Separate Appendices with Supplementary Material for:

The Costs of Agglomeration: House and Land Prices in French Cities

Pierre-Philippe Combes†
University of Lyon and Sciences Po

Gilles Duranton‡
University of Pennsylvania

Laurent Gobillon§
Paris School of Economics, CNRS

September 2018

ABSTRACT: This document contains a set of appendices with supplementary material.

Key words: urban costs, house prices, land prices, land use, agglomeration

JEL classification: R14, R21, R31

†University of Lyon, CNRS, GATE-LSE UMR 5824, 93 Chemin des Mouilles, 69131 Ecully, France and Sciences Po, Economics Department, 28, Rue des Saints-Pères, 75007 Paris, France (e-mail: ppcombes@gmail.com; website: https://www.gate.cnrs.fr/ppcombes/). Also affiliated with the Centre for Economic Policy Research.

‡Wharton School, University of Pennsylvania, 3620 Locust Walk, Philadelphia, PA 19104, USA (e-mail: duranton@wharton.upenn.edu; website: https://real-estate.wharton.upenn.edu/profile/21470/). Also affiliated with the Centre for Economic Policy Research and the National Bureau of Economic Research.

§Paris School of Economics, CNRS, 48 Boulevard Jourdan, 75014 Paris, France (e-mail: laurent.gobillon@psemail.eu; website: http://laurent.gobillon.free.fr/). Also affiliated with the Centre for Economic Policy Research and the Institute for the Study of Labor (IZA).

Introduction

This document complements "The Costs of Agglomeration: House and Land Prices in French Cities" by the same authors. It contains extensions and robustness checks not included in the main paper.

- Appendix A reports additional first-step results for all dwellings in the estimation of housing price at the centre of French urban areas.
- Appendix B provides evidence regarding the effect of urban area population on the distance gradients. It provides further support to our result that house prices at the centre increase with city population.
- Appendix C reports additional second-step results for the estimation of the population elasticity of the price of houses and land parcels. This appendix focuses on the possible sorting of residents across cities and within cities.
- Appendix D also reports further second-step results for the estimation of the population elasticity of the price of houses. This appendix replicates our main OLS results for all dwellings instead of only houses.
- Appendix E reports again further second-step results for the estimation of the population elasticity of the price of houses and land parcels. This appendix replicates our preferred OLS specification for alternatives samples of observations, definitions of urban centres, functional forms for distances within cities in the first step, and for alternative estimation techniques.
- Appendix F provides further econometric details about the FGLS and WLS estimation techniques used in Appendix E.
- Appendix G focuses on the estimation of possible non-constant elasticities of house and land prices with respect to urban area population.
- Appendix H reports second-step results for the estimation of the population elasticity from specifications that do not include land area.
- Appendix I reports IV results for our 2000-2012 difference estimations of the population elasticity of house prices.

Appendix Table 1: Summary statistics from the first step estimation regressions for all dwellings, 277 urban areas

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Municipality Controls									
House/Parcel charac.	Y			Y	Y	Y	Y	Y	Y
Geography and geology					Y				Y
Income, education						Y			Y
Urbanisation							Y		Y
Consumption amenities								Y	Y
All dwellings, price per m^2									
Urban area effect									
1st quartile		-.159	-.168	-.183	-.183	-.154	-.186	-.176	-.153
2nd quartile		.129	.151	.145	.143	.132	.151	.152	.128
log distance effect									
1st quartile			-.018	-.0228	-.0232	-.0104	-.0046	-.0032	-.0054
Median			-.018	-.0228	-.0232	-.0104	-.0046	-.0032	-.0054
2nd quartile			.0339	.0247	.0263	.0227	.0462	.0436	.0298
Observations	75,194	75,194	75,194	75,194	75,194	75,194	75,194	75,194	75,194
Within-time R^2	0.28	0.68	0.84	0.84	0.85	0.92	0.85	0.85	0.92

Notes: Same as for table 3 of the main text.

- Appendix J provides additional results regarding the estimation of the housing shares.
- Appendix K provides more complete results for the urban cost elasticity.

Appendix A. First-step results for all dwellings

Appendix table 1 duplicates the summary statistics of the first-step results reported panel A of table 3 in the main text for all dwellings instead of only houses. The fixed effects estimated in the regressions of appendix table 1 for all dwellings are strongly correlated with the fixed effects estimated in the corresponding regressions of table 3. For our preferred estimation in column 9, the correlation between the two tables is 0.92. Interestingly, we observe a slightly smaller dispersion of the fixed effects estimated in Appendix table 1 relative to table 3 of the main text. The estimated gradients estimated in appendix table 1 are also slightly smaller in absolute value relative to table 3 of the main text. This feature is consistent with the lower land intensity of apartments, which represent a large share of all dwellings in French urban areas.

Appendix B. Gradient analysis

In standard models of urban structure where land prices at the city fringe are identical for all cities, the higher prices of houses and land parcels at the centre of cities with greater population can be due to a greater distance to the urban fringe and/or to steeper gradients. The illustrative panels of figure 2 in the main text appear to support both explanations. To take a single example, it is easy to see that the higher intercept for house prices in Paris relative to Toulouse results from both a greater distance between the centre and the urban fringe and a steeper gradient for Paris.[1] In this appendix, we provide more systematic evidence that higher prices at the centre of urban areas with greater population can, at least in part, be accounted for by steeper distance gradients.

We implement the same two-step approach as in our estimation of the population elasticity of house prices except that our second-step dependent variable estimated in the first step is now the distance gradient instead of the urban area fixed effect. Results are reported in appendix table 2 which mirrors table 4 in the main text for this different dependent variable. A minor difference is that columns 1-3 of appendix table 2 use the output of column 3 of table 3 in the main text instead of column 2 since we need to use a first-step specification which estimates a distance gradient (unlike column 2 of table 3 of the main text).

The estimated coefficient on population is insignificant for both house and land prices in columns 1-3. For all subsequent columns, this coefficient is negative and significant for house prices. If we compare an urban area at the first quartile of population with an urban area at the third quartile, the difference in log population is 1.56. In, say, column 5 of appendix table 2, the coefficient of -0.015 then predicts a difference in distance gradient of 0.027 between the two quartiles. This corresponds to slightly more than a quarter of the interquartile range for the gradients in the corresponding first-step specification. For column 9 of appendix table 2, the population coefficient of -0.021 explains more than half the interquartile range of the distance gradients of the corresponding first-step estimation in column 9 of table 3 in the main text. The results for land prices are slightly weaker because of larger standard errors for the estimated coefficients.

Possible explanation for the steeper distance gradient of more populated urban areas include higher construction costs to build higher in larger cities and greater commuting costs per unit of distance, perhaps as a result of more congestion.

[1] For both cities, the price of houses at the urban fringe is somewhat similar.

Appendix Table 2: The determinants of the distance prices gradients for houses land parcels, OLS regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
First-step	Only fixed effects			Basic controls			Full set of controls		
Controls	N	Y	Ext.	N	Y	Ext.	N	Y	Ext.
Panel A. Houses									
Log population	-0.00956	-0.00697	-0.00812	-0.0151[b]	-0.0150[b]	-0.0170[b]	-0.0172[a]	-0.0184[a]	-0.0207[a]
	(0.00720)	(0.00771)	(0.00950)	(0.00594)	(0.00631)	(0.00790)	(0.00543)	(0.00575)	(0.00701)
Log land area	-0.0270[a]	-0.0223[a]	-0.0163[c]	-0.00739	-0.00382	0.00221	-0.00521	-0.00140	0.00522
	(0.00827)	(0.00831)	(0.00942)	(0.00681)	(0.00679)	(0.00783)	(0.00623)	(0.00619)	(0.00695)
R^2	0.17	0.23	0.30	0.12	0.19	0.23	0.14	0.21	0.30
Observations	277	277	277	277	277	277	277	277	277
Panel B. Land parcels									
Log population	0.00797	0.00747	-0.00611	-0.0128	-0.0151	-0.0265[b]	-0.0148[c]	-0.0192[b]	-0.0332[a]
	(0.0164)	(0.0175)	(0.0218)	(0.00881)	(0.00921)	(0.0115)	(0.00853)	(0.00901)	(0.0111)
Log land area	-0.0853[a]	-0.0772[a]	-0.0660[a]	-0.0292[a]	-0.0259[a]	-0.0147	-0.0197[b]	-0.0161	-0.00400
	(0.0188)	(0.0190)	(0.0217)	(0.0101)	(0.00997)	(0.0114)	(0.00980)	(0.00976)	(0.0110)
R^2	0.16	0.19	0.27	0.16	0.23	0.30	0.12	0.18	0.28
Observations	277	277	277	277	277	277	277	277	277

Notes: The dependent variable is the distance coefficient specific to the urban area estimated in the first step. Columns 1 to 3 use the output of column 3 of table 3 in the main text. Columns 4 to 6 use the output of column 4 of table 3 in the main text. Columns 7 to 9 use the output of column 9 of table 3 in the main text. All regressions include year effects. All reported R^2 are within-time. The superscripts a, b, and c indicate significance at 1%, 5%, and 10% respectively. Standard errors clustered at the urban area level are between brackets. For second-step controls, N, Y, and Ext. stand for no further explanatory variables beyond population, land area, and year effects, a set of explanatory variables, and a full set, respectively. Second-step controls include population growth of the urban area (as log of 1 + annualised population growth over the period), income and education variables for the urban area (log mean income, log standard deviation, and share of university degrees). Extended controls additionally include the urban-area means of the same 20 geography and geology controls as in table 3 in the main text and the same two land use variables (share of built-up land and average height of buildings) used in the same table.

Appendix C. Second-step: spatial heterogeneity

Appendix table 3 duplicates table 4 in the main text and includes interaction terms for population and income or education. Panel A considers house prices at the centre as dependent variable and includes the interaction between log city population and log mean city income as explanatory variable. Panel B also considers house prices as dependent variable and includes the interaction between log city population and the city share of university graduates as explanatory variable. Panels C and D mirror the previous two panels but use land prices instead of house prices as dependent variable.

Appendix Table 3: The determinants of unit house and land prices at the centre, OLS regressions with interactions between population and socioeconomic characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
First-step	Only fixed effects			Basic controls			Full set of controls		
Controls	N	Y	Ext.	N	Y	Ext.	N	Y	Ext.
Panel A. Houses, population and income interacted									
Log population	0.175[a]	0.174[a]	0.223[a]	0.204[a]	0.203[a]	0.288[a]	0.199[a]	0.199[a]	0.291[a]
	(0.0169)	(0.0141)	(0.0283)	(0.0183)	(0.0164)	(0.0357)	(0.0183)	(0.0167)	(0.0361)
Log pop. × log inc.	0.00779[a]	0.00171	0.000452	0.0102[a]	0.0102[a]	0.00816[a]	0.00993[a]	0.00816[a]	0.00624[a]
	(0.00093)	(0.00198)	(0.00163)	(0.000666)	(0.00113)	(0.00115)	(0.00067)	(0.00104)	(0.00109)
Log land area	-0.171[a]	-0.152[a]	-0.224[a]	-0.139[a]	-0.118[a]	-0.230[a]	-0.168[a]	-0.149[a]	-0.267[a]
	(0.0174)	(0.0136)	(0.0293)	(0.0205)	(0.0182)	(0.0364)	(0.0193)	(0.0168)	(0.0375)
R^2	0.54	0.65	0.72	0.64	0.69	0.74	0.62	0.67	0.73
Panel B. Houses, population and education interacted									
Log population	0.171[a]	0.173[a]	0.224[a]	0.205[a]	0.195[a]	0.281[a]	0.200[a]	0.194[a]	0.289[a]
	(0.0185)	(0.0141)	(0.0282)	(0.0230)	(0.0161)	(0.0372)	(0.0222)	(0.0166)	(0.0372)
Log pop. × educ.	0.321[a]	0.195	0.0133	0.374[a]	1.349[a]	1.147[a]	0.365[a]	0.948[a]	0.744[a]
	(0.0329)	(0.223)	(0.184)	(0.0465)	(0.176)	(0.164)	(0.0446)	(0.196)	(0.167)
Log land area	-0.172[a]	-0.154[a]	-0.224[a]	-0.138[a]	-0.125[a]	-0.233[a]	-0.167[a]	-0.154[a]	-0.271[a]
	(0.0173)	(0.0136)	(0.0293)	(0.0212)	(0.0181)	(0.0384)	(0.0199)	(0.0170)	(0.0389)
R^2	0.48	0.65	0.72	0.55	0.70	0.75	0.52	0.68	0.74
Panel C. Land parcels, population and income interacted									
Log population	0.716[a]	0.720[a]	0.908[a]	0.583[a]	0.571[a]	0.653[a]	0.581[a]	0.572[a]	0.704[a]
	(0.0469)	(0.0436)	(0.120)	(0.0354)	(0.0328)	(0.0841)	(0.0350)	(0.0337)	(0.0861)
Log pop. × log inc.	0.00874[a]	-0.00925[b]	-0.0148[a]	0.0140[a]	0.0236[a]	0.0197[a]	0.0120[a]	0.0182[a]	0.0135[a]
	(0.00183)	(0.00450)	(0.00510)	(0.00116)	(0.00369)	(0.00413)	(0.00106)	(0.00324)	(0.00434)
Log land area	-0.698[a]	-0.679[a]	-0.906[a]	-0.380[a]	-0.355[a]	-0.472[a]	-0.469[a]	-0.447[a]	-0.608[a]
	(0.0493)	(0.0449)	(0.131)	(0.0402)	(0.0368)	(0.0906)	(0.0399)	(0.0367)	(0.0936)
R^2	0.58	0.64	0.70	0.74	0.77	0.80	0.71	0.74	0.78
Panel D. Land parcels, population and education interacted									
Log population	0.695[a]	0.718[a]	0.892[a]	0.581[a]	0.572[a]	0.664[a]	0.579[a]	0.577[a]	0.715[a]
	(0.0472)	(0.0423)	(0.119)	(0.0402)	(0.0324)	(0.0861)	(0.0386)	(0.0332)	(0.0873)
Log pop. × educ.	0.489[a]	-0.655	-0.906[b]	0.598[a]	1.868[a]	1.686[a]	0.511[a]	1.228[a]	1.021[a]
	(0.0757)	(0.468)	(0.445)	(0.0744)	(0.338)	(0.350)	(0.0665)	(0.406)	(0.382)
Log land area	-0.703[a]	-0.673[a]	-0.888[a]	-0.378[a]	-0.370[a]	-0.493[a]	-0.467[a]	-0.457[a]	-0.623[a]
	(0.0469)	(0.0449)	(0.130)	(0.0393)	(0.0375)	(0.0929)	(0.0387)	(0.0374)	(0.0950)
R^2	0.59	0.64	0.70	0.70	0.77	0.80	0.68	0.74	0.78

Notes: 1,937 observations in all columns for panels A and B and 1,933 for panels C and D. This table duplicates table 4 in the main text but also includes an interaction between population and log income or education (share of university degrees). All reported R^2 are within-time. The superscripts *a*, *b*, and *c* indicate significance at 1%, 5%, and 10% respectively. Standard errors clustered at the urban area level are between brackets.

Appendix Table 4: The determinants of unit house and land prices at the centre, OLS regressions using a first step estimation where distance is interacted with income

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
First-step	Only fixed effects			Basic controls			Full set of controls		
Controls	N	Y	Ext.	N	Y	Ext.	N	Y	Ext.
Panel A. Houses									
Log population	0.262[a]	0.215[a]	0.302[a]	0.258[a]	0.213[a]	0.300[a]	0.253[a]	0.209[a]	0.306[a]
	(0.0274)	(0.0185)	(0.0424)	(0.0269)	(0.0184)	(0.0420)	(0.0262)	(0.0181)	(0.0408)
Log land area	-0.122[a]	-0.131[a]	-0.245[a]	-0.118[a]	-0.126[a]	-0.241[a]	-0.142[a]	-0.151[a]	-0.276[a]
	(0.0253)	(0.0191)	(0.0439)	(0.0247)	(0.0189)	(0.0433)	(0.0247)	(0.0190)	(0.0422)
R^2	0.44	0.68	0.73	0.44	0.67	0.73	0.40	0.65	0.72
Observations	1,937	1,937	1,937	1,937	1,937	1,937	1,937	1,937	1,937
Panel B. Land parcels									
Log population	0.869[a]	0.797[a]	0.980[a]	0.649[a]	0.587[a]	0.724[a]	0.650[a]	0.592[a]	0.751[a]
	(0.0592)	(0.0510)	(0.107)	(0.0445)	(0.0358)	(0.0911)	(0.0434)	(0.0361)	(0.0913)
Log land area	-0.472[a]	-0.489[a]	-0.717[a]	-0.369[a]	-0.382[a]	-0.560[a]	-0.429[a]	-0.440[a]	-0.640[a]
	(0.0603)	(0.0548)	(0.113)	(0.0473)	(0.0431)	(0.0936)	(0.0470)	(0.0428)	(0.0950)
R^2	0.60	0.68	0.71	0.61	0.72	0.77	0.60	0.71	0.76
Observations	1,933	1,933	1,933	1,933	1,933	1,933	1,933	1,933	1,933

Notes: This table duplicates table 4 of the main text but relies on a first-step estimation that also includes an interaction term of log distance and log municipal income with a separate coefficient estimated for each urban area.

Appendix table 4 duplicates table 4 in the main text but it relies on first-step estimates which also include an interaction term between log distance and log municipal income for which we estimate a specific coefficient for each urban area. Panel A considers house prices at the centre as dependent variable while panel B considers unit land prices. For our preferred specification in column 8, the estimated population elasticity is 0.209 for house prices and 0.592 for land prices. These elasticities are extremely close to 0.208 and 0.597, respectively, estimated in the corresponding column of table 4 of the main text. On average, in panel A the coefficients are about 0.03 higher than in the corresponding panel if table 4. We also note that the more noisy estimates for land prices in panel B. This is likely due to power issues in the first step as 277 extra coefficients are estimated.

We note finally that a first-step specification including an interaction term between distance and income group would coincide closely with the predictions of the monocentric urban model with discrete income groups that differ in size across cities and face different commuting costs (Duranton and Puga, 2015). Because sorting within cities is in reality less extreme than the perfect sorting predicted by this simple model and because we have a continuum of incomes instead of discrete income groups, in our specification we interact continuous income with distance instead of using indicator variables by income group interacted with distance.

Appendix D. Second-step: all dwellings

Appendix Table 5: The determinants of unit property prices at the centre, OLS regressions for all dwellings

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
First-step	Only fixed effects			Basic controls			Full set of controls		
Controls	N	Y	Ext.	N	Y	Ext.	N	Y	Ext.
Log population	0.200^a	0.163^a	0.170^a	0.222^a	0.184^a	0.237^a	0.182^a	0.151^a	0.187^a
	(0.0191)	(0.0119)	(0.0272)	(0.0257)	(0.0174)	(0.0379)	(0.0197)	(0.0134)	(0.0340)
Log land area	-0.129^a	-0.130^a	-0.157^a	-0.0995^a	-0.104^a	-0.181^a	-0.114^a	-0.117^a	-0.168^a
	(0.0198)	(0.0125)	(0.0287)	(0.0227)	(0.0176)	(0.0367)	(0.0184)	(0.0140)	(0.0351)
R^2	0.34	0.66	0.73	0.36	0.61	0.67	0.30	0.57	0.64
Observations	1,937	1,937	1,937	1,937	1,937	1,937	1,937	1,937	1,937

Notes: The dependent variable is an urban area-time fixed effect estimated in the first step using municipal prices for all dwellings instead of only houses. Otherwise, this table is similar to panel A of table 4 in the main text. The superscripts *a*, *b*, and *c* indicate significance at 1%, 5%, and 10% respectively. Standard errors clustered at the urban area level are between brackets. All R^2 are within time.

The specifications of Appendix table 5 duplicate those of panel A of table 4 in the main text for housing prices that pertain to all dwellings instead of only houses. We estimate population elasticities of the price at the centre that are somewhat lower than in table 4 of the main text where we consider only houses. As mentioned in the main text, this is possibly caused by the lower land intensity of apartments relative to houses.

To gain further insight into this question, it is interesting to consider the following back-of-the-envelop calculation. For our preferred specification of column 8 in appendix table 5, we estimate a population elasticity that is about 27% less for the price of all dwellings relative to the price of houses, 0.151 instead of 0.208 (estimated in the corresponding specification table 4 in the main text). More generally, in appendix table 5 we estimate population elasticities that are between 10% and 40% lower for all dwellings relative to the same elasticity for houses only.

Recall that our model interprets the ratio of the elasticity of housing prices to the elasticity of land prices as the share of land in housing (see Appendix 1 in the main text). Hence, for our preferred estimate the implicit share of land implied by our model for all dwellings is thus about $(1 - 0.27 =)$ 0.73 times the share of land for houses only (and between about 0.6 to 0.9 times when considering all specification of Appendix table 5 and table 4 of the main text). Put differently, with our preferred specification we have an implicit share of land for all dwellings of about 0.25 (and more generally between 0.2 and 0.3 for other specifications) instead of 0.36 for houses (which we know from new construction data).

With about 50% of apartments and 50% of single family homes in French urban areas (CGDD, 2011), this implies a share of land for apartments of 0.14 so that the average between apartments and houses reaches 0.25 (and more generally we obtain a range between 0.05 and 0.25 for other specifications regarding the share of land for apartments). While this calculation is subject to caveats (including applying the share of 0.36 observed in the data for new house constructions to all houses), these proportions do not strike us as implausible.

Appendix E. Second-step: further robustness checks

Appendix tables 6 and 7 report results for further robustness checks for house prices in panel A and for land prices in panel B.

The specifications of appendix table 6 experiment with a number of further specifications regarding the distance gradient using either alternative functional forms to measure distance in the first step, alternative definitions for centres, richer specifications for distance effects allowing gradients to vary across years for each urban area, or alternative samples of observations eliminating potentially more selected observations that are either particularly close to the centre or particularly cheap.

For house prices, the estimated population elasticity is between 0.180 and 0.228 for seven of the eight specifications of the table instead of 0.208 for our preferred estimate in table 4 of the main text. When allowing for two centres in column 4, the estimated population elasticity is lower at 0.134. For land prices, we find relatively similar patterns.

Appendix table 7 reports results for specifications that explore two further potential problems. The first two columns focus on samples of observations that do not contain urban areas with low or negative growth. As argued by Glaeser and Gyourko (2005), the housing supply curve is expected to be kinked and much steeper when population declines as the supply of housing is then inelastic

Appendix Table 6: The determinants of unit house prices, further robustness checks part 1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Houses							
Log population	0.188^a	0.228^a	0.180^a	0.134^a	0.207^a	0.194^a	0.211^a
	(0.0162)	(0.0251)	(0.0155)	(0.0439)	(0.0352)	(0.0177)	(0.0185)
Log land area	-0.149^a	-0.168^a	-0.135^a	-0.0352	-0.140^a	-0.146^a	-0.154^a
	(0.0163)	(0.0343)	(0.0155)	(0.0574)	(0.0339)	(0.0172)	(0.0181)
R^2	0.64	0.46	0.61	0.39	0.40	0.64	0.61
Observations	1,937	1,937	1,937	1,937	1,937	1,937	1936
Panel B. Land parcels							
Log population	0.535^a	0.546^a	0.542^a	0.513^a	0.605^a	0.620^a	0.696^a
	(0.0317)	(0.0433)	(0.0332)	(0.0512)	(0.0400)	(0.0348)	(0.0937)
Log land area	-0.451^a	-0.486^a	-0.468^a	-0.381^a	-0.389^a	-0.477^a	-0.599^a
	(0.0356)	(0.0739)	(0.0376)	(0.0658)	(0.0433)	(0.0360)	(0.144)
R^2	0.70	0.43	0.66	0.57	0.69	0.74	0.20
Observations	1,933	1,933	1,933	1,933	1,933	1,933	1,921

Notes: [a]: significant at 1% level; [b]: significant at 5% level; [c]: significant at 10% level. Standard errors clustered at the urban area level are between brackets. All R^2 are within time. All OLS regressions. Each column is a variant of our preferred OLS estimation reported in column 8 of table 4 of the main text. As explanatory variables, column 1 includes the distance to the centre of the urban area in level instead of its log. Column 2 includes log distance and its square (estimating a specific coefficient for each urban area for both variables). Column 3 defines the centre of an urban area as the centroid of the municipality with the highest residential density. Column 4 controls for distances to the the closest two municipalities with the highest population in the urban area. Column 5 drops the 25% of observations closest to the centre in each urban area. Column 6 drops the 25% of observations with the lowest price per square metre in each urban area. Column 7 uses as dependent variable urban-area fixed effects which are estimated allowing for year-specific gradients for each urban area in the first step.

and only adjust following slow depreciation. We either eliminate observations for urban areas when they experience negative growth during our study period or eliminate every year the lowest 20% of year-to-year population growth. In both cases, eliminating low-growth urban areas leaves the estimated population elasticity of housing prices essentially unchanged. For land prices, the estimated population elasticity is marginally higher (albeit statistically undistinguishable). The following six columns of table 7 experiment with alternative estimation methods that use either a different weighting scheme in the first-step, a different sample of urban areas, or a different econometric approach. In particular, recall that our second-step estimation relies on a dependent variable that is estimated (with error) in a first step. As made clear in Appendix F below, this problem can be addressed using FGLS and WLS techniques. We can also estimate the population elasticity of prices at the centre in a single step. Finally, we estimate the population elasticity on a larger sample of urban areas (324 instead of 277).

Appendix Table 7: The determinants of unit house prices, further robustness checks part 2

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Houses								
Log population	0.223[a]	0.214[a]	0.169[a]	0.160[a]	0.205[a]	0.224[a]	0.204[a]	0.186[a]
	(0.0229)	(0.0194)	(0.0142)	(0.0146)	(0.0067)	(0.0087)	(0.0299)	(0.0230)
Log land area	-0.168[a]	-0.157[a]	-0.149[a]	-0.0730[a]	-0.156[a]	-0.183[a]	-0.153[a]	-0.147[a]
	(0.0214)	(0.0176)	(0.0136)	(0.0159)	(0.0074)	(0.0097)	(0.0306)	(0.0195)
R^2	0.67	0.67	0.59	0.56	-	0.79	0.81	0.64
Observations	1,546	1,607	1,937	1,937	1,937	1,937	74,621	2,266
Panel B. Land parcels								
Log population	0.621[a]	0.616[a]	0.523[a]	0.499[a]	0.585[a]	0.664[a]	0.634[a]	0.537[a]
	(0.0448)	(0.0369)	(0.0323)	(0.0314)	(0.0128)	(0.0176)	(0.0441)	(0.0397)
Log land area	-0.477[a]	-0.473[a]	-0.502[a]	-0.479[a]	-0.369[a]	-0.519[a]	-0.493[a]	-0.409[a]
	(0.0458)	(0.0382)	(0.0343)	(0.0334)	(0.0124)	(0.0198)	(0.0513)	(0.0328)
R^2	0.71	0.73	0.67	0.66	-	0.78	0.81	0.68
Observations	1,572	1,603	1,933	1,933	1,933	1,933	204,656	2,261

Notes: [a]: significant at 1% level; [b]: significant at 5% level; [c]: significant at 10% level. Standard errors are between brackets. Except in columns 5 and 6 (see below), all OLS regressions with standard errors clustered at the urban area level. All R^2 are within time. Each column is a variant of our preferred OLS estimation reported in column 8 of table 4 of the main text. Column 1 drops all urban areas that lost population during the study period. Column 2 drops the 20% of urban areas with the lowest growth each year. Column 3 uses as dependent variable urban-area fixed effect which are estimated without weights in the first step. Column 4 uses as dependent variable urban-area fixed effect which are estimated with population weights in the first step (instead of using the number of transactions as weights). Column 5 estimates the regression using feasible generalised least squares (FGLS) as described in Appendix F. Column 6 estimates the regression using weighted least squares (WLS) as described in Appendix F. Column 7 estimates the elasticity of house and land prices with respect to population in one step instead of two separate steps. Column 8 considers a full sample of 324 urban areas for which we can estimate our preferred specification instead of our preferred sample of 277.

We estimate smaller population elasticities that are up to 0.05 smaller than our preferred elasticity of table 4 in the main text for both house and land prices when using alternative weighting schemes. We also estimate a slightly smaller elasticity when using a larger sample of urban areas, for which the added urban areas are mostly small. This is consistent with the possibility entertained below that this elasticity may be smaller for smaller urban areas. For our other variants, the results only differ marginally.

Appendix F. Second-step: FGLS and WLS estimators

In this appendix, we explain how we construct weighted least squares (WLS) and feasible general least squares (FGLS) estimators used in some second-stage regressions of the previous appendix.

The model is of the form:

$$C = X\varphi + \zeta + \eta\,, \tag{F1}$$

where C is a $JT \times 1$ vector stacking the estimated urban area-time fixed effects capturing unit house or land prices at the centre, $\ln C^P$ or $\ln C^R$, with J the number of urban areas and K the number of years, X is a $JT \times K$ matrix stacking the observations for urban area variables (area, population, population growth, etc.), ζ is a $JT \times 1$ vector of error terms supposed to be independently and identically distributed with variance σ^2, and η is a $JT \times 1$ vector of sampling errors with known covariance matrix V.

It is possible to construct a consistent FGLS estimator of φ as:

$$\widehat{\varphi}_{FGLS} = \left(X'\widehat{\Omega}^{-1}X\right)^{-1} X'\widehat{\Omega}^{-1}C\,, \tag{F2}$$

where $\widehat{\Omega}$ is a consistent estimator of the covariance matrix of $\zeta + \eta$, $\Omega = \sigma^2 I + V$. To compute this estimator, we use an unbiased and consistent estimator of σ^2 which can be computed from the OLS residuals of equation (F1) denoted $\widehat{\zeta + \eta}$:

$$\widehat{\sigma}^2 = \frac{1}{N-K}\left[\widehat{\zeta+\eta}'\widehat{\zeta+\eta} - tr\left(M_X V\right)\right], \tag{F3}$$

where $M_X = I - X\left(X'X\right)^{-1}X'$ is the projection orthogonally to X. We thus use $\widehat{\Omega} = \widehat{\sigma}^2 I + V$ in the computation of (F2). A consistent estimator of the covariance matrix of the FGLS estimator is:

$$\widehat{V}\left(\widehat{\varphi}_{FGLS}\right) = \left(X'\widehat{\Omega}^{-1}X\right)^{-1}. \tag{F4}$$

As the FGLS is said not to be always robust, we also compute a WLS estimator in line with Card and Krueger (1992), using the diagonal matrix of inverse of first-stage variances as weights, denoted Δ. The estimator is given by:

$$\widehat{\varphi}_{WLS} = \left(X'\Delta X\right)^{-1} X'\Delta C\,, \tag{F5}$$

with a consistent estimator of the covariance matrix given by:

$$\widehat{V}\left(\widehat{\varphi}_{WLS}\right) = \left(X'\Delta X\right)^{-1} X'\Delta\widehat{\Omega}_w\Delta X\left(X'\Delta X\right)^{-1},$$

where $\widehat{\Omega}_w = \widehat{\sigma}^2_w I + V$ with $\widehat{\sigma}^2_w$ a consistent estimator of σ^2 based on the residuals of WLS denoted $\widehat{\Delta^{1/2}\left(\zeta+\eta\right)}$ and given by:

$$\widehat{\sigma}^2_w = \frac{1}{tr\left(\Delta^{1/2} M_{\Delta^{1/2}X}\Delta^{1/2}\right)}\left[\widehat{\Delta^{1/2}\left(\zeta+\eta\right)}'\widehat{\Delta^{1/2}\left(\zeta+\eta\right)} - tr\left(\Delta^{1/2} M_{\Delta^{1/2}X}\Delta^{1/2} V\right)\right]. \tag{F6}$$

Appendix G. Second-step: non-constant elasticity

Appendix table 8 duplicates some OLS specifications of table 4 in the main text as well as some IV specifications in the same spirit as those of tables 8 and 9 in the main text and includes terms of higher order for population, namely the square and cube of log population. Panel A considers house prices at the centre as dependent variable while panel B uses land prices.

We find that when estimating specifications which only add a quadratic population term, its coefficient is generally positive and significant. This is suggestive of a convex relationship between log prices for houses or land and log population. As a caveat, we note that this convexity is driven by the three or four largest French urban areas. When we estimate specifications with both a quadratic and a cubic term for log populations, the estimated coefficients are generally not significant.

Adding a quadratic term for log population to our preferred specification of column 8 of table 4 in the main text implies an elasticity of house prices with respect to population of 0.205 for an urban area with 100,000 inhabitants, an elasticity of 0.288 for an urban area with a million inhabitants, and 0.378 for an urban area with the same population as Paris. Because, the non-linear estimate of the population elasticity for Paris is nearly twice as large as our preferred OLS estimate of 0.208, we keep this range in mind for our computation of the urban cost elasticity in section 7 of the main text.

Appendix H. Second-step results without land area

Appendix table 9 duplicates table 4 in the main text but omits land area as an explanatory variable. What is estimated here is the population elasticity of house and land prices when we allow for land area to adjust to population growth.

In appendix table 9, we find that for both house and land prices, the coefficient on population is smaller than when land area is included and typically larger than (or about equal to) the sum of the coefficients on population and land in table 4 in the main text. This is consistent with the standard prediction of land use models for monocentric cities: When cities grow in population, they physically expand. This expansion is slightly less than proportionate as cities also become denser (Duranton and Puga, 2015). When we regress log area on log population, we estimate a

Appendix Table 8: Non-linear effects of population on house and land prices

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
First step controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Second step controls	No	Yes	No	Yes	Yes	No	Yes	Yes
Panel A. House prices								
Log population	0.0370	0.116	-0.325^a	-0.208^b	0.0541	-0.635^a	-0.149	0.00399
	(0.123)	(0.133)	(0.0628)	(0.0935)	(0.832)	(0.228)	(0.122)	(1.453)
Log pop. squared	0.00774	0.00259	0.0248^a	0.0179^a	-0.00376	0.0353^a	0.0154^a	0.00271
	(0.00510)	(0.00582)	(0.00268)	(0.00395)	(0.0667)	(0.00887)	(0.00492)	(0.115)
Log pop. cubed					0.000592			0.000345
					(0.00175)			(0.00299)
Log land area	-0.150^a	-0.152^a	-0.139^a	-0.147^a	-0.147^a	-0.0696^c	-0.131^a	-0.131^a
	(0.0221)	(0.0138)	(0.00897)	(0.0175)	(0.0174)	(0.0322)	(0.0207)	(0.0206)
First-stage statistic						22.1	48.6	15.9
Overid. p-value						0.67	0.44	0.43
Observations	1,937	1,937	1,937	1,937	1,937	1,937	1,937	1,937
R^2	0.35	0.65	0.43	0.67	0.67	-	-	-
Panel B. land prices								
Log population	1.113^a	1.217^a	-0.236	-0.0837	3.265^b	-0.934^a	-0.0984	3.406
	(0.239)	(0.220)	(0.270)	(0.208)	(1.490)	(0.333)	(0.264)	(2.704)
Log pop. squared	-0.0145	-0.0219^b	0.0384^a	0.0293^a	-0.247^b	0.0639^a	0.0305^a	-0.254
	(0.00939)	(0.00881)	(0.0113)	(0.00863)	(0.119)	(0.0126)	(0.0102)	(0.212)
Log pop. cubed					0.00752^b			0.00760
					(0.00312)			(0.00547)
Log land area	-0.678^a	-0.680^a	-0.432^a	-0.447^a	-0.448^a	-0.322^a	-0.434^a	-0.434^a
	(0.0528)	(0.0448)	(0.0454)	(0.0377)	(0.0371)	(0.0571)	(0.0443)	(0.0433)
First-stage statistic						26.3	31.8	10.3
Overid. p-value						0.12	0.70	0.63
Observations	1,933	1,933	1,933	1,933	1,933	1,933	1,933	1,933
R^2	0.54	0.64	0.63	0.74	0.74	-	-	-

Notes: OLS regressions in column 1 to 5 and LIML regressions in column 6 to 8. The fixed effects for house and land prices are as estimated in column 2 of table 3 in the main text (no first-step controls) or as column 9 of the same table (with first-step controls). The second-step controls are either only year effects (no second-step controls) or the controls used in our preferred estimation of column 8 of table 4 of the main text (second-step controls). Instruments include: 1831 (log) urban population, and its square, and 1881 (log) urban population density in columns 6 to 8 of panel A. Column 6 additionally includes January temperature. Column 7 additionally includes (log) of number of hotel rooms. Column 8 additionally includes the (log) of number of hotel rooms and the cube of 1831 population. In panel B, column 6-8 include 1831 (log) urban population, and its square, and 1881 (log) urban population density. Columns 6 and 7 additionally include a Bartik industry employment growth predictor for 1990-1999. Column 8 additionally includes a Bartik industry employment growth predictor for 1990-1999 and the cube of log 1831 population. ^{a}: significant at 1% level; ^{b}: significant at 5% level; ^{c}: significant at 10% level. All R^2 are within time. Standard errors clustered at the urban area level are between brackets. The critical value for 10% maximal LIML size of Stock and Yogo (2005) weak identification test is below 5.44 for columns. Controls are first conditioned out before the estimation. The first-stage statistics is the Kleibergen-Paap rk Wald F.

Appendix Table 9: The determinants of unit house and land prices at the centre, OLS regressions without land area

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
First-step	Only fixed effects			Basic controls			Full set of controls		
Controls	N	Y	Ext.	N	Y	Ext.	N	Y	Ext.
Panel A. Houses									
Log population	0.110[a]	0.0775[a]	0.0234[b]	0.178[a]	0.136[a]	0.0883[a]	0.151[a]	0.109[a]	0.0561[a]
	(0.0110)	(0.0103)	(0.00936)	(0.0178)	(0.0128)	(0.0139)	(0.0157)	(0.0122)	(0.0124)
R^2	0.24	0.53	0.67	0.40	0.62	0.69	0.33	0.58	0.67
Observations	1,937	1,937	1,937	1,937	1,937	1,937	1,937	1,937	1,937
Panel B. Land parcels									
Log population	0.296[a]	0.262[a]	0.0671[b]	0.434[a]	0.365[a]	0.242[a]	0.352[a]	0.299[a]	0.163[a]
	(0.0252)	(0.0348)	(0.0303)	(0.0288)	(0.0252)	(0.0271)	(0.0262)	(0.0265)	(0.0256)
R^2	0.23	0.34	0.59	0.54	0.66	0.75	0.44	0.55	0.70
Observations	1,933	1,933	1,933	1,933	1,933	1,933	1,933	1,933	1,933

Notes: This table duplicates table 4 in the main text but omits land area as an explanatory variable.

coefficient of about 0.7, consistent with our comparison between appendix table 9 and table 4 in the main text.

The other remarkable result of appendix table 9 is that the population elasticity of land prices remains about three times as large as the population elasticity of house prices. This occurs despite sizeable fluctuations in the absolute value of these elasticities across specifications. This result is highly consistent with our theoretical model which predicts that the ratio of these two elasticities should be equal to the inverse of the share of land in the value of houses. This share is equal to 0.36 in our data for the new constructions associated with the land parcels that we observe.

Appendix I. Second-step: IV estimations for 2000-2012 differences

When regressing 2012-2000 changes in house prices at the centre on changes in population over the same period, the latter is potentially endogenous. An unobserved labour demand shock in an urban area may simultaneously determine house price growth and population growth. It is also possible that house price growth affects population growth. To address this worry, we follow a standard strategy initially proposed by Bartik (1991) and often used in subsequent literature (e.g., Diamond, 2016, among many others).

Appendix Table 10: The determinants of house prices at the centre, IV estimations in difference

	(1)	(2)	(3)	(4)
First-step controls	Yes	No	Yes	No
Second-step controls	No	No	Yes	Yes
Log population	0.905^{a}	0.916^{b}	1.964^{b}	2.022^{b}
	(0.344)	(0.369)	(0.843)	(0.924)
First-stage statistic	11.3	11.3	7.1	7.1
Overidentification p-value	0.18	0.20	0.91	0.88
Instruments				
Number of hotel rooms	Y	Y	N	N
Urban population in 1831	Y	Y	N	N
Bartik industry 1999-2011	Y	Y	Y	Y
Bartik occupation 2006-2011	Y	Y	Y	Y
Observations	275	275	275	275

Notes: ^{a}: significant at 1% level; ^{b}: significant at 5% level; ^{c}: significant at 10% level. White-robust standard errors reported between brackets. The first-step controls are the same as in column 9 of table 3 in the main text. The second-step controls correspond to the extended controls used in column 8 of table 4 that are time varying. All estimations are performed with limited information maximum likelihood (LIML). The critical value for 10% maximal LIML size of Stock and Yogo (2005) weak identification test is 5.44 in columns (1) and (2) and 8.68 in columns (3) and (4). For these columns, it is 5.33 for 15% maximal LIML size. The first-stage statistics is the the Kleibergen-Paap rk Wald F.

The idea of the 'Bartik instrument' is that we can predict the population growth of cities using their initial structure of sectoral employment interacted with the national growth of sectoral employment. Loosely put, a city with a high fraction of employment in high-end services in 2000 is expected to enjoy more growth from 2000 to 2012 than a city with a high initial share of employment in traditional manufacturing which kept declining over the period. We also develop a parallel approach using the initial structure of employment by occupations and national changes in employment by occupation. This approach is described in greater details in Appendix 2 in the main text.

The results are reported in appendix table 10. The point estimates are larger but noisy, in particular when we include changes in income, education, and inequality as controls in columns 3 and 4. While they do not contradict those of table 5 in the main text, these imprecise estimates are probably the consequence of our instruments being marginally weak for these specifications. This is perhaps unsurprising. Changes in labour demand may be tracked by changes in predicted employment (our instrument) but also by changes in local incomes (a control). Put differently, our controls may condition out much of the variation contained in the Bartik predictors. The estimates of columns 1 and 2, which do not include changes in income, education, and inequality for the urban area as controls lead to stronger instruments and relatively more precisely estimated coefficients. The point estimates are also more in line with those obtained without instrumenting in table 5 of the main text.

Appendix J. The share of housing in expenditure: supplementary results

In addition to the issues already discussed in the main text, we may also worry that our results regarding the share of housing in expenditure for the joint sample of homeowners and renters may mask some important heterogeneity between the two groups. To gain insight into this issue, we duplicate the results of table 6 in the main text separately for homeowners and renters in the two panels of appendix table 11. We first note that, unsurprisingly, renters are more prevalent than homeowners in larger urban areas. The difference is nonetheless modest as mean urban area population is 3.13 million for homeowners instead of 3.29 million for renters. A comparison of the two samples of renters and homeowners also indicates that renters devote a slightly larger share of their income to housing than homeowners.[2]

Turning to the coefficients on city population, we find that they are very close for renters and homeowners in most OLS specifications. Modest differences arise when we instrument for population. We estimate coefficients of 0.055 for homeowners and 0.034 for renters instead of 0.048 for the pooled sample of column 8 of table 6 of the main text. While the coefficients on population for renters and homeowners differ, they remain less than two standard deviations apart. They are also, for most of them, in the same range as our estimates for the pooled sample in table 6 of the main text.

In results not reported here, we also experimented with instrumenting for land area using 1881 population density in addition to population. This does not affect our results in any major way. For instance, we estimate a coefficient of 0.039 for city population instead of 0.048 in column 8 of table 6 of the main text when also instrumenting for land area. We also experimented with including education directly as a control variable to condition out elements of permanent income instead of instrumenting. This does not affect the coefficient on urban area population. Using education as

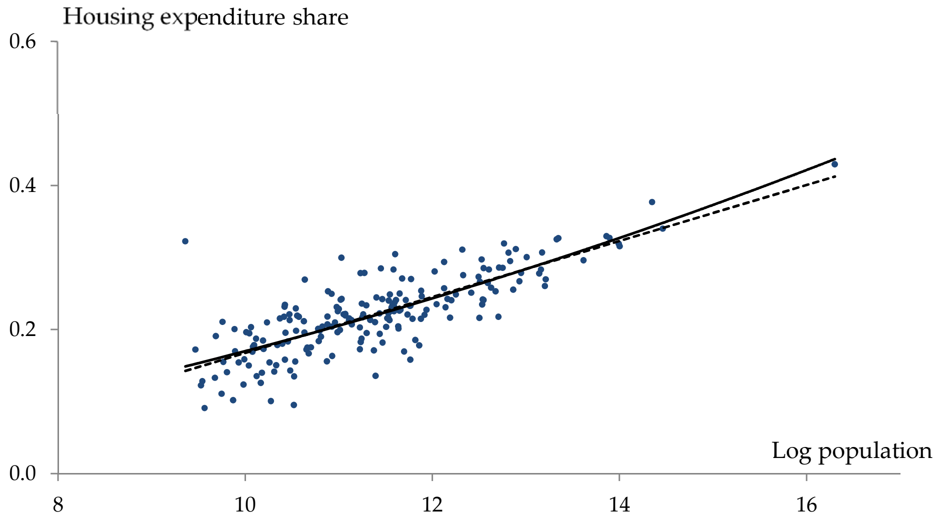
[2]This difference remains somewhat modest at about 4 percentage points after we account for the difference in mean city population. This difference even flips signs if we also account for income differences across both groups. Overall, these results suggest small differences between the two groups.

Appendix Table 11: The share of housing in expenditure for homeowners and private renters

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Homeowners								
Log population	0.027^{a}	0.029^{a}	0.041^{a}	0.045^{a}	0.044^{a}	0.055^{a}	0.076^{a}	0.055^{a}
	(0.001)	(0.002)	(0.005)	(0.008)	(0.008)	(0.014)	(0.013)	(0.012)
Log land area			-0.020	-0.028^{a}	-0.033^{a}	-0.038^{a}	-0.057^{a}	-0.038^{a}
			(0.007)	(0.008)	(0.007)	(0.013)	(0.012)	(0.011)
Population growth			2.593^{b}	2.662^{a}	2.443^{a}	2.470^{a}	2.084^{a}	2.471^{a}
			(0.610)	(0.727)	(0.743)	(0.763)	(0.780)	(0.740)
Log distance to city centre		-0.005	-0.004	-0.006^{c}	-0.002	-0.008^{b}	-0.013^{a}	-0.008^{b}
		(0.005)	(0.005)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)
Log income	-0.252^{a}	-0.253^{a}	-0.253^{a}	-0.256^{a}	-0.168^{a}	-0.256^{a}	-0.257^{a}	-0.256^{a}
	(0.012)	(0.011)	(0.011)	(0.010)	(0.013)	(0.009)	(0.009)	(0.009)
First-stage statistic					253.2	97.0	5.8	14.9
Overidentification p-value					0.33		0.26	0.05
Instruments								
Degree					X			
Urban population in 1831						X		X
Consumption amenities							X	X
Local controls	No	No	No	Yes	Yes	Yes	Yes	Yes
R^2	0.53	0.53	0.54	0.55				
Panel B. Renters								
Log population	0.030^{a}	0.033^{a}	0.038^{a}	0.028^{a}	0.021^{a}	0.028^{c}	0.056^{a}	0.034^{a}
	(0.002)	(0.002)	(0.009)	(0.009)	(0.008)	(0.014)	(0.017)	(0.012)
Log land area			-0.008	0.005	0.009	0.005	-0.021	-0.001
			(0.013)	(0.012)	(0.011)	(0.018)	(0.019)	(0.016)
Population growth			2.775^{b}	3.950^{a}	4.205^{a}	3.957^{a}	3.277^{b}	3.806^{a}
			(1.262)	(1.116)	(1.184)	(1.256)	(1.273)	(1.217)
Log distance to city centre		-0.009^{b}	-0.009^{b}	-0.005	-0.003	-0.005	-0.011^{b}	-0.006
		(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)
Log income	-0.342^{a}	-0.342^{a}	-0.341^{a}	-0.343^{a}	-0.184^{a}	-0.343^{a}	-0.343^{a}	-0.343^{a}
	(0.023)	(0.023)	(0.023)	(0.022)	(0.033)	(0.022)	(0.022)	(0.022)
First-stage statistic					31.6	157.4	8.1	22.0
Overidentification p-value					0.03		0.03	0.01
R^2	0.58	0.58	0.58	0.59				

Notes: ^{a}: significant at 1% level; ^{b}: significant at 5% level; ^{c}: significant at 10% level. All R^2 are within time. The same regressions are estimated in both panels. 5,984 observations in each regression of panel A corresponding to 177 urban areas. 2,464 observations in each regression of panel B corresponding to 177 urban areas (20 of which differ from the previous sample). All variables are centred and the estimated constant, which corresponds to the expenditure share in a city of average size (2.94 million inhabitants in panel A and 3.12 million in panel B), takes the value 0.314 in all specifications of panel A and 0.352 in all specifications of panel B. Regressions are weighted with sampling weights and include: age and dummies for year 2011 (ref. 2006), living in couple within the dwelling (ref. single), one child, two children, three children and more (ref. no child). Standard errors clustered at the urban area level are reported between brackets. Local controls include the same geography variables for urban areas as in table 4 of the main text and the same geology, land use, and amenity variables as in table 3 of the main text. OLS for columns (1) to (4). IV estimated with limited information maximum likelihood (LIML) in columns (5) (income instrumented), (6) and (7) (population instrumented) and (8) (income and population instrumented). The first-stage statistics is the Kleibergen-Paap rk Wald F. The critical value for 10% maximal LIML size of Stock and Yogo (2005) weak identification test is 4.45 for column (5), 16.38 for column (6), 3.50 for column (7), and 3.42 for column (8). The instruments are the same as in table 8 in the main text. The education instruments are five indicator variables corresponding to PhD and elite institution degree, master, lower university degree, high school and technical degree, lower technical degree, and primary school (reference).

Appendix Figure 1: Share of housing in household expenditure and log city population



Notes: The horizontal axis represents log urban area population. The vertical axis represents the urban area median of the residual of column 8 of table 6 in the main text plus log urban area population multiplied by its estimated coefficient. The plain continuous curve is a quadratic trend line. The dotted line is a linear trend.

a control variable to the specification of column 4 of table 6 of the main text leads to a coefficient 0.033 for population instead of 0.036 in column 5 where it is used as instrument.

Our last worry is about functional forms. Our (semi log) linear estimation of a share of expenditure on a log population may fail to capture important non-linearities as population increases. In figure 1, we provide a 'component plus residual' plot where we represent the share of housing in expenditure after conditioning out other controls on the vertical axis and log urban area population on the horizontal axis. The figure also contains two trend lines, linear and quadratic. As made clear by the figure, the two trends are virtually undistinguishable except at the very top of the distribution. For a city of the size of Paris, the difference between the linear and quadratic trends is a modest 2 percentage points. For a city of the size of Lyon (the second largest city), the difference is already less than half of a percentage point. Consistent with this, the difference in explanatory power between the quadratic and linear trends is small. We have an R^2 of 63.1% for the quadratic instead of 62.8% for the linear trend line. Hence, we conclude that our log linear specification provides an accurate first-order description of the relationship between housing expenditure and city population, except for Paris that deviates modestly.

Appendix Table 12: The elasticity of urban costs

	City 1 (pop. 100,000)			City 2 (pop. 1m)			City 3 (pop. Paris)		
Panel A. Population elasticity of prices									
Baseline (preferred OLS)	0.208	0.208	0.208	0.208	0.208	0.208	0.208	0.208	0.208
Non-linear population elasticity	0.205	0.205	0.205	0.288	0.288	0.288	0.378	0.378	0.378
12-year adjustment	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.780	0.780
Allowing for urban expansion	0.109	0.109	0.109	0.109	0.109	0.109	0.109	0.109	0.109
Panel B. Housing share									
Slope of the housing share	0.028	0.048	0.067	0.028	0.048	0.067	0.028	0.048	0.067
Share of housing in expenditure	0.093	0.159	0.228	0.247	0.269	0.293	0.363	0.390	0.415
Panel C. Urban costs elasticity using:									
Baseline	0.019	0.033	0.048	0.051	0.056	0.061	0.075	0.081	0.086
	(0.007)	(0.007)	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)	(0.008)
Non-linear population elasticity	0.019	0.032	0.047	0.071	0.078	0.084	0.137	0.147	0.157
	(0.002)	(0.007)	(0.005)	(0.007)	(0.007)	(0.007)	(0.015)	(0.017)	(0.018)
12-year adjustment	0.073	0.124	0.178	0.193	0.210	0.228	0.283	0.304	0.324
	(0.031)	(0.036)	(0.041)	(0.044)	(0.047)	(0.051)	(0.063)	(0.069)	(0.073)
Allowing for urban expansion	0.010	0.017	0.025	0.027	0.029	0.032	0.040	0.043	0.045
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)	(0.005)

Notes: In panel A, row 1, the estimate of 0.208 is our preferred OLS estimate from column 8 of table 4. In row 2, the three estimates are marginal effects computed from column 4 of appendix table 10. In row 3, the estimate of 0.780 is for the 2000-2012 difference from column 8 of table 5. In row 4, we use the elasticity of 0.109 estimated in column 8 of appendix table 11, which does not include land area as a control. In panel B, for the coefficient on log population for the housing share we report our preferred estimate from column 8 of table 6 as well as the largest and smallest coefficients for log population estimated in the same table. From these coefficients and the constant of the regression, we compute the predicted housing share in expenditure for our three hypothetical cities. Panel C reports the urban cost elasticity for all the combinations of housing share in expenditure and population elasticity of house prices. Standard errors in brackets are computed from the estimated coefficients and their variances using the following formula for the variance of their product: $var(XY) = var(X)var(Y) + var(X)E(Y)^2 + var(Y)E(X)^2$.

Appendix K. More complete results for the urban cost elasticity

In the main text, we focus on the share of housing in expenditure predicted from our preferred estimate for the coefficient on log city population of 0.048 in table 6. In this appendix, we also consider a lower estimate of 0.028 and a higher estimate of 0.067 corresponding to the lowest and highest estimated coefficients for log city population obtained in table 6 of the main text. The predicted shares of housing in expenditure associated with the three scenarios are reported in panel B of appendix table 12. We note that for a city like Paris or for a city with a million inhabitants, the predicted share of housing in expenditure is only modestly affected by the value that we consider for the population semi elasticity. Differences are larger for a city with 100,000 inhabitants.

Consistent with this result, we find that the exact way we predict the share of housing in expenditure only makes a modest difference to our estimated urban cost elasticity for the hypothetical cities with one or 12 million inhabitants like Paris. Appendix table 12 reports a full set of results. The differences are more sizeable for a smaller city with 100,000 inhabitants. For this hypothetical city, we prefer to rely on the predicted share of housing in expenditure of 0.159 coming from our preferred estimate of 0.048 for log population. This share of 0.159 is close to the share we observe in the data for actual urban areas of this size. Our more extreme values for the population coefficient predict housing shares of 0.228 or 0.093, which are out of line with the raw data.

References

Bartik, Timothy. 1991. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo (MI): W.E. Upjohn Institute for Employment Research.

Card, David and Alan B. Krueger. 1992. School quality and black-white relative earnings: A direct assessment. *Quarterly Journal of Economics* 107(1):151–200.

Commissariat Général au Développement Durable. 2011. *Comptes du Logement: Premiers Résultats 2010, le Compte 2009*. Paris: Ministère de l'Ecologie, du Développement Durable, des Transports et du Logement.

Diamond, Rebecca. 2016. The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000. *American Economic Review* 106(3):479–524.

Duranton, Gilles and Diego Puga. 2015. Urban land use. In Gilles Duranton, J. Vernon Henderson, and William C. Strange (eds.) *Handbook of Regional and Urban Economics*, volume 5A. Amsterdam: North-Holland, 467–560.

Glaeser, Edward L. and Joseph Gyourko. 2005. Urban decline and durable housing. *Journal of Political Economy* 113(2):345–375.

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.