

Chapitre 5 : Vérifications des hypothèses du modèle de régression linéaire multiple

I. AUTOCORRELATIONS DES ERREURS

5.1. Exposé du problème de l'autocorrélation des erreurs

Dans l'étude des modèles linéaires $Y = Xa + \varepsilon$, nous avons supposé l'hypothèse de non corrélation des erreurs : $E(\varepsilon_i \varepsilon_j) = 0 \quad \forall i \neq j$.

Autrement dit : nous avons supposé que la matrice des variances-covariances des erreurs est telle que :

$$\Omega_\varepsilon = E(\varepsilon \varepsilon') = \begin{pmatrix} E(\varepsilon_1 \varepsilon_1) & E(\varepsilon_1 \varepsilon_2) \cdots & E(\varepsilon_1 \varepsilon_n) \\ \vdots & \vdots & \vdots \\ \vdots & \ddots & \vdots \\ E(\varepsilon_n \varepsilon_1) & \cdots & E(\varepsilon_n \varepsilon_n) \end{pmatrix} = \begin{pmatrix} \sigma_\varepsilon^2 & 0 \cdots & 0 \\ \vdots & \vdots & \vdots \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_\varepsilon^2 \end{pmatrix} = \sigma_\varepsilon^2 I_n$$

Or, l'expérience montre que cette hypothèse est très coercitive (conditionnée) pour les erreurs successives. Dans ce cas, les concepts économiques considérés se présentent, le plus souvent, sous la forme de séries temporelles où il est plus approprié d'admettre la dépendance de l'erreur d'une période par rapport à la période antérieure (d'une année à l'année antérieure, ...etc.).

En d'autre terme, il faut admettre que : $E(\varepsilon_i \varepsilon_j) \neq 0 \quad \forall i \neq j$.

C'est-à-dire, que la matrice des variances-covariances des erreurs est telle que :

$$\Omega_\varepsilon = E(\varepsilon \varepsilon') \neq \sigma_\varepsilon^2 I_n$$

Ce problème est connu sous le nom de *l'autocorrélation des erreurs*.

5.2. Conséquences de l'autocorrélation des erreurs

Soit le modèle $Y = Xa + \varepsilon$, l'estimateur $\hat{a} = (X'X)^{-1}X'Y$ obtenu par les moindres carrés ordinaires (MCO) est sans biais, mais la matrice $\Omega_{\hat{a}}$ des variances covariances des coefficients est modifiée.

La variance de l'erreur n'est plus constante sur l'échantillon (Hypothèse H_3 n'est pas respectée) ; elle varie avec les observations d'une ou plusieurs variables explicatives, ou bien est différente sur un ou plusieurs sous échantillons. Analytiquement, ceci se traduit par :

$$\Omega_\varepsilon = E(\varepsilon \varepsilon') = \begin{pmatrix} \sigma_{\varepsilon_1}^2 & 0 \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{\varepsilon_n}^2 \end{pmatrix}$$

Ainsi, la conséquence sur l'estimateur des MCO demeure sans biais. En revanche, l'estimateur des MCO n'est plus convergent, il est dit inefficace. Les écarts types issus des MCO et par conséquent, les statistiques de *student* ne sont plus utilisables, et il devient impossible de juger de la significativité des estimations. En effet :

$$\Omega_{\hat{a}} = E((\hat{a} - a)(\hat{a} - a)') = (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1} \neq \sigma_{\varepsilon}^2(X'X)^{-1} \text{ (car } E(\varepsilon\varepsilon') \neq \sigma_{\varepsilon}^2 I_n \text{)}.$$

Il se pose alors le problème de la recherche d'un nouvel estimateur optimal pour « a » par les méthodes autre que la méthode des MCO, et la nécessité de détecter une éventuelle autocorrélation des erreurs.

5.3. Estimateur des moindres carrés généralisés (MCG)

Considérons le modèle linéaire $Y = Xa + \varepsilon$ dans lequel : $\Omega_{\varepsilon} = E(\varepsilon\varepsilon') \neq \sigma_{\varepsilon}^2 I_n$, il est démontré qu'un estimateur de « a » sans biais et à variance minimale est donné par la formule : $\hat{a} = (X' \Omega_{\varepsilon}^{-1} X)^{-1} (X' \Omega_{\varepsilon}^{-1} Y)$.

Cet estimateur est appelé estimateur des moindres carrés généralisés ou estimateur de Aitken. Il est aussi démontré que : $\Omega_{\hat{a}} = (X' \Omega_{\varepsilon}^{-1} X)^{-1}$.

5.4. Détection de l'autocorrélation des erreurs

La détection de l'autocorrélation des erreurs peut s'effectuer par une analyse graphique, par le test de Durbin et Watson et par le test de Breush Godfrey...etc.

5.4.1. Analyse graphique

On dit qu'on est en présence d'une autocorrélation des erreurs lorsque les erreurs sont liées par un processus de reproduction (processus à mémoire).

Selon la forme du graphique représentant l'évolution des erreurs, On peut distinguer une autocorrélation positive des erreurs (Cf. Graphique 5.1), dans ce cas, les erreurs sont pendant plusieurs périodes consécutives soit positives, soit négatives.

Une autocorrélation négative des erreurs est représentée sur le (Cf. Graphique 5.2), dans ce cas, les erreurs sont alternées.

Figure 5.1. Autocorrélation positive des erreurs

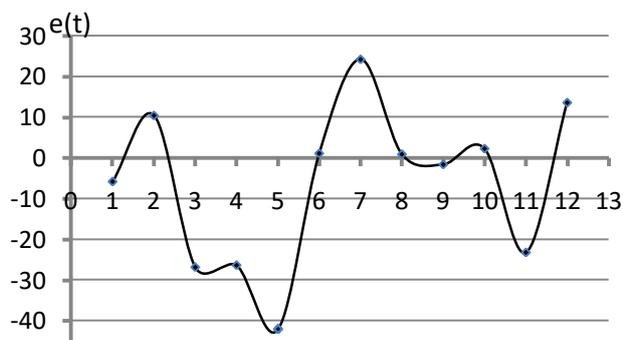
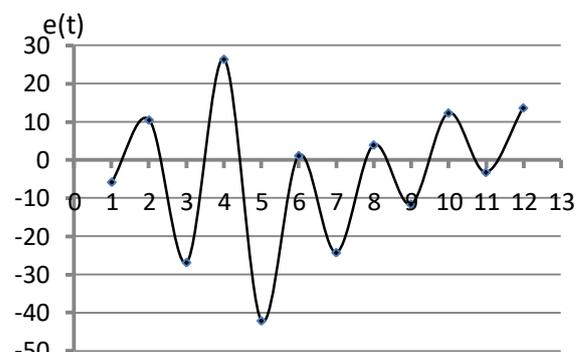


Figure 5.2. Autocorrélation négative des erreurs



5.4.2. Test de Durbin et Watson (DW)

Le test de Durbin et Watson permet de détecter une autocorrélation des erreurs d'ordre 1 selon la forme : $\varepsilon_i = \rho\varepsilon_{i-1} + \vartheta_i$ avec : $\vartheta_i \rightarrow N(0, \sigma_\vartheta^2)$.

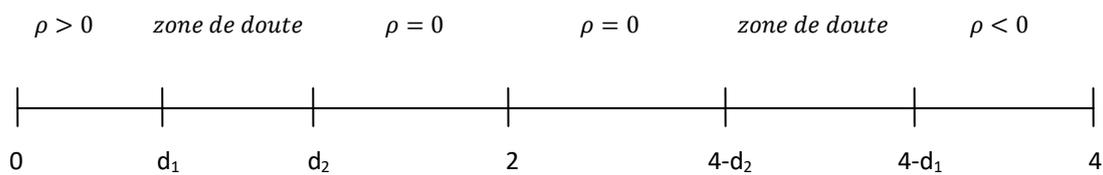
Le test d'hypothèse de DW est $H_0 : \rho=0$ contre $H_1 : \rho \neq 0$

Pour tester H_0 , (DW) utilisent la statistique suivante : $DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n (e_i)^2}$ où : $e_i = y_i - \hat{y}_i$

Par construction : $DW \approx 2(1 - \hat{\rho})$ avec $\hat{\rho} = \frac{\sum_{i=2}^n e_i e_{i-1}}{\sum_{i=1}^n e_i^2}$

Afin de tester H_0 , (DW) ont tabulé les valeurs critiques de DW au seuil de signification $\alpha = 5\%$ en fonction de n et k , avec k représente le nombre de variables explicatives.

La lecture de la table permet de déterminer deux points critiques d_1 et d_2 compris entre 0 et 2 qui subdivisent l'intervalle $[0, 4]$ selon le schéma suivant :



a) Règle de décision :

- Si $0 \leq DW < d_1$, on rejette H_0 , il y a une autocorrélation positive des erreurs ;
- Si $4 - d_1 < DW \leq 4$, on rejette H_0 , il y a une autocorrélation négative des erreurs ;
- Si $d_2 < DW \leq 4 - d_2$, on accepte H_0 , il y a absence d'autocorrélation des erreurs ;
- Si $d_1 \leq DW \leq d_2$ ou $4 - d_2 \leq DW \leq 4 - d_1$, on accepte H_0 , il y a doute.

b) Conditions d'utilisation du test de DW

- Le modèle de régression linéaire multiple doit comporter un terme constant ;
- La taille de l'échantillon $n \geq 15$;
- La variable à expliquer (Y_t) ne doit pas figurer parmi les variables explicatives entant que variable retardée (Y_{t-1} par exemple) ;
- Pour les modèles en coupe instantanée, les observations doivent être ordonnées en fonction de la variable à expliquer.

c) Interprétation du test de DW

Dans le cas où le test de DW admet l'hypothèse d'autocorrélation des erreurs, le statisticien considère :

- Soit une liaison temporelle des erreurs ;
- Soit une omission d'une variable explicative.

5.4.3. Test de Breusch-Godfrey (BG)

Ce teste permet de détecter l'autocorrélation des erreurs d'un ordre supérieur à 1. On peut même l'utiliser en présence d'une variable dépendante décalée en tant que variable explicative.

Il est fondé sur le test de fisher de nullité des coefficients du Multiplicateur de Lagrange (LM). Son principe est de chercher une éventuelle relation significative entre le résidu et ce même résidu décalé.

Une autocorrélation des erreurs d'un ordre p s'écrit :

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \dots + \rho_p \varepsilon_{t-p} + \vartheta_t$$

Soit le modèle général à erreurs autocorrélées d'ordre p :

$$Y_t = a_0 + a_1 X_{1t} + a_2 X_{2t} + \dots + \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \dots + \rho_p \varepsilon_{t-p} + \vartheta_t \dots \dots (5.1)$$

Ce test est mené en trois étapes :

- Estimation par les MCO du modèle (5.1) et calcul du résidu e_t .
- Estimation par les MCO de l'équation intermédiaire suivante :

$$e_t = a_0 + a_1 X_{1t} + a_2 X_{2t} + \dots + \rho_1 e_{t-1} + \rho_2 e_{t-2} + \dots + \rho_p e_{t-p} + \vartheta_t$$
- On teste sur l'équation intermédiaire l'hypothèse $H_0: " \rho_1 = \rho_2 = \dots = \rho_p = 0 "$ contre $H_1: \exists i, \text{ telque } \rho_i \neq 0 "$, $i = 1, \dots, p$.

Pour mener ce test, nous effectuons :

- i) Le test de fisher de nullité des coefficients ρ_i .
 - Si $F_{\text{calculé}} \leq F_{\text{tabulé}}$, on accepte H_0 , y a absence d'autocorrélation des erreurs.
 - Si $F_{\text{calculé}} > F_{\text{tabulé}}$, on rejete H_0 , y a présemption d'autocorrélation des erreurs.
- ii) Faire recours à la statistique LM qui est distribuée comme un Khi-2 à p degré de liberté ($X_\alpha^2(p)$).
 - Si $n \times R^2 > X_\alpha^2(p)$ (lu dans la table), on rejete l'hypothèse d'indépendance des erreurs.

Remarque 1: d'autres tests de détection de l'autocorrélation des erreurs existent, on donne comme exemple le test de Box-Ljung et le test de Box-Pierce.

5.5. Les procédures d'estimations en cas d'autocorrélation des erreurs

Si nous retenons que les erreurs sont liées entre elles par un processus autorégressif d'ordre (1), le modèle linéaire s'écrit $Y = X.a + \varepsilon$, avec : $\varepsilon_i = \rho \varepsilon_{i-1} + \vartheta_i$ avec $|\rho| \leq 1$.

L'erreur aléatoire ϑ_i est telle que :

- $\forall i, E(\vartheta_i) = 0$;
- $\forall i, E(\vartheta_i)^2 = \sigma_\vartheta^2$;
- $\forall i \neq j, E(\vartheta_i \vartheta_j) = 0$;

Dans ce cas, nous allons présenter le processus d'estimation à suivre, dans ce qui suit.

5.5.1. Utilisation de l'estimateur des moindres carrés généralisés (MCG)

Afin d'utiliser l'estimateur des MCG, on explicite la matrice des variances-covariances Ω_ε .

$$\Omega_\varepsilon = E(\varepsilon\varepsilon') = \begin{pmatrix} E(\varepsilon_1\varepsilon_1) & E(\varepsilon_1\varepsilon_2) & \cdots & E(\varepsilon_1\varepsilon_n) \\ \vdots & \ddots & & \vdots \\ E(\varepsilon_n\varepsilon_1) & \cdots & & E(\varepsilon_n\varepsilon_n) \end{pmatrix}$$

➤ **Calcul de $E(\varepsilon_i\varepsilon_i) = E(\varepsilon_i^2)$:**

$$\begin{aligned} E(\varepsilon_i^2) &= E((\rho\varepsilon_{i-1} + \vartheta_i)^2) = \rho^2 E(\varepsilon_{i-1}^2) + E(\vartheta_i^2) + 2\rho E(\varepsilon_{i-1}\vartheta_i) \\ &= \rho^2\sigma_\varepsilon^2 + \sigma_\vartheta^2 + 0 \end{aligned}$$

$$\Leftrightarrow \sigma_\varepsilon^2 - \rho^2\sigma_\varepsilon^2 = \sigma_\vartheta^2 \quad \Leftrightarrow \quad \sigma_\varepsilon^2 = \frac{\sigma_\vartheta^2}{1-\rho^2}$$

➤ **Calcul de $E(\varepsilon_i\varepsilon_{i+1})$:**

$$E(\varepsilon_i\varepsilon_{i+1}) = E(\varepsilon_i(\rho\varepsilon_i + \vartheta_{i+1})) = \rho E(\varepsilon_i^2) + E(\varepsilon_i\vartheta_{i+1}) = \rho\sigma_\varepsilon^2$$

➤ **Calcul de $E(\varepsilon_i\varepsilon_{i+2})$:**

$$E(\varepsilon_i\varepsilon_{i+2}) = E(\varepsilon_i(\rho\varepsilon_{i+1} + \vartheta_{i+2})) = \rho E(\varepsilon_i\varepsilon_{i+1}) + E(\varepsilon_i\vartheta_{i+2}) = \rho^2\sigma_\varepsilon^2$$

Ainsi, de proche en proche, on calcule $E(\varepsilon_i\varepsilon_{i+j})$. par récurrence, on montre que :

$$E(\varepsilon_i\varepsilon_{i+j}) = \rho^j\sigma_\varepsilon^2.$$

La matrice des variances-covariances Ω_ε s'écrit alors :

$$\Omega_\varepsilon = \begin{pmatrix} \sigma_\varepsilon^2 & \rho\sigma_\varepsilon^2 & \rho^2\sigma_\varepsilon^2 & \cdots & \rho^{n-1}\sigma_\varepsilon^2 \\ \rho\sigma_\varepsilon^2 & \sigma_\varepsilon^2 & \rho\sigma_\varepsilon^2 & \cdots & \rho^{n-2}\sigma_\varepsilon^2 \\ \vdots & \vdots & & & \vdots \\ & & \ddots & & \\ \rho^{n-1}\sigma_\varepsilon^2 & \rho^{n-2}\sigma_\varepsilon^2 & & \cdots & \sigma_\varepsilon^2 \end{pmatrix} \quad \text{avec } \rho \neq 1$$

$$\Omega_\varepsilon = \sigma_\varepsilon^2 \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \vdots & \vdots & \ddots & & \vdots \\ & & & \ddots & \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{pmatrix} \quad \text{avec } \sigma_\varepsilon^2 = \frac{\sigma_\vartheta^2}{1-\rho^2}$$

L'inverse de Ω_ε est égale à :

$$\Omega_\varepsilon^{-1} = \frac{1}{\sigma_\vartheta^2} \begin{pmatrix} 1 & -\rho & 0 & 0 & \dots & \dots & 0 \\ -\rho & 1 + \rho^2 & -\rho & 0 & \dots & \dots & 0 \\ 0 & -\rho & 1 + \rho^2 & -\rho & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \dots & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots & & 1 + \rho^2 & -\rho \\ 0 & 0 & 0 & 0 & 0 & -\rho & 1 \end{pmatrix}$$

Cependant, nous ne connaissons ni le terme ρ , ni la variance σ_ϑ^2 , nous allons donc chercher une transformation matricielle M telle que le modèle $MY = MX.a + M\varepsilon$ ait ses erreurs indépendantes et homoscédastiques.

5.5.2. Passage aux quasi-différences premières avec les moindres carrés quasi-généralisés (MCQG)

Sur le modèle $Y = X.a + \varepsilon \dots\dots (5.2)$

ou encore $Y_i = a_0 + a_1X_{1i} + a_2X_{2i} + \dots + a_kX_{ki} + \varepsilon_i \dots\dots (5.3)$ avec $\varepsilon_i = \rho \varepsilon_{i-1} + \vartheta_i$

L'équation (5.3) se transforme en :

$$\rho Y_{i-1} = \rho a_0 + \rho a_1 X_{1i-1} + \rho a_2 X_{2i-1} + \dots + \rho a_k X_{ki-1} + \rho \varepsilon_{i-1} \dots\dots (5.4)$$

En retranchant (5.4) de (5.3), on obtient :

$$Y_i - \rho Y_{i-1} = a_0(1 - \rho) + a_1(X_{1i} - \rho X_{1i-1}) + a_2(X_{2i} - \rho X_{2i-1}) + \dots + a_k(X_{ki} - \rho X_{ki-1}) + \varepsilon_i - \rho \varepsilon_{i-1} \dots\dots (5.5)$$

Une régression par les MCO sur cette dernière équation donnerait alors \hat{a} sans biais et à variance minimale.

Ecriture matricielle : en faisant varier dans (5.5) i de 2 à n , on aura :

$$\begin{aligned} Y_2 - \rho Y_1 &= a_0(1 - \rho) + a_1(X_{12} - \rho X_{11}) + \dots + a_k(X_{k2} - \rho X_{k1}) + \varepsilon_2 - \rho \varepsilon_1 \\ Y_3 - \rho Y_2 &= a_0(1 - \rho) + a_1(X_{13} - \rho X_{12}) + \dots + a_k(X_{k3} - \rho X_{k2}) + \varepsilon_3 - \rho \varepsilon_2 \\ &\vdots && \vdots && \vdots \\ Y_n - \rho Y_{n-1} &= a_0(1 - \rho) + a_1(X_{1n} - \rho X_{1n-1}) + \dots + a_k(X_{kn} - \rho X_{kn-1}) + \varepsilon_n - \rho \varepsilon_{n-1} \end{aligned}$$

$$\Leftrightarrow MY = MX.a + M\varepsilon \dots\dots (5.6)$$

avec :

$$M = \begin{pmatrix} -\rho & 1 & 0 & \dots & \dots & 0 \\ 0 & -\rho & 1 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \dots & \vdots \\ 0 & 0 & 0 & 0 & -\rho & 1 \end{pmatrix}$$

La méthode des MCO appliquée à l'équation transformée (5.6) donne :
 $\hat{a} = ((MX)'MX)^{-1}((MX)'MY)$ ou encore $\hat{a} = (X'M'MX)^{-1}(X'M'MY)$

avec :

$$M'M = \begin{pmatrix} \rho^2 & -\rho & 0 & \dots & \dots & 0 \\ -\rho & 1 + \rho^2 & -\rho & \dots & \dots & 0 \\ 0 & -\rho & 1 + \rho^2 & & & 0 \\ \vdots & \vdots & \vdots & \ddots & \dots & \vdots \\ 0 & & & & 1 + \rho^2 & -\rho \\ 0 & 0 & 0 & 0 & -\rho & 1 \end{pmatrix}$$

Remarque 2 :

Si l'on compare $M'M$ avec Ω_ε^{-1} obtenue précédemment, on constate au coefficient $\frac{1}{\sigma_\varepsilon^2}$ près, une grande ressemblance ; tous les éléments des deux matrices sont égaux sauf celui qui est à l'intersection de la 1^{ère} ligne et la 1^{ère} colonne :

- Pour $M'M$ c'est ρ^2 ;
- Pour Ω_ε^{-1} c'est 1.

Il est ainsi possible d'appliquer la méthode des MCO sur le modèle transformé en quasi-différences premières pour obtenir un estimateur sans biais et à variance minimale.

➤ Procédure d'estimation de ρ :

Afin de rendre les méthodes données en 5.5.1 et 5.5.2 opérationnelles, il convient d'estimer ρ .

- i) On remplace ρ par $\hat{\rho} = 1 - \frac{DW}{2}$;
- ii) On remplace ρ par $\hat{\rho}$ calculé par la régression directe de e_i sur e_{i-1} , $\hat{\rho} = \frac{\sum_{i=2}^n e_i e_{i-1}}{\sum_{i=1}^n e_i^2}$

II. HETEROSCEDASTICITE DES ERREURS

5.6. Exposé du problème d'hétéroscédasticité

Dans l'étude du modèle linéaire $Y = X.a + \varepsilon$, nous avons supposé l'hypothèse d'homoscédasticité des erreurs (hypothèse 4 du chapitre 3) $E(\varepsilon_i^2) = \sigma_\varepsilon^2$, ainsi, la matrice des variances-covariances des erreurs $\Omega_\varepsilon = E(\varepsilon\varepsilon') = \sigma_\varepsilon^2 I_n$.

Or, l'expérience montre que cette hypothèse n'est pas toujours vérifiée pour les séries en coupe instantanée, ou bien lorsque les observations sont représentatives des moyennes. On parle alors, d'hypothèse d'hétéroscédasticité, c'est-à-dire qu'on est obligé d'admettre que $E(\varepsilon_i^2) \neq \sigma_\varepsilon^2$. En présence d'hétéroscédasticité, la variance de l'erreur n'est plus constante.

Dans ce cas, la matrice des variances-covariances des erreurs Ω_ε n'est plus diagonale, $\Omega_\varepsilon \neq \sigma_\varepsilon^2 I_n$.

$$\Omega_\varepsilon = E(\varepsilon\varepsilon') = \begin{pmatrix} E(\varepsilon_1\varepsilon_1) & E(\varepsilon_1\varepsilon_2) \cdots & E(\varepsilon_1\varepsilon_n) \\ E(\varepsilon_2\varepsilon_1) & E(\varepsilon_2\varepsilon_2) & \\ \vdots & \ddots & \vdots \\ E(\varepsilon_n\varepsilon_1) & \cdots & E(\varepsilon_n\varepsilon_n) \end{pmatrix} = \begin{pmatrix} \sigma_{\varepsilon_1}^2 & 0 \cdots & 0 \\ 0 & \sigma_{\varepsilon_2}^2 & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{\varepsilon_n}^2 \end{pmatrix}$$

Les variances des erreurs ne sont plus constantes sur la première diagonale.

Les conséquences de l'hétéroscédasticité sont identiques à celles de l'autocorrélation des erreurs, c'est-à-dire, avec les MCO, on obtient des estimateurs sans biais mais qui ne sont pas à variance minimale.

5.7. Détection de l'hétéroscédasticité

La détection de l'hétéroscédasticité peut s'effectuer par le test de White et par le test de Gleisjer...etc.

5.7.1. Test de white

Avec le test de white on teste l'hypothèse nulle d'absence d'hétéroscédasticité (H_0) contre l'hypothèse d'hétéroscédasticité de forme inconnue (H_1).

La statistique de ce test est issue de la régression du carré des résidus sur toutes les variables explicatives et, leurs carrés, et des fois tous leurs produits croisés possibles).

$$e_t^2 = a_0 + a_1 X_{1t} + b_1 X_{1t}^2 + a_2 X_{2t} + b_2 X_{2t}^2 + \cdots + a_k X_{kt} + b_k X_{kt}^2 + \vartheta_t \dots \dots (5.7)$$

Soit n le nombre d'observations disponibles pour estimer les paramètres du modèle et R^2 le coefficient de détermination du modèle (5.7).

- Si l'un des coefficient de ce modèle est significativement différent de zéro, alors on accepte l'hypothèse d'hétéroscédasticité.

Nous pouvons procéder à ce test :

- i) Soit, à l'aide d'un test de Fisher de nullité des coefficients :

$$H_0: "a_1 = b_1 = a_2 = b_2 = \dots = a_k = b_k = 0"$$

- Si on rejette l'hypothèse H_0 , donc il existe un risque d'hétéroscédasticité ;

- ii) Soit, on va recourir à la statistique LM qui est distribuée comme un χ^2 à $p=2k$ degré de liberté (autant de coefficients que nous estimons, hormis le terme constant).

- Si $n \times R^2 > \chi^2_\alpha(p)$ (lu dans la table), on rejette l'hypothèse d'homoscédasticité des erreurs.

5.7.2. Test de Gleisjer

Ce test est fondé sur la relation entre le résidu de l'estimation par les MCO effectuée sur le modèle de base et la variable explicative supposée être la cause de l'hétéroscédasticité.

Étape1 : régression par les MCO de Y_j en X_j , le vecteur e_j est donc connu ;

Étape2 : régression de la valeur absolue $|e_j|$ des résidus sur X_j ;

Gleisjer suggère de tester différentes formes de relation, par exemple :

- $|e_j| = a_0 + a_1 X_j + \vartheta_j \rightarrow$ hétéroscédasticité de type : $\hat{\sigma}_{\varepsilon_j}^2 = k^2 X_j^2$;
- $|e_j| = a_0 + a_1 X_j^{\frac{1}{2}} + \vartheta_j \rightarrow$ hétéroscédasticité de type : $\hat{\sigma}_{\varepsilon_j}^2 = k^2 X_j$;
- $|e_j| = a_0 + a_1 X_j^{-1} + \vartheta_j \rightarrow$ hétéroscédasticité de type : $\hat{\sigma}_{\varepsilon_j}^2 = k^2 X_j^{-2}$;

Avec k une constante quelconque.

L'hypothèse d'homoscédasticité est rejetée si le coefficient a_1 d'une des spécifications est significativement différent de zéro.

La forme à retenir en cas d'hétéroscédasticité se refaire au coefficient affecté du ratio de *student* le plus élevé.

5.8. Méthodes d'estimations

En cas d'hétéroscédasticité, pour l'estimation du modèle, on utilise la méthode des moindres carrés généralisés (MCG) ou la méthode de régression pondérée (MCP).

5.8.1. Méthode des MCG

L'estimateur de \mathbf{a} sans biais et à variance minimale est celui des MCG :

$$\hat{\mathbf{a}} = (\hat{X}' \Omega_\varepsilon^{-1} X)^{-1} (\hat{X}' \Omega_\varepsilon^{-1} Y)$$

avec :

$$\Omega_\varepsilon = \begin{pmatrix} \sigma_{\varepsilon_1}^2 & 0 \dots & 0 \\ 0 & \sigma_{\varepsilon_2}^2 & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{\varepsilon_n}^2 \end{pmatrix} \quad \text{et} \quad \Omega_\varepsilon^{-1} = \begin{pmatrix} \frac{1}{\sigma_{\varepsilon_1}^2} & 0 \dots & 0 \\ 0 & \frac{1}{\sigma_{\varepsilon_2}^2} & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\sigma_{\varepsilon_n}^2} \end{pmatrix}$$

L'utilisation de cet estimateur nécessite l'estimation de Ω_ε^{-1} .

5.8.2. Méthode de régression pondérée (MCP)

Au lieu d'avoir pour tout $i=1, \dots, n$:

$var(\varepsilon_i) = E(\varepsilon_i^2) = \sigma_\varepsilon^2$, on aura $var(\varepsilon_i) = \frac{\sigma_\varepsilon^2}{w_i}$, où les poids w_i peuvent être différents pour chaque $i=1, \dots, n$.

La matrice des variances-covariances de ε est modifiée comme suit :

$$\Omega_\varepsilon = \begin{pmatrix} \frac{\sigma_\varepsilon^2}{w_1} & 0 \dots & 0 \\ 0 & \frac{\sigma_\varepsilon^2}{w_2} & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{\sigma_\varepsilon^2}{w_n} \end{pmatrix}$$

En posant : $v_1 = \sqrt{w_1} \varepsilon_1$,
 $v_2 = \sqrt{w_2} \varepsilon_2$,
 \vdots
 $v_n = \sqrt{w_n} \varepsilon_n$,

On aura : $var(v_1) = var(\sqrt{w_1} \varepsilon_1) = (\sqrt{w_1})^2 var(\varepsilon_1) = w_1 \frac{\sigma_\varepsilon^2}{w_1} = \sigma_\varepsilon^2$,
 $var(v_2) = var(\sqrt{w_2} \varepsilon_2) = (\sqrt{w_2})^2 var(\varepsilon_2) = w_2 \frac{\sigma_\varepsilon^2}{w_2} = \sigma_\varepsilon^2$,
 \vdots
 $var(v_n) = var(\sqrt{w_n} \varepsilon_n) = (\sqrt{w_n})^2 var(\varepsilon_n) = w_n \frac{\sigma_\varepsilon^2}{w_n} = \sigma_\varepsilon^2$,

Ainsi, pour obtenir un modèle homoscedastique, il faut procéder à la transformation suivante :

$$MY = MXa + M\varepsilon$$

$$\text{avec : } M = \begin{pmatrix} \sqrt{w_1} & 0 \dots & 0 \\ 0 & \sqrt{w_2} & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sqrt{w_n} \end{pmatrix}$$

La méthode des MCO appliquée au modèle homoscédastique $MY = MX\alpha + M\varepsilon$ nous donne un estimateur sans biais et à variance minimale :

$$\hat{\alpha} = ((MX)'MX)^{-1}((MX)'MY) \text{ ou encore } \hat{\alpha} = (X'M'MX)^{-1}(X'M'MY)$$

avec :

$$M'M = \begin{pmatrix} w_1 & 0 \cdots & 0 \\ 0 & w_2 & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_n \end{pmatrix}$$

III. MODELES A ERREURS SUR LES VARIABLES

5.9. Conséquences lorsque les variables sont entachées d'erreurs

Dans les hypothèses des modèles linéaires (Cf. Chapitre 3), nous avons supposé que la variable endogène et les variables exogènes étaient observables sans erreurs (H_1).

Dans la pratique, cette hypothèse est rarement vérifiée, cependant, nous pouvons admettre, généralement, que l'erreur de mesure des observations est faible par rapport à l'erreur de spécification.

Toutes fois, dans certains modèles, les variables économiques retenues peuvent être entachées d'une erreur de mesure relativement importante. C'est le cas, par exemple, lorsque les données proviennent non pas, d'une mesure directe, mais des données d'enquêtes et de sondage. Dans ce cas, il convient de distinguer les variables vraies (et inconnues) : $Y^*, X_1^*, X_2^*, \dots, X_k^*$ des valeurs observées Y, X_1, X_2, \dots, X_k et d'étudier les conséquences concernant les propriétés de l'estimateur obtenu par les MCO.

5.10. La méthode des variables instrumentales

La méthode des variables instrumentales est utilisée lorsqu'une ou des variables explicatives sont corrélées avec le terme d'erreur ($cov(x_t, \varepsilon_t) \neq 0$) ce qui est en contradiction avec les hypothèses respectées au cas de l'estimation par les MCO (H_5). Si cette hypothèse n'est pas vérifiée, l'estimateur des MCO ne sera pas convergent.

Cela se produit dans les modèles autoregressifs ou lorsque les variables explicatives sont entachées d'une erreur d'observation.

Le but de la technique des variables instrumentales est de déterminer une (ou des) variable(s) explicative(s) Z_t telle que :

$$E(z_t, \varepsilon_t) = 0 \text{ et } cov(z_t, x_t) \neq 0.$$

Le difficulté de mise en œuvre de cette méthode réside dans la sélection de la (ou des) variable(s) instrumentale(s) (*Proxy variable*) (appelées les instruments) z_t qui doit (doivent) être non corrélée(s) avec ε_t et fortement corrélée(s) avec x_t .

Dans certains cas, nous pouvons simplement retenir comme variable instrumentale la variable exogène décalée d'une période : $z_t = x_{t-1}$.

L'estimateur des variables instrumentales est donnée par : $\hat{a} = (Z'X)^{-1}(Z'Y)$

La matrice des variances-covariances des coefficients est égale à :

$$\hat{\Omega}_{\hat{a}} = \hat{\sigma}_{\varepsilon}^2 (Z'X)^{-1} (Z'Z) (X'Z)^{-1}$$

avec : Y : vecteur de la variable à expliquer ;

X : matrice des variables explicatives ;

Z : matrice des instruments.

5.11. Le test d'exogénéité d'Hausman

Ce test permet de détecter une éventuelle corrélation entre le terme d'erreur ε_t et une ou plusieurs variables explicatives x_{it} .

Dans cette hypothèse, nous ne pouvons plus utiliser l'estimateur des MCO qui est non convergent, il faut alors recourir à la méthode des variables instrumentales (VI) ou à la méthode des moments généralisés (MMG).

Soit le test d'hypothèse, $H_0: cov(x_t, \varepsilon_t) = 0$ (la variable x_t est exogène) contre l'hypothèse d'endogénéité $H_1: cov(x_t, \varepsilon_t) \neq 0$.

Sous l'hypothèse H_0 les estimateurs des MCO et des VI sont convergents alors que sous H_1 la covariance est non nulle et l'estimateur des MCO est biaisé et non convergent.

Ce test peut être mené de deux manières, soit un test de différence entre l'estimateur des VI et des MCO, soit un test à partir d'une régression augmentée.

5.11.1. Test de différence

Nous calculons la statistique : $H = (\hat{a}_{VI} - \hat{a}_{MCO})' [var(\hat{a}_{VI}) - var(\hat{a}_{MCO})]^{-1} (\hat{a}_{VI} - \hat{a}_{MCO})$

La statistique H est distribuée selon un χ^2 à k degré de liberté.

- Si $H < \chi^2_{\alpha}(k)$ pour un seuil α fixé, nous acceptons H_0 , l'estimateur des MCO est non biaisé.

5.11.2. Régression augmentée

La procédure proposée par Hausman est la suivante :

- Estimation d'un modèle par les MCO avec pour variable à expliquer la variable dont nous désirons tester l'exogénéité et comme variables explicatives le ou les instruments, le plus souvent les variables explicatives décalées d'une période ;
- Estimation de la ou des variables ajustées \hat{x}_{it} à partir de la ou des régressions précédentes ;
- Estimation du modèle augmenté (modèle initial dans lequel nous rajoutons la ou les variables explicatives ajustées \hat{x}_{it}) ;
- Tester la significativité du ou des coefficients de la ou des variables explicatives ajustées. Si ce ou ces coefficients ne sont pas significativement différents de zéro (test de student ou de fisher), alors nous retenons l'hypothèse $H_0: cov(x_t, \varepsilon_t) = 0$.

5.12. La méthode des moments généralisés (MMG)

Cette méthode est utilisée lorsque la ou les variables explicatives sont supposées exogènes $cov(x_t, \varepsilon_t) \neq 0$ et que, de plus, la matrice des variances-covariances des erreurs est quelconque ($E(\varepsilon\varepsilon') \neq \sigma_\varepsilon^2 I_n$).

L'estimateur des MMG combine alors la méthode des moindres carrés généralisés avec celle des variables instrumentales.

L'estimateur des MMG est donnée par :

$$\hat{a} = (X'Z(Z'\hat{\Omega}Z)^{-1}Z'X)^{-1}X'Z(Z'\hat{\Omega}Z)^{-1}Z'Y$$

avec : Y : variable à expliquer ;

X : les variables explicatives ;

Z : les instruments ;

$\hat{\Omega}$: la matrices des variances-covariances des résidus estimés dans une première étape par la méthode des variables instrumentales.

IV. VIOLATION DE L'HYPOTHESE DE NORMALITE DES ERREURS

Dans le modèle de régression linéaire multiple (Cf. Chapitre 3), nous avons pris par hypothèse que les erreurs sont distribuées normalement (pour les échantillons de petite taille).

Lorsque cette hypothèse n'est pas réalisée, il convient de procéder à la correction de l'échantillon de la manière suivante :

1. Calculer les résidus ;
2. examiner les résidus et repérer ceux dont la valeur absolue est exceptionnelle grande, par exemple les résidus dont la valeur absolue $|e_i|$ dépasse trois fois les moyennes des $|e_i|$;
3. rejeter les valeurs correspondantes ou les corriger de manière à ramener (e_i) à une valeur plus normale.

Ainsi, nous pouvons obtenir des tests plus robustes que ceux utilisés avant la correction.

Références bibliographiques

1. Brigitte Dormont, « Introduction à l'économétrie » Ed Montchrestien, Paris 1999.
2. Domadar N. Gujarati, « Econométrie ». Ed de boeck, Bruxelles, 2004.
3. Régis Bourbonnais, « Econométrie, manuel et exercices corrigés ». Ed Dunod, Paris 2015.
4. Régis Bourbonnais, « Exercices pédagogiques d'économétrie ». Ed Economica Paris 2015.
5. William Greene, « Econometric analysis ». Ed Pearson, New York, 2018.
6. William Greene, « Économétrie. Annexes : exercices et corrigés. » Ed Pearson, New York, 2006.