wife (Chappell, 1991, Freedman, 1996). In case of need, a widow has to turn to other family members, or to professionals financed by private or public insurance. The issue of long term care is related to the housing choices of the oldest old, as they choose between accommodation in nursing homes or personal care in their own dwelling. As the baby-boom generations reach age of widowhood, their impact on the housing market may be considerable. We study the residential mobility, housing and location choices of recent widows and widowers. Do they downsize? Do they relocate? Our goal is to get some evidence on the impact of the residential mobility of widows on the housing market and on the extent to which widows may rely on kinship for support.

The specific residential mobility and housing choices of the elderly have not often been analyzed in the economic literature except in a few empirical papers (Venti and Wise, 1987; Börsch-Supan, 1990; Ermisch and Jenkins, 1999; Venti and Wise, 2001; Tatsiramos, 2006; Laferrère, 2005 and 2006). These studies adopt a broad view, looking at the effect of all shocks - job change, retirement, widowhood - on mobility. They also analyze the change in housing characteristics and location when a move occurs. However, they do not usually disentangle the various causes of mobility.

Hence, the results are generated by a mix of several economic and socio-demographic effects. Conversely, the literature on widowhood does not look much at mobility (with the exception of Chevan, 1995) and housing choices, and focuses more on widows' living arrangements at a given point in time, but not on their dynamics (Macunovich et al., 1995, Costa, 1999, Iacovou, 2000).[1] The present paper tries to reconcile the two approaches and explain how widowhood may lead to mobility, housing adjustments and relocation.

We find that the loss of a spouse has a significant positive impact on residential mobility, especially at older ages. *Ceteris paribus*, when the partner dies, the probability of moving within the next four years is nearly 90 percent higher than if no death occurred. A childless widow is less likely to move. Mobile widows tend to live closer to their relatives but moving to co-reside with a child is extremely rare. Compared to mobile couples, mobile widows are more likely to decrease their number of rooms and to choose the rental sector. They also switch more often from a house to an apartment. Finally they move more often to larger municipalities. These results on housing and location adjustments are consistent with a tendency among single elderly people to move closer to health and personal care services.

We then simulate the housing demand of additional widows over the 2010-2030 period relying on official households projections. We find that this demand would represent nearly 8 percent of yearly new constructions. The demand would be especially important for apartments and small dwellings. For one or two room units, it would represent between 13 and 19 percent of the yearly construction in this submarket.

Section 2 provides descriptive statistics and presents some institutional features related to widowhood in France. In section 3, we discuss the effect of a transition to widowhood on housing and location choices. We then test some of the mechanisms on data from the French Housing Surveys, which are described in section 4. Section 5 delineates our empirical findings. Section 6 presents some simulations. Section 7 concludes.

[1] Another strand of the literature studies the living arrangements of the elderly in a dynamic setting but does not focus on widows (see Börsch-Supan et al. 1992; Heiss, Hurd and Börsch-Supan, 2003).

2 The French setting

We first present some stylized facts on widowhood after age 60. Figure 1 shows the proportion of widows and widowers by age, for five birth cohorts. The proportion increases with age, and is always larger for women than for men. The difference can be explained by the higher death rate of men and by the age gap between spouses, as wives are on average 2.5 years younger than their husbands. For instance, for women born in 1920, the rate of widowhood at age 80 is 60 percent, more than 3 times the rate for men (17 percent). It means that a large majority of married men live with their spouses until death whereas a large majority of women spend part of their life as widows. This justifies our use of the word widow instead of widow or widower in this paper. At a younger age, the rate of widowhood also decreases from one cohort to the next. This is due to the general increase in life expectancy which makes widowhood occur later in the life-cycle.[2] As a result, the loss of a spouse or a partner is more and more likely to be combined with old age problems.

[*Insert Figure* 1]

The death of a spouse induces a fall in household resources as it entails the loss of the partner's income. To compensate for the loss, widows in many countries are eligible for social security benefits in the form of a survivor's pension (see Burkhauser and alii, 2005). In France, the average survivor's pension varies between 50 percent and 60 percent of the deceased spouse's pension (COR, 2008). Hence, in many cases, it does not fully compensate for the income loss related to widowhood.[3] It is not possible to compute the change in income due to widowhood using the French Housing Surveys. Nevertheless, we can recover some indirect information from the average income of cohorts by age of the household head (see Figure 2).

[*Insert Figure* 2]

[2] Deaths among unmarried couples are not recorded here as widowhood. The bias is negligible for the current cohorts aged 60 or more. But for the future generations who form lasting partnerships outside marriage, a new word may have to be found for the loss of a partner outside marriage.

[3] Suppose the husband received a pension P_H and the wife has no pension. After the husband's death, the survivor pension will be $0.55P_H$. With the most commonly used equivalence scale, the living standard of the surviving spouse will decrease from $0.7P_H$ (ie. $\frac{P_H}{\sqrt{2}}$) to $0.55P_H$, i.e. by about 22 percent. Assuming that the wife received P_F equal to one third of P_H, the decrease in the living standard will be around 6 percent.

The average income does not decrease after age 70. This is surprising at a first glance because many couples experience the death of one partner at that age (see Figure 1). The observed stability of income may be generated by three main mechanisms:

- First, as mentioned above, the surviving spouse gets a survivor pension that is designed to help her maintain the same living standards. This pension may be complemented with income from assets.
- Second, mortality rates at older ages vary with education and income level. The life expectancy of the lowest income groups is lower and on average they die first (Jusot, 2004). Hence, there is a selection effect as the proportion of high-income households increases with age.
- Third, some poor widows may move to sheltered housing or nursing homes. They are excluded from our sample.[4] Delbès and Gaymu (2005), and Angelini and Laferrère (2008) find that entry into a nursing home is more likely for low-income groups.

Actually, most widows aged between 60 and 85 years old live independently (see Figure 3). Co-residing with children is rare, and even gets less frequent among younger generations (Flipo et al., 1999). Entries into nursing homes increase only above age 85. Nearly one third of widows between 90 and 94 years old live in residential care (Delbès and Gaymu, 2005).

[*Insert Figure* 3]

Widowhood can also influence a surviving spouse's wealth because of the rules governing marriage property and inheritance. Under the French marriage law,[5] all assets acquired during marriage are held in common, that is, half of them belong to each spouse. Hence after a death, half of the couple's common property belongs to the surviving spouse, but the other half is bequeathed to the heirs. The deceased partner's property is divided between the surviving spouse and their children.

[4] More precisely, only part of retirement homes and dwellings for the elderly are categorized as ordinary homes and included in the French Housing Surveys used for Figure 2. These are mainly dwellings for non-disabled elderly.

[5] As was applied until 2001. The law is now more favorable to the surviving spouse. See Laferrère (2001) for more details on French marriage contracts, and Arrondel and Laferrère (2001) on inheritance rules.

A surviving spouse may have to share with her children the property of the dwelling in which she lived with her husband. In most cases, the transfer of ownership rights to the children does not change much for the widowed mother who can go on living in her home. But depending on the overall size of the inheritance, she may be forced out. Typically if the couple's only asset was the home, the children might put some pressure on their surviving parent to sell the home and divide the money among the heirs, if only to pay inheritance tax. Besides, an altruistic surviving mother may agree to sell the dwelling to help her liquidity constrained children. This awkward situation can be prevented if the deceased spouse has made a will which gives the surviving spouse a life interest in the home (the usufruct) as long as she lives.[6] Due to this feature of the French law, we expect that the more children a widow has, the more likely she is to move out of her home.

Some further information is useful to fully understand the idiosyncrasies of the French housing market that may be relevant to our subject. In France, households cannot borrow on the value of their property as in the US. Typically, a couple saves for a downpayment while renting an apartment, then borrows to buy a house and repays all of the mortgages over the next 10 to 20 years. Hence, most of the elderly do not have any mortgages. In 2002, only 2.3 percent of persons aged 65 and more had a mortgage.[7]

One means of extracting equity from a home in France is to sell it *in viager*. The full ownership of the dwelling is exchanged for a given amount of money and a life annuity, which can be the right to stay in the dwelling. Such life annuity sales have been made famous by Jeanne Calment who lived to 122 years old and outlived the purchaser of her home. However, life annuity sales are rare and their number has been declining over time. New equity-release products such as reverse mortgage loans are currently being proposed but they have not yet become widespread.

Homeowners do not pay an income tax on the imputed rent of their home. There is no tax on capital gains when selling a family home, but stamp duty and a compulsory notary act amount to transaction costs of around 7 percent of the property value. Annual property tax is based on the rental value of the property,[8] but persons aged 75 and more who do not pay any income tax,

[6] For a dwelling, the usufruct is the right to use it. For a financial asset or a housing investment, it is its return. Since 2001, the survivor has a life interest in the deceased spouse's property even in the absence of a will.

[7] This figure was computed from the French Housing Survey.

[8] Property tax is typically one percent of the rental value. However, the rental values were established in 1970, and are only mechanically updated, without being revised: neighborhood gentrification is, for instance, not taken

are exempted from the property tax. There is also a local tax (called *taxe d'habitation*) based on the rental value which is paid both by owner-occupiers and by tenants.[9]

Tenants can rent a dwelling in the private sector where the yearly rent growth is capped by law. Rents are freely adjusted after a change of tenant. This discourages moves as they are associated with the loss of tenure discount. Tenants can also rent a dwelling in the public sector where rents are low.[10] When a tenant is aged 70 and above, and has a low income, he/she cannot easily be expelled.[11] Overall, most features of the French housing market tend to discourage the elderly from moving.

3 The effect of widowhood on housing choices

This section reviews the main mechanisms by which the death of a spouse can affect the residential mobility of the survivor, as well as housing and location choices.[12]

3.1 Changes in housing services and income

Housing has many specific characteristics compared to other consumption goods. There are some large economies of scale in consumption as housing is a partially public good (Nelson, 1988). Besides, there can be increasing returns in the household production of goods and services. For instance, cooking for two takes less than twice the time of cooking for one.[13] Sharing a home may also yield positive complementarities as some tasks, such as gardening, may be performed better by one of the spouses than by the other. All these benefits disappear when a partner dies and occupying a large dwelling becomes less attractive. On the other hand, if the home was too

into account.

[9] Low-income households also have a tax exemption. In 2002, the median annual local tax for persons aged 60 and above was 345 euros (authors' computation from the French Housing Survey).

[10] More details can be found in Le Blanc and Laferrère (2001).

[11] Except if the landlord is him/herself aged 60 and above, or has a low income.

[12] A simple model of the trade-offs determining the choices is proposed in Bonnet, Gobillon and Laferrère (2008).

[13] Another type of scale economies mentioned by Nelson (1988) are scale economies in price, when the marginal cost of housing is a decreasing function of its quantity. Scale economies in price remain the same when one partner dies, while scale economies in consumption disappear.

small, congestion disappears. An extra room may also be useful if the survivor wants to lodge a care-giver or visitors to overcome loneliness.

Overall, we expect that for a new widow, the benefits of occupying a large dwelling are small compared to the high occupation costs, especially when the survivor's pension does not fully compensate for the loss of the partner's income. A new widow is thus likely to reduce her housing consumption. This is all the more true if she is liquidity constrained and is forced to move. If the housing choices of the couple were made in anticipation of widowhood, the size of the dwelling is closer to the optimum for the widowed partner, and moving is less likely.

A decrease in housing consumption can be achieved by moving from a house to an apartment building, or by reducing the number of rooms. The issue of downsizing of the elderly has been widely discussed in the literature. Venti and Wise (2001) show that US elderly do not reduce their housing equity except when facing a shock such as widowhood. Ermisch and Jenkins (1999) on British panel data, and Angelini and Laferrère (2008) on European panel data find that residential mobility of the elderly is low and often leads to some downsizing, especially at older ages.

3.2 Mobility costs

Because of moving costs, the moving decision and housing adjustments follow a (s, S) rule (Grossman and Laroque, 1990; Gobillon and Le Blanc, 2004). For a new widow, if the optimal housing consumption is close to her current housing consumption, it is not worth adjusting it because of moving costs. She will move only if her optimal housing consumption is far enough, for her housing adjustment to more than compensate the moving cost. The low mobility of the elderly can be explained by high non-monetary moving costs. Indeed, the elderly are usually less healthy, and have acquired over time some habits and a knowledge of their neighborhood that would be lost if they moved. We expect owners to be less mobile than tenants as their moving costs are usually higher.[14]

Moreover, there is less time to recover the sunk cost of the investment as one gets older, and more maintenance tasks are required from an owner than from a renter. For all these reasons, we expect widows to switch from owning to renting more often than the opposite, especially at older ages.

[14] It is also easier for owners than for tenants to adapt their dwelling.

The mobility decision is also likely to depend on the trends in housing prices. If prices increased, a widow may want to realize the capital gains on her house. On the contrary, if prices decreased, she may want to stay in her dwelling.

3.3 Preferences and location

The loss of a spouse may also modify the household's preferences. Indeed, husband and wife may have had different preferences which led to a compromise when choosing a dwelling. Widowhood might allow a surviving spouse with low bargaining power to follow her own preferences and choose another home.

Preferences also change because the loss of a spouse means that the survivor faces the absence of a potential caregiver. A widow may be induced to relocate closer to her children, or to move to a place where consumption amenities allow her to live independently in old age. It is usually considered that consumption amenities are offset by low wages or high rents (Roback, 1982). As the elderly no longer get their income from the labour market, they should prefer locations where amenities are offset by low wages and rents are reasonable (Graves and Knapp, 1988).[15]

Local housing markets differ in urban and rural areas. Cities provide more apartment buildings and fewer houses, and dwellings are more often for rent. Hence, a new widow owning a house in a rural area is likely to move to an apartment in the rental sector if she relocates in a city.

4 The data

To investigate the housing adjustments made after widowhood, we need information on residential and family history, as well as on the characteristics of the former and current accommodation. Few datasets provide such information. Panel data would seem well suited to studying transitions. However their sample size is small. For instance in the European Community Household Panel, only 65 males and 192 females became widowed over the 1994-2001 period (Ahn, 2004).[16] Besides,

[15] See also Chen and Rosenthal (2008) for a discussion on how locations vary in their appeal to the elderly.

[16] In the US Panel Study on Income Dynamics (1980-1997), the German Socio-Economic Panel (1984-2000), the British Household Panel (1991-2000) and the Canadian Survey of Labour and Income Dynamics (1993-2000), 571, 345, 197 and 633 females aged 50 years old and over become widowed (Burkhauser and alii, 2005).

panel attrition is likely to be endogenous as mobile households are more difficult to retrieve. For those reasons we use the 1996 and 2002 French Housing Surveys (FHS) that offer large representative samples of the population. These cross-section surveys are also designed to study residential mobility. They offer a large choice of retrospective questions on the housing situation four years before the survey date, as well as questions on whether a move occurred within the last four years, and the reasons for the move. The data include the usual socio-demographic characteristics and income. Importantly, the 2002 survey also provides the total number of children outside the parents' home, which is likely to be an important component of preferences and constraints.

We define a mobile household as one who changed homes within the last four years. We restrict the sample to households whose head is retired or inactive and was aged between 60 and 85 four years before the survey date.[17] The exclusion of those who are employed is meant to reduce the impact of labor market transitions leading to residential mobility unrelated to the loss of a spouse.[18] We exclude the oldest old because people living in institutions are not included in our data. Entries into nursing homes are not frequent before age 85 (Delbès and Gaymu, 2005, and Figure 3).

In what follows, the date of the survey (1996 or 2002) is labelled t and the date four years before is labelled $t-1$. The surveys provide no information on matrimonial status in $t-1$, but the status in t and the number of household members in $t-1$ and t are known. We consider that there is a transition to widowhood if a person is widowed and lives alone in t, and the number of household members decreased from two to one between $t-1$ and t.

This definition ignores recently widowed persons moving to live with their children. However their number is negligible and ignoring them does not induce any significant bias (See Appendix). Neither do we study widowhood when it occurs in a couple living with their children, because we cannot identify them accurately enough.[19]

Our final sample comprises 14,257 households (6,610 in the 1996 FHS and 7,637 in the 2002 FHS) among whom 1,016 individuals experience a transition to widowhood (441 in the 1996 FHS and

[17] Household head is defined as the male for couples and as the individual living alone at the survey date otherwise.

[18] It does not eliminate the effect of retirement on mobility occurring after retirement. However only 8 mobile households gave this reason for their move in our sample (see Table 6).

[19] Their number can be approximated by the number of households including a widow in t whose size decreased from $n-1$ to n between $t-1$ and t. They represent only 6% of the recent widows.

575 in the 2002 FHS).[20] Descriptive statistics are presented in Table 1.

[*Insert Table* 1]

Table 2 gives the rates of transition to widowhood among couples. They increase with age. Between 1998 and 2002, around 30 percent of couples aged 80-84 experienced the loss of a spouse. Widowhood is less frequent before 64 and is more frequent at later ages in 2002 than in 1996. As was noted for the cohort effect in Figure 1, these differences are related to the rapid increase in life expectancy that makes widowhood happen later in the life cycle.

[*Insert Table* 2]

We define six non-overlapping family situations from marital status and shocks on household composition:

(1) *Couple*: a man and a woman living together in $t-1$, whether they are legally married or not and still living together as a couple in t.[21]

(2) *Single or divorced*: a person living alone in $t-1$, and single or divorced in t.

(3) *Widow*: a person living alone in $t-1$, and widowed and living alone in t.

(4) *3 people and more*: households with more than two members in $t-1$.[22]

(5) *Recently widowed*: a person in a two-person household in $t-1$, and widowed and living alone in t.

(6) *Recently separated*: a person in a two-person household in $t-1$, and divorced and living alone in t.

Recently separated couples (6) account for less than 1 percent of the sample. We exclude them from the analysis. Table 1 gives the population composition by family situation in 2002. Couples (1) are the largest group and account for 42 percent of the sample, long-term widows (3) are the second largest group at 26 percent. The percentage of recently widowed (5) is 8 percent.

[20] In 1996, 103 males and 338 females had experienced widowhood. In 2002, the corresponding figures are 144 males and 431 females.

[21] Most people over 60 years old living together are married.

[22] This group includes some couples with children who experience the death of one of the partners. We do not distinguish them as we focus on transitions to widowhood among couples (see above).

The mobility rate over the 1998-2002 period is reported for each group in col. 3 of Table 1. Recently widowed have the highest mobility rate (13.3 percent), which is more than twice the rate of couples. Interestingly, the mobility rate of long-term widows is far smaller (7.9 percent) than that of recently widowed.

5 Multivariate analysis

5.1 Mobility

We now assess empirically the effect of being recently widowed on mobility. We estimate a probit model where the dependent variable is a dummy equal to one in the event of a move and zero otherwise. Differences in mobility between family situations are captured by four dummies, each corresponding to one of the types (2)-(5) defined in the previous section. Couples (type 1) are the reference. As children are both potential providers of care and help, and potential claimants to the inheritance of the parental home, a dummy for the existence of children living outside the parents' home is introduced.[23] Regressors also include age, sex and education of the household head. Housing tenure and housing type are introduced as proxies for mobility costs, dwelling quality, and long term suitability to needs. Municipality size is measured by the 1999 Census population which was added to our dataset using a restricted access municipality code. The population of the municipality (less than 1,000 inhabitants; 1,000-5,000; 5,000-10,000; 10,000-50,000; and more than 50,000) captures effects related to the structure of the housing market and to amenities.
Finally, the income level after the partner's death may affect mobility.[24] On the one hand, income can have a positive effect as it helps finance moving costs. On the other hand, it can have a negative effect because low-income recently widowed may be unable to pay for their housing expenditure and be forced to downsize. The overall effect on mobility is an empirical issue and the arguments

[23] The information on independent children is only provided in the 2002 FHS. Hence, we restrict the sample to this survey in this subsection. As a robustness check, we ran a regression including all the other explanatory variables on the pooled 1996 and 2002 FHS. Results remain the same.

[24] The change in income due to the partner's death is likely to influence mobility. However, we only know income at the survey date, hence we cannot compute the change in income. As a result, the dummy for being recently widowed will capture the effect of the income change on mobility.

given above suggest that it may be non linear. We first introduced income and its square in our probit models. The effect of income was found to be an inverse U-shape, with the vast majority of observed households being on the increasing part of the parabola (the maximum of the parabola being as high as 86,000 euros). Hence, the income effect is positive and nearly linear, and we stick to a linear specification (in log).[25]

The first specification tests for differences in mobility between family categories (Table 3, column 1). Single or divorced, as well as recently widowed persons, are found to be more mobile than couples. Recently widowed are the most mobile. *Ceteris paribus*, their probability of moving is nearly 90 percent higher than for couples. Note that those who have been widowed for more than four years are no more likely to move than couples. It suggests that when widowhood induces mobility, it is mostly within the four-year period after the partner's death. This result is in line with that obtained on the US Panel Study of Income Dynamics (Chevan, 1995).

Mobility decreases with age until age 80 and then increases. Education has no significant effect. This is not surprising since residential mobility related to education choices would have occurred sooner in the life-cycle. The positive effect of income on mobility is in line with the need to pay for moving costs, but not with liquidity constraints that would force a move to reduce housing costs. Those who have children are significantly more mobile than those who are childless. This is consistent with parents relocating closer to their children either to receive support (Ogg and Renaut, 2005, Glaser and Tomassini, 2000, Laditka and Laditka, 2001) or to take care of their grand-children. The effect of tenure is also in line with expectations: owners are less mobile than tenants. We also find the usual result for France that private-sector tenants are more mobile than public housing tenants (Gobillon, 2001). Indeed, public housing tenants pay lower rents than in the private sector and would loose this benefit when moving to the private sector. Living in a house has a negative effect on mobility, probably because it is usually associated with higher quality. There is also a positive effect of living in a large municipality (more than 50,000 inhabitants) on

[25] Income after widowhood might be endogenous since new mobile widows may sell a dwelling, invest in a financial asset and get some extra income. To overcome this difficulty, we instrumented income with the overall pension level. For the recently widowed, the pension includes both her own pension and the survivor's pension. It is based by law on the level of the two partners' incomes before retirement and is thus exogenous. The results were very similar (they are available in Bonnet, Gobillon and Laferrère, 2008).

mobility. Finally the number of excess rooms, defined as the number of rooms minus the number of persons living in the dwelling, has no significant effect on mobility.

To shed more light on the specific behavior and constraints of the recently widowed, we then run separate probits for three main family situations: couples, long-term widows, and recently widowed (Table 3, columns 2 to 4). Overall, estimated parameters of couples and recently widowed are quite similar. A Khi-square test shows that the two sets of parameters are not different at a 5% level. By contrast, the sets of estimated parameters of long-term widows and recently widowed are significantly different.[26]

The variations of the effect of some variables between groups are worth commenting. Interestingly, the age profile and the effect of children differ for the group of recently widowed. The mobility of recent widows does not decline with age, and increases sharply above age 80, more than for couples and long-term widows. This is consistent with housing adjustments triggered by health problems. While couples can rely on a spouse for care and stay at home, an older widow may have to move to get care. She may want to relocate closer to her children or to a place where health services and medical care are more accessible. As people moving to institutions are excluded from our study, the high residential mobility between private dwellings above age 80 is consistent with new cohorts of elderly trying to live and age independently for as long as possible. Having children increases the propensity to move only for those recently widowed (the effect is significant at 10 percent), and has no effect for couples and long-term widows. It is hard to disentangle the reasons for this positive effect: it may point to the need for family support at close range, or to some pressure by the children at the time of inheritance. Some of the moves may be due to the necessity of sharing the deceased parent's estate. The pressure is likely to be stronger for widows than for widowers because the wives of the cohorts we study might own fewer personal assets than their husbands. Consistent with this idea, we find that recent widows are significantly more likely to move than recent widowers. A more convincing test would be to interact the children

[26] When conducting a comparison test for couples and recently widowed, we dropped the sex variable from the specification for recently widowed to have the same variables for the two probit specifications. The critical value of the χ^2 (19) statistic at the 5% level is 30.14. We get a value of 18.24 which is below the threshold. When comparing the results for stable and recently widowed, we get a χ^2 (20) statistic of 42.03. This value is above the 1% threshold 37.57.

dummy with the sex dummy. Unfortunately the sample does not include enough recent childless widowers to get convincing results. Females are more affected by disabilities than males of the same age (Cambois et al., 2003). The significant positive effect of the female dummy is thus also compatible with their having or anticipating more health problems.

The number of excess rooms has a positive effect for widows but not for couples. Their mobility is more likely than that of couples to be triggered by a disequilibrium in housing consumption. It may be a sign of the financial burden of a large dwelling. The next subsection analyzes the housing choices of movers and will provide some additional evidence that widows may be income constrained.

[*Insert Table* 3]

5.2 The housing choices of mobile widows

Recent widows may move to adjust their housing consumption, and more precisely to downsize by reducing their number of rooms. We now use both the 1996 and 2002 surveys to get a large enough sample of movers. In this sample, 30 percent of moving couples increase their number of rooms and 39 percent decrease it.[27] By contrast, only 9 percent of mobile recent widows increase their number of rooms, while 74 percent decrease it. Moreover, half of those who downsize, do so by two rooms or more. To get more insight into the determinants of downsizing, we model the simultaneous decision of mobility and housing adjustments using a multinomial logit with four categories: no move, a move with no change in the number of rooms (reference), an increase, or a decrease. This model is meant to describe the change in the number of rooms when moving, all other things being equal, in the spirit of Ermisch and Jenkins (1999).[28] Results are reported in Table 4. Conditionally on moving, downsizing increases after age 75. The number of excess rooms has a positive effect on downsizing and a negative effect on upsizing. Hence, moves tend to correct for a disequilibrium in housing quantity.[29] While income has no significant effect on downsizing, more income induces upsizing. Whereas recent widows are more likely to downsize than couples

[27] All figures in this section are weighted.

[28] The set of explanatory variables does not include the number of children as it is not available in the 1996 survey.

[29] Gobillon and Wolff (2009) find the same results for retiring French households.

when moving, there is no significant difference for upsizing.

[*Insert Table* 4]

We then examine whether households choose an apartment or a house when moving. Among mobile recent widows, 36 percent lived in an apartment before the move when they were still married and this proportion doubles to 73 percent after the move. By comparison, the increase is negligible for couples: 45 percent live in an apartment before the move and 47 percent do so after the move. We then have a more careful look at the subsample of households who lived in a house in $t-1$ for which we estimate a multinomial logit of mobility and housing type with three categories: no move, a move towards a house (reference), a move towards an apartment. Results are reported in Table 5. As expected, mobile recent widows are more likely than mobile couples to switch from a house to an apartment. So are mobile long-term widows, as well as mobile single and divorced individuals. Leaving a house for an apartment significantly increases with age. This is not surprising as living in a house usually involves maintenance tasks that are taken care of more collectively in apartment buildings. With increasing age, such tasks become more difficult to perform. Also, houses in France are mostly located in the suburbs and are quite far from town centers where amenities such as stores and health services are located. Moving from a house to an apartment may grant the elderly living on their own better access to these services.

[*Insert Table* 5]

Along the same lines, we investigate the effect of being widowed on a change in housing tenure when moving. Among moving owners, we expect recent widows to switch more often to the rental sector than couples as ownership is more demanding for a single person because of maintenance tasks and paperwork. Indeed, among recent widows, 51 percent of owners switch to renting when they move. Conversely, only 18 percent of renters switch to owning. The proportions for couples are respectively 19 percent and 29 percent. We check that the differences in the switches from ownership to rental hold ceteris paribus. For the subsample of owners in $t-1$, we estimate a multinomial logit with three categories: no move, a move within the ownerhip sector (reference) and a switch towards the rental sector. The results reported in Table 6 confirm that mobile widows, whether recent or not, switch more often from owning to renting than mobile couples. This is

consistent with widows simplifying housing management and with moves toward town centers where the rental market is larger. It could also result from estate sharing following the spouse's death (see section 2 on the influence of inheritance laws). Interestingly, among recent widows who move from owning to renting, one third choose the public sector, which is quite attractive as it provides some homes adapted to the elderly.

[*Insert Table* 6]

Finally, we test whether recent widows are more likely than couples to move to larger municipalities where health and other services are more easily available. Among movers, 40 percent of recent widows move to a larger municipality whereas this proportion is only 28 percent for couples. Conversely, only 17 percent of recent widows move to a smaller municipality whereas 28 percent of couples do so. We test whether these results still hold ceteris paribus by estimating a multinomial logit with four categories: no move, moving within the current municipality (reference), moving to a larger municipality and moving to a smaller municipality (see Table 7). As expected, mobile recent widows chose more often larger municipalities than mobile couples. Interestingly, this is not the case for long-term widows, and single or divorced individuals. They may have already moved to a location more suited to living alone. We also find that being a recent widow decreases the propensity to move to a smaller municipality compared to couples, but the effect is not significant. Overall, the results are consistent with widows moving to larger municipalities where there are more services. Using a file linking each municipality with local services (the so-called 1998 Municipal Inventory), it was possible to check that a larger municipality size goes along with more stores, care and health services.[30]

[*Insert Table* 7]

Our results suggest that the loss of a spouse leads to a relocation for reasons related to preferences. Reasons for moving can also be investigated by using direct questions on the motives for a move which were asked in the 1996 survey. More than one reason could be given. The primary reason for moving given by recent widows is to live close to relatives or to her birthplace. This reason is mentioned by 25.9 percent of mobile recent widows, compared to only 15.3 percent of long-term widows and 12.1 percent of couples (see Table 8). The second reason given by recent

[30] Descriptive statistics on this topic are available upon request.

widows is downsizing: 17.5 percent of them wanted to reduce the size of their dwelling. The corresponding proportion is lower for long-term widows (12.1 percent) and small for couples (4.9 percent). The third reason given by recent widows for moving is related to their neighborhood quality and location (12.8 percent). A larger proportion of couples mention these reasons (20.6 percent). It must be noted that more than one fifth of mobile recent widows declare moving for 'another reason'. Laferrère (2005) observes that this type of answer increases with age and suggests that it could reflect health-related reasons.

[*Insert Table* 8]

If living closer to their relatives is the main reason given by recent widows for moving, we may wonder how close they get to their children. This can be investigated using the 2002 Housing Survey which asks for the distance from the independent children. Mobile recent widows usually live very close to their children at the survey date: 84.5 percent of them live less than 25 kilometers from a child (Table 9, col.1) versus 71.8 percent of recent widows who did not move . By contrast, the figures are lower for couples (at 61.1% and 69.6% respectively). This again suggests that recent widows want to live close to their children.[31] We could verify that *ceteris paribus*, mobile recent widows live closer to a child than mobile couples (Table 9, col.2). Living closer to a child is a means to get more care. Fontaine et al. (2007) stress the importance of children to a widowed parent and show how the siblings step in to take care of a widowed disabled parent.[32]

[*Insert Table* 9]

6 Simulations

We now use our results to assess the effect of the increase in widows in the next twenty years on the French housing market. We rely on two additional sources of information: the official household projection by household type conducted by the French Institute of Statistics (Jacquot, 2007) and the projections by matrimonial status derived from the DESTINIE micro-simulation model. We

[31] Note however that we cannot look at the effect of mobility on the change in distance from the closest child as the distance before the move is not available.

[32] See also Roan and Raley (1996).

only propose some rough calculations that are meant to give an order of magnitude, rather than precise predictions based on an equilibrium model of housing which is beyond the scope of this paper.

According to the official household projection, the number of households will increase by $234,000$ per year in the next twenty years. Most extra households will be elderly single-person households. Persons aged 60 and over living alone will account for 45 percent of additional households between 2006 and 2010 and for 60 percent between 2026 and 2030. According to DESTINIE, 15 percent of the additional one-person households aged 60 and over will be widows. This represents around $18,500$ additional widows per year on average over the next 20 years.

In order to turn additional widows into a potential demand for new constructions, more assumptions have to be made. First we ignore the construction for second homes and replacement, and assume that each additional household needs one additional home. Under these assumptions, additional widows represent 8 percent of new constructions. We can also assess what kind of dwellings are needed. We first approximate the breakdown of new constructions by type over the next twenty years using the information we have on the breakdown of new constructions in 2002. In 2002, 34 percent of new dwellings were apartments (66 percent were houses) and 17 percent had one or two rooms. We then assume that the residential behavior of widows observed over the 1998-2002 period will remain the same over the period up to 2030. We propose two benchmark scenarios.

In scenario (1), we assume that the flow of additional widowed one-person households has the same housing demand as the mobile recently widowed in 2002. Hence, 37 percent of them choose apartments, which translates into an extra demand of $6,800$ apartment units per year. This demand accounts for 9 percent of the additional demand of apartment units on average over the next 20 years. This figure is an upper bound. In scenario (2), we assume that the flow of additional widowed one-person households behaves like the mobile widows in 2002. 20 percent of them would choose an apartment, which translates into a demand for $3,700$ aparments units per year. This demand accounts for 5 percent of the additional demand for apartment units on average over the next 20 years. This is a lower bound.

The same kind of computations are conducted by dwelling size. Units with one or two rooms account for 17 percent of new constructions. The proportion of additional widows occupying a

small dwelling is between 28 percent (scenario 2) and 40 percent (scenario 1). They correspond to 5,100 to 7,400 additional apartment units per year, that is between 13 and 19 percent of the new construction of small units.

Note that the proportion of widows among additional households is much lower in the beginning of the period when most new elderly one-person households will be divorced or single individuals, but it will reach 35 percent of the new one-person households aged 60+ around 2030. Hence the bulk of our widowhood effect on the housing market will take place after 2020 as the baby-boomers reach the age of widowhood. Indeed at the end of the period, up to a quarter of apartments and half of the units with one or two rooms will have to be built for widows.

Such rough computations remain tentative, as the types of new housing built are likely to evolve under the pressure of additional demand. Indeed, the proportion of apartments among new constructions has increased since 2002 reaching 47 percent in 2008, and the proportion of small units has increased to 24 percent. This reduces the relative weight of widows' demand on each sub-market. On the other hand, non-widowed single-person households tend to occupy apartments and small units even more than widows. Overall, the ageing of the baby-boomers together with the death of their spouses is likely to significantly affect the housing market.

7 Conclusion

We studied the effect of widowhood on mobility, housing and location choices. Empirical tests using the French Housing Surveys show that the residential mobility of recent widows is around 90 percent higher than for couples. It is also higher than for long-term widows, suggesting that housing adjustments occur within four years after the loss of the spouse. The mobility of recent widows increases after age 80 and is more likely when they have children.

When they move, recent widows are more likely than couples to downsize, to switch from owning to renting, to exchange a house for an apartment, and to live in a larger municipality. Finally, mobile recent widows mention more often that they moved to live closer to their family and to reduce the number of rooms. In fact, they tend to live closer to their children than non-mobile recent widows and couples, even if they seldom co-reside with their children.

Overall, these results suggest that widows downsize to adjust their dwelling to the income loss

due to widowhood and to their current or anticipated need for care. Downsizing usually cuts down housing maintenance tasks. Apartments are also easier to manage than houses, and so is renting compared to owning. Living closer to a child and in a larger municipality are some means of facilitating access to care. The higher residential mobility of the oldest recent widows may point to a need for more care as their health declines and disability risk increases.

As baby-boomers get older, their residential choices after the loss of their spouses will have an impact on the housing market. Our simulations show that a significant fraction of the demand for apartments and small units will come from widows, especially after 2020. This new demand will have an effect on construction and, if not fully anticipated, on the relative prices of the various types of housing units. Residential choices of widows will also have an impact on the way long term care of the elderly is financed and delivered. Accounting for the behaviour of widows in a general equilibrium model of the housing market including institutions remains a topic for future research.

A limit to our analysis is that we could not separately identify the various channels by which the existence of children may affect the mobility and housing choices of their widowed parent. A widow may move either to get closer to care-providing children, or because she has to move out to share the deceased spouse's estate. We found many indirect hints pointing towards care by children. However, it would be interesting to measure how the rules of intergenerational transfer may trigger mobility. This is another topic for future research.

References

Angelini, V. et A. Laferrère, 2008. Home, houses and residential mobility, in Health, Ageing and Retirement in Europe (2004-2007). Starting the Longitudinal Dimension (Börsch-Supan A. et al.ed.), Mannheim Research Institute for the Economics of Aging, 99-107.

Ahn N., 2004. Economic consequences of widowhood in Europe: cross-country and gender differences, FEDEA Working paper, 27.

Arrondel L. and A. Laferrère, 2001. Taxation and Wealth transmission in France, Journal of Public Economics, 79, 3-33.

Bonnet C., Gobillon L., and A. Laferrère 2008. The housing and location choices of widows, CREST working paper, 2008-12.

Börsch-Supan A., 1990. A Dynamic Analysis of Household Dissolution and Living Arrangement Transitions by Elderly Americans, in D.A. Wise (ed). Issues in the Economics of Aging. The University of Chicago Press: Chicago.

Börsch-Supan A., Hajivassiliou V., Kotlikoff L. and J. Norris (1992). Health, Children, and Elderly Living Arrangements; A Multiperiod-Multinomial Probit Model with Unobserved Heterogeneity and Autocorrelated Error, in D.A. Wise (ed). Topics in the Economics of Aging. The University of Chicago Press.

Burkhauser R.V., Giles P., Lillard D.R. and J. Schwarze, 2005. Until Death Do Us Part: An Analysis of the Economic Well-Being of Widows in Four Countries, The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 60, 238-246.

Cambois E., Désesquelles A. and J.F. Ravaud, 2003. The gender disability gap, Population and Societies, 386, 1-4.

Chappell N.L., 1991. Living arrangements and sources of caregiving, Journal of Gerontology, 46(1), 1-87.

Chen Y. and S. Rosenthal, 2008. Local amenities and life-cycle migration: Do people move for jobs or fun?, Journal of Urban Economics, 64(3), 519-537

Chevan A., 1995. Holding on and letting go. Residential mobility during widowhood, Research on Aging, 17(3), 278-302.

COR (2008), Retraite : droits familiaux et conjugaux, 6ème Rapport du Conseil d'Orientation des Retraites.

Costa D., 1999. A House of Her Own: Old Age Assistance and the Living Arrangements of Older Nonmarried Women, Journal of Public Economics, 72(1), 39-59.

Delbès C. and J. Gaymu, 2005. Qui vit en institution?, Gérontologie et Société, 112, 13-24.

Ermisch J. F. and S. P. Jenkins, 1999. Retirement and housing adjustment in later life: evidence from the British Household Panel Survey, Labour Economics, 6, 311-333.

Flavin M. and S. Nakagawa, 2008. A model of housing in the presence of adjustment costs: A structural interpretation of habit persistence, American Economic Review, 98(1), pp. 474-495.

Flipo A., Le Blanc D. and A. Laferrère, 1999. De l'histoire individuelle à la structure des ménages, Insee Première, 649.

Fontaine R., Gramain A. and J. Wittwer, 2007. Les configurations d'aide familiales mobilisées autour des personnes âgées dépendantes en Europe, Economie et Statistique, 403-404, 77-98.

Freedman, 1996. Family structure and the risk of nursing home admission, Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 51(2), S61-S69.

Glaser K. and C. Tomassini, 2000. Proximity of older women to their children. A comparison of Britain and Italy, The Gerontologist, 40, 729-737.

Gobillon L., 2001. Emploi, logement et mobilité résidentielle, Economie et Statistique, 349-350, 77-98.

Gobillon L. and D. Le Blanc, 2004. L'impact des contraintes d'emprunt sur la mobilité résidentielle et les choix entre location et propriété, Annales d'Economie et de Statistique, 74, 15-46.

Gobillon L. and D. Le Blanc, 2008. Economic effects of upfront subsidies to ownership: The case of the Prêt à Taux Zéro in France. Journal of Housing Economics, 17(1), 1-33.

Gobillon L. and F.C. Wolff, 2009. Housing and Location Choices of Retiring Households: Evidence from France, Working paper.

Graves P. and T. Knapp, 1988. Mobility behavior of the elderly, Journal of Urban Economics, 24(1), 1-8.

Grossman S. and G. Laroque, 1990. Asset pricing and optimal portfolio choice in the presence of illiquid durable consumption goods, Econometrica, 58(1), 25-52.

Heiss F., Hurd M. and A. Börsch-Supan, 2003. Healthy, wealthy and knowing where to live: predicted trajectories of health, wealth and living arrangements among the oldest old, NBER Working Paper, 9897.

Iacovou M., 2000. The living arrangements of Elderly Europeans, ISER Working Paper, 2000-09.

Ioannides Y.M. and K. Kan, 1996. Structural Estimation of Residential Mobility and Housing Tenure Choice, Journal of Regional Science, 36(3), 335-363.

Jacquot A., 2007. Projections de ménages pour la France métropolitaine à l'horizon 2030, Insee Résultats, 60. http://www.insee.fr/fr/ppp/ir/accueil.asp?page=projmen2030/dd/projmen2030.htm

Jusot F., 2004. Mortalité et inégalités de revenu en France, Working paper, 2004-32, DELTA.

Kalogirou S. and M. Murphy, 2006. Marital Status of people aged 75 and over in nine EU countries in the period 2000-2030, European Journal of Ageing, 3(2), 74-81.

Kean M.P., 1992. A Note on Identification in the Multinomial Probit Model, Journal of Business & Economic Statistics, 10(2), pp. 193-200

Laditka J.N. and S.B. Laditka, 2001. Adult Children Helping Older Parents, Research on Aging, 23(4), 429-456.

Laferrère A., 2001. Marriage Settlement, Scandinavian Journal of Economics, 103, 485-504.

Laferrère A., 2005. Old age and housing: dissaving, adjusting consumption, and the role of children, Working Paper.

Laferrère A., 2006. Vieillesse et logement: désépargne, adaptation de la consommation et rôle des enfants, Retraite et Société, 47, 66-108.

Le Blanc D. and A. Laferrère, 2001. The Effects of Public Social Housing on Households' consumption in France, Journal of Housing Economics, 10, 429-455.

Macunovich, D., Easterlin R., Schaeffer C., and E. Crimmins, 1995. Echoes of the Baby Boom and Bust : Recent and Prospective Changes in Living Alone among Elderly widows in the United States, Demography, 32(1), 17-28.

Nelson J., 1988. Household Economics of Scale in Consumption: Theory and Evidence, Econometrica, 56(6), 1301-1314.

Ogg J. and S. Renaut, 2005. Le soutien familial intergénérationnel dans l'Europe élargie, Retraite et Société, 46, 30-57.

Roan C.L. and R.K. Raley, 1996. Intergenerational Coresidence and Contact: A Longitudinal Analysis of Adult Children's Response to their Mother's Widowhood, Journal of Marriage and the Family, 58 (3), 708-717.

Roback J., 1982. Wages, Rents, and the Quality of Life, The Journal of Political Economy, 90(6),

1257-1278.

Tatsiramos K., 2006. Residential Mobility and Housing Adjustment of Older Households in Europe. IZA Working Paper, 2435.

Venti S.F. and D.A. Wise, 1987. Aging, Moving, and Housing Wealth, NBER Working Paper, 2324.

Venti S.F. and D.A. Wise, 2001. Aging and Housing Equity: Another Look, NBER Working Paper, 8608.

Zorn P.M., 1989. Mobility-Tenure Decisions and Financial Credit: Do Mortgage Qualification Requirements Constrain Homeownership? AREUA Journal, 17(1), 1-16.

Appendix: Widows do not move to live with their children

Moving to coreside with a child could be a way for a widow to adjust her housing consumption (Börsch-Supan, 1990). We ignore such moves in this paper, arguing that they are very rare. We can identify whether a household is likely to include a widowed mother who moved in by using three criteria. Firstly, the household must include a $60-84$ year-old widow who is not the reference person. Secondly, the household size must have increased by one in the four-year period before the survey date. Thirdly, this increase must not be due to obvious demographic reasons unrelated to the arrival of a widow, such as a birth or the household formation.

In our 2002 data, 258 households include a widow aged between 60 and 84 years old who is not the reference person (first criterion). Among them, only 33 households had increased their size by one (second criterion). Finally, only 14 of them are likely to have experienced the arrival of a widow (third criterion), and hence meet the three criteria. Only very few widows move in with their children after their spouse's death.

Table 1 - Descriptive statistics

Variables	Sample Size	Number of movers	Mobility rate (1998-2002) in percent
Age group in t-1			
60-64 years old	1776	170	9.4
65-69 years old	2112	170	7.8
70-74 years old	1889	134	6.8
75-79 years old	1302	83	6.0
80-84 years old	498	47	10.3
Sex			
Male	4636	312	6.6
Female	2941	292	9.6
Education			
Primary school	4953	400	7.8
Secondary School, Technical, High School	2028	153	7.5
=2 years at University	131	13	9.4
>2 years at University	465	38	7.7
Children outside household			
No	1239	77	5.9
Yes	6338	527	8.1
Housing Tenure in t-1			
Homeowner	5552	279	4.8
Private Renter	815	194	22.7
Public Renter	804	93	11.0
Rent free	406	38	9,4
Population in municipality in t-1[(1)]			
Less than 1,000	1485	73	5.2
1,000 – 5,000	1962	138	6.7
5,000 – 10,000	780	49	6.0
10,000 – 50,000	1800	164	8.9
More than 50,000	1416	175	11.5
Income (Quartiles) in t			
Q1	1859	162	8.4
Q2	1930	148	7.3
Q3	1906	154	8.1
Q4	1882	140	7.1
Housing type in t-1			
House	5320	314	5.6
Apartment	2257	290	12.3
Family Type in t-1 and t			
Couple in t-1 and t	3224	203	6.1
Single or divorced in t-1 and t	774	91	11.5
Widow in t-1 and t	1965	163	7.9
Three people or more in t-1	1039	71	6.3
Couple in t-1 widowed in t	575	76	13.3
Number of observations	7577	604	7.8

Source: Authors' computation from the 2002 Housing Survey, INSEE.
Note: Sample of households where head is retired or inactive and aged 60-84 in 1998, excluding recently separated (60 observations). Mobility rates are weighted.
(1) The sample size used for Population in municipality is smaller (7443 observations), due to missing values.

Table 2: Transitions to widowhood by age group

Age group	1992-1996		1998-2002	
	Rate	Number of observations	Rate	Number of observations
60 – 64	11.5%	98	8.3%	81
65 – 69	11.7%	115	14.2%	157
70 – 74	15.0%	112	17.4%	164
75 – 79	18.4%	63	20.6%	123
80 – 84	23.3%	53	29.8%	50
All	14.4 %	441	15.6 %	575

Source: Authors' computation from the 1996 and 2002 Housing Surveys, INSEE.

Note: The rate of transitions to widowhood is defined for the sample of couples (with head aged 60-84 and retired or inactive four years before the survey date), as the ratio of the number of couples experiencing a transition to widowhood to the total number of couples. This rate is weighted.

Table 3: Probability of moving between 1998 and 2002 (probit)

	Whole sample	Couple in t-1 and t	Widow in t-1 and t	Couple in t-1 widowed in t
Variables	(1)	(2)	(3)	(4)
Constant	-3.194***	-3.391***	-3.738***	-4.319***
	(0.454)	(0.786)	(0.841)	(1.535)
Age group in t-1				
60-64 years old	ref.	ref.	ref.	ref.
65-69 years old	-0.142**	-0.182*	-0.132	0.290
	(0.061)	(0.096)	(0.144)	(0.248)
70-74 years old	-0.234***	-0.212**	-0.311**	-0.034
	(0.065)	(0.100)	(0.144)	(0.261)
75-79 years old	-0.326***	-0.266**	-0.398***	0.088
	(0.075)	(0.124)	(0.150)	(0.270)
80-84 years old	-0.124	-0.103	-0.187	0.577*
	(0.097)	(0.201)	(0.166)	(0.307)
Sex				
Male	ref.		ref.	ref.
Female	0.085		0.089	0.399**
	(0.071)		(0.132)	(0.196)
Education				
Primary school	ref.	ref.	ref.	ref.
Secondary School, Technical, High School	-0.086	-0.158*	-0.057	-0.374*
	(0.056)	(0.090)	(0.111)	(0.198)
=2 years at University	-0.023	-0.011	-0.101	0.556
	(0.162)	(0.254)	(0.369)	(0.436)
>2 years at University		-0.175		
	-0.125	(0.155)	-0.257	-0.889
	(0.105)		(0.281)	(0.627)
Children outside household				
No	ref.	ref.	ref.	ref.
Yes	0.242***	0.051	0.207	0.485*
	(0.070)	(0.134)	(0.142)	(0.273)
Housing Tenure in t-1				
Homeowner	ref.	ref.	ref.	ref.
Private Renter	0.910***	1.057***	0.745***	1.125***
	(0.064)	(0.111)	(0.121)	(0.206)
Public Renter	0.338***	0.413***	0.102	0.482*
	(0.079)	(0.148)	(0.150)	(0.276)
Rent free	0.380***	0.296	0.244	0.655**
	(0.096)	(0.190)	(0.160)	(0.284)
Population in municipality in t-1				
Less than 1,000	ref.	ref.	ref.	ref.
1,000 – 5,000	0.118	0.190*	0.035	0.174
	(0.074)	(0.114)	(0.155)	(0.241)
5,000 – 10,000	0.020	-0.037	0.319*	-0.138
	(0.095)	(0.160)	(0.174)	(0.333)
10,000 – 50,000	0.137*	0.100	0.201	0.252
	(0.078)	(0.127)	(0.155)	(0.264)
More than 50,000	0.242***	0.214	0.205	0.296
	(0.084)	(0.139)	(0.168)	(0.270)
Housing type in t-1				
House	ref.	ref.	ref.	ref.
Apartment	0.136**	0.097	0.219*	-0.011
	(0.064)	(0.109)	(0.120)	(0.204)
Log-income in t	0.125***	0.168**	0.188**	0.175
	(0.045)	(0.079)	(0.085)	(0.155)
Number of Excess Rooms in t-1	0.027	0.005	0.055*	0.116**
	(0.017)	(0.029)	(0.033)	(0.057)
Family Type in t-1 and t				
Couple in t-1 and t	ref.			
Single or divorced in t-1 and t	0.204**			
	(0.094)			
Widow in t-1 and t	0.032			
	(0.086)			
Three people and more in t-1	-0.030			
	(0.078)			
Couple in t-1 widowed in t	0.385***			
	(0.096)			
Number of observations	7440	3172	1924	569

Source: Authors' computation from the 2002 Housing Survey, INSEE.
Note: Sample of households whose head is retired or inactive and aged 60-84 in 1998.
***: significant at 1%; **: significant at 5%; *: significant at 10%.

Table 4: Change in the number of rooms, multinomial logit
(reference: *Moving, no change*)

Category	No move	Downsizing	Upsizing
Constant	3.158***	-0.845	-1.364
	(0.681)	(0.825)	(1.001)
Age group in t-1			
60-64 years old	ref.	ref.	ref.
65-69 years old	0.096	-0.216	-0.044
	(0.157)	(0.197)	(0.220)
70-74 years old	0.158	-0.022	-0.595**
	(0.166)	(0.204)	(0.255)
75-79 years old	0.837***	0.568**	-0.109
	(0.229)	(0.268)	(0.329)
80-84 years old	0.692**	0.742**	-0.037
	(0.280)	(0.321)	(0.390)
Sex			
Male	ref.	ref.	ref.
Female	-0.085	0.056	-0.050
	(0.188)	(0.223)	(0.278)
Housing Tenure in t-1			
Homeowner	ref.	ref.	ref.
Private or Public Renter	-1.133***	0.193	-0.287
	(0.145)	(0.181)	(0.210)
Population in municipality in t-1			
Less than 1,000	ref.	ref.	ref.
1,000 – 5,000	-0.334	0.011	-0.434
	(0.249)	(0.288)	(0.356)
5,000 – 10,000	-0.300	0.031	-0.426
	(0.295)	(0.346)	(0.425)
10,000 – 50,000	-0.330	0.139	-0.392
	(0.254)	(0.295)	(0.361)
More than 50,000	-0.580**	-0.081	-0.529
	(0.261)	(0.308)	(0.372)
Housing type in t-1			
House	ref.	ref.	ref.
Apartment	-0.153	-0.271	0.203
	(0.172)	(0.210)	(0.253)
Log-income in t	0.030	-0.046	0.203**
	(0.060)	(0.072)	(0.088)
Number of Excess Rooms in t-1	0.465***	0.757***	-0.258***
	(0.056)	(0.062)	(0.083)
Family Type in t-1 and t			
Couple in t-1 and t	ref.	ref.	ref.
Single or divorced in t-1 and t	-0.042	0.006	-0.010
	(0.219)	(0.285)	(0.315)
Widow in t-1 and t	0.219	0.357	0.082
	(0.233)	(0.281)	(0.343)
Three people and more in t-1	0.616***	1.134***	-0.693**
	(0.208)	(0.253)	(0.316)
Couple in t-1 widowed in t	-0.103	1.322***	-0.172
	(0.271)	(0.312)	(0.406)
Number of exits	12879	558	243
Number of observations	13978	13978	13978

Source: Authors' computations from the 1996 and 2002 Housing Survey, INSEE.
Note: Sample of mobile households whose head is retired or inactive and aged 60-84 in t-1. Number of individuals moving with no change in the number of rooms: 298.
***: significant at 1%; **: significant at 5%; *: significant at 10%.

Table 5: Change in municipality size, multinomial logit
(reference: *Moving, no change*)

Category	No move	Smaller municipality size	Larger municipality size
Constant	3.317***	-3.938***	-1.209
	(0.515)	(1.026)	(0.823)
Age group in t-1			
60-64 years old	ref.	ref.	ref.
65-69 years old	0.087	-0.084	-0.335*
	(0.120)	(0.206)	(0.199)
70-74 years old	0.251*	-0.047	-0.021
	(0.129)	(0.223)	(0.205)
75-79 years old	0.694***	0.357	0.330
	(0.166)	(0.268)	(0.253)
80-84 years old	0.082	-0.147	-0.430
	(0.175)	(0.312)	(0.307)
Sex			
Male	ref.	ref.	ref.
Female	-0.100	-0.025	-0.143
	(0.132)	(0.244)	(0.216)
Housing Tenure in t-1			
Homeowner	ref.	ref.	ref.
Private or Public Renter	-1.586***	-0.972***	-0.714***
	(0.111)	(0.187)	(0.192)
Population in municipality in t-1			
Less than 1,000	ref.	ref.	ref.
1,000 – 5,000	-0.637***	1.302**	-0.778***
	(0.209)	(0.640)	(0.253)
5,000 – 10,000	-0.723***	1.654***	-1.257***
	(0.239)	(0.664)	(0.315)
10,000 – 50,000	-1.060***	1.669***	-2.310***
	(0.205)	(0.627)	(0.284)
More than 50,000	-1.302***	1.841***	-4.025***
	(0.211)	(0.631)	(0.402)
Housing type in t-1			
House	ref.	ref.	ref.
Apartment	0.181	0.580***	0.601***
	(0.125)	(0.211)	(0.217)
Log-income in t	0.106**	0.171**	0.212***
	(0.044)	(0.076)	(0.073)
Number of Excess Rooms in t-1	0.056	0.081	0.073
	(0.037)	(0.062)	(0.055)
Family Type in t-1 and t			
Couple in t-1 and t	ref.	ref.	ref.
Single or divorced in t-1 and t	-0.199	-0.487	-0.034
	(0.170)	(0.306)	(0.294)
Widow in t-1 and t	-0.013	-0.385	0.105
	(0.168)	(0.298)	(0.274)
Three people and more in t-1	-0.161	-0.476*	-0.243
	(0.151)	(0.270)	(0.249)
Couple in t-1 widowed in t	-0.707***	-0.321	0.739***
	(0.187)	(0.323)	(0.277)
Number of exits	12879	249	298
Number of observations	13978	13978	13978

Source: Authors' computations from the 1996 and 2002 Housing Survey, INSEE.
Note: Sample of mobile households whose head is retired or inactive and aged 60-84 in t-1. Number of individuals moving with no change in the municipality size: 552.
***: significant at 1%; **: significant at 5%; *: significant at 10%.

Table 6: Switches from house to apartment, multinomial logit,
subsample: households living in a house in t-1 (reference: *Moving, house in t*)

Category	No move	Apartment in t
Constant	3.719***	-1.556
	(0.665)	(0.947)
Age group in t-1		
60-64 years old	ref.	ref.
65-69 years old	0.172	0.021
	(0.150)	(0.233)
70-74 years old	0.442**	0.643***
	(0.170)	(0.241)
75-79 years old	0.949***	0.896***
	(0.249)	(0.324)
80-84 years old	0.333	0.485
	(0.262)	(0.355)
Sex		
Male	ref.	ref.
Female	-0.105	0.031
	(0.202)	(0.265)
Housing Tenure in t-1		
Homeowner	ref.	ref.
Private or Public Renter	-1.587***	-0.011
	(0.149)	(0.204)
Population in municipality in t-1		
Less than 1,000	ref.	ref.
1,000 – 5,000	-0.273*	-0.044
	(0.160)	(0.248)
5,000 – 10,000	0.157	0.763**
	(0.238)	(0.327)
10,000 – 50,000	-0.038	0.675**
	(0.189)	(0.268)
More than 50,000	-0.286	0.740**
	(0.228)	(0.312)
Log-income in t	-0.019	0.018
	(0.060)	(0.085)
Number of Excess Rooms in t-1	-0.020	0.061
	(0.041)	(0.057)
Family Type in t-1 and t		
Couple in t-1 and t	ref.	ref.
Single or divorced in t-1 and t	0.581*	1.147***
	(0.308)	(0.400)
Widow in t-1 and t	0.473	0.880***
	(0.246)	(0.331)
Three people and more in t-1	0.101	0.218
	(0.182)	(0.287)
Couple in t-1 widowed in t	-0.648***	1.047***
	(0.246)	(0.325)
Number of exits	9683	9683
Number of observations	9120	271

Source: Authors' computations from the 1996 and 2002 Housing Survey, INSEE.
Note: Sample of mobile households whose head is retired or inactive and aged 60-84 in t-1. Number of individuals moving with house in t: 292.
***: significant at 1%; **: significant at 5%; *: significant at 10%.

Table 7: Switches from ownership to rental, multinomial logit,
subsample: households owning in t-1 (reference: *Moving, owning in t*)

Category	No move	Renting in t
Constant	3.570***	-1.804*
	(0.585)	(0.975)
Age group in t-1		
60-64 years old	ref.	ref.
65-69 years old	0.360***	0.597**
	(0.136)	(0.266)
70-74 years old	0.491***	1.056***
	(0.150)	(0.269)
75-79 years old	0.958***	1.236***
	(0.210)	(0.332)
80-84 years old	0.476**	1.318***
	(0.233)	(0.349)
Sex		
Male	ref.	ref.
Female	-0.080	0.315
	(0.187)	(0.279)
Population in municipality in t-1		
Less than 1,000	ref.	ref.
1,000 – 5,000	-0.171	0.342
	(0.173)	(0.285)
5,000 – 10,000	-0.146	0.516
	(0.227)	(0.356)
10,000 – 50,000	-0.470***	0.041
	(0.175)	(0.298)
More than 50,000	-0.538***	0.132
	(0.195)	(0.327)
Housing type in t-1		
House	ref.	ref.
Apartment	-0.577***	-0.597**
	(0.146)	(0.245)
Log-income in t	-0.004	0.008
	(0.053)	(0.086)
Number of Excess Rooms in t-1	-0.077**	-0.154**
	(0.036)	(0.062)
Family Type in t-1 and t		
Couple in t-1 and t	ref.	ref.
Single or divorced in t-1 and t	0.424*	1.115***
	(0.249)	(0.397)
Widow in t-1 and t	0.547**	1.233***
	(0.227)	(0.352)
Three people and more in t-1	0.137	0.471
	(0.172)	(0.324)
Couple in t-1 widowed in t	-0.594***	1.038***
	(0.223)	(0.347)
Number of exits	10312	207
Number of observations	10883	10883

Source: Authors' computations from the 1996 and 2002 Housing Survey, INSEE.
Note: Sample of mobile households whose head is retired or inactive and aged 60-84 in t-1. Education dummies are included as controls, as in Table 3. Number of individuals moving and owning in t: 364.
***: significant at 1%; **: significant at 5%; *: significant at 10%.

Table 8: Reasons for moving, by family type

	Couple in t-1 widowed in t		Couple in t-1 and t		Widow in t-1 and t	
Type of reason						
Retirement	-		3.9		0.6	
Personal or family reasons[1]	**27.2**		13.1		16.5	
including: move closer to family or friends, return to birthplace		25.9		12.1		15.3
Environment or location[2]	12.8		**20.6**		16.0	
Dwelling size or comfort[3]	18.9		19.9		**27.2**	
including: poor dwelling quality		0.9		7.9		11.1
wanted a smaller dwelling		**17.5**		4.9		12.1
Type of dwelling[4]	7.1		6.7		4.7	
Housing tenure[5]	6.9		6.8		7.4	
Income constraint[6]	1.0		1.2		1.8	
Obligation to move[7]	3.5		7.0		6.4	
Other reason	23.6		20.8		20.4	
Number of observations	78		168		117	

Source: Authors' computation from the 1996 Housing Survey, INSEE.
Note: Sample of mobile households whose head is retired or inactive and aged 60-84 in 1992.

[1] Separated from partner, moved closer to family or friends, went back to birthplace, looked for a better climate (this item cannot be separated from the preceding reason).

[2] Unattractive or insecure neighborhood, unpleasant neighbors (too noisy, antisocial behavior), too far from town centre and community facilities, wanted to get closer to town centre, wanted to live in a less urbanized place.

[3] Wanted a larger/smaller dwelling, the dwelling quality was poor

[4] Wanted to live in a house/in an apartment.

[5] Wanted to become owner/tenant, found accommodation that could be used for free

[6] Wanted to reduce housing expenses (rent, utilities, maintenance cost)

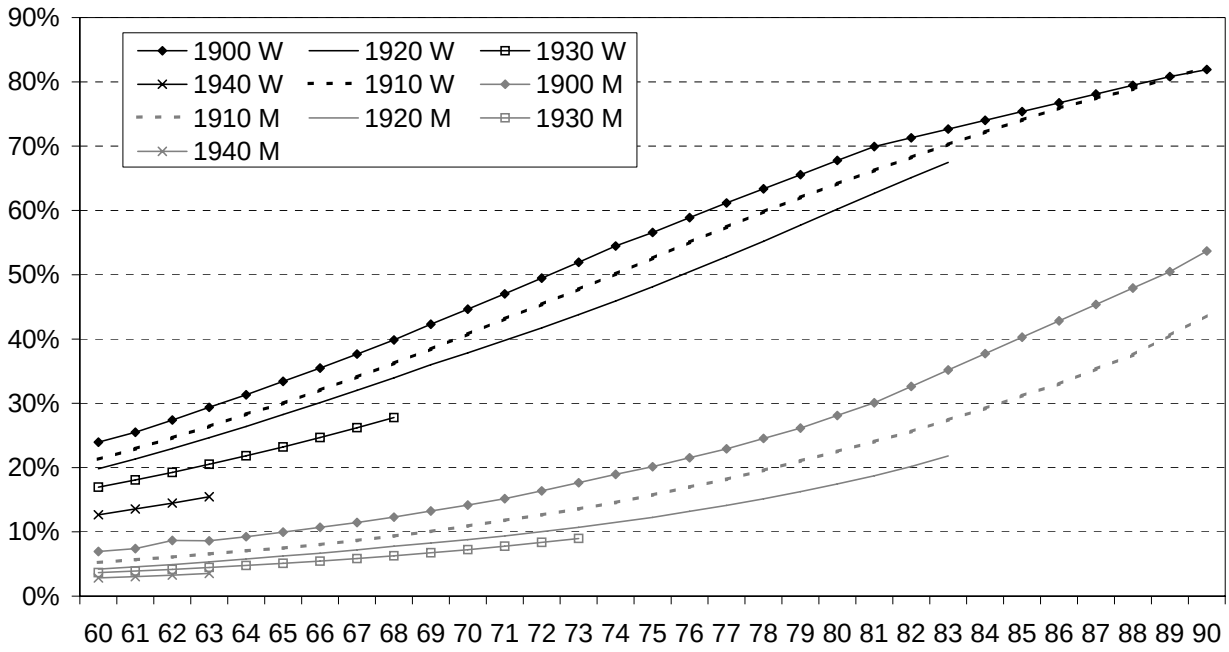
[7] Lived temporarily in the dwelling, was expelled by the owner

Table 9: Living less than 25 kilometres from closest independent child, by family type, for mobile and non mobile households

Family Type in t-1 and t		% living <25 km	Estimated parameter
		(1)	(2)
Couple in t-1 and t	Mobile	61.1	ref.
	Non mobile	69.6	0.476***
			(0.163)
Widow in t-1 and t	Mobile		-0.160
		76,9	(0.336)
	Non mobile		-0.278
		73,7	(0.214)
Single or divorced in t-1 and t	Mobile		0.334
		62,4	(0.270)
	Non mobile		0.205
		64,4	(0.189)
3 people or more in t-1	Mobile		0.706*
		82,1	(0.365)
	Non mobile		0.649***
		74,4	(0.180)
Couple in t-1 widowed in t	Mobile	84.5	0.827**
			(0.377)
	Non mobile	71.8	0.166
			(0.206)
Number of observations			6225

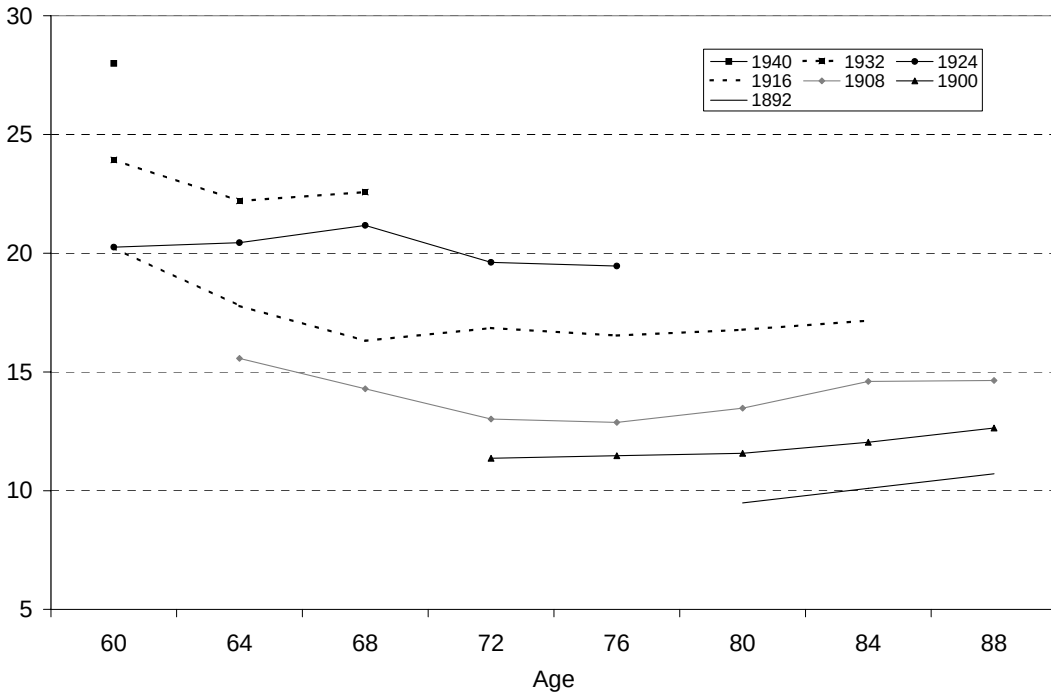
Source: Authors' computation from the 2002 Housing Survey, INSEE.
Note: Sample of households whose head is retired or inactive and aged 60-84 in 1998 with at least one child who lives independently, and no child at home. We estimate a logit model of having a child living less than 25 kilometres from the household (col. 2). Controls are age groups in t-1, sex, education level, housing type and tenure in t-1, population in municipality in t-1 and log income in t.

Figure 1 – Percentage of widows (W) and widowers (M) by age for five birth cohorts



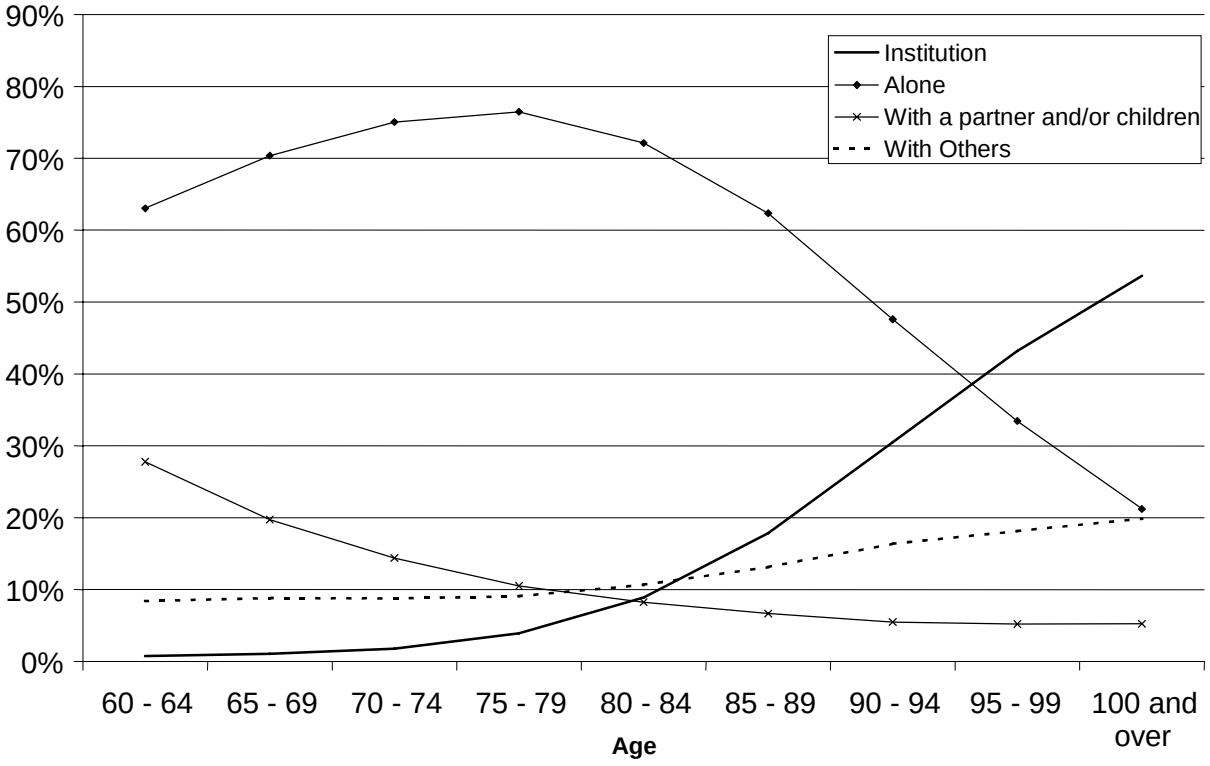
Source: French register of civil status (*Etat Civil*).

Figure 2 – Average household income by age and birth cohort



Source: Laferrère (2005), computation from the 1973, 1978, 1984, 1988, 1992, 1996 and 2002 Housing Surveys, INSEE.
Note: income is expressed in thousands of 2001 euros. Cohorts are four-year cohorts. For instance, the 1940 cohort includes all the households born during the 1937-1940 period.

Figure 3 – Living arrangements of widows by age, in 1999



Source: constructed from the 1999 French Census.
Note: the sample excludes widowers.
"With others" describes households with at least two members, other than a child or a partner.