The estimation of cluster effects in linear panel models

Laurent Gobillon
INED and LSE*

October 22, 2004
Preliminary Version

Abstract

In this paper, we study the impact of aggregate variables on individual outcome in linear panel models with fixed effects. Individuals are *mobile* and can change aggregate group at each period. We show how a two-stage estimation method can properly account for aggregate unobserved heterogeneity to avoid biases on standard errors. The method can deal with individual and aggregate heteroskedasticity and/or autocorrelation. It is also flexible enough to allow for instrumentation of both individual and aggregate variables. Using Monte-Carlo simulations, we study the properties of the estimators for different mobility patterns of individuals between aggregate groups.

Keywords: panel data; multilevel model; cluster sample; fixed effects
JEL Classification: C23

*Laurent Gobillon, Institut National d'Etudes Démographiques (INED), 133 Boulevard Davout, 75980 Parix Cedex 20, France. Email: laurent.gobillon@ined.fr.

1 Introduction

In many economic fields, researchers are interested in the measure of some group effects on an individual outcome. The related issues encompass the effect of industry structure on the individual wage in labour economics (Krueger and Summers, 1988; Gibbons and Katz, 1992; Abowd, Kramarz and Margolis, 1999); the impact of density and human externalities on the local productivity of workers in economic geography (Glaeser and al., 1992; Rauch, 1993; Ciccone and Hall, 1996; Ciccone, 2002; Combes, Duranton and Gobillon, 2003); the effect of local school characteristics on the market price of households' dwellings in urban economics (Gibbons and Machin, 2003); the consequences of local weather conditions on farmers' income in development economics (Gurgand, 2003).

As a consequence, econometric methods have been developped to estimate the effect of aggregate variables on an individual outcome when using linear cross-section models (see Goldstein, 1995; Wooldridge, 2002 and 2003). In this paper, we extend previous work to linear panel models including individual fixed effects. We focus on cases where individuals are *mobile* and can change group across time.[1] The group choice process of individuals is supposed to be strictly exogenous.

It is well-known from cross-section analysis that aggregate heterogeneity should be taken into account to avoid some potentially large biases on standard errors. For that purpose, many papers introduce *iid* aggregate random terms in their econometric specification and conduct a feasible general least square estimation (see for instance: Moulton, 1990; Pepper, 2002).[2] However, such an approach becomes unfeasible when using a panel model with fixed effects where individuals are *mobile*. Indeed, fixed effects should be differenced out and the structure of the covariance matrix becomes very complex as individuals move between groups.

A less widespread cross-section approach consists in estimating the model in the within-group dimension. This allows to recover some estimates of the coefficients of individual variables. The group-mean of residuals is then regressed on the aggregate variables to recover some estimates of group coefficients (see for instance: Hausman and Taylor, 1981; Donald and Lang, 2001). When the uncertainty on the dependent variable is properly accounted for in this group-mean regression, it is possible to compute some unbiased standard errors for the estimated coefficients of aggregate variables. The procedure is equivalent to estimating group fixed effects in a first stage, and regressing them on the aggregate variables in second stage.

This two-stage method can be extended to panel models with individual fixed effects: in first stage, all aggregate terms are replaced by group-year fixed effects in the outcome equation. Thus, the individual outcome is explained by

[1] The case where individuals are *immobile* is addressed briefly in section 3. We also give some more information on models where individual unobserved effects are random (and not fixed) in sections 3.

[2] A closely related approach is the iterated generalized least square estimator proposed by Goldstein (1986).

the individual explanatory variables, some individual fixed effects, some group-year fixed effects and some individual error terms. It is evaluated after individual fixed effects have been differenced out. The estimated group-year fixed effects are then regressed in second stage on the aggregate variables.

We explain how to construct an unbiased and consistent estimator of the variance of aggregate error terms. In particular, this estimator is the root mean square of second-stage residuals corrected to account for the uncertainty on the dependent variable. It is used to compute unbiased and consistent standard errors for the estimated coefficients of group variables. As a by-product, a feasible general least square estimation in second stage can also be performed.

We then present some extensions of the method that allow to account for heteroskedasticity and/or autocorrelation of individual and aggregate errors terms. We also explain how robust standard errors can be recovered when individual and aggregate variables are instrumented. We then discuss which properties of the estimators still hold if the exogeneity assumption made for the group choice process of individuals is relaxed.

The accuracy of all the estimators proposed in this paper depends on the group mobility pattern of individuals across time. We analyse two simple cases for which this mobility pattern differs. In the first case, some individuals depart from each group and go to all other groups between two dates. In the second case, all movers from group g go to group $g+1$, except those in the last group that move to the first group. We show that the estimated group-year fixed effects are measured on average with far more accuracy in the first case than in the second case when the number of groups is high.

We finally conduct some Monte-Carlo simulations to study more complex cases. Results suggest that the two-stage method allows to avoid a bias on standard errors that can be higher than 200% as in Moulton (1990). The estimator of the variance of aggregate error terms is accurate as long as groups are well interconnected across time by flows of movers. The corrective term accounting for the uncertainty on the second-stage dependent variable can easily represent more than 25% of the estimated variance. Lastly, the coefficient estimates obtained in two stages are as accurate as those obtained by a direct estimation of the model except when groups are badly interconnected.

The rest of the paper is as follows. We introduce the model in section 2. We present the estimation method in section 3. We give some properties of the estimators in section 4. We discuss some extensions and limits of the estimation method in section 5. We study the effect of the group mobility pattern on the accuracy of the estimators for two simple examples in section 6. We report some Monte Carlo simulation results for more complex configurations in section 7. Finally, section 8 concludes.

2 The model

The two-stage method is presented for balanced panel data but the extension to the unbalanced case is straightforward. We consider a model of the form:

$$y_{i,t} = x_{i,t}\alpha + z_{g(i,t),t}\gamma + \eta_{g(i,t),t} + u_i + \varepsilon_{i,t}$$

for $i = 1, ..., N$ and $t = 1, ..., T$. In this equation, $x_{i,t}$ is a $1 \times L$ vector of individual time-varying characteristics, u_i an individual fixed effects, and $\varepsilon_{i,t}$ an individual error term. We denote $g(i,t)$ the group to which the individual i belongs at time t. For $g = 1, ..., G$, $z_{g,t}$ is a $1 \times K$ vector of group-year characteristics including the scalar one, and $\eta_{g,t}$ is a group error term.

For an individual variable $s_{i,t}$, we denote S the stacked vector of the observations in the individual and time dimensions. Similarly, for an aggegate variable $q_{g,t}$, we denote Q the stacked vector in the group and time dimensions. The model in vector form is:

$$Y = X\alpha + FZ\gamma + F\eta + AU + \varepsilon \tag{1}$$

where F is the matrix associating to each individual her group at a given date, and A is the (non stochastic) matrix associating to each individual her fixed effect.

We introduce the following assumptions on error terms:

A1: $\varepsilon_{i,t}$ *i.i.d.*, $E(\varepsilon_{i,t} | \Phi) = 0$, $E(\varepsilon_{i,t}^2 | \Phi) = \kappa^2 < \infty$, $E(\varepsilon_{i,t}^4 | \Phi) = R < \infty$
A2: $\eta_{g,t}$ *i.i.d.*, $E(\eta_{g,t} | Z) = 0$, $E(\eta_{g,t}^2 | Z) = \sigma^2 < \infty$, $E(\eta_{g,t}^4 | Z) = Q < \infty$
A3: $E(\varepsilon | \eta, \Phi) = 0$

where $\Phi = \{X, F, Z\}$.

In particular, Assumptions *A1* and *A2* do not allow for endogeneity of some explanatory variables. They also rule out heteroskedasticity and autocorrelation of both individual and aggregate error terms. Finally, Assumption *A3* imposes that individual and group error terms are uncorrelated. All these issues are discussed in more details in Section 4.

An intuitive approach to estimate equation (1) is to compute the first-difference estimator or the within estimator because they allow to deal with individual fixed effects properly. However, the covariance matrix of the residuals rewritten in first-difference or in mean-difference is very complex as there exist some group error terms and some mobility of individuals between groups across time. As a consequence, the variance of group error terms and the standard errors of estimated coefficients cannot be computed.

It could be tempting to omit group error terms when estimating the model even if the standard errors of parameters are biased. This can lead in practice to an important underestimation of standard errors (see Moulton, 1990; Pepper, 2002). Hence, group error terms must be taken into account. We show that a two-stage method can deal with this issue.

The basic idea is to replace all aggregate terms in model (1) by group-year fixed effects. The resulting equation is estimated after individual fixed effects have been eliminated by mean-difference. This first-stage regression allows to recover some coefficient estimates for individual variables as well as their standard errors. The estimated group-year fixed effects are then regressed on group variables. This second regression provides some coefficient estimates for group variables. We show in next sections that it is possible to compute some unbiased and consistent estimates of their standard errors.

3 The estimation method

More formally, we specify a group-year fixed effect $\beta_{a,t}$ as the sum of all aggregate terms. This group-year fixed effect is then introduced in equation (1). The model rewrites in vector form:

$$Y = X\alpha + F\beta + AU + \varepsilon \tag{2}$$

$$\beta = Z\gamma + \eta \tag{3}$$

If all groups are properly interconnected by inflows or outflows of movers across time, the group year fixed-effects β are identified if one identifying restriction only is imposed to the model (see Abowd, Kramarz and Margolis, 1999; Combes, Duranton and Gobillon, 2003). We suppose that this property holds in the sequel and we impose $\beta_{1,1} = 0$. We explain in Appendix A how to adapt the analysis when there is no mobility or groups are not well interconnected by flows of movers.

The two-stage method can be described as follows (the properties of the estimators are detailed in the next subsection):

1. We first compute the within-individual estimator of the parameters in the first stage equation. We denote $\widehat{\beta} = BY$ with $B = \left(0, FM_H \left(F'M_H F\right)^{-1}\right)'$, the estimator of the group-year fixed effects ($\widehat{\beta}_{1,1}$ being fixed to zero for convenience), where $H = (X, A)$ and $M_H = I - H\left(H'H\right)^{-1} H'$.

2. Introducing $\Psi = \left(0, \left(\widehat{\beta} - \beta\right)'\right)'$ the uncertainty on the group-year fixed effects, the second-stage equation rewrites: $\widehat{\beta} = Z\gamma + \eta + \Psi$. We denote $\widehat{\gamma}_{OLS}$ the OLS estimator of γ derived from this equation.

3. It is then possible, under some assumptions (see next section), to compute an unbiased and consistent estimator of the variance of group error terms:

$$\widehat{\sigma}^2 = \frac{1}{GT - K}\left[\left(\widehat{\eta + \Psi}\right)'\left(\widehat{\eta + \Psi}\right) - tr\left[M_Z \widehat{V}(\widehat{\beta}\,|\Phi)\right]\right] \tag{4}$$

where $M_Z = I - Z\left(Z'Z\right)^{-1} Z'$, $\widehat{\eta + \Psi} = \widehat{\beta} - Z\widehat{\gamma} = M_Z(\eta + \Psi)$, and $\widehat{V}(\widehat{\beta}\,|\Phi\,)$ is the estimator of the covariance matrix of group-year fixed effects derived from the first stage bordered with a first line and a first column of zeros.

4. We then obtain an unbiased estimator of the covariance matrix $\Omega = V(\eta + \Psi\,|\Phi\,)$:

$$\widehat{\Omega} = \widehat{\sigma}^2 I + \widehat{V}(\widehat{\beta}\,|\Phi\,) \tag{5}$$

5. It can be used to compute an unbiased estimator of the covariance matrix of $\widehat{\gamma}_{OLS}$:[3]

$$\widehat{V}\left(\widehat{\gamma}_{OLS}\,|\Phi\,\right) = (Z'Z)^{-1} Z'\widehat{\Omega}Z\left(Z'Z\right)^{-1} \tag{6}$$

6. It is also possible to construct the second-stage general least square estimator of γ:

$$\widehat{\gamma}_{GLS} = \left(Z'\Omega^{-1}Z\right)^{-1} Z'\Omega^{-1}\widehat{\beta} \tag{7}$$

   Its conditional variance is given by:

$$V\left(\widehat{\gamma}_{GLS}\,|\Phi\,\right) = \left(Z'\Omega^{-1}Z\right)^{-1} \tag{8}$$

7. A feasible general least square estimator $\widehat{\gamma}_{FGLS}$ is obtained replacing Ω by $\widehat{\Omega}$ in equation (7).

It is easy to compute the estimator $\widehat{\sigma}^2$ since this requires only the covariance matrix of parameters obtained in first stage, the second-stage residuals, and the projector in the dimension orthogonal to group variables.
Note that the two-stage estimation method can be implemented when individual effects are random (see Appendix A).

Compared to a direct first-difference or within estimation, the two-stage method has two advantages. First, it is possible to take into account the aggregate random terms properly when computing standard errors. Second, it allows to conduct a variance analysis at the aggregate level using the second-stage equation. However, it cannot be applied when the number of groups is large as when studying the impact of firms fixed effects on individual wages (see Abowd, Kramarz and Margolis, 1999). Indeed, it cannot deal with too many group-year fixed effects in the first stage. More details on the properties of the estimators are given in the next section

[3] Strictly speaking, the formula is true for all coefficients but the constant that is derived from the normalization of $\beta_{1,1}$ to zero for each sample drawn (and is thus random). This issue is ignored in all formulas for readability. The constant is not identified anyway without any restrictions to the model.

4 Some properties of the estimators

In this section, we give some sufficient conditions for some second-stage estimators to be unbiased and consistent when N and G go to infinity simultaneously (T fixed). We also derive the asymptotic distribution of the estimated coefficients of group explanatory variables. All properties are given for stochastic regressors. The non-stochastic case is similar. We first focus on the estimator of the variance of group error terms and show its unbiasedness:

Property 1 *Under A2-A3, $\widehat{\sigma}^2$ is an unbiased estimator of σ^2.*
Proof: see Appendix B. ■

We now give some sufficient conditions on the mobility pattern of individuals between groups for the estimated variance of group error terms to be consistent. First, we consider that there exists a rule that links N to G, when G tends to infinity. This rule is supposed to verify the following assumption:

A4: there exists a strictly increasing function $f(\bullet)$ such that $\forall G > 0$, $N = f(G)$ with $\frac{G}{f(G)} \underset{G\to+\infty}{\rightarrow} 0$.

This condition ensures that the number of regressors in the first-stage equation becomes negligible compared to the sample size when the number of groups and individuals tend to infinity. It also reduces the convergence problem to a one-dimension issue. In the sequel, we consider that the condition is verified but do not replace N by $f(G)$ for a better readability. The key assumption for convergence, which ensures that the estimators of group-year fixed effects are estimated with enough accuracy when N and G tend to infinity, is then:

A5: $\frac{1}{GT} E\left[tr\left(F' M_H F\right)^{-1}\right] \underset{G\to+\infty}{\rightarrow} 0$ and $E\left[\frac{1}{GT} tr\left(F' M_H F\right)^{-1}\right]^2 \underset{G\to+\infty}{\rightarrow} 0.$

The second asymptotic condition in *A5* is technical and has to be made because regressors are stochastic. When the regressors are non stochastic, the two asymptotic conditions collapse into one only. It is straightforward to show that assumption *A5* is equivalent to:

$$\frac{1}{GT} E trV\left(\widehat{\beta}\,|\Phi\right) \underset{G\to+\infty}{\rightarrow} 0;\; E\left[\frac{1}{GT} trV\left(\widehat{\beta}\,|\Phi\right)\right]^2 \underset{G\to+\infty}{\rightarrow} 0 \qquad (9)$$

This suggests an empirical diagnosis for assumption *A5* to be approached by the data at finite distance when N and G are large. Indeed, estimates are in line with this assumption if the average variance of group-year fixed effects given by the within estimation of equation (2) is small. This is very intuitive since it means that on average the uncertainty on the second-stage dependent variable is small. Interestingly, even if the variance of some group-year fixed effects tends to infinity because all the groups are not well interconnected to the others, assumption *A5* may still hold. We then prove the following consistency property:

Property 2 *Under assumptions A1-A5, we have:* $\widehat{\sigma}^2 \underset{G\rightarrow+\infty}{\overset{P}{\rightarrow}} \sigma^2$ *(T* fixed).

Proof: see Appendix B. ■

We now turn to the properties of the estimated coefficients of the group explanatory variables. First note that, using the expression of $\widehat{\beta}$ as well as equations (2) and (3), we can write:

$$\widehat{\gamma}_{OLS} = \gamma + \left(Z'Z\right)^{-1} Z'\eta + \left(Z'Z\right)^{-1} Z'B\varepsilon \tag{10}$$

Thus, the OLS estimator of γ can be decomposed into the real parameter value, a component due to the use of a finite sample of groups, and a component that accounts for the uncertainty on the dependent variable. A similar decomposition holds for the GLS estimator that writes:

$$\widehat{\gamma}_{GLS} = \gamma + \left(Z'\Omega^{-1}Z\right)^{-1} Z'\Omega^{-1}\eta + \left(Z'\Omega^{-1}Z\right)^{-1} Z'\Omega^{-1}B\varepsilon \tag{11}$$

Property 3 *Under A1-A2,* $\widehat{\gamma}_{OLS}$ and $\widehat{\gamma}_{GLS}$ *are unbiased estimators of* γ.

Proof: It is straightforward using equations (10) and (11). ■

The property of unbiasedness does not hold for the FGLS estimator as usual in the literature because the estimated variance $\widehat{\Omega}$ enters the formula and is correlated with error terms. We now turn to asymptotic properties. To a given matrix v, associate $|v|$ the matrix where all terms of v are taken in absolute value. We need some further assumptions to prove the consistency of $\widehat{\gamma}_{OLS}$, $\widehat{\gamma}_{GLS}$ and $\widehat{\gamma}_{FGLS}$:

A6a: $\frac{Z'Z}{GT} \underset{G\rightarrow+\infty}{\overset{P}{\rightarrow}} Q_0$ *is finite and definite positive; there exists* θ_1 *such that all the elements of* $|Z|$ *are inferior to* θ_1 *for all* G; $\frac{1}{GT} trE\left[\left(F'M_HF\right)^{-1}\right] \underset{G\rightarrow+\infty}{\rightarrow} 0$.

A6b: $\frac{Z'\Omega^{-1}Z}{GT} \underset{G\rightarrow+\infty}{\overset{P}{\rightarrow}} Q_2$ *is finite and definite positive; there exists* θ_2 *such that all the elements of* $|Z|$ *are inferior to* θ_2 *for all* G; $\frac{1}{GT} trE\left(\Omega^{-1}\right) \underset{G\rightarrow+\infty}{\rightarrow} 0$.

A6c: $\frac{Z'\widehat{\Omega}^{-1}Z}{GT} \underset{G\rightarrow+\infty}{\overset{P}{\rightarrow}} Q_3$ *is finite and definite positive; there exists* θ_3 *such that all the elements of* $|Z|$ *are inferior to* θ_3 *for all* G; $\frac{1}{GT} trE\left(\widehat{\Omega}^{-1}\right) \underset{G\rightarrow+\infty}{\rightarrow} 0$.

In particular, these assumptions ensure that the uncertainty on the second-stage dependent variable becomes negligible in the expressions of the estimators of γ when G tends to infinity. It is then possible to apply a Chebychev's weak law of large numbers for triangular arrays (see Borovkov, 1998) to prove the consistency of the estimators. We have the following property:

Property 4 *Suppose that A1,A2,A4 are verified. We have:*
Under A6a, $\widehat{\gamma}_{OLS} \underset{G\to+\infty}{\overset{P}{\to}} \gamma$. *Under A6b,* $\widehat{\gamma}_{GLS} \underset{G\to+\infty}{\overset{P}{\to}} \gamma$.
Under A6c, $\widehat{\gamma}_{FGLS} \underset{G\to+\infty}{\overset{P}{\to}} \gamma$.

Proof: see Appendix B. ■

We now determine the limit distribution of $\widehat{\gamma}_{OLS}$, $\widehat{\gamma}_{GLS}$ and $\widehat{\gamma}_{FGLS}$. Recall that $B = \left(0, FM_H\left(F'M_HF\right)^{-1}\right)'$. Introduce ℓ a $GT \times 1$ vector of ones. We need the following assumptions:

A7a: There exists δ_1 *such that all the elements of* $\ell'.\left|B\right|$ *are bounded by* δ_1 *for all G.* $\frac{1}{GT}Z'\left(F'M_HF\right)^{-1}Z \underset{G\to+\infty}{\overset{P}{\to}} Q_1$ *is finite and definite positive.*
A7b: There exists δ_2 *such that all the elements of* $\ell'.\left|\Omega^{-1}B\right|$ *are bounded by* δ_2 *for all G.*
A7c: There exists δ_3 *such that all the elements of* $\ell'.\left|\widehat{\Omega}^{-1}B\right|$ *are bounded by* δ_3 *for all G.*

According to equation (10), the second stage OLS, GLS and FGLS estimators can be rewritten as a linear combination of the group and individual error terms. Assumptions *A7a-A7c* ensure that the contribution of every individual error term to the second-stage estimators is negligible when G tends to infinity. It is then possible to apply a central limit theorem for triangular arrays in the multivariate case (see Borovkov, 1998). We have:

Property 5 *Under A1-A4,* A6a *and A7a*:

$$\sqrt{GT}\left(\widehat{\gamma}_{OLS}-\gamma\right) \underset{G\longrightarrow+\infty}{\overset{L}{\to}} N\left(0, \sigma^2 Q_0^{-1} + \kappa^2 Q_0^{-1} Q_1 Q_0^{-1}\right) \tag{12}$$

Under A1-A4, *A6b* and *A7b*:

$$\sqrt{GT}\left(\widehat{\gamma}_{GLS}-\gamma\right) \underset{G\longrightarrow+\infty}{\overset{L}{\to}} N\left(0, Q_2^{-1}\right) \tag{13}$$

Under assumptions A1-A4, *A6c* and *A7c*:

$$\sqrt{GT}\left(\widehat{\gamma}_{FGLS}-\gamma\right) \underset{G\longrightarrow+\infty}{\overset{L}{\to}} N\left(0, Q_3^{-1}\right) \tag{14}$$

Proof: see Appendix. ■

When stated in the case of non-stochastic regressors, Property 5 provides a means to conduct tests conditionally to the data.

5 Extensions and limits

In this section, we discuss some extensions of the two-stage method. We explain how heteroskedasticity and/or autocorrelation of individual and/or group error terms can be taken into account when computing the estimators and their standard errors. We also show how the method can be extended if one wants to instrument some individual or group variables because they are potentially endogenous. We then present some limits of the two-stage method. We study the unbiasedness properties of some estimators when the allocation process of individuals between groups is no longer exogenous (the equality $E(\varepsilon, \eta | F) = 0$ coming from Assumptions *A1* and *A2* no longer holds). We also examine these unbiasedness properties when a correlation between the individual and group error terms is allowed (Assumption *A3*: $E(\varepsilon | \eta, \Phi) = 0$ is not imposed anymore). We compare the results with those obtained when the model is estimated directly in one stage only.

5.1 Heteroskedasticity and autocorrelation

In some cases, heteroskedasticity and/or autocorrelation of individual and/or group error terms should be accounted for to avoid some bias on the estimated standard errors. This issue, ignored for a long time in the empirical literature, has been explicitely recognized more recently. For instance, it has been shown that ignoring heteroskedasticity and autocorrelation of individual error terms can lead to highly biased standard errors when studying policies targeted on some subgroups of the population (see Bertrand, Duflo and Mullainathan, 2004). It is possible to extend the two-stage method to compute standard errors robust to heteroskedasticity and/or autocorrelation of individual and/or group error terms. Indeed, heteroskedasticity and/or autocorrelation of individual error terms can be taken into account in the first stage regression, whereas heteroskedasticity and/or autocorrelation of group error terms can be dealt with in the second-stage regression.

More formally, rewrite the equation (2) with vectors at the individual level:

$$Y_i = X_i \alpha + F_i \beta + u_i \ell_T + \varepsilon_i \tag{15}$$

where $Y_i = (y_{i,1}, ..., y_{i,T})'$, $X_i = \left(x'_{i,1}, ..., x'_{i,T}\right)'$, $\ell_T = (1, ..., 1)'$, F_i is the $T \times GT$ matrix associating to individual i her group in all years, and $\varepsilon_i = (\varepsilon_{i,1}, ..., \varepsilon_{i,T})'$. For a given variable v, denote $\widetilde{v}$ its projection in the within dimension. Consider the four cases: (0) homoskedasticity and no autocorrelation; (1) heteroskedasticity (across individuals *but not* time) and no autocorrelation; (2) homoskedasticity and autocorrelation; (3) heteroskedasticity (across individuals *and* time) and autocorrelation. The conditional covariance matrix of coefficients estimated in first stage writes $\left(\widetilde{H}'\widetilde{H}\right)^{-1} \Gamma \left(\widetilde{H}'\widetilde{H}\right)^{-1}$, where $H = (X, F)$ and Γ is a matrix depending on the case that is considered:

$$
\begin{aligned}
(0)\ \Gamma &= \kappa^2 \widetilde{H}' \widetilde{H} \\
(1)\ \Gamma &= \sum_i \kappa_i^2 \widetilde{H}_i' \widetilde{H}_i \\
(2)\ \Gamma &= \sum_i \widetilde{H}_i' \Lambda \widetilde{H}_i \\
(3)\ \Gamma &= \sum_i \widetilde{H}_i' \Lambda_i \widetilde{H}_i
\end{aligned}
$$

where $H_i = (X_i, F_i)$, κ_i^2 is the variance of the individual shock for i in case (1); Λ is the covariance matrix of individual shocks common to all i in case (2); and Λ_i is the covariance matrix of individual shocks for i in case (3).

In each case, supposing that G is fixed, an estimator of $\frac{1}{NT}\Gamma$ consistent when N tends to infinity can easily be constructed using the estimated shocks $\widehat{\widetilde{\varepsilon}}_i = \left(\widehat{\widetilde{\varepsilon}}_{i,1}, ..., \widehat{\widetilde{\varepsilon}}_{i,T}\right)'$ from the within regression (see Kezdi, 2002). In case (0), κ^2 should be replaced by $\frac{1}{N(T-1)}\widehat{\widetilde{\varepsilon}}'\widehat{\widetilde{\varepsilon}}$. In case (1), each term κ_i^2 should be replaced by $\frac{1}{T-1}\widehat{\widetilde{\varepsilon}}_i'\widehat{\widetilde{\varepsilon}}_i$ (this is a slight modification of White, 1980). In case (2), Λ should be replaced by $\frac{1}{N}\sum_i \widehat{\widetilde{\varepsilon}}_i\widehat{\widetilde{\varepsilon}}_i'$ (see Kiefer, 1980). In case (3), Λ_i should be replaced by $\widehat{\widetilde{\varepsilon}}_i\widehat{\widetilde{\varepsilon}}_i'$ (see Arellano, 1987).[4]

If there is no heteroskedasticity and no autocorrelation for group error terms, the two-stage estimation method proposed in section 2 can be applied directly without any change in the formulas, except for the covariance matrix of the estimated group-year fixed effects. The unbiasedness property of $\widehat{\sigma}^2$ still holds. Provided that assumptions *A5* to *A7* are modified, this is also the case for the consistency and distribution properties.

If there is heteroskedasticity and/or autocorrelation of group error terms, it is possible to construct robust estimators of the second-stage standard errors (whether individual error terms are heteroskedastic and/or autocorrelated or not). The approach is quite similar to that used in first stage except that there is some uncertainty with known variance on the dependent variable that must be taken into account. Consider the following four cases characterizing the group error terms: (0) homoskedasticity and no autocorrelation; (1) heteroskedasticity (across groups *but not* time) and no autocorrelation; (1') heteroskedasticity (across groups *and* time) and no autocorrelation; (2) homoskedasticity and autocorrelation; (3) heteroskedasticity (across groups *and* time) and autocorrelation. The conditional covariance matrix of the second-stage OLS estimator

[4] Tests for heteroskedasticity and autocorrelation are given by Kezdi (2002); Inoue and Solon (2004).

writes $V\left(\widehat{\gamma}_{OLS}\left|\Phi\right.\right)=\left(Z'Z\right)^{-1}\left[Z'V\left(\widehat{\beta}\left|\Phi\right.\right)Z+\Pi\right]\left(Z'Z\right)^{-1}$ where Π is a matrix depending on the case:

$$
\begin{aligned}
(0)\ \Pi &= \sigma^2 Z'Z \\
(1)\ \Pi &= \sum_g \sigma_g^2 Z_g' Z_g \\
(1')\ \Pi &= \sum_{g,t} \sigma_{g,t}^2 Z_{g,t}' Z_{g,t} \\
(2)\ \Pi &= \sum_g Z_g' \Sigma Z_g \\
(3)\ \Pi &= \sum_g Z_g' \Sigma_g Z_g
\end{aligned}
$$

with σ_g^2 the variance of group shocks for g in case (1); $\sigma_{g,t}^2$ the variance of group shock for g in year t in case (1'); Σ the covariance matrix of group shocks common to all g in case (2); and Σ_g the covariance matrix of group shocks for g in case (3).[5]

In each case, an estimator of $\frac{1}{GT}\Pi$ consistent when N and G tend to infinity simultaneously can easily be constructed using the estimated individual shocks $\widehat{\eta+\Psi}_g$ from the OLS regression where $\widehat{\eta+\Psi}_g=\left(\widehat{\eta+\Psi}_{g,1},...,\widehat{\eta+\Psi}_{g,T}\right)'$. In case (0), it is possible to construct an estimator of σ^2 that is not only consistent, but also unbiased (see Section 4). In case (1), each term σ_g^2 should be replaced by $\left[\frac{1}{T}\widehat{\eta+\Psi}_g'\widehat{\eta+\Psi}_g-trV\left(\widehat{\beta}_g\left|\Phi\right.\right)\right]$ where $\widehat{\beta}_g=\left(\widehat{\beta}_{g,1},...,\widehat{\beta}_{g,T}\right)'$. In case (1'), each term $\sigma_{g,t}^2$ should be replaced by $\left[\widehat{\eta+\Psi}_{g,t}^2-V\left(\widehat{\beta}_{g,t}\left|\Phi\right.\right)\right]$. In case (2), Σ should be replaced by $\frac{1}{G}\sum_g\left[\widehat{\eta+\Psi}_g\widehat{\eta+\Psi}_g'-V\left(\widehat{\beta}_g\left|\Phi\right.\right)\right]$. In case (3), Σ_g should be replaced by $\left[\widehat{\eta+\Psi}_g\widehat{\eta+\Psi}_g'-V\left(\widehat{\beta}_g\left|\Phi\right.\right)\right]$.

We have shown how to deal with heteroskedasticity and/or autocorrelation of group error terms. Researchers may also want to take into account spatial autocorrelation if groups are defined as countries or regions. This can be done in two different ways.[6] First, it is possible to introduce some group fixed effects in

[5] The conditional covariance matrix of the second-stage GLS estimator writes: $V\left(\widehat{\gamma}_{GLS}\left|\Phi\right.\right)=\left[Z'\left[V\left(\widehat{\beta}\left|\Phi\right.\right)+\Pi_0\right]^{-1}Z\right]^{-1}$ where Π_0 is a matrix that depends on the case: (0) $\Pi_0=\sigma^2 I$; (1) $\Pi_0=diag\left(\sigma_1^2 I_T,...,\sigma_G^2 I_T\right)$ where I_T is the $T\times T$ identity matrix; (1') $\Pi_0=diag\left(\sigma_{1,1}^2,...,\sigma_{G,T}^2\right)$; (2) $\Pi_0=diag\left(\Sigma,...,\Sigma\right)$; (3) $\Pi_0=diag\left(\Sigma_1,...,\Sigma_G\right)$. An estimator of the conditional covariance matrix can be constructed replacing unknown quantities by their empirical counterpart computed from OLS (see next paragraph in the main text).

[6] See Case (1991) for a discussion on the advantages and drawbacks of the two approaches.

equation (3). Coefficients of aggregate variables can be estimated after projecting this equation in the within-group dimension.[7] Second, if T is quite large, G is small, and one believes that there is no time autocorrelation, it is possible to compute consistent standard errors with a method in the same spirit as the approach used above for cases (2) and (3). Indeed, consider the cases: (2b) homoskedasticity and spatial autocorrelation; (3b) heteroskedasticity (across groups *and* time) and spatial autocorrelation. The Π matrix shoud rewrite in the two cases:

$$\begin{aligned}(2b)\ \Pi &= \sum_t Z_t'\Gamma Z_t \\ (3b)\ \Pi &= \sum_t Z_t\Gamma_t Z_t\end{aligned}$$

where $Z_t = (Z_{1,t}, ..., Z_{G,t})'$, Γ is the covariance matrix common to all years in case (2b), Γ_t is the covariance matrix for year t in case (3b).
Both approaches allow to take spatial autocorrelation into account without estimating the coefficients of a spatial underlying process as it is often done in the literature (See Cressie, 1993; Anselin and Florax, 2002).

5.2 Endogeneity issues

There are several types of endogeneity problems that can arise:

1) Some individual explanatory variables are correlated with group error terms.
2) Some individual explanatory variables are correlated with individual error terms.
3) Some group explanatory variables are correlated with group error terms.
4) The group choice of individuals is correlated with group error terms.
5) The group choice of individuals is correlated with individual error terms.

We first discuss the endogeneity problems 1), 2), and 3). The discussion applies even if there exists heteroskedasticity and/or autocorrelation of individual and/or group error terms.
The estimated parameters are robust to the first endogeneity issue. Indeed, all terms at the group level are replaced by group-year fixed effects that are estimated jointly with the coefficients of individual explanatory variables. Thus, even if group error terms are correlated with some individual explanatory variables, it is not the case of the residuals in the first stage regression. Note that if the model was estimated directly in one stage only, the estimated parameters would be biased as the group error terms would enter the first-stage residuals.
The second endogeneity issue can be handled if there exist some instruments for the individual variables. In that case, the first-stage estimation becomes

[7] Note that it is still possible to take into account heteroskedasticity and/or autocorrelation of group error terms when group fixed effects are introduced in equation (3). This can be done using the same kind of formulas as in first stage when heteroskedasticity and/or autocorrelation of individual error terms are accounted for in presence of individual fixed effects.

a standard 2SLS estimation for panel data. The second stage of the method remains nearly unchanged: formulas are the same except that the variance of group-year fixed effects now includes an additional component coming from instrumentation in first stage.
The third endogeneity issue can be tackled in second stage if there exist some instruments for group variables. In that case, equation (3) is estimated with 2SLS. It is still possible to use the formulas given in sections 2 and 3 to recover some consistent estimators of standard errors, except that an additional variance coming from instrumentation must be added to the variance of group-year fixed effects.[8]

The fourth and fifth endogeneity issues are discussed below in a more structural way, specifying the group-choice process of individuals. We do not try to take into account the group choice process in the estimations. Indeed, methods have been proposed for that purpose in cross-section[9], but they cannot be easily extended to panel models with individual fixed effects. Instead, we discuss the biases that can arise for the estimated coefficients. For simplicity, we consider that assumptions *A1-A3* are verified.
We consider now that the choice of individuals between groups is made on the basis of the expected outcome associated to each group conditional on the error terms which are observed. Denote $y_{i,g,t}$ the outcome that an individual could obtain if she was in group g in year t. It is supposed to verify the equation:

$$y_{i,g,t} = x_{i,t}\alpha + z_{g,t}\gamma + \eta_{g,t} + u_i + \xi_{i,g,t}$$

where $\xi_{i,g,t}$ is an individual-group error term. The individual error term introduced in equation (1) verifies $\varepsilon_{i,t} = \xi_{i,g(i,t),t}$. We examine which estimates are biased depending on the group choice process:

1. The individuals do not observe any error term. We have:

$$g(i,t) = \arg\max_g E_{\xi_{i,g,t},\eta_{g,t}}\left(y_{i,g,t}\right)$$

   Then, F is strictly exogenous. This case is in line with Assumptions *A1* and *A2*.

2. The individuals observe the group error terms but not the individual-group error terms. We then have:

$$g(i,t) = \arg\max_g E_{\xi_{i,g,t}}\left(y_{i,g,t}\right)$$

[8] For more on the endogeneity issues 1), 2) and 3), but in a cross-section setting, see Blundell and Windmeijer (1997); Rice, Andrew and Glodstein (2002).
[9] See Lee (1983); Durbin and Mc Fadden (1984); Dahl (2002); Bourginon, Fournier and Gurgand (2003).

The group choice of individuals then depends on the group error terms only and $F = F(\eta)$. When model (1) is estimated directly, the coefficients of all explanatory variables are biased because the group error term enters the random component. Interestingly, when the first stage of the method is performed, the estimated group-year fixed effects and the estimated coefficients of individual explanatory variables are not biased. Indeed, the random component in equation (2) only consists in the individual error term that is supposed to be uncorrelated with the group error term (see *A3*). Also, the coefficients of group variables obtained in second stage with OLS and GLS are unbiased. This arises because the group error terms enter additively with all the terms including F in the expression of the second-stage estimators (see formulas (10) and (11)). Finally, whereas the estimated variance of individual error terms is unbiased, the estimated variance of group error terms given by (4) is biased because of some multiplicative interactions between F and the group error terms. As a consequence, the estimated standard errors in second stage are biased.

3. The individuals observe the individual-group error term but not the group error terms. In that case, we have:

$$g(i,t) = \arg\max_g E_{\eta_{g,t}}(y_{i,g,t})$$

The group choice of individuals then depends on the individual error terms only and $F = F(\varepsilon)$.[10] When model (1) is estimated directly, FZ enters the set of explanatory variables and is correlated with the random component $\varepsilon + \eta$. Consequently, the estimated coefficients of all explanatory variables are biased. In the two-stage method, F also enters the set of explanatory variables in first stage and the random component is ε. Thus, the estimated group-year fixed effects and the estimated coefficients of individual explanatory variables are biased. Moreover, the uncertainty on the second-stage dependent variable is a multiplicative function of F and ε. As a consequence, the estimated coefficients of group variables are biased. Finally, the estimated variance of group error terms is biased as ε and F interact multiplicatively in its formula.

4. The individuals observe all individual-group error terms as well as all group error terms. In that case, we have:

$$g(i,t) = \arg\max_g (y_{i,g,t})$$

The group choice of individuals then depends on both types of error terms and $F = F(\varepsilon, \eta)$. The results given in the previous case also hold here.

Results on biases are summarized in Table 1:

[10] It also depends on the $\xi_{i,g,t}$, $g \neq g(i,t)$. However, these individual-group error terms are omitted from the function $F(\cdot)$ for readability.

Table 1: bias on parameters depending on the mobility process

	$\hat{\gamma}_{DW}$	$\hat{\gamma}_{OLS}$	$\hat{\gamma}_{GLS}$	$\hat{\gamma}_{FGLS}$	$\hat{\sigma}^2$	$\hat{\kappa}^2$
Case 1: All error terms unobserved	+ (-)	+ (+)	+ (+)	-* -*	+	+
Case 2: Group error terms observed	- (-)	+ (-)	+ (-)	-* (-)	+	+
Case 3: Individual error terms observed	- (-)	- (-)	- (-)	- (-)	-	-
Case 4: All error terms Observed	- (-)	- (-)	- (-)	- (-)	-	-

+: unbiased, -: biased, * (for FGLS): biased only because the variance of group error terms has been replaced by its estimator. The sign without parenthesis refers to the estimator. The sign in parenthesis refers to its variance. The direct *within* estimator of model (1) in one stage only is noted $\hat{\gamma}_{DW}$. Its variance is computed without accounting for group error terms. Thus, it is always biased.

5.3 Correlation between error terms

For cases 1) and 2), it is interesting to study how a correlation between individual and group error terms can change the results on biases. When these error terms are correlated, the group-year fixed effects and the estimated coefficients of individual explanatory variables obtained with the two-stage method are biased. This occurs because there is some sorting of individuals across groups according to the value of group-year fixed effects. Put differently, the expectation of individual error terms conditional on the group-year fixed effects is not zero. More formally, we have: $E\left(\widehat{\beta}\middle|\beta\right)-\beta=E_{F,H}\left[\left(F'M_HF\right)^{-1}F'M_HE\left(\varepsilon|F,H,\beta\right)\right]$.[11] As $cov\left(\varepsilon,\beta|F,H\right)=cov\left(\varepsilon,\eta|F,H\right)\neq 0$ because $E\left(\varepsilon|\eta\right)\neq 0$, we have $E\left(\varepsilon|F,H,\beta\right)\neq 0$, and thus $E\left(\widehat{\beta}\middle|\beta\right)\neq 0$. The same line of argument applies to show that the estimated coefficients of individual explanatory variables are biased. In comparison, model (1) is estimated directly, the existence of a bias on the estimated coefficients of individual explanatory variables depends on the assumption made on the group-choice process. If the process if strictly exogenous (case 1), these estimated coefficients are unbiased because the group error terms enter the residuals. However, if the group-choice process depends on the aggregate error terms (case 2), the estimated coefficients of individual variables are biased since FZ enters the set of explanatory variables.

Interestingly, the existence of a bias on the estimated coefficients of group explanatory variables obtained with the two-stage method depends on the as-

[11] All formulas given in other sections are also conditional on parameters even if it is not stated for notations to remain simple. In particular, the first-stage estimation is conducted conditionally on group-year fixed effects. We introduce some expectations conditional on parameters in this subsection because we need the conditionality to be explicit in the discussion.

sumptions made on the group-choice process. If F is strictly exogenous, the estimated coefficients are unbiased as we have:
$E(\widehat{\gamma}_{OLS})-\gamma = E_{Z,F,H}\left[(Z'Z)^{-1}Z'(F'M_H F)^{-1}F'M_H E(\varepsilon|F)\right]$ and $E(\varepsilon|F) = 0$. When $F = F(\eta)$, they are biased since $E(\varepsilon|F) \neq 0$. The same results hold when the model is estimated directly in one stage only. This can be shown using the same line of arguments as previously for individual explanatory variables.

6 Some analytical examples

In this section, we try to assess how the level of interconnection between groups can affect the consistency results for the estimated variance of group error terms. For that purpose, we study two particular cases characterized by different mobility patterns of individuals. The two cases are designed so that it is possible to compute a closed form for the estimators of group-year fixed effects. We derive from their expression, the speed at which N and G must increase relatively to each other for the assumption *A5* to be satisfied. All proofs are given in Appendix C.
For simplicity, we consider that there is no individual explanatory variable. We also suppose that all groups include the same number of individuals at each date: $n = N/G$ (with N proportional to G), and that $T = 2$.

1. In the first case, for each group g, $\frac{m}{G-1}$ individuals move to each other group $g' \neq g$ (m is supposed to be proportional to $G-1$) and $n-m$ stay in their group. In this configuration, the first-order conditions write:

$$\begin{aligned} n\beta_{g,2} - (n-m)\beta_{g,1} - \frac{m}{G-1}\sum_{z\neq g}\beta_{z,1} &= w_{g,2} \\ n\beta_{g,1} - (n-m)\beta_{g,2} - \frac{m}{G-1}\sum_{z\neq g}\beta_{z,2} &= w_{g,1} \end{aligned}$$

for all $g = 1,...,G$ and $t = 1,2$, with $t' = 3-t$ and $w_{g,t} = \sum_{i|g(i,t)=g}(y_{i,t} - y_{i,t'})$.

Using the identifying condition $\beta_{1,1} = 0$, we obtain:

$$\begin{aligned} \widehat{\beta}_{g,1} &= \frac{1}{n}\frac{\theta}{2\theta-1}\left[\theta(w_{g,1}-w_{1,1}) + (\theta-1)(w_{g,2}-w_{1,2})\right] \\ \widehat{\beta}_{g,2} &= w_g + (n-m)\widehat{\beta}_{g,1} + \frac{m}{G-1}\sum_{z\neq g}\widehat{\beta}_{z,1} \end{aligned}$$

with $\theta = \frac{G-1}{G}\frac{n}{m}$.

We can first notice that estimated group-year fixed effects are the same for all groups in each year. This is not surprising since the position of all groups relatively to the reference group ($g = 1$) is similar. We can study how the variance of estimated group fixed effects in each year varies depending on the number of movers. For $G > 3$, this variance is minimum in each year when all individuals move from their group ($m = n$). As soon as the number of movers is proportional to the number of individuals per group (i.e. $m = \mu n$ with $\mu > 0$ for all G), we have:

$$\frac{1}{GT} trV\left(\widehat{\beta}\right) = \frac{4 + 2\mu^2 (1-\mu)}{\mu^2 (2-\mu)} \sigma^2 \frac{G}{N} + o\left(\frac{G}{N}\right)$$

The number of individuals must converge at a rate at least equal to $G^{1+\lambda}$, $\lambda > 0$, for Assumption *A5* to hold.

2. In the second case, for each group g, m individuals move to group $g+1$ (except when $g = G$, in which case they move to group $g = 1$) and $n-m$ individuals stay in their group. In this configuration, the first-order conditions write:

$$\begin{aligned} n\beta_{g,2} - m\beta_{g-1,1} - (n-m)\beta_{g,1} &= w_{g,2} \\ n\beta_{g,1} - m\beta_{g+1,2} - (n-m)\beta_{g,2} &= w_{g,1} \end{aligned}$$

Using the identifying restriction $\beta_{1,1} = 0$, we obtain:

$$\begin{aligned} \widehat{\beta}_{g,1} &= \frac{1}{G}\sum_{b=1}^{g-1} (b-1)(G-g+1)\overline{w}_{b,1} + \frac{1}{G}\sum_{b=g}^{G} (g-1)(G-b+1)\overline{w}_{b,1} \\ \widehat{\beta}_{g,2} &= \frac{1}{n}\left[w_{g,2} + m\widehat{\beta}_{g-1,1} + (n-m)\widehat{\beta}_{g,1}\right] \end{aligned}$$

with $\overline{w}_{b,1} = \frac{1}{n-1}\left[nw_{b,1} + mw_{b+1,2} + (n-m)w_{b,2}\right]$.

In that case, the value of each estimated group fixed effect depends on the position of the group-year relative to the reference group 1. This is also the case for the variance of each estimated group-year fixed effect. In fact, the maximum variance in year 1 corresponds to the group $g = \frac{G}{2} + 1$ (for G even). This is not surprising as the corresponding group is the furthest away from the reference group 1. We study how the variance of each estimated group fixed effect in year 1 varies depending on the number of movers. We find that each variance is minimum when half the individuals move to the next group. To derive some simple sufficient conditions for Assumption *A5* to be verified, we write that the number of movers is proportional to the number of individuals per group with a factor μ (i.e. $m = \mu n$ with $\mu > 0$ for all G). We have:

$$\frac{1}{GT} trV\left(\widehat{\beta}\right) = \frac{1}{45} \frac{\sigma^2}{\mu(1-\mu)} \frac{G^4}{N} + o\left(\frac{G^4}{N}\right)$$

The number of individuals must converge at a rate at least equal to $G^{4+\lambda}$, $\lambda > 0$, for assumption *A5* to hold. This rate is far more important than in case 1. One could argue that the multiplying constant of the leading term is far larger in case 1 than in case 2, especially when the migration intensity is small. However, for instance when $\mu = 10\%$, the leading term is more important in case 1 only when $G < 10$ whereas the reverse is true for some higher G.

In conclusion, these two cases suggest that Assumption *A5* should be verified when G and N tend to infinity with N reasonably higher than G if the mobility pattern allows all groups to be well interconnected by flows of movers. A deeper insight of the link between the inter-group mobility pattern and the estimators is given in next section with Monte-Carlo simulations.

7 Monte Carlo results

7.1 Simulations with an exogenous mobility pattern

In this section, we conduct some simulations to assess how the inter-group mobility pattern affects the accuracy of the estimators. We also compute the bias on standard errors that the two-stage method allows to avoid compared to a direct estimation of the model that does not take into account aggregate error terms. For simplicity, we restrict the analysis to the setting given in section 2 where error terms are *iid*.

We focus on mobility patterns in which individuals move at most once to another group during the T years. As in previous section, we suppose that N is proportional to G, and that individuals are equally distributed across groups. Each group contains $n = N/G$ individuals and has the same number of out-movers in each year, noted m. We impose the restriction $n \geqslant mT$ so that there are enough individuals in each group for moves to occur in all years. The out-movers go to d destinations, d being the same for all groups, with $d < G$.

The mobility process is the following: in year 1, all individuals are affected to a group of origin. Then, $\frac{m}{d}$ individuals move from each group g to one of the next d groups between years 1 and 2 (m being supposed to be proportional to d). Once the individuals have moved, they stay in their group of destination until the end of the period. The process is renewed for each next pair of years.

We now describe the simulation procedure. We consider that there exist only two aggregate variables: the vector one accounting for the constant and another variable whose values are drawn independently in a uniform law $[-1, 1]$. Its coefficient is fixed to 1. We draw the group and individual error terms in some centered normal laws with variances equal to 1 and 15, respectively.[12] For Assumption *A3* to be verified, the two types of error terms must be drawn independently. However, we will sometimes allow for a correlation ρ between them to test the robustness of the results to a misspecification of the model. We then construct some group-year fixed effects as the sum of the effect of

[12] The relative order of magnitude of the parameters are fixed according to the empirical results obtained by Combes, Duranton and Gobillon (2003) on the effect of density on individual wages.

aggregate explanatory variables and the group error terms. The value of the constant derives from the normalization to zero of the first group fixed effect in the first year .

We proceed to 1000 simulations of the model for different numbers of groups and different mobility patterns. We report the median and mean estimated variance of individual and group error terms. We assess the importance of the corrective term $\frac{1}{GT-K}tr\left[M_Z\widehat{V}(\widehat{\beta}\,|\Phi\,)\right]$ for the median estimated variance of group error terms by computing its ratio with the variance. We then report the mean estimator of the aggregate variable coefficient when OLS and FGLS are used, as well as the median estimator and its standard error, the median standard error, the root of the mean estimated variance and the RMSE. We also give for the direct within estimation of the model: the mean estimator, the median estimator and its standard error uncorrected for the existence of group error terms, the uncorrected median standard error, the uncorrected root of the mean estimated variance and the RMSE. Results are reported in Table 2.

We now comment the results for $G = 50$ groups when $N = 10000$ and $T = 2$, for different mobility patterns and correlations between individual and group error terms.

We first consider the benchmark case in column (1) for the mobility pattern in which 100 individuals in each group move to the 5 next groups (20 movers per group). We find that the corrective term used to compute the median variance is quite important as it constitutes 25% of this variance. The variance of group error terms and the coefficients of the explanatory variable are estimated with a reasonable accuracy, there $RMSE$ being nearly 0.20. We can also note that the median standard error computed when the model is estimated directly by OLS is 2.5 times lower than its $RMSE$ (0.07 against 0.20). This arises from the effect described by Moulton (1990): when the aggregate error terms are not taken into account in the computation of standard errors, the latter can be highly biased.

We then change the mobility pattern, allowing for the migration of 100 individuals to the next group only. The estimates are reported in column (2). The variance of group error terms is estimated with a very bad accuracy and its $RMSE$ is huge (1.56). In 28.4% of the simulations, the estimated variance is even negative. The corrective term is very important as it accounts for 81% of the median estimated variance. These results are in line with the analysis of case 2 in the previous section. However, the different estimators of the explanatory variable coefficient still perform well. The $RMSE$ is very similar to the previous case for the direct within estimation (0.22) and is slightly higher for the two-stage OLS (0.33).

We then focus on the mobility pattern in which there exists a high number of destinations, with two individuals moving to each of the next 49 groups. Results are reported in column (3). In that case, the variance of group error terms is estimated with more accuracy than in the benchmark case, the $RMSE$ decreasing from 0.23 to 0.17. The corrective term is also less important, as it now constitutes only 16.7% of the median estimated variance. However, the coefficient of the explanatory variable is estimated with a similar accuracy whatever estimation method is used.

We then “misspecify” the model, allowing for a correlation of 0.2 between the individual and group error terms (see column 4). The estimator of the group variance is now biased, the mean value (3.1) being far above one. This is not surprising as the unbiasedness of this estimator is obtained only when the individual and group error terms are independent (see Proof of Property 1). However, as expected (see Property 2), the direct within estimator and the two-stage OLS estimator are still unbiased. The FGLS estimator also behaves nicely as its value is only slightly biased. Finally, the accuracy of the parameter estimates is less good than in the benchmark case. When the correlation between error terms increases to 0.5 (column 5), the mean estimated variance of group error terms goes up even more, taking the value 8.7, and is thus very biased.

Table 2: Simulation results when the mobility process is exogenous

	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
Parameters					
S	1000	1000	1000	1000	1000
Variance of individual error term (κ^2)	15	15	15	15	15
Variance of group error term (σ^2)	1	1	1	1	1
Group variable coefficient (γ)	1	1	1	1	1
Number of groups (Z)	50	50	50	50	50
Number of individuals (N)	10000	10000	10000	10000	10000
Number of periods (T)	2	2	2	2	2
Number of destinations (ZM)	5	1	50	5	5
Number of movers per group (NM)	20	100	2	20	20
Correlation between error terms (ρ)	0	0	0	0.2	0.5
Simulation results					
Estimated variance of individual error terms ($\hat{\kappa}^2$)					
Mean estimator	14.9929	14.9917	15.0011	14.4057	11.2370
Median estimator	14.9923	14.9888	15.0077	14.4006	11.2426
Standard error across S	0.2115	0.2105	0.2151	0.2162	0.1604
RMSE	0.2115	0.2106	0.2150	0.6323	3.7664
Estimated variance of group error terms ($\hat{\sigma}^2$)					
Mean estimator	0.9881	1.0257	1.0054	3.1457	8.6580
Median estimator	0.9671	0.5773	0.9911	3.1113	8.5951
Correction (%)	25.8224	81.3729	16.6914	9.4320	2.8494
Standard error across S	0.2203	1.6632	0.1810	0.5328	1.2627
RMSE	0.2205	1.6625	0.1810	2.2108	7.7613
Number of negative values	0	296	0	0	0
Two-stage OLS estimator of group variable coefficient ($\hat{\gamma}_{OLS}$)					
Mean estimator	0.9990	1.0177	0.9954	1.0120	0.9964
Median estimator	0.9971	1.0195	0.9895	1.0168	1.0069
Its uncorrected standard error	0.1862	0.4167	0.1796	0.4020	0.4915
Its corrected standard error	0.1870	0.3362	0.1787	0.4009	0.4886
Mean uncorrected standard error	0.2014	0.3289	0.1922	0.3249	0.5219
Mean corrected standard error	0.2017	0.3261	0.1922	0.3248	0.5221
Median uncorrected standard error	0.1998	0.3089	0.1903	0.3233	0.5167
Median corrected standard error	0.1999	0.3067	0.1898	0.3230	0.5173
RMSE	0.2018	0.3269	0.2005	0.3262	0.5150
Two-stage FGLS estimator of group variable coefficient ($\hat{\gamma}_{FGLS}$)					
Mean estimator	0.9991	1.1760	0.9953	1.0119	0.9971
Median estimator	0.9966	0.9832	0.9912	1.0150	1.0085
Its standard error	0.2237	0.2598	0.1780	0.3636	0.4702
Mean standard error	0.1922	0.1973	0.1915	0.3212	0.5211
Median standard error	0.1901	0.1671	0.1891	0.3196	0.5163
RMSE	0.1913	6.2278	0.2005	0.3244	0.5144
Direct within estimator of group variable coefficient in one stage only ($\hat{\gamma}_{DW}$)					
Mean estimator	1.0012	1.0047	0.9941	1.0124	1.0080
Median estimator	1.0044	0.9897	0.9942	1.0077	1.0046
Its standard error	0.0653	0.0700	0.0704	0.0786	0.0759
Mean standard error	0.0698	0.0698	0.0698	0.0730	0.0777
Median standard error	0.0694	0.0694	0.0696	0.0725	0.0772
RMSE	0.2046	0.2175	0.2207	0.3580	0.5872
Mean estimator	1.0012	1.0047	0.9941	1.0124	1.0080

By definition, the standard error across S differs from the RMSE for a given estimator only because the empirical mean is used instead of the true expectation.

7.2 Simulations with an endogenous mobility pattern

We now try to assess the bias in the estimates when the mobility process is endogenous. For that purpose, we conduct simulations in which the group choice of individuals at each date is taken on the basis of the (conditional) expected outcome as described in Section 4 and an observed group-specific individual random error term $\phi_{i,g,t}$ (hereafter called *choice error term* to avoid some confusion with the error terms included in the outcome equations):

$$\begin{aligned}
&\text{Case 1:} \quad g(i,t) = \arg\max_g E_{\xi_{i,g,t},\eta_{g,t}} \left(y_{i,g,t} + \phi_{i,g,t} \right) \\
&\text{Case 2:} \quad g(i,t) = \arg\max_g E_{\xi_{i,g,t}} \left(y_{i,g,t} + \phi_{i,g,t} \right) \\
&\text{Case 3:} \quad g(i,t) = \arg\max_g E_{\eta_{g,t}} \left(y_{i,g,t} + \phi_{i,g,t} \right) \\
&\text{Case 4:} \quad g(i,t) = \arg\max_g \left(y_{i,g,t} + \phi_{i,g,t} \right)
\end{aligned}$$

with $y_{i,g,t} = \alpha + z_{g,t}\gamma + \eta_{g,t} + \xi_{i,g,t}$.

The inclusion of the choice error terms is necessary to obtain group fixed effects that are identified when the model is too deterministic (i.e. when the group choice is made conditionally on $\xi_{i,g,t}$). In many cases these error terms have an economic meaning. For instance, in the case of migrations made on the basis of expected outcome, they can represent the individual-specific effect of local amenities that may affect the migration choice. The law of the error terms in the outcome equations and the parameters are chosen to be similar to those in the benchmark case when the mobility pattern is exogenous: $\xi_{i,g,t}$ and $\eta_{g,t}$ are drawn independently in centered normal laws such that $V\left(\xi_{i,g,t}\right) = 15$ and $V\left(\eta_{g,t}\right) = 1$; $z_{g,t}$ is drawn in a $[-1, 1]$ uniform law and its coefficient is $\gamma = 1$. We also make some *iid* draws of $\phi_{i,g,t}$ in a normal law with variance 15. In some simulations, we will increase this variance to assess how estimates change when the outcome has a less important role in the choice process.

The estimation results of the outcome equation are reported in Table 3. Our purpose here is to assess the bias in the estimates depending on which error terms (individual or aggregate) are observed by the individuals. In the benchmark case when the outcome shocks are unobserved (case 1, column 1), mobility is exogenous. Thus, the estimated variances of error terms as well as the estimated coefficient of the group variable are unbiased as in the previous subsection. When the group error terms only are observed (case 2, column 2), the two-stage OLS estimator, the variance of group error terms and the variance of individual error terms are still unbiased as shown in section 5. However, the two-stage FGLS estimator and the direct within estimator exhibit a bias of respectively 6% and 5%. There biases are quite small because the variance of the group error terms is also small.

When the individual error terms are observed but not the group error terms (case 3, column 3), all the estimators are biased except the estimated variance of group error terms that seems to be unbiased.[13] As we fixed the variance of the individual error terms such that they are the main driver of the outcome and a major determinant of the group choice process, the biases are very large. The estimated variance of individual error terms, as well as the estimated coefficient of the group variable, are biased by 40%. Results obtained when both the individual and group error terms are observed (case 4, column 4) turn to be similar except that the estimated variance of group error terms is now biased by 65%. We finally examine for case 4, how biases decrease when the outcome has a less important impact in the choice process. For that purpose, we increase the variance of choice error terms. When the variance goes up from 15 to 75 (column 5), the bias on the estimated coefficient of the group variable reduces from 40% to 15%. The bias on the estimated variance of group error terms reduces from 65% to 25%. When the variance reaches 150 (column 6), there is a 8% bias only on the estimated coefficient of the group variable whereas there is still a 15% bias on the estimated variance of group error terms.

[13] As explained in Section 3, the estimated variance of group error terms should be biased. However, the bias is numerically negligible here. This arises from the fact that only $tr\left[M_Z\widehat{V}(\widehat{\beta}\,|\Phi)\right]$ is biased in formula (4). Note however that we obtained sometimes a very small detectable bias when we conducted some other simulations corresponding to different mobility patterns of individuals.

Table 3: Simulation results when the mobility process is endogenous

	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)
Parameters						
Mobility scheme	1	2	3	4	4	4
S	1000	1000	1000	1000	1000	1000
Variance of idiosyncratic term	15	15	15	15	15	15
Variance of individual error term (κ^2)	15	15	15	15	75	150
Variance of group error term (σ^2)	1	1	1	1	1	1
Group variable coefficient (γ)	1	1	1	1	1	1
Number of groups (Z)	50	50	50	50	50	50
Number of individuals (N)	10000	10000	10000	10000	10000	10000
Number of periods (T)	2	2	2	2	2	2
Correlation between error terms (ρ)	0	0	0	0	0	0
Simulation results						
Mobility rate	98.01%	98.00%	98.01%	98.00%	98.00%	97.99%
Estimated variance of individual error terms ($\hat{\kappa}^2$)						
Mean estimator	15.0037	15.0050	9.1288	9.1758	13.0425	13.9261
Median estimator	15.0129	15.0005	9.1319	9.1776	13.0415	13.9251
Standard error across S	0.2141	0.2168	0.1286	0.1244	0.1897	0.1936
RMSE	0.2141	0.2167	5.8726	5.8255	1.9667	1.0912
Estimated variance of group error terms ($\hat{\sigma}^2$)						
Mean estimator	1.0036	1.0049	1.0020	0.3629	0.7586	0.8560
Median estimator	0.9914	0.9857	0.9896	0.3559	0.7548	0.8482
Correction (%)	14.5822	20.0230	8.9258	24.6683	15.8184	14.6906
Standard error across S	0.1707	0.2001	0.1632	0.0697	0.1309	0.1414
RMSE	0.1706	0.2001	0.1631	0.6409	0.2746	0.2018
Number of negative values	0	0	0	0	0	0
Two-stage OLS estimator of group variable coefficient ($\hat{\gamma}_{OLS}$)						
Mean estimator	1.0059	0.9971	0.6077	0.6080	0.8676	0.9208
Median estimator	1.0095	0.9972	0.6051	0.6046	0.8632	0.9239
Its uncorrected standard error	0.1719	0.2235	0.1925	0.0990	0.1459	0.1578
Its corrected standard error	0.1721	0.2244	0.1924	0.0986	0.1459	0.1574
Mean uncorrected standard error	0.1892	0.1956	0.1835	0.1210	0.1658	0.1748
Mean corrected standard error	0.1900	0.1967	0.1837	0.1214	0.1659	0.1748
Median uncorrected standard error	0.1875	0.1931	0.1819	0.1202	0.1651	0.1735
Median corrected standard error	0.1882	0.1944	0.1821	0.1206	0.1652	0.1737
RMSE	0.1919	0.1994	0.4351	0.4093	0.2098	0.1962
Two-stage FGLS estimator of group variable coefficient ($\hat{\gamma}_{FGLS}$)						
Mean estimator	1.0056	0.9434	0.6078	0.5890	0.8599	0.9170
Median estimator	1.0090	0.9382	0.6057	0.5880	0.8553	0.9190
Its standard error	0.1990	0.1679	0.2105	0.1366	0.1646	0.1649
Mean standard error	0.1897	0.1943	0.1837	0.1204	0.1658	0.1747
Median standard error	0.1880	0.1919	0.1820	0.1196	0.1650	0.1735
RMSE	0.1920	0.1980	0.4349	0.4268	0.2136	0.1971
Direct within estimator of group variable coefficient in one stage only ($\hat{\gamma}_{DW}$)						
Mean estimator	1.0032	0.9507	0.6116	0.6002	0.8600	0.9192
Median estimator	1.0007	0.9533	0.6107	0.5988	0.8610	0.9191
Its standard error	0.0658	0.0709	0.0586	0.0568	0.0684	0.0650
Mean standard error	0.0722	0.0719	0.0564	0.0547	0.0650	0.0669
Median standard error	0.0720	0.0716	0.0562	0.0543	0.0649	0.0666
RMSE	0.2011	0.2493	0.4339	0.4215	0.2203	0.2012

By definition, the standard error across S differs from the RMSE for a given estimator only because the empirical mean is used instead of the true expectation.

8 Conclusion

In this paper, we study the effect of aggregate variables on an individual outcome in linear panel models with individual fixed effects. We focus on cases where individuals are *mobile* and can change group across time. It has been shown in the literature that standard errors of aggregate effects can be highly biased if unobserved heterogeneity at the aggregate level is omitted from the econometric specification. Many related papers recommand to take aggregate unobservables into account through *iid* random terms and to use a FGLS approach to estimate the model. However, they deal with cross-section models only and their approach cannot be adapted to panel models where individuals are mobile.

Consequently, we explain how an alternative two-stage method can be implemented to solve the issue. In first stage, the individual outcome is specified as a function of the individual variables, some group-year fixed effects and the individual fixed effects. This specification is estimated after the equation has been projected in the within-individual dimension. In second stage, the estimated group-year fixed effects are regressed on the aggregate variables.

The estimates obtained for the coefficients of individual and aggregate explanatory variables are unbiased. They are also consistent under reasonable assumptions. Moreover, we are able to construct an unbiased and consistent estimator of the variance of aggregate error terms. This estimator is used to compute some unbiased and consistent standard errors for the estimated coefficients of aggregate variables.

The two-stage method has many advantages. The estimator of the variance of group error terms is easy to compute. It is possible to conduct a variance analysis at the aggregate level in second stage. Morover, the estimation procedure can be adapted to take into account heteroskedasticity and/or autocorrelation of individual and group error terms, as well as many endogeneity issues.

However, the method cannot be extended easily to cope with some selection effects in the group choice process of individuals. Indeed, it would be tempting to correct for the selection bias using multiple-choice models. Unfortunately, the related approches in the literature are designed for cross-section data and extensions to panel data are not straightforward. The correction of the selection bias in linear panel models with individual fixed effects constitutes a topic for further research. Another limit of the two-stage method is that it becomes burdensome or even unapplicable when the number of groups is large.

Finally, note that the two-stage method may be used to estimate nonlinear models (like duration models). In that case, the first-stage equation is non linear and the second-stage equation is linear. However, the results on unbiasedness do not hold at finite distance and the assumptions made to show the large sample properties of the estimators have to be modified. The use of the two-stage method to measure the effect of aggregate variables on the individual outcome in a nonlinear framework is thus a topic for further research.

9 Appendix

9.1 Appendix A: different assumptions on mobility and individual effects

9.1.1 The two-stage method when mobility is imperfect

We explain briefly how to apply the two-stage method when mobility is imperfect and one restriction on group-year fixed effects is not enough to identify the model.

We first consider the setting where individuals are *immobile*. In the first stage of the method, only the within-group variation of group-year fixed effects can be identified. Thus, we need to impose G identifying restrictions: one for each group. Suppose for instance that for all g, we have: $\beta_{g,1} = 0$. In second stage, the most convenient way to cope with the identifying restrictions is to project equation (3) in the within-group dimension. It makes the unidentified inter-group variations disappear.[14] We introduce F_G the matrix corresponding to group dummies and M_{F_G} the within-group projector. The second-stage equation rewrites $M_{F_G}\widehat{\beta} = M_{F_G}Z\gamma + M_{F_G}\eta + M_{F_G}\Psi$, where the uncertainty on group-year fixed effects is redefined such that $\Psi = (\Psi_1', ..., \Psi_G')'$ with $\Psi_g = (0, \Psi_{g,2}, ..., \Psi_{g,T})'$. After estimating this equation, it is possible to construct an estimator of the variance of group error terms similar to (4):

$$\widehat{\sigma}^2 = \frac{1}{tr\left(M_{F_G}M_{M_{F_G}Z}\right)}\left[M_{F_G}\widehat{Z\,(\eta + \Psi)}' M_{F_G}\widehat{Z\,(\eta + \Psi)} - tr\left[M_{F_G}M_{M_{F_G}Z}\widehat{V}(\widehat{\beta}\,|\Phi\,)\right]\right] \tag{16}$$

where $\widehat{\beta}_{g,1} = 0$ for all g and $\widehat{V}(\widehat{\beta}\,|\Phi\,)$ is constructed from the first-stage results. When F_G and Z are orthogonal, this formula simplies as $M_{F_G}Z = Z$ and $tr\left(M_{F_G}M_{M_{F_G}Z}\right) = GT - G - K$.

We now turn to the setting where individuals are *mobile* but groups are imperfectly interconnected by movers. Consider for instance the case where groups are properly interconnected within two subsets, but these subsets are not connected to each other. We denote $\{g_1, ..., g_{S_1}\}$ the groups in the first subset and $\{g_{S_1+1}, ..., g_G\}$ the groups in the second subset. In the first stage of the method, two identifying restrictions (one for each subset) have to be imposed, say: $\beta_{g_1,1} = 0$ and $\beta_{g_{S_1+1},1} = 0$. In second stage, it is possible to cope with the identifying restrictions by projecting the model in the within-subset dimension. Denote F_2 the matrix corresponding to subset dummies and M_{F_2} the within-subset projector. The second-stage equation rewrites $M_{F_2}\widehat{\beta} = M_{F_2}Z\gamma + M_{F_2}\eta + M_{F_2}\Psi$, where the uncertainty on group-year fixed effects is redefined such that $\Psi = (\Psi_{S_1}', \Psi_{S_2}')'$, with $\Psi_{S_1} = (0, g_2, ..., g_{S_1})'$ and $\Psi_{S_1} = (0, g_{S_1+2}, ..., g_G)'$. An

[14] Note that coefficients related to time-varying explanatory variables are not identified.

estimator of the variance of group error terms can be obtained by replacing F_G with F_2 in equation (16).
It is possible to generalize this procedure to any number of subsets. In practice, the subsets are defined by examining the flows of stayers and movers across time.
Note that a polar case is obtained when the number of subsets is G. Then, all individuals are immobile and we are in the first setting analyzed in this appendix.
Another polar case is obtained when the number of subsets is 1. Then, all groups are interconnected. There is a slight variation here compared to how the two-stage method was applied in section 3. Here, the second-stage equation (3) has been centered to make the constant disappear.

9.1.2 The two-stage method when individual effects are random

The two-stage estimation method can also be implemented when individual effects are random (and not fixed). In that case, there is no need to impose any identifying restriction on group-year fixed effects (like $\beta_{1,1} = 0$) provided that the constant is omitted in first stage. Equation (2) is a random coefficients panel model for which it is possible to apply general least squares (for instance, see Wooldridge, 2002). The GLS estimators of group-year fixed effects are used as dependent variables in second stage. The GLS estimator of γ derived from model (3) is then the GLS estimator of γ for model (1) according to Amemiya (1978)'s results. Interestingly, note that we obtain a FGLS estimator for model (1) replacing κ^2 by $\widehat{\kappa}^2$ (its first-stage estimator), and σ^2 by $\widehat{\sigma}^2$.

If there is no mobility, the two-stage method is not necessary when individual effects are random. It is possible to construct a FGLS estimator of coefficients for model (1). Indeed, some adequate projections of model (1) allow to recover the variance of all random terms. More specifically, projecting the model in the within-individual dimension makes individual random effects and group error terms disappear. Thus, the variance of individual shocks can be recovered. Then, projecting model (1) in the between-individual dimension leads to a simple two-level model that has been studied extensively in the literature (see Wooldridge, 2003). Projecting this two-level model in the within-group and between-group dimensions allows to recover the variance of individual random effects and group error terms.

9.2 Appendix B: properties of estimators

Proof of Property 1:

We have:

$$\left(\widehat{\eta + \Psi}\right)' \left(\widehat{\eta + \Psi}\right) = \eta' M_Z \eta + \Psi' M_Z \eta + \eta' M_Z \Psi + \Psi' M_Z \Psi \tag{17}$$

We then write:

$$E\left(\eta' M_Z \eta\right) = Etr\left[M_Z E\left(\eta\eta' \left|Z\right.\right)\right] = \left(GT - K\right)\sigma^2 \tag{18}$$

We note $B = \left(0, M_H F \left(F' M_H F\right)^{-1}\right)'$ the matrix such that $\Psi = B\varepsilon$. From $A3$, we obtain:

$$E\left(\Psi' M_Z \eta\right) = trE\left[B' M_Z E\left(\eta\varepsilon' \left|\Phi\right.\right)\right] = 0 \tag{19}$$

Similarly, we get $E\left(\eta' M_Z \Psi\right) = 0$. We also have:

$$E\left(\Psi' M_Z \Psi\right) = E\left[tr\left(M_Z E\left(\Psi\Psi' \left|\Phi\right.\right)\right)\right] = trE\left[M_Z V\left(\widehat{\beta}\left|\Phi\right.\right)\right] \tag{20}$$

Moreover,

$$\begin{aligned} E\left[tr\left[M_Z \widehat{V}\left(\widehat{\beta}\left|\Phi\right.\right)\right]\right] &= E\left[tr\left[M_Z E\left[\widehat{V}\left(\widehat{\beta}\left|\Phi\right.\right)\left|\Phi\right.\right]\right]\right] \\ &= trE\left[M_Z V\left(\widehat{\beta}\left|\Phi\right.\right)\right] \end{aligned} \tag{21}$$

We finally show the property. ■

Proof of Property 2:

We will need the following Chebychev's weak law of large numbers for triangular arrays to show consistency properties (see Borovkov, p153):

Theorem A: Let $\varphi(\bullet) : N \rightarrow N$ be a strictly increasing function; $\varsigma_{n,i}$, $i = 1, ..., \varphi(n)$, $n \in N$, forms a triangular array of 1×1 independent random variables with $E\left(\varsigma_{n,i}\right) = 0$ and $E\left(\varsigma_{n,i}^2\right) = \sigma_{n,i}^2 < +\infty$. Denote $\zeta_n = \sum_{i=1}^{n} \varsigma_{n,i}$ and suppose that the following condition, noted C1, *holds:*

$$V\left(\frac{1}{\varphi(n)}\zeta_{\varphi(n)}\right) = \frac{1}{\left[\varphi(n)\right]^2}\sum_{i=1}^{\varphi(n)} \sigma_{n,i}^2 \underset{n \longrightarrow +\infty}{\longrightarrow} 0 \tag{22}$$

Then, we have: $\frac{1}{\varphi(n)}\zeta_{\varphi(n)} \underset{n \longrightarrow +\infty}{\overset{P}{\longrightarrow}} 0$.

We will also use the two following lemmas:[15]

Lemma 1: Consider two matrices X and Y. $tr\left(XY\right) \leqslant \sqrt{tr\left(XX'\right)}\sqrt{tr\left(YY'\right)}$.
Lemma 2: Consider Ω a symmetric positive definite matrix. $tr\left(\Omega^2\right) \leqslant \left[tr\left(\Omega\right)\right]^2$.

We now prove Property 2. We have:

$$\left(\widehat{\eta + \Psi}\right)' \left(\widehat{\eta + \Psi}\right) = \eta' M_Z \eta + \Psi' M_Z \eta + \eta' M_Z \Psi + \Psi' M_Z \Psi \tag{23}$$

We can write:

$$\frac{\eta' M_Z \eta}{GT} = \frac{\eta'\eta}{GT} - \frac{1}{GT} tr\left(P_Z \eta\eta'\right) \tag{24}$$

[15] The proof of Lemma 1 uses the Cauchy-Schwartz inequality. The proof of Lemma 2 uses the fact that for any given positive definite matrix $v = \left(v_{i,j}\right)$, we have: $\left|v_{i,j}\right| \leqslant \sqrt{v_{i,i}}\sqrt{v_{j,j}}$.

As the shocks $\eta_{g,t}$ are *iid* with fourth moment Q, condition *C1* is verified by the first right-hand-side term. Consequently, this term tends to σ^2 when G tends to infinity according to the LGN given by Theorem A. It is possible to check that *C1* is also verified for the second right-hand-side term as:

$$V\left[\frac{1}{GT}tr\left(P_Z\eta\eta'\right)\right] \leqslant \frac{1}{(GT)^2}\left(Q+5\sigma^4\right)K \tag{25}$$

Consequently, we can apply theorem A again. Finally, we get: $\frac{\eta' M_Z \eta}{GT} \xrightarrow{P} \sigma^2$.

We also have:

$$\frac{\Psi' M_Z \eta}{GT} = \frac{\varepsilon' B' \eta}{GT} - \frac{\varepsilon' B' P_Z \eta}{GT} = tr\left[\left(\frac{B'}{GT}\right)\eta\varepsilon'\right] - tr\left[\left(\frac{B' P_Z}{GT}\right)\eta\varepsilon'\right] \tag{26}$$

Both quantities on the right-hand side can be rewritten as weighted sums of the *iid* centered residuals $\varepsilon_{i,t}\eta_{g,t}$. We want to apply the LGN to both these sums. We check that condition *C1* is verified. We have:

$$V\left[tr\left[\left(\frac{B'}{GT}\right)\eta\varepsilon'\right]\right] = E\left[tr\left[\left(\frac{B'}{GT}\right)\eta\varepsilon'\right]\right]^2 = \frac{\sigma^2\kappa^2}{(GT)^2}E\left[tr\left(F' M_H F\right)^{-1}\right]$$

Using *A5*, this term tends to zero when G tends to infinity. Consequently, condition *C1* is verified for the first sum in (26). It is possible to show the same way that:

$$V\left[tr\left[\left(\frac{B' P_Z}{GT}\right)\eta\varepsilon'\right]\right] = \frac{\sigma^2\kappa^2}{(GT)^2}E\left[tr\left(P_Z B B'\right)\right] \tag{27}$$

Using Lemma 1 and 2, we get:

$$tr\left(P_Z B B'\right) \leqslant \sqrt{tr\left(P_Z^2\right)}\sqrt{\left[tr\left(BB'\right)\right]^2} \leqslant \sqrt{K}tr\left(BB'\right) \tag{28}$$

Thus,

$$V\left[tr\left[\left(\frac{B' P_Z}{GT}\right)\eta\varepsilon'\right]\right] \leqslant \frac{\sigma^2\kappa^2}{(GT)^2}\sqrt{K}E\left[tr\left(F' M_H F\right)^{-1}\right] \tag{29}$$

and condition *C1* is verified for the second sum in (26). Applying Theorem A twice, we obtain:

$$\frac{\Psi' M_Z \eta}{GT} \xrightarrow{P} Etr\left[\left(\frac{B'}{GT}\right)\eta\varepsilon'\right] + Etr\left[\left(\frac{B' P_Z}{GT}\right)\eta\varepsilon'\right] = 0 \tag{30}$$

Finally, we have:

$$\frac{\Psi' M_Z \Psi}{GT} - \frac{1}{GT}tr\left[M_Z\widehat{V}(\widehat{\beta}\,|\Phi)\right] = \frac{1}{GT}\left[\Psi' M_Z \Psi - tr\left[M_Z V(\widehat{\beta}\,|\Phi)\right]\right] \tag{31}$$

$$+\frac{1}{GT}tr\left[M_Z\left[V(\widehat{\beta}\,|\Phi) - \widehat{V}(\widehat{\beta}\,|\Phi)\right]\right] \tag{32}$$

We can write:

$$\frac{1}{GT}\left[\Psi' M_Z\Psi - tr\left[M_Z V(\widehat{\beta}\,|\Phi\,)\right]\right] = tr\left[M_Z\left(B\varepsilon\varepsilon' B'\right)\right] - tr\left[M_Z E\left(B\varepsilon\varepsilon' B'\,|\Phi\,\right)\right]$$
$$= tr\left[B' M_Z B\left[\varepsilon\varepsilon' - E\left(\varepsilon\varepsilon'\,|\Phi\,\right)\right]\right] \qquad (33)$$

We want to apply the LGN given by Theorem A. For that purpose, we check that condition *C1* is verified. Using the fact that shocks $\varepsilon_{i,t}$ are *iid* with fourth moment R, we obtain:

$$V\left(\frac{1}{GT} tr\left[B' M_Z B\left[\varepsilon\varepsilon' - E\left(\varepsilon\varepsilon'\,|\Phi\,\right)\right]\right]\right) \leqslant \frac{R+\kappa^4}{(GT)^2} trE\left[B' M_Z B B' M_Z B\right] \qquad (34)$$

Then, we get after using two times Lemma 1 and 2:

$$V\left(\frac{1}{GT} tr\left[B' M_Z B\left[\varepsilon\varepsilon' - E\left(\varepsilon\varepsilon'\,|\Phi\,\right)\right]\right]\right) \leqslant \frac{R+\kappa^4}{(GT)^2}\left(1+\sqrt{K}\right)^2 E\left[tr\left(F' M_H F\right)^{-1}\right]^2 \qquad (35)$$

Using Assumption *A5*, this term tends to zero when G tends to infinity. Applying Theorem A, we obtain:

$$\frac{1}{GT}\left[\Psi' M_Z\Psi - tr\left[M_Z V(\widehat{\beta}\,|\Phi\,)\right]\right] \xrightarrow{P} 0 \qquad (36)$$

Under Assumption *A1*, we get $\widehat{V}(\widehat{\beta}\,|\Phi\,) = \frac{\widehat{\kappa}^2}{\kappa^2} V(\widehat{\beta}\,|\Phi\,)$ from the first-stage within estimation, with $\widehat{\kappa}^2 = \frac{1}{tr\left[M_{(M_H F)}\right]}\varepsilon' M_{(M_H F)}\varepsilon$. We show that: $\widehat{\kappa}^2 \underset{G\longrightarrow+\infty}{\xrightarrow{P}} \kappa^2$. Indeed, we have: $(N-G)\,T \leqslant tr\left[M_{(M_H F)}\right] \leqslant NT$. Using *A4*, we obtain: $\frac{1}{NT} tr\left[M_{(M_H F)}\right] \underset{G\longrightarrow+\infty}{\xrightarrow{P}} 1$. Finally, it is easy to show that $\frac{1}{NT}\varepsilon' M_{(M_H F)}\varepsilon \underset{G\longrightarrow+\infty}{\xrightarrow{P}} \kappa^2$ with a proof similar to the one used to show that $\frac{\eta' M_Z \eta}{GT} \xrightarrow{P} \sigma^2$. Using lemma 1 and 2, we then get:

$$\frac{1}{GT} tr\left[M_Z\left[V(\widehat{\beta}\,|\Phi\,) - \widehat{V}(\widehat{\beta}\,|\Phi\,)\right]\right] = \frac{1}{(GT)\kappa^2}\left(\widehat{\kappa}^2 - \kappa^2\right) tr\left[M_Z V(\widehat{\beta}\,|\Phi\,)\right] \qquad (37)$$
$$\leqslant \frac{1+\sqrt{K}}{GT}\left|\widehat{\kappa}^2 - \kappa^2\right| tr\left[\left(F' M_H F\right)^{-1}\right] \xrightarrow{P} 0 \qquad (38)$$

and we obtain: $\widehat{\sigma}^2 \underset{G\longrightarrow+\infty}{\xrightarrow{P}} \sigma^2$. ■

Proof of Property 4:

We have $\widehat{\gamma}_{OLS} = \gamma + \left(Z'Z\right)^{-1} Z'\eta + \left(Z'Z\right)^{-1} Z' B\varepsilon$. $\frac{Z'Z}{GT} \underset{G\rightarrow+\infty}{\xrightarrow{P}} Q_0$ a positive definite matrix, the inverse is a continous function, and $\frac{Z'\eta}{GT} \underset{G\rightarrow+\infty}{\xrightarrow{P}} 0$. According to Theorem A, we get: $\left(Z'Z\right)^{-1} Z'\eta \xrightarrow{P} Q_0^{-1}.0 = 0$. Moreover, each element of

$\frac{1}{GT}Z'B\varepsilon$ is a weighted sum of the *iid* centered residuals $\varepsilon_{i,t}$. We want to apply Theorem A to each of these sums. Thus, we check that condition *C1* is verified. Using Lemma 1 and 2, as well as *A6a*, we obtain:

$$trV\left[\frac{1}{GT}Z'B\varepsilon\right] = \frac{\kappa^2}{(GT)^2}trE\left[Z'\left[\begin{array}{c}0\\(F'M_HF)^{-1}\end{array}\right]Z\right] \tag{39}$$

$$\leqslant \frac{\kappa^2}{(GT)^2}E\left[\sqrt{tr\left(ZZ'ZZ'\right)}\sqrt{\left[tr\left(F'M_HF\right)^{-1}\right]^2}\right] \tag{40}$$

$$\leqslant \frac{K\theta_1^2}{GT}\kappa^2 trE\left[\left(F'M_HF\right)^{-1}\right] \tag{41}$$

Using *A6a*, this term tends to zero when G tends to infinity. Thus the variance of each element of $\frac{1}{GT}Z'B\varepsilon$ tends to zero. Consequently, $\widehat{\gamma}_{OLS} \underset{G\to+\infty}{\overset{P}{\to}} \gamma$.

We can write that $\widehat{\gamma}_{GLS} = \gamma + \left(\frac{Z'\Omega^{-1}Z}{GT}\right)^{-1}\frac{1}{GT}Z'\Omega^{-1}\left(\eta + B\varepsilon\right)$. Each element of $\frac{1}{GT}Z'\Omega^{-1}\left(\eta + B\varepsilon\right)$ is a weighted sum of the independent centered residuals $\varepsilon_{i,t}$ and $\eta_{g,t}$. We want to apply Theorem A to each of these sums. Thus, we check that condition *C1* is verified. Using Lemma 1 and 2, as well as *A6b*, we obtain:

$$\begin{aligned} trV\left[\frac{1}{GT}Z'\Omega^{-1}\left(\eta+\Psi\right)\right] &= \frac{1}{(GT)^2}trE\left(Z'\Omega^{-1}Z\right) \\ &\leqslant \frac{1}{(GT)^2}E\left[\sqrt{tr\left(ZZ'ZZ'\right)}\sqrt{tr\left(\Omega^{-1}\right)^2}\right] \\ &\leqslant \frac{K\theta^2}{GT}Etr\left(\Omega^{-1}\right) \end{aligned} \tag{42}$$

Using *A6b*, this term tends to zero when G tends to infinity. Thus the variance of each element of $\frac{1}{GT}Z'\Omega^{-1}\left(\eta + B\varepsilon\right)$ tends to zero. Consequently, $\widehat{\gamma}_{GLS} \underset{G\to+\infty}{\overset{P}{\to}} \gamma$. The same kind of argument can be applied using *A6c* to show that $\widehat{\gamma}_{FGLS} \underset{G\to+\infty}{\overset{P}{\to}} \gamma$. ■

Proof of Property 5:

We first prove the part of Property 5 related to the OLS estimator. We need to use a central limit theorem (CLT) for triangular arrays in the multivariate case that is given by Borovkov (1998, Theorem 11A, p174):

Theorem B. Let $\varphi(\bullet) : N \to N$ be a strictly increasing function, and let $X_{n,i}$, $i = 1, ..., \varphi(n)$, $n \in N$, form a triangular array of independent vectors of random variables with $E(X_{n,i}) = 0$ and $E(\|X_{n,i}\|) < +\infty$, where $\|.\|$ is the Euclidean *norm. Denote* $X_n = \sum_{i=1}^{\varphi(n)} X_{n,i}$, $\Sigma_{n,i} = E\left(X_{n,i}X'_{n,i}\right)$ *and* $\Sigma_n = \sum_{i=1}^{\varphi(n)} \Sigma_{n,i}$. *Suppose that the Lyapunov condition holds:*

$$\sum_{i=1}^{\varphi(n)} E\left(\|X_{n,i}\|^{2+\lambda}\right) \underset{n\longrightarrow+\infty}{\longrightarrow} 0 \text{ for some } \lambda > 0 \tag{43}$$

If $\Sigma_n \underset{n\longrightarrow+\infty}{\longrightarrow} \Sigma$ *a positive definite matrix, then* $X_n \overset{L}{\underset{n\to+\infty}{\to}} N(0,\Sigma)$.

To prove the Property 5, we can first write that:

$$\sqrt{GT}\left(\widehat{\gamma}_{OLS} - \gamma\right) = \left(\frac{Z'Z}{GT}\right)^{-1}\left(\frac{Z'\eta}{\sqrt{GT}} + \Theta\varepsilon\right) \tag{44}$$

with $\Theta = \frac{1}{\sqrt{GT}}Z'B$. Note that Θ depends on G even if it is not stated here. We want to apply Theorem B with $\varphi(\bullet)$ defined such as $\varphi(G) = GT + Tf(G)$ where $f(\bullet)$ verifies $N = f(G)$ (see Assumption *A4*) and $\lambda = 2$ to show that:

$$\frac{Z'\eta}{\sqrt{GT}} + \Theta\varepsilon \overset{L}{\underset{G\to+\infty}{\to}} N\left(0, \sigma^2 Q_0 + \kappa^2 Q_1\right) \tag{45}$$

Consequently, we check that all requirements are verified.

We first introduce the $K \times 1$ vectors $x_{g,t} = \frac{Z'_{g,t}\eta_{g,t}}{\sqrt{GT}}$ and $y_{i,t} = \Theta_{i,t}\varepsilon_{i,t}$ where $Z_{g,t}$ is the element of Z corresponding to group g in year t; $\Theta_{i,t}$ is the element of Θ corresponding to individual i in year t.
We have $E(x_{g,t}) = E(y_{i,t}) = 0$. Moreover, for $\|x_{g,t}\| > 1$, we get $\|x_{g,t}\| \leqslant \|x_{g,t}\|^2 = \eta^2_{g,t}\frac{Z_{g,t}Z'_{g,t}}{GT}$. The expectation of the right-hand side quantity exists as Z has a bounded support. Thus, we have $E(\|x_{g,t}\|) < +\infty$. Moreover, for $\|y_{i,t}\| > 1$, we have $\|y_{i,t}\| \leqslant \|y_{i,t}\|^2 \leqslant \varepsilon^2_{i,t}\Theta'_{i,t}\Theta_{i,t}$. As all elements of Z are bounded by θ_1 and all the elements of $\ell.|B|$ are bounded by δ_1, we get: $\Theta'_{i,t}\Theta_{i,t} \leqslant \frac{K\theta_1^2\delta_1^2}{GT}$. Consequently, $E(\|y_{i,t}\|) < +\infty$. Finally, we have:

$$\sum_{g,t} E\left(\|x_{g,t}\|^4\right) + \sum_{i,t} E\left(\|y_{i,t}\|^4\right) = Q\sum_{g,t} E\left[\left(\frac{Z_{g,t}Z'_{g,t}}{GT}\right)^2\right] + R\sum_{i,t} E\left[\left(\Theta'_{i,t}\Theta_{i,t}\right)^2\right] \tag{46}$$

Moreover, we have (using the inequality given by (39)):

$$\begin{aligned}
\sum_{i,t} E\left(\|y_{i,t}\|^4\right) &\leqslant \frac{K\theta_1^2\delta_1^2}{GT}\sum_{i,t} E\left(\Theta'_{i,t}\Theta_{i,t}\right) \\
&\leqslant \frac{K\theta_1^2\delta_1^2}{(GT)^2} E\left[tr\left(Z'BB'Z\right)\right] \\
&\leqslant \frac{K^2\theta_1^4\delta_1^2}{GT} Etr\left[\left(F'M_H F\right)^{-1}\right]
\end{aligned} \tag{47}$$

The right-hand-side quantity tends to zero as G tends to infinity according to *A6a*. Moreover, as $\left|\frac{Z_{g,t}Z'_{g,t}}{GT}\right| \leqslant \frac{K\theta_1^2}{GT}$, we have:

$$\sum_{g,t} E\left[\left(\frac{Z_{g,t}Z'_{g,t}}{GT}\right)^2\right] \leqslant \frac{K^2\theta_1^4}{GT} \tag{48}$$

The right-hand-side quantity tends to zero as G tends to infinity. Consequently, we obtain: $\sum_{g,t} E\left(\|x_{g,t}\|^4\right) + \sum_{i,t} E\left(\|y_{i,t}\|^4\right) \underset{G\longrightarrow+\infty}{\longrightarrow} 0$ and the property is shown.

The central limit theorems for the GLS and FGLS estimators can by shown similarly, noting that:

$$\sqrt{GT}\left(\widehat{\gamma}_{GLS} - \gamma\right) = \left(\frac{Z'\Omega^{-1}Z}{GT}\right)^{-1}\left(\frac{1}{\sqrt{GT}}Z'\Omega^{-1}\eta + \frac{1}{\sqrt{GT}}Z'\Omega^{-1}B\varepsilon\right) \tag{49}$$

$$\sqrt{GT}\left(\widehat{\gamma}_{FGLS} - \gamma\right) = \left(\frac{Z'\widehat{\Omega}^{-1}Z}{GT}\right)^{-1}\left(\frac{1}{\sqrt{GT}}Z\widehat{\Omega}^{-1}\eta + \frac{1}{\sqrt{GT}}Z'\widehat{\Omega}^{-1}B\varepsilon\right) \tag{50}$$

The proofs are formally similar to the one used for the OLS estimator, redefining in the GLS case: $\Theta = \frac{1}{\sqrt{GT}}Z'\Omega^{-1/2}$ and in the FGLS case: $\Theta = \frac{1}{\sqrt{GT}}Z'\widehat{\Omega}^{-1/2}$. ■

9.3 Appendix C: details on the two particular cases

9.3.1 Case 1

The first-order conditions rewrite:

$$n\beta_{g,2} - \left(n - \frac{G}{G-1}m\right)\beta_{g,1} - \frac{m}{G-1}\sum_z \beta_{z,1} = w_{g,2} \tag{51}$$

$$n\beta_{g,1} - \left(n - \frac{G}{G-1}m\right)\beta_{g,2} - \frac{m}{G-1}\sum_z \beta_{z,2} = w_{g,1} \tag{52}$$

Summing equations (51) and (52), we obtain an expression for $\beta_{g,2}$. Replacing $\beta_{g,2}$ by this expression in (52), we get:

$$\begin{aligned} &n\beta_{g,1} - \left(n - \tfrac{G}{G-1}m\right)\left[\tfrac{G-1}{mG}\left(w_{g,1} + w_{g,2}\right) + \tfrac{1}{G}\sum_z \left(\beta_{z,1} + \beta_{z,2}\right) - \beta_{g,1}\right] \\ &-\tfrac{m}{G-1}\sum_z \beta_{z,2} = w_{g,1} \end{aligned} \tag{53}$$

Differencing with respect to the equation for group 1 and using the fact that $\beta_{1,1} = 0$, we obtain:

$$\widehat{\beta}_{g,1} = \frac{1}{n}\frac{\theta}{2\theta - 1}\left[\theta\left(w_{g,1} - w_{1,1}\right) + (\theta - 1)\left(w_{g,2} - w_{1,2}\right)\right] \tag{54}$$

with $\theta = \frac{G-1}{G}\frac{n}{m}$.
We need the following equalities to compute the variance of group fixed effects in year 1:

$$\begin{aligned} cov\left(w_g, w_{g+k,1}\right) &= 1_{\{k=0\}} V\left(w_{g,1}\right) = 1_{\{k=0\}} 2n\sigma^2 & (55) \\ cov\left(w_{g,2}, w_{g+k,2}\right) &= 1_{\{k=0\}} V\left(w_{g,2}\right) = 1_{\{k=0\}} 2n\sigma^2 & (56) \\ cov\left(w_{g,1}, w_{g+k,2}\right) &= -1_{\{k=0\}} 2\left(n-m\right)\sigma^2 - 1_{\{k\neq 0\}} 2\frac{m}{G-1}\sigma^2 & (57) \end{aligned}$$

We then obtain:

$$\begin{aligned} V\left(\widehat{\beta}_{g,1}\right) &= \frac{1}{n^2}\frac{\theta^2\sigma^2}{\left(2\theta-1\right)^2}\left[\theta^2 4n + \left(\theta-1\right)^2 4n + 2\theta\left(\theta-1\right)\left[-4\left(n-m\right)+\frac{4m}{G-1}\right]\right] \\ &= \frac{4\sigma^2}{n}\frac{\theta^3}{2\theta-1} \end{aligned} \quad (58)$$

We can study the evolution of this variance as a function of θ. We define: $k\left(\theta\right) = \frac{\theta^3}{2\theta-1}$. We have $k'\left(\theta\right) = \frac{4}{\left(2\theta-1\right)^2}\theta^2\left(4\theta-3\right)$. As the value of θ is above $\frac{G-1}{G}$, the derivative of k is always positive for $G > 3$. In that case, the variance of group fixed effects in year 1 is minimum for $m = n$.

We now compute the variance of group fixed effects in year 2. Using the same line of arguments that leads to equation (54) in year 1, we obtain:

$$\widehat{\beta}_{g,2} - \widehat{\beta}_{1,2} = \frac{1}{n}\frac{\theta}{2\theta-1}\left[\theta\left(w_{g,2}-w_{1,2}\right) + \left(\theta-1\right)\left(w_{g,1}-w_{1,1}\right)\right] \quad (59)$$

Rewriting (52) for $g = 1$, we get:

$$\widehat{\beta}_{1,2} - \frac{1}{\theta G}\sum_z \widehat{\beta}_{z,1} = \frac{1}{n} w_{1,2} \quad (60)$$

Moreover using (54), we have:

$$\sum_z \widehat{\beta}_{z,1} = \frac{1}{n}\frac{\theta}{2\theta-1}\left[\theta\sum_z\left(w_{z,1}-w_{1,1}\right) + \left(\theta-1\right)\sum_z\left(w_{z,2}-w_{1,2}\right)\right] \quad (61)$$

Thus:

$$\widehat{\beta}_{1,2} = \frac{1}{n} w_{1,2} + \frac{1}{nG}\frac{1}{2\theta-1}\left[\theta\sum_z\left(w_{z,1}-w_{1,1}\right) + \left(\theta-1\right)\sum_z\left(w_{z,2}-w_{1,2}\right)\right] \quad (62)$$

Consequently with (59), we obtain:

$$\begin{aligned} \widehat{\beta}_{g,2} &= \frac{1}{n}\frac{\theta}{2\theta-1}\left[\theta\left(w_{g,2}-w_{1,2}\right) + \left(\theta-1\right)\left(w_{g,1}-w_{1,1}\right)\right] \\ &\quad + \frac{1}{n} w_{1,2} + \frac{1}{nG}\frac{1}{2\theta-1}\left[\theta\sum_z\left(w_{z,1}-w_{1,1}\right) + \left(\theta-1\right)\sum_z\left(w_{z,2}-w_{1,2}\right)\right] \end{aligned} \quad (63)$$

Using the formulas:

$$V\left[\theta\sum_z\left(w_{z,1}-w_{1,1}\right)+\left(\theta-1\right)\sum_z\left(w_{z,2}-w_{1,2}\right)\right]=2G\left(G-1\right)\left(2\theta-1\right)n\sigma^2$$

$$\theta\sum_z cov\left(w_{1,2},w_{z,1}-w_{1,1}\right)+\left(\theta-1\right)\sum_z cov\left(w_{1,2},w_{z,2}-w_{1,2}\right)=-2n\frac{2\theta-1}{\theta}\sigma^2$$

$$cov\left[\begin{array}{c}\theta\left(w_{g,2}-w_{1,2}\right)+\left(\theta-1\right)\left(w_{g,1}-w_{1,1}\right),\\ \theta\sum_z\left(w_{z,1}-w_{1,1}\right)+\left(\theta-1\right)\sum_z\left(w_{z,2}-w_{1,2}\right)\end{array}\right]=2nG\frac{(\theta-1)(2\theta-1)}{\theta}\sigma^2$$

we finally get:

$$V\left(\widehat{\beta}_{g,2}\right)=V\left(\widehat{\beta}_{g,1}\right)-\frac{2}{nG}\sigma^2+\frac{4}{n}\frac{\theta-1}{2\theta-1}\sigma^2 \tag{64}$$

We define $l\left(\theta\right)=\frac{\theta-1}{2\theta-1}$. We have $l'\left(\theta\right)=\frac{1}{(2\theta-1)^2}>0$. Consequently, the variance of group fixed effects in year 2 is also minimum for $m=n$.

9.3.2 Case 2

The first-order conditions write:

$$n\beta_{g,2}-m\beta_{g-1,1}-\left(n-m\right)\beta_{g,1} = w_{g,2} \tag{65}$$

$$n\beta_{g,1}-m\beta_{g+1,2}-\left(n-m\right)\beta_{g,2} = w_{g,1} \tag{66}$$

Substituting the expression of group fixed effects in year 2 given by (65) in (66), it gives:

$$\left(\beta_{g+1,1}-\beta_{g,1}\right)-\left(\beta_{g,1}-\beta_{g-1,1}\right)=-\overline{w}_{g,1} \tag{67}$$

with $\overline{w}_{g,1}=\frac{1}{m(n-m)}\left[nw_{g,1}+mw_{g+1,2}+\left(n-m\right)w_{g,2}\right]$. From this equation, we get:

$$\begin{aligned} & \beta_{g+1,1}-\beta_{g,1}=\beta_{2,1}-\sum_{b=2}^{g}\overline{w}_{b,1} \\ \Longrightarrow\quad & \beta_{g,1}=\left(g-1\right)\beta_{2,1}-\sum_{c=2}^{g-1}\sum_{b=2}^{c}\overline{w}_{b,1} \\ \Leftrightarrow\quad & \beta_{g,1}=\left(g-1\right)\beta_{2,1}-\sum_{b=2}^{g-1}\left(g-b\right)\overline{w}_{b,1}\end{aligned} \tag{68}$$

Using the first and last expressions for $g=G$, we obtain:

$$-\beta_{G,1}=-\beta_{2,1}+\sum_{b=2}^{G}\overline{w}_{b,1}=\left(G-1\right)\beta_{2,1}-\sum_{b=2}^{G-1}\left(G-b\right)\overline{w}_{b,1} \tag{69}$$

It implies that:

$$\beta_{2,1}=\frac{1}{G}\sum_{b=2}^{G}\left(G-b+1\right)\overline{w}_{b,1} \tag{70}$$

And finally, we get:

$$\widehat{\beta}_{g,1} = \frac{g-1}{G}\sum_{b=2}^{G}(G-b+1)\,\overline{w}_{b,1} - \sum_{b=2}^{g-1}(g-b)\,\overline{w}_{b,1} \tag{71}$$

$$= \frac{G-g+1}{G}\sum_{b=1}^{g-2} b\overline{w}_{b+1,1} + \frac{g-1}{G}\sum_{b=g-1}^{G-1}(G-b)\,\overline{w}_{b+1,1} \tag{72}$$

To compute the variance of this estimator, we need the following equalities:

$$cov\,(w_{g,1}, w_{g+k,1}) = 1_{\{k=0\}} V\,(w_{g,1}) = 1_{\{k=0\}} 2n\sigma^2 \tag{73}$$

$$cov\,(w_{g,2}, w_{g+k,2}) = 1_{\{k=0\}} V\,(w_{g,2}) = 1_{\{k=0\}} 2n\sigma^2 \tag{74}$$

$$cov\,(w_{g,1}, w_{g+k,2}) = -1_{\{k=0\}} 2\,(n-m)\,\sigma^2 - 1_{\{k=1\}} 2m\sigma^2 \tag{75}$$

Using these formulas, we obtain:

$$cov\,(\overline{w}_{g,1}, \overline{w}_{g+k,1}) = \frac{2\sigma^2}{m\,(n-m)}\left[2n1_{\{k=0\}} - 1_{\{k=1\}}\right] \tag{76}$$

We use expression (76) to get the variance of group fixed effects in year 1:

$$G^2 \frac{m\,(n-m)}{4\sigma^2} V\left(\widehat{\beta}_{g,1}\right) = \frac{1}{6} G\,(n-1)\,(1+2h\,(g))\,h\,(g) + \frac{1}{2} Gh\,(g) \tag{77}$$

with $h\,(g) = (g-1)\,(G-g+1)$.

We now compute the variance of group fixed effects in year 2. Using (65), we get:

$$\widehat{\beta}_{g,2} = \frac{1}{n}\left[w_{g,2} + m\widehat{\beta}_{g-1,1} + (n-m)\,\widehat{\beta}_{g,1}\right] \tag{78}$$

We have $V\,(w_{g,2}) = 2n\sigma^2$. Moreover:

$$cov\left(w_{g,2}, \widehat{\beta}_{g,1}\right) = \frac{G-g+1}{G}\sum_{b=1}^{g-2} b\,cov\,(w_{g,2}, \overline{w}_{b+1,1}) \tag{79}$$

$$-\frac{g-1}{G}\sum_{b=g-1}^{G-1}(G-b)\,cov\,(w_{g,2}, \overline{w}_{b+1,1}) \tag{80}$$

For any b, it is possible to check that $cov\,(w_{g,2}, \overline{w}_{b,1}) = 0$. Consequently, we get:

$$cov\left(w_{g,2}, \widehat{\beta}_{g-1,1}\right) = cov\left(w_{g,2}, \widehat{\beta}_{g,1}\right) = 0 \tag{81}$$

Moreover, we have:

$$
\begin{aligned}
G^2 cov\left(\widehat{\beta}_{g-1,1},\widehat{\beta}_{g,1}\right) &= (G-g+1)(G-g+2)\sum_{b=1}^{g-2}\sum_{c=1}^{g-3} bc cov\left(\overline{w}_{b+1,1},\overline{w}_{c+1,1}\right)\\
&+(g-1)(g-2)\sum_{b=g-1}^{G-1}\sum_{c=g-2}^{G-1}(G-b)(G-c)\, cov\left(\overline{w}_{b+1,1},\overline{w}_{c+1,1}\right)\\
&+(G-g+1)(g-2)\sum_{b=1}^{g-2}\sum_{c=g-2}^{G-1} b(G-c)\, cov\left(\overline{w}_{b+1,1},\overline{w}_{c+1,1}\right)\\
&+(g-1)(G-g+2)\sum_{b=g-1}^{G-1}\sum_{c=1}^{g-3}(G-c)\, b cov\left(\overline{w}_{b+1,1},\overline{w}_{c+1,1}\right) \quad (82)
\end{aligned}
$$

We then get:

$$
\begin{aligned}
&\frac{G^2 m(n-m)}{2\sigma^2} cov\left(\widehat{\beta}_{g-1,1},\widehat{\beta}_{g,1}\right)\\
=\;& (G-g+1)(G-g+2)\left[2n\sum_{b=1}^{g-3} b^2 - 2\sum_{b=1}^{g-4} b(b+1)-(g-2)(g-3)\right]\\
&+(g-1)(g-2)\left[2n\sum_{b=1}^{G-g+1} b^2 - 2\sum_{b=1}^{G-g+1} b(b-1)-(G-g+1)(G-g+2)\right]\\
&+(G-g+1)(g-2)\left[\begin{array}{c} 2n(g-2)(G-g+2)\\ -(g-3)(G-g+2)-(g-2)(G-g+1)\end{array}\right]\\
&+0 \quad (83)
\end{aligned}
$$

simplifying (83), we obtain:

$$
\begin{aligned}
\frac{G^2 m(n-m)}{2\sigma^2} cov\left(\widehat{\beta}_{g-1,1},\widehat{\beta}_{g,1}\right) &= \frac{2}{3}G(n-1)h(g)h(g-1) \quad (84)\\
&+\frac{G}{2}\left[h(g)+h(g-1)+1-G\right]
\end{aligned}
$$

Using formulas (78), (81) and (84), we compute the variance of group fixed effects in year 2:

$$
V\left(\widehat{\beta}_{g,2}\right)=\frac{4\sigma^2}{Gn^2}\left[\begin{array}{c}\frac{n-1}{3m(n-m)}\left[(n-m)h(g)+mh(g-1)\right]^2\\ +\frac{1}{2}\frac{n}{m(n-m)}\left[(n-m)h(g)+mh(g-1)\right]\\ +\frac{1}{6}(n-1)\left[\frac{(n-m)^2}{m(n-m)}h(g)+\frac{m^2}{m(n-m)}h(g-1)\right]+\frac{1}{2}(1-G)\end{array}\right] \quad (85)
$$

To obtain the relative convergence rate of N and G for Assumption $A5$ to be verified, we need to compute the leading term of $\sum_g h(g)^2$ when G tends to infinity. In fact, it is the leading term of:

$$G^2\sum_g g^2 + \sum_g g^4 - 2G\sum_g g^3 \tag{86}$$

Taking the leading term of each sum, it gives: $\frac{1}{3}G^5 + \frac{1}{5}G^5 - \frac{1}{2}G^5 = \frac{1}{30}G^5$. It is easy to check that the leading term is the same for $\sum_g h(g-1)\,h(g)$. Consequently, the leading term of $trV\left(\widehat{\beta}\right)$ writes: $\frac{1}{45}\frac{nG^3}{m(n-m)}\sigma^2$.

References

[1] Abowd J. Kramarz F. and D. Margolis (1999), "High Wage Workers and High Wage Firms", *Econometrica*, 67(2), pp. 251-334.

[2] Amemiya T. (1978), "A Note on Random Coefficients Model", *International Economic Review*, 19(3), pp. 793-796.

[3] Anselin L. and R. Florax (2002), *Advances in Spatial Econometrics*, Heildeberg: Stringer-Verlag.

[4] Arellano M. (1987), "Computing Robust Standard Errors for Within-Groups Estimators", *Oxford Bulletin of Economics and Statistics*, 49(4), pp. 431-434.

[5] Bertrand M., Duflo E. and S. Mullainathan (2004), "How Much Should We Trust Differences-in-differences Estimates?", *The Quaterly Journal of Economics*, 119(1), pp. 249-275.

[6] Blundell R. and F. Windmeijer (1997), "Cluster Effects and Simultaneity in Multilevel Models", *Health Economics Letters*, 6(4), pp. 439-443.

[7] Borovkov A.A. (1998), *Probability Theory*, Gordon and Breach Science Publishers.

[8] Bourgignon F., Fournier M. and M. Gurgand, "Selection Bias Corrections Based on the Multinomial Logit Model: Monte Carlo Comparisons", DELTA Working Paper n°2004-20.

[9] Case A.C. (1991), "Spatial Patterns in Household Demand", *Econometrica*, 59(4), pp. 953-965.

[10] Ciccone A. et R.E. Hall (1996), "Productivity and the Density of Economic Activity", *The American Economic Review*, 86(1), pp. 54-70.

[11] Ciccone A. (2002), "Agglomeration Effects in Europe", *European Economic Review*, 46(2), pp. 213-227.

[12] Combes P.Ph., Duranton G. and L. Gobillon (2003), "Wage Differences across French Local Labour Market: Endowments, Skills and Interactions", CEPR Discussion Paper Series 4240.

[13] Cressie N. (1993), *Statistics in Spatial Data* (revised version). Wiley. New York.

[14] Dahl G. B. (2002), "Mobility and the Returns to Education: Testing a Roy Model with Multiple Markets", *Econometrica*, 70(6), pp. 2367-2420.

[15] Donald S. and K. Lang (2001), "Inferences with Difference in Differences and Other Panel Data", Working Paper, Boston University.

[16] Dubin J. A. and D. L. McFadden (1984), "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption", *Econometrica*, 52(2), pp. 345-362.

[17] Gibbons R. and L. Katz (1992), "Does Unmeasured Ability Explain Inter-Industry Wage Differentials", *Review of Economic Studies*, 59(3), pp. 515-535.

[18] Gibbons S. and Machin S. (2003), "Valuing English Primary Schools", *Journal of Urban Economics*, 53(2), pp. 197-219.

[19] Glaeser E.L., Kallal H.D., Scheinkman J.A., Shleifer A. (1992), "Growth in Cities", *The Journal of Political Economy*, 100(6), pp. 1126-1152.

[20] Goldstein H. (1986), "Multilevel Mixed Linear Model Analysis Using Iterative Generalised Least Squares", *Biometrika*, 73, pp. 43-56.

[21] Goldstein H. (1995), *Multilevel Statistical Models*, Second Edition, London, Edward Arnold.

[22] Gurgand M. (2003), "Farmer Education and the weather: evidence from Taiwan (1976-1992)", *Journal of Development Economics*, 71(1), pp. 51-70.

[23] Hausman J. A. and W. E. Taylor (1981), "Panel Data and Unobservable Individual Effects", *Econometrica*, 49, pp. 1377-1398.

[24] Inoue A. and G. Solon (2004), "A Portemanteau Test for Serially Correlated Errors in Fixed Effects Models", Working Paper.

[25] Kezdi G. (2002), "Robust Standard Error Estimation In Fixed-Effect Panel Models", Working Paper, University of Michigan.

[26] Kiefer N. M. (1980), "Estimation of Fixed Effects Models for Time Series of Cross-Sections with Arbitrary Intertemporal Covariance", *Journal of Econometrics*, 14(2), pp. 195-202.

[27] Krueger A.B. and L.H. Summers (1988), "Efficiency Wages and the Inter-Industry Wage Structure", *Econometrica*, 56(2), pp. 259-293.

[28] Lee L. F. (1983), "Generalized Econometric Models with selectivity", *Econometrica*, 51(2), pp. 507-512.

[29] 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.

[30] Pepper, J. V. (2002), "Robust Inferences from Random Clustered Samples: an Application using Data from the Panel Study of Income Dynamics", *Economics Letters*, 75(3), pp341-345.

[31] Rauch J. (1993), "Productivity Gains fron Geographic Concentration of Human Capital: Evidence from Cities", *Journal of Urban Economics*, 34(3), pp. 380-400.

[32] Rice N., Andrew A. M. and H. Goldstein (2002), "Multilevel Models Where the Random Effects are Correlated With the Fixed Predictors", Working Paper.

[33] White H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", *Econometrica*, 48(4), pp. 817-838.

[34] Wooldridge J. M. (2002), *Econometric Analysis of Cross-Section and Panel Data*, Cambridge, MIT Press.

[35] Wooldridge J. M. (2003), "Cluster-Sample Methods in Applied Econometrics", *American Ecoonmic Review*, 93(2), pp. 133-138.