Économétrie II

Ch. 3. Hétéroscédasticité
L3 Économétrie – L3 MASS

Prof. Philippe Polomé, U. Lyon 2

Année 2016-2016

Rappel

1. $E(\epsilon_i) = 0 \; \forall i$: **Espérance nulle**
2. $var(\epsilon_i) = \sigma^2 \; \forall i$: **Homoscédasticité**
3. $cov(\epsilon_t, \epsilon_s) = 0 \; \forall t \neq s$: **Pas d'auto-corrélation**
4. $E(\epsilon_i x_i) = 0 \; \forall i$: **Exogénéité**
5. ✓ La matrice X est de plein rang : **Pas de multicolinéarité**
6. Le modèle est **correctement spécifié**
7. La variable dépendante Y est **continue**

Table des matières

Table des matières

Définition du problème

- L'hypothèse d'**homoscédasticité** requiert que la variance des termes d'erreur soit la **même** pour chaque observation
- Il y a hétéroscédasticité dans le modèle $Y = X\beta + \epsilon$ lorsque :

$$var(\epsilon) = E\left(\epsilon\epsilon^{'}\right) = \Sigma_\epsilon = \begin{bmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_N^2 \end{bmatrix} \neq \sigma^2 I_N$$

Représentation graphique

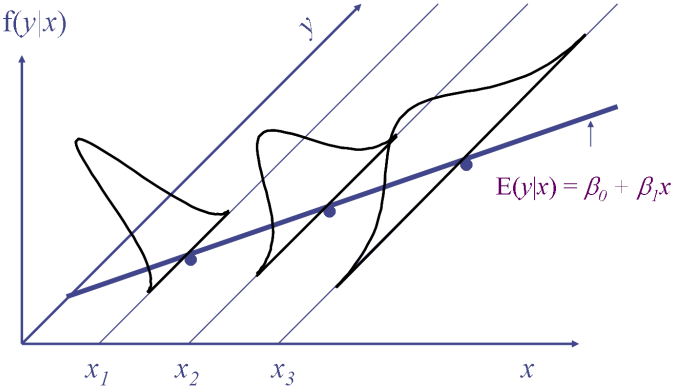

Table des matières

Données en moyenne

- S'il y a homoscédasticité dans les données de départ, les données en moyenne seront hétéroscédastiques
  - y_{it} avec $var(y_{it}) = 1 \forall i, t$
- Mais on ne dispose que des moyennes $\frac{1}{T_g} \sum_t y_{it} = y_i$ où T_g est la taille du groupe
  - Par ex : des moyennes régionales de données individuelles
  - $var(y_i) = \frac{1}{T_g^2} var\left(\sum_t y_{it}\right) = \frac{1}{T_g^2} \sum_t var(y_{it}) = \frac{T_g}{T_g^2} = \frac{1}{T_g}$: dépend de la taille du groupe

Modèle à “coefficients aléatoires”

- Si le modèle sous-jacent est $Y_i = \alpha + (\beta + \mu_i)\, x_i + \epsilon_i$
  - Par ex. effet de l'éducation sur le salaire
- Alors $Y_i = \alpha + \beta x_i + \mu_i x_i + \epsilon_i = \alpha + \beta x_i + \eta_i$
- Et, avec des termes d'erreurs ϵ_i et μ_i homoscédastiques et indépendant et un régresseur x_i non-stochastique, on trouve
  - $var(\eta_i) = \sigma_\epsilon^2 + \sigma_\mu^2 x_i$ hétéroscédastique
- Semblable au cas suivant

Régresseur manquant hétéroscédastique

- Si le modèle sous-jacent est $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$
- Mais le modèle estimé est $y_i = \beta_0 + \beta_1 x_{1i} + \mu_i$
- Alors $\mu_i = \beta_2 x_{2i} + \epsilon_i$ donc,
    - si x_{2i} n'est pas corrélé à ϵ_i, $var(\mu_i) = \beta_2^2 var(x_{2i}) + \sigma^2$
    - si x_{2i} n'est pas stochastique (analyse conditionnelle), $var(\mu_i) = \beta_2^2 x_{2i} + \sigma^2$
- Hétéroscédastique sauf cas particulier

Effet taille

- La variance est une mesure absolue, pas relative
  - Imaginons que le CA de toutes les entreprises varie de 10%
  - 10% est un grand nombre pour une grande entreprise
- Part du revenu disponible dépensé en loisirs
  - Les familles à faibles revenus dépensent relativement peu en loisirs. Les variations de ces dépenses entre ces familles sont donc faibles
  - Pour les familles avec des revenus importants, le montant moyen relatif dépensé en loisirs sera plus élevé, et il y aura une plus grande variabilité entre de telles familles

Variables explicatives **de la variance**

- Un régresseur définit des **groupes** de variances différentes dans la variable expliquée
- Ex. Rendement de l'éducation
  - **variance** en productivité propre inobservable ϵ diffère selon les niveaux h d'éducation atteints
  - $\ln(salaire_i) = \alpha + \gamma education_h + X_i\beta + \epsilon_i$
    - avec $i \in h$ et $var(\epsilon_i) = \sigma_h^2 = fonction(education_h)$
  - Similaire à l'effet taille
    - Faible éducation : salaire proche du minimum
- Également :
  - Qualité inobservée d'un bien par niveau de prix
  - Taux d'épargne par niveau de revenu

Table des matières

Propriétés de $\hat{\beta}$

- MCO restent **sans biais** (X non-stochastique par facilité)

$$E\left(\hat{\beta}\right) = E\left(\left(X'X\right)^{-1} X'\left(X\beta + \epsilon\right)\right) = \beta + \left(X'X\right)^{-1} X' E\left(\epsilon\right) = \beta$$

- MCO **consistants** / convergents (sans démonstration)
- Matrice de variance-covariance des coefficients estimés n'est plus $\sigma^2\left(X'X\right)^{-1}$, mais bien **(sandwich)**

$$\Sigma_{\hat{\beta}} = E\left(\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)'\right) = \left(X'X\right)^{-1} X' E\left(\epsilon\epsilon'\right) X \left(X'X\right)^{-1}$$

$$= \left(X'X\right)^{-1} X' \Sigma_{\epsilon} X \left(X'X\right)^{-1}$$

- Théorème de Gauss-Markov ne s'applique plus
  - MCO n'est plus **efficient**

Inférence

- Estimateur MCO $\dfrac{\hat{\epsilon}'\hat{\epsilon}}{N-k}\left(X'X\right)^{-1}$ est **biaisé** pour $\Sigma_{\hat{\beta}}$
  - Tests d'hypothèse usuels post-estimation (t-stat, F-stat ou LM) **invalides** dans leur forme classique
  - Le bootstrap reste par contre valide
    - asymptotiquement comme toujours
- Comment faire face à ces conséquences ? 2 approches
  - Estimer $\Sigma_{\hat{\beta}}$ à partir de MCO & refaire l'inférence
  - Proposer un estimateur **alternatif**
    - Moindres Carrés Pondérés
    - Vise à récupérer l'efficience

Table des matières

Estimateur robuste White (1980)

- On sait $\Sigma_{\hat{\beta}} = \left(X'X\right)^{-1} X'\Sigma_\epsilon X \left(X'X\right)^{-1}$ **sandwich**
- Sauf cas particulier, Σ_ϵ inconnue
- White : Pour obtenir un estimateur de $\Sigma_{\hat{\beta}}$, il suffit d'un estimateur de $X'\Sigma_\epsilon X$ (et pas de Σ_ϵ)
- Sous des conditions très générales, $S = \frac{1}{N}\sum_{i=1}^{N} \hat{\epsilon}_i^2 X_i X_i'$ est un estimateur consistant de $\frac{1}{N}X'\Sigma_\epsilon X$
  - $\hat{\epsilon}_i = y_i - X_i'\hat{\beta}_{MCO}$ résidu MCO i
  - X_i vecteur-colonne correspondant à l'observation i de X
    - Donc $X_i X_i'$ est bien $k \times k$
- Expliciter la matrice $X'\Sigma_\epsilon X$: intuition estimateur de White

La matrice $X'\Sigma_\epsilon X_{k\times k}$

$$
= \begin{bmatrix} x_{11} & x_{21} & \dots & x_{N1} \\ x_{12} & x_{22} & \dots & x_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \dots & x_{Nk} \end{bmatrix} \begin{bmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_N^2 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nk} \end{bmatrix}
$$

$$
= \begin{bmatrix} \sigma_1^2 x_{11} & \sigma_2^2 x_{21} & \dots & \sigma_N^2 x_{N1} \\ \sigma_1^2 x_{12} & \sigma_2^2 x_{22} & \dots & \sigma_N^2 x_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_1^2 x_{1k} & \sigma_2^2 x_{2k} & \dots & \sigma_N^2 x_{Nk} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nk} \end{bmatrix}
$$

La matrice $X'\Sigma_\epsilon X_{k\times k}$

$$
= \begin{bmatrix}
\sum_{i=1}^{N}\sigma_i^2 x_{i1}^2 & \sum_{i=1}^{N}\sigma_i^2 x_{i1}x_{i2} & \cdots & \sum_{i=1}^{N}\sigma_i^2 x_{i1}x_{ik} \\
 & \sum_{i=1}^{N}\sigma_i^2 x_{i2}^2 & \cdots & \sum_{i=1}^{N}\sigma_i^2 x_{i2}x_{i2} \\
 & & \ddots & \vdots \\
sym & & & \sum_{i=1}^{N}\sigma_i^2 x_{ik}^2
\end{bmatrix}
$$

Comparer avec $S = \frac{1}{N}\sum_{i=1}^{N}\hat{\epsilon}_i^2 X_i X_i'$

Matrice de variance-covariance "robuste"

- $\hat{\Sigma}_{\hat{\beta}} = N \left(X'X\right)^{-1} S \left(X'X\right)^{-1}$ consistant pour $\Sigma_{\hat{\beta}}$
  - Peut-être utilisé pour tests usuels post-estimation
- Écart-types issus de White : "robustes à l'hétéroscédasticité"
  - Suggestion : corriger la matrice de White par $n/(n-k-1)$
  - Lorsque $n \longrightarrow \infty$ les deux approches sont équivalentes
- L'estimateur de White est seulement consistant
  - Pas sans biais
  - Valable seulement asymptotiquement
  - Sur échantillons de petite taille
    - t de Student "de White" n'ont pas une distribution proche du t
    - Tests ont peu de puissance
    - Utile de voir si Bootstrap mène aux mêmes résultats

Matrice de White : logiciels

- La correction par la matrice de White est pré-programmée sur tous les logiciels d'économétrie.
  - Sous Gretl, cocher une case dans la boîte de dialogue d'estimation
    - Plusieurs variantes à la correction de White, manuel de Gretl pour les détails
- Dans GRETL
  - Prenez les données hprice1.gdt dans Gretl Wooldridge
  - Régressez *lprice* sur *llotsize*, *lsqrft*, *bdrms*, *colonial*
  - Pour obtenir l'estimation "robuste" des t-stats
    - Cocher "erreurs standards robustes" (la constante n'est plus significative)

Table des matières

Moindres carrés pondérés "Weighted Least Squares"

- Approche alternative à celle de White pour traiter l'hétéroscédasticité
  - + ancienne
- Disposer d'informations supplémentaires sur la forme de l'hétéroscédasticité rencontrée permet toujours de dériver un estimateur **plus efficient** que celui donné par l'estimation "robuste"
  - Donc si on connait la forme de l'hétéroscédasticité, on devrait pouvoir obtenir un gain en efficience
- L'idée générale est de transformer les données de sorte à ce que les erreurs deviennent homoscédastiques

Forme de l'hétéroscédasticité connue à une constante près

- Supposons que l'hétéroscédasticité puisse être modélisée sous la forme $var\left(\epsilon|X\right) = \sigma^2 h\left(X\right)$
- On peut alors écrire $\Sigma_\epsilon = \sigma^2 \begin{pmatrix} h_1 & 0 & \cdots & 0 \\ 0 & h_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & h_N \end{pmatrix} = \sigma^2 \Psi$

  avec $h_i = h\left(X_i\right) > 0$
- Si on réécrit le modèle $Y = X\beta + \epsilon$
  - sous la forme $\dfrac{Y_i}{\sqrt{h_i}} = \dfrac{X_i}{\sqrt{h_i}}\beta + \dfrac{\epsilon_i}{\sqrt{h_i}}$
  - alors le terme d'erreur est homoscédastique

Remarque

- Le résidu MCP est $\hat{\epsilon}^* = Y^* - X^* \hat{\beta}_{MCP}$
- Le but de MCP est minimiser la $\sum$ des carrés des résidus sur les données **transformées** :
  $\min \hat{\epsilon}^{*\prime} \hat{\epsilon}^* = \min \sum_i \left(Y_i^* - X_i^* \hat{\beta}_{MCP} \right)^2$
  $= \min \sum_i \left(\frac{Y_i}{\sqrt{h_i}} - \frac{X_i}{\sqrt{h_i}} \hat{\beta}_{MCP} \right)^2$
  $= \min \sum_i \left(Y_i - X_i \hat{\beta}_{MCP} \right)^2 / h_i$
- Chaque observation est **pondérée** par l'inverse de sa variance
  - Plus une observation a une variance élevée, moins elle **pèse** dans la somme des carrés des résidus
  - La qualité de l'ajustement (R^2) aux données originales n'est donc plus recherchée : R^2 n'est **plus une mesure intéressante**

MCP en pratique

- Il faut connaître la forme de l'hétéroscédasticité
- Dans la plupart des cas on ne sait rien sur cette forme
- Il faut donc un estimateur de $h(X)$ ou éventuellement d'autres formes $h(X, Z)$
  - Moindres Carrés Pondérés **Faisables** / ***Feasible*** *Weighted Least Squares*

MCP Faisable

- On suppose une forme simple de type $h(X_i) = \exp(\delta_0 + \delta_1 x_{1i} + \delta_2 x_{2i} + \ldots)$
  - exp garanti la positivité
  - $var(\epsilon_i | X_i) = \sigma^2 \exp(\delta_0 + \delta_1 x_{1i} + \delta_2 x_{2i} + \ldots)$
- Comme MCO est sans biais en présence d'hétéroscédasticité, $\hat{\epsilon}_i{}^2$ peut être vu comme une estimation de $var(\epsilon_i | X_i)$
  - $\hat{\epsilon}_i^2 = \sigma^2 \exp(\delta_0 + \delta_1 x_{1i} + \delta_2 x_{2i} + \ldots) \nu_i$; ν terme d'erreur
- Estimer $\ln\left(\hat{\epsilon}_i^2\right) = \alpha_0 + \delta_1 x_{1i} + \delta_2 x_{2i} + \ldots + \ln(\nu_i)$ par MCO
  - On peut rajouter des régresseurs $Z \not\supseteq X$ dans cette équation
- Estimation de h_i : $\hat{h}_i = \exp\left(\hat{\alpha}_0 + \hat{\delta}_1 x_{1i} + \hat{\delta}_2 x_{2i} + \ldots + \frac{1}{2}\hat{\sigma}_\nu^2\right)$
  - Il faut ajouter un terme en variance au carré parce que exp est non-linéaire (Verbeek) – sans démonstration

Mise en garde sur les MCP faisables

- Dans le doute sur la présence et la forme de l'hétéroscédasticité
  - il peut être tentant de prendre une forme usuelle et d'appliquer les MCP
  - D'autant plus tentant si le logiciel utilisé propose une procédure simple
- Mais
  - Si les termes d'erreurs sont homoscédastiques au départ,
    - l'estimateur des **MCPF pourra être biaisé et inconsistant**
    - Par monte-carlo, on voit que ce sont des configurations peu courantes
  - Si l'hétéroscédasticité dépend d'une variable inconnue,
    - ou ne dépend pas d'une variable,
    - il peut être difficile d'apporter une correction significative

Comparaison des deux approches

- Approche 1 : MCP
  - $\hat{\beta}_{MCP} \neq \hat{\beta}_{MCO}$
  - Approche historique, en principe meilleure que MCO
    - Si l'hyp sur forme de l'hétéroscédasticité est correcte
    - Permet alors un gain d'efficience
  - Mais risques l'hyp est fausse
- Approche 2 : $\widehat{var\left(\hat{\beta}_{MCO}\right)}$ robuste
  - On garde les $\hat{\beta}_{MCO}$
  - 2.a. White : basé sur un résultat plus récent (1980) que MCP ; requiert une plus grande puissance de calcul
  - 2.b. bootstrap : puissance de calcul encore plus grande
  - 2.a et 2.b : pas d'hypothèses supplémentaires par rapport à MCO mais seulement valables pour de grands échantillons
  - Évite de tester l'hétéroscédasticité
    - et de chercher quelle forme elle pourrait prendre
  - Renonce au gain potentiel d'efficience

Importance de l'hétéroscédasticité en pratique

- L'hétéroscédasticité est la norme avec les données micro en coupe transversale
  - L'homoscédasticité est l'exception
  - On va habituellement utiliser White / bootstrap
    - Plus rarement MCP / MCG
- Pas aussi évident pour les séries temporelles
  - Car c'est toujours la même unité qui est observée
  - White est aussi disponible
    - Mais existence de modèles alternatifs propres

Exemple dans Gretl

- Prenez les données hprice1.gdt dans Gretl Wooldridge
- Régressez *lprice* sur *llotsize*, *lsqrft*, *bdrms*, *colonial*
- Moindres Carrés Pondérés : "Modèle" – "autres modèles linéaires" – "MCP"
  - Un seul régresseur est associé à l'hét.
  - Une façon alternative est "Hétéroscédasticité corrigée" qui impose une correction "à la White"
    - c'est-à-dire : l'hét. est approximée à partir d'une regression contenant les régresseurs et leurs carrés
    - comme dans le test de White (plus loin)

Table des matières

Principe des tests d'hétéroscédasticité

- On n'observe jamais les vrais termes d'erreur.
- On les "estime" à partir des résidus de la régression par MCO $\hat{\epsilon}_i$
  - "car" MCO sans biais en présence d'hétéroscédasticité
- 3 tests populaires
  - Breusch-Pagan
  - White
  - Goldfeld-Quandt

Test de Breusch-Pagan

- On veut tester $H_0 : var(\epsilon_i | x_{1i}, ..., x_{ki}) = \sigma^2 \; \forall i$
  - Equivalent à tester $H_0 : E\left(\epsilon_i^2 | x_{1i}, ..., x_{ki}\right) = \sigma^2 \, \forall i$ car $E(\epsilon) = 0$
- Si on suppose que la relation entre ϵ_i^2 et X_i est suffisamment proche du linéaire
  - $\epsilon_i^2 = \delta_0 + \delta_1 x_{1i} + ... + \delta_k x_{ki} + \nu$
  - Alors tester H_0 revient à tester $H_0 : \delta_1 = \delta_2 = ... = \delta_k = 0$
- Régresser carré des résidus $\hat{\epsilon}_i^2$ sur toutes les variables explicatives X
  - Tester significativité globale via la procédure habituelle (F-test ou LM-test)

Test de White

- Test de Breusch-Pagan permet de détecter les formes linéaires d'hétéroscédasticité
- Test de White permet de prendre en compte certaines non-linéarités en utilisant les carrés et les produits croisés de toutes les variables explicatives
  - Même procédure que Breusch-Pagan
  - En introduisant tous les x_j^2 et les $x_j x_m$
  - Et en testant que les paramètres associés sont conjointement significatifs (F-test ou LM-test)
  - Rapidement : nombre de paramètres à estimer impraticable...
    - En proba (5%) erreur type I : $\neg R\ H_0$ fréquente

Forme alternative du test de White

- Les valeurs ajustées $\hat{y}_i = X_i\hat{\beta}$ sont fonction de toutes les X
  - $\hat{y}^2$ est une fonction des carrés x_j^2 et des produits croisés $x_j x_m$
  - $\hat{y}^2$ peut être utilisé pour représenter les non-linéarités
  - On peut utiliser $\hat{y}$ pour représenter tous les X à la fois
- Procédure
  1. Régresser le carré des résidus MCO $\hat{\epsilon}_i^2$ sur les valeurs ajustées $\hat{y}_i$ et $\hat{y}_i^2$
  2. F-test ou un LM-test sur la significativité globale de cette régression
- On ne teste plus que les coef. de ces 2 paramètres
- Ce test repose sur une hypothèse forte concernant la forme de l'hétéroscédasticité
  - Celle-ci est fonction des variables incluses
  - N'impose pas la linéarité de cette forme (on peut étendre aux cubes...)

Test de Goldfeld-Quandt

- Dans le cas où la théorie permet d'envisager une hétéroscédasticité causée par une seule explicative x_m

1. Trier les observations par les valeurs de la variable explicative soupçonnée source d'hétéroscédasticité
   - et donc corrélée avec $\hat{\epsilon}_i^2$
2. Supprimer la partie des données triées qui se trouve au milieu de l'échantillon
   - entre $1/3$ et $1/5$ des données
3. Estimations séparées (MCO) sur les deux sous échantillons : hautes et basses valeurs de x_m
4. Les rapport des variances estimées des termes d'erreur des deux régressions suit une distribution de Fisher
   - $GQ = \dfrac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \sim F_{N_1-K,N_2-K}$
   - $N_1 - K =$ nombre de degrés de liberté de la 1ère régression

Exemple dans Gretl

- Prenez les données hprice1.gdt dans Gretl Wooldridge
- Régressez *lprice* sur *llotsize*, *lsqrft*, *bdrms*, *colonial*
- Pour tester : MCO puis post-estimation
  - Gretl propose 2 tests : White (1º forme), Breusch-Pagan, mais pas Goldfeld-Quandt
  - Breusch-Pagan R homoscédasticité
  - White (1º forme)$\neg R$
  - Difficile de conclure

Devoir #3 Hétéroscédasticité

Conception d'une feuille de tableur pour montrer l'effet de **l'hétéroscédasticité** sur les coefficients estimés dans une régression linéaire MCO à deux variables x_1 et x_2 et une constante.

1. Générer le terme d'erreur ϵ ; par exemple pour chaque observation i, générer d'abord un nombre aléatoire $\alpha_i \in [1, 10]$, puis générer $\epsilon_i = \alpha_i n(0, 1)$
2. Illustrer que les MCO sont inconsistants ou non lorsqu'il y a hétéroscédasticité
3. Employer la formule classique de calcul de la matrice de variance-covariance des estimations MCO
   3.1 (pour les plus motivés) Montrer que la diagonale de cette matrice ne s'approche pas des variances des coefficients estimés en Monte-Carlo