

Introduction à l'économétrie

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

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

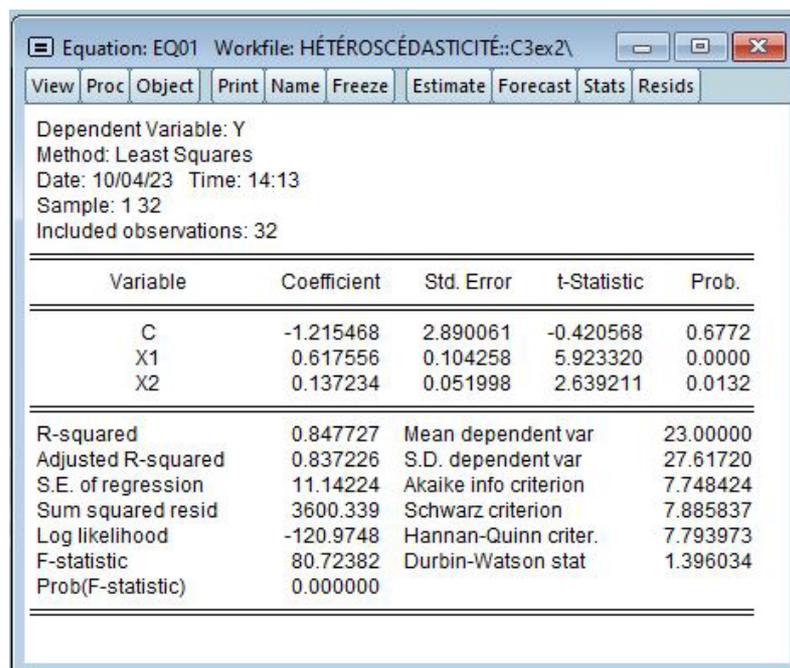
Exemple d'application sur l'hétéroscédasticité

Inspiré de Bourbonnais, R. « Exercices pédagogiques d'économétrie », *Economica*, 2015.

On dispose de 32 observations sur une variable dépendante Y et deux variables explicatives X_1 et X_2 . Nous cherchons à estimer la relation suivante :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \quad i = 1, \dots, 32$$

L'estimation du modèle précédent par la méthode des Moindres Carrés Ordinaires conduit aux résultats suivants :



Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.215468	2.890061	-0.420568	0.6772
X1	0.617556	0.104258	5.923320	0.0000
X2	0.137234	0.051998	2.639211	0.0132

R-squared	0.847727	Mean dependent var	23.00000
Adjusted R-squared	0.837226	S.D. dependent var	27.61720
S.E. of regression	11.14224	Akaike info criterion	7.748424
Sum squared resid	3600.339	Schwarz criterion	7.885837
Log likelihood	-120.9748	Hannan-Quinn criter.	7.793973
F-statistic	80.72382	Durbin-Watson stat	1.396034
Prob(F-statistic)	0.000000		

Figure 1 : estimation avec la méthode des MCO réalisée avec le logiciel Eviews

Les résultats de cette estimation sont assez satisfaisants. Le modèle explique plus de 84% de la variance de la variable dépendante. Les deux variables explicatives influencent significativement la variable dépendante (les *p-values* des tests de Student sont respectivement de 0 et 0,013 pour les coefficients β_1 et β_2).

On **soupçonne cependant qu'une situation d'hétéroscédasticité des perturbations puisse être présente** dans les données utilisées ce qui pourrait invalider les résultats des tests précédents. Plus précisément, on soupçonne que la variance du terme d'erreur est proportionnelle à l'une des deux variables explicatives incluses dans le modèle (effet taille).

Détection de l'hétéroscédasticité

Détection visuelle

Ne connaissant pas la variable explicative pouvant être la cause du problème, nous pouvons demander des graphiques croisant les résidus estimés et/ou leurs carrés avec les deux variables explicatives incluses dans le modèle.

Pour les graphiques croisant les résidus estimés avec les variables explicatives, il faut dans un premier temps classer les observations par valeurs croissantes de la variable explicative incriminée. On obtient les représentations suivantes :

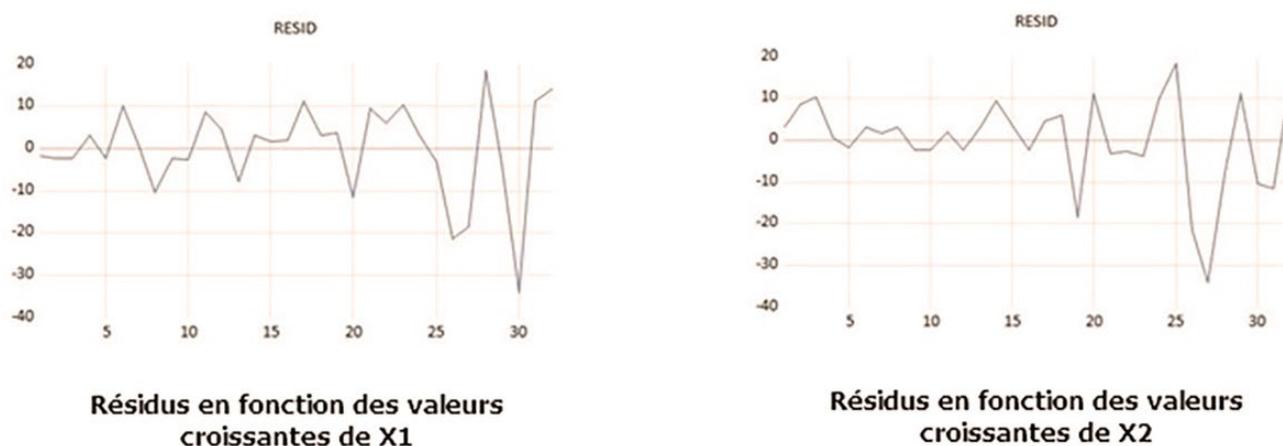


Figure 2 : détection visuelle de l'hétéroscédasticité

Dans un cas comme dans l'autre on observe une variabilité des résidus estimés qui augmente avec les valeurs prises par la variable explicative considérée.

On peut aussi demander des graphiques croisant les carrés des résidus estimés avec les deux variables explicatives du modèle. On obtient :

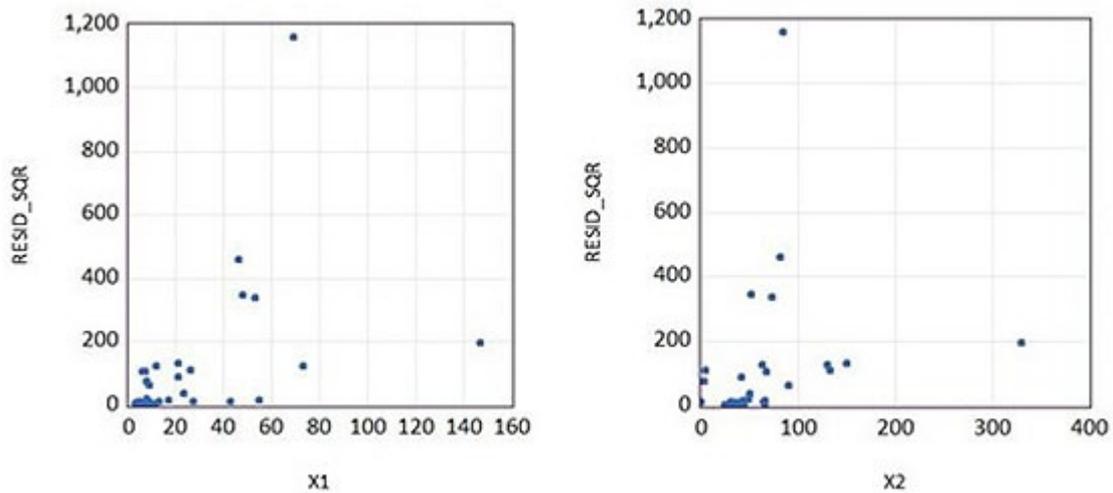


Figure 3 : représentation graphique des carrés des résidus croisés avec les deux variables

Ici encore, on observe une relation positive entre les carrés des résidus issus de l'estimation et les valeurs prises par les deux variables explicatives. Par contre, aucun élément ne permet de choisir la variable éventuellement à la source du problème potentiel d'hétéroscédasticité des perturbations. L'utilisation des tests classiques de détection de l'hétéroscédasticité devrait permettre de lever le doute qui subsiste à l'issue de cette analyse graphique.

Test de White

Comme nous l'avons dit au cours de la présentation détaillée, le test de White est très général et a l'avantage de pouvoir être mené même lorsqu'on ne connaît 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.

Pour rappel on régresse le carré des résidus estimés sur une constante, les variables explicatives du modèle, leurs carrés et leurs produits croisés. Le coefficient de détermination de cette régression auxiliaire permet de calculer la statistique LM permettant de prendre une décision. Dans l'exemple étudié, on obtient les résultats ci-dessous.

La conclusion du test de White est assez claire car, dans la partie supérieure de la page de résultats, on lit clairement le rejet de l'hypothèse nulle d'homoscédasticité. En notant R_{ε}^2 le coefficient de détermination de l'équation auxiliaire, la statistique $LM = n \cdot R_{\varepsilon}^2 = 32 * 0,5645 = 18,06$ qu'il faut comparer avec la valeur critique d'une variable du Khi-deux à 5 degrés de

liberté, qui pour un risque de première espèce égal à 5% vaut 11,07. La valeur calculée est largement supérieure à la valeur critique et la *p-value* du test est de 0,003.

D'autre part, on remarque que parmi les variables explicatives incluses dans la régression auxiliaire, seules deux sont significatives : X_1^2 et le produit croisé $X_1 \cdot X_2$. **La variable X_1 semble être la cause de l'hétéroscédasticité. Le test de Breusch-Pagan va confirmer cette première impression.**

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-10.14736	70.76396	-0.143397	0.8871
X1^2	0.392854	0.131929	2.977773	0.0062
X1*X2	-0.197178	0.069063	-2.855066	0.0083
X1	-3.410120	5.200760	-0.655696	0.5178
X2^2	0.011104	0.017214	0.645069	0.5245
X2	1.791954	2.143927	0.835828	0.4109

R-squared	0.564526	Mean dependent var	112.5106
Adjusted R-squared	0.480781	S.D. dependent var	221.8420
S.E. of regression	159.8523	Akaike info criterion	13.15374
Sum squared resid	664371.8	Schwarz criterion	13.42856
Log likelihood	-204.4598	Hannan-Quinn criter.	13.24484
F-statistic	6.741016	Durbin-Watson stat	1.543493
Prob(F-statistic)	0.000377		

Figure 4 : test de White réalisé avec le logiciel Eviews

Test de Breusch-Pagan

Dans ce test, on régresse le carré des résidus estimés sur une constante et l'ensemble des variables explicatives du modèle. La décision se prend sur la base d'un test du multiplicateur de Lagrange comme pour le test de White.

Les résultats sont fournis ci-dessous :

Equation: EQ01 Workfile: HÉTÉROSCÉDASTICITÉ::C3ex2\									
View	Proc	Object	Print	Name	Freeze	Estimate	Forecast	Stats	Resids
Heteroskedasticity Test: Breusch-Pagan-Godfrey									
Null hypothesis: Homoskedasticity									
F-statistic	5.206175	Prob. F(2,29)	0.0117						
Obs*R-squared	8.454081	Prob. Chi-Square(2)	0.0146						
Scaled explained SS	13.07510	Prob. Chi-Square(2)	0.0014						
Test Equation:									
Dependent Variable: RESID^2									
Method: Least Squares									
Date: 10/07/23 Time: 09:54									
Sample: 1 32									
Included observations: 32									
Variable	Coefficient	Std. Error	t-Statistic	Prob.					
C	52.70796	51.03201	1.032841	0.3102					
X1	5.261578	1.840971	2.858045	0.0078					
X2	-1.136922	0.918172	-1.238245	0.2256					
R-squared	0.264190	Mean dependent var	112.5106						
Adjusted R-squared	0.213445	S.D. dependent var	221.8420						
S.E. of regression	196.7471	Akaike info criterion	13.49078						
Sum squared resid	1122573.	Schwarz criterion	13.62819						
Log likelihood	-212.8524	Hannan-Quinn criter.	13.53632						
F-statistic	5.206175	Durbin-Watson stat	1.865854						
Prob(F-statistic)	0.011698								

Figure 5 : test de Breusch Pagan réalisé avec le logiciel Eviews

Ici encore, l'hypothèse d'homoscédasticité est rejetée et l'unique variable explicative significative de la régression auxiliaire est la variable X_1 . Pour finir, nous allons **confirmer cette conclusion à l'aide du test de Goldfeld et Quandt**.

Test de Goldfeld et Quandt

La première étape de ce test consiste à classer les observations en fonction des valeurs croissantes de la variable supposée être à l'origine de l'hétéroscédasticité, X_1 ici. S'il existe une situation d'hétéroscédasticité en fonction de cette variable, cela revient à classer les observations par ordre croissant des variances des perturbations. On élimine ensuite les observations centrales de ce fichier préalablement classé. Le nombre d'observations à éliminer doit représenter entre un quart et un tiers de l'échantillon complet. Compte tenu de la faible taille de l'échantillon utilisé dans cet exemple, nous choisissons d'éliminer les huit observations centrales (sur les 32 composant l'échantillon complet).

On réalise ensuite l'estimation du modèle sur les deux groupes d'observations afin de récupérer les sommes des carrés des résidus estimés. La statistique du test est construite à partir de ces deux sommes.

Les estimations réalisées sur les deux groupes sont résumées dans les tableaux ci-dessous :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.848678	4.496956	-0.633468	0.5422
X1	1.700839	0.746800	2.277503	0.0488
X2	0.036222	0.042765	0.847001	0.4189

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.800906	11.43787	0.769453	0.4613
X1	-0.115534	0.488693	-0.236413	0.8184
X2	0.465379	0.201141	2.313693	0.0460

Figure 6 : test de Goldfeld et Quandt réalisé avec le logiciel Eviews

Les résultats fournissent : $SCR_1 = 204,85$ et $SCR_2 = 2216,41$.

Pour rappel, le test de Goldfeld et Quandt est basé sur le résultat suivant :

Si H_0 est vraie (Homoscédasticité), la statistique :

$$f = \frac{SCR_2}{SCR_1} \rightarrow F(9; 9)$$

La valeur calculée de la statistique f est $f = \frac{2216,41}{204,85} = 10,82$ alors que pour un risque de première espèce égal à 5%, la valeur critique de la loi de Fisher considérée est égale à $f_{0,05}^* = 3,18$. On rejette donc l'hypothèse nulle d'homoscédasticité et on peut donc conclure que la **variable X_1 est la source de l'hétéroscédasticité.**

Correction de l'hétéroscédasticité

On vient de voir que la variable X_1 est la source de l'hétéroscédasticité dans les données étudiées. Compte tenu des résultats du test de white, on peut envisager une hétéroscédasticité prenant la forme suivante :

$$\text{Var}(\varepsilon_i) = \sigma_i^2 = \sigma^2 \cdot X_{1i}^2$$

La correction du problème rencontré passe par la pondération des observations, en pondérant chaque observation i par poids inversement proportionnel à la valeur prise par la variable explicative X_1 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. Compte tenu de la forme retenue, les poids p_i seront définis de la façon suivante :

$$p_i = \frac{1}{X_{1i}}$$

Le modèle transformé s'écrit alors :

$$p_i \cdot Y_i = p_i \cdot \beta_0 + \beta_1 \cdot p_i \cdot X_{1i} + \beta_2 \cdot p_i \cdot X_{2i} + p_i \cdot \varepsilon_i$$

$$\Leftrightarrow \frac{Y_i}{X_{1i}} = \beta_0 \cdot \frac{1}{X_{1i}} + \beta_1 \cdot \frac{X_{1i}}{X_{1i}} + \beta_2 \cdot \frac{X_{2i}}{X_{1i}} + u_i$$

$$\Leftrightarrow \frac{Y_i}{X_{1i}} = \beta_0 \cdot \frac{1}{X_{1i}} + \beta_1 + \beta_2 \cdot \frac{X_{2i}}{X_{1i}} + u_i$$

$$\Leftrightarrow Y_i^* = \beta_0 \cdot X_{1i}^* + \beta_1 + \beta_2 \cdot X_{2i}^* + u_i$$

Le modèle transformé fait intervenir une nouvelle variable dépendante $Y^* = \frac{Y}{X_1}$ et deux nouvelles variables explicatives notées $X_1^* = \frac{1}{X_1}$ et $X_2^* = \frac{X_2}{X_1}$

Calculons alors la variance de la perturbation du nouveau modèle :

$$\text{Var}(u_i) = \text{Var}(p_i \cdot \varepsilon_i) = p_i^2 \cdot \text{Var}(\varepsilon_i) = p_i^2 \cdot \sigma^2 \cdot X_{1i}^2 = \left(\frac{1}{X_{1i}}\right)^2 \cdot \sigma^2 \cdot X_{1i}^2 = \sigma^2 \quad \forall i = 1, \dots, n$$

On se retrouve donc, pour le modèle transformé, **dans une situation d'homoscédasticité globale des perturbations.**

Une fois les nouvelles variables créées, l'estimation du modèle transformé par Moindres Carrés Ordinaires donne les résultats suivants :

Equation: EQ02 Workfile: HÉTÉROSCÉDASTICITÉ::C3ex2\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: Y_STAR
 Method: Least Squares
 Date: 10/07/23 Time: 14:41
 Sample: 1 32
 Included observations: 32

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X1_STAR	-0.330784	1.435183	-0.230482	0.8193
C	0.914408	0.160201	5.707873	0.0000
X2_STAR	0.057468	0.030086	1.910142	0.0660

R-squared	0.155297	Mean dependent var	1.153132
Adjusted R-squared	0.097041	S.D. dependent var	0.600504
S.E. of regression	0.570624	Akaike info criterion	1.804887
Sum squared resid	9.442737	Schwarz criterion	1.942300
Log likelihood	-25.87819	Hannan-Quinn criter.	1.850435
F-statistic	2.665790	Durbin-Watson stat	1.918456
Prob(F-statistic)	0.086539		

Figure 7 : estimation du modèle transformé par Moindres Carrés Ordinaires-logiciel Eviews

Le logiciel Eviews permet bien entendu d'appliquer les Moindres Carrés Pondérés au modèle initial en spécifiant précisément les pondérations utilisées. Les résultats sont donnés ci-dessous :

Equation: WLS Workfile: HÉTÉROSCÉDASTICITÉ::C3ex2\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: Y
 Method: Least Squares
 Date: 10/07/23 Time: 14:38
 Sample: 1 32
 Included observations: 32
 Weighting series: 1/X1
 Weight type: Inverse standard deviation (EViews default scaling)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.330784	1.435183	-0.230482	0.8193
X1	0.914408	0.160201	5.707873	0.0000
X2	0.057468	0.030086	1.910142	0.0660

Weighted Statistics

R-squared	0.184982	Mean dependent var	11.00009
Adjusted R-squared	0.128773	S.D. dependent var	5.728402
S.E. of regression	5.443364	Akaike info criterion	6.315732
Sum squared resid	859.2763	Schwarz criterion	6.453144
Log likelihood	-98.05171	Hannan-Quinn criter.	6.361280
F-statistic	3.291009	Durbin-Watson stat	1.918456
Prob(F-statistic)	0.051514	Weighted mean dep.	2.852650

Unweighted Statistics

R-squared	0.785230	Mean dependent var	23.00000
Adjusted R-squared	0.770418	S.D. dependent var	27.61720
S.E. of regression	13.23271	Sum squared resid	5078.030
Durbin-Watson stat	1.331619		

Figure 8 : résultats des Moindres Carrés Pondérés appliquées au modèle initial-logiciel Eviews

Les résultats sont identiques et suggèrent le commentaire suivant : la correction de l'hétéroscédasticité a dégradé l'estimation des paramètres du modèle puisque le coefficient β_2 n'est plus significativement différent de zéro au seuil de 5% (la nouvelle *p-value* est égale à 6,6% contre 1,3% dans le modèle de départ).

On peut enfin s'assurer que l'hétéroscédasticité des perturbations n'est plus présente dans le modèle transformé en effectuant un test de White à partir des résidus estimés issu de cette dernière estimation. On obtient :

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	1.148832	Prob. F(5,26)	0.3604
Obs*R-squared	5.790454	Prob. Chi-Square(5)	0.3271
Scaled explained SS	17.89816	Prob. Chi-Square(5)	0.0031

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 10/07/23 Time: 14:46
Sample: 1 32
Included observations: 32

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.028302	0.362541	-0.078065	0.9384
X1_STAR^2	-59.03336	36.89652	-1.599971	0.1217
X1_STAR*X2_STAR	1.514905	1.104650	1.371388	0.1820
X1_STAR	8.362010	6.799791	1.229745	0.2298
X2_STAR^2	-0.005500	0.006161	-0.892749	0.3802
X2_STAR	-0.068110	0.177954	-0.382742	0.7050

R-squared	0.180952	Mean dependent var	0.295086
Adjusted R-squared	0.023442	S.D. dependent var	0.822540
S.E. of regression	0.812842	Akaike info criterion	2.590799
Sum squared resid	17.17849	Schwarz criterion	2.865625
Log likelihood	-35.45279	Hannan-Quinn criter.	2.681896
F-statistic	1.148832	Durbin-Watson stat	2.397338
Prob(F-statistic)	0.360364		

Figure 9 : test de White dans le modèle transformé-logiciel Eviews

Le test de White conduit clairement à retenir l'**hypothèse nulle d'homoscédasticité des perturbations pour le modèle transformé**.

Références

Comment citer ce cours ?

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



Cette œuvre est mise à disposition dans le respect de la législation française protégeant le droit d'auteur, selon les termes du contrat de licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). En cas de conflit entre la législation française et les termes de ce contrat de licence, la clause non conforme à la législation française est réputée non écrite. Si la clause constitue un élément déterminant de l'engagement des parties ou de l'une d'elles, sa nullité emporte celle du contrat de licence tout entier.

Table des illustrations

Figures

Figure 2 : estimation avec la méthode des MCO réalisée avec le logiciel Eviews	1
Figure 3 : détection visuelle de l'hétéroscédasticité	2
Figure 4 : représentation graphique des carrés des résidus croisés avec les deux variables	3
Figure 5 : test de White réalisé avec le logiciel Eviews	4
Figure 6 : test de Breusch Pagan réalisé avec le logiciel Eviews.....	5
Figure 7 : test de Goldfeld et Quandt réalisé avec le logiciel Eviews	6
Figure 8 : estimation du modèle transformé par Moindres Carrés Ordinaires-logiciel Eviews.....	8
Figure 9 : résultats des Moindres Carrés Pondérés appliquées au modèle initial-logiciel Eviews ...	8
Figure 10 : test de White dans le modèle transformé-logiciel Eviews	9