

Econométrie
Support de Cours
Licence 3 Economie-Gestion
Année 2018-2019

Catherine Laffineur*

21 novembre 2018

*GREDEG, CNRS UMR 7321 - Université Nice Sophia-Antipolis, e-mail : catherine.laffineur@unice.fr
Document à destination des étudiants de L3 ECO-GESTION. Interdiction de diffusion.

Table des matières

1	Introduction	4
1.1	Présentation	4
1.2	L'économétrie en pratique	7
1.2.1	Les différents types de données	7
1.2.2	Types de variables et échelle d'analyse	8
1.3	La méthode économétrique : contexte général	9
2	Le modèle linéaire simple	9
2.1	Les hypothèses du modèle linéaire simple	11
2.2	Les différentes méthodes d'estimation	12
2.3	L'estimateur des MCO et ses propriétés	14
2.3.1	Les estimateurs des coefficients de régression (a et b)	14
2.3.2	L'espérance des estimateurs	17
2.3.3	Variance des estimateurs	18
2.3.4	L'estimateur du coefficient σ^2	20
2.3.5	Résumé	21
2.4	Inférence statistique	22
2.4.1	Coefficient de détermination et puissance du modèle	22
2.4.2	Intervalle de confiance	23
2.4.3	Test sur la valeur d'un coefficient	26
2.4.4	Test de significativité d'un paramètre	26
2.4.5	Significativité globale d'un modèle et test de Fisher	27
2.5	La prévision	28
2.6	Interprétation des coefficients	30
3	Le modèle linéaire multiple	32
3.1	Le modèle et ses hypothèses	33
3.1.1	Ecriture du modèle	33
3.1.2	Les hypothèses	34
3.2	Les estimateurs et leurs propriétés	35
3.2.1	Estimation des coefficients de la régression et propriétés des estimateurs	35
3.3	Interprétation de l'estimateur des MCO dans le cas de la régression multiple	36
3.3.1	Interprétation avec variables d'interaction	38
3.4	Biais et efficacité des estimateurs MCO	42
3.4.1	Estimation du coefficient de la partie aléatoire	43
3.4.2	Analyse de la variance	43
3.4.3	Le coefficient de détermination	44
3.5	Les tests statistiques	45

3.5.1	Les tests individuels	45
3.5.2	Généralisation	46
3.6	La prévision	46
4	Violations des hypothèses de Gauss-Markov : conséquences et remèdes	47
4.1	La multicolinéarité	47
4.2	L'hétéroscédasticité	48
4.2.1	Détection visuelle de l'hétéroscédasticité	49
4.2.2	La méthode des MCG	49
4.2.3	Estimation du coefficient de la partie aléatoire	51
4.2.4	Emploi des MCO et correction de la matrice de variances-covariances des estimateurs	52
4.2.5	Tests d'hypothèses d'homoscédasticité	52
4.3	L'autocorrélation des erreurs	54
4.3.1	Le cas particulier des données groupées	55
4.4	L'endogénéité d'une ou de plusieurs explicatives	56
4.4.1	La méthode des variables instrumentales	56
5	Introduction à l'économétrie des séries temporelles	62
5.1	Les Processus Aléatoires Stationnaires et les Processus ARMA	62
5.1.1	Séries Temporelles	62
5.1.2	Indices descriptifs d'une série temporelle	62
5.1.3	Méthode économétrique et gestion de la tendance et des saisonalités	63
5.1.4	Les principaux modèles stationnaires	67
5.2	Les tests de stationarité ou Unit root Tests	69
6	Introduction à l'économétrie des données de panel	73
7	Introduction à l'économétrie des variables qualitatives	73
8	Annexe-Rappels	73

1 Introduction

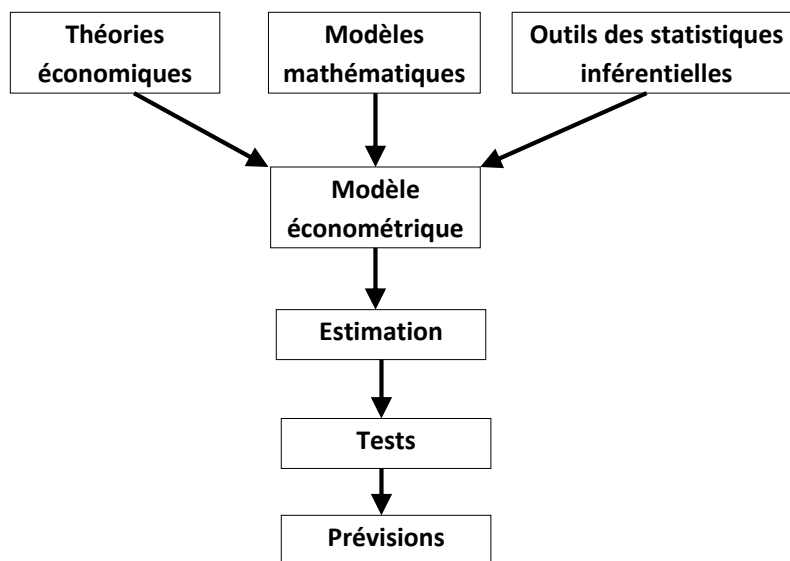
1.1 Présentation

L'économie se divise habituellement en deux branches :

- L'analyse théorique
- L'analyse empirique

L'analyse théorique vise à décrire le comportement des agents économiques et les mécanismes à l'oeuvre dans la formation des quantités agrégées. Les théories économiques sont donc des énoncés logiques, qui reposent sur des hypothèses plus ou moins réalistes. Dans un second temps ces théories économiques doivent être confrontées aux faits observés : c'est là que l'économétrie intervient.

Pour schématiser, l'analyse économique se fait de la manière suivante :



En d'autres termes, la théorie économique nous informe sur le meilleur modèle économétrique à utiliser, puis il s'agit de choisir les meilleures méthodes d'estimation pour obtenir un estimateur sans biais et efficace, c'est à dire un estimateur qui se rapproche de sa valeur réelle et avec une grande précision (variance minimale). Ensuite, il faut tester la conformité des résultats obtenus.

Histoire de l'économétrie. C'est en 1930 que Frisch et Fisher auraient baptisé la discipline économétrie, qui relève à la fois de l'économie et de la statistique. Les méthodes statistiques qu'elle emploie sont elles très anciennes et remontent au 18e siècle. Nombre d'historiens considèrent que l'avènement de l'économétrie date de 1944. Les économistes commencent à utiliser des méthodes probabilistes de la statistique inférentielle pour spécifier les relations entre les lois économiques et les données observées.

Fonctions de l'économétrie. L'économétrie peut se définir comme l'application des méthodes de statistiques inférentielles à l'étude des phénomènes économiques et sociaux. L'objectif premier de l'économétrie est de tester les résultats des modèles théoriques en économie. Un modèle théorique en économie est une représentation simplifiée du monde réel. L'économétrie permet donc de mesurer les liens de corrélation entre les variables et de mesurer leur ampleur. Ainsi il existe deux objectifs premier à l'économétrie :

- Evaluer les paramètres en jeu
- Tester la validité des théories économiques

a- Evaluer les paramètres en jeu :

1 : les déterminants de la croissance.

Supposons une fonction de production Cobb-Douglas $Y = AK^\alpha L^\beta$

L'économétrie permet de mesurer les paramètres α et β qui correspondent à l'élasticité de production par rapport aux facteurs de production. Autrement dit, ces variables mesurent la sensibilité de la production aux facteurs de production. Ainsi, ces estimations sont extrêmement importantes pour orienter les politiques économiques vers un soutien plus important du capital ou du travail (selon l'évolution des valeurs de α et β).

Exemple 2 : L'effet d'une baisse de la TVA

Supposons que le gouvernement envisage de baisser le taux de TVA sur les produits consommés afin de permettre une augmentation du revenu disponible. Pour connaître précisément l'effet de l'augmentation du revenu disponible sur la consommation il faut repartir de la théorie économique :

$$Y_d = \alpha C + \beta S$$

Pour connaître l'effet d'une baisse des prix de consommation sur Y_d il faut passer par l'économétrie pour mesurer α .

b- Tester les théories économiques :

Exemple 1 : L'intéressement permet-il d'améliorer la productivité des entreprises ?

La pratique de l'intéressement consiste à distribuer aux salariés des primes liées aux résultats de l'entreprise. En France, cette pratique est encouragée par une exonération de cotisations sociales des primes correspondantes, tant qu'elles ne dépassent pas 20% de la masse salariale. Compte tenu du manque à gagner que représentent ces exonérations

pour la collectivité, il importe de vérifier que l'intéressement a bien des retombées positives sur les performances des entreprises françaises en matière de productivité. L'intéressement devrait théoriquement augmenter l'effort individuel et la productivité. Cependant, puisque l'intéressement est une prime collective partagée entre les travailleurs, il peut exister des comportements de passager clandestins. Pour juger de l'opportunité d'un encouragement public à l'intéressement il a fallu procéder à une étude économétrique pour évaluer ses effets sur la productivité. Les résultats montrent que les entreprises qui pratiquent un intéressement ont, ceteris paribus, des gains de productivité supérieurs à celles qui ne la pratiquent pas.

Exemple 2 : Quelle est l'influence du coût salarial sur la demande de travail ?

Les économistes s'accordent en général pour considérer que le coût salarial joue négativement sur la demande de travail. Plus le travail coûte cher moins les entreprises embauchent. Pourtant les modèles macroéconomiques ont longtemps considéré des modèles où le coût du travail n'influence pas la demande de travail du fait d'une substitution capital/travail faible, combiné à une faible mobilité du travail.

L'économétrie permet de mesurer la complémentarité ou substitution du capital travail. Au niveau macroéconomique on observe sur le long terme une augmentation de la part du capital dans le PIB, il existerait donc une substituabilité. Par ailleurs, l'approche économétrique permettrait de mesurer précisément la sensibilité de la demande de travail au coût salarial et ainsi de mesurer l'élasticité emploi-salaire.

Exemple 3 : La consommation de soin est-elle influencée par des phénomènes de demande induite ?

Depuis plus de 40 ans les dépenses de santé augmentent très rapidement dans l'ensemble des pays développés. De nombreux facteurs contribuent à cette croissance continue : la progression du revenu disponible de ménages principalement, mais aussi de l'amélioration de la couverture par l'assurance maladie, le progrès technique de la médecine et le vieillissement de la population. Un autre facteur peut être évoqué : l'influence des médecins sur la demande de soins, appelée demande induite. En effet, l'ignorance du patient confère au médecin une certaine latitude pour manipuler à la hausse les quantités de soins consommées. Ce comportement a de fortes chances de se manifester dans un contexte comme celui de la France où le revenu du médecin dépend du nombre d'actes pratiqués et où les dépenses sont bien couvertes par la sécurité sociale. La demande induite est négativement influencée par la crainte de voir son patient changer de médecin, le point du bien être du patient dans la fonction d'utilité du médecin et du poids de son revenu dans sa fonction de bien être. Seule une estimation économétrique peut répondre à cette question. Lorsque les médecins subissent un rationnement sur leur nombre de consultations à cause d'une concurrence accrue, ils compensent le manque à gagner par une augmentation du volume de soin fournis au cours de chaque rencontre.

Ce ne sont que des exemples, l'économétrie permet de faire bien d'autres choses encore : prévision de cours boursiers, mesure de paramètre inobservables (par le statisticien), évaluation de politique publique.

Dans ce cours nous nous limiterons à des questions relativement simples qui reviennent à examiner si une variable influence ou non une autre variable.

1.2 L'économétrie en pratique

Tout l'enjeu de l'économétrie est de trouver des données qui permettent de tester ces différents modèles théoriques. Ces données peuvent être plus ou moins difficiles à trouver, il est parfois nécessaire de lancer des enquêtes, parfois nécessaire de passer par des machines virtuelles en lien avec un serveur de données (INSEE, Banque mondiale). Souvent les données utilisées ne sont pas exhaustives, mais elles représentent un échantillon d'une population. Un échantillon est dit représentatif lorsqu'il respecte la composition de la population en question sur différents critères (genre, lieu d'habitation, niveau d'éducation etc). Le fait d'avoir un échantillon représentatif permet de généraliser les conclusions sur l'ensemble de la population. Par ailleurs, nous verrons que les estimations sont affectées par une marge d'imprécision qui se réduit quand la taille de l'échantillon augmente.

1.2.1 Les différents types de données

Les données sont présentées sous différentes formes. Classiquement, on distingue trois catégories de données :

1) Données transversales ou en coupe instantanée

Dans ce cas, la base de données va contenir des informations sur N agents (ménage, pays, entreprise,...) à une date donnée. Dans le cas de l'exemple 1, cela correspondra à une base de données du type :

$$\begin{bmatrix} Y_1 & K_1 & L_1 \\ Y_2 & K_2 & L_2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ Y_N & K_N & L_N \end{bmatrix}$$

où Y_i représente le niveau de production de la firme industrielle i , K_i représente le nombre d'unités de capital utilisé par la firme industrielle i et L_i le nombre d'unités de travail utilisé par la firme i , $i = 1, \dots, N$.

2) Données temporelles ou en séries chronologiques

Ici, la base de données va contenir des informations concernant un agent sur plusieurs périodes. Dans le cas de l'exemple 1, cela correspondra à une base de données du type :

$$\begin{bmatrix} Y_1 & K_1 & L_1 \\ Y_2 & K_2 & L_2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ Y_T & K_T & L_T \end{bmatrix}$$

où Y_i représente le niveau de production de la firme industrielle à la date i , K_i représente le nombre d'unités de capital utilisé par la firme industrielle à la date i et L_i le nombre d'unités de travail utilisé par la firme à la date i , $i = 1, \dots, T$.

3) Données de panel (individuelles et temporelles)

Ici, la base de données va contenir des informations concernant N agents sur plusieurs périodes (c'est la combinaison des deux précédentes). Dans le case de l'exemple 1, cela correspondra à une base de données du type :

$$\begin{bmatrix} Y_{1,1} & K_{1,1} & L_{1,1} \\ Y_{1,2} & K_{1,2} & L_{1,2} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ Y_{1,T} & K_{1,T} & L_{1,T} \\ Y_{2,1} & K_{2,1} & L_{2,1} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ Y_{N,T} & K_{N,T} & L_{N,T} \end{bmatrix}$$

où $Y_{i,t}$ représente le niveau de production de la firme i à la date t , $K_{i,t}$ représente le nombre d'unités de capital utilisé par la firme i à la date t et $L_{i,t}$ le nombre d'unités de travail utilisé par la firme i à la date t , $i = 1, \dots, N$ et $t = 1, \dots, T$.

1.2.2 Types de variables et échelle d'analyse

Les variables peuvent être quantitatives, mais aussi qualitative (par exemple être un homme ou une femme, être ou non diplômé). Chaque modalité (homme ou femme par exemple) reçoit un code qui permet de déterminer si l'appartenance à ce code (par rapport à l'autre) permet d'obtenir un effet différencié sur la variable d'intérêt. Par exemple, être

une femme permet-il d'obtenir un meilleur salaire, ou une probabilité plus élevée d'obtenir un poste.

Par ailleurs, il existe différentes échelles d'analyse. L'analyse peut se faire au niveau microéconomique (individu) ou macroéconomique (pays, région). Selon l'échelle d'analyse et la nature des variables (quantitative ou qualitative) les techniques utilisées peuvent être différentes. Dans ce cours nous nous limiterons à des cas d'analyse microéconomique avec variable dépendante quantitative.

1.3 La méthode économétrique : contexte général

Dans un premier temps, on va utiliser la théorie économique, la modélisation mathématique et l'outil statistique afin de spécifier un modèle économétrique adéquat. Après cette étape, on va chercher à estimer les paramètres du modèle et, pour ce faire, choisir les méthodes de calculs les plus appropriées afin d'obtenir des estimateurs statistiquement efficaces. **Un estimateur de b noté \hat{b} est dit efficace si $E(\hat{b}) = b$ et si $V(\hat{b})$ est la variance minimale parmi tous les estimateurs sans biais.** Après avoir estimé le modèle, on passe aux tests du modèle en utilisant plusieurs outils statistiques afin de vérifier la conformité des résultats empiriques obtenus avec le sous-bassement théorique envisagé. Finalement la dernière étape (qui n'est pas toujours pertinente et/ou justifiée) consiste à exploiter le modèle à des fins de prévision et à élaborer un schéma de politique économique dans le domaine étudié.

Pour résumer la méthodologie économétrique que l'on va adopter est la suivante :

- 1- Définir le modèle économique théorique que l'on cherche à tester
- 2- Choisir le meilleur modèle économétrique
- 3- Choisir la meilleure méthode d'estimation
- 4- Tester la validité du modèle économique/économétrique

Dans ce cours, nous nous concentrerons principalement sur un seul modèle : le modèle linéaire. Nous allons principalement apprendre à choisir la meilleure méthode d'estimation et à tester la validité du modèle.

2 Le modèle linéaire simple

Dans ce cours, nous nous concentrerons principalement sur le modèle linéaire. Il s'agit d'une estimation d'une fonction linéaire de la forme $y = ax + b$ où y apparaît comme une fonction linéaire des coefficients a et b .

Une théorie économique simple comme la théorie de Fisher stipule une relation linéaire entre le taux d'intérêt et l'inflation de la forme $r = i - \pi$. r , le taux d'intérêt réel dépend

de manière linéaire du taux d'intérêt nominal i et de l'inflation π . Cette théorie peut ainsi facilement s'estimer par un modèle linéaire simple.

A l'inverse, une théorie de la croissance qui dépend d'une fonction de production Cobb-Douglas, n'est pas une fonction linéaire mais multiplicative de K et L dans laquelle $Y = K^\alpha L^\beta$

Les fonctions multiplicatives peuvent également être estimées par un modèle linéaire. Il suffit d'adopter une transformation logarithmique à la fonction pour obtenir une fonction linéaire :

$$Y = K^\alpha L^\beta$$

$$\ln(Y) = \ln(K^\alpha L^\beta)$$

$$\ln(Y) = \alpha \ln(K) + \beta \ln(L)$$

Ainsi, $\ln(Y)$ est une fonction linéaire de $\ln(K)$ et $\ln(L)$.

Par ailleurs, il est également possible d'intégrer des effets multiplicatifs dans un modèle linéaire. Ces effets sont appelés effets quadratiques ou effets non linéaires.

Supposons que l'on cherche à étudier le salaire : $W_i = a + c \exp_i$

Ce n'est pas très réaliste de penser qu'une année d'expérience supplémentaire a un effet linéaire sur le salaire. L'effet de l'expérience sur le salaire n'est pas le même selon que le salarié entre sur le marché du travail ou qu'il soit proche de la retraite.

La théorie économique suppose que le rendement marginal de l'expérience est décroissant. Autrement dit, les salaires augmentent rapidement avec l'expérience en début de carrière puis plus lentement par la suite.

Ainsi, lorsque l'on cherche un effet non linéaire on estime la fonction suivante :

$$W_i = a + c \exp_i + b \exp_i^2$$

Lorsque l'on cherche l'effet d'une année supplémentaire d'expérience sur le salaire on estime l'équation suivante :

$$\frac{\delta W_i}{\delta \exp} = c + 2b \exp$$

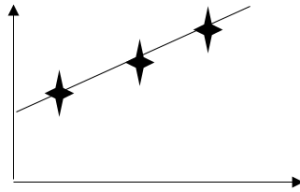
L'interprétation de cette équation est la suivante : lorsque l'on passe de zéro à une année d'expérience, l'effet sur le salaire est égale à c . Lorsqu'on passe de 10 à 11 années d'expérience l'effet sur le salaire est $c + 2b10$.

C'est donc b qui détermine l'effet marginal. Si b est positif l'effet augmente au fur et à mesure que l'expérience augmente, si b est négatif le rendement marginal de l'expérience est décroissant.

Il est également possible de déterminer le maximum c'est à dire le niveau d'expérience qui permet d'atteindre une rentabilité maximale sur le salaire. Ce point s'obtient lorsque la dérivée est nulle :

$$c + 2bexp = 0 \Rightarrow exp^* = \frac{c}{-2b}$$

Ainsi, un modèle linéaire estime une droite de régression linéaire de la forme $y = a + bx$. Si il existait une relation CERTAINE entre x et y , alors tous les points seraient sur la même droite linéaire.



Dans la réalité, d'autres facteurs peuvent fortement altérer y (dans le cas d'un modèle microéconomique cela peut être un déménagement, une naissance, un voyage prévu.)

En fait, il existe souvent pour n'importe quelle variable y une infinité de facteurs explicatifs qu'il est impossible d'intégrer dans le modèle. Ainsi, puisque nous sommes incapables de prendre en compte tous les facteurs explicatifs, la fonction y est affectée d'une incertitude. Chaque point de l'échantillon est un cas particulier dont le comportement s'écarte du modèle théorique d'un terme d'erreur noté u_i :

$$y_i = a + bx_i + u_i$$

Cette équation représente l'équation de base que l'on cherche à estimer dans un modèle linéaire. La variable a est la constante, la variable b est la pente, ou l'effet marginal (c'est-à-dire l'effet d'une variation de x sur y , ceteris paribus). La variable u_i est le terme d'erreur. Les variable y et x sont traitées de manière indépendante. La variable y est appelée variable dépendante et est supposée aléatoire ou stochastique. La variable x est la variable indépendante supposée avoir des valeurs fixes et certaines.

Lorsqu'il y a une seule variable indépendante x on parlera de modèle linéaire simple. Lorsqu'il y a plusieurs variables indépendantes x on parlera de modèle de régression multiple.

Dans ce chapitre, nous allons présenter la méthode de base de l'analyse économétrique. On étudie ici le cadre le plus simple où le modèle ne comporte qu'une seule variable explicative. Ce modèle constitue un point de départ nécessaire pour la compréhension de la procédure économétrique en matière d'estimation, de test et de prévision.

2.1 Les hypothèses du modèle linéaire simple

Le modèle linéaire simple est donné par :

$$y_i = a + bx_i + u_i \quad i = 1, 2, \dots, N \quad (1)$$

y_i est la variable que l'on cherche à expliquer, x_i est la variable explicative du modèle, u_i représente les résidus ou le terme d'erreur. a et b sont les paramètres à estimer. On appelle $a + bx_i$ la partie systématique du modèle et u_i est la partie aléatoire du modèle.

- Les hypothèses du modèle

H1) $E(u_i) = 0, \forall i \Rightarrow$ en moyenne le terme d'erreur est nul. Autrement dit, les variables spécifiées dans le modèle capturent bien y_i .

H2) x_i est une variable certaine (non stochastique)

H3) $V(x_i) = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \neq 0 \Rightarrow$ la variance de x est non nulle, i.e., les observations x_i ne prennent pas toutes la même valeur.

H4) $V(u_i) = E(u_i^2) = \sigma^2, \forall i \neq s \Rightarrow$ la variance est la même pour tout les u_i , on dit que les perturbations sont homoscédastiques.

$Cov(u_i, u_s) = E(u_i u_s) = 0 \Rightarrow$ la perturbation u_i n'est pas influencée par la perturbation u_s . Dans le cas d'une série temporelle, cela signifie que la perturbation à une période n'est pas influencée par la perturbation à une autre période, i.e., un choc qui s'est produit à une période n'a pas d'influence sur ce qui se passe dans les périodes suivantes.

H5) $u_i \sim N(0, \sigma^2) \Rightarrow$ les erreurs sont indépendantes et identiquement distribuées selon la loi normale. Cette hypothèse de normalité est nécessaire pour réaliser des tests statistiques sur la base des distributions normale, de Student et de Fisher.

H6) $Cov(x_i, u_i) = 0 \Rightarrow$ Cette hypothèse rend compte de l'indépendance entre la partie systématique et la partie aléatoire du modèle

2.2 Les différentes méthodes d'estimation

Estimer un paramètre θ consiste à donner une valeur approchée à ce paramètre à partir d'un sondage de la population. Nous allons analyser différents types d'estimation, qui serviront de rappel de notions de statistiques simples.

Supposons que l'on cherche à estimer θ , représentant l'espérance de la population X : $\theta = E(X)$. Dans ce cas, nous pouvons utiliser la moyenne empirique comme estimateur de l'espérance : $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$. Pour rappel, l'espérance d'une constante est égale à la constante de telle sorte que l'on peut écrire $E(a) = a$ et $E(aX) = aE(X)$. Par ailleurs lorsque X et Y sont indépendantes on peut écrire $E(XY) = E(X)E(Y)$

De manière symétrique, si l'on cherche à estimer θ comme étant la variance d'une population X : $\sigma^2 = V(X) = E((X - E(X))^2)$. Dans ce cas, nous pouvons utiliser

l'estimateur de la variance qui se définit par : $V(x) = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$. Pour rappel, $V(x) = E(x^2) - E(x)^2$.

Preuve :

Si $V(x) = E(x - E(x))^2$ avec $E(x) = \text{constante}$ (correspond à la moyenne), nous pouvons noter $E(x) = \mu$

$$\begin{aligned} V(x) &= E(x - E(x))^2 \\ V(x) &= E(x^2 - 2xE(x) + E(x)^2) \\ V(x) &= E(x^2 - 2x\mu + \mu^2) \\ V(x) &= E(x^2) - 2\mu^2 + \mu^2 \\ V(x) &= E(x^2) - E(x)^2 \end{aligned}$$

Le but d'un estimateur est d'être le plus précis possible pour avoir une erreur d'estimation la plus petite possible.

Une mesure de précision est l'erreur quadratique moyenne (MSE) :

$$MSE = E(\hat{b} - b)^2 = 0 \iff MSE = \underbrace{E(\hat{b} - E(\hat{b}))^2}_{\text{Variance de l'estimateur}} + \underbrace{E(E(\hat{b}) - b)^2}_{\text{Biais de l'estimateur}}$$

Cette erreur quadratique moyenne prend en compte le biais d'estimation et la précision de l'estimation. Un estimateur sans biais signifie que en moyenne la vraie valeur du paramètre s'établit autour du paramètre estimé. Un estimateur précis signifie que la variance du paramètre estimé est minimale. Un estimateur est d'autant meilleur que son biais est faible et sa précision est forte.

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.

D'après le théorème de Gauss-Markov (que nous prouverons dans les chapitres suivants), l'estimateur des MCO est le meilleur estimateur linéaire sans biais et à variance minimale. On dira que l'estimateur est **BLUE**, c'est-à-dire Best Linear Unbiased Estimator

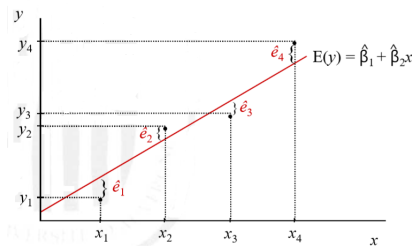
2.3 L'estimateur des MCO et ses propriétés

2.3.1 Les estimateurs des coefficients de régression (a et b)

On considère le modèle donné par l'équation (1). Pour estimer les paramètres a et b du modèle, on dispose d'un échantillon de données de la forme :

$$\begin{bmatrix} y_1 & x_1 \\ y_2 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ y_N & x_N \end{bmatrix}$$

Notons que chaque observation d'un individu peut-être représentée par un couple (y_i, x_i) . Représenté dans un plan (x, y) , on obtient un nuage de points.



Le problème est le suivant : comment choisir les valeurs de a et b tel qu'une droite de la forme $y = a + bx_i$ 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.

Les valeurs estimées de a et b notées \hat{a} et \hat{b} seront donc les valeurs qui minimisent la somme des carrés des résidus ($\sum u_i^2$). Pour déterminer \hat{a} et \hat{b} , nous devons chercher les solutions de la fonction objectif suivante :

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

Les solutions de ce programme sont :

$$\hat{a} = \bar{y} - \hat{b}\bar{x} \quad (3)$$

$$\hat{b} = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \quad (4)$$

Preuve résultats (3) et (4) :

En utilisant les conditions du premier ordre, on obtient :

$$\frac{\partial S}{\partial a} = -2 \sum (y_i - \hat{a} - \hat{b}x_i) = 0$$

$$\Leftrightarrow \sum y_i = N\hat{a} + \hat{b}N\bar{x}$$

$$\Leftrightarrow \hat{a} = \bar{y} - \hat{b}\bar{x} \quad (3)$$

$$\frac{\partial S}{\partial b} = -2 \sum x_i(y_i - \hat{a} - \hat{b}x_i) = 0$$

$$\frac{\partial S}{\partial b} = \sum x_i(y_i - \hat{a} - \hat{b}x_i) = 0$$

En insérant \hat{a} dans l'expression $\frac{\partial S}{\partial b}$, on peut récrire cette équation de la façon suivante :

$$\frac{\partial S}{\partial b} = \sum x_i(y_i - \bar{y} + \hat{b}\bar{x} - \hat{b}x_i) = 0$$

$$\sum x_i(y_i - \bar{y}) + \sum x_i(\hat{b}\bar{x} - \hat{b}x_i) = 0$$

$$\sum x_i(y_i - \bar{y}) = \sum x_i\hat{b}(x_i - \bar{x}) = 0$$

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

On peut récrire \hat{b} de la façon suivante :

Notons les deux identités remarquables suivantes : $\sum_{i=1}^N \bar{x}(y_i - \bar{y}) = 0$ et $\sum_{i=1}^N \bar{x}(x_i - \bar{x}) = 0$

Preuve :

En développant on obtient : $\sum \bar{x}y_i - N\bar{x}\bar{y}$.

Puisque \bar{x} et \bar{y} sont des constantes on peut écrire :

$$\bar{x} \sum y_i - N\bar{x}\bar{y}$$

En divisant par N on obtient :

$$\bar{x}\bar{y} - \bar{x}\bar{y} = 0$$

A partir de ces deux identités remarquables, on peut récrire \hat{b} de la façon suivante :

$$\hat{b} = \frac{\sum x_i(y_i - \bar{y})}{\sum x_i(x_i - \bar{x})} = \frac{\sum x_i(y_i - \bar{y}) - \underbrace{\sum \bar{x}(y_i - \bar{y})}_{=0}}{\sum x_i(x_i - \bar{x}) - \underbrace{\sum \bar{x}(x_i - \bar{x})}_{=0}}$$

En travaillant le numérateur on obtient :

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

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

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

En travaillant le dénominateur on obtient :

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

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

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

Il est possible d'obtenir une troisième écriture de \hat{b} de la façon suivante :

Puisque $\sum \bar{y}(x_i - \bar{x}) = 0$ on peut réécrire :

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

En factorisant le numérateur on obtient :

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

Ainsi, il existe 3 différentes définitions du paramètre \hat{b} , défini de la façon suivante :

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

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

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

Le paramètre \hat{a} se définit de la manière suivante :

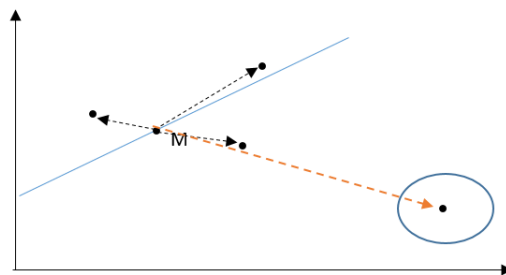
$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

Remarque : Cette dernière équation permet de voir l'importance des valeurs éloignées du point moyen dans l'estimation de \hat{b} .

En effet, nous pouvons écrire $\hat{b} = \frac{\sum(x_i - \bar{x})^2 \left(\frac{y_i - \bar{y}}{x_i - \bar{x}} \right)}{\sum(x_i - \bar{x})^2}$

D'où $\hat{b} = \sum \frac{y_i - \bar{y}}{x_i - \bar{x}}$

Cette équation représente la pente du point (x_i, y_i) rejoignant le point moyen de l'échantillon M



Ainsi, une observation x_i influence d'autant plus la pente \hat{b} qu'elle est éloignée du point moyen. Si ce point est une valeur aberrante, elle peut fortement biaiser les estimations. La procédure dans ce cas consiste à éliminer les outliers.

2.3.2 L'espérance des estimateurs

Pour rappel, un bon estimateur est un estimateur sans biais et à variance minimale. Vérifions d'abord que l'estimateur des MCO est sans biais, c'est à dire qu'en moyenne le paramètre estimé s'établit autour de sa vraie valeur.

Pour cela, calculons l'espérance des estimateurs.

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

$$\hat{b} = \frac{\sum(x_i - \bar{x})[b(x_i - \bar{x}) + (u_i - \bar{u})]}{\sum(x_i - \bar{x})^2}$$

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

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

Puisque $E(u_i) = 0$ et que $E(b) = b$ car b est une variable certaine on peut écrire :

$$\boxed{E(\hat{b}) = b}$$

De la même manière :

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

$$\hat{a} = a + b\bar{x} + \bar{u} - \hat{b}\bar{x}$$

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

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

Comme $E(\hat{b}) = b$ et $E(b) = b$, alors $E(\bar{x}) [E(b) - E(\hat{b})] = 0$. Par ailleurs, d'après H1, $E(u_i) = 0$. Ainsi :

$$E(\hat{a}) = E(a)$$

Puisque a est une variable certaine, on peut écrire :

$$\boxed{E(\hat{a}) = a}$$

En conclusion, l'estimateur des MCO est sans biais si les hypothèses du modèle linéaire sont vérifiées, en particulier si H1 est vérifiée.

2.3.3 Variance des estimateurs

- Variance de \hat{b}

$$V(\hat{b}) = E(\hat{b} - E(\hat{b}))^2$$

$$V(\hat{b}) = E(\hat{b} - b)^2$$

D'après l'équation de l'espérance de \hat{b} on sait que :

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

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

$$\text{Ainsi, } V(\hat{b}) = E \left(\frac{\sum(x_i - \bar{x})(u_i - \bar{u})}{\sum(x_i - \bar{x})^2} \right)^2$$

$$V(\hat{b}) = \frac{E(\sum(x_i - \bar{x})^2(u_i - \bar{u})^2)}{E(\sum(x_i - \bar{x})^4)}$$

Puisque x est une variable certaine $E(x_i - \bar{x}) = (x_i - \bar{x})$ et d'après H1 $E(u_i) = 0$. Par ailleurs, d'après H4 : $E(u_i^2) = \sigma^2$.

Ainsi, $V(\hat{b}) = \frac{\sum(x_i - \bar{x})^2 E(u_i^2)}{E(\sum(x_i - \bar{x})^4)}$ car $E(u_i - E(u_i))^2 = E(u_i^2) - E(u_i)^2 = E(u_i^2)$, puisque $E(u_i) = 0$. D'où :

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

• **Variance de \hat{a}**

$$V(\hat{a}) = E(\hat{a} - E(\hat{a}))^2$$

$$V(\hat{a}) = E(\hat{a} - a)^2$$

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

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

$$V(\hat{a}) = E(\bar{x}(b - \hat{b}))^2 + E(\bar{u})^2$$

$$V(\hat{a}) = \bar{x}^2 E((b - \hat{b}))^2 + E(\bar{u})^2$$

$$V(\hat{a}) = \bar{x}^2 V(\hat{b}) + E(\bar{u})^2$$

Avec $E(\bar{u})^2 = \left(\frac{1}{N} \sum u_i\right)^2 = \left(\frac{1}{N} \frac{1}{N} \sum u_i^2\right) = \frac{1}{N} E(u_i^2)$. Ainsi, $E(\bar{u})^2 = \frac{\sigma^2}{N}$

Ainsi :

$$\boxed{V(\hat{a}) = \bar{x}^2 V(\hat{b}) + \frac{\sigma^2}{N}}$$

Notons qu'en utilisant les nouvelles expressions de \hat{a} et \hat{b} , il vient de manière évidente que ces estimateurs sont sans biais, i.e, $E(\hat{a}) = a$ et $E(\hat{b}) = b$. On pourrait également montrer que ce sont les estimateurs linéaires de variance minimale. On dit alors que les estimateurs des MCO sous les hypothèses H1-H6 sont BLUE.

2.3.4 L'estimateur du coefficient σ^2

Nous avons estimé deux paramètres du modèle linéaire simple. Néanmoins, pour pouvoir calculer la variance des estimateurs, nous devons connaître σ^2 , i.e., la variance des résidus. Nous ne connaissons pas cette valeur, il nous faut donc l'estimer. La variance des résidus estimés est donnée par :

$$V(\sum \hat{u}_i) = E(\sum \hat{u}^2) - E(\sum \hat{u})^2 = \sum E(\hat{u}^2)$$

On sait que

$$\begin{aligned} \hat{u}_i &= y_i - \hat{y}_i = a + bx_i + u_i - \hat{a} - \hat{b}x_i \\ &= a + bx_i + u_i - \bar{y} + \hat{b}\bar{x} - \hat{b}x_i \\ &= a + bx_i + u_i - a - b\bar{x} - \bar{u} + \hat{b}\bar{x} - \hat{b}x_i \\ &= a + bx_i + u_i - a - b\bar{x} - \bar{u} + \hat{b}\bar{x} - \hat{b}x_i \\ &= u_i - \bar{u} - (x_i - \bar{x})(\hat{b} - b) \end{aligned}$$

donc on a

$$\hat{u}_i^2 = (u_i - \bar{u})^2 + (x_i - \bar{x})^2(\hat{b} - b)^2 - 2(x_i - \bar{x})(\hat{b} - b)(u_i - \bar{u})$$

et

$$\sum \hat{u}_i^2 = \sum (u_i - \bar{u})^2 + (\hat{b} - b)^2 \sum (x_i - \bar{x})^2 - 2(\hat{b} - b) \sum (x_i - \bar{x})(u_i - \bar{u})$$

$$\sum \hat{u}_i^2 = \sum u_i^2 + \sum \bar{u}^2 - 2 \sum u_i \bar{u} + (\hat{b} - b)^2 \sum (x_i - \bar{x})^2 - 2(\hat{b} - b) \sum (x_i - \bar{x})(u_i - \bar{u})$$

En utilisant le fait que :

$$\sum \bar{u}^2 = N\bar{u}^2 = N \left(\frac{1}{N^2} \sum u_i^2 \right) = N^{-1} \sum u_i^2,$$

$$\sum u_i \bar{u} = \sum u_i \left(\frac{1}{N} \sum u_i \right) = N^{-1} \sum u_i^2$$

et que $(u_i - \bar{u}) = [\hat{u}_i + (\hat{b} - b)(x_i - \bar{x})]$ (d'après l'équation du haut)

on obtient :

$$\sum \hat{u}_i^2 = \sum u_i^2 + N^{-1} \sum u_i^2 - 2N^{-1} \sum u_i^2 + (\hat{b} - b)^2 \sum (x_i - \bar{x})^2 - 2(\hat{b} - b) \sum (x_i - \bar{x})[\hat{u}_i + (\hat{b} - b)(x_i - \bar{x})]$$

On peut décomposer le dernier élément de l'équation précédente en deux :

$$2(\hat{b} - b) \sum (x_i - \bar{x})[\hat{u}_i + (\hat{b} - b)(x_i - \bar{x})] = 2(\hat{b} - b) \sum (x_i - \bar{x})\hat{u}_i + 2(\hat{b} - b)^2 \sum (x_i - \bar{x})^2$$

donc

$$\sum \hat{u}_i^2 = \sum u_i^2 + N^{-1} \sum u_i^2 - 2N^{-1} \sum u_i^2 - (\hat{b} - b)^2 \sum (x_i - \bar{x})^2 - 2(\hat{b} - b) \sum (x_i - \bar{x})\hat{u}_i$$

En utilisant le fait que $E(\sum \hat{u}_i^2) = \sum E(\hat{u}_i^2) = N\sigma^2$, que x est une variable certaine et que $E(u_i) = 0$ on peut écrire que :

$$E(\sum \hat{u}_i^2) = N\sigma^2 + \sigma^2 - 2\sigma^2 - \sum (x_i - \bar{x})^2 E((\hat{b} - b)^2)$$

Comme $E((\hat{b} - b)^2) = \hat{V}_{\hat{b}} = \sigma^2(\sum (x_i - \bar{x})^2)^{-1}$, il vient :

$$E(\sum \hat{u}_i^2) = N\sigma^2 + \sigma^2 - 2\sigma^2 - \sigma^2 = (N - 2)\sigma^2$$

donc on a :

$$E[\sum \hat{u}_i^2] = (N - 2)\sigma^2$$

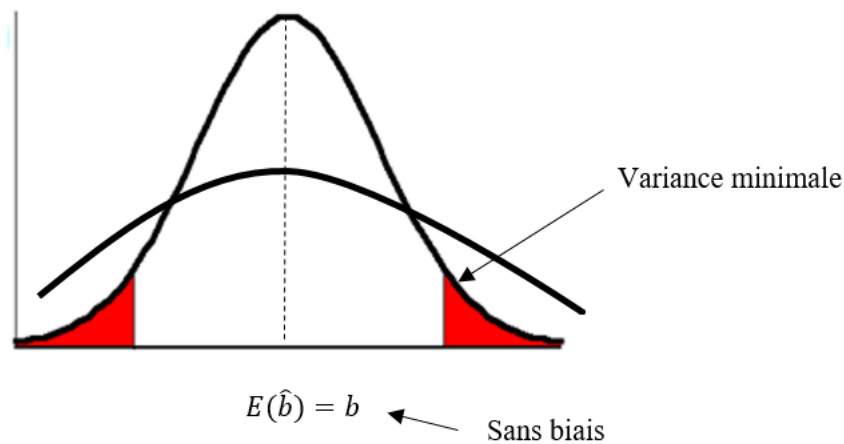
Par conséquent, un estimateur non biaisé (c'est-à-dire $E(\hat{u}_i) = u_i$) de σ^2 est :

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{N - 2} \quad (5)$$

2.3.5 Résumé

On parlera d'un estimateur sans biais si $E(\hat{b}) = b$ et d'un estimateur efficace si sa variance est minimale. D'après le théorème de Gauss Markov, l'estimateur des MCO est le meilleur estimateur sans biais, on dira qu'il s'agit d'un estimateur BLUE, car les 2 conditions sont vérifiées :

- Non biaisé (ou convergent)
- Variance minimale



On préférera toujours un estimateur non biaisé même si sa variance n'est pas minimale

2.4 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'infirmier 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

2.4.1 Coefficient de détermination et puissance du modèle

L'analyse de la variance est une procédure statistique qui consiste à reproduire le modèle linéaire simple en termes de variation. Cela permet d'apprécier la qualité de l'ajustement linéaire en calculant le coefficient de détermination noté R^2 et de donner une idée sur la significativité globale du modèle. On note par convention :

$$\text{Variation totale (SCT)} = \text{Variation expliquée par la régression (SCE)} + \text{Variation résiduelle (SCR)}$$

Ce qui correspond à :

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2 \quad (6)$$

Remarquons que $\sum (y_i - \hat{y}_i)^2 = \sum \hat{u}_i^2$

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

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} \quad (7)$$

Si le coefficient de détermination tend vers 1, l'ajustement linéaire est de bonne qualité alors que s'il tend vers 0, il est de mauvaise qualité. L'interprétation de ce coefficient est très simple. Supposons $R^2 = 0.9$, cela signifie que 90% de la variation totale de y est expliquée par le modèle. Nous verrons dans le prochain chapitre les limites de cet indicateur.

2.4.2 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. Il faut se souvenir que ce terme représente la racine carrée de la variance : il est donc directement lié à la précision de l'estimateur. En effet, 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 $1 - \alpha$.

Pour connaître la distribution des estimateurs, nous allons faire appel à l'hypothèse $u_i \sim N(0, \sigma^2)$. Cette hypothèse implique que :

$$E(y_i) = E(a + bx_i + u_i) = a + bx_i$$

$$V(y_i) = E[(y_i - E(a + bx_i + u_i))^2] = E(u_i^2) = V(u_i) = \sigma^2$$

X est non stochastique mais Y est une variable stochastique puisqu'elle dépend du terme d'erreur u_i . Sachant que $u_i \sim N(0, \sigma^2)$ on a

$$y \sim N(a + bx_i, \sigma^2) \quad (8)$$

On déduit également des résultats précédents :

$$\hat{a} \sim N(a, \sigma_{\hat{a}}^2) \quad (9)$$

$$\hat{b} \sim N(b, \sigma_{\hat{b}}^2) \quad (10)$$

En utilisant les résultats (9) et (10), les statistiques suivantes vont suivre une loi normale centrée réduite :

$$\frac{\hat{a} - a}{\sigma_{\hat{a}}} \sim N(0, 1) \quad \frac{\hat{b} - b}{\sigma_{\hat{b}}} \sim N(0, 1)$$

Rappelons que $V(\hat{b}) = \sigma_{\hat{b}}^2 = \frac{\sigma_u^2}{\sum(x_i - \bar{x})^2}$

Ainsi nous ne pouvons pas connaître $V(\hat{b})$ puisque nous ne connaissons pas l'estimation de l'erreur σ_u .

Nous procédons donc à une estimation :

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

Avec $\sigma_{\hat{u}}^2 = \frac{SCR}{N-2}$ où $N-2$ correspondent aux degrés de liberté (le 2 ici correspond au nombre de paramètres estimés).

Pour pouvoir définir la distribution de \hat{b} il faut connaître la distribution de \hat{u} . On sait par hypothèse que :

$\frac{u - E(u)}{\sigma_u} = \frac{u}{\sigma_u} \sim N(0, 1)$. Comme \hat{u} est une réalisation de u on peut écrire :

$$\frac{\hat{u}}{\sigma_u} \sim N(0, 1)$$

En passant au carré on obtient $\sum \left(\frac{u_i}{\sigma_u}\right)^2 = \frac{\sum \hat{u}^2}{\sigma_u^2}$. Puisque $\sum \hat{u}^2 = SCR = \hat{\sigma}_u^2(N-2)$. Ainsi, $\frac{\hat{\sigma}_u^2}{\sigma_u^2}(N-2) \sim \chi^2(N-2)$

Rappel sur la loi du χ^2 . Soit une variable aléatoire $X \sim N(0, 1)$ alors $Z = \sum_{i=1}^N X_i^2 \sim \chi_n^2$

Par ailleurs, on sait que $\sigma_b^2 = \frac{\sigma_u^2}{\sum(x_i - \bar{x})^2}$ et $\hat{\sigma}_b^2 = \frac{\hat{\sigma}_u^2}{\sum(x_i - \bar{x})^2}$.

Ainsi, $\frac{\hat{\sigma}_b^2}{\sigma_b^2} = \frac{\hat{\sigma}_u^2}{\sigma_u^2}$

$$\text{D'où : } \frac{\hat{\sigma}_b}{\sigma_b} \sim \sqrt{\frac{\chi^2(N-2)}{(N-2)}}$$

En se fixant, une marge d'erreur de $\alpha\%$, on peut déterminer un intervalle de confiance concernant la valeur des paramètres. Choisissons le paramètre b par exemple. On a dans ce cas :

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

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

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

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

Cependant, on ne peut pas directement calculer cet intervalle de confiance. En effet, on ne peut pas déterminer σ_b car on ne connaît pas la valeur de σ_u^2 . Il faudra donc calculer $\frac{\hat{b}-b}{\hat{\sigma}_b}$

On sait que la loi de Student est définie comme le rapport entre une loi normale et la racine carrée d'une loi du χ^2 normalisée par ses degrés de liberté.

Ainsi :

$$\frac{\frac{\hat{b}-b}{\hat{\sigma}_b}}{\frac{\hat{\sigma}_b}{\sigma_b}} = \frac{N(0,1)}{\sqrt{\frac{\chi^2(N-2)}{(N-2)}}}$$

En modifiant cette expression, on obtient :

$$t_c^* = \frac{\hat{b} - b}{\hat{\sigma}_b} \sim t_{N-2} \quad (11)$$

Bien évidemment, on peut montrer la même chose pour le paramètre a en utilisant les expressions correspondantes. La loi de Student a une densité de probabilité symétrique comme la loi normale mais elle est plus aplatie avec des queues de distribution plus épaisses. Notons également que la limite d'une loi de Student lorsque le nombre de degré de liberté

tend vers l'infini est une loi normale centrée réduite. Par conséquent, lorsque $N \rightarrow \infty$, on peut utiliser une loi normale à la place de la loi de Student. Nous allons maintenant présenter quelques tests statistiques.

2.4.3 Test sur la valeur d'un coefficient

Les coefficients testés prennent des valeurs réelles particulières selon la théorie économique. Dans ce cas on teste la validité de ces hypothèses par le test suivant :

$$\begin{cases} H0 : B_i = m \\ H1 : B_i \neq m \end{cases}$$

où $B_i \in \{a, b\}$ est un paramètre du modèle et m , la valeur du paramètre que l'on souhaite tester. 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 l'expression (11) notée 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 H0
 - Si $|t_c^*| < t_{tab}$, on accepte H0.

On peut également utiliser une approche alternative en déterminant l'intervalle de confiance de la valeur du paramètre testé. Dans ce cas, si la valeur assignée au paramètre (m) appartient à l'intervalle de confiance, on accepte H0 et inversement.

2.4.4 Test de significativité d'un paramètre

Contrairement au cas précédent, ici, on cherche à savoir si le paramètre est statistiquement significatif, i.e., s'il est significativement différent de 0.

$$\begin{cases} H0 : B_i = 0 \\ H1 : B_i \neq 0 \end{cases}$$

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 l'expression (11) notée 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 .

2.4.5 Significativité globale d'un modèle et test de Fisher

Pour rendre compte de la significativité globale d'un modèle, il faut dans un premier temps présenter le tableau d'analyse de la variance :

Source de variation	\sum des carrés	degré de liberté	carrés moyens
Régression	$SCE = \sum (\hat{y}_i - \bar{y})^2$	1	$SCE/1$
Résidu	$SCR = \sum (y_i - \hat{y}_i)^2$	N-2	$SCR/N - 2$
Totale	$SCT = \sum (y_i - \bar{y})^2$	N-1	

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. Plus précisément, on teste l'hypothèse que l'ensemble des coefficients soient nuls à l'exception de la constante, c'est à dire :

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

Etant donné que $SCE \sim \chi_1^2$ et $SCR \sim \chi_{N-2}^2$, on utilise la statistique suivante :

$$F = \frac{\frac{SCE}{1}}{\frac{SCR}{N-2}} = \frac{R^2}{1 - R^2} (N - 2) \sim F(1, N - 2) \quad (12)$$

Rappel : Pour une v.a X qui suit une loi du χ^2 à n_1 degré de liberté et une v.a Y qui suit une loi du χ^2 à n_2 degré de liberté, on a $\frac{\frac{X}{n_1}}{\frac{Y}{n_2}} \sim F(n_1, n_2)$

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 (12) notée F_c^*
- 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 .

2.5 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. Cela pourrait lui permettre de gérer de façon plus optimale son outil de production, ses besoins en main-d'oeuvre et ses stocks.

Bien évidemment, pour discuter la qualité de sa prévision, l'économètre doit être capable d'évaluer l'erreur de prévision. Cette erreur de prévision est notée $u_i^* = Y_i^* - \hat{Y}_i$. La valeur prévue de Y notée Y_i^* suit une loi normale d'espérance $a + bx_i^*$ où x_i^* est la valeur future de la variable explicative, i.e, $Y_i^* \sim N(a + bx_i^*, \sigma^2)$.

L'espérance de l'erreur de prévision est :

$$E(u_i^*) = E(a + bx_i^* + u_i - \hat{a} - \hat{b}x_i^*) = E(u_i) = 0$$

La variance de l'erreur de prévision est :

$$V(u_i^*) = V(Y_i^* - \hat{Y}_i)$$

Rappel sur la variance : $V(X+Y) = V(X) + V(Y) + 2COV(X, Y)$, lorsque $X \perp Y$ $V(X + Y) = V(X) + V(Y)$.

Remarquons que la valeur de Y_i^* dépend de l'erreur de prévision alors que la valeur de \hat{Y}_i dépend des erreurs observées pour les autres individus/périodes. Par conséquent on a $Y_i^* \perp \hat{Y}_i$. On a donc $V(u_i^*) = V(Y_i^*) + V(-\hat{Y}_i)$. Puisque $V(-\hat{Y}_i) = V(\hat{Y}_i)$ on a :

$$V(u_i^*) = V(Y_i^*) + V(\hat{Y}_i)$$

Par hypothèse des MCO, on a $V(Y_i^*) = \sigma^2$ et on peut écrire : $V(u_i^*) = \sigma^2 + V(\hat{a} + \hat{b}x_i^*) = \sigma^2 + V(\hat{a}) + V(\hat{b}x_i^*) + 2x_i^*Cov(\hat{a}, \hat{b})$

Rappel sur la variance : $V(aX) = a^2V(X)$

$$V(u_i^*) = \sigma^2 + V(\hat{a} + \hat{b}x_i^*) = \sigma^2 + V(\hat{a}) + (x_i^*)^2V(\hat{b}) + 2x_i^*Cov(\hat{a}, \hat{b})$$

En utilisant le fait que $Cov(\hat{a}, \hat{b}) = cov(\bar{y} - \hat{b}\bar{x}, \hat{b})$, $Cov(c + X, Y) = Cov(X, Y)$ et que $Cov(aX, Y) = aCov(X, Y)$, on peut écrire que :

$$Cov(\hat{a}, \hat{b}) = -\bar{x}Cov(\hat{b}, \hat{b}) = -\bar{x}V(\hat{b}) = -\bar{x}\frac{\sigma^2}{\sum(x_i - \bar{x})^2}$$

En réintroduisant cette dernière expression dans $V(u_i^*)$, on obtient :

$$V(u_i^*) = \sigma^2 + \frac{\sigma^2}{N} + \frac{\bar{x}^2\sigma^2}{\sum(x_i - \bar{x})^2} + x_i^2\frac{\sigma^2}{\sum(x_i - \bar{x})^2} - 2x_i\bar{x}\frac{\sigma^2}{\sum(x_i - \bar{x})^2}$$

$$V(u_i^*) = \sigma^2 \left[1 + \frac{1}{N} + \frac{(x_i^* - \bar{x})^2}{NV(X)} \right]$$

De là on déduit que :

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

Comme dans le modèle linéaire simple, on a :

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

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

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

\Leftrightarrow

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

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

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

2.6 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.

Selon la forme de chaque variable dépendante et indépendante, l'interprétation des coefficients n'est pas la même.

Prenons l'exemple d'une équation de Mincer où le salaire est estimé en fonction de l'expérience d'un individu. Selon la façon dont le salaire et l'expérience sont insérés dans l'équation, l'interprétation sera différente.

Modèle niveau-niveau. Considérons le modèle linéaire suivant, estimé par les moindres carrés :

$$Sal_i = a + bExp_i + u_i$$

Formellement, le coefficient b correspond à la dérivée partielle du salaire par rapport à l'expérience :

$$\frac{\partial Sal_i}{\partial Exp_i} = b$$

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.

Modèle log-log Considérons le même modèle, mais où les variables sont exprimées en logarithme. Le passage en logarithme se justifie de deux manières : (i) il permet de passer d'une forme multiplicative à une forme linéaire, (ii) il permet d'aplanir la variance des variables.

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

Ici la dérivée partielle $\frac{\partial Sal_i}{\partial Exp_i}$ ne peut se calculer directement, il faut procéder à certaines modifications :

$$Sal_i = e^{a+b\ln(Exp_i)+u_i}$$

Rappel : $e^{a+b} = e^a e^b$. Ainsi :

$$Sal_i = e^a e^{b\ln(Exp_i)} e^{u_i}$$

Rappel : $\ln(x^a) = a\ln(x)$. Ainsi :

$$Sal_i = e^a e^{\ln(Exp_i^b)} e^{u_i}$$

$$Sal_i = Exp_i^b e^{a+u_i} \quad (13)$$

$$\text{Ainsi, } \frac{\partial Sal_i}{\partial Exp_i} = bExp_i^{b-1} e^{a+u_i} = \frac{bExp_i^b e^{a+u_i}}{Exp_i}$$

Or, d'après (13) $Sal_i = Exp_i^b e^{a+u_i}$ d'où $\frac{\partial Sal_i}{\partial Exp_i} = \frac{b Sal_i}{Exp_i}$ et :

$$b = \frac{\partial Sal_i}{\partial Exp_i} \frac{Exp_i}{Sal_i}$$

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.

Modèle Log-niveau. Dans ce cas, la variable dépendante est sous forme logarithmique et la variable indépendante est en niveau. L'équation est alors la suivante :

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

On procède de la même manière :

$$Sal_i = e^a e^{bExp_i} e^{u_i}$$

On dérive ensuite par rapport à Exp_i :

$$\frac{\partial Sal_i}{\partial Exp_i} = b e^a e^{bExp_i} e^{u_i}$$

D'où :

$$\frac{\partial Sal_i}{\partial Exp_i} = b Sal_i. \text{ Ainsi,}$$

$$b = \frac{\partial Sal_i}{\partial Exp_i} \frac{1}{Sal_i} = \frac{\frac{\partial Sal_i}{\partial Exp_i}}{\frac{Sal_i}{Exp_i}} = \frac{\Delta Sal_i}{Exp_i}$$

En multipliant par 100 de chaque côté, on obtient :

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

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 $100 \times b$.

Modèle niveau-log Ici, il s'agit d'un modèle dans lequel la variable dépendante est en niveau et la variable indépendante a subi une transformation logarithmique :

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

On procède de la même manière que précédemment pour connaître l'interprétation du coefficient b :

$$\frac{\partial Sal_i}{\partial Exp_i} = \frac{b}{Exp_i}$$

$b = \frac{Exp_i}{\partial Exp_i} * \partial Sal_i =$. En inversant on obtient :

$\frac{1}{b} = \frac{\partial Exp}{Exp} * \frac{1}{\partial Sal}$. En multipliant par 100 de part et d'autre pour obtenir un raisonnement en pourcentage nous obtenons :

$\frac{100}{b} = \frac{\Delta Exp_i \%}{\partial Sal_i}$. D'où :

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

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

3 Le modèle linéaire multiple

• Introduction

On envisage un modèle plus général que dans le chapitre précédent. La forme fonctionnelle adoptée reste linéaire, seulement plusieurs variables peuvent intervenir pour décrire le comportement de la variable expliquée. Le modèle linéaire multiple est une extension du modèle linéaire simple à un nombre $k > 1$ de variables explicatives. Considérons à nouveau l'exemple 1, à savoir une fonction de production Cobb-Douglas :

$$Y = AK^\alpha L^\beta$$

que l'on peut log-linéariser

$$\log(Y) = \log(A) + \alpha \log(K) + \beta \log(L)$$

Notons que l'avantage de log-linéariser une fonction est que les paramètres de la régression deviennent des élasticités. En conséquence, en supposant que $\alpha = 0.5$, cela implique qu'une augmentation de 1% de la quantité de capital augmente la production de 0,5%. Dans ce modèle, il y a, en plus de la constante, deux variables explicatives. La méthode du chapitre 1 ne nous permet pas d'estimer un tel modèle. Nous allons maintenant décrire la méthode permettant d'estimer un modèle avec $k > 1$ variables explicatives (hors constante).

3.1 Le modèle et ses hypothèses

3.1.1 Ecriture du modèle

Soit un modèle linéaire de la forme :

$$y_i = b_0 + b_1x_{1i} + b_2x_{2i} + \dots + b_kx_{ki} + u_i, \quad i = 1, \dots, N$$

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 :

$$k \leq N$$

Pour aboutir à une écriture matricielle on empile les N observations de la façon suivante :

$$\begin{aligned} y_1 &= b_0 + b_1x_{11} + b_2x_{21} + \dots + b_kx_{k1} + u_1 \\ y_2 &= b_0 + b_1x_{12} + b_2x_{22} + \dots + b_kx_{k2} + u_2 \\ &\dots \\ y_N &= b_0 + b_1x_{1N} + b_2x_{2N} + \dots + b_kx_{kN} + u_N \end{aligned}$$

On peut le représenter sous forme matricielle comme :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = b_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + b_1 \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1N} \end{bmatrix} + \dots + b_k \begin{bmatrix} x_{k1} \\ x_{k2} \\ \vdots \\ x_{kN} \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}$$

En factorisant, on obtient :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1N} & x_{2N} & \dots & x_{kN} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}$$

ou encore

$$Y_{N \times 1} = X_{N \times (k+1)} B_{(k+1) \times 1} + u_{N \times 1} \quad (14)$$

On va donc avoir à estimer $k + 1$ paramètres relatifs à la partie systématique du modèle et un paramètre relatif à la partie aléatoire (σ^2).

3.1.2 Les hypothèses

On retient la même démarche que celle adoptée pour le modèle linéaire simple :

- a) énoncer des hypothèses potentiellement vérifiées
- b) étudier les propriétés des MCO
- c) Tester les hypothèses et au besoin changer de modèle

Les hypothèses du modèle sont analogues à celles énoncées au premier chapitre :

$$H1) \begin{matrix} E(u) = 0 \\ (N,1) \quad (N,1) \end{matrix}$$

$$H2) E(uu') = \sigma^2 I_N$$

c'est à dire que l'on suppose que la matrice de variance covariance des perturbations est une matrice scalaire : elle s'écrit comme le produit de la matrice identité par un scalaire.

$$E \begin{pmatrix} u & u' \\ (N,1) & (N,1) \end{pmatrix} = E \left[\begin{matrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \\ \times [u_1 u_2 \dots u_N] \end{matrix} \right]$$

$$E \begin{pmatrix} u & u' \\ (N,1) & (N,1) \end{pmatrix} = E \begin{bmatrix} u_1^2 & u_1 u_2 & \dots & u_1 u_N \\ u_2 u_1 & u_2^2 & \dots & u_2 u_N \\ \vdots & \vdots & \ddots & \vdots \\ u_N u_1 & u_N u_2 & \dots & u_N^2 \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

Par conséquent, comme pour le modèle linéaire simple, on suppose $E(u_i^2) = \sigma^2$ et $E(u_i u_s) = Cov(u_i u_s) = 0$

$$H3) u \sim N(0, \sigma^2 I_N)$$

H4) La matrice des variables explicatives X est une matrice certaine (exogène)

H5) $\begin{matrix} X \\ (N, k+1) \end{matrix}$ est de rang $Rg(X) = k + 1 < N$, c'est à dire de plein rang 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. Nous détaillerons cette condition plus loin.

H6) $Cov(X, u) = 0 \Rightarrow$ la covariance entre la partie systématique et celle aléatoire est nulle.

3.2 Les estimateurs et leurs propriétés

3.2.1 Estimation des coefficients de la régression et propriétés des estimateurs

Etant donné que la structure fondamentale du modèle n'a pas été modifiée, la méthode des MCO reste la méthode de calcul la plus appropriée permettant d'obtenir un vecteur \hat{B} efficient. Comme pour le modèle linéaire simple, la fonction objectif est :

$$\min_{b_0, \dots, b_k} \sum \hat{u}_i^2 = \min_{b_0, \dots, b_k} \sum (y_i - b_0 - b_1 x_{1i} - \dots - b_k x_{ki})^2$$

Dans la suite, on utilise des notations vectorielles et matricielles, lesquelles permettent d'aboutir à des écritures plus compactes.

Pour trouver \hat{b} le vecteur de dimension $(k, 1)$ on doit résoudre le programme suivant :

$$\boxed{\text{Min}_b \sum_{i=1}^N (y_i - X'_{(1,k)(k,1)} b)^2}$$

Ceci revient à écrire \hat{b} solution de $\text{Min}_b S(b)$, avec :

$$\boxed{S(b) = \begin{pmatrix} y_i \\ N,1 \end{pmatrix} - \begin{pmatrix} X \\ (N,k)(k,1) \end{pmatrix} \begin{pmatrix} b \\ N,1 \end{pmatrix} \begin{pmatrix} y_i \\ N,1 \end{pmatrix} - \begin{pmatrix} X \\ (N,k)(k,1) \end{pmatrix} \begin{pmatrix} b \\ N,1 \end{pmatrix}}$$

Pour trouver les valeurs estimées des paramètres, il faut satisfaire les conditions du premier ordre :

$$\frac{\partial S(b)}{\partial b_j} = 0, j = 0, \dots, k$$

Pour cela nous allons revenir sur les dérivées des vecteurs de formes quadratique et linéaire.

Proposition 1 : Soit v et a deux vecteurs $\in R^k$:

$$\frac{\partial a'v}{\partial v} = \frac{\partial v'a}{\partial v} = a$$

Proposition 2 : Soit un vecteur $v \in R^k$ et une matrice $M \in R^{k+k}$

$$\frac{\partial v' M v}{\partial v} = (M + M')v$$

En particulier, si M est symétrique $M' = M$ et

$$\frac{\partial v' M v}{\partial v} = 2 M v$$

Nous pouvons désormais résoudre $\frac{\partial S(b)}{\partial b_j} = 0 \rightarrow$

$$S(b) = y'y - b'X'y - y'Xb + b'X'Xb$$

On peut écrire $b'X'y = y'Xb$, donc :

$$S(b) = y'y - 2y'Xb + b'X'Xb$$

En appliquant les règles de dérivation vus précédemment on obtient :

$$\frac{\partial S(b)}{\partial b_j} = -2X'y + 2X'Xb = 0$$

Soit :

$$X'X\hat{b} = X'y$$

et par conséquent, l'estimateur des MCO est donné par :

$$\boxed{\hat{b} = (X'X)^{-1}X'y}$$

3.3 Interprétation de l'estimateur des MCO dans le cas de la régression multiple

Lorsqu'il n'y avait qu'une seule variable à expliquer, nous avons vu que b s'interprète comme l'effet d'une variation de X sur Y . L'effet peut s'exprimer en unité ou en pourcentage selon que les variables soient exprimées en log ou en niveau (voir section 2.6).

Lorsqu'il y a plusieurs variables l'interprétation d'un coefficient s'interprète "toutes choses égales par ailleurs".

Preuve :

La section précédente a montré que l'estimateur des MCO s'écrivait $X'X\hat{b} = X'y$ pour chaque variable associée.

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

$$\begin{bmatrix} X'_1X_1 & X'_1X_2 \\ X'_2X_1 & X'_2X_2 \end{bmatrix} \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \end{pmatrix} = \begin{pmatrix} X'_1y \\ X'_2y \end{pmatrix}$$

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

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

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 :

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

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 :

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

Avec $y - X_1\hat{b}_1$ les résidus estimés d'une régression sur la droite d'équation $y = bX_1 + u_i$

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

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

$$M_1' = M_1$$

$$\text{et } M_1M_1 = M_1$$

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 :

$$M_1y = M_1X_2b_2 + u_i$$

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 .

Ceci est obtenu en régressant :

a) M_1y , 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_1X_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)

Ainsi on voit bien que le coefficient X_2 est obtenu toute choses égales par ailleurs, c'est-à-dire en purgeant de l'effet de X_1

3.3.1 Interprétation avec variables d'interaction

En introduction, nous avons vu qu'il était possible d'introduire un effet non quadratique dans un modèle linéaire. Ceci implique que même si nous disposons que d'une seule variable, il faudra prendre son carré identifier un effet non linéaire.

Ceci correspond à un effet d'interaction simple qui consiste à multiplier une variable par elle même. Il est possible d'avoir des effets d'interaction plus complexes pour voir si il existe des effets hétérogènes d'une variable selon l'appartenance à un groupe (genre, éducation, âge par exemple).

Il existe plusieurs cas de figures selon le type de variables d'interaction : dichotomiques ou continues.

Interaction entre une variable indicatrice et une variable continue Partons d'un modèle sans interaction :

$$Salaire_i = b_0 + b_1Educ_i + b_2Sex$$

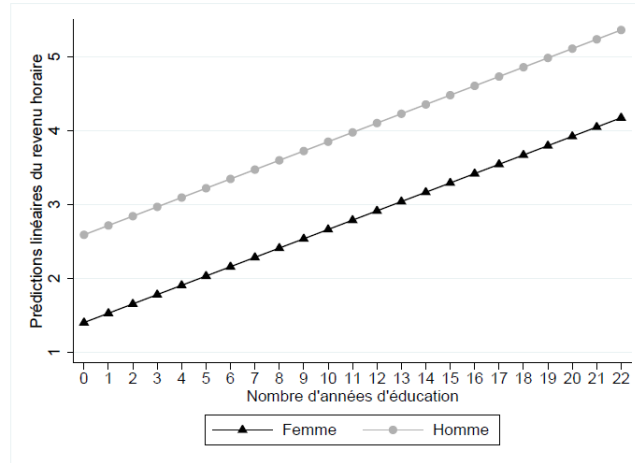
$Educ$ est une variable continue et Sex est une variable dichotomique qui prend la valeur 1 si l'individu est un homme et zéro si il s'agit d'une femme.

b_1 représente donc l'effet moyen d'une année supplémentaire d'éducation sur le salaire (quelque soit le genre de l'individu).

L'effet de l'éducation ne variant pas sur le genre, la pente ne change pas entre les hommes et les femmes, mais l'ordonnée à l'origine change. Elle est égale à b_0 pour les femmes et $b_0 + b_2$ pour les hommes.

Ainsi, b_2 indique le différentiel de salaire entre un homme et une femme, quel que soit le niveau d'éducation.

FIGURE 1 – Modèle sans interaction



Si l'on veut savoir si l'écart de revenu entre hommes et femmes varie en fonction de l'éducation, il faut mettre un terme d'interaction.

Dans cette nouvelle équation, nous autorisons l'effet de l'éducation à varier selon le genre. La pente peut donc être différente selon le genre :

$$Sal_i = b_0 + b_1 Educ_i + b_2 Sex_i + b_3 (Educ_i \times Sex_i)$$

De manière générale, l'interaction indique comment l'effet associé avec une variable X change selon les groupes considérés

$$y = b_0 + b_1 X_1 + b_2 d + b_3 (X_1 \times d)$$

L'interprétation se fait selon que d soit égal à zéro ou à 1.

Lorsque $d = 0$, il s'agit de l'équation dans l'échantillon des femmes. Cela correspond à

$$y = b_0 + b_1 X_1$$

b_1 correspond donc à l'effet d'une année supplémentaire d'étude pour les femmes.

Lorsque $d = 1$, il s'agit de l'équation dans l'échantillon des hommes. Cela correspond à

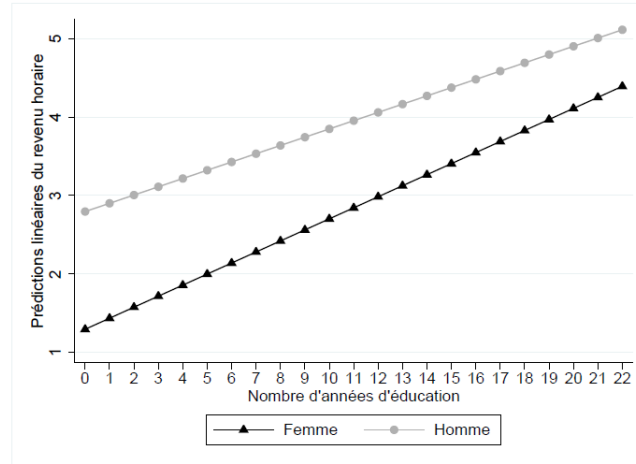
$$y = b_0 + b_1 X_1 + b_2 \times 1 + b_3 (X_1 \times 1)$$

L'effet de l'éducation chez les hommes est donc égal à $b_1 + b_3$

Ainsi, l'impact de X_1 varie pour $d = 1$ et $d = 0$. Quatre cas sont possibles.

— b_1 est positif et b_3 est positif : l'effet positif de X_1 est plus fort pour le groupe $d = 1$

FIGURE 2 – Modèle avec interaction



- b_1 est positif et b_3 est négatif : l'effet positif de X_1 est plus fort pour le groupe $d = 0$ (moins fort pour $d = 1$)
- b_1 est négatif et b_3 est positif : l'effet négatif de X_1 est moins fort pour le groupe $d = 1$
- b_1 est négatif et b_3 est négatif : l'effet négatif de X_1 est plus fort pour le groupe $d = 1$

Reprenons notre exemple précédent. Nous avons :

	(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

Interaction entre deux variables dichotomiques Regardons ce qu'il se passe quand deux variables dichotomiques sont interagies :

$$y = b_0 + b_1d_1 + b_2d_2 + b_3(d_2 \times d_3)$$

y représente le salaire, d_1 indique la scolarisation de l'individu (1 si oui, 0 si non) et d_2 le genre (1 si homme, 0 si femme).

Quatre profils peuvent être distingués :

- femmes non scolarisées : $y = b_0 + u_i$
- femmes scolarisées : $y = b_0 + b_1d_1 + u_i$
- hommes non scolarisés : $y = b_0 + b_2d_2 + u_i$
- hommes scolarisés : $y = b_0 + b_2d_2 + b_3(d_2 \times d_3) + u_i$

Ainsi b_0 représente le salaire pour les femmes non scolarisées. b_1 indique l'effet de la scolarisation chez les femmes. b_2 indique l'effet d'être un homme sur les salaires. $b_2 + b_3$ indique l'effet de la scolarisation chez les hommes.

Interaction entre deux variables continues Il est également possible d'interagir deux variables continues, par exemple l'éducation et l'expérience :

$$Sal_i = b_0 + b_1educ + b_2exp + b_3(educ \times exp)$$

$$Sal_i = b_0 + b_1educ + (b_2 + b_3educ)exp$$

$$Sal_i = b_0 + b_2exp + (b_1 + b_3exp)educ$$

b_0 représente l'effet sur le salaire quand l'expérience et l'éducation est nulle.

b_1 représente l'effet de l'éducation quand l'expérience est nulle.

b_2 représente l'effet de l'expérience quand l'éducation est nulle.

Avec un tel modèle, l'effet de l'expérience sur le revenu dépend de l'éducation : $b_2 + b_3 \times Educ$. L'effet de l'éducation sur le revenu dépend du niveau d'expérience : $b_1 + b_3 \times Exp$

3.4 Biais et efficacité des estimateurs MCO

Note : Définition d'un estimateur non biaisé et efficace

On dit qu'un estimateur de m noté \hat{m} est dit sans biais ssi $E(\hat{m}) = m$.

Supposons que nous ayons N estimateurs non biaisés de m notés : $\hat{m}_1, \hat{m}_2, \dots, \hat{m}_n$, $N = 1, \dots, n$. On appelle **estimateur efficace** / **estimateur BLUE** l'estimateur non biaisé dont la variance est minimale. Supposons trois estimateurs non biaisés de m notés $\hat{m}_1, \hat{m}_2, \hat{m}_3$ et que $V(\hat{m}_1) < V(\hat{m}_2) < V(\hat{m}_3)$. L'estimateur le plus efficace de m est \hat{m}_1 .

Montrons que l'estimateur des MCO est sans biais :

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

Calculons maintenant la variance de l'estimateur \hat{B} :

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

Comme dans le modèle linéaire simple, on retrouve dans la variance de \hat{B} la variance de u , σ^2 , et $X'X$, c'est à dire la quantité d'information apportée par X . D'après le Théorème de Gauss-Markov, sous les hypothèses H1 à H6, il n'existe pas d'estimateur linéaire sans biais des coefficients B ayant une variance plus petite que celle des estimateurs des moindres carrés. On dit que l'estimateur des MCO est BLUE (Best Linear Unbiased Estimator), ou efficace. C'est l'estimateur le plus précis de l'ensemble des estimateurs linéaires sans biais de B .

3.4.1 Estimation du coefficient de la partie aléatoire

Comme pour le modèle linéaire simple, pour calculer la variance des coefficients estimés \hat{B} , nous devons connaître σ^2 . Or, on ne connaît pas σ^2 et il faut donc l'estimer. Comme pour le modèle linéaire multiple, on a :

$$V(\hat{u}) = E(\hat{u}'\hat{u}) = E[\text{Tr}(\hat{u}\hat{u}')] = \text{Tr}(E[\hat{u}\hat{u}']) = \text{Tr}(\text{Var}(\hat{u})) = \text{Tr}(\sigma^2 P_X \perp)$$

Comme $P_X \perp$ est la matrice de la projection orthogonale sur un espace de dimension $(N - (k + 1))$, on a :

$$E(\hat{u}'\hat{u}) = (N - (k + 1))\sigma^2 \quad (15)$$

Par conséquent, un estimateur sans biais de σ^2 est donné par :

$$\hat{\sigma}^2 = \frac{E(\hat{u}'\hat{u})}{N - (k + 1)} \quad (16)$$

Notons que $E(u'u) = SCR = Y'Y - \hat{B}'X'Y$, en effet :

$$\begin{aligned} \hat{u}'\hat{u} &= (Y - X\hat{B})'(Y - X\hat{B}) \\ &= Y'Y - 2\hat{B}'X'Y + \hat{B}'X'X\hat{B} \\ &= Y'Y - 2\hat{B}'X'Y + \hat{B}'X'Y \\ &= Y'Y - \hat{B}'X'Y \end{aligned}$$

3.4.2 Analyse de la variance

L'analyse de la variance est une procédure statistique qui consiste à reproduire le modèle linéaire multiple en termes de variation. Cela permet d'apprécier la qualité de l'ajustement linéaire en calculant le coefficient de détermination noté R^2 et de donner une idée sur la significativité globale du modèle. On note par convention :

Variation totale (SCT) = Variation liée à la régression (SCE) + Variation résiduelle (SCR)

Ce qui correspond à :

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2 \quad (17)$$

Pour rendre compte de la significativité globale d'un modèle, il faut présenter le tableau d'analyse de la variance :

Source de variation	\sum des carrés	degré de liberté	carrés moyens
Régression	$SCE = \hat{Y}'\hat{Y} - N\bar{Y}^2$	k	SCE/k
Résidu	$SCR = Y'Y - B'X'Y$	$N - (k + 1)$	$SCR/N - (k + 1)$
Totale	$SCT = Y'Y - N\bar{Y}^2$	$N - 1$	

3.4.3 Le coefficient de détermination

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} = 1 - \frac{Y'Y - B'X'Y}{Y'Y - N\bar{Y}^2}$$

Si le R^2 est une statistique simple et intuitive qui apporte une information importante sur la qualité du modèle, il convient de préciser ses limites. La première est que le R^2 est fortement sensible aux transformations du vecteur Y (il va varier si on passe en logarithme, en taux de croissance,...). La seconde est que le R^2 augmente toujours quand on introduit une variable de plus, même si cette dernière a un très faible pouvoir explicatif. Par conséquent, le R^2 ne sera pas approprié pour comparer des modèles entre eux. Or, il arrive parfois que l'on désire comparer plusieurs équations de régression multiple comportant la même variable expliquée mais dont les équations diffèrent soit par le nombre d'observations soit par le nombre de variables explicatives. Le coefficient ajusté noté \bar{R}^2 permet de tenir compte du nombre de degrés de liberté associé à la SCR qui diminue au fur et à mesure qu'une nouvelle variable explicative est introduite dans le modèle. Ce coefficient est donné par :

$$\bar{R}^2 = 1 - \frac{SCR}{N - (k + 1)} \times \frac{N - 1}{SCT}$$

ou

$$\bar{R}^2 = 1 - (1 - R^2) \frac{N - 1}{N - (k + 1)} \quad (18)$$

Si on compare deux modèles, on prendra celui qui a le \bar{R}^2 le plus élevé. Néanmoins cette statistique peut-être insuffisante pour guider le modélisateur dans son choix des variables explicatives (doit-on préférer un modèle avec 12 variables explicatives au lieu de 4 si le \bar{R}^2 est légèrement plus élevé?). De plus, à l'instar du R^2 , le \bar{R}^2 présente l'inconvénient de ne pas pouvoir s'interpréter à partir de l'équation d'analyse de la variance. En outre, il peut être négatif si :

$$1 - R^2 > \frac{N - (k + 1)}{N - 1}$$

3.5 Les tests statistiques

3.5.1 Les tests individuels

Comme pour le modèle linéaire simple, les tests individuels sont élaborés à partir d'une statistique distribuée selon la loi de Student. On peut par exemple, 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_{ieme} composante sur la diagonale principale de la matrice de variance/covariance estimée :

$$\hat{V}(\hat{B}) = \hat{\sigma}^2(X'X)^{-1} = \begin{bmatrix} \hat{V}(\hat{b}_0) & \hat{Cov}(\hat{b}_0, \hat{b}_1) & \cdot & \cdot & \hat{Cov}(\hat{b}_0, \hat{b}_k) \\ \cdot & \hat{V}(\hat{b}_1) & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \hat{V}(\hat{b}_j) & \cdot \\ \cdot & \cdot & \cdot & \cdot & \hat{V}(\hat{b}_k) \end{bmatrix}$$

La décision se prend alors de la façon suivante :

- si $|t_c| > t_{tab}$ alors on rejette H_0
- si $|t_c| < t_{tab}$ alors on accepte H_0

Finalement, les tests d'égalité simple ainsi que la construction des intervalles de confiance pour les coefficients b_j et pour σ sont traités de la même manière que pour le modèle linéaire simple. La seule différence concerne le degré de liberté (on passe de $N - 2$ à $N - (k + 1)$).

3.5.2 Généralisation

La généralisation du modèle nous permet d'envisager d'autres tests. On pourrait postuler certaines relations linéaires qui existent entre deux ou plusieurs paramètres du modèle. Par exemple, on pourrait souhaiter tester :

$$\begin{cases} H0 : \rho b_0 + \beta b_1 = c \\ H1 : \rho b_0 + \beta b_1 \neq c \end{cases}$$

Pour réaliser ce test, on utilise la statistique suivante :

$$t_c = \frac{\rho \hat{b}_0 + \beta \hat{b}_1 - c}{\sqrt{\hat{V}(\rho \hat{b}_0 + \beta \hat{b}_1)}} \sim t_{N-(k+1)}$$

où

$$\hat{V}(\rho \hat{b}_0 + \beta \hat{b}_1) = \rho^2 \hat{V}(\hat{b}_0) + \beta^2 \hat{V}(\hat{b}_1) + 2\rho\beta \text{cov}(\hat{b}_0, \hat{b}_1)$$

La décision se prend alors de la façon suivante :

- si $|t_c| > t_{tab}$ alors on rejette H0
- si $|t_c| < t_{tab}$ alors on accepte H0

3.6 La prévision

De manière analogue au modèle linéaire simple, on peut calculer un intervalle de confiance d'une prévision réalisée avec un modèle linéaire multiple. A partir des valeurs prévues des variables explicatives notées $W' = (x_{1*}, x_{2*}, \dots, x_{k*})$ et de la prévision donnée par $\hat{Y}_i = \hat{b}_0 + \sum \hat{b}_k x_{k*}$, on définit l'intervalle de confiance de la prévision donné par :

$$Ic(Y_i^*) = \left[\hat{Y}_i - t_{\alpha/2} \sqrt{\sigma^2(1 + W'(X'X)^{-1}W)}; \hat{Y}_i + t_{\alpha/2} \sqrt{\sigma^2(1 + W'(X'X)^{-1}W)} \right]$$

4 Violations des hypothèses de Gauss-Markov : conséquences et remèdes

Nous allons nous intéresser aux conséquences de la violation des hypothèses de l'estimateur des MCO pour voir quelle stratégie pour corriger des biais et/ou de la réduction de l'efficacité induits par la violation des hypothèses.

4.1 La 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. Pour le prouver on peut supposer un modèle $y = x_1\beta_1 + X_{-1}\beta_{-1} + u_i$. β_1 et β_{-1} sont les paramètres à estimer de dimension respective $(1, 1)$ et $(k - 1, 1)$. Après quelques calculs on obtient que 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} . Donc plus la relation entre x_1 et les autres variables explicatives du modèle est étroite, plus $R_{1,-1}^2$ est important et plus la variance du coefficient estimé pour cette variable sera importante.

La multicolinéarité aurait donc tendance à augmenter le risque de deuxième espèce, c'est-à-dire d'accepter H_0 alors que celle-ci est fautive et réduit le risque de première espèce (rejeter H_0 alors que celle-ci est vraie). Minimiser le risque de première espèce est moins grave que minimiser le risque de deuxième espèce, puisque ça conforte les conclusions économiques lorsqu'un coefficient est significatif.

Il n'existe pas de solution miracle pour détecter la multicolinéarité, puisqu'il ne s'agit pas d'un problème statistique mais d'un problème numérique. 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

Lorsqu'on soupçonne une multicolinéarité, la solution peut être de retirer une ou plusieurs variables explicatives. Cependant, si ces variables doivent être incluses, il existe un

biais lorsqu'on retire les variables. Il existe donc un arbitrage entre efficacité (produire des estimateurs avec une variance minimale) et biais (produire des estimateurs qui reflètent la vraie valeur du β).

4.2 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. Or, souvent sur des données en coupe transversale (données individuelles), on a besoin de tenir compte d'une hétérogénéité du comportement. Si on tient à faire apparaître cet effet au niveau de la partie aléatoire du modèle, les erreurs ne peuvent plus être homoscédastiques mais seront plutôt hétéroscédastiques. Cela signifie que la variance résiduelle dépend de l'observation considérée. Par exemple, la variance résiduelle des salaires pour les personnes éduquées peut être plus élevée que pour les personnes moins éduquées.

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.

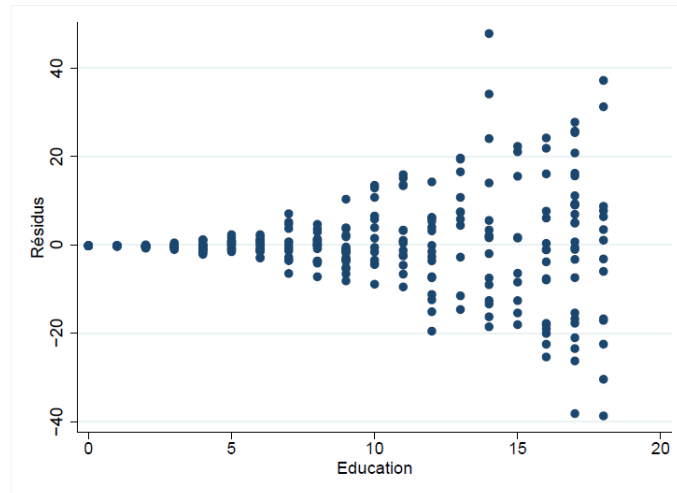
Sous cette hypothèse on peut établir que :

- 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. En effet sa matrice de variances-covariances n'est plus égale à $\sigma^2(X'X)^{-1}$ mais vaut $\sigma^2(X'X)^{-1}(X'\Omega X)(X'X)^{-1}$. Par conséquent les tests basés sur $\sigma^2(X'X)^{-1}$ ne sont pas valables.

Preuve :

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

FIGURE 3 – Représentation graphique des résidus



4.2.1 Détection visuelle de l'hétéroscédasticité

Supposons que l'on cherche à estimer le revenu en fonction de l'éducation. Lorsque l'on reporte les résidus de l'estimation en fonction de l'estimation, on observe que la plage de variation des résidus tend à s'élargir à mesure que l'éducation augmente. Le modèle est donc hétéroscédastique.

Quels remèdes ?

Lorsque le modèle est hétéroscédastique il existe deux solutions :

- Transformer le modèle pour le rendre homoscedastique : 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

4.2.2 La méthode des MCG

Les efforts menés afin de fournir de nouveaux estimateurs BLUE ont donné naissance à une procédure d'estimation générale appelée méthode des moindres carrés généralisés (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 :

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

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

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

(Rappel : $V(aX) = a^2V(X)$)

Le modèle transformé est donc homoscedastique. 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 :

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

Appliquons la méthode des MCO à ce modèle transformé. La fonction objectif consiste toujours à minimiser la somme des carrés des résidus :

$$\begin{aligned} & \min_B (\tilde{Y} - \tilde{X}B)'(\tilde{Y} - \tilde{X}B) \\ \Leftrightarrow & \min_B (\tilde{Y}'\tilde{Y} - \tilde{Y}'\tilde{X}B - B'\tilde{X}'\tilde{Y} + B'\tilde{X}'\tilde{X}B) \\ \Leftrightarrow & \min_B (\tilde{Y}'\tilde{Y} - 2B'\tilde{X}'\tilde{Y} + B'\tilde{X}'\tilde{X}B) \\ \Leftrightarrow & \min_B S(B) \end{aligned}$$

Les conditions de premier ordre impliquent :

$$\frac{\partial S}{\partial B} = 0 \Leftrightarrow -2\tilde{X}'\tilde{Y} + 2\tilde{X}'\tilde{X}B = 0$$

Par conséquent, l'estimateur des MCG est donné par :

$$\begin{aligned} \hat{B}_{MCG} &= (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y} \\ &= [(\Omega^{-1/2}X)'\Omega^{-1/2}X]^{-1}(\Omega^{-1/2}X)'\Omega^{-1/2}Y \\ &= (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y) \end{aligned} \tag{19}$$

On peut facilement vérifier que cet estimateur est sans biais :

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

Calculons maintenant la matrice de var/cov de \hat{B}_{MCG} .

Nous savons d'après l'équation précédente que $E(\hat{B}_{MCG}) - B = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}E(\epsilon)$.
Ainsi,

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

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 . Par exemple, on peut faire l'hypothèse que la variance du terme d'erreur d'une équation de revenus est une fonction de l'éducation ou du sexe, mais à priori on aimerait ne pas faire dépendre l'estimation du modèle.

4.2.3 Estimation du coefficient de la partie aléatoire

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

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

avec

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

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

L'estimateur des MCO est non biaisé avec de l'hétéroscédasticité mais il n'est pas efficace. Il est donc nécessaire de corriger le calcul de la matrice de variances-covariances.

Lorsque le modèle est hétéroscédastique, sa variance est égale à :

$$\begin{aligned}
V(\hat{\beta}_{MCO}) &= E(\hat{\beta}_{MCO} - \beta)(\hat{\beta}_{MCO} - \beta)' \\
&= (X'X)^{-1} X' E(\epsilon \epsilon') X (X'X)^{-1} \\
&= \sigma^2 (X'X)^{-1} X' \Omega X (X'X)^{-1}
\end{aligned}$$

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 $S_0 = \frac{1}{n} \sum_{i=1}^n \hat{\epsilon}_i^2 x_i x_i'$ est un estimateur convergence 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

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

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

4.2.5 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 homoscedastiques ou hétéroscedastiques

$$\begin{cases} H0 : \sigma_i^2 = \sigma^2, & \forall i \\ H1 : \sigma_i^2 \neq \sigma^2 \end{cases}$$

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. Sous l'hypothèse nulle d'homoscedasticité 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 H0 (les erreurs sont homoscedastiques)
- 2) si $F_c > F_{tab}$ alors on rejete H0 (les erreurs sont hétéroscedastiques)

• 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 :

$$\begin{cases} H0 : \sigma_i^2 = \sigma^2, V(\epsilon_i|X_i) = \sigma^2 & \forall i \\ H1 : \sigma_i^2 \neq \sigma^2 \end{cases}$$

Ce test se réalise en deux étapes :

1) On estime le modèle par les MCO.

2) On régresse les résidus estimés par les MCO sur l'ensemble des régresseurs, leur carré et leur produit.

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

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 .

La règle de décision est habituelle :

- si $\chi_{\frac{\alpha}{2}}^2 < \chi_c^2 < \chi_{1-\frac{\alpha}{2}}^2$ 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)

4.3 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, dans lesquelles on dispose d'observations répétées sur un échantillon d'individus, il peut arriver que les termes d'erreur correspondant aux observations d'un individu donné aux différentes dates soient corrélés entre eux (autocorrélation intertemporelle). 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), les termes d'erreur d'individus appartenant à un même groupe peuvent être corrélés (autocorrélation spatiale). 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.

4.3.1 Le cas particulier des données groupées

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.

$$\begin{aligned} y_{ig} &= x'_{ig}\beta + u_{ig}, i = 1, \dots, N_g; g = 1, \dots, G \\ y_g &= x'_g\beta + u_g, g = 1, \dots, G \\ y &= X'\beta + u \end{aligned}$$

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

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

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. 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.

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.

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.

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

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

La méthode est similaire avec données de panel.

4.4 L'endogénéité d'une ou de plusieurs 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 : $cov(X_i, \epsilon_i) = 0$ 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 (exemple : l'état de santé d'une personne explique le nombre d'heures travaillées et le nombre d'heures travaillées peut impacter l'état de santé)
- **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 (exemple : les aptitudes intellectuelles expliquent à la fois le revenu et le niveau d'étude)
- **Erreur de mesure** : une ou plusieurs variables explicatives du modèle sont mesurées avec erreur

4.4.1 La méthode des variables instrumentales

Deux types de méthodes peuvent être distinguées pour régler le problème de l'endogénéité d'une variable explicative. Soit la façon dont les données sont organisées permet de purger l'estimation du caractère endogène de la variable explicative (sauf si il s'agit d'une erreur de mesure). Par exemple, dans le cas de la causalité inverse, il est possible de lager la variable dépendante. Soit cela n'est pas possible et il faut alors recourir à une méthode dite de 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 et elle doit contribuer à expliquer les variations de la variable explicative instrumentée.

C'est ensuite cette source de variation qui expliquera les variations de la variable dépendante. Par exemple, le jour de naissance d'un individu permet d'expliquer les variations dans le niveau d'éducation, cette source exogène de variation peut ensuite expliquer le revenu.

- Estimateur à variables instrumentales

L'estimateur à variables instrumentales présente de bonnes propriétés asymptotiques. Lorsque l'échantillon avec lequel on travaille est petit, l'estimateur est en réalité biaisé.

Supposons que le modèle estimé s'écrive : $y = \beta X + \epsilon$. On suppose que $cov(\epsilon_i, X_i) \neq 0$ 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 :

$plim \frac{1}{N} Z' \epsilon = 0$ implique que la convergence en probabilité de $Z' \epsilon$ soit nulle. Cette hypothèse est équivalente à $cov(Z_i, \epsilon_i) = 0$. Par ailleurs $plim \frac{1}{N} Z' X = Q_{zx}$ et $plim \frac{1}{N} Z' Z = Q_{zz}$ où Q_{zx} est une matrice finie de dimension (k, k) et Q_{zz} une matrice finie de dimension (k, k) et définie positive. La deuxième condition exprime le fait que la corrélation entre l'instrument et la variable instrumentée est non nulle. La troisième condition est nécessaire au calcul de la matrice de variances-covariances de l'estimateur des variables instrumentales.

L'estimateur à variables instrumentales de β a pour expression :

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

On montre facilement que cet estimateur est convergent. En effet :

$$\begin{aligned} plim \hat{\beta}_{IV} &= plim (Z' X)^{-1} Z' (X \beta + \epsilon) \\ &= \beta + plim (Z' X)^{-1} plim (Z' \epsilon) \\ &= \beta \end{aligned}$$

Une application du théorème central limite permet de montrer que la matrice de variances-covariances est donnée par :

$$V(\hat{\beta}) = \sigma^2 (Z' X)^{-1} Z' Z (X' Z)^{-1}$$

Cette matrice de variances-covariances peut être très grande si la corrélation entre Z et X est très faible.

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

Il arrive parfois que l'on dispose de plus d'instruments que de variables explicatives. Comment alors faut-il procéder ?

Il est clair que plus il y a d'instruments et plus l'information pour expliquer X est riche. Néanmoins, dans ce cas, l'estimateur tel qu'il est présenté ci-dessus ne peut plus être utilisé car le produit $Z'X$ est une matrice non carrée ne pouvant être inversée.

Une meilleure solution consiste à se servir de l'hypothèse selon laquelle $cov(Z_i, \epsilon_i) = 0$, ce qui implique que chaque colonne de Z n'est pas corrélée avec le terme d'erreur. Par conséquent, toute combinaison linéaire des colonnes de Z n'est pas non plus corrélée avec le terme d'erreur de la régression. Ceci suggère par conséquent de choisir autant de combinaisons linéaires de colonnes Z qu'il y a de colonnes de X afin de construire une matrice d'instrument compatible avec X.

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 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

Ainsi la matrice Z dispose des instruments exclus (ceux qui expliquent les variables endogènes) et les instruments inclus (les X qui ne sont pas endogènes et qui sont de fait leur propre instrument)

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.

Les programmes statistiques procèdent directement par le calcul de $\hat{\beta}_{IV}$. Si l'on souhaite effectuer les doubles MCO par cette méthode, 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 :

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

L'identification des instruments est en général le résultat d'un raisonnement théorique. En général, l'écriture du modèle est la suivante :

$$\begin{aligned}y &= \alpha + X_1' \beta_1 + \beta_2 x_2 + \epsilon \\x_2 &= \lambda z + X_1' \mu + u\end{aligned}$$

Avec $cov(u, \epsilon) \neq 0$, $cov(u, z) = 0$, $cov(z, \epsilon) = 0$

On peut transformer la première équation du modèle en reportant l'expression de x_2 obtenue à partir de la seconde équation dans la première pour obtenir la forme réduite :

$$\begin{aligned}y &= \alpha + X_1' \beta_1 + \beta_2(\lambda z + X_1' \mu + u) + \epsilon \\&= \alpha + X_1'(\beta_1 + \beta_2 \mu) + \beta_2 \lambda z + \beta_2 u + \epsilon \\&= \alpha + X_1' \pi_1 + \pi_2 z + \eta\end{aligned}$$

Cette équation permet de voir que les coefficients de x_2 ne peuvent pas toujours être interprétés simplement. Par exemple z est fonction de β_2 et λ qui peuvent être de signes opposés. C'est la raison pour laquelle le coefficient de x_2 , β_2 dans l'équation (1) peut s'obtenir en calculant le ration de l'estimateur du coefficient de z dans la régression en forme réduite, sur le coefficient de z dans la régression instrumentale :

$$\hat{\beta}_{IV} = \frac{\hat{\pi}_2}{\hat{\lambda}}$$

- 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.

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

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

Avec $cov(\epsilon, u)$ 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.

Cette équation permet également de voir pour quelle raison il est préférable de limiter le nombre d'instrument si ils sont faibles :

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

k est le nombre de variables de la regression instrumentale. Si on ajoute à cette équation des variables qui n'ont aucun pouvoir explicatif R^2 augmentera moins vite que k et aura un effet négatif sur F , ce qui augmentera le biais $E(\hat{\beta}_{IV} - \beta)$

- Tests sur les instruments

1/ Instruments faibles :

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. 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. Il faut faire attention parce que la valeur critique du test de Fisher croit lentement avec le nombre d'instrument. Il est donc conseillé de faire plusieurs régressions avec différents instruments pour choisir les meilleurs.

2/ 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 :

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

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

3/ 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).

5 Introduction à l'économétrie des séries temporelles

5.1 Les Processus Aléatoires Stationnaires et les Processus ARMA

5.1.1 Séries Temporelles

Une série temporelle (ou série chronologique) à temps discret est une suite réelle finie y_t , $t = 1, \dots, T$, où t représente le temps (en minute, jour, année...). Un des objectifs principaux de l'étude d'une série temporelle est la prévision des réalisations futures (prévoir l'évolution de la vente d'un produit pour ajuster au mieux les moyens de production, prévoir l'évolution d'un marché financier,...).

◇◇◇ Tendances et composantes saisonnières

On parle de tendance lorsque la série y_t , $t = 1, \dots, T$ peut s'écrire, à une erreur d'ajustement ϵ_t près, comme une combinaison linéaire de m fonctions du temps, choisies a priori (par exemple fonction puissance, exponentielle, logarithmique...) :

$$y_t = \sum_{j=1}^m \alpha_j f_j(t) + \epsilon_t \quad (20)$$

→ La tendance est linéaire lorsque $y_t = \alpha t + \beta + \epsilon_t$.

→ La tendance est polynomiale lorsque $y_t = \alpha_1 t^p + \alpha_2 t^{p-1} + \dots + \alpha_0 + \epsilon_t$

On parle de composante périodique lorsque la série y_t , $t = 1, \dots, T$ peut se décomposer en :

$$y_t = s_t + \epsilon_t \quad (21)$$

où s_t est périodique, c'est-à-dire $s_{t+T} = s_t$, avec T la période (supposée entière). Lorsque la période est de 6 mois ou 1 an, on parle généralement de composante saisonnière.

5.1.2 Indices descriptifs d'une série temporelle

◇◇◇ Moment d'ordre 1 : Indice de tendance centrale

Usuellement on utilise la moyenne empirique :

$$\bar{y} = \frac{1}{T} \sum_{t=1}^{t=T} y_t \quad (22)$$

◇◇◇ Moment d'ordre 2 : Indice de dispersion

On utilise la variance empirique :

$$\hat{\sigma}^2(0) = \frac{1}{T} \sum_{t=1}^{t=T} [y_t - \bar{y}]^2 \quad (23)$$

◇◇◇ Indices de dépendance

• L'auto-covariance empirique d'ordre 1 renseigne sur la dépendance entre deux données successives :

$$\hat{\sigma}^2(1) = \frac{1}{T-1} \sum_{t=1}^{t=T-1} (y_t - \bar{y})(y_{t+1} - \bar{y}) \quad (24)$$

• L'auto-covariance empirique d'ordre 2 renseigne sur la dépendance entre deux données écartées de deux pas de temps :

$$\hat{\sigma}^2(2) = \frac{1}{T-2} \sum_{t=1}^{t=T-2} (y_t - \bar{y})(y_{t+2} - \bar{y}) \quad (25)$$

• Les auto-corrélations empiriques sont les quotients des covariances empiriques et de la variance empirique

$$\hat{\rho}(h) = \frac{\hat{\sigma}(h)}{\hat{\sigma}(0)} \quad (26)$$

Ce sont les auto-corrélations empiriques que nous utiliserons pour caractériser la dépendance entre les variables.

5.1.3 Méthode économétrique et gestion de la tendance et des saisonnalités

Une série temporelle y_t , $t = 1, \dots, T$ est l'observation des T premières réalisations d'un processus stochastique Y_t . C'est ce processus que l'on cherche désormais à modéliser. Pour cela, la démarche suivante doit être adoptée :

- représenter graphiquement la série afin de repérer les tendances et saisonnalités,
- estimer et supprimer les tendances et saisonnalités (partie déterministe du processus stochastique),
- choisir un modèle pour les résidus (partie aléatoire du processus stochastique) et l'estimer,
- prédire les réalisations futures à l'aide de ce modèle.

◇◇◇ Définitions

• bruit blanc

Un processus de bruit blanc est une suite de variables aléatoires Y_t indépendantes, d'espérance et de variance constantes. Si l'espérance est nulle, le bruit blanc est centré, et si les variables aléatoires sont gaussiennes, le bruit blanc est gaussien.

• Processus stationnaire

Un processus aléatoire Y_t est stationnaire s'il est d'espérance constante :

$$E[Y_t] = \mu, \forall t$$

et si les covariances sont stables par translation dans le temps, c'est-à-dire, pour tout h :

$$Cov[Y_t, Y_{t+h}] = \sigma(h), \forall t$$

On appelle fonction d'auto-covariance du processus stationnaire la suite $\sigma(h)$, et fonction d'auto-corrélation du processus stationnaire la suite $\rho(h) = \sigma(h)/\sigma(0)$.

• Processus non stationnaire déterministe

On dit que le processus Y_t est caractérisé par une non stationnarité déterministe, ou encore que le processus Y_t est TS (Trend stationary) s'il peut s'écrire :

$$Y_t = f(t) + \epsilon_t$$

où $f(t)$ est une fonction qui dépend du temps et ϵ_t est un processus stationnaire. Ainsi, ce processus est rendu stationnaire en lui enlevant sa tendance déterministe :

$$Y_t - f(t) = \epsilon_t \quad \text{est stationnaire}$$

Le processus ϵ_t peut être modélisé par un processus $ARMA(p, q)$ stationnaire (le bruit blanc étant un cas particulier). $f(t)$ est une fonction déterministe, par exemple $f(t) = a + bt$ (cas le plus couramment retenu), mais on pourrait aussi considérer, entre autres, une tendance quadratique $f(t) = a + bt + ct^2$.

Une première conséquence économique d'un processus TS est qu'un choc imprévu (ϵ_t) n'a pas d'effet persistant sur le processus puisqu'il ne peut pas modifier sa partie tendancielle (sa croissance), qui est ici exogène. Il n'aura donc d'effet que sur la partie cyclique, supposée être stationnaire, donc son effet sera forcément temporaire. Une deuxième conséquence économique est que la décomposition tendance-cycle est naturelle dans ce cas : la tendance est donnée par $f(t)$ et le cycle par les écarts de la série à sa tendance, soit ϵ_t . Les deux composantes ne sont pas corrélées.

• Processus non stationnaire stochastique

On dit que le processus Y_t est caractérisé par une non stationnarité stochastique, ou encore que le processus Y_t est DS (Difference stationary) si le processus différencié une fois $\Delta Y_t = Y_t - Y_{t-1}$ est stationnaire. On parle aussi de processus intégré d'ordre 1, on note $Y_t \sim I(1)$:

$$\Delta Y_t = \epsilon_t$$

Les exemples les plus connus de processus $I(1)$ sont, d'une part, la marche aléatoire pure :

$$Y_t = Y_{t-1} + \epsilon_t$$

et, d'autre part, la marche aléatoire avec dérive :

$$Y_t = c + Y_{t-1} + \epsilon_t$$

Une première conséquence importante (d'avoir un processus DS) est qu'un choc imprévu (ϵ_t) à une date donnée influence la tendance et le futur du processus. Le processus est caractérisé par de la persistance des chocs ou de l'hystérèse. Autrement dit, un choc temporaire à une date donnée a un effet permanent sur le niveau du processus puisque le processus ne rejoindra jamais sa valeur initiale suite à ce choc. Une deuxième conséquence est que la décomposition tendance-cycle n'est plus explicite dans cette formulation.

◇◇◇ Estimation paramétrique de la tendance (trend)

Nous supposons que la série temporelle étudiée soit la réalisation d'un processus stochastique composé d'une tendance déterministe m_t et d'une partie aléatoire ϵ_t (supposée de moyenne nulle) :

$$Y_t = m_t + \epsilon_t$$

Une hypothèse que nous pouvons formuler sur la tendance déterministe est celle de linéarité :

$$m_t = a + bt$$

Pour estimer un trend linéaire, il suffit d'appliquer la méthode des MCO vu aux chapitres précédents.

◇◇◇ Estimation non paramétrique de la tendance : moyenne mobile

Supposons que la tendance m_t soit linéaire dans un petit intervalle $[t - q, t + q]$ autour de t . Dans ce cas, un bon estimateur de la tendance est la moyenne sur cet intervalle :

$$\hat{m}_t = \frac{1}{2q + 1} \sum_{k=-q}^{t=q} y_{t+k}$$

On peut donc estimer la tendance à chaque temps t en calculant la moyenne sur les observations étant dans une fenêtre de largeur $2q + 1$ autour de t : c'est ce que l'on appelle une estimation par moyenne mobile.

◇◇◇ Tendance et saisonnalité

Supposons désormais que le processus ne comporte pas uniquement une tendance, mais également une saisonnalité :

$$Y_t = m_t + s_t + \epsilon_t$$

avec s_t une fonction T -périodique. Dans ce cas, le principe d'estimation est (en simplifiant légèrement) le suivant : on estime la tendance moyenne sur une période, puis on estime la composante saisonnière en moyennant sur toutes les périodes les écarts à la tendance moyenne de la période.

◇◇◇ Elimination de la tendance et de la saisonnalité par la méthode des différences

Cette méthode permet de supprimer la tendance et la saisonnalité d'une série temporelle sans les estimer. Soit Δ_T l'opérateur qui associe $(Y_t - Y_{t-T})$ à Y_t :

$$\Delta_T Y_t = (Y_t - Y_{t-T})$$

On note Δ l'opérateur Δ_1 et Δ_T^k l'opérateur Δ_T^k qui est égal à $\Delta_T \circ \dots \circ \Delta_T$ (k fois).

En appliquant k fois Δ , on élimine la tendance. Il est important de remarquer que si l'on applique Δ_t quelque soit t , le résultat est le même quant à l'élimination de la tendance. Comme en pratique il n'est pas évident de connaître le degré k , on appliquera l'opérateur Δ jusqu'à ce que la moyenne du processus soit nulle (k sera généralement 1, 2 ou 3).

Soit un processus admettant une tendance m_t et une saisonnalité, de période T :

$$Y_t = m_t + s_t + \epsilon_t.$$

Dans ce cas,

$$\Delta_T Y_t = (m_t - m_{t-T}) + (\epsilon_t - \epsilon_{t-T})$$

est un processus désaisonnalisé. De plus, si la tendance du processus est linéaire, elle est également supprimée.

◇◇◇ Test sur la série résiduelle

L'objectif des techniques présentées dans le point précédent est d'obtenir une série stationnaire (ou tout au moins le plus stationnaire possible), et en particulier sans tendance ou saisonnalité. L'étape suivante consiste à modéliser la série résiduelle obtenue. La première chose à faire est donc de tester s'il y a dépendance entre les termes de cette série. Si ce

n'est pas le cas, on dit que la série résiduelle (stationnaire) est un bruit blanc. Si la série résiduelle obtenue après désaisonnalisation et élimination de la tendance, est un bruit blanc, il n'est donc pas utile d'aller plus loin dans la modélisation si ce n'est d'estimer la moyenne et variance du bruit blanc.

- Tester la présence d'un bruit blanc

1) Par l'étude de la fonction d'auto-corrélation empirique : Lorsque T est assez grand, les auto-corrélations d'un bruit blanc sont approximativement indépendantes et de loi $\mathcal{N}(0, 1)$. Ainsi, 95% des auto-corrélations devraient se trouver dans l'intervalle $[-1.96/\sqrt{n}, 1.96/\sqrt{n}]$, et en traçant les 40 premières auto-corrélations il ne devrait pas y en avoir plus de 2 voir 3 en dehors de ces limites.

2) A l'aide du test du portemanteau : Plutôt que de regarder si chaque auto-corrélation est dans les bornes de l'intervalle précédent, nous considérons la statistique définie par la somme des h premières auto-corrélations au carré :

$$Q = T \sum_{j=1}^h \hat{\rho}^2(j)$$

D'après la remarque précédente sur la normalité des auto-corrélations, la statistique Q suit une loi du Khi-2 à h degrés de liberté. Il est donc possible de construire un test qui consistera à rejeter l'hypothèse nulle (la série est un bruit blanc) si Q est supérieur au quantile du Khi-2 à h degrés de liberté. Ljung et Box (1978) ont amélioré ce test en considérant la statistique

$$Q_{LB} = T(T+2) \sum_{j=1}^h \frac{\hat{\rho}^2(j)}{T-j}$$

dont la distribution est mieux approximée que la précédente par une loi du Khi-2 à h degrés de liberté.

5.1.4 Les principaux modèles stationnaires

◇◇◇ Les processus auto-régressifs AR_p

Les premiers modèles que nous présentons sont les processus auto-régressifs, construits à partir de l'idée que l'observation au temps t s'explique linéairement par les observations précédentes.

On dit que Y_t est un processus auto-régressif d'ordre p (centré) s'il s'écrit :

$$Y_t = \epsilon_t + \sum_{j=1}^p a_j Y_{t-j}$$

où ϵ_t est un bruit blanc centré de variance σ^2 . L'observation Y_t au temps t est alors la somme d'un choc aléatoire à l'instant t , ϵ_t , indépendant de l'historique, et d'une fonction linéaire de son passé $\sum_{j=1}^p a_j Y_{t-j}$, qui peut être vue comme la prédiction de Y_t à partir des p dernières observations passées.

◇◇◇ Les processus en moyenne mobile MA_q

La seconde catégorie de modèles classiques regroupe les processus en moyenne mobile. On appelle moyenne mobile (Moving Average) d'ordre q un processus de la forme :

$$Y_t = \epsilon_t + b_1 \epsilon_{t-1} + \dots + b_q \epsilon_{t-q}$$

qui est équivalent à :

$$Y_t = \sum_{j=0}^q b_j \epsilon_{t-j}, \quad b_0 = 1$$

où les ϵ_j sont des bruits blancs centrés de variance σ^2 . Précisons deux éléments : (1) un processus moyenne mobile est nécessairement centré et (2) un processus auto-régressif est un processus moyenne mobile d'ordre infini, et réciproquement un processus moyenne mobile est un processus auto-régressif d'ordre infini.

◇◇◇ Les processus mixtes $ARMA_{p,q}$

Cette classe plus générale de modèles définit des processus sous la forme d'une récurrence auto-régressive avec un second membre de type moyenne mobile. Un processus auto-régressif moyenne mobile d'ordres p et q est de la forme :

$$Y_t = \sum_{k=1}^p a_k Y_{t-k} + \sum_{j=1}^q b_j \epsilon_{t-j}$$

où les ϵ_j sont des bruits blanc centrés de variance σ^2 .

◇◇◇ Choix des modèles

L'étude de la matrice des variances/covariances ainsi que des corrélations partielles peut conduire à identifier certaines hypothèses sur la nature du modèle. Une fois quelques modèles choisis, et leur paramètres estimés, des critères vont être utilisés pour choisir le modèle qui effectue le meilleur compromis entre :

- ajustement à la série de données,
- complexité du modèle : Il est en effet très important de prendre en compte ce compromis, car si on ne s'intéressait qu'à coller au mieux aux données, on serait tenté de choisir un modèle ARMA avec un très grand nombre de paramètres. Or, plus il y a de paramètres, plus il faut de données pour les estimer. Et donc pour un nombre d'observations fixé de la série, plus le modèle sera complexe, moins bien seront estimés les paramètres.

Les critères de choix de modèles les plus courants sont :

- le critère AIC (Akaike Information Criterion), qui sera généralement préféré si l'objectif de l'étude est de faire de la prévision.
- le critère BIC (Bayesian Information Criterion) sera quant à lui généralement préféré si l'objectif de l'étude est de s'ajuster à la série observée.

Les modèles ayant la plus petite valeur du critère devront être choisis.

5.2 Les tests de stationarité ou Unit root Tests

Les tests de racine unitaire testent si une série temporelle est non-stationnaire en utilisant un modèle autorégressif. La plupart des tests de racine unitaire suppose sous H_0 l'existence d'une racine unitaire.

◇◇◇ Le test de Dickey-Fuller

Un test de non stationnarité largement utilisé et répandu est le test de racine unitaire proposé par Dickey et Fuller en 1979. L'hypothèse nulle du test est la présence de racine unitaire, soit la non stationnarité de type stochastique. Le test consiste à tester :

$$H_0 : \phi = 1$$

$$H_1 : \phi < 1$$

dans le modèle

$$Y_t = \phi Y_{t-1} + \epsilon_t$$

avec ϵ_t bruit blanc $\sim \mathcal{N}(0, \sigma^2)$. L'hypothèse nulle correspond au cas de marche aléatoire pure (processus DS, $I(1)$) et l'hypothèse alternative correspond au cas d'un modèle AR(1) stationnaire. Pour mener ce test, on calcule la statistique de Student, mais attention, cette

statistique ne suit plus sous l'hypothèse nulle une loi de Student, puisque, sous l'hypothèse nulle, le processus est non stationnaire de type DS et les propriétés asymptotiques ne sont plus standards. Ainsi, la différence avec un test standard repose sur les valeurs critiques à utiliser pour conclure sur le test. On ne peut plus utiliser 1.96 comme valeur critique pour un test à 5%. Il faut utiliser les valeurs critiques, qui ont été retabulées par Dickey et Fuller.

Ce test ne répond pas aux attentes de détection du type de non stationnarité dans les variables économiques, d'une part parce que l'hypothèse de processus TS n'est pas présente et d'autre part parce que les séries économiques sont caractérisées par de l'autocorrélation, qui conduira la plupart du temps à rejeter l'hypothèse de bruit blanc dans le test ci-dessus. Pour prendre en compte, d'une part la présence d'autocorrélation dans les séries économiques, et, d'autre part, l'hypothèse de tendance déterministe, on préfère le test Dickey-Fuller augmenté.

◇◇◇ Le test de Dickey-Fuller augmenté

C'est une version généralisée du test Dickey-Fuller pour des modèles plus complexes de séries temporelles. Pour prendre en compte, d'une part la présence d'autocorrélation dans les séries économiques, et, d'autre part, l'hypothèse de tendance déterministe, on mène les tests de racine unitaire dans les trois régressions suivantes :

$$(1) \quad \Delta Y_t = \rho Y_t + \alpha + \beta t + \sum_{j=1}^p \phi_j \Delta Y_{t-j} + \epsilon_t$$

$$(2) \quad \Delta Y_t = \rho Y_t + \alpha + \sum_{j=1}^p \phi_j \Delta Y_{t-j} + \epsilon_t$$

$$(3) \quad \Delta Y_t = \rho Y_t + \sum_{j=1}^p \phi_j \Delta Y_{t-j} + \epsilon_t$$

avec p le nombre de retards à ajouter dans la régression afin de prendre en compte l'autocorrélation et donc de "blanchir" les résidus.

Le test ADF consiste alors à tester :

$$H_0 : \rho = 0$$

contre

$$H_1 : \rho < 0$$

dans les modèles (1), (2) et (3).

- Dans le modèle (1) :

ΔY_t est $I(0) + T$ (il a une tendance déterministe et l'écart à cette tendance déterministe suit un modèle $AR(p)$ stationnaire) sous H_0 , c'est-à-dire que Y_t est $I(1) + T^2$. Sous H_1 , Y_t a une tendance déterministe et l'écart à cette tendance déterministe suit un modèle AR stationnaire, on note $I(0) + T$, soit un processus TS.

- Dans le modèle (2) :

ΔY_t est $I(0) + C$ (il suit un modèle $AR(p)$ stationnaire non centré) sous H_0 , c'est-à-dire que Y_t est $I(1) + T$. Sous H_1 , Y_t suit un modèle AR stationnaire non centré, on note $I(0) + C$.

- Dans le modèle (3) :

ΔY_t est $I(0)$ (il suit un modèle $AR(p)$ stationnaire centré) sous H_0 , c'est-à-dire que Y_t est $I(1)$. Sous H_1 , Y_t suit un modèle AR stationnaire, on note $Y_t I(0)$.

Comme les valeurs critiques dépendent de la présence ou non d'une constante ou d'une tendance (ddl), cela implique que le test de racine unitaire doit être mené dans le "bon" modèle. Ainsi, une possibilité de mettre en oeuvre les tests de racine unitaire est de procéder de manière emboîtée, selon la stratégie suivante : on teste la racine unitaire dans le modèle le plus général, puis on teste si le modèle utilisé pour mener le test était pertinent. Si tel n'est pas le cas, on doit mener à nouveau le test de racine unitaire dans le modèle contraint, etc.

◇◇◇ Le test de Phillips-Perron

Phillips-Perron (1988) proposent une méthode non paramétrique pour corriger la présence d'autocorrélation, sans avoir à ajouter des endogènes retardées comme dans la méthode ADF (méthode plus robuste en cas d'erreurs MA notamment). La procédure de test consiste à tester l'hypothèse de racine unitaire $H_0 : \rho = 0$ dans les modèles suivants :

$$(1) \quad \Delta Y_t = \rho Y_t + \alpha + \beta t + \epsilon_t$$

$$(2) \quad \Delta Y_t = \rho Y_t + \alpha + \epsilon_t$$

$$(3) \quad \Delta Y_t = \rho Y_t + \epsilon_t$$

La statistique de test de Phillips-Perron (PP) est une statistique de student corrigée de la présence d'autocorrélation par la prise en compte d'une estimation de la variance de long terme de ϵ_t , robuste à la présence d'autocorrélation et d'hétéroscédasticité.

◇◇◇ Le test de KPSS (Kwiatkowski-Phillips-Schmidt-Shin)

Contrairement aux tests de racine unitaire précédents, KPSS proposent un test où l'hypothèse nulle est celle de la stationnarité contre l'hypothèse alternative d'une racine unitaire. Le test considère le modèle suivant :

$$\Delta Y_t = \rho Y_t + \alpha + \beta t + \epsilon_t$$

L'hypothèse nulle est donc $H_0 : \rho < 0$ contre l'hypothèse alternative $H_1 : \rho = 0$.

◇◇◇ Stratégie de test de racine unitaire

1) On choisit le nombre de retards p à introduire dans la régression : on peut, pour cela, choisir l'ordre p de l' $AR(p)$ pour la variable Y_t sur la base des autocorrélations partielles de Y_t , et sur la base de la significativité du dernier retard de l'AR introduit dans la régression, tout en vérifiant que le résidu est bien un bruit blanc.

2) On teste la racine unitaire $H_0 : \rho = 0$ dans le modèle le plus général (3) :

$$\Delta Y_t = \rho Y_t + \alpha + \beta t + \sum_{j=1}^p \phi_j \Delta Y_{t-j} + \epsilon_t$$

- Si on accepte H_0 (la racine unitaire), alors on va tester ensuite $H_0 : \rho = 0$ et $\beta = 0$. Si on accepte H_0 , alors on passe à l'étape 3. Si on rejette H_0 , on conclut que le processus est $I(1) + T^2$ (tout en sachant que cette conclusion est peu crédible économiquement, et qu'elle cache peut être le cas d'une tendance déterministe plus complexe que linéaire)

- Si on rejette H_0 (rejet de la racine unitaire), alors le processus est stationnaire, mais on doit aller tester la pertinence d'avoir tester la racine unitaire dans un modèle avec tendance en testant la significativité de la tendance (par un student normal) dans le modèle suivant :

$$\Delta Y_t = \rho Y_t + \alpha + \beta t + \sum_{j=1}^p \phi_j \Delta Y_{t-j} + \epsilon_t$$

Si on accepte $H_0 : \beta = 0$, alors il est recommandé de passer à l'étape 3. Si on rejette $H_0 : \beta = 0$, on conclut que le processus est $I(0) + T$.

3) On teste la racine unitaire $H_0 : \rho = 0$ dans le modèle (2) :

$$\Delta Y_t = \rho Y_t + \alpha + \sum_{j=1}^p \phi_j \Delta Y_{t-j} + \epsilon_t$$

6 Introduction à l'économétrie des données de panel

TBA

7 Introduction à l'économétrie des variables qualitatives

TBA

8 Annexe-Rappels

Notons $A_{(m,n)}$ une matrice à m lignes et n colonnes composée d'éléments a_{ij}

Notons $A'_{(n,m)}$ la matrice transposée de $A_{m,n}$

- $(AB)' = B' A'$
- $(A^{-1})' = (A')^{-1}$ si A est inversible
- $(AB)^{-1} = B^{-1} A^{-1}$ si A et B sont inversibles
- $tr(A) = \sum_{i=1}^n a_{ii}$ i.e. somme des valeurs diagonales de la matrice
- $tr(\lambda A) = \lambda tr(A)$
- Si $A^2 = A$ alors la matrice est dit idempotente

Soit X une variable aléatoire défini sur \mathbb{R}

- $E(X) = \int_{-\infty}^{+\infty} x f(x) dx$ où $f(x)$ représente la densité de probabilité de X
- $E(g(x)) = \int_{-\infty}^{+\infty} g(x) f(x) dx$
- $E(x^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx$
- $f(x, y) = f(x) f(y)$ ssi x, y sont deux variables aléatoires indépendantes
- $E(aX + b) = aE(X) + b$
- $E(X + Y) = E(X) + E(Y)$
- $E(XY) = E(X)E(Y)$ ssi X, Y sont deux variables aléatoires indépendantes
- $V(X) = E[(X - E(X))^2] = E(X^2) - E(X)^2$
- $V(aX + b) = a^2 V(X)$

Si X et Y sont deux variables aléatoires indépendantes alors on a :

$$\Rightarrow V(X + Y) = V(X) + V(Y)$$

$$\Rightarrow V(aX + bY) = a^2V(X) + b^2V(Y)$$

$$\Rightarrow V(X - Y) = V(X) + V(Y)$$

$$\Rightarrow Cov(X, Y) = 0$$

Si X et Y sont deux variables aléatoires quelconques :

$$\Rightarrow V(X + Y) = V(X) + V(Y) + 2cov(X, Y)$$

$$\Rightarrow Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$$

$$\Rightarrow V(\sum_{i=1}^n X_i) = \sum_{i=1}^n V(X_i) + 2 \sum_i \sum_j cov(X_i, X_j)$$

Le coefficient de corrélation entre deux variables est défini par :

$$\rho = \frac{Cov(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Les Lois statistiques usuelles utilisées en économétrie :

1) La loi normale

Si $X \sim N(\mu, \sigma^2)$, alors la fonction de densité de X est donnée par :

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(X - \mu)^2\right)$$

et on peut écrire que d'après le théorème de la limite centrale :

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

Soit $X_1 \sim N(\mu_1, \sigma_1^2)$ et $X_2 \sim N(\mu_2, \sigma_2^2)$, alors

- si X_1, X_2 non indépendantes, on a $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\sigma_{12})$
- si X_1, X_2 indépendantes, on a $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

2) La loi Khi-deux

Soit une variable aléatoire $X \sim N(0, 1)$ alors

$$Z = X^2 \sim \chi_1^2$$

Soit N variables aléatoires $X_1 \sim N(0, 1), \dots, X_n \sim N(0, 1)$ indépendantes alors

$$Z = \sum_{i=1}^N X_i^2 \sim \chi_n^2$$

2) La loi de Student

Soit une variable aléatoire $X \sim N(0, 1)$ et une variable aléatoire $Y \sim \chi_n^2$ alors

$$Z = \frac{X}{\sqrt{Y/n}} \sim t_n$$

2) La loi de Fisher

Soit deux variables aléatoires $X_1 \sim \chi_p^2$ et $X_2 \sim \chi_n^2$ indépendantes alors

$$Z = \frac{X_1/p}{X_2/q} \sim F(p, q)$$