ÉCONOMÉTRIE
(UGA, S2)
CHAPITRE 8 :
MODÈLES LINÉAIRES POUR DONNÉES DE PANEL

Michal W. Urdanivia*

*UGA, Faculté d'Économie, GAEL,
e-mail : michal.wong-urdanivia@univ-grenoble-alpes.fr

6 avril 2022

1. Introduction

Données de panel

- Les données de panel sont issues de l'observation d'unités d'observation (e.g., ménages, entreprises, personnes, pays, ...) au cours de plusieurs périodes.
- Traditionnellement les unités sont appelées *individus* et nous adopterons cette pratique.
- Ainsi si **w** est un vecteur de variables dont les observation correspondent à un panel, les observations/données seront :

$$\{\mathbf{w}_{it} : i = 1, \dots N; t \in \{1, \dots, T\}\}$$

  où :

  - ⋆ i est l'indice des individus,
  - ⋆ t est l'indice des périodes,
  - ⋆ N est le nombre d'individus,
  - ⋆ T est le plus grand nombre de périodes d'observation dans le panel.

Remarques

Remarque 1

- ⋆ Quand tous les individus sont observés un même nombre de périodes, leur nombre de périodes d'observation T_i est donc le même et égal à T. Plus précisément on distingue :
  - les panels cylindrés qui correspondent $T_i = T$, pour tout $i = 1, \dots, N$,
  - les panels non-cylindrés qui correspondent $T_i \neq T$ pour au moins un individu.
- ⋆ Dans un panel non-cylindré les individus peuvent commencer à être observés à des périodes différentes :

	T			
$i=1$	$\cdot$	$\mathbf{w}_{11}$	$\mathbf{w}_{12}$	$\cdot$
$i=2$	$\mathbf{w}_{21}$	$\mathbf{w}_{22}$	$\mathbf{w}_{23}$	$\mathbf{w}_{24}$
$i=3$	$\cdot$	$\cdot$	$\mathbf{w}_{32}$	$\mathbf{w}_{34}$
$i=4$	$\cdot$	$\mathbf{w}_{42}$	$\mathbf{w}_{43}$	$\mathbf{w}_{44}$

- ⋆ Des extension à plus de 2 dimensions sont possibles, e.g., données employeur-employé où les individus sont observés au cours du temps et dans les différentes emplois occupés.
- ⋆ Il est possible d'ignorer la distinction entre panel cylindrés et non-cylindrés tant que les observations manquantes dans ce derniers le sont de façon aléatoire.

Remarques

Remarque 2

- ⋆ Il est commun dans un cours d'économétrie sur les données de panel de (commencer par) considérer les cas où N est grand relativement à T. Les raisonnements et résultats asymptotiques sont alors obtenus pour $N \to +\infty$. Cela concerne souvent des panel microéconomiques(e.g., sur des ménages, entreprises, personnes).
- ⋆ Pour ce qui concerne les modèles statiques que nous étudierons pour commencer, tous les résultats qui seront présentés s'appliquent au cas inverse où N est relativement petit par rapport T et où on suppose que $T \to +\infty$(il faut néanmoins interchanger N et T dans les formules).

Remarques

Remarque 3

- ⋆ Les données de panel ne sont pas les fait d'avoir des observations sur plusieurs périodes, c.à.d., des *données en coupe répétées*. Ce sont des données où *les mêmes individus sont suivis au cours de différentes périodes*.
- ⋆ Nous verrons que cette caractéristique offre la possibilité de contrôler l'hétérogénéité entre les individus et notamment les problèmes d'endogénéité qu'elle peut générer.
- ⋆ Par rapport au données en coupe et au séries temporelles, les données de panel permettent de corriger les corrélation fallacieuses résultant de variables omises/inobservables. Pour cela :
  - On distingue une hétérogénéité inobservée dans la dimension individuelle du panel, et une hétérogénéité inobservée dans la dimension temporelle.
  - Supposer que l'hétérogénéité individuelle inobservée est constante dans la dimension temporelle et que l'hétérogénéité temporelle est constante dans la dimension individuelle permet de contrôler les biais qu'elles peuvent générer.
  - Quand on suppose N grand relativement à T et $N \to +\infty$ pour les raisonnements asymptotiques il est courant de ne se concentrer que sur l'hétérogénéité individuelle.

Exemple : fonction de production agricole

- Soit une fonction de production agricole Cobb-Douglas 'log-linéairisée" :

$$y_i = a_{0,0} + a_{1,0} l_i + a_{2,0} k_i + u_i, \tag{1}$$

  où pour une entreprise i :
  - ⋆ y_i le (log) du niveau de production,
  - ⋆ l_i et k_i sont les logs de deux facteurs de production, travail et capital,
  - ⋆ u_i capte l'ensemble des facteurs de production inobservées(qualité du sol, qualités/capacités inobservées du travail, ...).
- Ces facteurs inobservées peuvent être corrélés avec les facteurs observées :

$$\text{Cov}[u_i; l_i] \neq 0, \ \text{Cov}[u_i; k_i] \neq 0$$

  Par exemple $\text{Cov}(u_i; l_i) > 0$ si u_i représente une qualité du travail associée à une plus grande productivité.
- On sait que dans ce cas(voir cours passés), l'estimateur des MCO des paramètres $a_{0,0}, , a_{1,0}, a_{2,0}$, ne sera pas convergent.

Exemple : fonction de production agricole

- Considérons la version suivante de l'équation précédente sur données de panel :

$$y_{it} = a_{0,0} + a_{1,0} l_{it} + a_{2,0} k_{it} + \underbrace{u_{it}}_{=\alpha_i + \gamma_t + \varepsilon_{it}}, \qquad (2)$$

  où dans la décomposition de u_{it} :
  - α_i les facteurs de production inobservés invariants dans la dimension temporelle,
  - γ_t des chocs temporels inobservés communs à toutes les entreprises,
  - ε_{it} est un choc temporel inobservée spécifique à la firme i.
- Les différences premières entre t et $t-1$ associées à l'équation de y_{it} dans (2) sont :

$$\Delta y_{it} = a_{1,0} \Delta l_{it} + a_{2,0} \Delta k_{it} + \Delta \gamma_t + \Delta \varepsilon_{it}. \qquad (3)$$

- Pour $N \to +\infty$ et T fixe l'estimateur des MCO des paramètres de (3) est convergent dès lors en particulier que les régresseurs en différences premières sont exogènes par rapport à $\Delta \varepsilon_{it}$.
- Notons que même avec des régresseurs en différences premières endogènes on peut utiliser les donner de panel pour obtenir des estimateurs de VIs(exploitant la causalité au sens de Granger) des paramètres.

Exemple : demande de travail dynamique

- Sargent (1978) considère l'équation suivante pour représenter la demande de travail d'une firme représentative(equation d'Euler) :

$$\Delta l_t = a_{0,0} + a_{1,0}\Delta l_{t-1} + a_{2,0}w_t + a_{3,0}\frac{y_t}{l_t} + \underbrace{u_t}_{=\gamma_t+\varepsilon_t},$$

  où :

  - ⋆ Δl_t, w_t, et y_t sont respectivement la variation de l'emploi, le taux de salaire, et le niveau de production,
  - ⋆ ε_t est un terme d'erreur orthogonal à l'information en t ou avant t.
  - ⋆ γ_t est une composante du profit marginal de la firme inobservé du point de vue de l'analyste mais observé par la firme.
  - ⋆ Cette dernière composante dans le terme d'erreur u_t peut être corrélée aux régresseurs de sorte que l'estimateur des MCO des paramètres ne sera pas convergent.
  - ⋆ Supposons que l'équation précédente soit spécifiée pour des données de panel :

$$\Delta l_{it} = a_{0,0} + a_{1,0}\Delta l_{i,t-1} + a_{2,0}w_{it} + a_{3,0}\frac{y_{it}}{l_{it}} + \underbrace{u_{it}}_{=\alpha_i+\gamma_t+\varepsilon_{it}},$$

    où l'on suppose l'absence de corrélation sérielle pour ε_{it}.

Exemple : demande de travail dynamique

⋆ Les différence premières sur cette équation donnent :

$$\Delta^2 l_{it} = a_{1,0}\Delta^2 l_{i,t-1} + a_{2,0}\Delta w_{it} + a_{3,0}\Delta \frac{y_{it}}{l_{it}} + \Delta\gamma_t + \Delta\varepsilon_{it},$$

que l'on peut estimer par VIs et/ou la MMG.

Un modèle général pour données de panel

- Considérons le modèle :

$$y_{it} = f(\mathbf{x}_{it}, u_{it}, \mathbf{a}_0) \tag{4}$$

  où :
  - ⋆ $f(\cdot)$ est une fonction connue,
  - ⋆ $\mathbf{a}_0$ est un vecteur de paramètres inconnus,
  - ⋆ $\mathbf{x}_{it}$ est un vecteur de variables explicatives observées ;
  - ⋆ u_{it} représente les déterminants inobservés.

- Suivant ce qui a été indiqué dans les exemples, une spécification courante pour u_{it} est :

$$u_{it} = \alpha_i + \gamma_t + \varepsilon_{it}. \tag{5}$$

Problème du paramètre incident

- Dans un modèle pour données de panel où les erreurs se décomposent comme dans (5), $\alpha_1, \ldots, \alpha_N$, et $\gamma_1, \ldots, \gamma_T$ peuvent être traités comme des paramètres à estimer.
- En ce sens il peuvent être vus comme mesurant les effets de $N-1$ dummies individuelles et $T-1$ dummies temporelles.
- Dans un cadre où l'on suppose que $N \to +\infty$ et que T est fixe, le nombre de paramètres $\alpha_1, \ldots, \alpha_N$ augmentera à la même vitesse que N ce qui empêche d'en obtenir des estimateurs convergents.
- Cela peut en outre compromettre l'estimation convergente de $\mathbf{a}_0$.

Problème du paramètre incident

- **Plus généralement** :
  - ⋆ Étant donné un modèle caractérisé par des paramètres $\boldsymbol{\theta}_0$ et un échantillon de N observations,
  - ⋆ on est en présence d'un problème de paramètre incident si :

$$\text{quand } N \to +\infty \text{ nous avons } \dim(\boldsymbol{\theta}_0) \to +\infty.$$

  - ⋆ Neyman and Scott (1948) établissent qu'en règle général on ne peut avoir d'estimateur convergent de $\boldsymbol{\theta}$.
  - ⋆ Supposons que $\boldsymbol{\theta}_0 = (\boldsymbol{\alpha}, \mathbf{a}_0)$ où $\boldsymbol{\alpha}$ est le vecteur d'éléments $\alpha_1, \ldots, \alpha_N$, et que $N \to +\infty$ alors bien que $\dim(\mathbf{a}_0)$ demeure fixe $\dim(\boldsymbol{\alpha}) \to +\infty$.
  - ⋆ La question qui se pose concerne la possibilité d'avoir des estimateurs convergents de $\mathbf{a}_0$ bien que ceux de α ne puissent pas l'être.

Types de modèles pour données de panel

- Il est utile de distinguer des types différents de modèles pour données de panel car les méthodes d'estimation présentent souvent des propriétés différentes selon le type de modèle considéré.
- Cette distinction s'appuie sur deux critères :
  1. additivité/non-additivité des erreurs,
  2. modèles statiques/dynamiques.

additivité/non-additivité des erreurs

- Par exemple :

$$y_{it} = h(\mathbf{x}_{it}, \mathbf{a}_0) + u_{it} \text{ (exemple d'erreurs additifs)},$$
$$y_{it} = \max\{\mathbf{x}_{it}\mathbf{a}_0 + u_{it}; 0\} \text{ (exemple d'erreurs non-additifs)}.$$

- Le problème du paramètre incident se présente de manière différentes selon que les erreurs sont additifs ou non.
- Un cas important sont les modèles non-linéaires statiques où les erreurs sont non-additifs et où l'estimateur à effets fixes qui est convergent dans le cas des modèles linéaires statiques ne l'est plus.

Modèles statiques/non-statiques

- Cette distinction dépends de la relation entre les variables $\{\mathbf{x}_{it}\}$ et les composantes $\{\varepsilon_{it}\}$ du modèle.
- Dans un **modèle statique** $\mathbf{x}_{it}$ est supposé **strictement exogène** par quoi on entends qu'en t il n'inclut pas des variables endogènes *retardées*(c.à.d, correspondant aux périodes précédentes :

$$\mathrm{E}[u_{it}\mathbf{x}_{i,t+s}] = 0 \text{ pour tout } s \in \{\ldots, -2, -1, 0, 1, 2, \ldots\}.$$

- Dans un **modèle dynamique** les variables endogènes retardées peuvent faire partie des régresseurs. Par exemple $y_{i,t-1}$ peut faire partie de $\mathbf{x}_{it}$. Comme alors $\mathbf{x}_{i,t+1}$ inclut y_{it},

$$\mathrm{E}[u_{it}\mathbf{x}_{i,t+s}] \neq 0.$$

  Et des estimateurs convergents et efficaces pour des modèles statiques ne sont même plus convergents dans le cas de modèles dynamique

2. Modèles Statiques

Modèle

- Considérons l'équation suivante d'un modèle pour données de panel :

$$y_{it} = \mathbf{x}_{it}'\mathbf{a}_0 + u_{it}, \; u_{it} = \alpha_i + \gamma_t + \varepsilon_{it}. \quad (6)$$

où :

- ⋆ $\mathbf{x}_{it}$ est vecteur $K \times 1$ de régresseurs, auquel on associe un vecteur de paramètres $\mathbf{a}_0$,
- ⋆ α_i représente l'hétérogénéité inobservée dans la dimension individuelle, souvent appelé *effet fixe individuel*,
- ⋆ γ_t représente l'hétérogénéité inobservée dans la dimension temporelle, souvent appelé *effet fixe temporel*,
- ⋆ ε_{it} est l'hétérogénéité inobservée variable dans le temps et entre les individus.

Dummies temporelles

- Nous allons considérer les cas où T et relativement petit par rapport à N avec $N \to +\infty$ pour les analyses asymptotiques.
- Ce faisant pour faciliter l'écriture du modèle, il nous pouvons adopter une notation qui permettent d'inclure les paramètres $\{\gamma_t\}_{t=1}^{T}$ dans $\mathbf{a}_0$.
- Pour cela nous allons considérons un indicatrice temporelle d_{it}^s pour la période s telle que :

$$d_{it}^s = 1 \quad \text{si} \quad t = s; \quad \text{et} \quad D_{it}^s = 0 \quad \text{si} \quad t \neq s$$

alors,

$$\gamma_t = \sum_{s=1}^{T} d_{it}^s \gamma_s = \left[d_{it}^1, d_{it}^2, \ldots, d_{it}^T\right] \boldsymbol{\gamma} = \mathbf{d}_{it}' \boldsymbol{\gamma}$$

où $\boldsymbol{\gamma}$ est le vecteur $T \times 1$ défini par $\boldsymbol{\gamma} \equiv (\gamma_1, \gamma_2, \ldots, \gamma_T)'$.

Dummies temporelles

- On peut alors écrire (6) :

$$y_{it} = \begin{bmatrix} \mathbf{x}_{it} & \mathbf{d}_{it} \end{bmatrix}' \begin{bmatrix} \mathbf{a}_0 \\ \gamma \end{bmatrix} + \alpha_i + \varepsilon_{it}$$

  qui est la version de (6) avec les effets temporels pris en compte par le bias d'un vecteur de régresseurs incluant des dummies temporelles.

- Pour ne pas alourdir les notations nous allons écrire cette équation simplement :

$$y_{it} = \mathbf{x}_{it}'\mathbf{a}_0 + \alpha_i + \varepsilon_{it}, \tag{7}$$

  pour laquelle il sera entendu que $\mathbf{x}_{it}$ inclue des dummies temporelles, et que $\mathbf{a}_0$ inclue γ.

- Comme nous supposons que T et petit relativement à N avec $N \rightarrow +\infty$ le problème du paramètre incident provient seulement de $\alpha_1, \ldots, \alpha_N$.

- Une démarche similaire peut être adopté dans le cas où T est grand relativement à N et $T \rightarrow +\infty$.

Dummies temporelles

- Le modèle peut être présenté comme un système de T equations(une equation pour $t = 1, 2, \ldots, T$) avec n observations pour chacune des équations du système.

$$\begin{aligned} y_{i1} &= \mathbf{x}'_{i1}\beta + \alpha_i + \varepsilon_{it} \\ y_{i2} &= \mathbf{x}'_{i2}\beta + \alpha_i + \varepsilon_{i2} \\ &\vdots \\ y_{iT} &= \mathbf{x}'_{iT}\beta + \alpha_i + \varepsilon_{iT} \end{aligned}$$

- Sous forme vectorielle on peut écrire :

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{1}_T\alpha_i + \boldsymbol{\varepsilon}_i$$

où

- $\mathbf{y}_i$ et $\boldsymbol{\varepsilon}_i$ sont de vecteurs $T \times 1$,
- $\mathbf{1}_T$ est un vecteur $T \times 1$ avec tous ses éléments égaux à 1 ;
- $\mathbf{X}_i$ est une matrice $T \times K$.

Dummies temporelles

- Une notation similaire peut être utilisée pour des panels non-cylindrés.
  - ⋆ Soit $w_{it} \in \{0,1\}$; une indicatrice pour l'événement "l'individu i est observé à la période t".
  - ⋆ Alors :

$$\mathbf{y}_i = \mathbf{X}_i \mathbf{a}_0 + \mathbf{w}_i \alpha_i + \boldsymbol{\varepsilon}_i,$$

où :

$$\underset{T\times 1}{\mathbf{y}_i} = \begin{bmatrix} w_{i1}y_{i1} \\ w_{i2}y_{i2} \\ \vdots \\ w_{iT}y_{iT} \end{bmatrix} \quad , \quad \underset{T\times K}{\mathbf{X}_i} = \begin{bmatrix} w_{i1}\mathbf{x}'_{i1} \\ w_{i2}\mathbf{x}'_{i2} \\ \vdots \\ w_{iT}\mathbf{x}'_{iT} \end{bmatrix} \quad , \quad \underset{T\times 1}{\mathbf{w}_i} = \begin{bmatrix} w_{i1} \\ w_{i2} \\ \vdots \\ w_{iT} \end{bmatrix}$$

  - ⋆ Les expressions des estimators qui seront vus plus loin s'appliquent à cette définition de $\mathbf{y}_i$, $\mathbf{X}_i$, et $\mathbf{w}_i$.

Exogénéité stricte(modèle statique)

Condition C1

(**Exogénéité stricte**)

$$\mathrm{E}[\mathbf{x}_{it}\varepsilon_{is}] = 0, \text{ pour tout } (t,s) \in \{1,2,\ldots,T\}.$$

- Cette condition ne sera pas vraisemblable sur des modèles incluant des régresseurs endogènes retardés : e.g.,$y_{i,t-1}$,$y_{i,t-2}$.
- Par exemple, supposons que $y_{i,t-1} \subset \mathbf{x}_{it}$.
- Le modèle énonce que $y_{i,t-1}$ depends de $\varepsilon_{i,t-1}$.
- Il est alors clair que $\mathrm{E}[\mathbf{x}_{it}\varepsilon_{i,t-1}] \neq 0$.

Effets fixes ou effets aléatoires

- Un des problème les plus importants que les données de panel permettent de traiter est celui de l'endogénéité due à l'hétérogénéité individuelle inobservée.
- Dans le cadre de la décomposition de l'erreur du modèle dans (6) cela concerne le problème posé par :

$$\mathrm{E}[\mathbf{x}_{it}\alpha_i] \neq 0.$$

- Deux approches alternatives sont en général adoptées :
  1. par les **effets fixes**,
  2. par les **effets aléatoires**.

Effets fixes

- Dans cette approche aucune condition contraignant la relation entre α_i et $\{\mathbf{x}_{it}\}_{t=1}^{T}$ n'est posée. Par exemple, on ne supposera pas que $\mathrm{E}[\mathbf{x}_{it}\alpha_i] = 0$.
- Dans la mesure où la loi conditionnelle de α_i sachant $\{\mathbf{x}_{it}\}_{t=1}^{T}$ n'est pas contrainte, un **modèle à effets fixes** peut être vu comme non-paramétrique par rapport à cette loi.
- Typiquement les modèles à effets fixes s'appuient sur une transformation du modèle qui élimine les effets individuels où les rend redondants dans une vraisemblance conditionnelle.

Effets aléatoires

- Cette approche se caractérise par les conditions qui contraignent la loi conditionnelle de α_i sachant $\{\mathbf{x}_{it}\}_{t=1}^{T}$.
- La condition la plus forte est l'indépendance de α_i entre $\{\mathbf{x}_{it}\}_{t=1}^{T}$ avec $\alpha_i \sim i.i.d.(0, \sigma_\alpha^2)$. Il est courant de qualifier cette cas d'approche à effets aléatoires.
- Néanmoins d'autres approches à effets aléatoires existent. Par exemple dans Chamberlain (1984) un modèle à **effets aléatoires corrélés** pose :

$$\alpha_i = \lambda_0 + \mathbf{x}_{i1}'\lambda_1 + \ldots + \mathbf{x}_{i1}'\lambda_T + e_i,$$

  où e_i est indépendant de $\{\mathbf{x}_{it}\}_{t=1}^{T}$, et $\{\lambda_t\}_{t=1}^{T}$ sont des paramètres à estimer avec $\mathbf{a}_0$.
- C'est une **approche paramétrique** car elle dépends sur une condition quant à la loi de $\{\mathbf{x}_{it}\}_{t=1}^{T}$ et α_i.

Avantages relatifs entre les approches à EF et à EA

- L'approche à EF est plus robuste qu'une approche à EA ou EA corrélés dans la mesure où ces dernières dépendent de la validité de la condition imposé sur la relation entre effets individuels et régresseurs. Si cette condition n'est pas satisfaite les estimateurs dans ces approches ne seront probablement pas convergents.
- Les transformations utilisées dans les approches à EF peuvent éliminer la variabilité dans l'échantillon des régresseurs exogènes. Cela peut rendre les résultats moins précis que ceux d'une approche à EA(quand les conditions dans celui-ci sont valides).
- Pour certains modèles, il n'y pas de méthode d'estimation convergent par EF. C'est par exemple le cas des modèles non-linéaires dynamiques. Voir par exemple Chamberlain (2010).

3. Estimation sous exogénéité par rapport aux effets individuels

Estimation par MCO

- On commence avec le cas le plus simple où l'effet individuel et les régresseurs ont une corrélation nulle :

$$\mathrm{E}[\mathbf{x}_{it}\alpha_i] = 0.$$

- Dans ce cas, un estimateur convergent est simplement l'estimateur des MCO :

$$\begin{aligned}\hat{\mathbf{a}}_0^{MCO} &= \left[\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}\mathbf{x}_{it}'\right]^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}y_{it} \\ &= \left[\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i\right]^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{y}_i.\end{aligned}$$

- Cet estimateur est souvent appelé **estimateur des MCO empilé** car nous empilons/traitons les observations comme s'il s'agissait de données en coupe.
- Cet estimateur n'est pas efficace en raison de la corrélation sérielle des erreurs u_{it} :

$$\mathrm{E}[u_{it}; u_{i,t-j}] = \mathrm{E}\left[(\alpha_i + \varepsilon_{it})(\alpha_i + \varepsilon_{i,t-j})\right] = \mathrm{E}[\alpha_i^2] = \sigma_\alpha^2$$

- L'estimateur asymptotiquement efficace est l'estimateur de moindres carrés généralisés.

Estimateur des MCG

- Pour présenter cet estimateur, écrivons d'abord le modèle dans la notation matricielle suivante :

$$\mathbf{y} = \mathbf{X}\mathbf{a}_0 + \mathbf{u}$$

où :

$$\underset{(NT\times 1)}{\mathbf{y}} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix}, \quad \underset{(NT\times K)}{\mathbf{X}} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix}, \quad \underset{(NT\times 1)}{\mathbf{u}} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_n \end{bmatrix}$$

- Soit $\mathbf{V_u}$ la variance de $\mathbf{u}$.
- Il est courant de définir l'estimateur des MCG comme l'estimateur des MCO du modèle transformé/pondéré suivant :

$$\left[\mathbf{V_u}^{-1/2}\mathbf{y}\right] = \left[\mathbf{V_u}^{-1/2}\mathbf{X}\right]\mathbf{a}_0 + \left[\mathbf{V_u}^{-1/2}\mathbf{u}\right]$$

- Nous obtenons des expressions de $\mathbf{V_u}$, de $\mathbf{V_u}^{-1/2}$, et des variables transformées $\left[\mathbf{V_u}^{-1/2}\mathbf{y}\right]$ et $\left[\mathbf{V_u}^{-1/2}\mathbf{X}\right]$.

Estimateur des MCG

- Nous avons :

$$\mathbf{V_u} = \mathrm{E}\left[\mathbf{u}\mathbf{u}'\right] = \mathrm{E}\begin{bmatrix} \mathbf{u}_1\mathbf{u}_1' & \mathbf{u}_1\mathbf{u}_2' & \dots & \mathbf{u}_1\mathbf{u}_N' \\ & \mathbf{u}_2\mathbf{u}_2' & \dots & \mathbf{u}_2\mathbf{u}_N' \\ & & \ddots & \vdots \\ & & & \mathbf{u}_N\mathbf{u}_N' \end{bmatrix}$$

- Quand $\mathbf{u}_i$ est homoscédastique sur la dimension individuelle et $\mathrm{E}\left[\mathbf{u}_i\mathbf{u}_j^prime\right] = 0$ pour $i \neq j$,

$$\mathbf{V_u} = \begin{bmatrix} \boldsymbol{\Omega} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega} & \dots & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \vdots \\ \mathbf{0} & \vdots & \dots & \boldsymbol{\Omega} \end{bmatrix} = \mathbf{I}_N \otimes \boldsymbol{\Omega}$$

où $\boldsymbol{\Omega}$ est la matrice $T \times T$ $\mathrm{E}\left[\mathbf{u}_i\mathbf{u}_i'\right]$.

QMCG quand les erreurs ne sont pas i.i.d.

- On considère que $\mathbf{\Sigma_u} = \mathbf{I}_N \otimes \mathbf{\Omega}$ où pour tout i la matrice des variances-covariances suivante ne fait pas l'objet de contraintes,

$$\mathbf{\Omega} = \mathrm{E}\begin{bmatrix} u_{i1}^2 & u_{i1}u_{i2} & \dots & u_{i1}u_{iT} \\ & u_{i2}^2 & \dots & u_{i2}u_{iT} \\ & & \ddots & \vdots \\ & & & u_{iT}^2 \end{bmatrix}$$

- L'estimateur des MCG est équivalent à l'estimateur des MCO du modèle transformé suivant :

$$\left[\mathbf{I}_n \otimes \Omega^{-1/2}\right]\mathbf{y} = \left[\mathbf{I}_N \otimes \Omega^{-1/2}\right]\mathbf{X}\mathbf{a}_0 + \mathbf{u}^\star$$

- L'estimateur des moindres carrés quasi généralisés(MCQG) est alors obtenu comme estimateur des MCO dans ce modèle transformé où Ω a été préalablement estimé par un estimateur convergent :
  - ⋆ Soit $\hat{\mathbf{u}}_i = [\hat{u}_{i1}, \hat{u}_{i2}, \dots, \hat{u}_{iT}]'$ le vecteur des résidus dans l'estimation par MCO du modèle non-transformé(MCO empilés) pour l"individu i.
  - ⋆ Un estimateur de $\mathbf{\Omega}$ est :

$$\hat{\mathbf{\Omega}} = N^{-1}\sum_{i=1}^{N}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'.$$

  - ⋆ L'estimateur des MCQG utilise cet estimateur de $\mathbf{\Omega}$.

4. Estimation sans contrainte sur les effets individuels : "différences premières", "within", etc

Non-convergence des MCO

- L'estimateur des MCO(ou sa version efficace avec les MCG) précédent n'est plus convergent car :

$$E\left[\mathbf{x}_{it}(\alpha_i + \epsilon_{it})\right] \neq 0.$$

- Nous allons étudier les estimateurs suivants :
  - ⋆ MCO du modèle en différences premières.
  - ⋆ Estimateur à effets fixes "within" : variables en différence par rapport aux moyennes dans la dimension individuelle.
  - ⋆ MCO avec dummies individuelles(LSDV).
  - ⋆ Estimateur à effets aléatoires corrélés de Chamberlain.

Modèle en différences premières

- Comme cela a été indiqué plus haut les méthodes dites à effets fixes(c.à.d., n'imposant pas de contrainte sur la relation entre α_i et $\{\mathbf{x}_{it}\}_{i=1}^{N}$) s'appuient sur une transformation du modèle qui supprime α_i.
- Un premier estimateur fondé sur cette approche est l'estimateur du modèle en différences premières contenu dans les exemples introductifs.
- Le modèle transformé considéré est :

$$\Delta y_{it} = \Delta \mathbf{x}_{it}' \mathbf{a}_0 + \Delta \varepsilon_{it}$$

où par Δw_{it} on représente ici la différence première dans la dimension temporelle d'une variable w_{it}, c.à.d., $\Delta w_{it} = w_{it} - w_{i,t-1}$. Pour des vecteurs $\mathbf{w}_{it}$, $\Delta \mathbf{w}_{it}$ est simplement l'opérateur de Δ appliqué à chaque élément du vecteur.

- L'exogénéité stricte de $\mathbf{x}_{it}$ implique que $\mathrm{E}\left[\Delta \mathbf{x}_{it} \Delta \varepsilon_{it}\right] = 0$:

$$\mathrm{E}\left[\Delta \mathbf{x}_{it} \Delta \varepsilon_{it}\right] = = \underbrace{\mathrm{E}[\mathbf{x}_{it}\varepsilon_{it}]}_{=0} + \underbrace{\mathrm{E}[\mathbf{x}_{it}\varepsilon_{i,t-1}]}_{=0} + \underbrace{\mathrm{E}[\mathbf{x}_{i,t-1} V_{it}]}_{=0} + \underbrace{\mathrm{E}[\mathbf{x}_{i,t-1}\varepsilon_{i,t-1}]}_{=0} = 0,$$

et par conséquent l'estimateur des MCO du modèle en differences premières est convergent.

$$\hat{\mathbf{a}}_N^{FD} = \left[\sum_{i=1}^{N}\sum_{t=2}^{T} \Delta \mathbf{x}_{it} \Delta \mathbf{x}_{it}'\right]^{-1} \left[\sum_{i=1}^{N}\sum_{t=2}^{T} \Delta \mathbf{x}_{it} \Delta y_{it}\right] \underset{N \to +\infty}{\overset{p}{\to}}$$

Modèle en différences premières

- Cet estimateur n'est pas cependant efficace (sauf quand ε_{it} est une marche aléatoire), et on peut définir une estimateur des MCQG du modèle en différences premières.

Estimateur du modèle en "within"

- L'estimateur en "within" du modèle est un estimateur à effets fixes sur la transformation suivante du modèle :

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \, \mathbf{a}_0 + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

où $\bar{y}_i := T^{-1}\sum_{t=1}^{T} y_t$, $\bar{\mathbf{x}}_i := T^{-1}\sum_{t=1}^{T} \mathbf{x}_{it}$, $\bar{\varepsilon}_i := T^{-1}\sum_{t=1}^{T} \varepsilon_{it}$.

- L'estimateur en within est simplement l'estimateur des MCO de ce modèle en différence par rapport aux moyennes dans la dimension temporelle.

$$\hat{\mathbf{a}}_N^{WG} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(y_{it} - \bar{y}_i)$$

Convergence de l'estimateur en "within"

- Cet estimateur est convergent dès lors qu'il n'y a pas de colinéarité de que $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$ et $\mathrm{E}\left((\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\varepsilon_{it} - \bar{\varepsilon}_i)\right) = 0$.
- Notons que :

$$\mathrm{E}\left[(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\varepsilon_{it} - \bar{\varepsilon}_i)\right] = \underbrace{\mathrm{E}[\mathbf{x}_{it}\varepsilon_{it}]}_{=0} + \underbrace{\mathrm{E}[\mathbf{x}_{it}\bar{\varepsilon}_i]}_{=0} + \underbrace{\mathrm{E}[\bar{\mathbf{x}}_i\varepsilon_{it}]}_{=0} + \underbrace{\mathrm{E}[\bar{\mathbf{x}}_i\bar{\varepsilon}_i]}_{=0} = 0$$

- L'exogénéité stricte des régresseurs implique que toutes ces espérances sont zéro et que par conséquent l'estimateur en within est convergent.
- Cependant, l'estimateur de α_i ne l'est pas pour T fixe. En effet :

$$\begin{aligned}\operatorname*{plim}_{N\to+\infty} \hat{\alpha}_i &= T^{-1}\sum_{t=1}^{T}\left(y_{it} - \mathbf{x}'_{it}\left(\operatorname*{plim}_{N\to+\infty} \hat{\mathbf{a}}_N^{WG}\right)\right) = T^{-1}\sum_{t=1}^{T}\left(y_{it} - \mathbf{x}'_{it}\beta\right) \\ &= \alpha_i + \bar{\varepsilon}_i \neq \alpha_i\end{aligned}$$

MCO avec dummies individuelle(LSDV)

- On traite maintenant les $\{\alpha_i\}_{i=1}^N$ comme des paramètres à estimator $\mathbf{a}_0$.
- Pour cela écrivons le modèle comme suit :

$$\mathbf{y} = \mathbf{X}\mathbf{a}_0 + \mathbf{D}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} = \begin{bmatrix}\mathbf{X} & : & \mathbf{Q}\end{bmatrix}\begin{bmatrix}\mathbf{a}_0 \\ \boldsymbol{\alpha}\end{bmatrix} + \boldsymbol{\varepsilon}$$

  - ⋆ $\mathbf{Q}$ est une matrice $NT \times N$ de dummies, une pour chaque individu,
  - ⋆ la ième colonne $\mathbf{d}$ contient les observations de la dummie pour l'individu i, i.e., 1 si l'observation appartient à i et 0 sinon
  - ⋆ $\boldsymbol{\alpha}$ est le vecteur associé aux $\{\alpha_i\}_{i=1}^N$, i.e., $\boldsymbol{\alpha} := [\alpha_1, \ldots, \alpha_N]'$.

- On appelle estimateur LSDV(abbreviation pour "least squares dummie variables") l'estimateur des MCO de $\begin{bmatrix}\mathbf{a}_0 \\ \boldsymbol{\alpha}\end{bmatrix}$, soit :

$$\begin{bmatrix}\hat{\mathbf{a}}_N^{LSDV} \\ \hat{\boldsymbol{\alpha}}_N^{LSDV}\end{bmatrix} = \left[\begin{bmatrix}\mathbf{X} & : & \mathbf{D}\end{bmatrix}'\begin{bmatrix}\mathbf{X} & : & \mathbf{Q}\end{bmatrix}\right]^{-1}\begin{bmatrix}\mathbf{X} & : & \mathbf{Q}\end{bmatrix}'\mathbf{y}$$

$$= \begin{bmatrix}\mathbf{X}^\top\mathbf{X} & \mathbf{X}^\top\mathbf{Q} \\ \mathbf{Q}^\top\mathbf{X} & \mathbf{Q}^\top\mathbf{Q}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{X}^\top\mathbf{y} \\ \mathbf{Q}^\top\mathbf{y}\end{bmatrix}$$

MCO avec dummies individuelle(LSDV)

- Cet estimateur peut sembler exigeant du point de vue calculatoire dans la mesure il repose sur l'inversion d'une matrice $(N+K)\times(N+K)$ avec N pouvant devenir très grand sur des panels typiques.
- Cet estimateur peut être obtenu sans besoin d'inverser "directement" cette matrice.
- On peut pour cela utiliser les propriétés des matrices par blocs/partitionnées $\mathbf{Q}'\mathbf{Q}$ et $\mathbf{Q}'\mathbf{X}$.
- Appuyons nous sur le théorème de Frish-Waugh-Lovell, qui permet d'écrire l'estimateur des MCO comme suit :

$$\hat{\mathbf{a}}_N^{LSDV} = \left[\left(\mathbf{X}'\mathbf{M_Q}\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{M_Q}\mathbf{y}\right)\right]$$

où $\mathbf{M_Q}$ est une matrice idempotente :

$$\begin{aligned}
\mathbf{M_Q} &= \mathbf{I}_{NT} - \mathbf{Q}\left[\mathbf{Q}'\mathbf{Q}\right]^{-1}\mathbf{Q}' \\
&= (\mathbf{I}_n \otimes \mathbf{I}_T) - \frac{1}{T}\left[\mathbf{I}_n \otimes \mathbf{1}_T\mathbf{1}_T'\right] \\
&= \mathbf{I}_n \otimes \left[\mathbf{I}_T - \frac{1}{T}\mathbf{1}_T\mathbf{1}_T'\right]
\end{aligned}$$

- Pré-multiplier $\mathbf{y}$ et $\mathbf{X}$ par $\mathbf{M_Q}$, le transforme en écart par rapport à leurs moyennes individuelles respectives :

MCO avec dummies individuelle(LSDV)

- ⋆ pour $\mathbf{M_Q y} =: \mathbf{y}^\star$, et
- ⋆ $\mathbf{M_Q X} =: \mathbf{X}^\star$,

- L'élément (i, t) étant donné par :

$$y_{it}^\star = y_{it} - \bar{y}_i, \quad \mathbf{x}_{it}^\star = \mathbf{x}_{it} - \bar{\mathbf{x}}_i$$

où $\bar{y}_i := T^{-1}\sum_{t=1}^{T} y_{it}$, $\bar{\mathbf{x}}_i := T^{-1}\sum_{t=1}^{T} \mathbf{x}_{it}$.

- Comme $\mathbf{M_Q}$ est idempotente, on peut écrire :

$$\hat{\mathbf{a}}_N^{LSDV} = \left[\mathbf{X}^{\star\prime}\mathbf{X}^\star\right]^{-1}\left[\mathbf{X}^{\star\prime}\mathbf{y}^\star\right]$$

- Par conséquent l'estimateur LSDV de $\mathbf{a}_0$ est numériquement équivalent à l'estimateur du modèle transformé suivant :

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)'\mathbf{a}_0 + \varepsilon_{it}^\star,$$

c.à.d., à l'estimateur en "within" lequel apparaît comme une procédure plus "économe" de calcul de l'estimateur LSDV.

MCO avec dummies individuelle(LSDV)

- Il est alors clair d'après ce qui a été dit sur l'estimateur en "within" que les les estimateur des effets fixes individuels suivants :

$$\hat{\alpha}_i^{LSDV} = T^{-1}\sum_{t=1}^{T}(y_{it} - x_{it}'\hat{\mathbf{a}}_N^{LSDV}), \; i = 1, \ldots, N,$$

ne seront pas convergents, et ce qui pose alors la question de la convergence de $\hat{\alpha}_{i,N}^{LSDV}$.

Paramètre incident et convergence de l'estimateur LSDV

- En **règle générale**(c.à.d., en dehors du seul cadre d'un modèle linéaire statique), l'estimateur LSDV d'un modèle de panel n'est pas convergent dans un raisonnement asymptotique avec $N \rightarrow +\infty$ et T fixe.
- La raison en est le problème du paramètre incident : le nombre de paramètres dans $\mathbf{a}$ augmente à la même vitesse que N, et nous n'avons que T observations pour chaque α_i, et ainsi $\hat{\alpha}_i^{LSDV}$ n'est pas convergent.
- Est-ce que cela affecte la convergence de $\hat{\mathbf{a}}_N^{LSDV}$?
- Comme indiqué plus haut, oui en **règle générale** mais non le cas(particulier) d'un modèle linéaire statique.
- Autrement dit, la convergence de $\hat{\mathbf{a}}_N^{LSDV}$ malgré la non convergence des estimateurs des effets fixes individuels, n'étend pas à des modèles dynamique et/ou non-linéaires.

Paramètre incident et convergence de l'estimateur LSDV

Remarque 4 (Intuitions sur la convergence de l'estimateur LSDV)

- ⋆ Malgré le fait que l'estimation LSDV des effets fixes individuels ne donne pas des estimateurs convergents pour T fixe, l'erreur d'estimation $\alpha_i - \hat{\alpha}_i^{LSDV}$ n'est pas corrélée avec les régresseurs quand ceux-ci sont supposés strictement exogènes.
- ⋆ Notons qu'asymptotiquement $\hat{\alpha}_i^{LSDV}$ est égal à $\alpha_i + \bar{\varepsilon}_i$ ce qui fait que l'erreur d'estimation $\alpha_i - \hat{\alpha}_i^{LSDV}$ est égal à $\bar{\varepsilon}_i$.
- ⋆ Quand le régresseurs sont strictement exogènes cet erreur d'estimation n'est pas corrélée avec les régresseurs $\mathbf{x}_{it}$.
- ⋆ De plus, quand les chocs transitoires ε_{it} sont i.i.d., l'estimateur "within" est aussi asymptotiquement efficace.

Références

Chamberlain, Gary. 1984. "Chapter 22 Panel data." Elsevier, 1247–1318. URL https://www.sciencedirect.com/science/article/pii/S1573441284020146.

———. 2010. "Binary response models for panel data : Identification and information." *Econometrica* 78 (1) :159–168.

Neyman, J and E. L. Scott. 1948. "Consistent estimates based on partially consistent observations." *Econometrica* 16 :1–32.

Sargent, Thomas. 1978. "Estimation of Dynamic Labor Demand Schedules under Rational Expectations." *Journal of Political Economy* 86 (6) :1009–44. URL https://EconPapers.repec.org/RePEc:ucp:jpolec:v:86:y:1978:i:6:p:1009-44.