Les Moindres Carrés ordinaires (suite)

Stéphane Adjemian

Université Maine, GAINS & CEPREMAP

11 mars 2008

– Soit le DGP suivant :

$$y_i = a + bx_i + cz_i + \epsilon_i$$

avec a, b et c trois paramètres réels, ϵ_i une variable aléatoire identiquement et indépendamment distribuée vérifiant :

(i) $\mathbb{E}\left[\epsilon_i|\mathbf{x}\right] = \mathbb{E}\left[\epsilon_i|\mathbf{z}\right] = 0$, et

(ii) $\mathbb{E}\left[\epsilon_i^2|\mathbf{x}\right] = \mathbb{E}\left[\epsilon_i^2|\mathbf{z}\right] = \sigma^2$ pour tout i.

– Notons que les propriétés de la perturbation ϵ_i sont définies conditionellement aux variables explicatives $\mathbf{x}$ et $\mathbf{z}$.

– On suppose aussi que $\mathbb{C}ov\left(z_i, x_i\right) = \gamma\mathbb{V}\left[x_i\right]$ pour tout $i = 1, \dots, N$.

- La dernière hypothèse nous dit que les deux variables explicatives sont corrélées.

- Nous estimons, par les MCO, le modèle suivant :

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

  Ainsi la seconde variable z_i est omise du modèle estimé, par exemple parce que celle-ci n'est pas observée, en ce sens le modèle estimé est mal spécifié.

- Pouvons-nous estimer sans biais l'effet marginal de x sur y (c'est-à-dire le paramètre b) en excluant la variable z de la régression ?

– L'esimateur des moindre carrés ordinaires est obtenu en minimisant la somme des carrés des résidus :

$$(\hat{\alpha}, \hat{\beta}) = \arg\min_{\alpha,\beta} \sum_{i=1}^{N} (y_i - \alpha - \beta x_i)^2$$

– L'estimateur de β est :

$$\hat{\beta} = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}$$

– Au dénominateur de l'estimateur de la pente, $\hat{\beta}$, nous retrouvons un estimateur (à N^{-1} près) de la variance de la variable explicative, au numérateur (toujours à N^{-1} près) un estimateur de la covariance entre la variable expliquée et la variable explicative.

– Afin d'étudier les propriétés de cet estimateur, on remplace le DGP dans $(y_i - \bar{y})$. Il vient :

$$\begin{aligned}\hat{\beta} &= \frac{\sum_{i=1}^{N}(x_i - \bar{x})((x_i - \bar{x})b + (z_i - \bar{z})c + \epsilon_i - \bar{\epsilon})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \\ &= b + \frac{\sum_{i=1}^{N}(x_i - \bar{x})(\epsilon_i - \bar{\epsilon})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} + \frac{\sum_{i=1}^{N}(x_i - \bar{x})(z_i - \bar{z})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}\end{aligned}$$

– Nous avons déjà vu que la première fraction est nulle en espérance. Nous avons donc que $\mathbb{E}\left[\hat{\beta}\right] = b$ si est seulement si la seconde fraction est nulle en espérance.

– Or il y a peu de chances pour que cette espérance soit nulle…

- … En effet :
  - Le dénominateur du second terme est, à un facteur N^{-1} près, un estimateur de la variance de x.
  - Le numérateur du second terme est, toujours à un facteur N^{-1} près, un estimateur de la covariance de x et z.

- Ainsi dès lors que la covariance entre les deux variables explicatives est non nulle nous avons $\mathbb{E}\left[\hat{\beta}\right] \neq b$.

- Plus précisément, d'après le DGP nous avons :

$$\hat{\beta} \underset{N\to\infty}{\overset{\text{proba}}{\longrightarrow}} b + \frac{\mathbb{C}ov(x,z)}{\mathbb{V}\left[x\right]} = b + \gamma$$

- $\hat{\beta}$ converge en probabilité vers le vrai effet marginal de x sur y si et seulement si $\gamma = 0$.
- L'omission d'une variable ne pose pas de problème tant que la variable omise n'est pas corrélée avec la(es) variable(s) incluse(s) dans la régression.
- Si cette condition n'est pas vérifiée alors :
  - On peut dans certains cas "corriger" du biais sur la base de considérations théoriques (par exemple si on connaît la valeur de γ alors un estimateur asymptotiquement sans biais est $\hat{\beta} - \gamma$).
  - Il ne faut pas utiliser les MCO !

– On considère le modèle estimé suivant :

$$y_i = x_{1,i}\beta_1 + x_{2,i}\beta_2 + \cdots + x_{K,i}\beta_K + \varepsilon_i$$

– Le DGP a la même forme…
– Pour estimer $\beta \equiv (\beta_1, \ldots, \beta_K)'$ nous allons réécrire le modèle sous une forme matricielle.
– On pose :

$$\mathbf{x}_i \equiv (x_{1,i}, \ldots, x_{K,i})$$

$$X \equiv (\mathbf{x}'_1, \ldots, \mathbf{x}'_K)$$

$$Y \equiv (y_1, \ldots, y_N)'$$

$$\varepsilon \equiv (\varepsilon_1, \ldots, \varepsilon_N)'$$

- On peut alors écrire le modèle linéaire sous la forme :

$$Y = X\beta + \varepsilon$$

où Y est un vecteur $N \times 1$, X une matrice $N \times K$, β un vecteur $K \times 1$ et ε un vecteur $N \times 1$.

- L'estimateur des MCO du vecteur β est défini par :

$$\hat{\beta} = \arg\min_{\beta}(Y - X\beta)'(Y - X\beta)$$

- En effet, le produit scalaire $\varepsilon'\varepsilon = \sum_{i=1}^{N} \varepsilon_i^2$ est la somme des carrés des résidus.

– On obtient la CNO en annulant la dérivée de l'objectif par rapport à β :

$$-2(Y - X\hat{\beta})'X = 0$$

– En développant, il vient :

$$Y'X = \hat{\beta}'X'X$$

ou de façon équivalente (en transposant et en notant que $X'X$ est une matrice symétrique) :

$$X'X\hat{\beta} = X'Y$$

– En supposant que la matrice X'X est de plein rang (inversible), nous obtenons finalement :

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

– Ce résultat généralise celui que nous avons obtenu dans le cas du modèle avec une seule variable explicative.

– On note que l'estimateur des MCO est égale à l'inverse d'un estimateur (à un facteur N^{-1} près) de la matrice de variance-covariance des variables explicatives post-multiplié par un estimateur (toujours à un facteur N^{-1} près) de la covariance entre les variables explicatives et la variable expliquée.

– On peut montrer que (sous certaines conditions d'orthogonalité et de rang) que $\hat{\beta}$ est un estimateur sans biais et qu'il converge en probabilité vers $\mathbb{V}[\mathbf{x}]^{-1}\mathbb{C}ov(\mathbf{x},\mathbf{y})$.