UMR 8184
Thema

THÉORIE ÉCONOMIQUE,
MODÉLISATION ET APPLICATIONS

Estimateurs de Modèles Linéaires de Panel

Aristide E. Houndetoungan

20 Octobre 2021

Introduction

- Considérons le modèle linéaire de panel suivant :

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}. \tag{1}$$

- Les données sont observées à chaque période pour chaque individu.
- y_{it} est la variable dépendante, $\mathbf{x}_{it}$ inclut les variables explicatives, ε_{it} est le terme d'erreur. $\mathbf{x}_{it}$ **n'inclut pas de variable constante**. Toutefois, certaines colonnes de $\mathbf{x}_{it}$ peuvent ne pas varier avec i ou avec t.
- $\boldsymbol{\beta}$ est la pente de y_{it} par rapport aux variables explicatives (même interprétation que dans le cas d'un modèle en coupe instantanée).
- L'intercept α_i varie selon l'individu. α_i est appelé **effet individuel**. Il induit une **d'hétérogénéité individuel**.
- Sous l'hypothèse d'exogénéité de $\mathbf{x}_{it}$, $\mathbb{E}(y_{it}|\mathbf{x}_{it}) = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta}$. Même si $\mathbf{x}_{it} = \mathbf{x}_{jt}$ pour deux individus i et j, on a $\mathbb{E}(y_{it}|\mathbf{x}_{it}) \neq \mathbb{E}(y_{jt}|\mathbf{x}_{jt})$.

- Cette formalisation se rapproche plus de la réalité que celle en coupe instantanée.
  - Par exemple, même si les caractéristiques observables de deux individus sont identiques (même éducation, même âge, même domaine de travail), ils n'ont probablement pas les mêmes salaires. Quelqu'un aurait peut-être mieux négocié son salaire au moment de son embauche que l'autre. Ce pouvoir de négociation individuel n'est pas observé par celui qui valorise le modèle. Il est donc exclu de $\mathbf{x}_{it}$ et pris en compte par α_i.
- Le terme α_i permet de capter l'ensemble des facteurs individuels (non pris en compte dans $\mathbf{x}_{it}$) et constants dans le temps qui influence y_{it}.
- Lorsque le modèle comporte des effets individuels, il faut nécessairement observer des données de panel pour les estimer.
- Par exemple, en coupe instantanée, $y_i = \alpha_i + \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$ ne peut pas être estimé sans faire d'hypothèses supplémentaires.

- **Motivation des effets individuels**

- Supposons que le vrai modèle est comme défini en (1) :

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}. \tag{1}$$

- Supposons que l'économètre estime,

$$y_{it} = \alpha + \mathbf{x}'_{it}\boldsymbol{\beta} + \tilde{\varepsilon}_{it}, \tag{2}$$

où $\tilde{\varepsilon}_{it} = \varepsilon_{it} + \alpha_i - \alpha$.

- Sous l'hypothèse d'exogénéité de $\mathbf{x}_{it}$ par rapport à ε_{it},

$$\mathbb{E}(\tilde{\varepsilon}_{it}|\mathbf{x}_{it}) = \mathbb{E}(\varepsilon_{it} + \alpha_i - \alpha|\mathbf{x}_{it}) = \underbrace{\mathbb{E}(\varepsilon_{it}|\mathbf{x}_{it})}_{=0} + \mathbb{E}(\alpha_i|\mathbf{x}_{it}) - \underbrace{\mathbb{E}(\alpha|\mathbf{x}_{it})}_{=\alpha},$$

$$\mathbb{E}(\tilde{\varepsilon}_{it}|\mathbf{x}_{it}) = \mathbb{E}(\alpha_i|\mathbf{x}_{it}) - \alpha.$$

- Si $\mathbb{E}(\alpha_i|\mathbf{x}_{it})$ varie en fonction de i, alors il exite nécessairement au moins un i pour lequel $\mathbb{E}(\alpha_i|\mathbf{x}_{it}) - \alpha \neq 0$. Donc, $\mathbb{E}(\tilde{\varepsilon}_{it}|\mathbf{x}_{it}) \neq 0 \implies \mathbf{x}_{it}$ n'est pas exogène par rapport à $\tilde{\varepsilon}_{it}$ et (2) ne peut pas être estimé par MCO.

- En revanche, si $\mathbb{E}(\alpha_i|\mathbf{x}_{it}) = \bar{\alpha}$, avec $\bar{\alpha}$ indépendant de i, alors $\mathbb{E}(\alpha_i|\mathbf{x}_{it}) - \alpha = 0$ avec $\alpha = \bar{\alpha}$ et (2) être estimé par MCO.

- La condition : $\mathbb{E}(\alpha_i|\mathbf{x}_{it}) = \bar{\alpha}$, avec $\bar{\alpha}$ indépendant de i implique, par la loi des espérances itérées, que $\mathbb{E}(\alpha_i) = \mathbb{E}_{\mathbf{x}}\left[\mathbb{E}(\alpha|\mathbf{x}_{it})\right] = \tilde{\alpha}$. Elle implique aussi que,

$$\mathbb{C}\mathbf{ov}(\alpha_i, \mathbf{x}_{it}) = \mathbb{E}(\alpha_i, \mathbf{x}_{it}) - \mathbb{E}(\alpha_i)\,\mathbb{E}(\mathbf{x}_{it}) = \mathbb{E}_{\mathbf{x}}\left[\mathbb{E}(\alpha_i\mathbf{x}_{it}|\mathbf{x}_{it})\right] - \tilde{\alpha}\,\mathbb{E}(\mathbf{x}_{it}),$$

$$\mathbb{C}\mathbf{ov}(\alpha_i, \mathbf{x}_{it}) = \mathbb{E}_{\mathbf{x}}\left[\mathbb{E}(\alpha_i|\mathbf{x}_{it})\mathbf{x}_{it}\right] - \tilde{\alpha}\,\mathbb{E}(\mathbf{x}_{it}),$$

$$\mathbb{C}\mathbf{ov}(\alpha_i, \mathbf{x}_{it}) = \mathbb{E}_{\mathbf{x}}(\tilde{\alpha}\mathbf{x}_{it}) - \tilde{\alpha}\,\mathbb{E}(\mathbf{x}_{it}) = \tilde{\alpha}\,\mathbb{E}(\mathbf{x}_{it}) - \tilde{\alpha}\,\mathbb{E}(\mathbf{x}_{it}),$$

$$\mathbb{C}\mathbf{ov}(\alpha_i, \mathbf{x}_{it}) = 0.$$

  Donc α_i et $\mathbf{x}_{it}$ sont indépendants. Une telle restriction est très forte et n'est généralement pas vérifiée en pratique.

  - Par exemple, être en mesure de mieux négocier son salaire peut dépendre de son âge, de son niveau d'éducation et d'autres facteurs qui sont inclus dans $\mathbf{x}_{it}$. Cela suffit pour admettre que α_i et $\mathbf{x}_{it}$ ne sont pas indépendants.

- La prise en compte des effets individuels est donc importante.

Modèle pooled

- Vrai modèle :

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}. \tag{3}$$

- Spécification Pooled :

$$y_{it} = \alpha + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}. \tag{4}$$

- La spécification (4) est estimée par la méthode des MCO.
- Si la spécification est vraie, l'estimateur des MCO est convergent sous les hypothèses standards.
- Comme démontré précédemment, même si la vraie spécification est $y_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}$, l'estimateur pooled est convergent si $\mathbb{C}\mathbf{ov}\left(\alpha_i, \mathbf{x}_{it}\right) = \mathbf{0}$.
- Toutefois, $\mathbb{C}\mathbf{ov}\left(\alpha_i, \mathbf{x}_{it}\right) = \mathbf{0}$ est rarement vérifié.

Estimateur Between

- Vrai modèle :

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}. \tag{5}$$

- Considérons le modèle avec la moyenne des variables pour chaque individu :

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} y_{it} &= \frac{1}{T}\sum_{t=1}^{T}\alpha_i + \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}'_{it}\boldsymbol{\beta} + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}, \\ \bar{y}_i &= \alpha_i + \bar{\mathbf{x}}'_i\boldsymbol{\beta} + \bar{\varepsilon}_i. \end{aligned} \tag{6}$$

- La spécification between assimile les données de chaque individu à la moyenne. Une observation par individu, et par conséquent les effets individuels ne peuvent plus être estimés. Ils sont remplacés par un effet commun.

$$\bar{y}_i = \alpha + \bar{\mathbf{x}}'_i\boldsymbol{\beta} + \bar{\varepsilon}_i. \tag{7}$$

- L'estimateur between est donc l'estimateur des MCO de la spécification (7). Pour qu'il soit convergent, il faut que $\mathbb{E}(\alpha_i|\mathbf{x}_{it}) = \alpha$; i.e., $\mathbb{C}\mathbf{ov}\,(\alpha_i, \mathbf{x}_{it}) = \mathbf{0}$.

Estimateur Within ou Estimateur à effets fixes

- La condition $\mathbb{C}\mathbf{ov}\left(\alpha_i, \mathbf{x}_{it}\right) = \mathbf{0}$ est rarement vérifiée. Il faut donc trouver un estimateur dont la convergence ne nécessite pas cette condition.
- Un tel estimateur est l'estimateur within ou estimateur à effets fixes. Le terme « effets fixes » signifie que α_i et $\mathbf{x}_{it}$ sont corrélés : $\mathbb{C}\mathbf{ov}\left(\alpha_i, \mathbf{x}_{it}\right) \neq \mathbf{0}$.
- Les effets fixes n'impliquent pas que α_i est une variable constante (si non on aurait $\mathbb{C}\mathbf{ov}\left(\alpha_i, \mathbf{x}_{it}\right) = \mathbf{0}$). C'est un terme utilisé pour signifier que α_i et $\mathbf{x}_{it}$ sont corrélés.
- L'estimateur within (ou à effets fixes) est un estimateur des MCO sur la spécification panel en déviation par rapport à la moyenne de chaque individu.

Spécification panel : $y_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}$,

Spécification moyenne : $\bar{y}_i = \alpha_i + \bar{\mathbf{x}}'_i\boldsymbol{\beta} + \bar{\varepsilon}_i$,

Spécification en déviation : $y_{it} - \bar{y}_i = (\mathbf{x}'_{it} - \bar{\mathbf{x}}'_i)\boldsymbol{\beta} + \varepsilon_{it} - \bar{\varepsilon}_i$.

- L'estimateur within est donc l'estimateur des MCO de la régression de $y_{it} - \bar{y}_i$ sur les variables explicatives $\mathbf{x}'_{it} - \bar{\mathbf{x}}'_i$.
- La différence entre la spécification panel et la spécification moyenne élimine les effets individuels. Il n'y a donc plus d'hypothèse à faire sur les effets individuels.
- L'estimateur within (ou à effets fixes) est donc valide si $\mathbb{C}\mathbf{ov}\left(\alpha_i, \mathbf{x}_{it}\right) \neq \mathbf{0}$ (effets fixes) et aussi si $\mathbb{C}\mathbf{ov}\left(\alpha_i, \mathbf{x}_{it}\right) = \mathbf{0}$ (effets aléatoires). C'est un estimateur général.
- L'estimateur est convergent sous les hypothèses standard de la méthode des MCO.
    - L'exogénéité des variables explicatives est maintenant appliquée à $\mathbf{x}_{it} - \bar{\mathbf{x}}_i$. C'est-à-dire que $\mathbb{E}(\varepsilon_{it} - \bar{\varepsilon}_i.|\mathbf{x}_{it} - \bar{\mathbf{x}}_i) = 0$.
    - Une condition suffisante pour que l'exogénéité soit vérifiée est $\mathbb{E}(\varepsilon_{it}|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}) = 0$. Cette condition est qualifiée d'exogénéité « forte ».

- Une autre manière pour calculer l'estimateur within lorsque le nombre d'individu n'est pas grand est de faire recours à des variables muettes.
- Les données de l'individu i peuvent être présentées sous la forme suivante :

$$\underbrace{\begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix}}_{\mathbf{y}_i} = \underbrace{\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}}_{\mathbf{e}_T} \alpha_i + \underbrace{\begin{pmatrix} \mathbf{x}'_{i1} \\ \vdots \\ \mathbf{x}'_{iT} \end{pmatrix}}_{\mathbf{X}_i} \boldsymbol{\beta} + \underbrace{\begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix}}_{\boldsymbol{\varepsilon}_i},$$

$$\mathbf{y}_i = \mathbf{e}_T \alpha_i + \mathbf{x}'_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i,$$

où $\mathbf{e}_T$ est un vecteur de dimension T composé de 1.

- On peut également étendre la représentation matricielle à l'échelle de la population (en empilant les données les unes à la suite des autres) :

$$\begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbf{e}_T & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{e}_T \end{pmatrix}}_{N \text{ variables muettes}} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix} + \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_N \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_N \end{pmatrix},$$

$$\underbrace{\begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} \mathbf{e}_T & \dots & \mathbf{0} & \mathbf{X}_1 \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \dots & \mathbf{e}_T & \mathbf{X}_N \end{pmatrix}}_{\ddot{X}} \underbrace{\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \\ \boldsymbol{\beta} \end{pmatrix}}_{\ddot{\boldsymbol{\beta}}} \underbrace{\begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_N \end{pmatrix}}_{\boldsymbol{\varepsilon}},$$

$$\mathbf{y} = \ddot{X}\ddot{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}. \tag{8}$$

- L'estimateur des MCO de la spécification (8) est connu sous le nom d'estimateur des moindres carrés avec variables muettes.

- L'estimateur des moindres carrés avec variables muettes correspond à l'estimateur within (ou à effets fixes).
- Un estimateur qui impose que N ne tend pas vers l'infini. En effet, si $N \rightarrow \infty$, alors la dimension de $\ddot{\beta} = \infty$. Pour des raisons de convergence, l'inférence économétrique suppose toujours que la dimension des paramètres à estimer est finie.
- Dans le cas où la dimension $= \infty$, l'estimateur ne converge pas. Ce problème est connu sous le nom de **paramètres incidents** et est **à éviter absolument !**
- L'estimateur within (avec le modèle en déviation par rapport à la moyenne) n'est pas affecté par ce problème parce que les effets sont neutralisés.

Estimateur en première différence

- Pour se débarrasser des effets individuels sans imposer l'indépendance entre α_i et $\mathbf{x}_{it}$, on peut aussi considérer le modèle en différence première :

  Spécification à t : $y_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}$,

  Spécification à $t-1$: $y_{i(t-1)} = \alpha_i + \mathbf{x}'_{i(t-1)}\boldsymbol{\beta} + \varepsilon_{i(t-1)}$,

  Spécification en diff 1ère : $y_{it} - y_{i(t-1)} = (\mathbf{x}'_{it} - \mathbf{x}'_{i(t-1)})\boldsymbol{\beta} + \varepsilon_{it} - \varepsilon_{i(t-1)}$.

- La spécification en différence première ne comporte pas des effets individuels. Elle peut être estimée de manière convergence par MCO.
- L'estimateur en différence première est moins efficace que l'estimateur à effets fixes lorsque $T > 2$ et les erreurs ε_{it} sont iid.

Estimateur à effets aléatoires

- Lorsque $\mathbb{C}\mathbf{ov}\left(\alpha_i, \mathbf{x}_{it}\right) = \mathbf{0}$, il est montré que les estimateurs pooled, between et within sont tous convergents. Toutefois, ils ne sont pas efficaces. Il existe bien une manière d'estimer le modèle sous la condition $\mathbb{C}\mathbf{ov}\left(\alpha_i, \mathbf{x}_{it}\right) = \mathbf{0}$ de manière efficace. Cet estimateur s'appelle estimateur à effets aléatoires.
- Le terme « effets aléatoires » renvoie à l'indépendance entre α_i et $\mathbf{x}_{it}$.
- On suppose que $\alpha_i = u + u_i$ avec $\mathbb{E}(u_i) = 0$ et $\mathbb{V}\mathbf{ar}(u_i) = \sigma_u^2$.
- L'estimateur à effets aléatoires est un estimateur de moindres carrés généralisés. Il consiste à appliquer la méthode des MCO à la spécification :

$$y_{it} - \hat{\lambda}\bar{y}_i = (1-\hat{\lambda})u + (\mathbf{x}_{it}' - \hat{\lambda}\bar{\mathbf{x}}_i')\boldsymbol{\beta} + v_{it}, \tag{9}$$

avec $v_{it} = (1-\hat{\lambda})u_i + (\varepsilon_{it} - \hat{\lambda}\bar{\varepsilon}_i)$ et $\hat{\lambda}$ un estimateur convergent de $\lambda = 1 - \dfrac{\sigma_\varepsilon}{\sqrt{T\sigma_u^2 + \sigma_\varepsilon^2}}$.

- σ_ε^2 est estimé en utilisant le modèle within :

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N(T-1)-K} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - \bar{y}_i - (\mathbf{x}'_{it} - \bar{\mathbf{x}}'_i)\hat{\boldsymbol{\beta}}_{\mathsf{W}} \right)^2, \tag{10}$$

avec $K = \dim(\mathbf{x}_{it})$ et $\hat{\boldsymbol{\beta}}_{\mathsf{W}}$ l'estimateur within.

- σ_u^2 est estimé à partir de la spécification between et σ_ε^2 :

$$\hat{\sigma}_u^2 = \frac{1}{N-(K+1)} \sum_{i=1}^{N} \left(y_{it} - \hat{\alpha}_{\mathsf{B}} - \bar{\mathbf{x}}'_i \hat{\boldsymbol{\beta}}_{\mathsf{B}} \right)^2 - \frac{1}{T}\hat{\sigma}_\varepsilon^2, \tag{11}$$

avec $\begin{bmatrix} \hat{\alpha}_{\mathsf{B}} \\ \hat{\boldsymbol{\beta}}_{\mathsf{B}} \end{bmatrix}$ est l'estimateur between.

Inférence

- Tous les estimateurs étudiés sont des estimateurs de MCO sur des formes transformées du modèle données par :

$$\tilde{y}_{it} = \tilde{\mathbf{w}}'_{it}\tilde{\boldsymbol{\beta}} + \tilde{\varepsilon}_{it}. \tag{12}$$

- Exemples
  - Pour l'estimateur between, $\tilde{y}_{i1} = \bar{y}_i$, $\tilde{\mathbf{x}}_{i1} = \begin{bmatrix} 1 \\ \bar{\mathbf{x}}_i \end{bmatrix}$, $\tilde{\boldsymbol{\beta}} = \begin{bmatrix} u \\ \boldsymbol{\beta} \end{bmatrix}$ et $\tilde{\varepsilon}_{i1} = \bar{\varepsilon}_i$ ($T = 1$ dans le modèle transformé).
  - Pour l'estimateur du modèle à effets fixes, $\tilde{y}_{it} = y_{it} - \bar{y}_i$, $\tilde{\mathbf{x}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i$, $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}$ et $\tilde{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_i$.
  - Pour l'estimateur en diff 1ère, $\tilde{y}_{it} = y_{it} - y_{it}$, $\tilde{\mathbf{x}}_{it} = \mathbf{x}_{it} - \mathbf{x}_{it}$, $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}$ et $\tilde{\varepsilon}_{it} = \varepsilon_{it} - \varepsilon_{it}$.
- L'inférence est établie sur la forme transformée générale (12)

- **Hypothèses du principe des MCO**

  H.1. Linéarité : $\tilde{y}_{it} = \tilde{\mathbf{w}}'_{it}\tilde{\boldsymbol{\beta}} + \tilde{\varepsilon}_{it}$.

  H.2. Indépendance linéaire des variables explicatives : $\mathrm{rang}(\widetilde{\mathbf{W}})$ est une matrice de plein rang.

  H.3. Exogénéité des variables explicatives : $\mathbb{E}(\tilde{\varepsilon}_{it}|\tilde{\mathbf{z}}_{i1}, \ldots, \tilde{\mathbf{z}}_{iT}) = 0$.

  H.4. Homoscédasticité et non autocorrélation des erreurs : pour tout i, $\mathbb{V}\mathbf{ar}(\tilde{\varepsilon}_{it}|\tilde{\mathbf{z}}_{i1}, \ldots, \tilde{\mathbf{z}}_{iT}) = \sigma^2_{\varepsilon}$. De plus, pour $i \neq j$ ou $t \neq s$, $\mathbb{E}(\tilde{\varepsilon}_{it}\tilde{\varepsilon}_{js}|\tilde{\mathbf{z}}_{i1}, \ldots, \tilde{\mathbf{z}}_{iT}, \tilde{\mathbf{z}}_{j1}, \ldots, \tilde{\mathbf{z}}_{jT}) = 0$.

- Les estimateurs de modèles de panel sont convergents si H.1.-H.3. sont vérifiées dans le cadre de la spécification (12).

- En revanche, H.4. n'est généralement pas vérifiée.
  - Pour un même individu, $\varepsilon_{i(t-1)}$ peut être corrélé ε_{it}.
  - Il faut faire recours à des méthodes robustes (e.g., bootstrap) pour calculer Est.Asy. $\mathbb{V}\mathbf{ar}\left(\hat{\tilde{\boldsymbol{\beta}}}_{\text{MCO}}\right)$.

- Estimateur (robuste) de la variance

$$\text{Est.Asy. } \mathbb{V}\mathbf{ar}\left(\hat{\tilde{\boldsymbol{\beta}}}_{\text{MCO}}\right) = \left[\sum_{i,t} \tilde{\mathbf{w}}_{it}\tilde{\mathbf{w}}_{it}'\right]^{-1} \sum_{i,t,s} \tilde{\mathbf{w}}_{it}\tilde{\mathbf{w}}_{is}'\hat{\tilde{\varepsilon}}_{it}\hat{\tilde{\varepsilon}}_{is} \left[\sum_{i,t} \tilde{\mathbf{w}}_{it}\tilde{\mathbf{w}}_{it}'\right]^{-1}$$

- Alternative (Bootstrap)
  1. Tirer avec remise N individus de la population de référence et réestimer le modèle de panel (lorsqu'un individu est tiré on considère ses T observations)
  2. Répliquer l'étape précédente B fois. Soit $\hat{\tilde{\boldsymbol{\beta}}}_1, \ldots, \hat{\tilde{\boldsymbol{\beta}}}_B$ les B estimateurs obtenus.
  3. Est.Asy. $\mathbb{V}\mathbf{ar}\left(\hat{\tilde{\boldsymbol{\beta}}}_{\text{MCO}}\right)$ est donné par,

$$\text{Est.Asy. } \mathbb{V}\mathbf{ar}\left(\hat{\tilde{\boldsymbol{\beta}}}_{\text{MCO}}\right) = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\tilde{\boldsymbol{\beta}}}_b - \hat{\tilde{\boldsymbol{\beta}}}_*\right)\left(\hat{\tilde{\boldsymbol{\beta}}}_b - \hat{\tilde{\boldsymbol{\beta}}}_*\right)',$$

avec $\hat{\tilde{\boldsymbol{\beta}}}_* = \frac{1}{B}\sum_{b=1}^{B}\hat{\tilde{\boldsymbol{\beta}}}_b$.

Effets fixes Vs Effets aléatoires

- Deux estimateurs résument TOUT.
  1. Estimateur à effets fixes valide (convergent) si $\mathbb{C}\mathbf{ov}\left(\alpha_i, \mathbf{x}_{it}\right) \neq \mathbf{0}$, mais aussi convergent si $\mathbb{C}\mathbf{ov}\left(\alpha_i, \mathbf{x}_{it}\right) = \mathbf{0}$.
  2. Estimateur à effets aléatoires convergent et efficace si $\mathbb{C}\mathbf{ov}\left(\alpha_i, \mathbf{x}_{it}\right) = \mathbf{0}$, mais biaisé lorsque $\mathbb{C}\mathbf{ov}\left(\alpha_i, \mathbf{x}_{it}\right) \neq \mathbf{0}$.
- Bien que l'estimateur à effets fixes soit valide dans tous les cas, il ne sera pas toujours préféré parce qu'il est moins efficace que l'estimateur à effets aléatoires lorsque $\mathbb{C}\mathbf{ov}\left(\alpha_i, \mathbf{x}_{it}\right) \neq \mathbf{0}$.
- Si $\mathbb{C}\mathbf{ov}\left(\alpha_i, \mathbf{x}_{it}\right) \neq \mathbf{0}$, on choisit les effets fixes. Si $\mathbb{C}\mathbf{ov}\left(\alpha_i, \mathbf{x}_{it}\right) = \mathbf{0}$ on préfère les effets aléatoires (mais l'estimateur à effets fixes est aussi convergent).
- Pour choisir entre effets fixes ou effets aléatoires, on peut utiliser **le test de Hausman**.

- Soit $\hat{\boldsymbol{\beta}}_{\text{FE}}$ l'estimateur à effets fixes et $\hat{\boldsymbol{\beta}}_{\text{RE}}$ l'estimateur à effets aléatoires, avec $\hat{\boldsymbol{\beta}}_{\text{FE}}$ et $\hat{\boldsymbol{\beta}}_{\text{RE}}$ uniquement composés des paramètres associés aux variables qui changent dans le temps. En effet, seules ces paramètres peuvent être estimés via la spécification within. Cela implique donc que $\dim(\hat{\boldsymbol{\beta}}_{\text{FE}}) = \dim(\hat{\boldsymbol{\beta}}_{\text{RE}})$.
- Le principe du test est le suivant. Si $\mathbb{C}\mathbf{ov}\left(\alpha_i, \mathbf{x}_{it}\right) = \mathbf{0}$, alors $\hat{\boldsymbol{\beta}}_{\text{FE}}$ et $\hat{\boldsymbol{\beta}}_{\text{RE}}$ devraient être très proches, car ils sont tous convergents. Mais si $\mathbb{C}\mathbf{ov}\left(\alpha_i, \mathbf{x}_{it}\right) \neq \mathbf{0}$, alors $\hat{\boldsymbol{\beta}}_{\text{FE}}$ et $\hat{\boldsymbol{\beta}}_{\text{RE}}$ sont très différents. Le principe du test est donc de comparer $\hat{\boldsymbol{\beta}}_{\text{FE}} - \hat{\boldsymbol{\beta}}_{\text{RE}}$ à $\mathbf{0}$.
- Hypothèse nulle : effets aléatoires.
- Statistique du test :

$$H = \left(\hat{\boldsymbol{\beta}}'_{\text{FE}} - \hat{\boldsymbol{\beta}}'_{\text{RE}}\right)\left\{\text{Est.Asy. } \mathbb{V}\mathbf{ar}\left(\hat{\boldsymbol{\beta}}_{\text{FE}}\right) - \text{Est.Asy. } \mathbb{V}\mathbf{ar}\left(\hat{\boldsymbol{\beta}}_{\text{RE}}\right)\right\}^{-1}\left(\hat{\boldsymbol{\beta}}_{\text{FE}} - \hat{\boldsymbol{\beta}}_{\text{RE}}\right)$$

- Sous l'hypothèse nulle, $H \sim \chi^2(\dim(\hat{\boldsymbol{\beta}}_{\text{FE}}))$.

Extension

- Modèles avec effets fixes individuels et temporels :

$$y_{it} = \alpha_i + \gamma_t + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}. \tag{13}$$

  On peut utiliser une transformation similaire à celle de la spécification within pour se débarrasser des effets.

- Panel dynamique

$$y_{it} = \alpha_i + \theta y_{i(t-1)} + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}. \tag{14}$$

  L'estimateur des MCO n'est pas convergent car $y_{i(t-1)}$ et ε_{it} sont corrélés. L'estimateur GMM est souvent utilisé.

- Applications avec R : script `panel.estim.R` et `panel.lineaire.R`