UNIVERSITE DE LISALA

CENTRE INTERUNIVERSITAIRE DE RECHERCHE PLURIDISCIPLINAIRE (CIREP)
STATUT : UNIVERSITE PUBLIQUE
Web : www.cirep.ac.cd
Email : info@cirep.ac.cd

NOTES DE COURS DE QUESTIONS SPECIALES D'ECONOMETRIE

OBJECTIFS DU COURS

Économétrie à trois objectifs essentiels :

- ✓ L'analyse, c'est-à-dire le test de la théorie ;
- ✓ L'aide à la prise de décision, c'est-à-dire fournir des estimateurs numériques auxcoefficients des relations économiques ;
- ✓ La prévision, les estimations numériques des coefficients sont utilisées dansl'optique de donner une prévision des valeurs futures.

I. Introduction

1.1 Le domaine de l'analyse économique

L'économie est une science sociale, elle étudie le problème des choix dans une société d'un point de vue scientifique, à partir d'une exploration systématique de ce problème. Cette exploration systématique passe aussi bien par la formulation de théorie que par l'examen de données empiriques. Une théorie est fondée sur une série d'hypothèses et de conclusion découlant de ces hypothèses. Les théories sont des exercices logiques : si les hypothèses sont correctes, alors les résultats se vérifient nécessairement.

La science économique a défini des théories pour formaliser les relations entre les variables dans leurs interrelations. Ainsi par exemples : **La théorie de la production** nous enseigne entre autres que la production (Y) d'un bien ou service est fonction d'un ensemble de facteurs de production entre autres le capital (K) et le travail (L) : $Y=f(K, L)$; **La théorie du consommateur** nous instruit quant à elle que la demande d'un bien X par un consommateur est fonction du revenu de ce consommateur R, du prix du bien Px et du prix des autres biens Py : $Q = f(R, Px, Py\ldots)$, etc.

Le rôle de l'économétrie est de tester les propositions théoriques qui fondent ces relations entre variables, d'obtenir des estimations numériques des coefficients de ces relations, de faire des prévisions par rapport à leurs valeurs futures et de faire des recommandations pertinentes à même d'aider à la prise de décision.

1.2 La théorie économique et l'économétrie

L'économie comme nous l'avons souligné plus haut est une science sociale. Son objet d'étude est la société, et le comportement des institutions et des ménages qui la composent. La théorie économique essaie d'expliquer les relations entre variables économiques et utilise les informations obtenues dans un cadre théorique général pour expliquer l'affectation des ressources, la production et la décision de répartition au sein d'un système qui doit fonctionner dans un contexte de rareté.
L'économétrie essaie quant à elle, à partir de l'analyse statistique de données limitées, de tirer des conclusions relatives au monde réel. Ainsi, la théorie économique tout comme l'économétrie essaye de produire une série d'information pour améliorer la prise de décision.

1.3 Définition, objectifs et nécessité de l'économétrie

a) Définition

Au sens littéral, l'économétrie signifie « mesure de l'économie ». Bien que la mesure soit une part importante de l'économétrie, le domaine de cette discipline est plus vaste ; les citations suivantes en témoignent :

- L'économétrie consiste à appliquer les mathématiques statistiques aux données économiques pour fournir une base empirique aux modèles construits par l'économie mathématique et obtenir des résultats mesurés.

- L'économétrie peut être définie comme l'analyse quantitative des phénomènes économiques actuels.

L'économétrie peut être définie comme l'application de méthodes statistique et mathématique à l'analyse de statistiques économiques en vue de donner un contenu empirique aux théories économiques et de les vérifier ou de les réfuter.

En fait, l'économétrie est une discipline essentiellement basée sur la construction de modèles économétriques.

b. La nécessité de l'analyse économétrique

Si l'objectif de l'opérateur économique est de prendre la meilleure décision parmi un ensemble de choix possibles, alors la connaissance de la direction de la relation et, dans plusieurs cas celle de l'ampleur de la relation est nécessaire pour ne pas dire indispensable. Ainsi, en fournissant à l'opérateur économique un support pour tester, modifier ou réfuter si possible les conclusions contenues dans les théories économiques, et d'affecter des signes, et des interprétations fiables aux coefficients des variables économiques, l'économétrie se positionne comme un outil indispensable à la prise de décision.

1.4 Méthodologie de la recherche économétrique

1.5 La construction de modèle économique

La formulation d'un modèle économique, de même que la détermination du nombre de variables à inclure dans ce modèle dépendront essentiellement des objectifs pour lesquels le modèle est bâtit. On peut noter aussi que la complexité du modèle est fonction du degré d'information que l'on voudrait obtenir pour le modèle.

Par exemple : le modèle économique de l'offre et de la demande cherche traditionnellement à expliquer la relation prix-quantité sur un marché particulier. Ce modèle comprendra trois (3) équations :

a. l'équation de la demande

$$Q_D = \alpha_0 - \alpha_1 P_1 + \alpha_2 Y_d \quad (1 - 1)$$

b. l'équation de l'offre

$$Q_D = \beta_0 + \beta_1 P_1 - \beta_2 R_2 \quad (1 - 2)$$

c. l'équation d'équilibre de marché

$$Q_D = Q_O \qquad (1 - 3)$$

(

Avec QD = Quantité de biens demandée ;QO = Quantité de biens offerte P1 = Prix du bien ; R2 = Prix moyen des facteurs ; Yd = Revenu disponible

Les modèles plus compliqués peuvent tenter d'expliquer le comportement de plusieurs variables. Disons cependant que les modèles économiques qui relèvent de la microéconomie ou de la macroéconomie qui s'adressent à l'économie d'un pays ou à une industrie, à une firme ou à un marché particulier ont des caractéristiques en commun.

En effet, les modèles sont d'abord formulés sur la base de comportement des variables économiques et sont déterminées par l'opération simultanée d'un certain nombre d'opérations économiques. Ensuite, celui qui construit le modèle part d'un principe, même si les modèles sont une signification du monde réel des plus complexes, ils retiendront des aspects importants du secteur économique ou du système économique étudié. Enfin, le constructeur du modèle a la conviction que les informations et le degré de compréhension fournit par les modèles permettront de prédire le comportement du système dans le futur et probablement de contrôler le comportement afin d'améliorer le fonctionnement du système économique.

A titre d'exemple considérons un modèle économique basé sur les hypothèses suivantes :

H1 : la consommation est fonction croissante du revenu disponible mais cette consommation croit dans une proportion inférieure à celle du revenu disponible ; la propension marginale à consommer du revenu disponible est positive mais inférieure àl'unité (0≤ Pmc ≤1).

H2 : l'investissement est fonction directe du revenu national, mais il est fonction décroissante du taux d'intérêt.

H3 : le revenu national est la somme de la consommation, de l'investissement et des dépenses gouvernementales pour les biens et services.

Avant de traduire les propositions que nous venons de faire en langage mathématique, il est essentiel de résoudre le problème de spécification, la nature des relations entre les variables endogènes que sont la consommation, l'investissement et le revenu national d'une part et les variables exogènes tels que l'impôt, les dépenses gouvernementales et le taux d'intérêt d'autre part. Ces relations devront être linéaires ou non linéaires ? Si elles sont non linéaires, devront elles avoir une forme quadratique ou même avoir une forme de degré ≥ 2. Par ailleurs, on devrait se poser la question de

Savoir si cet investissement est aussi fonction des périodes antérieures T-2 ; T-3 et les profits antérieurs ou anticipés. Pour simplifier notre modèle, nous formulerons de la façon suivante :

$$C_t = \alpha_0 + \alpha_1(Y_t - T_t) \quad \text{(1- 4) Equation de comportement}$$
$$I_t = \beta_1 Y_{t-1} + \beta_2 R_t \quad \text{(1- 5) Equation de comportement}$$
$$Y_t = C_t + I_t + G_t \quad \text{(1- 6)}$$

Avec les restrictions $0<\alpha_1<1$; $\beta_1>0$; $\beta_2<0$ Et C_t = Consommation de la période t

I_t = Investissement de la période t Y_t = Revenu national de la période t

G_t = Dépenses gouvernementales de la période

tT_t = Impôts sur le revenu de la période t

R_t = Taux d'intérêt de la période t

Notre modèle ainsi formulé comprend deux équations de comportement et une identité qui est l'équation (1 – 6).

Les variables C_t, I_t et G_t sont définies comme des variables endogènes de la période t. Quant aux variables R_t, Y_t, T_t elles sont classifiées comme variables exogènes de la période t. Y_{t-1} qui était une variable endogène, l'année antérieure à l'année t est considérée comme variable exogène de la période t.

Les équations (1-4), (1-5), (1-6) présentent le modèle sous sa forme structurelle. Quant à la forme réduite, elle présentera le modèle de sorte que les variables endogènes soient uniquement fonction des variables aux valeurs prédéterminées c'est-à-dire les variables exogènes.

La forme réduite de notre modèle s'obtiendra en substituant les équations (1-5), (1-6) dans l'équation (1-4) et nous obtiendrons :

$$C_t = \alpha_0 + \alpha_1(C_t + \beta_1 Y_{t-1} + \beta_2 R_t + G_t - T_t)$$
$$C_t = \alpha_0 + \alpha_1 C_t + \alpha_1\beta_1 Y_{t-1} + \alpha_1\beta_2 R_t + \alpha_1 G_t - \alpha_1 T_t$$
$$C_t = \frac{\alpha_0}{1-\alpha_1} + \frac{\alpha_1\beta_1}{1-\alpha_1} Y_{t-1} + \frac{\alpha_1\beta_2}{1-\alpha_1} R_t + \frac{\alpha_1}{1-\alpha_1} G_t - \frac{\alpha_1}{1-\alpha_1} T_t \quad \text{(1-7)}$$

L'équation (1-7) représente la forme réduite de la fonction de consommation avec 0<α1<1, quant à la fonction d'investissement, elle demeure inchangée :

$I_t = \beta_1 Y_{t-1} + \beta_2 R_t$ (1-8) parce qu'elle est déjà sous la forme réduite.

En effet, l'équation de l'investissement est uniquement exprimée en fonction des variables exogènes à la période t. Pour obtenir la forme réduite de l'équation (1-6), nous allons écrire que :

$$Y_t = \underbrace{\frac{\alpha_0}{1-\alpha_1} + \frac{\alpha_1\beta_1}{1-\alpha_1}Y_{t-1} + \frac{\alpha_1\beta_2}{1-\alpha_1}R_t + \frac{\alpha_1}{1-\alpha_1}G_t - \frac{\alpha_1}{1-\alpha_1}T_t}_{C_t \approx (1-7)} + \underbrace{\beta_1 Y_{t-1} + \beta_2 R_t}_{I_t \approx (1-8)} + G_t$$

$$Y_t = \frac{\alpha_0}{1-\alpha_1} + \frac{\beta_1}{1-\alpha_1}Y_{t-1} + \frac{\beta_2}{1-\alpha_1}R_t + \frac{G_t}{1-\alpha_1} - \frac{\alpha_1}{1-\alpha_1}T_t \quad (1\text{-}9)$$

(1-9) représente la forme réduite de la fonction du revenu national.

Le système d'équation (1-7), (1-8) et (1-9) représente la forme réduite de notre modèle macroéconomique initial représenté par les équations (1-4), (1-5) et (1-6).

Le terme de l'erreur (variable résiduelle)

La différence fondamentale entre l'économiste et l'économètre réside dans le fait que l'économètre se souci principalement du terme de l'erreur, de son importance et de son comportement. En effet, avant que l'économiste décidera que la fonction de consommation dépend du revenu disponible (Y_d) et écrira C_t=f(Y_d), l'économètre dira que cette relation devrait inclure un terme d'erreur. Ainsi donc il écrira sa fonction de consommation comme C_t = f(Y_d, O $_t$) = C_t O$_0$ Y_t

Sans la variable résiduelle ε_t, la fonction de consommation est une fonction déterministe, alors avec la prise en compte de la variable résiduelle elle devient stochastique.

Odésigne une variable aléatoire suivant une loi de probabilité déterminée. Le terme

Y ***mesure la différence entre les valeurs réellement observées de C_t et les valeurs***

qui auraient dues être observées si la relation fonctionnelle avait été rigoureusement exacte.

Pour mieux fixer les idées, supposons que nous avons des données en provenance d'une enquête sur les ménages, Y_d est le revenu disponible et C_t leurs dépenses de consommation. Il est clair que la dépense d'un ménage dépend, en plus de son revenu, de toute une série d'autres facteurs, tels sa taille, sa composition. Supposons donc que nous étudions la relation entre C_t et Yd_t pour les ménages ayant « *même taille et même composition* ». Malgré cela il ne serait pas réaliste de s'attendre à ce que tous les ménages avec un revenu Yd_t aient la consommation ($\alpha_0 + \alpha_1 Yd_t$). Premièrement, même parmi les ménages de taille et composition identique, il existera des différences dans l'âge des parents et des enfants. La consommation variera aussi en fonction des habitudes du mari (qui peut être un joueur de poker, un alcoolique…) ou celle de sa femme. Certains ménages ont un revenu qui augmente tandis que pour d'autres il diminue. Beaucoup de ces facteurs ne sont même pas qualifiables. Même si ces données existaient, le nombre de variables dépasserait le nombre d'observation. En outre, beaucoup de ces facteurs n'ont que des effets très faibles. C'est pourquoi nous représentons « l'effet net » de toutes ces influences possibles par une seule variable aléatoire ε_t.

Une deuxième raison pour ajouter le terme aléatoire à notre équation est qu'il y a peut être un élément de hasard, fondamental et imprévisible, dans le comportement humain.

Ce sont là deux raisons suffisantes pour introduire une variable aléatoire, leurs effets conjugués se répercutant au niveau de sa variance. La variable ε_t est souvent appelée élément de perturbation ou erreur de l'équation. On ne peut pas prédire la valeur de ε_t pour une quelconque observation, mais on peut proposer des hypothèses concernant les caractéristiques principales de sa loi de probabilité.

1.1 Définitions et concepts utilisés en économétrie

a. Variable aléatoire

Une variable aléatoire est une variable dont la valeur est inconnue jusqu'à ce quelle fasse l'objet d'une observation. La valeur d'une variable aléatoire résulte d'une expérimentation.

b. Variables aléatoires discrètes et continues

Une variable aléatoire est dite discrète si elle ne peut prendre qu'un nombre fini de valeur que l'on peut compter en utilisant des nombres entiers. Exemple :le nombre de fille dans un ménage : 1, 2, 3, …
Une variable est dite continue si elle peut prendre toute valeur réelle dans au moins un intervalle des nombres réels.

L'équation structurelle est une expression quantitative qui permet de déduire un modèle tel qu'il a été conceptualisé.

Equation de comportement : c'est une expression quantitative qui décrit le comportement des individus ou groupe d'individus. Exemple : fonction de consommation, d'investissement, d'épargne, etc.

Equation technique : c'est une expression qui montre comment les variables sont combinées pour donner naissance à une autre variable économique.

Exemple : fonction de production $Q \; AX^{b1} \; X^{b2}$ où X et X sont des inputs qui sont combinés pour donner Q qui est l'output ou le produit.

Variable économique

C'est une grandeur économique qui peut prendre des valeurs différentes d'une observation à une autre pour un problème économique donné. La variable économique peut être systématique,aléatoire, endogène ou exogène.

-La variable économique sera dite systématique si sa valeur peut être connue avec certitude selon des relations spécifiques

-Elle sera dite aléatoire lorsque la détermination de sa valeur ne découle pas d'un modèle prévisible avec certitude.

-Une variable économique est endogène lorsque sa valeur est déterminée par la structure économique considérée dans l'étude.

-Une variable économique exogène est celle dont la valeur économique est déterminée et connue à l'avance.

une série chronologique : une série chronologique est une série statistique dans laquelle les valeurs du caractère sont fonction du temps. Une série chronologique est encore appelée chronique ou série temporelle

Données en coupe transversale : ce sont des données qui sont observées au même moment pour un échantillon d'unités (ménages, firmes, villes, pays, etc.).

Population : Au sens économétrique, c'est l'ensemble de toutes les observations possibles sur une variable.

Echantillon : c'est un sous ensemble de la population

Constante : une constante est une magnitude qui ne change pas en par conséquent représente l'antithèse de la variable

Coefficient : Quand une constante est jointe à une variable, on l'appelle coefficient de cette variable. Exemple : Si l'on a l'expression 5R dans un modèle, alors 5 serait un coefficient et non pas une constante.

m Paramètre ou constante : Quand un nombre spécifique n'est pas assigné à un coefficient parce que ce coefficient est inconnu et par conséquent peut prendre virtuellement n'importe qu'elle valeur, on l'appelle constante paramétrique ou simplement paramètre. Exemple : Dans l'expression 5R, 5 est un coefficient. Mais dans l'expression aR, R étant une variable, (a) est un paramètre.

1.8 Notions simples de statistique

- Soit x une variable aléatoire, si x prend n valeurs $x_1, x_2, \cdots, x_n$ alors leur est donnée par : $\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$

Exemple, $x_i = 2, 5, 3, 7, 9, 6, 4$ $\quad \sum_{i=1}^{7} x_i = 2+5+3+7+9+6+4 = 36$

- Si a est une constante alors : $\sum_{i=1}^{n} a x_i = a \sum_{i=1}^{n} x_i$

- Si x et y sont deux variables aléatoires, alors : $\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$

- Si x et y sont deux variables aléatoires et a, b sont deux constantes, $\sum_{i=1}^{n} (a x_i + b y_i) = a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} y_i$

- La moyenne arithmétique de n valeurs de x $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n} (x_1 + x_2 + \cdots + x_n)$

Notons aussi que $\sum_{i=1}^{n} (x_i - \bar{x}) = 0$ en effet, $\sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x}$ or $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

On peut donc écrire que :

$n\bar{x} = \sum_{i=1}^{n} x_i$ ainsi on a :

$$\sum (x_i - \bar{x}) = n\bar{x} - \sum_{i=1}^{n} \bar{x} = n\bar{x} - n\bar{x} = 0$$

1.9 Moyenne d'une variable aléatoire

La moyenne ou espérance mathématique d'une variable aléatoire x est la moyenne arithmétique de cette variable aléatoire. Elle est notée E(X) et lue espérance mathématique de X.

Si X est une variable aléatoire discrète qui prend les valeurs $x_1, x_2, \cdots, x_n$ avec des densités de probabilité dont les valeurs sont : $f(x_1), f(x_2), \cdots, f(x_n)$, l'espérance mathématique de X est donnée par :

$$E(X) = x_1 f(x_1) + x_2 f(x_2) + \cdots + x_n f(x_n)$$

$$E(X) = \sum_{i=1}^{n} x_i f(x_i)$$

Quelques propriétés de l'espérance mathématique

- Si c est une constante, alors :

E(c)= c

- Si c est une constante, et X une variable aléatoire alors :

$$E(cX) = cE(X)$$

- Si a et c sont deux constantes, et X une variable aléatoire alors :

$$E(a + cX) = a + E(cX) = a + cE(X)$$

- Si X et Y sont deux variables aléatoires alors :

$$E(X + Y) = E(X) + E(Y)$$

$$E(X - Y) = E(X) - E(Y)$$

$E(XY) = E(X)E(Y)$ si X et Y sont indépendantes

1.10 Variance d'une variable aléatoire

Si X est une variable aléatoire, la variance de X notée par : $Var(X) = \sigma_X^2 \ est:$

$$Var(X) = E[X - E(X)]^2 = E\left[X^2 - 2XE(X) + [E(X)]^2\right]$$

$$Var(X) = E(X^2) - 2E(X)E(X) + [E(X)]^2$$

$$Var(X) = E(X^2) - 2[E(X)]^2 + [E(X)]^2$$

$$Var(X) = E(X^2) - [E(X)]^2$$

NB : LA racine carrée de la variance est appelée écart type.

Propriétés de la variance

- Var(a) = 0 pour toute constante a

- Si X et Y sont deux variables aléatoires alors :

$$Var(X + Y) = Var(X) + Var(Y) + 2\operatorname{cov}(XY)$$

en effet,

$$Var(X + Y) = E\left[(X + Y) - E(X + Y)\right]^2$$

$$Var(X+Y) = Var(X) + 2\operatorname{cov}(XY) + Var(Y)$$

de même :

$$Var(X-Y) = Var(X) - 2\operatorname{cov}(XY) + Var(Y)$$

Ainsi si X et Y sont deux variables aléatoires indépendantes : $C\operatorname{ov}(XY) = 0$ alors :

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

$$\mathrm{Var}(X+Y) = Var(X) + Var(Y)$$

//LES DONNEES QUALITATIVES

nous avons essayé de situer l'économétrie par rapport à la théorie économique, de définir ce qu'est l'économétrie et de définir son objectif. Nous avons ensuite présenté la méthodologie de recherche économétrique. La révision de quelques concepts statistiques à clos ce chapitre. Dans le présent chapitre, nous nous attèlerons à présenter le modèle de régression linéaire simple, les hypothèses qui sous-tendent ce modèle et l'estimation de ses

Le modèle de régression linéaire à deux variables

Le modèle de régression linéaire à deux variables est un modèle qui lie deux variables entre elles à savoir :

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \text{ ; t} = 1,2, \ldots,\text{n} \quad (2\text{-}1)$$

Où : Yt = variable dépendante ou expliquée

Xt = variable indépendante ou explicative

εt = variable aléatoire ou terme de l'erreur

β0 et β1 sont des paramètres constants inconnus que l'on se propose d'estimer à l'aide

des observations. On dispose de n observations sur Yt et Xt ; c'est-à-dire de n couples

(Yt ;Xt) qui sont les réalisations de X et Y.

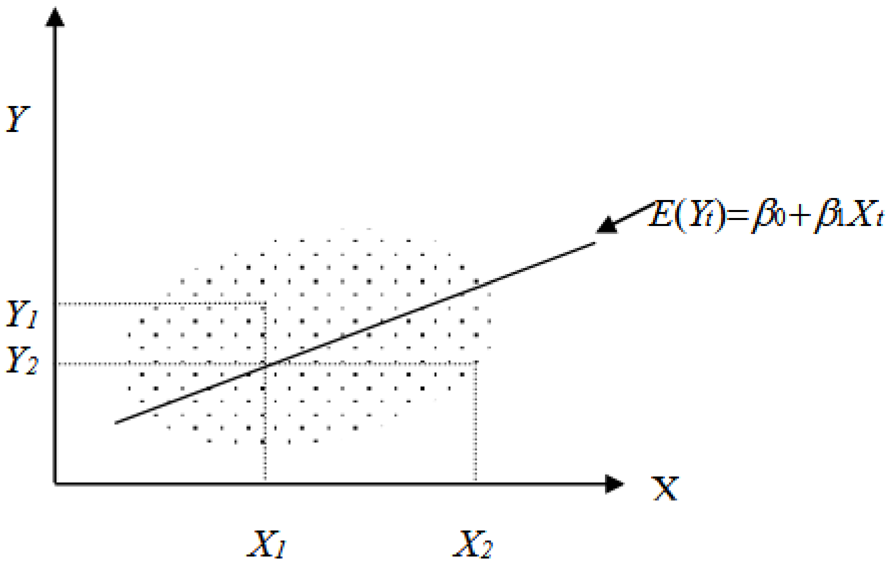
En terme graphique, l'équation (2-1) a deux composantes :

- Une composante systématique rendue par la ligne droite qui traverse le nuage

de points formé par les différentes observations c'est-à-dire0 1

ˆ( )t tE Y X

et ;- Une composante aléatoire rendue par l'écart entre la ligne droite et les points.



Comme la droiteˆtY passe par les points moyens, il va se trouver que des points seront au dessus et en dessous de cette dernière. Ainsi, les observations qui sont au dessus de la ligne donnent des valeurs positives de εt et les observations.des valeurs négatives de εt. Et comme la ligne représentée parˆtY passe par les points moyens, les valeurs négatives et positives de qui sont εt s'annuleront mutuellement de sorte que la sommation des erreurs sera nulle c'est-à-dire0i

Les hypothèses du modèle de régression Qualitative

Estimer le modèle présenté dans l'équation (2-1) implique la détermination des valeurs numériques des paramètres β0 et β1. Pour cela, il est nécessaire de connaitre Xt, Yt et εt. Les deux premières sont observables (Xt, Yt) et ne pose pas de problème. Là où il y a problème, c'est avec εt car l'erreur n'est pas observable. Ainsi, le mieux que nous pouvons faire afin de déterminer les valeurs numériques de β0 et β1 est d'émettre des hypothèses sur εt. Ces hypothèses

<u>Hypothèse 1 (H1)</u> : linéarité

Y_t est une fonction linéaire de X_t ou en n'importe qu'elle transformation de X_t. C'est-à-dire $Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$ (2-1).

<u>Hypothèse 2 (H2)</u> : Stochasticité de ε_t

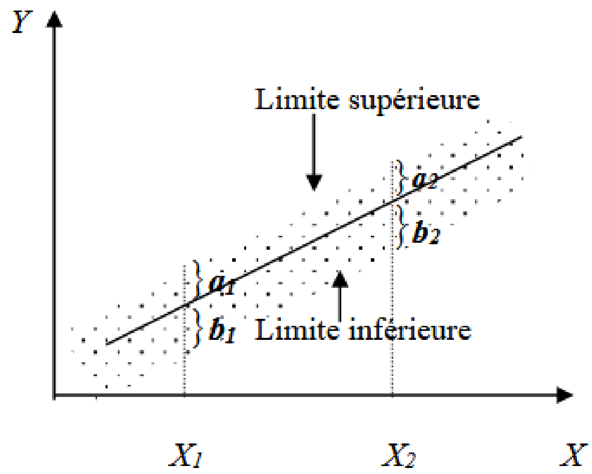
Les valeurs possibles de ε_t ne sont pas connues d'avance, elles peuvent être positives, négatives, ou nulles. Toutes fois, quelque soit la valeur de ε_t, cette dernière a une certaine probabilité de réalisation.

<u>Hypothèse 3 (H3)</u> : Nullité de la moyenne des erreurs

Quoique ε_t puisse prendre des valeurs aussi bien positives, négatives ou nulles, ces valeurs n'annulent mutuellement. En d'autres termes la moyenne des erreurs est nulle : $E(\varepsilon_t) = 0$ (2-2)

<u>Hypothèse 4 (H4)</u> : Homoscédasticité

L'homoscédasticité traduit la constance de la variance du terme de l'erreur c'est-à-dire que la dispersion de ε_t autour de sa moyenne est constante pour toutes les valeurs de X_t. Ce qui signifie que les limites supérieures et inférieures du nuage de points sont à égales distance de la ligne de régression : $Var(\varepsilon_t) = E\left[\varepsilon_t - E(\varepsilon_t)\right]^2 = E(\varepsilon_t)^2 = \sigma_\varepsilon^2$ (2-3)



<u>Hypothèse 5 (H5)</u> : Normalité

Nous faisons l'hypothèse que ε_t suit une loi normale avec comme moyenne zéro (0) et variance σ_ε^2 : $\varepsilon_t \to N(0, \sigma_\varepsilon^2)$ (2-4)
L'implication de la normalité est que la plupart des observations sont concentrées dans le voisinage immédiat de la ligne de régression.

<u>Hypothèse 6 (H6)</u> : Nullité de la covariance entre ε_t et X_t

Il n'ya pas d'association linéaire entre en ε_t et X_t par conséquent cov(ε_t ,X_t) = 0 (2-5)
En clair la variable résiduelle ε_t est indépendante des n variables explicatives X_t
Preuve :

$$Cov(\varepsilon_t, X_t) = E[(\varepsilon_t - E(\varepsilon_t))(X_t - E(X_t))]$$

$$Cov(\varepsilon_t, X_t) = E[\varepsilon_t(X_t - E(X_t)] \text{ car } E(\varepsilon_t) = 0$$

$$Cov(\varepsilon_t, X_t) = E[\varepsilon_t X_t - \varepsilon_t E(X_t)] \text{ or } X_t \text{ est non stochastique } E(X_t) = X_t$$

$$Cov(\varepsilon_t, X_t) = X_t E(\varepsilon_t) - X_t E(\varepsilon_t) = 0$$

$$Cov(\varepsilon_t, X_t) = 0$$

<u>Hypothèse 7 (H7)</u> : Absence d'autocorrélation entre les résidus

$$Cov(\varepsilon_t, \varepsilon_s) = 0 \quad \forall t \neq s \text{ (2-6) car}$$

$$Cov(\varepsilon_t, \varepsilon_s) = E[(\varepsilon_t - \underbrace{E(\varepsilon_t)}_{=0})(\varepsilon_s - \underbrace{E(\varepsilon_s)}_{=0})]$$

$$Cov(\varepsilon_t, \varepsilon_s) = E(\varepsilon_t, \varepsilon_s) = 0 \quad \forall t \neq s$$

Deux erreurs relatives à deux observations différentes (t) et (s) sont non corrélées ou indépendantes entre elles.

<u>Hypothèse 8 (H8)</u> : Rang plein

Il faut qu'il y ait au moins autant d'observations que de paramètres à estimer c'est-à-dire N>k, où k est le nombre de paramètres à estimer et N le nombre d'observations.

II-4 Estimation des paramètres du modèle de régression linéaire simple : la Méthode des Moindres Carrés Ordinaires (MCO)

Il existe plusieurs méthodes économétriques que l'on peut utiliser pour déterminer les
valeurs numériques des paramètres de l'équation (2-1). Toutes-fois nous allons focaliser notre étude sur la méthode dite des moindres carrés ordinaires pour des raisons suivantes :

- ✓ Elle est simple comparée aux autres méthodes ;
- ✓ Les paramètres obtenus à partir de cette méthode ont des propriétés optimales ;
- ✓ Elle est très utilisée et demeure de loin la plus utilisée dans l'estimation des

modèles économétriques ;Les principes qui sous tendent la méthode des MCO

sont simples à comprendre Le critère de la méthode des MCO peut se décomposer en deux parties : La première
exige que la ligne de régression passe par les points moyens du nuage de points de sorte que les déviations positives ou négatives s'annulent mutuellement c'est-à-dire . La deuxième partie du critère exige que la somme des carrés des erreurs soit minimale. Ainsi, si on écrit l'équation (2-1) on a :

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

Notons $\hat{\beta}_0$, $\hat{\beta}_1$ les paramètres estimés de β_0 et β_1 de sorte que l'on ait :

$$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 X_t \quad (2\text{-}7)$$

Posons : $\hat{\varepsilon}_t = Y_t - \hat{Y}_t = Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t$ (2-8)

Où ε_t représente les déviations résiduelles (résidus).

Le principe de la méthode des MCO revient à trouver les valeurs des paramètres qui minimisent la somme du carré des erreurs.

Elevons les résidus au carré et prenons la sommation. On a

$$\sum_{i=1}^{n} \hat{\varepsilon}_t^2 = \sum_{i=1}^{n} (Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t)^2 \quad (2\text{-}9)$$

Nous devons minimiser $\sum_{i=1} \hat{\varepsilon}_t^2$, c'est-à-dire déterminer $\hat{\beta}_0$ et $\hat{\beta}_1$ tel que $\sum_{i=1} \hat{\varepsilon}_t^2$ soit minimale.

Pour cela nous allons déterminer la dérivée de $\sum_{i=1}^{n} \hat{\varepsilon}_t^2$ par rapport à $\hat{\beta}_0$ et $\hat{\beta}_1$.

La condition nécessaire pour un minimum exige que les dérivées premières soient égales à zéro, ceci nous permet de déterminer $\hat{\beta}_0$ et $\hat{\beta}_1$.

$$\begin{cases} \dfrac{\partial \sum \hat{\varepsilon}_t^2}{\partial \hat{\beta}_0} = 0 \\ \dfrac{\partial \sum \hat{\varepsilon}_t^2}{\partial \hat{\beta}_1} = 0 \end{cases}$$

$$\frac{\partial \sum \hat{\varepsilon}_t^2}{\partial \hat{\beta}_0} = -2\sum (Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t) = 0 \quad (2\text{-}10)$$

$$\frac{\partial \sum \hat{\varepsilon}_t^2}{\partial \hat{\beta}_1} = -2\sum (Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t) X_t = 0 \quad (2\text{-}11)$$

En réarrangeant les termes des équations (2-10) et (2-11) nous obtenons les deux équations normales :

$$\sum Y_t = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_t \quad (2\text{-}12)$$

Les équations normales

$$\sum X_t Y_t = \hat{\beta}_0 \sum X_t + \hat{\beta}_1 \sum X_t^2 \quad (2\text{-}13)$$

En divisant l'équation (2-12) par « n » ensuite la multiplier par $\bar{X}$ nous obtenons :

$$\bar{X}\bar{Y} = \hat{\beta}_0 \bar{X} + \hat{\beta}_1 \bar{X}^2 \quad (2\text{-}14)$$

Divisons ensuite l'équation (2-13) par « n » pour ensuite lui soustraire l'équation (2-14) :

$$\frac{\sum X_t Y_t}{n} = \hat{\beta}_0 \frac{\sum X_t}{n} + \hat{\beta}_1 \frac{\sum X_t^2}{n} \quad (2\text{-}15)$$

$$\frac{\sum X_t Y_t}{n} - \bar{X}\bar{Y} = \hat{\beta}_1 \left(\frac{\sum X_t^2}{n} - \bar{X}^2\right)$$ et nous obtenons :

$$\boxed{\hat{\beta}_1 = \frac{\sum X_t Y_t - n\bar{X}\bar{Y}}{\sum X_t^2 - n\bar{X}^2}} \quad (2\text{-}16)$$

$\hat{\beta}_1$ peut s'écrire encore :

$$\boxed{\hat{\beta}_1 = \frac{\sum (X_t - \bar{X})(Y_t - \bar{Y})}{\sum (X_t - \bar{X})^2}} \quad (2\text{-}17)$$

Et comme $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ on a :

$$\boxed{\hat{\beta}_0 = \bar{Y} - \bar{X}\left[\frac{\sum (X_t - \bar{X})(Y_t - \bar{Y})}{\sum (X_t - \bar{X})^2}\right]} \quad (2\text{-}18)$$

Représentation des variables par les écarts : $x_t = X_t - \bar{X}$ et $y_t = Y_t - \bar{Y}$

$$\boxed{\hat{\beta}_1 = \frac{\sum x_t y_t}{\sum x_t^2}} \quad (2\text{-}19)$$

Et une fois $\hat{\beta}_1$ obtenue selon l'équation (2-19) nous pouvons avoir :

$$\boxed{\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}} \quad (2\text{-}20)$$

t égal à 0,9 ?

//LES DONNEES EN PANEL

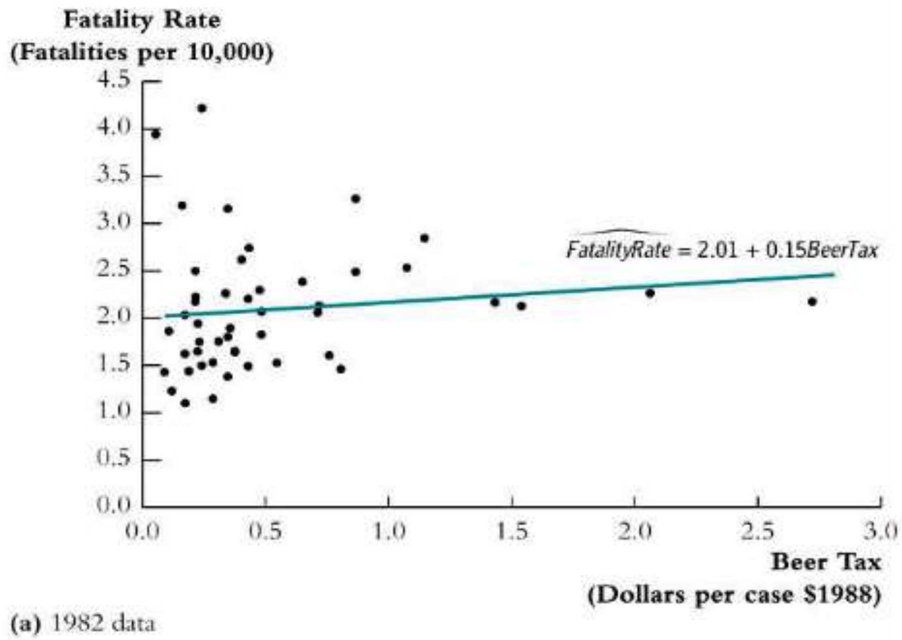
Les données en panel permettent de contrôler pour des facteurs qui varient entre les individus, mais ne varient pas au cours du temps, pourraient causer un biais d'omission si l'on en tenait pas compte, sont inobservables ou non disponibles et ne peuvent être inclus dans la régression. Si ces facteurs ne varient pas au cours du temps, ils ne peuvent pas influencer la variation de Y au cours du temps.

Exemple

Accidents mortels de la circulation et impôts sur l'alcool aux
Etats–Unis Unité d'observation : une année dans un Etat des
Etats–Unis : 48 Etats : n = 48,7 ans (1982, . . ., 1988) : T = 7, panel équilibré : 7 × 48 = 336 observations

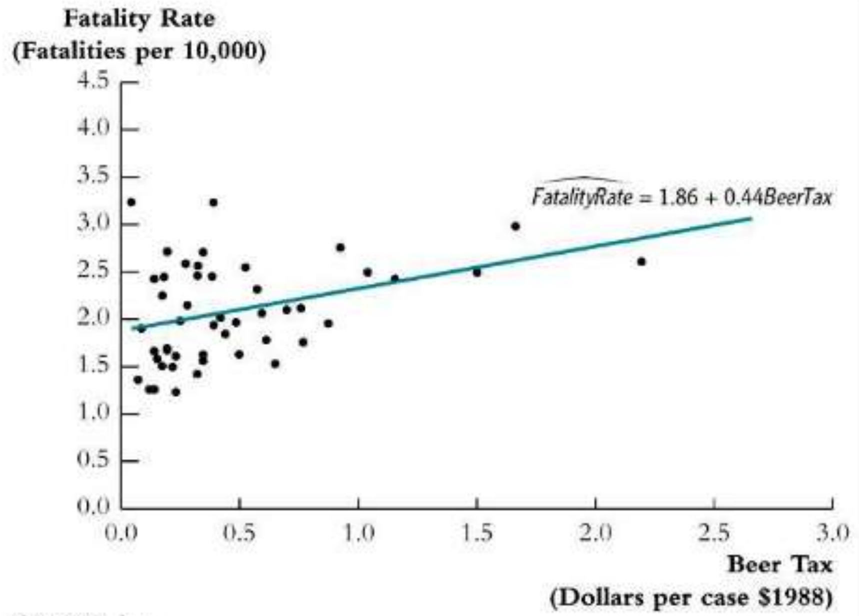
Variables: nombre de morts par accident de la circulation pour10000 habitants impôts sur un paquet de bières autres (âge minimum pour la conduite, lois contre l'alcool au volant, . . .)

DONNEES POUR 1982



Panel a is a scatterplot of traffic fatality rates and the real tax on a case of beer (in 1988 dollars) for 48 states in 1982. Panel b shows the data for 1988. Both plots show a positive relationship between the fatality rate and the real beer tax.

DONNEES POUR 1988



Panel a is a scatterplot of traffic fatality rates and the real tax on a case of beer (in 1988 dollars) for 48 states in 1982. Panel b shows the data for 1988. Both plots show a positive relationship between the fatality rate and the real beer tax.

Autres facteurs

Pourquoi y aurait–il davantage de morts par accident de la
circulation dans les Etats qui ont des impôts plus élevés sur
l'alcool? Qualité (âge) des automobiles
Qualité des routes Tolérance culturelle vis à vis de boire et conduire
Densité des voitures sur la route Ces facteurs peuvent causer un biais d'omission s'ils sont corrélé avec les impôts sur l'alcool.

DENSITE DU TRAFFIC

Une densité du traffic plus élevée est associée avec davantage d'accidents mortels sur la route Les Etats de l'Ouest, moins peuplés, ont des impôts
plus faibles sur l'alcool. Facteur peu susceptible des changer au cours du
temps (en 7 ans)

Tolérance culturelle

Une tolérance culturelle envers le fait de boire et
conduire peut être associée à des accidents mortels plus nombreux
Il peut y avoir une corrélation entre les impôts sur l'alcool et l'attitude culturelle Les attitudes culturelles ne changent en général pas
d'une année à l'autre

Panel sur deux périodes

Modèle

$F\,R_{it} = \beta_0 + \beta_1 BT_{it} + \beta_2 Z_i + u_{it}$ $i = 1, \ldots, 48; t = 1, 2$

avec $F\,R_{it}$ le taux d'accidents mortels (fatality rate), dans
l'Etat i, à la période t, et BT_{it}, les impôts sur la bière (beer
tax), dans l'Etat i, à la période t.

Z_i est un facteur qui ne varie pas au cours du temps.

Si Z n'est pas observé et que corr(BT, Z) 6 = 0, son omission
entraîne un biais de l'estimateur $\hat{\beta}_1$

Eliminer *Z*

L’équation pour 1982:

$F\ Ri82 = \beta 0 + \beta 1BTi82 + \beta 2Zi + ui82$

L’équation pour 1988

$F\ Ri88 = \beta 0 + \beta 1BTi88 + \beta 2Zi + ui88$

Supposon que $E\ (uit|BTit, Zi) = 0$ (Supposons qu’il n’y a pas d’autres facteurs importants).

Différence entre 1982 et 1988

$\Delta F\ Ri = \beta 1\Delta BTi + vi$

avec $vi = ui88 - ui82$. Bien que Z a disparu,

corr $(\Delta BT, v) = 0$.

Résultats

Données pour 1982

$\hat{F\ Ri} = 2.01$

$(0.15)+ 0.15(0.13)\ BTi\ (n = 48)$

Données pour 1988

$\hat{F\ Ri} = 1.86(0.11)+ 0.44(0.13)BTi\ (n = 48)$

Différences 1982–1988

$\hat{\Delta F\ Ri} = - 0.072(0.065)- 1.04(0.36)\Delta BTi\ (n = 48)$