Econométrie II

L'hétéroscédasticité

Prof: Saintilus Jn François

Contenu

- La nature de l'hétéroscédasticité
- L'estimation par les MCO en présence d'hétéroscédasticité
- Les conséquences de l'utilisation des MCO en présence d'hétéroscédasticité
- La détection de l'hétéroscédasticité
- Les mesures correctives
- Pratiques

Définition de l’hétéroscédasticité

- L’une des hypothèses importantes des Moindres Carrés Ordinaires (MCO) dans le cadre du modèle classique de régression linéaire suposse que les erreurs sont homocédastiques, c’est à dire que la variance du terme d’erreur est constante. En revanche (quand cette hypothèse n’est pas respectée) , on dit qu’on est en présence d’un cas d’hétéroscédasticité, donc les éléments se ituant sur la première diagonale de la matrice variance-covariance ne sont pas constants. Ce problème est plus fréquent dans les modèles spécifiées en coupe instantanée.

La Nature de l'hétéroscédasticité

- Nous avons noté que l'une des hypothèses importantes du modèle classique de régression linéaire était que la variance de chaque terme d'erreur $U_{i,}$ dépendant des valeurs choisies de variables explicatives, est un nombre constant égal à σ^2 . C'est l'hypothèse d'homoscédasticité, ou égale (homo) différence (scédasticité), c'est- à-dire variance églae.
- Symboliquement, $E(u^2) = \sigma^2$
- Graphiquement, dans le modèle de régression à deux

variables, l'homoscédasticité peut etre représentée par la figure 11.1 Comme l'indique celle – ci, la variance conditionnelle de Y_i (qui est égale à celle de $U_{i\ ),}$ dépendant du X_i donnée, reste identique quelle que soit la valeur prise par la variable X.

Figure 11.1

- Epargne et revenu

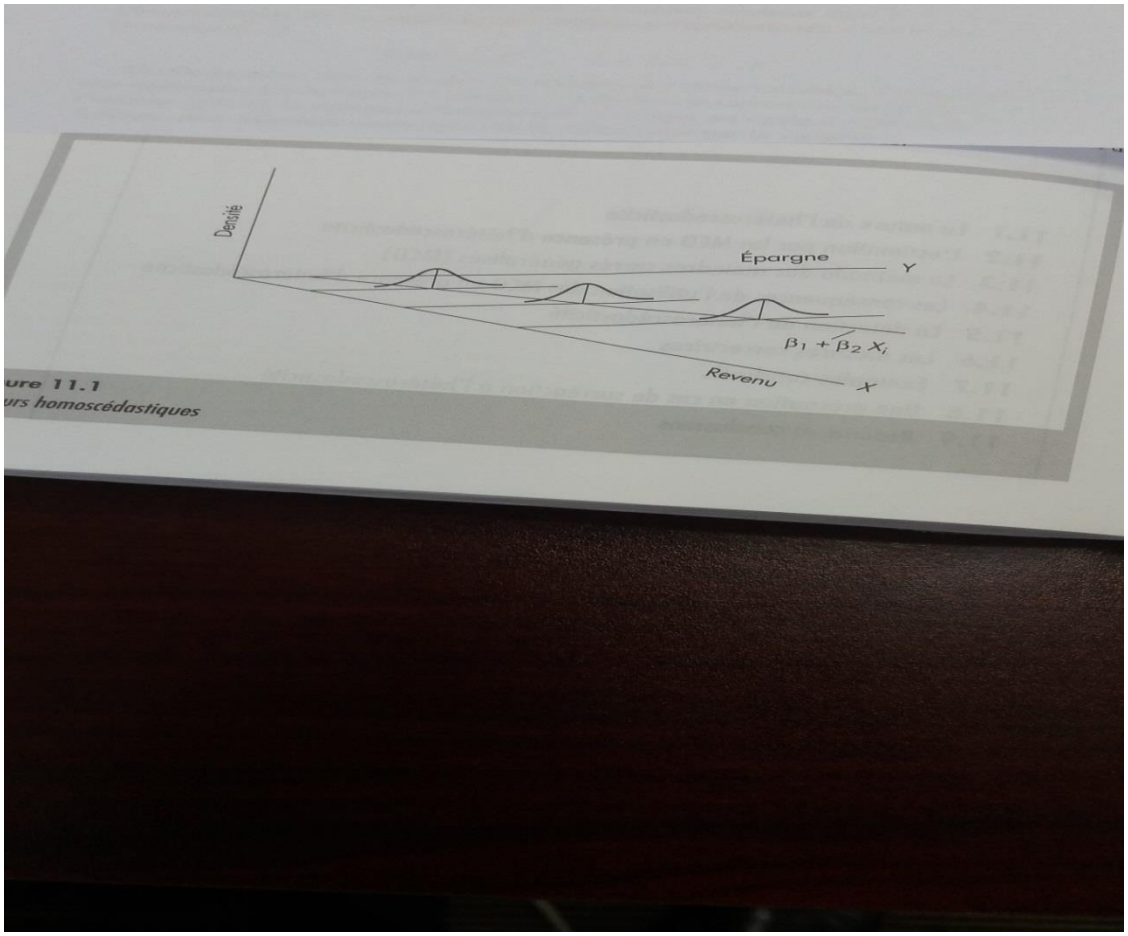


Figure 11.2

- Par opposition, considérons la figure 11.2 ; elle montre que la variance condionelle de Y_i croit lorsque X aumente. Dans ce cas, les variances de Y_i ne sont pas identiques:il y a hétéroscédasticité.

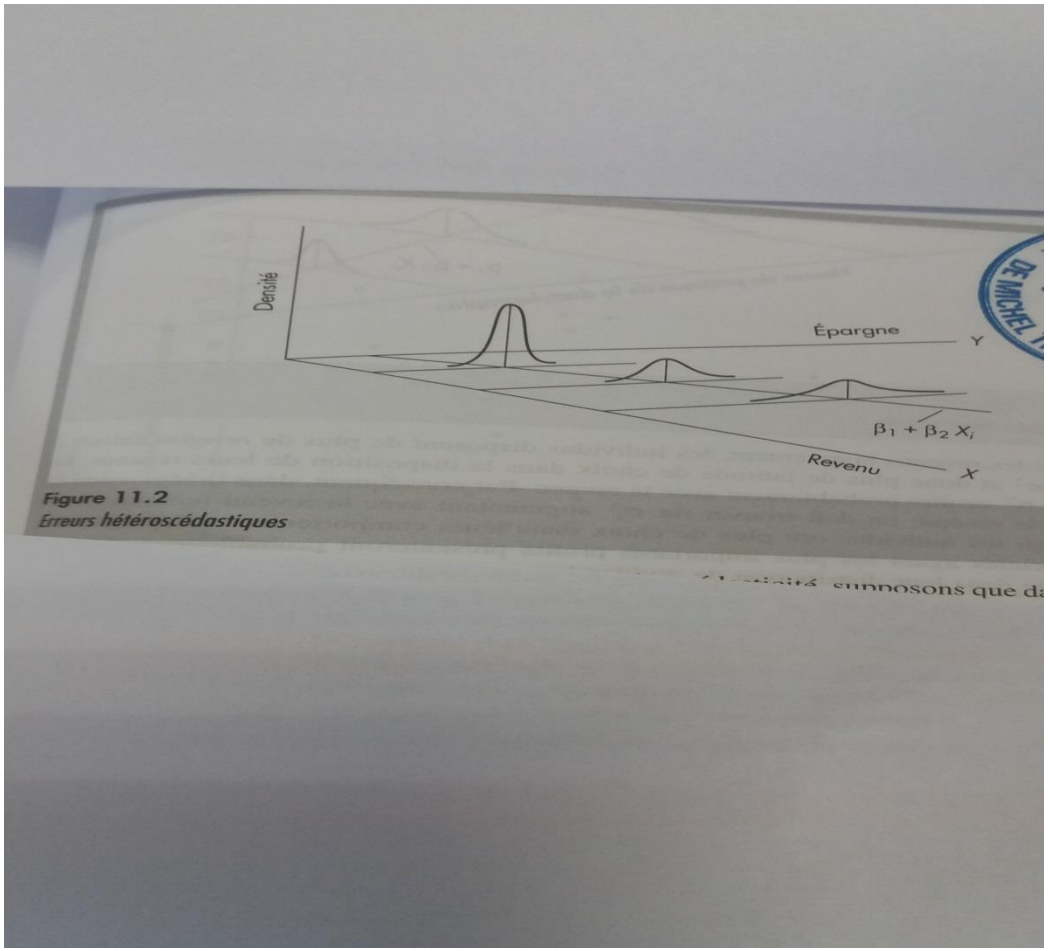


Figure 11.2
Erreurs hétéroscédastiques

Interprétation Figures 11.1 et 11.2

- Pour différencier l'homoscédasticité de l'hétéroscédasticité, supposons que dans le modèle à deux variables $Y_{i} = \beta_{1+} \beta_{2} X_{i} + U_{i, Y}$ représente l'épargne et X le revenu. Les figures 11.1 et 11.2 montrent que, la croissance du revenu est associée, en moyenne, à celle de l'épargne. Mais, sur la figure 11.1, la variance de l'épargne reste la meme pour tout niveau de revenu, alors que sur la figure 11.2, elle augmente avec le revenu.

Conséquences d'hétéroscédasticité : problèmes

Les causes d'hétéroscédasticité :

Lorsque la variance des erreurs ne sont plus constantes sur la première diagonale $V(UI) \neq \sigma^2 I$ on parle de présence d'hétéroscédasticité dans ce cas les estimateurs des MCO sont sans biais mais non efficace (la variance n'est plus minimale).Le niveau de signification et la puissance des tests usuels ainsi que les intervalles de confiance sont fortement affectés.

ce phénomène peut être expliqué par plusieurs raisons:

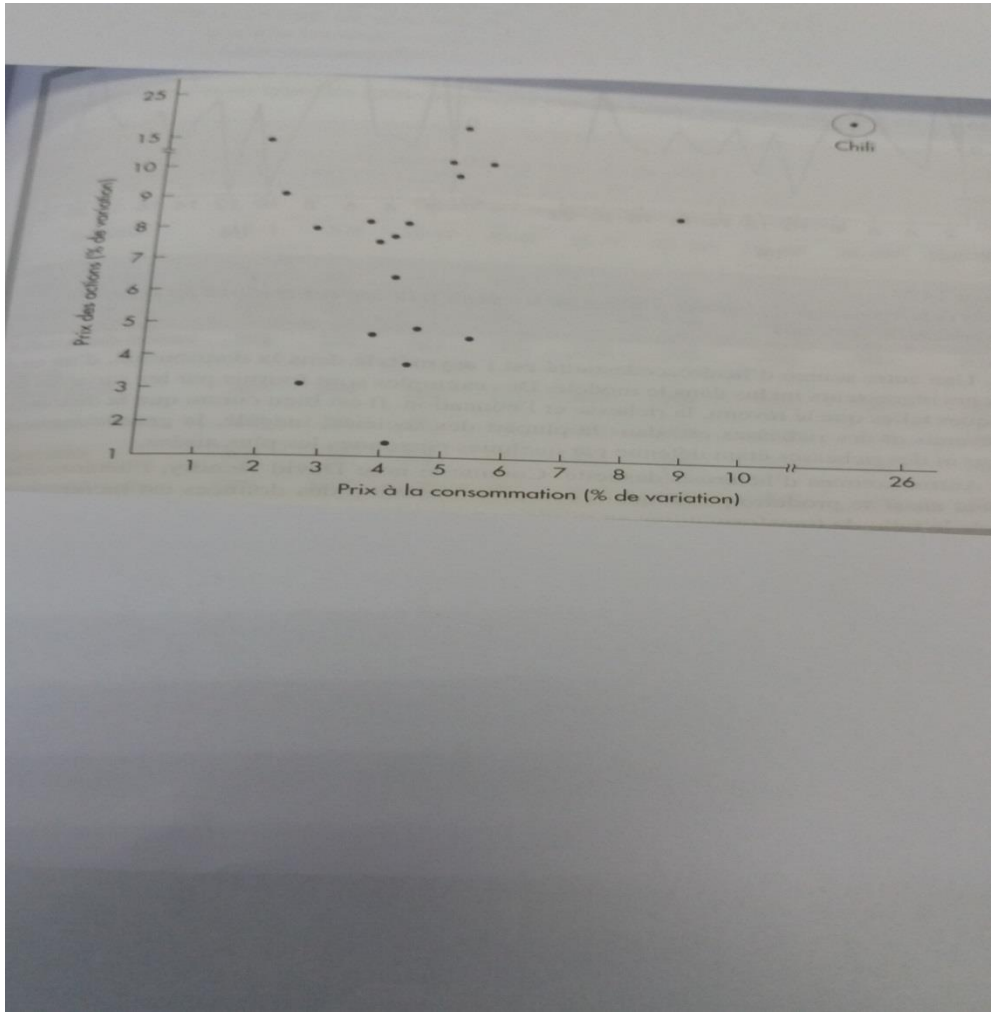
1) la répétition d'une même valeur de la variable à expliquer pour des valeurs différentes d'une variable explicative ;
2) la présence des moyennes calculée sur des échantillons de taille différente ;
3) lorsque les erreurs sont liées aux valeurs prises par une variable explicative, dans un modèle en coupe instantanée la variance de la consommation croit, par exemple, avec le revenu disponible, etc.…
4) la présence d'observations isolées. (une observation isolée est celle qui est très différente (soit plus élevée, soit plus faible) par rapport à celles de l'échantillon.
5) Des erreurs de spécification du modèle (conséquence de l'omission de variables importantes dans le modèle).

6) l'asymétrie dans la distribution d'un ou de plusieurs régresseurs inclus dans le modèle.

7) Autres sources: 1) la transformation des données est incorrecte, 2) la forme fonctionnelle est incorrecte.

Observation isolée

Observation isolée



La Détection de l'hétéroscédasticité :

Comment peut-on savoir si l'hétéroscédasticité est présente dans une situation donnée? Il n'existe de règles abosolues de détection, seulement des méthodes empirique.

Les méthodes informelles (la méthode graphique)

La méthode graphique. S'il n'existe pas d'a *priori* ou d'information empirique sur la nature de l'hétéroscédasticité , on peut, en pratique, faire une analyse de régression à partir de l'hypothèse qu'il n'y a pas l'hétéroscédasticité et faire un examen a posteriori des résidus au carré (U^2) pour voir s'ils présentent une tendance systématique. S'il n'y a pas de configuration systématique entre les deux variables, ce qui suggère que l'hétéroscédasticité est absente.

•

Détection d'hétéroscédasticité

- Il existe une multiplicité de méthode pour détecter la pertinence d'hétéroscédasticité des erreurs dans un modèle de régression linéaire. Nous citons les méthodes informelles (graphiques) et formelles des résidus.

Les méthodes formelles

Le test de Park. Park a formulé la méthode graphique en proposant que σ^2 est une fonction quelconque de la variable explicative X_i. La forme fonctionnelle proposée était

- $\sigma^2 = \sigma^2 X^{\beta} e^{v}$ ou $\ln \sigma^2 = \ln \sigma^2 + \beta \ln X_i + v_i$
- Où v_i est le terme d'erreur stochastique.
- Puisque σ^2 n'est généralement pas connue, Park suggéra d'utiliser u^2 comme substitut et posa la régression suivante:
- $\ln u^2 = \ln \sigma^2 + \beta \ln X_i + v_i$ (11.5.2)
- $\ln u^2 = \alpha + \beta \ln X_i + v_i$
- Si β s'avère etre statistiquement significatif, cela suggérerait la présence d'hétéroscédasticité dans les données. S'il n'est pas significatif, on peut admettre l'hypothèse d'homoscédasticité. Le test de Park est donc un procédé en deux étapes. Dans la 1ère, on execute la régression MCO sans tenir compte de l'hétéroscédasticité. De cette régression, on tire $\hat{U}$ estimé et dans la 2eme etape, on fait la régression.

Les méthodes formelles

- **Le test de Park.** Park a formulé la méthode graphique en proposant que σ^2 est une fonction quelconque de la variable explicative $X_{i.}$ La forme fonctionnelle proposée était
- $\sigma^2 = \sigma^2 X^{\beta} e^{v}$ ou $\ln \sigma^2 = \ln \sigma^2 + \beta \ln X_i + v_i$
- Où v_i est le terme d'erreur stochastique.
- Puisque σ^2 n'est généralement pas connue, Park suggéra d'utiliser u^2 comme substitut et posa la régression suivante:
- $\ln u^2 = \ln \sigma^2 + \beta \ln X_i + v_i$ (11.5.2)
- $\ln u^2 = \alpha + \beta \ln X_i + v_i$
- Si β s'avère etre statistiquement significatif, cela suggérerait la présence d'hétéroscédasticité dans les données. S'il n'est pas significatif, on peut admettre l'hypothèse d'homoscédasticité. Le test de Park est donc un procédé en deux étapes. Dans la 1ère, on execute la régression MCO sans tenir compte de l'hétéroscédasticité. De cette régression, on tire u estimé et dans la 2eme etape, on fait la régression.

Exemple Test de Park

Rémunération par fonctionnaires et tabela 7.3 Orlando

Rémunération moyenne (Y)	Productivité moyenne (X)
3396	9355
3787	8584
4013	7962
4104	8275
4146	8389
4241	9418
4388	9795
4538	10281
4843	11750

Exemple Test de Park

Les résultats du modèle sont les suivants:

- $\hat{y} = 1992{,}3452 + 0{,}2329\ X \quad (1)$
- $\quad\quad (936{,}4791) \quad (0{,}0998)$

$t = (2{,}1275) \quad\quad (2{,}333) \quad\quad R^2 = 0{,}4375$

Les résultats ont révélé que le coefficient angulaire estimé est significatif au niveau de 5%.

- Maintenant, calculons la régression des résidus obtenus dans l'équation (1) **.**
- $\ln u^2 = 35{,}817 - 2{,}8099 \ln X$
- $\quad\quad (38{,}319) \quad (4{,}216)$

$t = (0{,}934) \quad\quad (-0{,}667) \quad\quad R^2 = 0{,}0595$

Il n'y a pas de relation statistiquement significative entre les deux variables. Selon le test de Park, on peut conclure qu'il n'y a pas d'hétéroscédasticité dans la variance des résidus.

Le test de Glejser.

Le test de Glejser. Dans l'esprit, ce test est proche de celui de Park. Après avoir obtenu les résidus U de la régression MCO. Ce test permet de déterminer la forme de la corrélation qui existe entre la variable Xj et la variance des erreurs. Il se base sur la régression des résidus du modèle $Y_i = \beta_1 + \beta_2 X_{2i+} \beta_3 X_3 + ... + \beta_{ki} X_k + U_i$

On teste ensuite la significativité de α_1 dans l'estimation par MCO des modèles suivants :

$|U| = \alpha_0 + \alpha_1 X$. Si α_1 est statiquement significatif dans l'une des régressions citée on accepte alors la présence d'hétéroscédasticité dans le modèle du départ

La valeur absolue des résidus obtenus de la régression sont utilisés dans le calcul de la régression contre la productivité moyenne (X). On a obtenu les résultats

- $|u| = 407,2783 - 0,0203X$
- (633,1621) (0,0675) $R^2 = 0,0127$
- t = (0,6432) (-0,3012)
- Il n' y a pas beaucoup de relation systématique entre les valeurs absolues des résidus et les valeurs estimées de X puisque la valeur de t des coefficients de pente n'est pas systématiquement significative. Donc, l'hétéroscédasticité n'est pas présente.

Pratique test de Glejser

- Tabela 7.3 orlando

Le test de corrélation par ordre de Spearman

- Définition de coefficient de corrélation par ordre de Spearman.
- $r_s = 1 - 6[\Sigma d^2 / n(n^2 - 1)]$
- $d_{i\ = \text{différence}}$ des classifications attribuées à deux caractéristiques différentes (individu ou phénomène) et n = nombre d'individus ou phénomènes classifiés. Le coefficient de corrélation d'ordre précédent peut être utilisé pour détecter l'hétéroscédasticité.

Le test de corrélation par ordre de Spearman

- Supposons. $Y_i = \beta 0 + \beta_1 X_i + U_i$
- Etape 1: Ajuster la régression aux donnéees en y et X et obtenir les résidus (U).
- Etape 2: Ignorer le signe de U ou bien prendre sa valeur absolue |U| , ordonner |U| et X_i (ou Y estimé) de manière ascendante ou et descendante et calculer le coefficient de corrélation.
- Etape 3: page 385 (11.5.7)

Pratique:**Le test de corrélation par ordre de Spearman**

- Tabela 11.2 Damodar

Test de Goldfeld-Quandt

- Cette méthode populaire est applicable lorsqu'on suppose que la variance hétéroscédastique σ^2 ait une relation positive avec l'une des variables indépendantes dans le modèle de régression. Considérons le modèle habituel
- $Y_i = \beta_1 + \beta_2 X_{2i+} \beta_3 X_3 + ... + \beta_{ki} X_k + U_i$
- Supposons que la variable x_j soit la source de l'hétéroscédasticité de ce fait on pourra stipuler que $V(U_i) = f(X_j) = \sigma^2 X^2$ Ceci représente une violation de l'hypothèse d'homoscédasticité.

Test de Goldfeld-Quandt

- Goldfeld-Quandt, Pour tester cette hypothèse on procède comme suit :
  1ère étape : On classe par ordre croissant les données de la variable Xj

- **2ème étape :** On omet de l'échantillon, c observations centrales et on divise le reste en deux sous échantillons de même taille ($\frac{n-C}{2}$)

  . n étant la taille de l'échantillon initial, c est
  généralement le quart de l'ensemble des observations .

- **3ème étape :** On effectue séparément les estimations par MCO des deux sous échantillons et sauve les SCR de chacun des deux régressions (SCR1 et SCR2). Chacun a ($\frac{n-C}{2}$) - k son degré de liberté où k nombre de paramètres à estimer y compris le terme constant. Pour le cas de deux variables, évidemment k égale à 2.

- **4)** Sous l'hypothèse d'homoscédasticité H0 :($\sigma_1^2 = \sigma_2^2$) le rapport des variations résiduelles : F= (SCR_2/dl)/SCR_1/dl une distribution F le numérateur et le dénominateur
- ($\frac{n-C-2k}{2}$) degré de liberté.
- Si *F (calculé)* ≤ *F(tabulé)* on accepte H0 sinon il y a présence d'hétéroscédasticité des erreurs.
  Notons enfin que le numérateur prend toujours la valeur la plus élevée des SCR on pourra donc calculer SCR1/SCR2 si SCR1>SCR2.

Exemple : Goldfeld-Quandt,

- Page 388 tabela 11.3 Damodar et tabela 7.4 Orlando
-

Le test de White

- Soit le modèle de départ : $Y_i = \beta_1 + \beta_2 X_{2i+} \beta_3 X_3 + ... + \beta_{ki} X_k + U_i$ (11.5.21)
- C'est le test le plus utilisé et le plus général puisqu'on n'impose aucune forme a priori de l'hétéroscédasticité. On fait la régression suivante : Pour réaliser le test de White, on procède ainsi:
- **Etape1:** Avec les données, calculons l'équation (11.5.21) et obtenir les résidus U_i .
- **Etape 2**: Alors, réaliser la régression suivante (auxiliaire):
- $U^2_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X^2_{2i} + \alpha_5 X^2_{3i} + \alpha_6 X_{2i} X_{3i} + v^i$ (11.5.22)
- On voit donc le caractère général de ce test puisque dans les variables explicative on trouve les xi originales leurs carrés et leurs produits croisés.
- La statistique de test de White repose sur le calcul du coefficient de détermination R^2 tiré de la régression auxiliaire. (11.5.22)
- **Etape 3:** Sous l'hypothèse qu'il n'y a pas d'hétéroscédasticité, on peut montrer que la taille d'échantillon (n) multipliée par R^2 de la régression auxiliaire suit asyntotiquement la distribution de χ^2 avec degré de liberté égal au nombre de régresseurs (exclui le terme constant) dans la régression auxiliaire. C'est –à-dire, $n R^2 = \chi^2_{(p)}$ p étant le nombre de régresseurs dans la régression auxiliaire ($p = \frac{k(k+3)}{2}$) k est le nombre de variable explicative dans la régression du départ. La règle de décision est toujours la même.
  Etape 4: Si la valeur de χ^2 obtenue dans l'étape3 excède la valeur critique de χ^2 au niveau de α choisi, la conclusion est qu'il n'y a pas d'hétéroscédasticité.
  : Si la valeur de χ^2 obtenue dans l'étape3 n'excède pas la valeur critique de χ^2 au niveau de α choisi, la conclusion est qu'il n'y a pas d'hétéroscédasticité.

Pratique test de White

Le test de Pesaran

- Soit le modèle de départ :$Y_i = \beta_1 + \beta_2 X_{2i+} \beta_3 X_3 + ...+ \beta_{ki} X_k + U_i$ (11.5.21)
- La régression des résidus au carré sur le carré des valeurs estimées de la variable dépendante. $U^2_i = \alpha_1 + +\alpha_1(Yestimee)$.
- Après on teste la significativité de la variable estimée. Si elle est significative, l'hétéroscédasticité est présente.

Test de Breusch-Pagan- Godfrey

- Soit le modèle $Y_i = \beta_1 + \beta_2 X_{2i+} \beta_3 X_3 + ... + \beta_{ki} X_k + U_i$ (11.5.12)
- Supposons que la variance des résidus soit ecrite ainsi:
- $\sigma^2_i = f(\alpha_1 + \alpha_2 Z_{2i} + ... + \alpha_m Z_{mi})$ où σ^2_i est une fonction des variables non stochastiques Z; quelques ou toutes les X' peuvent servir comme Z'. Spécifiquement, supposons que $\sigma^2_i = \alpha_1 + \alpha_2 Z_{2i} + ... + \alpha_m Z_{mi}$ où σ^2_i est une fonction linéaire des Z. Si $\alpha_2 = \alpha_{3.....} = \alpha_m = 0$ et $\sigma^2_i = \alpha_1$ qui est une constante. Pour tester si σ^2_i est homoscédastique , nous pouvons tester l'hypothèse $\alpha_2 = \alpha_{3.....} = \alpha_{m.} = 0$. C'est l'idée de base du test de Breusch-Pagan-Godfrey (BPG). Suivons les procédures pour le test:
- **Etape 1:** Calculer l'équation par MCO et obtenir les résidus $\hat{u}_1, \hat{u}_2, ..., \hat{u}_n$
  Etape 2: Obtenir $\sigma^2_{i} = \Sigma \hat{u}^2_i / n$
- **Etape 3:** Construire les variables Pi définies comme $Pi = \hat{u}^2_i / \sigma^2_i$
- **Etape 4**: Faire la régression de Pi ainsi construites sur les Z comme

$Pi = = \alpha_1 + \alpha_2 Z_{2i} + ... + \alpha_m Z_{mi}$ (11 . 5.15)

Test de Breusch-Pagan- Godfrey

- Etape 5: Obtenir la SCE de l'équation (11.5.15) et définir
- $\Theta = \frac{1(SCE)}{2}$ supposons que les Ui aient une distribution normale, on peut montrer qu'il y a homoscédasticité si la taille d'échantillon augmente de manière indéfinie, alors $\Theta \sim \chi^2_{m-1}$ où Θ suit une distribution Khi-deux avec (m-1) degré de liberté .
- Si dans une application ou Θ ($=\chi^2$) calculée est supérieure à la valeur critique χ^2 au niveau α choisi, nous pouvons rejeter Ho d'homoscedasticité; dans le cas contraire Ho ne sera pas rejetée.

- Pratique tabela 11.3

Correction de l'hétéroscédasticité causée par Xj :

- Les éminents chercheurs à l'instar de Park et Glesjer supposent que la pertinence d'hétéroscédasticité soit due à la corrélation qui existe entre le vecteur des résidus et l'une des variables explicatives. Dans ce cas, leur test est fondé sur la relation entre le résidu du modèle initial en valeur absolue et la variable explicative retenue. Ils proposent différentes formes de relation, par exemple:

Modèles intermédiaires	**type d'hétéroscédasticité**
$\|U\| = a1 + b1Xi + Ui$ (1)	$\sigma^2_U = K X^2$ k est une constante
$\|U\| = a2 + b2\sqrt{Xi} + Ui$ (2)	$\sigma^2_U = KXi$
$\|U\| = a3 + b3\frac{1}{Xi} + Ui$ (3)	$\sigma^2_U = K X^{-2}$

- Si le parametre bj(j = 1, 2, 3) de l'une des spécifications est significativement différent de zéero par un test de Student, alors l'hypothèse d'homoscédasticité est rejetée.
- Si le parametre b est significatif dans plusieurs modèle intermédiaire, on retiendra la forme du modèle intermédiaire dont le t calculé de student est le plus élevé afin de corriger la présence d'hétéroscédasticité.

b_3

Correction de l'hétéroscédasticité causée par Xj :

La correction du modèle initial se fait dépendamment du type d'hétéroscédasticité constaté à la section précédente:

1. Si $\mathbf{b_1}$ est statistiquement significatif, ceci suppose que la variance du terme d'erreur est proportionnelle à X^2 , donc la varaince du terme aléatoire n'est pas constante, contrairement à la 2eme hypothèse des MCO. Pour remédier à ce problème, on doit diviser chaque terme du modèle initial par la variable.

2. Si $\mathbf{b_2}$ est statistiquement significatif selon la spécification du modèle intermédiaire 2, alors la variance du terme d'erreur est proportionnelle à la variance exogène X, la correction du modèle initial passe par la division de chaque terme par $\sqrt{X}$.

3. Si $\mathbf{b_3}$ est statistiquement significatif, alors nous déduisions que la variance du terme d'erreur est proportionnelle à X^{-2} , pour pallier à cet incovénient, divisons chaque terme du modèle initial par X^{-1} . On multiplie chaque terme par X.

.

Correction de l'hétéroscédasticité causée par Xj :

- **<u>Moindres Carrés Généralisés (MCG)</u> :** Reprenons le modèle de départ : $Y_i = \beta_1 + \beta_2 X_{2i+} \beta_3 X_3 + ... + \beta_{ki} X_k + U_i$ et
- et supposons que : $V(u) = \sigma^2 X^2_j$
- L'estimateur BLUE d'un modèle hétéroscédastique est alors celui des MCG.
- La méthodes des MCG permet d'obtenir des estimateurs sans biais quelque soit la matrice variance-covariance des résidus. Cette découverte nous permet également d'estimer n'importe quel modèle en présence du cas d'hétéroscédasticité.
- **<u>Moindres Carrés Pondérés (MCP) :</u>**
  Pour rendre les erreurs homoscédastiques il faut transformer le modèle du départ afin d'avoir une variance constante ceci est possible si on pose: $\hat{u} = u / X j$ t= 1, ...n