ÉCONOMÉTRIE
(UGA, S2)
CHAPITRE 3 :
ENDOGÉNÉITÉ ET VARIABLES INSTRUMENTALES(1)

Michal W. Urdanivia*

*UGA, Faculté d'Économie, GAEL,
e-mail : michal.wong-urdanivia@univ-grenoble-alpes.fr

24 avril 2022

1. Variables explicatives endogènes et biais d’endogénéité

Endogénéité

- Nous continuons à étudier un modèle linéaire de forme :

$$y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i,\ \mathrm{E}[u_i] \equiv 0.$$

où $\dim(\mathbf{x}_i) = K$, et on souhaite estimer $\mathbf{a}_0$.

- Dans le chapitre précédent il a été montré qu'une condition essentielle pour l'identification de $\mathbf{a}_0$ et son estimation convergente par les MCO est que $\mathbf{x}_i$ soit exogène :

$$\underbrace{\mathrm{E}[u_i|\mathbf{x}_i] = 0}_{\text{exogénéité forte}} \Rightarrow \underbrace{\mathrm{E}[\mathbf{x}_i u_i] = \mathrm{Cov}[\mathbf{x}_i; u_i] = 0}_{\text{exogénéité faible}}.$$

- Inversement quand $\mathrm{E}[\mathbf{x}_i u_i] \neq 0$ on parle d'endogénéité des variables explicatives.
- Il convient néanmoins de noter $\mathrm{E}[\mathbf{x}_i u_i] \neq 0$ peut être du à seulement certains éléments de $\mathbf{x}_i$, par exemple :
  - ⋆ On peut avoir $M \in \{0, \ldots, K\}$ éléments de $\mathbf{x}_i$ tels que $\mathrm{E}[x_{k,i} u_i] = 0,\ k = 1, \ldots, M.$, et ces variables seront qualifiées d'exogènes.
  - ⋆ Alors que les $K - M$ autres éléments de $\mathbf{x}_i$ sont tels que $\mathrm{E}[x_{k,i} u_i] \neq 0,\ k = M + 1, \ldots, K$, et ces variables seront qualifiées d'endogènes.

Non convergence de l'estimateur des MC en présence d'endogénéité

- Les résultats du chapitre précédent nous permettent d'écrire que pour l'estimateur des MCO de $\mathbf{a}_0$ que :

$$\underset{N\to+\infty}{\text{plim}}\ \hat{\mathbf{a}}_N^{MCO} = \begin{cases} \mathbf{a}_0 + \underbrace{\mathrm{E}[\mathbf{x}_i\mathbf{x}_i]^{-1}\,\mathrm{E}[\mathbf{x}_i u_i]}_{=0} = \mathbf{a}_0 & \text{quand } \mathbf{x}_i \text{ est exogène,} \\ \mathbf{a}_0 + \underbrace{\mathrm{E}[\mathbf{x}_i\mathbf{x}_i]^{-1}\,\mathrm{E}[\mathbf{x}_i u_i]}_{\neq 0} \neq \mathbf{a}_0 & \text{quand } \mathbf{x}_i \text{ est endogène,} \end{cases}$$

  où le terme $\mathrm{E}[\mathbf{x}_i\mathbf{x}_i]^{-1}\,\mathrm{E}[\mathbf{x}_i u_i]$ est le biais (asymptotique) d'endogénéité de l'estimateur des MCO de $\mathbf{a}_0$.
- Poursuivant ce qui a été remarqué plus haut quant au fait d'avoir $\mathrm{E}[\mathbf{x}_i u_i] \neq 0$, il est important de remarquer que même si une seule variable explicative est endogène, disons le K-ième élément de $\mathbf{x}_i$, $x_{K,i}$, avec donc :

$$\mathrm{E}[x_{K,i}u_i] \neq 0, \quad \text{et} \quad \mathrm{E}[x_{k,i}u_i] \neq 0, \quad \text{pour } k = 1, \ldots, K-1,$$

  alors :
  - ⋆ Le coefficient estimé de $x_{K,i}$, $\hat{a}_{K,N}^{MC0}$, n'est pas le seul élément de $\hat{\mathbf{a}}_N^{MCO}$ potentiellement biaisé.
  - ⋆ En général le problème d'endogénéité de $x_{K,i}$ « contamine » les les coefficients estimés des autres éléments de $\mathbf{x}_i$, $\hat{a}_{k,N}^{MC0}$ pour $k = 1, \ldots, K-1$.
  - ⋆ Ceci donne une idée de l'importance d'un problème d'endogénéité posé par une ou plusieurs variables explicatives.

Non convergence de l’estimateur des MC en présence d’endogénéité

Remarque 1

- ⋆ On peut préciser le dernier point précédent. Pour cela réécrivons le modèle linéaire :

$$y_i = \mathbf{x}'_{-K,i}\mathbf{a}_{-K,0} + a_{K,0}x_{K,i} + u_i, \quad \text{avec} \quad \mathrm{E}[\mathbf{x}_{-K,i}u_i] = \mathrm{E}[u_i] \equiv 0,$$

  où $\mathbf{x}_{-K,i}$ est le vecteur des $K-1$ variables explicatives autres que $x_{K,i}$, et $\mathbf{a}_{-K,0}$ leurs coefficients respectifs.

- ⋆ En utilisant le propriétés des projections linéaires(c.f., chapitre 2) on obtient l’équation suivante, $\hat{\mathbf{a}}_N^{MCO} \equiv (\hat{\mathbf{a}}_{-K,N}^{MCO}, \hat{a}_{K,N}^{MC0})$, de $\mathbf{a} \equiv (\mathbf{a}_{-K,0}, a_{K,0})$ dû à l’endogénéité de $x_{K,i}$:

$$\operatorname{plim} \hat{\mathbf{a}}_N^{MC0} = \mathbf{a}_0 + \underbrace{\mathrm{Cov}[x_{K,i}; u_i] \times \mathrm{V}[e_{K,i}] \times \begin{bmatrix} -\gamma_K \\ 1 \end{bmatrix}}_{\text{Biais as. de } \hat{\mathbf{a}}_N^{MC0}},$$

  où $e_{K,i} \equiv x_{K,i} - \mathrm{EL}[x_{K,i}|\mathbf{x}_{-K,i}]$ est le résidu de la projection linéaire de $x_{K,i}$ sur $\mathbf{x}_{-K,i}$: $\mathbf{x}'_{-K,i}\gamma_K \equiv \mathrm{EL}[x_{K,i}|\mathbf{x}_{-K,i}]$.

- ⋆ Sont donc biaisés asymptotiquement : $\hat{a}_{K,N}^{MC0}$, les éléments de $\hat{\mathbf{a}}_{-K,N}^{MCO}$ correspondant aux éléments de $\mathbf{x}_{-K,i}$ liés à $x_{K,i}$(au sens où leurs coefficients dans γ_K sont non nuls), et la constante du modèle.

2. Sources de l’endogénéité

Pourquoi peut t-il y avoir un problème d'endogénéité ?

- On présente par des exemples, les cas dans lesquels on peut avoir $\mathrm{E}[\mathbf{x}_i u_i] = 0$.
- On parle ici de trois formes/sources d'endogénéité :
  1. Variables explicatives pertinentes omises.
  2. Simultanéité.
  3. Erreurs de mesure sur les variables explicatives.
- Dans tous ces cas, qui ne sont pas exclusifs les uns par rapport aux autres, les MCO ne sont pas convergents.

Variables explicatives pertinentes omises

- On s'appuie ici sur les travaux de Angrist and Krueger (1991), et Card (1993) qui s'intéressent à la mesure de l'effet causal du niveau d'études sur le salaire.
- Pour cela ils considèrent une équation linéaire du type :

$$lwage_i = \alpha_0 + b_{1,0} educ_i + \mathbf{c}_i' \mathbf{b}_{-1,0} + u_i, \quad \text{avec} \quad \mathrm{E}[\mathbf{c}_i u_i] = \mathrm{E}[u_i] \equiv 0, \tag{1}$$

  où pour un individu i dans la population de référence :
  - ⋆ $lwage_i$ est son salaire annuel en logarithme.
  - ⋆ $educ_i$ le nombre d'années études réalisés. depuis l'entrée dans le primaire.
  - ⋆ $\mathbf{c}_i$ est un vecteur d'autres caractéristiques observées de l'individu et corrélées avec son salaire. Par exemple son lieu de résidence, son expérience professionnelle en nombre d'années, son origine ethnique, etc. Ces caractéristiques sont supposées exogènes.
- Un problème courant dans cette littérature concerne l'*ability bias*.

Variables explicatives pertinentes omises

- Citons Card (1993) :

> There are a variety of reasons why schooling may be correlated with the unobserved component of earnings. One that has received considerable attention in the literature is "ability bias" (see e.g. Griliches (1977)). Suppose that some individuals have an unobserved characteristic ("ability") that enables them to earn higher wages at any level of education. If these individuals acquire higher-than-average schooling then the OLS estimate of β will be upward-biased. The fact that individuals with higher test scores (on IQ or achievement tests) tend to have higher earnings and more schooling is often interpreted as evidence of ability bias.

- Où le terme β correspond au coefficient de $educ_i$ que nous avons noté $\mathbf{b}_{1,0}$ dans (1).
- Notons la caractéristique inobservée que l'auteur appelle *ability*(que l'on peut essayer de traduire par capacité ou compétence) $abil_i$.

Variables explicatives pertinentes omises

- Suivant l'auteur on s'attend à ce que $abil_i$:
  - ⋆ ait un effet positif sur $lwage_i$,
  - ⋆ soit corrélée positivement avec $educ_i$.
- On peut résumer cela par deux équations.
  - ⋆ La première inclut l'effet de la caractéristique inobservé $abil_i$ dans l'erreur du modèle (1) :

$$lwage_i = \alpha_0 + b_{1,0} educ_i + \mathbf{c}_i' \mathbf{b}_{-1,0} + \underbrace{\delta_0 abil_i + v_i}_{u_i}, \quad \text{avec} \quad \mathrm{E}[v_i | educ_i, \mathbf{c}_i, abil_i] = \mathrm{E}[v_i] \equiv 0, \tag{2}$$

  - ⋆ La seconde prend en compte la corrélation entre $abil_i$ et $educ_i$ en considérant l'équation de la projection linéaire de $educ_i$ sur $abil_i$ et une constante.

$$educ_i = \gamma_0 + \gamma_1 abil_i + r_i, \quad \text{avec} \quad \mathrm{E}[r_i abil_i] = 0. \tag{3}$$

    où r_i est le résidu de cette projection, et on sait que $\gamma_1 = \frac{\mathrm{Cov}[educ_i; abil_i]}{\mathrm{V}[abil_i]}$.
- En résumé, si on observait $abil_i$ un estimateur convergent de $b_{1,0}$ serait l'estimateur des MCO de l'équation dans (2), mais comme $abil_i$ est inobservée on estime l'équation dans (1) dont l'erreur inclut $abil_i$.
- Mais dans ce cas nous avons,

$$\begin{aligned} \mathrm{Cov}[educ_i; u_i] &= \mathrm{Cov}[\gamma_0 + \gamma_1 abil_i + r_i; \delta_0 abil_i + v_i] \\ &= \gamma_1 \delta_0 \, \mathrm{V}[abil_i]. \end{aligned}$$

- Par conséquent, $educ_i$ est endogène dans le modèle (1) sauf si :
  - ⋆ $\gamma_1 = 0$ auquel cas $abil_i$ et $educ_i$ ne sont pas corrélés,
  - ⋆ ou $\delta_0 = 0$ auquel cas $abil_i$ n'est pas pertinente comme explicative de $lwage_i$.

Simultanéité

- On considère comme exemple le travail de Angrist and Evans (1998) sur l'effet de la maternité(nombre d'enfants) sur l'offre de travail(nombre d'heures travaillées).
- Il est vraisemblable que les choix de maternité et les choix professionnels soient décidés conjointement, et que par conséquent les deux propositions suivantes soient « vraies » :
  - ★ Le salaire est un déterminant du nombre d'enfants selon un double effet « ressources » et « temps disponible » pour les enfants.
  - ★ Le nombre d'enfants est un déterminant du salaire selon double effet « ressources » et « temps disponible » pour la carrière professionnelle.
- Cette simultanéité peut être analysée dans le cadre du système d'équations linéaire suivant :

$$\begin{cases} heures_i & = a_{1,0} enfants_i + \mathbf{x}'_{-1,i}\mathbf{a}_{-1,0} + u_i, \\ enfants_i & = \alpha_{1,0} heures_i + \mathbf{z}'_{-1,i}\boldsymbol{\alpha}_{-1,0} + v_i, \end{cases}$$

où $heures_i$ est le nombre d'heures travaillées par semaine, $enfants_i$ est le nombre d'enfants dans une famille, $\mathbf{x}_{-1,i}$ et $\mathbf{z}_{-1,i}$ sont des vecteur de variables explicatives exogènes(incluant chacun un régresseur constant) pour respectivement le nombre d'heures travaillées par semaine, et le nombre d'enfants.

Simultanéité

- En substituant l'expression pour les heures dans l'équation pour le nombre d'enfants, nous obtenons(en supposant que $1 - a_{1,0}\alpha_{1,0} \neq 0$),

$$enfants_i = \mathbf{x}'_{-1,i}\left(\frac{\mathbf{a}_{-1,0}\alpha_{1,0}}{1 - a_{1,0}\alpha_{1,0}}\right) + \mathbf{z}'_{-1,i}\left(\frac{\boldsymbol{\alpha}_{-1,0}}{1 - a_{1,0}\alpha_{1,0}}\right) + \left(\frac{\alpha_{1,0}}{1 - a_{1,0}\alpha_{1,0}}\right)u_i + \left(\frac{1}{1 - a_{1,0}\alpha_{1,0}}\right)v_i$$

  En supposant que $\mathbf{x}_{-1,i}$, $\mathbf{z}_{-1,i}$, v_i ne sont pas corrélés avec u_i, nous obtenons,

$$\mathrm{E}(u_i enfants_i) = \left(\frac{\alpha_{1,0}}{1 - a_{1,0}\alpha_{1,0}}\right)\mathrm{E}(u_i^2) \neq 0$$

Erreurs de mesure sur les variables explicatives

- Pour simplifier la présentation on considère le modèle simple suivant :

$$y_i = \alpha_0 + b_0\tilde{x}_i + u_i, \quad \text{avec} \quad \mathrm{E}[u_i|\tilde{x}_i] = \mathrm{E}[u_i] \equiv 0.$$

- Le problème est qu'à la place de $\tilde{x}_i$ on observe une mesure imparfaite de cette variable $\tilde{x}_i^e$.
- En effet $\tilde{x}_i^e$ est affectée d'erreurs de mesure :

$$\tilde{x}_i^e = \tilde{x}_i + \underbrace{e_i}_{\substack{\text{erreur de}\\ \text{mesure}\\ \text{sur } \tilde{x}_i}}, \quad \text{avec} \quad \mathrm{E}[e_i|\tilde{x}_i] = \mathrm{E}[u_i|e_i] = \mathrm{E}[e_i] = 0.$$

- L'erreur de mesure est donc a priori tout-à-fait anodine : elle est sans biais et n'est liée ni à $\tilde{x}$,ni à u_i.
- Par la suite on utilisera le fait que :

$$\mathrm{V}[\tilde{x}_i^e] = \mathrm{V}[\tilde{x}_i] + \mathrm{V}[e_i]. \tag{4}$$

- Par substitution de $\tilde{x}_i$ par $\tilde{x}_i^e - e_i$ dans le modèle d'intérêt on obtient le « modèle observable » :

$$y_i = \alpha_0 + b_0\tilde{x}_i^e + \underbrace{v_i}_{u_i - b_0 e_i}, \quad \text{avec} \quad \mathrm{E}[v_i] = 0.$$

Erreurs de mesure sur les variables explicatives

- La corrélation entre la variable explicative observée $\tilde{x}_i^e$, et l'erreur v_i est :

$$\mathrm{Cov}[\tilde{x}_i^e; v_i] = \mathrm{Cov}[\tilde{x}_i + e_i; u_i - b_0 e_i] = -b_0 \,\mathrm{V}[e_i] \neq 0,$$

  et ainsi la variable explicative mesurée avec erreur est endogène dans le modèle observable, par « construction ».

- En outre pour l'estimateur des MCO dans le modèle observable nous avons :

$$\operatorname*{plim}_{N\to+\infty} \hat{b}_N^{MCO} = \frac{\mathrm{Cov}[\tilde{x}_i^e; y_i]}{\mathrm{V}[\tilde{x}_i^e]}, \tag{5}$$

  et comme,

$$\mathrm{Cov}[\tilde{x}_i^e; y_i] = \mathrm{Cov}[\tilde{x}_i^e; b_0\tilde{x}_i^e + v_i] = b_0 \,\mathrm{V}[\tilde{x}_i^e] + \mathrm{Cov}[\tilde{x}_i^e; v_i], \tag{6}$$

  on obtient en utilisant (4)-(6) :

$$\operatorname*{plim}_{N\to+\infty} \hat{b}_N^{MCO} = b_0 \times \frac{\mathrm{V}[\tilde{x}_i]}{\mathrm{V}[\tilde{x}_i + \mathrm{V}[e_i]]} \Rightarrow \left| \operatorname*{plim}_{N\to+\infty} \hat{b}_N^{MCO} \right| < |b_0| \,. \tag{7}$$

- Le résultat (7) est un exemple de biais d'atténuation, à savoir que l'estimateur des MCO de b_0 dans le modèle observable sous-estime systématiquement b_0 en valeur absolue.
- Ce type de biais ne concerne pas seulement l'endogénéité en raison d'erreurs de mesure pouvant aussi concerner les autres cas d'endogénéité.

3. Introduction à la notion de variable instrumentale

Endogénéité et (non-)identification dans un modèle linéaire simple

- On considère ici le modèle linéaire le plus simple :

$$y_i = \alpha_0 + b_0 x_i + u_i \text{ avec } \mathrm{E}[u_i] \equiv 0, \tag{8}$$

mais on considère que l'analyse du PGD indique que x_i est endogène dans ce modèle, i.e. que :

$$\mathrm{E}[u_i|x_i] \neq 0 \Rightarrow \mathrm{Cov}[x_i; u_i] \neq 0.$$

- Rappelons que lorsque x_i est exogène et que $\mathrm{Var}[x_i] \neq 0$, on a par l'exogénéité de x_i :

$$\mathrm{Cov}[x_i; u_i] = 0 \Leftrightarrow \mathrm{Cov}[x_i; y_i - \alpha_0 - b_0 x_i] = 0 \Rightarrow b_0 = \frac{\mathrm{Cov}[x_i; y_i]}{\mathrm{Var}[x_i]}.$$

- Autrement dit l'exogénéité de x_i permet d'identifier b_0(et aussi de α_0) comme une fonction de la distribution des variables observées y_i et x_i.
- Inversement dans la situation que nous considérons dans ce chapitre $\mathrm{Cov}[x_i; u_i] \neq 0$, rend impossible l'identification b_0(et aussi celle de α_0), et la construction d'un estimateur convergent.

Endogénéité et (non-)identification dans un modèle linéaire simple

- On peut représenter ce problème en utilisant un **graphe causal** :

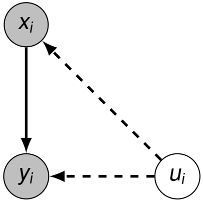

FIGURE 1 – Graphe causal du modèle : $y_i = \alpha_0 + b_0 x_i + u_i$, avec $\text{Cov}(x_i; u_i) \neq 0$. Les variables (x_i, y_i, u_i) sont les nœuds du graph et les nœuds foncés correspondent aux variables observées. Les arêtes représentent les relations entre les variables. Les relations observées sont en trait plein.

Identification avec une variable instrumentale

- L'intuition sous-jacente à la méthode des VIs consiste à répondre à la question de savoir si avec une variable, que nous notons z_i, il est possible d'obtenir une mesure de la relation causale entre x_i et y_i qui ne dépende pas de u_i.
- Autrement dit, z_i doit être exogène par rapport à u_i :

$$E[u_i|z_i] = 0 \Rightarrow \text{Cov}[z_i; u_i] = 0, \tag{9}$$

ce qui nous permets d'écrire :

$$\begin{aligned} \text{Cov}[z_i; u_i] = 0 &\Leftrightarrow \text{Cov}[z_i; y_i - \alpha_0 - b_0 x_i] = 0 \\ &\Leftrightarrow \text{Cov}[z_i; y_i - \alpha_0 + b_0 \text{Cov}[z_i; x_i] = 0 \end{aligned}$$

- Ceci indique que pour identifier b_0 on doit aussi supposer aussi que,

$$\text{Cov}[z_i; x_i] \neq 0, \tag{10}$$

et b_0 est identifié par :

$$b_0 = \frac{\text{Cov}[z_i; y_i]}{\text{Cov}[z_i; x_i]}, \tag{11}$$

Identification avec une variable instrumentale

- On peut résumer les conditions (9)-(10) ainsi :

Définition 1 (Conditions de validité de VIs dans un modèle simple)

Dans le modèle $y_i = \alpha + b_0 x_i + u_i$ avec $\mathrm{E}[u_i] \equiv 0$, une variable z_i est un instrument(de x_i) ssi :

(i) $\mathrm{Cov}[z_i; u_i] = 0$, i.e. z_i est exogène par rapport à u_i et :
(ii) $\mathrm{Cov}[z_i; x_i] \neq 0$, i.e. z_i et x_i sont liées.

- Dans la représentation en termes de graphe causal cela donne :

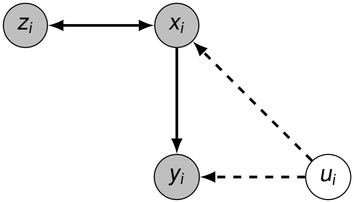

FIGURE 2 – Graphe causal du modèle : $y_i = \alpha_0 + b_0 x_i + u_i$, avec $\mathrm{Cov}(x_i; u_i) \neq 0$, $\mathrm{Cov}(z_i; u_i) = 0$. Les variables (z_i, x_i, y_i, u_i) sont les nœuds du graph et les nœuds foncés correspondent aux variables observées. Les arêtes représentent les relations entre les variables. Les relations observées sont en trait plein.

Estimateur des VIs dans le modèle simple

- L'identification de b_0 par (11) suggère l'estimateur :

$$\hat{b}_N^{VI} = \frac{N^{-1}\sum_{i=1}^{N}(z_i - \bar{z}_N)(y_i - \bar{y}_N)}{N^{-1}\sum_{i=1}^{N}(z_i - \bar{z}_N)(x_i - \bar{x}_N)} = \frac{N^{-1}\sum_{i=1}^{N}(z_{-}\bar{z}_N)y_i}{N^{-1}\sum_{i=1}^{N}(z_{-}\bar{z}_N)x_i}, \tag{12}$$

  où $\bar{z}_N$, $\bar{x}_N$, et $\bar{y}_N$ sont le moyennes empiriques respectives de z_i, x_i, et y_i.
- De plus, $\hat{b}_N^{VI}$ est convergent. Nous avons en effet :

$$\operatorname*{plim}_{N\to+\infty} N^{-1}\sum_{i=1}^{N}(z_{-}\bar{z}_N)y_i \to \mathrm{Cov}[z_i; y_i],$$
$$\operatorname*{plim}_{N\to+\infty} N^{-1}\sum_{i=1}^{N}(z_{-}\bar{z}_N)x_i \to \mathrm{Cov}[z_i; x_i],$$

  d'où :

$$\begin{aligned}\hat{b}_N^{VI} \underset{N\to+\infty}{\overset{p}{\to}} \frac{\mathrm{Cov}[z_i; y_i]}{\mathrm{Cov}[z_i; x_i]} &= \frac{\mathrm{Cov}[z_i; \alpha_0 + b_0 x_i + u_i]}{\mathrm{Cov}[z_i; x_i]},\\ &= b_0 + \frac{\mathrm{Cov}[z_i; u_i]}{\mathrm{Cov}[z_i; x_i]},\\ &= b_0.\end{aligned}$$

Estimateur des VIs dans le modèle simple

Remarque 2

- ⋆ On dit des variations de z_i qu'elles sont des variations exogènes : elles ne sont pas liées à u_i puisque $\text{Cov}[z_i; u_i] = 0$.
- ⋆ Ce sont les effets de ces variations exogènes sur x_i qui sont exploitées pour l'identification de b_0 grâce à $\text{Cov}[z_i; x_i] \neq 0$.
- ⋆ Noter qu'il n'est aucunement nécessaire que l'effet de z_i sur x_i soit causal.
- ⋆ L'effet de z_i sur y_i ne « transite » que via x_i. La variable instrumentale z_i n'est pas une variable explicative dans le modèle de y_i. On parle alors de relation d'exclusion (de la VI z_i vis-à-vis du modèle de y_i).
- ⋆ L'estimateur des VIs est parfois appelé estimateur des moindres carrés indirects. Cela provient de ce que b_0 dans (11) peut s'écrire :

$$b_0 = \frac{\text{Cov}[z_i; y_i] / \text{Var}[z_i]}{\text{Cov}[z_i; x_i] / \text{Var}[z_i]},$$

  qui est le rapport entre le coefficient de z_i dans la projection de y_i sur z_i, et le coefficient de z_i dans la projection de x_i sur z_i.

4. L'estimateur de VIs

Variables endogènes, exogènes, instruments

- L'objectif est maintenant de généraliser l'approche présentée dans le cas simple précédent au modèle linéaire général :

$$y_i = \mathbf{x}_i' \mathbf{a}_0 + u_i, \text{ avec } \mathrm{E}[u_i] \equiv 0. \tag{13}$$

- Plusieurs éléments du vecteur $\mathbf{x}_i$ peuvent être endogènes de sorte que dans l'estimateur des MCO de $\mathbf{a}_0$ plusieurs éléments sont potentiellement biaisés(c.f. cours précédent sur les VIs).
- Notons :

$$\mathbf{x}_i = \begin{bmatrix} \begin{bmatrix} 1 \\ \tilde{\mathbf{x}}_i^x \end{bmatrix} \\ \tilde{\mathbf{x}}_i^e \end{bmatrix} = \begin{bmatrix} \mathbf{x}_i^x \\ \tilde{\mathbf{x}}_i^e \end{bmatrix} \begin{array}{ll} \{\text{variables explicatives exogènes} & : \mathrm{E}[u_i | x_{k,i}^x] = 0 (k = 1, \ldots, M) \\ \{\text{variables explicatives endogènes} & : \mathrm{E}[u_i | x_{k,i}^x] \neq 0 (k = M+1, \ldots, K) \end{array}$$

Remarque 3

- ⋆ Il est clair que la variable constante 1 est "exogène" : $\mathrm{E}[u_i|1] = \mathrm{E}[u_i] = \mathrm{E}[1 \times ui] = 0$.
- ⋆ Comme pour l'estimateur des MCO nous utiliserons la Méthode des Moments pour construire un estimateur convergent de $\mathbf{a}_0$, l'estimateur des VI du modèle (13).
- ⋆ On considère ici que chaque élément de $\mathbf{x}_i^e$ a une variable instrumentale.

Définition 2 (Variable instrumentale)

$z_{k,i}$ est une variable instrumentale de $x_{k,i}$ dans le modèle linéaire (13) si :

(i) $\mathrm{Cov}[z_{k,i}; u_i] = 0$ i.e., $z_{k,i}$ est exogènes par rapport à u_i,
(ii) $z_{k,i}$ "suffisamment" liée à $x_{k,i}$.

Remarque 4

- ⋆ On verra dans la suite (analyse des conditions de rang) que la condition (ii) doit en fait être définie comme :

$$\mathrm{Cov}[z_{k,i}; e_{k,i}] \neq 0 \text{ pour } k > 1,$$

où $e_{k,i}$ est la partie spécifique de $x_{k,i}$ dans x_i , i.e. le résidu de la projection linéaire de $x_{k,i}$ sur les autres explicatives $\mathbf{x}_{-k,i}$.

$$e_{k,i} = x_{k,i} - \mathrm{EL}[x_{k,i}|\mathbf{x}_{-k,i}].$$

- ⋆ Dans la définition d'un VI précédente, on voit que lorsqu'une variable explicative $x_{k,i}$ est exogène alors c'est aussi une variable instrumentale d'elle même. En ce sens que non seulement elle vérifie (i) mais elle vérifie forcément (ii) (car ayant une corrélation de 1 avec elle même)

Variables endogènes, exogènes, instruments

- On construit le vecteur des variables instrumentales $\mathbf{z}_i$ avec :

$$\tilde{\mathbf{z}}_i^e = \begin{bmatrix} \tilde{z}_{M+1,i} \\ \tilde{z}_{M+2,i} \\ \vdots \\ \tilde{z}_{K,i} \end{bmatrix} \text{ et } \mathbf{z}_i = \begin{bmatrix} \begin{bmatrix} 1 \\ \tilde{\mathbf{x}}_i^x \end{bmatrix} \\ \tilde{\mathbf{z}}_i^e \end{bmatrix} = \begin{bmatrix} \mathbf{x}_i^x \\ \tilde{\mathbf{z}}_i^e \end{bmatrix} \begin{array}{ll} \{\text{variables exogènes de } \mathbf{x}_i & : \mathrm{E}[u_i|x_{k,i}^x] = 0 (k = 1, \ldots, M) \\ \{\text{variables instrumentales} & : \mathrm{E}[u_i|z_{k,i}] = 0 (k = M+1, \ldots, K) \end{array}$$

- Ce vecteur contient en fait toutes les variables exogènes du modèle. Ce sont ces variables qui assurent l'identification des paramètres du modèle.
- $\mathbf{z}_i$ est parfois nommé ensemble d'information du modèle.

Modèle linéaire à VIs

Définition 3 (Modèle linéaire à variables instrumentales)

Le modèle défini par :

$$y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i, \text{ avec } \mathrm{E}[u_i|\mathbf{z}_i] = \mathrm{E}[u_i] \equiv 0, \tag{14}$$

est un modèle linéaire à variables instrumentales. La condition d'identification de $\mathbf{a}_0$ dans ce modèle est donnée par :

$$\mathrm{Rang}\left(\mathrm{E}[\mathbf{z}_i\mathbf{x}_i']\right) = K = \dim(\mathbf{x}_i) = \dim(\mathbf{z}_i). \tag{15}$$

Remarque 5

La condition d'exogénéité de $\mathbf{z}_i$ est définie par $\mathrm{E}[u_i|\mathbf{z}_i] = 0$, et non par $\mathrm{Cov}[\mathbf{z}_i; u_i] = \mathbf{0}$. Ce n'est pas nécessaire pour un modèle linéaire où $\mathrm{Cov}[\mathbf{z}_i; u_i] = \mathbf{0}$ suffit mais c'est standard et cela simplifie la présentation des hypothèses d'homoscédasticité.

Estimateur des VIs

- Comme dans le cas où on a construit l'estimateur des MCO de $\mathbf{a}_0$ on part de la condition d'exogénéité des $\mathbf{z}_i$ (et non des $\mathbf{x}_i$ comme dans le cas des MCO), i.e. la condition d'orthogonalité donnée par :

$$\mathrm{E}[u_i|\mathbf{z}_i] = 0 \Rightarrow \mathrm{E}[\mathbf{z}_i u_i] = 0 \Leftrightarrow \mathrm{E}[z_i(y_i - \mathbf{x}_i'\mathbf{a}_0)] = \mathbf{0}.$$

- On a ici une condition de moment estimante pour $\mathbf{a}_0$ qui est $\mathrm{E}[\mathbf{z}_i(y_i - \mathbf{x}_i'\mathbf{a}_0)] = \mathbf{0}$. Celle-ci indique que :

$$\mathbf{a}_0 \text{ solution en } \mathbf{a} \text{ de : } \mathrm{E}[\mathbf{z}_i(y_i - \mathbf{x}_i'\mathbf{a})] = \mathbf{0}.$$

- Supposons que $\mathbf{a}_0$ soit l'unique solution en $\mathbf{a}$ de cette condition de moment :

$$\mathrm{E}[\mathbf{z}_i(y_i - \mathbf{x}_i'\mathbf{a})] = \mathbf{0} \Leftrightarrow \mathbf{a} = \mathbf{a}_0.$$

Estimateur des VIs

- Le principe d'analogie définit l'estimateur de la MM de $\mathbf{a}_0$ par :

$$\hat{\mathbf{a}}_N^{MM} \text{ solution en } \mathbf{a} \text{ de : } N^{-1}\sum_{i=1}^{N}\mathbf{z}_i(y_i - \mathbf{x}_i'\mathbf{a}) = \mathbf{0}_{K\times 1},$$

qui est un système de K équations linéaires à K inconnues (les éléments de $\hat{\mathbf{a}}_N^{MM}$). dont la solution en $\mathbf{a}$ est $\hat{\mathbf{a}}_N^{MM}$ qui vérifie donc :

$$N^{-1}\sum_{i=1}^{N}\mathbf{z}_i(y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{MM}) = \mathbf{0}_{K\times 1} \Leftrightarrow N^{-1}\sum_{i=1}^{N}\mathbf{z}_i y_i - \left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right]\hat{\mathbf{a}}_N^{MM} = \mathbf{0}_{K\times 1},$$

- Dès lors que $\left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right]$ est inversible(condition qui sera discutée par la suite)on obtient on obtient la forme explicite de $\hat{\mathbf{a}}_N^{MM}$,

$$\hat{\mathbf{a}}_N^{MM} = \left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right]^{-1} N^{-1}\sum_{i=1}^{N}\mathbf{z}_i y_i \equiv \hat{\mathbf{a}}_N^{VI},$$

qui définit ce qu'on appelle l'estimateur des VI.

Estimateur des VIs

Définition 4 (Estimateur de VIs)

Dans le modèle à variables instrumentales de la définition 3, soit :

$$y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i \text{ avec } \mathrm{E}[u_i|\mathbf{z}_i] = \mathrm{E}[u] \equiv 0, \text{ et } \mathrm{Rang}\left(\mathrm{E}[\mathbf{z}_i\mathbf{x}_i']\right) = K = \dim(\mathbf{x}_i),$$

l'estimateur des variables instrumentales(VIs) de $\mathbf{a}_0$ est défini par :

$$\hat{\mathbf{a}}_N^{VI} \equiv \left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right]^{-1} N^{-1}\sum_{i=1}^{N}\mathbf{z}_i y_i.$$

- Cet estimateur est convergent pour $\mathbf{a}_0$ et asymptotiquement normal,ce que nous allons démontrer (rapidement car « on fait toujours les mêmes démonstrations »).

Propriété 1 (Convergence de l'estimateur des VIs)

Soit $\{(y_i, \mathbf{x}_i \mathbf{z}_i); i = 1, \ldots N.\}$ *un échantillon de variables aléatoires vérifiant la définition 3 d'un modèle à VIs, soit :*

$$y_i = \mathbf{x}_i \mathbf{a}_0' + u_i, \ \textit{avec} \ \mathrm{E}[u_i | \mathbf{z}_i] = \mathrm{E}[u_i] \equiv 0, \ \textit{et} \ \mathrm{Rang}\left(\mathrm{E}[\mathbf{z}_i \mathbf{x}_i']\right) = K = \dim(\mathbf{x}_i).$$

L'estimateur des VI de $\mathbf{a}_0$ *(c.f., définition 4) :*

$$\hat{\mathbf{a}}_N^{VI} \equiv \left[N^{-1} \sum_{i=1}^{N} \mathbf{z}_i \mathbf{x}_i' \right]^{-1} N^{-1} \sum_{i=1}^{N} \mathbf{z}_i y_i.$$

(i) *existe avec une probabilité approchant 1, et,*

(ii) *est convergent, i.e. :*

$$\hat{\mathbf{a}}_N^{VI} \underset{N \to +\infty}{\overset{p}{\longrightarrow}} \mathbf{a}_0.$$

Convergence

Démonstration.

Le modèle de y nous donne que $y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i$, *et* $\hat{\mathbf{a}}_N^{VI}$ *peut s'écrire :*

$$\hat{\mathbf{a}}_N^{VI} = \left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right]^{-1} N^{-1}\sum_{i=1}^{N}\mathbf{z}_i y_i = \left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right]^{-1} N^{-1}\sum_{i=1}^{N}\mathbf{z}_i(\mathbf{x}_i'\mathbf{a}_0 + u_i)$$

$$= \mathbf{a}_0 + \left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right]^{-1} N^{-1}\sum_{i=1}^{N}\mathbf{z}_i u_i.$$

Par la LGN :

$$\left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right] \underset{N\to+\infty}{\overset{p}{\to}} \mathrm{E}[\mathbf{z}_i\mathbf{x}_i']; \; N^{-1}\sum_{i=1}^{N}\mathbf{z}_i u_i \underset{N\to+\infty}{\overset{p}{\to}} \mathrm{E}[\mathbf{z}_i u_i] = 0.$$

En utilisant les propriétés 9 et 10 du chapitre 1, on obtient que :

$$\left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right]^{-1} N^{-1}\sum_{i=1}^{N}\mathbf{z}_i u_i \underset{N\to+\infty}{\overset{p}{\to}} \mathrm{E}[\mathbf{z}_i\mathbf{x}_i']^{-1} \times \mathbf{0} = \mathbf{0},$$

et finalement que,

$$\hat{\mathbf{a}}_N^{VI} \underset{N\to+\infty}{\overset{p}{\to}} \mathbf{a}_0$$

□

Normalité asymptotique

Propriété 2 (Normalité asymptotique de l'estimateur des VIs)

Soit $\{(y_i, \mathbf{x}_i \mathbf{z}_i); i = 1, \ldots N.\}$ *un échantillon de variables aléatoires vérifiant la définition 3 d'un modèle à VIs, soit :*

$$y_i = \mathbf{x}_i \mathbf{a}_0' + u_i, \textit{ avec } \mathrm{E}[u_i|\mathbf{z}_i] = \mathrm{E}[u_i] \equiv 0, \textit{ et } \mathrm{Rang}\left(\mathrm{E}[\mathbf{x}_i \mathbf{x}_i']\right) = K = \dim(\mathbf{x}_i).$$

L'estimateur des VI de $\mathbf{a}_0$ *(c.f., définition 4) :*

$$\hat{\mathbf{a}}_N^{VI} \equiv \left[N^{-1} \sum_{i=1}^{N} \mathbf{z}_i \mathbf{x}_i' \right]^{-1} N^{-1} \sum_{i=1}^{N} \mathbf{z}_i y_i,$$

vérifie :

$$\sqrt{N}\left(\hat{\mathbf{a}}_N^{VI} - \mathbf{a}_0\right) \underset{N \to +\infty}{\overset{L}{\to}} \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}_0\right), \textit{ avec, } \mathbf{\Sigma}_0 \equiv \mathrm{E}[\mathbf{z}_i \mathbf{x}_i']^{-1} \, \mathrm{E}[u_i^2 \mathbf{z}_i \mathbf{z}_i'] \, \mathrm{E}[\mathbf{x}_i \mathbf{z}_i']^{-1}. \quad (16)$$

Remarque 6

La condition d'homoscédasticité éventuelle des u_i *est définie par rapport à* $\mathbf{z}_i$ *et non* $\mathbf{x}_i$*.*

Normalité asymptotique

Démonstration.

Partons de l'écriture suivante de $\hat{\mathbf{a}}_N^{VI}$ *(c.f., démonstration précédente) :*

$$\hat{\mathbf{a}}_N^{VI} = \mathbf{a}_0 + \left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right]^{-1} N^{-1}\sum_{i=1}^{N}\mathbf{z}_i u_i \Rightarrow \sqrt{N}\left(\hat{\mathbf{a}}_N^{VI} - \mathbf{a}_0\right) = \left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right]^{-1} \sqrt{N}\left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i u_i\right].$$

La LGN et la propriété 9 du chapitre 1 donne,

$$\left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right]^{-1} \underset{N\to+\infty}{\overset{p}{\to}} \mathrm{E}[\mathbf{z}_i\mathbf{x}_i']^{-1}.$$

Le TCL(c.f., propriété 2, chapitre 0) implique que :

$$\sqrt{N}\left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i u_i\right] \underset{N\to+\infty}{\overset{L}{\to}} \mathcal{N}\left(\mathrm{E}[\mathbf{z}_i u_i], \mathrm{E}[u_i^2\mathbf{z}_i\mathbf{z}_i']\right) \Rightarrow \sqrt{N}\left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i u_i\right] \underset{N\to+\infty}{\overset{L}{\to}} \mathcal{N}\left(\mathbf{0}, \mathrm{E}[u_i^2\mathbf{z}_i\mathbf{z}_i']\right),$$

car $\mathrm{E}[\mathbf{z}_i u_i] = 0$*, et nous avons ainsi,*

$$\sqrt{N}\left(\hat{\mathbf{a}}_N^{VI} - \mathbf{a}_0\right) \underset{N\to+\infty}{\overset{L}{\to}} \mathrm{E}[\mathbf{z}_i\mathbf{x}_i']^{-1}\mathcal{N}\left(\mathbf{0}, \mathrm{E}[u_i^2\mathbf{z}_i\mathbf{z}_i']\right) \Rightarrow \sqrt{N}\left(\hat{\mathbf{a}}_N^{VI} - \mathbf{a}_0\right) \underset{N\to+\infty}{\overset{L}{\to}} \mathcal{N}\left(\mathbf{0}, \mathrm{E}[\mathbf{z}_i\mathbf{x}_i']^{-1}\,\mathrm{E}[u_i^2\mathbf{z}_i\mathbf{z}_i']\,\mathrm{E}[\mathbf{x}_i\mathbf{z}_i']^{-1}\right)$$

□

Remarque 7

Dans cette démonstration on a appliqué les propriétés classiques suivantes :

- ⋆ *concernant la variance du produit d'un vecteur $\mathbf{m}_i$ et une matrice $\mathbf{A}_0$ de formats conformes :*

$$\mathrm{V}[\mathbf{A}_0\mathbf{m}_i] = \mathbf{A}_0\,\mathrm{V}[\mathbf{m}_i]\mathbf{A}_0' \Rightarrow \mathbf{\Sigma}_0 \equiv \mathrm{V}\left[\underbrace{\mathrm{E}[\mathbf{z}_i\mathbf{x}_i']^{-1}}_{\mathbf{A}_0}\underbrace{\mathcal{N}\left(\mathbf{0}, \mathrm{E}[u_i^2\mathbf{z}_i\mathbf{z}_i']\right)}_{\mathbf{m}_i}\right] = \mathrm{E}[\mathbf{z}_i\mathbf{x}_i']^{-1}\,\mathrm{E}[u_i^2\mathbf{z}_i\mathbf{z}_i']\left(\mathrm{E}[\mathbf{z}_i\mathbf{x}_i']^{-1}\right)'.$$

- ⋆ *que aussi :*

$$\left(\mathbf{A}_0^{-1}\right)' = (\mathbf{A}_0')^{-1} \Rightarrow \left(\underbrace{\mathrm{E}[\mathbf{z}_i\mathbf{x}_i']^{-1}}_{\mathbf{A}_0^{-1}}\right)' = (\mathrm{E}[\mathbf{z}_i\mathbf{x}_i']')^{-1}$$

- ⋆ *et(pour des matrice $\mathbf{A}$ et $\mathbf{B}$ de formats conformes, et une matrice $\mathbf{M}$)*

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}' \;\textit{et}\; \mathrm{E}[\mathbf{M}]' = \mathrm{E}[\mathbf{M}'] \Rightarrow \mathrm{E}[\mathbf{z}_i\mathbf{x}_i']' = \mathrm{E}[\mathbf{x}_i\mathbf{z}_i'].$$

Estimation de la variance de l'estimateur des VIs

- Comme pour l'estimateur des MCO(c.f. chapitre 2) l'estimateur de la variance de la loi approximative de l'estimateur des VIs, $\mathbf{\Sigma}_0$, est différent selon que :
  - ⋆ les u_i sont supposés hétéroscédastiques auquel cas on estime $\mathbf{\Sigma}_0$ telle que donné dans (16).
  - ⋆ les u_i sont supposés homoscédastiques auquel nous avons :

$$\underbrace{\mathrm{V}[u_i|\mathbf{z}_i] = \mathrm{E}[u_i^2|\mathbf{z}_i] = \sigma_0^2}_{\text{Homoscédasticité}} \Rightarrow \mathrm{E}[u_i^2\mathbf{z}_i\mathbf{z}_i'] = \sigma_0^2\,\mathrm{E}[\mathbf{z}_i\mathbf{z}_i'],$$

  conséquence de la loi de conditionnements successifs. La variance à estimer dans (16) devient alors :

$$\mathbf{\Sigma}_0 = \sigma_0^2\,\mathrm{E}[\mathbf{z}_i\mathbf{x}_i']^{-1}\,\mathrm{E}[\mathbf{z}_i\mathbf{z}_i']\,\mathrm{E}[\mathbf{x}_i\mathbf{z}_i']^{-1} \tag{17}$$

- Dans tous les cas, on utilise toujours les mêmes techniques : on remplace les espérances mathématiques par des moyennes et les paramètres inconnus par des estimateurs convergents.

Estimation de la variance de l'estimateur des VIs

- **Cas homoscédastique**
  - ⋆ Pour estimer $\boldsymbol{\Sigma}_0$ dans (17) on utilise le fait que :

$$\left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right]^{-1} \underset{N\to+\infty}{\overset{p}{\longrightarrow}} \mathrm{E}[\mathbf{x}_i\mathbf{z}_i']^{-1}, \quad \left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right]^{-1} \underset{N\to+\infty}{\overset{p}{\longrightarrow}} \mathrm{E}[\mathbf{z}_i\mathbf{x}_i']^{-1}, \quad N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i' \underset{N\to+\infty}{\overset{p}{\longrightarrow}} \mathrm{E}[\mathbf{z}_i\mathbf{z}_i'],$$

  - ⋆ Avec,

$$u_i = y_i - \mathbf{x}_i'\mathbf{a}_0, \quad \text{et} \quad \hat{\mathbf{a}}_N^{VI} \underset{N\to+\infty}{\overset{p}{\longrightarrow}} \mathbf{a}_0,$$

nous avons alors :

$$\hat{\sigma}_N^2 \equiv N^{-1}\sum_{i=1}^{N}\left(y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{VI}\right) \underset{N\to+\infty}{\overset{p}{\longrightarrow}} \sigma_0^2 \equiv \mathrm{E}[(y_i - \mathbf{x}_i'\mathbf{a}_0)],$$

et :

$$\hat{\boldsymbol{\Sigma}}_N \equiv \hat{\sigma}_N^2\left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right]^{-1}\left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right]\left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right]^{-1} \underset{N\to+\infty}{\overset{p}{\longrightarrow}} \sigma_0^2\,\mathrm{E}[\mathbf{z}_i\mathbf{x}_i']^{-1}\,\mathrm{E}[\mathbf{z}_i\mathbf{z}_i']\,\mathrm{E}[\mathbf{x}_i\mathbf{z}_i']^{-1} \equiv \boldsymbol{\Sigma}_0.$$

Estimation de la variance de l'estimateur des VIs

- **Cas hétéroscédastique**
  - ⋆ Ici on estime $\mathbf{\Sigma}_0$ dans "le cas général" où $\mathbf{\Sigma}_0$ est donné par (16).
  - ⋆ On utilise le fait que :

$$\left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right]^{-1} \underset{N\to+\infty}{\overset{p}{\to}} \mathrm{E}[\mathbf{x}_i\mathbf{z}_i']^{-1}, \quad \left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right]^{-1} \underset{N\to+\infty}{\overset{p}{\to}} \mathrm{E}[\mathbf{z}_i\mathbf{x}_i']^{-1}.$$

  - ⋆ Avec :

$$u_i = y_i - \mathbf{x}_i'\mathbf{a}_0 \quad \text{on a} \quad \mathrm{E}[u_i^2\mathbf{z}_i\mathbf{z}_i'] = \mathrm{E}[(y_i - \mathbf{x}_i'\mathbf{a}_0)^2\mathbf{z}_i\mathbf{z}_i'],$$

par conséquent ,

$$\hat{\mathbf{a}}_N^{VI} \underset{N\to+\infty}{\overset{p}{\to}} \mathbf{a}_0 \Rightarrow \sum_{i=1}^{N}(y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{VI})^2\mathbf{z}_i\mathbf{z}_i' \underset{N\to+\infty}{\overset{p}{\to}} \mathrm{E}[u_i^2\mathbf{z}_i\mathbf{z}_i'] = \mathrm{E}[(y_i - \mathbf{x}_i'\mathbf{a}_0)^2\mathbf{z}_i\mathbf{z}_i']$$

et :

$$\hat{\mathbf{\Sigma}}_N^{W} \equiv \left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right]^{-1}\left[\sum_{i=1}^{N}(y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{VI})^2\mathbf{z}_i\mathbf{z}_i'\right]\left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right]^{-1} \underset{N\to+\infty}{\overset{p}{\to}} \mathrm{E}[\mathbf{z}_i\mathbf{x}_i']^{-1}\,\mathrm{E}[u_i^2\mathbf{z}_i\mathbf{z}_i']\,\mathrm{E}[\mathbf{x}_i\mathbf{z}_i']^{-1} \equiv \mathbf{\Sigma}_0.$$

Estimation de la variance de l'estimateur des VIs

Propriété 3 (Estimateurs de la variance asymptotique de l'estimateur des VIs)

Des estimateurs convergent de la variance asymptotique $\mathbf{\Sigma}_0$, de l'estimateur des VI de $\mathbf{a}_0$, $\hat{\mathbf{a}}_N^{VI}$ dans le modèle à VIs(c.f., définition 3) sont :

(i) *dans le cas d'erreurs hétéroscédastiques :*

$$\hat{\mathbf{\Sigma}}_N^W \equiv \left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right]^{-1}\left[\sum_{i=1}^{N}\hat{u}_{i,N}^2\mathbf{z}_i\mathbf{z}_i'\right]\left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right]^{-1},$$

(ii) *et dans le cas d'erreurs homoscédastiques :*

$$\hat{\mathbf{\Sigma}}_N \equiv \hat{\sigma}_N^2\left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right]^{-1}\left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right]\left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right]^{-1}$$

où on note $\hat{u}_{i,N} \equiv y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{VI}$ le résidu de l'estimation et $\hat{\sigma}_N^2 \equiv N^{-1}\sum_{i=1}^{N}\hat{u}_{i,N}^2$.

Remarque 8

L'estimateur $\hat{\mathbf{\Sigma}}_N^W$ est un estimateur qualifié de robuste à l'hétérosécédasticité. Il est aussi appelé estimateur robuste de White. D'autres estimateurs robustes en présence d'hétérosécédasticité existent.

5. L’estimateur des 2MC

Sur-identification dans un modèle à VIs

- On continue d'étudier un modèle à VIs.
- Cependant par rapport au modèle de la définition 3, nous allons étudier le cas où $\dim(\mathbf{z}_i) > \dim(\mathbf{x}_i)$.
- Cela correspond au cas où pour certaines variables explicatives endogènes dans $\mathbf{x}_i^e \subseteq \mathbf{x}_i$, nous pouvons avoir plusieurs variables instrumentales dans $\mathbf{z}_i^e \subseteq \mathbf{z}_i$. Autrement dit une variable explicative endogène peut avoir plus d'un instrument.
- En effet, quand on cherche à « instrumenter » une variable explicative endogène, disons $x_{k,i}^e$ (avec $k > M$), on cherche des variables exogènes par rapport à u_i(ce que l'on ne peut pas vérifier u_i n'étant pas observé) et susceptibles d'être corrélées à $x_{k,i}^e$ (ce qu'on peut vérifier).
- Schématiquement une VI de $x_{k,i}^e$:
  1. Doit être un "bon" prédicteur de $x_{k,i}^e$ par exemple dans le cadre de la projection linéaire de $x_{k,i}^e$ sur la VI candidate.
  2. Doit pouvoir être supposée exogène dans le modèle d'intérêt, i.e., par rapport à u_i dans $y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i$, où $\mathbf{x}_i$ contient des élément endogènes. Cette condition n'étant pas testable empiriquement(car u_i n'est pas observée), sa vraisemblance doit être discutée en fonction de l'application empirique considérée.

Sur-identification dans un modèle à VIs

- En pratique on peut donc disposer de plusieurs VI pour une seule variable explicative endogène :
  - ⋆ C'est bon en matière d'inférence : plusieurs VIs contiennent plus d'information qu'une seule, ...
  - ⋆ ..., mais cela rend impossible l'utilisation directe de l'estimateur des VIs.
- La **méthode de doubles moindres carrés(2MC)** et l'estimateur des 2MC sont des outils classiques pour traiter ce type de cas où l'on parle de **modèle sur-identifié**.

Remarque 9

L'estimateur 2MC ainsi que celui des VIs vu précédemment(et d'autres encore comme les MCO avec explicatives exogènes), peuvent être étudiés dans le cadre de la **méthode des moments généralisés(MMG)**, qui a été expressément développé pour étudier des modèles potentiellement sur-identifiés. Cette méthode fera l'objet d'un autre cours, et ici nous présentons les 2MC d'une manière alternative.

Sur-identification dans un modèle à VIs

- On continue à se placer dans le cadre du modèle modèle linéaire à VI de forme générale :

$$y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i, \quad \text{avec} \quad \mathrm{E}[u_i|\mathbf{z}_i] = \mathrm{E}[u_i] \equiv 0,$$

  où $\tilde{\mathbf{x}}_i^e \subseteq \mathbf{x}_i$ variables explicatives sont endogènes, $\mathbf{x}_i^x \subseteq \mathbf{x}_i$ variables explicatives sont exogènes(dont un régresseur constant), avec $\dim(\mathbf{x}_i^x) = M$ et $\dim(\tilde{\mathbf{x}}_i^e) = K - M$.
- Nous avons un vecteur de VIs $\mathbf{z}_i$ composé comme précédemment :
  - ⋆ des variables explicatives exogènes $\mathbf{x}_i^x$(qui sont "leurs propres instruments"),
  - ⋆ et d'un vecteur $\tilde{\mathbf{z}}_i^e$ de VIs "externes" disponibles pour instrumenter le vecteur de variables explicatives endogènes $\tilde{\mathbf{x}}_i^e$, avec $\dim(\tilde{\mathbf{z}}_i^e) = L - M$ pour $L \geq K$.
- Donc :

$$\mathbf{z}_i = \begin{bmatrix} \begin{bmatrix} 1 \\ \tilde{\mathbf{x}}_i^x \end{bmatrix} \\ \tilde{\mathbf{z}}_i^e \end{bmatrix} = \begin{bmatrix} \mathbf{x}_i^x \\ \tilde{\mathbf{z}}_i^e \end{bmatrix} \begin{array}{ll} \{\text{variables exogènes de } \mathbf{x}_i & : \mathrm{E}[u_i|x_{k,i}^x] = 0 (k = 1, \ldots, M) \\ \{\text{variables instrumentales} & : \mathrm{E}[u_i|z_{k,i}] = 0 (k = M+1, \ldots, L) \end{array},$$

  où $\tilde{\mathbf{z}}_i^e$ contient toutes les VI «externes» disponibles pour instrumenter $\tilde{\mathbf{x}}_i^e$ et :

$$\underbrace{\dim(\tilde{\mathbf{x}}_i^e) = K - M}_{\text{Nombre de variables explicatives exogènes}} \leq \underbrace{\dim(\tilde{\mathbf{z}}_i^e) = L - M}_{\text{Nombre de VIs externes}} \Leftrightarrow \underbrace{\dim(\mathbf{x}_i) = K}_{\text{Nombre de variables explicatives}} \leq \underbrace{\dim(\mathbf{z}_i) = L}_{\text{Nombre de VIs}}$$

Sur-identification dans un modèle à VIs

- Quand nous avons construit l'estimateur des VIs avec $\dim(\mathbf{z}_i) = K$ nous avons utilisé la condition d'exogénéité des VIs qui nous donnait une condition de moment estimante pour $\mathbf{a}_0$:

$$\mathrm{E}[\mathbf{z}_i u_i] = 0 \Rightarrow \mathrm{E}[\mathbf{z}_i(y_i - \mathbf{x}_i'\mathbf{a}_0)] = \mathbf{0},$$

- Nous avons alors défini l'estimateur des VIs comme la solution de la contrepartie empirique de cette condition théorique :

$$N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\left(y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{VI}\right) = \mathbf{0}_{K\times 1},$$

qui est un système de $K = \dim(\mathbf{z}_i)$ équations à $K = \dim(\hat{\mathbf{a}}_N^{VI})$ inconnues, lequel possède une solution dès lors que $\mathrm{Rang}\left(\sum_{i=1}^N \mathbf{z}_i'\mathbf{x}_i\right) = K$, $\left[N^{-1}\sum_{i=1}^N \mathbf{z}_i'\mathbf{x}_i\right]$ étant alors inversible et comme nous l'avons vu :

$$\hat{\mathbf{a}}_N^{VI} \equiv \left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right]^{-1} N^{-1}\sum_{i=1}^{N}\mathbf{z}_i y_i.$$

Sur-identification dans un modèle à VIs

- Quand $\dim(\mathbf{z}_i) = L > K$ le système à considérer devient :

$$\sum_{i=1}^{N} \mathbf{z}_i \left(y_i - \mathbf{x}_i' \hat{\mathbf{a}}_N^{VI} \right) = \mathbf{0}_{L\times 1}, \tag{18}$$

  qui est un système de $L = \dim(\mathbf{z}_i)$ équations à $K = \dim(\hat{\mathbf{a}}_N^{VI})$ lequel n'admet pas, en général, de solution.

- Dans un modèle à VIs où le nombre de VIs est supérieur au nombre de variables explicatives est un modèle où l'on dit que les *VIs sur-identifient* les paramètres, et on parle aussi de *modèle sur-identifié* ou de *sur-identification*.
- Plus généralement, nous avons la définition suivante selon que $K \lesseqgtr L$:

Définition 5

Dans un modèle à VIs, le vecteur d'instruments $\mathbf{z}_i$

(i) identifie exactement $\mathbf{a}_0$ si $\dim(\mathbf{a}_0) = \dim(\mathbf{x}_i) = K = L = \dim(\mathbf{z}_i)$,
(ii) sur-identifie $\mathbf{a}_0$ si $\dim(\mathbf{a}_0) = \dim(\mathbf{x}_i) = K < L = \dim(\mathbf{z}_i)$,
(iii) n'identifie pas $\mathbf{a}_0$ si $\dim(\mathbf{a}_0) = \dim(\mathbf{x}_i) = K > L = \dim(\mathbf{z}_i)$.

Sur-identification dans un modèle à VIs

- En résumé, le problème de la sur-identification est ici que (18) contient "trop" d'équations pour avoir une solution exacte dans $\mathbb{R}^K$.
- Il y a plusieurs possibilités pour résoudre ce problème :
  (i) Éliminer des éléments de $\mathbf{z}_i$, en particulier $L - K$ pour revenir au cas où d'identification exacte avec $\dim(\mathbf{z}_i) = K$. Cependant, cela s'accompagnera en général d'une perte d'information(fournit par les VIs éliminées) et d'efficacité.
  (ii) Utiliser, comme mentionné plus haut, la Méthode des Moments Généralisée qui a été expressément développée pour des modèles potentiellement sur-identifiés.
  (iii) Utiliser une astuce pour "réduire" la dimension du vecteur de VIs utilisé comme dans la première possibilité, mais sans perdre d'information (ou tout au moins un minimum).
- C'est à cette dernière possibilité que correspond la méthode des 2MC.

Projection linéaire et 2MC

- On veut en fait définir à partir de $\mathbf{z}_i$, un vecteur d'instruments, noté $\mathbf{w}(\mathbf{z}_i)$, de dimension K tel que :

$$\mathrm{E}[\mathbf{w}(\mathbf{z}_i)(y_i - \mathbf{x}_i'\mathbf{a}_0)] = \mathbf{0},$$

ce qui nous permettra de définir un estimateur des VIs, $\hat{\mathbf{a}}_N^{VI}$, comme solution de la contrepartie empirique de la condition ci-dessus :

$$N^{-1}\sum_{i=1}^{N}\mathbf{w}(\mathbf{z}_i)(y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{VI}) = \mathbf{0}_{K\times 1}.$$

- Pour cela on part du principe qu'un bon vecteur d'instruments doit :
  - (i) être exogène par rapport à u_i et,
  - (ii) permettre de "bien prédire" $\mathbf{x}_i$ (ou avoir une "forte" corrélation avec $\mathbf{x}_i$).
- La projection linéaire(c.f., chapitre 2 pour un rappel) de $\mathbf{x}_i$ sur $\mathbf{z}_i$, $\mathrm{EL}[\mathbf{x}_i|\mathbf{z}_i]$ se présente comme un choix approprié. En effet :
  - (i) $\dim(\mathrm{EL}[\mathbf{x}_i|\mathbf{z}_i]) = K$ ce qu'on veux pour utiliser la MM.
  - (ii) $\mathrm{EL}[\mathbf{x}_i|\mathbf{z}_i]$ est exogène car c'est(par construction) une fonction des VIs.
  - (iii) $\mathrm{EL}[\mathbf{x}_i|\mathbf{z}_i]$ est(par définition) la meilleure prédiction linéaire de $\mathbf{x}_i$ par $\mathbf{z}_i$.

Remarque 10

$\mathrm{EL}[\mathbf{x}_i|\mathbf{z}_i]$ n'est pas cependant le meilleur choix. Ce dernier est, en termes d'efficacité as. (c.f., Chamberlain), $\mathrm{EL}[\mathbf{x}_i|\mathbf{z}_i]\,\mathrm{V}[u_i|\mathbf{z}_i]^{-1}$.

Projection linéaire et 2MC

- En conséquence $\mathbf{w}(\mathbf{z}_i) = \mathrm{EL}[\mathbf{x}_i|\mathbf{z}_i]$ qui est un empilement des projections de $\mathrm{EL}[x_{k,i}|\mathbf{z}_i]$ pour $k = 1, \ldots, K$:

$$\mathrm{EL}[\mathbf{x}_i|\mathbf{z}_i] \equiv \begin{bmatrix} \mathrm{EL}[x_{1,i}|\mathbf{z}_i] \\ \vdots \\ \mathrm{EL}[x_{K,i}|\mathbf{z}_i] \end{bmatrix} \equiv \begin{bmatrix} \mathbf{z}_i'\gamma_1 \\ \vdots \\ \mathbf{z}_i'\gamma_K \end{bmatrix} = \begin{bmatrix} \mathrm{EL}[1|\mathbf{z}_i] \\ \mathrm{EL}[x_{2,i}|\mathbf{z}_i] \\ \vdots \\ \mathrm{EL}[x_{M,i}|\mathbf{z}_i] \\ \mathrm{EL}[x_{M+1,i}|\mathbf{z}_i] \\ \vdots \\ \mathrm{EL}[x_{K,i}|\mathbf{z}_i] \end{bmatrix} = \begin{bmatrix} 1 \\ x_{2,i} \\ \vdots \\ x_{M,i} \\ \mathbf{z}_i'\gamma_{M+1} \\ \vdots \\ \mathbf{z}_i'\gamma_K \end{bmatrix} = \begin{bmatrix} \mathbf{x}_i^x \\ \mathbf{z}_i'\gamma_{M+1} \\ \vdots \\ \mathbf{z}_i'\gamma_K \end{bmatrix} \equiv \mathbf{\Gamma}\mathbf{z}_i,$$

où à condition que $\mathrm{E}[\mathbf{z}_i\mathbf{z}_i']$ soit inversible(on peut montrer que) :

$$\mathbf{\Gamma}\mathbf{z}_i = \underbrace{\mathrm{E}[\mathbf{x}_i\mathbf{z}_i']\,\mathrm{E}[\mathbf{z}_i\mathbf{z}_i']^{-1}}_{\mathbf{\Gamma}}\,\mathbf{z}_i, \tag{19}$$

Projection linéaire et 2MC

- Un estimateur convergent de $\mathbf{\Gamma}$, noté $\hat{\mathbf{\Gamma}}_N$, est simplement l'empilement des estimateurs des MCO $\hat{\gamma}_{k,N}^{MCO\prime}$ des γ_k pour la projection linéaire $\mathrm{EL}[x_{k,i}|\mathbf{z}_i]$. Celui-ci nous est donné par :

$$\hat{\mathbf{\Gamma}}_N \equiv \left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right]\left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right]^{-1} \underset{N\to+\infty}{\overset{p}{\to}} \mathbf{\Gamma}, \Rightarrow \hat{\mathbf{\Gamma}}_N\mathbf{z}_i \underset{N\to+\infty}{\overset{p}{\to}} \mathbf{\Gamma}\mathbf{z}_i, \tag{20}$$

où l'on suppose que la LGN et les propriétés des suites convergeant en probabilité s'appliquent.

Estimateur des 2MC

- Les instruments étant

$$\mathbf{w}(\mathbf{x}_i) = \mathrm{EL}[\mathbf{x}_i|\mathbf{z}_i] = \mathbf{\Gamma}\mathbf{z}_i = \mathrm{E}[\mathbf{x}_i\mathbf{z}_i'] \, \mathrm{E}[\mathbf{z}_i\mathbf{z}_i']^{-1}\mathbf{z}_i,$$

  la condition de moment estimante "modifiée" pour $\mathbf{a}_0$ est donné par :

$$\mathrm{E}[\mathbf{w}(\mathbf{z}_i)(y_i - \mathbf{x}_i'\mathbf{a}_0)] = \mathrm{E}\left[\underbrace{\mathrm{E}[\mathbf{x}_i\mathbf{z}_i'] \, \mathrm{E}[\mathbf{z}_i\mathbf{z}_i']^{-1}\mathbf{z}_i}_{\mathbf{w}(\mathbf{z}_i)}(y_i - \mathbf{x}_i'\mathbf{a}_0)\right] = \mathrm{E}[\mathbf{x}_i\mathbf{z}_i'] \, \mathrm{E}[\mathbf{z}_i\mathbf{z}_i']^{-1} \, \mathrm{E}\left[\mathbf{z}_i(y_i - \mathbf{x}_i'\mathbf{a}_0)\right] = \mathbf{0}$$

  qui est donc un système de $\dim(\mathbf{w}(\mathbf{z}_i)) = K$ équations à $\dim(\mathbf{a}_0) = K$ inconnues à partir duquel on peut construire un estimateur des moments.
- Notons que dans l'équation correspondant à la dernière égalité il apparaît plus explicitement que $\mathbf{z}_i$ identifie $\mathbf{a}_0$ parce qu'elle n'influence y_i que via son effet sur $\mathbf{x}_i$.

Estimateur des 2MC

- Pour construire un estimateur des moments à partir du principe d'analogie on part donc de cette dernière égalité qui indique que :

$$\mathbf{a}_0 \text{ solution en } \mathbf{a} \text{ de : } \mathrm{E}[\mathbf{x}_i\mathbf{z}_i'] \, \mathrm{E}[\mathbf{z}_i\mathbf{z}_i']^{-1} \, \mathrm{E}\left[\mathbf{z}_i(y_i - \mathbf{x}_i'\mathbf{a})\right] = \mathbf{0}$$

  et l'on suppose que $\mathbf{a}_0$ est l'unique solution de ce système :

$$\mathrm{E}[\mathbf{x}_i\mathbf{z}_i'] \, \mathrm{E}[\mathbf{z}_i\mathbf{z}_i']^{-1} \, \mathrm{E}\left[\mathbf{z}_i(y_i - \mathbf{x}_i'\mathbf{a})\right] = \mathbf{0} \Leftrightarrow \mathbf{a} = \mathbf{a}_0.$$

- L'utilisation du principe d'analogie permet alors de définir un estimateur des moments par :

$$\hat{\mathbf{a}}_N^{MM} \text{ solution en } \mathbf{a} \text{ de : } N^{-1} \sum_{i=1}^{N} \hat{\boldsymbol{\Gamma}}_N \mathbf{z}_i(y_i - \mathbf{x}_i'\mathbf{a}) = \mathbf{0}_{K\times 1},$$

Estimateur des 2MC

et qui implique :

$$
\begin{aligned}
N^{-1}\sum_{i=1}^{N}\hat{\boldsymbol{\Gamma}}_N\mathbf{z}_i(y_i-\mathbf{x}_i'\hat{\mathbf{a}}_N^{MM}) = \mathbf{0}_{K\times 1} \Rightarrow \hat{\mathbf{a}}_N^{MM} &= \left[N^{-1}\sum_{i=1}^{N}\hat{\boldsymbol{\Gamma}}_N\mathbf{z}_i\mathbf{x}_i'\right]^{-1} N^{-1}\sum_{i=1}^{N}\hat{\boldsymbol{\Gamma}}_N\mathbf{z}_i y_i,\\
&= \left\{\underbrace{\left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right]\left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right]^{-1}}_{\hat{\boldsymbol{\Gamma}}_N} N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right\}^{-1}\\
&\quad\times \underbrace{\left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right]\left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right]^{-1}}_{\hat{\boldsymbol{\Gamma}}_N} N^{-1}\sum_{i=1}^{N}\mathbf{z}_i y_i \equiv \hat{\mathbf{a}}_N^{2MC}.
\end{aligned}
$$

où l'on a utilisé l'expression de $\hat{\boldsymbol{\Gamma}}_N$ d'après (20) pour obtenir un estimateur des moments appelé estimateur des doubles moindres carrés.

Estimateur des 2MC

Définition 6 (Estimateur des 2MC dans un modèle à VIs)

Dans un modèle à variables instrumentales :

$$y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i, \quad \text{avec} \quad \mathrm{E}[u_i|\mathbf{z}_i] = \mathrm{E}[u_i] \equiv 0, \quad \text{et} \quad \dim(\mathbf{z}_i) \geq \dim(\mathbf{x}_i),$$

l'estimateur des 2MC de $\mathbf{a}_0$ est donné par :

$$\hat{\mathbf{a}}_N^{2MC} \equiv \left[N^{-1} \sum_{i=1}^{N} \hat{\boldsymbol{\Gamma}}_N \mathbf{z}_i \mathbf{x}_i' \right]^{-1} N^{-1} \sum_{i=1}^{N} \hat{\boldsymbol{\Gamma}}_N \mathbf{z}_i y_i,$$

où :

$$\hat{\boldsymbol{\Gamma}}_N \equiv \left[N^{-1} \sum_{i=1}^{N} \mathbf{x}_i \mathbf{z}_i' \right] \left[N^{-1} \sum_{i=1}^{N} \mathbf{z}_i \mathbf{z}_i' \right]^{-1}$$

Estimateur des 2MC

Remarque 11

L'estimateur $\hat{\mathbf{a}}_N^{2MC}$ a la structure d'un estimateur des VIs :

$$\hat{\mathbf{a}}_N^{2MC} \equiv \left[N^{-1} \sum_{i=1}^{N} \underbrace{\hat{\mathbf{\Gamma}}_N \mathbf{z}_i}_{\hat{\mathbf{w}}_N(\mathbf{z}_i)} \mathbf{x}_i' \right]^{-1} N^{-1} \sum_{i=1}^{N} \underbrace{\hat{\mathbf{\Gamma}}_N \mathbf{z}_i}_{\hat{\mathbf{w}}_N(\mathbf{z}_i)} y_i,$$

avec $\hat{\mathbf{w}}_N(\mathbf{z}_i) = \hat{\mathbf{\Gamma}}_N \mathbf{z}_i$ comme vecteur d'instruments "estimés" et qui est en fait un estimateur convergent de $\mathbf{w}(\mathbf{z}_i) \equiv \mathrm{EL}[\mathbf{x}_i | \mathbf{z}_i]$.

Estimateur des 2MC

Remarque 12

On peut aussi montrer (car $\hat{\mathbf{\Gamma}}_N$ est une matrice de projection et donc idempotente) que que l'estimateur $\hat{\mathbf{a}}_N^{2MC}$ a la structure d'un estimateur des MCO :

$$\hat{\mathbf{a}}_N^{2MC} \equiv \left[N^{-1} \sum_{i=1}^{N} \left(\underbrace{\hat{\mathbf{\Gamma}}_N \mathbf{z}_i}_{\hat{\mathbf{w}}_N(\mathbf{z}_i)} \right) \left(\underbrace{\hat{\mathbf{\Gamma}}_N \mathbf{z}_i}_{\hat{\mathbf{w}}_N(\mathbf{z}_i)} \right)' \right]^{-1} N^{-1} \sum_{i=1}^{N} \left(\underbrace{\hat{\mathbf{\Gamma}}_N \mathbf{z}_i}_{\hat{\mathbf{w}}_N(\mathbf{z}_i)} \right) y_i,$$

avec un vecteur de variables explicatives "estimées"(régresseurs estimés) $\hat{\mathbf{w}}_N(\mathbf{z}_i) = \hat{\mathbf{\Gamma}}_N \mathbf{z}_i$, l'estimateur (convergent) de $\mathbf{w}(\mathbf{z}_i) \equiv \mathrm{EL}[\mathbf{x}_i | \mathbf{z}_i]$.

Estimateur des 2MC

Remarque 13

L'estimateur des 2MC peut en fait être calculé en deux étapes chacune consistant en un estimation par MCO, ou "MCO successifs"(d'où son nom).

1. Calcul de $\hat{\mathbf{\Gamma}}_N$ qui, nous l'avons indiqué, est un empilement d'estimateurs des MCO $\hat{\gamma}_{k,N}^{MCO'}$ pour la projection $\mathrm{EL}[x_{k,i}|\mathbf{z}_i]$. Plus précisément :

$$\hat{\mathbf{\Gamma}}_N = \begin{bmatrix} \begin{bmatrix} \mathbf{I}_M & \mathbf{0} \end{bmatrix} \\ \begin{bmatrix} \hat{\gamma}_{M+1,N}^{MCO'} \\ \vdots \\ \hat{\gamma}_{K,N}^{MCO'} \end{bmatrix} \end{bmatrix}, \text{ avec } \hat{\gamma}_{K,N}^{MCO} \equiv \left[N^{-1} \sum_{i=1}^{N} \mathbf{z}_i \mathbf{z}_i' \right]^{-1} N^{-1} \sum_{i=1}^{N} \mathbf{z}_i x_{k,i}$$

2. MCO pour calculer :

$$\hat{\mathbf{a}}_N^{2MC} \equiv \left[N^{-1} \sum_{i=1}^{N} \left(\hat{\mathbf{\Gamma}}_N \mathbf{z}_i \right) \left(\hat{\mathbf{\Gamma}}_N \mathbf{z}_i \right)' \right]^{-1} N^{-1} \sum_{i=1}^{N} \left(\hat{\mathbf{\Gamma}}_N \mathbf{z}_i \right) y_i,$$

Estimateur des 2MC

Remarque (suite de la remarque 13)

- ⋆ Cette propriété été importante lorsque les moyens de calculs étaient limités. Elle n'a plus beaucoup d'intérêt maintenant.
- ⋆ Il est conseillé d'utiliser cette propriété avec précaution (et je changerais bien le nom de cet estimateur) car elle peut être "dangereuse" :
  - (i) Par exemple, un estimateur des 2MC non linéaires a été défini par extension au cas linéaire, mais il ne possède pas cette propriété des « MCO successifs » !
  - (ii) De même, les calculs par estimations successives sont à employer avec précaution, surtout lorsqu'on « estime » des variables explicatives. Cela est discuté plus en détail dans le cadre de la présentation du "test de la régression augmentée".
- ⋆ La technique des « MCO successifs » repose (en seconde étape) sur l'utilisation de « régresseurs estimés » (i.e. de variables explicatives estimées), ces derniers devant être utilisés avec précaution.
- ⋆ Notons que l'utilisation d'"instruments estimés" ne pose pas de problème.

Estimateur des VIs et estimateur des 2MC

- En pratique il est courant de parler d'estimateur des VIs et d'estimateur des 2MC sans en faire la distinction :
  - ⋆ L'estimateur des VI « n'existe pas » dans les logiciels/bibliothèques/modules de programmation pour l'économétrie (et il a tendance à disparaître des manuels d'économétrie).
  - ⋆ Ceci provient de ce que l'estimateur des VIs est un cas particulier de l'estimateur des 2MC dont le calcul est programmé dans tous les logiciels d'économétrie.
- Pour préciser cela réintroduisons et introduisons les notations matricielles suivante(c.f., chapitre 1 pour la matrice des régresseurs ou cours du S1) :

$$\mathbf{X} \equiv \begin{bmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_N' \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{2,1} & \dots & x_{K,1} \\ x_{1,2} & x_{2,2} & \dots & x_{K,2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,N} & x_{2,N} & \dots & x_{K,N} \end{bmatrix}_{N\times K}, \mathbf{Z} \equiv \begin{bmatrix} \mathbf{z}_1' \\ \vdots \\ \mathbf{z}_N' \end{bmatrix} = \begin{bmatrix} z_{1,1} & z_{2,1} & \dots & z_{L,1} \\ z_{1,2} & z_{2,2} & \dots & z_{L,2} \\ \vdots & \vdots & \ddots & \vdots \\ z_{1,N} & z_{2,N} & \dots & z_{L,N} \end{bmatrix}_{N\times L}, \mathbf{y} \equiv \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}_{N\times 1}.$$

- On peut écrire les estimateurs $\hat{\mathbf{a}}_N^{VI}$, $\hat{\mathbf{a}}_N^{2MC}$, ainsi que $\hat{\mathbf{a}}_N^{MCO}$ en utilisant ces notations :

$$\hat{\mathbf{a}}_N^{VI} \equiv \left[\mathbf{Z}'\mathbf{X}\right]^{-1}\mathbf{Z}'\mathbf{y},\ \hat{\mathbf{a}}_N^{2MC} \equiv \left\{\mathbf{X}'\mathbf{Z}\left[\mathbf{Z}'\mathbf{Z}\right]^{-1}\mathbf{Z}'\mathbf{X}\right\}^{-1}\mathbf{X}'\mathbf{Z}\left[\mathbf{Z}'\mathbf{Z}\right]^{-1}\mathbf{Z}'\mathbf{y},\ \hat{\mathbf{a}}_N^{MCO} \equiv \left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{y}$$

Estimateur des VIs et estimateur des 2MC

- Cette écriture permet de montrer que l'estimateur des VIs est un cas particulier d'estimateur des 2MC :
  - ⋆ Si $K = L$ alors les matrices $\mathbf{X}'\mathbf{Z}$, $\mathbf{Z}'\mathbf{Z}$ et $\mathbf{Z}'\mathbf{X}$ sont carrées et de même dimension : $K \times K$.
  - ⋆ En utilisant les propriétés des inverses de produits de matrices inversibles, $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$, on obtient :

$$\hat{\mathbf{a}}_N^{2MC} \equiv \left\{\mathbf{X}'\mathbf{Z}\left[\mathbf{Z}'\mathbf{Z}\right]^{-1}\mathbf{Z}'\mathbf{X}\right\}^{-1}\mathbf{X}'\mathbf{Z}\left[\mathbf{Z}'\mathbf{Z}\right]^{-1}\mathbf{Z}'\mathbf{y} = \left[\mathbf{Z}'\mathbf{X}\right]^{-1}\mathbf{Z}'\mathbf{Z}\left[\mathbf{X}'\mathbf{Z}\right]^{-1}\mathbf{X}'\mathbf{Z}\left[\mathbf{Z}'\mathbf{Z}\right]^{-1}\mathbf{Z}'\mathbf{y} = \left[\mathbf{Z}'\mathbf{X}\right]^{-1}\mathbf{Z}'\mathbf{y} \equiv \hat{\mathbf{a}}_N^{VI}$$

  - ⋆ De même, l'estimateur des MCO est un cas particulier d'estimateur des VIs, celui où les variables explicatives peuvent être utilisées comme VIs. On a alors $\mathbf{Z} = \mathbf{X}$ et :

$$\hat{\mathbf{a}}_N^{VI} = \left[\mathbf{Z}'\mathbf{X}\right]^{-1}\mathbf{Z}'\mathbf{y} = \left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{y} \equiv \hat{\mathbf{a}}_N^{MCO}.$$

Références

Angrist, Joshua D. and William N. Evans. 1998. “Children and Their Parents’ Labor Supply : Evidence from Exogenous Variation in Family Size.” *The American Economic Review* 88 (3) :pp. 450–477. URL http://www.jstor.org/stable/116844.

Angrist, Joshua D. and Alan B. Krueger. 1991. “Does Compulsory School Attendance Affect Schooling and Earnings ?” *The Quarterly Journal of Economics* 106 (4) :pp. 979–1014. URL http://www.jstor.org/stable/2937954.

Card, David. 1993. “Using geographic variation in college proximity to estimate the return to schooling.” Tech. rep., National Bureau of Economic Research. URL http://www.nber.org/papers/w4483.