THÉORIE ÉCONOMIQUE,
MODÉLISATION ET APPLICATIONS

Principe des MCO et Données de Panel

Aristide E. Houndetoungan

14 Septembre 2022

Notations

- b un scalaire ;
- $\mathbf{b}$ un vecteur
- $\mathbf{B}$ une matrix
- Si X_i est une variable aléatoire, x_i est utilisé pour désigner une réalisation de X_i.
- $\mathbb{E}$, $\mathbb{V}\mathbf{ar}$ et $\mathbb{C}\mathbf{ov}$ sont respectivement l'espérance mathématique, la variance et la covariance.
- $\mathbb{P}$ désigne une mesure de probabilité.

Quelques Rappels

- Si X est une variable discrète, $\mathbb{E}\left(X\right) = \sum_{k \in \Omega} k \times \mathbb{P}(X = k)$, où Ω est le support de X.
- Si X est une variable continue, $\mathbb{E}\left(X\right) = \int_{\Omega} x f(x) dx$, où f est la fonction de densité de X.
- $\mathbb{V}\mathbf{ar}\left(X\right) = \mathbb{E}\left((X - \mathbb{E}\left(X\right))^2\right)$ ou encore $\mathbb{V}\mathbf{ar}\left(X\right) = \mathbb{E}\left(X^2\right) - \left(\mathbb{E}(X)\right)^2$.
- $\mathbb{C}\mathbf{ov}\left(X, Y\right) = \mathbb{E}\left((X - \mathbb{E}\left(X\right))(Y - \mathbb{E}\left(Y\right))\right)$ ou encore $\mathbb{C}\mathbf{ov}\left(X, Y\right) = \mathbb{E}\left(XY\right) - \mathbb{E}\left(X\right)\mathbb{E}\left(Y\right)$.
- En échantillon fini (sur des données en pratique) de taille n, où les réalisations de X sont $x_i, \ldots, x_n$, l'espérance de X est approximée par,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- **Convergence en probabilité**

  Soit $\hat{\theta}_n$ un estimateur de θ calculé à partir d'un échantillon de taille n. On dit que $\hat{\theta}_n$ converge en probabilité vers θ (ou simplement $\hat{\theta}_n$ est convergent), et on note $\hat{\theta}_n \xrightarrow{p} \theta$ ou encore $\operatorname{plim} \hat{\theta}_n = \theta$, si pour tout $\mu > 0$, $\lim_{n\to\infty} \mathbb{P}\left(||\hat{\theta}_n - \theta|| > \mu\right) = 0$.

- **Loi faible des grands nombres**

  Si $x_1, \ldots, x_n$ sont indépendants et identiquement distribués (i.i.d), alors la Loi des Grands Nombres (LGN) stipule que $\operatorname{plim} \bar{X} = \mathbb{E}(X)$.

- **Loi des espérances itérées**

  Soit $\mathbf{X}$ et $\mathbf{Y}$ deux variables aléatoires. La loi des espérances itérées implique :

$$\mathbb{E}(\mathbf{X}) = \mathbb{E}_Y\left[\mathbb{E}(\mathbf{X}|\mathbf{Y})\right]. \tag{1}$$

  Analogiquement,

$$\mathbb{E}(\mathbf{XY}) = \mathbb{E}_X\left[\mathbb{E}(\mathbf{XY}|\mathbf{X})\right] = \mathbb{E}_X\left[\mathbf{X}\,\mathbb{E}(\mathbf{Y}|\mathbf{X})\right]. \tag{2}$$

Modèle linéaire en moyennes

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \varepsilon_i \tag{3}$$

- y_i est la variable dépendante ou encore variable expliquée.
- x_{i2}, ..., x_{iK} sont les variables indépendantes ou variables explicatives.
- ε_i est le terme d'erreur.
- i référence un individu. n individus donc i peut prendre les valeurs de 1 à n.
- β_k $(k \geq 2)$ est la pente de y_i par rapport à x_{ik}
- $\dfrac{\partial y}{\partial x_k} = \beta_k$ $(k \geq 2)$: La hausse d'une unité de x_k implique une hausse de β_k unité de y, toute chose étant égale par ailleurs.
- β_1 est l'intercept. Si $\mathbb{E}(\varepsilon_i) = 0$, alors $\mathbb{E}(y_i | x_{i2} = 0, \ldots, x_{iK} = 0) = \beta_1$.

Principe des MCO

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \varepsilon_i \tag{3}$$

- En général, l'objectif d'un économètre est d'estimer de façon "valide" les paramètres β_1, ..., β_K.
- Un des estimateurs les plus utilisés est l'estimateur des Moindres Carrés Ordinaires (MCO).
- Son principe est de minimiser la somme des carrées des résidus (et non des erreurs).
- Trouver β_1, ..., β_K pour que,

$$\sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_{i2} - \cdots - \beta_K x_{iK})^2$$

ait la plus petite valeur possible.

Ecriture du modèle sous forme matricielle

- Modèle

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{12} & x_{13} & \dots & x_{1K} \\ 1 & x_{22} & x_{23} & \dots & x_{2K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & x_{n3} & \dots & x_{nK} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{4}$$

$$\underset{(n,1)}{\mathbf{y}} = \underset{(n,K)}{\mathbf{X}} \underset{(K,1)}{\boldsymbol{\beta}} + \underset{(n,1)}{\boldsymbol{\varepsilon}} \tag{5}$$

Pour chaque individu i, on peut aussi écrire,

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i, \tag{6}$$

où $\mathbf{x}_i' = (1,\ x_{i2},\ \dots,\ x_{iK})$.

- Somme des carrés des résidus sous forme matricielle

$$SCR = \boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}$$

$$SCR = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \tag{7}$$

Hypothèses du principe des MCO

- **Hypothèses**

H.1. Linéarité : $y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$.

H.2. Indépendance linéaire des variables explicatives : $\mathbf{X}$ est une matrice de plein rang ; c'est-à-dire, $\mathrm{rang}(\mathbf{X}) = K$.

H.3. Exogénéité des variables explicatives : $\mathbb{E}(\varepsilon_i|\mathbf{x}_i) = 0$.

H.4. Homoscédasticité et non autocorrélation des erreurs : pour tout i, $\mathbb{V}\mathbf{ar}(\varepsilon_i|\mathbf{x}_i) = \sigma^2$. De plus, pour $i \neq j$, $\mathbb{E}(\varepsilon_i\varepsilon_j|\mathbf{x}_i, \mathbf{x}_j) = 0$.

H.5. Normalité des erreurs : chaque ε_i suit une distribution normale.

H.6. $(\varepsilon_1, \mathbf{x}_1)$, ..., $(\varepsilon_n, \mathbf{x}_n)$ sont i.i.d.

- **Implication des hypothèses**

- H.1. et H.3. impliquent que, $\mathbb{E}(y_i|\mathbf{x}_i) = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} = \mathbf{x}_i'\boldsymbol{\beta}$.
- Par la loi des espérances itérées, H.3. implique : $\mathbb{E}(\varepsilon_i) = \mathbb{E}_x\left[\mathbb{E}(\varepsilon_i|\mathbf{x}_i)\right] = 0$.
- Par la loi des espérances itérées, H.3. implique aussi :
  $\mathbb{E}(\mathbf{x}_i\varepsilon_i) = \mathbb{E}_x\left[\mathbf{x}_i\,\mathbb{E}(\varepsilon_i|\mathbf{x}_i)\right] = 0$. C'est pourquoi H.3. est aussi appelée condition d'orthogonalité ($\varepsilon_i \perp \mathbf{x}_i$).
- Aussi, $\mathbb{C}\mathbf{ov}\left(\mathbf{x}_i, \varepsilon_i\right) = \mathbb{E}(\mathbf{x}_i\varepsilon_i) - \mathbb{E}(\mathbf{x}_i)\,\mathbb{E}(\varepsilon_i) = 0$.
- Par la Loi des Grands Nombres (LGN), H.6. implique :
  $\operatorname{plim}\left(\dfrac{\mathbf{X}'\mathbf{X}}{n}\right) = \mathbb{E}\left(\mathbf{x}_i\mathbf{x}_i'\right) = \mathbf{Q_{xx}}$, une matrice finie définie positive.

Estimateur des MCO

- Estimateur de $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}}_{\text{MCO}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} \tag{8}$$

- Variable dépendante prédite :

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{MCO}} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} = \mathbf{P}_{\mathbf{x}}\mathbf{y} \tag{9}$$

  avec $\mathbf{P}_{\mathbf{x}} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$ est la matrice de projection dans l'espace de $\mathbf{X}$.

- Résidus :

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{MCO}} = (\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}')\mathbf{y} = \mathbf{M}_{\mathbf{x}}\mathbf{y}, \tag{10}$$

  avec $\mathbf{M}_{\mathbf{x}} = \mathbf{I} - \mathbf{P}_{\mathbf{x}}$ est la matrice de projection dans l'espace orthogonal à celui de $\mathbf{X}$.

- Estimateur de σ^2 :

$$\hat{\sigma}^2_{\text{MCO}} = \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n-K} \tag{11}$$

Propriétés de l'estimateur des MCO

- **Définition** : Un estimateur linéaire est un estimateur qui s'écrit sous la forme $\mathbf{Ay}$, avec $\mathbf{A}$ une matrice.
- $\hat{\boldsymbol{\beta}}_{\text{MCO}} = \underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'}_{\mathbf{A}}\mathbf{y}$ est linéaire.

- **Propriétés en échantillon fini** :

1. Sous les hypothèses H.1. à H.3., $\hat{\boldsymbol{\beta}}_{\text{MCO}}$ est sans biais : $\mathbb{E}\left(\hat{\boldsymbol{\beta}}_{\text{MCO}}\right) = \boldsymbol{\beta}$.
2. Sous les hypothèses H.1. à H.4., $\hat{\sigma}^2_{\text{MCO}}$ est sans biais : $\mathbb{E}\left(\hat{\sigma}^2_{\text{MCO}}\right) = \sigma^2$.
3. Sous les hypothèses H.1. à H.4., $\mathbb{V}\mathbf{ar}\left(\hat{\boldsymbol{\beta}}_{\text{MCO}}|\mathbf{X}\right) = \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}$.
4. Sous les hypothèses H.1. à H.4., $\hat{\boldsymbol{\beta}}_{\text{MCO}}$ est le meilleur estimateur linéaire. Il est sans biais et a la plus petite variance parmi tous les estimateurs linéaires sans biais (BLUE : Best Linear Unbiased Estimator).

5. Sous les hypothèses H.1. à H.5., $\hat{\boldsymbol{\beta}}_{\text{MCO}}|\mathbf{X} \sim \mathcal{N}\left(\boldsymbol{\beta}, \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)$.

- **Propriétés en grand échantillon** :

6. Sous les hypothèses H.1. à H.3. et H.6., $\hat{\boldsymbol{\beta}}_{\text{MCO}}$ est convergent : on note aussi $\operatorname{plim} \hat{\boldsymbol{\beta}}_{\text{MCO}} = \boldsymbol{\beta}$.
7. Sous les hypothèses H.1. à H.4. et H.6., $\operatorname{plim} \hat{\sigma}^2_{\text{MCO}} = \sigma^2$.
8. Sous les hypothèses H.1. à H.4. et H.6., $\hat{\boldsymbol{\beta}}_{\text{MCO}} \overset{a}{\sim} \mathcal{N}\left(\boldsymbol{\beta}, \frac{\sigma^2}{n}\mathbf{Q}_{\mathbf{xx}}^{-1}\right)$. En pratique la variance asymptotique de $\hat{\boldsymbol{\beta}}_{\text{MCO}}$ est estimée par,

$$\text{Est.Asy. } \mathbb{V}\mathbf{ar}\left(\hat{\boldsymbol{\beta}}_{\text{MCO}}\right) = \hat{\sigma}^2_{\text{MCO}} \left(\mathbf{X}'\mathbf{X}\right)^{-1}.$$

- Application avec R : script `convergence.mco.R`

Données de Panel (DP)/Données Longitudinales

- Données/Informations observées dans une population de référence (de plus de 2 unités) sur plusieurs (au moins 2) périodes.
- Exemples :
  - PIB trimestriel des pays de la Zone Euro de 2010-2020 ;
  - Taux d'admission aux examens de Baccalauréat dans les lycées français durant les 10 dernières années ;
  - (Données discrètes) Participation des pays de la Zone Euro aux 10 dernières coupes mondiales ;
  - (Ne constitue pas des DP) Nombre journalier de vaccins ≪ Pfizer-BioNTech COVID-19 ≫ administrés en France en 2021.

Données de Panel (DP)/Données Longitudinales

- Panel équilibré : même période d'observation pour toutes les unités.
- Exemple :

ID	Année	Age	Femme	Revenu
1	2015	26	0	1500
1	2016	27	0	1600
1	2017	28	0	1500
1	2018	29	0	1650
2	2015	42	1	1900
2	2016	43	1	2000
2	2017	44	1	2100
2	2018	45	1	2700
3	2015	29	0	2000
3	2016	30	0	2000
3	2017	31	0	2800
3	2018	32	0	2900

- Panel non équilibré : périodes d'observation non identiques.
- Exemple :

ID	Année	Age	Femme	Revenu
1	2016	27	0	1600
1	2017	28	0	1500
2	2015	42	1	1900
2	2016	43	1	2000
2	2017	44	1	2100
3	2015	34	0	3300
4	2015	29	0	2000
4	2016	30	0	2000
4	2017	31	0	2800
4	2018	32	0	2900

Estimateur des MCO sur données de panel

- Pour une même unité, chaque nouvelle période est considérée comme étant une nouvelle unité (la dimension panel est ignorée).
- Modèle pooled :

$$y_{it} = \beta_1 + \beta_2 \mathbf{x}'_{it2} + \cdots + \beta_K \mathbf{x}'_{itK} + \varepsilon_{it}. \tag{12}$$

- Le modèle pooled est aussi estimé par la méthode des MCO.
- Sous les hypothèses standard de l'estimateur des MCO, l'estimateur du modèle pooled a également des mêmes propriétés que l'estimateur MCO.
- Application avec R : script `pooled.R`