A.1. Introduction à la Méthode des Moments Généralisée : les bases

1. Motivation

2. Principe de l'estimation par la MMG

3. Définition et propriétés des estimateurs de la MMG

4. Les estimateurs des 2MC en tant qu'estimateurs de la MMG

5. MMG et MM

6. Test de sur-identification

7. Estimation séquentielle (MM)

1. Motivation

- La ***Méthode des Moments Généralisée*** (MMG) a été conçue par Hansen (1982) comme :

    - une ***extension de la Méthode des Moments*** pour les cas où les conditions de moment estimantes sur-identifient le vecteur de paramètres à estimer,

    - *i.e.* ***lorsqu'il y a plus de conditions de moment que de paramètres à estimer***.

    - Equations d'Euler et anticipations rationnelles

Elle est présentée ici pour 4 raisons :

(*i*) Elle permet de construire simplement des estimateurs pour les paramètres de ***systèmes d'équations*** (une des raisons de son succès). Les estimateurs SUR et des 3MC sont des cas (très) particuliers d'estimateurs MMG

(*ii*) Elle fournit un ***test de spécification*** simple : ***le test de sur-identification***.

(*iii*) Elle donne un cadre d'inférence statistique permettant de gérer assez simplement les questions liées à l'hétéroscédasticité :

- Estimateurs robustes à l'hétéroscédasticité
- Estimateurs des MC ou des 2MC, Généralisés ou Quasi-Généralisés.

(i*v*) La MMG permet de *relier tous les cadres d'inférence statistique* (***fréquentistes***) ***utilisés usuellement*** :

- Moindres Carrés (MCO, MCG, MCQG, …)
- Méthode des Moments et techniques de VI (2MC, 3MC, …)
- Maximum de Vraisemblance
- Chi-2 Minimum, Moindres Carrés Asymptotiques, …

i.e. tous les estimateurs définis dans ces cadres ont une interprétation « MMG ».

Rmq. Le véritable lien entre toutes ces méthodes est cependant la Méthode des Moments « Standard ».

Rmq. Avec la MMG, pas besoin de spécifier la loi de distribution du « terme d'erreur au bout à droite », *i.e*. modèles semi-paramétriques (à termes d'erreur additifs).

Rmq. Les estimateurs MMG sont des exemples d'estimateurs « extrêmum », *i.e*. fondés sur a minimisation ou maximisation d'un critère sur les paramètres à estimer.

1.1. Constat

Les estimateurs présentés jusqu'à présent MCO, VI et 2MC exploitent tous une ***condition de moment identifiante*** sous forme de condition d'orthogonalité.

- Dans le cas des MCO et VI, la condition de moment est juste-identifiante.

- Le paramètre $\mathbf{a}_0$ étant de dimension $K \times 1$, les conditions de moment :

$$E\left[\mathbf{x}_i (y_i - \mathbf{a}_0' \mathbf{x}_i)\right] = \mathbf{0} \text{ et } E\left[\mathbf{z}_i (y_i - \mathbf{a}_0' \mathbf{x}_i)\right] = \mathbf{0} \text{ avec } \dim \mathbf{z}_i = K$$

  sont des systèmes de K équations à K inconnues (les éléments de $\mathbf{a}_0$).

- Ces équations « juste-identifiant » $\mathbf{a}_0$, ***l'application du principe l'analogie est directe*** et permet d'obtenir les estimateurs des MCO et des VI à partir de la résolution des équations :

$$N^{-1} \sum_{i=1}^{N} \mathbf{x}_i (y_i - \mathbf{x}_i' \hat{\mathbf{a}}_N^{MCO}) = \mathbf{0} \text{ et } N^{-1} \sum_{i=1}^{N} \mathbf{z}_i (y_i - \mathbf{x}_i' \hat{\mathbf{a}}_N^{VI}) = \mathbf{0}$$

- Dans le cas des 2MC, la condition de moment :

$$E\left[\mathbf{z}_i(y_i - \mathbf{a}_0'\mathbf{x}_i)\right] = \mathbf{0} \text{ avec } \dim \mathbf{z}_i \geq K$$

  juste ou sur-identifie $\mathbf{a}_0$. Ce qui ne permet pas l'application directe du principe d'analogie en cas de sur-identification.

- L'idée est ici que la contrepartie empirique de la condition de moment précédent n'admet en général pas de solution :

$$\boxed{\text{Il n'existe, en général, pas de } \hat{\mathbf{a}}_N^{MM} \text{ tel que } N^{-1}\sum_{i=1}^N \mathbf{z}_i(y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{MM}) = \mathbf{0}.}$$

- ***Rappel*** : Eliminer des instruments pour se ramener au cas juste-identifié n'est pas optimal :
  - Moins d'instruments = moins d'information
  - Moins d'information = estimation moins efficace

- L'***astuce utilisée*** pour résoudre ce problème a été de remplacer $\mathbf{z}_i$ par la projection linéaire de $\mathbf{x}_i$ sur $\mathbf{z}_i$, $E[\mathbf{x}_i\mathbf{z}_i']E[\mathbf{z}_i\mathbf{z}_i']^{-1}\mathbf{z}_i$, qui est un vecteur d'instruments estimable de dimension K.

- L'application du principe d'analogie donne alors l'estimateur des 2MC comme solution de :

$$\left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right]\left[N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right]^{-1}N^{-1}\sum_{i=1}^{N}\mathbf{z}_i(y_i-\mathbf{x}_i'\hat{\mathbf{a}}_N^{2MC})=\mathbf{0}.$$

La MMG permet d'utiliser directement $\mathbf{z}_i$ comme vecteur d'instruments, même lorsque $\dim \mathbf{z}_i > K$, sans astuce pour se ramener au cas juste-identifié.

Se pose néanmoins la question de savoir si utiliser directement $\mathbf{z}_i$ comme vecteur d'instruments est « statistiquement » efficace :
question des ***instruments « statistiquement » optimaux***
(*e.g*. pourquoi ne pas employer des carrés des éléments de $\mathbf{z}_i$?).

1.2. Lien avec le Maximum de Vraisemblance

- Utilisation du cadre d'inférence du MV si modèle « complexe », e.g. :
  - Choix discret
  - Variables à expliquer avec des discontinuités « sévères » telles que des variables dépendantes limitées
  - Modèles à paramètres aléatoires avec intérêt pour la distribution des paramètres aléatoire (loi mélangeante)

- Problème avec le MV : choix (spécification) de la loi de distribution (conditionnelle) des variables à expliquer :
  - Econometres très réticents, d'où MV réservé à des modèles avec lesquels « c'est difficile de faire autrement ».

- Analyse du PGD permet de définir la distribution paramétrique de $\mathbf{y}_i | \mathbf{x}_i$ à partir de sa fonction de densité (de probabilité) :

$$p_0(\mathbf{y}_i | \mathbf{x}_i) = f(\mathbf{y}_i | \mathbf{x}_i ; \boldsymbol{\theta}_0).$$

- Inégalité d'information de Kullback-Leibler :

> **Propriété.** ***Inégalité d'information de Kullback-Leibler***
>
> Pour tout $(\mathbf{x}, \boldsymbol{\theta}) \in \mathcal{X} \times \Theta$ on a :
>
> $$E_{\mathbf{y}_i | (\mathbf{x}_i = \mathbf{x})} \left[\ln p_0(\mathbf{y}_i ; \mathbf{x}_i) - \ln f(\mathbf{y}_i ; \mathbf{x}_i ; \boldsymbol{\theta}) \right] \geq 0.$$

- **Condition d'identification paramétrique** :

$$E\left[\ln p_0(\mathbf{y}_i ; \mathbf{x}_i) - \ln f(\mathbf{y}_i ; \mathbf{x}_i ; \boldsymbol{\theta}) \right] > 0 \quad \text{pour tout } \boldsymbol{\theta} \in \Theta \text{ tel que } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0.$$

- Inégalité d'information de Kullback-Leibler et identification paramétrique donnent :

$$\boldsymbol{\theta}_0 = \arg\max_{\boldsymbol{\theta}\in\Theta} E\left[\ln f(\mathbf{y}_i;\mathbf{x}_i;\boldsymbol{\theta})\right]$$

- Application du principe d'analogie :

$$\boxed{\hat{\boldsymbol{\theta}}_N \equiv \arg\max_{\boldsymbol{\theta}\in\Theta} N^{-1}\sum_{i=1}^{N}\ln f(\mathbf{y}_i;\mathbf{x}_i;\boldsymbol{\theta}) = \hat{\boldsymbol{\theta}}_N^{MV}}$$

- Caractérisation de $\hat{\boldsymbol{\theta}}_N^{MV}$ par les CO1 du programme de maximisation de la log-vraisemblance :

$$\boxed{N^{-1}\sum_{i=1}^{N}\frac{\partial \ln f(\mathbf{y}_i;\mathbf{x}_i;\hat{\boldsymbol{\theta}}_N^{MV})}{\partial\boldsymbol{\theta}} = N^{-1}\sum_{i=1}^{N}\mathbf{s}(\mathbf{y}_i;\mathbf{x}_i;\hat{\boldsymbol{\theta}}_N^{MV}) = \mathbf{0}}$$

où :

$$\boxed{\mathbf{s}(\mathbf{y}_i;\mathbf{x}_i;\boldsymbol{\theta}) \equiv \frac{\partial \ln f(\mathbf{y}_i;\mathbf{x}_i;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \text{ : score de l'observation } i \text{ en } \boldsymbol{\theta}.}$$

- La condition :

$$N^{-1}\sum_{i=1}^{N}\frac{\partial \ln f(\mathbf{y}_i;\mathbf{x}_i;\hat{\boldsymbol{\theta}}_N^{MV})}{\partial \boldsymbol{\theta}} = N^{-1}\sum_{i=1}^{N}\mathbf{s}(\mathbf{y}_i;\mathbf{x}_i;\hat{\boldsymbol{\theta}}_N^{MV}) = \mathbf{0}$$

est la contre-partie empirique de la condition de Moment (juste-identifiée) :

$$E\left[\frac{\partial \ln f(\mathbf{y}_i;\mathbf{x}_i;\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right] = E\left[\mathbf{s}(\mathbf{y}_i;\mathbf{x}_i;\boldsymbol{\theta}_0)\right] = \mathbf{0}$$

qui est une condition induite par la définition de $\boldsymbol{\theta}_0$:

$$\boldsymbol{\theta}_0 = \arg\max_{\boldsymbol{\theta}\in\Theta} E\left[\ln f(\mathbf{y}_i;\mathbf{x}_i;\boldsymbol{\theta})\right].$$

- **Conclusion** : Un estimateur du MV peut également être interprétée comme un estimateur de la MM(G).

- Partir directement de :

$$E\left[\frac{\partial \ln f(\mathbf{y}_i;\mathbf{x}_i;\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right] = E\left[\mathbf{s}(\mathbf{y}_i;\mathbf{x}_i;\boldsymbol{\theta}_0)\right] = \mathbf{0}$$

  - Cadre d’inférence à « l’intersection » de ceux du MV et de la MM(G) : **Méthode des Scores**.

    - Méthode des scores utilisée si vraisemblance très compliquée

    - **Attention** : Cette condition ne caractérise pas que $\boldsymbol{\theta}_0$ si $E\left[\ln f(\mathbf{y}_i;\mathbf{x}_i;\boldsymbol{\theta})\right]$ a plusieurs optima (maxima locaux et minima locaux) en $\boldsymbol{\theta} \in \Theta$.

1.2. Lien avec les Moindres Carrés

- Les techniques de régression, *i.e*. de « moindres carrés », sont des outils de prédiction très « puissants »
- En économétrie : utilisation pour la prédiction ou pour l'estimation de modèles structurels si exogénéité des variables explicatives
- Régression liée à la notion d'espérance conditionnelle :

Propriété. ***Espérance conditionnelle, erreur quadratique moyenne et prédiction***

$$g(\mathbf{x}_i) \equiv E[y_i \mid \mathbf{x}_i] \quad \Leftrightarrow \quad g(\mathbf{x}_i) = \min_{\gamma(.)\in L^2} E[(y_i - \gamma(\mathbf{x}_i))^2].$$

- Identification paramétrique donnent :

$$E[y_i \mid \mathbf{x}_i] = g(\mathbf{x};\boldsymbol{\theta}_0) \quad \Leftrightarrow \quad \boldsymbol{\theta}_0 = \arg\min_{\boldsymbol{\theta}\in\Theta} E[(y_i - g(\mathbf{x}_i;\boldsymbol{\theta}))^2]$$

- Application du principe d'analogie :

$$\boxed{\hat{\boldsymbol{\theta}}_N \equiv \arg\min_{\boldsymbol{\theta}\in\Theta} N^{-1}\sum_{i=1}^{N}\left(y_i - g(\mathbf{x}_i;\boldsymbol{\theta})\right)^2 = \hat{\boldsymbol{\theta}}_N^{MC}}$$

- Caractérisation de $\hat{\boldsymbol{\theta}}_N^{MC}$ par les CO1 du programme de maximisation de la log-vraisemblance :

$$\boxed{(-2\times) \quad N^{-1}\sum_{i=1}^{N}\frac{\partial g(\mathbf{x}_i;\hat{\boldsymbol{\theta}}_N^{MC})}{\partial\boldsymbol{\theta}}\left(y_i - g(\mathbf{x}_i;\hat{\boldsymbol{\theta}}_N^{MC})\right) = \mathbf{0}}$$

où :

- La condition :

$$N^{-1}\sum_{i=1}^{N}\frac{\partial g(\mathbf{x}_i;\hat{\boldsymbol{\theta}}_N^{MC})}{\partial\boldsymbol{\theta}}\left(y_i - g(\mathbf{x}_i;\hat{\boldsymbol{\theta}}_N^{MC})\right)=\mathbf{0}$$

est la contre-partie empirique de la condition de Moment (juste-identifiée) :

$$E\left[\frac{\partial g(\mathbf{x}_i;\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\left(y_i - g(\mathbf{x}_i;\boldsymbol{\theta}_0)\right)\right]=\mathbf{0}$$

qui est une condition d'orthogonalité induite par la définition de $\boldsymbol{\theta}_0$:

$$\boldsymbol{\theta}_0 = \arg\min_{\boldsymbol{\theta}\in\Theta} E[(y_i - g(\mathbf{x}_i;\boldsymbol{\theta}))^2].$$

- **Conclusion** : Un estimateur des MC peut également être interprétée comme un estimateur de la MM(G).

- Lien avec les modèles économétriques :

$$E[y_i \mid \mathbf{x}_i] = g(\mathbf{x};\boldsymbol{\theta}_0) \quad \Leftrightarrow \quad y_i = g(\mathbf{x};\boldsymbol{\theta}_0) + u_i \text{ avec } E[u_i \mid \mathbf{x}_i] = 0$$

où :

$$u_i \equiv y_i - g(\mathbf{x};\boldsymbol{\theta}_0).$$

- L'exogénéité de $\mathbf{x}_i$, *i.e.* $E[u_i \mid \mathbf{x}_i] = 0$, implique :

$$E\left[\frac{\partial g(\mathbf{x}_i;\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\left(y_i - g(\mathbf{x}_i;\boldsymbol{\theta}_0)\right)\right] = \mathbf{0}.$$

- On verra dans la suite que $\partial g(\mathbf{x}_i;\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}$ est le « meilleur » instrument du terme d'erreur u_i dans le modèle :

$$y_i = g(\mathbf{x};\boldsymbol{\theta}_0) + u_i \text{ avec } E[u_i \mid \mathbf{x}_i] = 0.$$

1.3. Démarche suivie dans la Partie A

- En fait la MMG définit ***comment exploiter toute condition de moment identifiant*** un vecteur de paramètres $\mathbf{a}_0$, *e.g.* :

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

pour construire un ***estimateur de*** $\mathbf{a}_0$,

  - à partir échantillon aléatoire de variables aléatoires iid $\left\{\mathbf{w}_i; i=1,2,\ldots,N\right\}$,

  - que cette condition de moment ***juste- ou sur-identifie*** $\mathbf{a}_0$, *i.e.* qu'on ait :

$$\underbrace{G=\dim \mathbf{g}(.)=K=\dim \mathbf{a}}_{\text{Juste-identification}} \text{ ou } \underbrace{G=\dim \mathbf{g}(.)>K=\dim \mathbf{a}}_{\text{Sur-identification}}.$$

Dans un premier temps : définir les estimateurs de la MMG et leurs propriétés

Comment exploiter toute condition de la forme :

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

pour construire un ***estimateur de*** $\mathbf{a}_0$.

- Utiliser la MMG et ***construire des estimateurs*** de la MMG de $\mathbf{a}_0$.
    - Définition des estimateurs de la MMG
    - Propriétés générales :
        - Existence, convergence, normalité as.
    - Estimateurs MMG optimaux
    - Conditions de moments et efficacité

Rmq. Estimateurs MMG = estimateurs MM si juste-identification

- ***Problème*** : La plupart des modèles économétriques donnent des conditions de moments conditionnels de la forme :

$$\boxed{E\left[\mathbf{u}(\mathbf{y}_i,\mathbf{x}_i;\mathbf{a}_0)/\mathbf{z}_i\right]=\mathbf{0} \quad \text{avec} \quad \mathbf{w}_i \equiv (\mathbf{y}_i,\mathbf{x}_i,\mathbf{z}_i).}$$

  - Le problème est que si on sait définir des estimateurs de $E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right]$, c'est plus beaucoup plus difficile pour $E\left[\mathbf{u}(\mathbf{y}_i,\mathbf{x}_i;\mathbf{a})/\mathbf{z}_i\right]$
  - D'où des difficultés pour appliquer le principe d'analogie et donc le principe de la MM.
  - Il faut construire une condition de moment de la forme :

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

    à partir d'un condition de moment de la forme :

$$E\left[\mathbf{u}(\mathbf{y}_i,\mathbf{x}_i;\mathbf{a}_0)/\mathbf{z}_i\right]=\mathbf{0}.$$

- Les propriétés des espérances conditionnelles donnent :

$$E\left[\mathbf{u}(\mathbf{y}_i,\mathbf{x}_i;\mathbf{a}_0)/\mathbf{z}_i\right]=\mathbf{0} \quad \Rightarrow \quad E\left[\overbrace{\mathbf{R}(\mathbf{z}_i)\mathbf{u}(\mathbf{y}_i,\mathbf{x}_i;\mathbf{a}_0)}^{\mathbf{g}(\mathbf{w}_i;\mathbf{a}_0)}\right]=\mathbf{0}$$

pour toute fonction $\mathbf{R}(.)$ de taille conforme.

- Question. Comment choisir $\mathbf{R}(\mathbf{z})$, une matrices d'instruments pour $\mathbf{u}(\mathbf{y}_i,\mathbf{x}_i;\mathbf{a}_0)$?

Dans un second temps : choix optimal de $\mathbf{R}(\mathbf{z})$ ou, de manière équivalente, de $\mathbf{g}(\mathbf{w};\mathbf{a})$

Chapitre A.2. Instruments optimaux

Remarques générales

- Les estimateurs ***nouveaux*** introduits par la MMG sont tous des estimateurs fondés sur des conditions d'orthogonalité :

$$E\left[\mathbf{R}(\mathbf{z}_i)\mathbf{u}(\mathbf{y}_i,\mathbf{x}_i;\mathbf{a})\right]=\mathbf{0} \quad \Leftrightarrow \quad \mathbf{a}=\mathbf{a}_0$$

  surtout :
    - Pour les ***systèmes d'équations***, et les ***modèles de données de panel*** en particulier
    - Pour les modèles définis par des espérances conditionnelles.

- La MMG a permis de « systématiser » la construction d'estimateurs robustes à l'hétéroscédasticité.
    - Elle a donné un cadre général à des résultats déjà établis, notamment par White.

- Sinon, elle a permis de ***ré-interpréter beaucoup de résultats*** concernant les estimateurs des MC et des 2MC.

2. Principe d'estimation de la MMG

- La MMG définit ***comment exploiter toute condition de moment identifiant*** un vecteur de paramètres $\mathbf{a}_0$, *e.g.* :

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

pour construire un estimateur de $\mathbf{a}_0$ à partir échantillon aléatoire de variables aléatoires iid $\left\{\mathbf{w}_i; i=1,2,\ldots,N\right\}$, que cette condition de moment ***juste- ou sur-identifie*** $\mathbf{a}_0$.

$$\underbrace{\dim \mathbf{g}(.)=G=K=\dim \mathbf{a}}_{\text{Juste-identification}} \quad \text{ou} \quad \underbrace{\dim \mathbf{g}(.)=G>K=\dim \mathbf{a}}_{\text{Sur-identification}}.$$

- Elle repose sur une extension « naturelle » du principe d'analogie utilisé dans le cas de la Méthode des Moments « standard »,
    - cette extension reposant sur un résultat d'algèbre linéaire concernant les ***formes quadratiques*** construites à partir de ***matrices définies positives***.

Propriété. ***Motivation du Critère d'estimation de la MMG***

Les deux conditions suivantes sont équivalentes :

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

et :

$$(ii)\quad \mathbf{a}_0 \equiv \arg\min_{\mathbf{a}} E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right]' \mathbf{M}_0 E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right]$$

pour ***toute*** matrice $\mathbf{M}_0$ ***définie positive*** de dimension G.

- Avec :

$$\mathbf{M}_0 \equiv \begin{bmatrix} m_{0,11} & m_{0,12} & \cdots & m_{0,1G} \\ m_{0,12} & m_{0,22} & \cdots & m_{0,2G} \\ \vdots & \vdots & \ddots & \vdots \\ m_{0,1G} & m_{0,2G} & \cdots & m_{0,GG} \end{bmatrix} \text{ et } \mathbf{g}(\mathbf{w}_i;\mathbf{a}) \equiv \begin{bmatrix} g_1(\mathbf{w}_i;\mathbf{a}) \\ g_2(\mathbf{w}_i;\mathbf{a}) \\ \vdots \\ g_K(\mathbf{w}_i;\mathbf{a}) \end{bmatrix}$$

on a :

$$E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right]' \mathbf{M}_0 E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right]$$

$$||$$

$$\sum_{k=1}^{G} \sum_{\ell=1}^{G} m_{0,k\ell} \times E\left[g_k(\mathbf{w}_i;\mathbf{a})\right] E\left[g_\ell(\mathbf{w}_i;\mathbf{a})\right]$$

$$||$$

$$\sum_{\ell=1}^{G} m_{0,kk} \times E\left[g_k(\mathbf{w}_i;\mathbf{a})\right]^2 + 2 \times \sum_{k=1}^{G} \sum_{\ell>k}^{G} m_{0,k\ell} \times E\left[g_k(\mathbf{w}_i;\mathbf{a})\right] E\left[g_\ell(\mathbf{w}_i;\mathbf{a})\right]$$

- Puisque $\mathbf{M}_0$ est définie positive, la forme quadratique :

$$Q_0(\mathbf{a};\mathbf{M}_0) \equiv E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a})'\right]\mathbf{M}_0 E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right]$$

est une fonction scalaire de **a** qui :

(*i*) est positive pour tout **a** : $Q_0(\mathbf{a};\mathbf{M}_0) \geq 0$

et :

(*ii*) n'est nulle que si $E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right] = \mathbf{0}$, ce qui est équivalent ici à $\mathbf{a} = \mathbf{a}_0$.

- Concrètement, le critère d'estimation pour calculer l'estimateur de la MMG de $\mathbf{a}_0$ est la ***contre-partie empirique*** de $Q_0(\mathbf{a};\mathbf{M}_0) \geq 0$. L'estimateur de $\mathbf{a}_0$ est la valeur de **a** qui minimise ce critère.

- La matrice $\mathbf{M}_0$ définit une ***métrique*** qui permet de mesurer l'écart de $E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right]$ par rapport au vecteur **0**, *i.e*. de juger si $E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right]$ est proche ou non de **0**.

- Bien, entendu il existe une ***infinité de choix possibles*** pour $\mathbf{M}_0$ mais :
    - Il existe des estimateurs de $\mathbf{a}_0$, dit ***estimateurs MMG optimaux*** (fondés sur $E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a}_0)\right]=\mathbf{0}$).
    - Ces estimateurs sont définis par un choix particulier de $\mathbf{M}_0$.
    - Ces estimateurs sont ***les « meilleurs » pour exploiter la condition de moment*** $E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a}_0)\right]=\mathbf{0}$ ***et l'information contenue dans les données***.

3. Définition et propriétés des estimateurs de la MMG

- L'intérêt de cette propriété est immédiat pour l'application du principe d'analogie.

- La MM applique directement ce principe, *i.e.* définit $\hat{\mathbf{a}}_N^{MM}$ par :

$$\hat{\mathbf{a}}_N^{MM} \text{ est solution en } \mathbf{a} \text{ de } N^{-1}\sum_{i=1}^{N}\mathbf{g}(\mathbf{w}_i;\mathbf{a}) = \mathbf{0}$$

lorsque $G = K$. Le programme précédent est la contre-partie empirique du problème limite :

$$\mathbf{a}_0 \text{ est la solution unique en } \mathbf{a} \text{ de } E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right] = \mathbf{0}$$

L'équation définissant $\hat{\mathbf{a}}_N^{MM}$ n'a, en général, pas de solution lorsque $G > K$.

***En revanche on peut toujours* (*en théorie*) *résoudre un programme de minimisation* (*même si c'est parfois difficile empiriquement*).**

- Un estimateur de la MMG de $\mathbf{a}_0$ est défini comme la solution en $\mathbf{a}$ d'une contre-partie empirique du programme de minimisation caractérisant $\mathbf{a}_0$ à partir de $\mathbf{M}_0$.

Avec :

$$\mathbf{a}_0 \equiv \arg\min_{\mathbf{a}} E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a})'\right]\mathbf{M}_0 E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right],$$

le principe d'analogie définit $\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N)$ par :

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

où :

$$\tilde{\mathbf{M}}_N \xrightarrow[N\to+\infty]{p} \mathbf{M}_0 \text{ où } \mathbf{M}_0 \text{ est définie positive.}$$

- Dans le critère utilisé par la MMG :

$$Q_N(\mathbf{a};\tilde{\mathbf{M}}_N) \equiv \left(N^{-1}\sum_{i=1}^{N}\mathbf{g}(\mathbf{w}_i;\mathbf{a})'\right)\tilde{\mathbf{M}}_N\left(N^{-1}\sum_{i=1}^{N}\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right)$$

la matrice $\tilde{\mathbf{M}}_N$ est un estimateur convergent de $\mathbf{M}_0$, cette matrice :

  - définit une ***métrique*** qui ***pondère les conditions de moment estimantes*** et ***permet de*** « ***juger*** » si $N^{-1}\sum_{i=1}^{N}\mathbf{g}(\mathbf{w}_i;\mathbf{a})$ est proche de **0**

  et :

  - peut être choisie de manière ***optimale***, *i.e.* de sorte à ce que l’estimateur $\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N)$ soit le plus efficace parmi les estimateurs MMG.

- La fonction critère $Q_N(\mathbf{a};\tilde{\mathbf{M}}_N)$ converge (point par point) en probabilité vers :

$$Q_0(\mathbf{a};\mathbf{M}_0) \equiv E\left[\mathbf{g}_i(\mathbf{a})'\right]\mathbf{M}_0 E\left[\mathbf{g}_i(\mathbf{a})\right]$$

3.1. Définition des estimateurs MMG

- Avec les notations :

$$\mathbf{F}_i(\mathbf{a}) \equiv \mathbf{F}(\mathbf{w}_i;\mathbf{a}) \text{ et } \overline{\mathbf{F}}_N(\mathbf{a}) \equiv N^{-1}\sum_{i=1}^{N}\mathbf{F}_i(\mathbf{a})$$ (***moyenne empirique***).

Définition. ***Estimateurs de la MMG***

Les estimateurs la MMG de $\mathbf{a}_0$, $\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N)$, fondés sur la condition de moment $E\left[\mathbf{g}_i(\mathbf{a}_0)\right] = \mathbf{0}$ sont définis par :

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

où :

$$\tilde{\mathbf{M}}_N \xrightarrow[N\to+\infty]{p} \mathbf{M}_0$$ et $\mathbf{M}_0$ est une matrice définie positive.

- Des ***conditions de régularité***, relatives à $\mathbf{g}(.;.)$ et aux $\mathbf{w}_i$, assurent que les $\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N)$ sont ***convergents*** et ***as. normaux***.
  Ces conditions ne sont pas détaillées ici pour deux raisons.

  (*i*) Elles sont relativement « techniques ». *E.g.*, on utilise la notion de « convergence uniforme en probabilité ».

  (*ii*) Dans notre utilisation de la MMG, elles sont immédiates.

- Notons que les propriétés des estimateurs MMG sont aisées à démontrer si :
  - $\mathbf{g}(\mathbf{w};\mathbf{a})$ est continûment différentiable en $\mathbf{a}$,
  - le domaine de définition de $\mathbf{a}$ est un ensemble compact et que $\mathbf{a}_0$ appartient à l'intérieur de cet ensemble,
  - la distribution des $\mathbf{w}_i$ et $\mathbf{g}(.;.)$ doivent être telles que les $\mathbf{g}(\mathbf{w}_i;\mathbf{a})$ et leurs Jacobiens sont de variance finie.

- La démonstration de ces résultats n'offre pas d'intérêt particulier pour nous, sauf celui de la normalité asymptotique.
- Voir Gouriéroux et Monfort (1989) et McFadden et Newey (1994).

Rmq. Les conditions permettant de définir les propriétés des estimateurs sont de deux types :

- Les ***conditions d'identification*** (ou ***structurelles***) assurent que l'estimateur permet d'atteindre sa cible. Elles sont essentielles, la plus importante étant ici :

  $\mathbf{a}_0$ est la solution unique en $\mathbf{a}$ de $E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right]=\mathbf{0}$

- Les ***conditions de régularité*** assurent le « fonctionnement correct » du principe d'analogie.
  Sous ces conditions les moyennes convergent vers les espérances « comme il le faut ».

3.2. Existence et convergence des estimateurs MMG

Propriété. ***Existence et convergence des estimateurs de la MMG***

Soit $\left\{\mathbf{w}_i ; i=1,2,\ldots,N\right\}$ un échantillon de variables aléatoires iid telles que :

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

Les estimateurs de la MMG de $\mathbf{a}_0$ sont définis par :

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

où $\tilde{\mathbf{M}}_N \xrightarrow[N \to +\infty]{p} \mathbf{M}_0$ et $\mathbf{M}_0$ est une matrice définie positive.

On a, sous certaines conditions de régularité :

(*i*) $\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N)$ existe avec une probabilité approchant 1

et :

(*ii*) $\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N) \xrightarrow[N \to +\infty]{p} \mathbf{a}_0 .$

$$\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N) \equiv \arg\min_{\mathbf{a}} \bar{\mathbf{g}}_N(\mathbf{a})'\tilde{\mathbf{M}}_N\bar{\mathbf{g}}_N(\mathbf{a})$$

$$\Big\downarrow \; \boxed{\text{LGN et conditions de régularité}}$$

$$\mathbf{a}_0 = \arg\min_{\mathbf{a}} E\left[\mathbf{g}_i(\mathbf{a})'\right]\mathbf{M}_0 E\left[\mathbf{g}_i(\mathbf{a})\right]$$

$$\bar{\mathbf{g}}_N(\mathbf{a})'\tilde{\mathbf{M}}_N\bar{\mathbf{g}}_N(\mathbf{a})$$

$$\Big\downarrow \; \boxed{\text{LGN}}$$

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

- Les propriétés de convergence des estimateurs de la MMG sont ***remarquables*** en raison de leur ***caractère général***.

- Elles s'appliquent pour toute matrice $\mathbf{M}_0$ définie positive et toute matrice $\tilde{\mathbf{M}}_N$ convergent en probabilité vers $\mathbf{M}_0$.

3.3. Normalité asymptotique des estimateurs MMG

Propriété. ***Normalité as. des estimateurs de la MMG***

Sous les hypothèses définies ci-dessus et si de plus :

(*i*) $rang\mathbf{G}_0 = K$ où $\mathbf{G}_0 \equiv E\left[\partial \mathbf{g}_i(\mathbf{a}_0)/\partial \mathbf{a}'\right]$

(*ii*) la matrice $\mathbf{W}_0 \equiv E\left[\mathbf{g}_i(\mathbf{a}_0)\mathbf{g}_i(\mathbf{a}_0)'\right]$ est inversible alors :

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

avec :

$$\mathbf{\Sigma}(\mathbf{M}_0) \equiv (\mathbf{G}_0'\mathbf{M}_0\mathbf{G}_0)^{-1}\mathbf{G}_0'\mathbf{M}_0\mathbf{W}_0\mathbf{M}_0\mathbf{G}_0(\mathbf{G}_0'\mathbf{M}_0\mathbf{G}_0)^{-1}$$

sous certaines conditions de régularité.

- $\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N)$ est ***as. normal pour toute*** $\tilde{\mathbf{M}}_N \xrightarrow[N\to+\infty]{p} \mathbf{M}_0$ définie positive
- ***La distribution as. de*** $\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N)$ ***ne dépend que de*** $\mathbf{M}_0$, pas de $\tilde{\mathbf{M}}_N$.

- La condition (*ii*) qui dit que $\mathbf{W}_0 \equiv E\left[\mathbf{g}_i(\mathbf{a}_0)\mathbf{g}_i(\mathbf{a}_0)'\right]$ est inversible est vérifiée :
  - Si les éléments de $\mathbf{g}_i(\mathbf{a}_0)$ ne sont pas (trop) linéairement liés, *i.e*. si les conditions estimantes utilisées ne sont ***pas redondantes*** (linéairement).
- La condition (*i*) qui dit que $rangE\left[\partial\mathbf{g}_i(\mathbf{a}_0)/\partial\mathbf{a}'\right] = K$ est inversible est dite ***condition locale d'identification*** :
  - Elle est vérifiée si les éléments de $\mathbf{g}_i(\mathbf{a})$ varient suffisamment indépendamment les uns des autres en **a** autour de $\mathbf{a}_0$.
  - Cette condition garantit que $Q_0(\mathbf{a};\mathbf{M}_0)$ n'est pas plate **a** autour de $\mathbf{a}_0$.
  - Dans un modèle linéaire à VI elle est donnée par $rangE\left[\mathbf{z}_i\mathbf{x}_i'\right] = K$.

Propriété. ***Approximation des estimateurs de la MMG***

Sous les hypothèses définies ci-dessus on a :

$$\sqrt{N}\left(\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N)-\mathbf{a}_0\right)=-(\mathbf{G}_0'\mathbf{M}_0\mathbf{G}_0)^{-1}\mathbf{G}_0'\mathbf{M}_0\sqrt{N}N^{-1}\sum\nolimits_{i=1}^{N}\mathbf{g}(\mathbf{w}_i;\mathbf{a}_0)+\mathbf{o}_p(1)$$

où $\mathbf{o}_p(1)$ est un « infiniment » petit en probabilité, *i.e.* $\mathbf{o}_p(1)\xrightarrow[N\to+\infty]{p}\mathbf{0}$.

- Cette formule de $\sqrt{N}\left(\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N)-\mathbf{a}_0\right)$ s'avère être très pratique pour certains calculs :
  - Voir le chapitre sur l'estimation séquentielle.
  - Calcul d'estimateurs MMG optimaux en une étape (voir, *e.g.*, Gouriéroux et Monfort, 1989).
  - Optimisation numérique

$$(i)\quad \mathbf{A}_N \xrightarrow[N\to+\infty]{p} \mathbf{0} \quad \Leftrightarrow \quad \mathbf{A}_N = \mathbf{o}_p(1)$$

$$(ii)\quad \mathbf{A}_N \xrightarrow[N\to+\infty]{p} \mathbf{A} \quad \Leftrightarrow \quad \mathbf{A}_N = \mathbf{A} + \mathbf{o}_p(1)$$

$$(iii)\quad \mathbf{A}_N \xrightarrow[N\to+\infty]{p} \mathbf{A} \ \text{ et } \ \left\|\mathbf{A}\right\| < +\infty \quad \Rightarrow \quad \mathbf{A}_N \times \mathbf{o}_p(1) = \mathbf{o}_p(1)$$

$$(iv)\quad \mathbf{A}_N \xrightarrow[N\to+\infty]{p} \mathbf{A} \quad \Rightarrow \quad \mathbf{A}_N + \mathbf{o}_p(1) = \mathbf{A} + \mathbf{o}_p(1)$$

$$(v)\quad \mathbf{a}_N \xrightarrow[N\to+\infty]{L} \mathcal{N}\left(\mathbf{m}, \mathbf{\Lambda}\right) \quad \Rightarrow \quad \mathbf{a}_N \times \mathbf{o}_p(1) = \mathbf{o}_p(1)$$

$$(vi)\quad \begin{array}{c} \mathbf{a}_N \xrightarrow[N\to+\infty]{L} \mathcal{N}\left(\mathbf{m}, \mathbf{\Lambda}\right) \ \text{ et } \ \mathbf{A}_N \xrightarrow[N\to+\infty]{p} \mathbf{A} \\ \Downarrow \\ \mathbf{A}_N \mathbf{a}_N \xrightarrow[N\to+\infty]{L} \mathcal{N}\left(\mathbf{A}\mathbf{m}, \mathbf{A}\mathbf{\Lambda}\mathbf{A}'\right) \end{array}$$

Calcul de la distribution as. de $\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N)$

La définition de $\tilde{\mathbf{a}}_N \equiv \hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N)$ en tant que minimande :

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

où :

$$\overline{\mathbf{g}}_N(\mathbf{a}) \equiv N^{-1}\sum_{i=1}^{N}\mathbf{g}_i(\mathbf{a})$$

implique que (CO1) :

$$\boxed{2\times\overline{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\overline{\mathbf{g}}_N(\tilde{\mathbf{a}}_N) = \mathbf{0}_{K\times 1}}$$

où :

$$\overline{\mathbf{G}}_N(\mathbf{a}) \equiv N^{-1}\sum_{i=1}^{N}\frac{\partial\mathbf{g}_i(\mathbf{a})}{\partial\mathbf{a}'} = N^{-1}\sum_{i=1}^{N}\mathbf{G}_i(\mathbf{a}).$$

On suppose ici que si le domaine de définition de $\mathbf{a}_0$ est borné, alors $\tilde{\mathbf{a}}_N$ ne se trouve pas sur cette borne.

En multipliant par $\sqrt{N}$ la CO1 pour ensuite invoquer un TCL, on obtient :

$$\boxed{\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\sqrt{N}\bar{\mathbf{g}}_N(\tilde{\mathbf{a}}_N)=\mathbf{0}_{K\times 1}.}$$

Il est important de voir ici que si :

$$\sqrt{N}\bar{\mathbf{g}}_N(\mathbf{a}_0)=\sqrt{N}N^{-1}\sum_{i=1}^{N}\mathbf{g}_i(\mathbf{a}_0)\xrightarrow[N\to+\infty]{L}\mathcal{N}\left(E\left[\mathbf{g}_i(\mathbf{a}_0)\right],E\left[\mathbf{g}_i(\mathbf{a}_0)\mathbf{g}_i(\mathbf{a}_0)'\right]\right),$$

par le TCL, la loi limite de $\sqrt{N}\bar{\mathbf{g}}_N(\tilde{\mathbf{a}}_N)$ n'est pas connue *a priori*.

C'est d'ailleurs ce qu'on va voir ici. L'estimateur $\tilde{\mathbf{a}}_N$ est une variable aléatoire, précisément celle dont on cherche la loi de distribution asymptotique.

En fait si :

$$p\lim_{N\to+\infty}\bar{\mathbf{g}}_N(\tilde{\mathbf{a}}_N)=p\lim_{N\to+\infty}\bar{\mathbf{g}}_N(\mathbf{a}_0)$$

on a, en général :

$$L\lim_{N\to+\infty}\sqrt{N}\bar{\mathbf{g}}_N(\tilde{\mathbf{a}}_N)\neq L\lim_{N\to+\infty}\sqrt{N}\bar{\mathbf{g}}_N(\mathbf{a}_0).$$

$$\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\sqrt{N}\bar{\mathbf{g}}_N(\tilde{\mathbf{a}}_N) = \mathbf{0}_{K\times 1}.$$

On sait que :

$$\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)\xrightarrow[N\to+\infty]{p}\mathbf{G}_0 \text{ et } \tilde{\mathbf{M}}_N\xrightarrow[N\to+\infty]{p}\mathbf{M}_0,$$

on sait donc que :

$$\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\times\mathbf{o}_p(1) = \mathbf{o}_p(1)$$

où $\mathbf{o}_p(1)$ est un « infiniment » petit en probabilité, *i.e.*

$$\mathbf{o}_p(1)\xrightarrow[N\to+\infty]{p}\mathbf{0}.$$

On sait de plus que que $\mathbf{G}_0$ $(G\times K)$ et $\mathbf{M}_0$ $(K\times K)$ sont de rang *K*.

$$\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\sqrt{N}\bar{\mathbf{g}}_N(\tilde{\mathbf{a}}_N)=\mathbf{0}_{K\times 1}$$

Le développement en probabilité à l'ordre 1 de $\sqrt{N}\bar{\mathbf{g}}_N(\tilde{\mathbf{a}}_N)$ en $\tilde{\mathbf{a}}_N$ autour de $\mathbf{a}_0$ (approximation assez frustre par ailleurs) donne que :

$$\boxed{\sqrt{N}\bar{\mathbf{g}}_N(\tilde{\mathbf{a}}_N)=\sqrt{N}\bar{\mathbf{g}}_N(\mathbf{a}_0)+\sqrt{N}\bar{\mathbf{G}}_N(\mathbf{a}_0)(\tilde{\mathbf{a}}_N-\mathbf{a}_0)+\mathbf{o}_p(1)}$$

L'intérêt de ce DL est de faire apparaître le terme $(\tilde{\mathbf{a}}_N-\mathbf{a}_0)$.

Par substitution de $\sqrt{N}\bar{\mathbf{g}}_N(\tilde{\mathbf{a}}_N)$ par son DL, on obtient :

$$\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\left[\sqrt{N}\bar{\mathbf{g}}_N(\mathbf{a}_0)+\sqrt{N}\bar{\mathbf{G}}_N(\mathbf{a}_0)(\tilde{\mathbf{a}}_N-\mathbf{a}_0)+\mathbf{o}_p(1)\right]=\mathbf{0}_{K\times 1}$$

Avec $\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\times\mathbf{o}_p(1)=\mathbf{o}_p(1)$, on obtient alors :

$$\boxed{\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\bar{\mathbf{G}}_N(\mathbf{a}_0)\times\sqrt{N}(\tilde{\mathbf{a}}_N-\mathbf{a}_0)=-\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\sqrt{N}\bar{\mathbf{g}}_N(\mathbf{a}_0)+\mathbf{o}_p(1).}$$

$$\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\bar{\mathbf{G}}_N(\mathbf{a}_0)\times\sqrt{N}(\tilde{\mathbf{a}}_N-\mathbf{a}_0)=-\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\sqrt{N}\bar{\mathbf{g}}_N(\mathbf{a}_0)+\mathbf{o}_p(1)$$

Il faut maintenant multiplier à gauche cette équation par l'inverse de la matrice $\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\bar{\mathbf{G}}_N(\mathbf{a}_0)$ pour « isoler » le terme $\sqrt{N}(\tilde{\mathbf{a}}_N-\mathbf{a}_0)$, si possible.

On sait que :

$$\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\bar{\mathbf{G}}_N(\mathbf{a}_0)\xrightarrow[N\to+\infty]{p}\mathbf{G}_0'\mathbf{M}_0\mathbf{G}_0$$

La matrice $\mathbf{G}_0'\mathbf{M}_0\mathbf{G}_0$ étant carrée de dimension K et les matrices $\mathbf{M}_0$ et $\mathbf{G}_0$ étant de rang K, la matrice $\mathbf{G}_0'\mathbf{M}_0\mathbf{G}_0$ est carrée de plein rang, et donc inversible. On a donc :

$$\left(\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\bar{\mathbf{G}}_N(\mathbf{a}_0)\right)^{-1}\xrightarrow[N\to+\infty]{p}(\mathbf{G}_0'\mathbf{M}_0\mathbf{G}_0)^{-1},$$

tout en sachant que $\left(\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\bar{\mathbf{G}}_N(\mathbf{a}_0)\right)^{-1}$ existe avec une probabilité approchant 1.

$$\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\bar{\mathbf{G}}_N(\mathbf{a}_0)\times\sqrt{N}(\tilde{\mathbf{a}}_N-\mathbf{a}_0)=-\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\sqrt{N}\bar{\mathbf{g}}_N(\mathbf{a}_0)+\mathbf{o}_p(1)$$

On a alors :

$$\left(\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\bar{\mathbf{G}}_N(\mathbf{a}_0)\right)^{-1}\times\mathbf{o}_p(1)=\mathbf{o}_p(1).$$

Par le TCL, on sait que :

$$\sqrt{N}\bar{\mathbf{g}}_N(\mathbf{a}_0)=\sqrt{N}N^{-1}\sum_{i=1}^N\mathbf{g}_i(\mathbf{a}_0)\xrightarrow[N\to+\infty]{L}\mathcal{N}\left(E\left[\mathbf{g}_i(\mathbf{a}_0)\right],E\left[\mathbf{g}_i(\mathbf{a}_0)\mathbf{g}_i(\mathbf{a}_0)'\right]\right)$$

Ceci donne que :

$$\boxed{\sqrt{N}(\tilde{\mathbf{a}}_N-\mathbf{a}_0)=-\left(\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\bar{\mathbf{G}}_N(\mathbf{a}_0)\right)^{-1}\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\sqrt{N}\bar{\mathbf{g}}_N(\mathbf{a}_0)+\mathbf{o}_p(1).}$$

Avec $E\left[\mathbf{g}_i(\mathbf{a}_0)\right]=\mathbf{0}_{G\times1}$ et $\mathbf{W}_0\equiv E\left[\mathbf{g}_i(\mathbf{a}_0)\mathbf{g}_i(\mathbf{a}_0)'\right]$ on a :

$$\sqrt{N}\bar{\mathbf{g}}_N(\mathbf{a}_0)\xrightarrow[N\to+\infty]{L}\mathcal{N}\left(\mathbf{0}_{G\times1},\mathbf{W}_0\right).$$

Avec :

$$\left(\overline{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\overline{\mathbf{G}}_N(\mathbf{a}_0)\right)^{-1}\overline{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N \xrightarrow[N\to+\infty]{p} (\mathbf{G}_0'\mathbf{M}_0\mathbf{G}_0)^{-1}\mathbf{G}_0'\mathbf{M}_0$$

et :

$$\sqrt{N}\,\overline{\mathbf{g}}_N(\mathbf{a}_0)\xrightarrow[N\to+\infty]{L}\mathcal{N}\left(\mathbf{0}_{G\times 1},\mathbf{W}_0\right)$$

on a :

$$\begin{aligned}\sqrt{N}(\tilde{\mathbf{a}}_N-\mathbf{a}_0)&=-\left(\overline{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\overline{\mathbf{G}}_N(\mathbf{a}_0)\right)^{-1}\overline{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\sqrt{N}\,\overline{\mathbf{g}}_N(\mathbf{a}_0)+\mathbf{o}_p(1)\\&\xrightarrow[N\to+\infty]{L}\mathbf{G}_0'\mathbf{M}_0\mathbf{G}_0)^{-1}\mathbf{G}_0'\mathbf{M}_0\times\mathcal{N}\left(\mathbf{0}_{G\times 1},\mathbf{W}_0\right)\end{aligned}$$

et finalement :

$$\boxed{\sqrt{N}(\tilde{\mathbf{a}}_N-\mathbf{a}_0)\xrightarrow[N\to+\infty]{L}\mathcal{N}\left(\mathbf{0}_{G\times 1},(\mathbf{G}_0'\mathbf{M}_0\mathbf{G}_0)^{-1}\mathbf{G}_0'\mathbf{M}_0\mathbf{W}_0\mathbf{M}_0\mathbf{G}_0(\mathbf{G}_0'\mathbf{M}_0\mathbf{G}_0)^{-1}\right).}$$

Avec :

$$\left(\overline{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\overline{\mathbf{G}}_N(\mathbf{a}_0)\right)^{-1}\overline{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N = (\mathbf{G}_0'\mathbf{M}_0\mathbf{G}_0)^{-1}\mathbf{G}_0'\mathbf{M}_0 + \mathbf{o}_p(1)$$

et :

$$\sqrt{N}\overline{\mathbf{g}}_N(\mathbf{a}_0)\xrightarrow[N\to+\infty]{L}\mathcal{N}\left(\mathbf{0}_{G\times 1},\mathbf{W}_0\right),$$

on peut écrire :

$$\sqrt{N}(\tilde{\mathbf{a}}_N-\mathbf{a}_0) = -\left(\overline{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\overline{\mathbf{G}}_N(\mathbf{a}_0)\right)^{-1}\overline{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\sqrt{N}\overline{\mathbf{g}}_N(\mathbf{a}_0)+\mathbf{o}_p(1),$$

sous la forme :

$$\boxed{\sqrt{N}(\tilde{\mathbf{a}}_N-\mathbf{a}_0) = -(\mathbf{G}_0'\mathbf{M}_0\mathbf{G}_0)^{-1}\mathbf{G}_0'\mathbf{M}_0\sqrt{N}\overline{\mathbf{g}}_N(\mathbf{a}_0)+\mathbf{o}_p(1).}$$

puisque $\sqrt{N}\overline{\mathbf{g}}_N(\mathbf{a}_0)\times\mathbf{o}_p(1)=\mathbf{o}_p(1)$ par la convergence en loi de $\sqrt{N}\overline{\mathbf{g}}_N(\mathbf{a}_0)$ et la convergence en probabilité de $\mathbf{o}_p(1)$ vers $\mathbf{0}$.

Remarque importante, définition de la matrice $\mathbf{W}_0$

On a ici défini la matrice $\mathbf{W}_0$ par :

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

Pourtant, la démonstration de la normalité as. des estimateurs MMG montre que :

$\boxed{\mathbf{W}_0 \text{ est la matrice de variance-covariance de la loi limite de } \sqrt{N}\overline{\mathbf{g}}_N(\mathbf{a}_0).}$

Dans les cas considérés ici on a bien $\mathbf{W}_0 = E\left[\mathbf{g}_i(\mathbf{a}_0)\mathbf{g}_i(\mathbf{a}_0)'\right]$ car :

$$\sqrt{N}\overline{\mathbf{g}}_N(\mathbf{a}_0) \xrightarrow[N\to+\infty]{L} \mathcal{N}\left(\mathbf{0}_{G\times 1}, E\left[\mathbf{g}_i(\mathbf{a}_0)\mathbf{g}_i(\mathbf{a}_0)'\right]\right).$$

On n'a cependant pas toujours $\mathbf{W}_0 = E\left[\mathbf{g}_i(\mathbf{a}_0)\mathbf{g}_i(\mathbf{a}_0)'\right]$, notamment lorsqu'on travaille avec des estimateurs construits par étapes.

Estimation par étapes :

On utilise parfois des conditions de moment de la forme :

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

pour estimer $\mathbf{a}_0$.

Si valeur du paramètre auxiliaire $\mathbf{b}_0$ n'est pas connue, alors on utilise un estimateur as. normal de $\mathbf{b}_0$, $\tilde{\mathbf{b}}_N$.

Cet estimateur sert à définir la contrepartie empirique de $E\left[\mathbf{g}_i(\mathbf{a};\mathbf{b}_0)\right]$ sous la forme $\overline{\mathbf{g}}_N(\mathbf{a}_0;\tilde{\mathbf{b}}_N)$ puisque :

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

On a alors :

$$\sqrt{N}\overline{\mathbf{g}}_N(\mathbf{a}_0;\tilde{\mathbf{b}}_N)\xrightarrow[N\to+\infty]{L}\mathcal{N}\left(\mathbf{0}_{G\times 1},\mathbf{W}_0\right)$$

mais en général on a :

$$\mathbf{W}_0\neq E\left[\mathbf{g}_i(\mathbf{a}_0;\mathbf{b}_0)\mathbf{g}_i(\mathbf{a}_0;\mathbf{b}_0)'\right].$$

Remarque importante, cas linéaire

Dans certains cas, on peut déterminer explicitement $\tilde{\mathbf{a}}_N$. De même on peut écrire directement $\sqrt{N}(\tilde{\mathbf{a}}_N - \mathbf{a}_0)$ sous forme explicite, sans recours à un DL, *i.e.* sans approximation. C'est le cas, très fréquent en pratique, où on considère des équations de la forme :

$$y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i \quad \text{avec} \quad E\left[u_i / \mathbf{z}_i\right] = 0$$

ou des systèmes d'équations de la forme :

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

avec des conditions d'orthogonalité estimantes de la forme :

$$E\left[\mathbf{r}_i(y_i - \mathbf{x}_i'\mathbf{a}_0)\right] = \mathbf{0} \quad \text{avec} \quad \mathbf{r}_i \equiv \mathbf{r}(\mathbf{z}_i)$$

ou :

$$E\left[\mathbf{R}_i(\mathbf{y}_i - \mathbf{X}_i'\mathbf{a}_0)\right] = \mathbf{0} \quad \text{avec} \quad \mathbf{R}_i \equiv \mathbf{R}(\mathbf{z}_i).$$

Dans ce cas, on a :

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

ou :

$$\tilde{\mathbf{a}}_N = \left[\left(N^{-1} \sum_{i=1}^{N} \mathbf{R}_i \mathbf{X}_i' \right) \tilde{\mathbf{M}}_N \left(N^{-1} \sum_{i=1}^{N} \mathbf{R}_i \mathbf{X}_i' \right) \right]^{-1} \times \left(N^{-1} \sum_{i=1}^{N} \mathbf{R}_i \mathbf{X}_i' \right) \tilde{\mathbf{M}}_N N^{-1} \sum_{i=1}^{N} \mathbf{R}_i \mathbf{y}_i .$$

et :

$$\sqrt{N} (\tilde{\mathbf{a}}_N - \mathbf{a}_0) = \left[\left(N^{-1} \sum_{i=1}^{N} \mathbf{R}_i \mathbf{X}_i' \right) \tilde{\mathbf{M}}_N \left(N^{-1} \sum_{i=1}^{N} \mathbf{R}_i \mathbf{X}_i' \right) \right]^{-1} \times \left(N^{-1} \sum_{i=1}^{N} \mathbf{R}_i \mathbf{X}_i' \right) \tilde{\mathbf{M}}_N \sqrt{N} N^{-1} \sum_{i=1}^{N} \mathbf{R}_i \mathbf{u}_i .$$

On ne s’étend pas sur ce cas ici. Il sera traité en détails par la suite.

3.4. Estimateurs MMG optimaux

$\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N)$ est un ***estimateur MMG optimal*** de $\mathbf{a}_0$ s'il n'existe pas d'autre estimateur de la MMG de $\mathbf{a}_0$ fondé sur $E\left[\mathbf{g}_i(\mathbf{a}_0)\right]=\mathbf{0}$ qui soit plus efficace.

- Les propriétés précédentes indiquent que, s'ils existent, les estimateurs MMG ***optimaux*** de $\mathbf{a}_0$ fondés sur la condition de moment $E\left[\mathbf{g}_i(\mathbf{a}_0)\right]=\mathbf{0}$ sont :
  - caractérisés par le ***choix optimal de*** $\mathbf{M}_0$, disons $\mathbf{M}_0^*$. En effet on a :

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

    ce qui implique que seul le choix de $\mathbf{M}_0$ est susceptible de modifier la précision des estimateurs MMG de $\mathbf{a}_0$ fondés sur $E\left[\mathbf{g}_i(\mathbf{a}_0)\right]=\mathbf{0}$.
- Les propriétés qui suivent montrent que (*i*) ***il existe des estimateurs MMG optimaux*** et (*ii*) ces estimateurs sont ***toujours calculables***

$\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N^*)$ est ***optimal dans la classe des estimateurs MMG de*** $\mathbf{a}_0$ ***fondés sur*** $E\left[\mathbf{g}_i(\mathbf{a}_0)\right] = \mathbf{0}$

si :

$$(i)\ \tilde{\mathbf{M}}_N^* \xrightarrow[N \to +\infty]{p} \mathbf{M}_0^*$$

et

(*ii*) s'il n'existe pas de matrice $\mathbf{M}_0$ définie positive telle que $\boldsymbol{\Sigma}(\mathbf{M}_0) << \boldsymbol{\Sigma}(\mathbf{M}_0^*)$.

- On ne démontrera pas le résultat suivant :
  - Son interprétation est simple et intuitive
  - La démonstration est essentiellement un exercice d'algèbre linéaire.

Propriété. ***Estimateurs MMG optimaux de*** $\mathbf{a}_0$ ***fondés sur*** $E\left[\mathbf{g}_i(\mathbf{a}_0)\right]=\mathbf{0}$

Si :

(*i*) $\mathbf{W}_0 \equiv E\left[\mathbf{g}_i(\mathbf{a}_0)\mathbf{g}_i(\mathbf{a}_0)'\right] = V\left[\mathbf{g}_i(\mathbf{a}_0)\right]$ est inversible

et :

(*ii*) $rang\mathbf{G}_0 = K$ où $\mathbf{G}_0 \equiv E\left[\partial\mathbf{g}_i(\mathbf{a}_0)/\partial\mathbf{a}'\right]$,

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

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

où :

$$\tilde{\mathbf{W}}_N \xrightarrow[N\to+\infty]{p} \mathbf{W}_0.$$

On a :

$$\sqrt{N}(\hat{\mathbf{a}}_N^{MMG} - \mathbf{a}_0) \xrightarrow[N\to+\infty]{L} \mathcal{N}(\mathbf{0};\mathbf{\Sigma}_0) \quad \text{avec } \mathbf{\Sigma}_0 \equiv (\mathbf{G}_0'\mathbf{W}_0^{-1}\mathbf{G}_0)^{-1}.$$

$$\boxed{\hat{\mathbf{a}}_N^{MMG} \equiv \arg\min_{\mathbf{a}} \overline{\mathbf{g}}_N(\mathbf{a})' \tilde{\mathbf{W}}_N^{-1} \overline{\mathbf{g}}_N(\mathbf{a}) \text{ et } \tilde{\mathbf{W}}_N^{-1} \xrightarrow[N\to+\infty]{p} \mathbf{W}_0^{-1}}$$

Interprétation.

- La matrice $\mathbf{M}_0$ optimale est $\mathbf{W}_0^{-1} \equiv E\left[\mathbf{g}_i(\mathbf{a}_0)\mathbf{g}_i(\mathbf{a}_0)'\right]^{-1}$. C'est l'inverse de la matrice de variance-covariance des équations estimantes :

$$\mathbf{M}_0^* = \mathbf{W}_0^{-1} \equiv E\left[\mathbf{g}_i(\mathbf{a}_0)\mathbf{g}_i(\mathbf{a}_0)'\right]^{-1} = V\left[\mathbf{g}_i(\mathbf{a}_0)\right]^{-1}.$$

- La matrice de pondérations $\mathbf{W}_0^{-1}$ donne un poids d'autant plus élevé à un élément du vecteur $E\left[\mathbf{g}_i(\mathbf{a}_0)\right]$ que cet élément a une petite variance.

- Cela garantit l'efficacité de $\hat{\mathbf{a}}_N^{MMG}$:

- Les conditions de moment les plus « précises » sont « sur-pondérées » par rapport aux conditions de moment les moins précises dans le critère de construction de $\hat{\mathbf{a}}_N^{MMG}$.
  - C’est la logique de pondération qu’on trouve dans les estimateurs des MC Généralisés et MC Quasi-Généralisés.
    - Ces estimateurs pondèrent les observations (issues de variables iid) par l’inverse de la variance de leur terme d’erreur dans le critère des MC.

- La condition (*i*), $\mathbf{W}_0 \equiv E\left[\mathbf{g}_i(\mathbf{a}_0)\mathbf{g}_i(\mathbf{a}_0)'\right]$ est inversible, est satisfaite si les conditions de moments contenues dans $E\left[\mathbf{g}_i(\mathbf{a}_0)\right] = \mathbf{0}$ ne sont pas (linéairement) redondantes.
    - Cette condition n'est pas nécessaire, il suffit que $rang\mathbf{W}_0 = K$.
    - Cette condition est une forme de ***condition d'ordre***.

- La condition (*ii*) $rang\mathbf{G}_0 = rangE\left[\partial \mathbf{g}_i(\mathbf{a}_0)/\partial \mathbf{a}'\right] = K$ est une ***condition locale d'identification*** de $\mathbf{a}_0$ par $E\left[\mathbf{g}_i(\mathbf{a}_0)\right] = \mathbf{0}$.
    - Cette condition indique que $E\left[\mathbf{g}_i(\mathbf{a}_0)\right]$ doit suffisamment dépendre de $\mathbf{a}$ au voisinage de $\mathbf{a}_0$ (*i.e.* localement) pour permettre l'identification de $\mathbf{a}_0$.
    - Cette condition est une forme de ***condition de rang*** (locale)

Calcul par étapes des estimateurs de la MMG optimaux

- La matrice de pondération optimale est l'inverse de la matrice :

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

Cette matrice dépend en général de $\mathbf{a}_0$, qui est le paramètre à estimer.

- C'est ici qu'intervient l'intérêt des propriétés génériques des estimateurs de la MMG :

Quelque soit $\tilde{\mathbf{M}}_N \xrightarrow[N\to+\infty]{p} \mathbf{M}_0$ telle $\mathbf{M}_0$ est une matrice définie positive, l'estimateur $\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N)$ sera convergent quoique non optimal.

Le calcul de $\hat{\mathbf{a}}_N^{MMG}$ se fait généralement en trois étapes, les deux premières ayant pour objectif de construire un estimateur convergent de $\mathbf{W}_0^{-1}$ (Etape 2) à partir d'un estimateur convergent mais non-optimal de $\mathbf{a}_0$ (Etape 1).

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 $\mathbf{a}_0$:

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

Rmq. On itère parfois les étapes 2 et 3, ce qui n'a pas d'intérêt … en théorie.

Néanmoins :

Cette procédure est souvent employée en pratique …

Si cette procédure ne converge pas rapidement, …
… c'est souvent mauvais signe.

Calcul d'un estimateur de la variance as. de $\hat{\mathbf{a}}_N^{MMG}$, $\boldsymbol{\Sigma}_0$

Procédure : ***Calcul de*** $\hat{\boldsymbol{\Sigma}}_N \xrightarrow[N\to+\infty]{p} \boldsymbol{\Sigma}_0$

On a :

$$\hat{\boldsymbol{\Sigma}}_N \equiv (\hat{\mathbf{G}}_N' \hat{\mathbf{W}}_N^{-1} \hat{\mathbf{G}}_N)^{-1} \xrightarrow[N\to+\infty]{p} \boldsymbol{\Sigma}_0 \equiv (\mathbf{G}_0' \mathbf{W}_0^{-1} \mathbf{G}_0)^{-1}$$

avec :

$$\hat{\mathbf{G}}_N \equiv N^{-1} \sum_{i=1}^{N} \partial \mathbf{g}_i(\hat{\mathbf{a}}_N^{MMG}) / \partial \mathbf{a}' \xrightarrow[N\to+\infty]{p} \mathbf{G}_0 \equiv E\left[\partial \mathbf{g}_i(\mathbf{a}_0)/\partial \mathbf{a}'\right]$$

et :

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

Rmq. On « re-estime » $\mathbf{W}_0$ par $\hat{\mathbf{W}}_N$, *i.e.* en utilisant l'estimateur de $\mathbf{a}_0$ disponible le plus efficace, $\hat{\mathbf{a}}_N^{MMG}$.

3.5. Estimateurs MMG et nombre de conditions de moment

- On montre ici que l'***efficacité <u>asymptotique</u> des estimateurs MMG croît avec le nombre de conditions de moment estimantes*** employées pour les construire

- Cette propriété est une conséquence de ce que :

    - dès lors qu'elles sont valides, les conditions estimantes constituent un ***apport d'information théorique*** pour l'estimation des paramètres d'intérêt.

    - Cet apport d'information se traduisant par une ***précision d'estimation accrue***.

Propriété. ***Estimateurs MMG (optimaux) et conditions de moment***

On considère ici que :

$$(i)\quad E\left[\mathbf{g}_i(\mathbf{a})\right] = E\left[\begin{bmatrix}\mathbf{g}_{1,i}(\mathbf{a})\\ \mathbf{g}_{2,i}(\mathbf{a})\end{bmatrix}\right] = \mathbf{0}_{(G_1+G_2)\times 1} \quad \Leftrightarrow \quad \mathbf{a} = \mathbf{a}_0,$$

$$(ii)\quad E\left[\mathbf{g}_{1,i}(\mathbf{a})\right] = \mathbf{0}_{G_1\times 1} \quad \Leftrightarrow \quad \mathbf{a} = \mathbf{a}_0,$$

(*iii*) les conditions d'identification locales et les conditions de régularité usuelles pour le calcul des estimateurs MMG définis sont satisfaites.

L'estimateur MMG optimal fondé sur $E\left[\mathbf{g}_i(\mathbf{a}_0)\right] = \mathbf{0}$ est noté $\hat{\mathbf{a}}_N$, celui fondé sur $E\left[\mathbf{g}_{1,i}(\mathbf{a}_0)\right] = \mathbf{0}$ est noté $\hat{\mathbf{a}}_N^1$. On a alors :

$$\sqrt{N}(\hat{\mathbf{a}}_N - \mathbf{a}_0) \xrightarrow[N\to+\infty]{L} \mathcal{N}\left(\mathbf{0};\mathbf{\Sigma}_0(\hat{\mathbf{a}}_N)\right) \text{ avec } \mathbf{\Sigma}_0(\hat{\mathbf{a}}_N) \equiv (\mathbf{G}_0'\mathbf{W}_0^{-1}\mathbf{G}_0)^{-1}$$

et :

$$\sqrt{N}(\hat{\mathbf{a}}_N^1 - \mathbf{a}_0) \xrightarrow[N\to+\infty]{L} \mathcal{N}\left(\mathbf{0};\mathbf{\Sigma}_0(\hat{\mathbf{a}}_N^1)\right) \text{ avec } \mathbf{\Sigma}_0(\hat{\mathbf{a}}_N^1) \equiv (\mathbf{G}_{1,0}'\mathbf{W}_{11,0}^{-1}\mathbf{G}_{1,0})^{-1},$$

en notant :

$$\mathbf{W}_0 \equiv E\left[\mathbf{g}_i(\mathbf{a}_0)\mathbf{g}_i(\mathbf{a}_0)'\right] \equiv \begin{bmatrix} \mathbf{W}_{11,0} & \mathbf{W}_{12,0} \\ \mathbf{W}'_{12,0} & \mathbf{W}_{22,0} \end{bmatrix},$$

et :

$$\mathbf{G}_0 \equiv E\left[\partial \mathbf{g}_i(\mathbf{a}_0)/\partial \mathbf{a}'\right] \equiv \begin{bmatrix} \mathbf{G}_{1,0} \\ \mathbf{G}_{2,0} \end{bmatrix}.$$

On montre alors que :

$$\begin{aligned} \boldsymbol{\Sigma}_0(\hat{\mathbf{a}}_N)^{-1} - \boldsymbol{\Sigma}_0(\hat{\mathbf{a}}_N^1)^{-1} = {} & (\mathbf{G}_{2,0} - \mathbf{W}'_{12,0}\mathbf{W}_{11,0}^{-1}\mathbf{G}_{1,0})' \\ & \times(\mathbf{W}_{22,0} - \mathbf{W}'_{12,0}\mathbf{W}_{11,0}^{-1}\mathbf{W}_{12,0})^{-1}(\mathbf{G}_{2,0} - \mathbf{W}'_{12,0}\mathbf{W}_{11,0}^{-1}\mathbf{G}_{1,0}) \end{aligned}$$

et donc que $\boldsymbol{\Sigma}_0(\hat{\mathbf{a}}_N)$ est plus petite que $\boldsymbol{\Sigma}_0(\hat{\mathbf{a}}_N^1)$ dans le pré-ordre des matrices définies positives (l'inverse de $\boldsymbol{\Sigma}_0(\hat{\mathbf{a}}_N)$ est plus grande que celle de $\boldsymbol{\Sigma}_0(\hat{\mathbf{a}}_N^1)$).

Conclusion. Plus on utilise de conditions estimantes, … plus l'estimateur construit est as. efficace. ***Problème*** : asymptotique = théorique

Dem. Il suffit d'utiliser les formules d'inversion par blocs et de remarquer que $(\mathbf{W}_{22,0} - \mathbf{W}'_{12,0}\mathbf{W}_{11,0}^{-1}\mathbf{W}_{12,0})^{-1}$ est définie positive puisque $\mathbf{W}_0$ est une matrice de variance-covariance définie positive.

On utilise alors les résultats suivants :

$$(i)\quad \begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}' & \mathbf{B} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I} & -\mathbf{A}^{-1}\mathbf{C} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{B} - \mathbf{C}'\mathbf{A}^{-1}\mathbf{C})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{C}'\mathbf{A}^{-1} & \mathbf{I} \end{bmatrix},$$

$$(ii)\quad V\left[\mathbf{w}_i\right] \equiv \mathbf{A} \equiv \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} \end{bmatrix} \text{ est déf. pos. } \Rightarrow \mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \text{ est déf. pos.}$$

$$(iii)\quad \mathbf{A} \text{ est déf. pos. } \Leftrightarrow \mathbf{A}^{-1} \text{ est déf. pos.}$$

$$(iv)\quad \mathbf{A} \text{ est déf. pos. } \Leftrightarrow \mathbf{B}'\mathbf{A}\mathbf{B} \text{ est semi-déf. pos. } \forall\ \mathbf{B} \text{ de taille conforme}$$

- Cependant, il convient d'être prudent
    - Plus on spécifie de conditions, plus on risque d'en spécifier d'invalides, notamment en raison des choix de VI.
    - Utiliser des VI peu informative expose au problème dit des ***instruments faibles***.
    - Les estimateurs de la MMG sont ***biaisés à distance finie***. Et le biais croît mécaniquement avec la sur-identification des conditions estimantes

Conclusion

Ne pas « râcler les fonds de tiroir » pour trouver des VI ou des instruments, (*i*) n'utiliser que celles dont on est sûr de l'exogénéité et (*ii*) celles qui sont suffisamment informatives.

4. Les estimateurs des 2MC en tant qu'estimateurs de la MMG

- Deux objectifs ici :
  - Illustrer l'utilisation des résultats précédents
  - Discuter de l'estimation par le 2MC en général

- Présentation en 3 étapes :
  - Exploiter la condition d'orthogonalité sur-identifiante d'un modèle linéaire à VI à termes d'erreur homoscédastiques … et retrouver les 2MC
  - Considérer le cas des modèles non-linéaires … et discuter des MCO successifs
  - Exploiter la condition d'orthogonalité sur-identifiante d'un modèle linéaire à VI à termes d'erreur potentiellement hétéroscédastiques … et définir 2MCH

4.1. Modèle à VI à termes d'erreur homoscédastiques, MMG et 2MC

- On considère à nouveau le modèle à VI à termes d'erreur homoscédastiques :

$$\boxed{y_i = \mathbf{a}_0'\mathbf{x}_i + u_i \quad \text{avec} \quad E\left[u_i / \mathbf{z}_i\right] = E\left[u_i\right] \equiv 0 \text{ et } E\left[u_i^2 / \mathbf{z}_i\right] = \sigma_0^2,}$$

et on souhaite estimer $\mathbf{a}_0$ à partir de la condition d'orthogonalité :

$$E\left[\mathbf{z}_i(y_i - \mathbf{a}_0'\mathbf{x}_i)\right] = \mathbf{0}_{L\times 1}.$$

- En fait, ici on a :

$$\mathbf{g}_i(\mathbf{a}) = \mathbf{z}_i(y_i - \mathbf{a}'\mathbf{x}_i)$$

- Si $L > K$, l'estimateur des VI n'est pas utilisable puisque $\mathbf{a}_0$ est sur-identifié par les conditions de moment estimantes employées.

- On peut directement employer $\hat{\mathbf{a}}_N^{2MC}$ mais l'objectif est ici de montrer que dans ce cas $\hat{\mathbf{a}}_N^{MMG} = \hat{\mathbf{a}}_N^{2MC}$ en sus d'illustrer l'utilisation de la MMG.

Première chose à faire : l'examen de la forme de $\mathbf{W}_0 \equiv E\left[\mathbf{g}_i(\mathbf{a}_0)\mathbf{g}_i(\mathbf{a}_0)'\right]$

- C'est la forme de $\mathbf{W}_0 \equiv E\left[\mathbf{g}_i(\mathbf{a}_0)\mathbf{g}_i(\mathbf{a}_0)'\right]$ qui détermine la forme de l'estimateur MMG optimal.

- On a ici :

$$\mathbf{W}_0 \equiv E\left[\mathbf{z}_i(y_i - \mathbf{a}_0'\mathbf{x}_i)^2\mathbf{z}_i'\right] = E\left[u_i^2\mathbf{z}_i\mathbf{z}_i'\right],$$

et, étant donnée l'homoscédasticité (conditionnelle) des u_i :

$$\boxed{\mathbf{W}_0 = E\left[u_i^2\mathbf{z}_i\mathbf{z}_i'\right] = \sigma_0^2 E\left[\mathbf{z}_i\mathbf{z}_i'\right].}$$

- Bien entendu, σ_0^2 est inconnu mais cela ne pose pas de problème outre mesure comme cela sera vu par la suite. On définit ici $\tilde{\mathbf{W}}_N$ par:

$$\boxed{\tilde{\mathbf{W}}_N \equiv \sigma_0^2 \times N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i' \xrightarrow[N\to+\infty]{p} \mathbf{W}_0.}$$

- On a alors :

$$\hat{\mathbf{a}}_N^{MMG} \equiv \arg\min_{\mathbf{a}} N^{-1}\sum_{i=1}^{N}(y_i - \mathbf{x}_i'\mathbf{a})\mathbf{z}_i'\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1} N^{-1}\sum_{i=1}^{N}\mathbf{z}_i(y_i - \mathbf{x}_i'\mathbf{a})\times\sigma_0^{-2}$$

et, puisque σ_0^{-2} est un scalaire strictement positif :

$$\hat{\mathbf{a}}_N^{MMG} \equiv \arg\min_{\mathbf{a}} N^{-1}\sum_{i=1}^{N}(y_i - \mathbf{x}_i'\mathbf{a})\mathbf{z}_i'\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1} N^{-1}\sum_{i=1}^{N}\mathbf{z}_i(y_i - \mathbf{x}_i'\mathbf{a}).$$

- Les CO1 du programme de minimisation définissant $\hat{\mathbf{a}}_N^{MMG}$ donnent, à l’optimum :

$$-2\times\left(N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right)\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1} N^{-1}\sum_{i=1}^{N}\mathbf{z}_i(y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{MMG}) = \mathbf{0}$$

et, après simplifications :

$$\boxed{\begin{aligned}\mathbf{a}_N^{MMG} = \Bigg[\left(N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right)\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i'\right)\Bigg]^{-1} \\ \times\left(N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i'\right)\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1} N^{-1}\sum_{i=1}^{N}\mathbf{z}_i y_i = \mathbf{a}_N^{2MC}\end{aligned}}$$

- On peut par ailleurs vérifier que la distribution as. de $\mathbf{a}_N^{MMG} = \mathbf{a}_N^{2MC}$ donnée par la MMG est identique à celle obtenue précédemment pour $\mathbf{a}_N^{2MC}$.

- La MMG donne :

$$\sqrt{N}(\hat{\mathbf{a}}_N^{MMG} - \mathbf{a}_0) \xrightarrow[N\to+\infty]{L} \mathcal{N}(\mathbf{0};\boldsymbol{\Sigma}_0) \quad \text{avec } \boldsymbol{\Sigma}_0 \equiv (\mathbf{G}_0'\mathbf{W}_0^{-1}\mathbf{G}_0)^{-1}$$

et ici on a :

$$\mathbf{G}_0 \equiv E\left[\frac{\partial \mathbf{g}(\mathbf{w}_i;\mathbf{a}_0)}{\partial \mathbf{a}'}\right] = E\left[\frac{\partial\left[\mathbf{z}_i(y_i - \mathbf{a}_0'\mathbf{x}_i)\right]}{\partial \mathbf{a}'}\right] = -E\left[\mathbf{x}_i'\mathbf{z}_i\right] \ \text{ et } \ \mathbf{W}_0 = \sigma_0^2 E\left[\mathbf{z}_i\mathbf{z}_i'\right],$$

- Ce qui donne :

$$\mathbf{\Sigma}_0 \equiv (\mathbf{G}_0'\mathbf{W}_0^{-1}\mathbf{G}_0)^{-1} = \sigma_0^2 \left\{ E[\mathbf{x}_i\mathbf{z}_i'] E[\mathbf{z}_i\mathbf{z}_i']^{-1} E[\mathbf{z}_i\mathbf{x}_i'] \right\}^{-1}$$

et on retrouve bien ici le fait que :

$$\boxed{\sqrt{N}(\hat{\mathbf{a}}_N^{2MC} - \mathbf{a}_0) \xrightarrow[N\to+\infty]{L} \mathcal{N}\left(\mathbf{0}; \sigma_0^2 \left(E[\mathbf{x}_i\mathbf{z}_i'] E[\mathbf{z}_i\mathbf{z}_i']^{-1} E[\mathbf{z}_i\mathbf{x}_i'] \right)^{-1} \right)}$$

dans le modèle à VI à termes d’erreur homoscédastiques.

4.2. Modèle non-linéaire à VI, MMG et 2MC

Un ***estimateur des 2MC non-linéaires*** a été introduit dans la littérature avant le développement de la MMG.

- C'est également un estimateur de la MMG, pour le modèle suivant :

$$y_i = f(\mathbf{x}_i;\mathbf{a}_0) + u_i \ \text{ avec } \ E\left[u_i/\mathbf{z}_i\right] = E\left[u_i\right] \equiv 0 \ \text{ et } \ E\left[u_i^2/\mathbf{z}_i\right] = \sigma_0^2,$$

i.e. un modèle à VI à termes d'erreur homoscédastiques et potentiellement non-linéaire dans ses paramètres.

Exemples

$$y_i = \exp(\mathbf{x}_i'\mathbf{a}_0) + u_i$$

$$y_i = \mathbf{x}_i'\mathbf{b}(\mathbf{a}_0) + u_i \ \text{ et } \ \mathbf{b}(.) \text{ de forme connue}$$

- On souhaite estimer $\mathbf{a}_0$ à partir de la condition d'orthogonalité :

$$E\left[\mathbf{z}_i\left(y_i - f(\mathbf{x}_i;\mathbf{a}_0)\right)\right] = \mathbf{0}_{L\times 1}.$$

i.e., avec $\mathbf{g}_i(\mathbf{a}) \equiv \mathbf{z}_i\left(y_i - f(\mathbf{x}_i;\mathbf{a}_0)\right)$.

- Si $L > K$, ce qu'on suppose ici, la MM n'est pas utilisable puisque $\mathbf{a}_0$ est sur-identifié par les conditions de moment estimantes employées.

Première chose à faire : l'examen de la forme de $\mathbf{W}_0 \equiv E\left[\mathbf{g}_i(\mathbf{a}_0)\mathbf{g}_i(\mathbf{a}_0)'\right]$

- C'est la forme de $\mathbf{W}_0 \equiv E\left[\mathbf{g}_i(\mathbf{a}_0)\mathbf{g}_i(\mathbf{a}_0)'\right]$ qui détermine la forme de l'estimateur MMG optimal.

- On a ici :

$$\mathbf{W}_0 \equiv E\left[\mathbf{z}_i\left(y_i - f(\mathbf{x}_i;\mathbf{a}_0)\right)\mathbf{z}_i'\right] = E\left[u_i^2\mathbf{z}_i\mathbf{z}_i'\right],$$

et, étant donnée l'homoscédasticité (conditionnelle) des u_i :

$$\boxed{\mathbf{W}_0 = \sigma_0^2 E\left[\mathbf{z}_i\mathbf{z}_i'\right].}$$

- Bien entendu, σ_0^2 est inconnu mais, comme cela a été vu précédemment, ça ne pose pas de problème outre mesure. On définit ici $\tilde{\mathbf{W}}_N$ par:

$$\boxed{\tilde{\mathbf{W}}_N \equiv \sigma_0^2 \times N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i' \xrightarrow[N\to+\infty]{p} \mathbf{W}_0.}$$

- On a alors :

$$\hat{\mathbf{a}}_N^{MMG} \equiv \arg\min_{\mathbf{a}} \left\{ \begin{array}{c} N^{-1}\sum_{i=1}^{N}\left(y_i - f(\mathbf{x}_i;\mathbf{a})\right)\mathbf{z}_i' \\ \times\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1} \\ \times N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\left(y_i - f(\mathbf{x}_i;\mathbf{a})\right) \end{array} \right\} = \hat{\mathbf{a}}_N^{2MCNL}.$$

- Ici $\hat{\mathbf{a}}_N^{MMG} = \hat{\mathbf{a}}_N^{2MCNL}$ ne peut être exprimé sous forme explicite, les CO1 donnent, à l'optimum :

$$\left(N^{-1}\sum_{i=1}^{N}\mathbf{f}_i(\hat{\mathbf{a}}_N^{MMG})\mathbf{z}_i'\right)\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1} N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\left(y_i - f_i(\hat{\mathbf{a}}_N^{MMG})\right) = \mathbf{0}$$

avec :

$$f_i(\mathbf{a}) \equiv f(\mathbf{x}_i;\mathbf{a}),\ \mathbf{f}_i(\mathbf{a}) \equiv \partial f_i(\mathbf{a})/\partial\mathbf{a} \text{ et donc } \mathbf{G}_i(\mathbf{a}) = -\mathbf{z}_i\mathbf{f}_i(\mathbf{a})'.$$

Ceci dit les propriétés des estimateurs de la MMG donnent que $\hat{\mathbf{a}}_N^{2MCNL}$ est convergent et as. normal avec :

$$\sqrt{N}(\hat{\mathbf{a}}_N^{MMG} - \mathbf{a}_0) \xrightarrow[N\to+\infty]{L} \mathcal{N}(\mathbf{0};\boldsymbol{\Sigma}_0) \text{ avec } \boldsymbol{\Sigma}_0 \equiv (\mathbf{G}_0'\mathbf{W}_0^{-1}\mathbf{G}_0)^{-1}$$

et ici on a :

$$\mathbf{G}_0 \equiv E\left[\mathbf{G}_i(\mathbf{a}_0)\right] = -E\left[\mathbf{z}_i\mathbf{f}_i(\mathbf{a}_0)'\right] \text{ et } \mathbf{W}_0 = \sigma_0^2 E\left[\mathbf{z}_i\mathbf{z}_i'\right],$$

- Ce qui donne :

$$\boldsymbol{\Sigma}_0 \equiv (\mathbf{G}_0'\mathbf{W}_0^{-1}\mathbf{G}_0)^{-1} = \sigma_0^2\left(E\left[\mathbf{f}_i(\mathbf{a}_0)\mathbf{z}_i'\right]E\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}E\left[\mathbf{z}_i\mathbf{f}_i(\mathbf{a}_0)'\right]\right)^{-1}.$$

- ***Remarques* « *techniques* » :**

  - Le problème de minimisation définissant $\hat{\mathbf{a}}_N^{2MCNL}$ n'a pas nécessairement une solution unique.
    Si une solution est obtenue : (*i*) vérifier que c'est bien un minimum et (*ii*) recommencer la procédure d'optimisation avec différentes valeurs initiales pour $\mathbf{a}_0$. Si ça converge toujours vers le même minimum, c'est bon. Sinon, choisir le minimum des minimas obtenus.

  - L'estimateur $\hat{\mathbf{a}}_N^{2MCNL}$ tel qu'il est défini ici est peu utilisé en pratique.
    L'idée est que si on a :

$$y_i = f(\mathbf{x}_i;\mathbf{a}_0) + u_i \quad \text{avec} \quad E\left[u_i / \mathbf{z}_i\right] = E\left[u_i\right] \equiv 0$$

alors il existe des conditions de moment plus « riches » que :

$$E\left[\mathbf{z}_i\left(y_i - f(\mathbf{x}_i;\mathbf{a}_0)\right)\right] = \mathbf{0}$$

pour construire un estimateur de la MMG de $\mathbf{a}_0$.

4.3. Modèle à VI à termes d’erreur hétéroscédastiques, MMG et 2MCH

- On considère le modèle à VI à termes d’erreur homoscédastiques :

$$y_i = \mathbf{a}_0'\mathbf{x}_i + u_i \quad \text{avec } E\left[u_i / \mathbf{z}_i\right] = E\left[u_i\right] \equiv 0 \text{ et } E\left[u_i^2 / \mathbf{z}_i\right] = ?,$$

et on souhaite estimer $\mathbf{a}_0$ à partir de la condition d’orthogonalité :

$$E\left[\mathbf{z}_i(y_i - \mathbf{a}_0'\mathbf{x}_i)\right] = \mathbf{0}_{L\times1}.$$

- En fait, ici on a :

$$\mathbf{g}_i(\mathbf{a}) = \mathbf{z}_i(y_i - \mathbf{a}'\mathbf{x}_i)$$

- Si $L > K$, ce qu’on suppose ici, l’estimateur des VI n’est pas utilisable puisque $\mathbf{a}_0$ est sur-identifié par les conditions de moment estimantes employées.

Première chose à faire : l'examen de la forme de $\mathbf{W}_0 \equiv E\left[\mathbf{g}_i(\mathbf{a}_0)\mathbf{g}_i(\mathbf{a}_0)'\right]$

- C'est la forme de $\mathbf{W}_0 \equiv E\left[\mathbf{g}_i(\mathbf{a}_0)\mathbf{g}_i(\mathbf{a}_0)'\right]$ qui détermine la forme de l'estimateur MMG optimal.
- On a ici :

$$\mathbf{W}_0 \equiv E\left[\mathbf{z}_i(y_i - \mathbf{a}_0'\mathbf{x}_i)^2\mathbf{z}_i'\right] = E\left[u_i^2\mathbf{z}_i\mathbf{z}_i'\right].$$

Bien entendu, on n'a pas de mesure des u_i, mais on pourra en construire un estimateur.

- On définit ici $\tilde{\mathbf{M}}_N$ par:

$$\boxed{\tilde{\mathbf{M}}_N \equiv \left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1} \xrightarrow[N\to+\infty]{p} \mathbf{M}_0,}$$

$\mathbf{M}_0$ est la matrice (définie positive) la plus proche de $\mathbf{W}_0^{-1} \equiv E\left[u_i^2\mathbf{z}_i\mathbf{z}_i'\right]^{-1}$ dont on sache calculer un estimateur convergent, $\tilde{\mathbf{M}}_N$.

- On a alors :

$$\boxed{\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N) \equiv \arg\min_{\mathbf{a}} N^{-1}\sum_{i=1}^{N}(y_i - \mathbf{x}_i'\mathbf{a})\mathbf{z}_i'\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1} N^{-1}\sum_{i=1}^{N}\mathbf{z}_i(y_i - \mathbf{x}_i'\mathbf{a}).}$$

et donc :

$$\boxed{\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N) = \hat{\mathbf{a}}_N^{2MC}}$$

Seconde chose à faire : définir un estimateur convergent de
$$\mathbf{W}_0 \equiv E\left[\mathbf{g}_i(\mathbf{a}_0)\mathbf{g}_i(\mathbf{a}_0)'\right]$$

- Si $\tilde{\mathbf{a}}_N \xrightarrow[N\to+\infty]{p} \mathbf{a}_0$ alors :

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

- Ici on a $\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N) = \hat{\mathbf{a}}_N^{2MC} \xrightarrow[N\to+\infty]{p} \mathbf{a}_0$ et donc :

$$\boxed{\tilde{\mathbf{W}}_N \equiv N^{-1}\sum_{i=1}^{N}(y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{2MC})^2\mathbf{z}_i\mathbf{z}_i' \xrightarrow[N\to+\infty]{p} \mathbf{W}_0 = E\left[\left(y_i - \mathbf{x}_i'\mathbf{a}_0\right)^2\mathbf{z}_i\mathbf{z}_i'\right].}$$

Dernière chose à faire : calculer $\hat{\mathbf{a}}_N^{MMG}$ en utilisant $\tilde{\mathbf{W}}_N$

- On a ici :

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

et donc :

$$\hat{\mathbf{a}}_N^{MMG} = \left[\left(N^{-1} \sum_{i=1}^{N} \mathbf{x}_i \mathbf{z}_i' \right) \tilde{\mathbf{W}}_N^{-1} \left(N^{-1} \sum_{i=1}^{N} \mathbf{z}_i \mathbf{x}_i' \right) \right]^{-1} \times \left(N^{-1} \sum_{i=1}^{N} \mathbf{x}_i \mathbf{z}_i' \right) \tilde{\mathbf{W}}_N^{-1} N^{-1} \sum_{i=1}^{N} \mathbf{z}_i y_i \equiv \hat{\mathbf{a}}_N^{2MCH}$$

- $\hat{\mathbf{a}}_N^{2MCH}$ est l'***estimateur des 2MC robuste à l'hétéroscédasticité***

- La MMG donne :

$$\sqrt{N}(\hat{\mathbf{a}}_N^{MMG} - \mathbf{a}_0) \xrightarrow[N\to+\infty]{L} \mathcal{N}(\mathbf{0};\boldsymbol{\Sigma}_0) \quad \text{avec } \boldsymbol{\Sigma}_0 \equiv (\mathbf{G}_0'\mathbf{W}_0^{-1}\mathbf{G}_0)^{-1}$$

et ici on a :

$$\boxed{\sqrt{N}(\hat{\mathbf{a}}_N^{2MC} - \mathbf{a}_0) \xrightarrow[N\to+\infty]{L} \mathcal{N}\left(\mathbf{0};\boldsymbol{\Sigma}_0 \equiv \left(E\left[\mathbf{x}_i\mathbf{z}_i'\right]E\left[u_i^2\mathbf{z}_i\mathbf{z}_i'\right]^{-1}E\left[\mathbf{z}_i\mathbf{x}_i'\right]\right)^{-1}\right)}$$

$$\hat{\mathbf{W}}_N \equiv N^{-1}\sum_{i=1}^{N}(y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{MMG})^2\mathbf{z}_i\mathbf{z}_i' \xrightarrow[N\to+\infty]{p} \mathbf{W}_0$$

$$\hat{\mathbf{G}}_N \equiv -N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i' \xrightarrow[N\to+\infty]{p} \mathbf{G}_0$$

$$\hat{\boldsymbol{\Sigma}}_N \equiv (\hat{\mathbf{G}}_N'\hat{\mathbf{W}}_N^{-1}\hat{\mathbf{G}}_N)^{-1} \xrightarrow[N\to+\infty]{p} \boldsymbol{\Sigma}_0 \equiv (\mathbf{G}_0'\mathbf{W}_0^{-1}\mathbf{G}_0)^{-1}$$

- L’expérience accumulée tend à montrer que $\hat{\mathbf{a}}_N^{2MCH}$ est « instable » si N n’est pas assez grand, … moments d’ordre 4 dans $\hat{\mathbf{W}}_N$.

5. Méthode des Moment Généralisée et Méthode des Moments

- La MMG peut être utilisée dans le cas juste-identifié, *i.e.* avec $G = K$, mais les matrices de pondération $\tilde{\mathbf{M}}_N$ n'ont alors aucun intérêt.
  - Si on a autant de conditions que de paramètres à estimer, aucune « pondération n'est possible »
  - On doit utiliser « à plein » les K conditions de moments pour calculer les K éléments de $\hat{\mathbf{a}}_N^{MM}$.
- ***Conclusion*** : Si on a :

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

alors on utilise la MM et donc :

$$\hat{\mathbf{a}}_N^{MM} \text{ solution en } \mathbf{a} \text{ de } N^{-1}\sum_{i=1}^{N}\mathbf{g}_i(\mathbf{a}) = \mathbf{0}_{K\times 1}$$

- La MMG peut être utilisée dans le cas juste-identifié, *i.e.* avec $G = K$, mais dans ce cas les matrices de pondération $\tilde{\mathbf{M}}_N$ n'ont alors aucun intérêt.

- On peut aisément voir cela avec la CO1 du programme de minimisation :

$$\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\bar{\mathbf{g}}_N(\tilde{\mathbf{a}}_N) = \mathbf{0}_{K\times 1}$$

  qui définit un système de K équations à K inconnues lorsque $G = K$.

- Dès lors que $\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)$ et $\tilde{\mathbf{M}}_N$ sont de rang K, elles sont inversibles puisque carrées et on a :

$$\boxed{\bar{\mathbf{G}}_N(\tilde{\mathbf{a}}_N)'\tilde{\mathbf{M}}_N\bar{\mathbf{g}}_N(\tilde{\mathbf{a}}_N) = \mathbf{0}_{K\times 1} \quad\Leftrightarrow\quad \bar{\mathbf{g}}_N(\tilde{\mathbf{a}}_N) = \mathbf{0}_{K\times 1},}$$

  ce qui indique que $\tilde{\mathbf{a}}_N$ ne dépend pas de $\tilde{\mathbf{M}}_N$ si $G = K$.

- Cette remarque est importante car le cas juste-identifié est fréquent.

- De même, dans le cas juste-identifié on a :

$$\boxed{\sqrt{N}\,(\hat{\mathbf{a}}_N^{MMG} - \mathbf{a}_0) \xrightarrow[N\to+\infty]{L} \mathcal{N}(\mathbf{0};\boldsymbol{\Sigma}_0) \quad \text{avec } \boldsymbol{\Sigma}_0 = (\mathbf{G}_0)^{-1}\mathbf{W}_0(\mathbf{G}_0')^{-1}}$$

On peut retrouver ça aisément à partir du résultat de la MMG :

$$\boldsymbol{\Sigma}_0 = \left(\mathbf{G}_0'\mathbf{W}_0^{-1}\mathbf{G}_0\right)^{-1}$$

en remarquant que les matrices $\mathbf{G}_0$ et $\mathbf{W}_0$ sont carrées $(K \times K)$ et inversibles. On obtient alors :

$$\boldsymbol{\Sigma}_0 = \left(\mathbf{G}_0'\mathbf{W}_0^{-1}\mathbf{G}_0\right)^{-1} = (\mathbf{G}_0)^{-1}\mathbf{W}_0(\mathbf{G}_0')^{-1}$$

en utilisant les propriétés d'inversion des produits et des transposées de matrices :

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \quad \text{et} \quad (\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'.$$

6. Le test de sur-identification

Bien entendu, il est utile de pouvoir juger la pertinence du modèle utilisé en termes statistiques, *i.e.* à partir de tests statistiques.

Le test de la régression augmentée, et plus généralement les tests dits d'Hausman, permet de tester la pertinence d'hypothèses d'exogénéité de variables explicatives moyennant la disponibilité de VI.

La MMG fournit un autre test permet de juger, au moins en partie, la ***spécification du modèle*** proposé aux données.

Il s'agit du ***test de sur-identification***, dit encore ***test de Hansen*** ou ***test de Sargan*** voire ***test du Chi-2***.

Ce test est maintenant utilisé ***en routine*** par les économètres appliqués (même s'il n'est pas toujours bien interprété).

Présentation en trois temps :

- Rappels sur les tests asymptotiques
- Présentation du test de sur-identification
- Interprétation du test et remarques conclusives

5.1. Rappels sur les tests asymptotiques

Un test asymptotique pour l'hypothèse nulle *Hyp*0 ***contre l'hypothèse alternative*** *Hyp*A (pas nécessairement le complément de *Hyp*0) est composé de :

- Une ***statistique de test***, $\hat{T}_N$ ***dont on connaît la loi limite sous*** *Hyp*0 :

  e.g. une loi du Chi-2 à *P* degrés de liberté $\chi^2(P)$.

- Cette loi limite permet de définir une ***région critique as.***, *w* :

  $$e.g.\ w \equiv \left[\chi^2_{1-\alpha}(P);+\infty\right[$$

  où est le quantile d'ordre $1-\alpha$ du Chi-2 à *P* degrés de liberté $\chi^2(P)$, *i.e.* si $x \sim \chi^2(P)$ alors $P\left[x \leq \chi^2_{1-\alpha}(P)\right] = 1-\alpha$ sous *Hyp*0.

- Une ***règle de décision*** : *Hyp*0 n'est pas rejetée si la valeur calculée de $\hat{T}_N$ n'appartient pas à w, *Hyp*0 est rejetée sinon.

Définition. ***Niveau asymptotique d'un test***.
Un test as. défini par la statistique de test $\hat{T}_N$ et la région critique w est de ***niveau as.*** α si :

$$\text{Sous } Hyp0,\ \lim_{N\to+\infty} P\left[\hat{T}_N \in w\right] = \alpha.$$

La probabilité de rejeter à tort *Hyp*0 (***risque de première espèce***) tend vers α lorsque N tend vers l'infini.

Définition. ***Convergence d'un test asymptotique.***
Un test as. défini par la statistique de test $\hat{T}_N$ et la région critique w est ***convergent contre*** *Hyp*A si :

$$\text{Sous } HypA,\ P\left[\hat{T}_N \in w\right] \xrightarrow[N\to+\infty]{} 1.$$

La probabilité de détecter *Hyp*A lorsque *Hyp*A est vraie (accepter à tort *Hyp*0, ***risque de seconde espèce***) approche 1 lorsque N tend vers l'infini.

Propriété. ***Convergence en loi vers un*** Chi-2

Soient $\mathbf{m}_N$ une suite de variables aléatoires dans $\mathbb{R}^G$ et $\mathbf{\Psi}$ une matrice réelle semi-définie positive de dimension G et de rang $S \leq G$ telles que :

$$\sqrt{N}\mathbf{m}_N \xrightarrow[N\to+\infty]{L} \mathcal{N}(\mathbf{m};\mathbf{\Psi}).$$

Soient $\mathbf{\Psi}^-$ une matrice inverse généralisée de Moore-Penrose de $\mathbf{\Psi}$ et $\mathbf{\Psi}_N^-$ une suite de matrices aléatoires telles que :

$$\mathbf{\Psi}_N^- \xrightarrow[N\to+\infty]{p} \mathbf{\Psi}^-$$

Alors on a :

$$N\times\mathbf{m}'_N\mathbf{\Psi}_N^-\mathbf{m}_N \xrightarrow[N\to+\infty]{L} \chi^2(S;\mathbf{m}'\mathbf{\Psi}^-\mathbf{m}),$$

où $\mathbf{m}'\mathbf{\Psi}^-\mathbf{m}$ est le paramètre de décentrage de la loi limite de $N\times\mathbf{m}'_N\mathbf{\Psi}_N^-\mathbf{m}_N$.

- Si $\mathbf{m}=\mathbf{0}_{G\times1}$ alors la loi limite de $N\times\mathbf{m}'_N\mathbf{\Psi}_N^-\mathbf{m}_N$ est un $\chi^2(G)$ « standard ».
- Si $rang\mathbf{\Psi}=G$ alors $\mathbf{\Psi}^-=\mathbf{\Psi}^{-1}$.

Intuition

Un $\chi^2(G)$ est une somme de G carrés de lois normales centrées réduites indépendantes.

Si $\mathbf{x} \sim \mathcal{N}(\mathbf{0};\mathbf{\Psi})$ et $\mathbf{\Psi}$ est inversible ($rang\mathbf{\Psi} = \dim \mathbf{x}$) alors $\mathbf{x}'\mathbf{x}$ est une ***combinaison linéaire de*** $rang\mathbf{\Psi}$ carrés de variables $\mathcal{N}(0;1)$ indépendantes.

De plus, $\mathbf{\Psi}$ étant inversible, les matrices $\mathbf{\Psi}^{-1}$ et $\mathbf{\Psi}^{-1/2}$ telle que $\mathbf{\Psi}^{-1/2}\mathbf{\Psi}^{-1/2} = \mathbf{\Psi}^{-1}$existent. On a alors :

$$\mathbf{\Psi}^{-1/2}\mathbf{x} \sim \mathcal{N}(\mathbf{0};\mathbf{I})$$

et :

$$\mathbf{x}'\mathbf{\Psi}^{-1}\mathbf{x} = \mathbf{x}'\mathbf{\Psi}^{-1/2}\mathbf{\Psi}^{-1/2}\mathbf{x} \sim \chi^2(\dim \mathbf{x}) \quad "= \mathcal{N}(\mathbf{0};\mathbf{I})'\mathcal{N}(\mathbf{0};\mathbf{I})".$$

Si $\mathbf{x} \sim \mathcal{N}(\mathbf{0};\mathbf{\Psi})$ et $\mathbf{\Psi}$ n'est pas inversible, *i.e.* $rang\mathbf{\Psi} < \dim \mathbf{x}$. Dans ce cas $\dim \mathbf{x} - rang\mathbf{\Psi}$ éléments de **x** sont des combinaisons linéaires des autres éléments de **x**.

Il y a $rang\mathbf{\Psi}$ éléments indépendants dans **x** et $\mathbf{x}'\mathbf{x}$ est une ***combinaison linéaire de*** $rang\mathbf{\Psi}$ carrés de variables $\mathcal{N}(0;1)$ indépendantes.

Si $\mathbf{\Psi}^-$ est une inverse de Moore-Penrose de $\mathbf{\Psi}$, le terme $\mathbf{x}'\mathbf{\Psi}^-\mathbf{x}$ est ***une somme de*** $rang\mathbf{\Psi}$ carrés de variables $\mathcal{N}(0;1)$ indépendantes, *i.e.* suit une loi du $\chi^2(rang\mathbf{\Psi})$. (On a $\mathbf{\Psi}^- = \mathbf{\Psi}^{-1}$ si $\mathbf{\Psi}$ est inversible).

Si $\mathbf{x} \sim \mathcal{N}(\mathbf{m};\mathbf{\Psi})$ alors $\mathbf{x}'\mathbf{\Psi}^-\mathbf{x}$ suit une loi $\chi^2(rang\mathbf{\Psi},\mathbf{m}'\mathbf{\Psi}^-\mathbf{m})$, *i.e.* une loi du χ^2 décentrée, le paramètre de décentrage étant $\mathbf{m}'\mathbf{\Psi}^-\mathbf{m}$.

5.2. Le test de sur-identification

L'estimateur MMG optimal de $\mathbf{a}_0$ fondé sur la condition d'identification :

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

est défini par :

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

où :

$$\tilde{\mathbf{W}}_N \xrightarrow[N\to+\infty]{p} \mathbf{W}_0 .$$

Comme précédemment, on notera :

$$Q_N(\mathbf{a};\tilde{\mathbf{M}}_N)\equiv\left[N^{-1}\sum_{i=1}^{N}\mathbf{g}(\mathbf{w}_i;\mathbf{a})'\right]\tilde{\mathbf{M}}_N\left[N^{-1}\sum_{i=1}^{N}\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right]=\overline{\mathbf{g}}_N(\mathbf{a})'\tilde{\mathbf{M}}_N\overline{\mathbf{g}}_N(\mathbf{a})$$

Lorsque $\mathbf{a}_0$ est sur-identifié par les conditions de moment, *i.e.* lorsque $G > K$, on a :

$$N \times Q_N(\mathbf{a}; \tilde{\mathbf{W}}_N^{-1}) = N \times \overline{\mathbf{g}}_N(\mathbf{a})' \tilde{\mathbf{W}}_N^{-1} \overline{\mathbf{g}}_N(\mathbf{a}) \geq 0 \text{ pour tout } \mathbf{a}$$

et, puisque $\tilde{\mathbf{W}}_N^{-1}$ est définie positive, on a :

$$N \times Q(\hat{\mathbf{a}}_N^{MMG}; \tilde{\mathbf{W}}_N^{-1}) = 0 \quad \Leftrightarrow \quad \overline{\mathbf{g}}_N(\hat{\mathbf{a}}_N^{MMG}) = \mathbf{0}.$$

L'idée de Hansen est de ***juger la valeur de*** $N \times Q(\hat{\mathbf{a}}_N^{MMG}; \tilde{\mathbf{W}}_N^{-1})$ ***obtenue par rapport à*** 0, sachant que $N \times Q_N(\hat{\mathbf{a}}_N^{MMG}; \tilde{\mathbf{W}}_N^{-1})$ est positif par construction.

On considère ici les hypothèses nulle et alternative suivantes :

*Hyp*0 : On a $E\left[\mathbf{g}(\mathbf{w}_i; \mathbf{a}_0)\right] = \mathbf{0}_{G\times 1}$

*Hyp*A : On a $E\left[\mathbf{g}(\mathbf{w}_i; \mathbf{a})\right] \neq \mathbf{0}_{G\times 1}$ pour toute valeur de **a**.

Nous reviendrons plus tard sur l'interprétation de ces hypothèses.

Il nous faut déterminer la loi limite de $N \times Q_N(\hat{\mathbf{a}}_N^{MMG}; \tilde{\mathbf{W}}_N^{-1})$**.** Notons tout d'abord que, si *Hyp*0 est vraie alors :

$$\boxed{N \times Q_N(\mathbf{a}_0; \tilde{\mathbf{W}}_N^{-1}) \xrightarrow[N \to +\infty]{L} \chi^2(G)}$$

car :

$$N \times Q_N(\mathbf{a}_0; \tilde{\mathbf{W}}_N^{-1}) \equiv \sqrt{N}\overline{\mathbf{g}}_N(\mathbf{a}_0)' \tilde{\mathbf{W}}_N^{-1} \sqrt{N}\overline{\mathbf{g}}_N(\mathbf{a}_0)$$

et :

$$\sqrt{N}\overline{\mathbf{g}}_N(\mathbf{a}_0) \xrightarrow[N \to +\infty]{L} \mathcal{N}(\mathbf{0}; \mathbf{W}_0) \text{ et } \tilde{\mathbf{W}}_N^{-1} \xrightarrow[N \to +\infty]{p} \mathbf{W}_0^{-1}.$$

On montre cependant que :

$$\boxed{N \times Q_N(\hat{\mathbf{a}}_N^{MMG}; \tilde{\mathbf{W}}_N^{-1}) \xrightarrow[N \to +\infty]{L} \chi^2(G-K)}$$

car :

$$\sqrt{N}\overline{\mathbf{g}}_N(\hat{\mathbf{a}}_N^{MMG}) \xrightarrow[N \to +\infty]{L} \mathcal{N}\left(\mathbf{0}; \mathbf{I}_G - \mathbf{B}_0\right) \text{ et } rang(\mathbf{I}_G - \mathbf{B}_0) = G - K.$$

Ce résultat provient de ce que :

$$\begin{aligned}\sqrt{N}\overline{\mathbf{g}}_N(\hat{\mathbf{a}}_N^{MMG}) &= \sqrt{N}\overline{\mathbf{g}}_N(\mathbf{a}_0) + \overline{\mathbf{G}}_N(\mathbf{a}_0)\sqrt{N}(\hat{\mathbf{a}}_N^{MMG} - \mathbf{a}_0) + \mathbf{o}_p(1) \\ &= \sqrt{N}\overline{\mathbf{g}}_N(\mathbf{a}_0) + \mathbf{G}_0'\sqrt{N}(\hat{\mathbf{a}}_N^{MMG} - \mathbf{a}_0) + \mathbf{o}_p(1)\end{aligned}$$

On sait par ailleurs :

$$\sqrt{N}\left(\hat{\mathbf{a}}_N^{MMG} - \mathbf{a}_0\right) = -(\mathbf{G}_0'\mathbf{W}_0^{-1}\mathbf{G}_0)^{-1}\mathbf{G}_0'\mathbf{W}_0^{-1}\sqrt{N}\overline{\mathbf{g}}_N(\mathbf{a}_0) + \mathbf{o}_p(1)$$

ou :

$$\boxed{\sqrt{N}\overline{\mathbf{g}}_N(\hat{\mathbf{a}}_N^{MMG}) = (\mathbf{I}_G - \mathbf{B}_0)\sqrt{N}\overline{\mathbf{g}}_N(\mathbf{a}_0) + \mathbf{o}_p(1)}$$

avec :

$$\mathbf{B}_0 \equiv \mathbf{G}_0'(\mathbf{G}_0'\mathbf{W}_0^{-1}\mathbf{G}_0)^{-1}\mathbf{G}_0'\mathbf{W}_0^{-1}.$$

Avec :

$$\boxed{\sqrt{N}\,\overline{\mathbf{g}}_N(\hat{\mathbf{a}}_N^{MMG}) = (\mathbf{I}_G - \mathbf{B}_0)\sqrt{N}\,\overline{\mathbf{g}}_N(\mathbf{a}_0) + \mathbf{o}_p(1),}$$

on a :

$$\sqrt{N}\,\overline{\mathbf{g}}_N(\hat{\mathbf{a}}_N^{MMG}) \xrightarrow[N\to+\infty, Hyp0]{L} \mathcal{N}\left(\mathbf{0}_{G\times 1}; (\mathbf{I}_G - \mathbf{B}_0)\mathbf{W}_0(\mathbf{I}_G - \mathbf{B}_0)'\right)$$

Les matrices $\mathbf{B}_0$ et $\mathbf{I}_G - \mathbf{B}_0$ sont des projecteurs. $\mathbf{B}_0$ est de rang K et $\mathbf{I}_G - \mathbf{B}_0$ qui est de rang $G-K$. On montre également que $(\mathbf{I}_G - \mathbf{B}_0)\mathbf{W}_0(\mathbf{I}_G - \mathbf{B}_0)'$ est de rang $G-K$ et admet $\mathbf{W}_0^{-1}$ comme une inverse de Moore-Penrose.

Ceci qui permet de montrer que :

$$\boxed{N \times Q_N(\hat{\mathbf{a}}_N^{MMG}; \tilde{\mathbf{W}}_N^{-1}) \xrightarrow[N\to+\infty, Hyp0]{L} \chi^2(G-K).}$$

La perte des K degrés de liberté est dû au calcul de $\hat{\mathbf{a}}_N^{MMG}$ qui « annule » automatiquement l'équivalent de K conditions de moments dans $\overline{\mathbf{g}}_N(\hat{\mathbf{a}}_N^{MMG})$.

Interprétation

(i) L'estimateur $\hat{\mathbf{a}}_N^{MMG}$ est construit de sorte à ce que $\overline{\mathbf{g}}_N(\hat{\mathbf{a}}_N^{MMG})$ soit le plus proche possible de $\mathbf{0}$.

(ii) Le paramètre $\mathbf{a}$ étant de dimension K, on dispose de K degrés de liberté pour ce faire.
Aussi l' « ***équivalent*** » ***de*** K ***éléments de*** $\overline{\mathbf{g}}_N(\hat{\mathbf{a}}_N^{MMG})$ ***est garanti d'être effectivement nul***.

(iii) Ces K ***degrés de liberté sont perdus*** pour le test de $Hyp0$, car « ***consommés*** » ***par le calcul de*** $\hat{\mathbf{a}}_N^{MMG}$.

Il ne reste à tester empiriquement que les $G-K$ ***conditions sur-identifiantes***,
i.e. celles qui n'ont pas été « consommées » par le calcul de $\hat{\mathbf{a}}_N^{MMG}$

Exemple 1. ***Condition d'exogénéité de $\mathbf{x}_i$ et MCO***

Afin d'illustrer le point précédent, on peut prendre l'exemple d'un modèle linéaire :

$$y_i = \mathbf{a}_0'\mathbf{x}_i + u_i \quad \text{avec} \quad E\left[u_i\right] \equiv 0 \quad \text{et} \quad rangE\left[\mathbf{x}_i\mathbf{x}_i'\right] = K\,.$$

qu'on suppose être un modèle de régression. La construction de l'estimateur des MCO de $\mathbf{a}_0$ est fondé sur la condition de moment caractérisant l'exogénéité de $\mathbf{x}_i$:$E\left[\mathbf{x}_i(y_i - \mathbf{a}'\mathbf{x}_i)\right] = \mathbf{0}_{K\times 1} \quad \Leftrightarrow \quad \mathbf{a} = \mathbf{a}_0$.

De par la construction de $\hat{\mathbf{a}}_N^{MCO}$ on a :

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

Il n'y a donc rien à tester à partir de $\hat{\mathbf{a}}_N^{MCO}$ pour ce qui concerne la validité de $E\left[\mathbf{x}_i(y_i - \mathbf{a}_0'\mathbf{x}_i)\right] = \mathbf{0}_{K\times 1}$ puisqu'on a « automatiquement » :

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

Exemple 2. ***Condition d'orthogonalité avec les VI et 2MC***

Afin d'illustrer le point précédent, on peut prendre l'exemple d'un modèle linéaire à VI :

$$y_i = \mathbf{a}_0'\mathbf{x}_i + u_i \quad \text{avec} \quad E\left[u_i / \mathbf{z}_i\right] \equiv 0 \quad \text{et} \quad rangE\left[\mathbf{z}_i \mathbf{x}_i'\right] = K .$$

La construction de l'estimateur des 2MC de $\mathbf{a}_0$ est fondé sur la condition de moment caractérisant l'exogénéité de $\mathbf{z}_i$:

$$E\left[\mathbf{z}_i (y_i - \mathbf{a}'\mathbf{x}_i)\right] = \mathbf{0}_{L\times 1} \quad \Leftrightarrow \quad \mathbf{a} = \mathbf{a}_0 .$$

De par la construction de $\hat{\mathbf{a}}_N^{2MC}$ on a :

$$N^{-1}\sum_{i=1}^{N} \mathbf{z}_i (y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{2MC}) = \mathbf{0}_{L\times 1} \ \text{ si on a} L = K$$

Si $L > K$ alors, alors en général on a :

$$N^{-1}\sum_{i=1}^{N}\mathbf{z}_i(y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{2MC}) \neq \mathbf{0}_{L\times 1}.$$

Si $L > K$ et le modèle est correctement spécifié alors :

$$N^{-1}\sum_{i=1}^{N}\mathbf{z}_i(y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{2MC}) \simeq \mathbf{0}_{L\times 1}$$

dans la métrique définie par $\hat{\mathbf{W}}_N^{-1} = \left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1}$.

C'est ce qu'on cherche à tester ici :

- la proximité de $N^{-1}\sum_{i=1}^{N}\mathbf{z}_i(y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{2MC})$ par rapport à $\mathbf{0}_{L\times 1}$ dans la métrique définie par $\hat{\mathbf{W}}_N^{-1}$.
- sachant que $\hat{\mathbf{a}}_N^{2MC}$ a été calculé de sorte à minimiser la distance entre $\mathbf{0}_{L\times 1}$ et $N^{-1}\sum_{i=1}^{N}\mathbf{z}_i(y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{2MC})$ dans la métrique définie par $\hat{\mathbf{W}}_N^{-1}$.

Propriété. ***Test de sur-identification*.**

On considère ici la condition estimante de $\mathbf{a}_0$:

$$E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right]=\mathbf{0}_{G\times 1} \quad \Leftrightarrow \quad \mathbf{a}=\mathbf{a}_0 \text{ avec } G>\dim\mathbf{a}_0=K$$

et l'estimateur MMG optimal de $\mathbf{a}_0$ défini par :

$$\hat{\mathbf{a}}_N^{MMG}\equiv\arg\min_{\mathbf{a}} Q(\mathbf{a};\tilde{\mathbf{W}}_N^{-1}) \text{ avec } \tilde{\mathbf{W}}_N \xrightarrow[N\to+\infty]{p} E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a}_0)\mathbf{g}(\mathbf{w}_i;\mathbf{a}_0)'\right]$$

où :

$$Q_N(\mathbf{a};\tilde{\mathbf{W}}_N^{-1})\equiv\left[N^{-1}\sum_{i=1}^N\mathbf{g}(\mathbf{w}_i;\mathbf{a})'\right]\tilde{\mathbf{W}}_N^{-1}\left[N^{-1}\sum_{i=1}^N\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right].$$

On a :

(i) $\hat{T}_N\equiv NQ_N(\hat{\mathbf{a}}_N^{MMG};\tilde{\mathbf{W}}_N^{-1})\xrightarrow[N\to+\infty]{L}\chi^2(G-K)$ sous $Hyp0$: $E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a}_0)\right]=\mathbf{0}$.

(ii) Le test défini par la statistique $\hat{T}_N$ et la région critique $\left[\chi^2_{1-\alpha}(G-K);+\infty\right[$ est de niveau as. α.

(iii) Ce test est convergent contre HypA : $E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right]\neq\mathbf{0}_{G\times 1}$, $\forall\mathbf{a}\in\mathbb{R}^K$.

Le ***test de sur-identification*** est plus un ***test de compatibilité*** qu'un véritable test de spécification.

- Si ce test ***conduit à rejeter*** *Hyp*0 alors il est certain que le ***modèle*** défini pour l'estimation de $\mathbf{a}_0$ est ***incorrect***.
    - Le problème peut provenir du choix des VI ou de la forme du modèle choisi pour les variables à expliquer en tant que fonction des explicatives.
- Si ce test ***ne conduit pas à rejeter*** *Hyp*0, alors on ne doit ***pas nécessairement conclure que le modèle*** est ***correct*** :
    - Le test est convergent contre *Hyp*A : $E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right] \neq \mathbf{0}_{G\times 1},\ \forall \mathbf{a} \in \mathbb{R}^K$. Le non-rejet de *Hyp*0 indique donc simplement qu'il existe $\mathbf{a}$ tel que $E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right] = \mathbf{0}_{G\times 1}$. La solution en $\mathbf{a}$ de ce système étant ***supposée*** unique, $\mathbf{a}_0$, on en déduit alors que $E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a}_0)\right] = \mathbf{0}_{G\times 1}$.

La véritable hypothèse nulle est « il existe $\mathbf{a}$ tel que $E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right] = \mathbf{0}_{G\times 1}$ ».

- Ce n'est que ***notre interprétation de la condition de moment*** et les ***hypothèses relatives à son identification*** :

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

qui nous assurent que $\mathbf{a}_0$ ***est bien ce qu'on cherche à estimer***. Il y a ici deux points :

(*i*) L'équation $E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right]=\mathbf{0}_{G\times 1}$ peut avoir plusieurs solutions pour des modèles non-linéaires.
Dans ce cas il faut définir un ensemble de définition pour $\mathbf{a}_0$, $\mathcal{A}$, tel que :

$$\mathbf{a}\in \mathcal{A} \text{ et } E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right]=\mathbf{0}_{G\times 1} \quad \Leftrightarrow \quad \mathbf{a}=\mathbf{a}_0$$

(*ii*) *Hyp*0 peut ne pas être rejetée parce qu'il existe une solution de $E\left[\mathbf{g}(\mathbf{w}_i;\mathbf{a})\right]=\mathbf{0}_{G\times 1}$ bien que cette solution ne soit pas $\mathbf{a}_0$.
C'est un problème de spécification non détecté par le test de Hansen.

Exemple de problème de spécification non détecté

- On veut estimer $\mathbf{a}_0$ dans le modèle :

$$y_i = \mathbf{a}_0'\mathbf{x}_i + u_i \quad \text{avec} \quad E[u_i] \equiv 0 \ \text{où} \ \mathbf{x}_i \equiv (1, x_i^e, x_i^x) + u_i,$$

- On suppose que x_i^x est exogène par rapport à u_i et on instrumente x_i^e avec $\mathbf{z}_i^e$, *i.e.* on utilise le vecteur de VI $\mathbf{z}_i \equiv (1, \mathbf{z}_i^e, x_i^x)$ avec $rangE[\mathbf{z}_i\mathbf{x}_i'] = 3$. On emploie l'estimateur des 2MC de $\mathbf{a}_0$.

- Si l'analyse du PGD est erronée dans le sens où x_i^x est en réalité endogène. *E.g.*, en supposant (pour simplifier les calculs) que :

$$EL[u_i \mid \mathbf{z}_i] = \beta_0 x_i^x \ \text{avec} \ \beta_0 \neq 0,$$

on montre aisément que :

$$p\lim_{N \to +\infty} \hat{\mathbf{a}}_N^{2MC} = \mathbf{a}_0 + \mathbf{b}_0 \neq \mathbf{a}_0 \quad \text{avec} \ \mathbf{b}_0 \equiv (0, 0, \beta_0) \neq \mathbf{0}.$$

- Le problème est ici qu'on pense utiliser la condition de moment :

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

alors qu'on utilise la condition :

$$E\left[\mathbf{z}_i(y_i - (\mathbf{a}_0 + \mathbf{b}_0)'\mathbf{x}_i)\right] = E\left[\mathbf{z}_i e_i\right] = \mathbf{0} \text{ où } e_i \equiv u_i - \beta_0 x_i^x$$

suivante pour le calcul de $\hat{\mathbf{a}}_N^{2MC}$.

- Dans ce cas, le test de Hansen amènera vraisemblablement à accepter *Hyp*0. Ce test validera bien l'hypothèse :

$$Hyp0 : \text{« Il existe } \boldsymbol{\alpha}_0 \text{ telle que } E\left[\mathbf{z}_i(y_i - \boldsymbol{\alpha}'\mathbf{x}_i)\right] = \mathbf{0} \text{ ».}$$

Le problème est ici que :

$$\boldsymbol{\alpha}_0 = \mathbf{a}_0 + \mathbf{b}_0 \neq \mathbf{a}_0.$$

- De manière générale, tout test est construit à partir d’estimateurs.

    - ***Tout test repose donc sur un ensemble d’hypothèses dont certaines ne pourront pas être testées***.

    - On retrouve, une nouvelle fois ici, l’idée selon laquelle ***la phase de spécification d’un modèle***, et donc l’***analyse du PG des données utilisées***, est une phase au moins aussi importante que celle qui consiste à choisir et à utiliser des outils d’inférence statistique appropriés car …

    - … « on ne peut pas tout tester » d’un point de vue statistique.