

Econométrie des données de panel: Modèle à effet fixe

Guillaume Horny*

*Banque de France

M2 SE

Chapitre 2 : Modèle à effet fixe

Plan

- 1 Introduction
- 2 Modèle linéaire à effets inobservés
- 3 Modèle linéaire à effets fixes : modèle et hypothèses
- 4 Estimation par OLS du modèle à indicatrices
- 5 Estimateur intra-individuel
- 6 Modèle en différence
- 7 Les différences de différences
- 8 Conclusion

Plan

- 1 Introduction
- 2 Modèle linéaire à effets inobservés
- 3 Modèle linéaire à effets fixes : modèle et hypothèses
- 4 Estimation par OLS du modèle à indicatrices
- 5 Estimateur intra-individuel
- 6 Modèle en différence
- 7 Les différences de différences
- 8 Conclusion

Modèle le plus général possible

$$y_{it} = \alpha_{it} + x'_{it}\beta_{it} + \epsilon_{it},$$

où y_{it} est scalaire, x_{it} un vecteur de dimension $(K \times 1)$ 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.

⇒ 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

- 1 Introduction
- 2 Modèle linéaire à effets inobservés**
- 3 Modèle linéaire à effets fixes : modèle et hypothèses
- 4 Estimation par OLS du modèle à indicatrices
- 5 Estimateur intra-individuel
- 6 Modèle en différence
- 7 Les différences de différences
- 8 Conclusion

Modèle linéaire à effets inobservés

Un modèle linéaire pour données de panel à effets individuels inobservés s'écrit :

$$y_{it} = \alpha_i + x'_{it}\beta + \epsilon_{it},$$

où α_i est un scalaire, $x_{it} = (x_{1it}, \dots, x_{Kit})$ est un vecteur $K \times 1$ de variables exogènes, $\beta = (\beta_1, \dots, \beta_K)$ est un vecteur $K \times 1$ de paramètres et ϵ_{it} est un terme d'erreur i.i.d.

Les α_i

Les scalaires $\alpha_i, i = 1, \dots, N$ sont **constant** dans le temps et propre à chaque unité statistique. Leur influence sur la variable dépendante est la même à chaque date.

Les α_i peuvent être appelés de différentes manières : effets inobservés, effets individuels, composantes inobservées ou encore variables latentes.

Effets fixes ou aléatoires ?

- Une discussion lancinante, autour de laquelle s'est longtemps structurée l'économétrie des données de panel, est de savoir si les α_i doivent être traités comme des **effets fixes** ou des **effets aléatoires**
- La compréhension "historique" est :
 - ▶ lorsque les α_i sont des effets fixes, ils constituent N paramètres qui peuvent être estimés au même titre que les β
 - ▶ lorsque les α_i sont des effets aléatoires, ils sont tirés dans une distribution, avec une moyenne, une variance, etc.
- cette problématique est aujourd'hui dépassée, et nous verrons pourquoi
- toutefois, elle a si longtemps organisée l'économétrie des panels qu'elle est présente dans tous les manuels et bien enracinée dans les esprits. La suite du cours va donc reprendre très classiquement cette dichotomie.

Plan

- 1 Introduction
- 2 Modèle linéaire à effets inobservés
- 3 Modèle linéaire à effets fixes : modèle et hypothèses**
- 4 Estimation par OLS du modèle à indicatrices
- 5 Estimateur intra-individuel
- 6 Modèle en différence
- 7 Les différences de différences
- 8 Conclusion

Pourquoi s'intéresser à ce modèle ?

- relativement simple à estimer
- suffit à exploiter les différents avantages des modèles pour données de panel présentés dans l'introduction
- très répandu en pratique
- 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 $N \rightarrow \infty$ 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 formulée de plusieurs manières équivalentes.

Sachez qu'il existe trois hypothèses d'exogénéité : l'**exogénéité stricte**, l'**exogénéité stricte conditionnelle** aux inobservables, et l'**absence de corrélation**

On va travailler avec l'hypothèse d'**exogénéité stricte conditionnelle** aux inobservables

Exogénéité conditionnelle aux inobservables

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

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

Si : $y_{it} = \alpha_j + x'_{it}\beta + \epsilon_{it}$.

Sous l'exogénéité conditionnelle :

$$\begin{aligned} E(y_{it}|x_{i1}, \dots, x_{iT}, \alpha_j) &= \alpha_j + x'_{it}\beta \\ \iff E(\epsilon_{it}|x_{i1}, \dots, x_{iT}, \alpha_j) &= 0. \end{aligned}$$

Reformulation : Elle est donc parfois présentée de manière équivalente :

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

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

Implications communes des 3 hypothèses sur les explicatives

L'exogénéité stricte, l'exogénéité stricte conditionnelle aux inobservables ainsi que l'absence de corrélation excluent toutes 3 que la variable dépendante soit fonction de :

- variables dépendantes retardées
- comportement affectés par des anticipations sur ϵ_{it} , donc x_{it-1} est corrélée avec ϵ_{it} , qui ne soient pas mesurées par x_{it}
- variables mesurant des feedback (car alors x_{it+1} est corrélée avec ϵ_{it})
- plus généralement des variables endogènes

Si vous suspectez une de ces possibilités, il faut soit avoir recours à des procédures dédiées (ex : modèle dynamiques), soit trouver des proxies (probable en cas d'anticipation ou de feedback), ou encore réfléchir à la fréquence du modèle.

Plan

- 1 Introduction
- 2 Modèle linéaire à effets inobservés
- 3 Modèle linéaire à effets fixes : modèle et hypothèses
- 4 Estimation par OLS du modèle à indicatrices**
- 5 Estimateur intra-individuel
- 6 Modèle en différence
- 7 Les différences de différences
- 8 Conclusion

Les effets fixes individuels : une autre écriture

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 :

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

Exemple : une fonction de production

Considérons une fonction de production Cobb-Douglas, en log, pour $T = 3$ et $K = 2$. On a :

$$y_{it} = \alpha_i + \beta_k k_{it} + \beta_l l_{it} + \epsilon_{it}, \forall i, \forall t \in \{1, 2, 3\}.$$

L'écriture vectorielle pour l'entreprise i est :

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \alpha_i + \begin{pmatrix} k_{i1} & l_{i1} \\ k_{i2} & l_{i2} \\ k_{i3} & l_{i3} \end{pmatrix} \begin{pmatrix} \beta_k \\ \beta_l \end{pmatrix} + \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \epsilon_{i3} \end{pmatrix}.$$

Les effets fixes individuels : écriture matricielle

Une écriture matricielle de notre exemple serait :

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} k_{11} & l_{11} \\ k_{12} & l_{12} \\ k_{13} & l_{13} \\ k_{21} & l_{21} \\ k_{22} & l_{22} \\ k_{23} & l_{23} \end{pmatrix} \begin{pmatrix} \beta_k \\ \beta_l \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \end{pmatrix} .$$

Les effets fixes individuels : écriture matricielle

L'écriture matricielle serait :

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

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 : $\sum_{i=1}^N d_{ilt} = \iota$, où ι est un vecteur de dimension NT ne comprenant que des 1.

⇒ la matrice comprenant à la fois les indicatrices d'individus et la constante n'est plus de plein rang, donc $X'X$ n'est pas inversible

⇒ 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 :

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

Pour les autres :

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

Les α_i s'interprètent alors comme la déviation de $E(y_{it}|x_{it})$ 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

- 1 Introduction
- 2 Modèle linéaire à effets inobservés
- 3 Modèle linéaire à effets fixes : modèle et hypothèses
- 4 Estimation par OLS du modèle à indicatrices
- 5 Estimateur intra-individuel**
- 6 Modèle en différence
- 7 Les différences de différences
- 8 Conclusion

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

On peut estimer β après avoir transformé le modèle de manière à faire disparaître les α_j . 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_{it} d'une période à l'autre pour un individu i donné s'expliquent pas 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 $(N + K) \times (N + K)$ avec des indicatrices individuelles et $(K \times K)$ après transformation *within*.
- La transformation élimine les variables constantes dans le temps. Les coefficients ne sont estimés 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.

- **Efficacité**

Il est de variance minimale sous l'hypothèse d'exogénéité forte conditionnelle aux inobservables et d'homoscédasticité

$$E(\epsilon_i \epsilon_i' | x_i, \alpha_i) = \sigma_\epsilon^2 I_T.$$

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 :

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

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

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

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

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

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

D'où :

$$\begin{aligned} E(\hat{\beta}_W|X) &= \beta_2 + (X_2'MX_2)^{-1} X_2'ME(\epsilon|X) \\ &= \beta_2 \quad (\text{par l'exogénéité conditionnelle}) \end{aligned}$$

$$\Rightarrow E_X[E(\hat{\beta}_W|X)] = E(\hat{\beta}_W) = \beta_2.$$

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

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

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

$$\text{et } A A' = \left(X_2' M X_2 \right)^{-1} X_2' M M' X_2 \left(X_2' M X_2 \right)^{-1} = \left(X_2' M X_2 \right)^{-1}.$$

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 :

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

- Attention : le dénominateur n'est pas le même qu'avec l'estimateur OLS standard, où il est de $(NT - K)$. Le dénominateur *within* est de $NT - N - K$ car il fait intervenir le nombre d'effets fixes (N). Le ratio des deux dénominateurs est :

$$\frac{NT - K}{N(T-1) - K} \xrightarrow{N \rightarrow \infty} \frac{T}{T-1} > 1.$$

Si on estime le modèle après transformation *within* manuelle par une routine OLS, les $\hat{\sigma}_\epsilon^2$ seront sous-estimés et les t de Student surestimés. Le problème d'autant plus marqué que le panel est court.

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*

- l'estimation des coefficients des effets fixes est convergentes sous $T \rightarrow \infty$ (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 $\hat{\alpha}_i$, ou leur groupement, que les valeurs elles-mêmes.

Récapitulatif des propriétés

- l'estimateur est sans biais sous l'exogénéité conditionnelle aux inobservable \Rightarrow la présence d'endogénéité est potentiellement dévastatrice
- 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.

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.

Or, il peut y avoir une autocorrélation dans les erreurs, qui introduit une autocorrélation entre les y_{it} . En présence d'autocorrélation ou d'hétéroscédasticité, cet estimateur ne converge plus vers la vraie matrice de variance.

La bonne nouvelle est qu'il suffit d'adapter l'estimateur de la variance robuste à l'hétéroscédasticité et à l'autocorrélation du chapitre 1

Autocorrélation et hétéroscédasticité (2/3)

L'estimateur *within* a pour variance :

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

où $A = \left(X_2' M X_2 \right)^{-1} X_2' M$.

Arellano (1987) montre qu'on peut estimer $X_2' M E \left[\epsilon\epsilon' | X \right] M X_2$ par :

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

où \underline{X}_{2i} est un vecteur ($T \times K$) des explicatives de l'individu i et \hat{u}_i est le vecteur $T \times 1$ de ses résidus obtenus après estimation *within*.

Autocorrélation et hétéroscédasticité (3/3)

Cet estimateur de la variance est convergent sous :

- $N \rightarrow \infty$
- $E(\epsilon_{it}\epsilon_{js}) = 0$, pour $i \neq j$ (pas de corrélation des chocs d'un individu à l'autre).
- $E(\epsilon_{it}\epsilon_{is})$ 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 **unités panel** suivies dans le temps
- En pratique, vérifier dans les documentations que les commandes de type “robust” corrigent bien de l'autocorrélation est au niveau de l'**unité panel** (pas de problème sous Stata après `xtset`). Sinon, les logiciels ont la facheuse tendance à considérer par défaut que “robust” signifie seulement correction de l'hétéroscédasticité.
- Avec hétéroscédasticité, la matrice de variance reste diagonale, tandis qu'elle comprend des éléments non-nuls en dehors de la diagonale avec l'autocorrélation des erreurs
- Or, avec des données de panel, le problème est généralement plus l'autocorrélation des erreurs que leur hétéroscédasticité !

Plan

- 1 Introduction
- 2 Modèle linéaire à effets inobservés
- 3 Modèle linéaire à effets fixes : modèle et hypothèses
- 4 Estimation par OLS du modèle à indicatrices
- 5 Estimateur intra-individuel
- 6 Modèle en différence**
- 7 Les différences de différences
- 8 Conclusion

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'interprètent 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 $E(\Delta\epsilon_{it}|\Delta x_{i1}, \dots, \Delta x_{iT}) = 0$
- il n'est toutefois pas **efficace**.

En effet, si les erreurs du modèle sont homoscedastiques, celles du modèle en différence 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 :

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

⇒ les erreurs du modèle **en différence** sont négativement autocorrélées.

⇒ 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.

⇒ 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 : :

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

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

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

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

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 et d'absence d'autocorrélation des erreurs dans le modèle en différence, un estimateur convergent de la matrice de variance est fourni par :

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

Plan

- 1 Introduction
- 2 Modèle linéaire à effets inobservés
- 3 Modèle linéaire à effets fixes : modèle et hypothèses
- 4 Estimation par OLS du modèle à indicatrices
- 5 Estimateur intra-individuel
- 6 Modèle en différence
- 7 Les différences de différences**
- 8 Conclusion

Les différences de différences

- la majorité des travaux empiriques actuels cherchent à évaluer l'effet d'un choc (effet d'une taxe, d'un choc migratoire, d'une formation pour les personnes sans emploi...)
- les prix Nobel 2019 et 2021 ont d'ailleurs distingué les approches expérimentales et quasi-expérimentales, où on veut évaluer l'effet d'un traitement
- le formalisme le plus simple, le modèle "Diff-in-Diff" suppose qu'on observe un ensemble d'individus avant un traitement, qui est ensuite administré à une partie des individus
- c'est en fait un cas particulier du modèle à effet fixe pour données de panel

Le modèle à effet fixe avec traitement binaire

Le traitement est représenté par une indicatrice :

$$D_{it} = \begin{cases} 1 & \text{si l'individu } i \text{ reçoit le traitement en } t, \\ 0 & \text{sinon.} \end{cases}$$

Le modèle à effet fixe devient :

$$y_{it} = \alpha_j + \delta_t + \phi D_{it} + \mathbf{x}'_{it} \beta + \epsilon_{it}.$$

Le modèle en différence avec traitement binaire, a.k.a Diff-in-Diff

Par différence entre t et $t - 1$:

$$\Delta y_{it} = (\delta_t - \delta_{t-1}) + \phi \Delta D_{it} + \Delta \mathbf{x}'_{it} \beta + v_{it}.$$

Un estimateur convergent est l'estimateur OLS.

Le modèle en différence avec traitement binaire, a.k.a Diff-in-Diff

Ce modèle peut-être estimé dès que $T = 2$, avec l'ensemble de la population observée aux deux périodes et la traitement alloué uniquement en seconde période. On peut éliminer l'indice t :

$$\Delta y_i = \gamma + \phi \Delta D_i + \Delta x_i' \beta + v_{it}.$$

En l'absence de variables de contrôle, l'estimateur OLS se simplifie :

$$\hat{\phi} = \Delta \bar{y}^1 - \Delta \bar{y}^0.$$

On dit Diff-in-diff car on estime l'évolution d'un groupe en écart à l'évolution d'un autre groupe.

Plan

- 1 Introduction
- 2 Modèle linéaire à effets inobservés
- 3 Modèle linéaire à effets fixes : modèle et hypothèses
- 4 Estimation par OLS du modèle à indicatrices
- 5 Estimateur intra-individuel
- 6 Modèle en différence
- 7 Les différences de différences
- 8 Conclusion**

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 premièrespuis 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.
- L'estimateur du modèle à indicatrice et l'estimateur *within* efficaces sous l'hypothèse d'homoscédasticité, celui du modèle en différence 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 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 !
- la présence d'effets fixes fait qu'on ne peut pas identifier directement les coefficients des variables constantes dans le temps ...
- ... d'où l'intérêt de tester l'existence d'effets fixes ! (à suivre)
- on peut aussi récupérer l'influence de ces variables via des estimations des effets fixes $\hat{\alpha}_i$ par la régression :

$$\hat{\alpha}_i = z_i' \gamma + \phi_i,$$

où les z_i' sont observables (diplôme, genre, lieu de naissance...). On a ici N observations, et non pas NT .

Prévision et calcul des résidus

- Comment estimer les α_j ?
 - ▶ par les coefficients des indicatrices d'individus $\hat{\alpha}_{OLS}$, dont on peut montrer que ce sont les y_i .
 - ▶ ou par $\hat{\alpha}_i = y_i - x_i' \hat{\beta}_W$. La précision de leur estimation dépend de T
- on peut alors construire des prévisions des y_{it} :

$$\hat{y}_{it} = \hat{\alpha}_i + x_{it}' \hat{\beta}.$$

- le résidu est donné par $y_{it} - \hat{y}_{it}$. On peut étudier sa distribution pour analyser le modèle (recherche d'outliers, autocorrélation...)

Comparaison estimateur *within* ou différence première

- 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 préférera l'estimateur *within*. S'il est plus naturel de supposer qu'ils sont fortement autocorrélés, on préférera 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 ?

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 :

- $H_0 : \alpha_i = 0, \forall i.$
- $H_1 : \alpha_i \neq 0, \text{ pour au moins un } i$

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 H_0 .
- 2 Récupérer les résidus et calculer $SCR_0 = \sum_i \sum_t \hat{\epsilon}_{0,it}^2$.
- 3 Estimer le modèle sous H_1
- 4 Récupérer les résidus et calculer $SCR_1 = \sum_i \sum_t \hat{\epsilon}_{1,it}^2$.
- 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 degrés de libertés sous H_0 , et ddl_1 le nombre de degrés de libertés sous H_1 . 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 H_0 au seuil de α .