Econométrie 2

Chapitre 2 : données de panel.

ENSAE 2021-2022

Michael Visser

CREST-ENSAE

Plan

Introduction

Exogénéité des résidus mais autocorrélation

$E(X_{it}\alpha_i) \neq 0$ mais $E(X_{it}\varepsilon_{it'}) = 0,\ \forall (t, t')$ (exogénéité stricte)

$E(X_{it}\alpha_i) \neq 0$ mais $E(X_{it}\varepsilon_{it'}) = 0,\ \forall t' \geq t$ (exogénéité faible)

Exemple : effet de la police sur la criminalité

Définition et intérêt

- Panel = échantillon d'unités (individus, entreprises...) suivies dans le temps.
- Deux intérêts principaux :
  1. résoudre (ou limiter) le problème d'endogénéité ;
  2. étudier la dynamique et séparer la dépendance d'état de l'hétérogénéité inobservée.
- Un problème : l'attrition endogène. Cf. chapitre sur le problème de sélection.
- On notera Y_{it} la variable expliquée et X_{it} le vecteur (dimension $K \times 1$) des variables explicatives de l'unité i en t.
- On supposera toujours $(Y_{i1}, X_{i1}, ..., Y_{iT}, X_{iT})_{i=1...n}$ i.i.d. Mais (Y_{is}, X_{is}) sera en général corrélé à (Y_{it}, X_{it}), pour $s \neq t$.

Le modèle étudié

- Modèle de base :
$$Y_{it} = X_{it}'\beta_0 + \nu_{it}$$
pour $i = 1...n$ et $t = 1...T$.
- Asymptotique en n, pas en T. Hypothèse pas toujours évidente (en macroéconomie, on peut avoir $n = 70$ pays et $T = 50$ périodes).
- Le résidu sera souvent décomposé en deux termes : $\nu_{it} = \alpha_i + \varepsilon_{it}$.
- α_i est l'« effet individuel ». Agrège les facteurs inobservés constant dans le temps. .
- ε_{it}, souvent appelé « choc idiosyncratique », agrège les facteurs inobservés variables dans le temps.
- On supposera dans la suite au moins l'*exogénéité faible* :
$$E(X_{it}\varepsilon_{it'}) = 0, \ \forall t' \geq t. \qquad (1)$$

Le modèle étudié

- α_i est souvent appelé « effet fixe » ou « effet aléatoire ».
- Dans la littérature économétrique on considère α_i soit comme un paramètre fixe, comme β_0, soit comme une variable aléatoire, comme ε_{it}.
- Souvent plus simple de le supposer aléatoire : c'est l'approche adoptée dans ce cours.
- La question essentielle est alors : α_i est-il corrélé à X_{it} ?
- Si l'on suppose $E(X_{it}\alpha_i) = 0$, alors il n'y a pas d'endogénéité et les MCO sont convergents.
- Mais les estimateurs habituels des écarts-types ne sont pas convergents (à cause de l'autocorrélation des résidus) $\Rightarrow$ correction nécessaire sinon inférence incorrecte.
- Si, $E(X_{it}\alpha_i) \neq 0$, les panels peuvent permettre de résoudre le problème d'endogénéité.
- Mais on ne pourra pas inclure dans X_{it} des variables constantes dans le temps car on ne pourra séparer leur effet de α_i.

Plan

Introduction

Exogénéité des résidus mais autocorrélation

$E(X_{it}\alpha_i) \neq 0$ mais $E(X_{it}\varepsilon_{it'}) = 0$, $\forall (t, t')$ (exogénéité stricte)

$E(X_{it}\alpha_i) \neq 0$ mais $E(X_{it}\varepsilon_{it'}) = 0$, $\forall t' \geq t$ (exogénéité faible)

Exemple : effet de la police sur la criminalité

Discussion sur l'hypothèse et MCO

- Exemple : $i =$ classe, $t =$ élève, $Y_{it} =$ note à un examen, $X_{it} =$ âge, sexe, diplôme des parents...
- Probablement $E(\nu_{is}\nu_{it}) \neq 0$, du fait par exemple d'« effet profs ».
- Mais $E(X_{it}\nu_{it}) = 0$ reste crédible, en particulier si les professeurs sont affectés aléatoirement aux classes.
- On a :

$$\begin{aligned}
\widehat{\beta}_{MCO} &= \left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} X_{it}X_{it}'\right]^{-1}\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} X_{it}Y_{it}\right] \\
&= \beta_0 + \left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} X_{it}X_{it}'\right]^{-1}\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} X_{it}\nu_{it}\right].
\end{aligned}$$

- Il s'agit de l'estimateur standard obtenu à partir des nT observations (en anglais : pooled OLS estimator).

Propriétés des MCO

- Sous la seule hypothèse $E(X_{it}\nu_{it}) = 0$ et d'échantillonnage i.i.d. d'unités, l'estimateur est convergent et asymptotiquement normal mais il faut adapter l'estimateur des écarts-types.
- Soit (en omettant l'indice i) $J = E\left[\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right]$. Alors :

$$\sqrt{n}\left(\widehat{\beta}_{MCO} - \beta_{\mathbf{0}}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, J^{-\mathbf{1}} E\left[\left(\frac{\mathbf{1}}{T}\sum_{t=\mathbf{1}}^{T} X_t\nu_t\right)\left(\frac{\mathbf{1}}{T}\sum_{t=\mathbf{1}}^{T} X_t\nu_t\right)'\right] J^{-\mathbf{1}}\right).$$

- Preuve : on a

$$\sqrt{n}\left(\widehat{\beta}_{MCO} - \beta_{\mathbf{0}}\right) = \left[\frac{\mathbf{1}}{nT}\sum_{i=\mathbf{1}}^{n}\sum_{t=\mathbf{1}}^{T} X_{it}X_{it}'\right]^{-\mathbf{1}}\left[\frac{\mathbf{1}}{\sqrt{n}T}\sum_{i=\mathbf{1}}^{n}\sum_{t=\mathbf{1}}^{T} X_{it}\nu_{it}\right]$$

$$= J^{-\mathbf{1}}\left[\frac{\mathbf{1}}{\sqrt{n}T}\sum_{i=\mathbf{1}}^{n}\sum_{t=\mathbf{1}}^{T} X_{it}\nu_{it}\right] + \left[\left[\frac{\mathbf{1}}{nT}\sum_{i=\mathbf{1}}^{n}\sum_{t=\mathbf{1}}^{T} X_{it}X_{it}'\right]^{-\mathbf{1}} - J^{-\mathbf{1}}\right]\left[\frac{\mathbf{1}}{\sqrt{n}T}\sum_{i=\mathbf{1}}^{n}\sum_{t=\mathbf{1}}^{T} X_{it}\nu_{it}\right].$$

D'après le TCL on a

$$\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T} X_{it}\nu_{it} \xrightarrow{d} \mathcal{N}(0, B)$$

où

$$B = V\left(\frac{1}{T}\sum_{t=1}^{T} X_{it}\nu_{it}\right) = E\left[\left(\frac{1}{T}\sum_{t=1}^{T} X_{it}\nu_{it}\right)\left(\frac{1}{T}\sum_{t=1}^{T} X_{it}\nu_{it}\right)'\right].$$

Donc

$$\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T} X_{it}\nu_{it} = O_p(1)$$

ce qui donne

$$\sqrt{n}\left(\widehat{\beta}_{MCO} - \beta_0\right) = J^{-1}\left[\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T} X_{it}\nu_{it}\right] + o_p(1)O_p(1) = J^{-1}\left[\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T} X_{it}\nu_{it}\right] + o_p(1).$$

On en déduit finalement que

$$\sqrt{n}\left(\widehat{\beta}_{MCO} - \beta_0\right) \xrightarrow{d} \mathcal{N}\left(0, J^{-1}BJ^{-1}\right).$$

Propriétés des MCO

Dans cette preuve, nous avons utilisé plusieurs définitions et propriétés de la théorie asymptotique en stat/proba (voir par exemple Wooldridge, chapitre 3) :

1. Une séquence de variables aléatoires (scalaires) $\{a_n\}_{n=1}^{\infty}$ converge en probabilité vers une constante a si pour tout $\varepsilon > 0$ on a $P(|a_n - a| > \varepsilon) \to 0$ lorsque $n \to \infty$. On écrit alors $a_n \xrightarrow{P} a$ ou $\text{plim}(a_n) = a$. Lorsque $a = 0$, on dit que $\{a_n\}$ est $o_p(1)$, noté $a_n = o_p(1)$.
2. Une séquence $\{a_n\}_{n=1}^{\infty}$ est bornée en probabilité si pour tout $\varepsilon > 0$ il existe un nombre réel $b_\varepsilon < \infty$ et un nombre entier n_ε tel que $P(|a_n| \geq b_\varepsilon) < \varepsilon$ pour tout $n > n_\varepsilon$. On écrit alors $a_n = O_p(1)$.
3. Lorsque une séquence $\{a_n\}_{n=1}^{\infty}$ converge en loi vers une variable aléatoire a, noté $a_n \xrightarrow{d} a$, on a $a_n = O_p(1)$.
4. Lorsque les séquences $\{a_n\}_{n=1}^{\infty}$ et $\{b_n\}_{n=1}^{\infty}$ sont telles que $a_n = o_p(1)$ et $b_n = O_p(1)$, alors $a_n b_n = o_p(1)$.
5. Si $a_n \xrightarrow{d} a$ et $b_n - a_n \xrightarrow{P} 0$, alors $b_n \xrightarrow{d} a$.
6. Ces différentes définitions et propriétés s'appliquent également lorsque $\{a_n\}_{n=1}^{\infty}$ et $\{b_n\}_{n=1}^{\infty}$ sont des séquences de vecteurs ou matrices aléatoires.

Propriétés des MCO

- L'élément de la $(l+1)-$ème ligne et $(j+1)-$ème colonne de B s'écrit

$$B_{l+1,j+1} = E\left[\left(\frac{1}{T}\sum_{t=1}^{T} X_{ilt}\nu_{it}\right)\left(\frac{1}{T}\sum_{t=1}^{T} X_{ijt}\nu_{it}\right)\right]$$

où X_{ilt} (resp. X_{ijt}) est le $l-$ème (resp. $j-$ème) élément de X_{it}.

- On n'impose aucune restriction sur $B_{l+1,j+1}$. En particulier on ne fait pas l'hypothèse d'homoscédasticité, c'est à dire

$$E\left(X_{ilt}X_{ijt}\nu_{it}^2\right) = E\left(X_{ilt}X_{ijt}\right)E\left(\nu_{it}^2\right) = E\left(X_{ilt}X_{ijt}\right)\sigma^2$$

ni l'hypothèse de non-corrélation entre les termes d'erreur (conditionnellement aux variables explicatives), c'est à dire

$$E\left(X_{ilt}\nu_{it}X_{ijs}\nu_{is}\right) = 0$$

pour tout $s \neq t$.

Propriétés des MCO

- La variance asymptotique $J^{-1}BJ^{-1}$ peut s'estimer de manière convergente par

$$\widehat{V} = \widehat{J}^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}X_{it}\widehat{\nu}_{it}\right)\left(\frac{1}{T}\sum_{t=1}^{T}X_{it}\widehat{\nu}_{it}\right)'\right]\widehat{J}^{-1},$$

avec $\widehat{J} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}X_{it}X_{it}'$ et $\widehat{\nu}_{it} = Y_{it} - X_{it}\widehat{\beta}_{MCO}$.

- Cet estimateur est convergent et robuste (i) à l'hétéroscédasticité, et (ii) à l'autocorrélation de ν (i.e., $\text{Cov}(\nu_{is}, \nu_{it}) \neq 0$).
- On peut obtenir de tels écarts-types avec l'option `cluster` de Stata (voir l'application à la fin de ce chapitre).

Plan

Introduction

Exogénéité des résidus mais autocorrélation

$E(X_{it}\alpha_i) \neq 0$ mais $E(X_{it}\varepsilon_{it'}) = 0, \ \forall (t, t')$ (exogénéité stricte)

$E(X_{it}\alpha_i) \neq 0$ mais $E(X_{it}\varepsilon_{it'}) = 0, \ \forall t' \geq t$ (exogénéité faible)

Exemple : effet de la police sur la criminalité

Discussion sur l'hypothèse

- Jusqu'à la fin de ce chapitre on enlève explicitement la constante de X_{it}. Donc $X_{it} = (X_{i1t}, ..., X_{iK-1t})'$ est maintenant de dimension $(K-1) \times 1$. Le modèle s'écrit alors comme

$$Y_{it} = \delta_0 + X_{it}'\beta_0 + \alpha_i + \varepsilon_{it},$$

  où δ_0 est la constante et β_0 est un vecteur de dimension $(K-1) \times 1$.
- Supposons que $E(X_{it}'\alpha_i) \neq 0$ mais que l'hypothèse d'*exogénéité stricte* est vérifiée, i.e. $E(X_{it}\varepsilon_{it'}) = 0 \; \forall (t, t') \in \{1, ..., T\}^2$.
- Les 2 exemples suivants montrent que l'exogénéité stricte est parfois restrictive.
- 1er exemple : Y_{it}= salaire, X_{it}= expérience spécifique (temps passé dans la dernière entreprise), α_i = capacités inobservées de l'individu.
- X_{it} a priori non corrélé aux chocs futurs, non anticipables, de salaire $\Rightarrow$ exogénéité faible crédible.
- Mais X_{it} peut être corrélé aux chocs passés, si l'on décide de rester en fonction de son salaire passé.

Discussion sur l'hypothèse

- 2ème exemple : $X_{it} = Y_{it-1}$. Dans ce cas, on a, par exemple si ε_{it} i.i.d. et $E(\alpha_i \varepsilon_{it}) = 0\ \forall t$,

$$\begin{aligned} E(X_{it}\varepsilon_{it-1}) &= E(Y_{it-1}\varepsilon_{it-1}) = E\left[(\delta_0 + \beta_{01} Y_{it-2} + \alpha_i + \varepsilon_{it-1})\varepsilon_{it-1}\right] \\ &= E(\varepsilon^2_{it-1}) \neq 0 \end{aligned}$$

et de même $E(X_{it'}\varepsilon_{it-1}) \neq 0$ en général pour $t' > t$. En revanche, l'exogénéité faible est bien satisfaite :

$$E(X_{it}\varepsilon_{it'}) = E(Y_{it-1}\varepsilon_{it'}) = 0,\ \forall t' \geq t$$

car Y_{it-1} ne dépend que de α_i et des termes d'erreur passés ε_{it} pour $t < t'$.

Estimateur des différences premières : idée

- Deux estimateurs classiques : l'estimateur within et l'estimateur par différences premières.
- Idée commune : obtenir une équation où l'on s'est débarrassé du α_i.
- Différences premières : on a, en notant $\Delta U_{it} = U_{it} - U_{it-1}$ pour toute variable U,
$$\Delta Y_{it} = \Delta X_{it}'\beta_0 + \Delta\varepsilon_{it}. \tag{2}$$
- Remarquons que sous l'hypothèse d'exogénéité stricte :
$$E\left[\Delta X_{it}\Delta\varepsilon_{it}\right] = E[X_{it}\varepsilon_{it}] - E[X_{it}\varepsilon_{it-1}] - E[X_{it-1}\varepsilon_{it}] + E[X_{it-1}\varepsilon_{it-1}] = 0.$$
- Dans (2), les régresseurs sont exogènes et on peut les estimer de manière convergente par les MCO des données empilées (en utilisant $n(T-1)$ observations car on perd la première observation pour chaque unité i).

Estimateur des différences premières : propriétés

- Omettons l'indice i. On a $E\left[\Delta X_t \Delta Y_t\right] = E\left[\Delta X_t \Delta X_t'\right]\beta_0$ donc, si $J_{FD} = E\left[\frac{1}{T-1}\sum_{t=2}^{T} \Delta X_t \Delta X_t'\right]$ est de plein rang,

$$\beta_0 = J_{FD}^{-1} E\left[\frac{1}{T-1}\sum_{t=2}^{T} \Delta X_t \Delta Y_t\right].$$

- Comme les MCO, l'estimateur des différences premières vérifie

$$\begin{aligned}
\widehat{\beta}_{FD} &= \left[\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T} \Delta X_{it} \Delta X_{it}'\right]^{-1}\left[\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T} \Delta X_{it} \Delta Y_{it}\right] \\
&= \beta_0 + \left[\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T} \Delta X_{it} \Delta X_{it}'\right]^{-1}\left[\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T} \Delta X_{it} \Delta \varepsilon_{it}\right].
\end{aligned}$$

- Sous les hypothèses d'exogénéité stricte et d'échantillonnage i.i.d. d'unités, l'estimateur est convergent et (preuve identique à la preuve page 8/9) :

$$\sqrt{n}\left(\widehat{\beta}_{FD} - \beta_0\right) \xrightarrow{d} \mathcal{N}\left(0, J_{FD}^{-1} E\left[\left(\frac{1}{T-1}\sum_{t=2}^{T} \Delta X_t \Delta \varepsilon_t\right)\left(\frac{1}{T-1}\sum_{t=2}^{T} \Delta X_t \Delta \varepsilon_t\right)'\right] J_{FD}^{-1}\right).$$

Estimateur des différences premières : propriétés

- On peut estimer la matrice de variance asymptotique par

$$\widehat{V} = \left[\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta X_{it}\Delta X_{it}'\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T-1}\sum_{t=2}^{T}\Delta X_{it}\widehat{\Delta\varepsilon}_{it}\right)\left(\frac{1}{T-1}\sum_{t=2}^{T}\Delta X_{it}\widehat{\Delta\varepsilon}_{it}\right)'\right]$$
$$\times\left[\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta X_{it}\Delta X_{it}'\right]^{-1}.$$

  avec $\widehat{\Delta\varepsilon}_{it} = \Delta Y_{it} - \Delta X_{it}'\widehat{\beta}_{FD}$.

- Cet estimateur est convergent et robuste (i) à l'hétéroscédasticité, (ii) à l'autocorrélation de $\Delta\varepsilon$ (i.e., $\text{Cov}(\Delta\varepsilon_s, \Delta\varepsilon_t) \neq 0$).

- $\widehat{\beta}_{FD}$ est efficace asymptotiquement sous l'hypothèse beaucoup plus forte que i) $E(\Delta X_t \Delta\varepsilon_t^2 \Delta X_t') = E(\Delta\varepsilon_t^2)E(\Delta X_t \Delta X_t')$ avec $E(\Delta\varepsilon_t^2)$ indépendant de t (hypothèse d'homoscédasticité) ; ii) $E(\Delta X_t \Delta\varepsilon_t \Delta\varepsilon_s \Delta X_s') = 0$ pour tout $s \neq t$ (hypothèse de non-corrélation entre $\Delta\varepsilon_s$ et $\Delta\varepsilon_t$ conditionnellement à ΔX_t et ΔX_s).

Estimateur des différences premières : propriétés

- Sous ces deux hypothèses on obtient

$$E\left[\left(\frac{1}{T-1}\sum_{t=2}^{T}\Delta X_t\Delta\varepsilon_t\right)\left(\frac{1}{T-1}\sum_{t=2}^{T}\Delta X_t\Delta\varepsilon_t\right)'\right]=E[(\Delta\varepsilon_t)^2]E\left[\frac{1}{(T-1)^2}\sum_{t=2}^{T}\Delta X_t\Delta X_t'\right].$$

- Dans ce cas, la variance asymptotique de $\sqrt{n}\left(\widehat{\beta}_{FD}-\beta_0\right)$ se simplifie en $V=E[(\Delta\varepsilon_t)^2]J_{FD}^{-1}/(T-1)$.

- La non-corrélation entre $\Delta\varepsilon_s$ et $\Delta\varepsilon_t$ est satisfaite si $(\varepsilon_t)_t$ suit une marche aléatoire :

$$\varepsilon_{t+1}=\varepsilon_t+\eta_{t+1},$$

  où les $(\eta_t)_t$ sont indépendants et $E(\eta_t)=0$ pour tout t. En effet, dans ce cas $E(\Delta\varepsilon_s\Delta\varepsilon_t|\Delta X_s,\Delta X_t)=E(\eta_{s+1}\eta_{t+1}|\Delta X_s,\Delta X_t)=0$. Mais si les $(\varepsilon_t)_t$ sont indépendants,

$$\begin{aligned}E(\Delta\varepsilon_t\Delta\varepsilon_{t+1}|\Delta X_t,\Delta X_{t+1}) &= E((\varepsilon_t-\varepsilon_{t-1})(\varepsilon_{t+1}-\varepsilon_t)|\Delta X_t,\Delta X_{t+1})\\ &= -E(\varepsilon_t^2|\Delta X_t,\Delta X_{t+1})<0.\end{aligned}$$

Estimateur within : idée

- Deuxième estimateur classique : l'estimateur within. On a, en notant, pour toute v.a. U, $\overline{U}_i = \frac{1}{T}\sum_{t=1}^{T} U_{it}$ et $\tilde{U}_{it} = U_{it} - \overline{U}_i$,

$$\overline{Y}_i = \delta_0 + \overline{X}_i'\beta_0 + \alpha_i + \overline{\varepsilon}_i$$

et donc

$$\tilde{Y}_{it} = \tilde{X}_{it}'\beta_0 + \tilde{\varepsilon}_{it}. \tag{3}$$

- Là aussi, de par l'exogénéité stricte,

$$E\left[\tilde{X}_{it}\tilde{\varepsilon}_{it}\right] = 0.$$

- Donc les régresseurs de (3) sont exogènes et β_0 peut être estimé de manière convergente par les MCO empilés : estimateur within.

Estimateur within : propriétés

- Comme pour les différences premières, on a, en notant $J_W = E\left[\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_t\tilde{X}_t'\right]$ (supposée de plein rang),

$$\beta_0 = J_W^{-1}E\left[\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_t\tilde{Y}_t\right].$$

- Et l'estimateur within vérifie

$$\begin{aligned}\widehat{\beta}_W &= \left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right]^{-1}\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{Y}_{it}\right]\\ &= \beta_0 + \left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right]^{-1}\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\varepsilon_{it}\right].\end{aligned}$$

N.B. : on a remplacé $\tilde{\varepsilon}_{it}$ par ε_{it} car $\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}\overline{\varepsilon}_i = 0$.

- Ainsi (preuve similaire à celle de la page 8/9) :

$$\sqrt{n}\left(\widehat{\beta}_W - \beta_0\right) \xrightarrow{d} \mathcal{N}\left(0, J_W^{-1}E\left[\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_t\varepsilon_t\right)\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_t\varepsilon_t\right)'\right]J_W^{-1}\right).$$

- Comme pour les différences premières, on peut estimer de manière robuste cette matrice de variance par

$$\widehat{V} = \left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}\widehat{\varepsilon}_{it}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}\widehat{\varepsilon}_{it}\right)'\right]$$
$$\times\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right]^{-1}.$$

  avec $\widehat{\varepsilon}_{it} = \tilde{Y}_{it} - \tilde{X}_{it}'\widehat{\beta}_W$.

- $\widehat{\beta}_W$ est efficace asymptotiquement sous l'hypothèse beaucoup plus forte que

$$E\left[\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_t\varepsilon_t\right)\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_t\varepsilon_t\right)'\right] = E[\varepsilon_t^2]E\left[\frac{1}{T^2}\sum_{t=1}^{T}\tilde{X}_t\tilde{X}_t'\right],$$

  qui combine homoscédasticité et non corrélation entre ε_s et ε_t (conditionnellement aux variables explicatives).

- La variance asymptotique s'écrit alors $V = E[\varepsilon_t^2]J_W^{-1}/T$.

Estimateur within : propriétés

- Pour estimer $E[\varepsilon_t^2] = \sigma_\varepsilon^2$ on utilise que

$$\begin{aligned} E(\tilde{\varepsilon}_{it}^2) &= E(\varepsilon_{it}^2) + E(\overline{\varepsilon}_i^2) - 2E(\varepsilon_{it}\overline{\varepsilon}_i) \\ &= \sigma_\varepsilon^2 + \sigma_\varepsilon^2/T - 2\sigma_\varepsilon^2/T = \sigma_\varepsilon^2(1 - 1/T). \end{aligned}$$

  sous l'hypothèse de non-corrélation entre ε_{it} et ε_{is}.
- Donc on a $\frac{1}{T-1}\sum_{t=1}^T E(\tilde{\varepsilon}_{it}^2) = \sigma_\varepsilon^2$.
- Un estimateur convergent (et sans biais) de σ_ε^2 est alors (comme il n'y pas de constante dans le modèle (3), le nombre de paramètres à estimer vaut $K-1$)) :

$$\widehat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^n \sum_{t=1}^T \widehat{\tilde{\varepsilon}}_{it}^2}{(n(T-1) - K + 1)}.$$

- Attention, cet estimateur diffère de l'estimateur habituel de la variance (du terme d'erreur) qu'on obtient à l'issue de MCO de $\tilde{Y}_{it}$ sur $\tilde{X}_{it}$: $\frac{\sum_{i=1}^n \sum_{t=1}^T \widehat{\tilde{\varepsilon}}_{it}^2}{(nT-K+1)}$ (voir Econométrie 1, Chapitre 3).

Comparaison entre les différences premières et le within

- $\widehat{\beta}_{FD}$ et $\widehat{\beta}_{W}$ coïncident quand $T = 2$ mais pas quand $T > 2$.
- $\widehat{\beta}_{W}$ sera plus précis lorsque les $(\varepsilon_t)_t$ sont très peu corrélés, $\widehat{\beta}_{FD}$ sera meilleur lorsque les $(\varepsilon_t)_t$ sont très corrélés.
- Les deux estimateurs sont tous deux convergents et asymptotiquement normaux sous l'hypothèse d'exogénéité stricte.
- Mais ils sont en général non convergents sous la seule hypothèse d'exogénéité faible, et convergent vers des limites en général différentes.

$\Rightarrow$ s'ils sont très différents, l'hypothèse d'exogénéité stricte est suspecte.

Un test simple de la condition d'exogénéité stricte

- Idée : considérons $T = 2$ et le modèle en différence première :

$$\begin{aligned} \Delta Y_2 &= \Delta X_2' \beta_0 + \Delta \varepsilon_2 \\ &= \Delta X_2' \beta_0 + X_2' \gamma_0 + \Delta \varepsilon_2, \end{aligned} \tag{4}$$

  avec $\gamma_0 = 0$.

- Sous l'exogénéité stricte, $E[\Delta X_2 \Delta \varepsilon_2] = 0$ et $E[X_2 \Delta \varepsilon_2] = 0$.
- Donc si l'on régresse ΔY_2 sur ΔX_2 et X_2, les coefficients de X_2 devraient être non significatifs.
- Mais sous la seule hypothèse d'exogénéité faible, $E[X_2 \Delta \varepsilon_2] = -E[X_2 \varepsilon_1] \neq 0$ en général.

$\Rightarrow$ X_2 endogène dans (4) et $\widehat{\gamma}_{MCO}$ ne convergera pas vers 0.

- Il suffit donc de tester $\gamma_0 = 0$ dans (4).
- Idem si $T > 2$: on inclut $X_2, ..., X_T$ (ou un sous-ensemble de ces régresseurs) dans l'équation en différences premières et on teste leur significativité.

Approche par les GMM

- Sous l'hypothèse d'exogénéité stricte, on a :

$$E\left[X_{t'}\left(\Delta Y_t - \Delta X_t'\beta_0\right)\right] = E\left[X_{t'}\Delta\varepsilon_t\right] = 0$$

  $\forall\,(t,t') \in \{2,...,T\} \times \{1,...,T\}$.
- Cette écriture permet (i) de construire l'estimateur optimal de β_0 sous la condition d'exogénéité stricte et (ii) de tester le modèle, puisqu'on a $(K-1)(T-1)T > (K-1)$ conditions de moment.
- Cette approche est peu utilisée en pratique, car l'estimateur en deux étapes peut être instable quand le nombre de conditions de moments est grand p/r aux nombres de paramètres, comme ici dès que $T \geq 5$.

Plan

Introduction

Exogénéité des résidus mais autocorrélation

$E(X_{it}\alpha_i) \neq 0$ mais $E(X_{it}\varepsilon_{it'}) = 0$, $\forall (t, t')$ (exogénéité stricte)

$E(X_{it}\alpha_i) \neq 0$ mais $E(X_{it}\varepsilon_{it'}) = 0$, $\forall t' \geq t$ (exogénéité faible)

Exemple : effet de la police sur la criminalité

Définition et exemples

- On suppose maintenant seulement l'exogénéité faible :

$$E(X_{it}\varepsilon_{it'}) = 0 \quad \forall t' \geq t.$$

- Exemple 1 : modèles dynamiques :

$$Y_{it} = W_{it}'\gamma_0 + Y_{it-1}\rho_0 + \alpha_i + \varepsilon_{it},$$

avec $X_{it} = (W_{it}', Y_{it-1})'$. Dans ce modèle, par construction l'exogénéité stricte n'est pas vérifiée car $E(X_{it}\varepsilon_{it-1}) \neq 0$.

- Mais l'exogénéité faible l'est si :
  1. les $(W_{it})_t$ sont faiblement exogènes ;
  2. les $(\varepsilon_{it})_t$ sont non corrélés entre eux, et non corrélés à (α_i, Y_{i0}).
- Dans ce cas en effet, on a, pour $t \geq 2$,

$$Y_{it-1} = \left(\sum_{j=0}^{t-2} \rho_0^j W_{it-1-j}'\right)\gamma + Y_{i0}\rho_0^{t-1} + \alpha_i\left(\frac{1-\rho_0^{t-1}}{1-\rho_0}\right) + \sum_{j=0}^{t-2}\rho_0^j \varepsilon_{it-1-j}.$$

- Par conséquent,

$$E\left[Y_{it-1}\varepsilon_{it'}\right] = 0 \quad \forall t' \geq t.$$

Définition et exemples

- Le modèle précédent permet de distinguer l'hétérogénéité inobservée (α_i) de la dépendance d'état (effet de Y_{it-1}).
- Exemple 2 : modèle statique avec effets de retour. $X_{it} = (V_{it}, W'_{it})'$ avec :

$$\begin{array}{rcl} Y_{it} & = & W'_{it}\gamma_0 + V_{it}\delta_0 + \alpha_i + \varepsilon_{it} \\ V_{it} & = & W'_{it}\lambda_0 + Y_{it-1}\rho_0 + \zeta_i + \eta_{it}. \end{array}$$

- V_{it} peut par exemple représenter le taux de SIDA dans une ville et Y_{it} les ventes individuelles de préservatifs . Les ventes passées ont a priori un effet sur le taux de SIDA en t.
- Dans ce modèle, l'exogénéité stricte n'est pas satisfaite sauf si $\rho_0 = 0$ et W_{it} est strictement exogène. Mais l'exogénéité faible l'est toujours sous les mêmes conditions que précédemment.

La faillite des within et différences premières

- L'estimateur des différences premières vérifie :

$$\text{plim}\,\widehat{\beta}_{FD} = \beta_0 + \left[\frac{1}{T-1}\sum_{t=2}^{T} E\left(\Delta X_t \Delta X_t'\right)\right]^{-1} \left[\frac{1}{T-1}\sum_{t=2}^{T} E\left(\Delta X_t \Delta \varepsilon_t\right)\right].$$

- On a par ailleurs :

$$E\left(\Delta X_t \Delta \varepsilon_t\right) = E(X_t\varepsilon_t) - E(X_t\varepsilon_{t-1}) - E(X_{t-1}\varepsilon_t) + E(X_{t-1}\varepsilon_{t-1}) = -E(X_t\varepsilon_{t-1}).$$

- Si (X_t, ε_t) est stationnaire (dans ce cas $E\left(\Delta X_t \Delta X_t'\right)$ et $E\left(\Delta X_t \Delta \varepsilon_t\right)$ ne dépendent pas de t), on a alors

$$\text{plim}\,\widehat{\beta}_{FD} = \beta_0 - E\left(\Delta X_t \Delta X_t'\right)^{-1} E\left(X_t \varepsilon_{t-1}\right) \neq \beta_0 \text{ en général.}$$

- N.B. : le biais est indépendant de T.

La faillite des within et différences premières

- L'estimateur within vérifie quant à lui :

$$\text{plim}\widehat{\beta}_W = \beta_0 + \left[\frac{1}{T}\sum_{t=1}^{T} E\left(\tilde{X}_t\tilde{X}_t'\right)\right]^{-1} \left[\frac{1}{T}\sum_{t=1}^{T} E\left(\tilde{X}_t\varepsilon_t\right)\right].$$

- On a par ailleurs :

$$E\left(\tilde{X}_t\varepsilon_t\right) = E[(X_t - \overline{X})\varepsilon_t] = -E[\overline{X}\varepsilon_t].$$

- Donc, si (X_t, ε_t) est stationnaire,

$$\text{plim}\widehat{\beta}_W = \beta_0 - E\left(\tilde{X}_t\tilde{X}_t'\right)^{-1} E[\overline{X}\overline{\varepsilon}] \neq \beta_0 \text{ en général.}$$

- Cette fois le biais dépend de T. Par Cauchy-Schwartz, $|E[\overline{X}\overline{\varepsilon}]| = |E[(\overline{X} - E(\overline{X}))\,\overline{\varepsilon}]| \leq \left(V(\overline{X})V(\overline{\varepsilon})\right)^{1/2} = O(1/T)$.
- Dans ce cas, le biais tend vers 0 quand $T \to \infty$. Mais il peut être grand à distance finie.

- Principe : par exogénéité faible on a $E(X_s \Delta\varepsilon_t) = 0$ pour $s < t$ donc on peut utiliser les conditions de moments :

$$E\left[X_s\left(\Delta Y_t - \Delta X_t'\beta_0\right)\right] = 0, \quad \forall s < t. \tag{5}$$

- Exemple simple : pour $T = 2$ on instrumente, dans l'équation de différences premières, ΔX_2 par X_1.
- N.B. : pour le modèle dynamique, il faut au moins observer Y_t sur 3 périodes : on estime l'effet de $Y_2 - Y_1$ sur $Y_3 - Y_2$, en utilisant Y_1 comme instrument de $Y_2 - Y_1$.
- Pour $T > 2$, on peut utiliser des GMM basés sur (5). Comme précédemment l'estimateur des GMM basé sur l'ensemble des conditions de moment peut être difficile à calculer et n'est pas toujours très fiable à distance finie.
- En pratique, on utilise souvent certaines conditions de moment seulement, pour se ramener à des 2MC sur les données empilées.

Choix des instruments

- Première possibilité : utiliser X_{t-1} comme instrument de ΔX_t.
- Deuxième possibilité : utiliser ΔX_{t-1} comme instrument de ΔX_t (N.B. : on perd alors une date !).
- Pour choisir, on peut s'appuyer sur la qualité de la régression de 1ère étape et la variance estimée des 2MC.

Plan

Introduction

Exogénéité des résidus mais autocorrélation

$E(X_{it}\alpha_i) \neq 0$ mais $E(X_{it}\varepsilon_{it'}) = 0, \forall (t, t')$ (exogénéité stricte)

$E(X_{it}\alpha_i) \neq 0$ mais $E(X_{it}\varepsilon_{it'}) = 0, \forall t' \geq t$ (exogénéité faible)

Exemple : effet de la police sur la criminalité

Problématique

- Une augmentation des forces de police permet-elle de réduire la criminalité ?
- On se focalise ici sur les crimes violents. On considère la régression :

$$\log(crime_{it}) = \log(police_{it})\beta_1 + unem_{it}\beta_2 + incpc_{it}\beta_3 + black_{it}\beta_4 + \theta_t + \alpha_i + \varepsilon_{it}.$$

  où $crime_{it}=$ nb de crimes violents pour 10^5 hab. dans l'état i à la date t, $police_{it}=$nb. de policiers pour 10^5 hab., $unem_{it}=$ taux de chômage, $incpc_{it}=$ revenu moyen, $black_{it}=$proportion de noirs.
- Problème potentiel : un niveau de criminalité élevé en t peut amener les autorités à augmenter le nombre de policiers après t $\Rightarrow$ corrélation positive entre ε_{it} et $police_{it+1}$.
- Modèle estimé sur un sous-échantillon des données de Levitt (1996), PRISON.DTA, utilisé par Wooldridge : 51 états américains entre 1980 et 1993.

Code Stata

```
gen log_police = log(polpc)

* Command xtset: for panel data, to indicate that state=i, year=t
xtset state year
foreach x of varlist lcriv log_police unem incpc black y81-y93{
    gen d_`x'= d.`x'
}

* Pooled OLS without clusters
regress lcriv log_police unem incpc black y81-y93

* Pooled OLS with clusters
regress lcriv log_police unem incpc black y81-y93, cluster(state)

* Within estimator
* Without the cluster option
xtreg lcriv log_police unem incpc black y81-y93, fe
* With the cluster option
xtreg lcriv log_police unem incpc black y81-y93, fe cluster(state)

* First difference estimator
reg d_lcriv d_log_police d_unem d_incpc d_black d_y82-d_y93, cluster(state)
```

Résultats : MCO empilés

Source	SS	df	MS
Model	173.192926	17	10.1878192
Residual	132.384693	696	.190207892
Total	305.577619	713	.428580111

Number of obs	=	714
F(17, 696)	=	53.56
Prob > F	=	0.0000
R-squared	=	0.5668
Adj R-squared	=	0.5562
Root MSE	=	.43613

lcriv	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
log_police	1.095963	.101124	10.84	0.000	.8974189	1.294508
unem	7.100307	.9282458	7.65	0.000	5.277809	8.922804
incpc	.0000402	8.68e-06	4.63	0.000	.0000232	.0000573
black	1.716159	.1661151	10.33	0.000	1.390013	2.042306

(Std. Err. adjusted for 51 clusters in state)

lcriv	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
log_police	1.095963	.3639663	3.01	0.004	.3649157	1.827011
unem	7.100307	2.761588	2.57	0.013	1.553494	12.64712
incpc	.0000402	.0000257	1.57	0.124	-.0000114	.0000918
black	1.716159	.6635824	2.59	0.013	.383315	3.049004

- Différence entre les deux ?
- Quid du signe des coefficients ? Intérêt des méthodes à effets fixes p/r aux MCO ?

Résultats : estimateur within

```
Fixed-effects (within) regression               Number of obs      =       714
Group variable: state                           Number of groups   =        51

R-sq:  within  = 0.4676                         Obs per group: min =        14
       between = 0.0031                                        avg =      14.0
       overall = 0.0253                                        max =        14

                                                F(17,646)          =     33.38
corr(u_i, Xb)  = -0.0540                        Prob > F           =    0.0000

------------------------------------------------------------------------------
       lcriv |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  log_police |   .3695031   .0720416     5.13   0.000      .228039    .5109671
        unem |  -1.548982   .4138484    -3.74   0.000    -2.361632   -.7363314
       incpc |   9.75e-07   5.63e-06     0.17   0.863    -.0000101     .000012
       black |  -.6217821    1.26768    -0.49   0.624    -3.111052    1.867488
```

```
Fixed-effects (within) regression               Number of obs      =       714
Group variable: state                           Number of groups   =        51

R-sq:  within  = 0.4676                         Obs per group: min =        14
       between = 0.0031                                        avg =      14.0
       overall = 0.0253                                        max =        14

                                                F(17,50)           =     36.23
corr(u_i, Xb)  = -0.0540                        Prob > F           =    0.0000

                                  (Std. Err. adjusted for 51 clusters in state)
------------------------------------------------------------------------------
             |               Robust
       lcriv |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  log_police |   .3695031   .1567384     2.36   0.022     .0546847    .6843214
        unem |  -1.548982   .6720916    -2.30   0.025    -2.898918   -.1990462
       incpc |   9.75e-07   .0000115     0.08   0.933    -.0000222    .0000241
       black |  -.6217821   1.835126    -0.34   0.736    -4.307742    3.064178
```

Résultats : différences premières

(Std. Err. adjusted for 51 clusters in state)

d_lcriv	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
d_log_police	.0542456	.0538304	1.01	0.318	-.0538758	.1623671
d_unem	-.0163343	.3722453	-0.04	0.965	-.7640111	.7313424
d_incpc	.0000319	.0000115	2.79	0.007	8.92e-06	.0000549
d_black	-1.743021	2.704599	-0.64	0.522	-7.175368	3.689325

- Les valeurs estimées sont assez différents du within $\Rightarrow$ exogénéité stricte suspecte.
- On considère alors la régression :

$$\Delta \log(crime_{it}) = \Delta \log(police_{it})\beta_1 + \log(police_{it})\gamma + controls + \Delta \varepsilon_{it}.$$

(Std. Err. adjusted for 51 clusters in state)

d_lcriv	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
d_log_police	.0726276	.0511293	1.42	0.162	-.0300685	.1753238
log_police	-.0255225	.0145134	-1.76	0.085	-.0546736	.0036286
d_unem	.0883762	.3927475	0.23	0.823	-.7004804	.8772328
d_incpc	.0000382	.000013	2.93	0.005	.000012	.0000644
d_black	-3.763961	2.337207	-1.61	0.114	-8.45838	.9304588

- Conclusion ?

Estimation avec exogénéité faible

- Instruments $= \log(police_{it-1})$ ou $\Delta \log(police_{it-1})$.

```
by state: gen lag_log_police = log_police[_n-1]
gen d_lag_log_police = d.lag_log_police

ivreg d lcriv d unem d incpc d black (d log police = lag log police ), first robust
ivreg d lcriv d unem d incpc d black (d log police = d lag log police), first robust
```

```
First-stage regressions

      Source |       SS       df       MS              Number of obs =     663
-------------+------------------------------           F(  4,   658) =    1.25
       Model |  .014473114     4  .003618278           Prob > F      =  0.2895
    Residual |  1.90863889   658  .002900667           R-squared     =  0.0075
-------------+------------------------------           Adj R-squared =  0.0015
       Total |    1.923112   662  .002905003           Root MSE      =  .05386

------------------------------------------------------------------------------
  d_log_police |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------+----------------------------------------------------------------
        d unem |   -.233953   .1839071    -1.27   0.204    -.5950686    .1271626
       d incpc |   1.21e-06   6.47e-06     0.19   0.852    -.0000115    .0000139
       d_black |  -3.516948   2.484921    -1.42   0.157    -8.396278    1.362381
lag log police |  -.0127569   .0097783    -1.30   0.192    -.0319574    .0064436
         _cons |   .0786864   .0532305     1.48   0.140    -.0258358    .1832086
--------------------------------------------------------------------------------

Instrumental variables (2SLS) regression               Number of obs =     663
                                                       F(  4,   658) =    2.89
                                                       Prob > F      =  0.0217
                                                       R-squared     =       .
                                                       Root MSE      =   .1137

------------------------------------------------------------------------------
             |               Robust
     d_lcriv |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
d_log_police |   1.476505   1.723324     0.86   0.392    -1.907372    4.860382
      d unem |   .2405441   .5399006     0.45   0.656    -.8195917     1.30068
     d incpc |   .0000481   .0000154     3.12   0.002     .0000178    .0000783
     d black |   1.994339   5.937232     0.34   0.737    -9.663866    13.65254
       _cons |  -.0351123   .0241679    -1.45   0.147    -.0825678    .0123431
------------------------------------------------------------------------------
Instrumented:  d log police
Instruments:   d_unem d_incpc d_black lag_log_police
```

Estimation avec exogénéité faible

First-stage regressions

Source	SS	df	MS
Model	.02569767	4	.006424418
Residual	1.83956165	607	.003030579
Total	1.86525932	611	.003052798

Number of obs =	612
F(4, 607) =	2.12
Prob > F =	0.0769
R-squared =	0.0138
Adj R-squared =	0.0073
Root MSE =	.05505

d_log_police	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
d_unem	-.1549976	.1923305	-0.81	0.421	-.5327116	.2227164
d_incpc	2.12e-06	6.63e-06	0.32	0.749	-.0000109	.0000152
d_black	-3.365325	2.517615	-1.34	0.182	-8.309618	1.578967
d_lag_log_police	-.1136732	.0466006	-2.44	0.015	-.2051911	-.0221552
_cons	.0094642	.0057333	1.65	0.099	-.0017954	.0207238

Instrumental variables (2SLS) regression

Number of obs =	612
F(4, 607) =	5.38
Prob > F =	0.0003
R-squared =	.
Root MSE =	.08771

d_lcriv	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
d_log_police	-.3505592	.6220623	-0.56	0.573	-1.572215	.8710963
d_unem	-.1251975	.3083298	-0.41	0.685	-.7307201	.4803252
d_incpc	.0000485	.000011	4.41	0.000	.0000269	.0000701
d_black	-2.543389	4.749559	-0.54	0.593	-11.87095	6.784174
_cons	-.0162598	.0114898	-1.42	0.158	-.0388243	.0063047

Instrumented: d_log_police
Instruments: d_unem d_incpc d_black d_lag_log_police

L'essentiel

- Estimation de variance des MCO avec *clustering*.
- Exogénéité stricte.
- Estimateur des différences premières et estimateur within.
- Test de l'exogénéité stricte.
- Exogénéité faible.
- Estimateur des variables instrumentales sous l'exogénéité faible.