

# **Econométrie 1**

---

F. Karamé

# Introduction

## 1 Les données

Les données sont au centre de la réflexion économétrique. En effet, ce sont elles qui vont permettre de mesurer le phénomène étudié et ses déterminants. **C'est la « réalité » que le modèle économétrique cherche à représenter.** Dès lors, **un vrai travail de recherche, de construction quand c'est possible, mais surtout de sélection, d'interprétation, de compréhension et d'analyse critique des données doit être mené en préalable à toute étude,** afin de connaître **les richesses et les limites des données que l'on utilise. Nous verrons également plus tard qu'elles conditionneront les choix méthodologiques effectués.**

On considérera ici la convention suivante : **les données sont des tableaux avec les observations en ligne et les variables en colonnes.**

Il existe différents types de données permettant différents types d'analyse.

### 1.1 La nature des observations

- **Les séries temporelles ou chronologiques** correspondent à des observations répétées de variables (généralement des agrégats macro-économiques) à intervalles temporels réguliers (le mois, le trimestre, le semestre, l'année). On les note généralement avec un indice  $t$  faisant référence à la date (avec  $t = 1, \dots, T$ ).
- **Les coupes instantanées ou transversales ou encore en coupe** correspondent à l'observation à un moment donné de différents individus (entreprises, ménages, secteurs, pays, ...). On les note généralement avec un indice  $i$  faisant référence au numéro de l'individu observé dans l'échantillon (avec  $i = 1, \dots, N$ ).
- Enfin, **les données de panel** ou encore **individuelles-temporelles** intègrent les deux dimensions individuelles et temporelles et permettent le suivi des variables caractérisant des individus au cours du temps. Les données sont double-indicées en  $i$  pour les individus et en  $t$  pour les dates (avec  $t = 1, \dots, T$  et  $i = 1, \dots, N$ ).

La première et dernière catégories de données feront l'objet de techniques spécifiques. Certaines données, comme les données financières, peuvent ne pas être observées à intervalles réguliers. Elles feront également l'objet de méthodes spécifiques.

## 1.2 La nature des variables

Beaucoup de variables sont **quantitatives**, c'est-à-dire mesurent des phénomènes en prenant des valeurs numériques réelles. Par exemple, le montant de consommation, de l'investissement, ... D'autres sont **qualitatives** et mesurent les phénomènes au moyen de valeurs numériques entières. Par exemple, le sexe : codage à 1 pour les hommes et 2 pour les femmes.

Une variable qualitative peut très bien être utilisée pour expliquer un phénomène quantitatif. Par exemple, la disparité de salaires s'explique en fonction du sexe, du diplôme, ... Cela ne soulève pas de problème particulier d'un point de vue méthodologique.

A l'inverse, on peut aussi chercher à expliquer une variable qualitative. Par exemple, les déterminants de l'obtention d'un CDI. Cela pose alors des problèmes spécifiques qui seront traités avec des méthodes spécifiques.

## 1.3 Collecte et construction des données

Les données individuelles peuvent concerner des individus, ménages ou entreprises. Elles peuvent être exhaustives (obtenues par recensement) ou d'enquêtes. Dans ce dernier cas, il faut s'interroger sur la signification des réponses, à mettre en relation avec

- la rédaction du questionnaire et le type de questions posées : questions parfois sensibles du type salaires, impôts, ... qui vont conditionner l'honnêteté de la réponse
- la méthode d'échantillonnage utilisée et la fréquence des interrogations : par exemple problème de l'impôt statistique pour les petites entreprises.
- le mode d'interview utilisé : face à face, téléphone, courrier, ...

Enfin, on s'expose au problème de traitement, lourd selon la taille de l'échantillon, coûteux en temps d'exploitation et économiquement, aux problèmes de non réponses, d'erreur de mesure, .... Ces différents points feront l'objet d'un cours spécifique de théorie des sondages au second semestre.

Les données agrégées peuvent être à la fois temporelles ou en coupe, et concerner des agrégats macro-économiques (production, investissement, exportations ou importations, ...) ou des regroupements d'individus (régions, secteurs, ...). A ce niveau, on peut s'intéresser à des modèles macro-économiques permettant d'effectuer des simulations de politiques économiques. Cependant plusieurs inconvénients apparaissent. D'une part, les données agrégées sont moins précises (et moins potentiellement moins riches puisque qu'en sommant sur les individus, on perd l'information capturée par l'hétérogénéité individuelle). D'autre part, on s'expose à des biais d'agrégation dans la mesure où les comportements des individus agrégés sont hétérogènes.

## 2 La notion de modèle économétrique

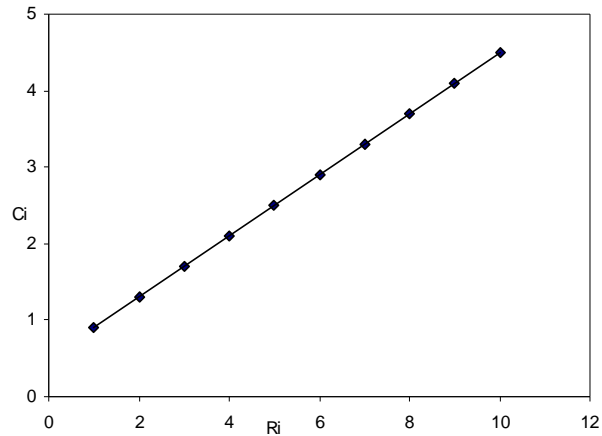
Prenons le cas de deux variables :  $C$  la consommation et  $R$  le revenu. Et supposons que l'on observe ces 2 informations pour  $N$  individus, notés génériquement  $i$ .

### 2.1 Les différents types de relations entre variables

On peut dire qu'il existe trois types de relations possibles entre ces deux variables.

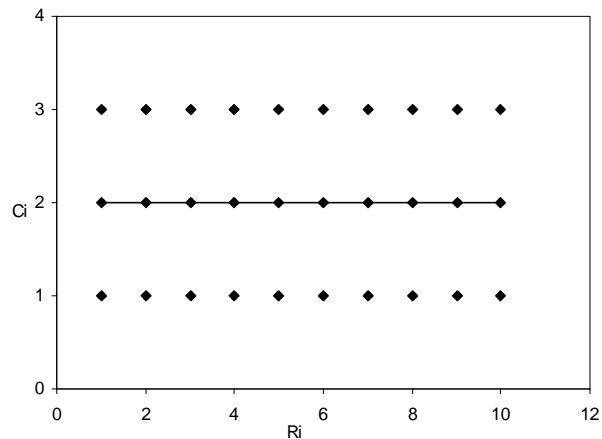
La première est la **relation fonctionnelle déterministe** : à une valeur de  $R_i$  correspond une valeur de  $C_i$ .

La représentation graphique associée dans le cas où la relation est linéaire, est alors :



**Graphique 1 : relation déterministe fonctionnelle**

La deuxième est l'absence de toute relation entre les deux variables. On parle alors d'**indépendance**. La représentation graphique donne alors :

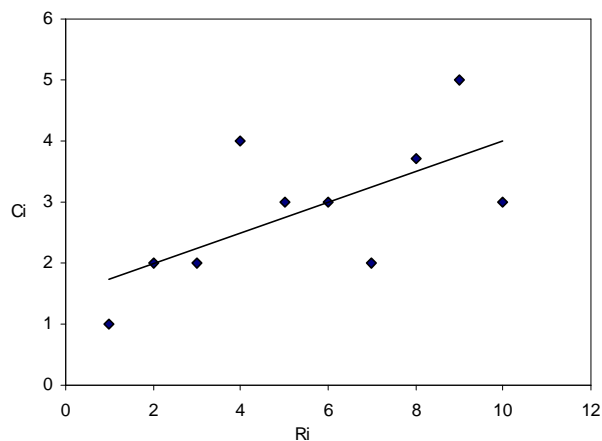


**Graphique 2 : indépendance**

A toute valeur de  $R_i$  peut correspondre une infinité de valeurs de  $C_i$ .

Le troisième type de relation se situe entre les deux. On fait l'hypothèse d'une relation fonctionnelle, mais celle-ci n'est pas parfaitement vérifiée pour chaque observation **parce que la réalité économique est trop complexe à appréhender**.

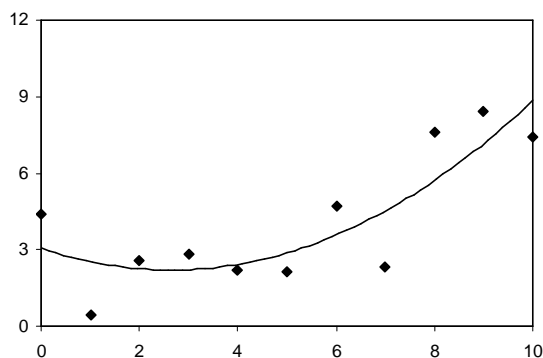
La représentation graphique est alors :



**Graphique 3 : corrélation linéaire positive**

On dit alors que les deux variables sont corrélées. Le graphique précédent montre que la relation représentée par la droite ne passe pas exactement par tous les points de l'échantillon, mais que les 2 variables semblent montrer « un certain degré de dépendance » et que la droite passe « au milieu du nuage de points ».

Il existe deux types de corrélations : **la corrélation linéaire et la corrélation non linéaire**. Le graphique 3 fournit une représentation de corrélation linéaire positive. Le graphique 4 présente une corrélation non linéaire positive.



**Graphique 4 : corrélation non linéaire positive**

Il existe bien évidemment les cas de corrélation négative (linéaire et non linéaire). Le graphique 2 représente un cas de corrélation nulle. Le graphique 1 représente le cas d'une corrélation linéaire parfaite.

Pour mesurer la corrélation linéaire entre deux variables  $x$  et  $y$ , on utilise **le coefficient de corrélation linéaire simple** :

$$r_{x,y} = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Ce coefficient est compris entre  $-1$  et  $1$ .

- Ainsi, pour un coefficient de corrélation linéaire simple proche de  $1$ , les deux variables sont très linéairement corrélées positivement (quand une variable évolue dans un sens, l'autre évolue dans le même sens).
- Pour un coefficient de corrélation linéaire simple proche de  $-1$ , les deux variables sont très linéairement corrélées négativement (quand une variable évolue dans un sens, l'autre évolue dans le sens contraire).
- Enfin, pour un coefficient de corrélation linéaire nul, les deux variables sont non corrélées linéairement. Quand une variable évolue dans un sens, on ne peut rien dire sur le sens d'évolution de l'autre variable (dans le cadre d'une relation linéaire).

Il convient de noter qu'une corrélation linéaire nulle ne signifie pas l'absence de corrélation non linéaire entre les deux variables. **Autrement dit, l'absence de corrélation linéaire n'implique pas l'indépendance.**

## 2.2 Implications

Plaçons-nous dans le cas de la corrélation, qui permet de retrouver les 2 cas extrêmes de relation mathématique et d'indépendance. On considère que chaque ménage est ici un cas particulier d'une règle générale spécifiée dans l'équation suivante :

$$C_i = f(R_i) + u_i$$

On observe  $R_i$  et  $C_i$  pour chaque ménage  $i$  (ou plus généralement observation  $i$ ), la fonction  $f(.)$  est supposée. On a également supposé que  $C_i$  est la variable à expliquer et  $R_i$  la variable explicative. Un modèle peut comporter plusieurs variables explicatives. **Ici, on a fait un pas supplémentaire vers un schéma explicatif puisque contrairement à la corrélation qui est une relation symétrique, on suppose un sens de causalité de la variable  $R$  vers la variable  $C$ .**

**Pour gérer les erreurs et l'incertitude inhérentes au modèle que l'on s'est donné, on utilise une approche probabiliste qui considère  $u_i$  comme une variable aléatoire appelée perturbation aléatoire.  $u_i$  représente et capture :**

- l'oubli de variables dans le modèle,
- l'absence de variables explicatives remplacées par des approximations,
- tous les facteurs du second ordre qui ne sont pas pris en compte dans le schéma explicatif de notre modèle,
- les erreurs de mesure sur les variables,
- l'incertitude sur la forme du modèle  $f$ ,

en un mot, **tout ce qui nous éloigne de la vraie équation du processus qu'on cherche à expliquer.**

$u_i$  est un processus théorique, non observable et il en existe une réalisation particulière pour chaque observation  $i$ .

On fixera certaines propriétés de  $u_i$ . Ainsi, lorsqu'on spécifie comme loi des perturbations, une loi normale  $N(0, \sigma^2)$  par exemple, on qualifie l'approche de **paramétrique**. Dans le cas contraire, elle est **semi-paramétrique**.

### Remarques :

- Comme  $u_i$  est aléatoire,  $C_i$  l'est aussi d'après le modèle. On considère alors que l'observation de la consommation de l'observation  $i$  est fournie par la réalisation de la variable aléatoire  $C_i$  lors du tirage aléatoire de notre échantillon dans la population.
- Si on connaissait toutes les variables explicatives de  $C_i$ , on pourrait **en théorie construire la relation fonctionnelle déterministe**, c'est-à-dire non aléatoire. Même si de toutes façons cela n'est pas possible, on ne le voudrait pas forcément, dans la mesure où on recherche **une approximation acceptable de la réalité**.

D'où un premier essai de définition...

## 2.3 Définition d'un modèle économétrique

Un modèle économétrique est une équation dont le rôle est « **d'expliquer** » un phénomène grâce à des variables que l'on juge déterminantes au premier plan. L'objectif en est de capturer le ou les faits les plus marquants de la réalité qu'il cherche à représenter.

Le modèle économétrique est une « histoire » qui s'applique à chacune des observations de l'échantillon, à une erreur possible près représentée par la perturbation aléatoire.

Les paramètres inconnus du modèle mesurent l'impact des variables explicatives sur la variable à expliquer. **C'est l'économétrie qui va permettre une évaluation de ces**

**paramètres en utilisant l'information contenue dans toutes les observations de l'échantillon et donc d'analyser et d'utiliser les résultats obtenus.**

**Du fait des différentes sources d'incertitude qui l'entourent, il faut comprendre et accepter que tout modèle économétrique est une représentation simplifiée voire simpliste d'une réalité complexe.** On pourrait donc le considérer dès le départ et quels que soient nos efforts, comme erroné. L'idée étant cependant que cette représentation soit acceptable au regard de critères techniques et interprétatifs. Le but étant, au travers du modèle économétrique et de son évaluation, de « **raconter une histoire convaincante** » pour expliquer le phénomène étudié.

## 2.4 Discussion sur la notion de modèle économétrique

**Un modèle économétrique est donc une représentation forcément simplifiée d'un phénomène** (expl : la consommation d'orange dépend du prix des oranges). En effet, on pourrait dire aussi que cela dépend de la consommation de café, du prix des pommes, du temps pour la récolte, du prix de l'essence ...

Cette approche simplifiée dépend aussi de façon cruciale des données dont on dispose ou dont on peut disposer. Il faut y inclure toutes celles dont on pense qu'elles jouent un rôle vraiment important et ne pas s'occuper des autres. Ces dernières sont alors incluses dans la perturbation aléatoire.

**Une question se pose alors concernant la fiabilité d'un modèle économétrique pour raconter des histoires et s'il peut constituer une représentation acceptable de la « réalité ».**

Selon Popper [1959] et Friedman [1953], un modèle est forcément quelque chose de simple car c'est plus facile à comprendre, à faire comprendre et à tester. Mais cela conduit cependant à deux critiques importantes.

- **La sur-simplification :** dans l'exemple précédent, la spécification du modèle économétrique est trop simpliste. En règle générale, deux écoles s'affrontent pour spécifier un modèle :
  - la première dit qu'il faut commencer par un modèle simple et le compliquer progressivement (Koopmans [1957], conception ascendante de la modélisation).
  - La seconde dit qu'il faut partir d'un modèle très général et le simplifier progressivement sur la base des données utilisables et de tests statistiques (Sargan puis Hendry, conception descendante de la modélisation).
  - Les deux s'accordent pour retenir le modèle présentant un **arbitrage acceptable entre parcimonie et réalisme**.
- **Les hypothèses irréalistes :** sur ce point, Friedman réplique en soulignant que la question n'est pas tant de savoir si les hypothèses formulées par les théories sont réalistes sur le plan descriptif (ce qu'elles sont rarement) mais plutôt si elles constituent des approximations suffisamment bonnes pour répondre à la question que l'on se pose. Et l'on ne peut y répondre qu'en regardant si une théorie fonctionne, c'est-à-dire si elle fournit des prédictions suffisamment précises.

## 3 Le modèle linéaire

La première étape de l'analyse économétrique consiste à dégager les mécanismes théoriques à l'œuvre pour expliquer le phénomène qui nous intéresse.

La deuxième étape de l'analyse consiste ensuite à en déduire la forme de la relation entre les variables explicatives supposées et la variable à expliquer. **C'est généralement une forme linéaire qui est retenue, du fait de sa simplicité.** Notre modèle s'écrit donc :

$$C_i = a + bR_i + u_i$$

Dans la nature, il n'y a absolument aucune raison qu'une relation soit linéaire en fonction des coefficients. En effet, les implications d'une telle hypothèse sont fortes. Ici, si le revenu varie positivement d'une unité, la consommation varie de  $b$  unités. Inversement, si le revenu varie négativement d'une unité, alors la consommation varie de  $-b$ . Retenir une telle spécification implique donc une réponse symétrique de la consommation à une variation du revenu, ce que l'on pourrait considérer comme une hypothèse forte.

Cependant, retenir une forme linéaire pour la spécification du modèle économétrique a aussi le mérite d'éviter d'avoir recours à des méthodes économétriques plus complexes, tout en n'interdisant pas forcément de s'intéresser à des phénomènes fortement non linéaires. **L'important pour la suite étant que le modèle soit linéaire en les coefficients ; rien n'interdit de faire apparaître des variables ayant subi des transformations non linéaires...**

**Expl1** : la relation entre la formation du salaire et l'expérience professionnelle.

$$w_i = a s_i + b \text{étude}_i + c \text{exp}_i + d + u_i$$

Un modèle linéaire pour cette représentation n'est pas réaliste dans la mesure où il revient à dire qu'une année d'expérience supplémentaire en début ou en fin de carrière va induire une augmentation identique du salaire. Or on sait que le rendement marginal de l'expérience est décroissant : ainsi, les salaires croissent rapidement avec l'expérience en début de carrière et de façon beaucoup plus lente par la suite. Ainsi, le modèle estimé sera plutôt :

$$w_i = a s_i + b \text{étude}_i + c \text{exp}_i + c' (\text{exp}_i)^2 + d + u_i$$

On constate qu'il s'agit encore d'un modèle linéaire sur le logarithme du salaire. On conclura ici que le rendement marginal de l'expérience est décroissant si l'estimation de  $c'$  est négative.

**Expl2** : L'effet sur le niveau de consommation des ménages du nombre des enfants.

$$C_i = \alpha + \beta R_i + \gamma \text{nenf}_i + u_i$$

Il n'y a aucune raison de supposer qu'un enfant supplémentaire va avoir un effet identique s'il y en a déjà 1 ou 8. Mais comment prendre en compte cet effet non linéaire car il existe probablement des effets de discontinuités rendant la prise en compte de ces effets par une forme quadratique difficile (achat d'une voiture plus grande, d'une maison plus grande, ...). Une possibilité est ici de poser les variables  $\text{nenf}0_i$  égale à 1 pour des ménages n'ayant pas d'enfant et à 0 sinon,  $\text{nenf}12_i$  égale à 1 pour les ménages ayant 1 ou 2 enfants et à 0 sinon et  $\text{nenf}3_i$  égal à 1 pour les ménages ayant 3 enfants et plus et à 0 sinon. Le modèle devient alors :

$$C_i = \beta R_i + \gamma_1 \text{nenf}0_i + \gamma_2 \text{nenf}12_i + \gamma_3 \text{nenf}3_i + u_i$$

Le modèle est toujours linéaire en ces coefficients, alors qu'on a modélisé plusieurs discontinuités dans les données. Les effets différenciés portent sur le niveau moyen de la consommation et sont pris en charge par les estimations de  $\gamma_1, \gamma_2$  et  $\gamma_3$  (voir exercice du TD1).

**Expl3** : la fonction de production de Cobb-Douglas : c'est une relation non linéaire entre le niveau du produit  $Q_t$  et les facteurs de production capital  $K_t$  et travail  $L_t$  :

$$Q_t = A L_t^\alpha K_t^\beta$$

Ce modèle non linéaire est équivalent à un modèle linéaire lorsqu'on en prend le logarithme :

$$\ln Q_t = \ln(A) + \alpha \ln(L_t) + \beta \ln(K_t) + u_t$$



Cependant, l'exemple suivant montre que se ramener à une forme linéaire n'est pas toujours possible.

**Expl4** : l'estimation d'une fonction de production CES :

$$Q_t = \left[ \mu L_t^{-\rho} + (1 - \mu) K_t^{-\rho} \right]^{\frac{1}{\rho}}$$

Il n'y a pas de traduction linéaire de ce modèle. Il faudra donc avoir recours à d'autres méthodes que celles présentées ici pour l'estimer.

## 4 L'estimation des paramètres inconnus

La question posée, les données utilisées et la complexité de la spécification du modèle à estimer conditionnent la méthode d'estimation à utiliser.

Plaçons-nous dans un cas simple et supposons que l'on cherche à présent à estimer le modèle suivant :

$$y_i = a + bx_i + u_i$$

On dispose des données concernant la consommation et le revenu pour  $N$  ménages à une date donnée. Les données dont on dispose sont en réalité un échantillon supposé représentatif tiré dans une population dont on cherche à caractériser les comportements de consommation au moyen des paramètres inconnus  $a$  et  $b$ . Ainsi cette population peut être l'ensemble des ménages belges de Los Angeles, des célibataires de Clermond-Ferrand, ... Pour cette population, on suppose qu'il existe une loi qu'il s'agit de connaître le mieux possible au travers de l'échantillon tiré. Pour cela, on va calculer des estimations des paramètres inconnus  $a$  et  $b$  à partir des observations de cet échantillon.

### 4.1 La méthode d'estimation des MCO

Les estimations sont obtenues grâce à une **méthode d'estimation**. Ici, on retiendra **la méthode des moindres carrés ordinaires (MCO ou Ordinary Least Squares en anglais) car le modèle est linéaire en les coefficients**.

Si ça n'avait pas été le cas, on aurait choisi une autre méthode plus adaptée (moindres carrés non linéaires, maximum de vraisemblance, méthode des moments généralisée, ...).

Le principe de la méthode est le suivant. Pour une droite permettant d'ajuster le nuage de points, on va minimiser la distance entre chaque point de la variable à expliquer et le modèle. **On va donc chercher une valeur pour les paramètres qui confère un rôle (total quadratique) minimum aux perturbations aléatoires à chaque observation**. Le critère d'estimation est donc de trouver  $a$  et  $b$  tels que :

$$\{\hat{a}, \hat{b}\} = \arg \min_{a,b} \sum_{i=1}^N u_i^2 = \arg \min_{a,b} \sum_{i=1}^N (y_i - a - bx_i)^2$$

Les estimations sont obtenues à partir de **formules théoriques** qui dépendent des variables aléatoires caractérisant le problème étudié (en particulier  $y_i$ ) : **ce sont des estimateurs, qui sont donc eux-mêmes des variables aléatoires**. En effet, comme  $u_i$  est aléatoire,  $y_i$  l'est aussi, et comme l'estimateur est fonction de  $y_i$ ... On pourra noter les estimateurs  $\hat{A}$  et  $\hat{B}$ .

La valeur des estimations, qu'on note  $\hat{a}$  et  $\hat{b}$ , dépend des observations de l'échantillon. Avec un autre échantillon, l'application des formules des estimateurs donnerait des estimations **numériquement** différentes mais d'un ordre de grandeur relativement comparable si les deux échantillons sont représentatifs.

### Remarque :

On manipule donc trois types de paramètres :

- $a$  et  $b$  les « vrais » paramètres inconnus que l'on n'observe jamais, mais qui peuvent être utilisés dans les calculs théoriques.
- $\hat{A}$  et  $\hat{B}$  les estimateurs, qui sont des variables aléatoires.
- $\hat{a}$  et  $\hat{b}$  les estimations des paramètres inconnus, c'est-à-dire des valeurs numériques particulières associées à l'échantillon, des réalisations particulières des estimateurs associées à l'échantillon.

Dans les cours d'économétrie traditionnels et à l'inverse des cours de probabilités, on fait rarement la différence dans les notations entre les estimations et les estimateurs. Ce sera donc à vous de bien faire attention lors des calculs concernant les propriétés théoriques des estimateurs de savoir quel objet vous manipulez.

## 4.2 Les propriétés des estimateurs

Les estimateurs que l'on utilise pour évaluer les paramètres inconnus sont associés à la méthode d'estimation choisie et sont intéressants pour les propriétés qu'ils possèdent et qui peuvent varier d'une méthode d'estimation à une autre. Ainsi :

- lorsqu'ils fournissent en moyenne la valeur des paramètres inconnus, on dit alors qu'ils sont **sans biais (unbiased)**.

**Définition :** l'estimateur  $\hat{A}$  est sans biais si  $E(\hat{A}) = a$ .

- Les estimations sont réalisées avec des marges d'erreur dues aux imprécisions associées aux incertitudes entourant le modèle. Ces marges d'erreur diminuent avec la taille de l'échantillon, et on pourrait même dire que lorsque la taille de l'échantillon tend vers l'infini (c'est-à-dire la taille de la population), on pourrait connaître parfaitement la valeur des paramètres inconnus. On dit alors que les estimateurs sont **convergen**t (**consistent**).

**Définition :** l'estimateur  $\hat{A}$  est convergent si  $V(\hat{A}) \xrightarrow{N \rightarrow \infty} 0$ .

- Lorsque parmi deux estimateurs du même paramètre, un possède une marge d'erreur inférieure à l'autre, on dit qu'il est **meilleur (best)**.

**Définition :** l'estimateur  $\hat{A}$  est meilleur que  $\hat{A}'$  si  $V(\hat{A}) < V(\hat{A}')$ .

- Enfin, si dans une classe d'estimateurs sans biais, un estimateur possède la plus petite variance possible, on dit qu'il est **efficace (efficient)**.

**Si la méthode des MCO est plus utilisée que les autres, c'est parce que ses estimateurs possèdent, sous certaines conditions, les meilleures propriétés possibles.**

## 5 La démarche économétrique

Dans l'absolu, la démarche économétrique doit être la suivante.

- 1) Formuler clairement la question que l'on se pose.
- 2) Construire ou partir d'une théorie expliquant les différents mécanismes que l'on cherche à mettre en évidence.
- 3) Sélectionner les données dont on a besoin, les construire et les étudier.
- 4) Formuler le modèle économétrique à partir de la théorie, de la disponibilité des données et de ses propres *a priori* (interactions avec l'étape 3 pour ce qui est de la collecte et de la construction des données dont on a besoin).
- 5) Estimer le modèle.
- 6) Tester la pertinence (statistique et logique) du modèle. S'il est rejeté par les tests statistiques, revenir à l'étape 4 et éventuellement à l'étape 3.

- 7) S'il est accepté, l'utiliser pour faire de la prévision ou des recommandations. Cette étape permet donc de « valider » la théorie utilisée pour formuler le modèle, et éventuellement d'orienter les évolutions théoriques à partir de ce que disent les données. On est alors en présence d'une vraie interaction entre la théorie et l'économétrie appliquée.

# Chapitre 1

## Le modèle de régression linéaire

### Introduction et notations

Soit le modèle de régression linéaire multiple fourni par la théorie économique, et dans le cadre de la démarche économétrique énoncée précédemment :

$$y_t = a + b_1 x_{1,t} + b_2 x_{2,t} + \dots + b_{k-1} x_{k-1,t} + u_t$$

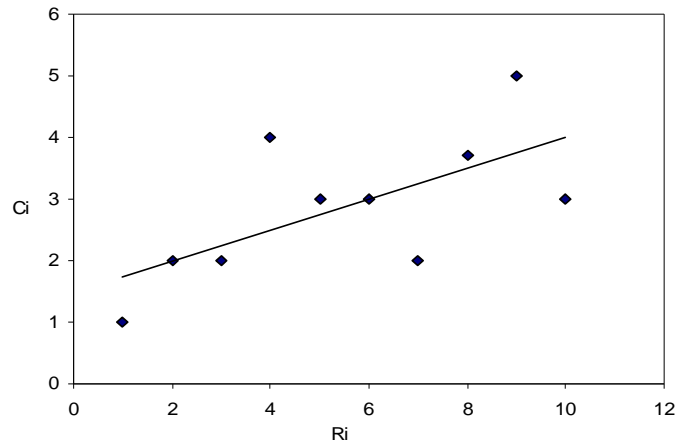
- $y_t$  est la réalisation observée en  $t$  de la variable à expliquer, appelée encore variable endogène ou variable dépendante.
- $x_{1t}, x_{2t}, \dots, x_{k-1t}$  sont les réalisations observées en  $t$  des variables explicatives, encore appelées encore variables exogènes ou variables indépendantes. Le modèle est multiple s'il y en a plus d'une.
- $u_t$  est la réalisation non observée en  $t$  de la perturbation aléatoire.
- $a$  et  $b_1, \dots, b_{k-1}$  sont les paramètres inconnus dont on recherche la valeur.

On a vu dans l'introduction générale que ce modèle est considéré comme **linéaire**, car  $y_t$  est une fonction linéaire des paramètres inconnus.

Il est qualifié de simple lorsqu'il ne comporte qu'une seule variable explicative.

$$y_t = a + bx_t + u_t$$

Dans ce cas particulier, on va donc ajuster le nuage des observations par une droite dans le plan  $(x_t, y_t)$  et les paramètres inconnus s'interprètent donc comme la pente de la droite pour  $b$  et comme l'ordonnée à l'origine pour  $a$ .



Dans le cas du modèle de régression linéaire multiple, ce modèle comporte  $k$  variables explicatives ( **$y$  compris la constante**) et l'ajustement du nuage des  $T$  observations se fera par un (hyper-)plan de dimension  $k$  (le nombre des variables explicatives).

L'objectif est ici de fournir une estimation pour les paramètres  $a, b_1, \dots, b_{k-1}$ . Pour cela nous utiliserons la méthode des MCO pour déduire les formules des estimateurs et leurs propriétés.

**Le plan du chapitre est le suivant.**

- Nous allons envisager les différentes écritures possibles du modèle et particulièrement l'écriture générale matricielle (section 1).
- Il sera ensuite nécessaire de formuler différentes hypothèses concernant les perturbations aléatoires. Nous allons donc les énumérer et les interpréter (section 2).
- Nous allons ensuite trouver la solution du problème des MCO et l'expression des estimateurs (section 3).
- Nous en étudierons ensuite les propriétés théoriques à distance finie (section 4) puis asymptotiquement (section 5) en utilisant les différentes hypothèses formulées sur les perturbations aléatoires.
- Nous étudierons la possibilité d'estimateurs alternatifs comme celui du maximum de vraisemblance et ses relations avec l'estimateur des MCO (section 6) pour nous convaincre que c'est le meilleur estimateur possible.
- Enfin nous discuterons des prévisions dans un modèle économétrique (section 7).

## 1 Les différentes formulations du modèle de régression linéaire multiple

Pour passer aux notations matricielles, il est nécessaire d'appliquer l'écriture générique du modèle pour chaque observation :

$$\begin{aligned}
 y_1 &= a + b_1 x_{1,1} + b_2 x_{2,1} + \dots + b_{k-1} x_{k-1,1} + u_1 \\
 y_2 &= a + b_1 x_{1,2} + b_2 x_{2,2} + \dots + b_{k-1} x_{k-1,2} + u_2 \\
 &\dots \\
 y_t &= a + b_1 x_{1,t} + b_2 x_{2,t} + \dots + b_{k-1} x_{k-1,t} + u_t \\
 &\dots \\
 y_T &= a + b_1 x_{1,T} + b_2 x_{2,T} + \dots + b_{k-1} x_{k-1,T} + u_T
 \end{aligned}$$

On peut réécrire vectoriellement le modèle linéaire en gardant bien à l'esprit les formats des vecteurs :

$$\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_t \\ \dots \\ y_T \end{pmatrix} = a \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \\ \dots \\ 1 \end{pmatrix} + b_1 \begin{pmatrix} x_{1,1} \\ x_{1,2} \\ \dots \\ x_{1,t} \\ \dots \\ x_{1,T} \end{pmatrix} + \dots + b_{k-1} \begin{pmatrix} x_{k-1,1} \\ x_{k-1,2} \\ \dots \\ x_{k-1,t} \\ \dots \\ x_{k-1,T} \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \dots \\ u_t \\ \dots \\ u_T \end{pmatrix}$$

On peut donc réécrire le modèle linéaire multiple comme une combinaison linéaire dans  $\mathbb{R}^T$  :

$$y = a e_T + b_1 x_1 + \dots + b_{k-1} x_{k-1} + u$$

$(T \times 1) \quad (1 \times 1)(T \times 1) \quad (1 \times 1)(T \times 1) \quad (1 \times 1) \quad (T \times 1) \quad (T \times 1)$

avec

- $y$  le vecteur des  $T$  observations de la variable à expliquer.
- $e_T$  le vecteur de dimension  $T$  ne comportant que des 1. C'est une « variable » qui prend la même valeur pour toutes les observations. **Elle capture donc ce qui est commun à toutes les observations dans le phénomène à expliquer.**
- $x_1, \dots, x_{k-1}$  les  $(k-1)$  vecteurs des  $T$  observations des variables explicatives. Chaque variable prend des valeurs différentes entre les observations. **C'est ce qui caractérise l'hétérogénéité entre les observations dans le phénomène à expliquer.**
- **L'influence de chaque variable explicative sur le phénomène à expliquer est mesurée par les  $k$  paramètres inconnus  $a, b_1, \dots, b_{k-1}$  dont on recherche la valeur.** C'est l'hétérogénéité des déterminants dans l'ensemble des observations qui va expliquer l'hétérogénéité du phénomène à expliquer.
- $u$  le vecteur des  $T$  réalisations non observées de la perturbation aléatoire.

Posons à présent la matrice  $X$  composée des observations des  $k$  variables explicatives ( $y$  compris la constante) ainsi que  $\beta$  le vecteur contenant les paramètres inconnus à estimer :

$$X_{(T \times k)} = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{k-1,1} \\ \dots & \dots & \dots & \dots \\ 1 & x_{1,t} & \dots & x_{k-1,t} \\ \dots & \dots & \dots & \dots \\ 1 & x_{1,T} & \dots & x_{k-1,T} \end{pmatrix} \quad \text{et} \quad \beta_{(k \times 1)} = \begin{pmatrix} a \\ b_1 \\ \dots \\ b_{k-1} \end{pmatrix}$$

Il est important à ce stade d'envisager **les deux sens de lecture** de la matrice  $X$  :

- d'une part, une ligne correspond à une observation, caractérisée dans l'espace des variables  $\mathbb{R}^k$  (puisque ici une observation se définit avec  $k$  coordonnées). L'ajustement du nuage de points des observations se fera dans l'espace des variables, comme on l'a vu dans le cadre du modèle de régression linéaire simple.
- D'autre part, une colonne correspond à une variable, caractérisée dans l'espace des observations  $\mathbb{R}^T$  (puisque ici une variable se définit avec  $T$  coordonnées). Ce sont ces variables qu'on étudie dans cet espace lors de l'interprétation géométrique des MCO.

Le modèle se réécrit alors :

$$y = X \beta + u$$

$(T \times 1) \quad (T \times k)(k \times 1) \quad (T \times 1)$

On a présenté précédemment la matrice  $X$ , soit de façon encore condensée sous sa forme en colonne :

$$X_{(T \times k)} = \begin{pmatrix} e_T & x_1 & \dots & x_{k-1} \\ (T,1) & (T,1) & & (T,1) \end{pmatrix}$$

Sous cette forme, une colonne correspond à toutes les observations d'une variable.

Pour la suite, il est nécessaire de bien savoir manipuler ces différentes écritures et de bien garder à l'esprit le format des matrices afin d'éviter les erreurs, même si dans la suite nous nous concentrerons majoritairement sur l'écriture la plus compacte.

## 2 Les premières hypothèses sur les perturbations aléatoires

Avant de pouvoir tirer une conclusion quelconque sur la question que l'on se pose grâce au modèle économétrique que l'on a formulé, il faut avoir dépassé les étapes techniques de spécification du modèle. Sur le plan technique, une fois le modèle formulé, la démarche est la suivante :

- Enoncer un certain nombre d'hypothèses concernant les perturbations aléatoires (non observées).
- Utiliser une méthode d'estimation sous ces hypothèses.
- Tester la validité des hypothèses formulées une fois l'estimation du modèle réalisée, à partir d'une estimation des perturbations.
- Envisager la meilleure méthode d'estimation lorsque telle ou telle hypothèse n'est pas vérifiée.

Détaillons la liste et l'interprétation des hypothèses sur les perturbations.

### 2.1 La perturbation est d'espérance nulle

$$H_1 : E(u_{(T,1)} | X) = 0_{(T,1)}$$

L'espérance des perturbations conditionnellement aux variables explicatives est nulle **pour chaque réalisation**. L'ensemble des déterminants non retenus dans le modèle (et regroupés dans les perturbations) est d'espérance nulle, c'est-à-dire que **leurs effets sur la variable à expliquer des variables de second ordre se compensent entre eux à chaque observation**. Autrement dit, l'approximation constituée par le modèle correspond à la loi moyenne de la variable aléatoire à expliquer :

$$E(y | X) = X\beta$$

Dans le cas contraire, on ferait le modèle ferait une erreur systématique à chaque observation pour expliquer  $y$ .

### 2.2 Les variables exogènes sont des variables certaines

$$H_2 : E(X' \cdot u) = 0$$

Cette hypothèse implique que  $X$  et  $u$  ne sont pas linéairement corrélés. Cela signifie que l'approximation constituée par le modèle est telle que les déterminants de seconde importance de  $y$  que l'on a négligés et qui figurent donc dans la perturbation aléatoire ne sont pas liés aux variables explicatives de première importance  $X$  retenues dans le modèle. C'est cette implication qui est essentielle et qui doit être respectée si  $X$  est finalement une matrice constituée de variables aléatoires, car elle permet d'obtenir des estimateurs non biaisés.

Pour simplifier les calculs qui vont suivre sans avoir d'impact fondamental sur les résultats, on supposera :

$H_{2\text{-bis}}$  :  $X$  est une matrice de variables certaines.

## 2.3 $X$ est une matrice de plein rang colonne

$H_3$  :  $X$  est une matrice de rang égal à  $k$  c'est-à-dire de plein rang colonne.

On fait ici l'hypothèse que les colonnes de la matrice  $X$  (c'est-à-dire les variables) sont linéairement indépendantes entre elles.

Dans le cas contraire, cela signifierait qu'au moins une variable explicative pourrait s'écrire comme une combinaison linéaire d'autres variables explicatives du modèle. Dans un tel cas de figure, cette variable serait alors redondante et n'apporterait rien au modèle. De plus, cela nous empêcherait même d'estimer ses paramètres puisque le même phénomène interviendrait deux fois dans l'explication. On est alors dans le cas de multicollinéarité stricte et le modèle n'est pas identifiable.

Dans le cas du modèle de régression linéaire simple, cela signifie que les observations  $x_t$  possèdent une variance ( $s_x^2 = \frac{1}{T} \sum_{t=1}^T (x_t - \bar{x})^2 \neq 0$  i.e. elles ne sont pas toutes égales à  $\bar{x}$ ) et se

comportent de façon différente de la constante. Elles apportent donc une information supplémentaire et non redondante relativement à la constante du modèle (ce point sera plus compréhensible dans la section sur l'interprétation géométrique du modèle linéaire simple). C'est cette variabilité de comportement de la variable  $x$  qui va permettre d'expliquer la variable  $y$ .

## 2.4 L'hypothèse d'homoscédasticité et de non covariance des perturbations

$$H_4 : E \begin{pmatrix} u \cdot u' \\ (T,1) \quad (1,T) \end{pmatrix} = \sigma^2 I_T.$$

La matrice de variances-covariances des perturbations est une matrice scalaire, c'est-à-dire qu'elle s'écrit comme le produit d'un scalaire par la matrice identité. En effet, pour  $H_1$  vérifiée :

$$\begin{aligned} V \begin{pmatrix} u \\ (T,1) \end{pmatrix} &= E \begin{pmatrix} u \cdot u' \\ (T,1) \quad (1,T) \end{pmatrix} = E \left[ \begin{pmatrix} u_1 \\ \dots \\ u_t \\ \dots \\ u_T \end{pmatrix} \cdot \begin{pmatrix} u_1 & \dots & u_t & \dots & u_T \end{pmatrix} \right] = E \begin{bmatrix} u_1^2 & \dots & u_1 u_t & \dots & u_1 u_T \\ \dots & \dots & \dots & \dots & \dots \\ u_1 u_t & & u_t^2 & & u_t u_T \\ \dots & & \dots & \dots & \dots \\ u_1 u_T & \dots & u_t u_T & \dots & u_T^2 \end{bmatrix} \\ &= \begin{bmatrix} E(u_1^2) & \dots & E(u_1 u_t) & \dots & E(u_1 u_T) \\ \dots & \dots & \dots & \dots & \dots \\ E(u_1 u_t) & & E(u_t^2) & & E(u_t u_T) \\ \dots & & \dots & \dots & \dots \\ E(u_1 u_T) & \dots & E(u_t u_T) & \dots & E(u_T^2) \end{bmatrix} = \begin{bmatrix} \sigma^2 & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & & \sigma^2 & & 0 \\ \dots & & \dots & \dots & \dots \\ 0 & \dots & 0 & \dots & \sigma^2 \end{bmatrix} = \sigma^2 I_T \end{aligned}$$

Envisageons plus précisément ces deux hypothèses.

- **L'homoscédasticité des perturbations** : la variance des perturbations est identique pour tout  $t$ . En effet,



$$\forall t, \quad V(u_t) = E[u_t - E(u_t)]^2 = E(u_t^2) = \sigma^2$$

Cela revient à dire que l'ampleur des erreurs (ou l'approximation réalisée par le modèle) est constante pour toute observation. Cela revient encore à dire que cette ampleur **ne dépend pas d'une variable ou d'un facteur ayant rapport avec l'observation  $t$** , auquel cas le modèle « oublierait » d'expliquer quelque chose de **systematique** qui demeure dans l'erreur et ne serait donc pas acceptable. **C'est dès lors l'hypothèse la plus simple à formuler concernant la variance des erreurs, puisqu'elle ne peut pas être supposée nulle.** On ne formule donc pas d'hypothèse concernant la valeur de  $\sigma^2$ . C'est un paramètre inconnu supplémentaire, mais on sent bien que les résultats de l'estimation seront d'autant plus précis que  $\sigma^2$  sera faible.

- **La non-autocorrélation des perturbations :**

$$\forall t_1 \neq t_2, \quad \text{Cov}(u_{t_1}, u_{t_2}) = E[u_{t_1} - E(u_{t_1})] \cdot [u_{t_2} - E(u_{t_2})] = E[u_{t_1} \cdot u_{t_2}] = 0$$

**Une erreur faite sur une observation ne dépend linéairement d'aucune autre erreur faite sur une autre observation.** Dans le cas contraire, cela signifierait en effet que le modèle « oublierait » **systematiquement** un facteur explicatif important de  $y_t$  dans la partie de la perturbation aléatoire et ne serait donc pas acceptable. De façon rigoureuse, cela signifie qu'il n'existe pas de relation linéaire entre les perturbations. On a vu en effet en introduction que le coefficient de corrélation linéaire entre deux variables s'exprime comme le rapport de la covariance entre les variables et le produit des écart-types de variables (ici la variance puisqu'on est sous l'hypothèse d'homoscédasticité). Pour une covariance nulle, le coefficient de corrélation linéaire est nul. Cependant, cela n'implique pas forcément l'indépendance entre les perturbations, comme on l'a vu en introduction et dans le cours de probabilités (à moins bien sûr que les perturbations soient normalement distribuées).

## 3 L'estimation par la méthode des moindres carrés ordinaires

Il faut à présent rechercher une méthode d'estimation qui fournit des estimateurs pour le vecteur de paramètres inconnus possédant des propriétés intéressantes (non biaisés, convergents, ...). Nous allons donc commencer par étudier la méthode des moindres carrés ordinaires (MCO).

### 3.1 Le critère des MCO

Le modèle linéaire multiple s'écrit :

$$y_t = a + b_1 x_{1,t} + b_2 x_{2,t} + \dots + b_{k-1} x_{k-1,t} + u_t$$

On est en présence d'un nuage de  $T$  points (que l'on ne peut représenter) dans  $\mathbb{R}^k$ , l'espace des variables. En effet, chaque point a besoin de  $k$  coordonnées pour être défini dans cet espace. L'ajustement du nuage dans  $\mathbb{R}^k$  se fera donc par un hyperplan dont l'équation s'écrit :

$$\hat{y}_t = \hat{a} + \hat{b}_1 x_{1,t} + \hat{b}_2 x_{2,t} + \dots + \hat{b}_{k-1} x_{k-1,t}$$

et qui passe le plus près possible de tous les points de l'échantillon. La généralisation du programme des MCO se fait dans la droite ligne du cas de la régression linéaire simple : on recherche les paramètres  $\hat{a}, \hat{b}_1, \hat{b}_2, \dots$  et  $\hat{b}_{k-1}$  tels que :

$$\begin{aligned}
\{\hat{a}, \hat{b}_1, \dots, \hat{b}_{k-1}\} &= \arg \min_{(a, b_1, \dots, b_{k-1})} S(a, b_1, \dots, b_{k-1}) \\
&= \arg \min_{(a, b_1, \dots, b_{k-1})} \sum_{t=1}^T u_t^2 \\
&= \arg \min_{(a, b_1, \dots, b_{k-1})} \sum_{t=1}^T (y_t - a - b_1 x_{1,t} - b_2 x_{2,t} - \dots - b_{k-1} x_{k-1,t})^2
\end{aligned}$$

On est en présence d'une fonction scalaire à  $k$  paramètres que l'on cherche à minimiser. On doit donc résoudre le système formé par les  $k$  équations du premier ordre (CPO) calculées en les solutions  $\hat{a}, \hat{b}_1, \hat{b}_2, \dots, \hat{b}_{k-1}$ , c'est-à-dire la valeur particulièrement permettant d'égaliser chaque CPO à 0.

$$\begin{aligned}
\left. \frac{\partial S(a, b_1, \dots, b_{k-1})}{\partial a} \right|_{\substack{a=\hat{a} \\ b_1=\hat{b}_1 \\ \dots \\ b_{k-1}=\hat{b}_{k-1}}} &= 0 \Leftrightarrow -2 \sum_{t=1}^T (\hat{y}_t - \hat{a} - \hat{b}_1 x_{1,t} - \dots - \hat{b}_{k-1} x_{k-1,t}) = 0 \\
\left. \frac{\partial S(a, b_1, \dots, b_{k-1})}{\partial b_1} \right|_{\substack{a=\hat{a} \\ b_1=\hat{b}_1 \\ \dots \\ b_{k-1}=\hat{b}_{k-1}}} &= 0 \Leftrightarrow -2 \sum_{t=1}^T (\hat{y}_t - \hat{a} - \hat{b}_1 x_{1,t} - \dots - \hat{b}_{k-1} x_{k-1,t}) \cdot x_{1,t} = 0 \\
&\dots \\
\left. \frac{\partial S(a, b_1, \dots, b_{k-1})}{\partial b_{k-1}} \right|_{\substack{a=\hat{a} \\ b_1=\hat{b}_1 \\ \dots \\ b_{k-1}=\hat{b}_{k-1}}} &= 0 \Leftrightarrow -2 \sum_{t=1}^T (\hat{y}_t - \hat{a} - \hat{b}_1 x_{1,t} - \dots - \hat{b}_{k-1} x_{k-1,t}) \cdot x_{k-1,t} = 0
\end{aligned}$$

Ensuite, on doit valider cette solution en calculant les conditions du second ordre (CSO) pour caractériser un minimum.

Mais écrivons directement la solution pour le cas général.

### 3.2 Expression matricielle de l'estimateur des MCO

On peut également calculer l'expression matricielle de l'estimateur des MCO. Pour cela, il suffit d'écrire le programme des MCO sous forme matricielle. Ainsi, en repartant du modèle

$$y = X \beta + u$$

$(T \times 1) \quad (T \times k) \quad (k \times 1) \quad (T \times 1)$

le programme précédent se réécrit :

$$\hat{\beta} = \arg \min_{\beta} S(\beta) = \arg \min_{\beta} (u' u) = \arg \min_{\beta} (y - X\beta)' (y - X\beta)$$

Développons l'expression à minimiser :

$$S(\beta) = (y - X\beta)' \cdot (y - X\beta) = \begin{pmatrix} y' \\ (1 \times T) \end{pmatrix} - \begin{pmatrix} \beta' \\ (1 \times k) \end{pmatrix} \begin{pmatrix} X' \\ (k \times T) \end{pmatrix} \cdot \begin{pmatrix} y \\ (T \times 1) \end{pmatrix} - \begin{pmatrix} X \\ (T \times k) \end{pmatrix} \begin{pmatrix} \beta \\ (k \times 1) \end{pmatrix} = \begin{pmatrix} y' y \\ (1 \times 1) \end{pmatrix} - \begin{pmatrix} y' X \\ (1 \times 1) \end{pmatrix} \beta - \beta' \begin{pmatrix} X' y \\ (1 \times 1) \end{pmatrix} + \beta' \begin{pmatrix} X' X \beta \\ (1 \times 1) \end{pmatrix}$$

Comme  $y'X\beta$  et  $\beta'X'y$  sont des scalaires et qu'ils sont la transposée l'un de l'autre, ils sont égaux. Le critère se réécrit donc :

$$S(\beta) = \underset{(1 \times 1)}{y'y} - 2 \underset{(1 \times 1)}{y'X\beta} + \underset{(1 \times 1)}{\beta'X'X\beta}$$

Dérivons le critère afin d'obtenir les CPO. La fonction  $S(\beta)$  est minimum en  $\hat{\beta}$  si :

$$\frac{\partial S(\beta)}{\partial \beta} \Big|_{\beta=\hat{\beta}} = 0$$

Pour pouvoir résoudre ce programme, il est nécessaire de connaître les règles de dérivation matricielle des combinaisons linéaires et des formes quadratiques.

**Rappels :** pour une forme linéaire du type  $a'z = z'a = \sum_{i=1}^n a_i z_i$ , avec  $a$  et  $z$  deux vecteurs colonnes de taille  $n$ , la dérivation par rapport au vecteur de variables  $z$  donne :

$$\frac{\partial(a'z)}{\partial z} = \frac{\partial(z'a)}{\partial z} = a$$

Pour une forme quadratique  $z'Az$  avec  $A$  une matrice carrée symétrique, la dérivation toujours par rapport au vecteur  $z$  donne :

$$\frac{\partial(z'Az)}{\partial z} = 2Az$$

On a alors :

$$-2 \underset{(k \times T)}{X'} \underset{(T \times 1)}{y} + 2 \underset{(k \times T)}{X'} \underset{(T \times k)}{X} \underset{(k \times 1)}{\hat{\beta}} = \underset{(k \times 1)}{0}$$

En réécrivant l'expression, on obtient :

$$\underset{(k \times k)}{X'X} \underset{(k \times 1)}{\hat{\beta}} = \underset{(k \times 1)}{X'y}$$

Ce système correspond au système de  $k$  équations évoqué au point précédent. C'est **le système des équations normales**. Il admet une solution unique si la matrice  $X'X$  est inversible (c'est-à-dire si elle est de rang égal à  $k$ ), ce qui est une implication de l'hypothèse  $H_3$ .

Le vecteur  $\hat{\beta}$  se définit donc comme :

$$\hat{\beta} = \underset{(k \times 1)}{(X'X)^{-1}} \underset{(k \times 1)}{X'y}$$

C'est l'expression matricielle de l'estimateur des MCO dans le cadre de la régression linéaire multiple.

L'expression de la dérivée seconde ( $2X'X$ ) du programme étant une matrice définie positive, on est bien à un minimum en le point solution du programme fourni par les conditions du premier ordre.

On en déduit alors la valeur ajustée par le modèle :

$$\hat{y} = X \hat{\beta}$$

$(T \times 1)$      $(T \times k)$      $(k \times 1)$

qui est donc une combinaison linéaire des variables explicatives, ainsi que le vecteur des résidus estimés :

$$\hat{u} = y - \hat{y}$$

$(T \times 1)$      $(T \times 1)$      $(T \times 1)$

### 3.3 Remarques

- D'autres critères pourraient être envisagés pour minimiser la distance entre l'hyperplan estimé et les observations. Cela pourrait être par exemple la somme de la valeur absolue des écarts (plutôt que des carrés comme ici) ou une distance de projection orthogonale des observations sur la droite estimée (plutôt que verticale comme ici). **Cependant, on va bientôt voir que c'est le critère des MCO qui permet de fournir les estimateurs possédant les meilleures propriétés.**
- $\hat{\beta} = (\hat{a}, \hat{b}_1, \dots, \hat{b}_{k-1})'$  est l'estimateur des moindres carrés ordinaires du vecteur de paramètres inconnus  $\beta = (a, b_1, \dots, b_{k-1})'$ .  $\hat{\beta}$  est un vecteur de variables aléatoires puisque c'est une fonction linéaire du vecteur  $y$ , qui dépend lui-même du vecteur des perturbations aléatoires  $u$ , les poids de la combinaison linéaire étant rassemblés dans la matrice  $(X'X)^{-1}X'$ .
- On appelle estimations les valeurs particulières prises par les fonctions  $\hat{\beta} = (\hat{a}, \hat{b}_1, \dots, \hat{b}_{k-1})'$  pour les réalisations des variables de l'échantillon. Comme pour les variables aléatoires, on ne distinguera pas formellement la variable aléatoire de sa réalisation particulière, mais il s'agit bien de deux choses différentes.
- $\hat{y}_t = \hat{a} + \hat{b}x_t$  est la valeur ajustée ou prédiction ou encore valeur estimée de la variable  $y_t$  dans le cadre de la régression linéaire simple. C'est la droite de régression de  $y$  sur  $x$ . Ce terme est dû à Sir Francis Galton pour ses travaux concernant la relation entre la taille des parents et de leurs enfants. Il a en effet observé que plus (moins) les parents étaient grands, plus (moins) leurs enfants l'étaient, mais plus ils se rapprochaient de la moyenne. D'où le terme de régression (vers le point moyen de l'échantillon) mais qui désigne aujourd'hui couramment l'estimation générale d'un modèle linéaire.
- On a vu que la perturbation aléatoire  $u_t$ , encore appelée aléa ou erreur, est une variable aléatoire dont on a supposé certaines propriétés et dont les réalisations ne sont pas observées. En revanche, on peut calculer :  $\hat{u}_t = y_t - \hat{y}_t$  appelé résidu ou résidu estimé. On a donc :

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

$$y = X \cdot \hat{\beta} + \hat{u} = \hat{y} + \hat{u}$$

La première formulation sera utilisée pour démontrer les propriétés **théoriques** des estimateurs sous les hypothèses 1 à 6 formulées.

- Lorsque l'hypothèse  $H_3$  n'est pas vérifiée, le système n'a pas de solution unique. On dit que le modèle n'est pas identifiable : il y a une infinité de solutions  $\hat{\beta}$  et l'influence de chaque variable explicative sur la variable à expliquer ne peut être mesurée.

## 4 Propriétés de l'estimateur des MCO à distance finie

On va s'intéresser aux propriétés du vecteur aléatoire  $\hat{\beta}$  dans le cadre du modèle de régression linéaire multiple :

$$y = X \beta + u$$

$(T \times 1) \quad (T \times k) \quad (k \times 1) \quad (T \times 1)$

On a montré que c'est un vecteur aléatoire comme fonction de la variable aléatoire  $y$ . On va donc étudier comme au chapitre précédent s'il vérifie les propriétés standards qu'on recherche chez un estimateur lorsque les hypothèses classiques  $H_1$ ,  $H_2$ ,  $H_3$  et  $H_4$  sont vérifiées. On a vu que  $H_3$  était déjà forcément vérifiée, puisqu'elle nous a permis de fournir l'expression de l'estimateur.

## 4.1 L'espérance de l'estimateur des MCO

L'estimateur se définit comme :

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

$(k \times 1) \quad (k \times k) \quad (k \times 1)$

Montrons qu'il est sans biais. Pour cela, développons selon l'expression de  $y$  :

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1} X' \cdot (X \cdot \beta + u) = \underbrace{(X'X)^{-1} X' \cdot X}_{I_k} \cdot \beta + (X'X)^{-1} X' \cdot u \\ &= \beta + (X'X)^{-1} X' \cdot u \end{aligned}$$

Prenons l'espérance de cette expression : il est facile en utilisant les hypothèses 2 puis 1 de voir que :

$$E(\hat{\beta}) = E[\beta + (X'X)^{-1} X' \cdot u] = \beta + (X'X)^{-1} X' \cdot \underbrace{E(u)}_{=0} = \beta$$

**Conclusion : sous  $H_1$ ,  $H_2$  et  $H_3$ , l'estimateur des MCO est sans biais.**

## 4.2 La variance de l'estimateur

### 4.2.1 Calcul de la variance

On s'intéresse à présent à la mesure de la précision de l'estimateur. Comme il s'agit d'un vecteur, on sera en présence d'une matrice de variances-covariances carrée, symétrique, de taille  $k$ .

On part donc de la définition de la variance :

$$\begin{aligned} V(\hat{\beta}) &= E[(\hat{\beta} - E(\hat{\beta})) \cdot (\hat{\beta} - E(\hat{\beta}))'] = E\left\{[(X'X)^{-1} X' \cdot u] \cdot [(X'X)^{-1} X' \cdot u]'\right\} \\ &= E[(X'X)^{-1} X' \cdot u \cdot u' \cdot X \cdot (X'X)^{-1}] \\ &= (X'X)^{-1} X' \cdot E[u \cdot u'] \cdot X \cdot (X'X)^{-1} \\ &= (X'X)^{-1} X' \cdot \sigma^2 I_T \cdot X \cdot (X'X)^{-1} \\ &= \sigma^2 (X'X)^{-1} \underbrace{X'X}_{I_k} \cdot (X'X)^{-1} \\ &= \sigma^2 (X'X)^{-1}_{(k \times k)} \end{aligned}$$

sous les hypothèses  $H_1$ ,  $H_2$ ,  $H_3$  et  $H_4$ . On pourra vérifier que cette matrice correspond à la matrice :

$$\sigma^2 \begin{pmatrix} \frac{1}{T} + \frac{\bar{x}^2}{\sum_{t=1}^T (x_t - \bar{x})^2} & \frac{-\bar{x}}{\sum_{t=1}^T (x_t - \bar{x})^2} \\ \frac{-\bar{x}}{\sum_{t=1}^T (x_t - \bar{x})^2} & \frac{1}{\sum_{t=1}^T (x_t - \bar{x})^2} \end{pmatrix}$$

dans le cas du modèle de régression linéaire simple.

### 4.2.2 Estimation de la variance

Comme le paramètre scalaire  $\sigma^2$  est inconnu, il est nécessaire de l'estimer afin de disposer d'une estimation de la matrice de variances-covariances de  $\hat{\beta}$ . L'estimateur retenu est le suivant :

$$\hat{\sigma}^2 = \frac{SCR}{T-k} = \frac{\hat{u}'\hat{u}}{T-k}$$

**Remarque :** le lecteur prendra bien garde de s'interroger sur la différence entre  $V(u)$ ,  $\hat{\sigma}^2$  et  $V(\hat{u})$ .

## 4.3 Le théorème de Gauss-Markov

**Théorème :**

**Soit un modèle de régression simple  $y = X \cdot \beta + u$  avec les hypothèses associées 1 à 4. L'estimateur  $\hat{\beta}$  des MCO est l'estimateur le plus précis dans l'ensemble des estimateurs linéaires sans biais de  $\beta$ . Il fournit donc la meilleure information possible sur ce vecteur de paramètres inconnus à distance finie.**

## 4.4 L'équation d'analyse de la variance

La première équation normale implique :

$$e_T' \cdot \hat{u} = 0 \Leftrightarrow \sum_{t=1}^T \hat{u}_t = 0 \Leftrightarrow \bar{\hat{u}} = 0$$

Ce premier résultat provient toujours du fait qu'il y a une constante dans le modèle. **La moyenne empirique des résidus estimés est donc égale à 0.** Ce résultat est la contrepartie empirique de l'hypothèse  $H_1$ .

De cela, on peut déduire comme précédemment que la moyenne de la variable à expliquer est égale à la moyenne de la valeur ajustée. En effet, comme :

$$y = \hat{y} + \hat{u} \Rightarrow \sum_{t=1}^T y_t = \sum_{t=1}^T \hat{y}_t + \underbrace{\sum_{t=1}^T \hat{u}_t}_0 \Rightarrow \bar{y} = \bar{\hat{y}}$$

**On en déduit donc que l'hyperplan ajustant le nuage des observations passe par le point moyen.**

$$\bar{\hat{y}} = \bar{y} = \hat{a} + \hat{b}_1 \bar{x}_1 + \dots + \hat{b}_{k-1} \bar{x}_{k-1}$$

L'analyse de la variance découle de ces deux résultats :

$$\underset{(T \times 1)}{y} = \underset{(T \times 1)}{\hat{y}} + \underset{(T \times 1)}{\hat{u}}$$

En retranchant  $\bar{y}e_T$  de chaque coté de l'égalité, il vient :

$$y - \bar{y}e_T = (\hat{y} - \bar{y}e_T) + \hat{u}$$

Les deux termes du membre de droite sont orthogonaux car soustraire le scalaire  $\bar{y}e_T$  de  $\hat{y}$  ne modifie pas sa propriété d'orthogonalité avec  $\hat{u}$ . On peut donc appliquer le théorème de Pythagore :

$$\|y - \bar{y}e_T\|^2 = \|\hat{y} - \bar{y}e_T\|^2 + \|\hat{u}\|^2$$

soit encore :

$$\boxed{\underbrace{\sum_{t=1}^T (y_t - \bar{y})^2}_{SCT} = \underbrace{\sum_{t=1}^T (\hat{y}_t - \bar{y})^2}_{SCE} + \underbrace{\sum_{t=1}^T \hat{u}_t^2}_{SCR} \Leftrightarrow V(y) = V(\hat{y}) + V(\hat{u})}$$

**La variance de  $y$  est égale à la variance expliquée par le modèle plus la variance résiduelle.**

**Remarque :** encore une fois on remarque que tout provient de la première équation normale (qui implique que  $\bar{u} = 0$  et  $\bar{y} = \bar{\hat{y}}$ ). **Dans le cas où il n'y aurait pas de constante dans le modèle, l'analyse de la variance serait impossible.**

## 4.5 Le coefficient de détermination

### 4.5.1 Définition

Grâce à l'analyse de la décomposition de la variance, ce coefficient se définit très simplement comme **la part de la variance expliquée par le modèle rapportée à la variance totale :**

$$\boxed{R^2 = \frac{V(\hat{y})}{V(y)} = \frac{SCE}{SCT} = \frac{\sum_{t=1}^T (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^T (y_t - \bar{y})^2}}$$

Il est facile de voir dans l'équation de la variance que :

$$\frac{\sum_{t=1}^T (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^T (y_t - \bar{y})^2} + \frac{\sum_{t=1}^T \hat{u}_t^2}{\sum_{t=1}^T (y_t - \bar{y})^2} = 1$$

d'où :

$$\boxed{R^2 = 1 - \frac{SCR}{SCT} = 1 - \frac{\sum_{t=1}^T \hat{u}_t^2}{\sum_{t=1}^T (y_t - \bar{y})^2}}$$

Le coefficient de détermination prend donc des valeurs comprises entre 0 et 1. Au pire, le modèle n'explique rien, au mieux il explique toute la variance de la variable  $y$ .

### 4.5.2 Le $R^2$ comme coefficient de corrélation

Dans le modèle de régression linéaire simple, le coefficient de détermination est égal au carré du coefficient de corrélation linéaire  $r_{x,y}^2$  entre les variables explicative et expliquée  $x$  et  $y$ .

**Preuve :** repartons de la définition du coefficient de corrélation linéaire.

$$r_{x,y}^2 = \frac{S_{x,y}^2}{S_x^2 S_y^2}$$

Repartons de la définition de  $R^2$  :

$$R^2 = \frac{\sum_{t=1}^T (\hat{a} + \hat{b}x_t - \hat{a} - \hat{b}\bar{x})^2}{\sum_{t=1}^T (y_t - \bar{y})^2} = \hat{b}^2 \frac{\sum_{t=1}^T (x_t - \bar{x})^2}{\sum_{t=1}^T (y_t - \bar{y})^2} = \hat{b}^2 \frac{S_x^2}{S_y^2}$$

Or, on sait que  $\hat{b} = \frac{S_{x,y}}{S_x^2}$ . Dès lors,

$$R^2 = \frac{S_{x,y}^2}{S_x^2 S_x^2} \frac{S_x^2}{S_y^2} = \frac{S_{x,y}^2}{S_x^2 S_y^2} = r_{x,y}^2$$

### 4.5.3 Interprétation géométrique du $R^2$

On peut directement réécrire :

$$R^2 = \frac{\sum_{t=1}^T (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^T (y_t - \bar{y})^2} = \frac{\|\hat{y}_t - \bar{y}e_T\|^2}{\|y_t - \bar{y}e_T\|^2} = \cos^2 w$$

avec  $w$  l'angle formé par les deux vecteurs dans le triangle rectangle. D'où des interprétations très intuitives. Plus le  $R^2$  est élevé, plus les variables sont corrélées. Ou encore, plus le pouvoir explicatif du modèle est élevé, plus  $y$  est proche de  $\hat{y}$ , plus l'angle est faible et plus le cosinus et donc le  $R^2$  est proche de 1.

### 4.5.4 Les limites du $R^2$ : premier exemple

Il ne faut cependant pas attacher trop d'importance à cette mesure pour juger de la qualité de l'ajustement d'un modèle. En effet, elle est très dépendante de la forme des variables dans le modèle.

**Illustration :** Soit le modèle suivant :

$$y_t = a + bx_t + u_t$$

Réécrivons-le en posant :  $z_t = y_t - x_t$ . On a alors :

$$y_t - x_t = a + (b - 1)x_t + u_t \Leftrightarrow z_t = \alpha + \beta x_t + w_t$$



Ces deux modèles sont équivalents mais si  $\hat{b} < \frac{1}{2}$ , alors on obtiendra un  $R^2$  supérieur avec l'estimation du second modèle.

#### 4.5.5 Les limites du $R^2$ : second exemple

On a vu à la section précédente que le coefficient de détermination n'était pas un critère sur lequel on pouvait fonder seul l'analyse d'un modèle économétrique. En effet, on a vu qu'**on pouvait, sous une certaine condition, faire augmenter ce coefficient en manipulant les données. Ici, nous constatons qu'il peut également augmenter de façon mécanique en ajoutant des variables explicatives dans le modèle. Ainsi, il suffit d'introduire des variables supplémentaires dans la régression pour faire augmenter le coefficient de détermination de façon mécanique et cela quelque soit la pertinence du modèle considéré.**

#### 4.5.6 Le coefficient de détermination ajusté

Cet indicateur est noté  $\bar{R}^2$ . Il vise à fournir une indication concernant la part de la variance totale expliquée par le modèle mais corrigée de l'effet mécanique induit par l'ajout de variables supplémentaires. Il est défini comme :

$$\bar{R}^2 = 1 - \frac{\sum_{t=1}^T \hat{u}_t^2}{T - k} \frac{\sum_{t=1}^T (y_t - \bar{y})^2}{T - 1}$$

L'ajout d'une variable supplémentaire implique comme on l'a vu la diminution de la somme des carrés des résidus, mais également celle de  $T - k$  (car  $k$  augmente). On n'a dès lors plus une augmentation systématique de cet indicateur lors de l'ajout d'une variable explicative supplémentaire.

## 5 Propriétés asymptotiques de l'estimateur des MCO

Que deviennent l'estimateur des MCO et ses propriétés lorsque la taille de l'échantillon tend vers l'infini ?

### 5.1 Une hypothèse supplémentaire : $H_5$

Aux quatre premières hypothèses standards utilisées jusqu'à présent, on en ajoute une nouvelle.

$$H_5 : \lim_{T \rightarrow \infty} S_X = \lim_{T \rightarrow \infty} \frac{1}{T} X'X = V_X, \quad (k, k)$$

avec  $V_X$  une matrice finie définie positive et inversible.

Il suffit de poser la structure de la matrice  $\frac{1}{T} X'X$ . Cette hypothèse signifie que lorsque  $T$  tend vers l'infini :

1. **les moyennes, les variances et les covariances sont finies.** En effet, en présence d'une constante dans le modèle, cette matrice contient des éléments

égaux à  $\sum_{t=1}^T x_{j,t}$  à la  $(j+1)$ ème de la première colonne (avec  $j \in \llbracket 1, k-1 \rrbracket$ ),  $\sum_{t=1}^T x_{j,t}^2$  à l'intersection de la diagonale principale et de la  $(j+1)$ ème ligne (avec  $j \in \llbracket 1, k-1 \rrbracket$ ), et enfin  $\sum_{t=1}^T x_{i,t} \cdot x_{j,t}$  sur les éléments non diagonaux de la matrice, à l'intersection entre  $(i+1)$ ème ligne et la  $(j+1)$ ème colonne (avec  $i \in \llbracket 1, k-1 \rrbracket, j \in \llbracket 1, k-1 \rrbracket$  et  $i \neq j$ ).

2. **les variables explicatives restent linéairement indépendantes.** On a alors la conservation de l'hypothèse  $H_3$  lorsque  $T$  tend vers l'infini.

**L'idée est donc comme précédemment que les variables explicatives conservent toujours une certaine variance lorsque  $T$  tend vers l'infini.**

Partons du fait que les variables  $x_{it}$  sont des variables aléatoires indépendamment et identiquement distribuées de variances  $\sigma_{x_i}^2$ . L'idée est ici que les variables  $x_{it}$  conservent toujours une certaine variance lorsque la taille de l'échantillon tend vers l'infini (

$\lim_{T \rightarrow \infty} s_{x_i}^2 = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T (x_{it} - \bar{x})^2 = \sigma_{x_i}^2 \neq 0$ ). En effet, dans le cas contraire, cela voudrait dire que

les  $x_{it}$  convergent vers leur moyenne à partir d'une certaine date. Dès lors, une observation supplémentaire n'apporterait aucune information. **Ainsi, disposer d'observations supplémentaires apporte de l'information.**

N'oublions pas que même s'il paraît paradoxal de parler des explicatives comme une variable aléatoire au regard de l'hypothèse  $H_2$ , cette hypothèse n'est là que pour simplifier les calculs et permettre de modéliser  $y_t$  conditionnellement aux  $x_{it}$  observés dans l'échantillon.

## 5.2 Les différents types de convergence et leurs relations.

**Rappel :**

$$X_n \xrightarrow{m.q.} X \Rightarrow X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{L} X$$

## 5.3 Convergence en probabilité de l'estimateur des MCO

L'estimateur se définit à présent comme  $\hat{\beta}_T$  :

$$\hat{\beta}_T = \begin{pmatrix} X' & X \end{pmatrix}_{(k \times T) \ (T \times k)}^{-1} \begin{pmatrix} X' & y \end{pmatrix}_{(k \times T) \ (T \times 1)}$$

et dépend de la taille de l'échantillon  $T$ . On s'intéresse donc à la suite de variables aléatoires  $\hat{\beta}_T, \hat{\beta}_{T+1}, \dots$  et sa convergence lorsque  $T$  tend vers l'infini.

**Preuve :** montrons que l'estimateur des MCO converge en probabilité vers la vraie valeur du paramètre inconnu. On sait que la convergence en moyenne quadratique est une condition **suffisante** de la convergence en probabilité. **Autrement dit, la convergence en moyenne quadratique implique la convergence en probabilité.** Pour que  $\hat{\beta}_T$  converge en moyenne quadratique vers  $\beta$ , il faut que :

$$E[(\hat{\beta}_T - \beta) \cdot (\hat{\beta}_T - \beta)'] \xrightarrow{T \rightarrow +\infty} \underset{(k \times k)}{0}$$

Comme  $\hat{\beta}_T$  est un estimateur sans biais, cela revient à montrer que :

$$V(\hat{\beta}_T) \xrightarrow{T \rightarrow +\infty} 0_{(k \times k)}$$

On a déjà calculé l'expression de la matrice de variances-covariances de  $\hat{\beta}_T$  :

$$V(\hat{\beta}_T) = \sigma^2 (X'X)^{-1} = \frac{\sigma^2}{T} \left( \frac{X'X}{T} \right)^{-1} \xrightarrow{T \rightarrow +\infty} 0 \cdot (V_X)^{-1} = 0_{(k \times k)}$$

d'après  $H_5$ .

**Conclusion :** l'estimateur des MCO converge en moyenne quadratique donc en probabilité vers la vraie valeur du paramètre inconnu.

**Remarque :** Cette hypothèse  $H_5$  peut être considérée comme restrictive dans la mesure où elle ne permet pas d'inclure des variables telles que le temps dans le modèle de régression, car la moyenne et la variance d'une telle variable augmentent avec le nombre d'observations. A la formulation de cette hypothèse, on pourra préférer par exemple :  $H_{5bis}$  :

$$\lim_{T \rightarrow \infty} (X'X)^{-1} = 0_{(k,k)}$$

## 5.4 La normalité asymptotique de l'estimateur des MCO

Pour pouvoir étudier la convergence en loi de l'estimateur des MCO, on va utiliser le théorème central-limite. Pour cela, on va redéfinir  $H_4$  en introduisant l'hypothèse d'indépendance des perturbations.

$H_{4bis}$  : Les perturbations  $u_t$  sont **indépendamment** et identiquement distribuées (*iid*) d'espérance 0 et de variance  $\sigma^2$ .

On en déduit donc le résultat suivant.

Sous les hypothèses  $H_1$ ,  $H_2$ ,  $H_3$ ,  $H_{4bis}$  et  $H_5$ , l'estimateur centré dilaté  $\sqrt{T}(\hat{\beta}_T - \beta)$  suit asymptotiquement une loi normale :

$$\sqrt{T}(\hat{\beta}_T - \beta) \xrightarrow{L} N(0, \sigma^2 (V_X)^{-1})$$

**Preuve :** appliquons le théorème central limite. On a vu que l'estimateur des MCO est une combinaison linéaire avec des poids non aléatoires de lois indépendantes :

$$\hat{\beta}_T = \beta + (X'X)^{-1} X' \cdot u$$

On peut donc avoir recours au théorème central-limite. On a vu au point précédent qu'il est nécessaire de dilater  $\hat{\beta}_T - \beta$ , puisque cet estimateur converge en probabilité vers 0 (dans le cas contraire, sa distribution asymptotique serait dégénérée puisque égale à une constante). L'expression de l'espérance et de la variance de cet estimateur étant connue :

$$E[\sqrt{T}(\hat{\beta}_T - \beta)] = \sqrt{T}[E(\hat{\beta}_T) - \beta] = 0$$

$$V[\sqrt{T}(\hat{\beta}_T - \beta)] = TV(\hat{\beta}_T) = T \frac{\sigma^2}{T} \left( \frac{X'X}{T} \right)^{-1} = \sigma^2 \left( \frac{X'X}{T} \right)^{-1} \xrightarrow{T \rightarrow +\infty} \sigma^2 \cdot (V_X)^{-1}$$

On en déduit donc le résultat énoncé précédemment.

### Conclusion :

- A distance finie ( $T$  est fixé), les MCO fournissent des estimateurs sans biais et les plus précis parmi l'ensemble des estimateurs linéaires sans biais.
- Au niveau asymptotique ( $T$  tend vers l'infini), les MCO fournissent des estimateurs convergents en probabilité et les estimateurs centrés-dilatés suivent asymptotiquement une loi normale.

## 6 L'hypothèse de normalité des perturbations et ses conséquences

Envisageons à présent les conséquences d'une sixième hypothèse :

$$H_6 : u \sim N(0, \sigma^2 I_T).$$

**Cette hypothèse implique en premier que les perturbations sont indépendantes sous  $H_4$ .** Avec l'hypothèse  $H_4$ , les perturbations sont à présent indépendantes, plus besoin d' $H_4$  bis.

Cette hypothèse assez exigeante permettra d'effectuer les tests sur le modèle. Sa pertinence repose sur le fait que nombre de variables économiques observées suivent des lois log-normales et qu'on modélise souvent les variables dans le cadre d'une spécification log-linéaire. Dans ce cas, les perturbations, représentant le logarithme de variables économiques, peuvent être normalement distribuées.

Cependant, **grâce au théorème central-limite**, il n'est pas nécessaire de supposer la normalité des composantes de  $u$  pour justifier la normalité des perturbations. La convergence en loi vers la normalité le permet également. En effet,  $u_t$  peut comprendre une infinité de facteurs indépendamment et identiquement distribués qui jouent additivement. Trois conséquences majeures apparaissent :

- La distribution de l'estimation des MCO
- La distribution de  $y$ . On va donc pouvoir écrire la vraisemblance.
- **Les estimateurs du maximum de vraisemblance sont ceux des MCO.**

### 6.1 La loi de l'estimateur des paramètres inconnus du premier ordre

A distance finie, il est assez simple de voir que :

$$\hat{\beta} \square N(\beta, \sigma^2 (X'X)^{-1})$$

**Preuve :** c'est trivial quand on applique les propriétés de combinaison linéaire de lois normales. On peut exprimer l'estimateur comme une combinaison linéaire non aléatoire de  $y$ , qui dépend de  $u$ , un vecteur de lois normales. C'est donc un vecteur de lois normales. Dès lors, comme on connaît ses moments (espérance et variance), ce vecteur aléatoire est parfaitement défini.

## 6.2 La loi de yt

$$y_t \sim N(X_t' \cdot \beta, \sigma^2 I_T)$$

## 6.3 Le critère du maximum de vraisemblance

L'idée est de retenir l'estimateur permettant de maximiser la vraisemblance d'un échantillon donné. L'estimateur obtenu est convergent, asymptotiquement efficace et asymptotiquement normal.

Comme on a vu que  $y_t \sim N(X_t' \cdot \beta, \sigma^2 I_T)$ , il est possible d'écrire la densité de probabilité de l'observation en  $t$  :

$$f(y_t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y_t - X_t'\beta)^2}$$

L'indépendance des perturbations fait que les  $y_t$  sont aussi indépendantes. On écrit alors la vraisemblance de l'échantillon comme le produit des densités de probabilités de toutes les réalisations  $y_t$  :

$$l(\beta, \sigma^2; y_1, \dots, y_T) = \prod_{t=1}^T f(y_t) = \frac{1}{(\sigma\sqrt{2\pi})^T} e^{-\frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - X_t'\beta)^2} = \frac{1}{(\sigma\sqrt{2\pi})^T} e^{-\frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta)}$$

La vraisemblance est donc ici une fonction à plusieurs arguments à valeurs dans  $\mathbb{R}^+$ . Les estimateurs du maximum de vraisemblance, notés  $\tilde{\beta}$  et  $\tilde{\sigma}^2$ , sont solutions du programme suivant :

$$\max_{(\beta, \sigma^2)} l(\beta, \sigma^2; y_1, \dots, y_T)$$

Il est équivalent de maximiser cette fonction ou le logarithme népérien de cette fonction, puisque la transformation est définie, continue et croissante. Le programme devient alors :

$$\max_{(\beta, \sigma^2)} \ln l(\beta, \sigma^2; y_1, \dots, y_T)$$

avec :

$$\ln l(\beta, \sigma^2; y_1, \dots, y_T) = -\frac{T}{2} \ln(2\pi) - T \ln \sigma - \frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta)$$

Maximiser  $-\frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta)$  par rapport à  $\beta$  revenant à minimiser  $(y - X\beta)'(y - X\beta)$  par rapport à  $\beta$ , **les estimateurs du maximum de vraisemblance sont identiques à ceux des MCO.**

### Remarques :

- L'estimateur des MCO étend donc la portée des propriétés à distance finie mises en évidence précédemment par le théorème de Gauss-Markov. En effet, l'estimateur des MCO possède en plus **la propriété d'efficacité asymptotique des estimateurs du maximum de vraisemblance**, et cela, quel que soit l'ensemble des estimateurs auxquels on se réfère (linéaires ou pas). C'est donc l'estimateur le plus précis possible !

- Il convient de garder à l'esprit que l'estimateur des MCO ne correspond à celui du maximum de vraisemblance que sous l'hypothèse  $H_6$ .

Il reste à déterminer l'estimateur du maximum de vraisemblance de  $\sigma^2$ . Au maximum de vraisemblance, on sait qu'on vérifie la condition du premier ordre :

$$\left. \frac{\partial \ln l(\beta, \sigma^2; y_1, \dots, y_T)}{\partial \sigma} \right|_{\substack{\beta=\hat{\beta} \\ \sigma^2=\hat{\sigma}^2}} = 0$$

Ainsi, et d'après l'équivalence entre les estimateurs des MCO et du maximum de vraisemblance :

$$-\frac{T}{\hat{\sigma}} + \frac{1}{\hat{\sigma}^3} (y - X\hat{\beta})'(y - X\hat{\beta}) = 0$$

d'où :

$$\hat{\sigma}^2 = \frac{1}{T} (y - X\hat{\beta})'(y - X\hat{\beta}) = \frac{1}{T} \hat{u}'\hat{u} = \frac{SCR}{T}$$

On constate ici que l'estimateur du maximum de vraisemblance de la variance des perturbations  $\hat{\sigma}^2 \neq \sigma^2$  puisqu'on a précédemment défini un estimateur de la variance des perturbations comme :

$$\hat{\sigma}^2 = \frac{SCR}{T-k}$$

L'estimateur  $\hat{\sigma}^2$  du maximum de vraisemblance est donc biaisé négativement à distance finie puisque  $\hat{\sigma}^2$  est sans biais :

$$\begin{aligned} E\left(\frac{SCR}{T-k}\right) &= \sigma^2 \Leftrightarrow \frac{T-k}{T} E\left(\frac{SCR}{T-k}\right) = \frac{T-k}{T} \sigma^2 \\ \Leftrightarrow E\left(\frac{T-k}{T} \frac{SCR}{T-k}\right) &= \frac{T-k}{T} \sigma^2 \Leftrightarrow E\left(\frac{SCR}{T}\right) = \left(1 - \frac{k}{T}\right) \sigma^2 < \sigma^2 \end{aligned}$$

**La méthode du maximum de vraisemblance conduit donc à une sous-estimation systématique de la vraie valeur  $\sigma^2$  mais fournit un estimateur cependant plus précis.** A contrario, le biais tend asymptotiquement vers 0 et l'estimateur est convergent. On a donc un arbitrage à faire selon que l'on est à distance finie ou non entre un estimateur biaisé mais plus précis et un estimateur non biaisé mais moins précis.

## Chapitre 2

# Inférence dans le modèle linéaire

Dans le chapitre précédent, nous avons envisagé l'estimation du modèle linéaire général. Nous avons étudié un estimateur particulier, celui des moindres carrés ordinaires, et montré que sous les hypothèses standards de la régression, il possédait les meilleures propriétés possibles.

Nous avons donc à présent à notre disposition une estimation **ponctuelle** pour les paramètres du modèle, ainsi qu'une mesure de la précision de cette estimation, au travers des variances fournies par la matrice de variances-covariances des estimations.

Cependant, cette estimation est intimement liée à l'échantillon utilisé et serait certainement différente numériquement pour un autre échantillon. En effet, l'estimateur donne en espérance la vraie valeur du paramètre inconnu mais l'estimation obtenue est une réalisation particulière de la variable aléatoire que constitue cet estimateur.

Plusieurs types de question se posent à présent.

- D'abord, à la place d'une estimation ponctuelle, ne pourrait-on pas fournir plutôt **un intervalle pour le paramètre inconnu**, qui tiendrait compte de l'incertitude liée au modèle estimé, à l'échantillon utilisé, ... ?
- Ensuite, au regard de la valeur estimée du paramètre inconnu, ne pourrait-on pas envisager **l'hypothèse que la valeur de ce paramètre inconnu n'est pas statistiquement différente d'une valeur prédite par la théorie** ? (Exemple : dans une équation de salaires, ne pourrait-on pas envisager que le taux de croissance des salaires nominaux est parfaitement indexé sur le taux de croissance des prix ?) Ou encore, si la variable explicative n'est pas pertinente, **ne pourrait-on pas tester la nullité du paramètre qui lui est associée et donc l'exclure du modèle** ?
- Enfin, **ne pourrait-on pas généraliser cette dernière procédure à plusieurs paramètres estimés simultanément, et ainsi réaliser un test joint** ? (Exemple : dans une équation de Cobb-Douglas estimée, la somme des coefficients estimés sur le travail et le capital est-il significativement différent de 1 ?). Ou encore, si plusieurs variables explicatives ne semblent pas pertinentes, ne pourrait-on pas tester leur

nullité simultanément ? Est-il alors différent de tester individuellement ou de façon jointe ?

Nous allons donc envisager dans ce chapitre les résultats concernant l'estimation des paramètres par intervalle de confiance, et la réalisation de tests simples et de tests joints sur les paramètres, ainsi que l'estimation d'un modèle sous contraintes et ses implications.

## 1 Résultats préliminaires

$$\hat{\beta}_{(k \times 1)} \sim N(\beta, \sigma^2 (X'X)^{-1}) \Leftrightarrow \forall j \in [1, k] \quad \hat{\beta}_j \sim N(\beta_j, [\sigma^2 (X'X)^{-1}]_{jj})$$

$$(T - k) \frac{\hat{\sigma}^2}{\sigma^2} = \frac{SCR}{\sigma^2} \sim \chi^2(T - k)$$

**R1 :** Supposons que le vecteur  $X$  soit composé de  $n$  lois normales centrées réduites, identiquement et indépendamment distribuées :

$$X \sim N(0, I_n) \Leftrightarrow x_i \sim N(0, 1) \quad i \in [1, n]$$

Alors la forme quadratique suit une loi du Chi-deux :

$$X'X = \sum_{i=1}^n x_i^2 = \sum_{i=1}^n (N(0, 1))^2 \sim \chi^2(n)$$

Student

Fisher

## 2 L'estimation par intervalles de confiance

La démarche à suivre est toujours la même, en deux temps.

- trouver une fonction de l'estimateur et du paramètre inconnu dont on peut identifier la loi, classique de préférence.
- utiliser ce résultat et la table de la loi pour un niveau de confiance donné pour construire l'intervalle de confiance.

### 2.1 Intervalles de confiance pour $\beta$

#### 2.1.1 Loi suivie par une fonction de l'estimateur et du paramètre inconnu

On a vu précédemment que sous les hypothèses classiques de la régression, la distribution de l'estimateur  $\hat{\beta}$  du vecteur de paramètres inconnus  $\beta$  suivait (asymptotiquement ou exactement) des lois normales :

$$\hat{\beta}_{(k \times 1)} \sim N(\beta, \sigma^2 (X'X)^{-1}) \Leftrightarrow \forall j \in [1, k] \quad \hat{\beta}_j \sim N(\beta_j, [\sigma^2 (X'X)^{-1}]_{jj})$$



où  $[\sigma^2(X'X)^{-1}]_{jj}$  désigne l'élément de la  $j^{\text{ème}}$  ligne et colonne de la matrice de variances-covariances des paramètres estimés, c'est-à-dire la variance de  $\hat{\beta}_j$  le  $j^{\text{ème}}$  paramètre du vecteur estimé  $\hat{\beta}$ . Notons-le  $\sigma_{\hat{\beta}_j}^2$  pour alléger les notations.

Pour tout paramètre inconnu  $\beta_j$  du vecteur  $\beta$ , on peut écrire, d'après les propriétés de la loi normale :

$$\forall j \in [1, k] \quad \hat{\beta}_j \sim N(\beta_j, \sigma_{\hat{\beta}_j}^2) \Leftrightarrow \frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma_{\hat{\beta}_j}^2}} \sim N(0, 1)$$

### 2.1.2 Construction de l'intervalle de confiance

Dès lors, il est possible de construire un intervalle de confiance **pour les paramètres inconnus** à partir de ce résultat.

Définissons cet intervalle **comme la probabilité que les réalisations de la loi normale centrée réduite aient une probabilité égale à  $1-\alpha$  d'appartenir respectivement à l'intervalle  $[-t_\alpha, t_\alpha]$**  inconnus pour le moment :

$$\Pr \left\{ -t_\alpha \leq \frac{\hat{\beta}_j - \beta_j}{\sigma_{\hat{\beta}_j}} \leq t_\alpha \right\} = 1 - \alpha$$

L'intervalle pour la loi de  $\hat{\beta}$  est ici forcément symétrique, puisque la loi normale est une loi symétrique autour de 0. **La valeur de  $t_\alpha$  dépend donc de la valeur choisie pour  $1-\alpha$ . Ainsi, si la probabilité recherchée est  $1-\alpha$ , on choisit dans la table de la loi normale la valeur de  $t_\alpha$  associée à la surface à gauche de  $1 - \alpha/2$ .**

**Exemple :** Si on désire qu'il y ait 95% de chances que les réalisations de la loi normale appartiennent à l'intervalle de confiance, on choisit dans la table de la loi normale la valeur de  $t_\alpha$  associée à la surface à gauche de 97,5%, soit 1,95. Pour 90% de chances, on choisira 1,64.

Dès lors, comme on connaît à présent la valeur de  $t_\alpha$ , il est possible de déduire un intervalle de confiance pour les paramètres inconnus  $\beta$  :

$$\Pr \left\{ -t_\alpha \leq \frac{\hat{\beta}_j - \beta_j}{\sigma_{\hat{\beta}_j}} \leq t_\alpha \right\} = 1 - \alpha \Leftrightarrow -t_\alpha \leq \frac{\hat{\beta}_j - \beta_j}{\sigma_{\hat{\beta}_j}} \leq t_\alpha$$

On peut donc en déduire la définition de l'intervalle de confiance de  $\beta_j$  :

$$\beta_j \in \left[ \hat{\beta}_j - t_\alpha \sigma_{\hat{\beta}_j}, \hat{\beta}_j + t_\alpha \sigma_{\hat{\beta}_j} \right]$$

au niveau de confiance de  $1-\alpha$ .

Cet intervalle est donc symétrique autour de l'estimation ponctuelle. La largeur de l'intervalle dépend :

- du niveau de confiance exigé  $1-\alpha$  (plus  $1-\alpha$  est grand, plus  $t_\alpha$  sera important et donc plus l'intervalle sera large)

- de la précision de l'estimation ponctuelle du paramètre inconnu, mesurée par l'écart-type de l'estimateur  $\hat{\sigma}_{\hat{\beta}_j}$  (plus l'imprécision de l'estimation ponctuelle est importante, plus l'intervalle sera large).

**Application :** fournir un intervalle de confiance pour  $\hat{\beta}_j = 6$  et  $\hat{\sigma}_{\hat{\beta}_j} = 2$ .

## 2.2 Construction d'un intervalle de confiance opérationnel pour $\beta$

Un problème se pose ici. La formule obtenue pour l'estimateur par intervalle de confiance de  $\beta_j$  n'est pas opérationnelle car on ne connaît pas la valeur de  $\sigma^2$ . L'approche précédente est donc peu réaliste. On va cependant la mettre à profit pour en déduire un intervalle de confiance opérationnel de  $\beta$ .

### 2.2.1 Loi suivie par une modification de la fonction initiale

Remplaçons  $\sigma^2$  par son estimation  $\hat{\sigma}^2$ .

On obtient alors l'estimation pour la variance de  $\beta_j$  égale à  $\sqrt{[\hat{\sigma}^2(X'X)^{-1}]_{jj}}$  que l'on notera pour simplifier  $\hat{\sigma}_{\hat{\beta}_j}$ . **Notez bien la présence des deux chapeaux pour marquer l'estimateur de la variance de l'estimateur.**

Cependant, la variable aléatoire formée à présent par  $\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}}$  ne suit plus une loi normale centrée réduite mais une loi de Student. En effet, on peut réécrire cette variable aléatoire comme :

$$\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} = \frac{\hat{\beta}_j - \beta_j}{\sqrt{[\hat{\sigma}^2(X'X)^{-1}]_{jj}}} = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\left[\frac{\sigma^2}{\hat{\sigma}^2} \hat{\sigma}^2(X'X)^{-1}\right]_{jj}}} = \frac{\frac{\hat{\beta}_j - \beta_j}{\sqrt{[\sigma^2(X'X)^{-1}]_{jj}}}}{\sqrt{(T-k) \frac{\hat{\sigma}^2}{\sigma^2} \frac{1}{(T-k)}}}$$

On voit bien qu'il s'agit du rapport entre une loi normale centrée réduite et la racine carrée d'une loi du  $\chi^2$  rapporté à son nombre de degrés de liberté. **On est donc en présence d'une loi de Student si ces deux lois sont indépendantes.** Pour démontrer l'indépendance, on a recours au résultat préliminaire 2 de la section 1, sous les hypothèses standards.

### 2.2.2 Construction pratique de l'intervalle de confiance

Dès lors, cette nouvelle statistique suit par définition une loi de Student, d'un nombre de degrés de liberté égal au nombre de degrés de liberté de la loi du  $\chi^2$  du dénominateur.

$$\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \square T_{T-k}$$

Comme précédemment, il est alors possible de construire un intervalle de confiance opérationnel **pour les paramètres inconnus du vecteur  $\beta$**  à partir de cette nouvelle distribution. **Définissons cet intervalle comme la probabilité que les réalisations de la loi de Student aient une probabilité égale à  $1-\alpha$  d'appartenir à l'intervalle  $[-t_{\alpha}, t_{\alpha}]$  :**

$$\Pr \left\{ -t_\alpha \leq \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \leq t_\alpha \right\} = 1 - \alpha$$

L'intervalle pour la loi de  $\hat{\beta}$  est ici forcément symétrique, puisque la loi de Student est une loi symétrique autour de 0. **La valeur de  $t_\alpha$  dépend donc de la valeur choisie pour  $1-\alpha$ . Ainsi, si la probabilité recherchée est  $1-\alpha$ , on choisit dans la table de la loi de Student la valeur de  $t_\alpha$  associée à la surface  $p/2$  égale à  $\alpha/2$  et pour un nombre de degrés de liberté égal à  $T-k$ .**

On peut donc en déduire la définition de l'intervalle de confiance de  $\beta_j$  :

$$\beta_j \in \left[ \hat{\beta}_j - t_\alpha \hat{\sigma}_{\hat{\beta}_j}, \hat{\beta}_j + t_\alpha \hat{\sigma}_{\hat{\beta}_j} \right]$$

au niveau de confiance de  $1-\alpha$ .

**Exemple :** si on désire qu'il y ait 95% de chances que les réalisations d'une loi de Student à 25 degrés de liberté appartiennent à l'intervalle de confiance, on choisit dans la table de la loi de Student la valeur de  $t_\alpha$  associée à la surface à gauche de 97,5% ou de 2,5% à droite, soit une valeur de  $p$  de 0,05 et donc la valeur de  $t_\alpha$  2,06.

## 2.3 Construction des intervalles de confiance pour $\sigma^2$

### 2.3.1 Loi suivie par la statistique

On a vu précédemment que sous les hypothèses classiques de la régression, la distribution de l'estimateur  $\hat{\sigma}^2$  du paramètre inconnu  $\sigma^2$  suit une loi du  $\chi^2$  :

$$(T - k) \frac{\hat{\sigma}^2}{\sigma^2} = \frac{SCR}{\sigma^2} \sim \chi^2(T - k)$$

### 2.3.2 Construction de l'intervalle de confiance

**Définissons cet intervalle comme la probabilité que les réalisations de la loi du  $\chi^2$  aient une probabilité égale à  $1-\alpha$  d'appartenir à l'intervalle  $[\chi_{\text{inf}}^2, \chi_{\text{sup}}^2]$  :**

$$\Pr \left\{ \chi_{\text{inf}}^2 \leq (T - k) \frac{\hat{\sigma}^2}{\sigma^2} \leq \chi_{\text{sup}}^2 \right\} = 1 - \alpha$$

La loi du  $\chi^2$  n'étant pas symétrique, il n'y a aucune relation entre les valeurs formant l'intervalle contenant les réalisations de la loi avec une probabilité égale à  $1-\alpha$ . **La surface à l'extérieur de l'intervalle est égale à  $\alpha$ , répartie en deux surfaces de taille identique égales à  $\alpha/2$ . On doit donc lire la valeur de  $\chi_{\text{inf}}^2$  dans la table du  $\chi^2$  à  $T-k$  degrés de liberté, pour une surface à gauche égale à  $\alpha/2$  ou une surface à droite égale à  $1-\alpha/2$ . De même, on lira la valeur de  $\chi_{\text{sup}}^2$  dans la table du  $\chi^2$  à  $T-k$  degrés de liberté, pour une surface à gauche égale à  $1-\alpha/2$  ou une surface à droite égale à  $\alpha/2$ .**

**Exemple :** Pour une loi du  $\chi^2$  à 30 degrés de liberté, l'intervalle de confiance à 90% est [18.49 , 43.77].

L'intervalle de confiance concernant le paramètre inconnu  $\sigma^2$  :

$$\Pr \left\{ \chi_{\text{inf}}^2 \leq (T-k) \frac{\hat{\sigma}^2}{\sigma^2} \leq \chi_{\text{sup}}^2 \right\} = 1 - \alpha \Leftrightarrow \chi_{\text{inf}}^2 \leq (T-k) \frac{\hat{\sigma}^2}{\sigma^2} \leq \chi_{\text{sup}}^2$$

avec  $[\chi_{\text{inf}}^2, \chi_{\text{sup}}^2]$  connu grâce à  $1-\alpha$ . On peut donc en déduire :

$$\sigma^2 \in \left[ \frac{(T-k)\hat{\sigma}^2}{\chi_{\text{sup}}^2}, \frac{(T-k)\hat{\sigma}^2}{\chi_{\text{inf}}^2} \right]$$

au seuil de confiance de  $1-\alpha$ . Ici, l'intervalle n'a aucune raison d'être symétrique autour de l'estimation ponctuelle du paramètre inconnu. Comme précédemment, la largeur de l'intervalle dépend du niveau de confiance requis pour construire l'intervalle.

### 3 Test sur plusieurs paramètres

L'idée est d'ici de se demander si le ou les paramètres inconnus que l'on cherche à estimer sont statistiquement égaux ou non à telle valeur supposée.

#### 3.1 Rappels généraux sur la méthodologie des tests

L'utilisation des intervalles de confiance comme moyen de décision est possible. Cependant, bien que conscient des erreurs qu'il peut commettre, l'économètre-statisticien ne peut évaluer de façon directe le risque d'erreur.

La théorie des tests le permet, en se ramenant uniquement au choix entre deux hypothèses antagonistes, notées  $H_0$  et  $H_1$ . L'hypothèse  $H_0$ , encore appelée **hypothèse nulle**, est privilégiée par l'économètre-statisticien jusqu'au moment où elle est infirmée par l'observation. Ainsi, le test a pour but de mesurer l'adéquation d'une hypothèse à la réalité observée à travers l'information apportée par notre échantillon.

On retient plusieurs étapes dans la démarche des tests.

##### 3.1.1 La formulation des hypothèses

Il est d'abord nécessaire de formuler les hypothèses à tester, et par voie de conséquence, les erreurs de décision en termes de ces hypothèses.

Supposons que l'on fasse un test sur le paramètre inconnu  $b$ . On distinguera types de tests sur un paramètre.

Dans le cas des **tests unilatéraux**, on teste l'égalité du paramètre inconnu à la valeur  $b_0$  sous l'hypothèse nulle contre la stricte supériorité (respectivement infériorité) à cette même valeur sous l'hypothèse alternative :

$$\begin{array}{l} H_0 : b = b_0 \\ H_1 : b > b_0 \end{array} \quad \text{ou} \quad \begin{array}{l} H_0 : b = b_0 \\ H_1 : b < b_0 \end{array}$$

Dans le cas des **tests bilatéraux**, on teste l'égalité du paramètre inconnu à la valeur  $b_0$  sous l'hypothèse nulle contre la différence (c'est-à-dire la stricte supériorité ou infériorité) à cette même valeur sous l'hypothèse alternative :

$$\begin{array}{l} H_0 : b = b_0 \\ H_1 : b \neq b_0 \end{array}$$

C'est généralement ces deux derniers types de tests qu'on formulera dans nos analyses.

### 3.1.2 La définition des risques et leurs probabilités

A partir de là, nous avons deux actions possibles, et donc deux possibilités de se tromper. Ainsi, « **se tromper** » revient à « **choisir une hypothèse alors que l'autre est vraie** ». On peut alors définir deux types de risques :

- **le risque de première espèce** est le risque de refuser l'hypothèse nulle alors qu'elle est vraie. On définit la probabilité de ce risque comme  $\alpha = \Pr\{\bar{H}_0 | H_0\}$ .
- **Le risque de seconde espèce** est le risque d'accepter l'hypothèse nulle alors qu'elle est fautive. On définit la probabilité de ce risque comme  $\beta = \Pr\{H_0 | H_1\}$ .

On notera que le conditionnement ne se fait pas par rapport à un événement mais par rapport à un état de la nature que l'on n'observe pas. De plus, alors qu'on peut rencontrer des notations laissant entendre qu'on choisit l'hypothèse alternative, il doit être bien clair que l'on accepte ou que l'on rejette l'hypothèse nulle et rien d'autre ! L'hypothèse alternative ne permet que de définir la zone de rejet.

### 3.1.3 La règle de décision

Il s'agit ensuite de trouver un outil, c'est-à-dire une statistique, nous permettant de mesurer l'adéquation entre l'hypothèse formulée et la réalité observée, et ainsi d'apporter une réponse à la question posée. Il s'agit d'une **fonction discriminante (ou pivotale) construite sous l'hypothèse nulle, dont on va comparer la valeur observée à la valeur théorique calculée à partir de la loi de probabilités et définie à partir de la zone de rejet de l'hypothèse nulle construite à partir de l'hypothèse alternative pour une probabilité de risque de première espèce  $\alpha$  donnée** (méthodologie de Neymann-Pearson). Une règle de décision est ensuite formulée, permettant de comparer la valeur particulière de la fonction discriminante pour l'échantillon utilisé à la valeur théorique issue de la table statistique.

### 3.1.4 Exemples

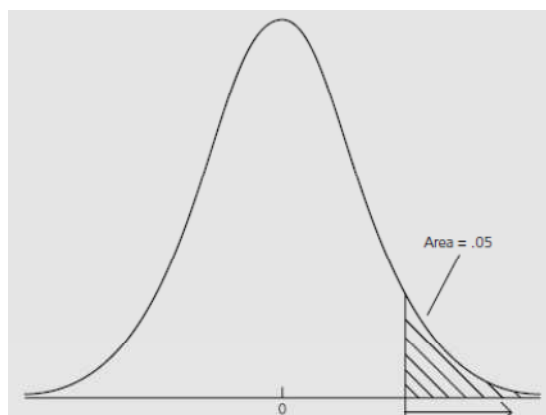
Dans le cas du test unilatère sur le paramètre inconnu  $b$  :

$$H_0 : b = b_0$$

$$H_1 : b > b_0$$

on utilisera comme fonction discriminante la loi suivie par l'estimateur de ce paramètre formulée sous l'hypothèse nulle. Ainsi, pour  $\hat{b} \square N(b, \sigma_b)$  (on supposera ici  $\sigma_b$  connu), on écrira sous  $H_0$  :

$$\frac{\hat{b} - b_0}{\sigma_{\hat{b}}} \square N(0,1)$$



L'hypothèse alternative permet de définir la zone critique, c'est à dire la zone de rejet de l'hypothèse nulle connaissant la probabilité du risque de première espèce  $\alpha$ . On raisonne de la manière suivante : du fait de l'incertitude due au modèle, à l'échantillonnage, ..., on ne pourra probablement pas observer exactement l'égalité à la valeur  $b_0$  même si c'est bien le cas. On va donc se laisser **une certaine marge** pour accepter l'hypothèse nulle. Posons que cette marge d'acceptation va jusqu'à une valeur  $c$ , pour le moment inconnue mais supérieure à  $b_0$  (du fait de l'hypothèse alternative). L'action d'accepter ou de rejeter l'hypothèse nulle va donc se traduire dans la position du  $\hat{b}$  constatée relativement à cette marge :

- à l'intérieur de cette marge, on peut statistiquement accepter l'hypothèse d'égalité du paramètre inconnu  $b$  à la valeur  $b_0$  supposée dans l'hypothèse nulle.
- Au delà de cette marge, l'hypothèse n'est plus statistiquement acceptable.

C'est ainsi que l'on va confronter la pertinence de notre hypothèse à la réalité des données. La probabilité du risque de première espèce, à savoir rejeter l'hypothèse nulle alors qu'elle est vraie, se traduit par l'égalité suivante et par l'observation de  $\hat{b}$  qui est au delà de la marge acceptable  $c$  :

$$\alpha = \Pr\{\overline{H_0} | H_0\} = \Pr\{\hat{b} > c | b = b_0\} = \Pr\left\{ \underbrace{\frac{\hat{b} - b_0}{\sigma_{\hat{b}}}}_{=Z \sim N(0,1)} > \underbrace{\frac{c - b_0}{\sigma_{\hat{b}}}}_t \mid b = b_0 \right\}$$

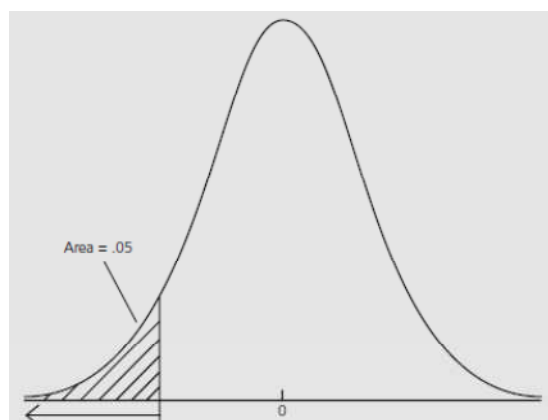
où  $c$  est la valeur critique au delà de laquelle il n'est plus tenable de défendre l'hypothèse nulle et où on doit donc la rejeter. Il ne reste alors plus qu'à trouver la valeur de  $t$  dans la table de la loi normale centrée réduite pour cette probabilité  $\alpha$  et à la comparer à  $\frac{\hat{b} - b_0}{\sigma_{\hat{b}}}$

pour savoir si on se place dans la zone d'acceptation ( $\frac{\hat{b} - b_0}{\sigma_{\hat{b}}} < t$ ) ou de rejet ( $\frac{\hat{b} - b_0}{\sigma_{\hat{b}}} > t$ ) de l'hypothèse nulle.

Dans le cas du test unilatère sur le paramètre inconnu  $b$  :

$$\begin{aligned} H_0 &: b = b_0 \\ H_1 &: b < b_0 \end{aligned}$$

Ici, la marge d'acceptation est toujours au voisinage de  $b_0$  jusqu'à une autre valeur  $c$ , pour le moment inconnue, mais inférieure à  $b_0$  (du fait de l'hypothèse alternative).



L'action d'accepter ou de rejeter l'hypothèse nulle va donc se traduire dans la position du  $\hat{b}$  constatée relativement à cette marge :

- A l'intérieur de cette marge, on peut statistiquement accepter l'hypothèse d'égalité du paramètre inconnu  $b$  à la valeur  $b_0$  supposée dans l'hypothèse nulle.
- Au delà de cette marge, l'hypothèse n'est plus statistiquement acceptable.

C'est ainsi que l'on va confronter la pertinence de notre hypothèse à la réalité des données. Comme précédemment, la probabilité du risque de première espèce ne change pas. Ce qui change, c'est la façon de rejeter l'hypothèse nulle :

$$\alpha = \Pr\{\overline{H_0} | H_0\} = \Pr\{\hat{b} < c | b = b_0\} = \Pr\left\{ \underbrace{\frac{\hat{b} - b_0}{\sigma_{\hat{b}}}}_{=Z \sim N(0,1)} < \underbrace{\frac{c - b_0}{\sigma_{\hat{b}}}}_t \mid b = b_0 \right\}$$

Il ne reste alors plus qu'à trouver la valeur de  $t$  dans la table de la loi normale centrée réduite pour la probabilité  $\alpha$  et à le comparer à  $\frac{\hat{b} - b_0}{\sigma_{\hat{b}}}$  pour savoir si on se place dans la

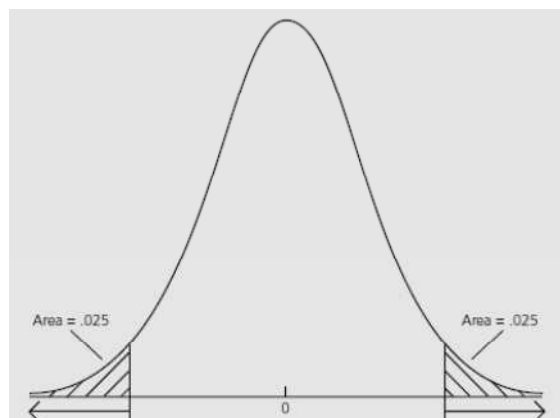
zone d'acceptation ( $\frac{\hat{b} - b_0}{\sigma_{\hat{b}}} > t$ ) ou de rejet ( $\frac{\hat{b} - b_0}{\sigma_{\hat{b}}} < t$ ) de l'hypothèse nulle.

Enfin, dans le cas du test bilatère sur le paramètre inconnu  $b$  :

$$H_0 : b = b_0$$

$$H_1 : b \neq b_0$$

Ici, la marge d'acceptation va se situer autour de  $b_0$ , jusqu'à deux valeurs  $c_1$  et  $c_2$ , pour le moment inconnues situées de part et d'autre de  $b_0$  (toujours du fait de l'hypothèse alternative).



L'action d'accepter ou de rejeter l'hypothèse nulle va donc se traduire dans la position du  $\hat{b}$  constatée relativement à cette marge d'acceptation :

- A l'intérieur de cette marge, on peut statistiquement accepter l'hypothèse d'égalité du paramètre inconnu  $b$  à la valeur  $b_0$  supposée dans l'hypothèse nulle.
- Au delà de cette marge (d'un coté ou d'un autre), l'hypothèse n'est plus statistiquement acceptable.

En appliquant la définition de la probabilité du risque de première espèce, et en adaptant la règle de rejet de l'hypothèse nulle :

$$\alpha = \Pr\{\overline{H_0} | H_0\} = \Pr\{\hat{b} \notin [c_1, c_2] | b = b_0\} \Leftrightarrow 1 - \alpha = \Pr\{\hat{b} \in [c_1, c_2] | b = b_0\}$$

$$\Leftrightarrow 1 - \alpha = \Pr\left\{ \underbrace{\frac{c_1 - b_0}{\sigma_{\hat{b}}}}_{-t} < \underbrace{\frac{\hat{b} - b_0}{\sigma_{\hat{b}}}}_{=Z \square N(0,1)} < \underbrace{\frac{c_2 - b_0}{\sigma_{\hat{b}}}}_t \mid b = b_0 \right\}$$

Il ne reste alors plus qu'à trouver la valeur de  $t$  associée la probabilité  $\alpha$  dans la table de la loi normale centrée réduite et à les comparer à  $\frac{\hat{b} - b_0}{\sigma_{\hat{b}}}$  pour savoir si on se place dans la zone d'acceptation ( $t < \frac{\hat{b} - b_0}{\sigma_{\hat{b}}} < t$ ) ou de rejet de l'hypothèse nulle (dans le cas contraire).

### 3.1.5 La performance du test

Enfin, et nous nous y intéresserons dans une moindre mesure pour ce qui est de ce cours, il est nécessaire d'étudier la performance du test mis en œuvre. Ainsi, la puissance, c'est-à-dire la probabilité de rejeter l'hypothèse nulle lorsqu'elle est fautive, devra être forte pour un test performant.

## 3.2 Test sur la valeur d'un paramètre du modèle

### 3.2.1 Description générale

Pour les paramètres estimés dans le cadre d'un modèle économétrique, on a vu que la distribution opérationnelle de  $\hat{\beta}_j$  était une loi de Student à  $T-k$  degrés de liberté :

$$\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \square T_{T-k}$$

On veut tester l'égalité du paramètre inconnu à une valeur particulière  $\beta_0$ . Les hypothèses s'écrivent donc :

$$H_0 : \beta_j = \beta_0$$

$$H_1 : \beta_j \neq \beta_0$$

La fonction discriminante sous  $H_0$  est donc :

$$\frac{\hat{\beta}_j - \beta_0}{\hat{\sigma}_{\hat{\beta}_j}} \square T_{T-k}$$

pour une probabilité du risque de première espèce de  $\alpha$ . La règle de décision s'écrit :

$$\Pr\left\{ \underbrace{\frac{c_1 - \beta_0}{\hat{\sigma}_{\hat{\beta}_j}}}_{-t} < \underbrace{\frac{\hat{\beta}_j - \beta_0}{\hat{\sigma}_{\hat{\beta}_j}}}_{\square T_{T-k}} < \underbrace{\frac{c_2 - \beta_0}{\hat{\sigma}_{\hat{\beta}_j}}}_t \mid \beta = \beta_0 \right\} = 1 - \alpha$$

Il est possible de lire la valeur de  $t$  dans la table de Student en fonction de la probabilité du risque de première espèce et du nombre de degrés de liberté de la loi, et donc d'en déduire la



zone d'acceptation. Si  $\frac{\hat{\beta}_j - \beta_0}{\hat{\sigma}_{\hat{\beta}_j}} \in [-t, t]$ , alors l'hypothèse nulle est acceptable. Dans le cas contraire, elle n'est pas soutenable.

### 3.2.2 Application particulière

Un test très répandu dit **de significativité** cherche à tester la pertinence de chaque variable explicative dans la régression. On envisage pour cela la nullité du paramètre inconnu qui lui est associé :

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

Ainsi, si  $\frac{\hat{\beta}_j}{\hat{\sigma}_{\hat{\beta}_j}} \in [-t, t]$ , alors l'hypothèse nulle est acceptable et le paramètre inconnu n'est pas significativement différent de 0 au seuil de  $\alpha$ . Dans le cas contraire, l'hypothèse nulle est rejetée et la variable explicative est statistiquement pertinente.

### 3.2.3 De l'utilisation des intervalles de confiance pour les tests d'hypothèses bilatères

On peut utiliser les intervalles de confiance pour répondre aux questions posées par les tests dans le cas particulier où :

- le test est bilatéral
- le niveau de confiance de l'intervalle est égal à  $1 - \alpha$ , la probabilité de risque de première espèce avec lequel le test est réalisé.

Dans ces conditions en effet, on constate que ***l'intervalle de confiance correspond à la zone d'acceptation du test.***

Il est alors inutile de mener un test si ces conditions sont remplies. ***Il suffit de regarder si la valeur supposée dans l'hypothèse nulle appartient ou non à l'intervalle de confiance.*** Si c'est le cas, on est dans la zone d'acceptation et on accepte l'hypothèse nulle ; dans le cas contraire, on la rejette.

## 3.3 Tests joints sur un sous-ensemble de paramètres

Supposons que l'on se place dans le modèle linéaire de régression multiple avec  $k$  variables explicatives (y compris la constante) et formulé sous les hypothèses standards. On s'intéresse à poser  $r$  hypothèses sur les paramètres inconnus. Ces hypothèses vont prendre la forme de contraintes linéaires sur les paramètres inconnus du modèle.

### 3.3.1 Quelques exemples de tests joints

**Exemple 1 :** supposons la fonction de production suivante :

$$y_t = a + \alpha l_t + \beta k_t + u_t$$

On va par exemple chercher à tester l'hypothèse de rendements constants. Pour cela, on va poser :

$$H_0 : \alpha + \beta = 1$$

$$H_1 : \alpha + \beta \neq 1$$

**Exemple 2 :** soit l'équation de salaire suivante :

$$w_t = a + \alpha_0 p_t + \alpha_1 p_{t-1} + \alpha_2 p_{t-2} + \gamma Z_t + u_t$$

avec  $Z$  les autres variables explicatives du modèle. On va par exemple chercher à tester l'hypothèse de parfaite indexation des salaires sur les prix. Pour cela, on va poser :

$$H_0 : \alpha_0 + \alpha_1 + \alpha_2 = 1$$

$$H_1 : \alpha_0 + \alpha_1 + \alpha_2 \neq 1$$

**Exemple 3 :** soit le modèle de régression multiple suivant :

$$y_t = a + b_1 x_{1,t} + b_2 x_{2,t} + b_3 x_{3,t} + b_4 x_{4,t} + u_t$$

On va par exemple chercher à tester l'hypothèse de l'égalité des paramètres  $b_1$  et  $b_2$  et le fait que  $b_3 = -b_4$ . Pour cela, on va poser :

$$H_0 : b_1 = b_2 \text{ et } b_3 = -b_4$$

$$H_1 : H_0 \text{ fautive}$$

**Exemple 4 :** soit le modèle de régression multiple suivant :

$$y_t = a + b_1 x_{1,t} + b_2 x_{2,t} + \dots + b_{k-1} x_{k-1,t} + u_t$$

On va par exemple chercher à tester l'hypothèse de non pertinence de la régression. Pour cela, on va poser :

$$H_0 : b_1 = b_2 = \dots = b_{k-1} = 0$$

$$H_1 : H_0 \text{ fautive}$$

### 3.3.2 La méthodologie

On a besoin d'un cadre général où formuler ces différentes hypothèses. L'idée est donc d'exprimer le test de la façon suivante :

$$H_0 : \begin{cases} y = X \cdot \beta + u \\ C \cdot \beta = c \end{cases} \begin{matrix} (r \times k) & (k \times 1) & (r \times 1) \end{matrix}$$

$$H_1 : y = X \cdot \beta + u$$

La matrice  $C$  résume les  $r$  relations supposées entre les paramètres inconnus du modèle. La matrice  $C$  et le vecteur  $c$  permettent de réécrire les contraintes linéaires de la façon suivante :

$$C \cdot \beta = c \begin{matrix} (r \times k) & (k \times 1) & (r \times 1) \end{matrix}$$

où  $r$  est le nombre d'hypothèses formulées. Le rang de la matrice  $C$  est donc égal à  $r$ .

Ainsi, dans l'exemple 1, on a formulé  $r = 1$  contrainte pour 3 paramètres inconnus. La matrice  $C$  est de format  $(1 \times 3)$  et  $c$  est un scalaire. On écrira alors l'hypothèse nulle comme :

$$(0 \quad 1 \quad 1) \cdot \begin{pmatrix} a \\ \alpha \\ \beta \end{pmatrix} = 1$$

Dans l'exemple 2, on a formulé  $r = 1$  contrainte pour au moins 4 paramètres inconnus plus un vecteur de paramètres inconnus. La matrice  $C$  est de format  $(1 \times 5)$  et  $c$  est un scalaire. On écrira alors l'hypothèse nulle comme :

$$(0 \ 1 \ 1 \ 1 \ 0) \cdot \begin{pmatrix} a \\ \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \gamma \end{pmatrix} = 1$$

Dans l'exemple 3, on a formulé  $r = 2$  contraintes pour 5 paramètres inconnus. La matrice  $C$  est de format  $(2 \times 5)$  et  $c$  est un vecteur de taille 2. On écrira alors l'hypothèse nulle comme :

$$\begin{pmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} a \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Enfin, dans l'exemple 4, on a formulé  $r = k-1$  contraintes pour  $k$  paramètres inconnus. La matrice  $C$  est de format  $(k-1 \times k)$  et  $c$  est un vecteur de taille  $k-1$ . On écrira alors l'hypothèse nulle comme :

$$\begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{k-1} \\ a \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Il est alors nécessaire de trouver une fonction discriminante pour mener à bien ce test.

### 3.3.3 Le test

Comme précédemment, on sait que :

$$\hat{\beta} \underset{(k \times 1)}{\square} N(\beta, \sigma^2 \underset{(k \times k)}{(X'X)^{-1}})$$

ce qui implique :

$$\underset{(r \times 1)}{(C \cdot \hat{\beta} - c)} \underset{(r \times 1)}{\square} N \left[ \underset{(r \times 1)}{(C \cdot \beta - c)}, \sigma^2 \underset{(r \times r)}{C(X'X)^{-1}C'} \right]$$

d'après les propriétés de la loi normale et des opérateurs espérance et variance. Or sous  $H_0$ , on suppose que  $C \cdot \beta = c$ . Dès lors, la distribution de  $C \cdot \hat{\beta} - c$  sous  $H_0$  peut donc se réécrire :

$$C \cdot \hat{\beta} - c \underset{(r \times 1)}{\square} N(0, \sigma^2 C(X'X)^{-1}C')$$

dont on déduit :

$$[C \cdot \hat{\beta} - c]' [\sigma^2 C(X'X)^{-1}C']^{-1} [C \cdot \hat{\beta} - c] \underset{(r)}{\square} \chi^2(r)$$

d'après le résultat R3 sur les formes quadratiques du chapitre précédent (section 6).

Comme à la section précédente, cette statistique n'est pas opérationnelle puisqu'on ne connaît pas la vraie valeur de  $\sigma^2$ .

On va donc remplacer  $\sigma^2$  par son estimation  $\hat{\sigma}^2$ .

$$[C \cdot \hat{\beta} - c]' [\hat{\sigma}^2 C(X'X)^{-1} C']^{-1} [C \cdot \hat{\beta} - c]$$

Ce choix modifie la loi suivie par la statistique. **Il est donc nécessaire de se ramener à une loi connue.** Modifions la statistique en divisant l'expression précédente par  $r$ . On peut alors écrire :

$$F = \frac{[C \cdot \hat{\beta} - c]' [C(X'X)^{-1} C']^{-1} [C \cdot \hat{\beta} - c]}{r \cdot \hat{\sigma}^2}$$

En faisant apparaître les termes adéquats ( $\sigma^2$  et  $T-k$ ) au numérateur et au dénominateur de l'expression, on obtient alors facilement :

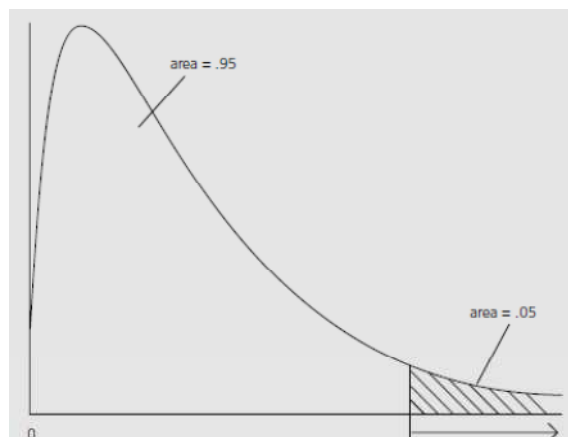
$$F = \frac{\frac{[C \cdot \hat{\beta} - c]' [C(X'X)^{-1} C']^{-1} [C \cdot \hat{\beta} - c]}{r \cdot \hat{\sigma}^2}}{\frac{1}{\sigma^2 (T-k)}} = \frac{\frac{[C \cdot \hat{\beta} - c]' [\sigma^2 C(X'X)^{-1} C']^{-1} [C \cdot \hat{\beta} - c]}{r}}{\frac{1}{\sigma^2 (T-k)}} = \frac{\chi^2(r)}{\frac{\chi^2(T-k)}{k}}$$

On reconnaît un rapport de formes quadratiques suivant des lois du  $\chi^2$  divisées par leur nombre de degrés de liberté. **On est donc en présence d'une loi de Fisher si ces deux lois sont indépendantes.** Pour démontrer l'indépendance, il suffit d'utiliser le résultat préliminaire n°1 de la section 1 concernant l'indépendance de deux formes quadratiques.

**La fonction discriminante pour les tests impliquant des combinaisons linéaires entre les paramètres inconnus est donc la statistique**

$$F = \frac{[C \cdot \hat{\beta} - c]' [C(X'X)^{-1} C']^{-1} [C \cdot \hat{\beta} - c]}{r \cdot \hat{\sigma}^2}$$

**qui suit une loi de Fisher ici à  $r$  et  $T-k$  degrés de liberté.**



**Questions :** que constate-t-on si on pose  $C$  comme l'identité de taille  $k$  et  $c = \beta_0$ ? Ce test peut-il également s'appliquer pour le cas d'une contrainte sur un seul paramètre ?

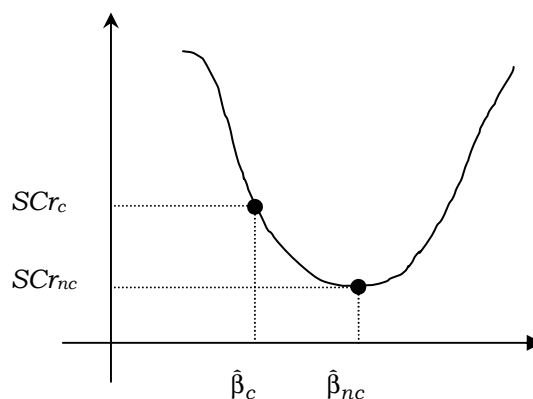
## 4 L'estimation sous contraintes linéaires

On va dans cette section présenter l'estimateur  $\hat{\beta}_c$  des MC sous contraintes (MCC) et étudier ses propriétés.

**On va voir qu'il est équivalent**

- **d'estimer un modèle à  $k$  paramètres inconnus par les MCC sous  $r$  contraintes, ce qui donnera l'estimateur  $\hat{\beta}_c$**
- **et d'intégrer ces  $r$  contraintes dans le modèle et d'estimer ce modèle contraint à  $k - r$  paramètres par les MCO.**

Intuitivement, on se dit que si la contrainte était parfaitement valable, l'estimateur des MCO (non contraint) se positionnerait naturellement de façon à respecter la contrainte. Cependant, il faut aussi envisager, pour les mêmes raisons que pour les tests réalisés précédemment, que les caractéristiques de l'échantillon utilisé ne permettent pas de mettre en exergue complètement ces contraintes. Dès lors, il est nécessaire de se demander si ces contraintes sont statistiquement soutenables au regard des données ou pas.



On pourra envisager cette question au travers de la distance entre les quantités  $SCR_{nc}$  et  $SCR_c$ , qui caractérise la distance entre le modèle sous contraintes et le modèle non contraint, c'est-à-dire entre l'hypothèse nulle formulée et les données. **Cette approche permettra donc d'aborder le problème des tests d'hypothèses simples ou jointes vu à la section précédente de façon plus simple.**

### 4.1 Retour sur les exemples précédents

En examinant les exemples présentés précédemment, on constate que les hypothèses peuvent le plus souvent être examinées dans le cadre d'un modèle contraint, c'est-à-dire en estimant des modèles modifiés.

Ainsi, en ce qui concerne l'exemple 1, une solution possible consisterait à estimer le modèle contraint (écrit sous l'hypothèse nulle) suivant :

$$y_t - k_t = a + \alpha(l_t - k_t) + \varepsilon_t$$

On n'a donc plus que 2 paramètres à estimer. Ce modèle possède  $T-2$  degrés de liberté.

Pour l'exemple 2, l'hypothèse nulle consiste à estimer le modèle suivant :

$$w_t - p_{t-2} = a + \alpha_0(p_t - p_{t-2}) + \alpha_1(p_{t-1} - p_{t-2}) + \gamma Z_t + \varepsilon_t$$

On n'a donc plus que 4 paramètres à estimer. Ce modèle possède  $T-4$  degrés de liberté.

Pour l'exemple 3, l'hypothèse nulle consiste à estimer le modèle suivant :

$$y_t = a + b_1(x_{1,t} + x_{2,t}) + b_3(x_{3,t} - x_{4,t}) + \varepsilon_t$$

On n'a donc plus que 2 paramètres à estimer. Ce modèle possède  $T-2$  degrés de liberté. Enfin, pour l'exemple 4, l'hypothèse nulle consiste à estimer le modèle suivant :

$$y_t = a + \varepsilon_t$$

On n'a donc plus qu'un paramètre à estimer. Ce modèle possède  $T-1$  degrés de liberté. Il reste à présent à comparer la « distance » entre le modèle contraint et le modèle non contraint. Il va falloir bien évidemment pour cela construire un test.

## 4.2 La distance entre les modèles contraint et non contraint

### 4.2.1 Réinterprétation du test joint sur les paramètres inconnus

On a vu que la statistique de test joint sur une partie des paramètres inconnus pouvait s'écrire :

$$F = \frac{[C \cdot \hat{\beta} - c]' [C(X'X)^{-1}C']^{-1} [C \cdot \hat{\beta} - c]}{r \cdot \hat{\sigma}^2}$$

Cette expression peut se traduire comme la distance entre le modèle contraint et le modèle non contraint. Ainsi :

$$[C \cdot \hat{\beta} - c]' [C(X'X)^{-1}C']^{-1} [C \cdot \hat{\beta} - c] = SCr_c - SCr_{nc}$$

Dès lors, comme :

$$\hat{\sigma}^2 = \frac{SCr_{nc}}{T-k} = \frac{\hat{u}'\hat{u}}{T-k}$$

On peut réécrire :

$$F = \frac{[C \cdot \hat{\beta} - c]' [C(X'X)^{-1}C']^{-1} [C \cdot \hat{\beta} - c]}{r \cdot \hat{\sigma}^2} = \frac{SCr_c - SCr_{nc}}{\frac{r}{T-k} SCr_{nc}}$$

On a vu que le nombre de degrés de liberté associé à chaque somme des carrés des résidus est égal à  $dl_c = T - (k - r)$  dans le cas de la régression contrainte et à  $dl_{nc} = T - k$  dans celui de la régression non contrainte. On peut alors réécrire :

$$F = \frac{[C \cdot \hat{\beta} - c]' [C(X'X)^{-1}C']^{-1} [C \cdot \hat{\beta} - c]}{r \cdot \hat{\sigma}^2} = \frac{SCr_c - SCr_{nc}}{\frac{dl_c - dl_{nc}}{dl_{nc}} SCr_{nc}} \square F(dl_c - dl_{nc}, dl_{nc})$$

## 4.2.2 Réécriture et applications

On montre que cette expression peut se réécrire :

$$F = \frac{\frac{SCR_c - SCR_{nc}}{dl_{nc}}}{\frac{SCR_{nc}}{dl_{nc}}} = \frac{1 - 1 + \frac{SCR_c}{SCT} - \frac{SCR_{nc}}{SCT}}{\frac{dl_c - dl_{nc}}{dl_{nc}}} = \frac{R_{nc}^2 - R_c^2}{1 - R_{nc}^2} \frac{T - k}{r}$$

car  $SCT = SCE_{nc} + SCR_{nc} = SCE_c + SCR_c$  s'il y a des constantes dans les modèles.

**Une application directe est le test de significativité globale, qui permet de tester la pertinence globale d'une régression, c'est-à-dire la nullité des paramètres de toutes les variables explicatives à l'exception de la constante.**

On montre alors que la statistique de test se réécrit :

$$F = \frac{R_{nc}^2 - R_c^2}{1 - R_{nc}^2} \frac{T - k}{r} = \frac{R_{nc}^2}{1 - R_{nc}^2} \frac{T - k}{k - 1} \square F(k - 1, T - k)$$

puisqu'il est facile de voir que pour le modèle contraint,  $SCT = SCR_c$  et donc  $SCE_c = 0$  d'après l'équation de la variance. Cette écriture du test est très pratique et très souvent utilisée car elle ne requiert finalement que le calcul du  $R^2$  de la régression à tester.  $\square$

## Références bibliographiques

Dormont B., Introduction à l'Econométrie, Editions Montchrétien.

Greene W.H., Econometric Analysis, Mac Millan.

Maddala G., Econometric methods and applications,