

Hétéroscédasticité

GSF-6053

Hiver 2025

Table des matières

1	Hétéroscédasticité	2
1.1	Hétéroscédasticité groupée	2
1.2	Variance varie en fonction d'une variable exogène	2
1.3	Variance varie en fonction de plusieurs variables exogènes	3
2	Tests Diagnostiques d'Hétéroscédasticité	4
2.1	Le test de Goldfeld-Quandt (GQ)	4
2.2	Le test de Breusch-Pagan (BP)	5
2.3	Le test de White	5
2.4	Le correcteur de White	6
3	Conclusion	7

1 Hétéroscédasticité

wooldridge2010

Comme on connaît rarement Ω , il faut spécifier une forme pour cette matrice qui est gérable. Par exemple :

$$\Omega = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ \vdots & \vdots & \ddots & \sigma_T^2 \end{bmatrix}$$

n'est pas une forme estimable, car le nombre de paramètres à estimer $T + K > T$.

Il faut donc spécifier un modèle (une paramétrisation) pour la forme de la matrice Ω afin de réduire le nombre de paramètres inconnus.

1.1 Hétéroscédasticité groupée

On a une variance hétéroscédastique, mais en sous-groupes, c'est-à-dire qu'à l'intérieur de chaque sous-groupe, la variance est constante.

Par exemple :

$$\begin{aligned} \sigma_t^2 &= \sigma_{10}^2 \quad \text{pour } t = 1, \dots, T_1 \\ \sigma_t^2 &= \sigma_{20}^2 \quad \text{pour } t = T_1 + 1, \dots, T \end{aligned}$$

$$\Omega = \begin{bmatrix} \sigma_{10}^2 I_{T_1} & 0 & \dots & 0 \\ 0 & \sigma_{20}^2 I_{T_2} & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \end{bmatrix}$$

Dans un cas plus général :

$$\Omega = \begin{bmatrix} \sigma_{10}^2 I_{T_1} & 0 & \dots & 0 \\ 0 & \sigma_{20}^2 I_{T_2} & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \end{bmatrix}$$

1.2 Variance varie en fonction d'une variable exogène

Soit le modèle suivant :

$$\begin{aligned} Y &= X\beta + u \\ E(uu') &= \sigma^2 \Omega \end{aligned}$$



$$V(u_t) = \sigma^2 Z_t$$

Dans [wooldridge2010](#), ce cas revient à une estimation où l'hétéroscédasticité est connue à une constante multiplicative près, où Z_t est connu et observé (ATTN : connu et observé \neq estimé). C'est le seul cas où on fait un GLS et non un FGLS, car Ω est vraiment connu.

$$\Omega = \begin{bmatrix} Z_1 & 0 & 0 \\ 0 & Z_2 & 0 \\ \vdots & \vdots & \ddots \end{bmatrix}$$

$$\Omega^{-1} = \begin{bmatrix} 1/Z_1 & 0 & 0 \\ 0 & 1/Z_2 & 0 \\ \vdots & \vdots & \ddots \end{bmatrix}$$

Le modèle transformé est donc :

$$\left(\frac{Y_1}{\sqrt{Z_1}}, \frac{Y_2}{\sqrt{Z_2}}, \dots, \frac{Y_T}{\sqrt{Z_T}} \right) = \begin{bmatrix} \frac{1}{\sqrt{Z_1}} & \frac{1}{\sqrt{Z_2}} & \dots \end{bmatrix} X\beta + \text{résidus}$$

1.3 Variance varie en fonction de plusieurs variables exogènes

Soit le modèle suivant :

$$Y_t = \beta_0 + \beta_1 X_{1t} + \dots + \beta_{K-1} X_{K-1,t} + u_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 Z_{1t} + \dots + \alpha_H Z_{Ht} + e_t$$

Cette paramétrisation de la variance demande d'estimer $\hat{\sigma}_t^2$ par une régression artificielle. L'approche est la suivante :

1. On utilise les résidus au carré, \hat{u}_t^2 , pour avoir une proxy de σ_t^2 .
2. On régresse les \hat{u}_t^2 sur les variables exogènes pour trouver les paramètres α_h définissant la variance.

$$\hat{\sigma}_t^2 = \hat{\alpha}_0 + \hat{\alpha}_1 Z_{1t} + \dots + \hat{\alpha}_H Z_{Ht}$$

Cette technique demande des données supplémentaires sur les exogènes. Le modèle transformé sera :

$$\frac{Y_t}{\hat{\sigma}_t} = \beta_0 + \frac{X_1}{\hat{\sigma}_t} + \dots + \frac{X_K}{\hat{\sigma}_t} + \frac{u_t}{\hat{\sigma}_t}$$

Où $\hat{\sigma}_t = \sqrt{\hat{\alpha}_0 + \hat{\alpha}_1 Z_{1t} + \dots + \hat{\alpha}_H Z_{Ht}}$. Ceux-ci sont estimés par OLS.



2 Tests Diagnostiques d'Hétéroscédasticité

Dans cette section, nous introduisons des tests pour déterminer si les régressions considérées souffrent de problèmes d'hétéroscédasticité des erreurs (et résidus). Une approche intuitive serait de regarder un graphique des résidus de régression dans le but de trouver des anomalies. Cependant, une approche par analyse graphique n'est pas un test statistique formel de la présence d'hétéroscédasticité. Nous présentons donc des tests diagnostiques d'hétéroscédasticité. Il faut être prudent, car il existe de nombreuses formes d'hétéroscédasticité et les tests sont généralement conçus pour détecter une seule spécification.

2.1 Le test de Goldfeld-Quandt (GQ)

Soit le modèle :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_{K-1} X_{Ki} + u_i$$

On teste l'hypothèse nulle d'une variance constante dans le temps contre l'alternative que la variance augmente de manière monotone selon une exogène Z_1 .

$$H_0 : \sigma_i^2 = \sigma^2 \quad \forall i$$

contre l'alternative

$$H_A : \sigma_i^2 \text{ augmente de manière monotone selon } Z_1.$$

Marche à suivre :

1. Ordonner les observations selon la variable Z_1 .
2. Omettre c observations centrales.
3. Effectuer deux régressions séparées sur les $\frac{N-c}{2}$ premières et les $\frac{N-c}{2}$ dernières observations et calculer $\hat{\sigma}_1^2$ et $\hat{\sigma}_2^2$.
4. La statistique de test est :

$$GQ = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} \sim F \left(\frac{N-c-2K}{2}, \frac{N-c-2K}{2} \right)$$

Ici, les estimateurs de $\hat{\sigma}_1^2$ et $\hat{\sigma}_2^2$ sont calculés à partir des deux régressions en sous-groupe. Le nombre d'observations sera $\frac{N-c}{2}$ et le nombre de degrés de liberté sera $\frac{N-c-2K}{2}$.



Le test est exact et non asymptotique. Nous perdons les observations du centre, car on veut observer ce qui se passe aux extrémités. Par contre, si les erreurs sont normales, le test suit exactement une loi F et non asymptotiquement. On n'a donc pas besoin d'une grande taille d'échantillon pour ce test.

2.2 Le test de Breusch-Pagan (BP)

Ce test est plus général que le précédent et plus utilisé. En présence d'exogènes, le modèle est :

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_{K-1} X_{K-1,t} + u_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 Z_{1t} + \cdots + \alpha_H Z_{Ht} + e_t.$$

$$H_0 : \alpha_1 = \cdots = \alpha_H = 0 \quad \forall h$$

contre l'alternative qu'il existe un $\alpha_h \neq 0$.

Le test est basé sur une régression artificielle suivante :

$$\hat{u}_t^2 = \hat{\alpha}_0 + \hat{\alpha}_1 Z_{1t} + \cdots + \hat{\alpha}_H Z_{Ht} + \epsilon_t.$$

Interprétation : La statistique de test est :

$$BP = TR^2 \sim \chi^2(H).$$

Où R^2 est le coefficient de détermination de la régression artificielle. Si la statistique BP est supérieure à la valeur critique de la distribution χ^2 avec H degrés de liberté, nous rejetons l'hypothèse nulle d'homoscédasticité.

Application intéressante en finance : Test d'effet ARCH sur les résidus de régression. Le test de Breusch-Pagan peut également être utilisé pour tester l'hétéroscédasticité conditionnelle, comme dans le modèle ARCH où la variance des erreurs dépend des erreurs passées.

2.3 Le test de White

Ce test s'applique pour tester l'hétéroscédasticité de forme générale. On ne spécifie pas de forme pour la variance. L'hypothèse nulle est que la variance est homoscédastique contre une alternative hétéroscédastique générale.

Soit le modèle :

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$$



Le test est basé sur une régression artificielle :

$$\hat{u}_t^2 = \hat{\alpha}_0 + \hat{\alpha}_1 X_{1t} + \hat{\alpha}_{11} X_{1t}^2 + \hat{\alpha}_2 X_{2t} + \hat{\alpha}_{22} X_{2t}^2 + \hat{\alpha}_{12} X_{1t} X_{2t} + \epsilon_t.$$

Interprétation : La statistique de test est :

$$WHITE = TR_{\text{reg art.}}^2 \sim \chi^2(H).$$

Où H est le nombre de variables explicatives dans la régression artificielle. Si la statistique WHITE est supérieure à la valeur critique de la distribution χ^2 , nous rejetons l'hypothèse nulle d'homoscédasticité.

L'avantage du test de White est qu'il est très général. Par contre, on perd des degrés de liberté dans la régression artificielle, car on doit estimer les produits croisés et les régresseurs au carré.

2.4 Le correcteur de White

Lorsque l'on n'est pas prêt à spécifier une forme paramétrique pour la variance hétéroscédastique des erreurs, on ne peut pas utiliser les estimateurs GLS et FGLS définis plus haut. Une approche alternative est d'utiliser l'estimateur de $\hat{\beta}_{OLS}$, mais de corriger l'estimateur de sa variance pour tenir compte de l'hétéroscédasticité. On se souvient que $\hat{\beta}_{OLS}$ était toujours sans biais pour les cas hétéroscédastiques, mais que sa variance était incorrecte. On trouve donc une correction pour la variance des OLS. C'est le correcteur de White.

$$V(\hat{\beta}_{OLS}) = (X'X)^{-1} \left(\frac{1}{T} \sum_{t=1}^T \hat{u}_t^2 x_t x_t' \right) (X'X)^{-1}$$

Où

$$\Lambda = \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2 (x_t x_t').$$

Le correcteur de White est donc une approximation de la variance qui pondère les observations des résidus en fonction des régresseurs dans le calcul de la variance. En général, si on a des indices sur la structure de la matrice Ω , il est préférable de passer par l'estimateur FGLS si la qualité de l'information ajoutée est bonne. Le correcteur de White est très utile lorsqu'on n'a aucune idée de la forme fonctionnelle à donner à Ω . De plus, il a le mérite de ne pas estimer un modèle transformé, mais il reste une valeur approximative.



3 Conclusion

L'hétéroscédasticité est une problématique courante dans les modèles de régression linéaire. Les tests diagnostiques tels que ceux de Goldfeld-Quandt, Breusch-Pagan et White permettent de détecter sa présence et de choisir les méthodes appropriées pour la corriger, assurant ainsi la fiabilité des estimations des paramètres du modèle.

