

Econométrie: le cours

Institut Supérieur d'Economie et de Management

Licence 3ème année EG

*Catherine Laffineur, Maître de conférences en sciences
économiques*

catherine.laffineur@unice.fr

Introduction

- Le rôle de l'économétrie est d'analyser la corrélation entre deux variables. Cela permet
 - De vérifier la validité de certaines théories économiques
 - Estimer les paramètres d'un modèle économique
- Une régression économétrique permet de décrire et d'évaluer la relation entre une variable dépendante (y) et une ou plusieurs variables indépendantes (x):
 - Lorsqu'il y a une variable indépendante on parlera de modèle de régression simple
 - Lorsqu'il y a plusieurs variables indépendantes on parlera de modèle de régression multiple
- Le modèle est dit linéaire car y_i est une fonction linéaire de a et b :

$$y_i = a + bx_i + u_i$$

- Même si x_i peuvent prendre d'autres formes ($x_i = x_i + x_i^2$,
 $x_i = \frac{K_t}{L_t}$)

Introduction

- Afin d'obtenir des informations sur la relation pour l'ensemble d'une population, on fait de l'inférence statistique:
 - Obtient de l'information sur la population à partir d'un échantillon
 - Echantillon: sous ensemble représentatif la population étudiée
- Dans une régression, la variable y et x sont traitées de manière asymétrique:
 - La variable y est supposée aléatoire ou stochastique
 - La variable x est supposée avoir des valeurs fixes d'un échantillon à l'autre
- En raison du caractère aléatoire de y , il peut exister des déviations de l'espérance conditionnelle de y par rapport à x .

$$y_i = E(y_i|x_i, u_i) = a + bx_i + u_i$$

- a , ordonnée à l'origine (constante, valeur de y lorsque $x=0$)
- b , pente, mesure l'impact marginal, ceteris paribus, de x sur y

Introduction

- La relation spécifiée entre x et y n'est pas déterministe car il existe un terme d'erreur u , appelé résidu
- En effet, il est souvent impossible d'observer la totalité des variables x et y d'une population. Par ailleurs, cette relation n'est pas certaine et peut évoluer:
 - Erreur de spécification: la seule variable explicative n'est pas suffisante pour rendre compte de la totalité du phénomène expliqué
 - Erreur de mesure: les données ne représentent pas exactement le phénomène

Méthode d'estimation

- En économétrie, on s'intéresse à la caractéristique X d'une population
- On dispose d'un sondage de taille n de la population, noté $(x_1, x_2 \dots x_N)$
- Estimer un paramètre θ consiste à donner une valeur approchée à ce paramètre à partir d'un sondage de la population.
- Dans le cas de la régression linéaire, on cherchera à estimer \hat{b} et \hat{a}
- Nous allons analyser différents types d'estimation, qui serviront de rappel de notions de statistiques simples

Estimation de l'espérance

- Supposons que l'on cherche à estimer θ , représentant l'espérance de la population X
- $\theta = E(X) \rightarrow \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$
- \bar{x} est une estimation de θ
- Ainsi, l'estimateur de l'espérance est la moyenne empirique
- Rappel:
 - $E(a) = a$
 - $E(aX) = aE(X)$
 - Lorsque X et Y sont indépendantes $E(XY) = E(X)E(Y)$

Estimation de la variance

- Supposons que l'on cherche à estimer la variance
 $\sigma^2 = V(X) = E((X - E(X))^2)$
- L'estimateur de la variance se définit par: $V(x) = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$
- Rappel: $V(x) = E(x^2) - E(x)^2$

L'estimateur doit être le meilleur possible

- Nous verrons ceci plus en détails dans les chapitres suivants, mais une mesure de précision est l'erreur quadratique moyenne (MSE)

- Le but étant d'avoir l'erreur d'estimation la plus petite possible

- $MSE = E(\hat{b} - b)^2 = 0 \iff$

$$MSE = \underbrace{E(\hat{b} - E(\hat{b}))^2}_{\text{Variance de l'estimateur}} + \underbrace{E(E(\hat{b}) - b)^2}_{\text{Biais de l'estimateur}}$$

- Un bon estimateur doit être précis : il l'est d'autant plus que son erreur quadratique est faible:

- C'est à dire que l'estimateur est sans biais: l'estimation du paramètre s'établit, en moyenne, sont autour de ce paramètre (**BIAIS**)

- Et qu'il a une variance minimale (la plus petite variance possible) (**PRECISION**)

- Cette propriété est importante parce qu'un estimateur sans biais et de variance asymptotiquement nulle est convergent.

Méthodes d'estimation

- Il existe différentes méthodes d'estimation du modèle linéaire:
 - **Méthode des moments**: vise à estimer les moments de l'échantillon (moyenne, variance). Avantage: aucune hypothèse particulière concernant la distribution des résidus. Exemple: GMM, méthodes des moments généralisés
 - **Maximum de vraisemblance**: vise à maximiser la probabilité d'observer les y_i sachant la valeur de x_i
 - **Méthode des moindres carrés (MCO)**: vise à ajuster le nuage de point à l'aide d'une droite en minimisant la distance au carré entre chaque valeur observée et la droite d'estimation. Formellement, on cherche la somme des carrés des écarts aléatoires.

Théorème de Gauss-Markov

- Si les hypothèses du modèle linéaire sont vérifiées l'estimateur des MCO est le meilleur estimateur linéaire sans biais (BLUE)
 - Estimateur sans biais
 - Estimateur de variance minimale
- Ce résultat est important parce que les hypothèses permettent de ne pas faire d'hypothèse sur la loi suivie par la perturbation
- Nous prouverons ce théorème dans les chapitres suivants.
- Il est donc important de spécifier les **hypothèses du modèle linéaire**
- Puis tester la validité de ces hypothèses, pour utiliser (ou non) les MCO

Chapitre 1: Le modèle linéaire simple

Hypothèses sur le modèle linéaire

- Soit le modèle linéaire suivant $y_i = a + bx_i + u_i$ pour $i = 1, \dots, N$
- H1: $E(u_i) = 0 \forall i$.
→ En moyenne, le terme d'erreur est nul. Les variables spécifiées dans le modèle capturent bien y_i
- H2: x_i est une variable certaine (non stochastique).
→ On cherche à modéliser y conditionnellement aux réalisations de x observées dans notre échantillon. Dès lors que x_i est indépendant de u_i : $E(x_i | u_i) = x_i$
- La variance x est non nulle $V(x) = \sigma_x^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x}_i)^2 \neq 0$
avec $\bar{x}_i = \frac{1}{N} \sum_{i=1}^N x_i$
→ les observations de x ne sont pas toutes égales

Hypotèses sur le modèle linéaire

- H4: $V(u_i) = E(u_i^2) = \sigma^2 \forall i \neq s$
→ La variance est la même pour tous les u_i , on dira que les perturbations sont homoscédastiques
 $E(u_i u_s) = cov(u_i, u_s) = 0$
→ La perturbation de i n'est pas influencée par la perturbation de s
- H5: $u_i \sim N(0; \sigma^2)$
→ Les erreurs sont indépendantes et identiquement distribuées selon la loi normale
- H6: $cov(x_i, u_i) = 0$
→ Il existe une indépendance entre la partie systématique et aléatoire du modèle

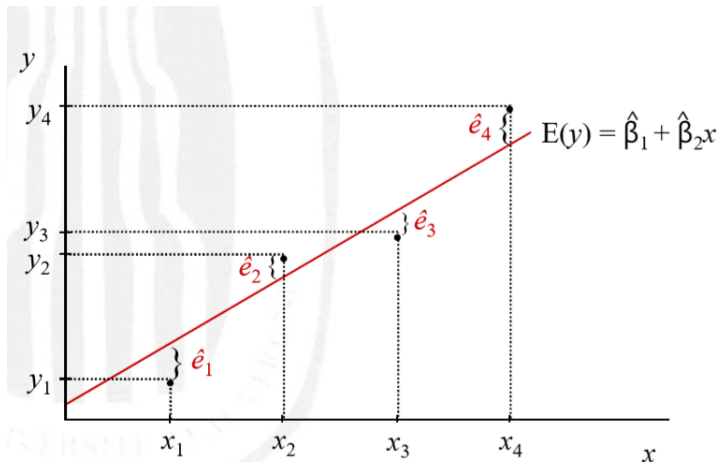
Définition de l'estimateur des MCO

- Il s'agit d'estimer les paramètres a et b du modèle:
$$y_i = a + bx_i + u_i$$
- Pour cela on dispose de données d'un échantillon:

$$\begin{pmatrix} y_1 & x_1 \\ y_2 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ y_N & x_N \end{pmatrix} (N, 2)$$

Définition graphique de l'estimateur des MCO

- Chaque observation peut être représentée par un couple (y_i, x_i) représenté dans un plan (x,y) :



Résolution graphique pour la constante

- Le problème est le suivant: comment choisir les valeurs de a et b telles qu'une droite de la forme $y_i = a + bx_i$ passe le plus près possible de tous les points du nuage?
- La solution graphique consiste à minimiser l'écart entre chaque point et la droite estimée:

$$\min_{a,b} \sum u_i^2 = \min_{a,b} \sum (y_i - a - bx_i)^2 = \min_{a,b} S$$

- $\frac{\delta S}{\delta a} = 0 \iff \bar{y}_i = \hat{a} + \hat{b}\bar{x}_i$

- $\hat{a} = \bar{y} - \hat{b}\bar{x}$

Résolution graphique pour la pente

- $\frac{\delta S}{\delta b} = 0 \longleftrightarrow$

- $\hat{b} = \frac{\sum x_i (y_i - \bar{y})}{\sum x_i (x_i - \bar{x})}$

- $\hat{b} = \frac{\sum y_i (x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$

- $\hat{b} = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$

Poids des estimations dans l'estimateur des MCO

- $$\hat{b} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{\sum(x_i - \bar{x})^2 \left(\frac{y_i - \bar{y}}{x_i - \bar{x}} \right)}{\sum(x_i - \bar{x})^2}$$
- $$\hat{b} = \sum \frac{y_i - \bar{y}}{x_i - \bar{x}}$$
- Ceci correspond à la pente du point $M(x_i, y_i)$ rejoignant le point moyen de l'échantillon
- Ainsi une observation x_i influence d'autant plus la pente \hat{b} qu'elle est éloignée du point moyen
- La procédure pour minimiser ces effets, est d'éliminer les outliers

Précision des estimateurs

- Nous avons vu en introduction que la précision d'un estimateur dépend de sa variance et de son espérance.
- Le biais: un estimateur est sans biais si $E(\hat{b}) = b$ et $E(\hat{a}) = a$
- La précision: un estimateur est à variance minimale si $V(\hat{b}) = E(\hat{b} - E(\hat{b}))^2$ et $V(\hat{a}) = E(\hat{a} - E(\hat{a}))^2$ est le plus petit possible

Espérance de b

- $\hat{b} = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum x_i (x_i - \bar{x})^2}$
- $\iff \hat{b} = b + \frac{\sum (x_i - \bar{x}) u_i}{\sum (x_i - \bar{x})^2}$
- $\iff E(\hat{b}) = E(b) + E\left(\frac{\sum (x_i - \bar{x}) u_i}{\sum (x_i - \bar{x})^2}\right)$
- $E(\hat{b}) = b$

Espérance de a

- $\hat{a} = \bar{y} - \hat{b}\bar{x}$
- $\iff E(\hat{a}) = E(a) + E(\bar{x}(b - \hat{b})) + E(\bar{u})$
- $\iff E(\hat{a}) = a$

Variance de \hat{a} et \hat{b}

- $V(\hat{b}) = \frac{\sigma^2}{\sum_{i=1}^N (x_i - \bar{x})^2}$
- $V(\hat{a}) = \bar{x}^2 V(\hat{b}) + \frac{\sigma^2}{N}$
- On a vu que la variance de \hat{b} et \hat{a} dépendent de la variance des résidus σ^2 :
- $\hat{\sigma}^2 = V(\sum \hat{u}_i) = E(\sum u_i - E(\sum u_i))^2 = E(\sum u_i^2) - E(\sum u_i)^2$
- $\hat{\sigma}^2 = \frac{SCR}{N-2}$ avec $SCR = \sum_{i=1}^N \hat{u}_i^2$

Résumé

- D'après le théorème de Gauss-Markov, les estimateurs des MCO sont les estimateurs les plus précis dans l'ensemble des estimateurs linéaires sans biais de a et b
- Sous les hypothèses associées d'un modèle linéaire, l'estimateur MCO est dit BLUE:
 - Variance minimale
 - Sans biais

Résumé

Propriétés sur petit échantillon	Propriétés sur échantillon de taille infinie (propriétés asymptotiques)
<u>Sans biais</u> si $E(\hat{\beta}) = \beta$	<u>Asymptotiquement sans biais</u> si $\lim_{n \rightarrow \infty} E(\hat{\beta}) = \beta$
<u>Efficace</u> si les 2 conditions suivantes sont satisfaites: <ul style="list-style-type: none">- Non-biaisé- Variance minimale	<u>Convergent</u> si $p \lim \hat{\beta} = \beta$ (consistant en « français »)
<u>Meilleur Estimateur linéaire sans biais</u> (BLUE) si les 3 conditions sont satisfaites: <ul style="list-style-type: none">- Fonction linéaire des observation de l'échantillon- Non-biaisé- Variance minimale	<u>Efficiéce asymptotique</u> , si les 3 conditions sont satisfaites <ul style="list-style-type: none">- Distribution asymptotique avec moyenne et variance finies- Convergent- Variance asymptotique minimale

Résumé

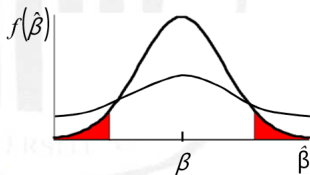
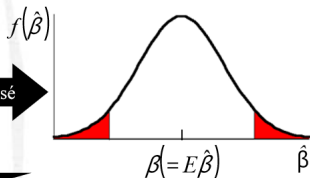
3.

Identification et propriété des estimateurs

Illustrations graphiques

Non-biaisé

Efficace



Inférence statistique

- Dans le modèle de régression, on postule l'existence d'une corrélation entre les variables x et y . Il est de ce fait légitime à travers l'échantillon considéré de confirmer ou d'infirmer empiriquement l'influence de la variable x sur la variable y .
- Pour faire de l'inférence statistique, la moyenne et l'écart type ne suffisent pas. Il est important de déterminer leur loi de distribution pour utiliser des outils de statistiques inférentielles habituels:
 - Test d'hypothèse
 - Intervalle de confiance
 - Puissance du modèle

Coefficient de détermination et puissance du modèle

- On note par convention:

Variation totale (SCT) = Variation expliquée par la régression
(SCE) + Variation résiduelle (SCR)

Ce qui correspond à:

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2 \quad (1)$$

- Le coefficient de détermination correspond à la part de la variation totale qui est expliquée par le modèle:

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} \quad (2)$$

- Supposons $R^2 = 0.9$, cela signifie que 90% de la variation totale de y est expliquée par le modèle.

Intervalle de confiance

- Dans la réalité, on ne dispose que d'un seul échantillon, de taille finie, et donc d'un seul estimateur. Il est nécessaire de pouvoir établir un "diagnostic" à partir de cet estimateur: est ce qu'on est très loin de la vraie valeur?
- Pour cela, la moyenne, ou l'espérance du paramètre ne suffit pas : il faut connaître toute la distribution du paramètre. Il sera alors possible de calculer un intervalle de confiance, c'est-à-dire un intervalle de valeurs où le vrai paramètre appartient avec une probabilité donnée.
- Cet intervalle de confiance va dépendre en particulier de l'écart-type de l'estimateur.

Intervalle de confiance

- Un intervalle de confiance consiste à trouver une estimation par intervalle d'un paramètre θ , c'est-à-dire de construire une "fourchette de valeurs" numériques permettant de situer θ dans un intervalle $\underline{\theta}$ et $\bar{\theta}$ avec une probabilité $1 - \alpha$.

$$P \left[-Z_{\alpha/2} \leq \frac{\hat{b} - b}{\hat{\sigma}_{\hat{b}}} \leq Z_{\alpha/2} \right] = 1 - \alpha$$

Avec

$$t_c^* = \frac{\hat{b} - b}{\hat{\sigma}_{\hat{b}}} \sim t_{N-2} \quad (3)$$

L'intervalle de confiance pour la valeur de b avec une erreur de $\alpha\%$ est:

$$IC(\hat{b}) = \left[\hat{b} - Z_{\alpha/2} \hat{\sigma}_{\hat{b}}; \hat{b} + Z_{\alpha/2} \hat{\sigma}_{\hat{b}} \right]$$

Test de significativité d'un paramètre

- on cherche à savoir si le paramètre est statistiquement significatif, i.e., s'il est significativement différent de 0.

$$\begin{cases} H_0 : B_i = 0 \\ H_1 : B_i \neq 0 \end{cases}$$

Test de significativité d'un paramètre

- Voici les différentes étapes du test:
 - 1) On fixe un risque d'erreur de première espèce α (en pratique on choisit $\alpha = 5\%$)
 - 2) On calcule la statistique de Student donnée par t_c^*
 - 3) On compare la valeur obtenue avec la valeur lue dans la table de la loi de Student notée t_{tab} . Cette valeur correspond à la valeur de la loi de Student à $(N - 2)$ degré de liberté avec un risque d'erreur de $\alpha\%$.
 - 4) Décision du test statistique:
 - Si $|t_c^*| > t_{tab}$, on rejette l'hypothèse H_0
 - Si $|t_c^*| < t_{tab}$, on accepte H_0 .

Significativité globale du modèle

- Pour tester la significativité globale du modèle, on utilise le test de Fisher. L'objectif de ce test est de déterminer si le modèle explique ou non le phénomène étudié.
- Voici les différentes étapes du test:
 - 1) On fixe un risque d'erreur de première espèce α (en pratique on choisit $\alpha = 5\%$)
 - 2) On calcule la statistique de Fisher donnée par l'expression
$$F = \frac{R^2}{1-R^2} (N-2) \sim F(1, N-2)$$
 - 3) On compare la valeur obtenue avec la valeur lue dans la table de la loi de Fisher notée F_{tab} .
 - 4) Décision du test statistique:
 - Si $|F_C^*| > F_{tab}$, on rejette l'hypothèse H_0
 - Si $|F_C^*| < F_{tab}$, on accepte H_0 .

La prévision

- Un des objectifs de l'économétrie est de servir à des fins de prévision.
- Par exemple, si on réalise une modélisation économétrique des ventes d'une entreprise au cours du temps et que l'ajustement linéaire est bon (R^2 élevé), on peut souhaiter utiliser cet outil pour prévoir les ventes futures de l'entreprise.
- Cependant pour pouvoir donner un intervalle de confiance.
- Cet intervalle de confiance dépend de la variance de l'erreur de prévision $V(u_i^*) = V(Y_i^* - \hat{Y}_i)$

La prévision

- On peut prouver que $V(u_i^*) = \sigma^2 \left[1 + \frac{1}{N} + \frac{(x_i^* - \bar{x})^2}{NV(X)} \right]$
- De là on déduit que:

$$u_i^* = Y_i^* - \hat{Y}_i \sim N \left(0, \sigma^2 \left[1 + \frac{1}{N} + \frac{(x_i^* - \bar{x})^2}{NV(X)} \right] \right)$$

- Dès lors, on détermine l'intervalle de confiance pour notre prévision Y_i^* de la façon suivante:

$$P \left(-t_{\alpha/2} \leq \frac{Y_i^* - \hat{Y}_i}{\sqrt{\sigma^2 \left[1 + \frac{1}{N} + \frac{(x_i^* - \bar{x})^2}{NV(X)} \right]}} \leq t_{\alpha/2} \right) = 1 - \alpha$$

La prévision

$$P \left(\hat{Y}_i - t_{\alpha/2} \sqrt{\sigma^2 \left[1 + \frac{1}{N} + \frac{(x_i^* - \bar{x})^2}{NV(X)} \right]} \leq Y_i^* \leq \hat{Y}_i + t_{\alpha/2} \sqrt{\sigma^2 \left[1 + \frac{1}{N} + \frac{(x_i^* - \bar{x})^2}{NV(X)} \right]} \right) = 1 - \alpha$$

- On en déduit l'intervalle de confiance de notre prévision:

$$IC(Y_i^*) = \left[\hat{Y}_i - t_{\alpha/2} \sqrt{\sigma^2 \left[1 + \frac{1}{N} + \frac{(x_i^* - \bar{x})^2}{NV(X)} \right]}; \hat{Y}_i + t_{\alpha/2} \sqrt{\sigma^2 \left[1 + \frac{1}{N} + \frac{(x_i^* - \bar{x})^2}{NV(X)} \right]} \right]$$

Interprétation des coefficients

- Le modèle linéaire peut s'écrire de différentes manières. La forme fonctionnelle adoptée est toujours linéaire mais les variables peuvent être incluses sous différentes formes: en niveau et en logarithme.
- Modèle niveau-niveau:
 - $Sal_i = a + bExp_i + u_i$
 - $\frac{\partial Sal_i}{\partial Exp_i} = b$
 - Puisque la fonction est linéaire, le coefficient s'interprète comme l'effet marginal d'une année supplémentaire d'expérience sur le salaire, toutes choses égales par ailleurs.

Interprétation des coefficients

- Modèle log-log

- $\ln(\text{Sal}_i) = a + b\ln(\text{Exp}_i) + u_i$

- $b = \frac{\partial \text{Sal}_i}{\partial \text{Exp}_i} \frac{\text{Sal}_i}{\text{Exp}_i}$

- Ainsi, dans un modèle log-log, le coefficient b s'interprète comme une élasticité. Elle s'interprète comme la variation de b % de salaire suite à une variation de 1% de l'expérience, toutes choses égales par ailleurs.

Interprétation des coefficients

- Modèle Log-niveau

- $\ln(Sal_i) = a + bExp_i + u_i$

- $100 \times b = \frac{\% \Delta Sal_i}{Exp_i}$

- Ainsi, on peut interpréter $100 \times b$ comme le changement en pourcentage du salaire lorsque le niveau d'expérience augmente d'une unité, toutes choses étant égales par ailleurs: lorsque le niveau d'expérience augmente d'une unité, le salaire augmente de $100 \times b$.

Interprétation des coefficients

- Modèle niveau-log

- $Sal_i = a + b \ln(Exp_i) + u_i$

- $\frac{b}{100} = \frac{\partial Sal_i}{\Delta Exp_i \%}$

- Par conséquent, l'interprétation est la suivante. Une augmentation de 1% du niveau d'étude modifie le salaire augmente de $\frac{b}{100}$ toutes choses égales par ailleurs.

Chapitre 2: Le modèle linéaire multiple

Introduction

- Le modèle reste linéaire mais il existe plusieurs variables explicatives
- Soit un modèle linéaire de la forme:
- $y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \dots + b_k x_{ki} + u_i, i = 1, \dots, N$
- Pour estimer ce modèle on dispose de N observations et k variables. Pour pouvoir estimer ce modèle il est important de respecter la règle suivante:

$$k \leq N$$

- On peut le représenter sous forme matricielle comme:

$$Y_{N \times 1} = X_{N \times (k+1)} B_{(k+1) \times 1} + u_{N \times 1}$$

Hypothèses du modèle

- H1: la valeur de X_i fixe la valeur moyenne de Y_i et toute variation autour de cette moyenne est nulle. Cette condition est la plus importante de toute car elle permet d'obtenir un estimateur sans biais: $E(u) = \begin{matrix} 0 \\ (N,1) \end{matrix} \begin{matrix} (N,1) \end{matrix}$
- H2: Condition **d'homoscédasticité**. En l'absence de cette condition l'estimateur n'est pas efficace. Lorsque cette condition est violée on parle d'hétéroscedasticité: $V(u_i) = \sigma^2 I_N$
- H3: absence d'autocorrélation: $E(u_i u_s) = Cov(u_i u_s) = 0$. Par conséquent: $u \sim N(0, \sigma^2 I_N)$

Les hypothèses du modèle

- La matrice des variables explicatives X est une matrice certaine (exogène)
- X est de rang $Rg(X) = k + 1 < N$, c'est à dire de plein rang $(N, k+1)$ colonne. Cette condition est appelée condition de rang. Elle signifie que parmi les variables qui entrent dans la liste des variables explicatives aucune n'est redondante, c'est-à-dire qu'il n'existe pas, parmi les variables explicatives, de variables dont les valeurs peuvent être déduites de celles prises par les autres variables.

Les estimateurs et leurs propriétés

- $\min_{b_0, \dots, b_k} \sum \hat{u}_i^2 = \min_{b_0, \dots, b_k} \sum (y_i - b_0 - b_1 x_{1i} - \dots - b_k x_{ki})^2$
- $\text{Min}_b \sum_{i=1}^N (y_i - \underset{(1,k)}{X'} \underset{(k,1)}{b})^2$
- $S(b) = \underset{N,1}{(y_i - \underset{(N,k)}{X} \underset{(k,1)}{b})}' \underset{N,1}{(y_i - \underset{(N,k)}{X} \underset{(k,1)}{b})}$
- $S(b) = y'y - b'X'y - y'Xb + b'X'Xb$
- $\frac{\partial S(b)}{\partial b_j} = -2X'y + 2X'Xb = 0$
- $\hat{b} = (X'X)^{-1}X'y$

Interprétation des coefficients

- Supposons que l'on dispose d'un modèle avec deux variables explicatives. En faisant apparaître les deux sous éléments on obtient:

$$\begin{bmatrix} X_1' X_1 & X_1' X_2 \\ X_2' X_1 & X_2' X_2 \end{bmatrix} \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \end{pmatrix} = \begin{pmatrix} X_1' y \\ X_2' y \end{pmatrix}$$

- On peut alors obtenir les valeurs de \hat{b}_1 et \hat{b}_2 en résolvant un système de deux équations:

$$\begin{cases} X_1' X_1 \hat{b}_1 + X_1' X_2 \hat{b}_2 = X_1' y \\ X_2' X_1 \hat{b}_1 + X_2' X_2 \hat{b}_2 = X_2' y \end{cases}$$

Interprétation des coefficients

- On peut alors obtenir les valeurs de \hat{b}_1 et \hat{b}_2 en résolvant un système de deux équations:

$$\begin{cases} X_1' X_1 \hat{b}_1 + X_1' X_2 \hat{b}_2 = X_1' y \\ X_2' X_1 \hat{b}_1 + X_2' X_2 \hat{b}_2 = X_2' y \end{cases}$$

- De la première équation on tire: $\hat{b}_1 = (X_1' X_1)^{-1} (X_1' y - X_1' X_2 \hat{b}_2)$
- En reportant dans la seconde équation, nous obtenons:
 $X_2' X_1 (X_1' X_1)^{-1} (X_1' y - X_1' X_2 \hat{b}_2) + X_2' X_2 \hat{b}_2 = X_2' y$
- Mettons maintenant les termes de y à droite et de \hat{b}_2 à gauche:

$$\left[X_2' X_2 - X_2' X_1 (X_1' X_1)^{-1} X_1' X_2 \right] \hat{b}_2 = X_2' y - X_2' X_1 (X_1' X_1)^{-1} X_1' y$$

Intepétation des coefficients

- En factorisant l'équation à gauche et à droite par $M_1 = I - X_1(X_1'X_1)^{-1}X_1'$ nous obtenons:

$$X_2'M_1X_2\hat{b}_2 = X_2'M_1y$$

- En multipliant la valeur de M_1 par y , nous remarquons que:

$$M_1y = y - X_1(X_1'X_1)^{-1}X_1'y = y - X_1\hat{b}_1$$

- Autrement dit, M_1 appliqué à y conduit au résidu estimé de la régression de y sur X_1 .

Interprétation des coefficients

- En utilisant les propriétés de la matrice M_1 , qui est symétrique et idempotente:

$$M_1' = M_1$$

$$\text{et } M_1 M_1 = M_1$$

- On peut écrire: $X_2' M_1' M_1 X_2 \hat{b}_2 = X_2' M_1' M_1 y$
- Ainsi: $\hat{b}_2 = (X_2' M_1' M_1 X_2)^{-1} X_2' M_1' M_1 y$
- On constate donc que \hat{b}_2 est l'estimateur des MCO dans la régression:

$$M_1 y = M_1 X_2 b_2 + u_i$$

- Ce résultat est celui du **Théorème de Frish-Waugh**. Pour obtenir l'estimation de l'impact de X_2 sur y , il faut "purger" au préalable de l'impact de X_1 .

Interprétation des variables

- Lorsqu'il y a plusieurs variables l'interprétation d'un coefficient s'interprète "toutes choses égales par ailleurs".
- Pour obtenir l'estimation de l'impact de X_2 sur y , il faut "purger" au préalable de l'impact de X_1 .
Ceci est obtenu en régressant:
 - a) $M_1 y$, c'est à dire le résidu de la régression de y sur X_1 (tout ce qui dans y ne vient pas de X_1)
 - b) sur $M_1 X_2$ c'est à dire le résidu de la régression de X_2 sur X_1 (tout ce qui dans X_2 ne vient pas de X_1)

Interprétation avec variables d'interaction

- Interaction entre une variable indicatrice et une variable continue
- $Sal_i = b_0 + b_1 Educ_i + b_2 Sex_i + b_3 (Educ_i \times Sex_i)$
- $y = b_0 + b_1 X_1 + b_2 d + b_3 (X_1 \times d)$

	(1)	(2)
Variable dépendante	Salaire horaire	
Homme	1.19*** (0.40)	1.50** (0.76)
Années d'éducation	0.13*** (0.04)	0.14*** (0.05)
Homme*années d'éducation		-0.04 (0.07)
R2	0.0136	0.0137
Observations	1984	1984

Autre exemple

```
. reg income educ jobexp i.black i.black#c.jobexp
```

Source	SS	df	MS	Number of obs = 500		
Model	33352.2559	4	8338.06397	F(4, 495)	=	604.39
Residual	6828.99339	495	13.7959462	Prob > F	=	0.0000
-----+-----				R-squared	=	0.8300
Total	40181.2493	499	80.5235456	Adj R-squared	=	0.8287
-----+-----				Root MSE	=	3.7143
-----+-----						
income	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	1.834776	.0463385	39.60	0.000	1.743732	1.925821
jobexp	.7128145	.0395293	18.03	0.000	.6351486	.7904805
1.black	.4686862	1.040728	0.45	0.653	-1.576103	2.513475
black#c.jobexp						
1	-.2556117	.0786289	-3.25	0.001	-.4100993	-.1011242
_cons	-5.514076	.9464143	-5.83	0.000	-7.373561	-3.654592
-----+-----						

Interaction entre deux variables dichotomiques

- $y = b_0 + b_1 d_1 + b_2 d_2 + b_3 (d_2 \times d_1)$
- Quatre profils peuvent être distingués:
 - femmes non scolarisées: $y = b_0 + u_i$
 - femmes scolarisées: $y = b_0 + b_1 d_1 + u_i$
 - hommes non scolarisés: $y = b_0 + b_2 d_2 + u_i$
 - hommes scolarisés: $y = b_0 + b_2 d_2 + b_3 (d_2 \times d_1) + u_i$

job_prestige	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
married						
yes	4.3	0.7	5.87	0.000	2.9	5.7
sex						
male	-1.9	0.7	-2.52	0.012	-3.3	-0.4
married#sex						
yes#male	3.6	1.1	3.34	0.001	1.5	5.7
_cons	41.7	0.5	86.62	0.000	40.8	42.7

Interaction entre deux variables continues

```
. reg health age weight c.age#c.weight
```

Source	SS	df	MS			
Model	2059.09026	3	686.36342			
Residual	12975.9311	10331	1.25601889			
Total	15035.0214	10334	1.4549082			

Number of obs = 10335
F(3, 10331) = 546.46
Prob > F = 0.0000
R-squared = 0.1370
Adj R-squared = 0.1367
Root MSE = 1.1207

health	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
age	-.0196621	.0030909	-6.36	0.000	-.0257208 -.0136034
weight	.0018507	.0021221	0.87	0.383	-.0023089 .0060104
c.age#c.weight	-.0000865	.000043	-2.01	0.044	-.0001708 -2.24e-06
_cons	4.512782	.1522368	29.64	0.000	4.214368 4.811196

Biais des estimateurs

$$\begin{aligned} E(\hat{B}) &= E[(X'X)^{-1}X'Y] \\ &= E[(X'X)^{-1}X'(XB + u)] \\ &= E[(X'X)^{-1}X'XB + (X'X)^{-1}X'u] \\ &= E[B + (X'X)^{-1}X'u] \\ &= E(B) + (X'X)^{-1}X'E(u) \\ &= B \end{aligned}$$

$$\begin{aligned}V(\hat{B}) &= E[(\hat{B} - B)(\hat{B} - B)'] \\&= E[(X'X)^{-1}X'u((X'X)^{-1}X'u)'] \\&= E[(X'X)^{-1}X'uu'X(X'X)^{-1}] \\&= (X'X)^{-1}X'E(uu')X(X'X)^{-1} \\&= (X'X)^{-1}X'\sigma^2I_NX(X'X)^{-1} \\&= \sigma^2(X'X)^{-1}\end{aligned}$$

Coefficient de détermination

- $E(u'u) = SCR = Y'Y - \hat{B}'X'Y$
- Coefficient de détermination:
$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} = 1 - \frac{Y'Y - \hat{B}'X'Y}{Y'Y - NY^2}$$

Tests Statistiques

- On peut tester la valeur d'un des paramètres estimés:

$$\begin{cases} H_0 : b_j = m \\ H_1 : b_j \neq m \end{cases}$$

- En utilisant la statistique:

$$t_c = \frac{\hat{b}_j - m}{\sqrt{\hat{V}(\hat{b}_j)}} \sim t_{N-(k+1)}$$

- où $\hat{V}(\hat{b}_j)$ est la j ème composante sur la diagonale principale de la matrice de variance/covariance estimée:

$$\hat{V}(\hat{B}) = \begin{bmatrix} \hat{V}(\hat{b}_0) & \text{Cov}(\hat{b}_0, \hat{b}_1) & \cdot & \cdot & \text{Cov}(\hat{b}_0, \hat{b}_k) \\ \cdot & \hat{V}(\hat{b}_1) & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \hat{V}(\hat{b}_j) & \cdot \\ \cdot & \cdot & \cdot & \cdot & \hat{V}(\hat{b}_k) \end{bmatrix}$$

Violations des hypothèses de Gauss-Markov

- Rappel:

- H1) $E(u) = \begin{matrix} 0 \\ (N,1) \end{matrix}$
- H2) $E(uu') = \sigma^2 I_N$
- H3) $u \sim N(0, \sigma^2 I_N)$
- H4) La matrice des variables explicatives X est une matrice certaine (exogène)
- H5) X est de rang $Rg(X) = k + 1 < N$. Elle signifie que parmi les variables qui entrent dans la liste des variables explicatives aucune n'est redondante.
- H6) $Cov(X, u) = 0 \Rightarrow$ la covariance entre la partie systématique et celle aléatoire est nulle.

Multicolinéarité

- Si les colonnes de la matrice X sont reliées par une relation linéaire, alors la matrice $x'X$ peut être inversée, mais la proximité des variables entre elles rend difficile l'identification de leur effet propre.
- Il est important de se rendre compte que le problème de multicolinéarité ne provient pas de la corrélation entre les variables explicatives prises deux à deux, mais lorsque la corrélation est trop forte ce qui engendre une variance très forte.
- la variance de β_1 est définie de la façon suivante:

$$V(\hat{\beta}_1) = \frac{\sigma^2}{S_{11}(1 - R_{1,-1}^2)}$$

où $S_{1,1} = nV(x_1)$ et $R_{1,-1}^2$ est le R^2 de la régression de x_1 sur X_{-1} .

La multicolinéarité

- Solutions

- Une méthode consiste à calculer la série des ratios $vif_j = \frac{1}{(1-R_j^2)}$
où R_j^2 est le R^2 de la régression de la variable j sur toutes les autres variables, y compris la constante (laquelle est donc présente dans toutes les régressions partielles).
- On soupçonne la présence de multicolinéarité lorsque:
 - Le vif maximum est supérieur à 10
 - La moyenne des vif_j est plus grande que 1

L'hétéroscédasticité

- Le calcul de la variance des estimateurs est important pour pouvoir estimer leur significativité. La variance des estimateurs repose sur la variance du terme d'erreur σ^2 .
- N'oublions pas que σ^2 est une variance estimée par l'économètre.
- Lorsque l'on fait l'hypothèse d'homoscédasticité on fait l'hypothèse que σ^2 est constant pour chaque observation.
- On dit que le modèle est hétéroscédastique lorsque l'hypothèse $V(u_i) = \sigma^2 * I_n$ n'est pas vérifiée.
- La matrice de variance covariance du terme d'erreur s'écrit $V(u_i) = \sigma^2 \Omega$ où Ω est toujours une matrice diagonale mais pas égale à l'identité.

L'hétéroscédasticité

- Sous cette hypothèse on peut établir que:
 - L'estimateur des MCO est sans biais (puisque nous faisons toujours l'hypothèse d'indépendance entre le terme d'erreur et le vecteur des variables explicatives)
 - Mais il n'est plus efficace parmi l'ensemble des estimateurs linéaires sans biais.

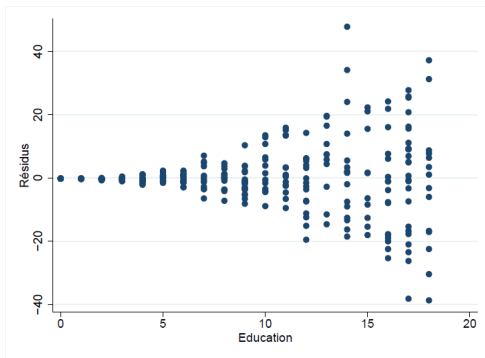
- Preuve:

$$\begin{aligned}V(\hat{B}_{MCO}) &= E[(\hat{B}_{MCO} - B)(\hat{B}_{MCO} - B)'] \\&= E[(X'X)^{-1}X'\varepsilon((X'X)^{-1}X'\varepsilon)'] \\&= E[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}] \\&= (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1} \\&= (X'X)^{-1}X'\sigma^2\Omega X(X'X)^{-1} \\&= \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1} \neq \sigma^2(X'X)^{-1}\end{aligned}$$

Détection visuelle de l'hétéroscédasticité

- On observe que la plage de variation des résidus tend à s'élargir à mesure que l'éducation augmente. Le modèle est donc hétéroscédastique.

Figure: Représentation graphique des résidus



Quels remèdes?

- Lorsque le modèle est hétéroscédastique il existe deux solutions:
 - Transformer le modèle pour le rendre homoscédastique: c'est la méthode des moindres carrés quasi généralisés (MCG)
 - Continuer d'utiliser les MCO mais en corrigeant de la matrice de variances-covariances des estimateurs

La méthode des MCG

- L'idée consiste à transformer le modèle en vue de ramener l'hypothèse $H2^*$ à sa forme initiale $H2$.
- Si la matrice Ω est connue, on peut définir l'inverse de la matrice Ω^{-1} et la racine carrée de cet inverse $\Omega^{-1/2}$.
- La méthode des MCG consiste à multiplier tout le modèle par $\Omega^{-1/2}$ ce qui donne:

$$y\Omega^{-1/2} = \Omega^{-1/2}X\beta + \Omega^{-1/2}\varepsilon$$

- Dans ce modèle transformé le terme d'erreur $\Omega^{-1/2}\varepsilon$ a une matrice de variance covariance égale à:

$$V(\Omega^{-1/2}\varepsilon) = \Omega^{-1/2} V(\varepsilon)\Omega^{-1/2} = \sigma^2\Omega^{-1/2}\Omega\Omega^{-1/2} = \sigma^2I_n$$

La méthode des MCG

- Le modèle transformé est donc homoscédastique. L'estimateur des MCG est obtenu en calculant l'estimateur des MCO du modèle transformé:

$$\beta_{MCG} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}y)$$

- Preuve:
- Notons $\tilde{Y} = \Omega^{-1/2}y$ et $\tilde{X} = \Omega^{-1/2}x$
- Le modèle estimé est donc:

$$\tilde{Y} = \tilde{X}B + \tilde{\varepsilon}$$

La méthode des MCG

- On peut facilement vérifier que cet estimateur est sans biais:

$$\begin{aligned} E(\hat{B}_{MCG}) &= E[(X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y] \\ &= E[(X' \Omega^{-1} X)^{-1} X' \Omega^{-1} (XB + \varepsilon)] \\ &= B + (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} E(\varepsilon) \\ &= B \end{aligned}$$

La méthode des MCG

- Calculons maintenant la matrice de var/cov de \hat{B}_{MCG} .
- Nous savons d'après l'équation précédente que $E(\hat{B}_{MCG}) - B = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} E(\varepsilon)$. Ainsi,

$$\begin{aligned} V(\hat{B}_{MCG}) &= E[(\hat{B}_{MCG} - E(\hat{B}_{MCG}))(\hat{B}_{MCG} - E(\hat{B}_{MCG}))'] \\ &= E[(X' \Omega^{-1} X)^{-1} X' \Omega^{-1} \varepsilon (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} \varepsilon'] \\ &= (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} E(\varepsilon \varepsilon') (\Omega^{-1})' X (X' \Omega^{-1} X)^{-1} \\ &= (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} \sigma^2 \Omega (\Omega^{-1})' X (X' \Omega^{-1} X)^{-1} \text{ (car } E(\varepsilon \varepsilon') \neq 0) \\ &= \sigma^2 (X' \Omega^{-1} X)^{-1} X' (\Omega^{-1})' X (X' \Omega^{-1} X)^{-1} \\ &= \sigma^2 (X' \Omega^{-1} X)^{-1} < V(\hat{B}_{MCO}) \end{aligned}$$

La méthode des MCG

- L'estimateur des moindres carrés généralisés est un estimateur efficace (non biaisé et de variance minimale).
- L'inconvénient de cette méthode est qu'il faut forcément faire des hypothèses pour calculer une estimation de la matrice Ω .

La méthode des MCG

- Comme pour le modèle linéaire multiple, une estimation non biaisée de la partie aléatoire est donnée par:

$$\hat{\sigma}^2 = \frac{SCR}{N - (k + 1)}$$

- avec

$$\begin{aligned} SCR &= \hat{\tilde{\varepsilon}}' \hat{\tilde{\varepsilon}} = (\tilde{Y} - \tilde{X}\hat{B})' (\tilde{Y} - \tilde{X}\hat{B}) \\ &= \tilde{Y}' \tilde{Y} - 2\hat{B}' \tilde{X}' \tilde{Y} + \hat{B}' \tilde{X}' \tilde{X} \hat{B} \\ &= \tilde{Y}' \tilde{Y} - 2\hat{B}' \tilde{X}' \tilde{Y} + \hat{B}' \tilde{X}' \tilde{X} (\tilde{X}' \tilde{X})^{-1} \tilde{X}' \tilde{Y} \\ &= \tilde{Y}' \tilde{Y} - 2\hat{B}' \tilde{X}' \tilde{Y} + \hat{B}' \tilde{X}' \tilde{Y} \\ &= \tilde{Y}' \tilde{Y} - \hat{B}' \tilde{X}' \tilde{Y} \\ &= Y' \Omega^{-1/2'} \Omega^{-1/2} Y - \hat{B}' X' \Omega^{-1/2'} \Omega^{-1/2} Y \\ &= Y' \Omega^{-1} Y - \hat{B}' X' \Omega^{-1} Y \end{aligned}$$

Emploi des MCO et correction de la matrice de variances-covariances des estimateurs

- White a montré qu'il n'est pas nécessaire de connaître la valeur de la matrice Ω pour obtenir la valeur estimée de $V(\hat{\beta}_{MCO})$.
- Plus précisément, il montre que la matrice $S_0 = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 x_i x_i'$ est un estimateur convergent de $\Gamma = \frac{1}{n} \sigma^2 X' \Omega X$.
- Par conséquent, un estimateur convergent de la variance de l'estimateur des MCO sous hétéroscédasticité est donné par

$$\hat{V}(\hat{\beta}) = n(X'X)^{-1} S_0 (X'X)^{-1}$$

- Cette matrice est connue sous le nom de la matrice de variance-covariances robuste à l'hétéroscédasticité selon la forme de White.

Tests d'hypothèses d'homoscédasticité

- Le Test de Goldfield et Quandt
 - Ce test n'est applicable que si l'un des régresseurs est la cause de l'hétéroscédasticité. De ce fait on postule l'existence d'une dépendance entre un régresseur quelconque du modèle et la variance des erreurs. Ceci autorise à formuler:

$$\sigma_i^2 = \sigma^2 x_{ki}^2$$

- L'objectif de ce test est de savoir si les termes d'erreurs sont homoscédastiques ou hétéroscédastiques

$$\left\{ \begin{array}{l} H0 : \sigma_i^2 = \sigma^2, \quad \forall i \\ H1 : \sigma_i^2 \neq \sigma^2 \end{array} \right.$$

Le Test de Goldfield et Quandt

- Le test se réalise en trois étapes:
 - 1) On classe les observations de l'échantillon considéré selon l'ordre croissant de la variable x_k
 - 2) On omet de l'échantillon c observations centrales et on divise le reste en deux sous-échantillon de même taille $((N - c)/2)$. Généralement, le nombre de valeurs centrales retiré de l'échantillon est environ égal au quart de l'ensemble des observations
 - 3) On effectue des estimations séparées par les MCO sur les deux sous-échantillons.

Le test de Goldfield et Quandt

- Sous l'hypothèse nulle d'homoscédasticité des erreurs, le rapport des variations résiduelles respectives permet d'établir la statistique suivante:

$$F_c = \frac{SCR_2}{\frac{N-c}{2} - (k+1)} \times \frac{\frac{N-c}{2} - (k+1)}{SCR_1}$$
$$\sim F\left(\frac{N-c}{2} - (k+1); \frac{N-c}{2} - (k+1)\right)$$

- SCR_1 indique la variation résiduelle estimée à partir du premier échantillon et SCR_2 la variation résiduelle estimée du second échantillon. La règle de décision est habituelle:
 - 1) si $F_c < F_{tab}$ alors on accepte H_0 (les erreurs sont homoscédastiques)
 - 2) si $F_c > F_{tab}$ alors on rejete H_0 (les erreurs sont hétéroscédastiques)

Le Test de White

- Ce test est plus général que le test précédent dans le sens où il n'impose aucune forme à priori de l'hétéroscédasticité.
- Comme pour le test précédent, on cherche à savoir si:

$$\left\{ \begin{array}{l} H0 : \sigma_i^2 = \sigma^2, V(\varepsilon_i|X_i) = \sigma^2 \quad \forall i \\ H1 : \sigma_i^2 \neq \sigma^2 \end{array} \right.$$

Test de White

- Ce test se réalise en deux étapes:
 - On estime le modèle par les MCO.
 - On régresse les résidus estimés par les MCO sur l'ensemble des régresseurs, leur carré et leur produit.

$$\varepsilon_i^2 = b_0 + \sum_k b_k x_{ki} + \sum_k \lambda_k x_{ki}^2 + \sum_{k \neq l} \mu_{kl} x_{ki} x_{li} + v_i$$

- La statistique du test de White repose sur le calcul du coefficient de détermination R^2 tiré de la régression ci-dessus. Sous l'hypothèse nulle d'homoscédasticité des erreurs, on montre que:

$$\chi_c^2 = NR^2 \sim \chi_P^2$$

- où P représente le nombre total de variables de la régression effectuée au cours de la deuxième étape, y compris la constante. on rejette l'hypothèse nulle lorsque la valeur critique lue dans la table du Chi-deux est dépassée par celle de nR^2 .

Test de White

- La règle de décision est habituelle:
 - si $\chi_c^2 < \chi_{\frac{\alpha}{2}}^2 < \chi_{1-\frac{\alpha}{2}}^2$ alors on accepte H_0 (les erreurs sont homoscédastiques)
 - si $\chi_c^2 < \chi_{\frac{\alpha}{2}}^2$ ou $\chi_c^2 > \chi_{1-\frac{\alpha}{2}}^2$ alors on rejete H_0 (les erreurs sont hétéroscédastiques)

L'autocorrélation des erreurs

- Une autre violation intervient lorsque les termes d'erreurs de la régression ne sont pas indépendants. On dit qu'il y a autocorrélation des erreurs.
- L'autocorrélation peut survenir dans plusieurs cas de figure.
 - Si l'on travaille avec des données de panel
 - En coupe transversale, si les données sont regroupées (par exemple si elles proviennent d'observations sur des individus regroupés dans un même ménage ou un même village)
- La conséquence de cette autocorrélation est, comme pour l'hétéroscédasticité une perte d'efficacité de l'estimateur des MCO et un mauvais calcul de la matrice de variances-covariances.

Autocorrélation des erreurs

- En coupe transversale, si les données sont groupées, il est possible qu'il existe une corrélation entre deux observations prises au hasard.
- Supposons que le modèle à estimer a la forme suivante:

$$y_{ig} = x'_{ig}\beta + u_{ig}, i = 1, \dots, N_g; g = 1, \dots, G$$

- Il y a donc G groupes regroupant chacun N_g individus. Les groupes peuvent être des zones géographiques et les individus des ménages. Ainsi, pour la suite il est utile d'écrire le modèle à différents niveaux d'agrégation.

$$y_{ig} = x'_{ig}\beta + u_{ig}, i = 1, \dots, N_g; g = 1, \dots, G$$

$$y_g = x'_g\beta + u_g, g = 1, \dots, G$$

$$y = X'\beta + u$$

Autocorrélation des erreurs

- On suppose que les termes d'erreur sont indépendants entre les groupes mais corrélés à l'intérieur des groupes:

$$\text{cov}(u_{ig}, u_{jg'}) = 0$$

sauf si $g=g'$

- On peut montrer dans ce cas que l'estimateur de la variance du coefficient de la variable j obtenue par la méthode des MCO doit être multipliée par un facteur égal à:

$$\tau_j = 1 + \rho_{x_j} \rho_u (\bar{N}_g - 1)$$

- Où \bar{N}_g le nombre moyen d'observation par groupe, ρ_{x_j} est le coefficient de corrélation intragroupe de la variable x_j et ρ_u le coefficient de corrélation du terme d'erreur.

Autocorrélation des erreurs

- Un coefficient de corrélation intragroupe se calcule comme le ratio entre la variance intergroupes de la variable et la somme des variances intra et inter groupes:

$$\rho_{x_j} = \frac{s_{inter}^2}{s_{inter}^2 + s_{intra}^2}$$

- On voit que plus la variance intra groupe est faible (les observations sont de plus en plus similaires) et plus la corrélation intragroupe tend vers l'unité.

Autocorrélation des erreurs

- Ainsi, lorsque la variance intragroupe est faible, le coefficient de corrélation intragroupe est égal à 1, on doit donc multiplier la variance par un coefficient $\tau_j = 1 + \rho_u(\bar{N}_g - 1)$, ce qui signifie que la vraie valeur de la variance doit être multipliée par un facteur positif.
- Ainsi, les écarts-types calculés par les méthodes standards sont sous-estimés. Dans la réalité, les écarts-types sont souvent plus grand et donc la significativité peut être réduite.

Autocorrélation des erreurs

- Comme pour l'hétéroscédasticité, il est possible de calculer la valeur de la matrice de variances-covariances de l'estimateur des MCO lorsque les erreurs sont autocorrélées.

$$V(\hat{\beta}) = (X'X)^{-1} \left(\sum_{g=1}^G x'_g \tilde{u}_g \tilde{u}'_g x_g \right) (X'X)^{-1}$$

- avec $\tilde{u}_g = \frac{G}{G-1} \frac{N-1}{N-k} (y_g - x'_g \hat{\beta})$

L'endogénéité d'une ou plusieurs variables explicatives

- Tous les résultats établis jusqu'à présent l'ont été sous l'hypothèse d'une absence de corrélation entre le terme d'erreur et la liste des variables explicatives: $cov(X_i, \varepsilon_i) = 0$ pour tout i .
- Supposons donc que cette hypothèse ne soit plus vérifiée. Dans ce cas, l'estimateur des MCO est biaisé.
- On dira qu'une, ou plusieurs, variable explicative est endogène. Ceci peut se produire dans trois cas de figure:
 - **Causalité inverse:** X explique y, mais y explique également X
 - **Cause commune:** y n'explique pas X, mais il existe une variable, non incluse dans la liste des variables explicatives du modèle, qui explique à la fois y et X
 - **Erreur de mesure:** une ou plusieurs variables explicatives du modèle sont mesurées avec erreur

Estimateur des variables instrumentales

- Le principe de la méthode des variables instrumentales est d'employer une source exogène de variation pour identifier le modèle.
- Cette source doit avoir deux propriétés essentielles:
 - Elle ne doit pas impacter directement la variable expliquée
 - Elle doit contribuer à expliquer les variations de la variable explicative instrumentée.

Estimateur des variables instrumentales

- Supposons que le modèle estimé s'écrive: $y = \beta X + \varepsilon$. On suppose que $cov(\varepsilon_i, X_i) \neq 0$ de sorte que l'emploi des MCO rende l'estimation biaisée.
- Les propriétés asymptotiques supposent que l'on fasse des hypothèses sur le comportement des variables. Supposons qu'il existe une matrice Z de même dimension que X et telle que:
 - $plim \frac{1}{N} Z' \varepsilon = 0$ implique que la convergence en probabilité de $Z' \varepsilon$ soit nulle. Cette hypothèse est équivalente à $cov(Z_i, \varepsilon_i) = 0$.
 - $plim \frac{1}{N} Z' X = Q_{zx}$ où Q_{zx} est une matrice finie de dimension (k, k) (la corrélation entre l'instrument et la variable instrumentée est non nulle)
 - $plim \frac{1}{N} Z' Z = Q_{zz}$ où Q_{zz} une matrice finie de dimension (k, k) et définie positive (condition nécessaire au calcul de la matrice de variances-covariances de l'estimateur des variables instrumentales.)

Estimateur des variables instrumentales

- La matrice Z est soit plus grande soit de la même taille que la matrice des observations X .
- Cela signifie que si une seule variable explicative est endogène, la matrice Z contient toutes les autres variables explicatives du modèle (qui sont alors leurs propres instruments) et une variable supplémentaire, appelée variable instrumentale.
- On dit que le modèle est juste identifié, parce qu'il comporte autant d'instruments que de variables explicatives. Le modèle sera dit sur-identifié s'il y a plus d'instruments que de variables explicatives endogènes.

Estimateur des doubles moindres carrés

- On peut se servir de l'hypothèse selon laquelle $cov(Z_i, \varepsilon_i) = 0$, ce qui implique que chaque colonne de Z n'est pas corrélée avec le terme d'erreur.
- Il suffit donc d'instrumenter le modèle par la projection de X sur les colonnes de Z :

$$\hat{X} = Z(Z'Z)^{-1}Z'X$$

Ceci correspond à l'estimation de X purgé de l'effet de Z .

L'estimateur des doubles MC

- L'estimateur à variables instrumentales est donc:

$$\hat{\beta}_{IV} = (\hat{X}'X)^{-1} \hat{X}'y = (X'Z(Z'Z)^{-1}Z'X)^{-1} (X'Z(Z'Z)^{-1})Z'y$$

- Cet estimateur est l'estimateur des doubles moindres carrés. Il possède asymptotiquement toutes les propriétés désirables à savoir:
 - Il est convergent
 - Il est efficace dans la classe des estimateurs des doubles MCO

L'estimateur des doubles MC

- En pratique, il suffit de suivre deux étapes pour mesurer l'estimateur de doubles MC:
 - 1/ On régresse la ou les variables endogènes sur les instruments (inclus et exclus).
 - 2/ On remplace ces variables par leur régression issue de la première étape.
- il faudra faire attention au fait que les résidus de l'estimation de la deuxième étape ne peuvent pas être utilisés pour calculer la variance du terme d'erreur. Celle-ci doit être basée sur le calcul des résidus du modèle d'origine:

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (y_i - x_i' \hat{\beta}_{IV})^2$$

Validité des instruments

- Les travaux théoriques sur le sujet ont permis d'établir quelques faits importants:
 - Les problèmes de biais sont accentués lorsque les instruments sont faibles
 - Mieux vaut ne pas avoir trop d'instruments lorsque ceux ci sont trop faibles
- La force de la corrélation entre l'instrument et la variable endogène est donc un aspect important de la question.

Biais de l'estimateur

- Le biais de l'estimateur à distance finie peut s'écrire de la façon suivante:

$$E(\hat{\beta}_{IV} - \beta) = \frac{\text{cov}(\varepsilon, u)}{\sigma_u^2} \frac{1}{F + 1}$$

- Avec $\text{cov}(\varepsilon, u)$ la covariance entre la régression principale et la régression instrumentale et σ_u^2 la variance du terme d'erreur de la régression instrumentale.
- F correspond à la statistique de Fisher de la régression instrumentale, qui mesure la significativité jointe de l'ensemble des instruments.
- Lorsque les instruments sont faibles, la statistique de Fisher l'est également et le biais tend alors vers le biais de l'estimateur de MCO.

Test de la force des instruments

- Dans un premier temps, il est important de montrer que l'instrument est assez fort (c'est à dire qu'il est suffisamment corrélé).
- La statistique de Fisher peut être utilisée pour identifier les bons instruments.

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

- k est le nombre de variables de la regression instrumentale.
- La pratique courante est de considérer que l'instrumentation est suffisamment forte lorsque la statistique de Fisher calculée pour tester la significativité jointe de l'ensemble des instruments prend au moins la valeur de 10.

Test de la validité des instruments

- Le principe du test est le suivant: si les instruments sont valides, alors ils sont non corrélés avec le terme d'erreur de l'équation instrumentée.
- Le test de Sargan repose sur ce principe: il s'agit d'examiner si la corrélation entre le ou les instruments en surnombre et le terme d'erreur de l'équation est non nulle. Le test repose alors sur la statistique suivante:

$$S = n \frac{\varepsilon' P \varepsilon}{\varepsilon' \varepsilon}$$

ε est le vecteur des résidus de l'équation instrumentée. P est la matrice égale à $Z(Z'Z)^{-1}Z'$. Ainsi le produit $P\hat{\varepsilon}$ est la projection du résidu sur le sous-espace vectoriel engendré par les Z .

Test de la validité des instruments

- Cette projection est nulle si le résidu est orthogonal à Z .
- On s'attend donc à ce que S soit nul si le résidu de l'équation instrumentée est effectivement orthogonal aux instruments.
- Malheureusement le test ne peut être utilisé que lorsqu'il n'existe plus d'instruments que de variables endogènes.

Vérifier le caractère endogène d'une variable

- On peut tester si une ou plusieurs variables explicatives sont endogènes en utilisant la procédure proposée par Hausman.
- Le test repose sur la différence $\beta - \beta_{IV}$. Sous l'hypothèse nulle cette différence a une limite en probabilité nulle, alors que sous l'hypothèse alternative, cette limite est non nulle. La statistique proposée est alors:

$$W = (\hat{\beta} - \hat{\beta}_{IV})' [V_1 - V_0]^{-1} (\hat{\beta} - \hat{\beta}_{IV})$$

- Avec V_1 la matrice de variances-covariances asymptotique de l'estimateur de doubles MC et V_0 la matrice $\hat{\sigma}^2 (X'X)^{-1}$. Ce test est connu sous le nom de Durbin-Wu-Hausman.
- Sous H_0 , toutes les variables explicatives sont exogènes. La statistique W est distribuée selon une loi du Chi-deux à k degrés de liberté (nombre d'éléments de β).