A.3. Estimation par la MMG dans les systèmes d’équations, en particulier linéaires

1. Approche générale

2. Systèmes d’équations linéaires

3. Estimateurs MMG optimaux dans les systèmes d’équations linéaires

4. Estimateurs « en système » *versus* « équation par équation »

5. Systèmes d’équations simultanées linéaires : estimateur des 3MC

6. Systèmes de régressions linéaires empilées : estimateur SUR

1. Approche générale

De fait, la MMG a deux très grands avantages :

- Le premier a été illustré dans le chapitre précédent, il s'agit de l'estimation de paramètres à partir de conditions de moment sur-identifiante.

C'est en ce sens qu'elle est une Méthode des Moments ***Généralisée***.

- Le second, qui est utilisé ici, concerne la simplicité avec laquelle elle permet de construire les estimateurs de paramètres de systèmes d'équations.
    - Probable raison de sa très large utilisation en économétrie appliquée, notamment pour les modèles de panels

Principe de la construction des estimateurs MMG « en système »

(*i*) On définit les ***conditions de moment estimantes équation par équation***

(*ii*) On les ***empile en un*** (***grand***) ***vecteur de conditions estimantes***, celui du système « complet ».

(*iii*) Ce dernier s’utilise comme dans le cas d’une seule équation.

A’y’est !

Remarque importante

On utilise des instruments efficaces (au sens de Chamberlain) ou des instruments proches de ces instruments efficaces autant que possible.

Système d’équations de forme générale

On considère des ***systèmes à M équations*** de la forme :

$$\begin{cases} y_{1,i} = f_1(\mathbf{x}_{1,i};\mathbf{a}_0)+u_{1,i} & \text{et } E\left[u_{1,i}/\mathbf{z}_{1,i}\right]=0 \\ \vdots & \vdots \\ y_{M,i} = f_M(\mathbf{x}_{M,i};\mathbf{a}_0)+u_{M,i} & \text{et } E\left[u_{M,i}/\mathbf{z}_{M,i}\right]=0 \end{cases}$$

Ces systèmes peuvent être également notés :

$$\begin{cases} y_{m,i} = f_m(\mathbf{x}_{m,i};\mathbf{a}_0)+u_{m,i} \quad \text{et } E\left[u_{m,i}/\mathbf{z}_{m,i}\right]=0 \\ \qquad\qquad\qquad \text{pour } m=1,\ldots,M \end{cases}$$

Il n’y a pas plus général …
… tant qu’on est dans des ***modèles à termes d’erreur additifs***.

$$\begin{cases} y_{1,i} = f_1(\mathbf{x}_{1,i};\mathbf{a}_0) + u_{1,i} & \text{et } E\left[u_{1,i}/\mathbf{z}_{1,i}\right] = 0 \\ \vdots & \vdots \\ y_{M,i} = f_M(\mathbf{x}_{M,i};\mathbf{a}_0) + u_{M,i} & \text{et } E\left[u_{M,i}/\mathbf{z}_{M,i}\right] = 0 \end{cases}$$

- Les $\mathbf{x}_{m,i}$ peuvent contenir des $y_{\ell,i}$ avec $\ell \neq m$.

- Les $\mathbf{z}_{m,i}$ peuvent contenir des éléments des $\mathbf{x}_{\ell,i}$, …
  - … voire des $y_{\ell,i}$ avec $\ell \neq m$.

- La forme des $f_m(.)$ et le contenu des $\mathbf{z}_{m,i}$ ne sont contraints que par les conditions d'identification de $\mathbf{a}_0$

1. Construire des conditions d'orthogonalité, pour $m = 1, \ldots, M$

$$y_{m,i} = f_m(\mathbf{x}_{m,i};\mathbf{a}_0) + u_{m,i} \quad \text{et} \quad E\left[u_{m,i} / \mathbf{z}_{m,i}\right] = 0$$

$$\Downarrow$$

$$E\left[\mathbf{r}_m(\mathbf{z}_{m,i})u_{m,i}(\mathbf{a}_0)\right] = \mathbf{0} \quad \text{où} \quad u_{m,i}(\mathbf{a}) \equiv y_{m,i} - f_m(\mathbf{x}_{m,i};\mathbf{a})$$

$$\Updownarrow$$

$$E\left[\mathbf{g}_{m,i}(\mathbf{a}_0)\right] = \mathbf{0} \quad \text{où} \quad \mathbf{g}_{m,i}(\mathbf{a}) \equiv \mathbf{r}_m(\mathbf{z}_{m,i})u_{m,i}(\mathbf{a})$$

- Pour assurer l'identification de $\mathbf{a}_0$, s'assurer que (Chamberlain) :

$$\mathbf{r}_m(\mathbf{z}_{m,i}) \text{ est « proche » de } E\left[\frac{\partial f_m(\mathbf{x}_{m,i};\mathbf{a}_0)}{\partial \mathbf{a}} \Big/ \mathbf{z}_{m,i}\right] E\left[u_{m,i}^2 / \mathbf{z}_{m,i}\right]^{-1}$$

- Si on choisit des $\mathbf{r}_m(\mathbf{z}_{m,i};\mathbf{b}_{m,0})$, veiller à ce que les $\mathbf{b}_{m,0}$ puissent être estimés de manière convergente.

Cas particulier important : $\mathbf{z}_{m,i} = \mathbf{z}_i$ pour $m = 1,\ldots,M$

- Définir :

$$\mathbf{f}(\mathbf{x}_i;\mathbf{a}_0) \equiv \begin{bmatrix} f_1(\mathbf{x}_{1,i};\mathbf{a}_0) \\ \vdots \\ f_M(\mathbf{x}_{M,i};\mathbf{a}_0) \end{bmatrix}, \ \ \mathbf{u}_i(\mathbf{a}) \equiv \begin{bmatrix} u_{1,i}(\mathbf{a}) \\ \vdots \\ u_{M,i}(\mathbf{a}) \end{bmatrix} \ \text{ et } \ \mathbf{u}_i \equiv \begin{bmatrix} u_{1,i} \\ \vdots \\ u_{M,i} \end{bmatrix}.$$

- Définir $\mathbf{g}_i(\mathbf{a})$ par :

$$\mathbf{g}_i(\mathbf{a}) \equiv \mathbf{R}(\mathbf{z}_i)\mathbf{u}_i(\mathbf{a}) \quad \text{ou} \quad \mathbf{g}_i(\mathbf{a}) \equiv \mathbf{R}(\mathbf{z}_i;\mathbf{b}_0)\mathbf{u}_i(\mathbf{a}) \text{ avec } \mathbf{b}_0 \text{ identifiable.}$$

- Choisir $\mathbf{R}(\mathbf{z}_i)$ (ou $\mathbf{R}(\mathbf{z}_i;\mathbf{b}_0)$) de telle sorte à ce que (Chamberlain) :

$$\mathbf{R}(\mathbf{z}_i) \text{ est « proche » de } E\left[\left.\frac{\partial \mathbf{f}(\mathbf{x}_i;\mathbf{a}_0)'}{\partial \mathbf{a}}\right/\mathbf{z}_i\right] E\left[\mathbf{u}_i\mathbf{u}_i'/\mathbf{z}_i\right]^{-1}.$$

2. Empiler les $\mathbf{g}_{m,i}(\mathbf{a})$ ***, pour*** $m = 1,\ldots,M$

$$\mathbf{g}_i(\mathbf{a}) \equiv \begin{bmatrix} \mathbf{g}_{1,i}(\mathbf{a}) \\ \vdots \\ \mathbf{g}_{M,i}(\mathbf{a}) \end{bmatrix} \text{ et } E\left[\mathbf{g}_{m,i}(\mathbf{a}_0)\right] = \mathbf{0} \text{ pour } m = 1,\ldots,M$$

$$\Updownarrow$$

$$E\left[\mathbf{g}_i(\mathbf{a}_0)\right] = \mathbf{0}$$

Cas particulier important : $\mathbf{z}_{m,i} = \mathbf{z}_i$ pour $m = 1,\ldots,M$

$$E\left[\mathbf{g}_i(\mathbf{a}_0)\right] \equiv E\left[\mathbf{R}(\mathbf{z}_i)\mathbf{u}_i(\mathbf{a}_0)\right] = \mathbf{0}$$

Si tout se passe bien, alors on a :

$$E\left[\mathbf{g}_i(\mathbf{a})\right] = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{a} = \mathbf{a}_0$$

- Cette condition d'identification globale peut être examinée *ex ante* dans le cas des modèles linéaires.

- C'est plus difficile dans le cas des modèles (à formes fonctionnelles) non-linéaires …

  - … et souvent $E\left[\mathbf{g}_i(\mathbf{a})\right] = \mathbf{0}$ a plusieurs solutions en **a**.

Astuce : Rester dans le cadre de modèles linéaires (ou presque)

- ***Si plusieurs solutions*** (bien identifiées, *i.e.* avec des matrices de variance-covariance « correctes » !).

    - Définir un ensemble $\mathcal{A}$ compact (domaine des valeurs admissibles pour $\mathbf{a}_0$) tel que :

$$E\left[\mathbf{g}_i(\mathbf{a})\right] = \mathbf{0} \text{ et } \mathbf{a} \in \mathcal{A} \quad \Leftrightarrow \quad \mathbf{a} = \mathbf{a}_0.$$

    - Si $E\left[\mathbf{g}_i(\mathbf{a}_0)\right] = \mathbf{0}$ décrit une équation de score (ou une partie d'équation de score), choisir la solution en $\mathbf{a}$ de $\overline{\mathbf{g}}_N(\mathbf{a}) = \mathbf{0}_{K\times 1}$ qui donne la valeur maximum de la vraisemblance de l'échantillon …

      … cette solution correspond à l'estimateur du Maximum de Vraisemblance de $\mathbf{a}_0$.

- On « balaie » l'ensemble des solutions de $\overline{\mathbf{g}}_N(\mathbf{a}) = \mathbf{0}_{K\times 1}$ ou des *minima* de $\overline{\mathbf{g}}_N(\mathbf{a})' \tilde{\mathbf{W}}_N^{-1} \overline{\mathbf{g}}_N(\mathbf{a})$ en lançant la procédure de calcul à partir de différents points initiaux pour $\mathbf{a}$.

3. Exploiter la condition de moment :

$$E\left[\mathbf{g}_i(\mathbf{a})\right] = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{a} = \mathbf{a}_0$$

pour construire un estimateur de la MM(G) de $\mathbf{a}_0$

Voir ce qui précède

Adapter la procédure au modèle considéré …

… dans la suite cas des modèles linéaires.

Procédure : ***Les 3 étapes du calcul de*** $\hat{\mathbf{a}}_N^{MMG}$

Etape 1. On détermine une matrice $\mathbf{M}_0$, aussi proche que possible de $\mathbf{W}_0^{-1}$, dont on sache directement calculer un estimateur, $\tilde{\mathbf{M}}_N \xrightarrow[N\to+\infty]{p} \mathbf{M}_0$.
On calcule alors un estimateur convergent de $\mathbf{a}_0$, $\tilde{\mathbf{a}}_N \equiv \hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N)$:

$$\tilde{\mathbf{a}}_N \equiv \arg\min_{\mathbf{a}} \overline{\mathbf{g}}_N(\mathbf{a})'\tilde{\mathbf{M}}_N \overline{\mathbf{g}}_N(\mathbf{a}).$$

Etape 2. On construit, à partir de $\tilde{\mathbf{a}}_N^{MMG}$, un estimateur de $\mathbf{W}_0$:

$$\tilde{\mathbf{W}}_N \equiv N^{-1}\sum_{i=1}^{N}\mathbf{g}_i(\tilde{\mathbf{a}}_N)\mathbf{g}_i(\tilde{\mathbf{a}}_N)' \xrightarrow[N\to+\infty]{p} \mathbf{W}_0.$$

Etape 3. La matrice $\tilde{\mathbf{W}}_N$ permet alors de calculer un estimateur MMG optimal de $\mathbf{a}_0$:

$$\hat{\mathbf{a}}_N^{MMG} \equiv \arg\min_{\mathbf{a}} \overline{\mathbf{g}}_N(\mathbf{a})'\tilde{\mathbf{W}}_N^{-1}\overline{\mathbf{g}}_N(\mathbf{a})$$

2. Systèmes d'équations linéaires

On considère des systèmes à M équations de la forme :

$$\begin{cases} y_{1,i} = \mathbf{a}'_{1,0}\mathbf{x}_{1,i} + u_{1,i} & \text{et } E\left[u_{1,i} / \mathbf{z}_{1,i}\right] = 0 \\ \vdots & \vdots \\ y_{M,i} = \mathbf{a}'_{M,0}\mathbf{x}_{M,i} + u_{M,i} & \text{et } E\left[u_{M,i} / \mathbf{z}_{M,i}\right] = 0 \end{cases}$$

Ces systèmes peuvent être également notés :

$$\begin{cases} y_{m,i} = \mathbf{a}'_{m,0}\mathbf{x}_{m,i} + u_{m,i} \quad \text{et } E\left[u_{m,i} / \mathbf{z}_{m,i}\right] = 0 \\ \qquad \text{pour } m = 1,\ldots,M \end{cases}$$

$$\begin{cases} y_{1,i} = \mathbf{a}'_{1,0}\mathbf{x}_{1,i} + u_{1,i} & \text{et } E\left[u_{1,i} / \mathbf{z}_{1,i}\right] = 0 \\ \qquad\vdots & \qquad\vdots \\ y_{M,i} = \mathbf{a}'_{M,0}\mathbf{x}_{M,i} + u_{M,i} & \text{et } E\left[u_{M,i} / \mathbf{z}_{M,i}\right] = 0 \end{cases}$$

On note $\mathbf{a}_0$ le vecteur obtenu par empilement des éléments distincts des $\mathbf{a}_{m,0}$ pour $m = 1,...,M$.

On note $\boldsymbol{\alpha}_0$ le vecteur obtenu par empilement des $\mathbf{a}_{m,0}$:

$$\boldsymbol{\alpha}_0 \equiv \begin{bmatrix} \mathbf{a}_{1,0} \\ \vdots \\ \mathbf{a}_{M,0} \end{bmatrix}, \quad \mathbf{a}_0 = \mathbf{s}(\boldsymbol{\alpha}_0) \text{ et } \dim \boldsymbol{\alpha}_0 \equiv K_\alpha = \sum\nolimits_{m=1}^{M} K_m \geq \dim \mathbf{a}_0 \equiv K.$$

$$\begin{cases} y_{1,i} = \mathbf{a}'_{1,0}\mathbf{x}_{1,i} + u_{1,i} & \text{et } E\left[u_{1,i}/\mathbf{z}_{1,i}\right] = 0 \\ \vdots & \vdots \\ y_{M,i} = \mathbf{a}'_{M,0}\mathbf{x}_{M,i} + u_{M,i} & \text{et } E\left[u_{M,i}/\mathbf{z}_{M,i}\right] = 0 \end{cases}$$

Les systèmes considérés ici sont très généraux :

- ***Chaque équation a son vecteur*** de K_m variables explicatives, de K_m paramètres et ***de*** L_m ***variables instrumentales***.
- Les vecteurs de variables explicatives et de variables instrumentales peuvent avoir des ***éléments communs***, dans une et entre les équations.
- Des paramètres peuvent être liés par des contraintes (linéaires) dans une équation ou entre équation.
  - Un paramètre peut être commun à plusieurs équations.

De tels systèmes se rencontrent fréquemment en économétrie appliquée :

- ***Systèmes de régressions empilées***
  - Formes réduites de systèmes d'équations simultanées
  - Systèmes de fonctions de demande de bien, …
- ***Systèmes d'équations simultanées***
- ***Modèles/systèmes d'équations plus spécifiques*** :
  - Modèles de données de panel (statique ou dynamique)
  - CO1 de programmes d'optimisation (CO1 standard en statique, équations d'Euler en dynamique ; …).

Exemple : Système de fonctions de demande de biens

Systèmes estimés pour l'analyse des effets de réformes fiscales :

$$\begin{cases} coef\_budget_{1,it} = d(\mathbf{prix}_{it}, revenu_{it}, \mathbf{c}_{it}; \mathbf{a}_{1,0}) + u_{1,it} \\ coef\_budget_{2,it} = d(\mathbf{prix}_{it}, revenu_{it}, \mathbf{c}_{it}; \mathbf{a}_{2,0}) + u_{2,it} \\ \quad\quad\vdots \quad\quad\quad\quad\quad\quad\quad\quad \vdots \\ coef\_budget_{M,it} = d(\mathbf{prix}_{it}, revenu_{it}, \mathbf{c}_{it}; \mathbf{a}_{M,0}) + u_{M,it} \end{cases}$$

$revenu_{it}$: revenu du ménage i l'année t

$coef\_budget_{m,it}$: part des dépenses de bien m dans le revenu

$\mathbf{prix}_{it}$: vecteur des indices des prix des biens

$\mathbf{c}_{it}$: vecteur de variables décrivant le ménage i l'année t

m : agrégats de biens (alimentation, transport, logement, ...)

Supposer que les $\mathbf{a}_{m,0}$ peuvent avoir des ***éléments communs pour différentes équations*** implique qu'***on sort du simple cadre de l' « empilement » de modèles linéaires***.

- Si les $\mathbf{a}_{m,0}$ ont des éléments communs pour différentes équations, alors un ***ensemble de contraintes*** est implicitement imposé sur les éléments de $\boldsymbol{\alpha}_0$.
- Les estimateurs « classiques », SUR et 3MC, « excluent » ces contraintes.
  - Ceci dit, le système demeure linéaire.

L'intérêt d'une estimation « en système » par rapport à une estimation « équation par équation » est double :

- L'***estimation « en système » est plus efficace as.*** car elle permet de tenir compte des corrélations entre les termes d'erreur des différentes équations du système, *i.e.* les corrélations entre les $u_{m,i}$.
- L'***estimation « en système » est naturelle*** (et parfois nécessaire) ***si les*** $\mathbf{a}_{m,0}$ ***ont des éléments communs pour différentes équations***.

Lorsque le vecteur des paramètres d'intérêt du modèle, $\mathbf{a}_0$, est simplement l'empilement des $\mathbf{a}_{m,0}$, *i.e.* $\mathbf{a}_0 = \boldsymbol{\alpha}_0$, le système est dit « linéaire simple » :

Système linéaire simple

$$\mathbf{y}_i = \mathbf{X}_i'\mathbf{a}_0 + \mathbf{u}_i \quad \text{avec} \quad E[\mathbf{u}_i] = \mathbf{0}$$

avec :

$$\mathbf{a}_0 \equiv \begin{bmatrix} \mathbf{a}_{1,0} \\ \mathbf{a}_{2,0} \\ \vdots \\ \mathbf{a}_{M,0} \end{bmatrix},\ \mathbf{y}_i \equiv \begin{bmatrix} y_{1,i} \\ y_{2,i} \\ \vdots \\ y_{M,i} \end{bmatrix},\ \mathbf{u}_i \equiv \begin{bmatrix} u_{1,i} \\ u_{2,i} \\ \vdots \\ u_{M,i} \end{bmatrix} \text{ et } \mathbf{X}_i \equiv \begin{bmatrix} \mathbf{x}_{1,i} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{2,i} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{x}_{M,i} \end{bmatrix}.$$

La matrice $[\mathbf{X}_i\mathbf{X}_i']$ étant bloc-diagonale, on peut aisément vérifier que :

$$rangE[\mathbf{X}_i\mathbf{X}_i'] = K = \sum_{m=1}^{M} K_m \quad \Leftrightarrow \quad rangE[\mathbf{x}_{m,i}\mathbf{x}_{m,i}'] = K_m \text{ pour } m = 1,\ldots,M\,.$$

Dans le cas où des éléments des $\mathbf{a}_{m,0}$ sont « partagés » par différentes équations du système, $\mathbf{a}_0$ n'est plus l'empilement des $\mathbf{a}_{m,0}$ pour $m = 1,\ldots,M$.

Le vecteur des paramètres d'intérêt $\mathbf{a}_0$ contient alors les éléments distincts des $\mathbf{a}_{m,0}$ pour $m = 1,\ldots,M$.

On peut alors considérer que $\mathbf{a}_0$ est un vecteur de paramètres dits ***contraints***, *i.e.* résultant de l'imposition de contraintes sur un vecteur de paramètres dits ***libres***.

On définit ici ce vecteur de paramètres libres $\boldsymbol{\alpha}_0$ en empilant simplement les vecteurs $\mathbf{a}_{m,0}$ définis pour chacune des équations, bien entendu on a :

$$\dim \boldsymbol{\alpha}_0 = K_\alpha > K = \dim \mathbf{a}_0 .$$

Le vecteur $\boldsymbol{\alpha}_0$ est lié à $\mathbf{a}_0$, le vecteur de paramètres d'intérêt, par une fonction linéaire représentée par une matrice **S** de dimension $K_\alpha \times K$:

$$\boldsymbol{\alpha}_0 \equiv \begin{bmatrix} \boldsymbol{\alpha}_{1,0} \\ \boldsymbol{\alpha}_{2,0} \\ \vdots \\ \boldsymbol{\alpha}_{M,0} \end{bmatrix}_{K_\alpha \times 1} \quad \text{et} \quad \boldsymbol{\alpha}_0 = \mathbf{S}\mathbf{a}_0 = \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_2 \\ \vdots \\ \mathbf{S}_M \end{bmatrix} \mathbf{a}_0 = \begin{bmatrix} \mathbf{a}_{1,0} \\ \mathbf{a}_{2,0} \\ \vdots \\ \mathbf{a}_{M,0} \end{bmatrix} = \begin{bmatrix} \mathbf{S}_1\mathbf{a}_0 \\ \mathbf{S}_2\mathbf{a}_0 \\ \vdots \\ \mathbf{S}_2\mathbf{a}_0 \end{bmatrix}$$

La matrice **S** est une matrice de sélection composée de 0 et de 1. La matrice **S** permet de construire $\boldsymbol{\alpha}_0$ à partir de $\mathbf{a}_0$, en particulier on a :

$$\mathbf{a}_{m,0} = \mathbf{S}_m \mathbf{a}_0 \text{ pour } m = 1, \dots, M \,,$$

i.e. la matrice $\mathbf{S}_m$ $(K_m \times K)$ sélectionne les éléments de $\mathbf{a}_0$ qui se trouvent dans $\mathbf{a}_{m,0}$.

La même technique peut définir $\mathbf{a}_0$ à partir de contraintes linéaires sur $\boldsymbol{\alpha}_0$.

Bien entendu :

$$rang\mathbf{S}_m = \dim \mathbf{a}_{m,0} = K_m \leq \dim \mathbf{a}_0 = K$$

L'intérêt de cette formulation est qu'on peut alors écrire le système d'équations :

$$\begin{cases} y_{m,i} = \mathbf{x}'_{m,i}\mathbf{a}_{m,0} + u_{m,i} \quad \text{et} \quad E\left[u_{m,i} / \mathbf{z}_{m,i}\right] = 0 \\ \qquad\qquad \text{pour } m = 1,\ldots,M \end{cases}$$

sous la forme :

$$\begin{cases} y_{m,i} = \mathbf{x}'_{m,i}\boldsymbol{\alpha}_{m,0} = \mathbf{x}'_{m,i}\mathbf{S}_m\mathbf{a}_0 + u_{m,i} \quad \text{et} \quad E\left[u_{m,i} / \mathbf{z}_{m,i}\right] = 0 \\ \qquad\qquad \text{pour } m = 1,\ldots,M \end{cases}$$

On a alors :

$$\boxed{\mathbf{y}_i = \mathbf{X}'_i\mathbf{S}\mathbf{a}_0 + \mathbf{u}_i \quad \text{avec} \quad E\left[\mathbf{u}_i\right] = \mathbf{0},}$$

Rmq. Il n'est pas toujours possible d'estimer $\mathbf{a}_0$ à partir de chaque équation.

$$\mathbf{y}_i = \mathbf{X}_i'\mathbf{S}\mathbf{a}_0 + \mathbf{u}_i \quad \text{avec} \quad E\left[\mathbf{u}_i\right] = \mathbf{0},$$

i.e. même lorsque des éléments des $\mathbf{a}_{m,0}$ sont « partagés » par différentes équations du système, le système peut s'écrire sous la forme général d'un système linéaire, moyennant la définition de la matrice de sélection **S**.

La matrice **S** est de dimension $K_\alpha \times K$ et de rang *K*. Ceci implique que :

$$rang E\left[\mathbf{S}'\mathbf{X}_i\mathbf{X}_i'\mathbf{S}\right] = rang\left(\mathbf{S}'E\left[\mathbf{X}_i\mathbf{X}_i'\right]\mathbf{S}\right) = K \,.$$

On montrera que l'estimateur MMG optimal de $\mathbf{a}_0$ peut être écrit sous forme explicite, que ce soit dans le cas linéaire ou dans le ***cas linéaire avec égalités inter-équations*** (sur les paramètres).

Rmq. Il est possible que :

(i) $\mathbf{S}'E\left[\mathbf{X}_i\mathbf{X}_i'\right]\mathbf{S}$ soit de rang K et que les $\mathbf{S}_m'E\left[\mathbf{x}_{m,i}\mathbf{x}_{m,i}'\right]\mathbf{S}_m$ soient également tous de rang K. Dans ce cas $\mathbf{a}_0$ peut être identifiable dans chacune des équations du système.

(ii) $\mathbf{S}'E\left[\mathbf{X}_i\mathbf{X}_i'\right]\mathbf{S}$ soit de rang K sans qu'aucun des $\mathbf{S}_m'E\left[\mathbf{x}_{m,i}\mathbf{x}_{m,i}'\right]\mathbf{S}_m$ soit de rang K. Dans ce cas, $\mathbf{a}_0$ n'est identifiable dans aucune équation du système mais peut l'être dans le système.

$$\mathbf{S}'E\left[\mathbf{X}_i\mathbf{X}_i'\right]\mathbf{S} = \begin{bmatrix} \mathbf{S}_1'E\left[\mathbf{x}_{1,i}\mathbf{x}_{1,i}'\right]\mathbf{S}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_2'E\left[\mathbf{x}_{2,i}\mathbf{x}_{2,i}'\right]\mathbf{S}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{S}_M'E\left[\mathbf{x}_{M,i}\mathbf{x}_{M,i}'\right]\mathbf{S}_M \end{bmatrix}$$

3. Estimateurs MMG optimaux dans les systèmes d'équations linéaires

On reprend ici le principe de construction des estimateurs MMG optimaux.

On obtient un estimateur de la MMG de $\mathbf{a}_0$ en :

i. construisant les conditions d'orthogonalité équation par équation,

ii. en empilant les conditions de moment ainsi obtenues en une seule condition de moment

et :

iii. en construisant un estimateur de $\mathbf{a}_0$ à partir de cette condition de moment dans le cadre d'inférence que procure la MMG.

3.1. Construction des conditions d'orthogonalité

Dans tous les cas, les conditions d'orthogonalité sont d'abord construites équation par équation. Ici on choisit :

$$E\left[\mathbf{z}_{m,i}(y_{m,i} - \mathbf{x}'_{m,i}\mathbf{a}_{m,0})\right] = \mathbf{0}_{L_m \times 1} \quad \text{pour} \quad m = 1,\ldots,M$$

ou :

$$E\left[\mathbf{g}_{m,i}(\mathbf{a}_{m,0})\right] = \mathbf{0}_{L_m \times 1} \quad \text{avec} \quad \mathbf{g}_{m,i}(\mathbf{a}_m) \equiv \mathbf{z}_{m,i}(y_{m,i} - \mathbf{x}'_{m,i}\mathbf{a}_m) \quad \text{pour} \quad m = 1,\ldots,M\ .$$

Puis on empile les conditions de moment estimantes obtenues pour obtenir :

$$E\left[\mathbf{g}_i(\mathbf{a}_0)\right] = \mathbf{0}_{L \times 1} \quad \text{avec} \quad \mathbf{g}_i(\mathbf{a}) \equiv \begin{bmatrix} \mathbf{g}_{1,i}(\mathbf{a}_2) \\ \vdots \\ \mathbf{g}_{M,i}(\mathbf{a}_M) \end{bmatrix}_{L \times 1}$$

dans le cas général. Dans les cas linéaire et linéaire « avec égalités inter-équations », on a :

Condition de moment, cas du système linéaire simple

$$E\left[\mathbf{Z}_i(\mathbf{y}_i - \mathbf{X}_i'\mathbf{a}_0)\right] = \mathbf{0}_{L\times 1}$$

et :

Condition de moment, cas du système linéaire avec égalités inter-équations

$$E\left[\mathbf{Z}_i(\mathbf{y}_i - \mathbf{X}_i'\mathbf{S}\mathbf{a}_0)\right] = \mathbf{0}_{L\times 1}$$

avec :

$$\mathbf{Z}_i \equiv \begin{bmatrix} \mathbf{z}_{1,i} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{2,i} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{z}_{M,i} \end{bmatrix}.$$

3.2. Examen de la forme de $\mathbf{W}_0$

Dans le cas général on a :

$$\mathbf{W}_0 \equiv E\left[\mathbf{g}_i(\mathbf{a}_0)\mathbf{g}_i(\mathbf{a}_0)'\right]$$

et :

$$\mathbf{W}_0 \equiv E\left[\mathbf{Z}_i \mathbf{u}_i \mathbf{u}_i' \mathbf{Z}_i'\right]$$

dans les cas linéaires, avec ou sans égalités inter-équations dans les paramètres.

Dans ces derniers cas, l'expression de se simplifie sous l'hypothèse d'homoscédasticité des $\mathbf{u}_i$, *i.e.* si :

$$E\left[\mathbf{u}_i \mathbf{u}_i' / \mathbf{z}_i\right] = E\left[\mathbf{u}_i \mathbf{u}_i'\right] \equiv \mathbf{\Omega}_0$$

où $\mathbf{z}_i$ est le vecteur contenant les éléments distincts des $\mathbf{z}_{m,i}$ pour $m = 1, \ldots, M$.

Remarque. ***Forme de*** $\mathbf{W}_0 \equiv E\left[\mathbf{Z}_i \mathbf{u}_i \mathbf{u}_i' \mathbf{Z}_i'\right]$

$$E\left[\mathbf{Z}_i \mathbf{u}_i \mathbf{u}_i' \mathbf{Z}_i'\right] = \begin{bmatrix} E\left[\mathbf{z}_{1,i} u_{1,i}^2 \mathbf{z}_{1,i}'\right] & E\left[\mathbf{z}_{1,i} u_{1,i} u_{2,i} \mathbf{z}_{2,i}'\right] & \cdots & E\left[\mathbf{z}_{1,i} u_{1,i} u_{M,i} \mathbf{z}_{M,i}'\right] \\ E\left[\mathbf{z}_{2,i} u_{2,i} u_{1,i} \mathbf{z}_{1,i}'\right] & E\left[\mathbf{z}_{2,i} u_{2,i}^2 \mathbf{z}_{2,i}'\right] & \cdots & E\left[\mathbf{z}_{2,i} u_{2,i} u_{M,i} \mathbf{z}_{M,i}'\right] \\ \vdots & \vdots & \ddots & \mathbf{0} \\ E\left[\mathbf{z}_{M,i} u_{M,i} u_{1,i} \mathbf{z}_{1,i}'\right] & E\left[\mathbf{z}_{M,i} u_{M,i} u_{2,i} \mathbf{z}_{2,i}'\right] & \cdots & E\left[\mathbf{z}_{M,i} u_{M,i}^2 \mathbf{z}_{M,i}'\right] \end{bmatrix}$$

- Les corrélations entre les éléments de $\mathbf{u}_i$ font que $\mathbf{W}_0 \equiv E\left[\mathbf{Z}_i \mathbf{u}_i \mathbf{u}_i' \mathbf{Z}_i'\right]$ est très rarement bloc-diagonale.

- L'hypothèse d'homoscédasticité de $\mathbf{u}_i$ apparaît relativement « étrange » si les vecteurs d'instruments, les $\mathbf{z}_{m,i}$, diffèrent d'une équation à l'autre.

- Si $E\left[u_{\ell,i}/\mathbf{z}_{m,i}\right]=\eta_\ell(\mathbf{z}_{m,i})\neq 0$, *i.e*. si $\mathbf{z}_{m,i}$ est (au moins partiellement) endogène par rapport à $u_{\ell,i}$ alors il est vraisemblable que :

$$E\left[\mathbf{z}_{m,i}u_{m,i}u_{\ell,i}\mathbf{z}'_{\ell,i}\right]\neq \omega_{m\ell}E\left[\mathbf{z}_{m,i}\mathbf{z}'_{\ell,i}\right].$$

- L'homoscédasticité de $\mathbf{u}_i$ est plus vraisemblable si $\mathbf{z}_{m,i}=\mathbf{z}_i$ pour $m=1,\ldots,M$.

- Pour cette dernière raison, nous considérerons le ***cas hétéroscédastique*** pratiquement tout au long de ce chapitre,
  - Sauf pour présenter deux estimateurs classiques : l'estimateur SUR et l'estimateur des 3MC.

$$\boxed{\mathbf{W}_0 \equiv E\left[\mathbf{Z}_i \mathbf{u}_i \mathbf{u}_i' \mathbf{Z}_i'\right]}$$

- On a :

$$\mathbf{W}_0 \equiv E\left[\mathbf{Z}_i \mathbf{u}_i \mathbf{u}_i' \mathbf{Z}'\right] = E\left[\mathbf{Z}_i (\mathbf{y}_i - \mathbf{X}_i' \mathbf{S} \mathbf{a}_0)(\mathbf{y}_i - \mathbf{X}_i' \mathbf{S} \mathbf{a}_0)' \mathbf{Z}_i'\right],$$

aussi on pourra estimer $\mathbf{W}_0 \equiv E\left[\mathbf{Z}_i \mathbf{u}_i \mathbf{u}_i' \mathbf{Z}_i'\right]$ si on dispose d'un estimateur convergent de $\mathbf{a}_0$, $\tilde{\mathbf{a}}_N$, avec :

$$\boxed{\tilde{\mathbf{W}}_N \equiv N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i (\mathbf{y}_i - \mathbf{X}_i' \mathbf{S} \tilde{\mathbf{a}}_N)(\mathbf{y}_i - \mathbf{X}_i' \mathbf{S} \tilde{\mathbf{a}}_N)' \mathbf{Z}_i' \xrightarrow[N \to +\infty]{p} \mathbf{W}_0}$$

- La matrice la plus proche de $E\left[\mathbf{Z}_i \mathbf{u}_i \mathbf{u}_i' \mathbf{Z}_i'\right]$ dont on sache calculer directement un estimateur convergent est $E\left[\mathbf{Z}_i \mathbf{Z}_i'\right]$ puisque :

$$N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i \mathbf{Z}_i' \xrightarrow[N \to +\infty]{p} E\left[\mathbf{Z}_i \mathbf{Z}_i'\right].$$

3.3. Estimateurs MMG optimaux des systèmes linéaires

Dans le cas général, les estimateurs MMG optimaux de $\mathbf{a}_0$ sont de la forme :

$$\hat{\mathbf{a}}_N^{MMG} \equiv \arg\min_{\mathbf{a}} \left(N^{-1} \sum_{i=1}^{N} \mathbf{g}(\mathbf{w}_i;\mathbf{a})' \right) \tilde{\mathbf{W}}_N^{-1} \left(N^{-1} \sum_{i=1}^{N} \mathbf{g}(\mathbf{w}_i;\mathbf{a}) \right)$$

avec :

$$\tilde{\mathbf{W}}_N \xrightarrow[N \to +\infty]{p} E\left[\mathbf{g}_i(\mathbf{a}_0) \mathbf{g}_i(\mathbf{a}_0)' \right].$$

Dans le cas linéaire, on a :

$$\boxed{\hat{\mathbf{a}}_N^{MMG} \equiv \arg\min_{\mathbf{a}} \left(N^{-1} \sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{X}_i'\mathbf{S}\mathbf{a})'\mathbf{Z}_i' \right) \tilde{\mathbf{W}}_N^{-1} \left(N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i(\mathbf{y}_i - \mathbf{X}_i'\mathbf{S}\mathbf{a}) \right)}$$

avec :

$$\tilde{\mathbf{W}}_N \xrightarrow[N \to +\infty]{p} E\left[\mathbf{Z}_i \mathbf{u}_i \mathbf{u}_i' \mathbf{Z}' \right] = E\left[\mathbf{Z}_i (\mathbf{y}_i - \mathbf{X}_i'\mathbf{S}\mathbf{a}_0)(\mathbf{y}_i - \mathbf{X}_i'\mathbf{S}\mathbf{a}_0)'\mathbf{Z}_i' \right]$$

et :

$\mathbf{S} = \mathbf{I}_K$ dans le cas linéaire simple.

Les CO1 de :

$$\min_{\mathbf{a}} \left(N^{-1} \sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{X}_i'\mathbf{S}\mathbf{a})'\mathbf{Z}_i' \right) \tilde{\mathbf{W}}_N^{-1} \left(N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i (\mathbf{y}_i - \mathbf{X}_i'\mathbf{S}\mathbf{a}) \right)$$

sont données par :

$$-2 \times \left(N^{-1} \sum_{i=1}^{N} \mathbf{S}'\mathbf{X}_i \mathbf{Z}_i' \right) \tilde{\mathbf{W}}_N^{-1} \left(N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i (\mathbf{y}_i - \mathbf{X}_i'\mathbf{S}\hat{\mathbf{a}}_N^{MMG}) \right) = \mathbf{0}_{K\times 1}.$$

On en déduit que :

$$\boxed{\begin{aligned} \hat{\mathbf{a}}_N^{MMG} &= \left[\left(N^{-1} \sum_{i=1}^{N} \mathbf{S}'\mathbf{X}_i \mathbf{Z}_i' \right) \tilde{\mathbf{W}}_N^{-1} \left(N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i \mathbf{X}_i'\mathbf{S} \right) \right]^{-1} \\ &\quad \times \left(N^{-1} \sum_{i=1}^{N} \mathbf{S}'\mathbf{X}_i \mathbf{Z}_i' \right) \tilde{\mathbf{W}}_N^{-1} N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i \mathbf{y}_i \end{aligned}}$$

On retrouve ici la ***structure d'un estimateur des 2MC***.

Propriété. ***Estimateurs MMG optimaux de $\mathbf{a}_0$ dans les systèmes linéaires***

Soit un système d'équations linéaires tel que :

$$\mathbf{y}_i = \mathbf{X}_i'\mathbf{S}\mathbf{a}_0 + \mathbf{u}_i \quad \text{avec} \quad E[\mathbf{u}_i] = \mathbf{0}$$

et :

$$E[\mathbf{Z}_i(\mathbf{y}_i - \mathbf{X}_i'\mathbf{S}\mathbf{a}_0)] = \mathbf{0}_{L\times 1} \text{ avec } rang E[\mathbf{Z}_i\mathbf{X}_i'\mathbf{S}] = K.$$

Alors les estimateurs de la MMG optimaux de $\mathbf{a}_0$ sont définis par :

$$\hat{\mathbf{a}}_N^{MMG} = \left[\left(N^{-1}\sum_{i=1}^{N}\mathbf{S}'\mathbf{X}_i\mathbf{Z}_i'\right)\tilde{\mathbf{W}}_N^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i\mathbf{X}_i'\mathbf{S}\right)\right]^{-1} \times\left(N^{-1}\sum_{i=1}^{N}\mathbf{S}'\mathbf{X}_i\mathbf{Z}_i'\right)\tilde{\mathbf{W}}_N^{-1}N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i\mathbf{y}_i$$

où :

$$\tilde{\mathbf{W}}_N \xrightarrow[N\to+\infty]{p} E[\mathbf{Z}_i\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}'] = E[\mathbf{Z}_i(\mathbf{y}_i - \mathbf{X}_i'\mathbf{S}\mathbf{a}_0)(\mathbf{y}_i - \mathbf{X}_i'\mathbf{S}\mathbf{a}_0)'\mathbf{Z}_i'].$$

La matrice $\tilde{\mathbf{W}}_N$ peut être calculée à partir des deux étapes suivantes :

***Etape 1* : *Calcul de* $\tilde{\mathbf{a}}_N$**, estimateur MMG non optimal de $\mathbf{a}_0$

$$\tilde{\mathbf{a}}_N = \left[\left(N^{-1} \sum_{i=1}^{N} \mathbf{S}'\mathbf{X}_i\mathbf{Z}_i' \right) \left(N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i\mathbf{Z}_i' \right)^{-1} \left(N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i\mathbf{X}_i'\mathbf{S} \right) \right]^{-1}$$

$$\times \left(N^{-1} \sum_{i=1}^{N} \mathbf{S}'\mathbf{X}_i\mathbf{Z}_i' \right) \left(N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i\mathbf{Z}_i' \right)^{-1} N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i\mathbf{y}_i$$

***Etape 2* : *Calcul de* $\tilde{\mathbf{W}}_N$**

$$\tilde{\mathbf{W}}_N \equiv N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i\tilde{\mathbf{u}}_{N,i}\tilde{\mathbf{u}}_{N,i}'\mathbf{Z}_i' \xrightarrow[N\to+\infty]{p} E\left[\mathbf{Z}_i\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i'\right]$$

avec :

$$\tilde{\mathbf{u}}_{N,i} \equiv \mathbf{y}_i - \mathbf{X}_i'\mathbf{S}\tilde{\mathbf{a}}_N \xrightarrow[N\to+\infty]{p} \mathbf{u}_i$$

Etape 3 : Calcul de $\hat{\mathbf{a}}_N^{MMG}$

$$\hat{\mathbf{a}}_N^{MMG} = \left[\left(N^{-1}\sum_{i=1}^{N}\mathbf{S}'\mathbf{X}_i\mathbf{Z}_i'\right)\tilde{\mathbf{W}}_N^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i\mathbf{X}_i'\mathbf{S}\right)\right]^{-1}$$

$$\times\left(N^{-1}\sum_{i=1}^{N}\mathbf{S}'\mathbf{X}_i\mathbf{Z}_i'\right)\tilde{\mathbf{W}}_N^{-1}N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i\mathbf{y}_i$$

avec :

$$\tilde{\mathbf{u}}_{N,i} \equiv \mathbf{y}_i - \mathbf{X}_i'\mathbf{S}\tilde{\mathbf{a}}_N \xrightarrow[N\to+\infty]{p} \mathbf{u}_i$$

D'après les propriétés des estimateurs MMG optimaux on sait :

$$\sqrt{N}(\hat{\mathbf{a}}_N^{MMG} - \mathbf{a}_0) \xrightarrow[N \to +\infty]{L} \mathcal{N}(\mathbf{0}; \mathbf{\Sigma}_0)$$

avec :

$$\mathbf{\Sigma}_0 \equiv \left(E\left[\mathbf{S}'\mathbf{X}_i\mathbf{Z}_i'\right] E\left[\mathbf{Z}_i\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i'\right]^{-1} E\left[\mathbf{Z}_i\mathbf{X}_i'\mathbf{S}\right] \right)^{-1}.$$

et :

$$\hat{\mathbf{\Sigma}}_N \equiv \left[\left(N^{-1} \sum_{i=1}^N \mathbf{S}'\mathbf{X}_i\mathbf{Z}_i' \right) \hat{\mathbf{W}}_N^{-1} \left(N^{-1} \sum_{i=1}^N \mathbf{Z}_i\mathbf{X}_i'\mathbf{S} \right) \right]^{-1} \xrightarrow[N \to +\infty]{p} \mathbf{\Sigma}_0$$

avec :

$$N^{-1} \sum_{i=1}^N \mathbf{S}'\mathbf{X}_i\mathbf{Z}_i' \xrightarrow[N \to +\infty]{p} E\left[\mathbf{S}'\mathbf{X}_i\mathbf{Z}'\right]$$

et :

$$\hat{\mathbf{W}}_N \equiv N^{-1} \sum_{i=1}^N \mathbf{Z}_i\hat{\mathbf{u}}_{N,i}\hat{\mathbf{u}}_{N,i}'\mathbf{Z}_i' \xrightarrow[N \to +\infty]{p} E\left[\mathbf{Z}_i\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i'\right] \text{ où } \hat{\mathbf{u}}_{N,i} \equiv \mathbf{y}_i - \mathbf{X}_i'\mathbf{S}\hat{\mathbf{a}}_N^{MMG}.$$

Rmq. On recalcule un estimateur de $\mathbf{W}_0 \equiv E\left[\mathbf{Z}_i\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i'\right]$, $\hat{\mathbf{W}}_N$, à partir de $\hat{\mathbf{a}}_N^{MMG}$.

Cet estimateur est (en général) plus efficace que $\tilde{\mathbf{a}}_N$.

4. Estimation « en système » *versus* « équation par équation »

Le premier cas particulier important est celui du système linéaire sans contraintes inter-équations sur les paramètres. Ce système est caractérisé par :

$$\boxed{\mathbf{y}_i = \mathbf{X}_i'\mathbf{a}_0 + \mathbf{u}_i \quad \text{avec} \quad E\left[\mathbf{u}_i\right] = \mathbf{0} \quad \text{et} \quad E\left[\mathbf{Z}_i(\mathbf{y}_i - \mathbf{X}_i'\mathbf{a}_0)\right] = \mathbf{0}_{L\times 1}}$$

Ces systèmes sont importants pour deux raisons :

- Ils permettent de ***comparer les logiques*** d'estimation « en système » et « équation par équation », cette dernière n'ayant pas grand intérêt lorsqu'il existe des égalités inter-équations de paramètres.

- ***Deux systèmes d'équations linéaires classiques***, analysés dans la suite, sont des cas particuliers de ces systèmes : les ***systèmes d'équations simultanées*** « standards » et les ***systèmes de régression empilées*** « standards ».

L'estimateur MMG de $\mathbf{a}_0$ dans ces systèmes est donné par :

$$\hat{\mathbf{a}}_N^{MMG} \equiv \arg\min_{\mathbf{a}} \left(N^{-1} \sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{X}_i'\mathbf{a})'\mathbf{Z}_i' \right) \tilde{\mathbf{W}}_N^{-1} \left(N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i (\mathbf{y}_i - \mathbf{X}_i'\mathbf{a}) \right)$$

avec :

$$\tilde{\mathbf{W}}_N \xrightarrow[N\to+\infty]{p} E\left[\mathbf{Z}_i\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}'\right],$$

c'est un ***estimateur « en système »***.

- Il exploite ***conjointement*** l'information contenue dans l'ensemble des données et dans ***l'ensemble des équations du système***.

- En particulier, il tient compte des ***corrélations potentielles entre les éléments de*** $\mathbf{u}_i$ *via* $\tilde{\mathbf{W}}_N \xrightarrow[N\to+\infty]{p} E\left[\mathbf{Z}_i\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}'\right]$

En revanche, l'estimateur utilisé pour le calcul de $\tilde{\mathbf{W}}_N$:

$$\boxed{\begin{aligned}\tilde{\mathbf{a}}_N = &\left[\left(N^{-1}\sum_{i=1}^{N}\mathbf{S}'\mathbf{X}_i\mathbf{Z}_i'\right)\left(N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i\mathbf{Z}_i'\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i\mathbf{X}_i'\mathbf{S}\right)\right]^{-1}\\ &\times\left(N^{-1}\sum_{i=1}^{N}\mathbf{S}'\mathbf{X}_i\mathbf{Z}_i'\right)\left(N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i\mathbf{Z}_i'\right)^{-1}N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i\mathbf{y}_i\end{aligned}}$$

est un estimateur « équation par équation » ***lorsque*** $\mathbf{S} = \mathbf{I}_K$.

En effet, on montre que $\tilde{\mathbf{a}}_N$ est l'empilement des estimateurs des 2MC des $\mathbf{a}_{m,0}$:

$$\begin{aligned}\hat{\mathbf{a}}_{m,N}^{2MC} = &\left[\left(N^{-1}\sum_{i=1}^{N}\mathbf{x}_{m,i}\mathbf{z}_{m,i}'\right)\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_{m,i}\mathbf{z}_{m,i}'\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_{m,i}\mathbf{x}_{m,i}'\right)\right]^{-1}\\ &\times\left(N^{-1}\sum_{i=1}^{N}\mathbf{x}_{m,i}\mathbf{z}_{m,i}'\right)\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_{m,i}\mathbf{z}_{m,i}'\right)^{-1}N^{-1}\sum_{i=1}^{N}\mathbf{z}_{m,i}y_{m,i}\end{aligned}$$

pour $m = 1,\ldots,M$, en utilisant :

$$N^{-1}\sum_{i=1}^{N}\mathbf{X}_i\mathbf{Z}_i' \equiv \begin{bmatrix} N^{-1}\sum_{i=1}^{N}\mathbf{x}_{1,i}\mathbf{z}_{1,i}' & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & N^{-1}\sum_{i=1}^{N}\mathbf{x}_{M,i}\mathbf{z}_{M,i}' \end{bmatrix},$$

$$N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i\mathbf{y}_i \equiv \begin{bmatrix} N^{-1}\sum_{i=1}^{N}\mathbf{z}_{1,i}y_{1,i} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & N^{-1}\sum_{i=1}^{N}\mathbf{z}_{M,i}y_{M,i} \end{bmatrix}$$

et :

$$\left[N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i\mathbf{Z}_i'\right]^{-1} \equiv \begin{bmatrix} \left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_{1,i}\mathbf{z}_{1,i}'\right]^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_{M,i}\mathbf{z}_{M,i}'\right]^{-1} \end{bmatrix}.$$

L'estimateur $\tilde{\mathbf{a}}_N$ n'est pas l'estimateur MMG « équation par équation » optimal de $\mathbf{a}_0$. De fait, en présence d'hétéroscédasticité potentielle des $\mathbf{u}_i$, l'estimateur MMG « équation par équation » optimal de $\mathbf{a}_0$ est $\hat{\mathbf{a}}_{\mathbf{E},N}^{MMG}$:

Estimateur MMG « équation par équation » optimal de $\mathbf{a}_0$

$$\hat{\mathbf{a}}_{\mathbf{E},N}^{MMG} \equiv \begin{bmatrix} \hat{\mathbf{a}}_{1,N}^{2MCH} \\ \vdots \\ \hat{\mathbf{a}}_{M,N}^{2MCH} \end{bmatrix}$$

où :

$$\hat{\mathbf{a}}_{m,N}^{2MCH} = \left[\left(N^{-1} \sum_{i=1}^{N} \mathbf{x}_{m,i} \mathbf{z}'_{m,i} \right) \tilde{\mathbf{W}}_{m,N}^{-1} \left(N^{-1} \sum_{i=1}^{N} \mathbf{z}_{m,i} \mathbf{x}'_{m,i} \right) \right]^{-1} \times \left(N^{-1} \sum_{i=1}^{N} \mathbf{x}_{m,i} \mathbf{z}'_{m,i} \right) \tilde{\mathbf{W}}_{m,N}^{-1} N^{-1} \sum_{i=1}^{N} \mathbf{z}_{m,i} y_{m,i}$$

avec :

$$\tilde{\mathbf{W}}_{m,N} \equiv N^{-1} \sum_{i=1}^{N} \tilde{u}_{N,m,i}^2 \mathbf{z}_{m,i} \mathbf{z}'_{m,i} \xrightarrow[N \to +\infty]{p} E\left[u_{m,i}^2 \mathbf{z}_{m,i} \mathbf{z}'_{m,i} \right] \text{ et } \tilde{u}_{N,i} \equiv y_i - \mathbf{x}'_i \mathbf{a}_N^{2MC}.$$

Propriété. ***Estimation « en système » et « équation par équation »***

(i) L'estimateur « en système » $\hat{\mathbf{a}}_N^{MMG}$, ***est égal*** à l'estimateur « équation par équation $\hat{\mathbf{a}}_{\mathbf{E},N}^{MMG}$, *i.e.* $\hat{\mathbf{a}}_{\mathbf{E},N}^{MMG} = \hat{\mathbf{a}}_N^{MMG}$, si $\dim \mathbf{z}_{m,i} = \dim \mathbf{a}_{m,0}$ pour $m = 1,\ldots,M$. C'est-à-dire si chaque $\mathbf{a}_{m,0}$ est juste-identifié dans son équation.

(ii) On a : $\hat{\mathbf{a}}_N^{MMG} - \hat{\mathbf{a}}_{\mathbf{E},N}^{MMG} \xrightarrow[N\to+\infty]{p} \mathbf{0}_{K\times 1}$, *i.e.* les estimateurs $\hat{\mathbf{a}}_N^{MMG}$ et $\hat{\mathbf{a}}_{\mathbf{E},N}^{MMG}$ sont ***as. équivalents***, si :

$$E\left[\mathbf{Z}_i\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i'\right] = \begin{bmatrix} E\left[\mathbf{z}_{1,i}u_{1,i}^2\mathbf{z}_{1,i}'\right] & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & E\left[\mathbf{z}_{2,i}u_{2,i}^2\mathbf{z}_{2,i}'\right] & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & E\left[\mathbf{z}_{M,i}u_{M,i}^2\mathbf{z}_{M,i}'\right] \end{bmatrix}$$

(iii) Dans les autres cas $\hat{\mathbf{a}}_N^{MMG}$ est as. plus efficace que $\hat{\mathbf{a}}_{\mathbf{E},N}^{MMG}$.

Condition (i).

Les estimateurs $\hat{\mathbf{a}}_N^{MMG}$ et $\hat{\mathbf{a}}_{\mathbf{E},N}^{MMG}$ sont des estimateurs MMG qui ne diffèrent que par la matrice de pondération du critère MMG, *i.e.* avec l'inverse de $\tilde{\mathbf{W}}_N$ pour $\hat{\mathbf{a}}_N^{MMG}$ Pour $\hat{\mathbf{a}}_{\mathbf{E},N}^{MMG}$ on a :

$$\hat{\mathbf{a}}_{\mathbf{E},N}^{MMG} = \left[\left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i\mathbf{Z}_i'\right)\tilde{\mathbf{W}}_{E,N}^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i\mathbf{X}_i'\right)\right]^{-1} \times\left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i\mathbf{Z}_i'\right)\tilde{\mathbf{W}}_{E,N}^{-1}N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i\mathbf{y}_i$$

avec :

$$\tilde{\mathbf{W}}_{\mathbf{E},N} = \begin{bmatrix} \tilde{\mathbf{W}}_{1,N} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{W}}_{2,N} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \tilde{\mathbf{W}}_{M,N} \end{bmatrix}.$$

Dans les cas juste-identifié, ces matrices n’ont aucun effet. En fait, on a dans ce cas :

$$\hat{\mathbf{a}}_N^{MMG} = \hat{\mathbf{a}}_{\mathbf{E},N}^{MMG} = \left(N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i\mathbf{X}_i'\right)^{-1} N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i\mathbf{y}_i\,,$$

i.e. $\hat{\mathbf{a}}_N^{MMG}$ et $\hat{\mathbf{a}}_{\mathbf{E},N}^{MMG}$ sont des estimateurs de la MM (« non généralisée »).

De fait, les estimateurs $\hat{\mathbf{a}}_N^{MMG}$ et $\hat{\mathbf{a}}_{\mathbf{E},N}^{MMG}$ sont obtenus comme l’empilement des estimateurs des VI des $\mathbf{a}_{m,0}$ pour $m = 1,\ldots,M$:

$$\hat{\mathbf{a}}_{m,N}^{VI} = \left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_{m,i}\mathbf{x}_{m,i}'\right)^{-1} N^{-1}\sum_{i=1}^{N}\mathbf{z}_{m,i}y_{m,i}\,.$$

Condition (*ii*).

Les estimateurs $\hat{\mathbf{a}}_N^{MMG}$ et $\hat{\mathbf{a}}_{\mathbf{E},N}^{MMG}$ ont la même matrice de variance covariance as. lorsque :

$$\mathbf{W}_0 \equiv E\left[\mathbf{Z}_i\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i'\right] = \begin{bmatrix} \mathbf{W}_{1,0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_{2,0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{W}_{M,0} \end{bmatrix}$$

où $\mathbf{W}_{m,0} \equiv p\lim_{N\to+\infty} \tilde{\mathbf{W}}_{m,N}$ pour $m = 1,\ldots,M$ en effet dans ce cas :

$$p\lim_{N\to+\infty} \tilde{\mathbf{W}}_{\mathbf{E},N} = p\lim_{N\to+\infty} \tilde{\mathbf{W}}_N = \mathbf{W}_0 .$$

Rmq. Cette condition (*ii*) se rencontre de manière ***exceptionnelle en pratique***.

Condition (iii).

Dans tous les autres cas, l'estimateur $\hat{\mathbf{a}}_N^{MMG}$ est as. plus efficace que $\hat{\mathbf{a}}_{\mathbf{E},N}^{MMG}$ car $\hat{\mathbf{a}}_{\mathbf{E},N}^{MMG}$ n'est pas un estimateur MMG optimal de $\mathbf{a}_0$:

(a) $\mathbf{a}_0$ est sur-identifié par les conditions de moment utilisé, ce qui implique que les matrices de pondération utilisées pour le calcul des estimateurs de $\mathbf{a}_0$ affectent leur distribution as.

(b) $\hat{\mathbf{a}}_{\mathbf{E},N}^{MMG}$ est calculé à partir d'une matrice de pondération, $\tilde{\mathbf{W}}_{\mathbf{E},N}$, bloc-diagonale alors que $\mathbf{W}_0 \equiv E\left[\mathbf{Z}_i \mathbf{u}_i \mathbf{u}_i' \mathbf{Z}_i'\right]$ n'est pas bloc-diagonale. $\hat{\mathbf{a}}_{\mathbf{E},N}^{MMG}$ ne peut donc être optimal au sens de la MMG.

et :

(c) $\hat{\mathbf{a}}_N^{MMG}$ est optimal par construction.

Remarques : « système » *vs* « équation par équation »

- Estimation « en système » ***plus efficace*** que « équation par équation »
  - Parce que l'estimation « en système » tient compte de la corrélation potentielle (et réelle dans l'immense majorité des cas rencontrés en pratique) des éléments de $\mathbf{u}_i$.

- Les estimateurs « en système » exploitent, *via* le calcul du critère utilisé par la MMG $Q_N(\mathbf{a};\tilde{\mathbf{W}}_N^{-1})$, l'***ensemble des conditions de moment*** déterminées à partir des équations du système pour construire un estimateur optimal de chacun des $\mathbf{a}_{m,0}$
  - Si une seule équation du système, disons *m*, est incorrecte alors $\hat{\mathbf{a}}_N^{MMG}$ peut être as. biaisé pour l'ensemble des éléments de $\mathbf{a}_0$. $\hat{\mathbf{a}}_{\mathbf{E},N}^{MMG}$ n'est biaisé que pour $\hat{\mathbf{a}}_{m,N}^{2MCH}$.

Arbitrage habituel : efficacité (+ risqué) ***versus*** **robustesse** (- précis)

Remarques

Sur-identification : contraintes sur les paramètres et hétéroscédasticité

- Lorsqu'on choisit les :

$$\mathbf{g}_{m,i}(\mathbf{a}_m) \equiv \mathbf{z}_{m,i}(y_{m,i} - \mathbf{x}'_{m,i}\mathbf{a}_m) \text{ pour } m = 1,\ldots,M$$

il convient de faire attention lorsque les éléments des $\mathbf{a}_{m,0}$ sont liés par des contraintes inter-équations.

- Plus les $\mathbf{a}_{m,0}$ sont liés par des contraintes inter-équations et plus la condition de moment :

$$E\left[\mathbf{g}_i(\mathbf{a}_0)\right] = \mathbf{0}_{G\times 1}$$

sur-identifie $\mathbf{a}_0$.

  - Cette sur-identification peut générer des biais d'estimation importants comme le suggèrent les études sur les 2MC.

- On peut essayer de limiter la sur-identification de $\mathbf{a}_0$ en jouant sur le fait que certains de des éléments de $\mathbf{a}_0$:
    - Sont « naturellement » identifiés dans certaines équations du système…
    - … alors qu'ils sont simplement utilisés dans d'autres, ces autres équations permettant d'identifier d'autres éléments de $\mathbf{a}_0$.

- On peut essayer de limiter la sur-identification de $\mathbf{a}_0$ en utilisant des instruments inspirés des instruments optimaux au sens de Chamberlain.

- Autrement, lorsque l'estimation équation par équation est possible (et simple) il peut être préférable :
    - D'estimer une version sans contrainte de $\mathbf{a}_0$, *i.e.* $\boldsymbol{\alpha}_0$,
    - Puis d'imposer les contraintes sur $\boldsymbol{\alpha}_0$ pour obtenir $\mathbf{a}_0$ (voir le Chapitre 28).

- Un des avantages de la juste-identification est qu'elle permet d'éviter le calcul de $\tilde{\mathbf{W}}_N$.

- Autrement, il convient d'être prudent dans l'utilisation des estimateurs robustes à l'hétéroscédasticité de $\mathbf{a}_0$ lorsque ce paramètre est sur-identifié par les conditions de moment employées.

    - Cette robustesse à l'hétéroscédasticité peut générer des biais d'estimation importants comme le suggèrent les études sur les 2MCH.

    - Il est peut être préférable d'utiliser des versions simples des estimateurs en système, *i.e.* utilisant des estimateurs de $E\left[\mathbf{Z}_i E\left[\mathbf{u}_i \mathbf{u}_i'\right]\mathbf{Z}_i'\right]^{-1}$ pour matrice de pondération plutôt que des estimateurs de $\mathbf{W}_0^{-1} \equiv E\left[\mathbf{Z}_i \mathbf{u}_i \mathbf{u}_i' \mathbf{Z}_i'\right]^{-1}$ (***mais c'est juste une intuition***).

5. Equations simultanées et estimateur des 3MC

Les ***systèmes d'équations simultanées (linéaires) standards*** sont définis par :

$$\begin{cases} y_{1,i} = \mathbf{a}'_{1,0}\mathbf{x}_{1,i} + u_{1,i} & \text{et } E\left[u_{1,i}/\mathbf{z}_i\right] = 0 \\ y_{2,i} = \mathbf{a}'_{2,0}\mathbf{x}_{2,i} + u_{2,i} & \text{et } E\left[u_{2,i}/\mathbf{z}_i\right] = 0 \\ \quad\vdots & \quad\vdots \\ y_{M,i} = \mathbf{a}'_{M,0}\mathbf{x}_{M,i} + u_{M,i} & \text{et } E\left[u_{M,i}/\mathbf{z}_i\right] = 0 \end{cases} \quad \text{et } \; E\left[\mathbf{u}_i\mathbf{u}'_i/\mathbf{z}_i\right] = \mathbf{\Omega}_0$$

Par rapport aux cas traités ci-dessus, ils ont ***trois caractéristiques*** :

- Les paramètres des différentes équations ne sont pas liés.
- Toutes les équations du système ont le ***même vecteur d'instruments*** $\mathbf{z}_i$.
- Le vecteur des termes d'erreur du système est ***homoscédastique***.

Dans ce cas, on a :

$$\mathbf{Z}_i = \mathbf{I}_M \otimes \mathbf{z}_i$$

et donc :

$$E\left[\mathbf{Z}_i\mathbf{Z}_i'\right] = \mathbf{I}_M \otimes E\left[\mathbf{z}_i\mathbf{z}_i'\right] \text{ et } N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i\mathbf{Z}_i' = \mathbf{I}_M \otimes \left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right),$$

$$E\left[\mathbf{Z}_i\mathbf{u}_i\mathbf{u}'\mathbf{Z}_i'\right] = \mathbf{\Omega}_0 \otimes E\left[\mathbf{z}_i\mathbf{z}_i'\right] \text{ et } E\left[\mathbf{Z}_i\mathbf{u}_i\mathbf{u}'\mathbf{Z}_i'\right]^{-1} = \mathbf{\Omega}_0^{-1} \otimes E\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}$$

et :

$$N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i\tilde{\mathbf{\Omega}}_N^{-1}\mathbf{Z}_i' = \tilde{\mathbf{\Omega}}_N^{-1} \otimes \left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1}.$$

L'estimateur MMG optimal de $\mathbf{a}_0$ est un estimateur connu depuis bien avant la MMG.

Il s'agit de l'***estimateur des Triples Moindres Carrés*** (***3MC***).

Cet estimateur est obtenu en remplaçant les formules précédentes dans celle de $\hat{\mathbf{a}}_N^{MMG}$.

Définition. ***Estimateur des 3MC***

Soit un système d’équations simultanées linéaires tel que :

$$\mathbf{y}_i = \mathbf{X}_i'\mathbf{a}_0 + \mathbf{u}_i \quad \text{avec} \quad E\left[\mathbf{u}_i / \mathbf{z}_i\right] = \mathbf{0} \text{ et } E\left[\mathbf{u}_i\mathbf{u}_i' / \mathbf{z}_i\right] = \mathbf{\Omega}_0.$$

L’estimateur MMG optimal de $\mathbf{a}_0$ est l’***estimateur des 3MC*** :

$$\hat{\mathbf{a}}_N^{3MC} = \left[\left(N^{-1}\sum_{i=1}^N \mathbf{X}_i\mathbf{Z}_i'\right)\tilde{\mathbf{W}}_N^{-1}\left(N^{-1}\sum_{i=1}^N \mathbf{Z}_i\mathbf{X}_i'\right)\right]^{-1} \times\left(N^{-1}\sum_{i=1}^N \mathbf{X}_i\mathbf{Z}_i'\right)\tilde{\mathbf{W}}_N^{-1}N^{-1}\sum_{i=1}^N \mathbf{Z}_i\mathbf{y}_i$$

où :

$$\mathbf{Z}_i \equiv \mathbf{I}_M \otimes \mathbf{z}_i \quad \text{et} \quad \tilde{\mathbf{W}}_N^{-1} \equiv \tilde{\mathbf{\Omega}}_N^{-1} \otimes \left(N^{-1}\sum_{i=1}^N \mathbf{z}_i\mathbf{z}_i'\right)^{-1}$$

avec :

$$\tilde{\mathbf{\Omega}}_N \equiv N^{-1}\sum_{i=1}^N \tilde{\mathbf{u}}_{i,N}\tilde{\mathbf{u}}_{i,N}' \quad \text{et} \quad \tilde{\mathbf{u}}_{i,N} \equiv \mathbf{y}_i - \mathbf{X}_i'\hat{\mathbf{a}}_{\mathbf{E},N}^{2MC},$$

$\hat{\mathbf{a}}_{\mathbf{E},N}^{2MC}$ étant l’estimateur de $\mathbf{a}_0$ obtenu en empilant les estimateurs des 2MC des $\mathbf{a}_{m,0}$ pour $m = 1,\ldots,M$.

L'estimateur $\hat{\mathbf{a}}_N^{3MC}$ est donc calculé en trois étapes.

Etape 1. Calcul des $\hat{\mathbf{a}}_{m,N}^{2MC}$ pour $m=1,\ldots,M$.

Etape 2. Calcul de $\tilde{\mathbf{\Omega}}_N \equiv N^{-1}\sum_{i=1}^{N}\tilde{\mathbf{u}}_{i,N}\tilde{\mathbf{u}}'_{i,N}$ à partir des :

$$\tilde{u}_{m,i,N} \equiv y_{m,i} - \mathbf{x}'_{m,i}\hat{a}_{m,N}^{2MC} \quad \text{et} \quad \tilde{\mathbf{u}}_{i,N} \equiv \begin{bmatrix} \tilde{u}_{1,i,N} \\ \vdots \\ \tilde{u}_{M,i,N} \end{bmatrix}$$

Etape 3. Calcul de $\hat{\mathbf{a}}_N^{3MC}$.

- Le Système d'Equations Simultanées Linéaires simple s'écrit sous la forme :

$$\boxed{\mathbf{y}_i = \mathbf{X}_i'\mathbf{a}_0 + \mathbf{u}_i \quad \text{avec} \quad E\left[\mathbf{u}_i \mid \mathbf{z}_i\right] = \mathbf{0} \quad \text{et} \quad E\left[\mathbf{u}_i\mathbf{u}_i' / \mathbf{z}_i\right] = \mathbf{\Omega}_0}$$

- Il peut se résumer sous la forme de la condition de moment conditionnelle :

$$E\left[(\mathbf{y}_i - \mathbf{X}_i'\mathbf{a}_0) \mid \mathbf{z}_i\right] = \mathbf{0}_{M\times 1}$$

pour laquelle l'instrument optimal de Chamberlain est de la forme :

$$\mathbf{R}^+(\mathbf{z}_i) = E\left[\mathbf{X}_i \mid \mathbf{z}_i\right]\mathbf{\Omega}_0^{-1}.$$

- En notant $\mathbf{X}_i = \mathbf{S}(\mathbf{I}_M \otimes \mathbf{x}_i)$ où S est une matrice de sélection on montre alors que :

$$\mathbf{R}^+(\mathbf{z}_i) = \mathbf{S}E\left[(\mathbf{I}_M \otimes \mathbf{x}_i) \mid \mathbf{z}_i\right]\mathbf{\Omega}_0^{-1} = \mathbf{S}\left(\mathbf{\Omega}_0^{-1} \otimes E\left[\mathbf{x}_i \mid \mathbf{z}_i\right]\right).$$

- En utilisant :

$$\hat{\mathbf{a}}_N^{3MC} = \left[\left(N^{-1} \sum_{i=1}^{N} \mathbf{X}_i \mathbf{Z}_i' \right) \tilde{\mathbf{W}}_N^{-1} \left(N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i \mathbf{X}_i' \right) \right]^{-1}$$

$$\times \left(N^{-1} \sum_{i=1}^{N} \mathbf{X}_i \mathbf{Z}_i' \right) \tilde{\mathbf{W}}_N^{-1} N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i \mathbf{y}_i$$

avec :

$$\mathbf{X}_i = \mathbf{S}(\mathbf{I}_M \otimes \mathbf{x}_i), \quad \mathbf{Z}_i \equiv \mathbf{I}_M \otimes \mathbf{z}_i \quad \text{et} \quad \tilde{\mathbf{W}}_N^{-1} \equiv \tilde{\boldsymbol{\Omega}}_N^{-1} \otimes \left(N^{-1} \sum_{i=1}^{N} \mathbf{z}_i \mathbf{z}_i' \right)^{-1}$$

on obtient :

$$\hat{\mathbf{a}}_N^{3MC} = \left(N^{-1} \sum_{i=1}^{N} \mathbf{S} \left(\tilde{\boldsymbol{\Omega}}_N^{-1} \otimes \widehat{EL}_N \left[\mathbf{x}_i \mid \mathbf{z}_i \right] \right) \mathbf{X}_i') \right)^{-1}$$

$$\times N^{-1} \sum_{i=1}^{N} \mathbf{S} \left(\tilde{\boldsymbol{\Omega}}_N^{-1} \otimes \widehat{EL}_N \left[\mathbf{x}_i \mid \mathbf{z}_i \right] \right) \mathbf{y}_i$$

où :

$$\widehat{EL}_N \left[\mathbf{x}_i \mid \mathbf{z}_i \right] \equiv \left(N^{-1} \sum_{i=1}^{N} (\mathbf{x}_i \mathbf{z}_i') \right) \left(N^{-1} \sum_{i=1}^{N} \mathbf{z}_i \mathbf{z}_i' \right)^{-1} \mathbf{z}_i \,.$$

- L’estimateur $\hat{\mathbf{a}}_N^{3MC}$ est également un estimateur de la MM qui utilise une matrice d’instruments estimés qui est un estimateur convergent de la matrice :

$$\mathbf{S}\left(\mathbf{\Omega}_0^{-1} \otimes EL\left[\mathbf{x}_i \mid \mathbf{z}_i\right]\right).$$

6. Régressions empilées et estimateur SUR

Les ***systèmes de régression empilées (linéaires) standards*** sont définis par :

$$\begin{cases} y_{1,i} = \mathbf{a}'_{1,0}\mathbf{x}_{1,i} + u_{1,i} & \text{et } E\left[u_{1,i}/\mathbf{x}_i\right] = 0 \\ y_{2,i} = \mathbf{a}'_{2,0}\mathbf{x}_{2,i} + u_{2,i} & \text{et } E\left[u_{2,i}/\mathbf{x}_i\right] = 0 \\ \quad\vdots & \quad\vdots \\ y_{M,i} = \mathbf{a}'_{M,0}\mathbf{x}_{M,i} + u_{M,i} & \text{et } E\left[u_{M,i}/\mathbf{x}_i\right] = 0 \end{cases} \quad \text{et } E\left[\mathbf{u}_i\mathbf{u}'_i/\mathbf{x}_i\right] = \mathbf{\Omega}_0$$

où $\mathbf{x}_i$ est le vecteur composé des éléments distincts des $\mathbf{x}_{m,i}$ pour $m = 1,\ldots,M$.

Par rapport aux cas traités ci-dessus, ils ont ***quatre caractéristiques*** :

- Les paramètres des différentes équations ne sont pas liés.
- Le système est composé de ***modèles de régression***.
- Les variables explicatives de toutes les équations sont exogènes dans le système complet, et non seulement dans l'équation où elles apparaissent. C'est-à-dire qu'on a :

$$E\left[u_{m,i} / \mathbf{x}_{\ell,i}\right] = 0 \text{ pour } m = 1,\ldots,M \text{ et } \ell = 1,\ldots,M \text{ .}$$

  Ce qui implique que $\mathbf{x}_i$ ***est un vecteur d'instruments valide pour les*** M ***équations du système***.

- Le vecteur des termes d'erreur du système est ***homoscédastique***.

Ceci implique que :

- Ce système est un cas (très) particulier de système d'équations simultanées : celui où $\mathbf{z}_i = \mathbf{x}_i$ (*i.e.* un système d'équations simultanées sans variables explicatives endogènes)

- L'estimateur MMG optimal de $\mathbf{a}_0$ dans un système de régressions empilées est un cas (très) particulier d'estimateur des 3MC.

Dans ce cas, on a :

$$\mathbf{Z}_i = \mathbf{I}_M \otimes \mathbf{x}_i$$

et l'estimateur MMG optimal de $\mathbf{a}_0$ est un estimateur connu depuis bien avant la MMG.

Il s'agit de l'***estimateur des régressions empilées ou estimateur SUR, pour « seemingly unrelated regressions »***.

Le nom estimateur SUR, dû à Zellner, provient de ce que les régressions du système :

- ne sont pas liées par des contraintes inter-équations sur les paramètres (relations « évidentes » entre les équations)

mais

- peuvent être liées *via* les corrélations de leurs termes d'erreur.

Cet estimateur est obtenu en remplaçant les formules précédentes dans celle de $\hat{\mathbf{a}}_N^{MMG}$.

Il est important de noter que $\mathbf{X}_i = \mathbf{I}_M \otimes \mathbf{x}_i$ si seulement si $\mathbf{x}_i = \mathbf{x}_{m,i}$ pour $m = 1,...,M$.

Définition. ***Estimateur SUR ou des régressions empilées.***

Soit un système de régressions linéaires empilées tel que :

$$\mathbf{y}_i = \mathbf{X}_i'\mathbf{a}_0 + \mathbf{u}_i \quad \text{avec} \quad E[\mathbf{u}_i / \mathbf{x}_i] = \mathbf{0} \text{ et } E[\mathbf{u}_i\mathbf{u}_i' / \mathbf{x}_i] = \mathbf{\Omega}_0.$$

L'estimateur MMG optimal de $\mathbf{a}_0$ est l'***estimateur SUR*** :

$$\hat{\mathbf{a}}_N^{SUR} = \left\{ \left(N^{-1}\sum_{i=1}^N \mathbf{X}_i(\mathbf{I}_M \otimes \mathbf{x}_i') \right) \tilde{\mathbf{W}}_N^{-1} \left(N^{-1}\sum_{i=1}^N (\mathbf{I}_M \otimes \mathbf{x}_i)\mathbf{X}_i' \right) \right\}^{-1}$$
$$\times \left(N^{-1}\sum_{i=1}^N \mathbf{X}_i(\mathbf{I}_M \otimes \mathbf{x}_i') \right) \tilde{\mathbf{W}}_N^{-1} N^{-1}\sum_{i=1}^N (\mathbf{I}_M \otimes \mathbf{x}_i)\mathbf{y}_i$$

où :

$$\tilde{\mathbf{W}}_N^{-1} \equiv \tilde{\mathbf{\Omega}}_N^{-1} \otimes \left(N^{-1}\sum_{i=1}^N \mathbf{x}_i\mathbf{x}_i' \right)^{-1}, \; \tilde{\mathbf{\Omega}}_N \equiv N^{-1}\sum_{i=1}^N \tilde{\mathbf{u}}_{i,N}\tilde{\mathbf{u}}_{i,N}' \text{ et } \tilde{\mathbf{u}}_{i,N} \equiv \mathbf{y}_i - \mathbf{X}_i'\hat{\mathbf{a}}_{\mathbf{E},N}^{MCO},$$

$\hat{\mathbf{a}}_{\mathbf{E},N}^{MCO}$ étant l'estimateur de $\mathbf{a}_0$ obtenu en empilant les estimateurs des MCO des $\mathbf{a}_{m,0}$ pour $m = 1,...,M$.

L'estimateur $\hat{\mathbf{a}}_N^{SUR}$ est donc calculé en trois étapes.

Etape 1. Calcul des $\hat{\mathbf{a}}_{m,N}^{MCO}$ pour $m=1,\ldots,M$.

Etape 2. Calcul de $\tilde{\mathbf{\Omega}}_N \equiv N^{-1}\sum_{i=1}^{N}\tilde{\mathbf{u}}_{i,N}\tilde{\mathbf{u}}_{i,N}'$ à partir des :

$$\tilde{u}_{m,i,N} \equiv y_{m,i} - \mathbf{x}_{m,i}'\hat{a}_{m,N}^{MCO} \quad \text{et} \quad \tilde{\mathbf{u}}_{i,N} \equiv \begin{bmatrix} \tilde{u}_{1,i,N} \\ \vdots \\ \tilde{u}_{M,i,N} \end{bmatrix}$$

Etape 3. Calcul de $\hat{\mathbf{a}}_N^{SUR}$.

Un résultat, connu depuis bien avant la MMG, décrit les cas où les MCO équation par équation sur as. efficaces dans un système de régressions empilées.

Propriété. ***Théorème de Zellner***

(i) L'estimateur SUR $\hat{\mathbf{a}}_N^{SUR}$, ***est égal*** à l'estimateur des MCO « équation par équation » $\hat{\mathbf{a}}_{\mathbf{E},N}^{MCO}$ si $\mathbf{x}_{m,i} = \mathbf{x}_i$ pour $m = 1,\ldots,M$.

(ii) On a : $\sqrt{N}(\hat{\mathbf{a}}_N^{SUR} - \hat{\mathbf{a}}_{\mathbf{E},N}^{MCO}) \xrightarrow[N\to+\infty]{p} \mathbf{0}_{K\times 1}$, *i.e.* les estimateurs $\hat{\mathbf{a}}_N^{SUR}$ et $\hat{\mathbf{a}}_{\mathbf{E},N}^{MCO}$ sont ***as. équivalents***, si $\mathbf{\Omega}_0$ est diagonale. Dans ce cas, les régressions sont vraiment « *unrelated* ».

(iii) Dans les autres cas $\hat{\mathbf{a}}_N^{SUR}$ est as. plus efficace que $\hat{\mathbf{a}}_{\mathbf{E},N}^{MCO}$.

Cette propriété est un cas particulier du résultat général présenté ci-avant.

Rmq. Le Théorème de Zellner est un résultat important car les estimateurs MCO ont de bonnes propriétés à distance finie, *i.e.* pour de petits « N ».

Remarque. ***Définition de*** $\hat{\mathbf{a}}_N^{SUR}$

- En remarquant que $\mathbf{X}_i$ peut s'écrire à partir de la matrice $\mathbf{I}_M \otimes \mathbf{x}_i$ sous la forme :

$$\mathbf{X}_i = \mathbf{S}(\mathbf{I}_M \otimes \mathbf{x}_i)$$

avec **S** pour matrice de sélection on peut aisément montrer que :

$$\mathbf{X}_i \tilde{\mathbf{\Omega}}_N^{-1} = \mathbf{S}(\mathbf{I}_M \otimes \mathbf{x}_i)\tilde{\mathbf{\Omega}}_N^{-1} = \mathbf{S}(\tilde{\mathbf{\Omega}}_N^{-1} \otimes \mathbf{x}_i)$$

et :

$$\mathbf{X}_i \tilde{\mathbf{\Omega}}_N^{-1} \mathbf{X}_i' = \mathbf{S}(\mathbf{I}_M \otimes \mathbf{x}_i)\tilde{\mathbf{\Omega}}_N^{-1}(\mathbf{I}_M \otimes \mathbf{x}_i')\mathbf{S}' = \mathbf{S}\left(\tilde{\mathbf{\Omega}}_N^{-1} \otimes (\mathbf{x}_i \mathbf{x}_i')\right)\mathbf{S}'.$$

Il suffit d'utiliser $\tilde{\mathbf{\Omega}}_N^{-1} = \tilde{\mathbf{\Omega}}_N^{-1} \otimes 1$ et la propriété de « coagulation » de $\otimes$.

- On montre alors que :

$$\hat{\mathbf{a}}_N^{SUR} = \left(N^{-1}\sum_{i=1}^{N} \mathbf{X}_i \tilde{\mathbf{\Omega}}_N^{-1} \mathbf{X}_i'\right)^{-1} N^{-1}\sum_{i=1}^{N} \mathbf{X}_i \tilde{\mathbf{\Omega}}_N^{-1} \mathbf{y}_i\,,$$

ce qui est la forme la plus « standard » de $\hat{\mathbf{a}}_N^{SUR}$.