

Econométrie et applications

Olivier Donni

Octobre 2014

Contents

I	Préliminaires	1
1	Qu'est-ce que l'économétrie?	2
1.1	Définition	2
1.2	Les étapes d'une analyses empirique	2
1.3	Structure des données	4
1.4	La causalité en économétrie	5
2	Quelques rappels de statistique	6
2.1	Variables aléatoires	6
2.2	Distributions jointes, conditionnelles, et marginales	7
2.3	Espérance, variance, et covariance	8
2.4	Espérance et variance conditionnelle	11
2.5	La loi normale et les distributions dérivées	13
2.6	Population, paramètres et estimateurs	14
2.7	Propriétés asymptotiques des estimateurs	17
2.8	Estimation: la méthode des moments	18
3	Le modèle de régression simple	20
3.1	Définition	20
3.2	Dérivation des estimateurs des MCO	22
3.3	Propriétés algébriques des MCO	25
3.4	Propriétés statistiques des MCO	28
3.5	Le modèle sans constante	35

II	Le modèle de régression multiple	37
4	Définition, calcul et interprétation	38
4.1	Définition	38
4.2	Mécanique et interprétation des MCO	40
4.3	Propriétés algébriques des MCO	41
4.4	Une formule utile	43
5	La distribution des estimateurs	46
5.1	L'espérance des estimateurs des MCO	46
5.2	La variance des estimateurs des MCO	52
5.3	L'efficacité des estimateurs des MCO	55
5.4	La distribution des estimateurs des MCO	55
6	Inférence statistique	58
6.1	Test d'une seule restriction: le t-test	58
6.2	Tests d'une multiplicité de restrictions: le F-test	66
6.3	Prédiction et analyse de résidus	69
7	Asymptotique et hétéroscédasticité	72
7.1	Convergence	72
7.2	Normalité asymptotique	75
7.3	Tests de l'hétéroscédasticité	76
7.4	La méthode des MC Pondérés	78
8	Echantillon et spécification	83
8.1	Le choix de la forme fonctionnelle	83
8.2	Les variables qualitatives	87
8.3	Test de Chow	90
8.4	Le modèle de probabilité linéaire	91
8.5	Le R carré ajusté et la sélection de modèles	93
8.6	Problèmes de spécification	94
A	Formules usuelles	96

<i>CONTENTS</i>	1
B Démonstration.	98

Part I

Préliminaires

Chapter 1

Qu'est-ce que l'économétrie?

1.1 Définition

L'économétrie est un ensemble de méthodes statistiques développées afin d'étudier des questions économiques qui permet de tester des théories économiques, de prédire des comportement économiques et d'évaluer des politiques économiques. En tant que branche des statistiques, ses principales caractéristiques sont doubles:

1. Les données utilisées ne sont généralement pas expérimentales; elles sont souvent collectées à l'aide d'enquêtes; on parle de données observationnelles;
2. Les modèles estimés sont structurels (et non descriptifs); ils représentent généralement des relations causales entre des variables économiques.

Par ailleurs, les économistes ont quelques fois mis au point des techniques originales pour modéliser la spécificité des variables économiques.

1.2 Les étapes d'une analyses empirique

Dans une première étape, un modèle économique est construit (ou emprunté à la littérature existante) sous une forme très générale; il associe une variable

expliquée, désignée souvent par la lettre y , et une ou plusieurs variables explicatives, désignée souvent par les lettres x_1, \dots, x_K , où K est le nombre de variables explicatives. Dans une seconde étape, un modèle économétrique est construit en définissant précisément les variables, en choisissant une forme fonctionnelle (c'est-à-dire une relation paramétrique entre les variables), et en introduisant un terme d'erreur qui représentera toutes les variables pertinentes mais omises du modèle. Considérons les exemples suivants.

1. **Le modèle d'investissement en capital humain.** Selon ce modèle, le salaire est une fonction du niveau d'éducation, de l'expérience et de l'ancienneté, qui représentent les différentes formes de capital humain d'un travailleur,

$$\text{sal} = f(\text{edu}, \text{exp}, \text{anc}),$$

où sal est le salaire horaire du travailleur, edu , exp et anc l'éducation, l'expérience et l'ancienneté du travailleur mesurées en années, et où les dérivées de $f(\cdot)$ sont positives. Un exemple de modèle économétrique est le suivant:

$$\text{sal} = \beta_0 + \beta_1 \text{edu} + \beta_2 \text{exp} + \beta_3 \text{anc} + u,$$

où u est un terme d'erreur. Les paramètres β_0 , β_1 , β_2 et β_3 sont estimés à l'aide de données sur les variables sal , edu , exp et anc .

2. **Le modèle du consommateur.** Le consommateur choisit le panier de biens qui maximise son utilité parmi l'ensemble de biens qu'il peut acheter sachant le niveau des prix et de son revenu. La quantité demandée d'un bien quelconque est une fonction de l'ensemble des prix et du revenu du ménage,

$$\text{qbien1} = f(\text{pbien1}, \text{pbien2}, \dots, \text{pbienK}, \text{rev})$$

où qbien1 est la quantité d'un bien particulier (par exemple, le café) achetée par le consommateur, pbien1 , pbien2 , \dots et pbienK le prix de

biens 1 à K (qui sont substitués ou compléments du bien acheté et qui incluent le prix de ce bien), rev le revenu du consommateur, et $f(\cdot)$ une fonction ayant certaines propriétés générales (négativité et homogénéité).

1.3 Structure des données

Les données sont la matière première de l'économètre. Plusieurs types de données existent qui doivent être abordées de manière différente. On distingue, entre autres,

1. **Des données transversales (cross-sectional data).** Les données sont collectées pour un ensemble d'unités (individus, firmes, communes, pays, par exemple) à un moment précis du temps. Il s'agit souvent, mais pas toujours, de données microéconomiques; en général, elles constituent un échantillon aléatoire issu d'une certaine population. Un échantillon aléatoire est un ensemble d'observations indépendantes les unes des autres et issues d'une même population.
2. **Des séries temporelles (time series data).** Les données sont collectées pour une même unité (un individu ou un pays, par exemple) au cours du temps. Il peut s'agir de données collectées tous les jours, toutes les semaines, tous les mois, tous les ans; ces données sont souvent macroéconomiques (comme le taux de chômage, le taux d'inflation, le produit intérieur brut) et ne constituent pas, en général, un échantillon aléatoire; cela peut rendre leur analyse nettement plus compliquée.

Un panel est une base de données qui mélange les caractéristiques des données transversales et des séries temporelles.

1.4 La causalité en économétrie

L'un des principaux objectifs d'une étude empirique est souvent de mesurer une relation "causale" entre deux variables. Par exemple, l'économètre s'attachera à mesurer l'effet de l'éducation sur le salaire des travailleurs. En fait, on dira qu'une variable x a un effet causal sur une autre variable y si une variation de x – toutes autres choses étant égales (*ceteris paribus*) – implique (en moyenne) un changement de y . Lorsque les données ne sont pas expérimentales, l'estimation d'effet causal peut s'avérer délicate.

Pour illustrer cela, considérons le cas d'un agronome qui désire mesurer l'effet d'un engrais sur la productivité d'une terre. Pour cela, il divisera un terrain en un certain nombre de parcelles. Sur certaines parcelles, choisies au hasard, il mettra de l'engrais. Sur les autres, il ne mettra rien. Si, en moyenne, les parcelles traitées ont un rendement plus élevé que les autres parcelles, il conclura que l'engrais a un effet positif.

Pour un économiste, un problème assez similaire consiste à mesurer le rendement de l'éducation. S'il le pouvait, l'économiste effectuerait une expérience assez similaire à celle de l'agronome. Il choisirait des individus au hasard, auxquels il donnerait un certain niveau d'éducation. Les autres resteraient non éduqués. Il suffirait alors de comparer le niveau de salaire des uns et des autres pour connaître le rendement de l'éducation.

Même si les données non expérimentales ne permettent de contrôler les variables explicatives, les techniques économétriques simulent un effet *ceteris paribus*.

Chapter 2

Quelques rappels de statistique

2.1 Variables aléatoires

Une variable aléatoire est une variable qui prend une valeur numérique et dont le résultat est déterminé par une expérience aléatoire. Une variable aléatoire est désignée par une lettre en majuscule X , Y ou Z ; et une réalisation particulière par une lettre en minuscule x , y ou z .

Variable aléatoire discrète

Une variable aléatoire discrète est une variable qui peut prendre seulement un nombre fini ou dénombrable de valeurs, et dont la distribution est décrite par un ensemble de probabilités. Si X peut prendre les valeurs x_1, \dots, x_K , alors

$$\Pr(X = x_k) = p_k,$$

où

$$0 \leq p_k \leq 1 \quad \text{et} \quad p_1 + p_2 + \dots + p_K = 1.$$

La fonction cumulative de probabilité, qui est définie par

$$F(x_k) = \Pr(X \leq x_k),$$

résume l'information concernant les valeurs possibles que peut prendre la variable X et les probabilités correspondantes.

Variable aléatoire continue

Une variable X est une variable aléatoire continue si elle peut prendre n'importe quelle valeur réelle avec une probabilité égale à zéro. La fonction cumulative de probabilité est définie par

$$F(x) = \Pr(X \leq x).$$

Sous certaines conditions de régularité, il est possible d'écrire:

$$F(x) = \int_{-\infty}^x f(t) \cdot dt,$$

où $f(x)$ est la fonction de densité de probabilité, avec

$$\frac{\partial}{\partial x} F(x) = f(x).$$

2.2 Distributions jointes, conditionnelles, et marginales

Soit X et Y une paire de variables aléatoires.

Dans le cas discret, les variables (X, Y) ont une distribution jointe représentée par un ensemble de probabilités,

$$\Pr(X = x_k, Y = y_j) = p_{jk},$$

et une fonction cumulative de probabilité jointe:

$$F(x_k, y_j) = \Pr(X \leq x_k, Y \leq y_j) = \sum_{k'=1}^k \sum_{j'=1}^j p_{j'k'}.$$

La probabilité conditionnelle d'obtenir y_j sachant que x_k a été obtenu est donnée par

$$p_{j|k} = \frac{p_{j,k}}{p_k}.$$

On dit que X et Y sont indépendantes si et seulement si

$$p_{j,k} = p_k \cdot p_j,$$

pour tout x_k et tout y_j .

Dans le cas continu, la distribution jointe est représentée par une fonction cumulative de probabilité telle que:

$$F(x, y) = \Pr(X \leq x, Y \leq y),$$

et, sous certaines conditions de régularité, une fonction de densité de probabilité jointe définie de manière implicite de la manière suivante:

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(x, y) \cdot dx dy.$$

On dit que X et Y sont indépendantes si et seulement si

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y),$$

pour tout x et tout y , où les distributions $f_X(x)$ et $f_Y(y)$ sont appelées fonctions de distribution de densité marginales de X et de Y . En économétrie, on est souvent intéressé à la manière dont une variable varie lorsque la valeur d'autres variables se modifie. Cette information est contenue dans la fonction de densité de probabilité conditionnelle, définie par

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

pour toutes les valeurs de X telles que $f_X(x) > 0$.

2.3 Espérance, variance, et covariance

Espérance

L'espérance, ou moyenne de la population, est une mesure de la tendance centrale de la distribution, désignée par $E(X)$ ou μ_X ou μ . Dans le cas discret, l'espérance de X est égale à

$$E(X) = x_1 p_1 + x_2 p_2 + \dots + x_K p_K = \sum_{k=1}^K x_k p_k,$$

et dans le cas continu, à

$$E(X) = \int_{-\infty}^{+\infty} xf(x) \cdot dx,$$

où cette intégrale est supposée exister.

L'opérateur espérance a quelques propriétés simples.

Propriété 2.1:

Pour toute constante c ,

$$E(c) = c.$$

L'espérance d'une constante est égale à cette constante.

Propriété 2.2: *Pour toutes constantes a et b ,*

$$E(aX + b) = aE(X) + b.$$

En d'autres termes, l'espérance est un opérateur linéaire. Donc, si $E(X) = \mu \neq 0$, on peut définir $Y = X - \mu$ de telle manière que $E(Y) = 0$.

Propriété 2.3: *Si a_1, \dots, a_K sont des constantes, alors*

$$E(a_1X_1 + \dots + a_KX_K) = a_1E(X_1) + \dots + a_KE(X_K).$$

En d'autres termes, l'espérance d'une somme de variables aléatoires est égale à la somme des espérances de ces variables aléatoires.

Variance

Pour une variable aléatoire, définissons $\mu = E(X)$. La variance (de population) est une mesure de la dispersion des variables aléatoires dans une distribution, définie par

$$\text{Var}(X) = E \left[(X - \mu)^2 \right] \geq 0.$$

La variance est aussi désignée par σ_X^2 . De plus,

$$\begin{aligned}\text{Var}(X) &= \text{E}(X^2 + \mu^2 - 2X\mu) \\ &= \text{E}(X^2) + \mu^2 - 2\mu^2 = \text{E}(X^2) - \mu^2.\end{aligned}$$

L'opérateur variance a aussi certaines propriétés.

Propriété 2.4: *Pour toute constante c ,*

$$\text{Var}(c) = 0.$$

Propriété 2.5: *Pour toutes constantes a et b ,*

$$\text{Var}(aX + b) = a^2 \text{Var}(X).$$

L'écart-type d'une variable aléatoire est défini par

$$\text{sd}(X) = \sqrt{\text{Var}(X)}.$$

Covariance

Pour une paire de variables aléatoires, définissons $\mu_X = \text{E}(X)$ et $\mu_Y = \text{E}(Y)$.

La covariance est une mesure de la dépendance linéaire entre deux variables aléatoires, définie par

$$\text{Cov}(X, Y) = \text{E}[(X - \mu_X)(Y - \mu_Y)].$$

Elle est aussi représentée par σ_{XY} . Si $\sigma_{XY} > 0$, alors les variables X et Y varient dans la même direction, et si $\sigma_{XY} < 0$, alors elles varient dans des directions opposées. La covariance peut s'écrire de manière alternative:

$$\begin{aligned}\text{Cov}(X, Y) &= \text{E}[(X - \mu_X)(Y - \mu_Y)] \\ &= \text{E}[X(Y - \mu_Y)] \\ &= \text{E}[(X - \mu_X)Y] \\ &= \text{E}(YX) - \mu_Y\mu_X.\end{aligned}$$

Donc, si $\mu_X = 0$ ou $\mu_Y = 0$, alors $\text{Cov}(X, Y) = \text{E}(YX)$.

L'opérateur covariance a des propriétés simples.

Propriété 2.6: Si X et Y sont indépendantes, alors,

$$\text{Cov}(X, Y) = 0.$$

Il est important de noter que l'inverse de cette propriété n'est pas vraie.

Propriété 2.7: Pour toutes constantes a_1, a_2, b_1, b_2 ,

$$\text{Cov}(a_1X + b_1, a_2Y + b_2) = a_1a_2\text{Cov}(X, Y).$$

Le coefficient de corrélation est défini de la manière suivante:

$$\text{Corr}(X, Y) \equiv \frac{\text{Cov}(X, Y)}{\text{sd}(X)\text{sd}(Y)} = \frac{\sigma_{XY}}{\sigma_Y\sigma_X}.$$

Il s'agit d'une mesure normalisée de la dépendance linéaire qui est toujours comprise entre -1 et $+1$.

Une fois que la covariance est définie, il est possible de définir la variance de la somme de deux variables.

Propriété 2.8: Pour toutes constantes a et b ,

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y).$$

Donc, si X et Y sont non corrélées,

$$\text{Var}(X + Y) = \text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y).$$

2.4 Espérance et variance conditionnelle

La notion de distribution conditionnelle a déjà été introduite; elle permet de représenter l'effet d'une variable X sur une autre variable Y . L'espérance conditionnelle de Y pour X , désignée par $E(Y|X = x)$ ou $E(Y|x)$, est définie par

$$E(Y|X = x_k) = \sum_{j=1}^K y_j p_{j|k}$$

dans le cas discret, et par

$$E(Y|X = x) = \int_{-\infty}^{+\infty} y f_{Y|X}(y|x) \cdot dy,$$

dans le cas continu. Les propriétés de l'espérance conditionnelle sont les suivantes.

Propriété 2.9: Pour toute fonction $c(X)$,

$$E[c(X)|X] = c(X).$$

Propriété 2.10. Pour toutes fonctions $a(X)$ et $b(X)$,

$$E[a(X)Y + b(X)|X] = a(X)E[Y|X] + b(X).$$

Propriété 2.11. Si X et Y sont indépendantes,

$$E(Y|X) = E(Y).$$

Propriété 2.12:

$$E[E(Y|X)] = E(Y).$$

En conséquence, si $E(Y|X) = 0$, alors $E(Y) = 0$.

Propriété 2.13: Si $E(Y|X) = 0$, alors $\text{Cov}(Y, X) = 0$.

La variance conditionnelle de Y pour X donnée est égale à

$$\text{Var}(Y|X = x) = E\left[\left(Y - \mu_{Y|X}\right)^2 \middle| X = x\right]$$

Propriété 2.14: Pour toute fonction $c(X)$,

$$\text{Var}[c(X)|X] = 0.$$

Propriété 2.15: Pour toutes fonctions $a(X)$ et $b(X)$,

$$\text{Var}[a(X)Y + b(X)|X] = [a(X)]^2 \text{Var}(Y|X).$$

2.5 La loi normale et les distributions dérivées

La loi normale

Une variable aléatoire normale est une variable aléatoire continue pouvant prendre n'importe quelle valeur de $-\infty$ à $+\infty$, dont la fonction de densité de probabilité est:

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

où $\mu = E(X)$ et $\sigma^2 = \text{Var}(X)$. On dit que X a une distribution normale d'espérance μ et de variance σ^2 , et on écrit:

$$X \sim \mathcal{N}(\mu, \sigma^2).$$

La distribution normale est aussi appelée distribution gaussienne. La distribution normale standardisée est obtenue quand $\mu = 0$ et $\sigma^2 = 1$. Sa fonction de densité de probabilité est

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

La distribution normale a les propriétés suivantes.

Propriété 2.16: Si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors

$$\left(\frac{X - \mu}{\sigma}\right) \sim \mathcal{N}(0, 1).$$

Propriété 2.17: Si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors

$$(aX + b) \sim \mathcal{N}(\mu + b, a^2\sigma^2).$$

Propriété 2.18: Soit $\{X_k : k = 1, \dots, K\}$ un ensemble de variables aléatoires normales indépendantes d'espérance μ_k et de variance σ_k^2 et $\{a_k : k = 1, \dots, K\}$ un ensemble de constantes, alors

$$\sum_{k=1}^K a_k X_k \sim \mathcal{N}\left(\sum_{k=1}^K \mu_k, \sum_{k=1}^K a_k^2 \sigma_k^2\right).$$

En d'autres termes, toute combinaison linéaire de variables aléatoires normales indépendantes est distribuée selon une loi normale.

Les distributions dérivées de la loi normale

Soit $\{X_k : k = 1, \dots, K\}$ un ensemble de variables aléatoires distribuées selon la loi normale standardisée, indépendantes les unes des autres. La variable aléatoire Z définie de la manière suivante:

$$Z \equiv \sum_{k=1}^K X_k^2$$

est distribuée selon une loi du Chi-carré à K degrés de liberté ou

$$Z \sim \chi^2(K).$$

Soit X une variable aléatoire distribuée selon une loi normale standardisée et Z une variable aléatoire distribuée selon une loi du Chi-carré à K degrés de liberté. La variable aléatoire T définie de la manière suivante:

$$T \equiv \frac{X}{\sqrt{Z/K}}.$$

est distribuée selon une loi de Student à K degrés de liberté ou

$$T \sim \mathcal{T}(K).$$

Soit Z_1 et Z_2 deux variables aléatoires indépendantes distribuées selon des lois de Fisher avec K_1 et K_2 degrés de liberté, respectivement. La variable aléatoire F définie de la manière suivante:

$$F \equiv \frac{Z_1/K_1}{Z_2/K_2}$$

est distribuée selon une loi de Fisher à (K_1, K_2) degrés de liberté ou

$$F \sim \mathcal{F}(K_1, K_2).$$

2.6 Population, paramètres et estimateurs

Un des objectifs de la statistique est de dégager, à partir d'un échantillon, des résultats qui s'appliqueront à une population inconnue. On parle d'**inférence statistique**. Un autre objectif est de partir de la population connue et de dériver les propriétés de l'échantillon.

Soit Y une variable aléatoire issue d'une population avec une fonction de densité de probabilité $f(y; \theta)$, qui dépend d'un seul paramètre θ . La fonction de densité de probabilité est supposée de Y connue à l'exception de la valeur de θ .

Définition 2.1: *Un échantillon aléatoire $\{Y_n : n = 1, \dots, N\}$ de la population représentée par $f(y; \theta)$ est un ensemble de variables aléatoires indépendantes et de distribution $f(y; \theta)$.*

Etant donné un échantillon aléatoire $\{Y_n : n = 1, \dots, N\}$ issus d'une distribution dépendant d'un paramètre inconnu θ , un estimateur est une règle qui assigne à chaque réalisation possible de l'échantillon une valeur $\hat{\theta}$.

Un estimateur est une variable aléatoire dont la distribution est appelée distribution d'échantillonnage. Pour avoir une utilité pratique, cette distribution doit avoir certaines propriétés.

Définition 2.2: *Un estimateur $\hat{\theta}$ est non biaisé si*

$$E(\hat{\theta}) = \theta,$$

pour toutes les valeurs de θ .

Par exemple, la moyenne d'échantillon

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

est un estimateur non biaisé de la moyenne de population μ_X (ou espérance).

En effet,

$$\begin{aligned}
 E(\bar{X}) &= E\left(\frac{1}{N} \sum_{i=1}^N X_i\right) \\
 &= \frac{1}{N} E\left(\sum_{i=1}^N X_i\right) \\
 &= \frac{1}{N} \sum_{i=1}^N E(X_i) \\
 &= \frac{1}{N} \sum_{i=1}^N \mu_X = \mu_X
 \end{aligned}$$

La variance d'échantillon

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2$$

est un estimateur non biaisé de la variance de population σ_X (la démonstration est plus compliquée et omise).

Un estimateur a une distribution d'échantillonnage et, en particulier, une variance d'échantillonnage. Par exemple, la variance de la moyenne d'échantillon est donnée par

$$\begin{aligned}
 \text{Var}(\bar{X}) &= \text{Var}\left(\frac{1}{N} \sum_{i=1}^N X_i\right) \\
 &= \frac{1}{N^2} \text{Var}\left(\sum_{i=1}^N X_i\right) \\
 &= \frac{1}{N^2} \sum_{i=1}^N \text{Var}(X_i) \\
 &= \frac{1}{N} \text{Var}(X_i) = \frac{\sigma_X^2}{N}.
 \end{aligned}$$

Définition 2.3: Soit $\hat{\theta}_1$ and $\hat{\theta}_2$ deux estimateurs non biaisés de θ . Alors $\hat{\theta}_1$ est efficace relativement à $\hat{\theta}_2$ pour tout θ si

$$\text{Var}(\hat{\theta}_1) \leq \text{Var}(\hat{\theta}_2),$$

avec une inégalité stricte pour au moins une valeur de θ .

2.7 Propriétés asymptotiques des estimateurs

Les propriétés asymptotiques d'un estimateur concernent le comportement de cet estimateur lorsque la taille de l'échantillon tend vers l'infini.

Définition 2.4: Soit $\hat{\theta}_N$ un estimateur of θ basé sur un échantillon $(Y_n : n = 1, \dots, N)$ de taille N . Alors $\hat{\theta}_N$ est un estimateur convergent si pour tout ε

$$\lim_{N \rightarrow \infty} \Pr \left(\left| \hat{\theta}_N - \theta \right| > \varepsilon \right) = 0.$$

On écrit aussi

$$\text{plim } \hat{\theta}_N = \theta,$$

ou la probabilité limite de $\hat{\theta}_N$ est θ . Par exemple, la moyenne d'échantillon \bar{X} calculée pour un échantillon aléatoire est un estimateur convergent. En effet, $E(\bar{X}) = \mu$ et $\text{Var}(\bar{X}) = \sigma^2/N$ pour n'importe quel échantillon de taille N . Donc,

$$\lim_{N \rightarrow \infty} \text{Var}(\bar{X}) = 0.$$

Plus généralement, le résultat suivant a un rôle très important dans la dérivation des propriétés des estimateurs.

Loi des grands nombres. Soit Y_1, Y_2, \dots, Y_N une séquence de variables aléatoires indépendantes et identiquement distribuées d'espérance μ . Alors,

$$\text{plim } \bar{Y}_N = \mu.$$

L'opérateur plim a certaines propriétés.

Propriété 2.19: Soit θ un paramètre et définissons: $\gamma = g(\theta)$, pour une certaine fonction $g(\cdot)$. Supposons que $\text{plim } \hat{\theta}_N = \theta$. Définissons un estimateur $\hat{\gamma}_N = g(\hat{\theta}_N)$. Alors,

$$\text{plim } \hat{\gamma}_N = \gamma.$$

Propriété 2.20: Si $\text{plim}(T_N) = \alpha$ et $\text{plim}(U_N) = \beta$, alors

1. $\text{plim}(T_N + U_N) = \alpha + \beta$;
2. $\text{plim}(T_N U_N) = \alpha\beta$;
3. $\text{plim}(T_N/U_N) = \alpha/\beta$ sous la condition que $\beta \neq 0$.

Sous certaines conditions, la distribution des estimateurs peut converger vers la distribution normale.

Définition 2.5: Soit Y_1, Y_2, \dots, Y_N une séquence de variables aléatoires telle que pour tout y ,

$$\lim_{N \rightarrow \infty} \Pr(Y_N \leq y) = \Phi(y).$$

Alors Y_N est dit asymptotiquement normal.

Le théorème suivant détermine les conditions sous lesquelles un estimateur peut être asymptotiquement normal.

Théorème de la limite centrale. Soit Y_1, Y_2, \dots, Y_N une séquence de variables aléatoires indépendantes et identiquement distribuées d'espérance μ et de variance σ^2 . Alors,

$$Z_N = \frac{\bar{Y}_N - \mu}{\sigma/\sqrt{N}}$$

est asymptotiquement normal.

2.8 Estimation: la méthode des moments

La méthode des moments est une méthode permettant d'estimer les paramètres dans un cadre très général; elle procède de la manière suivante.

Soit $\{Y_n : n = 1, \dots, N\}$ un échantillon aléatoire de la population représentée par $f(y, \theta)$. En général, les moments théoriques de la distribution de

Y sont des fonctions connues de θ . La méthode des moments consiste à égaliser certains moments théoriques et leurs contreparties empiriques calculées à partir de l'échantillon, et à résoudre les équations qui en résultent.

Par exemple, supposons que le paramètre à estimer est l'espérance de la distribution $\theta = \mu$ et considérons le moment théorique $E(Y) = \theta$. Puisque le moment empirique correspondant est $m = \frac{1}{N} \sum_{n=1}^N Y_n$, on obtient la condition suivante:

$$\hat{\theta} = \frac{1}{N} \sum_{n=1}^N Y_n.$$

Il s'agit de l'estimateur par la méthode des moments de l'espérance d'une distribution.

Chapter 3

Le modèle de régression simple

3.1 Définition

Le modèle de régression (linéaire) simple permet d'étudier l'effet d'une variable x sur une autre variable y . Il s'écrit de la manière suivante:

$$y = \beta_0 + \beta_1 x + u$$

où β_0 est la constante, β_1 est la pente et u le terme d'erreur. On parle de la régression de y sur x , où les variables y et x sont appelées:

$$y = \left\{ \begin{array}{l} \text{variable dépendante} \\ \text{variable expliquée} \\ \text{variable de réponse ,} \\ \text{variable prédite} \\ \text{régressant} \end{array} \right.$$

et

$$x = \left\{ \begin{array}{l} \text{variable indépendante} \\ \text{variable explicative} \\ \text{variable de contrôle} \\ \text{variable prédictrice} \\ \text{régresseur} \end{array} \right. .$$

On considère les exemples suivants.

Exemple 3.1: La relation entre le rendement de parcelles de terre et la quantité d'engrais utilisée s'écrit:

$$\text{rend} = \beta_0 + \beta_1 \cdot \text{engr} + u.$$

Ce modèle est linéaire car l'effet "ceteris paribus" de la variable x (engrais) sur y (rendement) est linéaire:

$$\Delta y = \beta_1 \cdot \Delta x \quad \text{si} \quad \Delta u = 0.$$

Estimer correctement le paramètre β_1 est souvent plus important d'un point de vue économique que d'estimer β_0 . Pour les raisons expliquées ci-dessus, si l'économètre choisit au hasard la quantité d'engrais sur les différentes parcelles, le paramètre β_1 mesurera l'effet causal de l'engrais sur le rendement. Par construction, le terme Δu sera en moyenne égal à zéro.

Exemple 3.2: La relation entre le salaire et le niveau d'éducation (mesuré en années) s'écrit:

$$\text{sal} = \beta_0 + \beta_1 \cdot \text{edu} + u.$$

Si cette relation est estimée à l'aide de données observationnelles (issues d'une enquête, par exemple), la clause "ceteris paribus" ne sera pas nécessairement garantie car des variables incluses dans le terme d'erreur peuvent varier simultanément avec le niveau d'éducation. Pour éviter cela, une condition sur le terme d'erreur doit être satisfaite.

Pour mesurer correctement l'effet causal de x sur y , la condition suivante doit être satisfaite:

$$E(u|x) = \text{constante}.$$

Intuitivement, cette condition signifie que, en moyenne, l'effet de toutes les variables représentées par le terme d'erreur u reste constant quand la

variable x se modifie; elle implique, en vertu de la Propriété 2.13, que la covariance entre u et x est nulle,

$$\text{Cov}(u, x) = 0. \quad (3.1)$$

Pour estimer le modèle, une autre condition est utilisée, à savoir,

$$\text{E}(u) = 0. \quad (3.2)$$

Cette condition est une simple normalisation; elle n'a généralement pas une grande importance. Sous ces deux hypothèses, et en vertu de la Propriété 2.10, l'espérance conditionnelle de y pour x donnée est égale à

$$\text{E}(y|x) = \beta_0 + \beta_1 x + \text{E}[u|x] = \beta_0 + \beta_1 x.$$

Cette relation est appelée **droite de régression de la population**.

3.2 Dérivation des estimateurs des MCO

On dispose d'un échantillon aléatoire $\{(x_i, y_i) : i = 1, \dots, N\}$ de N observations. Les estimateurs des Moindres Carrés Ordinaires (MCO) sont désignés par $\hat{\beta}_0$ et $\hat{\beta}_1$. Pour les calculer, une manière de procéder consiste à utiliser la méthode des moments et à égaliser les moments théoriques (3.1) et (3.2) et les moments empiriques qui correspondent:

$$\frac{1}{N} \sum_{i=1}^N (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = \frac{1}{N} \sum_{i=1}^N \hat{u}_i = 0,$$

$$\frac{1}{N} \sum_{i=1}^N x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = \frac{1}{N} \sum_{i=1}^N x_i \hat{u}_i = 0,$$

où

$$\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

est le **résidu des MCO**, la première expression est la la moyenne empirique du résidu, et la seconde expression est la covariance empirique entre le résidu et la variable x .

De manière alternative, les estimateurs des MCO, $\hat{\beta}_0$ et $\hat{\beta}_1$, peuvent être obtenus par la minimisation du carré des résidus:

$$\min_{\{b_0, b_1\}} \sum_{i=1}^N (y_i - b_0 - b_1 x_i)^2.$$

En effet, les conditions de premier ordre sont:

$$\begin{aligned} -\frac{1}{2} \sum_{i=1}^N (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0, \\ -\frac{1}{2} \sum_{i=1}^N x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0; \end{aligned}$$

en d'autres termes, elles sont équivalentes à celles obtenues par la méthode des moments. La première équation devient:

$$\begin{aligned} \bar{y} &= \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \\ \text{avec } \bar{y} &= \frac{1}{N} \sum_{i=1}^N y_i \text{ et } \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \end{aligned}$$

et donne $\hat{\beta}_0$:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

La deuxième équation devient:

$$\begin{aligned} \sum_{i=1}^N x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0, \\ \sum_{i=1}^N x_i (y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i) &= 0. \end{aligned}$$

Elle devient ensuite:

$$\begin{aligned} \sum_{i=1}^N x_i (y_i - \bar{y}) &= \sum_{i=1}^N x_i (\hat{\beta}_1 x_i - \hat{\beta}_1 \bar{x}), \\ \sum_{i=1}^N x_i (y_i - \bar{y}) &= \hat{\beta}_1 \sum_{i=1}^N x_i (x_i - \bar{x}). \end{aligned}$$

Donc, si

$$\sum_{i=1}^N (x_i - \bar{x})^2 \neq 0,$$

alors

$$\begin{aligned}\widehat{\beta}_1 &= \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2} = \frac{\widehat{\text{Cov}}(x, y)}{\widehat{\text{Var}}(x)} \\ \widehat{\beta}_0 &= \bar{y} - \widehat{\beta}_1 \bar{x}\end{aligned}$$

La **valeur prédite des MCO** de la variable y pour $x = x_i$ est:

$$\hat{y}_i \equiv \widehat{\beta}_0 + \widehat{\beta}_1 x_i,$$

tandis que la **droite de régression des MCO** est:

$$\hat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x.$$

Celle-ci est une estimation de la droite de régression de la population.

Exemple 3.3: En utilisant un échantillon de 209 observations sur le salaire des P.D.G., Wooldridge (2010) obtient la droite de régression suivante:

$$\widehat{\text{sal}} = 963,191 + 18,501 \cdot \text{rend.}$$

où **sal** est la rémunération totale annuelle en milliers de dollars du PDG et **rend** le rendement moyen de l'action de la firme qu'il dirige. Selon cette équation, la rémunération moyenne des P.D.G. augmente de 18,500 dollars chaque fois que le rendement de l'action augmente d'un point. De plus, la constante représente la rémunération moyenne d'un PDG dont l'action aurait un taux de rendement égal à zero.

Exemple 3.4: En utilisant un échantillon de 526 observations sur les employés des Etats-Unis en 1976, Wooldridge (2010) obtient la droite de régression suivante:

$$\widehat{\text{sal}} = -0,90 + 0,54 \cdot \text{edu},$$

où **sal** est le salaire horaire du travailleur et **edu** est son niveau d'éducation mesurée en années. Selon cette équation, le salaire des travailleurs augmente, en moyenne, de 54 cents pour une augmentation d'une année du niveau d'éducation. La constante n'a pas d'interprétation satisfaisante.

Le modèle de régression simple est linéaire dans les paramètres. En dépit de cela, des relations non linéaires entre les variables explicatives et expliquées peuvent être modélisées en changeant assez simplement la définition des variables en question.

Exemple 3.5: En partant de l'échantillon utilisé dans l'exemple 3.3 et en exprimant le salaire horaire en logarithme – une transformation assez commode qui est souvent utilisée en économétrie appliquée – Wooldridge (2010) obtient la droite de régression suivante:

$$\widehat{\log(\text{sal})} = 0,584 + 0,083 \cdot \text{edu}.$$

Dans le cas présent, le paramètre β_1 est une semi-élasticité qui mesure l'effet en pourcentage sur le salaire horaire d'une augmentation d'une année d'éducation. Ainsi, le taux moyen de rendement de l'éducation sera de 8,3%.

3.3 Propriétés algébriques des MCO

Les résidus satisfont un certain nombre de propriétés algébriques. Celles-ci découlent directement de la manière dont les estimateurs des MCO sont construits, et ne nécessitent généralement pas de démonstration.

Propriété 3.1: *La moyenne des résidus est nulle:*

$$\sum_{i=1}^N \hat{u}_i = 0$$

Propriété 3.2: *La covariance entre les résidus et les valeurs de la variable explicative est nulle:*

$$\sum_{i=1}^N x_i \hat{u}_i = 0$$

Propriété 3.3: *La régression passe par le point moyen de l'échantillon:*

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

Propriété 3.4: La covariance entre les résidus et les valeurs prédites est nulle:

$$\begin{aligned}\sum_{i=1}^N \hat{y}_i \hat{u}_i &= \sum_{i=1}^N (\hat{\beta}_0 + \hat{\beta}_1 x_i) \hat{u}_i \\ &= \hat{\beta}_0 \sum_{i=1}^N \hat{u}_i + \hat{\beta}_1 \sum_{i=1}^N x_i \hat{u}_i = 0.\end{aligned}$$

La dérivation de la Propriété 3.5 nécessite de définir les concepts suivants:

$$\begin{aligned}\text{SCT} &= \sum_{i=1}^N (y_i - \bar{y})^2, \\ \text{SCE} &= \sum_{i=1}^N (\hat{y}_i - \bar{y})^2, \\ \text{SCR} &= \sum_{i=1}^N \hat{u}_i^2.\end{aligned}$$

Propriété 3.5: La somme des carrés totaux est égale à la somme de somme des carrés expliqués et de la somme des carrés résiduels:

$$\text{SCT} = \text{SCE} + \text{SCR}$$

Cette propriété permet de calculer le coefficient de détermination (ou R-carré) de la manière suivante

$$R^2 = \frac{\text{SCE}}{\text{SCT}} = 1 - \frac{\text{SCR}}{\text{SCT}} \in [0, 1].$$

Le R-carré est une mesure de l'ajustement de la droite de régression au nuage de points, sa valeur est égale à un lorsque les points des observations sont parfaitement alignés. Il est possible de montrer que le coefficient de détermination est égal à:

$$R^2 = \frac{(\widehat{\text{Cov}(x, y)})^2}{\widehat{\text{Var}(x)} \widehat{\text{Var}(y)}};$$

c'est-à-dire le carré du coefficient de corrélation défini dans le Chapitre 2.

Démonstration de la Propriété 3.5:

$$\begin{aligned}
\text{SCT} &= \sum_{i=1}^N (y_i - \bar{y})^2 \\
&= \sum_{i=1}^N ((y_i - \hat{y}_i) + (\hat{y}_i - \bar{y}))^2 \\
&= \sum_{i=1}^N (\hat{u}_i + (\hat{y}_i - \bar{y}))^2 \\
&= \sum_{i=1}^N (\hat{u}_i^2 + 2\hat{u}_i(\hat{y}_i - \bar{y}) + (\hat{y}_i - \bar{y})^2) \\
&= \sum_{i=1}^N \hat{u}_i^2 + 2 \sum_{i=1}^N \hat{u}_i(\hat{y}_i - \bar{y}) + \sum_{i=1}^N (\hat{y}_i - \bar{y})^2 \\
&= \text{SCR} + 2 \sum_{i=1}^N \hat{u}_i(\hat{y}_i - \bar{y}) + \text{SCE}
\end{aligned}$$

Or

$$\begin{aligned}
\sum_{i=1}^N u_i(\hat{y}_i - \bar{y}) &= \sum_{i=1}^N \hat{u}_i \hat{y}_i - \bar{y} \sum_{i=1}^N \hat{u}_i \\
&= 0 \text{ en vertu de la Propriété 3.4.}
\end{aligned}$$

Exemple 3.6: En utilisant les estimations faites précédemment dans l'exemple 3.3, le coefficient de détermination de la régression du salaire des P.D.G. sur le rendement des actions,

$$\widehat{\text{sal}} = 963,191 + 18,501 \cdot \text{rend},$$

est égal à 0,0132. Le coefficient de détermination est proche de zero; en d'autres termes, la régression n'explique qu'une très petite fraction de la variabilité des salaires des P.D.G.

Changement d'unités de mesure

Si les unités de mesure des variables sont modifiées, les estimations des paramètres changent mais d'une manière cohérente de sorte que l'interprétation des effets reste la même.

Exemple 3.7: Dans le cas de la régression du salaire des P.D.G. sur le rendement des actions de l'exemple précédent, la variable expliquée est exprimée en milliers de dollars, et la variable explicative en points de pourcentage. La droite de régression est:

$$\widehat{\text{sal}} = 963,191 + 18,501 \cdot \text{rend.}$$

Si la variable expliquée est exprimée en dollars plutôt qu'en milliers de dollars, elle devient:

$$\widehat{\text{sal_dollars}} = 963.191 + 18.501 \cdot \text{rend.}$$

Autrement dit, les estimations de la constante et de la pente sont multipliées par 1000. Si la variable expliquée est exprimée en milliers d'euros, et la variable explicative en décimales, la droite de régression devient:

$$\widehat{\text{sal}} = 963,191 + 1850,1 \cdot \text{rend_dec.}$$

Seul l'estimation de la pente est multipliée par 100. Dans tous les cas, le coefficient de détermination ne se modifie pas lorsque l'on change les unités de mesure.

3.4 Propriétés statistiques des MCO

Les bonnes propriétés statistiques des estimateurs des MCO nécessitent qu'un ensemble d'hypothèses soient satisfaites.

Hypothèse 3.1: Le modèle dans la population peut se décrire par une relation linéaire à une seule variable explicative telle que:

$$y = \beta_0 + \beta_1 x + u$$

où β_0 et β_1 sont des paramètres, et u est un terme d'erreur.

Hypothèse 3.2: Le terme u a une espérance de zéro pour toute valeur des variables indépendantes. En d'autres termes,

$$E(u|x) = 0.$$

Hypothèse 3.3: Un échantillon aléatoire de N observations, $\{(y_i, x_i) : i = 1, \dots, N\}$, issu du modèle de population décrit par les Hypothèses 3.1 et 3.2 est mis à disposition de l'économètre.

Hypothèse 3.4: Dans l'échantillon (et donc dans la population), le régresseur n'est pas une constante:

$$\sum_{i=1}^N (x_i - \bar{x})^2 \neq 0.$$

Les démonstrations qui suivent sont basées sur la transformation suivante de l'estimateur de la pente:

$$\begin{aligned} \widehat{\beta}_1 &= \frac{\sum_{i=1}^N (x_i - \bar{x}_i) y_i}{\sum_{i=1}^N (x_i - \bar{x})^2} \\ &= \frac{\sum_{i=1}^N (x_i - \bar{x}_i) (\beta_0 + \beta_1 x_i + u_i)}{\sum_{i=1}^N (x_i - \bar{x})^2} \\ &= \beta_0 \frac{\sum_{i=1}^N (x_i - \bar{x}_i)}{\sum_{i=1}^N (x_i - \bar{x})^2} + \beta_1 \frac{\sum_{i=1}^N (x_i - \bar{x}_i) x_i}{\sum_{i=1}^N (x_i - \bar{x})^2} + \frac{\sum_{i=1}^N (x_i - \bar{x}_i) u_i}{\sum_{i=1}^N (x_i - \bar{x})^2}. \end{aligned}$$

En se rappelant que $\sum_{i=1}^N (x_i - \bar{x}_i) x_i = \sum_{i=1}^N (x_i - \bar{x})^2$ et que $\sum_{i=1}^N (x_i - \bar{x}_i) = 0$, on obtient:

$$\widehat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^N (x_i - \bar{x}_i) u_i}{\sum_{i=1}^N (x_i - \bar{x})^2}.$$

L'estimateur $\widehat{\beta}_1$ est donc égal à la somme de la vraie valeur de β_1 dans la population et une combinaison de termes aléatoires. L'estimateur $\widehat{\beta}_1$ est donc une variable aléatoire.

Théorème 3.1: *Sous les hypothèses 3.1–3.4, les estimateurs des MCO sont non biaisés:*

$$\mathbf{E}(\widehat{\beta}_0 | \mathbf{x}) = \beta_0 \quad \text{et} \quad \mathbf{E}(\widehat{\beta}_1 | \mathbf{x}) = \beta_1,$$

où $\mathbf{x} = (x_1, x_2, \dots, x_N)$ désigne l'ensemble des variables indépendantes.

En vertu de la Propriété 2.12, le théorème implique aussi que $\mathbf{E}(\widehat{\beta}_0) = \beta_0$ et $\mathbf{E}(\widehat{\beta}_1) = \beta_1$.

Démonstration: (Dans la démonstration, les espérances sont conditionnelles aux valeurs x de l'échantillon.) La démonstration est en deux parties:

Partie 1: De ce qui précède, on a:

$$\widehat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^N (x_i - \bar{x}) u_i}{\sum_{i=1}^N (x_i - \bar{x})^2}.$$

Si l'on prend les espérances des membres de droite et de gauche, et que l'on utilise les propriétés de l'opérateur espérance, on a:

$$\begin{aligned} \mathbf{E}(\widehat{\beta}_1 | \mathbf{x}) &= \beta_1 + \mathbf{E} \left(\frac{\sum_{i=1}^N (x_i - \bar{x}) u_i}{\sum_{i=1}^N (x_i - \bar{x})^2} \middle| \mathbf{x} \right) \\ &= \beta_1 + \frac{\sum_{i=1}^N \mathbf{E}((x_i - \bar{x}) u_i | \mathbf{x})}{\sum_{i=1}^N (x_i - \bar{x})^2} \\ &= \beta_1 + \frac{\sum_{i=1}^N (x_i - \bar{x}) \mathbf{E}(u_i | \mathbf{x})}{\sum_{i=1}^N (x_i - \bar{x})^2}, \end{aligned}$$

en vertu de la Propriété 2.10, où \mathbf{x} désigne les données sur toutes les variables indépendantes. Puisque $\mathbf{E}(u_i | \mathbf{x}) = 0$, on démontre que $\mathbf{E}(\widehat{\beta}_1 | \mathbf{x}) = \beta_1$.

Partie 2: La définition de $\widehat{\beta}_0$ est:

$$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x}$$

En utilisant la propriété $\bar{y} = \beta_0 + \beta_1 \bar{x}$,

$$\widehat{\beta}_0 = \beta_0 + (\beta_1 - \widehat{\beta}_1) \bar{x} + \bar{u}$$

Si l'on prend les espérances des membres de droite et de gauche, et que l'on utilise les propriétés de l'opérateur espérance, on a:

$$\begin{aligned} \mathbf{E}(\widehat{\beta}_0 | \mathbf{x}) &= \beta_0 + \mathbf{E} \left((\beta_1 - \widehat{\beta}_1) \bar{x} \middle| \mathbf{x} \right) + \mathbf{E}(\bar{u} | \mathbf{x}) \\ &= \beta_0 + \bar{x} \mathbf{E} \left(\beta_1 - \widehat{\beta}_1 \middle| \mathbf{x} \right), \end{aligned}$$

en vertu de la Propriété 2.10. Puisque $\mathbf{E}(\widehat{\beta}_1 | \mathbf{x}) = \beta_1$, on démontre que $\mathbf{E}(\widehat{\beta}_0 | \mathbf{x}) = \beta_0$. \square

Exemple 3.8: Le Michigan Educational Assessment Program test est un test standardisé effectué dans les écoles secondaires réalisées en 1992-93 dans l'Etat du Michigan. La variable `math10` représente la proportion d'élèves réussissant le test dans les écoles. Wooldridge (2010) considère la régression de cette variable sur la proportion d'élèves qui bénéficient d'une subvention pour les repas:

$$\widehat{\text{math10}} = 32,14 - 0,319 \cdot \log(\text{prosub}),$$

$$R^2 = 0,171.$$

En d'autres termes, une augmentation de 10% des élèves bénéficiant de la subvention mène à une réduction d'environ 3,2 points de pourcentage du taux de réussite. Les estimateurs sont certainement biaisés car l'hypothèse d'espérance conditionnelle n'est pas satisfaite.

Pour calculer la variance des estimateurs, l'hypothèse suivante, qui assure que les termes aléatoires ont une variance constante, est nécessaire.

Hypothèse 3.5: Le terme d'erreur est homoscédastique. En d'autres termes,

$$\text{Var}(u|x) = \sigma^2,$$

où σ^2 est une constante positive.

Remarque: Les hypothèses 3.1, 3.2 et 3.5 peuvent être mises sous la forme conditions sur la moyenne et la variance conditionnelle de la variable dépendante:

$$\mathbf{E}(y|x) = \beta_0 + \beta_1 x,$$

$$\text{Var}(y|x) = \sigma^2.$$

On a alors le théorème suivant.

Théorème 3.2: *Sous les hypothèses 3.1 à 3.5, les variances des estimateurs de la constante et de la pente sont définies par*

$$\begin{aligned}\text{Var}(\widehat{\beta}_0|\mathbf{x}) &= \frac{\sigma^2 \sum_{i=1}^N x_i^2}{N \sum_{i=1}^N (x_i - \bar{x})^2}, \\ \text{Var}(\widehat{\beta}_1|\mathbf{x}) &= \frac{\sigma^2}{\sum_{i=1}^N (x_i - \bar{x})^2},\end{aligned}$$

où $\mathbf{x} = (x_1, x_2, \dots, x_K)$ désigne l'ensemble des variables indépendantes.

Ces expressions de la variances sont conditionnelles aux valeurs des variables explicatives. Par souci de simplicité, nous omettrons généralement celles-ci dans la notation de la variance des estimateurs.

Démonstration: (Dans la démonstration, les espérances sont conditionnelles aux valeurs x de l'échantillon.) La démonstration est en deux parties:

Partie 1: On a:

$$\widehat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^N (x_i - \bar{x})u_i}{\sum_{i=1}^N (x_i - \bar{x})^2}.$$

En prenant la variance des membres de droite et de gauche pour des valeurs des variables explicatives données, et en utilisant les propriétés de l'opérateur variance, on obtient:

$$\begin{aligned}\text{Var}(\widehat{\beta}_1|\mathbf{x}) &= \text{Var}\left(\frac{\sum_{i=1}^N (x_i - \bar{x})u_i}{\sum_{i=1}^N (x_i - \bar{x})^2} \middle| \mathbf{x}\right) \\ &= \frac{\text{Var}\left(\sum_{i=1}^N (x_i - \bar{x})u_i \middle| \mathbf{x}\right)}{\left(\sum_{i=1}^N (x_i - \bar{x})^2\right)^2},\end{aligned}$$

en utilisant la Propriété 2.15. Puisque l'échantillon est aléatoire et que donc les termes d'erreur sont indépendants les uns des autres, on a:

$$\begin{aligned}\text{Var}(\widehat{\beta}_1|\mathbf{x}) &= \frac{\sum_{i=1}^N (x_i - \bar{x})^2 \text{Var}(u_i|\mathbf{x})}{\left(\sum_{i=1}^N (x_i - \bar{x})^2\right)^2} \\ &= \frac{\sum_{i=1}^N (x_i - \bar{x})^2 \sigma^2}{\left(\sum_{i=1}^N (x_i - \bar{x})^2\right)^2}.\end{aligned}$$

En simplifiant, finalement, on obtient:

$$\text{Var}(\widehat{\beta}_1|\mathbf{x}) = \frac{\sigma^2}{\sum_{i=1}^N (x_i - \bar{x})^2}.$$

Partie 2: Comme on l'a vu précédemment, la définition de $\widehat{\beta}_0$ peut s'écrire

$$\widehat{\beta}_0 = \beta_0 + (\beta_1 - \widehat{\beta}_1)\bar{x} + \bar{u}.$$

Donc,

$$\begin{aligned} \text{Var}(\widehat{\beta}_0|\mathbf{x}) &= \text{Var}\left(\left(\beta_1 - \widehat{\beta}_1\right)\bar{x} + \bar{u} \mid \mathbf{x}\right) \\ &= \mathbf{E}\left(\left(\left(\beta_1 - \widehat{\beta}_1\right)\bar{x} + \bar{u}\right)^2 \mid \mathbf{x}\right) \\ &= \bar{x}^2 \mathbf{E}\left(\widehat{\beta}_1^2 \mid \mathbf{x}\right) + \mathbf{E}\left(\bar{u}^2 \mid \mathbf{x}\right) + 2\bar{x} \mathbf{E}\left(\widehat{\beta}_1 \bar{u} \mid \mathbf{x}\right) \end{aligned}$$

Or

$$\begin{aligned} \mathbf{E}\left(\widehat{\beta}_1 \bar{u} \mid \mathbf{x}\right) &= \mathbf{E}\left(\left(\widehat{\beta}_1 - \beta_1\right)\bar{u} \mid \mathbf{x}\right) \\ &= \mathbf{E}\left(\frac{\sum_{i=1}^N (x_i - \bar{x})u_i}{\sum_{i=1}^N (x_i - \bar{x})^2} \left(\frac{1}{N} \sum_{i=1}^N u_i\right) \mid \mathbf{x}\right) \\ &= \frac{\mathbf{E}\left(\frac{1}{N} \sum_{j=1}^N u_j \sum_{i=1}^N (x_i - \bar{x})u_i \mid \mathbf{x}\right)}{\sum_{i=1}^N (x_i - \bar{x})^2} \\ &= \frac{\mathbf{E}\left(\frac{1}{N} \sum_{j=1}^N \sum_{i=1}^N (x_i - \bar{x})u_i u_j \mid \mathbf{x}\right)}{\sum_{i=1}^N (x_i - \bar{x})^2} \\ &= 0 \end{aligned}$$

Donc:

$$\begin{aligned} \text{Var}(\widehat{\beta}_0|\mathbf{x}) &= \bar{x}^2 \text{Var}\left(\widehat{\beta}_1|\mathbf{x}\right) + \text{Var}\left(\bar{u}|\mathbf{x}\right) \\ \text{Var}(\widehat{\beta}_0|\mathbf{x}) &= \frac{\bar{x}^2 \sigma^2}{\sum_{i=1}^N (x_i - \bar{x})^2} + \frac{\sigma^2}{N} \end{aligned}$$

Or

$$\sum_{i=1}^N (x_i - \bar{x})^2 = \sum_{i=1}^N x_i^2 - N\bar{x}^2$$

Donc:

$$\begin{aligned}\text{Var}(\widehat{\beta}_0|\mathbf{x}) &= \frac{\sum_{i=1}^N x_i^2 - \frac{(\sum_{i=1}^N x_i)^2}{N}}{N \sum_{i=1}^N (x_i - \bar{x})^2} \sigma^2 + \frac{\sigma^2}{N} \\ &= \frac{\sigma^2}{N \sum_{i=1}^N (x_i - \bar{x})^2} \sum_{i=1}^N x_i^2. \quad \square\end{aligned}$$

Puisque les formules de la variance des estimateurs dépendent de la variance des termes d'erreur σ^2 , qui est non observée, il faut trouver un estimateur de cette dernière. Pour cela, remarquons préalablement que les résidus sont des "approximations" des vrais termes d'erreur. En effet,

$$\begin{aligned}y_i &= \beta_0 + \beta_1 x_i + u_i \\ y_i &= \widehat{\beta}_0 + \widehat{\beta}_1 x_i + \widehat{u}_i\end{aligned}$$

et donc,

$$\begin{aligned}\widehat{u}_i &= y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i \\ &= (\beta_0 + \beta_1 x_i + u_i) - \widehat{\beta}_0 - \widehat{\beta}_1 x_i \\ &= u_i - (\widehat{\beta}_0 - \beta_0) - (\widehat{\beta}_1 - \beta_1) x_i.\end{aligned}$$

L'estimateur de σ^2 est basé sur ces résidus. En fait, deux estimateurs sont envisageables, mais seul le second est non biaisé.

Estimateur 1: Un premier estimateur que l'on peut considérer est le suivant:

$$\widehat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N \widehat{u}_i^2 = \frac{\text{SCR}}{N}$$

mais cet estimateur sera biaisé car les résidus doivent satisfaire des contraintes:

$$\sum_{i=1}^N \widehat{u}_i = 0, \quad \sum_{i=1}^N x_i \widehat{u}_i = 0.$$

Cet estimateur est appelé estimateur du Maximum de Vraisemblance (MV).

Estimateur 2: Un second estimateur est le suivant:

$$\hat{\sigma}^2 = \frac{1}{N-2} \sum_{i=1}^N \hat{u}_i^2 = \frac{\text{SCR}}{N-2}.$$

Celui-ci est non biaisé comme le montre le théorème suivant. C'est l'estimateur traditionnel des MCO.

Théorème 3.3: *Sous les hypothèses 3.1 à 3.5, l'estimateur des MCO de la variance des termes d'erreur est non biaisé:*

$$\mathbb{E}(\hat{\sigma}^2 | \mathbf{x}) = \sigma^2,$$

où $\mathbf{x} = (x_1, x_2, \dots, x_K)$ désigne l'ensemble des variables indépendantes.

Démonstration: Voir annexe. \square

L'estimateur de la variance de l'estimateur des paramètres β_0 et β_1 est donné par:

$$\widehat{\text{Var}}(\hat{\beta}_0) = \frac{\hat{\sigma}^2 \sum_{i=1}^N x_i^2}{N \sum_{i=1}^N (x_i - \bar{x})^2}, \quad \widehat{\text{Var}}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\sum_{i=1}^N (x_i - \bar{x})^2}.$$

3.5 Le modèle sans constante

Dans certains cas, l'économètre désire imposer la condition que la valeur espérée de y est zero quand $x = 0$. Formellement, le modèle estimé est le suivant:

$$y = \beta_1 x + u, \quad \text{avec } \mathbb{E}(u|x) = 0$$

si bien que

$$\mathbb{E}(y|x) = \beta_1 x.$$

La méthode d'estimation repose également sur la minimisation du carré des résidus, à savoir,

$$\min_{\{b_1\}} \sum_{i=1}^N (y_i - b_1 x_i)^2.$$

La condition de premier ordre est:

$$\sum_{i=1}^N x_i(y_i - \hat{\beta}_1 x_i) = 0.$$

En conséquence,

$$\sum_{i=1}^N x_i y_i - \hat{\beta}_1 \sum_{i=1}^N x_i^2 = 0.$$

et

$$\hat{\beta}_1 = \frac{\sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^2},$$

sous la condition qu'au moins une valeur de x est différente de zero. L'estimateur de la pente du modèle de régression linéaire avec constante coincide avec l'estimateur ci-dessus dans le cas où $\bar{x} = 0$. Il est facile de montrer que les propriétés algébriques dérivées précédemment ne sont plus satisfaites (à l'exception de la Propriété 3.2).

Part II

Le modèle de régression multiple

Chapter 4

Définition, calcul et interprétation

4.1 Définition

Le modèle de régression simple présenté dans le chapitre précédent est souvent insatisfaisant. En effet, pour obtenir de bonnes prédictions d'une variable quelconque, il est nécessaire de prendre en compte non pas un seul facteur explicatif mais plusieurs. De plus, dans le modèle de régression simple, le terme d'erreur peut potentiellement inclure un grand nombre de variables explicatives. Si ces variables sont corrélées avec le régresseur, l'hypothèse 5.2 ne sera pas satisfaite et les estimateurs des MCO seront biaisés. Le modèle de régression multiple permet alors de "sortir" ces variables du terme d'erreur et de les mettre parmi les régresseurs. Ce problème est illustré dans l'exemple suivant.

Exemple 4.1: Dans le modèle d'investissement en capital humain, le salaire horaire est expliqué par le niveau d'éducation (mesuré en nombre d'années). Il s'écrit de la manière suivante:

$$\text{sal} = \beta_0 + \beta_1 \cdot \text{edu} + u,$$

où le terme d'erreur inclut des variables telles que exp , le nombre d'années d'expérience du travailleur. Toutefois, puisque les personnes qui prolongent leurs études débutent leur carrière professionnelle plus tard, l'expérience est susceptible d'être corrélée avec le terme d'erreur. L'économètre peut alors considérer le modèle plus général suivant:

$$\text{sal} = \beta_0 + \beta_1 \cdot \text{edu} + \beta_2 \cdot \text{exp} + u.$$

Dans ce modèle, le terme d'erreur n'incorpore plus l'expérience puisque cette variable figure parmi les régresseurs. Cependant, nous verrons par la suite un problème de variable omise dans les équations de salaire qui est sans doute plus important que celui-ci.

Enfin, le modèle de régression multiple permet de représenter des relations non linéaires complexes entre une variable expliquée et une ou plusieurs variables explicatives, comme on peut le voir dans l'exemple suivant.

Exemple 4.2: Dans le modèle du consommateur (voir exemple 1.2), les dépenses consacrées à un bien quelconque, par exemple les vêtements, par un ménage sont exprimés en fonction du revenu. Si l'on omet les prix pour simplifier, le modèle peut s'écrire:

$$\text{depvet} = \beta_0 + \beta_1 \cdot \text{rev} + \beta_2 \cdot \text{rev}^2 + u,$$

où rev est le revenu du ménage, et depvet sont les dépenses consacrées aux vêtements. L'effet marginal du revenu sur les dépenses de vêtements est donc égal à

$$\frac{\partial \text{depvet}}{\partial \text{rev}} = \beta_1 + 2\beta_2 \cdot \text{rev} \quad \text{si } \Delta u = 0.$$

Cet effet est non linéaire puisqu'il dépend du niveau de consommation. En particulier, si $\beta_1 > 0$ et $\beta_2 < 0$, l'effet du revenu sera positif (pour des niveaux de revenu suffisamment faibles du moins) et diminuera avec le niveau de revenu. Le modèle peut également s'écrire de manière plus générale comme suit:

$$\text{depvet} = \beta_0 + \beta_1 \cdot \text{rev} + \beta_2 \cdot \text{rev}^2 + \beta_3 \cdot \text{rev}^3 + u.$$

En fin de compte, des relations d'une grande flexibilité entre variable expliquée et variables explicatives peuvent être estimées.

4.2 Mécanique et interprétation des MCO

On considère un échantillon $\{(x_{1i}, \dots, x_{Ki}, y_i) : i = 1, \dots, N\}$ de N observations. Les estimateurs des MCO sont obtenus par la minimisation du carré des résidus:

$$\min_{b_0, b_1, \dots, b_K} \sum_{i=1}^N (y_i - b_0 - b_1 x_{1i} \dots - b_K x_{Ki})^2$$

Les conditions de premier ordre ont la forme suivante:

$$\begin{aligned} -2 \sum_{i=1}^N (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} \dots - \hat{\beta}_K x_{Ki}) &= 0, \\ -2 \sum_{i=1}^N x_{1i} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} \dots - \hat{\beta}_K x_{Ki}) &= 0, \\ &\vdots \\ -2 \sum_{i=1}^N x_{Ki} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} \dots - \hat{\beta}_K x_{Ki}) &= 0. \end{aligned}$$

Sous certaines conditions qui seront examinées très prochainement, ce système de $K + 1$ équations a une solution unique, qui correspondent aux estimateurs des MCO.

Comme précédemment la **valeur prédite des MCO** de y pour $x_1 = x_{1i}$, \dots , et $x_K = x_{Ki}$ est donnée par

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \dots + \hat{\beta}_K x_{Ki}$$

et le **résidu des MCO** par

$$\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \dots - \hat{\beta}_K x_{Ki} = y_i - \hat{y}_i.$$

L'intérêt du modèle de régression multiple et de la méthode des MCO est de permettre de mesurer les effets "toutes autres choses étant égales" des variables explicatives même si les données ne sont pas de nature expérimentale

et que les variables explicatives ne sont pas contrôlées par l'économètre. Enfin, il est possible aussi d'estimer un modèle sans constante en posant $\widehat{\beta}_0 = 0$ dans les équations qui précèdent et en supprimant la première d'entre elles. L'exemple suivant illustre l'interprétation des estimations obtenues avec le modèle de régression multiple.

Exemple 4.3: En utilisant l'échantillon de 526 observations sur les employés des Etats-Unis en 1976 de l'exemple 3.4, et en ajoutant des variables explicatives, on obtient la droite de régression suivante:

$$\widehat{\log(\text{sal})} = 0,284 + 0,092 \cdot \text{edu} + 0,0041 \cdot \text{exp} + 0,022 \cdot \text{anc}.$$

où **exp** et **anc** désignent le nombre d'années d'expérience et d'ancienneté du travailleur. Une année d'éducation supplémentaire, en maintenant toutes les autres choses égales, dont l'expérience et l'ancienneté, représente un accroissement du salaire d'environ 9,2%. Si plusieurs variables indépendantes se modifient simultanément, les effets sont sommés. Par exemple, si l'expérience et l'ancienneté augmentent d'une année chacune, l'effet total sera donné par

$$\begin{aligned} \Delta \widehat{\log(\text{sal})} &= 0,0041 \cdot \Delta \text{exp} + 0,022 \cdot \Delta \text{anc} \\ &= 0,0261. \end{aligned}$$

Ainsi une année d'expérience et d'ancienneté supplémentaires implique un accroissement de salaire d'environ 2,6%.

4.3 Propriétés algébriques des MCO

Les estimateurs des MCO du modèle de régression multiple possèdent des propriétés algébriques analogues à celles dérivées pour le modèle de régression simple. Les démonstrations sont élémentaires et sont omises.

Propriété 4.1: *La moyenne des résidus est nulle:*

$$\sum_{i=1}^N \widehat{u}_i = 0;$$

Propriété 4.2: La covariance entre les résidus et les valeurs de la variable explicative est nulle:

$$\sum_{i=1}^N x_{ki} \hat{u}_i = 0 \text{ pour } k = 1, \dots, K;$$

Propriété 4.3: La régression passe par le point moyen de l'échantillon:

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \dots + \hat{\beta}_K \bar{x}_K;$$

Propriété 4.4: La covariance entre les résidus et les valeurs prédites est nulle:

$$\sum \hat{u}_i \hat{y}_i = 0.$$

De plus, si l'on définit

$$\begin{aligned} \text{SCT} &= \sum_{i=1}^N (y_i - \bar{y})^2, \\ \text{SCE} &= \sum_{i=1}^N (\hat{y}_i - \bar{y})^2, \\ \text{SCR} &= \sum_{i=1}^N \hat{u}_i^2, \end{aligned}$$

le propriété suivante peut être démontrée.

Propriété 4.5: La somme des carrés totaux est égale à la somme des carrés expliqués et la somme des carrés résiduels:

$$\text{SCT} = \text{SCE} + \text{SCR}$$

Le coefficient de détermination ou R-carré est alors défini par

$$R^2 = \frac{\text{SCE}}{\text{SCT}} = 1 - \frac{\text{SCR}}{\text{SCT}} \in [0, 1]$$

On peut aussi montrer que

$$R^2 = \frac{\widehat{\text{Cov}}(y, \hat{y})^2}{\widehat{\text{Var}}(y) \widehat{\text{Var}}(\hat{y})}.$$

Cette égalité signifie que le coefficient de détermination est égal au carré du coefficient de corrélation entre y_i et \hat{y}_i .

Remarque: Lorsque une ou plusieurs variables explicatives sont ajoutées dans le modèle, le coefficient de détermination ne décroît jamais et il augmente presque toujours. Cette caractéristique peut rendre son interprétation délicate comme on le verra plus loin.

4.4 Une formule utile

Pour dériver certains résultats et affiner l'intuition des MCO, il est utile de trouver une formule simple pour caractériser les estimateurs. La droite de régression inclut une constante et s'écrit de la manière suivante:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_K x_K.$$

Dans ce cas, on peut montrer que l'estimateur de β_k est simplement égal à:

$$\hat{\beta}_k = \frac{\sum_{i=1}^N \hat{r}_{ki} y_i}{\sum_{i=1}^N \hat{r}_{ki}^2}, \text{ avec } k = 0, 1, \dots, K, \quad (4.1)$$

où \hat{r}_k pour $k = 1, \dots, K$ est le résidu obtenu par la régression (avec constante) de la variable x_k sur l'ensemble des autres variable explicatives du modèle, à savoir, $x_1, x_2, \dots, x_{k-1}, x_{k+1}, \dots, x_K$, et \hat{r}_k pour $k = 0$ est le résidu obtenu par la régression (sans constante) de la constante unitaire sur l'ensemble des autres variable explicatives du modèle, à savoir, x_1, x_2, \dots, x_K , c'est-à-dire:

$$\hat{r}_{ki} = x_{ki} - \hat{x}_{ki}.$$

Donc, selon la formule ci-dessus, l'estimateur de β_k peut être obtenu par la régression de la variable y sur le résidu \hat{r}_k . Cette formulation a une interprétation très simple dans le cas de l'estimateur d'une pente. En effet, \hat{r}_k représente la part de x_k qui n'est pas corrélée avec les autres variables explicatives du modèle ou, en d'autres termes, \hat{r}_k correspond à x_k dont on aurait purgé l'effet des autres variables explicatives. L'estimateur mesure donc l'effet de x_k sur y en maintenant les autres variables explicatives constantes.

On peut facilement prouver ce résultat dans le cas $K = 2$ et en calculant la valeur de $\widehat{\beta}_1$. La généralisation de cette démonstration est évidente.

Démonstration partielle de la formule (4.1): D'abord, les estimateurs des MCO doivent être tels que la Propriété 4.2 est satisfaite:

$$\sum_{i=1}^N x_{1i} \widehat{u}_i = 0.$$

Si l'on remplace x_{1i} par $\widehat{x}_{1i} + \widehat{r}_{1i}$, l'on obtient:

$$\sum_{i=1}^N (\widehat{x}_{1i} + \widehat{r}_{1i}) \widehat{u}_i = 0.$$

On note que

$$\sum_{i=1}^N \widehat{x}_{1i} \widehat{u}_i = \sum_{i=1}^N (\widehat{\delta}_0 + \widehat{\delta}_2 x_{2i}) \widehat{u}_i = 0, \quad (4.2)$$

puisque \widehat{u}_i est un résidu qui satisfait les Propriétés 4.1 et 4.2, où $\widehat{\delta}_0$ et $\widehat{\delta}_2$ sont les estimateurs obtenus en régressant x_1 sur x_2 . L'expression ci-dessus se simplifie donc de la manière suivante:

$$\sum_{i=1}^N \widehat{r}_{1i} \widehat{u}_i = 0.$$

En remplaçant \widehat{u}_i par $y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_{1i} - \widehat{\beta}_2 x_{2i}$ et en utilisant les Propriétés 4.1 et 4.2, on obtient:

$$\sum_{i=1}^N \widehat{r}_{1i} (y_i - \widehat{\beta}_1 x_{1i}) = 0.$$

En remplaçant à nouveau x_{1i} par $\widehat{x}_{1i} + \widehat{r}_{1i}$, cette expression devient:

$$\sum_{i=1}^N \widehat{r}_{1i} (y_i - \widehat{\beta}_1 (\widehat{x}_{1i} + \widehat{r}_{1i})) = 0,$$

ou encore, en utilisant la propriété (4.2),

$$\sum_{i=1}^N \widehat{r}_{1i} (y_i - \widehat{\beta}_1 \widehat{r}_{1i}) = 0.$$

Cette dernière expression peut s'écrire de la manière suivante:

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^N \widehat{r}_{1i} y_i}{\sum_{i=1}^N \widehat{r}_{1i} \widehat{r}_{1i}},$$

afin de donner une décomposition de l'estimateur $\widehat{\beta}_1$ (l'estimateur $\widehat{\beta}_2$ peut naturellement s'écrire d'une manière décomposable). La démonstration peut être adaptée pour traiter le cas de la constante et elle peut être généralisée à un nombre quelconque de variables explicatives. \square

Chapter 5

La distribution des estimateurs

Ce chapitre a pour objectif de dériver la distribution des estimateurs en fonction des caractéristiques de l'échantillon.

5.1 L'espérance des estimateurs des MCO

Le calcul de l'espérance des estimateurs des MCO nécessite un ensemble d'hypothèses qui généralisent les hypothèses utilisées précédemment dans le cadre du modèle de régression simple.

Hypothèse 5.1: Le modèle dans la population peut se décrire par une relation linéaire à K variables explicatives telle que:

$$y = \beta_0 + \beta_1 \cdot x_1 + \dots + \beta_K \cdot x_K + u$$

où $\beta_0, \beta_1, \dots, \beta_K$ sont des paramètres, et u est un terme d'erreur.

On peut aussi écrire la droite de régression de la manière suivante:

$$y = \beta_0 \cdot 1 + \beta_1 \cdot x_1 + \dots + \beta_K \cdot x_K + u.$$

Donc, le paramètre β_0 peut être vu comme celui associé à une variable qui prendrait une valeur constante toujours égale à 1.

Hypothèse 5.2: Le terme d'erreur u a une espérance de zéro pour toute valeur des variables indépendantes. En d'autres termes,

$$E(u|x_1, \dots, x_K) = 0$$

Hypothèse 5.3: Un échantillon aléatoire de N observations, $\{(y_i, x_{1i}, \dots, x_{Ki}) : i = 1, \dots, N\}$ issu du modèle de population décrit par les hypothèses 5.1 et 5.2 est mis à disposition de l'économètre.

Note: les variables explicatives pour l'observation i peuvent écrire de manière compacte par $\mathbf{x}_i = (x_{1i}, \dots, x_{Ki})$ avec $i = 1, \dots, N$.

Hypothèse 5.4: Dans l'échantillon (et donc dans la population), les variables indépendantes ne sont pas colinéaires. En d'autres termes, il n'existe pas un ensemble de scalaires c_0, \dots, c_K avec au moins un de ces scalaires différent de zéro tels que

$$c_0 + c_1 x_{1i} + \dots + c_K x_{Ki} = 0 \tag{5.1}$$

pour tout $i = 1, \dots, N$.

Cette hypothèse généralise l'hypothèse 3.4: elle serait violée si l'une des variables explicatives était constante.

Remarques:

1. Selon une interprétation particulière de l'hypothèse 3.4, le coefficient de détermination obtenu si l'on régresse une variable explicative quelconque sur l'ensemble des autres variables explicatives doit être différent de l'unité.
2. Les variables explicatives peuvent être imparfaitement corrélées sans que l'hypothèse 5.4 soit violée. Par exemple, dans la régression suivante,

$$\text{depvet} = \beta_0 + \beta_1 \cdot \text{rev} + \beta_2 \cdot \text{rev}^2 + u,$$

où depvet représente les dépenses en vêtements d'un ménage et rev son revenu, les variables explicatives seront fortement, mais imparfaitement, corrélées l'une avec l'autre.

3. Quelques cas particuliers où l'hypothèse 5.4 est violée:

- Une variable est la multiple d'une autre variable (si, par exemple, les unités de mesure sont différentes). En modifiant le modèle précédent de la manière suivante:

$$\log(\text{depvet}) = \beta_0 + \beta_1 \cdot \log(\text{rev}) + \beta_2 \cdot \log(\text{rev}^2) + u,$$

l'hypothèse 5.4 n'est plus respectée car $\log(\text{rev}^2) = 2 \times \log(\text{rev})$.

- Une variable est la somme de deux autres variables. Dans la régression suivante:

$$\text{depvet} = \beta_0 + \beta_1 \cdot \text{revh} + \beta_2 \cdot \text{revf} + \beta_3 \cdot \text{revtot} + u,$$

la variable revh représente le revenu de l'homme dans un couple, revf le revenu de la femme, revtot le revenu total du ménage, égal à la somme des revenus de l'homme et de la femme.

- La taille de l'échantillon est trop petite: si $N < K + 1$, la condition (5.1) sera toujours satisfaite pour un ensemble particulier de scalaires c_0, \dots, c_K .

Les démonstrations qui vont suivre reposent sur le résultat suivant. Comme nous l'avons vu, l'estimateur $\hat{\beta}_k$ peut s'écrire de la manière suivante:

$$\hat{\beta}_k = \frac{\sum_{i=1}^N \hat{r}_{ki} y_i}{\sum_{i=1}^N \hat{r}_{ki}^2} \quad \text{pour } k = 0, \dots, K.$$

En remplaçant y_i par $\beta_0 + \beta_1 x_{1i} + \dots + \beta_K x_{Ki} + u_i$, et en développant l'expression, on obtient:

$$\begin{aligned} \hat{\beta}_k &= \beta_0 \frac{\sum_{i=1}^N \hat{r}_{ki}}{\sum_{i=1}^N \hat{r}_{ki}^2} + \beta_1 \frac{\sum_{i=1}^N \hat{r}_{ki} x_{1i}}{\sum_{i=1}^N \hat{r}_{ki}^2} + \dots \\ &+ \beta_k \frac{\sum_{i=1}^N \hat{r}_{ki} x_{ki}}{\sum_{i=1}^N \hat{r}_{ki}^2} + \dots + \beta_K \frac{\sum_{i=1}^N \hat{r}_{ki} x_{Ki}}{\sum_{i=1}^N \hat{r}_{ki}^2} + \frac{\sum_{i=1}^N \hat{r}_{ki} u_i}{\sum_{i=1}^N \hat{r}_{ki}^2}. \end{aligned}$$

En utilisant les Propriétés 4.1 et 4.2, on montre que $\sum_{i=1}^N \hat{r}_{ki} = 0$ et $\sum_{i=1}^N \hat{r}_{ki} x_{ji} = 0$ pour $j \neq k$, et on obtient:

$$\hat{\beta}_k = \beta_k \frac{\sum_{i=1}^N \hat{r}_{ki} x_{ki}}{\sum_{i=1}^N \hat{r}_{ki}^2} + \frac{\sum_{i=1}^N \hat{r}_{ki} u_i}{\sum_{i=1}^N \hat{r}_{ki}^2}.$$

De plus, puisque $x_{ki} = \hat{x}_{ki} + \hat{r}_{ki}$, on obtient:

$$\begin{aligned} \hat{\beta}_k &= \beta_k \left(\frac{\sum_{i=1}^N \hat{r}_{ki} \hat{x}_{ki}}{\sum_{i=1}^N \hat{r}_{ki}^2} + \frac{\sum_{i=1}^N \hat{r}_{ki}^2}{\sum_{i=1}^N \hat{r}_{ki}^2} \right) + \frac{\sum_{i=1}^N \hat{r}_{ki} u_i}{\sum_{i=1}^N \hat{r}_{ki}^2} \\ &= \beta_k + \frac{\sum_{i=1}^N \hat{r}_{ki} u_i}{\sum_{i=1}^N \hat{r}_{ki}^2} \end{aligned}$$

Cette formule exprime le relation existantes entre les estimateurs, d'une part, et les paramètres et les résidus, d'autre part.

Théorème 5.1: *Sous les hypothèses 5.1–5.4, les estimateurs des MCO sont non biaisés:*

$$\mathbb{E}(\hat{\beta}_k | \mathbf{X}) = \beta_k \quad \text{où } k = 0, \dots, K$$

où $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N)$ désigne l'ensemble des variables explicatives dans les données.

Démonstration. (Dans la démonstration, les espérances sont conditionnelles aux valeurs $\mathbf{x}_1, \dots, \mathbf{x}_N$ de l'échantillon.) Les estimateurs et les paramètres dans la population sont liés par la formule suivante:

$$\hat{\beta}_k = \beta_k + \frac{\sum_{i=1}^N \hat{r}_{ki} u_i}{\sum_{i=1}^N \hat{r}_{ki}^2}.$$

Si l'on calcule l'espérance des membres de droite et de gauche pour les valeurs des variables explicatives données, et que l'on simplifie, on obtient:

$$\begin{aligned} \mathbb{E}(\hat{\beta}_k | \mathbf{X}) &= \beta_k + \mathbb{E} \left(\frac{\sum_{i=1}^N \hat{r}_{ki} u_i}{\sum_{i=1}^N \hat{r}_{ki}^2} \middle| \mathbf{X} \right) \\ &= \beta_k + \frac{\sum_{i=1}^N \hat{r}_{ki} \mathbb{E}(u_i | \mathbf{X})}{\sum_{i=1}^N \hat{r}_{ki}^2} = \beta_k \end{aligned}$$

puisque $\mathbb{E}(u_i | \mathbf{X}) = 0$. \square

Remarques:

1. **Inclusion de variables non-pertinentes.** On suppose que le modèle suivant satisfait les hypothèses 5.1–5.4:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + u. \quad (5.2)$$

En vertu du théorème 5.1, les estimateurs des MCO des paramètres sont non biaisés. Toutefois, l'économètre incorpore une variable supplémentaire et estime la relation suivante:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + u. \quad (5.3)$$

Dans ce cas, l'inclusion de la variable non-pertinente n'aura pas d'implications en termes de biais: les estimateurs des MCO de β_0 , β_1 et β_2 seront toujours non biaisés (mais leur variance pourra être affectée par l'introduction de la variable supplémentaire). En effet, si le modèle (5.2) satisfait les hypothèses 5.1–5.4, le modèle (5.3) satisfait également ces hypothèses. Le paramètre β_3 dans l'équation (5.3) sera, en principe, égal à zéro auquel cas les deux modèles sont identiques. Cependant, cela pourrait ne pas être le cas. Les deux modèles représenteraient alors, tous les deux, une relation causale correcte entre les variables.

2. **Exclusion de variables pertinentes.** On suppose que le modèle suivant satisfait les hypothèses 5.1–5.4:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + u,$$

alors que le modèle suivant, où une variable explicative a priori pertinente a été omise, est estimé:

$$y = \beta_0 + \beta_1 \cdot x_1 + u.$$

Dans ce cas, les estimateurs seront, en général, biaisés. On parle d'un problème de mauvaise spécification. Pour le modèle mal spécifié, l'estimateur de β_1 est donné par:

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^N (x_{1i} - \bar{x}) y_i}{\sum_{i=1}^N (x_{1i} - \bar{x})^2}.$$

Or:

$$y_i = \beta_0 + \beta_1 \cdot x_{1i} + \beta_2 \cdot x_{2i} + u_i.$$

L'introduction de cette équation dans la précédente donne:

$$\begin{aligned} \tilde{\beta}_1 &= \beta_0 \frac{\sum_{i=1}^N (x_{1i} - \bar{x}_1)}{\sum_{i=1}^N (x_{1i} - \bar{x}_1)^2} + \beta_1 \frac{\sum_{i=1}^N (x_{1i} - \bar{x}_1) x_{1i}}{\sum_{i=1}^N (x_{1i} - \bar{x}_1)^2} \\ &\quad + \beta_2 \frac{\sum_{i=1}^N (x_{1i} - \bar{x}_1) x_{2i}}{\sum_{i=1}^N (x_{1i} - \bar{x}_1)^2} + \frac{\sum_{i=1}^N (x_{1i} - \bar{x}_1) u_i}{\sum_{i=1}^N (x_{1i} - \bar{x}_1)^2}. \end{aligned}$$

En utilisant les propriétés $\sum_{i=1}^N (x_{1i} - \bar{x}_1) = 0$ et $\sum_{i=1}^N (x_{1i} - \bar{x}_1) \cdot x_{1i} = \sum_{i=1}^N (x_{1i} - \bar{x}_1)^2$, on obtient:

$$\tilde{\beta}_1 = \beta_1 + \beta_2 \frac{\sum_{i=1}^N (x_i - \bar{x}) x_{2i}}{\sum_{i=1}^N (x_i - \bar{x})^2} + \frac{\sum_{i=1}^N (x_{1i} - \bar{x}_1) u_i}{\sum_{i=1}^N (x_{1i} - \bar{x}_1)^2}$$

En calculant l'espérance des membres de droite et de gauche, on obtient:

$$\mathbf{E}(\tilde{\beta}_1 | \mathbf{X}) = \beta_1 + \beta_2 \frac{\sum_{i=1}^N (x_{1i} - \bar{x}_1) x_{2i}}{\sum_{i=1}^N (x_{1i} - \bar{x}_1)^2}.$$

Le dernier terme peut s'interpréter comme étant la pente de la régression de x_2 sur x_1 à savoir

$$x_2 = \hat{\delta}_0 + \hat{\delta}_1 x_1.$$

Donc:

$$\mathbf{E}(\tilde{\beta}_1 | \mathbf{X}) = \beta_1 + \beta_2 \hat{\delta}_1$$

Les estimateurs seront non biaisés si x_1 et x_2 sont non corrélés dans l'échantillon ($\hat{\delta}_1 = 0$) et/ou si la variable x_2 n'a pas d'effet sur y ($\beta_2 = 0$). Dans les autres cas, les estimateurs seront biaisés. Le biais est positif

$$\mathbf{E}(\tilde{\beta}_1 | \mathbf{X}) - \beta_1 > 0$$

si le signe de β_2 est le même que le signe de $\hat{\delta}_1$, et négatif

$$\mathbf{E}(\tilde{\beta}_1 | \mathbf{X}) - \beta_1 < 0$$

si le signe de β_2 et le signe de $\hat{\delta}_1$ sont différents.

Exemple 5.1: Dans l'équation de Mincer, le modèle qui satisfait les hypothèses 5.1–5.4 est supposé être le suivant:

$$\log(\text{sal}) = \beta_0 + \beta_1 \cdot \text{edu} + \beta_2 \cdot \text{habil} + u$$

où la variable **habil** représente les habiletés de l'individu (sous cette terminologie vague, peuvent être regroupées l'intelligence, la santé, la force physique, l'adresse, par exemple). Toutefois, puisque cette variable n'est généralement pas observable, le modèle suivant est estimé:

$$\log(\text{sal}) = \beta_0 + \beta_1 \cdot \text{edu} + u.$$

En utilisant les données précédentes, on obtient:

$$\widehat{\log(\text{sal})} = 0,584 + 0,083 \cdot \text{edu}$$

où la variable **habil** a été omise, l'évaluation du taux de rendement de l'éducation sera vraisemblablement surestimée. En effet, il est probable que l'habileté et le niveau d'éducation de l'individu sont corrélés.

Dans le cas général, avec un nombre arbitraire de variables explicatives, la direction des biais indiquée par le résultat précédent ne sera plus exacte. Ce résultat peut toutefois servir de première indication.

5.2 La variance des estimateurs des MCO

Pour calculer la variance des estimateurs, on a besoin d'une hypothèse supplémentaire.

Hypothèse 5.5: Le terme d'erreur est homoscedastique. En d'autres termes,

$$\text{Var}(u|x_1, \dots, x_K) = \sigma^2,$$

où σ^2 est une constante positive.

Remarque: Les hypothèses 5.1–5.5 sont appelées **hypothèses de Gauss-Markov**.

Théorème 5.2: *Sous les hypothèses 5.1–5.5, la variance de l'estimateur des MCO est donnée par*

$$\text{Var}(\hat{\beta}_k | \mathbf{X}) = \frac{\sigma^2}{\sum_{i=1}^N \hat{r}_{ki}^2} \quad \text{pour } k = 0, 1, \dots, K,$$

où \hat{r}_k pour $k = 1, \dots, K$ est le résidu obtenu par la régression (avec constante) de la variable x_k sur l'ensemble des autres variable explicatives du modèle, à savoir, $x_1, x_2, \dots, x_{k-1}, x_{k+1}, \dots, x_K$, et \hat{r}_k pour $k = 0$ est le résidu obtenu par la régression (sans constante) de la constante unitaire sur l'ensemble des variable explicatives du modèle, à savoir, x_1, x_2, \dots, x_K .

Démonstration: (Dans la démonstration, les espérances sont conditionnelles aux valeurs $\mathbf{x}_1, \dots, \mathbf{x}_N$ de l'échantillon.) Les estimateurs et les paramètres dans la population sont liés par la formule suivante:

$$\hat{\beta}_k = \beta_k + \frac{\sum_{i=1}^N \hat{r}_{ki} u_i}{\sum_{i=1}^N \hat{r}_{ki}^2}$$

Si l'on prend la variance des membres de gauche et de droite de cette expression et que l'on simplifie, on obtient:

$$\text{Var}(\hat{\beta}_k | \mathbf{X}) = \text{Var} \left(\frac{\sum_{i=1}^N \hat{r}_{ki} u_i}{\sum_{i=1}^N \hat{r}_{ki}^2} \middle| \mathbf{X} \right) = \frac{\sum_{i=1}^N \hat{r}_{ki}^2 \text{Var}(u_i | \mathbf{X})}{\left(\sum_{i=1}^N \hat{r}_{ki}^2 \right)^2}.$$

En utilisant l'hypothèse 5.5, on obtient:

$$\text{Var}(\hat{\beta}_k | \mathbf{X}) = \frac{\sum_{i=1}^N \hat{r}_{ki}^2 \sigma^2}{\left(\sum_{i=1}^N \hat{r}_{ki}^2 \right)^2} = \frac{\sigma^2 \sum_{i=1}^N \hat{r}_{ki}^2}{\left(\sum_{i=1}^N \hat{r}_{ki}^2 \right)^2} = \frac{\sigma^2}{\sum_{i=1}^N \hat{r}_{ki}^2}.$$

□

Remarques:

1. En utilisant la définition du $R_k^2 = 1 - \text{SCR}_k / \text{SCT}_k$, et donc,

$$R_k^2 = 1 - \frac{\sum_{i=1}^N \hat{r}_{ki}^2}{\sum_{i=1}^N (x_{ki} - \bar{x}_k)^2},$$

pour $k = 1, \dots, K$, la variance de l'estimateur d'une pente peut également s'écrire

$$\text{Var}(\hat{\beta}_k | \mathbf{X}) = \frac{\sigma^2}{(1 - R_k^2) \sum_{i=1}^N (x_{ki} - \bar{x}_k)^2}.$$

Les composantes de la variance des estimateurs des moindres carrés ordinaires sont les suivantes:

- La variance du terme aléatoire σ^2 : les estimations sont moins précises quand la variance du terme aléatoire est grande.
- La variance des variables explicatives et le nombre d'observations (le problème de micronumérosité) $\sum_{i=1}^N (x_{ki} - \bar{x}_k)^2$: les estimations sont plus précises quand le nombre d'observations est grand et quand la variance des variables explicatives est grande;
- La collinéarité entre les variables explicatives (problème de multicollinéarité) $(1 - R_k^2)$: les estimateurs sont plus précis quand les variables explicatives sont peu corrélées entre elles.

2. Comme dans le cas du modèle de régression simple, la variance du terme aléatoire doit être estimée.

Théorème 5.3: *Sous les hypothèses 5.1–5.5, l'estimateur des MCO de la variance des termes d'erreur est non biaisé:*

$$\mathbb{E}(\hat{\sigma}^2) = \mathbb{E} \left(\frac{\sum_{i=1}^N \hat{u}_i^2}{N - K - 1} \right) = \sigma^2.$$

Démonstration: Voir Hayashi (2000). \square

L'estimateur de la variance de l'estimateur des paramètres est donné par:

$$\widehat{\text{Var}}(\hat{\beta}_k | \mathbf{X}) = \frac{\hat{\sigma}^2}{(1 - R_k^2) \sum_{i=1}^N (x_{ki} - \bar{x}_k)^2},$$

tandis que l'estimateur de l'erreur-type de estimateurs des MCO est:

$$\text{se}(\hat{\beta}_k | \mathbf{X}) = \frac{\hat{\sigma}}{\sqrt{(1 - R_k^2) \sum_{i=1}^N (x_{ki} - \bar{x}_k)^2}}.$$

5.3 L'efficacité des estimateurs des MCO

Il existe une infinité d'estimateurs sans biais, mais l'estimateur des MCO a une propriété très attractive.

Théorème 5.4 (Théorème de Gauss-Markov): *Sous les hypothèses 5.1–5.5, les estimateurs des MCO sont les estimateurs les plus efficaces dans la classe des estimateurs linéaires non biaisés de $\beta_0, \beta_1, \dots, \beta_K$.*

Démonstration: Voir Hayashi (2000). \square

Remarques:

1. Un estimateur est linéaire s'il peut s'écrire sous la forme d'une combinaison linéaire des variables dépendantes:

$$\tilde{\beta}_k = \sum_{i=1}^N w_{ik} y_i$$

où les w_{ik} sont des pondérations qui peuvent dépendre de toutes les variables indépendantes. Or, l'estimateur des MCO est linéaire puisque

$$\hat{\beta}_1 = \frac{\sum_{i=1}^N \hat{r}_{1i} y_i}{\sum_{i=1}^N \hat{r}_{1i}^2}$$

2. Un estimateur est le plus efficace si sa variance est plus petite que celle de tous les estimateurs disponibles dans une classe donnée.
3. Les estimateurs qui ont la propriété décrite dans le Théorème 5.4 sont dits "BLUE" (pour "Best Linear Unbiased Estimator").

5.4 La distribution des estimateurs des MCO

Pour le moment, on ne connaît que deux moments de la distribution des estimateurs (l'espérance et la variance). Or, pour s'assurer de la précision des estimations, et pour faire des tests d'hypothèses, il est nécessaire de connaître l'intégralité de la distribution des estimateurs. Pour cela, on utilisera donc l'hypothèse suivante.

Hypothèse 5.6: Les termes aléatoires u sont distribués conditionnellement à \mathbf{x} selon une loi normale.

En d'autres termes, si on suppose que $E(u|x) = 0$, la fonction de densité du terme d'erreur u peut s'écrire:

$$f(u|\mathbf{x}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{u^2}{2\sigma^2}\right).$$

Remarques:

1. La loi normale est aussi appelée loi de Gauss. Sa fonction de densité a une forme en cloche caractéristique.
2. La normalité est justifiée par le Théorème de la limite centrale (voir section 2), selon lequel la somme (divisée par \sqrt{N}) d'un très grand nombre N de variables aléatoires d'espérance nulle et de variance donnée suit approximativement une loi normale.
3. Si l'on ajoute les hypothèses 5.2 et 5.5 à l'hypothèse 5.6, on caractérise complètement la distribution du terme d'erreur:

$$u|\mathbf{x} \rightsquigarrow \mathcal{N}(0, \sigma^2)$$

et la distribution de la variable expliquée:

$$y|\mathbf{x} \rightsquigarrow \mathcal{N}(\beta_0 + \beta_1 \cdot x_1 + \dots + \beta_K \cdot x_K, \sigma^2)$$

4. Les hypothèses 5.1–5.6 constituent les **hypothèses classiques** du modèle linéaire, et elles impliquent une forme plus forte du théorème de Gauss-Markov.

Théorème 5.5 (Cramer-Rao): *Sous les hypothèses 5.1–5.6, les estimateurs des MCO sont les plus efficaces dans la classe des estimateurs non biaisés de $\beta_0, \beta_1, \dots, \beta_K$.*

Remarque: Les estimateurs des MCO sont donc les meilleurs y compris dans la classe des estimateurs non linéaires. Les variances de $\widehat{\beta}_0, \widehat{\beta}_1, \dots, \widehat{\beta}_K$ sont donc plus petites (ou du moins pas plus grande) que celles de tout autre estimateur non biaisé. On dit parfois que le estimateurs des MCO sont BUE ("Best Unbiased Estimators").

Théorème 5.6: *Sous les hypothèses 5.1–5.6, la distribution des estimateurs des MCO est normale:*

$$\widehat{\beta}_k | \mathbf{X} \rightsquigarrow \mathcal{N}(\beta_k, \text{Var}(\widehat{\beta}_k | \mathbf{X}))$$

où $\text{Var}(\widehat{\beta}_k | \mathbf{X})$ est donné par $\sigma^2 / [\sum_{i=1}^N \widehat{r}_{ki}^2]$. En particulier,

$$\frac{\widehat{\beta}_k - \beta_k}{\text{sd}(\widehat{\beta}_k)} \Big| \mathbf{X} \rightsquigarrow \mathcal{N}(0, 1),$$

où $\text{sd}(\widehat{\beta}_k) = \sqrt{\text{Var}(\widehat{\beta}_k | \mathbf{X})}$.

Démonstration: L'estimateur des MCO, conditionnellement aux variables explicatives, est une combinaison linéaire des termes d'erreur car il peut s'écrire sous la forme suivante:

$$\widehat{\beta}_k = \beta_k + \frac{\sum_{i=1}^N \widehat{r}_{ki} u_i}{\sum_{i=1}^N \widehat{r}_{ki}^2}$$

où \widehat{r}_{ki} est une fonction des variables explicatives. Or, les termes aléatoires sont distribués selon une loi normale. En vertu de la Propriété 2.18, la combinaison linéaire de variables aléatoires distribuées selon des lois normales sera distribuée elle-même selon une loi normale. \square

Remarque: Avec ce théorème, la distribution des estimateurs est totalement caractérisée. Toutefois, cette distribution repose sur la variance du terme aléatoire qui n'est pas connue.

Chapter 6

Inférence statistique

L'inférence statistique consiste à induire les caractéristiques inconnues d'une population à partir d'un échantillon issu de cette population. Les caractéristiques de l'échantillon, une fois connues, reflètent avec une certaine marge d'erreur possible celles de la population.

6.1 Test d'une seule restriction: le t-test

Le principe d'un test d'hypothèse consiste à construire une statistique, à partir des données, dont la distribution dépendra de la véracité de l'hypothèse testée, et de confronter la valeur de cette statistique calculée à la distribution en question. De cette confrontation, on conclura que l'hypothèse testée doit être rejetée ou non.

On considère d'abord des tests d'une seule restriction, où l'hypothèse nulle à tester a la forme suivante:

$$H_0 : \beta_k = b_k$$

où b_k est une certaine constante. On distingue des tests unilatéraux et des tests bilatéraux. Dans le premier cas, l'hypothèse alternative (qui est vraie si l'hypothèse nulle est fausse) a l'une des formes suivantes:

$$H_1 : \beta_k > b_k \quad \text{ou} \quad H_1 : \beta_k < b_k,$$

et dans le second cas l'hypothèse a la forme suivante:

$$H_1 : \beta_k \neq b_k.$$

Le t-test (ou test de Student) s'appuie sur le résultat suivant.

Théorème 6.1: *Sous les hypothèses 5.1–5.6, et sous l'hypothèse nulle, la statistique*

$$\frac{\widehat{\beta}_k - b_k}{\text{se}(\widehat{\beta}_k)} \rightsquigarrow \mathcal{T}(N - K - 1)$$

où N est le nombre d'observations, et K le nombre de variables explicatives, et

$$\text{se}(\widehat{\beta}_k) = \frac{\widehat{\sigma}}{\sqrt{(1 - R_k^2) \sum_{i=1}^N (x_{ki} - \bar{x}_k)^2}}$$

est l'erreur-type de l'estimateur $\widehat{\beta}_k$.

Démonstration: On ne présentera pas une démonstration complète de ce résultat (pour cela, voir Hayashi (2000)). La démonstration s'appuie sur la définition de la loi de Student donnée dans la section 2. Le terme $\widehat{\beta}_k - b_k / \text{sd}(\widehat{\beta}_k)$ a une distribution normale centrée réduite et le terme $\widehat{\sigma}^2 / \sigma^2$ a une distribution du Chi-carré à $N - K - 1$ degrés de liberté. \square

Test unilatéral.

Pour discuter les tests unilatéraux, on considère un test unilatéral à droite (un test unilatéral à gauche serait le symétrique) et on teste l'hypothèse nulle suivante:

$$H_0 : \beta_k = 0.$$

Cette hypothèse constitue une forme particulière du test qui est très importante dans la pratique. Dans ce cas, l'hypothèse alternative s'écrit:

$$H_1 : \beta_k > 0.$$

Le t-test ou test de Student est basé sur le t-ratio, statistique définie de la manière suivante:

$$t = \frac{\widehat{\beta}_k}{\text{se}(\widehat{\beta}_k)}.$$

De manière intuitive, la tendance à rejeter l'hypothèse nulle ($H_0 : \beta_k = 0$) sera d'autant plus élevée que la valeur de cette statistique est grande, donc que la valeur de l'estimateur est grande et son erreur-type est petite. Pour un test unilatéral à droite, la **règle de décision** est donnée par

si $t > c$, H_0 est rejetée,

si $t \leq c$, H_0 n'est pas rejetée,

où c est un **seuil critique** qui est lui-même déterminée par le **seuil de signification** choisi par l'économètre. Le seuil de signification est la probabilité de rejeter l'hypothèse nulle dans le cas où cette dernière est vraie. Si l'on choisit un seuil de signification égal à α , la valeur critique c est le quantile de la loi de Student telle que la probabilité d'obtenir une valeur supérieure à c est égale à α , c'est-à-dire

$$\Pr(t > c | H_0 \text{ est vraie}) = \alpha.$$

Le seuil de signification sera souvent choisi de manière arbitraire égal à 10%, 5% ou 1%. Intuitivement, si le t-ratio est supérieure au seuil c , la vraisemblance que la distribution du t-ratio soit effectivement une loi de Student est faible.

Exemple 6.1: On reconsidère le modèle d'investissement en capital humain et on prend la spécification suivante:

$$\log(\text{sal}) = \beta_0 + \beta_1 \cdot \text{edu} + \beta_2 \cdot \text{exp} + \beta_3 \cdot \text{anc} + u.$$

On désire tester l'hypothèse nulle $H_0 : \beta_1 = 0$ contre l'hypothèse alternative $H_1 : \beta_1 > 0$. La régression du logarithme du salaire sur l'éducation, l'expérience et l'ancienneté donne les résultats suivants:

$$\widehat{\log(\text{sal})} = 0,284 + 0,092 \text{ edu} + 0,0041 \text{ exp} + 0,022 \text{ anc}$$

(0,104)	(0,007)	(0,0017)	(0,003)
---------	---------	----------	---------

avec $N = 526$ et $R^2 = 0,316$. Le t-ratio pour cette hypothèse nulle est égal à

$$t = \frac{0,092}{0,007} \approx 13,143.$$

Or, sachant que le nombre de degrés de liberté est égal à $522 = 526 - 4$, la valeur critique pour un test unilatéral avec un seuil de signification de 5% est de 1,645. L'hypothèse nulle est donc rejetée.

Test bilatéral.

Dans le cas d'un test bilatéral, l'hypothèse alternative s'écrit:

$$H_1 : \beta_k \neq 0.$$

L'hypothèse nulle peut donc être rejetée si le t-ratio est positif et plus grand que la valeur critique ou s'il est négatif et plus petit que la valeur critique. Donc, pour un test bilatéral, la règle de rejet est donnée par

$$\begin{aligned} \text{si } |t| > c, & \quad H_0 \text{ est rejetée,} \\ \text{si } |t| \leq c, & \quad H_0 \text{ n'est pas rejetée,} \end{aligned}$$

où c est le seuil critique. Le seuil critique c est également déterminé par le seuil de signification α . Formellement,

$$\Pr(|t| > c | H_0 \text{ est vraie}) = \alpha.$$

Puisque la loi de Student est symétrique, $\Pr(|t| > c | H_0 \text{ est vraie}) = 2 \times \Pr(t > c | H_0 \text{ est vraie})$.

Tests généraux

Plus généralement, le t-test permet de tester des hypothèses nulles telles que:

$$H_0 : \beta_k = b_k.$$

où b_k est une valeur quelconque que peut prendre β_k . Le t-ratio dans ce contexte s'écrit:

$$t = \frac{\widehat{\beta}_k - b_k}{\text{se}(\widehat{\beta}_k)}.$$

La procédure de test est alors exactement la même que celles qui viennent d'être décrites ci-dessus. En particulier, il est possible de faire des tests unilatéraux ou bilatéraux.

Exemple 6.2: J. Wooldridge (2010) s'intéresse à la relation entre le nombre de délits commis sur les universités et collèges américains et le nombre d'étudiants inscrits dans ces établissements. Les données viennent des FBI's Uniform Crime Reports pour 1992. Le nombre moyen de délits dans l'échantillon est de 394 et le nombre d'étudiants est de 16.076. La relation estimée est la suivante:

$$\widehat{\log(\text{nbdelits})} = -6,63 + 1,27 \log(\text{nbetud})$$

$$(1,03) \quad (0,11)$$

avec $N = 97$ et $R^2 = 0,585$, où `nbdelits` est le nombre de délits dans l'établissement et `nbetud` est le nombre d'étudiants. La relation est naturellement positive: le nombre de délits est plus élevé sur les grands campus où le nombre d'étudiants est également plus élevé. Plus intéressant est de savoir si le nombre de délits augmente proportionnellement ou non avec le nombre d'étudiants sur le campus. Puisque le modèle est sous forme logarithmique, l'hypothèse nulle correspondant à cette question est $H_0 : \beta_2 = 1$. En effet, si cette restriction est satisfaite, l'élasticité du nombre de délits par rapport au nombre d'étudiants sera égale à un. Le t-ratio est alors égal à

$$t = \frac{1,27 - 1}{0,11} \approx 2.4545.$$

Le nombre de degrés de liberté est égal à $95 = 97 - 2$ et la valeur critique pour un test bilatéral avec un seuil de signification de 5% est de 1,960. L'hypothèse nulle est donc rejetée.

Tests impliquant plusieurs paramètres.

Le test de Student permet également de tester une restriction unique portant sur une combinaison linéaire de paramètres. Le plus simple pour cela est de reparamétriser le modèle.

Exemple 6.3: Dans le modèle suivant, les dépenses consacrées aux vêtements des enfants (qui peuvent être vus comme un indicateur du bien-être des enfants dans le ménage) sont supposées dépendre du revenu de la mère et de celui du père:

$$\text{depvetenf} = \beta_0 + \beta_1 \cdot \text{revm} + \beta_2 \cdot \text{revp} + u.$$

où depvetenf désigne les dépenses de vêtements pour les enfants, revm le revenu de la mère, et revp le revenu du père. Toutefois, dans certains modèles simples de comportement, la répartition du revenu total du ménage entre les parents est supposée ne pas affecter le comportement du ménage, auquel cas les dépenses ne seront pas affectées par une redistribution des ressources entre les parents. On désire tester cette condition. L'hypothèse nulle s'écrit:

$$H_0 : \beta_1 = \beta_2.$$

Pour tester cette condition, on peut procéder en calculant la différence entre ces estimateurs et la variance de cette différence, et en dérivant le t-ratio correspondant. Cette procédure est correcte. Toutefois il est plus simple de procéder à un changement de paramétrisation du modèle. On définit un nouveau paramètre de la manière suivante:

$$\theta = \beta_1 - \beta_2 \quad \text{de sorte que} \quad \beta_1 = \theta + \beta_2.$$

L'hypothèse nulle peut naturellement s'écrire en utilisant le nouveau paramètre:

$$H_0 : \theta = 0.$$

En substituant ce nouveau paramètre dans l'équation initiale, on obtient:

$$\begin{aligned} \log(\text{depvetenf}) &= \beta_0 + (\theta + \beta_2) \cdot \text{revm} + \beta_2 \cdot \text{revp} + u \\ &= \beta_0 + \theta \cdot \text{revm} + \beta_2 \cdot \text{revtot} + u \end{aligned}$$

où $\text{revtot} = \text{revm} + \text{revp}$ est le revenu total du ménage. Ce modèle reparamétré peut être estimé. Il suffit alors de tester si le paramètre θ est égal à zéro en utilisant un t-test traditionnel.

Exemple 6.4: La théorie économique enseigne que les équations de demande doivent être homogènes de degré zéro. Cela signifie que si tous les prix et le revenu du ménage sont multipliés par une même constante positive, la quantité de biens demandée par le ménage ne se modifiera pas. En d'autres termes, les agents ne sont pas affectés par une illusion monétaire. Si l'on veut tester cette propriété, il faut construire un modèle économétrique et dériver les implications de la contrainte d'homogénéité sur les paramètres de ce modèle. On choisit la forme fonctionnelle suivante pour la demande pour le bien 1:

$$\log(\text{qbien1}) = \beta_0 + \beta_1 \cdot \log(\text{pbien1}) + \beta_2 \cdot \log(\text{pbien2}) + \beta_3 \cdot \log(\text{rev}) + u.$$

La propriété d'homogénéité implique, en effet, une restriction simple sur les paramètres. La propriété d'homogénéité signifie que l'équation suivante

$$\beta_0 + \beta_1 \cdot \log(k \times \text{pbien1}) + \beta_2 \cdot \log(k \times \text{pbien2}) + \beta_3 \cdot \log(k \times \text{rev})$$

doit être égale à

$$\beta_0 + \beta_1 \cdot \log(\text{pbien1}) + \beta_2 \cdot \log(\text{pbien2}) + \beta_3 \cdot \log(\text{rev})$$

quelle que soit la valeur de k , ce qui sera le cas si et seulement si

$$\beta_1 + \beta_2 + \beta_3 = 0.$$

En d'autres termes, l'équation de demande sera homogène de degré zéro si cette condition est satisfaite. Pour tester cette condition, on définit donc comme précédemment un paramètre $\theta = \beta_1 + \beta_2 + \beta_3$ et on procède comme dans l'exemple précédent pour faire le test.

Tests et intervalles de confiance

Les intervalles de confiance permettent également de faire des tests. En vertu du théorème 6.1, on sait que les inégalités suivantes

$$t_{\alpha/2} < \frac{\widehat{\beta}_k - \beta_k}{\text{se}(\widehat{\beta}_k)} < t_{1-\alpha/2},$$

où $t_{\alpha/2}$ est le $(\alpha/2 \times 100)$ ème percentile de la loi de Student à $N - K - 1$ degrés de liberté, et $t_{1-\alpha/2}$ est le $((1 - \alpha/2) \times 100)$ ème percentile, est satisfaite dans $(1 - \alpha) \times 100$ pourcents des cas. Donc, en transformant ces inégalités, l'intervalle de confiance est donné par:

$$\widehat{\beta}_k - t_{\alpha/2} \times \text{se}(\widehat{\beta}_k) < \beta_k < \widehat{\beta}_k + t_{1-\alpha/2} \times \text{se}(\widehat{\beta}_k).$$

En d'autres termes, le paramètre β_k est compris entre ces deux bornes dans $(1 - \alpha) \times 100$ pourcents des cas.

Exemple 6.5: On considère la régression du logarithme du prix des maisons sur diverses variables qui décrivent cette maison. Les données proviennent de l'enregistrement des ventes de maisons à Waltham, Massachusetts en 1990, collectées par J. Wooldridge (2010). L'équation estimée est donnée par:

$$\log(\widehat{\text{prixmais}}) = 7,46 + 0,634 \log(\text{surfh}) - 0,066 \text{ nch} + 0,158 \text{ nsb}$$

$$(1,15) \quad (0,184) \quad (0,059) \quad (0,075)$$

où est le $\widehat{\text{prixmais}}$ est le prix de la maison, surfh la surface habitable de la maison (en pieds carrés), nch le nombre de chambres, et nsb le nombre de salles de bains, avec $N = 19$, et $R^2 = 0,806$. Cette équation constitue **un modèle de prix hédonique**, dans laquelle le prix d'un bien (ici une maison) est expliqué par les caractéristiques de ce bien. Les équations de prix hédonique, dont les paramètres représentent le prix des caractéristiques du bien, sont utilisées notamment pour construire des indices de prix. De cette estimation, on peut calculer l'intervalle de confiance à 95% de l'estimation de l'effet de la superficie. Le nombre de degrés de liberté est égal à 15 et le 97,5ème percentile à 2,131. Donc, l'intervalle de confiance est:

$$0,634 - 2,131 \times 0,184 < \beta_k < 0,634 + 2,131 \times 0,184$$

ou

$$0,24190 < \beta_k < 1,0261.$$

Les intervalles de confiance sont liés aux tests de Student. Ainsi, puisque la valeur 1 n'appartient pas à l'intervalle de confiance calculé ci-dessus, on conclut que l'hypothèse $H_0 : \beta_k = 1$ n'est pas rejetée.

Tests et probabilité critique

Le résultat d'un test dépend du seuil de signification dont le choix est arbitraire. La probabilité critique (ou p-value) est le plus petit seuil de signification auquel l'hypothèse nulle serait rejetée. Formellement, dans le cas où

$$H_0 : \beta_k = b_k \quad \text{et} \quad H_1 : \beta_k \neq b_k,$$

la probabilité critique est donnée par

$$p = \min\{\alpha \text{ tel que } |t| > c_{\alpha/2}\}$$

où

$$t = \frac{\widehat{\beta}_k - b_k}{\text{se}(\widehat{\beta}_k)},$$

et $c_{\alpha/2}$ est le $(\alpha/2)$ ème quantile de la loi de Student à $N - K - 1$ degrés de liberté. Si $p < 0.05$ (resp. $p < 0.01$, $p < 0.10$) alors l'hypothèse nulle est rejetée au seuil de 5% (resp. 1%, 10%).

6.2 Tests d'une multiplicité de restrictions: le F-test

Dans certaines circonstances, il est désirable de tester un ensemble de restrictions au lieu d'une seule. On considère le modèle général suivant:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_K x_K + u,$$

et on suppose que l'hypothèse nulle est constituée d'un ensemble des restrictions linéaires. Si (sans perte de généralité) les restrictions portent sur les Q derniers paramètres, l'hypothèse nulle peut écrire sous la forme suivante:

$$H_0 : \beta_{K-Q+1} = b_{K-Q+1}, \beta_{K-Q+2} = b_{K-Q+2}, \dots, \beta_K = b_K,$$

où $b_{K-Q+1}, b_{K-Q+2}, \dots, b_K$ sont des constantes. D'abord, on définit le modèle restreint comme étant le modèle dans lequel les restrictions à tester sont imposées. Ainsi, si les conditions sont imposées, le modèle contraint est

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_{K-Q} x_{K-Q} + b_{K-Q+1} x_{K-Q+1} + \dots + b_K x_K + u.$$

La statistique de Fisher se construit à partir de la somme des carrés résiduels du modèle restreint et du modèle non restreint. Le F-test (ou test de Fisher) s'appuie alors sur le résultat suivant.

Théorème 6.2: Soit SCR_{NR} la somme des carrés résiduels du modèle non restreint, et SCR_R la somme des carrés résiduels du modèle restreint. Sous les hypothèses 5.1–5.6, et sous l'hypothèse nulle

$$F = \frac{(SCR_R - SCR_{NR})/Q}{(SCR_{NR})/(N - K - 1)} \rightsquigarrow \mathcal{F}(Q, N - K - 1)$$

où Q le nombre de restrictions, N le nombre d'observations, et K le nombre de variables explicatives.

Remarques:

1. Le test de Fisher appliqué à une restriction unique est parfaitement valable et donne alors exactement le même résultat que le test de Student. En effet, on peut montrer que la statistique de Fisher dans le cas d'une seule restriction est égale au carré de la statistique de Student. Or, la distribution de Fisher à $(1, N - K - 1)$ degrés de liberté est égale à la distribution du carré d'une variable qui suit une loi de Student à $(N - K - 1)$.

2. Le test de Fisher sous la forme de R^2 est très pratique. Soulignons que $SCR_R = SCT \times (1 - R_R^2)$ et $SCR_{NR} = SCT \times (1 - R_{NR}^2)$. Donc:

$$F = \frac{(SCR_R - SCR_{NR})/Q}{(SCR_{NR})/(N - K - 1)} = \frac{(R_{NR}^2 - R_R^2)/Q}{(1 - R_{NR}^2)/(N - K - 1)}.$$

3. Un test de Fisher appliqué à l'ensemble des paramètres est fourni par la plupart des logiciels d'économétrie. L'hypothèse nulle est alors:

$$H_0 : \beta_1 = 0, \beta_2 = 0, \dots, \beta_K = 0$$

et la statistique de Fisher se simplifie de la manière suivante:

$$F = \frac{R^2/K}{(1 - R^2)/(N - K - 1)}.$$

puisque $Q = K$, $R_{NR}^2 = R^2$ et $R_R^2 = 0$.

Exemple 6.6: J. Wooldridge (2010) considère la relation entre le salaire des joueurs de base-ball est expliqué par **années** (le nombre d'années passé dans la ligue), **nbmatch** (le nombre de matchs joués en moyenne), et le niveau du joueur représenté par trois indicateurs: **batav** ("career batting average"), **hrun** ("home runs per year") et **runbat** ("runs batted in per year"). Les résultats de cette estimation sont donnés ci-dessous:

$$\begin{aligned} \widehat{\log(\text{sal})} = & 11,10 + 0,0689 \cdot \text{années} + 0,126 \cdot \text{nbmatch} \\ & (0,29) \quad (0,0121) \quad (0,0026) \\ & + 0,00098 \cdot \text{batav} + 0,0144 \cdot \text{hrun} + 0,0108 \cdot \text{runbat} \\ & (0,0110) \quad (0,0161) \quad (0,0072) \end{aligned}$$

avec $N = 353$, $SCR = 183,16$, $R^2 = 0,627$. On désire tester l'hypothèse que le niveau des joueurs représenté donc par **batav**, **hrun** et **runbat**, n'influence leur salaire. On pourrait, à cette fin, effectuer des tests de Student pour chacune des variables en question mais le test fourni ne serait pas très efficace. On préfère estimer le modèle contraint suivant:

$$\begin{aligned} \widehat{\log(\text{sal})} = & 11,22 + 0,0713 \cdot \text{années} + 0,0202 \cdot \text{nbmatch} \\ & (0,11) \quad (0,0125) \quad (0,0013) \end{aligned}$$

avec $N = 353$, $\text{SCR} = 198, 31$, $R^2 = 0, 597$. En utilisant la SCR pour chacun des modèles, la statistique de Fisher est égale à

$$F = \frac{(198, 31 - 183, 16) / 3}{(183, 16) / 347} \approx 9, 55.$$

En utilisant la table statistique de Fisher, on constate que la seuil critique calculé pour un seuil de signification de 5% avec 3 et 347 degrés de liberté est de 2, 61. L'hypothèse nulle selon laquelle le salaire des joueurs n'est pas affecté par leur niveau des joueurs est donc rejetée.

6.3 Prédiction et analyse de résidus

Estimation de la moyenne conditionnelle

On veut trouver une prédiction de la moyenne conditionnelle de la variable expliquée pour des valeurs données des variables explicatives. Pour cela, on effectue une régression de y sur un ensemble de variables explicatives x_1, \dots, x_K , à savoir:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \dots + \beta_K \cdot x_K + u, \quad (6.1)$$

et on obtient les estimateurs des paramètres $\widehat{\beta}_1, \widehat{\beta}_2, \dots, \widehat{\beta}_K$. Soit x_1^*, \dots, x_K^* des valeurs particulières pour chacune des variables explicatives. La moyenne conditionnelle est définie par

$$\begin{aligned} y^* &= \mathbf{E}(y \mid x_1 = x_1^*, x_2 = x_2^*, \dots, x_K = x_K^*) \\ &= \beta_0 + \beta_1 \cdot x_1^* + \beta_2 \cdot x_2^* + \dots + \beta_K \cdot x_K^*. \end{aligned} \quad (6.2)$$

La prédiction de la moyenne conditionnelle est donc:

$$\widehat{y}^* = \widehat{\beta}_0 + \widehat{\beta}_1 \cdot x_1^* + \widehat{\beta}_2 \cdot x_2^* + \dots + \widehat{\beta}_K \cdot x_K^*.$$

Pour calculer la variance de la prédiction \widehat{y}^* , le plus simple est de procéder en reparamétrant le modèle. On écrit l'équation (6.2) de la manière suivante:

$$\beta_0 = y^* - \beta_1 \cdot x_1^* - \beta_2 \cdot x_2^* - \dots - \beta_K \cdot x_K^*,$$

et l'on remplace dans l'équation (6.1). On obtient le modèle suivant:

$$y = y^* - \beta_1 \cdot (x_1 - x_1^*) + \beta_2 \cdot (x_2 - x_2^*) + \dots + \beta_K \cdot (x_K - x_K^*) + u.$$

L'estimation de cette équation donne un estimateur de la moyenne conditionnelle y^* et une estimation de la variance de cet estimateur, à savoir, $\widehat{\text{Var}}(\hat{y}^*)$. Cette procédure permet de procéder à des tests et de calculer un intervalle de confiance pour la moyenne conditionnelle.

Variance de la prédiction

Pour calculer un intervalle de confiance de la prédiction (et non de la moyenne conditionnelle), il faut également tenir compte de la variance du terme aléatoire. Soit u^* une valeur (non observée) pour le terme aléatoire:

$$y^* + u^* = \beta_0 + \beta_1 \cdot x_1^* + \beta_2 \cdot x_2^* + \dots + \beta_K \cdot x_K^* + u^*.$$

L'erreur de prédiction qui est générée en utilisant \hat{y}^* pour prédire y^* s'écrit:

$$\hat{e}^* = y^* + u^* - \hat{y}^*.$$

Or, on sait que

$$\text{E}(\hat{y}^*) = \text{E}(\hat{\beta}_0) + \text{E}(\hat{\beta}_1) \cdot x_1^* + \text{E}(\hat{\beta}_2) \cdot x_2^* + \dots + \text{E}(\hat{\beta}_K) \cdot x_K^* = y^*$$

puisque les estimateurs sont non biaisés, et

$$\text{E}(u^*) = 0, \quad \text{et donc} \quad \text{E}(\hat{e}^*) = 0.$$

L'espérance de l'erreur de prédiction est ainsi nulle. La variance de l'erreur de prédiction est:

$$\text{Var}(\hat{e}^*) = \text{Var}(u^* - \hat{y}^*) = \text{Var}(\hat{y}^*) + \sigma_u^2$$

car u^* et \hat{y}^* sont non corrélés (en effet, le terme d'erreur u^* est non corrélé avec les observations de l'échantillon utilisées pour estimer \hat{y}^*). Donc, la variance de erreur de prédiction se décompose en deux éléments: la variance de la moyenne conditionnelle (elle-même liée à la variance ces estimateurs)

et la variance du terme d'erreur. Cependant, la première composante peut être négligée si l'échantillon est grand. L'erreur-type de la prédiction est donc donnée par

$$\text{se}(\hat{e}^*) = \left[\widehat{\text{Var}}(\hat{y}^*) + \hat{\sigma}_u^2 \right]^{1/2}$$

Si on suppose que le terme d'erreur a une distribution normale, suivant en cela l'hypothèse 5.6, l'erreur de prédiction aura également une distribution normale. Puisque $\hat{e}^*/\text{se}(\hat{e}^*)$ a une distribution de Student avec $N - K - 1$ degrés de liberté, l'intervalle de confiance de la prédiction $y^* + u^*$ est donné par

$$\left[\hat{y}^* - t_{\alpha/2} \cdot \text{se}(\hat{e}^*), \hat{y}^* + t_{\alpha/2} \cdot \text{se}(\hat{e}^*) \right]$$

où $t_{\alpha/2}$ est le $(\alpha/2) \times 100$ ème percentile de la loi de Student à $N - K - 1$ degrés de liberté.

Analyse de résidus

L'analyse des résidus est utilisée afin d'identifier les observations caractérisées par un résidu très grand ou très petit. Cette analyse peut être utile dans diverses circonstances:

1. **Pour mesurer l'efficacité d'une firme.** Une fonction de production des firmes est estimée, décrivant la relation entre le niveau de production et les facteurs de production utilisés. Un résidu positif signifie que, pour une quantité donnée des facteurs de production, la production est particulièrement élevée. On peut conclure que la firme est efficace.
2. **Pour classer les écoles de droit.** Le salaire médian mesuré en début de carrière des étudiants qui ont fini leurs études dans une école de droit est régressé sur les caractéristiques de ces étudiants. Un résidu positif signifie que le salaire médian de l'étudiant est particulièrement élevé au vu des caractéristiques de celui-ci. On peut conclure que l'école est performante.

Chapter 7

Asymptotique et hétéroscédasticité

Ce chapitre traite des propriétés des estimateurs des MCO lorsque le nombre d'observations devient très grand (et, formellement, tend vers l'infini).

7.1 Convergence

If you can't get it right as n goes to infinity, you shouldn't be in this business. (C.W.J. Granger).

Quand le modèle devient compliqué, il est parfois difficile de montrer que les estimateurs sont non biaisés. Toutefois, ils auront souvent une propriété de convergence.

Intuition: Soit $\hat{\beta}_k$ l'estimateur des MCO du paramètre β_k . Pour tout N , cet estimateur est distribué selon une certaine loi de probabilité. La convergence d'un estimateur $\hat{\beta}_k$ signifie que la probabilité que $|\hat{\beta}_k - \beta_k| > \varepsilon$, où ε est une constante positive arbitraire, tend vers zéro lorsque la taille de l'échantillon tend vers l'infini. La fonction de densité de l'estimateur dégénère alors et se concentre sur la valeur du paramètre de la population.

Notation: La valeur vers laquelle converge un estimateur est appelée "limite en probabilité". Elle représentée par l'opérateur plim. Ainsi, $\text{plim } \hat{\theta}$ est la valeur vers laquelle la statistique $\hat{\theta}$ converge.

Théorème 7.1: *Supposons que $\text{Var}(\hat{r}_k)$ existe et est finie pour tout k . Sous les hypothèses 5.1-5.4, les estimateurs des MCO sont convergents:*

$$\text{plim } \hat{\beta}_k = \beta_k, \quad \text{pour } k = 0, \dots, K.$$

Démonstration: L'estimateur des MCO s'écrit de la manière suivante:

$$\hat{\beta}_k = \beta_k + \frac{\sum_{i=1}^N \hat{r}_{ki} u_i}{\sum_{i=1}^N \hat{r}_{ki}^2} = \beta_k + \frac{(1/N) \sum_{i=1}^N \hat{r}_{ki} u_i}{(1/N) \sum_{i=1}^N \hat{r}_{ki}^2}.$$

Si l'on calcule la limite en probabilité des membres de gauche et de droite, on obtient:

$$\text{plim } \hat{\beta}_k = \beta_k + \frac{\text{Cov}(\hat{r}_k, u)}{\text{Var}(\hat{r}_k)},$$

en vertu de la Loi des grands nombres. [La Loi des grands nombres que, dans le cas d'un échantillonnage aléatoire, les moments empiriques convergent vers les moments théoriques. Voir Chapitre 2.] Donc,

$$\text{plim } \hat{\beta}_k = \beta_k.$$

En effet, l'hypothèse 5.2 implique que $E(u|\hat{r}_k) = 0$ et donc $\text{Cov}(\hat{r}_k, u) = 0$.
□

Remarques:

1. Le résultat de convergence est toujours valide si l'hypothèse 5.2 est remplacée par la condition suivante:

$$\text{Cov}(\hat{r}_k, u) = 0, \text{ pour tout } k.$$

Cette condition est plus faible que celle de l'hypothèse 5.2.

2. De la corrélation entre u et \hat{r}_k cause la non-convergence des estimateurs des MCO. Le terme

$$\text{plim} \hat{\beta}_k - \beta_k = \frac{\text{cov}(\hat{r}_k, u)}{\text{var}(\hat{r}_k)}$$

désigne le **biais asymptotique**. Ce terme est positif si \hat{r}_k et u sont positivement corrélés.

3. L'omission d'une variable pertinente dans le modèle cause généralement un biais asymptotique. On suppose que le modèle suivant satisfait les hypothèses 5.1-5.4:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u.$$

Le modèle estimé est

$$y = \beta_0 + \beta_1 x_1 + u.$$

Or nous avons vu dans le chapitre 5 que:

$$\tilde{\beta}_1 = \beta_1 + \beta_2 \frac{(1/N) \sum_{i=1}^N (x_{1i} - \bar{x}_1) x_{2i}}{(1/N) \sum_{i=1}^N (x_{1i} - \bar{x}_1)^2} + \frac{(1/N) \sum_{i=1}^N (x_{1i} - \bar{x}_1) u_i}{(1/N) \sum_{i=1}^N (x_{1i} - \bar{x}_1)^2}.$$

Si l'on calcule la limite en probabilité des membres de gauche et de droite sous certaines conditions de régularité, on obtient:

$$\text{plim} \tilde{\beta}_1 = \beta_1 + \beta_2 \frac{\text{Cov}(x_1, x_2)}{\text{Var}(x_1)} + \frac{\text{Cov}(x_1, u)}{\text{Var}(x_1)}.$$

En vertu de l'hypothèse 5.2, le dernier terme du membre de droite est égal à zero de sorte que

$$\text{plim} \tilde{\beta}_1 = \beta_1 + \beta_2 \frac{\text{Cov}(x_1, x_2)}{\text{Var}(x_1)}.$$

ou encore,

$$\text{plim} \hat{\beta}_1 = \beta_1 + \beta_2 \delta_1$$

où

$$\delta_1 = \frac{\text{Cov}(x_1, x_2)}{\text{Var}(x_1)} = \text{plim} \hat{\delta}_1$$

est la limite en probabilité du paramètre de la pente de la régression de x_2 sur x_1 .

7.2 Normalité asymptotique

En grand échantillon, l'hypothèse 5.6 sur la normalité sur les termes d'erreur peut être abandonnée. C'est important puisque, souvent, cette hypothèse est peu crédible.

Théorème 7.2: *Supposons que $\text{Var}(\hat{r}_k)$ existe et est finie pour tout k . Sous les hypothèses 5.1-5.5,*

$$\sqrt{N} \times (\hat{\beta}_k - \beta_k) \overset{a}{\rightsquigarrow} \mathcal{N}(0, \sigma^2/a_k^2)$$

où σ^2/a_k^2 est la variance asymptotique de $\sqrt{N} \times (\hat{\beta}_k - \beta_k)$; et

$$a_k^2 = \text{plim} \left(\frac{1}{N} \sum_{i=1}^N \hat{r}_{ki}^2 \right),$$

où \hat{r}_{ki} est le résidu de la régression de x_k sur l'ensemble des variables autres explicatives; et $\hat{\sigma}^2$ est un estimateur convergent de σ^2 .

Démonstration: L'estimateur des MCO s'écrit de la manière suivante:

$$\hat{\beta}_k = \frac{\sum_{i=1}^N \hat{r}_{ki} y_i}{\sum_{i=1}^N \hat{r}_{ki}^2} = \beta_k + \frac{\sum_{i=1}^N \hat{r}_{ki} u_i}{\sum_{i=1}^N \hat{r}_{ki}^2}.$$

et peut donc être réécrit de la manière suivante:

$$\sqrt{N} \times (\hat{\beta}_k - \beta_k) = \frac{(1/\sqrt{N}) \sum_{i=1}^N \hat{r}_{ki} u_i}{(1/N) \sum_{i=1}^N \hat{r}_{ki}^2}.$$

Si la taille de l'échantillon tend vers l'infini, le dénominateur de cette expression converge vers a_k^2 par application de la Loi des grands nombres. Puisque $E(\hat{r}_{ki} u_i) = 0$ en vertu de l'hypothèse 5.2, le numérateur converge en probabilité vers une variable aléatoire distribuée selon une loi normale par application du Théorème de la limite centrale. \square

Théorème 7.3: *Supposons que $\text{Var}(\hat{r}_k)$ est finie pour tout k . Sous les hypothèses 5.1-5.5, l'estimateur des MCO de la variance du terme d'erreur est un estimateur convergent:*

$$\text{plim} \hat{\sigma}^2 = \sigma^2.$$

Démonstration: Voir Hayashi (2000). \square

Remarques:

1. Selon les formules traditionnelles (en petit échantillon), la variance de $\sqrt{N} \times (\hat{\beta}_k - \beta_k)$ est égale à

$$N \times \frac{\sigma^2}{(1 - R_k^2) \sum_{i=1}^N (x_{ki} - \bar{x}_k)^2} = \frac{\sigma^2}{(1/N) \sum_{i=1}^N \hat{r}_{ki}^2}.$$

Quand la taille de l'échantillon tend vers l'infini, le dénominateur de cette expression converge vers a_k^2 .

2. Pour tout k , puisque $\text{plim } \hat{\sigma}^2 = \sigma^2$,

$$\frac{\sqrt{N} \times (\hat{\beta}_k - \beta_k)}{\sqrt{N\hat{\sigma}^2 / \sum_{i=1}^N \hat{r}_{ki}^2}} = \frac{\hat{\beta}_k - \beta_k}{\text{se}(\hat{\beta})} \xrightarrow{a} \mathcal{N}(0, 1).$$

En effet, la distribution de Student converge vers la distribution Normale centrée réduite lorsque le nombre de degrés de liberté tend vers l'infini. Donc, quand l'échantillon est très grand, les tests de Student peuvent être réalisés de la même manière que précédemment en remplaçant la distribution de Student par la distribution Normale. Ces tests ne reposent pas sur l'hypothèse 5.6.

3. Les tests de Fisher sont asymptotiquement valide. En d'autres termes, si l'échantillon est grand, les procédures d'inférences décrites précédemment pour tester une multiplicité de restrictions peuvent être appliquées telles quelles.

7.3 Tests de l'hétéroscédasticité

L'hétéroscédasticité des termes d'erreur ne remet pas en question le caractère non biaisé des estimateurs des MCO. Toutefois, elle invalide les estimateurs de la variance de ceux-ci et fausse les procédure classiques de tests; de plus, le théorème de Gauss-Markov ne peut plus être appliqué et les estimateurs des MCO ne sont plus nécessairement les meilleurs. Il est donc intéressant

de pouvoir de détecter la présence d'hétéroscédasticité. Pour commencer, on considère le modèle de régression suivant:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_K x_K + u.$$

Afin de tester la présence d'hétéroscédasticité, on veut tester l'hypothèse nulle suivante:

$$H_0 : \text{Var}(u|\mathbf{x}) = \sigma^2 = \text{constante},$$

où, pour simplifier la notation, $\mathbf{x} = (x_1, \dots, x_K)$ est le vecteur des variables explicatives. Puisque $E(u|\mathbf{x}) = 0$, et en vertu d'une identité bien connue, cette hypothèse peut s'écrire de manière équivalente:

$$H_0 : E(u^2|\mathbf{x}) = \sigma^2.$$

Le test d'homoscédasticité consiste à tester si le carré du terme d'erreur est corrélé avec certaines variables explicatives. Si l'hypothèse nulle est satisfaite, on peut écrire

$$u^2 = \sigma^2 + v$$

avec $E(v|\mathbf{x}) = 0$. Le **test de Breusch-Pagan** suppose une relation linéaire entre le carré du terme aléatoire et les variables explicatives:

$$u^2 = \sigma^2 + \delta_1 x_1 + \dots + \delta_K x_K + v$$

avec $E(v|\mathbf{x}) = 0$. La procédure se réduit à tester l'hypothèse suivante:

$$H_0 : \delta_1 = \dots = \delta_K = 0.$$

Cependant, comme le terme aléatoire n'est pas observé, il doit être remplacé par le résidu \hat{u}^2 obtenu par les MCO. La statistique de Fisher de la significativité globale de la régression est:

$$F = \frac{R_{u^2}^2 / K}{(1 - R_{u^2}^2) / (N - K - 1)}$$

où $R_{u^2}^2$ est le coefficient de détermination calculé à partir de la régression du carré du résidu sur les variables explicatives. Le **test de White** suppose

une relation quadratique entre le carré du terme aléatoire et les variables explicatives:

$$u^2 = \sigma^2 + \delta_1 x_1 + \dots + \delta_K x_K + \Delta_{11} x_1^2 + \dots + \Delta_{KK} x_K^2 + v$$

avec $E(v|\mathbf{x}) = 0$. Le test de White se réduit alors à tester l'hypothèse suivante:

$$H_0 : \delta_1 = \dots = \delta_K = \Delta_{11} = \dots = \Delta_{KK} = 0. \quad (7.1)$$

La statistique de Fisher est la même que celle dérivée précédemment. Toutefois, le test de White se justifie par le fait que, si la forme de l'hétéroscédasticité est différente de celle testée par (7.1), l'inférence reste correcte.

7.4 La méthode des MC Pondérés

Si les tests de Breusch-Pagan ou White détectent la présence d'hétéroscédasticité, les tests de Student et de Fisher ne sont plus valides. La méthode des MCO n'est plus efficace, et une autre méthode doit être envisagée. Soit le modèle suivant qui satisfait les hypothèses 5.1-5.4:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_K x_{Ki} + u_i \quad (7.2)$$

où le terme aléatoire est hétéroscédastique. Sans perte de généralité, la variance du terme aléatoire peut s'écrire sous la forme suivante:

$$\text{Var}(u|\mathbf{x}) = \sigma^2 \times h(\mathbf{x}),$$

où $h(\mathbf{x})$ est une fonction positive des variables explicatives. Pour estimer le modèle, on peut considérer les deux cas suivants.

Cas 1: observation de $h(\mathbf{x})$.

On suppose ici que la fonction $h(\mathbf{x})$ est connue par l'économètre. En d'autres termes, la variance du terme aléatoire est connue à une constante multiplicative près. Puisque $\text{Var}(u_i|\mathbf{x}_i) = \sigma^2 \times h(\mathbf{x}_i)$, et en utilisant les propriétés de

l'opérateur variance, la variance de $u_i/\sqrt{h(\mathbf{x}_i)}$ est égale à

$$\text{Var} \left(\left(\frac{u_i}{\sqrt{h(\mathbf{x}_i)}} \right) \middle| \mathbf{x}_i \right) = \frac{1}{h(\mathbf{x}_i)} \text{var}(u_i | \mathbf{x}_i) = \sigma^2.$$

Si nous divisons les variables de l'équation de régression (7.2), on obtient:

$$\frac{y_i}{\sqrt{h(\mathbf{x}_i)}} = \beta_1 \frac{1}{\sqrt{h(\mathbf{x}_i)}} + \beta_2 \frac{x_{i1}}{\sqrt{h(\mathbf{x}_i)}} + \dots + \beta_K \frac{x_{iK}}{\sqrt{h(\mathbf{x}_i)}} + \frac{u_i}{\sqrt{h(\mathbf{x}_i)}}.$$

ou

$$y_i^* = \beta_1 x_{0i}^* + \beta_1 x_{i1}^* + \beta_2 x_{i2}^* + \dots + \beta_K x_{iK}^* + u_i^*,$$

avec

$$y_i^* = \frac{y_i}{\sqrt{h(\mathbf{x}_i)}}, \quad x_{0i}^* = \frac{1}{\sqrt{h(\mathbf{x}_i)}}, \quad x_{ki}^* = \frac{x_{ki}}{\sqrt{h(\mathbf{x}_i)}}, \quad u_i^* = \frac{u_i}{\sqrt{h(\mathbf{x}_i)}}.$$

On peut calculer les estimateurs des MCO pour ce modèle transformé. Les estimateurs obtenus sont une variante, appelée **Moindres Carrés Pondérés** (MCP), de la méthode des **Moindres Carrés Généralisés** (MCG). En effet, la somme des résidus du modèle transformé est égale à

$$\sum_{i=1}^N (y_i^* - \beta_0 x_{0i}^* - \beta_1 x_{i1}^* - \beta_2 x_{i2}^* - \dots - \beta_K x_{iK}^*)^2,$$

et coïncide avec la somme des résidus du modèle initial pondérés par l'inverse de la fonction $h(\mathbf{x}_i)$:

$$\sum_{i=1}^N \frac{1}{h(\mathbf{x}_i)} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \dots - \beta_K x_{iK})^2.$$

Les tests de Student et de Fisher peuvent être effectués en utilisant le modèle transformé. Les conditions du Théorème de Gauss-Markov sont également remplies de sorte que l'estimateur est le meilleur dans la classe des estimateurs linéaires non biaisés.

Exemple 8.1: On considère le modèle initial suivant sur le comportement d'épargne des ménages:

$$\text{epar} = \beta_0 + \beta_1 \cdot \text{rev} + u,$$

où epar désigne le montant épargné par le ménage et rev le revenu de ce même ménage, avec

$$\text{Var}(u | \text{rev}) = \sigma^2 \times \text{rev}.$$

Le terme d'erreur est donc hétéroscédastique, et la fonction $h(\mathbf{x})$ est simplement la fonction identité. Le modèle transformé est donné par:

$$\frac{\text{epar}}{\sqrt{\text{rev}}} = \beta_0 \cdot \frac{1}{\sqrt{\text{rev}}} + \beta_1 \cdot \sqrt{\text{rev}} + u^*.$$

Le terme d'erreur de ce modèle transformé est naturellement homoscedastique.

Cas 2: non observation de $h(\mathbf{x})$.

En général, le fonction $h(\mathbf{x})$ n'est pas observée; elle doit alors être estimée. Pour cela, il faut choisir une forme fonctionnelle particulière pour l'expression de la variance du terme d'erreur:

$$\text{Var}(u | \mathbf{x}) = \sigma^2 \cdot h(\mathbf{x}) = \delta_0 + \delta_1 x_1 + \dots + \delta_K x_K. \tag{7.3}$$

(Notons que cette forme fonctionnelle linéaire choisie ici n'exclut pas, en général, la possibilité d'une variance négative.) La première étape consiste donc à estimer les paramètres $\delta_0, \delta_1, \dots, \delta_K$. Pour cela, on se rappelle que, de manière équivalente, la variance du terme d'erreur peut s'écrire de la manière suivante:

$$\text{E}(u^2 | \mathbf{x}) = \delta_0 + \delta_1 x_1 + \dots + \delta_K x_K,$$

ou encore:

$$u^2 = \delta_0 + \delta_1 x_1 + \dots + \delta_K x_K + v,$$

où $\text{E}(v | \mathbf{x}) = 0$. Les termes d'erreur u ne sont naturellement pas observables. Cependant, on peut les remplacer par les résidus \hat{u} obtenus par l'estimation du modèle par la méthode des MCO permettant ainsi d'obtenir les estimateurs $\hat{\delta}_0, \hat{\delta}_1, \dots, \hat{\delta}_K$. Les estimations des paramètres permettent d'estimer

la fonction $h(\mathbf{x})$ de la manière suivante:

$$\sigma^2 \cdot \widehat{h(x)} = \widehat{\delta}_0 + \widehat{\delta}_1 x_1 + \dots + \widehat{\delta}_K x_K$$

où σ^2 est une constante quelconque qui peut être, sans perte de généralité, normalisée à 1. En utilisant cette estimation de la fonction $h(\mathbf{x})$, il est alors possible de transformer le modèle comme précédemment et de calculer les estimateurs des MCO sur ce modèle transformé. Les estimateurs obtenus sont une variante des estimateurs des **Moindres Carrés Généralisés Faibles**. Ils sont convergents (et éventuellement non biaisés); ils sont asymptotiquement distribués selon une loi normale; et ils sont plus efficaces que les estimateurs des MCO à condition que la spécification de la variance (7.3) soit correcte et que l'échantillon soit suffisamment grand. Les procédures de tests précédemment décrites (Student et Fisher) peuvent être appliquées.

Exemple 8.2: J. Wooldridge (2010) s'intéresse à la consommation quotidienne de cigarettes et dispose d'un échantillon sur le comportement des adultes célibataires aux Etats-Unis. Le modèle estimé par la méthode des MCO donne les résultats suivants:

$$\begin{aligned} \widehat{\text{cigs}} = & - 3,64 + 0,880 \log(\text{rev}) - 0,751 \log(\text{prix cig}) - 0,501 \text{ edu} \\ & (24,08) \quad (0,728) \quad (5,773) \quad (0,167) \\ & + 0,771 \text{ age} - 0,099 \text{ age}^2 - 2,83 \text{ restaurn} \\ & (0,167) \quad (0,002) \quad (1,11) \end{aligned}$$

avec $N = 807$ et $R^2 = 0.0526$, où **cigs** est le nombre de cigarettes fumées par jour, **rev** est le revenu annuel, **prix cig** est le prix du paquet de cigarette (en cents), **edu** est le nombre d'années d'éducation, **age** est l'âge en années et **restaurn** est une variable prenant une valeur égale à 1 si la personne réside dans un Etat dans lequel les restaurants imposent des restrictions sur l'usage du tabac et 0 autrement (ce genre de variables qui prennent seulement les valeurs 1 ou 0 sera examiné dans le prochain chapitre). Le test de Breusch-Pagan est effectué afin de tester l'éventuelle présence d'hétéroscédasticité. La régression du carré des résidus sur les variables explicatives donne $R_{u^2}^2 =$

0.040. Le F-ratio correspondant à ce test est égal à

$$F = \frac{R_{u^2}^2/K}{(1 - R_{u^2}^2)/(N - K - 1)} = \frac{0.040/6}{(1 - 0.040)/(807 - 6 - 1)} \simeq 5.555.$$

Sous l'hypothèse nulle, cette statistique est distribuée selon une distribution de Fisher à 6 et 800 degrés de liberté. Pour un seuil de signification à 5%, la valeur critique est de 2,10. L'homoscédasticité est donc rejetée. Estimé par la méthode des MCP, le même modèle donne donc les résultats suivants:

$$\begin{aligned} \widehat{\text{cigs}} = & 5,64 + 1,30 \log(\text{rev}) - 2,94 \log(\text{prix cig}) - 0,463 \text{ edu} \\ & (17,80) \quad (0,44) \quad (4,46) \quad (0,120) \\ & + 0,482 \text{ age} - 0,006 \text{ age}^2 - 3,46 \text{ restaurn} \\ & (0,097) \quad (0,001) \quad (0,80) \end{aligned}$$

avec $N = 807$ et $R^2 = 0.1134$.

Chapter 8

Echantillon et spécification

8.1 Le choix de la forme fonctionnelle

En dépit du fait que les méthodes d'estimation sont linéaires, des relations non linéaires entre variables explicatives et variables expliquées peuvent être estimées. On peut distinguer des formes fonctionnelles linéaires en niveau, linéaires en logarithmes, et des formes fonctionnelles quadratiques. Le choix de la forme fonctionnelle doit permettre d'améliorer l'ajustement du modèle au nuage de points; il peut aussi obéir à certaines raisons de commodité.

Formes linéaires et logarithmiques

Lorsque les variables explicatives et expliquées sont en niveau, les estimateurs des paramètres dépendent des unités de mesure:

$$y = \beta_0 + \beta_1 x + u$$

Si les variables explicatives et expliquées sont positives, elles peuvent être exprimées en logarithmes. Dans ce cas, les paramètres sont des élasticités:

$$\log(y) = \beta_0 + \beta_1 \log(x) + u$$

où β_1 représente la variation en pourcentages de y dû à une variation en pourcentages de x . Lorsque seule la variable expliquée est en logarithme, ou

lorsque seule la variable explicative est en logarithme, les paramètres sont des demi-élasticités:

$$\log(y) = \beta_0 + \beta_1 x + u$$

où β_1 représente la variation en pourcentages de y dû à une variation en niveau de x , et

$$y = \beta_0 + \beta_1 \log(x) + u$$

où β_1 représente la variation en niveau de y dû à une variation en pourcentages de x . Les variables monétaires, ou plus généralement celles dont la valeur est large, sont souvent exprimées en logarithmes. En revanche, les variables mesurées en années sont souvent exprimées en niveaux. De même, les taux, les proportions, les pourcentages sont souvent exprimées en niveaux. Si la variable prend des valeurs négatives, cependant, elle ne peut pas être exprimée en logarithmes. Si elle est parfois égale à 0, alors on peut utiliser la transformation: $\log(x + 1)$. Notons enfin que les élasticités (ou demi-élasticités) sont des approximations si on considère des variations discrètes de la variable explicative.

Exemple 9.1: Dans l'exemple ci-dessous, J. Wollridge (2010) s'intéresse à la relation entre le prix des maisons et le niveau de pollution dans son environnement. Il dispose d'un échantillon de 506 communautés dans la région de Boston. Le modèle estimé est le suivant:

$$\log(\widehat{\text{prixmedmais}}) = 9,23 - 0,718 \log(\text{oxnit}) + 0,306 \text{ nbpièces},$$

$$(0,19) \quad (0,066) \quad (0,019)$$

avec $N = 506$, $R^2 = 0,514$, où prixmedmais est le prix médian des maisons de la communauté, oxnit désigne la quantité d'oxyde nitreux dans l'air par million (une mesure de la pollution) dans la communauté et nbpièces est le nombre moyen de pièces des maisons dans la communauté. L'effet exact en pourcentage de l'accroissement d'une pièce sur le prix peut être obtenu de la manière suivante. L'effet d'une variation du nombre de pièces est donné

par:

$$\Delta \log(\widehat{\text{prixmais}}) = 0,306 \cdot \Delta \text{nbpièces}.$$

En conséquence,

$$\log \left(\frac{\text{prix d'une maisons avec } r + 1 \text{ pièces}}{\text{prix d'une maisons avec } r \text{ pièces}} \right) = 0,306,$$

et donc l'accroissement du prix des maisons est donné par

$$\frac{\text{prix d'une maisons avec } r + 1 \text{ pièces}}{\text{prix d'une maisons avec } r \text{ pièces}} = \exp(0,306) = 1,358.$$

Le prix de la maison augmente de 35% si une pièce de plus est ajoutée. Notons que cet ajustement pour calculer le pourcentage n'est pas important si le paramètre correspondant est proche de zéro.

Formes quadratiques

Afin de représenter des relations non linéaires, certaines variables explicatives peuvent également être exprimées sous forme quadratique. La forme fonctionnelle est alors caractérisée par une partie croissante et une partie décroissante en fonction de la valeur de la variable explicative. Les termes quadratiques peuvent être mélangés avec des termes logarithmiques. Ils peuvent également être associés à des termes d'interaction.

Exemple 9.2: En utilisant les données sur les Etats-Unis déjà mentionnées J. Wooldridge (2010) estime la relation suivante:

$$\widehat{\text{sal}} = 3,73 + 0,298 \text{ exp} - 0,0061 \text{ exp}^2.$$

(0,35) (0,041) (0,0009)

avec $N = 526$, $R^2 = 0,093$. L'effet de l'expérience sur le salaire est donné par

$$\Delta \widehat{\text{sal}} = (0,298 - 2 \times 0,0061 \cdot \text{exp}) \times \Delta \text{exp}.$$

La première année d'expérience rapporte environ 29,8 centimes, la seconde année environ 28,6 centimes, la onzième année environ 17,6 centimes. Le

rendement de l'expérience est donc décroissant. L'année à partir de laquelle l'effet d'exper est négatif est donnée par

$$(0,298 - 2 \times 0,0061 \cdot \text{exp}^*) = 0 \Rightarrow \text{exp}^* = \frac{0,298}{2 \times 0,0061} \approx 24,4.$$

Notons que, si l'on désire tester que l'expérience d'un travailleur a un effet sur son salaire, il convient d'utiliser un test de Fisher.

Exemple 9.3: En utilisant les mêmes données sur le prix des maisons J. Wooldridge (2010) estime la relation suivante:

$$\begin{aligned} \log(\widehat{\text{prixmedmais}}) = & 13,39 - 0,902 \log(\text{oxnit}) - 0,545 \text{ nbpièces} \\ & (0,57) \quad (0,115) \quad (0,165) \\ & + 0,062 \text{ nbpièces}^2 - 0,087 \log(\text{dist}) - 0,048 \text{ sratio}, \\ & (0,013) \quad (0,043) \quad (0,006) \end{aligned}$$

avec $N = 506$, $R^2 = 0,603$, où dist est la distance pondérée entre la communauté de résidence et cinq zones d'emplois et sratio est le rapport moyen du nombre d'étudiants et du nombre d'enseignants dans les écoles de la communauté. L'effet du nombre moyen de pièces dans la communauté sur le prix médian des maisons est donné par

$$\Delta \widehat{\text{prixmedmais}} = (-0,545 + 2 \times 0,062 \cdot \text{nbpièces}) \times \Delta \text{nbpièces}.$$

Cet effet est donc croissant. L'année à partir de laquelle l'effet d'exper est négatif est donnée par

$$(-0,545 + 2 \times 0,062 \cdot \text{nbpièces}^*) = 0 \Rightarrow \text{nbpièces}^* = \frac{0,545}{2 \times 0,062} \approx 4,4.$$

Exemple 9.4: En reprenant les données sur les salaires, on estime l'équation suivante incluant le croisement des années d'éducation et des années d'expérience:

$$\begin{aligned} \widehat{\log(\text{sal})} = & 5,95 + 0,0440 \text{ edu} - 0,0215 \text{ exp} + 0,00320 \text{ edu} \times \text{exp} \\ & (0,24) \quad (0,0174) \quad (0,0200) \quad (0,00153) \end{aligned}$$

avec $N = 935$, $R^2 = 0,135$. L'effet de l'expérience sur le salaire est donné par:

$$\frac{\Delta \widehat{\log(\text{sal})}}{\Delta \text{exp}} = -0,0215 + 0,00320 \cdot \text{edu}.$$

et donc

$$-0,0215 + 0,00320 \cdot \text{edu}^* = 0 \Rightarrow \text{edu}^* = \frac{0,0215}{0,00320} \approx 6,71.$$

L'effet de l'expérience sur le salaire est négatif pour les personnes ayant 6 années et moins d'éducation, et est positif pour les personnes ayant 7 années et plus d'éducation.

8.2 Les variables qualitatives

Les variables qualitatives sont des variables qui peuvent prendre une forme binaire. Par exemple, le fait d'être un homme ou une femme, français ou étranger, etc dans une enquête. On parle également de variables "dichotomique" ou de variable "dummy". Pour illustrer l'usage de variables dichotomiques, on va présenter une série d'exemples.

Personne	SAL	EDUC	EXPER	FEM	MAR
1	3.10	11	2	1	0
2	3.24	12	22	1	1
3	3.00	11	2	0	0
4	6.00	8	44	0	1
5	5.30	12	7	0	1
⋮	⋮	⋮	⋮	⋮	⋮
525	11.56	16	5	0	1
526	3.50	14	8	1	0

Exemple 9.5: J. Wooldridge (2010) considère la régression du salaire sur les années d'éducation, les années d'expérience, les années d'ancienneté et sur une variable binaire qui prend la valeur 1 lorsque le travailleur est une femme et zéro sinon:

$$\text{sal} = \beta_0 + \delta_0 \cdot \text{fem} + \beta_1 \cdot \text{edu} + \beta_2 \cdot \text{exp} + \beta_3 \cdot \text{anc} + u.$$

En fait, ce modèle général incorpore deux modèles particuliers avec des constantes différentes:

- Si $fem = 1$,

$$sal = (\beta_0 + \delta_0) + \beta_1 \cdot edu + \beta_2 \cdot exp + \beta_3 \cdot anc + u$$

- Si $fem = 0$,

$$sal = \beta_0 + \beta_1 \cdot edu + \beta_2 \cdot exp + \beta_3 \cdot anc + u$$

Une variable explicative dichotomique implique un déplacement de la constante de la droite de régression, la pente restant la même. Dans le cas présent, la catégorie "homme" est la catégorie de référence, par rapport à laquelle l'effet sur le salaire du fait d'être une femme est mesuré. Le choix de la catégorie de référence ("homme" ou "femme") est arbitraire. L'estimation de ce modèle est la suivantes:

$$\widehat{sal} = -1,57 - 1,81 fem + 0,572 edu + 0,025 exp + 0,141 anc$$

$$(0,72) \quad (0,26) \quad (0,049) \quad (0,012) \quad (0,021)$$

avec $N = 526$, $R^2 = 0,364$. Ce genre d'estimation permet notamment de mesurer la discrimination en termes de salaire: dans les données utilisées, et à un même niveau d'éducation, d'expérience, et d'ancienneté, une femme gagne environ un dollar et quatre-vingt cents de moins qu'un homme. Grâce à la régression suivante,

$$\widehat{sal} = 7,10 - 2,51 fem,$$

$$(0,21) \quad (0,30)$$

$$N = 526, \quad R^2 = 0,116$$

avec $N = 526$, $R^2 = 0,116$, on remarque que, si on ne contrôle pas pour l'éducation, l'expérience et l'ancienneté, la différence de salaires est encore plus élevée et s'élève à environ deux dollars et demi.

Exemple 9.6: Les variables dichotomiques peuvent être introduites dans des modèles logarithmiques. Dans l'exemple suivant,

$$\log(\widehat{prixmais}) = 5,56 + 0,168 \log(surfteer) + 0,707 \log(surfhab)$$

$$(0,65) \quad (0,038) \quad (0,093)$$

$$+ 0,027 nbcha + 0,054 col$$

$$(0,029) \quad (0,045)$$

avec $N = 88$, $R^2 = 0,649$, le fait que la maison soit de style colonial soit de style colonial augmente son prix de 5% environ.

Exemple 9.7: Les variables dichotomiques peuvent être croisées entre elles afin de former de nouvelles variables dichotomiques. Dans l'exemple suivant,

$$\begin{aligned} \widehat{\log(\text{sal})} = & 0,321 + 0,213 \text{ hom} \cdot \text{mar} - 0,198 \text{ fem} \cdot \text{mar} \\ & (0,100) \quad (0,055) \quad (0,058) \\ & - 0,110 \text{ fem} \cdot \text{cel} + 0,079 \text{ edu} + 0,027 \text{ exp} \\ & (0,056) \quad (0,007) \quad (0,005) \\ & - 0,00054 \text{ exp}^2 + 0,029 \text{ anc} - 0,00053 \text{ anc}^2 \\ & (0,00011) \quad (0,007) \quad (0,00023) \end{aligned}$$

avec $N = 526$, $R^2 = 0,461$. Le produit de la variable dichotomique "hom" et de celle "mar" génère une variable dichotomique qui prend la valeur 1 pour les hommes mariés et 0 pour les autres.

Exemple 9.8 (Hamermesh et Bidle, 1994): Une enquête est réalisée sur les salaires; les enquêteurs récoltent également de l'information sur la beauté physique des personnes interrogées. Hamermesh et Bidle classent ensuite les individus entre beauté inférieure à la moyenne, moyenne, supérieure à la moyenne, et créent des variables dichotomiques. Ils effectuent les estimations suivantes pour les hommes et les femmes:

$$\widehat{\log(\text{salh})} = \hat{\beta}_0 - 0,164 \text{ infmoy} + 0,016 \text{ supmoy} + \text{autres}$$

$$(0,046) \quad (0,033)$$

avec $N = 700$,

$$\widehat{\log(\text{salf})} = \hat{\beta}_0 - 0,124 \text{ infmoy} + 0,035 \text{ supmoy} + \text{autres}$$

$$(0,066) \quad (0,049)$$

avec $N = 409$, où **salh** est le salaire des hommes et **salf** le salaire des femmes. Les autres variables introduites dans la régression sont l'éducation, l'expérience, l'ancienneté, le statut marital et l'origine ethnique. Les hommes et les femmes avec une beauté physique en-dessous de la moyenne obtiennent donc, en général, un salaire significativement moins élevé.

Exemple 9.9: J. Wooldridge (2010) explique le salaire médian des diplômés d'écoles de droit en fonction des caractéristiques de l'école de droit et notamment du rang de celle-ci. Il obtient la régression suivante:

$$\begin{aligned} \widehat{\log(\text{sal})} = & 9,17 + 0,700 \text{ top10} + 0,594 \text{ r11-25} + 0,375 \text{ r26-40} \\ & (0,41) \quad (0,053) \quad (0,039) \quad (0,034) \\ & + 0,263 \text{ r41-60} + 0,132 \text{ r61-100} + 0,0057 \text{ LSAT} \\ & (0,028) \quad (0,021) \quad (0,0031) \\ & + 0,0014 \text{ GPA} + 0,036 \text{ log(libvol)} + 0,0008 \text{ log(cost)} \\ & (0,074) \quad (0,026) \quad (0,0251) \end{aligned}$$

avec $N = 136$, $R^2 = 0.905$.

Exemple 9.10: Dans l'exemple suivant où J. Wooldridge (2010) explique le salaire des joueurs de base-ball par diverses variables, des variables dichotomiques sont croisées avec des variables continues:

$$\begin{aligned} \widehat{\log(\text{sal})} = & 10,34 + 0,0673 \text{ années} + 0,0089 \text{ nbmatches} \\ & (2,18) \quad (0,0129) \quad (0,0034) \\ & + 0,00095 \text{ batav} + 0,0146 \text{ hrun} + 0,0045 \text{ runbat} \\ & (0,00151) \quad (0,0164) \quad (0,0076) \\ & + 0,0072 \text{ run} + 0,0011 \text{ fldperc} + 0,0075 \text{ allstar} \\ & (0,0046) \quad (0,0021) \quad (0,0029) \\ & - 0,198 \text{ noir} - 0,190 \text{ hisp} + 0,0125 \text{ noir} \cdot \text{percnoir} \\ & (0,125) \quad (0,153) \quad (0,0050) \\ & + 0,0201 \text{ hisp} \cdot \text{perchisp} \\ & (0,0098) \end{aligned}$$

avec $N = 330$, $R^2 = 0.368$, où *hisp* et *noir* sont des variables dichotomiques pour les joueurs noirs et hispaniques. les variables dichotomiques, *percnoir* et *perchisp* sont les pourcentages de noirs et d'hispaniques dans la ville du club.

8.3 Test de Chow

Le test de Chow permet de tester que des populations différentes sont caractérisées par les mêmes valeurs de paramètres. L'estimation de la régression suivante,

$$\text{sal} = \beta_0 + \delta_0 \cdot \text{fem} + \beta_1 \cdot \text{edu} + \delta_1 \cdot \text{fem} \cdot \text{edu} + u$$

est équivalente à l'estimation de deux modèles particuliers avec des constantes et des pentes différentes:

- Si fem= 1,

$$\begin{aligned} \text{sal} &= \beta_{1,0} + \beta_{1,1} \cdot \text{edu} + u \\ &= (\beta_0 + \delta_0) + (\beta_1 + \delta_1) \cdot \text{edu} + u \end{aligned} \quad (8.1)$$

où $\beta_{1,0} = \beta_0 + \delta_0$ et $\beta_{1,1} = \beta_1 + \delta_1$.

- Si fem= 0,

$$\text{sal} = \beta_{2,0} + \beta_{2,1} \cdot \text{edu} + u. \quad (8.2)$$

Si l'on veut tester que la même droite de régression s'applique aux hommes et aux femmes, l'hypothèse nulle s'écrit:

$$H_0 : \delta_0 = 0, \quad \delta_1 = 0.$$

Pour tester cette hypothèse, il convient d'utiliser un test de Fisher traditionnel. Toutefois, ce dernier prend une forme particulière ici. Le modèle contraint suivant:

$$\text{sal} = \beta_{0,0} + \beta_{0,1} \cdot \text{edu} + u. \quad (8.3)$$

est estimé sur l'ensemble de l'échantillon afin de calculer SCR_0 . Le modèle non contraint est estimé par l'intermédiaire de (8.1) et de (8.2) afin d'obtenir SCR_1 et SCR_2 . Le test de Fisher de cette hypothèse nulle s'écrit:

$$F = \frac{(\text{SSR}_0 - (\text{SSR}_1 + \text{SSR}_2)) N - 2(K + 1)}{(\text{SSR}_1 + \text{SSR}_2) K + 1}$$

où K est le nombre de variables dans 8.1), (8.2) et (8.3).

8.4 Le modèle de probabilité linéaire

Une variable dichotomique peut également être utilisée comme variable dépendante de la régression. On suppose que la variable y prend deux valeurs: $y = 0$ et $y = 1$. Le modèle de régression multiple s'écrit:

$$y = \beta_0 + \beta_1 \cdot x_1 + \dots + \beta_K \cdot x_K + u.$$

Si on suppose que l'hypothèse 5.2 est satisfaite, alors

$$E(y|\mathbf{x}) = \beta_0 + \beta_1 \cdot x_1 + \dots + \beta_K \cdot x_K.$$

Dans ce cas particulier, $E(y = 1|\mathbf{x}) = \Pr(y = 1|\mathbf{x})$, la probabilité que $y = 1$ sachant les valeurs de \mathbf{x} . En d'autres termes, la probabilité que $y = 1$ est une fonction linéaire de \mathbf{x} . Ce modèle est appelé **modèle de probabilité linéaire**. Les paramètres représentent donc l'effet des variables explicatives sur la probabilité que y soit égal à 1,

$$\Delta \Pr(y|\mathbf{x}) = \beta_k \cdot \Delta x_k.$$

Exemple 9.11: En utilisant les données de T. Mroz (1987) pour 1975, on peut estimer la probabilité pour les femmes de participer au marché du travail. On définit la variable **part** qui est égale à 1 si la femme participe au marché du travail et à zéro sinon. Les estimations donnent les résultats suivants:

$$\begin{aligned} \widehat{\text{part}} = & 0,586 - 0,0034 \text{ nwifeinc} + 0,038 \text{ edu} \\ & (0,154) \quad (0,0014) \quad (0,007) \\ & + 0,039 \text{ exp} - 0,00060 \text{ exp}^2 - 0,016 \text{ age} \\ & (0,006) \quad (0,00018) \quad (0,002) \\ & + 0,262 \text{ kidslt6} + 0,0130 \text{ kidsge6} \\ & (0,034) \quad (0,0132) \end{aligned}$$

avec $N = 753$, $R^2 = 0.264$.

Le modèle de probabilité linéaire a un certain nombre de défauts. Le premier de ceux-ci c'est que les valeurs prédites de la probabilité \hat{y} peuvent être inférieures à 0 ou supérieures 1, ce qui est naturellement impossible pour une probabilité. Le lien entre les variables explicatives et la variable expliquée ne peut donc pas être globalement linéaire. Pour résoudre ce problème, toutefois, il faudrait faire appel à des techniques d'estimations non linéaires qui sont au delà des objectifs de ce cours. Il serait, néanmoins, possible de prédire y en utilisant la règle suivant: $\tilde{y}_i = 1$ si $\hat{y}_i \geq 0,5$ et $\tilde{y}_i = 0$ si $\hat{y}_i < 0,5$. On peut alors calculer les fréquences de \tilde{y}_i en fonction des variables explicatives. Enfin, un autre problème est que le modèle linéaire de

probabilité viole les conditions de Gauss-Markov. En particulier,

$$\text{Var}(y|\mathbf{x}) = p(\mathbf{x})(1 - p(\mathbf{x}))$$

where $p(\mathbf{x}) = \beta_0 + \beta_1 \cdot x_1 + \dots + \beta_K \cdot x_K$. Donc, il y aura nécessairement de l'hétéroscédasticité dont il devrait être tenu compte.

8.5 Le R carré ajusté et la sélection de modèles

Le coefficient de détermination peut s'écrire de la manière suivante

$$R^2 = 1 - \frac{\text{SCR}/N}{\text{SCT}/N},$$

et il peut être interprété comme un estimateur convergent du R carré de la population, à savoir, $1 - \sigma_u^2/\sigma_y^2$, où σ_u^2 est la variance du terme d'erreur et σ_y^2 la variance de la variable dépendante. Si le nombre de paramètres à estimer converge vers le nombre d'observations, le coefficient de détermination tend vers 1 (puisque le SCR tend vers 0). Ce problème est dû à ce que les estimateurs de σ_u^2 et de σ_y^2 utilisés dans le formule du R carré sont des estimateurs biaisés. Si on remplace ces estimateurs biaisés par des estimateur non biaisés, on obtient le R-barre carré (ou R carré ajusté). Formellement, il est défini par

$$\bar{R}^2 = 1 - \frac{\text{SCR}/(N - K - 1)}{\text{SCT}/(N - 1)}.$$

Cette version du R-carré n'est pas nécessairement croissante avec le nombre de variables explicatives; elle peut même être négative. En fait, on peut montrer que l'ajout d'une variable explicative fait augmenter le R-barre carré si et seulement si le t-ratio correspondant à cette variable est supérieur à 1 en valeur absolue.

Exemple 9.12: Le R-barre carré peut être utilisé pour sélectionner des modèles non emboîtés.avec des nombre différents de variables explicatives. Soit deux modèles qui expliquent l'intensité en R&D (c'est-à-dire le rapport

entre les dépenses en R&D et le chiffre d'affaire de la firme) en fonction du chiffre d'affaire de la firme:

$$\text{intensrd} = \beta_0 + \beta_1 \cdot \log(\text{ventes}) + u,$$

$$\text{intensrd} = \beta_0 + \beta_1 \cdot \text{ventes} + \beta_2 \cdot \text{ventes}^2 + u.$$

Ces deux modèles sont non emboîtés dans le sens que l'un ne peut pas être vu comme le cas particulier de l'autre et ils n'ont pas le même nombre de variables explicatives. La sélection du modèle peut se faire sur la base du R-barre carré: la spécification qui sera choisie sera celle ayant le R-barre carré le plus élevé.

8.6 Problèmes de spécification

On parle de problème de spécification lorsque la forme fonctionnelle du modèle estimé n'est pas la même que celle du vrai modèle qui satisfait les hypothèses 5.1-5.4 (et éventuellement 5.5 et 5.6). Une erreur de forme fonctionnelle mènera à des estimateurs qui sont biaisés et non convergents.

Exemple 9.13: Le modèle suivant est le vrai modèle en ce sens que les hypothèses 5.1-5.4 sont satisfaites:

$$\log(\text{sal}) = \beta_0 + \beta_1 \cdot \text{edu} + \beta_2 \cdot \text{exp} + \beta_3 \cdot \text{exp}^2 + u, \quad (8.4)$$

avec $\beta_3 \neq 0$. En particulier, les années d'expérience ont un effet quadratique sur le logarithme du salaire. Si le modèle estimé est

$$\log(\text{sal}) = \beta_0 + \beta_1 \cdot \text{edu} + \beta_2 \cdot \text{exp} + u, \quad (8.5)$$

par exemple, la forme fonctionnelle ne sera pas correcte et les hypothèses 5.1-5.4 ne seront pas satisfaites.

Pour identifier ce type de problème, il convient d'estimer un modèle plus général que (8.5) ou (??) en incorporant un polynôme d'ordre suffisamment élevé en exp . La pertinence de ces termes quadratiques peut être testée à

l'aide de tests de Student ou de Fisher. Plus généralement, les problèmes de spécification peuvent être identifiés grâce à des tests plus sophistiqués comme le **test RESET** (pour REgression Specification Error Test). On considère le modèle suivant:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_K x_K + u$$

et on calcule les valeurs prédites \hat{y} . On estime le modèle étendu à un polynôme des valeurs prédites suivant:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_K x_K + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + u.$$

Le test RESET consiste alors à tester l'hypothèse $\delta_1 = \delta_2 = 0$ à l'aide d'un test de Fisher. Enfin, soulignons qu'il existe également des tests permettant de tester l'un contre l'autre des modèles qui ne sont pas emboîtés.

Appendix A

Formules usuelles

Les formules suivantes sont régulièrement utilisées dans le cours:

$$\begin{aligned}
 \sum_{i=1}^N x_i(x_i - \bar{x}) &= \sum_{i=1}^N x_i(x_i - \bar{x}) - \bar{x} \left(\sum_{i=1}^N x_i - \sum_{i=1}^N x_i \right) \\
 &= \sum_{i=1}^N x_i(x_i - \bar{x}) - \bar{x} \left(\sum_{i=1}^N x_i - N\bar{x} \right) \\
 &= \sum_{i=1}^N x_i(x_i - \bar{x}) - \bar{x} \left(\sum_{i=1}^N (x_i - \bar{x}) \right) \\
 &= \sum_{i=1}^N (x_i(x_i - \bar{x}) - \bar{x}(x_i - \bar{x})) \\
 &= \sum_{i=1}^N (x_i - \bar{x})^2 \\
 \sum_{i=1}^N x_i(y_i - \bar{y}) &= \sum_{i=1}^N x_i(y_i - \bar{y}) - \bar{x} \left(\sum_{i=1}^N y_i - \sum_{i=1}^N y_i \right) \\
 &= \sum_{i=1}^N x_i(y_i - \bar{y}) - \bar{x} \left(\sum_{i=1}^N y_i - N\bar{y} \right) \\
 &= \sum_{i=1}^N x_i(y_i - \bar{y}) - \bar{x} \left(\sum_{i=1}^N (y_i - \bar{y}) \right) \\
 &= \sum_{i=1}^N (x_i(y_i - \bar{y}) - \bar{x}(y_i - \bar{y})) \\
 &= \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})
 \end{aligned}$$

Appendix B

Démonstration.

Les résidus sont définis par

$$\hat{u}_i = u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1)x_i.$$

Or:

$$\sum_{i=1}^n \hat{u}_i = 0 = \bar{u} - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1)\bar{x}$$

en vertu des propriétés algébriques des MCO (la somme des résidus est nulle), où $\bar{u} = \sum_{i=1}^n u_i$ et $\bar{x} = \sum_{i=1}^n x_i$. Donc, en soustrayant la seconde expression à la première, on obtient:

$$\hat{u}_i = (u_i - \bar{u}) - (\hat{\beta}_1 - \beta_1)(x_i - \bar{x}).$$

En prenant le carré des membres de droite et gauche, on obtient:

$$\hat{u}_i^2 = (u_i - \bar{u})^2 + (\hat{\beta}_1 - \beta_1)^2 (x_i - \bar{x})^2 - 2(\hat{\beta}_1 - \beta_1)(u_i - \bar{u})(x_i - \bar{x}).$$

En sommant sur i , cette expression devient:

$$\begin{aligned} \sum_{i=1}^n \hat{u}_i^2 &= \sum_{i=1}^n (u_i - \bar{u})^2 + \sum_{i=1}^n (\hat{\beta}_1 - \beta_1)^2 (x_i - \bar{x})^2 \\ &\quad - 2 \sum_{i=1}^n (\hat{\beta}_1 - \beta_1)(u_i - \bar{u})(x_i - \bar{x}). \end{aligned}$$

Et en prenant l'espérance,

$$E\left(\sum_{i=1}^n \widehat{u}_i^2\right) = A + B + C.$$

avec

$$\begin{aligned} A &= E\left(\sum_{i=1}^n (u_i - \bar{u})^2\right), \\ B &= E\left(\sum_{i=1}^n (\widehat{\beta}_1 - \beta_1)^2 (x_i - \bar{x})^2\right), \\ C &= -2E\left(\sum_{i=1}^n (\widehat{\beta}_1 - \beta_1) (u_i - \bar{u}) (x_i - \bar{x})\right) \end{aligned}$$

Enfin, par un résultat bien connu de statistique, on a:

$$A = (n - 1)\sigma^2.$$

De plus:

$$\begin{aligned} B &= \sum_{i=1}^n (x_i - \bar{x})^2 E\left((\widehat{\beta}_1 - \beta_1)^2\right) \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 E\left(\left(\frac{\sum_{i=1}^n (x_i - \bar{x}) u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)^2\right) \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} E\left(\left(\sum_{i=1}^n (x_i - \bar{x}) u_i\right)^2\right) \\ &= \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \sum_{i=1}^n (x_i - \bar{x})^2 E(u_i^2) \\ &= \sigma^2 \end{aligned}$$

où l'avant dernière ligne utilise le fait que $E((x_i - \bar{x})(x_j - \bar{x})u_i u_j) = 0$.

Finalement,

$$\begin{aligned}
C &= -2E \left(\sum_{i=1}^n \left(\frac{\sum_{i=1}^n (x_i - \bar{x})(u_i - \bar{u})}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) (u_i - \bar{u})(x_i - \bar{x}) \right) \\
&= \frac{-2}{\sum_{i=1}^n (x_i - \bar{x})^2} E \left(\sum_{i=1}^n \sum_{j=1}^n (x_i - \bar{x})(u_i - \bar{u})(u_j - \bar{u})(x_j - \bar{x}) \right) \\
&= \frac{-2}{\sum_{i=1}^n (x_i - \bar{x})^2} E \left(\sum_{i=1}^n \sum_{j=1}^n (x_i - \bar{x})(u_i - \bar{u})(u_j - \bar{u})(x_j - \bar{x}) \right) \\
&= -2\sigma^2
\end{aligned}$$

Et en fin de compte,

$$E \left(\sum_{i=1}^n \hat{u}_i^2 \right) = (n-1)\sigma^2 + \sigma^2 - 2\sigma^2 = (n-2)\sigma^2.$$

Bibliography

- [1] Hamermesh D.S. & J.E. Biddle (1994), "Beauty and the Labor Market", *American Economic Review*, 84, 1174-1194.
- [2] Hayashi F. (2000), *Econometrics*, Princeton University Press.
- [3] Mroz T. (1987), "The Sensitivity of An Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions", *Econometrica*, 55, 765-799.
- [4] Wooldridge J. (2010), *Introduction to Econometrics: A Modern Approach*, MIT Press.