Econométrie des données de panel: Rappels

Guillaume Horny*

*Banque de France

Master 2 MASERATI

Rappels

Plan

Plan

Les sommes

Soit $\{x_i : i = 1, \ldots, n\}$ une suite de n nombres. Leur somme s'écrit :

$$\sum_{i=1}^{n} x_i \equiv x_1 + x_2 + \ldots + x_n,$$

La somme a les propriété suivantes :

- Pour toute constante c, on a :

$$\sum_{i=1}^{n} c = nc,$$

$$\sum_{i=1}^{n} cx_i = c \sum_{i=1}^{n} x_i$$

- Pour a et b constants et $\{(x_i, y_i) : i = 1, \ldots, n\}$

$$\sum_{i=1}^{n} (ax_i + by_i) = a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} y_i$$

La moyenne

Leur moyenne s'écrit :

$$x_. \equiv \frac{1}{n}\sum_{i=1}^{n} x_i$$

Propriétés :

- la somme des écarts à la moyenne est nulle :

$$\sum_i (x_i - x_.) = \sum_i x_i - \sum_i x_. = \sum_i x_i - nx_. = nx_. - nx_. = 0.$$

- on peut aussi montrer :

$$\sum_i (x_i - x_.)^2 = \sum_i x_i^2 - nx_.^2$$

- ou encore :

$$\sum_i (x_i - x_.)(y_i - y_.) = \sum_i x_i y_i - nx_. y_.$$

Plan

Notations

- On indice par i les individus ($i = 1, \ldots, N$) et t les périodes ($t = 1, \ldots, T$). On a donc NxT observations
- On note y_{it} la variable dépendante (ici un scalaire).
- On note x_{it} un vecteur de $Kx1$ variables, dont le k-ème élément est noté x_{it}^k.

Notations

- La moyenne **inter-temporelle** de la $k$ème variable pour l'individu i est :

$$x_{i.}^k = \frac{1}{T} \sum_{t=1}^{T} x_{it}^k$$

- La moyenne **inter-individuelle** de la $k$ème variable à la date t est :

$$x_{.t}^k = \frac{1}{N} \sum_{i=1}^{N} x_{it}^k$$

- La moyenne **inter-individuelle** et **inter-temporelle** de la $k$ème variable est :

$$x_{..}^k = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}^k$$

Décompositions de la variance

Savoir qu'elle est l'origine de la variance est souvent informatif.
On dispose par exemple de données de salaire pour un groupe de personnes suivies plusieurs années. La dispersion des salaires est-elle due à des différences permanentes entre individus (qualification, grilles de salaires) ou bien à des fluctuations au fil du temps (par exemple liées au cycle économique) ?

La variabilité des salaires est :

$$\text{varb } y = \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - y_{..})^2$$

On peut écrire :

$$\text{varb } y = \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - y_{i.} + y_{i.} - y_{..})^2$$

Décompositions de la variance

$$\text{varb}\, y = \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-y_{i.})^2+\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i.}-y_{..})^2+2\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-y_{i.})(y_{i.}-y_{..})$$

Le dernier terme est nul car $\sum_t (y_{it} - y_{i.}) = Ty_{i.} - Ty_{i.} = 0$. D'où :

$$\text{varb } y = \text{varb}_{\text{intra-individuelle}}\ y + \text{varb}_{\text{inter-individuelle}}\ y$$

Dans l'exemple des salaires, cette formule nous permet de décomposer la variabilité totale en :

- variabilité temporelle propre à l'individu (part variable, mobilité...)
- variabilité permanente entre individus (formation initiale, talent...)

Décompositions de la variance

D'autres décompositions sont possibles

- On a :

$$\text{varb } y = \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - y_{.t})^2 + \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{.t} - y_{..})^2$$

  D'où :

$$\text{varb } y = \text{varb}_{\text{intra-temporelle}}\ y + \text{varb}_{\text{inter-temporelle}}\ y$$

- Ou encore des décompositions en variabilité intra-individuelle-temporelle, inter-individuelle et inter-temporelle (exercice !)

Plan

Espérance

Soit X une variable discrète prenant les valeurs $\{x_1, \ldots, x_k\}$. Sa densité de probabilité est $f_X()$. Son espérance est :

$$\mathrm{E}(X) = x_1 f_X(x_1) + x_2 f_X(x_2) + \ldots + x_k f_X(x_k) \equiv \sum_{j=1}^{K} x_j f_X(x_j).$$

Lorsque X est continue :

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty} x f_X(x) dx.$$

Pour toute fonction $g(.)$ de X, on a :

$$\mathrm{E}[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx.$$

Densité conditionnelle

On s'intéresse généralement aux relations entre une variable aléatoire Y et une ou plusieurs autres. Supposons qu'il n'y ait qu'une seule autre variable aléatoire appelée X. L'influence de X sur Y est visible dans la distribution conditionnelle de Y sachant X.

La densité de probabilité de Y conditionnellement à X est :

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \forall x \text{ tel que } f_X(x) > 0.$$

Lorsque X et Y sont discrètes :

$$f_{Y|X}(y|x) = \mathrm{P}(Y = y|X = x).$$

Espérance conditionnelle

Supposons que l'on connaisse $f_{Y/X}(.)$ et la valeur x prise par X. On peut alors calculer $\mathrm{E}(Y|X=x)$, souvent notée $\mathrm{E}(Y|X)$. Généralement, lorsque x change, $\mathrm{E}(Y|X)$ change également.

Pour Y discrète, on a :

$$\mathrm{E}(Y|X) = \sum_{j=1}^{K} y_j f_{Y|X}(y_j|x).$$

Il s'agit toujours d'une moyenne pondérée des y_j, mais les poids dépendent ici de la valeur prise par X.

Pour Y continue :

$$\mathrm{E}(Y|X) = \int_{-\infty}^{+\infty} y f_{Y|X}(y|x) dy.$$

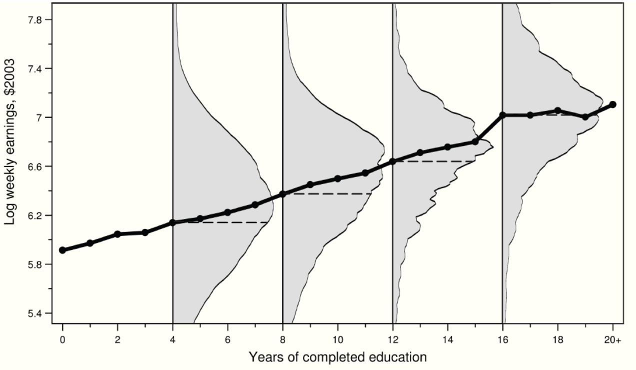
Espérance conditionnelle : Exemple

Supposons que (X, Y) représente l'ensemble des salariés, avec X une mesure des années d'étude et Y le niveau de salaire. Le salaire moyen des personnes ayant atteint le bac (étudié 12 ans) est alors mesuré par $E(Y|X = 12)$. On peut faire ce calcul pour différentes valeurs de X et voir ainsi comment le salaire est relié au niveau d'études.

On peut faire le calcul pour toutes les durées d'éducation et reporter les résultats dans un tableau.

Espérance conditionnelle : Exemple

On peut parfois supposer que l'espérance conditionnelle prend la forme d'une fonction standard.



Source : Angrist and Pischke (2009) : *Mostly Harmless Econometrics*, Princeton University Press.

Espérance conditionnelle : Exemple

Supposons que l'espérance conditionnelle est linéaire :

$$\mathrm{E}(salaire|etudes) = \beta_0 + \beta_1 etudes.$$

Si cette hypothèse est raisonnable, alors chaque année supplémentaire d'étude rapporte en moyenne β_1. Le salaire moyen après 17 ans d'études est $\beta_0 + 17\beta_1$.

Propriétés des espérances conditionnelles

- $\mathrm{E}(Y|X) = \mathrm{E}(Y)$ ssi X et Y sont indépendants
- **Loi des espérances itérées :** $\mathrm{E}_X[\mathrm{E}_{Y|X}(Y|X)] = \mathrm{E}_Y(Y)$.

Preuve de la loi des espérances itérées

$$
\begin{aligned}
\mathrm{E}_X[\mathrm{E}_{Y|X}(Y|X)] &= \int_{-\infty}^{+\infty} \mathrm{E}_{Y|X}(Y|X=u) f_X(u)du. \\
&= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} t f_Y(t|X=u)dt \right] f_X(u)du \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} t f_Y(t|X=u) f_X(u)dudt \\
&= \int_{-\infty}^{+\infty} t \left[\int_{-\infty}^{+\infty} f_Y(t|X=u) f_X(u)du \right] dt \\
&= \int_{-\infty}^{+\infty} t \left[\int_{-\infty}^{+\infty} f_{XY}(u,t)du \right] dt \\
&= \int_{-\infty}^{+\infty} t f_Y(t)dt \\
&= \mathrm{E}(Y).
\end{aligned}
$$

Illustration de la loi des espérances itérées

Soit $Y = salaire$ et $X = etudes$. Supposons que :

$$\mathrm{E}(salaire|etudes) = 1130 + 50etudes.$$

La loi des espérances itérées implique :

$$\begin{aligned}\mathrm{E}(salaire) &= \mathrm{E}_{etudes}(1130 + 50etudes) \\ &= 1130 + 50\mathrm{E}_{etudes}(etudes).\end{aligned}$$

On a de plus $\mathrm{E}(etudes) = 12$. Alors :

$$\mathrm{E}(salaire) = 1130 + 50 * 12 = 1730.$$

Note : $\mathrm{E}(salaire|etudes) = g(etudes)$. Elle dépend de *etude* et pas de *salaire* !

La décomposition en espérance conditionnelle

Grâce à loi des espérances itérées, on peut décomposer n'importe quelle variable aléatoire de la manière suivante :

$$Y = \mathrm{E}(Y|X) + \epsilon,$$

où :

1. $\mathrm{E}(\epsilon|X) = 0$,
2. ϵ est sans corrélation avec toute fonction de X.

Preuve de la décomposition en espérance conditionnelle

1. **Preuve de $\mathbf{E}(\epsilon|X) = 0$**

$$E(\epsilon|X) = E\left[Y - E(Y|X)|X\right] = E(Y|X) - E(Y|X) = 0.$$

2. **Preuve de l'absence de corrélation**
Pour toute fonction $h(.)$ quelconque, on a :

$$\begin{aligned} E(h(X)\epsilon) &= E\left[E(h(X)\epsilon|X)\right] \qquad \text{(loi des espérances itérées)} \\ &= E\left[h(X)E(\epsilon|X)\right] \\ &= 0 \qquad (\text{car } E(\epsilon|X) = 0) \end{aligned}$$

Propriétés des espérances conditionnelles (suite)

Théorème des projection successives :

$$\mathrm{E}(Y|X) = \mathrm{E}_{Z|X}[\mathrm{E}_{Y|X,Z}(Y|X,Z)|X].$$

On peut donc obtenir $\mathrm{E}(Y|X)$ en deux étapes :

1. Trouver l'expression de $\mathrm{E}_{Y|X,Z}(Y|X,Z)$, où Z est n'importe quelle variable aléatoire
2. Prendre son espérance conditionnelle à X.

Ce qui est important ici est que $X \subset (X,Z)$. Intuitivement, la première espérance dépend à la fois de X et de Z, elle est plus "détaillée" que la seconde qui ne retient que l'influence de X. En prenant la seconde espérance, on prend la moyenne par rapport à Z, ce qui permet de l'évacuer du calcul.

Plan

Modèle

Partant de la décomposition d'une variable en son espérance conditionnelle ($Y = \mathrm{E}(Y|X) + \epsilon$), le modèle linéaire revient à considérer que l'espérance conditionnelle est une fonction linéaire de X.

Le modèle linéaire s'écrit pour l'individu $i(i = 1, \ldots, N)$:

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_K x_{iK} + \epsilon_{it}.$$

En empilant les N observations, on obtient son écriture matricielle :

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

où Y est un vecteur ($Nx1$), X une matrice $Nx(K+1)$, β un vecteur de dimension $((K+1)x1)$ et ϵ un vecteur ($Nx1$).

Estimateur OLS

Comment trouver les valeurs de β permettant de minimiser $\sum_i (y_i - x_i\beta)^2$?

La condition de premier ordre de la minimisation par rapport à β de $\sum_i (y_i - x_i\beta)^2$ s'écrit sous forme matricielle :

$$X'(Y - X\widehat{\beta}) = 0.$$

Elle a pour solution :

$$\widehat{\beta} = (X'X)^{-1}X'Y.$$

Note : c'est uniquement lorsque X est une matrice carrée qu'on peut écrire

$$\widehat{\beta} = (X'X)^{-1}X'Y = X^{-1}(X')^{-1}X'Y = X^{-1}Y.$$

Estimateur OLS

$$
\begin{aligned}
\widehat{\beta} &= (X'X)^{-1}X'Y \\
&= \begin{bmatrix} \sum_{i=1}^{N}(x_i^1)^2 & \sum_{i=1}^{N} x_i^1 x_i^2 & \cdots & \sum_{i=1}^{N} x_i^1 x_i^K \\ \sum_{i=1}^{N} x_i^2 x_i^1 & \sum_{i=1}^{N}(x_i^2)^2 & & \vdots \\ \vdots & & & \\ \sum_{i=1}^{N} x_i^K x_i^1 & & \cdots & \sum_{i=1}^{N}(x_i^K)^2 \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{N} x_i^1 y_i \\ \sum_{i=1}^{N} x_i^2 y_i \\ \vdots \\ \sum_{i=1}^{N} x_i^K y_i \end{bmatrix}
\end{aligned}
$$

Propriété : absence de corrélation après estimation entre les X et les résidus

Par la décomposition de n'importe quelle variable aléatoire en $Y = \mathrm{E}(Y|X) + \epsilon$, on sait déjà que les résidus sont sans corrélation avec les X.

On peut le vérifier également avec les conditions de premier ordre pour $\beta = \widehat{\beta}$, qui s'écrivent :

$$X'(Y - X\widehat{\beta}) = 0.$$
$$\Leftrightarrow X'\widehat{\epsilon} = 0.$$

Après estimation, la covariance entre n'importe quelle variable explicative et les résidus OLS est toujours nulle, par construction.

Plan

Propriétés à **distance finie** d'un estimateur

- un estimateur est **sans biais** (*unbiased*) ssi :

$$\mathsf{E}(\widehat{\beta}) = \beta.$$

  Si on évalue $\widehat{\beta}$ avec différents échantillons, chacun comprenant un nombre fini d'observation, la moyenne des estimations correspondra à la vraie valeur du paramètre.

- un estimateur sans biais est **efficace** (*efficient*) s'il est de variance minimale.

Illustration de l'absence de biais

On demande au hasard à des gens dans la rue le prix d'une voiture que vous leur montrez. Le prix de la voiture est X, et on note X_n la moyenne des prix donnés par les différentes personnes. Cette moyenne se rapprochera de X au fur et à mesure que vous interrogez de plus en plus de personnes

Propriétés **asymptotiques** : convergence en probabilité

Convergence en probabilités : Une suite $\{X_n\}$ de variables aléatoires converge en probabilité vers une constante (non-aléatoire) X si, pour tout $\epsilon > 0$:

$$\lim_{n\to\infty} \Pr(|X_n - X| \geq \epsilon) = 0.$$

La constante X est appelée la **probabilité limite** de X_n, que l'on note également :

$$p \lim_{n\to\infty} X_n = X.$$

L'intuition est que la probabilité d'un événement inhabituel se rapproche de 0 au fur et à mesure que l'on augmente le nombre de tirages.

Illustration de la convergence en probabilité

Exemples : Une personne s'entraine avec un ballon de basket à marquer des paniers à 3 points. On note X_n son score au n-ème tire. Au fil des entraînements, son adresse augmente et il en marque de plus en plus souvent. La probabilité qu'il ne marque pas 3 points diminue au fil du temps, et X_n converge en probabilité vers $X = 3$ (après des années de pratique !).

Note : on retrouve la convergence en probabilité dans des contextes qui ne sont pas nécessairement reliés à la loi des grands nombres.

Propriétés **asymptotiques** d'un estimateur

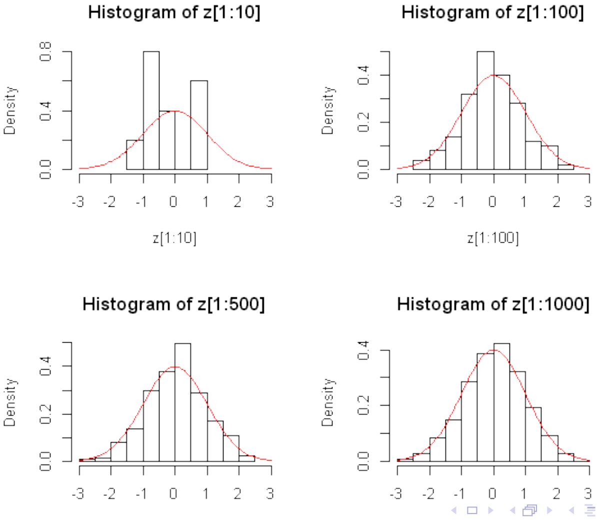
- un estimateur est **convergent** (*consistent*) ssi :

  $$p \lim_{n\to\infty} (\widehat{\beta} - \beta) = 0,$$

  où $p \lim$ est la convergence en probabilités
- Il existe différente notions de convergence. On se concentre ici sur la convergence en probabilité, car lorsqu'un estimateur converge en probabilité vers la vraie valeur, il est **asymptotiquement sans biais**
- un estimateur asymptotiquement sans biais est **efficace** (*efficient*) s'il est de variance minimale.

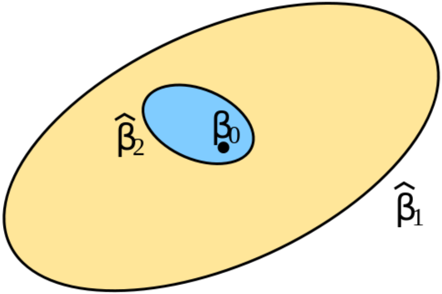
Illustration de la convergence en loi

La convergence en probabilités entraine la convergence en loi



Pourquoi distinguer propriétés asymptotiques et à distance finie ?

Certains estimateurs sont biaisés mais convergent, avec un biais négligeable en présence de grands échantillons. Lorsqu'ils sont (beaucoup) plus précis que des estimateurs sans biais, ils méritent qu'on s'y intéresse.



Propriété : Absence de biais de l'estimateur OLS

$$
\begin{aligned}
\widehat{\beta} &= (X'X)^{-1}X'Y \\
&= (X'X)^{-1}X'(X\beta + \epsilon) \\
&= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\epsilon \\
&= \beta + (X'X)^{-1}X'\epsilon.
\end{aligned}
$$

Si on suppose $\mathrm{E}(\epsilon|X) = 0$, alors :

$$
\begin{aligned}
\mathrm{E}(\widehat{\beta}|X) &= \beta + (X'X)^{-1}X'\mathrm{E}(\epsilon|X) \\
&= \beta.
\end{aligned}
$$

Variance de l'estimateur OLS (1/2)

Sous l'hypothèse d'**homoscédasticité** ($\text{var}(\epsilon) = \sigma^2 Id$) :

$$\begin{aligned}
\text{var}(\widehat{\beta}|X) &= \text{var}[(X'X)^{-1}X'Y] \\
&= \text{var}[\beta + (X'X)^{-1}X'\widehat{\epsilon}] \\
&= \text{var}[(X'X)^{-1}X'\widehat{\epsilon}] \\
&= (X'X)^{-1}X'\text{var}(\widehat{\epsilon})X(X'X)^{-1} \\
&= (X'X)^{-1}X'(\widehat{\sigma}^2 Id)X(X'X)^{-1} \\
&= \widehat{\sigma}^2(X'X)^{-1}X'X(X'X)^{-1} \\
&= \widehat{\sigma}^2(X'X)^{-1}.
\end{aligned}$$

Un estimateur sans biais de la variance est :

$$\widehat{\sigma}^2 = \frac{\widehat{\epsilon}'\widehat{\epsilon}}{N - K - 1}.$$

Variance de l'estimateur OLS (2/2)

Sous l'hypothèse d'**hétéroscédasticité** ($var(\epsilon_i) = \sigma_i^2 \neq \sigma_j^2$ pour $i \neq j$), $\sigma^2 Id$ n'est plus un estimateur convergent de $var(\epsilon)$.

White (1980) montre qu'un estimateur convergent de $var(\epsilon)$ est :

$$\frac{1}{N-K-1}\sum_i \epsilon_i^2 x_i x_i',$$

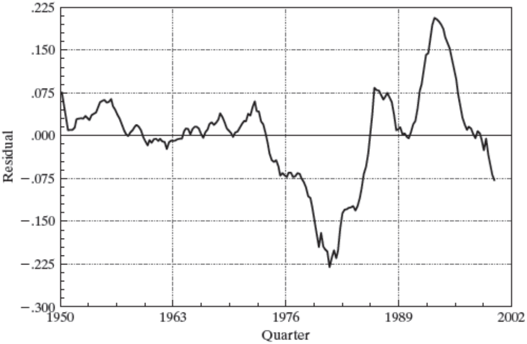
où x_i est $(Kx1)$. D'où :

$$\begin{aligned} var(\widehat{\beta}|X) &= (X'X)^{-1}X' var(\widehat{\epsilon})X(X'X)^{-1} \\ &= (X'X)^{-1}\left(\frac{N}{N-K-1}\sum_i \epsilon_i^2 x_i x_i'\right)(X'X)^{-1} \end{aligned}$$

On l'appelle l'estimateur **robuste**.

Qu'est ce que l'autocorrélation des erreurs ?

On a une autocorrélation des erreurs lorsqu'elles tendent à être corrélées entre elles dans le temps.
Exemple d'autocorrélation positive des résidus :



Source : Greene (2003), Econometric analysis, Prentice Hall.

Théorème de Gauss-Markov

Sous les hypothèses $E(\epsilon|X) = 0$ et d'homoscédasticité, on a :

Théorème : L'estimateur OLS est l'estimateur linéaire sans biais de plus petite variance.

Propriété : Normalité asymptotique

Sous l'hypothèse :

$$\epsilon \sim N(0, \sigma^2).$$

On a :

$$\widehat{\beta}|X \sim N(\beta, \sigma^2(X'X)^{-1}).$$

Sous l'hypothèse de modèle linéaire à erreurs gaussiennes homoscédastiques, on peut montrer que l'estimateur OLS est un estimateur du maximum de vraisemblance.

Annexes

Illustration de la convergence en loi

Plus un échantillon est grand, plus les observations tendent à se conformer à la distribution sous-jacente

```
R> z <- rnorm(1000)
R> par(mfrow = c(2, 2))
R> hist(z[1:10], freq = FALSE, xlim = c(-3, 3))
R> curve(dnorm, add = TRUE, col = 2)
R> hist(z[1:100], freq = FALSE, xlim = c(-3, 3))
R> curve(dnorm, add = TRUE, col = 2)
R> hist(z[1:500], freq = FALSE, xlim = c(-3, 3))
R> curve(dnorm, add = TRUE, col = 2)
R> hist(z[1:1000], freq = FALSE, xlim = c(-3, 3))
R> curve(dnorm, add = TRUE, col = 2)
```