# Introduction à l'économétrie

## Méthode des moindres carrés ordinaires en présence d'hétéroscédasticité ou d'autocorrélation des erreurs – Leçon 3

*Ce cours vous est proposé par Olivier Baron, Maître de conférences, Université de Bordeaux et par AUNEGe, l'Université Numérique en Économie Gestion.*

## Table des matières

***Préambule*** ........ **2**

***Introduction*** ........ **2**

***Prolongements : les Moindres Carrées Généralisés*** ........ **3**

***Transformation du modèle*** ........ **3**

***L'estimateur des Moindres Carrés Généralisés (MCG)*** ........ **5**

***Comparaison des estimateurs des MCO et des MCG*** ........ **5**

**Le critère d'estimation des Moindres Carrés Généralisés** ........ **6**

Conséquences ........ 6

**Estimateur de la variance de la perturbation** ........ **7**

**Les Moindres Carrés Quasi Généralisés (MCQG)** ........ **7**

***Références*** ........ **8**

AUNEGe
L'université numérique
Economie Gestion

# Préambule

**Objectifs :**

- Comprendre et étudier les conséquences de la non vérification de l'hypothèse $H_4$ sur la méthode d'estimation par moindres carrés ;
- Lever certaines interrogations concernant l'estimation des paramètres du modèle lorsque les situations décrites dans les deux premières sections de ce chapitre se rencontrent au niveau des données utilisées.

# Introduction

L'hypothèse $H_4$ présentée au cours du premier chapitre du cours suppose que la matrice de variances – covariances des perturbations est une matrice scalaire, soit :

$$\underset{(n,n)}{\operatorname{Var}} \varepsilon = E\left(\underset{(n,n)}{\varepsilon.\varepsilon'}\right) = \sigma^2.I_n \quad \Leftrightarrow \begin{cases} E(\varepsilon_i^2) = \operatorname{Var}(\varepsilon_i) = \sigma^2 & \forall\, i = 1, \dots, n \\ E(\varepsilon_i \varepsilon_j) = \operatorname{Cov}(\varepsilon_i\,, \varepsilon_j) = 0 & \forall\, i\,, \forall\, j, i \neq j \end{cases}$$

Selon cette hypothèse, la variance des perturbations est constante et égale à $\sigma^2$ (les erreurs sont homoscédastiques) et les perturbations ne sont pas corrélées entre elles (absence d'autocorrélation des erreurs).

Nous allons nous interroger dans ce chapitre sur les conséquences de la non vérification de cette hypothèse sur la méthode d'estimation par moindres carrés. Clairement, deux situations peuvent conduire à la remise en cause de l'hypothèse $H_4$ :

- Soit la variance des perturbations n'est pas constante (situation qualifiée d'hétéroscédasticité et plus couramment observée sur des données en coupe transversale) ;
- Soit les erreurs sont corrélées entre elles (situation qualifiée d'autocorrélation et plus souvent observée sur des séries temporelles ou longitudinales).

En pratique, les deux situations peuvent pourtant être simultanément rencontrées. Il faudra toujours tester la présence d'autocorrélation avant l'hétéroscédasticité car les tests permettant d'identifier cette dernière sont sensibles à l'autocorrélation alors que les tests d'autocorrélation (Durbin – Watson en particulier) sont robustes à l'hétéroscédasticité.

# Prolongements : les Moindres Carrées Généralisés

Cette section aborde des problèmes d'estimation plus complexes et ne fera pas l'objet d'un développement exhaustif, ni complètement détaillé. Elle permettra de lever certaines interrogations concernant l'estimation des paramètres du modèle lorsque les situations décrites dans les deux premières sections de ce chapitre se rencontrent au niveau des données utilisées.

On considère le modèle linéaire multiple habituel avec une matrice de variances-covariances des perturbations non scalaire, soit :

$$\boxed{\underset{(n,1)}{y} = \underset{(n,k)}{X} . \underset{(k,1)}{\beta} + \underset{(n,1)}{\varepsilon} \quad \text{avec} \quad \underset{(n,n)}{\text{Var}\ \varepsilon} = \underset{(n,n)}{\Omega} \neq \sigma^2 . I_n}$$

# Transformation du modèle

Par hypothèse on suppose donc que $\underset{(n,n)}{\text{Var}\ \varepsilon} = \underset{(n,n)}{\Omega}$ où $\underset{(n,n)}{\Omega}$ est une matrice symétrique, définie positive contenant $\frac{n(n+1)}{2}$ éléments inconnus différents.

Sans perte de généralité, on peut toujours factoriser un élément de cette matrice de variances-covariances, par exemple :

$$\underset{(n,n)}{\text{Var}\ \varepsilon} = E(\varepsilon.\varepsilon') = \sigma^2 . \underset{(n,n)}{\Psi}$$

En conséquence, on pourra obtenir une matrice carrée d'ordre $n$, triangulaire inférieure, notée $\Psi^{\frac{1}{2}}$ et appelée, par convention, racine carrée de $\Psi$, de telle sorte que :

$$\Psi = \Psi^{\frac{1}{2}} . \left(\Psi^{\frac{1}{2}}\right)'$$

Puisque $\Psi$ est une matrice symétrique définie positive, il s'agit de la décomposition de Cholesky, résultat qui stipule que toute matrice symétrique définie positive peut s'exprimer comme le produit d'une matrice triangulaire inférieure et de sa transposée (qui elle est triangulaire supérieure).

Cette matrice est inversible et son inverse $\left(\Psi^{\frac{1}{2}}\right)^{-1} = \Psi^{-\frac{1}{2}}$ est telle que : $\Psi^{-1} = \left(\Psi^{-\frac{1}{2}}\right)' . \Psi^{-\frac{1}{2}}$.

On notera plus loin cette matrice inverse $P = \Psi^{-\frac{1}{2}}$ et donc $\Psi^{-1} = P'.P$.

### Exemples de telles décompositions pour $n = 4$ :

**1.** $$\Psi = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 9 \end{pmatrix} \Rightarrow \Psi^{\frac{1}{2}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{3} & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \Rightarrow \Psi^{-\frac{1}{2}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & 0 & \frac{1}{3} \end{pmatrix}$$

et $$\Psi^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & 0 & \frac{1}{3} \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & 0 & \frac{1}{3} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{9} \end{pmatrix}$$

**2.** $$\Psi = \begin{pmatrix} 1{,}000 & 0{,}800 & 0{,}640 & 0{,}512 \\ 0{,}800 & 1{,}000 & 0{,}800 & 0{,}640 \\ 0{,}640 & 0{,}800 & 1{,}000 & 0{,}800 \\ 0{,}512 & 0{,}640 & 0{,}800 & 1{,}000 \end{pmatrix}$$

$$\Rightarrow \Psi^{\frac{1}{2}} = \begin{pmatrix} 1{,}000 & 0 & 0 & 0 \\ 0{,}800 & 0{,}600 & 0 & 0 \\ 0{,}640 & 0{,}480 & 0{,}600 & 0 \\ 0{,}512 & 0{,}384 & 0{,}480 & 0{,}600 \end{pmatrix} \Rightarrow \Psi^{-\frac{1}{2}} = \begin{pmatrix} 1{,}000 & 0 & 0 & 0 \\ -1{,}333 & 1{,}667 & 0 & 0 \\ 0 & -1{,}333 & 1{,}667 & 0 \\ 0 & 0 & -1{,}333 & 1{,}667 \end{pmatrix}$$

et $$\Psi^{-1} = \begin{pmatrix} 1{,}000 & -1{,}333 & 0 & 0 \\ 0 & 1{,}667 & -1{,}333 & 0 \\ 0 & 0 & 1{,}667 & -1{,}333 \\ 0 & 0 & 0 & 1{,}667 \end{pmatrix} \cdot \begin{pmatrix} 1{,}000 & 0 & 0 & 0 \\ -1{,}333 & 1{,}667 & 0 & 0 \\ 0 & -1{,}333 & 1{,}667 & 0 \\ 0 & 0 & -1{,}333 & 1{,}667 \end{pmatrix}$$

$$\Rightarrow \Psi^{-1} = \begin{pmatrix} 2{,}777 & -2{,}222 & 0 & 0 \\ -2{,}222 & 4{,}556 & -2{,}222 & 0 \\ 0 & -2{,}222 & 4{,}556 & -2{,}222 \\ 0 & 0 & -2{,}222 & 2{,}777 \end{pmatrix}$$

On va transformer le modèle initial pour retrouver les hypothèses classiques d'application des Moindres Carrés Ordinaires et donc satisfaire le Théorème de Gauss-Markov. Ainsi, on pré-multiplie le modèle par la matrice $P = \Psi^{-\frac{1}{2}}$ :

$$\underset{(n,1)}{P.y} = \underset{(n,k)}{P.X}.\underset{(k,1)}{\beta} + \underset{(n,1)}{P.\varepsilon} \quad \Leftrightarrow \quad \underset{(n,1)}{\tilde{y}} = \underset{(n,k)}{\tilde{X}}.\underset{(k,1)}{\beta} + \underset{(n,1)}{\tilde{\varepsilon}}$$

avec : $\begin{cases} \underset{(n,1)}{\tilde{y}} = P.\underset{(n,1)}{y} \\ \underset{(n,k)}{\tilde{X}} = P.\underset{(n,k)}{X} \\ \underset{(n,1)}{\tilde{\varepsilon}} = P.\underset{(n,1)}{\varepsilon} \end{cases}$

Dans ce nouveau modèle, il est simple de montrer que :

. $E\left(\underset{(n,1)}{\tilde{\varepsilon}}\right) = \underset{(n,1)}{0}$ ; Rang $\underset{(n,k)}{\tilde{X}} = k$

et les perturbations $\tilde{\varepsilon}$ sont sphériques. En effet :

$$\underset{(n,n)}{\text{Var}}\ \tilde{\varepsilon} = E(\tilde{\varepsilon}.\tilde{\varepsilon}') = E(P.\varepsilon(P.\varepsilon)') = E(P.\varepsilon\varepsilon'.P') = P.E(\varepsilon\varepsilon').P' = P.\sigma^2\Psi.P'$$

$$\underset{(n,n)}{\text{Var}}\ \tilde{\varepsilon} = \sigma^2.P\Psi P' = \sigma^2.P(P'P)^{-1}P' = \sigma^2.PP^{-1}(P')^{-1}P' = \sigma^2.I_n$$

On peut donc **appliquer les MCO au modèle transformé**.

# L'estimateur des Moindres Carrés Généralisés (MCG)

Les paramètres du modèle transformé vont être estimés par MCO :

$$\underset{(n,1)}{\tilde{y}} = \underset{(n,k)}{\tilde{X}}.\underset{(k,1)}{\beta} + \underset{(n,1)}{\tilde{\varepsilon}} \quad \text{avec} \quad \begin{cases} E(\tilde{\varepsilon}) = 0_n \\ \underset{(n,n)}{\text{Var}}\ \tilde{\varepsilon} = \sigma^2.I_n \end{cases}$$

$$\underset{(k,1)}{\hat{\beta}_{MCG}} = \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{y} = (X'P'PX)^{-1}X'P'Py$$

$$\underset{(k,1)}{\hat{\beta}_{MCG}} = (X'\Psi^{-1}X)^{-1}X'\Psi^{-1}y$$

On montre que $\hat{\beta}_{MCG}$ est sans biais et que $\text{Var}\ \hat{\beta}_{MCG} = \sigma^2.(X'\Psi^{-1}X)^{-1}$

C'est le meilleur estimateur linéaire sans biais pour le modèle avec erreurs hétéroscédastiques ou autocorrélées.

# Comparaison des estimateurs des MCO et des MCG

On cherche donc à estimer les paramètres du modèle :

$$\boxed{\underset{(n,1)}{y} = \underset{(n,k)}{X}.\underset{(k,1)}{\beta} + \underset{(n,1)}{\varepsilon} \quad \text{avec} \quad \begin{cases} E(\varepsilon) = 0_n \\ \underset{(n,n)}{\text{Var}}\ \varepsilon = \sigma^2.\Psi \end{cases}}$$

L'estimateur des MCG, $\hat{\beta}_{MCG} = (X'\Psi^{-1}X)^{-1}X'\Psi^{-1}y$ est sans biais et de variance minimale.

L'estimateur des MCO, $\hat{\beta}_{MCO} = (X'X)^{-1}X'y$ est lui aussi sans biais mais sa matrice de variances-covariances n'est plus : $\mathrm{Var}\ \hat{\beta}_{MCO} = \sigma^2(X'X)^{-1}$ mais :

$\mathrm{Var}\ \hat{\beta}_{MCO} = \sigma^2\left(X'X\right)^{-1}.X'\Psi X.\left(X'X\right)^{-1}$, qui est la vraie matrice de variances-covariances de l'estimateur des MCO, si les erreurs sont hétéroscédastiques ou autocorrélées. On parle ici d'une formule « sandwich » pour cette écriture, car la matrice $X'\Psi X$ est encadrée à gauche et à droite par $(X'X)^{-1}$.

On montre que la vraie matrice de variances-covariances de l'estimateur MCO est « plus grande » (au sens matriciel) que celle de l'estimateur MCG.

**Remarque**

Pour montrer ce résultat il suffit de montrer que $\mathrm{Var}\ \hat{\beta}_{MCO}$ - $\mathrm{Var}\ \hat{\beta}_{MCG}$ est une matrice définie positive.

**L'estimateur des MCO est donc moins précis que l'estimateur MCG**.

## Le critère d'estimation des Moindres Carrés Généralisés

Le critère de minimisation des Moindres Carrés Ordinaires s'applique au modèle transformé. $\hat{\beta}_{MCG}$ est donc solution de :

$$\underset{\beta}{\mathrm{Min}} \sum_{i=1}^{n} (\tilde{y}_i - \hat{\tilde{y}}_i)^2 = \underset{\beta}{\mathrm{Min}} \underset{(1,n)}{(\tilde{y} - \tilde{X}\hat{\beta})'}.\underset{(n,1)}{(\tilde{y} - \tilde{X}\hat{\beta})} = \underset{\beta}{\mathrm{Min}} \underset{(1,n)}{\left(y - X\hat{\beta}\right)'}.\underset{(n,n)}{\Psi^{-1}}.\underset{(n,1)}{(y - X\hat{\beta})}$$

On peut en conclure que le critère des Moindres Carrés Généralisés consiste à minimiser une somme pondérée des carrés des résidus. Géométriquement, on effectue une **projection oblique** et non plus orthogonale sur l'hyperplan engendré par les régresseurs.

### Conséquences

1. Même si le modèle comporte une constante, l'hyperplan de régression ne passe plus par le point moyen de l'échantillon ;
2. La somme des résidus estimés n'est plus nulle ;
3. Le coefficient de détermination $R^2$ n'est plus interprétable et peut même devenir négatif.

## Estimateur de la variance de la perturbation

La variance de l'estimateur des MCG est égale à $\sigma^2.(X'\Psi^{-1}X)^{-1}$. Elle est, bien entendu, inconnue, puisqu'elle dépend de $\sigma^2$, paramètre inconnu. Il faudra donc estimer $\sigma^2$ pour estimer la variance de $\hat{\beta}_{MCG}$.

On montre que $\hat{\sigma}^2 = \frac{\hat{\varepsilon}'_{MCG}.\Psi^{-1}.\hat{\varepsilon}_{MCG}}{n-k}$ est un estimateur sans biais de $\sigma^2$.

## Les Moindres Carrés Quasi Généralisés (MCQG)

Nous avons construit et calculé l'estimateur des Moindres Carrés Généralisés et avons abouti à l'expression : $\hat{\beta}_{MCG} = (X'\Psi^{-1}X)^{-1}X'\Psi^{-1}y$

Cet estimateur est néanmoins souvent impossible à calculer car on ne connaît pas la matrice $\Psi$. Il faut donc l'estimer. Comme nous l'avons déjà précisé, cette matrice comporte au maximum $\frac{n(n+1)}{2}$ éléments différents. Il est donc impossible d'estimer, dans la plupart des cas, tous ces paramètres avec seulement $n$ observations, à moins de contraindre la structure de cette matrice. On utilise alors des algorithmes itératifs non linéaires pour obtenir $\hat{\Psi}$.

On en déduit alors : $\hat{\beta}_{MCQG} = \left(X'\hat{\Psi}^{-1}X\right)^{-1}X'\hat{\Psi}^{-1}y$.

On parlera dans cette situation de **moindres carrés quasi généralisés** ou encore de **Feasible Generalised Least Squares** (FGLS) en anglais.

# Références

### Comment citer ce cours ?

Introduction à l'économétrie, Olivier Baron, AUNEGe (http://aunege.fr), CC – BY NC ND (http://creativecommons.org/licenses/by-nc-nd/4.0/).