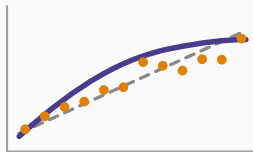


Économétrie avancée

Chapitre 8 : Hétéroscédasticité



Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger · 27 août 2026

Avant de commencer

Conséquences de l'hétéroscédasticité

Erreurs-types robustes à l'hétéroscédasticité

Détecter l'hétéroscédasticité

Corriger l'hétéroscédasticité : moindres carrés pondérés

Retour sur le modèle de probabilité linéaire

Synthèse

Avant de commencer

Pourquoi ce chapitre ?

L'hypothèse MLR.5 (homoscédasticité) n'a jamais servi à établir le non-biais ni la convergence des MCO — mais elle est indispensable à la validité des tests t et F usuels (chapitre 4). Que faire quand la variance de l'erreur dépend des régresseurs, comme c'est presque toujours le cas avec des données individuelles (ménages, firmes, actifs) ?

Conséquences de l'hétéroscédasticité

Non-biais préservé, inférence invalidée

L'hétéroscédasticité — $\text{Var}(u \mid x)$ non constante — **ne biaise pas** et **ne rend pas incohérents** les estimateurs MCO (MLR.1–4 suffisent). En revanche, les formules usuelles de $\text{Var}(\hat{\beta}_j)$ deviennent fausses : les statistiques t et F usuelles ne suivent plus leurs lois de référence, même asymptotiquement, et le théorème de Gauss-Markov (MCO = BLUE) ne s'applique plus.

Un problème d'inférence, pas d'estimation

Il est essentiel de bien situer le problème : les coefficients estimés restent des estimateurs valides de l'effet moyen ; ce sont leurs erreurs-types, et donc la significativité qu'on leur attribue, qui deviennent trompeuses en présence d'hétéroscédasticité non corrigée.

Erreurs-types robustes à l'hétéroscédasticité

Erreurs-types robustes (White/Huber/Eicker)

Sans supposer de forme particulière à $\text{Var}(u_i \mid x_i) = \sigma_i^2$, on peut estimer $\text{Var}(\hat{\beta}_j)$ de façon convergente à partir des résidus MCO au carré $\hat{u}_i^2$:

$$\widehat{\text{Var}}(\hat{\beta}_j) = \frac{\sum_i \hat{r}_{ij}^2 \hat{u}_i^2}{SSR_j^2},$$

où $\hat{r}_{ij}$ est le résidu de la régression de x_j sur tous les autres régresseurs. La racine carrée donne l'**erreur-type robuste**, valable asymptotiquement que l'hétéroscédasticité soit présente ou non.

Salaires : robuste vs usuel

Dans une équation de $\log(\text{wage})$ avec indicatrices croisées de genre/statut marital, l'erreur-type usuelle sur *educ* est .0067 contre .0074 en robuste — un écart modeste qui ne change aucune conclusion de significativité. Ce n'est pas toujours le cas : les deux versions peuvent diverger fortement, dans un sens comme dans l'autre.

Pourquoi ne pas toujours utiliser les erreurs-types robustes ?

Sous MLR.1–6 (normalité incluse), les statistiques t usuelles sont **exactes** à tout n ; les versions robustes

Détecter l'hétéroscédasticité

Test de Breusch-Pagan (BP)

On teste $H_0 : \text{Var}(u \mid x) = \sigma^2$ en régressant $\hat{u}_i^2$ sur les régresseurs originaux :

$$\hat{u}_i^2 = \delta_0 + \delta_1 x_{i1} + \cdots + \delta_k x_{ik} + \text{erreur},$$

puis on teste la significativité globale par F ou par $LM = n R_{\hat{u}^2}^2 \sim \chi_k^2$ sous H_0 . Un rejet indique une variance de l'erreur liée aux régresseurs.

Prix immobiliers : niveaux vs logs

En niveau, $\widehat{price} = f(lotsize, sqrft, bdrms)$ donne $R_{\hat{u}^2}^2 = .160$, soit $F \approx 5.34$ ($p = .002$) : hétéroscédasticité nette. En passant $price$, $lotsize$ et $sqrft$ en logarithme, $R_{\hat{u}^2}^2$ tombe à .048 ($p = .245$) : on ne rejette plus l'homoscédasticité. C'est une illustration directe d'un point du chapitre 6 — la forme logarithmique réduit souvent l'hétéroscédasticité.

Test de White et sa version simplifiée

Le test de **White** ajoute au test BP les carrés et produits croisés de tous les régresseurs — plus général, mais coûteux en degrés de liberté. Une version économe régresse plutôt $\hat{u}_i^2$ sur les valeurs ajustées $\hat{y}_i$ et $\hat{y}_i^2$:

$$\hat{u}_i^2 = \delta_0 + \delta_1 \hat{y}_i + \delta_2 \hat{y}_i^2 + \text{erreur},$$

avec toujours 2 degrés de liberté quel que soit le nombre de régresseurs originaux — un cas particulier du test de White complet.

Une mise en garde : confusion avec une mauvaise spécification

Un rejet du test d'hétéroscédasticité peut en réalité signaler une **mauvaise forme fonctionnelle** (terme quadratique omis, niveau au lieu du log) plutôt qu'une véritable hétéroscédasticité. Il vaut mieux traiter la spécification fonctionnelle en premier (chapitre 9) et ne tester l'hétéroscédasticité qu'une fois la forme du modèle jugée satisfaisante.

Corriger l'hétéroscédasticité :
moindres carrés pondérés

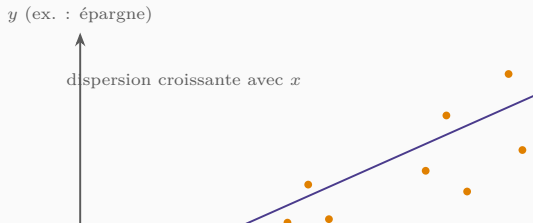
Hétéroscédasticité connue à une constante près

Transformation GLS

Si $\text{Var}(u \mid x) = \sigma^2 h(x)$ avec $h(x) > 0$ connue, diviser toute l'équation par $\sqrt{h_i}$ donne une équation transformée à erreur homoscédastique, estimable par MCO — ce sont les **moindres carrés généralisés (MCG)**, ou, de façon équivalente, les **moindres carrés pondérés (MCP)** avec poids $1/h_i$:

$$\min_b \sum_i \frac{(y_i - b_0 - b_1 x_{i1} - \dots - b_k x_{ik})^2}{h_i}.$$

Sous les hypothèses de Gauss-Markov appliquées à l'équation transformée, l'estimateur MCG est **BLUE** — strictement plus efficace que les MCO.



Hétéroscédasticité à estimer : les MCG faisables

Procédure FGLS

Quand $h(x)$ est inconnue, on la modélise (par exemple $\text{Var}(u \mid x) = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \dots + \delta_k x_k)$, qui garantit une variance positive) puis on l'estime en quatre étapes : (i) régression MCO de y sur les x_j , résidus $\hat{u}_i$; (ii) régression de $\log(\hat{u}_i^2)$ sur les x_j , valeurs ajustées $\hat{g}_i$; (iii) poids $\hat{h}_i = \exp(\hat{g}_i)$; (iv) MCP avec ces poids estimés. L'estimateur FGLS n'est plus exactement non biaisé, mais reste **convergent** et asymptotiquement plus efficace que les MCO.

Demande de cigarettes

Le test BP rejette fortement l'homoscédasticité ($LM = 32.3, p < .00002$). En FGLS, le coefficient sur $\log(\text{income})$ devient significatif et plus élevé qu'en MCO (1.30 contre 0.88) — signe que les MCO, en pondérant également des observations très bruitées, masquaient une partie de l'effet.

MCO ou FGLS : que faire en cas de désaccord ?

Un écart important et de signe opposé entre estimations MCO et FGLS n'est généralement pas dû à l'hétéroscédasticité elle-même, mais suggère une violation plus grave — typiquement $E(u \mid x) \neq 0$ (endogénéité). Les deux estimateurs sont alors biaisés, de façon différente; ce diagnostic informel doit alerter avant d'adopter aveuglément le résultat FGLS.

Retour sur le modèle de probabilité linéaire

Pondération naturelle du MPL

Dans le modèle de probabilité linéaire (chapitre 7), $\text{Var}(y \mid x) = p(x)[1 - p(x)]$ est de forme **connue** : on peut estimer les poids par $\hat{h}_i = \hat{y}_i(1 - \hat{y}_i)$ à partir des valeurs ajustées MCO, puis appliquer les MCP directement, sans étape FGLS intermédiaire.

La limite pratique des MCP pour le MPL

Le MCP échoue dès qu'une valeur ajustée sort de $[0, 1]$, ce qui rend $\hat{h}_i$ négatif — situation fréquente avec des probabilités de succès faibles ou beaucoup d'observations. En pratique, il est le plus souvent plus simple et plus robuste de conserver les MCO avec erreurs-types robustes plutôt que de forcer des MCP délicats à mettre en œuvre.

Synthèse

L'essentiel du chapitre 8

- L'hétéroscédasticité ne biaise pas les MCO mais invalide l'inférence usuelle; les erreurs-types robustes de White corrigent ce problème sans connaître la forme de l'hétéroscédasticité.
- Le test de Breusch-Pagan (et sa généralisation, le test de White) détectent l'hétéroscédasticité en régressant les résidus au carré sur les régresseurs (ou sur $\hat{y}$, $\hat{y}^2$).
- Les moindres carrés pondérés (MCG/MCP), et leur version faisable FGLS, restaurent l'efficacité en donnant moins de poids aux observations les plus bruitées — au prix de devoir modéliser la variance.
- Le modèle de probabilité linéaire a une hétéroscédasticité de forme connue, mais son application pratique en MCP est limitée par les valeurs ajustées hors $[0, 1]$.

Vers le chapitre 9

L'hétéroscédasticité n'est qu'une des façons dont un modèle peut s'écarter des hypothèses idéales. Le chapitre suivant élargit la focale : mauvaise forme fonctionnelle, variables proxy, erreurs de mesure, données manquantes et valeurs aberrantes — autant de problèmes de spécification et de données à diagnostiquer avant de faire confiance à une régression.