

**ÉCONOMÉTRIE  
(UGA S2)  
CHAPITRE 6:  
MOMENTS CONDITIONNELS, CHOIX DES MOMENTS,  
INSTRUMENTS EFFICACES**

Michal W. Urdanivia\*

\* Université de Grenoble Alpes, Faculté d'Économie, GAEL,  
e-mail: [michal.wong-urdanivia@univ-grenoble-alpes.fr](mailto:michal.wong-urdanivia@univ-grenoble-alpes.fr)

15 mars 2022

# Contenu

1. Motivation
2. La propriété des instruments estimés
3. Instruments efficaces, cas scalaire
4. Instruments efficaces, cas multivarié
5. Instruments efficaces et MMG

## 1. Motivation

# 1. Motivation

- La *Méthode des Moments Généralisée* (MMG) a été conçue par Hansen (1982) pour *exploiter toute condition de moment identifiant* un vecteur de paramètres  $\mathbf{a}_0$ , e.g. :

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

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

- à partir échantillon de variables aléatoires iid  $\{\mathbf{w}_i; i = 1, 2, \dots, N\}$ ,
  - que cette condition de moment *juste- ou sur-identifie*  $\mathbf{a}_0$ .
- **Problème** : beaucoup de modèles économétriques donnent des conditions de moments conditionnels de la forme :

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

- Dans la suite on notera :

$$u_i(\mathbf{a}) \equiv u(y_i, \mathbf{x}_i; \mathbf{a}) \quad \text{et} \quad \mathbf{u}_i(\mathbf{a}) \equiv \mathbf{u}(\mathbf{y}_i, \mathbf{x}_i; \mathbf{a})$$

(cas scalaire)                      (cas multivarié)

pour alléger les notations, comme on utilisait  $\mathbf{g}_i(\mathbf{a}) \equiv \mathbf{g}(\mathbf{w}_i; \mathbf{a})$ .

- On prend ici les notations  $u_i(\mathbf{a})$  et  $\mathbf{u}_i(\mathbf{a}) \equiv \mathbf{u}(\mathbf{y}_i, \mathbf{x}_i; \mathbf{a})$  car, dans la plupart des cas rencontrés en pratique, ces fonctions définissent des résidus :

$$y_i = f(\mathbf{x}_i; \mathbf{a}_0) + u_i \quad \Rightarrow \quad u_i(\mathbf{a}) \equiv y_i - f(\mathbf{x}_i; \mathbf{a}) \quad \text{et} \quad u_i(\mathbf{a}_0) = u_i$$

(Cas d'*une équation*)

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i; \mathbf{a}_0) + \mathbf{u}_i \quad \Rightarrow \quad \mathbf{u}_i(\mathbf{a}) \equiv \mathbf{y}_i - \mathbf{f}(\mathbf{x}_i; \mathbf{a}) \quad \text{et} \quad \mathbf{u}_i(\mathbf{a}_0) = \mathbf{u}_i$$

(Cas d'un *système d'équations*)

## Rappels. *Résidus et termes d'erreur*

- Si on considère un modèle dont la forme fonctionnelle est définie par l'équation :

$$y_i = f(\mathbf{x}_i; \mathbf{a}_0) + u_i$$

alors :

$u_i(\mathbf{a}) \equiv y_i - f(\mathbf{x}_i; \mathbf{a})$  est le ***résidu*** de l'équation donnant  $y_i$  ***calculé en a***,  
*i.e.* une construction mathématique dérivée du modèle de  $y_i$ .

$u_i = u_i(\mathbf{a}_0)$  est le ***terme d'erreur*** du modèle, *i.e.* une variable aléatoire  
qui fait partie du modèle.

- On a l'***égalité*** :

$$u_i = u_i(\mathbf{a}_0)$$

par le modèle de  $y_i$  qui donne  $u_i = y_i - f(\mathbf{x}_i; \mathbf{a}_0)$  et par la définition de la  
fonction  $u_i(.)$  qui donne  $u_i(.) \equiv y_i - f(\mathbf{x}_i; .)$ .

**Problème :** *Passer de moments conditionnels à des moments « standards »*

Construire une condition de moment de la forme :

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

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

$$E[u(y_i, \mathbf{x}_i; \mathbf{a}_0)/\mathbf{z}_i] = E[u_i(\mathbf{a}_0)/\mathbf{z}_i] = \mathbf{0},$$

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

**Problème :** Comment choisir  $\mathbf{g}(\mathbf{w}_i; \mathbf{a})$  ?

- On analysera d'abord en détail le cas scalaire, *i.e.* le cas correspondant à une équation
  - Puis on donnera les résultats correspondant au cas multivarié, *i.e.* pour des systèmes d'équations, et ses spécificités

- On part de la caractérisation de  $E[u_i(\mathbf{a}_0)/\mathbf{z}_i] = \mathbf{0}$  :

**Propriété. Caractérisation de  $E[u_i(\mathbf{a}_0)/\mathbf{z}_i] = \mathbf{0}$**

$$E[u_i(\mathbf{a}_0)/\mathbf{z}_i] = 0$$



$$E[\rho(\mathbf{z}_i)u_i(\mathbf{a}_0)] = 0$$

pour toute fonction  $\rho(\cdot)$  telle que  $E[\rho(\mathbf{z}_i)u_i(\mathbf{a}_0)] = 0$  existe  
(conditions de régularité).

- Le terme  $E[\rho(\mathbf{z}_i)u_i(\mathbf{a}_0)]$  est *un moment non conditionnel*.
- L'équation  $E[\rho(\mathbf{z}_i)u_i(\mathbf{a}_0)] = 0$  est une condition de moment, une *condition d'orthogonalité*, potentiellement identifiante pour  $\mathbf{a}_0$ .



- Dans l'équation  $E[\rho(\mathbf{z}_i)u_i(\mathbf{a}_0)] = 0$  le terme  $\rho(\mathbf{z}_i)$  peut jouer le rôle d'*instrument*, i.e. une variable aléatoire utilisée pour construire une condition d'orthogonalité potentiellement identifiante pour  $\mathbf{a}_0$ .
- La *condition d'ordre* du problème d'identification de  $\mathbf{a}_0$  implique qu'il faut au moins autant d'instruments que de paramètres à estimer.

**Problème : Choisir un vecteur d'instruments (cas scalaire)**

Choisir un vecteur d'instruments  $\mathbf{r}(\mathbf{z}_i)$  tel que :

$$E[\mathbf{r}(\mathbf{z}_i)u_i(\mathbf{a})] = \mathbf{0} \Leftrightarrow \mathbf{a} = \mathbf{a}_0$$

ou

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

avec  $\mathbf{g}_i(\mathbf{a}) \equiv \mathbf{r}(\mathbf{z}_i)u_i(\mathbf{a}) = \mathbf{g}(\mathbf{w}_i; \mathbf{a}) = \mathbf{r}(\mathbf{z}_i)u(y_i, \mathbf{x}_i; \mathbf{a})$

à partir d'une condition de moment de la forme  $E[u_i(\mathbf{a}_0)/\mathbf{z}_i] = 0$

- Définir un *instrument* (ou un *vecteur d'instruments*)  $\mathbf{r}(\mathbf{z}_i)$  tel que :

$$E[\mathbf{r}(\mathbf{z}_i)u_i(\mathbf{a})] = \mathbf{0} \Leftrightarrow \mathbf{a} = \mathbf{a}_0$$

suppose que plusieurs conditions soient réunies. Ces conditions concernent plusieurs éléments du modèle résumé par :

$$E[u_i(\mathbf{a}_0)/\mathbf{z}_i] = 0,$$

*e.g.* le modèle défini par :

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

1. La forme de  $f(., \mathbf{a})$
2. Les variations et les relations entre les éléments de  $\mathbf{x}_i$
3. Le contenu informatif des VI  $\mathbf{z}_i$  par rapport aux explicatives  $\mathbf{x}_i$ .
4. ... et la forme de la fonction  $\mathbf{r}(.)$

- Il n'existe pas de fonction  $\mathbf{r}(.)$  pour identifier  $\mathbf{a}_0$  si les conditions 1-3 ne sont pas réunies.

- De fait, on a déjà choisi des instruments pour construire des modèles de la forme :

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

**Exemple. *Modèle linéaire à VI***

Un modèle linéaire à VI est de la forme :

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

On a alors :

$$u_i(\mathbf{a}) \equiv y_i - \mathbf{a}' \mathbf{x}_i \text{ et } u_i = u_i(\mathbf{a}_0)$$

avec :

$$E[u_i/\mathbf{z}_i] = E[u_i(\mathbf{a}_0)/\mathbf{z}_i] = 0 = E[u_i] \text{ avec } u_i(\mathbf{a}) \equiv y_i - \mathbf{a}' \mathbf{x}_i.$$

- *Dans le cadre de la MMG* on a utilisé :

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

- *Dans le cadre de la MM* on a utilisé :

$$\mathbf{r}(\mathbf{z}_i) = EL[\mathbf{x}_i/\mathbf{z}_i] \text{ et donc } \mathbf{g}_i(\mathbf{a}) = EL[\mathbf{x}_i/\mathbf{z}_i](y_i - \mathbf{a}'\mathbf{x}_i)$$

- Dans les deux cas, il est nécessaire et suffisant que  $\text{rang}E[\mathbf{x}_i\mathbf{z}_i'] = K$  pour avoir :

$$E[\mathbf{g}_i(\mathbf{a})] = E[\mathbf{r}(\mathbf{z}_i)u_i(\mathbf{a})] = \mathbf{0} \Leftrightarrow \mathbf{a} = \mathbf{a}_0,$$

ce qui suppose que  $\text{rang}E[\mathbf{x}_i\mathbf{x}_i'] = K$ .

### Terminologie non établie

Ici on distingue les *variables instrumentales*,  $\mathbf{z}_i$ , des *instruments*,  $\mathbf{r}(\mathbf{z}_i)$ .

Wooldridge (2010) désigne  $\mathbf{r}(\mathbf{z}_i)$  sous le terme de VI généralisée.

- Les résultats qui suivent vont montrer qu'il existe un meilleur instrument que  $\mathbf{r}(\mathbf{z}_i) = EL[\mathbf{x}_i/\mathbf{z}_i]$  à exploiter dans le cadre de la MM ou que  $\mathbf{r}(\mathbf{z}_i) = \mathbf{z}_i$  dans le cadre de la MMG
- On pourrait choisir un instrument,  $\boldsymbol{\rho}(\mathbf{z}_i)$ , de très grande dimension et fonder un estimateur MMG de  $\mathbf{a}_0$  sur la condition :

$$E[\boldsymbol{\rho}(\mathbf{z}_i)u_i(\mathbf{a})] = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{a} = \mathbf{a}_0.$$

- Il suffit de choisir, de manière plus ou moins *ad hoc*, des fonctions de  $\mathbf{z}_i$  pour définir  $\boldsymbol{\rho}(\mathbf{z}_i)$  mais ... on ne fait pas ça car :
  - Si la dimension de  $\boldsymbol{\rho}(\mathbf{z}_i)$  accroît l'efficacité de l'estimateur MMG de  $\mathbf{a}_0$  ... ce n'est que d'un *point de vue asymptotique*.
  - Tout estimateur MMG de  $\mathbf{a}_0$  est *biaisé à distance finie* ... et ce biais croît avec la dimension de  $\boldsymbol{\rho}(\mathbf{z}_i)$ .

**Question** : quel est le meilleur choix pour  $\mathbf{r}(\cdot)$  ?

*Il doit permettre de construire les estimateurs de  $\mathbf{a}_0$  les plus précis.*

Chamberlain (1987) a apporté une réponse définitive à cette question en définissant la notion d'

« *Instruments optimaux* » ou « *Instruments efficaces* »

- On donne ici un rôle pivotale aux « **Instruments efficaces** » ou « **Instruments optimaux** » de Chamberlain (1987), plus que dans les manuels habituels :
  - (i) Très utilisé en recherche appliquée, pour *guider les choix d'instruments* dans les modèles à variables explicatives endogènes.
  - (ii) Introduction directe du *gain d'efficacité lié à la prise en compte de l'hétéroscédasticité et des corrélations des termes d'erreur dans les systèmes d'équations*, dans un cadre général.
    - *Limite* : mêmes VI pour toutes les équations du système

## 2. La propriété des instruments estimés

## 2. La propriété des instruments estimés

- L'estimation des paramètres  $\mathbf{a}_0$  d'un modèle de la forme :

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

par la MMG passe par le choix d'instruments  $\mathbf{r}(\mathbf{z}_i)$ , *i.e.* un vecteur de fonctions des VI  $\mathbf{z}_i$  qui permettent de construire des conditions d'orthogonalité qu'on espère être estimantes :

$$E[\mathbf{r}(\mathbf{z}_i)u_i(\mathbf{a})] = \mathbf{0} \Leftrightarrow \mathbf{a} = \mathbf{a}_0$$

avec :

$$u_i(\mathbf{a}) \equiv y_i - f(\mathbf{x}_i; \mathbf{a}) \text{ et } u_i = u_i(\mathbf{a}_0).$$

- **Problème pratique.** Les instruments les plus efficaces pour l'estimation de  $\mathbf{a}_0$  sont souvent des fonctions de paramètres inconnues, *i.e.* de la forme :

$$\mathbf{r}_0(\mathbf{z}_i) \equiv \mathbf{r}(\mathbf{z}_i; \mathbf{b}_0) \text{ avec } \mathbf{b}_0 \text{ inconnu.}$$



- En particulier, l'instrument optimal au sens de Chamberlain pour le modèle précédent est souvent une fonction non triviale de  $\mathbf{a}_0$ .
- Mais une propriété sur les estimateurs utilisant des *instruments estimés* s'avère très utile.
- On part ici du problème d'estimation de  $\mathbf{a}_0$  à partir de :

$$E[u_i(\mathbf{a}_0)/\mathbf{z}_i] = \mathbf{0}$$

On sait qu'il existe un vecteur de paramètres (dits auxiliaires)  $\mathbf{b}_0$  et une fonction  $\mathbf{r}(\mathbf{z}_i; \mathbf{b}_0)$  telle que :

$$E[\mathbf{r}(\mathbf{z}_i; \mathbf{b}_0)u_i(\mathbf{a})] = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{a} = \mathbf{a}_0,$$

*i.e.*  $\mathbf{r}(\mathbf{z}_i; \mathbf{b}_0)$  est un instrument permettant d'identifier  $\mathbf{a}_0$  ( $\mathbf{b}_0$  peut contenir tout ou partie de  $\mathbf{a}_0$ ).

- On ne connaît pas  $\mathbf{b}_0$  mais on dispose d'un estimateur asymptotiquement normal (donc convergent) de  $\mathbf{b}_0$ ,  $\tilde{\mathbf{b}}_N$ .

### Propriété. *Instruments estimés*

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

$$E[\mathbf{r}(\mathbf{z}_i; \mathbf{b}_0)u(y_i, \mathbf{x}_i; \mathbf{a})] = \mathbf{0} \Leftrightarrow \mathbf{a} = \mathbf{a}_0,$$

On suppose en outre que  $\tilde{\mathbf{b}}_N$  est un estimateur as. normal de  $\mathbf{b}_0$  et on note  $\mathbf{g}(\mathbf{w}_i; \mathbf{a}; \mathbf{b}) \equiv \mathbf{r}(\mathbf{z}_i; \mathbf{b})u(y_i, \mathbf{x}_i; \mathbf{a})$  et :

$$\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N; \mathbf{b}) \equiv \arg \min_{\mathbf{a}} \bar{\mathbf{g}}_N(\mathbf{a}; \mathbf{b})' \tilde{\mathbf{M}}_N^{-1} \bar{\mathbf{g}}_N(\mathbf{a}; \mathbf{b}) \text{ où } p \lim_{N \rightarrow +\infty} \tilde{\mathbf{M}}_N = \mathbf{M}_0$$

avec  $\mathbf{M}_0$  définie positive. Les conditions de régularité usuelles sont supposées satisfaites.

Les estimateurs  $\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N; \mathbf{b}_0)$  et  $\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N; \tilde{\mathbf{b}}_N)$  ont des propriétés asymptotiques identiques.

Dit autrement, on peut « remplacer »  $\mathbf{b}_0$  par  $\tilde{\mathbf{b}}_N$  dans l'instrument  $\mathbf{r}(\mathbf{z}_i; \mathbf{b}_0)$  sans que cela n'affecte les propriétés *asymptotiques* de l'estimateur utilisé.

### ***Interprétation***

Qu'on utilise directement l'instrument  $\mathbf{r}(\mathbf{z}_i; \mathbf{b}_0)$  ou son estimateur  $\mathbf{r}(\mathbf{z}_i; \tilde{\mathbf{b}}_N)$ , on obtient des estimateurs de la MM(G) de  $\mathbf{a}_0$  qui ont les mêmes propriétés asymptotiques.

En particulier, ces estimateurs ont la même distribution asymptotique, *i.e.* ils sont as. équivalents.

Concrètement, on peut négliger le fait que  $\mathbf{b}_0$  soit remplacé par un estimateur, pourvu que cet estimateur soit as. normal.

- Cette propriété sera démontrée plus tard, lors de la présentation de l'estimation par étapes.

- La *propriété des instruments estimés* est *une des exceptions à une règle* qui veut que la distribution as. d'un estimateur (dit de second étape) construit à partir d'un autre estimateur (dit de première étape) dépend :
  - De la distribution as. de l'estimateur de première étape
  - De la manière dont le paramètre estimé en première affecte le problème définissant le calcul de l'estimateur de seconde étape
- De manière générale, si un estimateur de seconde étape est construit à partir d'un estimateur convergent et as. normal de première étape alors :
  - L'estimateur de seconde étape est toujours convergent.
  - La distribution asymptotique de l'estimateur de seconde étape est affectée par le processus d'estimation par étapes ...
  - ... sauf dans des cas particuliers, dont celui des instruments estimés.

### **Propriété. Estimation par étapes, introduction**

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

$$E[\mathbf{g}(\mathbf{w}_i; \mathbf{b}_0; \mathbf{a})] = \mathbf{0} \Leftrightarrow \mathbf{a} = \mathbf{a}_0,$$

On suppose en outre que  $\tilde{\mathbf{b}}_N$  est un estimateur as. normal de  $\mathbf{b}_0$  et on note :

$$\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N; \mathbf{b}) \equiv \arg \min_{\mathbf{a}} \bar{\mathbf{g}}_N(\mathbf{a}; \mathbf{b})' \tilde{\mathbf{M}}_N^{-1} \bar{\mathbf{g}}_N(\mathbf{a}; \mathbf{b}) \text{ où } p \lim_{N \rightarrow +\infty} \tilde{\mathbf{M}}_N = \mathbf{M}_0$$

avec  $\mathbf{M}_0$  définie positive. Les conditions de régularité usuelles sont supposées satisfaites. Bien entendu,  $\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N; \mathbf{b}_0)$  a les propriétés habituelles d'un estimateur de la MMG de  $\mathbf{a}_0$ .

(i) **Convergence et normalité as.**  $\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N; \tilde{\mathbf{b}}_N)$  est un estimateur asymptotiquement normal, et donc convergent, de  $\mathbf{a}_0$ .

(ii) *Distributions as. différentes.*  $\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N; \tilde{\mathbf{b}}_N)$  n'a, en général, pas la même distribution asymptotique que  $\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N; \mathbf{b}_0)$ .

(iii) *Distributions as. identiques.* Les estimateurs  $\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N; \tilde{\mathbf{b}}_N)$  et  $\hat{\mathbf{a}}_N^{MMG}(\tilde{\mathbf{M}}_N; \mathbf{b}_0)$  ont la même distribution asymptotique si et seulement si :

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

(iv) *Instruments estimés.* Dans le cas où  $E[\mathbf{g}(\mathbf{w}_i; \mathbf{b}_0; \mathbf{a}_0)] = \mathbf{0}$  décrit une condition d'orthogonalité où :

$$\mathbf{g}(\mathbf{w}_i; \mathbf{a}; \mathbf{b}) \equiv \mathbf{r}(\mathbf{z}_i; \mathbf{b})u(y_i, \mathbf{x}_i; \mathbf{a}) \text{ et } E[u(y_i, \mathbf{x}_i; \mathbf{a}_0)/\mathbf{z}_i] = 0$$

on a alors :

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

▪ Pour le point (iv) :

- On sait que  $E[\mathbf{g}(\mathbf{w}_i; \mathbf{b}_0; \mathbf{a}_0)] = \mathbf{0}$  décrit une condition d'orthogonalité où :

$$\mathbf{g}(\mathbf{w}_i; \mathbf{a}; \mathbf{b}) \equiv \mathbf{r}(\mathbf{z}_i; \mathbf{b})u(y_i, \mathbf{x}_i; \mathbf{a})$$

avec :

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

- On a alors :

$$E[\mathbf{r}(\mathbf{z}_i; \mathbf{b})u(y_i, \mathbf{x}_i; \mathbf{a}_0)] = \mathbf{0}$$

et finalement :

$$E\left[\frac{\partial \mathbf{g}(\mathbf{w}_i; \mathbf{b}_0; \mathbf{a}_0)}{\partial \mathbf{b}'}\right] = E\left[\frac{\partial \mathbf{r}(\mathbf{z}_i; \mathbf{b}_0)}{\partial \mathbf{b}'}u(y_i, \mathbf{x}_i; \mathbf{a}_0)\right] = \mathbf{0}$$

par l'exogénéité des  $\mathbf{z}_i$  par rapport à  $u(y_i, \mathbf{x}_i; \mathbf{a}_0)$ .

### 3. Instruments efficaces, cas scalaire



### 3. Instruments efficaces, cas scalaire

#### 3.1. Motivations : rappels

##### Problème général

Comment exploiter une condition de moment conditionnel :

$$E[u(y_i, \mathbf{x}_i; \mathbf{a}_0)/\mathbf{z}_i] = E[u_i(\mathbf{a}_0)/\mathbf{z}_i] = \mathbf{0},$$

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

##### Cadre d'analyse

On utilise le cadre de la MMG et on utilise des instruments  $\mathbf{r}(\mathbf{z}_i)$  tels que:

$$E[\mathbf{g}_i(\mathbf{a})] = E[\mathbf{r}(\mathbf{z}_i)u_i(\mathbf{a})] = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{a} = \mathbf{a}_0$$

### Recherche de l'efficacité d'estimation

On cherche à déterminer l'instrument  $\mathbf{r}^+(\mathbf{z}_i)$  tel que l'estimateur MMG de  $\mathbf{a}_0$  fondé sur la condition de moment :

$$E[\mathbf{g}_i^+(\mathbf{a})] = E[\mathbf{r}^+(\mathbf{z}_i)u_i(\mathbf{a})] = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{a} = \mathbf{a}_0$$

est *efficace dans la classe des estimateurs MMG fondés sur une condition d'orthogonalité construite avec  $u_i(\mathbf{a})$* .

- Estimateurs MMG optimaux, en principe ...
  - ... on ne se servira que d'estimateurs **MM** car juste-identification.
- L'efficacité est définie par rapport une classe d'estimateurs MM(G) ...
  - ... mais le résultat de Chamberlain est bien plus général.
- L'« instrument efficace » ... n'est pas unique.
- L'efficacité sera toujours relative aux VI,  $\mathbf{z}_i$ , utilisées comme ensemble d'information du modèle considéré.

### 3.2. Les instruments efficaces pour $E[u_i(\mathbf{a}_0)/\mathbf{z}_i] = \mathbf{0}$

#### Propriété. *Instruments efficaces, cas scalaire*

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

$$(i) \quad E[u(y_i, \mathbf{x}_i; \mathbf{a}_0)/\mathbf{z}_i] = E[u_i(\mathbf{a}_0)/\mathbf{z}_i] = 0 \text{ et } E[u_i(\mathbf{a}_0)^2/\mathbf{z}_i] > 0,$$

$$(ii) \quad E\left[\frac{\partial u_i(\mathbf{a}_0)}{\partial \mathbf{a}} \frac{\partial u_i(\mathbf{a}_0)}{\partial \mathbf{a}'} / \mathbf{z}_i\right] \text{ est de rang } K \equiv \dim \mathbf{a},$$

alors l'estimateur de la MM de  $\mathbf{a}_0$  fondé sur la condition de moment :

$$E\left[E\left[\frac{\partial u_i(\mathbf{a}_0)}{\partial \mathbf{a}} / \mathbf{z}_i\right] E\left[u_i(\mathbf{a}_0)^2 / \mathbf{z}_i\right]^{-1} u_i(\mathbf{a}_0)\right] = \mathbf{0}_{K \times 1}$$

est as. efficace dans la classe des estimateurs as. normaux de  $\mathbf{a}_0$  sous les conditions de régularité usuelles.

## Remarques générales

- Le terme :

$$\mathbf{r}^+(\mathbf{z}_i) \equiv E \left[ \frac{\partial u_i(\mathbf{a}_0)}{\partial \mathbf{a}} / \mathbf{z}_i \right] E \left[ u_i(\mathbf{a}_0)^2 / \mathbf{z}_i \right]^{-1}$$

est désigné sous le terme d'*instrument efficace* (ou *optimal*).

- $\mathbf{r}^+(\mathbf{z}_i)$  est exogène car c'est une fonction des variables exogènes  $\mathbf{z}_i$
- La condition  $E[\mathbf{r}^+(\mathbf{z}_i)u_i(\mathbf{a}_0)] = \mathbf{0}$  juste-identifie  $\mathbf{a}_0$ , pas besoin de la MMG.
- Cette propriété est très « puissante » car l'estimateur MM de  $\mathbf{a}_0$  défini par :

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

*est l'estimateur as. normal le plus efficace de  $\mathbf{a}_0$ .*

**Définition. Borne d'efficacité semi-paramétrique**

$$\sqrt{N}(\hat{\mathbf{a}}_N^{MM} - \mathbf{a}_0) \xrightarrow[N \rightarrow +\infty]{L} \mathcal{N}(\mathbf{0}; \Sigma_0^*)$$

où :

$$\Sigma_0^* \equiv E \left[ \frac{\partial u_i(\mathbf{a}_0)}{\partial \mathbf{a}} E \left[ u_i(\mathbf{a}_0)^2 / \mathbf{z}_i \right]^{-1} \frac{\partial u_i(\mathbf{a}_0)}{\partial \mathbf{a}'} \right]^{-1}.$$

$\Sigma_0^*$  est la *borne minimax* ou *borne d'efficacité semi-paramétrique* (des estimateurs as. normaux de  $\mathbf{a}_0$ ).

**Propriété. Transformation affine non singulière des instruments efficaces**

Si  $\mathbf{B}$  est une matrice carrée de dimension  $K$  et de rang  $K$ , alors  $\mathbf{B}\mathbf{r}^+(\mathbf{z}_i)$  est également un instrument efficace.

- La démonstration de ces propriétés est ardue et peu instructive. Ceci dit, la forme de  $\mathbf{r}^+(\mathbf{z}_i)$  est relativement intuitive.
- Dans la suite on notera :

$$\gamma_i(\mathbf{a}) \equiv \frac{\partial u_i(\mathbf{a})}{\partial \mathbf{a}},$$

$$\gamma_0(\mathbf{z}_i) \equiv E\left[\frac{\partial u_i(\mathbf{a}_0)}{\partial \mathbf{a}} / \mathbf{z}_i\right] = E[\gamma_i(\mathbf{a}_0) / \mathbf{z}_i]$$

et :

$$\omega_0(\mathbf{z}_i) \equiv E[u_i(\mathbf{a}_0)^2 / \mathbf{z}_i]$$

et finalement :

$$\mathbf{r}^+(\mathbf{z}_i) = \gamma_0(\mathbf{z}_i) \omega_0(\mathbf{z}_i)^{-1}.$$

**Interprétation,  $\omega_0(\mathbf{z}_i) \equiv E[u_i(\mathbf{a}_0)^2 / \mathbf{z}_i]$**

$$\mathbf{g}_i^+(\mathbf{a}) = \underbrace{\gamma_0(\mathbf{z}_i)\omega_0(\mathbf{z}_i)^{-1}}_{\mathbf{r}^+(\mathbf{z}_i)} u_i(\mathbf{a})$$

- $\omega_0(\mathbf{z}_i)$  est la variance conditionnelle en  $\mathbf{z}_i$  de  $u_i = u_i(\mathbf{a}_0)$ 
  - Si  $y_i = f(\mathbf{x}_i; \mathbf{a}_0) + u_i$  avec  $E[u_i / \mathbf{z}_i] = E[u_i] \equiv 0$  alors  $u_i(\mathbf{a}_0) = u_i$ .
  - $\omega_0(\mathbf{z}_i)$  tient compte de l'éventuelle *hétéroscédasticité conditionnelle* de  $u_i(\mathbf{a}_0) = u_i$ .
    - Correction pour l'hétéroscédasticité des **MCG** ou des **MCQG**.

**Intuition.** Multiplier  $u_i(\mathbf{a})$  par  $\omega_0(\mathbf{z}_i)^{-1}$  revient à « sur-pondérer » les observations les moins bruitées, *i.e.* les plus « fiables ».

**Rmq.** Dans le critère de la MMG, *la matrice  $\tilde{\mathbf{W}}_N^{-1}$  pondère des conditions de moment*, les  $\bar{\mathbf{g}}_N(\mathbf{a}) \equiv N^{-1} \sum_{i=1}^N \mathbf{g}_i(\mathbf{a})$

*Ici on pondère des observations* : les  $\omega_0(\mathbf{z}_i)^{-1}$  pondèrent les  $u_i(\mathbf{a})$  dans :

$$\bar{\mathbf{g}}_N(\mathbf{a}) \equiv N^{-1} \sum_{i=1}^N \gamma_0(\mathbf{z}_i) \omega_0(\mathbf{z}_i)^{-1} u_i(\mathbf{a}).$$

Les  $\omega_0(\mathbf{z}_i)^{-1}$  pondèrent les  $\gamma_0(\mathbf{z}_i) u_i(\mathbf{a})$  dans :

$$\bar{\mathbf{g}}_N(\mathbf{a}) \equiv N^{-1} \sum_{i=1}^N \gamma_0(\mathbf{z}_i) \omega_0(\mathbf{z}_i)^{-1} h_i(\mathbf{a}).$$



$$\text{Interprétation de } \gamma_0(\mathbf{z}_i) \equiv E \left[ \frac{\partial u_i(\mathbf{a}_0)}{\partial \mathbf{a}} \middle/ \mathbf{z}_i \right]$$

$$\mathbf{g}_i^+(\mathbf{a}) = \underbrace{\gamma_0(\mathbf{z}_i) \omega_0(\mathbf{z}_i)^{-1}}_{\mathbf{r}^+(\mathbf{z}_i)} u_i(\mathbf{a})$$

- Le terme  $\omega_0(\mathbf{z}_i)^{-1}$  joue essentiellement un rôle de *pondération*, i.e. de prise en compte de l'*hétéroscédasticité* de  $u_i(\mathbf{a}_0)$ .

Le terme  $\gamma_0(\mathbf{z}_i)$  porte la « *structure* » de  $\mathbf{r}^+(\mathbf{z}_i) = \gamma_0(\mathbf{z}_i) \omega_0(\mathbf{z}_i)^{-1}$ , *il est essentiel pour l'identification de  $\mathbf{a}_0$* .

- $\gamma_0(\mathbf{z}_i)$  est un bon instrument car c'est la *fonction de  $\mathbf{z}_i$  qui a la plus « forte » corrélation avec  $u_i(\mathbf{a})$*  pour  $\mathbf{a}$  autour de  $\mathbf{a}_0$
- On peut donner l'intuition de l'« optimalité » de  $\gamma_0(\mathbf{z}_i)$  en tant qu'instrument en linéarisant le problème d'estimation.

- On doit choisir un instrument  $\mathbf{r}(\mathbf{z}_i)$  pour  $u_i(\mathbf{a}_0)$  de telle sorte à ce que l'estimateur de la MM de  $\mathbf{a}_0$  fondé sur :

$$E[\mathbf{r}(\mathbf{z}_i)u_i(\mathbf{a}_0)] = \mathbf{0} = Cov[\mathbf{r}(\mathbf{z}_i); u_i(\mathbf{a}_0)]$$

(i) identifie  $\mathbf{a}_0$

et :

(ii) soit le plus efficace possible.

- Pour l'identification il faut que  $Cov[\mathbf{r}(\mathbf{z}_i); u_i(\mathbf{a})] \neq \mathbf{0}$  si  $\mathbf{a} \neq \mathbf{a}_0$ , donc il faut que  $\mathbf{r}(\mathbf{z}_i)$  soit *corrélé au mieux* à  $u_i(\mathbf{a})$  lorsque  $\mathbf{a} \neq \mathbf{a}_0$ , et ce même lorsque  $\mathbf{a}$  est très proche de  $\mathbf{a}_0$ .
  - **Intuition** :  $\mathbf{r}(\mathbf{z}_i)$  doit permettre de « bien » détecter si  $\mathbf{a} \neq \mathbf{a}_0$ .
  - **Intuition** : plus  $\mathbf{r}(\mathbf{z}_i)$  est corrélé à  $u_i(\mathbf{a})$  pour  $\mathbf{a} \neq \mathbf{a}_0$ , plus  $\mathbf{r}(\mathbf{z}_i)$  est efficace en tant qu'instrument d'identification.

- On veut  $\mathbf{r}(\mathbf{z}_i)$  tel que :

$$Cov[\mathbf{r}(\mathbf{z}_i); u_i(\mathbf{a})] = \mathbf{0} \Leftrightarrow \mathbf{a} = \mathbf{a}_0$$

lorsque  $\mathbf{a}$  est au voisinage de  $\mathbf{a}_0$ .

- Développement limité au premier ordre en  $\mathbf{a}$  de  $u_i(\mathbf{a})$  autour de  $\mathbf{a}_0$  :

$$u_i(\mathbf{a}) \simeq u_i(\mathbf{a}_0) + \frac{\partial u_i(\mathbf{a}_0)}{\partial \mathbf{a}'} (\mathbf{a} - \mathbf{a}_0)$$

or  $E[u_i(\mathbf{a}_0)/\mathbf{z}_i] = 0$  implique que  $Cov[\mathbf{r}(\mathbf{z}_i); u_i(\mathbf{a}_0)] = \mathbf{0}$  et donc que :

$$Cov[\mathbf{r}(\mathbf{z}_i); u_i(\mathbf{a})] \simeq Cov\left[\mathbf{r}(\mathbf{z}_i); \frac{\partial u_i(\mathbf{a}_0)}{\partial \mathbf{a}}\right] (\mathbf{a} - \mathbf{a}_0)$$

- On a :

$$\gamma_0(\mathbf{z}_i) = E\left[\frac{\partial u_i(\mathbf{a}_0)}{\partial \mathbf{a}} \middle/ \mathbf{z}_i\right] \Rightarrow Cov\left[\mathbf{r}(\mathbf{z}_i); \frac{\partial u_i(\mathbf{a}_0)}{\partial \mathbf{a}}\right] = Cov[\mathbf{r}(\mathbf{z}_i); \gamma_0(\mathbf{z}_i)].$$

- Avec :

$$\text{Cov}[\mathbf{r}(\mathbf{z}_i); u_i(\mathbf{a})] \simeq \text{Cov}[\mathbf{r}(\mathbf{z}_i); \gamma_0(\mathbf{z}_i)](\mathbf{a} - \mathbf{a}_0)$$

on voit que :

$$\text{un des « meilleurs » candidats pour } \mathbf{r}^+(\mathbf{z}_i) \text{ est } \gamma_0(\mathbf{z}_i) = E \left[ \frac{\partial u_i(\mathbf{a}_0)}{\partial \mathbf{a}} / \mathbf{z}_i \right].$$

- On veut  $\text{Cov}[\mathbf{r}^+(\mathbf{z}_i); \gamma_0(\mathbf{z}_i)](\mathbf{a} - \mathbf{a}_0) \neq \mathbf{0}$  lorsque  $\mathbf{a} \neq \mathbf{a}_0$ . Or on a :

$$\text{Cov}[\mathbf{r}^+(\mathbf{z}_i); \gamma_0(\mathbf{z}_i)] = V[\gamma_0(\mathbf{z}_i)] \text{ avec } \mathbf{r}^+(\mathbf{z}_i) = \gamma_0(\mathbf{z}_i)$$

et si  $V[\gamma_0(\mathbf{z}_i)]$  est inversible alors :

$$\text{Cov}[\mathbf{r}^+(\mathbf{z}_i); \gamma_0(\mathbf{z}_i)](\mathbf{a} - \mathbf{a}_0) = V[\gamma_0(\mathbf{z}_i)](\mathbf{a} - \mathbf{a}_0) = \mathbf{0} \Leftrightarrow \mathbf{a} = \mathbf{a}_0.$$

$\gamma_0(\mathbf{z}_i) = E\left[\frac{\partial u_i(\mathbf{a}_0)}{\partial \mathbf{a}} / \mathbf{z}_i\right]$  est un bon instrument car c'est la fonction de  $\mathbf{z}_i$  qui a la plus « forte » corrélation avec  $u_i(\mathbf{a})$  pour  $\mathbf{a}$  autour de  $\mathbf{a}_0$ .

- La condition d'inversibilité de  $E[\gamma_0(\mathbf{z}_i)\gamma_0(\mathbf{z}_i)']$  (condition (ii) de la propriété) est équivalente à celle de  $V[\gamma_0(\mathbf{z}_i)]$ . C'est une condition de rang locale, *i.e.* une ***condition locale d'identification***.
  - C'est la condition locale de rang, ***condition locale d'identification***, sur le terme  $\mathbf{G}_0$  dans la MM(G).
  - Elle est analogue à la condition de rang,  $\text{rang}E[\mathbf{z}_i\mathbf{x}_i'] = K$ , nécessaire à l'identification de  $\mathbf{a}_0$  par l'estimateur des 2MC dans le modèle linéaire à VI.
  - Pour un modèle linéaire à VI cette condition est équivalente à  $\text{rang}E[E[\mathbf{x}_i/\mathbf{z}_i]E[\mathbf{x}_i'/\mathbf{z}_i]] = K$ .

### 3.3. Quelques exemples

- **Objectifs** : Montrer comment utiliser le résultat de Chamberlain, en particulier :
  - Illustrer le rôle simplificateur des hypothèses
    - d'exogénéité
    - d'homoscédasticité (cas une seule équation)
    - de linéarité
  - Comment gérer les termes d'erreur hétéroscéastiques lorsque l'*hétéroscéasticité* est *de forme inconnue*
  - Montrer que l'instrument optimal dépend en général de paramètres inconnus, ce qui montre l'utilité de la *propriété des instruments estimés*

## Remarques générales

- 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, le résultat de Chamberlain permet de ré-interpréter beaucoup de résultats concernant les estimateurs des MC, les MCG, les MCQG et des 2MC.

### Exemple 1. *Modèle de régression linéaire*

- Un modèle de régression linéaire est de la forme :

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

- On a alors :

$$E[y_i - \mathbf{a}'_0 \mathbf{x}_i / \mathbf{x}_i] = 0$$

avec :

$$u_i(\mathbf{a}) \equiv y_i - \mathbf{a}' \mathbf{x}_i, \mathbf{z}_i = \mathbf{x}_i \text{ et } u_i = y_i - \mathbf{a}'_0 \mathbf{x}_i = u_i(\mathbf{a}_0).$$

- On a :

$$E[u_i(\mathbf{a}_0) / \mathbf{x}_i] = E[u_i / \mathbf{x}_i] = 0$$

- La forme de l'instrument optimal est donnée par :

$$\mathbf{r}_i^+ \equiv E \left[ \frac{\partial u_i(\mathbf{a}_0)}{\partial \mathbf{a}} / \mathbf{x}_i \right] E[u_i(\mathbf{a}_0)^2 / \mathbf{x}_i]^{-1} = -\mathbf{x}_i E[u_i^2 / \mathbf{x}_i]^{-1}$$



- Il y a plusieurs cas à considérer selon que :

- $E[u_i^2 / \mathbf{x}_i] = \omega_0$  : *homoscédasticité* des  $u_i = y_i - \mathbf{a}_0' \mathbf{x}_i = u_i(\mathbf{a}_0)$

- $E[u_i^2 / \mathbf{x}_i] = \omega(\mathbf{x}_i; \mathbf{b}_0)$  avec  $\omega(\cdot)$  et  $\mathbf{b}_0$  connus : *hétéroscédasticité conditionnelle de forme connue* (éventuellement à un terme positif multiplicatif près)

- $E[u_i^2 / \mathbf{x}_i] = \omega(\mathbf{x}_i)$  et  $\omega(\cdot)$  inconnue : *hétéroscédasticité conditionnelle de forme inconnue*

- Le dernier cas utilise la propriété des instruments estimés :

- $E[u_i^2 / \mathbf{x}_i] = \omega(\mathbf{x}_i; \mathbf{b}_0)$ ,  $\omega(\cdot)$  connue,  $\mathbf{b}_0$  inconnu mais  $\tilde{\mathbf{b}}_N$  est un estimateur as. normal de  $\mathbf{b}_0$  : *hétéroscédasticité conditionnelle de forme connue estimée*.

### 1a. MRL et homoscedasticité

- On a :

$$y_i = \mathbf{a}'_0 \mathbf{x}_i + u_i \text{ avec } E[u_i/\mathbf{x}_i] = E[u_i] \equiv 0 \text{ et } E[u_i^2/\mathbf{x}_i] = \omega_0$$

et :

$$\mathbf{r}_i^+ = -\mathbf{x}_i \omega_0^{-1}$$

- L'estimateur de la MM de  $\mathbf{a}_0$  fondé sur :

$$E[\mathbf{r}_i^+ u_i(\mathbf{a}_0)] = \mathbf{0} \Leftrightarrow E[\mathbf{x}_i (y_i - \mathbf{x}'_i \mathbf{a}_0)] = \mathbf{0} \Leftrightarrow E[\mathbf{g}_i^+(\mathbf{a}_0)] = \mathbf{0}$$

est l'estimateur des MCO de  $\mathbf{a}_0$ ,  $\hat{\mathbf{a}}_N^{MCO}$ .

- **Chamberlain.** Dans le modèle précédent,  $\hat{\mathbf{a}}_N^{MCO}$  est as. efficace dans la classe des estimateurs as. normaux de  $\mathbf{a}_0$ .

- **MMG.** On a :

$$\sqrt{N}(\hat{\mathbf{a}}_N^{MCO} - \mathbf{a}_0) \xrightarrow[N \rightarrow +\infty]{L} \mathcal{N}(\mathbf{0}; \Sigma_0) \text{ avec } \Sigma_0 \equiv (\mathbf{G}_0' \mathbf{W}_0^{-1} \mathbf{G}_0)^{-1}$$

avec :

$$\mathbf{W}_0 \equiv E[\mathbf{g}_i(\mathbf{a}_0) \mathbf{g}_i(\mathbf{a}_0)'] = E[\mathbf{x}_i u_i^2 \mathbf{x}_i'] = \omega_0 E[\mathbf{x}_i \mathbf{x}_i'] \text{ (homoscédasticité)}$$

$$\mathbf{G}_0 \equiv E\left[\frac{\partial \mathbf{g}_i(\mathbf{a}_0)'}{\partial \mathbf{a}}\right] = -E[\mathbf{x}_i \mathbf{x}_i']$$

et donc :

$$\Sigma_0 \equiv (\mathbf{G}_0' \mathbf{W}_0^{-1} \mathbf{G}_0)^{-1} = \omega_0 E[\mathbf{x}_i \mathbf{x}_i']^{-1} \text{ avec } \omega_0 = E[(y_i - \mathbf{x}_i' \mathbf{a}_0)^2].$$

- **MM.** (Juste-identification). On a :

$$\sqrt{N}(\hat{\mathbf{a}}_N^{MCO} - \mathbf{a}_0) \xrightarrow[N \rightarrow +\infty]{L} \mathcal{N}(\mathbf{0}; \Sigma_0) \text{ avec } \Sigma_0 \equiv (\mathbf{G}_0)^{-1} \mathbf{W}_0 (\mathbf{G}_0')^{-1}$$

avec :

$$\Sigma_0 \equiv (\mathbf{G}_0' \mathbf{W}_0^{-1} \mathbf{G}_0)^{-1} = \omega_0 E[\mathbf{x}_i \mathbf{x}_i']^{-1}.$$

### 1b. MRL et hétéroscédasticité de forme connue

- On a :

$$y_i = \mathbf{a}'_0 \mathbf{x}_i + u_i \text{ avec } E[u_i / \mathbf{x}_i] = E[u_i] \equiv 0 \text{ et } E[u_i^2 / \mathbf{x}_i] = \omega(\mathbf{x}_i; \mathbf{b}_0)$$

et :

$$\mathbf{r}_i^+ = -\mathbf{x}_i \omega(\mathbf{x}_i; \mathbf{b}_0)^{-1}$$

- L'estimateur de la MM de  $\mathbf{a}_0$  fondé sur :

$$E[\mathbf{r}_i^+ u_i(\mathbf{a}_0)] = \mathbf{0} \Leftrightarrow E[\mathbf{x}_i \omega(\mathbf{x}_i; \mathbf{b}_0)^{-1} (y_i - \mathbf{x}'_i \mathbf{a}_0)] = \mathbf{0} \Leftrightarrow E[\mathbf{g}_i^+(\mathbf{a}_0)] = \mathbf{0}$$

est l'estimateur des MCG de  $\mathbf{a}_0$  :

$$\hat{\mathbf{a}}_N^{MCG} = \left( N^{-1} \sum_{i=1}^N \mathbf{x}_i \omega(\mathbf{x}_i; \mathbf{b}_0)^{-1} \mathbf{x}'_i \right)^{-1} N^{-1} \sum_{i=1}^N \mathbf{x}_i \omega(\mathbf{x}_i; \mathbf{b}_0)^{-1} y_i.$$

- **Chamberlain.** Dans le modèle précédent,  $\hat{\mathbf{a}}_N^{MCG}$  est as. efficace dans la classe des estimateurs as. normaux de  $\mathbf{a}_0$ .

- **MMG**. On a :

$$\sqrt{N}(\hat{\mathbf{a}}_N^{MCG} - \mathbf{a}_0) \xrightarrow[N \rightarrow +\infty]{L} \mathcal{N}(\mathbf{0}; \Sigma_0) \text{ avec } \Sigma_0 \equiv (\mathbf{G}'_0 \mathbf{W}_0^{-1} \mathbf{G}_0)^{-1}$$

avec :

$$\mathbf{W}_0 \equiv E[\mathbf{g}_i(\mathbf{a}_0) \mathbf{g}_i(\mathbf{a}_0)'] = E[\mathbf{x}_i \omega(\mathbf{x}_i; \mathbf{b}_0)^{-1} u_i^2 \omega(\mathbf{x}_i; \mathbf{b}_0)^{-1} \mathbf{x}_i'] = E[\mathbf{x}_i \omega(\mathbf{x}_i; \mathbf{b}_0)^{-1} \mathbf{x}_i']$$

$$\mathbf{G}_0 \equiv E\left[\frac{\partial \mathbf{g}_i(\mathbf{a}_0)'}{\partial \mathbf{a}}\right] = -E[\mathbf{x}_i \omega(\mathbf{x}_i; \mathbf{b}_0)^{-1} \mathbf{x}_i']$$

et donc :

$$\Sigma_0 = E[\mathbf{x}_i \omega(\mathbf{x}_i; \mathbf{b}_0)^{-1} \mathbf{x}_i']^{-1}.$$

- **MM** (Juste-identification). On a :

$$\sqrt{N}(\hat{\mathbf{a}}_N^{MCG} - \mathbf{a}_0) \xrightarrow[N \rightarrow +\infty]{L} \mathcal{N}(\mathbf{0}; \Sigma_0) \text{ avec } \Sigma_0 \equiv (\mathbf{G}_0)^{-1} \mathbf{W}_0 (\mathbf{G}_0')^{-1}$$

avec :

$$\Sigma_0 \equiv (\mathbf{G}'_0 \mathbf{W}_0^{-1} \mathbf{G}_0)^{-1} = E[\mathbf{x}_i \omega(\mathbf{x}_i; \mathbf{b}_0)^{-1} \mathbf{x}_i']^{-1}.$$

### 1c. MRL et hétéroscédasticité de forme inconnue

- On a :

$$y_i = \mathbf{a}_0' \mathbf{x}_i + u_i \text{ avec } E[u_i / \mathbf{x}_i] = E[u_i] \equiv 0 \text{ et } E[u_i^2 / \mathbf{x}_i] = \omega(\mathbf{x}_i) \neq \omega_0$$

- **Problème** : On ne connaît pas la forme de  $\omega(\cdot)$ , on ne peut donc utiliser  $\mathbf{r}_i^+$ , ou un estimateur de  $\mathbf{r}_i^+$
- **Solution** : On choisit l'instrument  $\mathbf{r}_i$  le plus proche de  $\mathbf{r}_i^+$  qu'on puisse utiliser,  $\mathbf{r}_i = \mathbf{x}_i$  ici. On utilise alors l'estimateur de la MM de  $\mathbf{a}_0$  fondé sur :

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

i.e. l'estimateur  $\hat{\mathbf{a}}_N^{MCO}$ .

Mais *on garde à l'esprit* que :  $E[u_i^2 / \mathbf{x}_i] = \omega(\mathbf{x}_i) \neq \omega_0$

- **Chamberlain.** Dans le modèle précédent,  $\hat{\mathbf{a}}_N^{MCO}$  n'est pas as. efficace dans la classe des estimateurs as. normaux de  $\mathbf{a}_0$ .
- Il est cependant *robuste à l'hétéroscédasticité (conditionnelle) des  $u_i$*
- **MMG et MM.** On a :

$$\sqrt{N}(\hat{\mathbf{a}}_N^{MCO} - \mathbf{a}_0) \xrightarrow[N \rightarrow +\infty]{L} \mathcal{N}(\mathbf{0}; \Sigma_0)$$

avec :

$$\Sigma_0 \equiv (\mathbf{G}_0' \mathbf{W}_0^{-1} \mathbf{G}_0)^{-1} = (\mathbf{G}_0)^{-1} \mathbf{W}_0 (\mathbf{G}_0')$$

avec :

$$\mathbf{W}_0 \equiv E[\mathbf{g}_i(\mathbf{a}_0) \mathbf{g}_i(\mathbf{a}_0)'] = E[\mathbf{x}_i (y_i - \mathbf{a}_0' \mathbf{x}_i)^2 \mathbf{x}_i']$$

$$\mathbf{G}_0 \equiv E\left[\frac{\partial \mathbf{g}_i(\mathbf{a}_0)'}{\partial \mathbf{a}}\right] = -E[\mathbf{x}_i \mathbf{x}_i']$$

- On a donc :

$$\Sigma_0 = E[\mathbf{x}_i \mathbf{x}_i']^{-1} E[\mathbf{x}_i (y_i - \mathbf{a}_0' \mathbf{x}_i)^2 \mathbf{x}_i'] E[\mathbf{x}_i \mathbf{x}_i']^{-1},$$

*i.e.* on retrouve donc la matrice de variance covariance as. de  $\hat{\mathbf{a}}_N^{MCO}$  dite de White ou robuste à l'hétéroscédasticité.

- On parle alors d'*inférence robuste par rapport à l'hétéroscédasticité*.



### 1d. MRL et hétéroscédasticité de forme connue estimée

- On a :

$$y_i = \mathbf{a}_0' \mathbf{x}_i + u_i \text{ avec } E[u_i / \mathbf{x}_i] = E[u_i] \equiv 0 \text{ et } E[u_i^2 / \mathbf{x}_i] = \omega(\mathbf{x}_i; \mathbf{b}_0)$$

et :

$$\mathbf{r}_i^+ = -\mathbf{x}_i \omega(\mathbf{x}_i; \mathbf{b}_0)^{-1} \text{ et } \tilde{\mathbf{b}}_N \text{ est un estimateur as. normal de } \mathbf{b}_0$$

- L'estimateur de la MM de  $\mathbf{a}_0$  fondé sur :

$$E[\mathbf{r}_i^+ u_i(\mathbf{a}_0)] = \mathbf{0} \Leftrightarrow E[\mathbf{x}_i \omega(\mathbf{x}_i; \mathbf{b}_0)^{-1} (y_i - \mathbf{x}_i' \mathbf{a}_0)] = \mathbf{0} \Leftrightarrow E[\mathbf{g}_i^+(\mathbf{a}_0)] = \mathbf{0}$$

utilisant  $\tilde{\mathbf{b}}_N$  en lieu et place de  $\mathbf{b}_0$  est l'estimateur des MCQG de  $\mathbf{a}_0$  :

$$\hat{\mathbf{a}}_N^{MCQG} = \left( N^{-1} \sum_{i=1}^N \mathbf{x}_i \omega(\mathbf{x}_i; \tilde{\mathbf{b}}_N)^{-1} \mathbf{x}_i' \right)^{-1} N^{-1} \sum_{i=1}^N \mathbf{x}_i \omega(\mathbf{x}_i; \tilde{\mathbf{b}}_N)^{-1} y_i.$$

- **Instruments estimés.**  $\tilde{\mathbf{b}}_N$  étant un estimateur as. normal de  $\mathbf{b}_0$  on sait que :

$$\hat{\mathbf{a}}_N^{MCQG} = \left( N^{-1} \sum_{i=1}^N \mathbf{x}_i \omega(\mathbf{x}_i; \tilde{\mathbf{b}}_N)^{-1} \mathbf{x}_i' \right)^{-1} N^{-1} \sum_{i=1}^N \mathbf{x}_i \omega(\mathbf{x}_i; \tilde{\mathbf{b}}_N)^{-1} y_i$$

a les mêmes propriétés as. que :

$$\hat{\mathbf{a}}_N^{MCG} = \left( N^{-1} \sum_{i=1}^N \mathbf{x}_i \omega(\mathbf{x}_i; \mathbf{b}_0)^{-1} \mathbf{x}_i' \right)^{-1} N^{-1} \sum_{i=1}^N \mathbf{x}_i \omega(\mathbf{x}_i; \mathbf{b}_0)^{-1} y_i.$$

- Cette propriété des estimateurs MCG et MCQG est en fait une application de la *propriété des instruments estimés*.
- **Propriétés as. de  $\hat{\mathbf{a}}_N^{MCG}$ .** Les propriétés de  $\hat{\mathbf{a}}_N^{MCQG}$  sont celles de  $\hat{\mathbf{a}}_N^{MCG}$  :

$$\sqrt{N}(\hat{\mathbf{a}}_N^{MCQG} - \mathbf{a}_0) \xrightarrow[N \rightarrow +\infty]{L} \mathcal{N}\left(\mathbf{0}; E\left[\mathbf{x}_i \omega(\mathbf{x}_i; \mathbf{b}_0)^{-1} \mathbf{x}_i'\right]^{-1}\right)$$

et :

$$N^{-1} \sum_{i=1}^N \mathbf{x}_i \omega(\mathbf{x}_i; \tilde{\mathbf{b}}_N)^{-1} \mathbf{x}_i' \xrightarrow[N \rightarrow +\infty]{p} E\left[\mathbf{x}_i \omega(\mathbf{x}_i; \mathbf{b}_0)^{-1} \mathbf{x}_i'\right].$$

### Exemple 2. *Modèle de régression non linéaire*

- Un modèle de régression non linéaire est de la forme :

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

- On a alors :

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

avec :

$$u_i(\mathbf{a}) \equiv y_i - f(\mathbf{x}_i; \mathbf{a}), \mathbf{z}_i = \mathbf{x}_i \text{ et } u_i = y_i - f(\mathbf{x}_i; \mathbf{a}_0) = u_i(\mathbf{a}_0)$$

- On a :

$$E[u_i(\mathbf{a}_0) / \mathbf{x}_i] = E[u_i / \mathbf{x}_i] = 0$$

- La forme de l'instrument optimal est donnée par :

$$\mathbf{r}_i^+ \equiv E\left[\frac{\partial u_i(\mathbf{a}_0)}{\partial \mathbf{a}} / \mathbf{x}_i\right] E[u_i(\mathbf{a}_0)^2 / \mathbf{x}_i]^{-1} = -\frac{\partial f(\mathbf{x}_i; \mathbf{a}_0)}{\partial \mathbf{a}} E[u_i^2 / \mathbf{x}_i]^{-1}$$

- Il y a plusieurs cas selon que :

- $E[u_i^2 / \mathbf{x}_i] = \omega_0$  : *homoscédasticité* des  $u_i = y_i - \mathbf{a}_0' \mathbf{x}_i = h_i(\mathbf{a}_0)$
  - $E[u_i^2 / \mathbf{x}_i] = \omega(\mathbf{x}_i; \mathbf{b}_0)$  avec  $\omega(\cdot)$  et  $\mathbf{b}_0$  connus : *hétéroscédasticité conditionnelle de forme connue* (éventuellement à un terme positif multiplicatif près)
  - $E[u_i^2 / \mathbf{x}_i] = \omega(\mathbf{x}_i; \mathbf{b}_0)$ ,  $\omega(\cdot)$  connue,  $\mathbf{b}_0$  inconnu mais  $\tilde{\mathbf{b}}_N$  est un estimateur as. normal de  $\mathbf{b}_0$  : *hétéroscédasticité conditionnelle de forme connue estimée*
  - $E[u_i^2 / \mathbf{x}_i] = \omega(\mathbf{x}_i)$  et  $\omega(\cdot)$  inconnue : *hétéroscédasticité conditionnelle de forme connue*
- On ne s'intéresse ici qu'au « ***cas homoscédastique*** »

## *Modèle de régression non linéaire avec homoscédasticité*

- On a :

$$y_i = f(\mathbf{x}_i; \mathbf{a}_0) + u_i \quad \text{avec} \quad E[u_i / \mathbf{x}_i] = E[u_i] \equiv 0 \quad \text{et} \quad E[u_i^2 / \mathbf{x}_i] = \omega_0$$

et :

$$\mathbf{r}_i^+ = - \frac{\partial f(\mathbf{x}_i; \mathbf{a}_0)}{\partial \mathbf{a}} \omega_0^{-1}$$

- L'estimateur de la MM de  $\mathbf{a}_0$  fondé sur :

$$\begin{aligned} E[\mathbf{r}_i^+ u_i(\mathbf{a}_0)] = \mathbf{0} &\Leftrightarrow E\left[\frac{\partial f(\mathbf{x}_i; \mathbf{a}_0)}{\partial \mathbf{a}} (y_i - f(\mathbf{x}_i; \mathbf{a}_0))\right] = \mathbf{0} \\ &\Leftrightarrow E[\mathbf{g}_i(\mathbf{a}_0)] = \mathbf{0} \end{aligned}$$

est l'estimateur des MCO non linéaire de  $\mathbf{a}_0$ ,  $\hat{\mathbf{a}}_N^{MCO}$ .

- Dans le cadre de la MM(G), l'estimateur de  $\mathbf{a}_0$  fondé sur :

$$E[\mathbf{g}_i(\mathbf{a}_0)] = E\left[\frac{\partial f(\mathbf{x}_i; \mathbf{a}_0)}{\partial \mathbf{a}} (y_i - f(\mathbf{x}_i; \mathbf{a}_0))\right] = \mathbf{0}_{K \times 1}$$

est caractérisé par :

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

- C'est bien l'estimateur des MCO non linéaires :

$$\hat{\mathbf{a}}_N^{MCO} \equiv \arg \min_{\mathbf{a}} \sum_{i=1}^N (y_i - f(\mathbf{x}_i; \mathbf{a}))^2$$

car les CO1 de ce programme de minimisation sont données par :

$$-2 \times \sum_{i=1}^N \frac{\partial f(\mathbf{x}_i; \hat{\mathbf{a}}_N^{MCO})}{\partial \mathbf{a}} (y_i - f(\mathbf{x}_i; \hat{\mathbf{a}}_N^{MCO})) = \mathbf{0}_{K \times 1}.$$

- **Chamberlain.** Dans le modèle précédent,  $\hat{\mathbf{a}}_N^{MCO}$  est as. efficace dans la classe des estimateurs as. normaux de  $\mathbf{a}_0$ .

- **MM.** On a :

$$\sqrt{N}(\hat{\mathbf{a}}_N^{MCO} - \mathbf{a}_0) \xrightarrow[N \rightarrow +\infty]{L} \mathcal{N}(\mathbf{0}; \Sigma_0) \text{ avec } \Sigma_0 \equiv (\mathbf{G}_0)^{-1} \mathbf{W}_0 (\mathbf{G}_0')^{-1}$$

avec :

$$\mathbf{W}_0 \equiv E[\mathbf{g}_i(\mathbf{a}_0) \mathbf{g}_i(\mathbf{a}_0)'] = \omega_0 E\left[\frac{\partial f(\mathbf{x}_i; \mathbf{a}_0)}{\partial \mathbf{a}} \frac{\partial f(\mathbf{x}_i; \mathbf{a}_0)}{\partial \mathbf{a}'}\right]$$

$$\mathbf{G}_0 \equiv E\left[\frac{\partial \mathbf{g}_i(\mathbf{a}_0)'}{\partial \mathbf{a}}\right] = -E\left[\frac{\partial f(\mathbf{x}_i; \mathbf{a}_0)}{\partial \mathbf{a}} \frac{\partial f(\mathbf{x}_i; \mathbf{a}_0)}{\partial \mathbf{a}'}\right]$$

et donc :

$$\Sigma_0 = \omega_0 E\left[\frac{\partial f(\mathbf{x}_i; \mathbf{a}_0)}{\partial \mathbf{a}} \frac{\partial f(\mathbf{x}_i; \mathbf{a}_0)}{\partial \mathbf{a}'}\right]^{-1} \text{ avec } \omega_0 = E\left[\left(y_i - f(\mathbf{x}_i; \mathbf{a}_0)\right)^2\right].$$

- On peut décliner tous les cas d'hétéroscédasticité de la même manière que dans le cas linéaire.
  - Estimateur des MC non-linéaires Généralisé
  - Estimateur des MC non-linéaires Quasi-Généralisé
  - Estimateur des MCO non-linéaires avec matrice de variance-covariance as. robuste à l'hétéroscédasticité

$$\sqrt{N}(\hat{\mathbf{a}}_N^{MCO} - \mathbf{a}_0) \xrightarrow[N \rightarrow +\infty]{L} \mathcal{N}(\mathbf{0}; \Sigma_0) \text{ avec } \Sigma_0 \equiv (\mathbf{G}_0)^{-1} \mathbf{W}_0 (\mathbf{G}_0')^{-1}$$

avec :

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

et :

$$\mathbf{G}_0 \equiv E\left[ \frac{\partial \mathbf{g}_i(\mathbf{a}_0)'}{\partial \mathbf{a}} \right] = -E\left[ \frac{\partial f(\mathbf{x}_i; \mathbf{a}_0)}{\partial \mathbf{a}} \frac{\partial f(\mathbf{x}_i; \mathbf{a}_0)}{\partial \mathbf{a}'} \right].$$



### Exemple 3. *Modèle linéaire à VI*

- Un modèle linéaire à VI est de la forme :

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

- On a alors :

$$E[y_i - \mathbf{a}'_0 \mathbf{x}_i / \mathbf{z}_i] = 0$$

avec :

$$u_i(\mathbf{a}) \equiv y_i - \mathbf{a}' \mathbf{x}_i \quad \text{et} \quad u_i = y_i - \mathbf{a}'_0 \mathbf{x}_i = u_i(\mathbf{a}_0)$$

- On a :

$$E[u_i(\mathbf{a}_0) / \mathbf{z}_i] = E[u_i / \mathbf{z}_i] = 0$$

- La forme de l'instrument optimal est donnée par :

$$\mathbf{r}_i^+ \equiv E \left[ \frac{\partial u_i(\mathbf{a}_0)}{\partial \mathbf{a}} / \mathbf{z}_i \right] E[u_i(\mathbf{a}_0)^2 / \mathbf{z}_i]^{-1} = -E[\mathbf{x}_i / \mathbf{z}_i] E[u_i^2 / \mathbf{z}_i]^{-1}$$

- Avec :

$$\mathbf{r}_i^+ = -E[\mathbf{x}_i/\mathbf{z}_i] E[u_i^2/\mathbf{z}_i]^{-1}$$

on a deux problèmes pratiques :

- l'hétéroscédasticité potentielle des  $u_i$ , ce qui se traite comme ci-dessus selon les cas rencontrés
- la forme de  $E[\mathbf{x}_i/\mathbf{z}_i]$  qui est inconnue
- Ici, présentation en deux étapes :
  - *D'abord* le problème de  $E[\mathbf{x}_i/\mathbf{z}_i]$ , en prenant le cas  $E[u_i^2/\mathbf{z}_i] = \omega_0$
  - *Ensuite* le problème de  $E[u_i^2/\mathbf{z}_i] \neq \omega_0$
- Assez grand *décalage entre la « pratique »* et la « *théorie* », en tous cas pour ce qui concerne la recherche de l'efficacité d'estimation.

### 3a. Modèle linéaire à VI avec homoscedasticité

- Les *instruments optimaux* d'un modèle linéaire à VI de la forme :

$$y_i = \mathbf{a}'_0 \mathbf{x}_i + u_i \quad \text{avec} \quad E[u_i / \mathbf{z}_i] = E[u_i] \equiv 0 \quad \text{et} \quad E[u_i^2 / \mathbf{z}_i] = \omega_0$$

sont définis par :

$$\mathbf{r}_i^+ = -E[\mathbf{x}_i / \mathbf{z}_i] \omega_0^{-1}.$$

- Ils donnent des *conditions de moment* (juste-)identifiantes de la forme :

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

- *Problème* : l'espérance conditionnelle  $E[\mathbf{x}_i / \mathbf{z}_i]$  est inconnue *a priori*.
- *Deux approches potentielles* : l'une très utilisée, l'autre quasiment pas

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

▪ **Première approche :**

- On calcule des estimateurs non paramétriques des  $E\left[\mathbf{x}_i/\mathbf{z}_i\right]$ , les  $\tilde{\mathbf{x}}_N(\mathbf{z}_i)$ , et on résout :

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

ce qui donne :

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

- ***Chamberlain + instruments estimés*** : l'estimateur obtenu est as. efficace
- ***Estimation non-paramétrique*** : difficile à mettre en œuvre

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

▪ **Seconde approche :**

- On n'utilise pas  $E[\mathbf{x}_i/\mathbf{z}_i]$ , mais :

$$EL[\mathbf{x}_i/\mathbf{z}_i] = E[\mathbf{x}_i\mathbf{z}_i']E[\mathbf{z}_i\mathbf{z}_i']^{-1}\mathbf{z}_i,$$

*i.e.* pas l'espérance conditionnelle de  $\mathbf{x}_i$  en  $\mathbf{z}_i$  mais la **projection linéaire** de  $\mathbf{x}_i$  sur  $\mathbf{z}_i$ .

▪ **Idées sous-jacentes**

- $E[\mathbf{x}_i/\mathbf{z}_i]$  est le meilleur prédicteur (EQM) de  $\mathbf{x}_i$  par  $\mathbf{z}_i$ ,  
 $EL[\mathbf{x}_i/\mathbf{z}_i]$  est le meilleur prédicteur (EQM) linéaire de  $\mathbf{x}_i$  par  $\mathbf{z}_i$  et c'est pas mal quand-même.
- Utiliser  $EL[\mathbf{x}_i/\mathbf{z}_i]$  nous ramène sur des « terrains connus ».

- Deux interprétations « techniques » l'estimateur MM de  $\mathbf{a}_0$  exploitant la condition de moment :

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

- *Estimateur de la MM avec instrument estimé :*

$$N^{-1} \sum_{i=1}^N \underbrace{\left[ N^{-1} \sum_{i=1}^N \mathbf{x}_i \mathbf{z}'_i \left( N^{-1} \sum_{i=1}^N \mathbf{x}_i \mathbf{z}'_i \right)^{-1} \mathbf{z}_i \right]}_{\text{Estimateur convergent de } EL[\mathbf{x}_i/\mathbf{z}_i]} (y_i - \mathbf{x}'_i \hat{\mathbf{a}}_N^{MM}) = \mathbf{0}_{K \times 1}$$

- *Estimateur de la MMG :*

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

- On obtient :

$\hat{\mathbf{a}}_N^{MM} = \hat{\mathbf{a}}_N^{MMG} = \hat{\mathbf{a}}_N^{2MC}$
---

▪ **Conclusions : estimateur  $\hat{\mathbf{a}}_N^{2MC}$  dans un modèle linéaire à VI**

- **Chamberlain** : L'estimateur  $\hat{\mathbf{a}}_N^{2MC}$  est as. efficace **si** les termes d'erreur du modèle sont homoscédastiques et **si** on a :

$$EL[\mathbf{x}_i/\mathbf{z}_i] = E[\mathbf{x}_i/\mathbf{z}_i]$$

- L'estimateur de la MMG fondé sur la condition d'orthogonalité :

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

calcule « *automatiquement* » un estimateur de  $EL[\mathbf{x}_i/\mathbf{z}_i]$  et fournit le même estimateur que celui de la MM fondé sur :

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

- Il se peut que  $EL[\mathbf{x}_i/\mathbf{t}(\mathbf{z}_i)] \approx E[\mathbf{x}_i/\mathbf{z}_i]$  pour un **choix judicieux** de la fonction  $\mathbf{t}(\cdot)$ , e.g. termes carrés, croisés, ... des VI.  
On a ici un gain d'efficacité « pas cher ».

### 3b. Modèle linéaire à VI avec hétéroscédasticité de forme inconnue

- Les *instruments optimaux* d'un modèle linéaire à VI de la forme :

$$y_i = \mathbf{a}'_0 \mathbf{x}_i + u_i \text{ avec } E[u_i/\mathbf{z}_i] = E[u_i] \equiv 0 \text{ et } E[u_i^2/\mathbf{z}_i] = \omega(\mathbf{z}_i)$$

sont définis par :

$$\mathbf{r}_i^* = -E[\mathbf{x}_i/\mathbf{z}_i] \omega(\mathbf{z}_i)^{-1}.$$

- **Problème** lorsque  $\omega(\cdot)$  est de forme inconnue
- **En pratique**: on remplace l'espérance conditionnelle par  $E[\mathbf{x}_i/\mathbf{z}_i]$  par des projections linéaires :

$$EL[\mathbf{x}_i/\mathbf{z}_i] \text{ ou } EL[\mathbf{x}_i/\mathbf{t}(\mathbf{z}_i)],$$

ce qui permet d'utiliser les 2MC (voir ci-avant) .



$$\mathbf{r}_i^* = -E[\mathbf{x}_i/\mathbf{z}_i] \omega(\mathbf{z}_i)^{-1} \text{ avec } \omega(.) \text{ de forme inconnue}$$

- **Solution** : On se place dans le cadre de la MMG et on travaille directement avec la condition de moment (généralement sur-identifiante) :

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

*en tenant compte de ce que :*

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

avec :

$$\mathbf{W}_0 \neq \omega_0 E[\mathbf{z}_i \mathbf{z}'_i] \text{ où } \omega_0 \equiv E[u_i^2] = E[(y_i - \mathbf{x}'_i \mathbf{a}_0)^2].$$

**Deux options** :  $\hat{\mathbf{a}}_N^{2MCH}$  ou  $\hat{\mathbf{a}}_N^{2MC}$

- $\hat{\mathbf{a}}_N^{2MCH}$ , *estimateur des 2MC robuste à l'hétéroscédasticité*, si beaucoup de données et de bons instruments.
  - $\hat{\mathbf{a}}_N^{2MCH}$  est l'estimateur MMG optimal fondé sur la condition de moment  $E[\mathbf{z}_i(y_i - \mathbf{x}'_i \mathbf{a}_0)] = \mathbf{0}_{L \times 1}$  sachant que  $\mathbf{W}_0 = E[(y_i - \mathbf{x}'_i \mathbf{a}_0)^2 \mathbf{z}_i \mathbf{z}'_i]$ .
  - Il se calcul en trois étapes et sa matrice de variance covariance as. est de la forme :

$$\Sigma_0 = \left( E[\mathbf{x}_i \mathbf{z}'_i] E[(y_i - \mathbf{x}'_i \mathbf{a}_0)^2 \mathbf{z}_i \mathbf{z}'_i]^{-1} E[\mathbf{z}_i \mathbf{x}'_i] \right)^{-1}.$$

- $\hat{\mathbf{a}}_N^{2MC}$ , *estimateur des 2MC* « usuel » avec calcul de sa variance as. en tenant compte de l'hétéroscédasticité des  $u_i$ .
  - $\hat{\mathbf{a}}_N^{2MC}$  se calcule directement en une étape.
  - $\hat{\mathbf{a}}_N^{2MC}$  n'est pas un estimateur MMG optimal. Il utilise  $\mathbf{M}_0 = E[\mathbf{z}_i \mathbf{z}_i']^{-1}$  pour matrice de pondération limite alors que  $\mathbf{W}_0 = E[(y_i - \mathbf{x}_i' \mathbf{a}_0)^2 \mathbf{z}_i \mathbf{z}_i']$ .
  - Sa matrice de variance as. est de forme « compliquée » :

$$\Sigma(\mathbf{M}_0) = \begin{bmatrix} \left( E[\mathbf{x}_i \mathbf{z}_i'] E[\mathbf{z}_i \mathbf{z}_i']^{-1} E[\mathbf{z}_i \mathbf{x}_i'] \right)^{-1} \\ \times E[\mathbf{x}_i \mathbf{z}_i'] E[\mathbf{z}_i \mathbf{z}_i']^{-1} E[(y_i - \mathbf{x}_i' \mathbf{a}_0)^2 \mathbf{z}_i \mathbf{z}_i'] E[\mathbf{z}_i \mathbf{z}_i']^{-1} E[\mathbf{z}_i \mathbf{x}_i'] \times \\ \left( E[\mathbf{x}_i \mathbf{z}_i'] E[\mathbf{z}_i \mathbf{z}_i']^{-1} E[\mathbf{z}_i \mathbf{x}_i'] \right)^{-1} \end{bmatrix}.$$

## Rappels sur les estimateurs des 2MC

- Les estimateurs des 2MC sont convergents ... mais **biaisés** à distance finie
  - ⇒ Utiliser avec de grands échantillons ... et veiller aux points aberrants
- Le biais de  $\hat{\mathbf{a}}_N^{2MC}$  croît avec le **degré de sur-identification** du modèle par les instruments utilisés
  - ⇒ Se concentrer sur des VI informatives, y.c. les transformations  $\mathbf{t}(\mathbf{z}_i)$ .
- Le **biais de  $\hat{\mathbf{a}}_N^{2MCH}$**  est très important à distance finie, créant de l'instabilité dans les estimations obtenues
  - ⇒ Utiliser avec de très grands échantillons ... et veiller aux points aberrants
- Une VI peu informative et « un peu endogène » peut créer une divergence des estimateurs 2MC (problème dit « des **instruments faibles** »)
  - ⇒ Se concentrer sur des VI informatives, y.c. les transformations  $\mathbf{t}(\mathbf{z}_i)$ .

#### 4. Instruments efficaces, cas multivarié

## 4. Instruments efficaces, cas multivarié

### 4.1. Motivations : rappels

#### Problème général

Comment exploiter une condition de moment conditionnel :

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

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

- *Cas classique*, un système d'équations à termes d'erreur additifs :

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i; \mathbf{a}_0) + \mathbf{u}_i \Rightarrow \mathbf{u}_i(\mathbf{a}) \equiv \mathbf{u}_i(\mathbf{y}_i, \mathbf{x}_i; \mathbf{a}) \equiv \mathbf{y}_i - \mathbf{f}(\mathbf{x}_i; \mathbf{a}) \text{ et } \mathbf{u}_i(\mathbf{a}_0) = \mathbf{u}_i$$

avec

$$E[\mathbf{u}_i/\mathbf{z}_i] \Rightarrow E[\mathbf{u}_i(\mathbf{a}_0)/\mathbf{z}_i] = \mathbf{0}.$$

### Cadre d'analyse

On utilise le cadre de la MMG et on utilise des instruments  $\mathbf{R}(\mathbf{z}_i)$  tels que:

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

- On utilise ici une matrice d'instruments. Si  $\mathbf{u}_i(\mathbf{a})$  est de dimension  $M \times 1$  alors  $\mathbf{R}(\mathbf{z}_i)$  est de dimension  $G \times M$ ,  $\mathbf{g}_i(\mathbf{a})$  étant de dimension  $G \times 1$ .
  - Bien entendu, la condition d'ordre d'indentification de  $\mathbf{a}_0$  implique qu'on doit avoir  $G \geq K \equiv \dim \mathbf{a}_0$ .
- On utilisera ici les notations :

$$\mathbf{g}_i(\mathbf{a}) \equiv \begin{bmatrix} g_{1,i}(\mathbf{a}) \\ \vdots \\ g_{G,i}(\mathbf{a}) \end{bmatrix}, \mathbf{u}_i(\mathbf{a}) \equiv \begin{bmatrix} u_{1,i}(\mathbf{a}) \\ \vdots \\ u_{M,i}(\mathbf{a}) \end{bmatrix} \text{ et } \mathbf{R}(\mathbf{z}_i) \equiv \begin{bmatrix} r_{11}(\mathbf{z}_i) & \cdots & r_{1M}(\mathbf{z}_i) \\ \vdots & \ddots & \vdots \\ r_{G1}(\mathbf{z}_i) & \cdots & r_{GM}(\mathbf{z}_i) \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1(\mathbf{z}_i)' \\ \vdots \\ \mathbf{r}_G(\mathbf{z}_i)' \end{bmatrix}$$

- L'hypothèse  $E[\mathbf{u}_i(\mathbf{a}_0)/\mathbf{z}_i] = \mathbf{0}$  implique que :

$$E[r_{gm}(\mathbf{z}_i) \times u_{m,i}(\mathbf{a}_0)] = 0 \text{ pour tout } m = 1, \dots, M \text{ et tout } g = 1, \dots, G.$$

- La condition de moment utilisant la matrice d'instruments  $\mathbf{R}(\mathbf{z}_i)$  combine linéairement les conditions d'orthogonalité précédentes :

$$E[\mathbf{g}_i(\mathbf{a})] = E[\mathbf{R}(\mathbf{z}_i)\mathbf{u}_i(\mathbf{a})] = \begin{bmatrix} \sum_{m=1}^M E[r_{1m,i}(\mathbf{a}) \times u_{m,i}(\mathbf{a})] \\ \vdots \\ \sum_{m=1}^M E[r_{Gm,i}(\mathbf{a}) \times u_{m,i}(\mathbf{a})] \end{bmatrix} = \begin{bmatrix} E[\mathbf{r}_1(\mathbf{z}_i)' \mathbf{u}_i(\mathbf{a})] \\ \vdots \\ E[\mathbf{r}_G(\mathbf{z}_i)' \mathbf{u}_i(\mathbf{a})] \end{bmatrix}$$

avec :

$$E[g_{\ell,i}(\mathbf{a})] = E[\mathbf{r}_\ell(\mathbf{z}_i)' \mathbf{u}_i(\mathbf{a})] = \sum_{m=1}^M E[r_{\ell m,i}(\mathbf{a}) \times u_{m,i}(\mathbf{a})] \text{ pour } g = 1, \dots, G.$$



- **Cas particulier fréquent.** Si  $\mathbf{R}(\mathbf{z}_i)$  est bloc-diagonale, *i.e.* de la forme :

$$\mathbf{R}(\mathbf{z}_i) \equiv \begin{bmatrix} \mathbf{r}_1(\mathbf{z}_i) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{r}_2(\mathbf{z}_i) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{r}_M(\mathbf{z}_i) \end{bmatrix}$$

alors on a :

$$E[\mathbf{g}_i(\mathbf{a})] = E[\mathbf{R}(\mathbf{z}_i)\mathbf{u}_i(\mathbf{a})] = \begin{bmatrix} E[\mathbf{r}_1(\mathbf{z}_i)u_{1,i}(\mathbf{a})] \\ \vdots \\ E[\mathbf{r}_M(\mathbf{z}_i)u_{M,i}(\mathbf{a})] \end{bmatrix},$$

*i.e.* chaque terme résiduel  $u_{m,i}(\mathbf{a})$  a « son » vecteur d'instruments  $\mathbf{r}_m(\mathbf{z}_i)$ .

L'instrument optimal n'a en général pas cette forme.

### Recherche de l'efficacité d'estimation

On cherche à déterminer l'instrument  $\mathbf{R}^+(\mathbf{z}_i)$  tel que l'estimateur MMG de  $\mathbf{a}_0$  fondé sur la condition de moment :

$$E[\mathbf{g}_i^+(\mathbf{a})] = E[\mathbf{R}^+(\mathbf{z}_i)\mathbf{u}_i(\mathbf{a})] = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{a} = \mathbf{a}_0$$

est *efficace dans la classe des estimateurs MMG fondés sur une condition d'orthogonalité construite avec  $\mathbf{u}_i(\mathbf{a})$* .

- Estimateurs MMG optimaux, en principe ...
  - ... on ne se servira que d'estimateurs **MM** car juste-identification.
- L'efficacité est définie par rapport une classe d'estimateurs MM(G) ...
  - ... mais le résultat de Chamberlain est bien plus général.
- L'« instrument efficace » ... n'est pas unique.

*L'efficacité sera toujours relative aux  $\mathbf{VI}$ ,  $\mathbf{z}_i$ , utilisées comme ensemble d'information du modèle considéré.*

### ***Remarque importante***

La condition de moment conditionnel considérée ici est assez particulière. En effet :

$$E[\mathbf{u}_i(\mathbf{a}_0)/\mathbf{z}_i] = \mathbf{0} \Leftrightarrow E[u_{m,i}(\mathbf{a}_0)/\mathbf{z}_i] = 0 \text{ pour } m = 1, \dots, M.$$

Ici  $\mathbf{z}_i$  sert de VI pour tout élément  $u_{m,i}(\mathbf{a}_0)$  de  $\mathbf{u}_i(\mathbf{a}_0)$ . Ce cas est fréquent.

Mais on rencontre des conditions de moment conditionnel plus générales que celle considérée ici (*e.g.*, cas des modèles avec données de panel), *i.e.* :

$$\begin{bmatrix} E[u_{1,i}(\mathbf{a}_0)/\mathbf{z}_{1,i}] \\ \vdots \\ E[u_{M,i}(\mathbf{a}_0)/\mathbf{z}_{M,i}] \end{bmatrix} = \mathbf{0} \Leftrightarrow \left( E[u_{m,i}(\mathbf{a}_0)/\mathbf{z}_{m,i}]; m = 1, \dots, M \right) = \mathbf{0}.$$

Chaque terme d'erreur  $u_{m,i}(\mathbf{a}_0)$  a « son » vecteur de VI  $\mathbf{z}_{m,i}$ .

La forme des instruments efficaces n'est pas connue pour ces modèles.

## 4.2. Les instruments efficaces pour $E[\mathbf{u}_i(\mathbf{a}_0)/\mathbf{z}_i] = \mathbf{0}$

### Propriété. *Instruments efficaces, cas multivarié*

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

$$(i) \quad E[\mathbf{u}(y_i, \mathbf{x}_i; \mathbf{a}_0)/\mathbf{z}_i] = E[\mathbf{u}_i(\mathbf{a}_0)/\mathbf{z}_i] = \mathbf{0}_{M \times 1}$$

$$(iii) \quad E[\mathbf{u}_i(\mathbf{a}_0)\mathbf{u}_i(\mathbf{a}_0)'/\mathbf{z}_i] \text{ est inversible}$$

$$(iii) \quad E\left[\frac{\partial \mathbf{u}_i(\mathbf{a}_0)'}{\partial \mathbf{a}} \frac{\partial \mathbf{u}_i(\mathbf{a}_0)}{\partial \mathbf{a}'} / \mathbf{z}_i\right] \text{ est de rang } K \equiv \dim \mathbf{a},$$

alors l'estimateur de la MM de  $\mathbf{a}_0$  fondé sur la condition de moment :

$$E\left[E\left[\frac{\partial \mathbf{u}_i(\mathbf{a}_0)'}{\partial \mathbf{a}} / \mathbf{z}_i\right] E[\mathbf{u}_i(\mathbf{a}_0)\mathbf{u}_i(\mathbf{a}_0)'/\mathbf{z}_i]^{-1} \mathbf{u}_i(\mathbf{a}_0)\right] = \mathbf{0}_{K \times 1}$$

existe avec une probabilité approchant 1 et est as. efficace dans la classe des estimateurs as. normaux de  $\mathbf{a}_0$  sous les conditions de régularité usuelles.

## Remarques générales

- Le terme :

$$\begin{aligned}\mathbf{R}^+(\mathbf{z}_i) &\equiv E\left[\frac{\partial \mathbf{u}_i(\mathbf{a}_0)'}{\partial \mathbf{a}} / \mathbf{z}_i\right] E[\mathbf{u}_i(\mathbf{a}_0)\mathbf{u}_i(\mathbf{a}_0)' / \mathbf{z}_i]^{-1} \\ &= E\left[\frac{\partial \mathbf{u}_i(\mathbf{a}_0)'}{\partial \mathbf{a}} / \mathbf{z}_i\right] V[\mathbf{u}_i(\mathbf{a}_0) / \mathbf{z}_i]^{-1}\end{aligned}$$

est désigné sous le terme d'*instrument efficace* (ou *optimal*).

- $\mathbf{R}^+(\mathbf{z}_i)$  est exogène car c'est une fonction des variables exogènes  $\mathbf{z}_i$
- La condition  $E[\mathbf{R}^+(\mathbf{z}_i)\mathbf{u}_i(\mathbf{a}_0)] = \mathbf{0}$  juste-identifie  $\mathbf{a}_0$ , on utilise la MM.
- L'estimateur MM de  $\mathbf{a}_0$  défini à avec la condition  $E[\mathbf{R}^+(\mathbf{z}_i)\mathbf{u}_i(\mathbf{a}_0)] = \mathbf{0}$  est l'estimateur as. normal le plus efficace de  $\mathbf{a}_0$ .

**Définition. Borne d'efficacité semi-paramétrique**

$$\sqrt{N}(\hat{\mathbf{a}}_N^{MM} - \mathbf{a}_0) \xrightarrow[N \rightarrow +\infty]{L} \mathcal{N}(\mathbf{0}; \Sigma_0^*)$$

où :

$$\Sigma_0^* \equiv E \left[ \frac{\partial \mathbf{u}_i(\mathbf{a}_0)'}{\partial \mathbf{a}} E[\mathbf{u}_i(\mathbf{a}_0) \mathbf{u}_i(\mathbf{a}_0)' / \mathbf{z}_i]^{-1} \frac{\partial \mathbf{u}_i(\mathbf{a}_0)}{\partial \mathbf{a}'} \right]^{-1}.$$

$\Sigma_0^*$  est la *borne minimax* ou *borne d'efficacité semi-paramétrique* (des estimateurs as. normaux de  $\mathbf{a}_0$ ).

**Propriété. Transformation affine non singulière des instruments efficaces**

Si  $\mathbf{B}$  est une matrice carrée de dimension  $K$  et de rang  $K$ , alors  $\mathbf{B}\mathbf{R}^+(\mathbf{z}_i)$  est également un instrument efficace.

- La démonstration de ces propriétés est ardue et peu instructive. Ceci dit, la forme de  $\mathbf{R}^+(\mathbf{z}_i)$  est relativement intuitive.
- Dans la suite on notera :

$$\mathbf{\Gamma}_0(\mathbf{z}_i) \equiv E \left[ \frac{\partial \mathbf{u}_i(\mathbf{a}_0)'}{\partial \mathbf{a}} / \mathbf{z}_i \right]$$

et :

$$\mathbf{\Omega}_0(\mathbf{z}_i) \equiv E [\mathbf{u}_i(\mathbf{a}_0) \mathbf{u}_i(\mathbf{a}_0)' / \mathbf{z}_i] = V [\mathbf{u}_i(\mathbf{a}_0) / \mathbf{z}_i]$$

et finalement :

$$\mathbf{R}^+(\mathbf{z}_i) = \mathbf{\Gamma}_0(\mathbf{z}_i) \mathbf{\Omega}_0(\mathbf{z}_i)^{-1}.$$

**Interprétation,  $\Omega_0(\mathbf{z}_i) \equiv E[\mathbf{u}_i(\mathbf{a}_0)\mathbf{u}_i(\mathbf{a}_0)'/\mathbf{z}_i] = V[\mathbf{u}_i(\mathbf{a}_0)/\mathbf{z}_i]$**

$$\mathbf{g}_i^+(\mathbf{a}) = \underbrace{\Gamma_0(\mathbf{z}_i)\Omega_0(\mathbf{z}_i)^{-1}}_{\mathbf{R}^+(\mathbf{z}_i)}\mathbf{u}_i(\mathbf{a})$$

- $\Omega_0(\mathbf{z}_i)$  est la matrice de variance-covariance conditionnelle en  $\mathbf{z}_i$  de  $\mathbf{u}_i = \mathbf{u}_i(\mathbf{a}_0)$ 
  - Si  $\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i; \mathbf{a}_0) + \mathbf{u}_i$  avec  $E[\mathbf{u}_i/\mathbf{z}_i] = E[\mathbf{u}_i] \equiv \mathbf{0}$  alors  $\mathbf{u}_i(\mathbf{a}_0) = \mathbf{u}_i$ .
  - $\Omega_0(\mathbf{z}_i)$  tient compte de l'éventuelle *hétéroscédasticité conditionnelle* et des *relations entre les éléments* de  $\mathbf{u}_i = \mathbf{u}_i(\mathbf{a}_0)$ .
    - Correction pour l'*hétéroscédasticité*, e.g. des *MCG* ou des *MCQG*.
    - Contrôle pour les *relations entre les termes d'erreur*, e.g. des *MCG* ou des *MCQG*.



$$\text{Interprétation de } \Gamma_0(\mathbf{z}_i) \equiv E \left[ \frac{\partial \mathbf{u}_i(\mathbf{a}_0)'}{\partial \mathbf{a}} / \mathbf{z}_i \right]$$

$$\mathbf{g}_i^+(\mathbf{a}) = \underbrace{\Gamma_0(\mathbf{z}_i)\mathbf{\Omega}_0(\mathbf{z}_i)^{-1}}_{\mathbf{R}^+(\mathbf{z}_i)} \mathbf{u}_i(\mathbf{a})$$

- Le terme  $\mathbf{\Omega}_0(\mathbf{z}_i)^{-1}$  joue essentiellement un rôle de *pondération*.

Le terme  $\Gamma_0(\mathbf{z}_i)$  porte la « *structure* » de  $\mathbf{R}^+(\mathbf{z}_i) = \Gamma_0(\mathbf{z}_i)\mathbf{\Omega}_0(\mathbf{z}_i)^{-1}$ , *il est essentiel pour l'identification de  $\mathbf{a}_0$* .

- $\Gamma_0(\mathbf{z}_i)$  est un bon instrument car c'est la *fonction de  $\mathbf{z}_i$  qui a la plus « forte » corrélation avec  $\mathbf{u}_i(\mathbf{a})$*  pour  $\mathbf{a}$  autour de  $\mathbf{a}_0$
- On peut donner l'intuition de l'« optimalité » de  $\Gamma_0(\mathbf{z}_i)$  en tant qu'instrument en linéarisant la condition d'identification de  $\mathbf{a}_0$ .

- On doit choisir un instrument  $\mathbf{R}(\mathbf{z}_i)$  pour  $\mathbf{u}_i(\mathbf{a}_0)$  de telle sorte à ce que l'estimateur de la MM de  $\mathbf{a}_0$  fondé sur :

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

(i) identifie  $\mathbf{a}_0$

et :

(ii) soit le plus efficace possible.

- Pour l'identification il faut que  $E[\mathbf{R}(\mathbf{z}_i)\mathbf{u}_i(\mathbf{a})] \neq \mathbf{0}$  si  $\mathbf{a} \neq \mathbf{a}_0$ , donc il faut que  $\mathbf{R}(\mathbf{z}_i)$  soit *corrélé au mieux* à  $\mathbf{u}_i(\mathbf{a})$  lorsque  $\mathbf{a} \neq \mathbf{a}_0$ , et ce même lorsque  $\mathbf{a}$  est très proche de  $\mathbf{a}_0$ .
  - *Intuition* :  $\mathbf{R}(\mathbf{z}_i)$  doit permettre de « bien » détecter si  $\mathbf{a} \neq \mathbf{a}_0$ .
  - *Intuition* : plus  $\mathbf{R}(\mathbf{z}_i)$  est corrélé à  $\mathbf{u}_i(\mathbf{a})$  pour  $\mathbf{a} \neq \mathbf{a}_0$ , plus  $\mathbf{R}(\mathbf{z}_i)$  est efficace en tant qu'instrument d'identification.

- Développement limité au premier ordre en  $\mathbf{a}$  de  $\mathbf{u}_i(\mathbf{a})$  autour de  $\mathbf{a}_0$  :

$$\mathbf{u}_i(\mathbf{a}) \simeq \mathbf{u}_i(\mathbf{a}_0) + \frac{\partial \mathbf{u}_i(\mathbf{a}_0)}{\partial \mathbf{a}'} (\mathbf{a} - \mathbf{a}_0)$$

or  $E[\mathbf{u}_i(\mathbf{a}_0)/\mathbf{z}_i] = 0$  implique que  $E[\mathbf{R}(\mathbf{z}_i)\mathbf{u}_i(\mathbf{a}_0)] = \mathbf{0}$  et donc que :

$$E[\mathbf{R}(\mathbf{z}_i)\mathbf{u}_i(\mathbf{a}_0)] \simeq E\left[\mathbf{R}(\mathbf{z}_i)\frac{\partial \mathbf{u}_i(\mathbf{a}_0)}{\partial \mathbf{a}'}\right](\mathbf{a} - \mathbf{a}_0)$$

- On a :

$$\Gamma_0(\mathbf{z}_i) \equiv E\left[\frac{\partial \mathbf{u}_i(\mathbf{a}_0)'}{\partial \mathbf{a}} / \mathbf{z}_i\right] \Rightarrow E\left[\mathbf{R}(\mathbf{z}_i)\frac{\partial \mathbf{u}_i(\mathbf{a}_0)}{\partial \mathbf{a}'}\right] = E[\mathbf{R}(\mathbf{z}_i)\Gamma_0(\mathbf{z}_i)'].$$

- Avec :

$$E[\mathbf{R}(\mathbf{z}_i)\mathbf{u}_i(\mathbf{a}_0)] \simeq E[\mathbf{R}(\mathbf{z}_i)\mathbf{\Gamma}_0(\mathbf{z}_i)'](\mathbf{a} - \mathbf{a}_0)$$

on voit que :

$$\text{un des « meilleurs » candidats pour } \mathbf{R}^+(\mathbf{z}_i) \text{ est } \mathbf{\Gamma}_0(\mathbf{z}_i) \equiv E\left[\frac{\partial \mathbf{u}_i(\mathbf{a}_0)'}{\partial \mathbf{a}} \middle/ \mathbf{z}_i\right].$$

- En effet si  $E[\mathbf{\Gamma}_0(\mathbf{z}_i)\mathbf{\Gamma}_0(\mathbf{z}_i)']$  est inversible alors :

$$E[\mathbf{R}^+(\mathbf{z}_i)\mathbf{\Gamma}_0(\mathbf{z}_i)'](\mathbf{a} - \mathbf{a}_0) = E[\mathbf{\Gamma}_0(\mathbf{z}_i)\mathbf{\Gamma}_0(\mathbf{z}_i)'](\mathbf{a} - \mathbf{a}_0) = \mathbf{0} \Leftrightarrow \mathbf{a} = \mathbf{a}_0.$$

- La condition d'inversibilité de  $E[\Gamma_0(\mathbf{z}_i)\Gamma_0(\mathbf{z}_i)']$  (condition (iii) de la propriété) est une condition de rang locale, *i.e.* une ***condition locale d'identification***.
  - C'est la condition locale de rang, ***condition locale d'identification***, sur le terme  $\mathbf{G}_0$  dans la MM(G).
  - Elle est garantie si la fonction critère de la MMG n'est pas plate en  $\mathbf{a}$  autour de  $\mathbf{a}_0$ .
- La condition d'inversibilité de  $\mathbf{\Omega}_0(\mathbf{z}_i) \equiv E[\mathbf{u}_i(\mathbf{a}_0)\mathbf{u}_i(\mathbf{a}_0)'/\mathbf{z}_i] = V[\mathbf{u}_i(\mathbf{a}_0)/\mathbf{z}_i]$  (condition (ii) de la propriété) est une condition locale d'ordre.
  - Elle est garantie si les  $\mathbf{u}_i(\mathbf{a})$  sont linéairement indépendant.
  - Il faut faire attention lorsqu'on travaille avec des systèmes de parts budgétaires qui somment à 1 par exemple.

## Remarques générales

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

$$E[\mathbf{R}(z_i)\mathbf{u}(y_i, \mathbf{x}_i; \mathbf{a})] = \mathbf{0} \Leftrightarrow \mathbf{a} = \mathbf{a}_0$$

surtout pour :

- les *systèmes d'équations* (implicites ou explicites)
- les *modèles utilisés avec des données de panel* en particulier
- Le résultat de Chamberlain permet une ré-interprétation des propriétés des estimateurs usuels tels que les estimateurs SUR, des 3MC, GIV, ...

### 4.3. Un exemple : *Systèmes de régressions empilées (ou système SUR)*

- On considère des systèmes à  $M$  équations de la forme :

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

- Ces systèmes peuvent être également notés :

$$\begin{cases} y_{m,i} = \mathbf{a}'_{m,0} \mathbf{x}_{m,i} + u_{m,i} & \text{et } E[u_{m,i} / \mathbf{x}_i] = 0 \\ \text{pour } m = 1, \dots, M \end{cases}$$

- Le vecteur  $\mathbf{x}_i$  contient les éléments non redondants des  $\mathbf{x}_{m,i}$ 
  - Chaque équation du système est une équation de régression
  - $\mathbf{x}_i$  exogène par rapport à tout  $u_{m,i}$ , plus fort que  $E[u_{m,i} / \mathbf{x}_{m,i}] = 0$

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

- On suppose ici que les vecteurs de paramètres des équations, les  $\mathbf{a}_{m,0}$ , n'ont aucun lien entre eux:
  - *Seemingly Unrelated Regression System* (Système SUR), Zellner
  - Le vecteur des paramètres d'intérêt est l'empilement des  $\mathbf{a}_{m,0}$  :

$$\mathbf{a}_0 \equiv \begin{bmatrix} \mathbf{a}_{1,0} \\ \vdots \\ \mathbf{a}_{M,0} \end{bmatrix}.$$

- **Exemples :**
  - Systèmes de fonctions de demande de biens, consommation
  - Formes réduites de systèmes d'équations simultanées



- Ecriture matricielle :

*Système linéaire simple*

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

avec :

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

- Ce modèle se résume sous la forme :

$$E[\mathbf{u}_i(\mathbf{a}_0) / \mathbf{x}_i] = \mathbf{0}_{M \times 1} \quad \text{avec} \quad \mathbf{u}_i(\mathbf{a}) \equiv \mathbf{y}_i - \mathbf{X}_i' \mathbf{a}$$

$$E[\mathbf{u}_i(\mathbf{a}_0)/\mathbf{x}_i] = \mathbf{0}_{M \times 1} \text{ avec } \mathbf{u}_i(\mathbf{a}) \equiv \mathbf{y}_i - \mathbf{X}'_i \mathbf{a}$$

- *Instrument optimal* pour ces moments conditionnels :

$$\mathbf{R}_i^+ \equiv E\left[\frac{\partial \mathbf{u}_i(\mathbf{a}_0)'}{\partial \mathbf{a}} / \mathbf{x}_i\right] E[\mathbf{u}_i(\mathbf{a}_0) \mathbf{u}_i(\mathbf{a}_0)' / \mathbf{x}_i]^{-1} = -\mathbf{X}_i E[\mathbf{u}_i \mathbf{u}_i' / \mathbf{x}_i]^{-1}$$

- Problème (technique) si  $E[\mathbf{u}_i \mathbf{u}_i' / \mathbf{x}_i]$  dépend de  $\mathbf{x}_i$ , *i.e.* si hétéroscédasticité. On suppose, pour simplifier, ici que :

$$E[\mathbf{u}_i \mathbf{u}_i' / \mathbf{x}_i] = \mathbf{\Omega}_0 \equiv E[\mathbf{u}_i \mathbf{u}_i'], \text{ i.e. homoscélasticité des } \mathbf{u}_i.$$

- Estimation efficace = estimation « en système »
  - Si  $\mathbf{\Omega}_0 \equiv E[\mathbf{u}_i \mathbf{u}_i']$  n'est pas diagonale, *i.e.* pas de relations entre termes d'erreur des équations du système (*truly unrelated*)

- Si  $\mathbf{\Omega}_0 \equiv E[\mathbf{u}_i \mathbf{u}_i']$  diagonale on a :

$$\mathbf{R}_i^+ = -\mathbf{X}_i' \mathbf{\Omega}_0^{-1} = \begin{bmatrix} \omega_{11,0}^{-1} \mathbf{x}_{1,i} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \omega_{22,0}^{-1} \mathbf{x}_{2,i} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \omega_{MM,0}^{-1} \mathbf{x}_{M,i} \end{bmatrix}$$

et la condition de moment estimante optimale s'écrit sous la forme :

$$E[\mathbf{g}_i^+(\mathbf{a}_0)] = E[\mathbf{R}_i^+ \mathbf{u}_i(\mathbf{a}_0)] = E \begin{bmatrix} \omega_{11,0}^{-1} \mathbf{x}_{1,i} (y_{1,i} - \mathbf{a}_{1,0}' \mathbf{x}_{1,i}) \\ \omega_{22,0}^{-1} \mathbf{x}_{2,i} (y_{2,i} - \mathbf{a}_{2,0}' \mathbf{x}_{2,i}) \\ \vdots \\ \omega_{MM,0}^{-1} \mathbf{x}_{M,i} (y_{M,i} - \mathbf{a}_{M,0}' \mathbf{x}_{M,i}) \end{bmatrix} = \mathbf{0},$$

ce qui est équivalent à :

$$E[\mathbf{x}_{m,i} (y_{m,i} - \mathbf{a}_{m,0}' \mathbf{x}_{m,i})] = \mathbf{0} \text{ pour } m = 1, \dots, M.$$

- **Conclusion** : Si  $\mathbf{\Omega}_0 \equiv E[\mathbf{u}_i \mathbf{u}_i']$  diagonale l'estimateur as. efficace de  $\mathbf{a}_0$  dans le système :

$$\begin{cases} y_{m,i} = \mathbf{a}'_{m,0} \mathbf{x}_{m,i} + u_{m,i} & \text{et } E[u_{m,i} / \mathbf{x}_i] = 0 \\ \text{pour } m = 1, \dots, M \end{cases}$$

est l'estimateur constitué par « l'empilement » des estimateurs MCO équation par équation.

En effet l'estimateur as. efficace de  $\mathbf{a}_0$  est l'estimateur MM fondé sur :

$$E[\mathbf{x}_{m,i} (y_{m,i} - \mathbf{a}'_{m,0} \mathbf{x}_{m,i})] = \mathbf{0} \quad \text{pour } m = 1, \dots, M$$

est défini par l'empilement des :

$$\hat{\mathbf{a}}_{m,N} \equiv \left( N^{-1} \sum_{i=1}^N \mathbf{x}_{m,i} \mathbf{x}_{m,i}' \right) N^{-1} \sum_{i=1}^N \mathbf{x}_{m,i} y_{m,i} \quad \text{pour } m = 1, \dots, M$$

- Cet estimateur est même sans biais ...
- ... mais on a rarement  $\mathbf{\Omega}_0 \equiv E[\mathbf{u}_i \mathbf{u}_i']$  diagonale.

- Dans le cas général,  $\mathbf{\Omega}_0 \equiv E[\mathbf{u}_i \mathbf{u}_i']$  *n'est pas diagonale* et la condition d'orthogonalité estimante fondée sur l'instrument optimal est donnée par :

$$E[\mathbf{R}^+(\mathbf{z}_i) \mathbf{u}_i(\mathbf{y}_i, \mathbf{x}_i; \mathbf{a}_0)] = E[\mathbf{X}_i \mathbf{\Omega}_0^{-1} (\mathbf{y}_i - \mathbf{X}_i' \mathbf{a}_0)] = \mathbf{0}$$

- Avec :

$$\mathbf{\Omega}_0^{-1} = \begin{bmatrix} v_{11,0} & \cdots & v_{1M,0} \\ \vdots & \ddots & \vdots \\ v_{M1,0} & \cdots & v_{MM,0} \end{bmatrix} \quad \text{où } v_{m\ell,0} = v_{\ell m,0}$$

on a :

$$\mathbf{X}_i \mathbf{\Omega}_0^{-1} (\mathbf{y}_i - \mathbf{X}_i' \mathbf{a}) = \begin{bmatrix} \sum_{m=1}^M \mathbf{x}_{1,i} v_{1m,0} (y_{m,i} - \mathbf{a}'_m \mathbf{x}_{m,i}) \\ \vdots \\ \sum_{m=1}^M \mathbf{x}_{M,i} v_{Mm,0} (y_{m,i} - \mathbf{a}'_m \mathbf{x}_{m,i}) \end{bmatrix} = \begin{bmatrix} \sum_{m=1}^M \mathbf{x}_{1,i} v_{1m,0} u_{m,i}(\mathbf{a}_m) \\ \vdots \\ \sum_{m=1}^M \mathbf{x}_{M,i} v_{Mm,0} u_{m,i}(\mathbf{a}_m) \end{bmatrix}$$

$$E\left[\mathbf{R}^+(\mathbf{z}_i)\mathbf{u}_i(\mathbf{y}_i, \mathbf{x}_i; \mathbf{a}_0)\right] = E\left[\mathbf{X}_i\mathbf{\Omega}_0^{-1}(\mathbf{y}_i - \mathbf{X}_i'\mathbf{a}_0)\right] = \mathbf{0}$$

- Bien entendu,  $\mathbf{\Omega}_0$  est inconnue *a priori*, mais on traitera ce problème dans un second temps.
- La condition  $E\left[\mathbf{X}_i\mathbf{\Omega}_0^{-1}(\mathbf{y}_i - \mathbf{X}_i'\mathbf{a}_0)\right] = \mathbf{0}$  est juste-identifiante. On peut donc utiliser la MM est définir un estimateur as. efficace de  $\mathbf{a}_0$  par :

$$\hat{\mathbf{a}}_N^{MM}(\mathbf{\Omega}_0) \text{ solution en } \mathbf{a} \text{ de } N^{-1} \sum_{i=1}^N \mathbf{X}_i \mathbf{\Omega}_0^{-1} (\mathbf{y}_i - \mathbf{X}_i' \mathbf{a}) = \mathbf{0},$$

ce qui donne :

$$\hat{\mathbf{a}}_N^{MM}(\mathbf{\Omega}_0) = \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i \mathbf{\Omega}_0^{-1} \mathbf{X}_i' \right)^{-1} N^{-1} \sum_{i=1}^N \mathbf{X}_i \mathbf{\Omega}_0^{-1} \mathbf{y}_i,$$

*i.e.* un estimateur de type MCG.

- L'estimateur  $\hat{\mathbf{a}}_N^{MM}(\mathbf{\Omega}_0)$  ne peut être utilisé en général car  $\mathbf{\Omega}_0$  est généralement inconnue.
- Ceci dit, puisque :

$$\mathbf{\Omega}_0 \equiv E[\mathbf{u}_i \mathbf{u}_i'] = E[(\mathbf{y}_i - \mathbf{X}_i' \mathbf{a}_0)(\mathbf{y}_i - \mathbf{X}_i' \mathbf{a}_0)']$$

on peut en construire un estimateur convergent :

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

si on dispose d'un estimateur as. normal (convergent suffit ici) de  $\mathbf{a}_0$ ,  $\tilde{\mathbf{a}}_N$ .

- Par la propriété des instruments estimés (et des MCQG) on sait que :

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

est as. équivalent à  $\hat{\mathbf{a}}_N^{MM}(\mathbf{\Omega}_0)$  qui est as. efficace. Aussi  $\hat{\mathbf{a}}_N^{MM}(\tilde{\mathbf{\Omega}}_N)$  est as. efficace pour l'estimation de  $\mathbf{a}_0$  à partir de  $E[\mathbf{y}_i - \mathbf{X}_i' \mathbf{a}_0 / \mathbf{x}_i] = \mathbf{0}$ .

- Ce résultat est une conséquence de ce que  $\mathbf{X}_i \tilde{\boldsymbol{\Omega}}_N^{-1}$  est un estimateur convergent de l'instrument efficace  $-\mathbf{R}_i^+ = -\mathbf{X}_i \boldsymbol{\Omega}_0^{-1}$ .
- Il reste maintenant à construire un estimateur as. normal de  $\mathbf{a}_0$ ,  $\tilde{\mathbf{a}}_N$ , pour pouvoir construire  $\tilde{\boldsymbol{\Omega}}_N$ .
- Si  $\mathbf{R}_i^+ = -\mathbf{X}_i \boldsymbol{\Omega}_0^{-1}$  est un instrument qui identifie  $\mathbf{a}_0$  alors l'instrument  $\mathbf{X}_i$  peut également identifier  $\mathbf{a}_0$ , *i.e.* :

$$E[\mathbf{X}_i \mathbf{u}_i] = E[\mathbf{X}_i (\mathbf{y}_i - \mathbf{X}_i' \mathbf{a}_0)] = \mathbf{0}.$$

Noter que choisir l'instrument  $\mathbf{X}_i$  revient à « remplacer »  $\boldsymbol{\Omega}_0$  par  $\mathbf{I}_M$ .

- L'estimateur de la MM fondé sur cette condition d'orthogonalité est l'estimateur :

$$\hat{\mathbf{a}}_N^{MM}(\mathbf{I}_M) = \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i \mathbf{X}_i' \right)^{-1} N^{-1} \sum_{i=1}^N \mathbf{X}_i \mathbf{y}_i$$

qui est à la structure d'estimateur des MCO.



- Un peu de calcul matriciel permet de montrer que :

$$\hat{\mathbf{a}}_N^{MM}(\mathbf{I}_M) = \left( N^{-1} \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i' \right)^{-1} N^{-1} \sum_{i=1}^N \mathbf{x}_i \mathbf{y}_i$$

s'écrit sous la forme :

$$\hat{\mathbf{a}}_N^{MM}(\mathbf{I}_M) = \begin{bmatrix} \left( N^{-1} \sum_{i=1}^N \mathbf{x}_{1,i} \mathbf{x}_{1,i}' \right)^{-1} N^{-1} \sum_{i=1}^N \mathbf{x}_{1,i} y_{1,i} \\ \vdots \\ \left( N^{-1} \sum_{i=1}^N \mathbf{x}_{M,i} \mathbf{x}_{M,i}' \right)^{-1} N^{-1} \sum_{i=1}^N \mathbf{x}_{M,i} y_{M,i} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{a}}_{1,N}^{MCO} \\ \vdots \\ \hat{\mathbf{a}}_{M,N}^{MCO} \end{bmatrix},$$

i.e. que  $\hat{\mathbf{a}}_N^{MM}(\mathbf{I}_M)$  se définit comme l'empilement des estimateurs des MCO des  $\mathbf{a}_{m,0}$ , les  $\hat{\mathbf{a}}_{m,N}^{MCO}$  pour  $m = 1, \dots, M$ .

- On retrouve ici l'*estimateur des MCO* « équation par équation » de  $\mathbf{a}_0$ . Cet estimateur est as. normal mais n'est pas as. efficace lorsque  $\mathbf{\Omega}_0$  n'est pas diagonale.

## Pour résumer

L'estimateur :

$$\hat{\mathbf{a}}_N^{MM}(\mathbf{\Omega}_0) = \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i \mathbf{\Omega}_0^{-1} \mathbf{X}_i' \right)^{-1} N^{-1} \sum_{i=1}^N \mathbf{X}_i \mathbf{\Omega}_0^{-1} \mathbf{y}_i$$

est un estimateur de as. de  $\mathbf{a}_0$  dans un système de régressions empilées :

$$\begin{cases} y_{m,i} = \mathbf{a}_{m,0}' \mathbf{x}_{m,i} + u_{m,i} & \text{et } E[u_{m,i} / \mathbf{x}_i] = 0 \\ \text{pour } m = 1, \dots, M \end{cases}$$

ou :

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

avec :

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

car il est construit à partir d'une condition d'orthogonalité optimale au sens de Chamberlain, *i.e.* utilisant l'instrument optimal  $\mathbf{R}_i^+ = -\mathbf{X}_i \mathbf{\Omega}_0^{-1}$ .

Cet estimateur ne peut être calculé si  $\mathbf{\Omega}_0$  est inconnue (cas général).

L'estimateur :

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

est as. équivalent à  $\hat{\mathbf{a}}_N^{MM}(\mathbf{\Omega}_0)$  car  $\mathbf{X}_i \tilde{\mathbf{\Omega}}_N^{-1}$  est un estimateur convergent de l'instrument efficace  $-\mathbf{R}_i^+ = -\mathbf{X}_i \mathbf{\Omega}_0^{-1}$ .

L'estimateur  $\hat{\mathbf{a}}_N^{MM}(\tilde{\mathbf{\Omega}}_N)$  peut être construit en trois étapes.

## Construction de $\hat{\mathbf{a}}_N^{MM}(\tilde{\mathbf{\Omega}}_N)$

**Etape 1.** Calcul de  $\hat{\mathbf{a}}_N^{MM}(\mathbf{I}_M) \equiv (\hat{\mathbf{a}}_{m,N}^{MCO}, m=1, \dots, M)$ :

$$\hat{\mathbf{a}}_N^{MM}(\mathbf{I}_M) = \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i \mathbf{X}_i' \right)^{-1} N^{-1} \sum_{i=1}^N \mathbf{X}_i \mathbf{y}_i \equiv \tilde{\mathbf{a}}_N$$

l'*estimateur des MCO* « *équation par équation* » de  $\mathbf{a}_0$ .

**Etape 2.** Calcul de :

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

**Etape 3.** Calcul de :

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

## Remarques

- L'estimateur  $\hat{\mathbf{a}}_N^{MM}(\tilde{\mathbf{\Omega}}_N)$  est plus connu sous le nom d'Estimateur « SUR » (*Seemingly Unrelated Regression Estimator*) après le travail de Zellner.
- De manière générale :
  - Un estimateur des MCQG, tel que  $\hat{\mathbf{a}}_N^{MM}(\tilde{\mathbf{\Omega}}_N) = \hat{\mathbf{a}}_N^{MCQG}$ , est désigné sous le terme de Feasible *Generalized Least Square Estimator*
  - Un estimateur des MCG, tel que  $\hat{\mathbf{a}}_N^{MM}(\mathbf{\Omega}_0) = \hat{\mathbf{a}}_N^{MCG}$  est désigné sous le terme de *Generalized Least Square Estimator* (il est *unfeasible* lorsque  $\mathbf{\Omega}_0$  est inconnue)

### **Propriété. Théorème de Zellner**

Soit un système de régressions empilées :

$$\begin{cases} y_{m,i} = \mathbf{a}'_{m,0} \mathbf{x}_{m,i} + u_{m,i} & \text{et } E[u_{m,i}/\mathbf{x}_i] = 0 \\ \text{pour } m = 1, \dots, M \end{cases}$$

ou :

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

avec :

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

(i)  $\hat{\mathbf{a}}_N^{MM}(\tilde{\mathbf{\Omega}}_N)$  et  $\hat{\mathbf{a}}_N^{MM}(\mathbf{I}_M) \equiv (\hat{\mathbf{a}}_{m,N}^{MCO}, m=1, \dots, M)$  sont as. équivalents si  $\mathbf{\Omega}_0$  est diagonale

(ii)  $\hat{\mathbf{a}}_N^{MM}(\tilde{\mathbf{\Omega}}_N)$  et  $\hat{\mathbf{a}}_N^{MM}(\mathbf{I}_M)$  sont égaux si  $\mathbf{x}_{m,i} = \mathbf{x}_i$  pour  $m=1, \dots, M$ .

## *Démonstration du Théorème de Zellner*

### Résultat (i)

On a vu précédemment que  $\hat{\mathbf{a}}_N^{MM}(\mathbf{I}_M) \equiv (\hat{\mathbf{a}}_{m,N}^{MCO}, m=1, \dots, M)$  est as. efficace si  $\mathbf{\Omega}_0$  est diagonale, tout comme  $\hat{\mathbf{a}}_N^{MM}(\tilde{\mathbf{\Omega}}_N)$  l'est parce qu'il est construit avec un estimateur convergent d'instrument optimal.

Les estimateurs  $\hat{\mathbf{a}}_N^{MM}(\tilde{\mathbf{\Omega}}_N)$  et  $\hat{\mathbf{a}}_N^{MM}(\mathbf{I}_M)$  étant as. efficaces, ils ont la même distribution as. et sont donc as. équivalents.

Les estimateurs  $\hat{\mathbf{a}}_N^{MM}(\tilde{\mathbf{\Omega}}_N)$  et  $\hat{\mathbf{a}}_N^{MM}(\mathbf{I}_M)$  ne sont pas égaux numériquement pour  $N < +\infty$ , on a « seulement » :

$$\tilde{\mathbf{\Omega}}_N \xrightarrow[N \rightarrow +\infty]{p} \mathbf{\Omega}_0 \text{ diagonale.}$$

### Résultat (ii)

Si  $\mathbf{x}_{m,i} = \mathbf{x}_i$  pour  $m = 1, \dots, M$  alors :

$$\mathbf{X}_i \equiv \begin{bmatrix} \mathbf{x}_i & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_i & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{x}_i \end{bmatrix} = \mathbf{I}_M \otimes \mathbf{x}_i$$

On va utiliser les propriétés suivantes des produits de Kronecker (ou tenseurs) :

**Coagulation** :  $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$  (si matrices conformes)

**Transposition** :  $(\mathbf{A} \otimes \mathbf{B})' = (\mathbf{A}') \otimes (\mathbf{B}')$

**Inversion** :  $(\mathbf{A} \otimes \mathbf{B})^{-1} = (\mathbf{A}^{-1}) \otimes (\mathbf{B}^{-1})$  (si matrices inversibles)

**Somme** :  $(\mathbf{A} \otimes \mathbf{B}) + (\mathbf{A} \otimes \mathbf{C}) = \mathbf{A} \otimes (\mathbf{B} + \mathbf{C})$  et  $(\mathbf{A} \otimes \mathbf{B}) + (\mathbf{C} \otimes \mathbf{B}) = (\mathbf{A} + \mathbf{C}) \otimes \mathbf{B}$



Avec  $\mathbf{X}_i = \mathbf{I}_M \otimes \mathbf{x}_i$  on a ici :

$$\begin{aligned}\hat{\mathbf{a}}_N^{MM}(\tilde{\mathbf{\Omega}}_N) &= \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i \tilde{\mathbf{\Omega}}_N^{-1} \mathbf{X}_i' \right)^{-1} N^{-1} \sum_{i=1}^N \mathbf{X}_i \tilde{\mathbf{\Omega}}_N^{-1} \mathbf{y}_i \\ &= \left( N^{-1} \sum_{i=1}^N (\mathbf{I}_M \otimes \mathbf{x}_i) \tilde{\mathbf{\Omega}}_N^{-1} (\mathbf{I}_M \otimes \mathbf{x}_i)' \right)^{-1} N^{-1} \sum_{i=1}^N (\mathbf{I}_M \otimes \mathbf{x}_i) \tilde{\mathbf{\Omega}}_N^{-1} \mathbf{y}_i\end{aligned}$$

Puisque  $\tilde{\mathbf{\Omega}}_N^{-1}$  est carrée de taille  $M \times M$  on a, avec  $\tilde{\mathbf{\Omega}}_N^{-1} = \tilde{\mathbf{\Omega}}_N^{-1} \otimes \mathbf{1}$  :

$$\begin{aligned}\hat{\mathbf{a}}_N^{MM}(\tilde{\mathbf{\Omega}}_N) &= \left( N^{-1} \sum_{i=1}^N (\tilde{\mathbf{\Omega}}_N^{-1} \otimes \mathbf{x}_i) (\mathbf{I}_M \otimes \mathbf{x}_i') \right)^{-1} N^{-1} \sum_{i=1}^N (\tilde{\mathbf{\Omega}}_N^{-1} \otimes \mathbf{x}_i) \mathbf{y}_i \\ &= \left( N^{-1} \sum_{i=1}^N \tilde{\mathbf{\Omega}}_N^{-1} \otimes (\mathbf{x}_i \mathbf{x}_i') \right)^{-1} N^{-1} \sum_{i=1}^N (\tilde{\mathbf{\Omega}}_N^{-1} \otimes \mathbf{x}_i) \mathbf{y}_i \\ &= \left( \tilde{\mathbf{\Omega}}_N^{-1} \otimes N^{-1} \sum_{i=1}^N (\mathbf{x}_i \mathbf{x}_i') \right)^{-1} N^{-1} \sum_{i=1}^N (\tilde{\mathbf{\Omega}}_N^{-1} \otimes \mathbf{x}_i) \mathbf{y}_i\end{aligned}$$

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

$$\begin{aligned} \hat{\mathbf{a}}_N^{MM}(\tilde{\mathbf{\Omega}}_N) &= \tilde{\mathbf{\Omega}}_N \otimes \left( N^{-1} \sum_{i=1}^N (\mathbf{x}_i \mathbf{x}_i') \right)^{-1} N^{-1} \sum_{i=1}^N (\tilde{\mathbf{\Omega}}_N^{-1} \otimes \mathbf{x}_i) \mathbf{y}_i \\ &= \tilde{\mathbf{\Omega}}_N \otimes \left( N^{-1} \sum_{i=1}^N (\mathbf{x}_i \mathbf{x}_i') \right)^{-1} (\tilde{\mathbf{\Omega}}_N^{-1} \otimes \mathbf{I}_K) (\tilde{\mathbf{\Omega}}_N \otimes \mathbf{I}_K) N^{-1} \sum_{i=1}^N (\tilde{\mathbf{\Omega}}_N^{-1} \otimes \mathbf{x}_i) \mathbf{y}_i \\ &= \mathbf{I}_M \otimes \left( N^{-1} \sum_{i=1}^N (\mathbf{x}_i \mathbf{x}_i') \right)^{-1} N^{-1} \sum_{i=1}^N (\mathbf{I}_M \otimes \mathbf{x}_i) \mathbf{y}_i \\ &= \left( N^{-1} \sum_{i=1}^N \mathbf{I}_M \otimes (\mathbf{x}_i \mathbf{x}_i') \right)^{-1} N^{-1} \sum_{i=1}^N (\mathbf{I}_M \otimes \mathbf{x}_i) \mathbf{y}_i \\ &= \left( N^{-1} \sum_{i=1}^N (\mathbf{I}_M \otimes \mathbf{x}_i) (\mathbf{I}_M \otimes \mathbf{x}_i') \right)^{-1} N^{-1} \sum_{i=1}^N (\mathbf{I}_M \otimes \mathbf{x}_i) \mathbf{y}_i \\ &= \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i \mathbf{X}_i' \right)^{-1} N^{-1} \sum_{i=1}^N \mathbf{X}_i \mathbf{y}_i \\ &= \hat{\mathbf{a}}_N^{MM}(\mathbf{I}_M) \end{aligned}$$

## 5. Instruments efficaces et MMG

## 5. Instruments efficaces et MMG

- Puisque l'instrument efficace juste-identifie les paramètres à estimer ...  
... pourquoi aurait-on encore besoin de la MMG ?
- Plusieurs réponses ici :
  - On n'a pas défini la forme des instruments efficaces dans le cas de systèmes d'équations avec des VI qui diffèrent par équation (cas de l'estimation des modèles « de panel » en particulier).
  - Le résultat de Chamberlain s'applique si on a  $E[\mathbf{u}_i / \mathbf{x}_i] = \mathbf{0}$  , il ne s'applique pas si on a « seulement »  $E[\mathbf{u}_i \mathbf{x}_i'] = \mathbf{0}$  .
  - La MMG s'applique à des conditions de moment autres que des conditions d'orthogonalité

- Mais même dans les cas où le résultat de Chamberlain s'applique :
  - Le calcul ou l'estimation des instruments efficaces peut être délicate, voire quasiment impossible en pratique (cas de beaucoup de VI).
  - Ceci dit, il me semble qu'on ne se réfère pas assez au résultat de Chamberlain pour choisir les instruments pour l'estimation des modèles à variables explicatives endogènes.

***MMG, MM et instruments : remarques finales***

- On considère ici le modèle :

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

et on note :

$$u_i(\mathbf{a}) \equiv y_i - f(\mathbf{x}_i; \mathbf{a}) \text{ avec } u_i = u_i(\mathbf{a}_0).$$

- On considère la condition de moment estimante (supposée identifiante) :

$$E[\boldsymbol{\rho}(\mathbf{z}_i)u_i(\mathbf{a}_0)] = \mathbf{0},$$

*i.e.* on utilise comme instrument une transformation des VI  $\mathbf{z}_i$ ,  $\boldsymbol{\rho}(\mathbf{z}_i)$ .

- L'estimateur  $\tilde{\mathbf{a}}_N^{MMG}$  de  $\mathbf{a}_0$  fondé sur cette condition est défini par :

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

- L'estimateur  $\tilde{\mathbf{a}}_N^{MMG}$  est caractérisé par les CO1 du programme de minimisation qui le définit :

$$N^{-1} \sum_{i=1}^N \left( N^{-1} \sum_{i=1}^N \mathbf{f}_i(\tilde{\mathbf{a}}_N^{MMG}) \boldsymbol{\rho}(\mathbf{z}_i)' \right) \tilde{\mathbf{M}}_N \boldsymbol{\rho}(\mathbf{z}_i) u_i(\tilde{\mathbf{a}}_N^{MMG}) = \mathbf{0}_{K \times 1}$$

où :

$$\mathbf{f}_i(\mathbf{a}) \equiv \frac{\partial f(\mathbf{x}_i; \mathbf{a})}{\partial \mathbf{a}}.$$

- En notant :

$$\tilde{\mathbf{r}}_N(\mathbf{z}_i) \equiv \left( N^{-1} \sum_{i=1}^N \mathbf{f}_i(\tilde{\mathbf{a}}_N^{MMG}) \boldsymbol{\rho}(\mathbf{z}_i)' \right) \tilde{\mathbf{M}}_N \boldsymbol{\rho}(\mathbf{z}_i)$$

on obtient que  $\tilde{\mathbf{a}}_N^{MMG}$  est défini comme un estimateur de la MM avec l'instrument estimé  $\hat{\mathbf{r}}_N(\mathbf{z}_i)$  :

$$N^{-1} \sum_{i=1}^N \tilde{\mathbf{r}}_N(\mathbf{z}_i) u_i(\tilde{\mathbf{a}}_N^{MMG}) = \mathbf{0}_{K \times 1}.$$

- Puisque :

$$\tilde{\mathbf{r}}_N(\mathbf{z}_i) \xrightarrow[N \rightarrow +\infty]{p} E[\mathbf{f}_i(\mathbf{a}_0)\boldsymbol{\rho}(\mathbf{z}_i)']\mathbf{M}_0\boldsymbol{\rho}(\mathbf{z}_i),$$

on obtient que  $\tilde{\mathbf{a}}_N^{MMG}$  peut être interprété comme un estimateur de la MM défini par la condition de moment juste-identifiante :

$$E\left[E[\mathbf{f}_i(\mathbf{a}_0)\boldsymbol{\rho}(\mathbf{z}_i)']\mathbf{M}_0\boldsymbol{\rho}(\mathbf{z}_i)u_i(\mathbf{a}_0)\right] = \mathbf{0}_{K \times 1},$$

cet estimateur utilisant un instrument estimé, *i.e.* un estimateur de :

$$\mathbf{r}(\mathbf{z}_i) \equiv E[\mathbf{f}_i(\mathbf{a}_0)\boldsymbol{\rho}(\mathbf{z}_i)']\mathbf{M}_0\boldsymbol{\rho}(\mathbf{z}_i).$$

Bien entendu, on suppose ici que :

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



- Dans le cas d'un *modèle linéaire* :

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

on a :

$$\mathbf{f}_i(\mathbf{a}) \equiv \mathbf{x}_i.$$

- Ce qui précède nous dit que l'*estimateur de la MMG* de  $\mathbf{a}_0$  fondé sur la condition d'orthogonalité :

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

et utilisant la matrice de pondération  $\tilde{\mathbf{M}}_N \xrightarrow[N \rightarrow +\infty]{p} \mathbf{M}_0$  est un *estimateur de la MM* fondé sur :

$$E[E[\mathbf{x}_i \boldsymbol{\rho}(\mathbf{z}_i)'] \mathbf{M}_0 \boldsymbol{\rho}(\mathbf{z}_i)(y_i - \mathbf{x}_i' \mathbf{a}_0)] = \mathbf{0}_{K \times 1}$$

i.e. :

$$E[\mathbf{r}(\mathbf{z}_i)u_i(\mathbf{a}_0)] = \mathbf{0}_{K \times 1} \text{ avec } \mathbf{r}(\mathbf{z}_i) \equiv E[\mathbf{x}_i \boldsymbol{\rho}(\mathbf{z}_i)'] \mathbf{M}_0 \boldsymbol{\rho}(\mathbf{z}_i).$$

$$E[\mathbf{r}(\mathbf{z}_i)u_i(\mathbf{a}_0)] = \mathbf{0}_{K \times 1} \text{ avec } \mathbf{r}(\mathbf{z}_i) \equiv E[\mathbf{x}_i \boldsymbol{\rho}(\mathbf{z}_i)'] \mathbf{M}_0 \boldsymbol{\rho}(\mathbf{z}_i).$$

- *Si on choisit*  $\mathbf{M}_0 = E[\boldsymbol{\rho}(\mathbf{z}_i)\boldsymbol{\rho}(\mathbf{z}_i)']^{-1}$  *on obtient*  $\hat{\mathbf{a}}_N^{2MC}$  *avec*  $\boldsymbol{\rho}(\mathbf{z}_i)$  *pour instruments et :*

$$\mathbf{r}(\mathbf{z}_i) = E[\mathbf{x}_i \boldsymbol{\rho}(\mathbf{z}_i)'] E[\boldsymbol{\rho}(\mathbf{z}_i)\boldsymbol{\rho}(\mathbf{z}_i)']^{-1} \boldsymbol{\rho}(\mathbf{z}_i) = EL[\mathbf{x}_i | \boldsymbol{\rho}(\mathbf{z}_i)].$$

*C'est bien si*  $EL[\mathbf{x}_i | \boldsymbol{\rho}(\mathbf{z}_i)] \simeq E[\mathbf{x}_i | \mathbf{z}_i]$  *d'après le résultat de Chamberlain.*

- *Si on choisit*  $\mathbf{M}_0 = E[\boldsymbol{\rho}(\mathbf{z}_i)u_i(\mathbf{a}_0)^2 \boldsymbol{\rho}(\mathbf{z}_i)']^{-1}$  *on obtient*  $\hat{\mathbf{a}}_N^{2MCH}$  *avec*  $\boldsymbol{\rho}(\mathbf{z}_i)$  *pour instruments et :*

$$\mathbf{r}(\mathbf{z}_i) = E[\mathbf{x}_i \boldsymbol{\rho}(\mathbf{z}_i)'] E[\boldsymbol{\rho}(\mathbf{z}_i)u_i(\mathbf{a}_0)^2 \boldsymbol{\rho}(\mathbf{z}_i)']^{-1} \boldsymbol{\rho}(\mathbf{z}_i).$$

*On pondère bien les conditions de moment mais on s'éloigne de*  $E[\mathbf{x}_i | \mathbf{z}_i]$

.

- Si on choisit l'instrument optimal, *i.e.* :

$$\boldsymbol{\rho}(\mathbf{z}_i) = E[\mathbf{x}_i | \mathbf{z}_i] V[u_i | \mathbf{z}_i]^{-1},$$

on n'a pas besoin de la MMG puisque  $E[\mathbf{x}_i | \mathbf{z}_i] V[u_i | \mathbf{z}_i]^{-1}$  juste-identifie  $\mathbf{a}_0$  dans :

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