Introduction à l'économétrie

Méthode des moindres carrés ordinaires en présence d'hétéroscédasticité ou d'autocorrélation des erreurs – Leçon 1

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

Objectifs :

- Comprendre et étudier les conséquences de la non vérification de l'hypothèse H_4 sur la méthode d'estimation par moindres carrés.

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 :

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

Selon cette hypothèse, la variance des perturbations est constante et égale à σ^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.

Hétéroscédasticité des perturbations

Dans cette situation, les variances des erreurs ne sont plus constantes et la matrice de variances-covariances des perturbations a la forme suivante :

$$\underset{(n,n)}{\operatorname{Var} \varepsilon} = \underset{(n,n)}{\Omega} = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix} \Leftrightarrow \operatorname{Var}(\varepsilon_i) = \sigma_i^2 \quad \forall\, i = 1, \dots, n$$

On suppose donc ici qu'il y a toujours absence d'autocorrélation des erreurs mais que les variances de celles-ci varient. En notant σ^2 un paramètre strictement positif, cette matrice s'écrit :

$$\underset{(n,n)}{\operatorname{Var} \varepsilon} = \sigma^2 \begin{pmatrix} \frac{\sigma_1^2}{\sigma^2} & 0 & \dots & 0 \\ 0 & \frac{\sigma_2^2}{\sigma^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\sigma_n^2}{\sigma^2} \end{pmatrix} = \sigma^2 . \underset{(n,n)}{\Psi}$$

On peut remarquer que dans le cas d'homoscédasticité, $\sigma_i^2 = \sigma^2$, et on retrouve $\underset{(n,n)}{\Psi} = I_n$.

Exemples d'hétéroscédasticité des perturbations

Plusieurs causes peuvent être à l'origine d'une hétéroscédasticité des perturbations. Cette situation est fréquente lorsqu'on travaille sur des données en coupe transversale mais peut aussi se rencontrer sur des données longitudinales (séries temporelles).

Effet taille

Une première situation produisant cet effet est liée au fait que la variance est une mesure absolue et non relative. Imaginons que le chiffre d'affaires de l'ensemble des entreprises varie de 10%. Cette quantité de 10% représente beaucoup plus pour une grande entreprise que pour une petite.

Une autre situation caractéristique d'un effet taille est la suivante : soit une fonction de consommation à estimer sur une coupe de *n* ménages. Supposons que la spécification estimée soit la suivante :

$$C_i = \beta_0 + \beta_1 . R_i + \varepsilon_i$$

où C_i et R_i représentent respectivement la consommation et le revenu du ménage i.

La consommation des ménages à faibles revenus se caractérise par des variations d'ampleurs relativement faibles. Partant d'un niveau de consommation déjà peu élevé, ces ménages ne peuvent pas le réduire de beaucoup (proche du minimum vital, du niveau de consommation incompressible), ni l'augmenter fortement car ils ne s'attendent pas à d'importants accroissements de leurs revenus. Par conséquent, la **variance des perturbations associées aux ménages à faibles revenus sera relativement peu élevée**.

À l'inverse, les ménages à revenus élevés n'étant pas soumis aux mêmes contraintes, les **variations de leur consommation pourront être de beaucoup plus grande ampleur**. Dans une telle situation, on aura :

$$\mathrm{Var}(\varepsilon_i) = \sigma_i^2 = \sigma^2 R_i$$

La variance du terme d'erreur associé au ménage i, ε_i, est proportionnelle au revenu R_i de ce ménage.

Données en moyennes

On peut aussi rencontrer cette situation si on ne dispose que des données en moyennes et non des données individuelles pour la variable expliquée et les variables explicatives (moyennes régionales de données individuelles par exemple). Et ceci, même si les données individuelles sont homoscédastiques.

Supposons que $\mathrm{Var}(\varepsilon_{ig}) = \sigma^2$ où ε_{ig} représente la perturbation associée à l'observation individuelle i du groupe g. On ne dispose en fait que des moyennes sur chaque groupe $(g = 1, \dots, G)$. Chaque groupe g est composé de n_g données individuelles. Pour les données en moyennes on aura donc :

$$\mathrm{Var}(\bar{\varepsilon}_g) = \mathrm{Var}\left(\frac{1}{n_g}\sum_{i=1}^{n_g} \varepsilon_{ig}\right) = \frac{1}{n_g^2}\mathrm{Var}\left(\sum_{i=1}^{n_g} \varepsilon_{ig}\right) = \frac{1}{n_g^2}\sum_{i=1}^{n_g} \mathrm{Var}(\varepsilon_{ig}) = \frac{\sigma^2}{n_g}$$

La variance de la perturbation des données en moyennes dépend de la taille de chaque groupe. Si ces tailles diffèrent entre les groupes, les données en moyennes sont entachées d'un problème d'hétéroscédasticité des perturbations.

Oubli d'une variable explicative importante

Supposons que le vrai modèle explicatif du phénomène y soit :

$$y_i = \beta_0 + \beta_1.x_{1,i} + \beta_2.x_{2,i} + \varepsilon_i \quad \text{avec } \mathrm{Var}(\varepsilon_i) = \sigma^2$$

On estime un modèle en omettant la variable x_2, soit le modèle suivant :

$$y_i = \beta_0 + \beta_1.x_{1,i} + u_i$$

Dans ce cas, la perturbation du modèle estimé est $u_i = \beta_2.x_{2,i} + \varepsilon_i$ et donc, si $x_{2,i}$ n'est pas corrélé à ε_i, on aura :

$$\mathrm{Var}(u_i) = \mathrm{Var}(\beta_2.x_{2,i} + \varepsilon_i) = \beta_2^2.\mathrm{Var}(x_{2,i}) + \sigma^2 \neq \sigma^2$$

Cette situation est souvent qualifiée **d'hétéroscédasticité impure** car le vrai modèle est lui, homoscédastique. Dans ce cas, le remède au problème rencontré est de **trouver la variable explicative omise et de l'intégrer dans la spécification à estimer**. Il sera donc essentiel de toujours s'assurer de la pertinence de la spécification choisie avant de chercher à détecter et à éliminer l'hétéroscédasticité qualifiée de pure.

Données financières

Cet autre cas fréquent se rencontre sur des données financières lorsqu'on a affaire à des grappes successives de fortes et de faibles perturbations, caractérisant des périodes de forte spéculation sur les marchés, suivies de périodes d'accalmie. Dans une telle situation, la variance de la perturbation actuelle dépend de la taille de la perturbation précédente, soit :

$$\mathrm{Var}(\varepsilon_t) = \sigma^2 + \gamma.\varepsilon_{t-1}$$

Conséquences de l'hétéroscédasticité sur la méthode des MCO

Si le terme d'erreur d'une estimation est hétéroscédastique, une telle situation aura trois conséquences majeures :

- Les estimateurs des Moindres Carrés Ordinaires définis par $\hat{\beta} = (X'X)^{-1}X'y$ **restent sans biais en présence d'hétéroscédasticité des perturbations**. Ce résultat est immédiat puisque nous n'avons pas besoin d'utiliser l'hypothèse H_4 pour montrer l'absence de biais de $\hat{\beta}$,

- Par contre, puisque $\underset{(n,n)}{\text{Var}}\ \varepsilon \neq \sigma^2 . I_n$, l'expression de la matrice de variances-covariances du vecteur des paramètres estimés $\left(\underset{(k,k)}{\text{Var}}\ \hat{\beta}\ = \sigma^2 . (X'X)^{-1}\right)$ **ne permet plus d'obtenir de bons estimateurs des variances des coefficients estimés.** Cela a pour conséquence d'invalider les résultats des tests qui les utilisent (test de Student, test de Fisher). De façon typique, le biais sur le calcul des écart-types estimés des coefficients estimés est négatif, ce qui signifie que la méthode des MCO va sous-estimer les écart-types estimés et en conséquence surestimer les valeurs des statistiques de Student associées. On aura donc tendance à rejeter trop souvent l'hypothèse nulle et donc à considérer qu'une variable explicative est significative alors qu'elle ne l'est pas,
- Enfin, en présence d'hétéroscédasticité, les **estimateurs des Moindres Carrés Ordinaires ne sont plus des estimateurs BLUE** (tels qu'aucun autre estimateur linéaire des paramètres ne possède une variance inférieure)[1].

A propos des estimateurs de variances des coefficients estimés

Si $\underset{(n,n)}{\text{Var}}\ \varepsilon = \sigma^2 . \underset{(n,n)}{\Psi}$, les tests de significativité doivent être basés sur la vraie matrice de variances – covariances de l'estimateur $\hat{\beta}$, soit : $\underset{(k,k)}{\text{Var}}\ \hat{\beta}\ = \sigma^2 . (X'X)^{-1} X' \Psi X . (X'X)^{-1} \neq \sigma^2 . (X'X)^{-1}$

Remarque initiale sur l'estimation par MCO en présence d'hétéroscédasticité des perturbations

Les estimateurs des Moindres Carrés Ordinaires sont obtenus à partir de la minimisation de la somme des carrés des résidus de l'estimation. Dans la définition de cette somme, le même poids est associé au carré de chaque résidu estimé. Autrement dit, le résidu associé à chaque observation est considéré apporter la même quantité d'information pour la recherche et l'estimation de l'hyperplan de régression.

En présence d'hétéroscédasticité, une pondération égale de tous les résidus n'est pas raisonnable puisque les points caractérisés par des résidus élevés apportent relativement moins

[1] Pour une illustration de ce résultat, voir W.H. Green, Econometric Analysis, pp 543-545.

d'information utile à l'estimation que les points caractérisés par de faibles résidus. De plus, les carrés des résidus élevés vont dominer cette somme qui est à la base de la méthode des MCO.

Une procédure plus raisonnable d'estimation devrait donc se baser sur une somme pondérée des résidus estimés où :

- on pondérerait plus les points auxquels sont associés les faibles résidus ;
- on pondérerait moins les points auxquels sont associés les résidus élevés.

Cette remarque est **à la base de la conception de la méthode des Moindres Carrés Pondérés** (MCP). Il reste néanmoins à résoudre le problème de définition des pondérations ce qui nécessite la connaissance exacte de la forme prise par l'hétéroscédasticité.

Exemple d'application des moindres carrés pondérés

Supposons que l'on dispose d'un échantillon de n observations sur une variable dépendante y et trois variables explicatives notées x_1, x_2 et x_3.

Le modèle estimé par MCO peut s'écrire :

$$y_i = \beta_0 + \beta_1.x_{1,i} + \beta_2.x_{2,i} + \beta_2.x_{3,i} + \varepsilon_i$$

On sait que la perturbation ε_i souffre d'un problème d'hétéroscédasticité en lien avec la variable x_3 (effet taille) qui prend la forme suivante :

$$\mathrm{Var}(\varepsilon_i) = \sigma_i^2 = \sigma^2.x_{3,i} \qquad \forall i = 1, \dots, n$$

Compte tenu de cette dernière expression, on peut définir les poids de la façon suivante :

$$p_i = \frac{1}{\sqrt{x_{3,i}}} \qquad \forall i = 1, \dots, n$$

Ainsi, à chaque observation i on affecte un poids inversement proportionnel à la valeur prise par la variable explicative x_3 pour cette observation et donc inversement proportionnel à la variance de la perturbation qui lui est associée.

L'étape suivante consiste à transformer le modèle initial en multipliant chaque observation i par le poids p_i qui lui a été affecté. Le modèle transformé s'écrit alors :

$$p_i.y_i = p_i.\beta_0 + \beta_1.p_i.x_{1,i} + \beta_2.p_i.x_{2,i} + \beta_2.p_i.x_{3,i} + p_i.\varepsilon_i$$

$$\Leftrightarrow \quad \frac{y_i}{\sqrt{x_{3,i}}} = \beta_0.\frac{1}{\sqrt{x_{3,i}}} + \beta_1.\frac{x_{1,i}}{\sqrt{x_{3,i}}} + \beta_2.\frac{x_{2,i}}{\sqrt{x_{3,i}}} + \beta_2.\frac{x_{3,i}}{\sqrt{x_{3,i}}} + u_i$$

$$\Leftrightarrow \quad \frac{y_i}{\sqrt{x_{3,i}}} = \beta_0.\frac{1}{\sqrt{x_{3,i}}} + \beta_1.\frac{x_{1,i}}{\sqrt{x_{3,i}}} + \beta_2.\frac{x_{2,i}}{\sqrt{x_{3,i}}} + \beta_2.\sqrt{x_{3,i}} + u_i$$

$$\Leftrightarrow \quad y'_i = \beta_0.\frac{1}{\sqrt{x_{3,i}}} + \beta_1.x'_{1,i} + \beta_2.x'_{2,i} + \beta_2.\sqrt{x_{3,i}} + u_i$$

Le modèle transformé fait intervenir une nouvelle variable dépendante $y' = \frac{y}{\sqrt{x_3}}$ et deux nouvelles variable explicatives notées $x'_1 = \frac{x_1}{\sqrt{x_3}}$ et $x'_2 = \frac{x_2}{\sqrt{x_3}}$.

Calculons alors la variance de la perturbation du nouveau modèle :

$$\mathrm{Var}(u_i) = \mathrm{Var}(p_i.\varepsilon_i) = p_i^2.\mathrm{Var}(\varepsilon_i) = p_i^2.\sigma^2.x_{3,i} = \left(\frac{1}{\sqrt{x_{3,i}}}\right)^2.\sigma^2.x_{3,i} = \sigma^2 \quad \forall i = 1, \ldots, n$$

On se retrouve donc, pour le modèle transformé, dans une situation d'homoscédasticité globale des perturbations. L'estimation par MCO du modèle transformé est équivalente à l'estimation par MCG du modèle initial[2].

Il faut noter que, puisque le modèle transformé ne comporte plus de terme constant, le coefficient de détermination associé à cette dernière estimation n'est donc pas interprétable.

Détection de l'hétéroscédasticité

Problématique générale

Le terme d'erreur étant par définition non observé et inconnu, sa variance l'est aussi. Pour estimer le terme d'erreur, la démarche consiste à utiliser les résidus estimés de l'estimation par MCO. Cette démarche est cohérente car on sait que la méthode d'estimation conduira à des estimateurs sans biais, même en présence d'hétéroscédasticité des perturbations. En effet, on sait que :

$$\underset{(n,1)}{\varepsilon} = \underset{(n,1)}{y} - \underset{(n,k)}{X}.\underset{(k,1)}{\beta} \quad \text{et que :}$$

$$\underset{(n,1)}{\hat{\varepsilon}} = \underset{(n,1)}{y} - \underset{(n,1)}{\hat{y}} = \underset{(n,1)}{y} - \underset{(n,k)}{X}.\underset{(k,1)}{\hat{\beta}}$$

---

[2] Voir la section 3 de ce chapitre et W.H. Green, Econometric Analysis, pp 555-556.

On voit immédiatement que si le vecteur $\hat{\beta}$ est suffisamment proche du vecteur β, le vecteur des résidus estimé $\hat{\varepsilon}$ devrait aussi être proche du vecteur des perturbations ε. Les résidus estimés de l'estimation par MCO vont donc nous renseigner sur les perturbations, par définition inconnues.

D'autre part :

$\mathrm{Var}(\varepsilon_i) = E(\varepsilon_i^2) - [E(\varepsilon_i)]^2 = E(\varepsilon_i^2)$ puisque sous l'hypothèse H_1, $E(\varepsilon_i) = 0 \quad \forall i = 1, \ldots, n$.

Ceci poussera la majorité des auteurs à utiliser les carrés des résidus estimés pour estimer les variances des perturbations associées à chaque observation i.

Détection graphique

Il est possible de détecter une situation d'hétéroscédasticité en construisant des graphiques croisant les carrés des résidus estimés $((\hat{\varepsilon}_i)^2)$ avec les différentes variables explicatives du modèle. Les deux cas extrêmes peuvent être schématisés de la façon suivante :

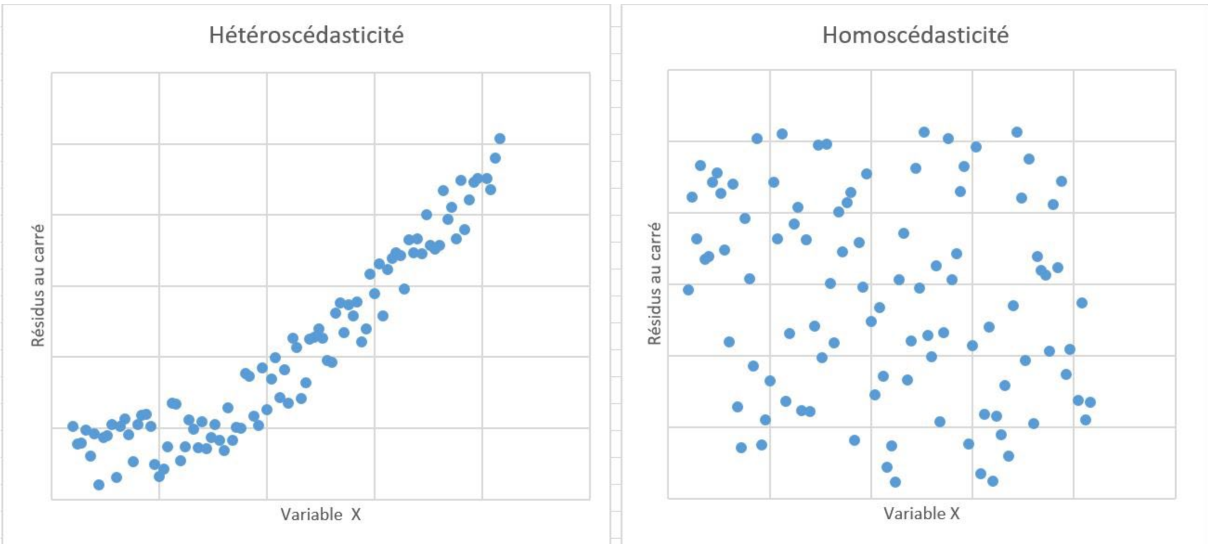

Figure 1 : détection graphique de l'hétéroscédasticité

Dans le graphique de gauche on observe que les carrés des résidus estimés augmentent avec les valeurs prises par la variable explicative considérée. On peut donc s'attendre à ce qu'il y ait un facteur de proportionnalité entre la variance de la perturbation et cette même variable.

Dans le graphique de droite, aucune liaison ne semble exister entre les carrés des résidus estimés et la variable explicative impliquée. Cela semble indiquer qu'il n'y a pas matière à rejeter l'hypothèse d'homoscédasticité dans ce cas.

Ces deux graphiques sont volontairement parfaitement lisibles et peuvent donc être considérés comme des cas d'école. Dans la pratique, la détection graphique est souvent beaucoup moins claire à analyser et nous aurons besoin d'utiliser des méthodes statistiques plus poussées pour statuer sur l'éventualité d'une situation d'hétéroscédasticité des perturbations.

Le test de White[3]

Ce test est très général et a l'avantage de pouvoir être mené même lorsqu'on ne connait pas la variable (ou les variables) à l'origine de l'hétéroscédasticité. L'hypothèse nulle testée est l'absence d'hétéroscédasticité, l'hypothèse alternative, sa présence. Le test se présente donc de la façon suivante :

$$H_0 : \ \mathrm{Var}(\varepsilon_i) = \sigma^2 \qquad \forall i = 1, \dots, n$$
$$contre\ \overline{H} : \quad \mathrm{Non}\ \ H_0$$

Supposons que le modèle soit de la forme :

$$y_i = \beta_0 + \beta_1 . x_{1,i} + \beta_2 . x_{2,i} + \varepsilon_i \qquad \forall i = 1, \dots, n$$

Volontairement nous présentons ce test sur cet exemple avec deux variables explicatives. La méthode se généralise bien entendu sans problème si le nombre de variables explicatives est plus élevé.

La procédure à adopter pour mener le test de White est la suivante :

1. Estimer le modèle avec les Moindres Carrés Ordinaires et en déduire les résidus estimés (εˆ_i),
2. Élever les résidus au carré pour obtenir les carrés des résidus estimés $(\hat{\varepsilon}_i)^2$,
3. Estimer la relation suivante :

$$\hat{\varepsilon}_i^2 = \gamma_0 + \gamma_1 . x_{1,i} + \gamma_2 . x_{2,i} + \gamma_3 . x_{1,i} . x_{2,i} + \gamma_4 . x_{1,i}^2 + \gamma_5 . x_{2,i}^2 + u_i$$

De manière générale, on régresse $\hat{\varepsilon}_i^2$ sur une constante, les variables explicatives initiales, leurs produits croisés et leurs carrés. Cette régression est souvent appelée la régression auxiliaire car elle sert un objectif statistique mais n'a pas d'intérêt économique.

4. Calculer la statistique $n.R_\varepsilon^2$ où R_ε^2 est le coefficient de détermination du modèle estimé lors de l'**étape 3**,
5. Le test est basé sur le résultat suivant :

Si H_0 est vraie (les erreurs sont homoscédastiques), alors la statistique précédente suit une loi du Khi-deux à 5 degrés de liberté, dans l'exemple présenté ici. Le degré de liberté est égal au

[3] White, H. « *A heteroscedasticity-consistent covariance matrix and a direct test for heteroscedasticity* », Econometrica, 48, 1980.

nombre de paramètres présents dans l'équation estimée à l'étape 3 moins 1, soit le nombre de régresseurs présents dans l'équation auxiliaire.

6. Décision : On ne rejette pas l'hypothèse nulle si la statistique $n.R_{\varepsilon}^{2}$ est inférieure à la valeur critique de la loi du Khi-deux, pour un risque de première espèce fixé à α.

Remarques concernant le test de White

- Comme nous l'avons précisé en introduction, le principal avantage de ce test est d'être très général et de pouvoir être mené même si on ne connait pas la cause potentielle de l'hétéroscédasticité. Ainsi, il ne fait aucune hypothèse sur la nature du problème éventuel. L'inconvénient majeur du test de White est qu'on peut le qualifier de non constructif. En effet, si à l'issue du test on rejette l'hypothèse nulle d'homoscédasticité, le test ne nous donne aucune indication sur la marche à suivre par la suite.
- La statistique utilisée, $n.R_{\varepsilon}^{2}$ est souvent notée LM pour Multiplicateur de Lagrange (Lagrange Multiplier en anglais). On aurait pu penser pouvoir utiliser un test de Fisher pour tester la nullité de l'ensemble des paramètres de pente de l'équation auxiliaire. Or, dans cette régression, la variable dépendante est le carré des résidus estimés. Si, sous les hypothèses habituelles, on peut considérer que les résidus estimés sont gaussiens, leur carré ne l'est évidemment pas, ce qui interdit l'utilisation de la loi de Fisher pour prendre une décision dans cette situation.

Le test de Breusch-Pagan[4]

Le test de Breusch-Pagan est très proche du test de White qui vient d'être présenté. Il diffère principalement par la forme de l'équation auxiliaire estimée lors de la mise en œuvre du test. Ici encore, il s'agit de tester l'hypothèse selon laquelle la variance des erreurs ne dépend pas des variables explicatives du modèle.

Soit le modèle de régression multiple :

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

[4] Breusch, T. et A. Pagan. « *A simple test for Heteroscedasticity and random coefficient variation* », Econometrica, 47, 1979.

ou encore :

$$y_i = \beta_0 + \beta_1 . x_{1,i} + \beta_2 . x_{2,i} + \cdots + \beta_{k-1} . x_{k-1,i} + \varepsilon_i \qquad \forall i = 1, \dots, n$$

L'hypothèse nulle d'homoscédasticité est : H_0 : $\mathrm{Var}(\varepsilon_i) = \sigma^2 \qquad \forall i = 1, \dots, n$

Pour tester la violation de cette hypothèse nous allons tester si les carrés des résidus estimés du modèle initial dépendent d'une ou de plusieurs variables explicatives. Si H_0 est rejetée alors on considèrera que la variance de la perturbation est une fonction des variables explicatives. La spécification la plus simple de cette fonction est linéaire, soit :

$$\mathrm{Var}(\varepsilon_i) = \gamma_0 + \gamma_1 . x_{1,i} + \gamma_2 . x_{2,i} + \cdots + \gamma_{k-1} . x_{k-1,i} + u_i$$

où la perturbation u_i vérifie toutes les hypothèses des MCO.

L'hypothèse nulle d'homoscédasticité devient H_0 : $\gamma_1 = \gamma_2 = \cdots = \gamma_{k-1} = 0$

Si, à l'exception de la constante, les paramètres γ_j sont tous nuls, alors la dispersion des erreurs ne dépend pas des variables explicatives et nous retrouvons l'hypothèse d'homoscédasticité. Comme dans le cas du test de White, nous procéderons à un test du multiplicateur de Lagrange pour tester la nullité des paramètres de l'équation auxiliaire.

La procédure à adopter pour mener le test de Breusch-Pagan est la suivante :

1. Estimer le modèle initial avec les Moindres Carrés Ordinaires et en déduire les résidus estimés $(\hat{\varepsilon}_i)$,

2. Élever les résidus au carré pour obtenir les carrés des résidus estimés $(\hat{\varepsilon}_i)^2$,

3. Estimer la relation suivante :

$$\hat{\varepsilon}_i^2 = \gamma_0 + \gamma_1 . x_{1,i} + \gamma_2 . x_{2,i} + \cdots + \gamma_{k-1} . x_{k-1,i} + u_i$$

On régresse donc $\hat{\varepsilon}_i^2$ sur une constante et l'ensemble des variables explicatives initiales. Ici encore, cette régression est appelée la régression auxiliaire.

4. Calculer la statistique $n. R_\varepsilon^2$ où R_ε^2 est le coefficient de détermination du modèle estimé lors de l'**étape 3**,

5. Le test est basé sur le résultat suivant :

Si H_0 est vraie (les erreurs sont homoscédastiques), alors la statistique précédente suit une loi du Khi-deux à $(k-1)$ degrés de liberté.

6. Décision : On ne rejette pas l'hypothèse nulle si la statistique $n.R_\varepsilon^2$ est inférieure à la valeur critique de la loi du Khi-deux, pour un risque de première espèce fixé à α.

Le test de Goldfeld et Quandt[5]

Le test de Goldfeld et Quandt est basé sur l'hypothèse que les observations peuvent être scindées en deux groupes de telle sorte que sous l'hypothèse d'homoscédasticité les variances des perturbations seraient identiques dans les deux groupes alors que sous l'hypothèse alternative d'hétéroscédasticité les variances des perturbations seraient significativement différentes entre les deux sous-échantillons.

On utilisera le test de Goldfeld et Quandt principalement dans le cas où la théorie permet d'envisager une hétéroscédasticité causée par une des variables explicatives du modèle.

La procédure à adopter pour mener le test de Goldfeld et Quandt est la suivante :

1. Ordonner les observations en fonction d'une variable explicative X_i dans un ordre croissant ou décroissant de celle-ci. S'il existe une situation d'hétéroscédasticité en fonction de cette variable X_i, cela revient à classer les observations par ordre croissant ou décroissant des variances des perturbations.

2. Éliminer c observations centrales dans l'échantillon ainsi classé, observations pour lesquelles les variances seraient de moyenne importance. Le nombre c d'observations à omettre doit environ représenter le quart ou le cinquième de l'échantillon initial. On obtient ainsi deux échantillons contenant l'un, les valeurs élevées de X_i et l'autre les faibles valeurs de cette variable. On s'arrangera pour que les deux échantillons ainsi constitués soient de taille égale à $\frac{(n-c)}{2}$.

[5] Goldfeld, S. et R. Quandt. « *Some tests for homoscedasticity* », Journal of the American Statistical Association, 60, 1965.

3. Effectuer la régression envisagée pour chacun des deux échantillons ainsi obtenus et retenir la somme des carrés des résidus estimés pour chacun d'entre eux. On note ces quantités SCR_1 et SCR_2 où $SCR_i = \sum_{j=1}^{\frac{(n-c)}{2}} \hat{\varepsilon}_{ij}^2 \quad$ avec $i = 1{,}2$.

4. Si les erreurs étaient homoscédastiques, les variances des erreurs seraient les mêmes dans les deux sous-échantillons et donc leur rapport serait égal à 1. L'idée centrale de ce test est donc d'estimer ce rapport par le rapport des sommes des carrés des résidus et de voir si ce dernier rapport s'éloigne significativement de 1. Dans un tel cas, l'hypothèse d'homoscédasticité sera rejetée.

En notant $\mathrm{Var}(\varepsilon_1)$ et $\mathrm{Var}(\varepsilon_2)$ les variances des perturbations dans les deux groupes d'observations, on estime le rapport $\frac{\mathrm{Var}(\varepsilon_1)}{\mathrm{Var}(\varepsilon_2)} = \frac{\sigma_1^2}{\sigma_2^2}$ par le rapport des variances estimées soit :

$$\widehat{\left(\frac{\sigma_1^2}{\sigma_2^2}\right)} = \frac{\widehat{\sigma_1}^2}{\widehat{\sigma_2}^2} = \frac{\dfrac{SCR_1}{\frac{(n-c)}{2} - k}}{\dfrac{SCR_2}{\frac{(n-c)}{2} - k}} = \frac{SCR_1}{SCR_2}$$

5. Pour prendre une décision, on utilise le résultat suivant :

Si H_0 est vraie (Homoscédasticité), la statistique :

$$f = \frac{SCR_1}{SCR_2} = \frac{\sum_{j=1}^{\frac{(n-c)}{2}} \hat{\varepsilon}_{1j}^2}{\sum_{j=1}^{\frac{(n-c)}{2}} \hat{\varepsilon}_{2j}^2} \rightarrow F\left(\frac{(n-c)}{2} - k, \frac{(n-c)}{2} - k\right)$$

Pour un risque de première espèce égal à α, on rejette l'hypothèse d'homoscédasticité si $f > f_\alpha^*$, où f_α^* est la valeur critique de la loi de Fisher considérée.

Remarque

Il y a forcément une somme des carrés des résidus estimés supérieure à l'autre. Lorsqu'on construit la statistique *f*, il faut mettre au numérateur la plus grande des deux valeurs de sorte que *f* soit supérieure à 1. Dans le cas contraire ($f < 1$), la valeur calculée ne permettra jamais de rejeter H_0 (la plus petite valeur que peut prendre une variable de Fisher est égale à 1. Cela correspond au cas où les deux degrés de liberté sont infinis).

Le test de Park[6]

Le test de Park formalise une situation d'hétéroscédasticité dans laquelle la variance de la perturbation a la forme suivante :

$$\mathrm{Var}(\varepsilon_i) = \sigma_i^2 = \sigma^2 . X_i^{\beta} . e^{u_i}$$

soit :

$$\ln \sigma_i^2 = \ln \sigma^2 + \beta \ln X_i + u_i$$

où X est une des variables explicatives du modèle et u une perturbation sphérique.

Comme dans le cas des tests de White et de Breusch-Pagan, puisque la variance des perturbations est en général inconnue, Park recommande d'utiliser les carrés des résidus estimés, obtenus lors de l'estimation du modèle de départ, comme proxy de ce paramètre. Par contre, la forme log-log de l'équation auxiliaire permet d'envisager d'utiliser un test de Student pour prendre la décision quant à la présence ou non de l'hétéroscédasticité des perturbations.

On estime donc l'équation auxiliaire sous la forme :

$$\ln \hat{\varepsilon}_i^2 = \ln \sigma^2 + \beta \ln X_i + u_i = \alpha + \beta \ln X_i + u_i$$

Si le paramètre β s'avère être significativement différent de zéro, cela traduit une situation d'hétéroscédasticité dans les données. Si par contre, β est non significativement différent de zéro, on peut accepter l'hypothèse nulle d'homoscédasticité. Comme pour les deux tests précités, le test de Park est un test à **deux étapes** :

- Dans la première étape on estime le modèle initial pour obtenir les résidus estimés ;
- Dans la seconde, on utilise ces résidus pour estimer l'équation auxiliaire et prendre sur cette base une décision.

Une variante de ce test propose de choisir comme variable explicative de l'équation auxiliaire, non pas une des variables initiales, mais un **facteur de proportionnalité inhérent aux données étudiées.** Sur des données en coupes transversales, un tel facteur serait celui mesurant le mieux la taille des observations. Sur des séries temporelles, ce facteur serait le temps ou encore le numéro de l'observation.

[6] Park, R. « *Estimation with heteroscedastic error terms* », Econometrica, 34,1966.

Bien qu'empiriquement assez attirant, le test de Park rencontre certains problèmes. Goldfeld et Quandt ont montré que la perturbation u_i de l'équation auxiliaire pouvait ne pas satisfaire les hypothèses des Moindres carrés ordinaires et en particulier pouvait, elle-même, être hétéroscédastique[7] . Néanmoins, il est tout à fait possible d'utiliser ce test en tant que méthode strictement exploratoire.

Le test de Glesjer[8]

Le test de Glesjer est dans l'esprit assez similaire au test de Park.

Après avoir obtenu les résidus estimés de la régression initiale par moindres carrés ordinaires, Glesjer propose de régresser la valeur absolue de ces résidus estimés $(\hat{\varepsilon}_i)$ sur la variable explicative (X) qui est supposée être liée à la variance de la perturbation et donc être la cause de l'hétéroscédasticité. L'avantage de ce test est qu'il **permet, non seulement de déceler une éventuelle hétéroscédasticité, mais aussi d'identifier la forme qu'elle revêt**. Lors de ses expérimentations, Glesjer suggère de tester les différentes formes fonctionnelles suivantes et en déduit chaque fois, la forme supposée prise par l'hétéroscédasticité :

$$(1) \qquad |\hat{\varepsilon}_i| = \alpha + \beta . X_i + u_i \qquad \Rightarrow \qquad \mathrm{Var}(\varepsilon_i) = \sigma_i^2 = k^2 X_i^2$$

$$(2) \qquad |\hat{\varepsilon}_i| = \alpha + \beta . \sqrt{X_i} + u_i \qquad \Rightarrow \qquad \mathrm{Var}(\varepsilon_i) = \sigma_i^2 = k^2 X_i$$

$$(3) \qquad |\hat{\varepsilon}_i| = \alpha + \beta . \frac{1}{X_i} + u_i \qquad \Rightarrow \qquad \mathrm{Var}(\varepsilon_i) = \sigma_i^2 = k^2 X_i^{-2}$$

$$(4) \qquad |\hat{\varepsilon}_i| = \alpha + \beta . \frac{1}{\sqrt{X_i}} + u_i \qquad \Rightarrow \qquad \mathrm{Var}(\varepsilon_i) = \sigma_i^2 = k^2 X_i^{-1}$$

où u_i est une perturbation sphérique. Dans chacun des cas, l'hypothèse d'homoscédasticité est rejetée si le coefficient β est significativement différent de zéro.

La principale critique que l'on peut adresser au test de Glesjer, ainsi qu'au test de Park d'ailleurs, est qu'il est nécessaire de connaître a priori la variable explicative à l'origine de l'hétéroscédasticité. De plus, Goldfeld et Quandt ont montré que la perturbation u_i n'était probablement pas d'espérance nulle et de variance non constante[9].

[7] Goldfeld, S. et R. Quandt. *Nonlinear Methods in Econometrics*, North Holland Publishing Company, Amsterdam, 1972, pp. 93–94
[8] Glesjer, H. « *A new test for heteroscedasticity* », Journal of the American Statistical Association, 64, 1969.
[9] Pour des précisions, voir Goldfeld, S. et R. Quandt., op. cit., chapitre 3.

Pourtant Glesjer montre que, sur de grands échantillons, les quatre spécifications ci-dessus sont relativement efficaces pour détecter l'hétéroscédasticité des perturbations.

Le test A.R.C.H[10]

Les modèles de type ARCH (AutoRegressive Conditional Heteroscedasticity) permettent de **modéliser des séries temporelles qui ont une volatilité qui dépend du passé**, comme beaucoup de séries financières. On constate en effet pour ce type de séries des successions de périodes de forte variabilité et de périodes d'accalmie.

Considérons le modèle linéaire multiple habituel :

$$\underset{(n,1)}{y} = \underset{(n,k)}{X} . \underset{(k,1)}{\beta} + \underset{(n,1)}{\varepsilon} \qquad \text{(A)}$$

On soupçonne que les erreurs sont hétéroscédastiques de la forme :

$$\text{Var}(\varepsilon_i) = \sigma_i^2 = \sigma^2 + \sum_{j=1}^{p} a_j \varepsilon_{i-j}^2$$

La procédure à adopter pour mener le test A.R.C.H est la suivante

1. Estimer le modèle avec les Moindres Carrés Ordinaires et en déduire les résidus estimés $(\hat{\varepsilon}_i)$,
2. Élever les résidus au carré pour obtenir les carrés des résidus estimés $(\hat{\varepsilon}_i)^2$,
3. Effectuer la régression autorégressive des carrés des résidus estimés sur ses p retards (résidus décalés), où seuls les retards significatifs sont conservés :

$$\hat{\varepsilon}_i^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j \hat{\varepsilon}_{i-j}^2 + u_i \qquad \text{(B)}$$

4. Calculer la statistique LM (multiplicateur de Lagrange) avec $LM = m. R^2$ où :

m : Nombre d'observations servant à l'estimation de la régression (B)

R^2 : Coefficient de détermination obtenu lors de l'estimation réalisée à **l'étape 3**.

[10] Engel, R. « *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations* », Econometrica, 50,1982.

5. Si LM est supérieure à la valeur critique d'une variable du Khi-deux à p degrés de liberté, pour un risque de première espèce α, on rejette l'hypothèse nulle d'homoscédasticité et on considère donc que le processus est justifiable d'un modèle ARCH d'ordre p.

Ce sont les tests de significativité des coefficients α_j de la régression (B) qui permettent de déterminer l'ordre p du processus ARCH, sachant qu'un processus ARCH d'ordre 3 semble être un maximum.

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/).

Table des illustrations

Figures