

## Chapter 6

# Hétéroscédasticité

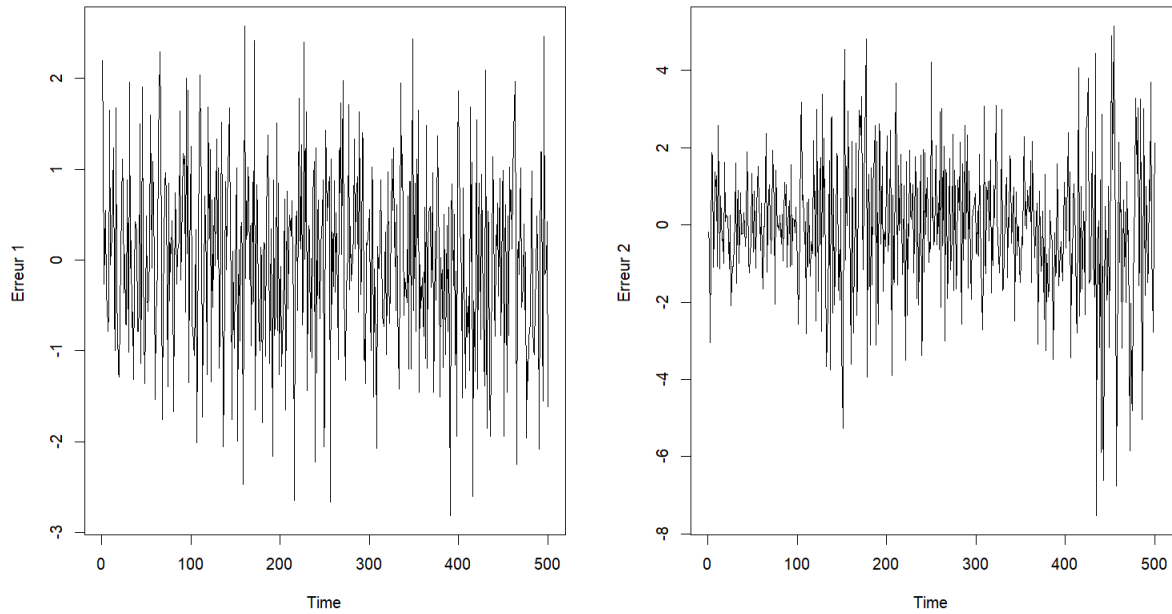
### 6.1 Introduction et définition générale

L'hétéroscédasticité est une caractéristique fondamentale des séries temporelles financières et économiques, où la variance conditionnelle des erreurs n'est pas constante au fil du temps. Contrairement à l'homoscédasticité, qui suppose une variance conditionnelle constante, l'hétéroscédasticité permet de mieux modéliser les phénomènes où les erreurs fluctuent en intensité en fonction de différentes périodes ou événements.

Dans ce chapitre, nous allons explorer la nature de l'hétéroscédasticité et son importance dans l'analyse des séries temporelles. Nous commencerons par définir ce concept et discuter des implications qu'il peut avoir sur les estimations et les prévisions économiques. Ensuite, nous aborderons plusieurs tests statistiques permettant de détecter la présence d'hétéroscédasticité, tels que le test de White et le test de Breusch-Pagan.

Enfin, nous introduirons le modèle GARCH (Generalized Autoregressive Conditional Heteroskedasticity), l'un des modèles les plus utilisés pour capturer la dynamique de la variance conditionnelle dans les séries temporelles. Ce modèle, développé par Robert Engle en 1982 et généralisé par Tim Bollerslev en 1986, offre un cadre robuste pour modéliser et prévoir la volatilité des séries financières.

Pour illustrer cette idée, nous prendrons deux exemples portant sur deux types d'erreurs. Il est facile de constater que les fluctuations, qui nous donnent une idée de la variance conditionnelle d'une série, sont constantes dans la figure à gauche. En revanche, les erreurs dans la figure à droite présentent des fluctuations qui varient au cours du temps. Ce phénomène est appelé hétéroscédasticité.



**Definition 15.** En statistique, une séquence de variables aléatoires est dite conditionnellement hétéroscédastique si ces variables ont des variances différentes. Au contraire, une séquence de variables est appelée homoscedastique si la variance est constante pour chaque variable.

Par exemple, la séquence  $Y_t, t = 1, \dots, n$ , est dite hétéroscédastique si la variance conditionnelle de  $Y_t$  sachant  $Y_t$  change avec le temps. **Ainsi, la variance conditionnelle n'est plus constante et doit être modélisée.**

Historiquement, l'analyse de la régression linéaire en présence d'hétéroscédasticité a été étudiée par Engle (1982). L'hétéroscédasticité n'affecte pas seulement les paramètres estimés par la méthode des moindres carrés, mais aussi l'écart-type des estimateurs.

Pour illustrer cela, considérons l'exemple suivant : soit

$$Y_t = \alpha + \beta Y_{t-1} + \varepsilon_t$$

où  $\text{Var}(\varepsilon_t/Y_{t-1}, \dots) = \sigma_t^2 = \sigma_\varepsilon^2 Z_t^2$ , et  $Z_t$  est un BB.  $\text{Var}(\varepsilon_t/Y_{t-1}, \dots)$  n'est plus constante, mais dépend du carré de  $Y_{t-1}, \dots$  (si  $Z_t \nearrow \implies \sigma_t \nearrow$ ).

## 6.2 Détection de l'hétéroscédasticité - Test de White

Ce test respecte les 3 étapes suivant:

1. On estime un modèle de régression multiple:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \varepsilon_t$$

2. On estime la regression auxiliaire. On regresse le carré des résidus  $\hat{\varepsilon}_t^2$  sur les variables  $X_{it}$ ,  $i = 1, \dots, k$ , sur leurs carrés  $X_{it}^2$  et sur leur produit 2 à 2.

$$\begin{aligned}\hat{\varepsilon}_t^2 = & a_0 + a_1 X_{1t} + a_2 X_{2t} + \dots + a_k X_{kt} \\ & + b_1 X_{1t}^2 + b_2 X_{2t}^2 + \dots + b_k X_{kt}^2 \\ & + c_{1,2} X_{1t} X_{2t} + \dots + c_{1,k} X_{1t} X_{kt} \\ & + c_{2,3} X_{2t} X_{3t} + \dots + c_{2,k} X_{2t} X_{kt} \\ & + \dots \\ & + c_{k-1,k} X_{k-1,t} X_{kt} \\ & + u_t, \quad \text{où } u_t \sim BB(0, \sigma_u^2).\end{aligned}$$

On inspecte  $R^2$  le coef. de détermination

$$R^2 = 1 - \frac{\sum u_t^2}{\sum y_t^2} = 1 - \frac{SCR}{SCT}$$

3. On teste l'hypothèse nulle de l'hétéroscédasticité

$$\begin{aligned}H_0 : & a_1 = a_2 = \dots = a_k = b_1 = b_2 = \dots = b_k \\ & = c_{1,2} = \dots = c_{k-1,k} = 0\end{aligned}$$

contre

$H_1$  : il existe au moins un paramètre non nulle.

On calcule la stat de White

$$L = TR^2 \sim \chi^2(p)$$

où  $T$  est la taille de la série chronologique des erreurs et  $p$  est le nombre des variables.

Si  $L_{obs} < \chi_{theo}^2 \implies$  on rejette  $H_0$  et **on conclut la présence d'hétéroscédasticité.**

*Dans ce cours, nous allons concentrer sur un modèle de séries chronologiques pour le cas d'hétéroscédasticité qui est le modèle **ARCH**.*

## 6.3 Test de Breusch-Pagan

Le test de Breusch-Pagan est un test statistique qui permet de détecter la présence d'hétéroscédasticité dans un modèle de régression linéaire. Ce test repose sur l'idée que, sous l'hypothèse nulle d'homoscédasticité, les résidus au carré devraient être indépendants des variables explicatives. Le test se déroule en deux étapes :

1. On estime le modèle de régression initial et on récupère les résidus  $\hat{\varepsilon}_i$ .

2. On effectue ensuite une régression auxiliaire où les résidus au carré sont régressés sur les variables explicatives du modèle initial. Le test de Breusch-Pagan se base sur la statistique de Lagrange Multiplier (LM), qui est asymptotiquement distribuée selon une loi du chi-deux sous l'hypothèse nulle d'homoscédasticité.

La statistique du test de Breusch-Pagan, notée LM, est utilisée pour tester l'hypothèse nulle d'homoscédasticité contre l'alternative où la variance des erreurs est une fonction linéaire des variables explicatives.

1. **Estimation des résidus :** Considérons le modèle de régression linéaire

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i,$$

où  $y_i$  est la variable dépendante,  $x_{i1}, x_{i2}, \dots, x_{ik}$  sont les variables explicatives,  $\beta_0, \beta_1, \dots, \beta_k$  sont les coefficients à estimer, et  $\varepsilon_i$  est le terme d'erreur.

Après avoir estimé les coefficients du modèle par moindres carrés ordinaires (MCO), on obtient les résidus estimés  $\hat{\varepsilon}_i = y_i - \hat{y}_i$ .

2. **Régression auxiliaire :** On effectue ensuite une régression auxiliaire où les carrés des résidus  $\hat{\varepsilon}_i^2$  sont régressés sur les variables explicatives du modèle original :

$$\hat{\varepsilon}_i^2 = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \cdots + \alpha_k x_{ik} + u_i.$$

Ici,  $\alpha_0, \alpha_1, \dots, \alpha_k$  sont les nouveaux coefficients à estimer, et  $u_i$  est le terme d'erreur de cette régression auxiliaire.

3. **Calcul de la statistique LM :** La statistique du test de Breusch-Pagan est calculée comme suit :

$$LM = \frac{n \times R^2}{2},$$

où  $n$  est le nombre d'observations et  $R^2$  est le coefficient de détermination de la régression auxiliaire. Cette statistique suit asymptotiquement une distribution du chi-deux avec  $k$  degrés de liberté sous l'hypothèse nulle d'homoscédasticité.

4. **Interprétation :** Si la statistique LM est grande, ou si la p-valeur associée est inférieure à un seuil de signification (généralement 0,05), on rejette l'hypothèse nulle d'homoscédasticité, suggérant la présence d'hétéroscédasticité.

## 6.4 Modèle ARCH(q)

Le modèle autorégressif conditionnellement hétéroscédastique (ARCH) est utilisé dans l'analyse des séries présentant des fluctuations irrégulières, notamment dans les domaines financiers, les sciences sociales, la santé publique, et l'épidémiologie, entre autres.

Le modèle ARCH s'applique lorsque la variance courante du terme d'erreur (innovation) *n'est plus constante*, mais dépend des carrés des innovations des périodes antérieures. Par conséquent, il s'agit d'une forme particulière d'hétéroscédasticité (voir Engle (1982)).

$$\sigma_\varepsilon^2 = f(\varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \dots, \varepsilon_{t-q}^2)$$

Les modèles ARCH sont couramment utilisés pour modéliser des séries financières chronologiques présentant une volatilité marquée, c'est-à-dire des périodes de grandes oscillations suivies par des périodes relativement calmes.

*Prenons un exemple introductif :*

Dans l'analyse traditionnelle de la prévision, la construction des valeurs prévues est fondée sur la moyenne conditionnelle de la série utilisée. Ainsi, la prévision de  $Y_t$  à la date  $t + 1$  compte tenu du passé est donnée par  $\mathbb{E}(Y_{t+1} \mid y_t, y_{t-1}, \dots)$ .

*Exemple numérique:*

Soit  $Y_t \sim AR(1)$  de moyenne 0. En d'autres termes,

$$y_t = \phi y_{t-1} + \varepsilon_t \quad \text{où } \varepsilon_t \sim \text{BB}(0, \sigma_\varepsilon^2).$$

Il est facile de voir que la moyenne non conditionnelle de  $y_t$  est

$$\mathbb{E}(y_t) = 0,$$

et la moyenne conditionnelle de  $y_t$  est :

$$\begin{aligned} \mathbb{E}(y_t \mid y_{t-1}) &= \mathbb{E}(\phi y_{t-1} \mid y_{t-1}) + \mathbb{E}(\varepsilon_t \mid y_{t-1}) \quad \text{car } \varepsilon_t \text{ est indépendant de } y_{t-1}, \\ &= \phi y_{t-1}. \end{aligned}$$

L'idée de Engle (1982) est de tenir compte des autres moments conditionnels de ce processus :

- La variance non conditionnelle de  $AR(1)$  est :

$$\mathbb{V}ar(y_t) = \frac{\sigma_\varepsilon^2}{1 - \phi^2}, \quad \phi \in [0, 1].$$

- La variance conditionnelle est égale à :

$$\begin{aligned} \mathbb{V}ar(y_t \mid y_{t-1}) &= \mathbb{V}ar(\phi y_{t-1} \mid y_{t-1}) + \mathbb{V}ar(\varepsilon_t \mid y_{t-1}) \\ &= \mathbb{V}ar(\varepsilon_t) \\ &= \sigma_\varepsilon^2. \end{aligned}$$

*Il est facile de voir que  $\mathbb{V}ar(y_t) > \mathbb{V}ar(y_t \mid y_{t-1})$ .*

C'est pourquoi Engle (1982) a proposé de modéliser simultanément la moyenne (représentée par le modèle  $AR(1)$ ) et la variance d'une série chronologique.

Le principe consiste à supposer que la variance dépend des informations antérieures. Il propose une spécification ARCH(q) où le carré des perturbations suit un processus autorégressif AR(q).

La famille des modèles ARCH peut se décomposer en deux sous ensemble

Modèle ARCH linéaire	Modèle ARCH non-linéaire
ARCH(q)	EGARCH(p,q)
GARCH(p,q)	TARCH(q)
IGARCH(p,q)	TGARCH (p,q)
caractérisé par des spécification quadratique de la variance conditionnelle des perturbations	caractérisé par des spécification asymétriques des perturbations

#### 6.4.1 Définition et représentation

Un processus  $\varepsilon_t$  est dit ARCH(q) s'il satisfait les deux conditions suivantes:

1.  $\mathbb{E}(\varepsilon_t/\varepsilon_u, u < t) = 0, \quad \forall t \in \mathbb{Z}$
2.  $\sigma_t^2 = \text{Var}(\varepsilon_t/\varepsilon_u, u < t) = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2$

Le modèle ARCH sera écrit de la façon suivante:

$$\varepsilon_t = \sigma_t z_t \quad \text{où } z_t \sim BB(0, \sigma_\varepsilon^2)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^n \alpha_i \varepsilon_{t-i}^2.$$

Autre écriture:

$$\varepsilon_t = z_t \sqrt{\alpha_0 + \sum_{i=1}^n \alpha_i \varepsilon_{t-i}^2}$$

**Exemple:**

Le modèle ARCH(1) s'écrit sous la forme:

$$\varepsilon_t = \sigma_t z_t$$

avec

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \quad \text{et} \quad z_t \sim BB(0, \sigma^2),$$

où  $\sigma^2$  est une constante.

On peut écrire le modèle aussi de la façon suivante

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + u_t,$$

où  $u_t$  est le processus d'innovation pour  $\varepsilon_t^2$  vérifiant  $\mathbb{E}(u_t/\varepsilon_{t-1}) = 0$ .

## 6.5 Autre Modèle

### 1. GARCH (Generalized Autoregressive Conditional Heteroskedasticity)

- Développé par Bollerslev (1987).
- Généralise le modèle ARCH en incluant également des termes autorégressifs.
- Formule générale pour un modèle GARCH(p, q) :

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

- Propriétés :
  - Capable de modéliser une dépendance à long terme dans la variance.
  - La variance conditionnelle dépend des erreurs passées et des variances conditionnelles passées.
  - Souvent utilisé dans la modélisation des séries financières pour capturer des dynamiques complexes de la volatilité.

### 2. EGARCH (Exponential GARCH)

- Proposé par Nelson (1992).
- Utilise une fonction logarithmique pour assurer la positivité de la variance conditionnelle.
- Formule générale pour un modèle EGARCH(p, q) :

$$\log(\sigma_t^2) = \omega + \sum_{i=1}^p \beta_i \log(\sigma_{t-i}^2) + \sum_{j=1}^q \alpha_j \left( \frac{\epsilon_{t-j}}{\sigma_{t-j}} - E \left[ \frac{\epsilon_{t-j}}{\sigma_{t-j}} \right] \right)$$

- Propriétés :
  - Capable de modéliser l'asymétrie dans la volatilité.
  - La variance conditionnelle reste positive grâce à la transformation logarithmique.
  - Utilisé pour des séries temporelles où les chocs positifs et négatifs ont des impacts différents sur la volatilité.

### 3. TGARCH (Threshold GARCH)

- Proposé par Francq and Zakoian (2019).
- Modélise la volatilité en tenant compte des effets seuils asymétriques.
- Formule générale pour un modèle TGARCH(p, q) :

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 + \sum_{k=1}^r \gamma_k \epsilon_{t-k}^2 I(\epsilon_{t-k} < 0)$$

- Propriétés :
  - Modèle bien adapté pour capturer les asymétries dans les séries temporelles financières.
  - La variance conditionnelle dépend des erreurs passées, des variances conditionnelles passées, et des chocs négatifs.

*Le domaine des séries temporelles est très vaste et trouve de nombreuses applications dans divers domaines. Pour plus de détails, voir Brockwell and Davis (2002). Autre références: Autre références:*

- *Campbell et al. (1997)*
- *Chatfield and Xing (2019)*
- *Gouriéroux and Monfort (1995)*

# Bibliography

- Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics*, 542–547.
- Brockwell, P. J. and R. A. Davis (2002). *Introduction to time series and forecasting*. Springer.
- Campbell, J., A. Lo, and A. MacKinlay (1997). *The econometrics of financial markets*. Princeton, nj: Princeton univ.
- Chatfield, C. and H. Xing (2019). *The analysis of time series: an introduction with R*. Chapman and hall/CRC.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, 987–1007.
- Francq, C. and J.-M. Zakoian (2019). *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons.
- Gouriéroux, C. and A. Monfort (1995). *Séries temporelles et modèles dynamiques*. FeniXX.
- Granger, C. W. J. and P. Newbold (2014). *Forecasting economic time series*. Academic press.
- Nelson, D. B. (1992). Filtering and forecasting with misspecified arch models i: Getting the right variance with the wrong model. *Journal of econometrics* 52(1-2), 61–90.