Econométrie des données de panel

Guillaume Horny*

*Banque de France

Master 2 MASERATI

Chapitre 1

Plan

Plan

Modèle général

Le modèle le plus général possible est :

$$y_{it} = \alpha_{it} + x_{it}^{'}\beta_{it} + \epsilon_{it},$$

où y_{it} est scalaire, x_{it} un vecteur de dimension $(Kx1)$ contenant les variables, et ϵ_{it} un terme d'erreur scalaire.

Problème : Ce modèle idéal comprend $NT + NTK$ paramètres, bien plus que le nombre d'observations.
$\Rightarrow$ pour que le modèle soit identifié, des contraintes doivent être posées sur la manière dont α_{it} et β_{it} évoluent dans les deux dimensions, de même que sur la distribution des ϵ_{it}.

Plan

Modèle empilé (*pooled model*)

L'approche la plus simple consiste à empiler les NT observations et à reprendre un modèle linéaire pour données en coupe :

$$y_{it} = \alpha + x_{it}^{'}\beta + \epsilon_{it},$$

Si :

- il n'y a pas d'hétérogénéité inobservée,
- et que les variablesexplicatives sont sans corrélation avec le terme d'erreur ($\text{cov}(x_{it}, \epsilon_{it}) = 0$),

alors l'estimateur OLS est convergent.

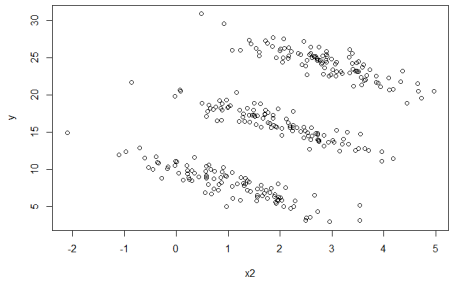
Modèle empilé (*pooled model*)

Attention :

- Les résidus de l'individu i sont très certainement autocorrélés. L'approximation habituelle de la variance de l'estimateur OLS, basée sur l'hypothèse d'observations i.i.d, n'est plus valide ici. Elle conduit à des t de Student surévalués.
  $\Rightarrow$ besoin d'utiliser des estimateurs de la matrice de variance corrigés.
- Si le vrai modèle comprend un terme d'hétérogénéité inobservée α_i corrélé avec x, alors l'estimateur OLS est biaisé (exercice !).

Exemple de biais des OLS dans le modèle empilé

On a simulé dans l'exemple de l'introduction les données suivantes :



Le "truc" est que deux variables corrélées, x1 et x2, ont été utilisées pour simuler les données. On n'en a toutefois utilisé qu'une seule (x2) pour la régression OLS. Du coup, l'influence de la variable omise a été capturée par le terme d'erreur, qui s'est retrouvé corrélée avec la seule explicative du modèle.

$\Rightarrow$ problème d'endogénéité

$\Rightarrow$ estimateur OLS biaisé

Plan

Pourquoi s'intéresser à ce modèle ?

- modélisation qui vient spontanément à l'esprit
- relativement simple à estimer
- suffit à exploiter les différents avantages des modèles pour données de panel présentés dans l'introduction
- très répandu en pratique (renouveau)
- Attention, bien que simple en apparence, le modèle a de nombreuses implications qui ne sont pas toutes évidentes

Le modèle à effets fixes individuels

$$y_{it} = \alpha_i + x_{it}^{'}\beta + \epsilon_{it},$$

où α_i est **constant** dans le temps et propre à chaque unité statistique. Son influence sur la variable dépendante est la même à chaque date

Le modèle à effets fixes individuels et indicatrices temporelles

On peut réécrire $x_{it}^{'}\beta$ pour faire apparaître des effets temporels (tendance, saisonnalité...). En effet, si x contient des indicatrices de période :

$$\begin{aligned} y_{it} =& \alpha_i + x_{it}^{'}\beta + \epsilon_{it} \\ =& \alpha_i + d_t^{'}\beta_1 + z_{it}^{'}\beta_2 + \epsilon_{it}, \end{aligned}$$

où $x = (d, z)$ et $\beta = (\beta_1, \beta_2)$.

Dans la suite du cours, on ne fera généralement pas apparaître explicitement les effets temporels. Ils peuvent néanmoins être introduits simplement au moyen d'indicatrices (annuelles, trimestrielles...). Pour $N \to \infty$ et T est fini, l'estimations des coefficients des indicatrices temporelles ne pose pas de difficulté particulière.

Les hypothèses du modèle à effets fixes : les explicatives

Pour éviter différents problèmes d'endogénéité, les variables explicatives doivent être exogènes. Cependant :

- Plusieurs hypothèses sur les variables explicatives sont possibles. Elles varient dans les contraintes qu'elles imposent et dans les résultats qu'elles permettent d'obtenir.
- La pratique n'est pas harmonisée, les différents cours et manuels utilisent des hypothèses différentes. Pour ne rien simplifier, chacune peut être fomulée de plusieurs manières équivalentes.

$\Rightarrow$ Tout ceci peut donner une impression de flou et de confusion, alors que les résultats sont pourtant bien établis !

On va voir trois hypothèses d'exogénéité : l'**exogénéité stricte**, l'**exogénéité stricte conditionnelle** aux inobservables, et l'**absence de corrélation**.

L'exogénéité stricte

La plus contraignante est l'hypothèse **d'exogènéité stricte** :

$$\mathrm{E}(y_{it}|x_{i1},\ldots,x_{iT}) = \mathrm{E}(y_{it}|x_{it}) = x_{it}^{'}\beta.$$

Interprétation :

- lorsque x_{it} est prit en compte, les $x_{is}, s \neq t$, n'ont plus d'effet sur y_{it}
- le modèle est linéaire dans les paramètres

Les hypothèses sur les explicatives

Reformulation : Elle est parfois présentée de la manière suivante :

$$\mathrm{E}(\epsilon_{it}|x_{i1},\ldots,x_{iT}) = \mathrm{E}(\epsilon_{it}|x_{it}) = 0.$$

La première espérance est conditionnelle aux explicatives à toute date. C'est en général une fonction non-linéaire de $x_{i1},\ldots,x_{iT}$. L'exogénéité stricte suppose que cette fonction est une constante égale à 0.

Implications de l'exogénéité stricte (1/2)

- $\mathrm{E}(\epsilon_{it}) = 0$ car $\mathrm{E}_X[\mathrm{E}(\epsilon_{it}|x_{i1}, \ldots, x_{iT})] = \mathrm{E}(\epsilon_{it})$
- $\mathrm{E}(x_{is}^{'}\epsilon_{it}) = 0, \forall s, \forall t$. En effet :

$$\begin{aligned} \mathrm{E}(\epsilon_{it}|x_{i1}, \ldots, x_{iT}) &= 0 \\ \iff x_{is}^{'}\mathrm{E}(\epsilon_{it}|x_{i1}, \ldots, x_{iT}) &= 0 \\ \iff \mathrm{E}(x_{is}^{'}\epsilon_{it}|x_{i1}, \ldots, x_{iT}) &= 0 \\ \Rightarrow \mathrm{E}_X\left[\mathrm{E}(x_{is}^{'}\epsilon_{it}|x_{i1}, \ldots, x_{iT})\right] &= \mathrm{E}(x_{is}^{'}\epsilon_{it}) = 0 \end{aligned}$$

$\Rightarrow$ les **erreurs** sont sans corrélation avec chacune des explicatives **à toute date** (hypothèse plus forte qu'à la seule date t !).
$\Rightarrow$ Elles sont sans corrélation avec n'importe quelle fonction des explicatives : $x_{i1}^2, x_{i1}x_{i2}$ ou encore $\ln(x_{it} + 1)$. En d'autres termes, la manière dont y_{it} dépend des x est complètement spéficiée.

Implications de l'exogénéité stricte (2/2)

Sous :

$$y_{it} = \alpha_i + x_{it}^{'}\beta + \epsilon_{it}.$$

On a :

$$\begin{aligned}
\mathsf{E}(y_{it}|x_{i1},\ldots,x_{iT}) &= x_{it}^{'}\beta + \mathsf{E}(\alpha_i|x_{i1},\ldots,x_{iT}) + \mathsf{E}(\epsilon_{it}|x_{i1},\ldots,x_{iT}) \\
&= x_{it}^{'}\beta + \mathsf{E}(\alpha_i|x_{i1},\ldots,x_{iT}) \qquad \text{(par déf. exo. stricte)} \\
&= x_{it}^{'}\beta \qquad \text{(par définition exo. stricte)} \\
\iff \mathsf{E}(\alpha_i|x_{i1},\ldots,x_{iT}) &= 0 \\
\Rightarrow \mathsf{E}(x_{it}^{'}\alpha_i) &= 0, \forall t.
\end{aligned}$$

L'exogénéité stricte implique également que les **inobservables** ne soient pas corrélés avec les explicatives.

Exogénéité conditionnelle aux inobservables

Certains auteurs préfèrent utiliser l'hypothèse **d'exogènéité stricte conditionnelle aux inobservables** :

$$\mathsf{E}(y_{it}|x_{i1},\ldots,x_{iT},\alpha_i) = \mathsf{E}(y_{it}|x_{it},\alpha_i) = \alpha_i + x_{it}'\beta.$$

Interprétation :

- lorsque x_{it} et α_i sont pris en compte, $x_{is}, s \neq t$, n'a plus d'effet sur y_{it}
- le modèle est linéaire dans les paramètres

Implications de l'exo. conditionnelle aux inobservables (1/2)

Si :

$$y_{it} = \alpha_i + x_{it}'\beta + \epsilon_{it}.$$

Sous l'exogénéité stricte conditionnelle, on a :

$$\begin{aligned} & \mathrm{E}(y_{it}|x_{i1}, \ldots, x_{iT}, \alpha_i) = \alpha_i + x_{it}'\beta \\ \Longleftrightarrow\ & \mathrm{E}(\epsilon_{it}|x_{i1}, \ldots, x_{iT}, \alpha_i) = 0. \end{aligned}$$

Reformulation : Elle est parfois présentée de manière équivalente :

$$\mathrm{E}(\epsilon_{it}|x_{i1}, \ldots, x_{iT}, \alpha_i) = \mathrm{E}(\epsilon_{it}|x_{it}, \alpha_i) = 0.$$

- Les erreurs sont sans corrélation avec les explicatives et l'hétérogénéité inobservée, car $\mathrm{E}(\epsilon_{it}|x_{it}, \alpha_i) = 0 = \mathrm{E}(\epsilon_{it})$,
- on ne suppose rien sur les relations entre les inobservables et les explicatives : ils peuvent être corrélés !

Implications de l'exo. conditionnelle aux inobservables (2/2)

Hypothèse plus générale que l'exogénéité stricte

Sous exogénéité stricte, on a pour un modèle linéaire à effets fixes :

$$\begin{aligned}
\mathsf{E}(y_{it}|x_{i1},\ldots,x_{iT}) &= x_{it}'\beta + \mathsf{E}(\alpha_i|x_{i1},\ldots,x_{iT}) + \mathsf{E}(\epsilon_{it}|x_{i1},\ldots,x_{iT}) \\
&= x_{it}'\beta + \mathsf{E}(\alpha_i|x_{i1},\ldots,x_{iT}) \\
&= x_{it}'\beta \\
&\iff \mathsf{E}(\alpha_i|x_{i1},\ldots,x_{iT}) = 0.
\end{aligned}$$

- L'exogénéité stricte implique donc que les α_i soient sans corrélation avec les explicatives à toute date.
- À l'inverse, sous l'exogénéité stricte conditionnelle aux inobservables, on ne fait aucune hypothèse sur $\mathsf{E}(\alpha_i|x_{i1},\ldots,x_{iT})$.

Conditionnelle aux inobservables ou non ?

En pratique, l'hypothèse d'exogénéité stricte est moins crédible que celle d'exogénéité stricte conditionnelle

Exemple : des fermes i produisent du maïs en quantité y_{it} aux années t,

- x_{it} mesure le capital, travail, engrais, pluviométrie...
- les inobservables α_i mesurent le talent du fermier, la qualité des sols...
- la production en t dépend de x_{it} et de α_i

Il est difficile de croire que les agriculteurs ne vont pas chercher à compenser un manque de talent par plus d'input, comme des engrais.

Les hypothèses sur les explicatives : absence de corrélation

D'autres auteurs préfèrent supposer **l'absence de corrélation** entre les explicatives et les erreurs :

$$\mathrm{E}(x_{is}^{'}\epsilon_{it}) = 0, \forall s, \forall t.$$

C'est une conséquence de l'exogénéité stricte conditionnelle aux inobservables. En effet :

$$\begin{aligned}
\mathrm{E}(\epsilon_{it}|x_{i1},\ldots,x_{iT},\alpha_i) &= 0 \\
\iff x_{is}^{'}\mathrm{E}(\epsilon_{it}|x_{i1},\ldots,x_{iT},\alpha_i) &= 0 \\
\iff \mathrm{E}(x_{is}^{'}\epsilon_{it}|x_{i1},\ldots,x_{iT},\alpha_i) &= 0 \\
\Rightarrow \mathrm{E}_{X,\alpha}\left[\mathrm{E}(x_{is}^{'}\epsilon_{it}|x_{i1},\ldots,x_{iT},\alpha_i)\right] &= \mathrm{E}(x_{is}^{'}\epsilon_{it}) = 0
\end{aligned}$$

Implications communes des hypothèses sur les explicatives (1/2)

L'exogénéité stricte, l'exogénéité stricte conditionnelle aux inobservables ainsi que l'absence de corrélation impliquent toutes 3 que certaines variables ne peuvent pas être parmi les x :

- les variables dépendantes retardées
- les variables influencées par :
  - des anticipations (car alors x_{it-1} est corrélée avec ϵ_{it})
  - des feedback (car alors x_{it+1} est corrélée avec ϵ_{it})
- et plus généralement les variables endogènes

Implications communes des hypothèses sur les explicatives (2/2)

Lorsque le modèle comprend des variables endogènes, des techniques dédiées doivent être utilisées (variables instrumentales...)

Dans la suite du chapitre, sauf mention contraire, on suppose **l'exogénéité stricte conditionnelle** aux inobservables

Plan

Les effets fixes individuels : une autre écriture (1/2)

Pour souligner le fait que les effets fixes sont des **paramètres** à estimer, on voit parfois l'écriture faisant explicitement apparaître des indicatrices d'individus.

Soit $d_{ilt} = 1, \forall t$ si $l = i$, et 0 sinon. On peut écrire le modèle à effets individuels :

$$\begin{aligned} y_{it} &= \alpha_{i1} d_{i1t} + \ldots + \alpha_{iN} d_{iNt} + x_{it}^{'} \beta + \epsilon_{it} \\ &= \sum_{l=1}^{N} \alpha_{il} d_{ilt} + x_{it}^{'} \beta + \epsilon_{it}. \end{aligned}$$

Les effets fixes individuels : une autre écriture (1/2)

L'écriture matricielle serait :

$$\begin{pmatrix} y_{11} \\ \vdots \\ y_{1T} \\ y_{21} \\ \vdots \\ y_{NT} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & 0 \\ 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix} + \begin{pmatrix} x_{11}^1 & \dots & x_{11}^K \\ \vdots & \dots & \vdots \\ x_{1T}^1 & \dots & x_{1T}^K \\ x_{21}^1 & \dots & x_{21}^K \\ \vdots & \dots & \vdots \\ x_{NT}^1 & \dots & x_{NT}^K \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_K \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1T} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{NT} \end{pmatrix}.$$

Attention toutefois à la contrainte d'identification (le fort bien nommé *dummy trap*).

La contrainte d'identification (a.k.a *dummy trap*) (1/2)

Si on a N effets individuels, leurs indicatrices sont parfaitement colinéaires avec la constante : $\sum_{l=1}^{N} d_{ilt} = \iota$, où ι est un vecteur de dimension NT ne comprenant que des 1.
$\Rightarrow$ la matrice comprenant à la fois les indicatrices d'individus et la constante n'est plus de plein rang, donc elle n'est pas inversible
$\Rightarrow$ les méthodes de moindres carrés ne s'appliquent plus.

La contrainte d'identification (a.k.a *dummy trap*) (2/2)

Deux solutions :

- éliminer la constante
- garder la constante et éliminer l'indicatrice d'un individu (disons celle de l'individu j). On dit alors que j est **l'individu de référence**.

En effet, considérons le cas où x est une matrice ne contenant qu'une constante. On a pour l'individu j :

$$\mathrm{E}(y_{jt}|x_{jt}) = \beta_0.$$

Pour les autres :

$$\mathrm{E}(y_{it}|x_{it}) = \beta_0 + \alpha_i, \forall i \neq j.$$

Les α_i s'interprêtent alors comme la déviation de $\mathrm{E}(y_{it}|x_{it})$ par rapport à l'individu j, qui devient de ce fait l'individu de référence.

Estimation directe par OLS

Comme l'écriture du modèle avec des indicatrices d'individu le laisse penser, on peut estimer (α, β) simplement par OLS appliqué au modèle avec indicatrices.

Deux difficultés en pratique :

1. difficultés numériques lorsqu'il y a beaucoup d'individus
2. lorsque T est petit, l'estimation des indicatrices est peu précise

Plan

Alternative : l'estimateur intra-individuel (*within*)

On peut estimer β après avoir transformé le modèle de manière à faire disparaître les α_i. Plusieurs techniques sont possibles, on va voir ici celle reposant sur la transformation *within*.

L'estimateur intra-individuel (*within*)

Le modèle est :

$$y_{it} = \alpha_i + x_{it}^{'}\beta + \epsilon_{it}.$$

En prenant les moyenne intra-individuelles, on a ...

$$y_{i.} = \alpha_i + x_{i.}^{'}\beta + \epsilon_{i.}.$$

... que l'on peut soustraire du modèle initial...

$$\begin{aligned} y_{it} - y_{i.} &= (\alpha_i - \alpha_i) + (x_{it}^{'} - x_{i.}^{'})\beta + (\epsilon_{it} - \epsilon_{i.}) \\ &= (x_{it}^{'} - x_{i.}^{'})\beta + (\epsilon_{it} - \epsilon_{i.}) \end{aligned}$$

... de manière à faire disparaître les inobservables !

L'estimateur intra-individuel (*within*)

L'estimateur *within* est l'estimateur OLS appliqué au modèle :

$$y_{it} - y_{i.} = (x_{it}^{'} - x_{i.}^{'})\beta + u_{it}.$$

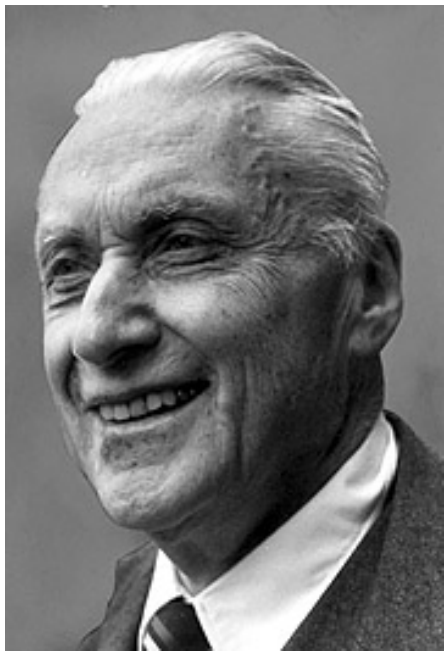
Intuition : Comme les inobservables ne varient pas dans le temps, toutes les variations de y d'une période à l'autre pour un individu donné peuvent être imputées à des changements dans les explicatives ou à de l'aléa.

L'estimateur *within* : avantages et inconvénient

+ Facile à calculer, il suffit de transformer les données avant d'appliquer les OLS.
+ Numériquement beaucoup plus stable que l'estimation par OLS du modèle avec des indicatrices individuelles. Les OLS demandent d'inverser $X'X$, qui est de dimension $(N+K)x(N+K)$ avec des indicatrices individuelles et (KxK) après transformation *within*.
- La transformation élimine les variables constantes dans le temps. On ne donc peut estimer de coefficient que pour les explicatives qui varient dans le temps (le problème est propre au modèle à effets fixes, il se retrouve donc aussi avec des indicatrices d'individu).

Going deeper...

Les OLS sur modèle avec indicatrices et l'estimation *within* donnent les mêmes résultats du fait du **théorème de Frisch-Waugh-Lovell**.



Ragnar Frisch (1895-1973), norvégien et cotitulaire avec Jan Tinbergen du premier prix Nobel d'économie. L'un des pères fondateurs de l'économétrie, aurait créé les termes "macroéconomie", "microéconomie", "économétrie" (les deux premiers en 1933 !).

Le théorème de Frisch-Waugh-Lovell

- L'estimation de β_2 dans la régression :

$$Y = X_1\beta_1 + X_2\beta_2 + \epsilon.$$

sera identique à son estimation dans la régression :

$$M_{X_1} Y = M_{X_1} X_2 \beta_2 + M_{X_1} \epsilon,$$

où $M_{X_1} = I - X_1(X_1'X_1)^{-1}X_1'$.

- En remplaçant X_1 par les indicatrices d'individus, on retrouve la transformation *within* (exercice !)
- Le théorème est aussi utilisé dans d'autres contextes pour éliminer des indicatrices de période

Propriétés de l'estimateur *within*

- **Convergence**
  C'est l'estimateur OLS, il est donc :
  - sans biais sous l'hypothèse d'exogénéité forte conditionnelle aux inobservables
  - convergent dès lors que les erreurs sont sans corrélation avec les variables explicatives
- **Efficacité**
  Il est de variance minimale sous l'hypothèse d'exogénéité forte conditionnelle aux inobservables et d'homoscédasticité
  $\mathrm{E}(\epsilon_i \epsilon_i^{'} | x_i, \alpha_i) = \sigma_\epsilon^2 I_T$.

Démonstration : absence de biais de l'estimateur *within* (1/2)

Après transformation *within*, le modèle s'écrit :

$$MY = MX_2\beta_2 + M\epsilon,$$

où $M = I - X_1(X_1'X_1)^{-1}X_1'$. L'estimateur *within* a pour expression :

$$\widehat{\beta}_W = \left((MX_2)'MX_2\right)^{-1}(MX_2)'MY.$$

Comme M est symétrique et idempotente ($M = MM'$), on peut simplifier en :

$$\widehat{\beta}_W = \left(X_2'MX_2\right)^{-1}X_2'MY.$$

Absence de biais de l'estimateur *within* (2/2)

Si on remplace MY par $MX_2\beta_2 + M\epsilon$, on a :

$$\widehat{\beta}_W = \left(X_2'MX_2\right)^{-1} X_2'MX_2\beta_2 + \left(X_2'MX_2\right)^{-1} X_2'M\epsilon.$$

D'où :

$$\begin{aligned} \mathrm{E}(\widehat{\beta}_W|X) &= \beta_2 + \left(X_2'MX_2\right)^{-1} X_2'M\mathrm{E}(\epsilon|X) \\ &= \beta_2 \qquad \text{(par l'exogénéité conditionnelle)} \end{aligned}$$

$$\Rightarrow \mathrm{E}_X[\mathrm{E}(\widehat{\beta}_W|X)] = \mathrm{E}(\widehat{\beta}_W) = \beta_2.$$

Variance de l'estimateur *within* (1/3)

$$\begin{aligned}
\text{Var}(\widehat{\beta}_W|X) &= \text{Var}\left[\left(X_2'MX_2\right)^{-1}X_2'MX_2\beta_2|X\right] + \text{Var}\left[\left(X_2'MX_2\right)^{-1}X_2'M\epsilon|X\right] \\
&= 0 + A\text{Var}\left[\epsilon|X\right]A' \\
&= A\text{E}\left[\epsilon\epsilon'|X\right]A' \\
&= \widehat{\sigma}_\epsilon^2 AA' \\
&= \widehat{\sigma}_\epsilon^2 (X_2'MX_2)^{-1}
\end{aligned}$$

car $A = \left(X_2'MX_2\right)^{-1}X_2'M$,

et $AA' = \left(X_2'MX_2\right)^{-1}X_2'MM'X_2\left(X_2'MX_2\right)^{-1} = (X_2'MX_2)^{-1}$.

On peut simplifier :

$$\text{Var}(\widehat{\beta}_W|X) = \widehat{\sigma}_\epsilon^2\left(\sum_i\sum_t (X_{2it} - X_{2i.})'(X_{2it} - X_{2i.})\right)^{-1}.$$

Variance de l'estimateur *within* (2/3)

- Un estimateur convergent et sans biais de σ_ϵ^2 est :

$$\widehat{\sigma}_\epsilon^2 = \frac{1}{N(T-1)-K} \sum_i \sum_t \widehat{\epsilon}_{it}^2.$$

- Attention : le dénominateur fait intervenir le nombre d'effets fixes (N). Si on estime le modèle après transformation *within* manuelle par une routine OLS standard, le denominateur utilisé sera $NT - K$, d'où des $\widehat{\sigma}_\epsilon^2$ sous-estimés et des t de Student surestimés.

Variance de l'estimateur *within* (3/3)

Optimalité :
Comme l'estimateur *within* est un estimateur OLS, le théorème de Gauss-Markov s'applique et il est de plus petite variance parmi les estimateurs linéaires sans biais. Ce résultat requiert l'exogénéité forte conditionnelle aux inobservables et l'homoscédasticité.

Convergence de l'estimateur *within*

Dans le cas des panels de données microéconomiques, on a souvent $T < \infty$ et $N \to \infty$. On s'intéresse donc avant tout à la convergence avec le nombre d'individus :

$$p \lim_{N \to \infty} (\widehat{\beta} - \beta).$$

Convergence de l'estimateur *within*

Comme :

$$Y = X_2\beta + \epsilon,$$

où Y est de dimension $(NTx1)$, X de dimension $(NTxK)$, β de dimension $(Kx1)$ et ϵ de dimension $(NTxK)$.

$$\begin{aligned}\widehat{\beta}_W &= \left(X_2' M X_2\right)^{-1} X_2' M Y \\ &= \left(X_2' M X_2\right)^{-1} X_2' M (X_2\beta + \epsilon).\end{aligned}$$

Convergence de l'estimateur *within*

D'où :

$$p \lim_{N\to\infty}(\widehat{\beta}-\beta) = p \lim_{N\to\infty}\left[\left(X_2' M X_2\right)^{-1} X_2' M X_2 \beta\right] + p \lim_{N\to\infty}\left[\left(X_2' M X_2\right)^{-1} X_2' M \epsilon\right] - \beta.$$

- le premier terme vaut β
- le second vaut :

$$p \lim_{N\to\infty}\left(\frac{1}{NT}\sum_i\sum_t x_{2it} m_{it} x_{2it}'\right)^{-1}\left(\frac{1}{NT}\sum_i\sum_t x_{2it} m_{it}\epsilon_{it}\right).$$

Convergence de l'estimateur *within*

Ainsi, l'estimateur *within* est convergent ssi :

$$p \lim_{N\to\infty} \left(\frac{1}{NT} \sum_i \sum_t x_{2it} m_{it} \epsilon_{it} \right) \to 0.$$

Or :

$$p \lim_{N\to\infty} \left(\frac{1}{NT} \sum_i \sum_t x_{2it} m_{it} \epsilon_{it} \right) \to M\mathrm{E}(X_2 \epsilon),$$

qui est nul dès que les erreurs sont sans corrélation avec les explicatives.

Convergence de l'estimateur *within*

- On peut montrer que l'estimation des coefficients des effets fixes est convergentes sous $T \to \infty$ (intuitivement, plus on a d'observations pour un individu, plus on est à même d'estimer son effet fixe)
- En pratique, T est souvent petit et il est plus prudent de ne commenter que les classements des $\widehat{\alpha}_i$, ou leur groupement, que les valeurs elles-mêmes.

Récapitulatif des propriétés

- l'estimateur est sans biais sous l'exogénéité conditionnelle aux inobservable
- convergent sous l'absence de corrélation entre explicatives et erreurs
  $\Rightarrow$ la présence d'endogénéité est potentiellement devastatrice
- de variance minimale sous l'hypothèse d'homoscédasticité et d'exogénéité conditionnelle
  $\Rightarrow$ la présence d'hétéroscédasticité ne remet pas en cause les valeurs estimés des coefficients, seulement leurs écarts-type.

Conclusion intermédiaire

- si l'une des hypothèses d'exogénéité est vérifiée
  $\Rightarrow$ l'estimateur est sans biais et convergent
- si seule l'absence de corrélation est vérifié
  $\Rightarrow$ l'estimateur est convergent (cad asymptotiquement sans biais)
- si on a de plus l'homoscédasticité
  $\Rightarrow$ l'estimateur est efficace

... et si on n'a pas l'homoscédasticité ?

Autocorrélation et hétéroscédasticité (1/3)

L'estimateur de la variance que nous avons vu a été obtenu sous l'hypothèse d'erreurs i.i.d et homoscédastiques. En présence d'autocorrélation ou d'hétéroscédasticité, cet estimateur ne converge plus vers la vraie matrice de variance.

Or, il peut y avoir une autocorrélation dans les erreurs, qui introduit une autocorrélation entre les y_{it}.

La bonne nouvelle est qu'il suffit d'un ajustement de la matrice de White (1980) pour obtenir un estimateur de la matrice de variance robuste à l'hétéroscédasticité.

Autocorrélation et hétéroscédasticité (2/3)

L'estimateur *within* a pour variance :

$$\text{Var}(\widehat{\beta}_W|X) = A\text{E}\left[\epsilon\epsilon^{'}|X\right]A^{'}$$

où $A = \left(X_2^{'}MX_2\right)^{-1}X_2^{'}M$. Arellano (1987) montre qu'on peut estimer $X_2^{'}M\text{E}\left[\epsilon\epsilon^{'}|X\right]MX_2$ par :

$$\frac{1}{N}\sum_i(\underline{X}_{2i} - \underline{X}_{2i.})^{'}\widehat{u}_i\widehat{u}_i^{'}(\underline{X}_{2i} - \underline{X}_{2i.}),$$

où $\underline{X}_{2i}$ est un vecteur (TxK) des explicatives de l'individu i et $\widehat{u}_i$ est le vecteur $Tx1$ de ses résidus obtenus après estimation *within*.

Autocorrélation et hétéroscédasticité (3/3)

Cet estimateur de la variance est convergent sous :

- $N \to \infty$
- $\mathrm{E}(\epsilon_{it}\epsilon_{js}) = 0,$ pour $i \neq j$ (pas de corrélation des chocs d'un individu à l'autre).
- $\mathrm{E}(\epsilon_{it}\epsilon_{is})$ non spécifié (pas de forme particulière d'autocorrélation pour un individu donné)

Implication pratique

- Fondamentalement, la question est de savoir si ce sont les **observations** qui sont indépendantes ou bien les **unitées** suivies dans le temps
- En pratique, vérifier dans les documentations que les commandes de type "robust" comprennent bien que l'autocorrélation est au niveau de l'**individu** (pas de problème sous Stata après `xtset`). Sinon, les logiciels ont la facheuse tendance à considérer par défaut que l'autocorrélation est niveau des **observations**, ce qui aboutit à corriger seulement pour l'hétéroscédasticité. Or, avec des données de panel, le problème est généralement plus l'autocorrélation des erreurs individuelles que leur hétéroscédasticité !

Plan

Intuition

C'est la présence de N paramètres, les effets fixes individuels, qui rend le modèle à effet fixe difficile à estimer. On peut les éliminer :

- avec la transformation *within*,
- en calculant les différences premières

Le modèle en différence première

$$y_{it} - y_{it-1} = (\alpha_i - \alpha_i) + (x'_{it} - x'_{it-1})\beta + (\epsilon_{it} - \epsilon_{it-1})$$
$$\Delta y_{it} = \Delta x'_{it}\beta + \Delta\epsilon_{it}.$$

Implications :

- les β s'interprêtent exactement de la même manière qu'après transformation *within* : le modèle sous-jacent reste un modèle linéaire à effets fixes individuels
- cette écriture montre explicitement qu'on ne peut pas identifier les β des explicatives qui ne varient pas dans le temps
- on perd les observations de la première date, il reste $N(T-1)$ observations

L'estimateur du modèle en différence première

C'est l'estimateur OLS du modèle en différence première.

À première vue, comme il est plus facile de calculer différences d'une période à l'autre que les moyennes intra-individuelles, on serait tenté de privilégier l'estimateur en différence première par rapport à la transformation *within*. Malheureusement, la comparaison des deux estimateurs est un peu plus compliquée...

Ecriture matricielle

$$DY = DX\beta + D\epsilon,$$

où :

$$D = \begin{pmatrix} -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & \cdots & 0 & 0 & -1 & 1 \end{pmatrix}$$

Propriétés de l'estimateur du modèle en différence première

- il est **sans biais** si $E(\Delta\epsilon_{it}|\Delta x_{i1}, \ldots, \Delta x_{iT}) = 0$
  C'est l'hypothèse d'exogénéité stricte dans le modèle en différence.
- il est **convergent** si $E(\Delta x_{it}^{'} \Delta\epsilon_{it}) = 0, \forall t$.
  Ainsi, il suffit d'avoir $E(\epsilon_{it} - \epsilon_{it-1}|x_{it} - x_{it-1}) = 0$, ce qui est moins exigeant que l'exogénéité stricte.
- il n'est toutefois pas **efficace**.
  En effet, les erreurs sont autocorrélées (voir prochaines slides).
  L'estimateur efficace est de type MCG.

Autocorrélation des erreurs dans le modèle en différence

Sous l'hypothèse d'erreurs ϵ_{it} homoscédastiques et i.i.d dans le modèle **avant** différence :

$$\begin{aligned}
\text{cov}(\Delta\epsilon_{it}, \Delta\epsilon_{it-1}) &= \text{E}(\Delta\epsilon_{it}\Delta\epsilon_{it-1}) \\
&= \text{E}\left[(\epsilon_{it} - \epsilon_{it-1})(\epsilon_{it-1} - \epsilon_{it-2})\right] \\
&= \text{E}\left[\epsilon_{it}\epsilon_{it-1} - \epsilon_{it}\epsilon_{it-2} - \epsilon_{it-1}^2 + \epsilon_{it-1}\epsilon_{it-2}\right] \\
&= -\text{E}[\epsilon_{it-1}^2] \\
&= -\sigma_\epsilon^2 \leq 0.
\end{aligned}$$

$\Rightarrow$ les erreurs du modèle **en différence** sont négativement autocorrélées.
$\Rightarrow$ L'estimateur OLS du modèle **en différence** ne sera pas efficace.

Efficacité de l'estimateur du modèle en différence première

L'estimateur OLS est efficace si les erreurs du modèle **en différence** sont homoscédastiques et i.i.d.
Réécrivons le modèle en différence avec $\Delta\epsilon_{it} = e_{it}$:

$$\Delta y_{it} = \Delta x_{it}^{'}\beta + e_{it}.$$

Si les e_{it} sont indépendants, alors $\epsilon_{it} = \epsilon_{it-1} + e_{it}$ est une marche aléatoire.

$\Rightarrow$ L'estimateur du modèle en différence première est (presque) de plus petite variance si les erreurs du modèle avant différence (ϵ_{it}) sont fortement autocorrélées.

Variance de l'estimateur du modèle en différence première

Lorsque les erreurs du modèle en différence sont homoscédastiques (cad lorsque les erreurs du modèle avant différence sont fortement autocorrélées), on peut utiliser l'estimateur de la variance des OLS : :

$$\text{var}(\widehat{\beta}_{FD}) = \widehat{\sigma}_{\epsilon}^2 (\Delta X' \Delta X)^{-1},$$

où $\widehat{\sigma}_{\epsilon}^2$ est estimé par :

$$\widehat{\sigma}_{\epsilon}^2 = \frac{1}{N(T-1)-K} \sum_i \sum_t \widehat{e}_{it}^2,$$

avec $\widehat{e}_{it} = \Delta y_{it} - \Delta x_{it} \widehat{\beta}_{FD}$.

Sous ces hypothèses sur les erreurs, les écarts-types fournis par l'estimation OLS sont directement utilisables après calcul manuel des différences premières, à l'inverse de la transformation *within* manuelle.

Variance robuste

En cas de violation de l'hypothèse d'homoscédasticité des erreurs dans le modèle en différence, un estimateur convergent de la matrice de variance est fournit par :

$$\text{var}(\widehat{\beta}_{FD}) = (\Delta X' \Delta X)^{-1} \left(\sum_i \Delta X_i' \widehat{e}_{it} \widehat{e}_{it}' \Delta X_i \right) (\Delta X' \Delta X)^{-1}.$$

Plan

Principaux résultats

- Grâce au suivi des individus, on peut inclure dans le modèle des paramètres mesurant l'influence des caractéristiques individuelles inobservées constantes dans le temps
- Le modèle linéaire à effets fixes individuels peut être estimé :
  - en créant une indicatrice par individu
  - en appliquant la transformation *within*
  - en calculant les différences premières

  puis en estimant le modèle transformé par OLS

Principaux résultats

- Sous l'hypothèse d'exogénéité conditionnelle aux inobservables, les estimateurs sont tous 3 sans biais et convergent. Elle exclue toutefois les variables dépendantes retardées, et plus généralement les variables endogènes (erreurs de mesure, simultanéité...)
- Les deux premiers sont efficaces sous l'hypothèse d'homoscédasticité, le dernier sous l'hypothèse que les erreurs suivent une marche aléatoire
- On peut facilement obtenir des estimations de la matrice de variance en cas d'hétéroscédasticité ou d'autocorrélation des erreurs

Principaux résultats

- la présence d'effets fixes fait qu'on ne peut pas identifier les coefficients associés à des variables constantes dans le temps ...
- ... d'où l'intérêt de tester l'existence d'effets fixes ! (à suivre)
- la coïncidence de l'estimateur avec indicatrices d'individus et de l'estimateur *within* est fortuite. Dans les modèles non-linéaires, ajouter aux explicatives N indicatrices ne conduit généralement pas à un estimateur convergent !
- Il est parfois utile d'estimer les α_i. On peut les retrouver soit par les coefficients des indicatrices d'individus, soit par $\widehat{\alpha}_i = y_{it} - x_{it}^{'}\widehat{\beta}$, où $\widehat{\beta}$ est l'estimateur *within* ou du modèle en différence première. La précision de leur estimation dépend de T.

Comparaison estimateur *within* ou différence première (1/2)

- Pour $T = 2$, les estimations sont identiques
- Pour $T > 2$, on peut faire son choix selon :
  - la taille de l'échantillon : les deux estimateurs convergent mais l'estimateur *within* converge plus vite que celui en différence première
  - le comportement des résidus auquel on s'attend. S'il est raisonnable penser qu'ils sont homoscédastiques, on préfèrera l'estimateur *within*. S'il est plus naturel de supposer qu'ils sont fortement autocorrélés, on préfèrera l'estimateur en différence première.
- Dans la pratique, les cas sont rarement aussi tranchés *ex ante* ! La modélisation dépendra alors plutôt de la **question** posée. Est-il plus naturel de formuler le problème en terme de niveau ou d'évolution ?

Comparaison estimateur *within* ou différence première (2/2)

- On a vu que les estimateurs sont convergent sous l'hypothèse d'absence de corrélation entre erreurs et explicatives. En cas de violation de cette hypothèse, ils convergent vers des limites différentes. Comparer les deux estimations des β est une manière de tester l'hypothèse d'endogénéité des explicatives (c'est le point de départ d'un test d'Haussman).
- Une manière de tester l'hypothèse d'exogénéité forte est d'inclure dans la spécification, en plus des explicatives, certaines de leurs valeurs avancées. Si leur coefficients sont différents de 0, l'exogénéité forte est peu crédible.

Test d'existence des effets individuels (1/3)

La prise en compte d'effets individuels, en compléments des variables explicatives contenues dans X, est-elle utile ?
Revient à tester :

- H0 : $\alpha_i = 0, \forall i$.
- H1 : $\alpha_i \neq 0$, pour au moins un i

Dans un modèle linéaire, le test d'égalité de plusieurs coefficients est un test de Fisher.

Test d'existence des effets individuels (2/3)

Démarche :

1. Estimer le modèle sous H0.
2. Récupérer les résidus et calculer $SCR_0 = \sum_i \sum_t \widehat{\epsilon}^2_{0,it}$.
3. Estimer le modèle sous H1
4. Récupérer les résidus et calculer $SCR_1 = \sum_i \sum_t \widehat{\epsilon}^2_{1,it}$.
5. Calculer la statistique de test :

$$S = \frac{SCR_0 - SCR_1}{SCR_1} \frac{ddl_1}{ddl_0 - ddl_1},$$

où ddl_0 est le nombre de degés de libertés sous H0, et ddl_1 le nombre de degrés de libertés sous H1. On a $ddl_0 = N - K$ et $ddl_1 = NT - N - (K - 1)$, d'où :

$$S = \frac{SCR_0 - SCR_1}{SCR_1} \frac{N(T-1) - K}{N - 1}.$$

Test d'existence des effets individuels (3/3)

6. Comparer S avec le quantile d'ordre $1-\alpha$ d'une loi de Fisher à $(N-1, N(T-1)-K)$ degrés de liberté.
7. Si $S > f$, on rejette H0 au seuil de α.

Annexes

Exemple de non-respect de l'exogénéité stricte

```
R> rm(list = ls())
R> id <- rep(1:3, each = 100)
R> date <- rep(1:100, 3)
R> x1 <- id * 10
R> x2 <- id + rnorm(300)
R> y <- x1 - 2 * x2 + rnorm(300)

R> Lid <- c(NA, id[-300])
R> Lx2 <- c(NA, x2[-300])
R> Lx2 <- ifelse(id == Lid, Lx2, NA)
R> summary(lm(y ~ x2 + Lx2))
```

Exemple de non-respect de l'exogénéité stricte

```
Call:
lm(formula = y ~ x2 + Lx2)
<SNIP>
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    8.0551     0.6773   11.89  < 2e-16 ***
x2             0.8660     0.2673    3.24  0.00133 **
Lx2            2.8830     0.2672   10.79  < 2e-16 ***
<SNIP>
   (3 observations deleted due to missingness)
```

- Ici, on a omit du modèle réduit l'hétérogénéité inobservée. L'influence de ces variables omises s'est retrouvée dans le résidu, d'où un problème d'endogénéité et donc une violation de l'exogénéité stricte
- **Implication pratique** : Des coefficients significatifs pour des variables retardées traduisent parfois l'omission d'effets individuels

Cas 1 : Impossible calculer OLS

```
R> rm(list = ls())
R> N <- 10^4
R> T <- 10
R> id <- rep(1:N, each = T)
R> date <- rep(1:T, N)
R> x1 <- id * 10
R> x2 <- id + rnorm(N * T)
R> y <- x1 - 2 * x2 + rnorm(N * T)
R> z <- lm(y ~ 0 + x2 + as.factor(id))
R> Error: cannot allocate vector of size 762.9 Mb
```

Cas 2 : Exemple avec T petit (1/2)

```
R> N <- 100
R> T <- 10
R> id <- rep(1:N, each = T)
R> date <- rep(1:T, N)
R> x1 <- id
R> x2 <- rnorm(N * T)
R> eps <- rnorm(N * T)
R> y <- x1 + 1 * x2 + eps
R> data <- data.frame(id, date, y, x2)
R> head(data)
R> z1 <- lm(y ~ 0 + x2 + as.factor(id), data = data)
R> summary(z1)
```

Cas 2 : Exemple avec T petit (2/2)

```
R> N <- 500
R> T <- 2
R> id <- rep(1:N, each = T)
R> date <- rep(1:T, N)
R> x1 <- id
R> y <- x1 + 1 * x2 + eps
R> data <- data.frame(id, date, y, x2)
R> head(data)
R> z2 <- lm(y ~ 0 + x2 + as.factor(id), data = data)
R> summary(z2)
```

Cas 2 : Explication (1/2)

La variance de l'estimateur OLS est $(X'X)^{-1}\sigma_\epsilon^2$. Pour une variance des erreurs donnée, elle va dépendre des éléments de X. Or, plus chaque indicatrice contiendra de 1, plus les éléments de $(X'X)^{-1}$ seront petits.

Illustration : Indicatrices "équilibrées"

```
R> #N = 2, T = 2
R> x1 <- matrix(c(1, 1, 0, 0, 0,
R>                0, 0, 1, 1, 0), nrow = 5, ncol = 2)
R> solve(t(x1) %*% x1)
     [,1] [,2]
[1,]  0.5  0.0
[2,]  0.0  0.5
```

Cas 2 : Explication (2/2)

Illustration : Indicatrices "déséquilibrées"

```
R> #N = 4, T = 1
R> x2 <- diag(4)
R> solve(t(x2) %*% x2)
     [,1] [,2] [,3] [,4]
[1,]    1    0    0    0
[2,]    0    1    0    0
[3,]    0    0    1    0
[4,]    0    0    0    1
```

Illustration (1/5)

```
rm(list = ls())

N <- 50
T <- 5
id <- rep(1:N, each = T)
date <- rep(1:T, length = N * T)
set.seed(1234)
x <- rnorm(N * T)
a <- rep(rnorm(N), each = T)
eps <- rnorm(N * T)
y <- a + x * 1 + eps
donnees <- data.frame(id, date, y, a, x)
```

Illustration (2/5)

```
# dummy variable estimator
md <- lm(y ~ 0 + x + as.factor(id), data = donnees)
summary(md)
...
<SNIP>
...
Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
x                   1.015373   0.074336  13.659  < 2e-16
as.factor(id)1      0.306501   0.469740   0.652 0.514838
...
<SNIP>
...
```

Illustration (3/5)

```
# FD manuel
shift <- function(x, id, data = donnees) {
            attach(data)
            Lx <- c(NA, x[-length(x)])
            Lid <- c(NA, id[-length(id)])
            Lx <- ifelse(id == Lid, Lx, NA)
            detach(data)
            return(Lx)
          }

Ly <- shift(y, id)
Lx <- shift(x, id)
Dy <- y - Ly
Dx <- x - Lx
```

Illustration (4/5)

```
mfd_man <- lm(Dy ~ 0 + Dx)
summary(mfd_man)
...
<SNIP>
...
Coefficients:
    Estimate Std. Error t value Pr(>|t|)
Dx   1.08668    0.07584   14.33   <2e-16
...
<SNIP>
...
```

Illustration (5/5)

```
# FD estimator
library(plm)
mfd <- plm(y ~ 0 + x, data = donnees,
                effect = "individual", model = "fd",
                index = c("id", "date"))
summary(mfd)
Oneway (individual) effect First-Difference Model
<SNIP>
Balanced Panel: n=50, T=5, N=250
<SNIP>
Coefficients :
  Estimate Std. Error t-value  Pr(>|t|)
x 1.086681   0.075836  14.329 < 2.2e-16
---
<SNIP>
```

Exemple de test d'effets individuels (1/2)

```
R> library("plm")
R> data("Grunfeld", package = "plm")
R> head(Grunfeld, 5)
  firm year   inv  value capital
1    1 1935 317.6 3078.5     2.8
2    1 1936 391.8 4661.7    52.6
3    1 1937 410.6 5387.1   156.9
4    1 1938 257.7 2792.2   209.2
5    1 1939 330.8 4313.2   203.4
R> range(Grunfeld$firm)
[1]  1 10
R> range(Grunfeld$year)
[1] 1935 1954
R> nrow(Grunfeld)
[1] 200
```

Exemple de test d'effets individuels (2/2)

```
R> g0 <- plm(inv ~ value+capital , data=Grunfeld ,\\
model = "pooling")
R> g1 <- plm(inv ~ value+capital , data=Grunfeld ,\\
model = "within")
R> df.residual(g0)
[1] 197
R> df.residual(g1)
[1] 188
R> pFtest(g0, g1)
F test for individual effects\\

data:  inv ~ value + capital\\
F = 49.1766, df1 = 9, df2 = 188, p-value < 2.2e-16\\
alternative hypothesis: significant effects
```