

UNIVERSITE HASSAN II - Casablanca
FACULTE DES SCIENCES JURIDIQUES ECONOMIQUES ET SOCIALES
CASABLANCA

INTRODUCTION AUX METHODES ECONOMETRIQUES

(LES FONDEMENTS DE BASE DE L'ÉCONOMETRIE)

Compléments du cours
De
Mr Fouzi Mourji

Filière : Économie et Gestion

Semestre 6

Année Universitaire 2019/2020

Ensemble 2

AVANT PROPOS / AVERTISSEMENT

Ce polycopié a été établi à partir de notes prises au cours du professeur Monsieur Fouzi Mourji durant l'année universitaire 2014-2015 et actualisé depuis. Elles ont été mises en forme par deux étudiants : Mademoiselle Fatima Zahra Mouhsini et Monsieur Bouazizi Youssef.

Il est conseillé de suivre les séances de cours pour procéder aux corrections éventuelles à apporter à ces notes et pour les compléter grâce aux explications et aux exemples donnés. Ces derniers permettront de mieux saisir les concepts exposés, l'économétrie étant par nature, une discipline d'application.

Certains compléments utiles comme préalables aux estimations des modèles sont présentés ici dans les annexes (il s'agit notamment de l'usage et l'intérêt des modèles ainsi que des rappels sur les indices). Enfin des rappels de statistique descriptive et des méthodes de spécifications, sont traités pendant les séances de cours et ne sont pas repris dans ce polycopié.

Bonne chance à toutes et à tous et bon travail, pour acquérir non seulement un diplôme, mais surtout des connaissances et un savoir-faire, déterminants pour votre insertion professionnelle.

PLAN DU POLYCOPIE

Le cœur du cours d'introduction aux méthodes économétriques est constitué de deux chapitres : le 1^{er} porte sur le modèle linéaire de régression simple et le 2nd sur la régression multiple. Nous y développons les méthodes basiques d'estimation des paramètres d'un modèle et les propriétés des estimateurs et introduisons également les techniques de validation des modèles.

Mais pour bien comprendre l'intérêt de cet apprentissage, nous présentons dans **une annexe 1**, les concepts de modèles et en particulier les modèles de politique économique. Nous montrons que les modèles que nous allons apprendre à construire dans ce cours, constituent **des outils d'aide à la décision**, que ce soit au niveau microéconomique ou macroéconomique.

Dans **l'annexe 2**, nous effectuons des rappels statistiques indispensables, nous y expliquons qu'avant tout travail de modélisation, il importe de procéder à des **traitements préalables** sur les séries statistiques ; en l'occurrence il importe de passer des séries en dirhams courants à des séries en dirhams constants.

L'annexe III comporte un certain nombre d'exercices complémentaires. Il s'agit d'exemples d'application, pour bien comprendre les techniques exposées dans le cours.

Notez que **les éléments contenus dans ces trois annexes font partie intégrante du cours** ; nous y ferons référence explicitement et les développerons en temps opportun au cours des différentes séances.

INTRODUCTION :

Voir les séances d'introduction de ce cours avec les définitions de base et la démarche en économétrie et lire également l'annexe I qui traite de la typologie des modèles. Cet ensemble constitue le Chapitre Introductif.

Chapitre I - Le MODELE LINEAIRE DE REGRESSION SIMPLE

SECTION I – Présentation et hypothèses de base

SECTION II – L'Estimation des paramètres et étude de leurs propriétés (cf. le cours et l'annexe II pour les traitements préalables sur les séries statistiques)

SECTION III : Introduction à la validation des modèles

CHAPITRE II : MODELE LINEAIRE DE REGRESSION MULTIPLE

SECTION I – Présentation

SECTION II – Hypothèses dans le M.L.R.M

SECTION III – Détermination et propriétés des estimateurs

SECTION IV – Tests dans le Modèle Linéaire de Régression Multiple

CHAPITRE I.

Le MODELE LINEAIRE DE REGRESSION SIMPLE (MLRS)

INTRODUCTION

Définition :

Du point de vue étymologique, le terme « **économétrie** », comporte à la fois le terme « économie » (en tant que science) et le terme « mesure ». Elle applique les méthodes statistiques aux données empiriques issues de l'observation de faits économiques, pour expliquer et prévoir l'évolution des variables.

En reprenant Malinvaud (1970) : « l'économétrie utilise l'outil mathématique et l'induction statistique, pour vérifier des théories économiques ».

L'objet de l'économétrie est donc de confronter une explication théorique à un ensemble de données. Ces données peuvent être : i) temporelles, ii) en coupe transversale (données d'enquêtes) ou iii) de panel etc...). Voir les exemples dans le cours.

L'économétrie vise à quantifier les relations entre grandeurs économiques dont la théorie ou l'expérience affirme l'existence. Elle permet de connaître **le sens et l'intensité** des liens entre variables. (Ex: voir le cours)

L'économétrie permet de construire des modèles de prévision ou d'analyse / d'aide à la décision.

Les principales phases de la modélisation en économétrie sont au nombre de quatre, le schéma suivant les résume:



Dans le cours, nous répondons à deux questions principales au niveau de chaque phase :

- 1- En quoi elle consiste ?
- 2- Comment y procéder ?

On souligne que le sens des « flèches » n'est pas unique dans la mesure où le travail de l'économétrie est un travail itératif.

Cf. l'annexe I sur la typologie des modèles

SECTION I – Présentation et hypothèses de base

I - PRESENTATION GENERALE du MLRS:

Soit $Y_t = aX_t + b + U_t \quad t=1,2,\dots,n$

Y : variable endogène (variable à expliquer), elle est déterminée par le modèle

X : variable exogène (variable explicative)

U : variable aléatoire appelée erreur, il permet de prendre en compte le fait que la variable y est affectée par d'autres variables que la variable X.

On dispose de « n » observations sur Y et X. Nous avons donc « n » couples (Y_t, X_t) qui sont des réalisations des variables Y et X.

a et b sont des paramètres réels et inconnus que l'on se propose d'estimer à l'aide des observations y_t et x_t . **Exemples : voir les développements du cours**

Hypothèse 1 : Le modèle est correctement spécifié.

Il y a une mauvaise spécification lorsqu'on effectue un mauvais choix des variables, la forme de la liaison entre la variable endogène et la variable exogène n'est pas correctement établie ou bien la définition des variables (Niveau, Variation, Indice,...etc.) ne correspond pas au problème traité.

Autrement dit, il faut que la variable explicative retenue soit la « meilleure » sans omission d'autres variables, la vraie relation soit une relation linéaire dans ou par rapport aux paramètres à estimer et enfin la variable aléatoire intervienne de manière additive.

Remarque :

Généralement, la régression par l'introduction du Logarithme présente deux avantages ;

1. Elle permet de rendre linéaire une relation non linéaire.

Exemple 1 :

Au lieu d'écrire $C = a R + b + U$, on écrira : $C = a \text{Log } R + b + U$

Exemple 2:

Considérons la fonction de production du type Cobb Douglas :

$$Q_i = c \cdot K_i^a \cdot L_i^b$$

Avec

Q : la production, K : le facteur capital et L : le facteur travail

Une transformation logarithmique conduit à :

$$\text{Log}(Q_i) = \text{Log}(c \cdot K_i^a \cdot L_i^b)$$

$$\text{Log}(Q_i) = \text{Log}(c) + a \text{Log}(K_i) + b \text{Log}(L_i)$$

En ajoutant le terme d'erreur, on obtient :

$$\text{Log}(Q_i) = \text{Log}(c) + a \text{Log}(K_i) + b \text{Log}(L_i) + U_i$$

Ce modèle est identique à n'importe quel autre modèle linéaire sous forme $Y_i = aX_{1i} + bX_{2i} + c + U_i$

2. Elle fait aussi d'interpréter les coefficients qui affectent les variables exogènes (les paramètres de la droite de régression) comme des coefficients d'élasticité.

$$\hat{a} = \frac{\Delta \text{Log}(Q)}{\Delta \text{Log}(K)} = \frac{Q'/Q}{K'/K} = \frac{\Delta Q/Q}{\Delta K/K} = e_{Q/K}$$

$$\hat{b} = \frac{\Delta \text{Log}(Q)}{\Delta \text{Log}(L)} = \frac{Q'/Q}{L'/L} = \frac{\Delta Q/Q}{\Delta L/L} = e_{Q/L}$$

Les coefficients \hat{a} et \hat{b} sont donc des élasticités

Hypothèse 2 : Les Y_t et X_t sont des grandeurs numériques observées sans erreur.

Y est une variable aléatoire par l'introduction de U_t .

$E(U_t) = 0$ quelque soit X_t ($t=1,2,\dots,n$)

Ce qui importe est que l'espérance mathématique de U_t soit nulle ou la même pour tout i . Cette hypothèse est une hypothèse de permanence structurelle.

La prise en considération de cette hypothèse de permanence structurelle rend compte de l'élimination de l'effet des fluctuations dans le cas d'un modèle estimé sur séries chronologiques.

Dans le cadre d'un modèle estimé sur données individuelles, cette hypothèse rend compte de la représentativité de la population étudiée ainsi que de la stabilité des "comportements".

La permanence structurelle doit être prise en considération dans le temps et dans l'espace.

En d'autres termes, on considère que les comportements ne changent pas de façon significative. Cette hypothèse est évidemment acceptable dans le court terme. A moyen et long terme, l'intensité des liens (la valeur des coefficients de liaison) peut changer ; mais le sens des liaisons demeure.

D'un autre côté, dans un espace donné (un pays par exemple), on considère que la logique à la base de la spécification est valable pour expliquer les comportements des individus de cet espace.

$$E(Y_i) = a E(X_i) + b$$

Hypothèse 3 : L'homoscédasticité

U_i est distribuée selon une loi de probabilité indépendante de " i " et de X_i .

$$V(U_i) = E(U_i^2) = \sigma_u^2 \text{ Quantité finie}$$

Hypothèse 3 reprend l'hypothèse 2 mais elle est plus forte. Si H_3 n'est pas réalisée, on parle d'hétéroscédasticité.

Hypothèse 4 : Hypothèse d'indépendance des erreurs (ou résidus)

On suppose que les U_i et les U_j erreurs relatives à 2 observations différentes sont indépendantes entre elles c'est-à-dire

$$\text{Cov}(U_i, U_j) = 0 \quad \forall i \neq j$$

Plus tard, on utilisera le test de DURBIN & WASTON pour savoir s'il faut ou non refuser l'hypothèse 4.

Hypothèse 5 : Hypothèse de normalité

On suppose que les U_t sont distribuées selon une loi normale.

Hypothèse 6 : Cette Hypothèse concerne la variable exogène

Lorsque n tend vers l'infini, la suite des X_i est telle que $\frac{\sum_{t=1}^n X_t}{n} = \bar{X}$ tend vers X_0
 $\frac{\sum_{t=1}^n (X_t - \bar{X})^2}{n}$ tend vers S^2 avec X_0 et S^2 des quantités finies.

Cette hypothèse est utile pour l'étude des propriétés des estimateurs de a et b . \hat{a} et \hat{b} sont des estimateurs convergents.

Hypothèse 7 : On n'introduit aucune restriction sur les paramètres a et b à estimer.

Ils peuvent prendre n'importe quelle valeur réelle positive, négative ou nulle. La théorie suggère seulement le signe attendu pour \hat{a} .

Avant la clôture de cette partie réservée aux hypothèses classiques pour les M.R.S on met en exergue un concept important :

- **la variabilité** : « On n'étudie pas la variabilité d'un phénomène qui ne présente pas de variation ». Autrement dit, il importe que $\sum_{i=1}^n (Y_i - \bar{Y})^2$ soit différent de zéro. Pour des raisons identiques, il faut que $\sum_{i=1}^n (X_i - \bar{X})^2$ soit différente de 0.

Exemples : 1) La quantité consommée de sel n'est pas significativement différente d'un ménage à l'autre (**voir les détails et explications dans le cours**). .

2) L'investissement en période t (I_t) d'une entreprise est fonction du taux d'intérêt (i_t) du profit de la période $t-1$ (π_{t-1}) et du taux de l'impôt sur la société (IS) $I_t = f(i_t, \pi_{t-1}, IS)$

Mais en pratique (dans la réalité), le taux de l'IS ne change pas (significativement) d'une année à l'autre.

SECTION II – L'Estimation des paramètres et étude de leurs propriétés

I Détermination \hat{a} et \hat{b} par les MOINDRES CARRES ORDINAIRES : (M.C.O)

Les estimateurs des coefficients \hat{a} et \hat{b} sont obtenus en minimisant la distance au carré entre chaque observation et la droite estimée ($\sum_{t=1}^n (Y_t - \hat{Y}_t)^2$), d'où le nom de l'estimateur des moindres carrés ordinaires.

Soit $F(\hat{a}, \hat{b}) = \sum_{t=1}^n (Y_t - \hat{a}X_t - \hat{b})^2$ lorsqu'on remplace \hat{Y}_t par sa formule.

Puisqu'il s'agit de la **minimisation**, on va appliquer les règles mathématiques de l'existence d'un minimum à savoir : les conditions nécessaires et les conditions suffisantes.

Rappel mathématique :

Un fonction $g(x)$ définie sur un intervalle $[e, f]$, admet un minimum si :

$g'(x)=0$ les conditions nécessaires

$g''(x)>0$ les conditions suffisantes

On applique les conditions d'existence d'un minimum à la fonction $F(\hat{a}, \hat{b})$ telle que :
les inconnus de la fonction sont \hat{a} et \hat{b}

$$\frac{\partial F}{\partial \hat{a}} = \sum_{t=1}^n [2(Y_t - \hat{a}X_t - \hat{b})'_a \cdot (Y_t - \hat{a}X_t - \hat{b})] = 0$$

$$\frac{\partial F}{\partial \hat{b}} = \sum_{t=1}^n [2(Y_t - \hat{a}X_t - \hat{b})'_b \cdot (Y_t - \hat{a}X_t - \hat{b})] = 0$$

$$\frac{\partial F}{\partial \hat{a}} = \sum_{t=1}^n [2(-X_t) \cdot (Y_t - \hat{a}X_t - \hat{b})] = 0$$

$$\frac{\partial F}{\partial \hat{b}} = \sum_{t=1}^n [2(-1) \cdot (Y_t - \hat{a}X_t - \hat{b})] = 0$$

$$\frac{\partial F}{\partial \hat{a}} = -2 \sum_{t=1}^n [(X_t) \cdot (Y_t - \hat{a}X_t - \hat{b})] = 0$$

$$\frac{\partial F}{\partial \hat{b}} = \sum_{t=1}^n (Y_t - \hat{a}X_t - \hat{b}) = 0$$

$$\frac{\partial F}{\partial \hat{a}} = \sum_{t=1}^n [(X_t) \cdot (Y_t - \hat{a}X_t - \hat{b})] = 0 \quad (\text{la 1}^{\text{ère}} \text{ équation : eq1})$$

$$\frac{\partial F}{\partial \hat{b}} = \sum_{t=1}^n (Y_t - \hat{a}X_t - \hat{b}) = 0 \quad (\text{la 2}^{\text{ème}} \text{ équation : eq2})$$

Ces équations simultanées (eq1 et eq2) sont appelées **les équations normales**.
 Pour trouver \hat{a} et \hat{b} , il ne reste qu'à résoudre le système.

L'équation 2 donne :

$$\sum_{t=1}^n (Y_t - \hat{a}X_t - \hat{b}) = 0$$

$$\sum_{t=1}^n Y_t - \sum_{t=1}^n \hat{a}X_t - \sum_{t=1}^n \hat{b} = 0$$

$$\sum_{t=1}^n Y_t - \hat{a} \sum_{t=1}^n X_t - n\hat{b} = 0$$

$$n\hat{b} = \sum_{t=1}^n Y_t - \hat{a} \sum_{t=1}^n X_t$$

En divisant le tout par n on aura :

$$\hat{b} = \frac{\sum_{t=1}^n Y_t}{n} - \hat{a} \frac{\sum_{t=1}^n X_t}{n}$$

Finalement le \hat{b} du modèle estimé est :

$$\hat{b} = \bar{Y} - \hat{a}\bar{X}$$

Avec :

\bar{Y} : la moyenne arithmétique de l'échantillon pour Y_t

\bar{X} : la moyenne arithmétique de l'échantillon pour X_t

n : la taille de l'échantillon

Rappel statistique :

La moyenne arithmétique est un indicateur de valeur centrale que l'on calcule en divisant la somme des valeurs d'une série statistique par leur nombre.

En utilisant l'expression de \hat{b} dans l'équation 1, on obtient \hat{a} après quelques manipulations algébriques :

$$\sum_{t=1}^n (X_t)(Y_t - \hat{a}X_t - \hat{b}) = 0$$

$$\sum_{t=1}^n (X_t Y_t - \hat{a}X_t^2 - \hat{b}X_t) = 0$$

$$\sum_{t=1}^n X_t Y_t - \hat{a} \sum_{t=1}^n X_t^2 - \hat{b} \sum_{t=1}^n X_t = 0$$

$$\sum_{t=1}^n X_t Y_t = \hat{a} \sum_{t=1}^n X_t^2 - n\hat{b}\bar{X}$$

$$\sum_{t=1}^n X_t Y_t = \hat{a} \sum_{t=1}^n X_t^2 - n(\bar{Y} - \hat{a}\bar{X})\bar{X}$$

$$\sum_{t=1}^n X_t Y_t = \hat{a} \sum_{t=1}^n X_t^2 - n\bar{Y} + \hat{a}n\bar{X}^2$$

$$\text{car } \sum_{t=1}^n X_t = n\bar{X}$$

$$\sum_{t=1}^n X_t Y_t = \hat{a} \sum_{t=1}^n X_t^2 - n\bar{X}\bar{Y} - \hat{a}n\bar{X}^2$$

$$\sum_{t=1}^n X_t Y_t = \hat{a} \sum_{t=1}^n (X_t^2 - n\bar{X}^2) - n\bar{X}\bar{Y}$$

$$\hat{a} \sum_{t=1}^n (X_t^2 - n\bar{X}^2) = \sum_{t=1}^n X_t Y_t - n\bar{X}\bar{Y}$$

$$\hat{a} = \frac{\sum_{t=1}^n X_t Y_t - n\bar{X}\bar{Y}}{\sum_{t=1}^n (X_t^2 - n\bar{X}^2)} = \frac{\text{Cov}(X, Y)}{V(x)}$$

$$\hat{a} = \frac{\sum_{t=1}^n (X_t - \bar{X})(Y_t - \bar{Y})/n}{\sum_{t=1}^n (X_t - \bar{X})^2/n}$$

$$\boxed{\hat{a} = \frac{\sum_{t=1}^n (X_t - \bar{X})(Y_t - \bar{Y})}{\sum_{t=1}^n (X_t - \bar{X})^2}}$$

Soit \hat{a} la pente du modèle estimé.

Nous considérons les conditions du 2^{ème} ordre comme vérifiées car la fonction est convexe.

\hat{a} n'est rien d'autre que la dérivée de \hat{Y}_t par rapport à X_t . En effet :

$$\frac{\partial Y_t}{\partial X_t} = (\hat{a}X_t + \hat{b})'_{X_t} = \hat{a}$$

Ce coefficient exprime l'impact de la variation d'une unité de X_t sur la variation de Y_t , autrement dit, lorsque X_t varie d'une unité, la variable Y_t évolue de \hat{a} points.

Remarque :

Désormais, pour simplifier les expressions, on note $(X_t - \bar{X})$ par x_t et $(Y_t - \bar{Y})$ par y_t .

D'après ces notations :

$$\hat{a} = \frac{\sum_{t=1}^n x_t y_t}{\sum_{t=1}^n x_t^2}$$

II Démonstration de la propriété de la linéarité de \hat{a}

On sait que \hat{a} est telle que :

$$\hat{a} = \frac{\sum_{t=1}^n x_t y_t}{\sum_{t=1}^n x_t^2}$$

On remplace Y_t par son expression, \hat{a} devient :

$$\hat{a} = \frac{\sum_{t=1}^n x_t y_t}{\sum_{t=1}^n x_t^2} = \frac{\sum_{t=1}^n x_t (Y_t - \bar{Y})}{\sum_{t=1}^n x_t^2} = \frac{\sum_{t=1}^n x_t Y_t - \bar{Y} \sum_{t=1}^n x_t}{\sum_{t=1}^n x_t^2}$$

$$\hat{a} = \frac{\sum_{t=1}^n x_t Y_t}{\sum_{t=1}^n x_t^2} \quad \text{car } \bar{Y} \sum_{t=1}^n x_t = 0$$

$$\text{D'où } \frac{\sum_{t=1}^n x_t Y_t}{\sum_{t=1}^n x_t^2} = \sum_{t=1}^n k_t Y_t$$

en posant $k_t = \frac{x_t}{\sum_{t=1}^n x_t^2}$ qui exprime le poids des MCO

$$\text{Finalement } \hat{a} = \sum_{t=1}^n k_t Y_t$$

On peut même montrer que \hat{b} est aussi un estimateur linéaire.

Notons les propriétés des pondérations k_t :

1) Puisque les X_t sont supposés être non stochastiques, les k_t le sont aussi.

$$2) \sum_{t=1}^n k_t = 0$$

$$\sum_{t=1}^n \left(\frac{x_t}{\sum_{t=1}^n x_t^2} \right) = \frac{1}{\sum_{t=1}^n x_t^2} \cdot \sum_{t=1}^n x_t = 0$$

$$\text{car } \sum_{t=1}^n X_t = \sum_{t=1}^n (X_t - \bar{X}) = \sum_{t=1}^n X_t - n\bar{X} = 0$$

$$\text{avec } \sum_{t=1}^n X_t = n\bar{X}$$

$$3) \sum_{t=1}^n k_t^2 = \frac{1}{\sum_{t=1}^n x_t^2} \left(\frac{1}{(\sum_{t=1}^n x_t^2)^2} \cdot \sum_{t=1}^n x_t^2 \right) \cdot \left(\frac{1}{(\sum_{t=1}^n x_t^2)^2} \cdot \sum_{t=1}^n x_t^2 \right)$$

$$4) \sum_{t=1}^n k_t x_t = \sum_{t=1}^n k_t X_t = 1$$

III- Étude des propriétés de \hat{a} et \hat{b} :

1°) Calcul de $E(\hat{a})$

On sait que

$$\hat{a} = \frac{\sum_{t=1}^n (X_t - \bar{X})(Y_t - \bar{Y})}{\sum_{t=1}^n (X_t - \bar{X})^2}$$

$$\text{Avec } Y_t = aX_t + b + U_t$$

$$\text{Et } \bar{Y} = a\bar{X} + b + \bar{U}$$

$$Y_t - \bar{Y} = aX_t + b + U_t - a\bar{X} - b - \bar{U}$$

$$Y_t - \bar{Y} = a(X_t - \bar{X}) + (U_t - \bar{U})$$

$$\hat{a} = \frac{\sum_{t=1}^n [(X_t - \bar{X})(a(X_t - \bar{X}) + (U_t - \bar{U}))]}{\sum_{t=1}^n (X_t - \bar{X})^2}$$

$$\hat{a} = \frac{\sum_{t=1}^n [a(X_t - \bar{X})^2 + (U_t - \bar{U})(X_t - \bar{X})]}{\sum_{t=1}^n (X_t - \bar{X})^2}$$

$$\hat{a} = a \frac{\sum_{t=1}^n (X_t - \bar{X})^2}{\sum_{t=1}^n (X_t - \bar{X})^2} + \frac{\sum_{t=1}^n U_t (X_t - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2} + \frac{\bar{U} \sum_{t=1}^n (X_t - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2}$$

$$\hat{a} = a + \sum_{t=1}^n U_t k_t + \bar{U} \sum_{t=1}^n k_t$$

$$\hat{a} = a + \sum_{t=1}^n U_t k_t$$

$$(\text{car évidemment } \bar{U} \sum_{t=1}^n k_t = 0)$$

$$\text{D'où } E(\hat{a}) = E(a) + \sum_{t=1}^n k_t E(U_t)$$

$$\text{or } E(U_t) = 0 \text{ selon l'hypothèse H2}$$

$$\text{Comme } E(a) = a$$

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

\hat{a} est donc un estimateur sans biais de a

2°) Calcul de $E(\hat{b})$

\hat{b} par définition est : $\hat{b} = \bar{Y} - \hat{a}\bar{X}$

$$\hat{b} = a\bar{X} + b + \bar{U} - \hat{a}\bar{X}$$

$$\hat{b} = (a - \hat{a})\bar{X} + b + \bar{U}$$

$$E(\hat{b}) = E(a - \hat{a})\bar{X} + b + E(\bar{U})$$

$$E(\hat{b}) = [E(a) - E(\hat{a})]\bar{X} + b + \frac{1}{n}E\left(\sum_{t=1}^n U_t\right)$$

or \hat{a} étant un estimateur sans biais, $E(a) - E(\hat{a}) = 0$

Par ailleurs $E(U_t) = 0 \quad \forall t$ selon l'hypothèse H2

On conclut que

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

\hat{b} est donc un estimateur sans biais de b

3°) Moment du 2^{ème} ordre de \hat{a} et \hat{b}

a-Variance de \hat{a}

Par définition de la variance, on peut écrire :

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

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

On sait déjà que :

$$\hat{a} = a + \frac{\sum_{t=1}^n (X_t - \bar{X})U_t}{\sum_{t=1}^n (X_t - \bar{X})^2}$$

$$\hat{a} = a + \sum_{t=1}^n k_t U_t$$

$$\hat{a} - a = \sum_{t=1}^n k_t U_t$$

$$\hat{a} - a = \sum_{t=1}^n k_t U_t$$

$$(\hat{a} - a)^2 = \left(\sum_{t=1}^n k_t U_t\right)^2$$

$$E(\hat{a} - a)^2 = E\left(\sum_{t=1}^n k_t U_t\right)^2$$

$= E(k_1^2 U_1^2 + k_2^2 U_2^2 + \dots + k_n^2 U_n^2 + 2k_1 k_2 U_1 U_2 + 2k_2 k_3 U_2 U_3 + \dots + 2k_{n-1} k_n U_{n-1} U_n)$
 L'hypothèse H3 stipule que $E(U_t^2) = \sigma_u^2 \forall t$ Quantité finie

et que $E(U_i, U_j) = 0$ selon l'hypothèse H4 $\forall i \neq j$.

Il s'ensuit que

$$V(\hat{a}) = \sigma_u^2 \sum_{t=1}^n k_t^2$$

$$V(\hat{a}) = \frac{\sigma_u^2}{\sum_{t=1}^n (X_t - \bar{X})^2}$$

b-Variance de \hat{b}

$$\hat{b} = (a - \hat{a})\bar{X} + b + \bar{U}$$

$$\hat{b} - b = (a - \hat{a})\bar{X} + \bar{U}$$

$$(\hat{b} - b)^2 = ((a - \hat{a})\bar{X} + \bar{U})^2$$

$$(\hat{b} - b)^2 = (\hat{a} - a)^2 \bar{X}^2 + 2(a - \hat{a})\bar{X}\bar{U} + \bar{U}^2$$

$$E(\hat{b} - b)^2 = E[(\hat{a} - a)^2 \bar{X}^2] + 2E[(a - \hat{a})\bar{X}\bar{U}] + E(\bar{U}^2)$$

On rappelle que :

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

et

$$E(\bar{U}) = \frac{1}{n} E\left(\sum_{t=1}^n U_t\right) = 0$$

$$\text{Donc } E[(a - \hat{a})\bar{X}\bar{U}] = 0$$

$$E(\bar{U}^2) = \frac{1}{n^2} E\left(\sum_{t=1}^n U_t^2\right) = \frac{1}{n^2} n \sigma_u^2$$

$$V(\hat{b}) = \frac{\sigma^2}{\sum_{t=1}^n (X_t - \bar{X})^2} \bar{X}^2 + \frac{1}{n} \sigma_u^2$$

$$V(\hat{b}) = \sigma_u^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_{t=1}^n (X_t - \bar{X})^2} \right]$$

4°) Calcul de la covariance

$$\text{cov}(\hat{a}, \hat{b}) = E[(\hat{a} - a)(\hat{b} - b)]$$

or

$$\hat{b} - b = (a - \hat{a})\bar{X} + \bar{U}$$

$$\hat{b} - b = -(\hat{a} - a)\bar{X} + \bar{U}$$

$$\text{donc } (\hat{a} - a)(\hat{b} - b) = -(\hat{a} - a)^2 \bar{X} + (\hat{a} - a) \bar{U}$$

$$E[(\hat{a} - a)(\hat{b} - b)] = -E[(\hat{a} - a)^2] \bar{X} + E[(\hat{a} - a) \bar{U}]$$

on sait déjà que $E[(\hat{a} - a) \bar{U}] = 0$ (voir démonstration de $V(\hat{b})$)

$$\text{d'où } \text{cov}(\hat{a}, \hat{b}) = -E[(\hat{a} - a)^2] \bar{X} = -V(\hat{a}) \bar{X}$$

$$\text{donc } \text{cov}(\hat{a}, \hat{b}) = -\frac{\sigma_u^2}{\sum_{t=1}^n (X_t - \bar{X})^2} \bar{X}$$

Remarques :

- On voit grâce à la formule de la variance de \hat{a} , que cette dernière dépend de deux éléments : σ_u^2 et $\sum_{t=1}^n (X_t - \bar{X})^2$.

$\lim_{n \rightarrow \infty} \sum_{t=1}^n (X_t - \bar{X})^2 = \infty$ d'où $\lim_{n \rightarrow \infty} V(\hat{a}) = 0$ puisque σ_u^2 est constant. En effet, cette variance est d'autant plus faible que $\sum_{t=1}^n (X_t - \bar{X})^2$ sera élevé et cela n'est possible que si le nombre d'observations (n) est important (hypothèse H6) et que les données sur X sont très dispersées (il y a une variabilité).
-Idem pour la formule de la variance de \hat{b} .

-Si \bar{X} est positive (ce qui est le cas de la plupart des séries économiques mesurées en niveau), alors \hat{a} et \hat{b} sont corrélés négativement.

- Si l'hypothèse d'indépendance des résidus n'est pas satisfaite (H 4 non satisfaite), la formule de $V(\hat{a})$ sous-estime celle-ci (donc surestime le T de Student $\frac{\hat{a}}{\sigma_{\hat{a}}}$ pouvant nous amener à conclure hâtivement que la variable explicative X_i est satisfaisante).

« Le lecteur constate qu'il est faux de dire, comme on le fait souvent, que l'estimateur par les moindres carrés suppose une distribution normale des aléas U_t . En fait, la normalité des aléas est nécessaire seulement pour la justification de certains tests associés à la méthode des moindres carrés et non pour des formules des estimations.

IV- Détermination d'un estimateur sans biais de σ_u^2 :

$V(U_t) = \sigma^2$ est en général inconnue. Il est nécessaire de l'estimer pour avoir les variances estimées de \hat{a} et \hat{b} .

Soit \hat{U} le résidu de l'estimation tel que :

$$\hat{U}_t = Y_t - \hat{a}X_t - \hat{b}$$

$$= (a - \hat{a})X_t + (b - \hat{b}) + U_t$$

et on sait que : $E(U_t) = 0$

\hat{U}_t diffère de U_t pour les termes $(a - \hat{a})X_t$ et $(b - \hat{b})$ qui dépendent des résidus de l'estimation.

Comme \hat{a} converge en probabilité vers a et \hat{b} converge en probabilité vers b alors la distribution des \hat{U}_t converge en probabilité vers celle des U_t et la valeur $S^2 = \frac{\sum_{t=1}^n \hat{U}_t^2}{n-2}$ estimateur de la variance des résidus.

n : le nombre d'observations

$n-2$: le nombre de degré de liberté

2 : le nombre des paramètres (a et b).

SECTION III : Introduction à la validation des modèles

I – Le coefficient de détermination

On appelle variation de Y (et non variance) par rapport à sa moyenne \bar{Y} la quantité : $\sum_{t=1}^n (Y_t - \bar{Y})^2$.
On parlera de la variance lorsqu'on la divise par n.

$$\begin{aligned} Y_t - \bar{Y} &= Y_t - \bar{Y} + \hat{Y}_t - \hat{Y}_t \\ Y_t - \bar{Y} &= (Y_t - \hat{Y}_t) + (\hat{Y}_t - \bar{Y}) \\ \sum_{t=1}^n (Y_t - \bar{Y})^2 &= \sum_{t=1}^n [(Y_t - \hat{Y}_t) + (\hat{Y}_t - \bar{Y})]^2 \end{aligned}$$

si bien que

$$\sum_{t=1}^n (Y_t - \bar{Y})^2 = \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 + \sum_{t=1}^n (\hat{Y}_t - \bar{Y})^2 + 2 \sum_{t=1}^n (Y_t - \hat{Y}_t) \sum_{t=1}^n (\hat{Y}_t - \bar{Y})$$

$$\text{or } \sum_{t=1}^n (Y_t - \hat{Y}_t) \sum_{t=1}^n (\hat{Y}_t - \bar{Y}) = \sum_{t=1}^n \hat{U}_t \sum_{t=1}^n (\hat{Y}_t - \bar{Y}) = \sum_{t=1}^n \hat{U}_t (\hat{Y}_t - \bar{Y})$$

$$= \sum_{t=1}^n \hat{U}_t \hat{Y}_t - \bar{Y} \sum_{t=1}^n \hat{U}_t = 0$$

$$\text{car } \sum_{t=1}^n \hat{U}_t = 0$$

donc :

$\sum_{t=1}^n (Y_t - \bar{Y})^2 = \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 + \sum_{t=1}^n (\hat{Y}_t - \bar{Y})^2$ <div style="display: flex; justify-content: space-around; width: 100%;">SCTSCRSCE</div>

L'équation ci-dessus s'appelle **équation d'analyse de la variance**, ou il apparait que la variabilité totale (SCT) est égale à la variabilité des résidus (SCR) + la variabilité expliquée (SCE).

$$\text{On dit que le modèle reproduit bien la réalité si : } \sum_{t=1}^n (Y_t - \bar{Y})^2 = \sum_{t=1}^n (\hat{Y}_t - \bar{Y})^2$$

Attention :

On ne peut pas dire qu'une variable (X) explique un modèle, mais la variation de X explique la variation de Y.

Par exemple : la variation du taux d'intérêt d'une période à une autre explique la variation de l'investissement.

Il semble alors naturel de mesurer le pourcentage de la variation de Y expliquée par la régression.

L'équation d'analyse de la variance s'écrit en divisant à gauche et à droite par la variabilité totale :

$$\begin{aligned} \frac{\sum_{t=1}^n (Y_t - \bar{Y})^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2} &= \frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2} + \frac{\sum_{t=1}^n (\hat{Y}_t - \bar{Y})^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \\ 1 &= \frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2} + \frac{\sum_{t=1}^n (\hat{Y}_t - \bar{Y})^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \end{aligned}$$

L'équation décompose la variation de Y en deux parts :

la part expliquée par la régression : $\frac{\sum_{t=1}^n (\hat{Y}_t - \bar{Y})^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2}$

la part résiduelle : $\frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2}$

On définit maintenant R^2 comme :

$$R^2 = \frac{\sum_{t=1}^n (\hat{Y}_t - \bar{Y})^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2} = \frac{SCE}{SCT} = 1 - \frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2} = 1 - \frac{SCR}{SCT}$$

Le terme R^2 ainsi défini porte le nom de coefficient de détermination.

Il faut noter que le R^2 n'est pas une quantité négative.

Exemple : si l'on obtient $R^2 = 0.72$ cela voudrait dire que le modèle permet d'expliquer 72% de la variabilité de Y.

Dans le cas de la régression simple, on montre aisément que le R^2 est égale au carré du coefficient de corrélation r_{xy} entre X et Y. Son signe est défini par le coefficient \hat{a} .

$$\begin{aligned} R^2 &= \frac{\sum_{t=1}^n (\hat{Y}_t - \bar{Y})^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2} = \frac{\sum_{t=1}^n [(\hat{a}X_t + \hat{b}) - (\hat{a}\bar{X} + \hat{b})]^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \\ &= \frac{\sum_{t=1}^n (\hat{a}X_t + \hat{b} - \hat{a}\bar{X} - \hat{b})^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2} = \frac{\sum_{t=1}^n [\hat{a}(X_t - \bar{X})]^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2} = \frac{\sum_{t=1}^n [\hat{a}(X_t - \bar{X})]^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2} = \hat{a}^2 \frac{\sum_{t=1}^n (X_t - \bar{X})^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \\ &= \hat{a}^2 \frac{\sum_{t=1}^n (X_t - \bar{X})^2 / n}{\sum_{t=1}^n (Y_t - \bar{Y})^2 / n} = \hat{a}^2 \frac{V(X)}{V(Y)} \end{aligned}$$

$$\text{or } \hat{a} = \frac{\text{cov}(X,Y)}{V(X)} = r_{xy} \frac{\sigma_y}{\sigma_x}$$

$$\text{donc } \hat{a}^2 = r_{xy}^2 \frac{V(Y)}{V(X)}$$

De fait, tester la significativité de \hat{a} revient à tester la significativité de la corrélation entre X et Y.

$$R^2 = r_{xy}^2 \frac{V(Y)}{V(X)} \frac{V(X)}{V(Y)}$$

On relève ainsi que :

$$R^2 = r_{xy}^2$$

et

$$r_{xy} = \text{signe}(\hat{a}) \cdot \sqrt{R^2}$$

Attention :

Bien que r_{xy} soit une mesure de la relation linéaire entre deux variables, il n'implique pas nécessairement une liaison de cause à effet (**causalité**).

(voir les détails et explications dans le cours).

II Brève présentation des tests dans le modèle linéaire de régression simple:

1°) Construction du test :

On centre et on réduit la variable gaussienne \hat{a} :

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

$$\frac{\hat{a} - a}{\sqrt{\frac{\sigma_u^2}{\sum_{t=1}^n (X_t - \bar{X})^2}}} \sim N(0,1)$$

Au lieu d'utiliser la distribution normale, on peut recourir à la distribution t de Student à n-2 degrés de liberté.

$$T_c = \frac{\frac{\hat{a} - a}{\sqrt{\frac{\sigma_u^2}{\sum_{t=1}^n (X_t - \bar{X})^2}}}}{\sqrt{\frac{(n-2)S^2}{\sigma_u^2}}}$$

Par définition d'une loi de Student qui est le rapport d'une loi normale centrée réduite

à la racine carrée d'un Khi2 divisé par son degré de liberté.

D'où

$$T_c = \frac{\hat{a} - a}{\sqrt{\frac{S^2}{\sum_{t=1}^n (X_t - \bar{X})^2}}} \text{ avec } S^2 = \frac{\sum_{t=1}^n u_t^2}{n-2} \text{ estimateur de la variance des résidus.}$$

De même pour \hat{b} :

$$T_c = \frac{\hat{b} - b}{\sqrt{S^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_{t=1}^n (X_t - \bar{X})^2} \right]}}$$

Remarque :

Comme on peut le constater, les hypothèses classiques sont nécessaires à la construction des tests et à la possibilité de répondre aux posées ci-dessous.

2°) Énoncé du problème :

Deux types de problème :

1^{er} type : $Y_i = aX_i + b + U_i \quad i=1,2,\dots,n$

Faut-il refuser l'hypothèse $a=k$?

Pour $k=0$, situation la plus fréquente : elle permet de tester la pertinence de la variable, 2 réponses possibles :

*il faut refuser $a=0$

*il ne faut pas refuser $a=0$

$$\text{Donc } T_c = \frac{\hat{a}-a}{\sqrt{\frac{s^2}{\sum_{t=1}^n (X_t-\bar{X})^2}}} = \frac{\hat{a}-0}{\sqrt{\frac{s^2}{\sum_{t=1}^n (X_t-\bar{X})^2}}} = \frac{\hat{a}}{\sqrt{\frac{s^2}{\sum_{t=1}^n (X_t-\bar{X})^2}}}$$
$$T_c = \frac{\hat{a}}{\widehat{\sigma_{\hat{a}}}}$$

2^{ème} type : 1^{er} modèle $Y_{1i} = a_1 X_{1i} + b_1 + U_{1i} \quad i=1,2,\dots,n_1$
2^{ème} modèle $Y_{2i} = a_2 X_{2i} + b_2 + U_{2i} \quad i=1,2,\dots,n_2$

*Faut-il refuser $a_1 = a_2$?

* Faut-il refuser $b_1 = b_2$?

* Faut-il refuser simultanément $a_1 = a_2$ et $b_1 = b_2$?

Remarque :

On cherche toujours à tester $a = k$ et pas $\hat{a}=k$, cependant on utilise \hat{a} pour la commodité des calculs du fait que “a” est inconnu.

L'Intérêt de tester $a=0$ est pour savoir la pertinence de la variable X et sa capacité à expliquer Y. Cependant, on peut être amené à tester “a” par rapport à une quantité donnée différente de zéro.

Exemple : soit le modèle suivant :

$$C_i = a_{1i}D_i + a_{2i}I_i + a_{3i}IG_i + b + U_i$$

C : Cours de l'action d'une entreprise ;

D : Dividende ;

I : L'investissement de cette entreprise ;

IG : L'indice général de la bourse.

Ici, on ne cherchera pas forcément à tester a_1, a_2 et a_3 par rapport à zéro mais par rapport à une autre quantité donnée (valeur obtenue pour une entreprise d'un autre secteur).

3°) Intervalle de confiance et solution du test :

L'intervalle de confiance (IC) pour \hat{a} s'écrit :

Sous $\{H_0 : a=0\}$

$$P[-t_\alpha < T_c < +t_\alpha] = 0.95 \text{ au seuil } \alpha = 0.05$$

$$\text{soit } P\left[-t_\alpha < \frac{\hat{a}}{\hat{\sigma}_{\hat{a}}} < +t_\alpha\right] = 0.95$$

$$\text{finalement } P[-t_\alpha \cdot \hat{\sigma}_{\hat{a}} < \hat{a} < +t_\alpha \cdot \hat{\sigma}_{\hat{a}}] = 0.95$$

Solution du test :

Deux cas possibles :

* \hat{a} appartient à l'intervalle $\bullet \longrightarrow$ on ne rejette pas l'hypothèse H_0 .

* \hat{a} n'appartient pas à cet intervalle $\bullet \longrightarrow$ on rejette H_0 .

Remarque et exemple :

Soit le modèle suivant, estimé sur 26 ans et expliquant la consommation Y, par le revenu X.

$$Y_t = 0.4X_t - 1.2 \quad \hat{a} = 0.4 \text{ (propension marginale à consommer) et } \hat{b} = -1.2 \text{ (consommation autonome)}$$

(0.06) (0.4)

(.) écarts-types estimés des estimateurs

Sous $\{H_0 : a=0\}$

$$P\left[-t_\alpha < \frac{\hat{a}}{\hat{\sigma}_{\hat{a}}} < +t_\alpha\right] = 0.95 ; \text{ s'écrit dans notre cas :}$$

$$P\left[-t_\alpha < \frac{0.4}{0.06} < +t_\alpha\right] = 0.95 \text{ (dans la table pour } n=26, t_\alpha=2.064)$$

$$\text{donc } P[-2.064 \times 0.06 < 0.4 < +2.064 \times 0.06] = 0.95$$

$$\text{ou encore } P[-0.12384 < 0.4 < +0.12384] = 0.95 \text{ sous } H_0.$$

« 0.4 » n'appartient pas à l'intervalle, on rejette l'hypothèse H_0 qui accordait à cet événement, au risque $\alpha = 0.05$. Donc on rejette l'hypothèse de nullité de a.

On peut donc dire que le revenu X explique bien la consommation Y

Souvent on effectue une approximation : elle consiste à rapporter l'estimateur à son écart-type estimé. Si ce rapport est supérieur à 2, on rejette H_0 . S'il est inférieur à 2, on ne rejette pas H_0 .

$$\text{Sous } H_0 : T_c = \frac{\hat{a}}{\hat{\sigma}_{\hat{a}}} = \frac{0.4}{0.06} = 6.67 > 2$$

CHAPITRE II

MODELE LINEAIRE DE REGRESSION MULTIPLE (M.L.R.M)

Section I- PRESENTATION

Les modèles linéaires sont fiables à court et moyen terme car le domaine de variation des variables est dans ce cas relativement restreint. La fiabilité de ces modèles (modèles linéaires) sera plus ou moins réduite dans le temps et ce en fonction de la rapidité avec laquelle les variables prises en compte subissent des transformations.

Généralement en économie, un phénomène est rarement explicable par un seul facteur, il est nécessaire d'introduire plusieurs variables explicatives susceptibles de l'expliquer.

Les modèles linéaires de régression multiple sont du type :

$$Y_t = a_1 X_{1t} + a_2 X_{2t} + \dots + a_k X_{kt} + a_0 + U_t \quad t=1,2,\dots,n \quad (1)$$

Cette équation comporte :

- Une variable endogène ou expliquée Y
- K variables exogènes ou explicatives : X_1, X_2, \dots, X_k
- Une variable aléatoire U appelée l'erreur ou le bruit.

A partir des coefficients estimés $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_k$ et en utilisant les différents tests, on peut apprécier parmi les variables du vecteur X celles qui sont les plus significatives.

Afin d'alléger l'écriture, le modèle peut être également présenté sous une forme matricielle. Cette écriture rend plus aisée sa manipulation et facilite aussi l'expression de certains résultats :

$$\begin{matrix} Y \\ (N, 1) \end{matrix} = \begin{matrix} X \\ (N, K + 1) \end{matrix} \begin{matrix} a \\ (K + 1, 1) \end{matrix} + \begin{matrix} U \\ (N, 1) \end{matrix}$$

avec :

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_t \\ \vdots \\ Y_n \end{bmatrix}; X = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{k1} \\ 1 & X_{12} & X_{22} & \dots & X_{k2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{1t} & X_{2t} & \dots & X_{kt} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{1n} & X_{2n} & \dots & X_{kn} \end{bmatrix}; a = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix}; U = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_t \\ \vdots \\ U_n \end{bmatrix}$$

La première colonne de la matrice X, composée de 1, qui correspond au coefficient a_0 .

La dimension de la matrice X est de n lignes et p colonnes ($p=k+1$; k étant le nombre de variables exogènes réelles, c'est-à-dire constante exclue).

Remarque:

Des hypothèses sont nécessaires pour justifier l'utilisation au MCO

Section II- Hypothèses dans le modèle linéaire de régression multiple :

Hypothèse 1 : Cf section 1 paragraphe II.

Hypothèse 2 : Les Y_{it} et les X_{it} sont des grandeurs numériques observées sans erreur.
 $E(U_t) = 0$ quelque soient X_{it} et pour $i=1,2,\dots,p$

Hypothèse 3 : Hypothèse d'homoscédasticité
 U_t est distribuée selon une loi indépendante de t et des X_{it} , pour $t=1,\dots,n$ et $i = 1,\dots,p$
 $V(U_t) = E(U_t^2) = \sigma_u^2$ est une quantité finie.

Hypothèse 4 : Indépendance des erreurs $Cov(U_t, U'_t) = 0$

Hypothèse 5 : La loi de distribution de l'aléa est une loi gaussienne de moyenne nulle et l'écart-type fini.

Hypothèse 6 : Hypothèse sur les variables exogènes:
Absence de colinéarité des variables X_1, X_2, \dots, X_k et E (vecteur unité).

Soit le modèle suivant : $I_t = f(i_t, \text{PIB}, \dot{C})$ où I_t est l'investissement public, i_t est le taux d'intérêt, PIB la variation du PIB et enfin \dot{C} la variation de la consommation.

Ici, on a une forte colinéarité entre PIB et \dot{C} dans la mesure où $\text{PIB} + C + I + X - M$.

On souligne l'importance de la statistique descriptive où l'étude de la covariance des variables deux à deux est un outil pertinent pour soulever ce problème.

$X(n,p)$ est de rang p (avec $n > p$) cela veut dire que $(X'X)^{-1}$ existe.
Avec n : ligne, p : colonne.

Cette hypothèse est utile pour l'étude des propriétés de \hat{a} .
Quand n tend vers l'infini, $(X'X)^{-1}$ reste non singulière.

Hypothèse 7 : On n'introduit pas de restriction sur les estimateurs. Ils peuvent être positifs, négatifs ou nuls.

Récapitulons :

* $X_{pt} = 1$ pour tout t permet d'introduire une constante dans le M.L.R.M. L'exclusion d'une constante dans un MLRM donne souvent un R^2 incorrect.

*Les p vecteurs $X_1 \dots X_{p-1}$ et E ne sont pas colinéaires.

Ces vecteurs forment une variété linéaire non dégénérée à p dimensions dans R^n qu'on appellera plan de régression.

La matrice $X(n,p) = (X_{1t}, X_{2t}, \dots, X_{p-1t}, E)$ pour t variant de 1 à n est de rang p .

*Écriture matricielle :

s'écrit pour chaque t et pour $t = 1 \dots n$

$$Y_t = X_t a + U_t \quad (2)$$

Enfin, on aura :

$$Y = Xa + U$$

$$\hat{Y} = X\hat{a} \text{ et } \hat{U} = Y - \hat{Y} \quad (3)$$

$$\text{D'où } \hat{U} = Y - X\hat{a} \quad (4) \text{ avec } \hat{Y} = X\hat{a}$$

Section III – Détermination et propriétés des estimateurs

1°) Calcul de \hat{a} par les M.C.O :

La méthode consiste à chercher les paramètres \hat{a}_i tels que :

$$\sum_{t=1}^n \hat{U}_t^2$$

Revenons à (4), le vecteur \hat{U} des écarts est orthogonal au plan de régression. Les équations normales s'écrivent alors :

$$X_1' \hat{U} = X_2' \hat{U} = \dots = X_j' \hat{U} = \dots = X_{p-1}' \hat{U} = E' \hat{U} = 0$$

Elles expriment l'orthogonalité de \hat{U} et des p vecteurs lignes $X_1', X_2', \dots, X_{p-1}'$ et E' . Ces équations peuvent s'écrire :

$$X' \hat{U} = 0 \rightarrow \text{le vecteur de la matrice } X'.$$

X' = transposé de la matrice X .

$$X' \hat{U} = 0 \text{ (le vecteur de la matrice } X').$$

X' = transposé de la matrice X .

En tenant compte de (4), on aura :

$$\begin{aligned} X' (Y - X\hat{a}) &= 0 \\ (X'Y) - (X'X\hat{a}) &= 0 \\ X'Y &= X'X\hat{a} \\ (X'X)^{-1}X'Y &= (X'X)^{-1}X'X\hat{a} \end{aligned}$$

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

2°) Calcul de $E(\hat{a})$:

Nous allons démontrer que \hat{a} est l'estimateur efficace dans l'ensemble des estimateurs linéaires et sans biais de a . \hat{a} sera dit BLUE (Best linear unbiased estimators).

Calcul de $E(\hat{a})$:

$$E(\hat{a}) = E[(X'X)^{-1}X'Y] = (X'X)^{-1}X'E(Y)$$

or

$$Y = Xa + U$$

donc

$$E(Y) = E(Xa + U) = Xa + E(U) = Xa \text{ car } E(U) = 0$$

En remplaçant dans $E(\hat{a})$, il vient :

$$E(\hat{a}) = (X'X)^{-1}X'Xa$$

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

\hat{a} est donc un estimateur sans biais de a

3°) Calcul de $V(\hat{a})$:

$$V(\hat{a}) = E[(\hat{a} - E(\hat{a}))(\hat{a} - E(\hat{a}))']$$

C'est donc une matrice carrée d'ordre p , en tenant de $E(\hat{a}) = a$, elle s'écrit :

$$V(\hat{a}) = E[(\hat{a} - a)(\hat{a} - a)']$$

$$\begin{aligned} \text{or } \hat{a} &= (X'X)^{-1}X'Y \\ \hat{a} &= (X'X)^{-1}X'(Xa + U) \end{aligned}$$

$$\hat{a} = (X'X)^{-1}X'Xa + (X'X)^{-1}X'U$$

soit :

$$\hat{a} = a + (X'X)^{-1}X'U$$

donc

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

et

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

donc

$$\begin{aligned} V(\hat{a}) &= E[(X'X)^{-1}X'UU'X(X'X)^{-1}] \\ V(\hat{a}) &= (X'X)^{-1}X'E(UU')X(X'X)^{-1} \text{ avec } E(UU') = \sigma_u^2 I_n = \Omega_U \end{aligned}$$

ce qui permet d'écrire :

$$\begin{aligned} V(\hat{a}) &= (X'X)^{-1}X'\sigma_u^2 I_n X(X'X)^{-1} \\ V(\hat{a}) &= \sigma_u^2 (X'X)^{-1}X'I_n X(X'X)^{-1} \end{aligned}$$

$$V(\hat{a}) = \sigma_u^2 (X'X)^{-1}$$

On montre qu'une condition nécessaire est suffisante pour que \hat{a} soit un estimateur convergent de a est que les vecteurs des variables exogènes ne tendent pas à être colinéaires quand n tend vers l'infini. Autrement dit H_6 reste valable quand n tend vers l'infini.

4°) Détermination de S^2 estimateur de σ_u^2 :

La variance de l'erreur est inconnue, il importe donc de trouver un estimateur S^2 sans biais pour σ_u^2 .

$$\hat{U} \text{ s'écrit : } \hat{U} = Y - X\hat{a}$$

$$\hat{U} = Xa + U - X\hat{a}$$

on sait déjà que :

$$\begin{aligned} \hat{a} &= (X'X)^{-1}X'Y \\ &= (X'X)^{-1}[X'(Xa + U)] \end{aligned}$$

on remplace dans \hat{U} :

$$\begin{aligned}\hat{U} &= Xa + U - X[(X'X)^{-1}X'(Xa + U)] \\ \hat{U} &= Xa + U - X(X'X)^{-1}X'Xa - X(X'X)^{-1}X'U \\ \hat{U} &= Xa + U - Xa - X(X'X)^{-1}X'U \\ \hat{U} &= U - X(X'X)^{-1}X'U \\ \hat{U} &= [I_n - X(X'X)^{-1}X']U\end{aligned}$$

Les résidus apparaissent ainsi comme une forme linéaire du bruit inconnu U , soit :

$$\hat{U} = MU$$

On remarque que la matrice M est symétrique et idempotente : $M'M = M$

En effet :

$$[I_n - X(X'X)^{-1}X']U \times [[I_n - X(X'X)^{-1}X']U]' = [I_n - X(X'X)^{-1}X']U$$

La somme des carrés des résidus est :

$$\begin{aligned}SCR &= \hat{U}'\hat{U} \\ &= U'M'MU\end{aligned}$$

ou encore : $\hat{U}'\hat{U} = U'MU$

$$E(\hat{U}'\hat{U}) = E[U'[I_n - X(X'X)^{-1}X']U]$$

Trace de M :

$$\begin{aligned}\text{tr}M &= \text{tr}[I_n - X(X'X)^{-1}X'] \\ &= \text{tr}I_n - \text{tr}[X(X'X)^{-1}X'] \\ &= n - \text{tr}[X(X'X)^{-1}X']\end{aligned}$$

or $n - \text{tr}I_p = p$

donc $\text{tr}M = n - p$

En utilisant le fait que la trace d'un scalaire est un scalaire, il vient :

$$E(\hat{U}'\hat{U}) = E[\text{tr}(U'MU)] = \sigma_u^2 (n - p)$$

D'où l'estimateur de la variance de l'erreur s'écrit donc :

$$S^2 = \frac{\hat{U}'\hat{U}}{n - p} = \frac{\sum_{t=1}^n \hat{U}_t^2}{n - p}$$

S^2 est un estimateur sans biais de σ_u^2

p est le nombre de paramètres à estimer

n est le nombre d'observations

Section IV – Tests dans le modèle linéaire de régression multiple

Pour apprécier ou valider un modèle, on dispose d'indicateurs fournis par l'économétrie. Ils permettent de juger la qualité du modèle dans son ensemble, des paramètres obtenus par la régression et de chacune des variables explicatives retenues. Ils permettent également de tester certaines hypothèses économiques.

I – Le coefficient de détermination

1°) Formulation :

Soit un modèle simple M1 de la forme : $Y_t = b + U_t$

Soit un modèle M2 plus sophistiqué de la forme : $Y_t = a_1 X_{1t} + a_2 X_{2t} + \dots + a_k X_{kt} + a_0 + U_t$

Pour M1 : $\bar{Y} = b$ M1 rend compte de la réalité avec une incertitude mesurée par $(Y_t - \bar{Y})$ ou $(Y_t - b)$. Si l'on considère l'ensemble des observations on aura $\sum_{t=1}^n (Y_t - \bar{Y})^2$ cela revient à considérer la variation totale du phénomène Y.

Pour M2 : on a pour une observation donnée

$$Y_t = X_t a + U_t$$

$$\hat{Y}_t = X_t \hat{a}$$

L'imprécision ou l'incertitude au niveau de M2 sera mesurée par

$$\sum_{t=1}^n (Y_t - X_t \hat{a})^2$$

Le coefficient de détermination R^2 d'un modèle représente le carré du coefficient de corrélation multiple R ; il mesure la qualité de l'ajustement des Y_t par la méthode des M.C.O. appliquée sur M2. Ici, il mesure le gain relatif de précision quand on utilise M2 au lieu de M1.

$$R^2 = \frac{\sum_{t=1}^n (Y_t - \bar{Y})^2 - \sum_{t=1}^n (Y_t - X_t \hat{a})^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2} = 1 - \frac{\sum_{t=1}^n (Y_t - X_t \hat{a})^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2} = 1 - \frac{\hat{U}'\hat{U}}{Y'Y - N\bar{Y}^2}$$

2°) Interprétation :

Au numérateur, $\sum_{t=1}^n (Y_t - \bar{Y})^2 - \sum_{t=1}^n (Y_t - X_t \hat{a})^2$ mesure la somme des carrés expliqués par le modèle. Plus elle sera élevée, plus elle se rapprochera de la variance totale.

Remarque :

R^2 proche de 1 n'est pas une preuve de causalité totale. De même R^2 proche de 0 ne signifie pas absence de lien entre la variable expliquée et les variables explicatives, mais peut être la forme de liaison n'est pas adaptée.

Bref, il faut être prudent durant l'interprétation de tout résultat.

II – Tests d'hypothèses dans le modèle linéaire de régression multiple

1°) Test de Fisher :

Test de l'hypothèse : $\{H_0: a_1 = a_2 = \dots = a_k = 0\}$

a- Construction du test :

Dire que $a_1 = a_2 = \dots = a_k = 0$, c'est dire que toutes les variables retenues ne sont pas explicatives.

$\hat{Y}'\hat{Y}$ suit la loi de Khi2 k et $\hat{U}'\hat{U}$ suit la loi de Khi2 à n-p. Ces deux Khi2 sont indépendants en probabilité. Or, on sait que le rapport de deux Khi2 est un F de Fisher. Donc pour établir le test de $a_1 = a_2 = \dots = a_k = 0$, on construit la variable F_c (Fisher calculé).

$$F_c = \frac{\frac{\hat{Y}'\hat{Y}}{k}}{\frac{\hat{U}'\hat{U}}{n-p}} = \frac{\frac{R^2}{k}}{\frac{1-R^2}{n-p}} = \frac{R^2}{1-R^2} \times \frac{n-p}{k} \text{ tend vers un } F_{k;n-p} \text{ ddl}$$

b- Solution du test :

On compare le F théorique càd le F lu sur la table de distribution de Fisher et le F calculé à partir de nos observations

2 cas peuvent se présenter :

- * F calculé > F théorique ; on rejette H_0 . Cela veut dire que les variables retenues sont explicatives.
- * F calculé < F théorique ; non rejet de H_0 .

Exemple : Soit un modèle à 5 variables exogènes (y compris la constante) et on suppose qu'il a été estimé sur 25 observations. Supposons enfin que le résultat obtenu est :

$$F_c = \frac{\frac{\hat{Y}'\hat{Y}}{k}}{\frac{\hat{U}'\hat{U}}{n-p}} = \frac{\hat{Y}'\hat{Y}}{5} \times \frac{19}{\hat{U}'\hat{U}} = 3.0$$

La valeur lue sur la table du F de Fisher pour 5 et 19 ddl

Et 2,74 au seuil de 5% et 4,17 au seuil de 1%

Nous allons rejeter H_0 au seuil de 5% (car $3 > 2,74$) mais on ne pourra pas la rejeter au seuil de 1% (car $3 < 4,17$).

Remarque :

Ce qui importe, c'est de tester la portée explicative l'influence des « vraies variables exogènes ». Ce qu'il faut tester c'est la nullité du vecteur formé par les (p-1) premières composantes de a. On a toujours :

$\hat{U}'\hat{U}$ —————→ Khi2 à (n-p) ddl

$\hat{Y}'\hat{Y}$ —————→ Khi2 à p-1 ddl

Le rapport des 2 khi deux sera donc un $F_{p-1;n-p}$

Exemple :

Au Maroc, les importations sont fonction de la consommation, de la FBCF et des exportations.

$$M = f(C, FBCF, X)$$

$$\log(M_t) = 0.469 \log(C_t) + 0.034 \log(FBCF_t) + 0.471 \log(X_t) - 0.993$$

n=14

$R^2=82.44\%$

$F_c=15.649$

n-p=10

k=3 donc $F_{3;10}=6.55$ au seuil $\alpha=1\%$

Donc on rejette l'hypothèse H_0 : le modèle comporte au moins une variable explicative de significative.

c- Généralisation :

Soit 2 modèles M1 et M2 tels que :

M1 : $Y = \check{X}\check{a} + \check{U}$ avec $Y_{(n,1)}, \check{X}_{(n,q)}, \check{a}_{(q,1)}$ et $\check{U}_{(q,1)}$

M2 : $Y = Xa + U$ avec $Y_{(n,1)}, X_{(n,p)}, a_{(p,1)}$ et $U_{(p,1)}$

On pose $n > q > p$

Faut-il refuser H_0 qui considère M2 comme meilleur ; cela veut dire : dans M1 les $(q-p)$ variables ne sont-elles pas en trop ?

Pour répondre à cette question, on définit un F qui suit une loi de probabilité donnée ;

$$F_c = \frac{\frac{\hat{U}'\hat{U} - \check{\hat{U}}'\check{\hat{U}}}{q-p}}{\frac{\check{\hat{U}}'\check{\hat{U}}}{n-p}} = \frac{\hat{U}'\hat{U} - \check{\hat{U}}'\check{\hat{U}}}{q-p} \times \frac{n-p}{\check{\hat{U}}'\check{\hat{U}}} = \frac{\hat{U}'\hat{U} - \check{\hat{U}}'\check{\hat{U}}}{\check{\hat{U}}'\check{\hat{U}}} \times \frac{n-p}{q-p}$$

Ce rapport suit une loi de Fisher à $(q-p)$ et $(n-p)$ ddl.

Ensuite, on compare le F calculé (F_c) et le F théorique (F_t). Si $F_c > F_t$, on rejette H_0 .

2°) Test de de Student : Significativité de chaque élément du vecteur a :

Test de l'hypothèse : $\{H_0: a_k = 0 \text{ la } k \text{ ème variable n'explique pas significativement le modèle}\}$

a- Construction du test :

Ce test permet en fait d'apprécier la portée explicative de chaque variable exogène retenue.

Soit le modèle : $Y = Xa + U$

Supposons que l'on s'intéresse à un coefficient déterminé a_k du modèle. Nous avons démontré que $E(\hat{a}) = a$ ou encore $E(\hat{a}_k) = a_k$ et nous avons démontré aussi que $V(\hat{a}) = \sigma_u^2 (X'X)^{-1}$.

Soit V_k le K ème terme de la diagonale principale de $(X'X)^{-1}$, la variance d'un coefficient quelconque de régression s'écrit :

$$V(\hat{a}_k) = \sigma_u^2 V_k \text{ et } \hat{a}_k \sim N(a_k, \sigma_u \sqrt{V_k})$$

$$\text{Donc la quantité } \frac{\hat{a}_k - a_k}{\sigma_u \sqrt{V_k}} \sim N(0,1)$$

Mais on ne connaît pas la variance de U. Il convient de l'estimer, nous savons que :

$$S^2 = \frac{\sum_{t=1}^n \hat{U}_t^2}{n-p} \text{ est un estimateur sans biais de la variance de U}$$

Comme $\frac{(n-p)S^2}{\sigma_u^2} \sim \text{Khi2}(n-p)$ et que \hat{a}_k et U sont indépendants il en découle que \hat{a}_k et S^2 sont indépendants et la statistique :

$$T_c = \frac{\frac{\hat{a}_k - a_k}{\sigma_u \sqrt{V_k}}}{\sqrt{\frac{(n-p)S^2}{\frac{\sigma_u^2}{(n-p)}}}} \sim \text{Student}(n-p)$$

ou encore : $T_c = \frac{\hat{a}_k}{\sqrt{S^2 \cdot V_k}} = \frac{\hat{a}_k}{\sigma_{\hat{a}_k}}$ sous H_0

b- Solution du test :

Sous H_0 :

On écrit que $P[-t_\alpha < T_c < +t_\alpha] = 0.95$

$$P\left[-t_\alpha < \frac{\hat{a}_k}{\sigma_{\hat{a}_k}} < +t_\alpha\right] = 0.95$$

ou encore : $P[-t_\alpha \times \sigma_{\hat{a}_k} < \hat{a}_k < +t_\alpha \times \sigma_{\hat{a}_k}] = 0.95$

2 cas possibles :

* \hat{a}_k appartient à l'intervalle $P[-t_\alpha \times \sigma_{\hat{a}_k}; +t_\alpha \times \sigma_{\hat{a}_k}] = 0.95$ alors on ne refuse pas H_0

* \hat{a}_k n'appartient pas à cet intervalle alors on refuse H_0 .

Remarque :

Nous avons utilisé le constat que la variable de Student T à (n-p) ddl converge en loi vers une loi normale centrée réduite quand (n-p) augmente. On peut sans commettre d'erreur notable utiliser l'approximation normale quand l'échantillon est grand.

T de Student ddl	95%
15	2.131
20	2.086
30	2.042
40	2.020
60	2.000

La table de la loi normale donne une valeur de 1.96 ; donc, en faisant une approximation, l'hypothèse de nullité d'un coefficient (ou son égalité à une quantité donnée) sera rejetée pour un $|T_c| > 2$.

Exemple :

$$M_t = aI_t + bC_t + c + e_t$$

pour n=22

$$M_t = \frac{0.371I_t}{(0.081)} + \frac{0.192C_t}{(0.125)} + \frac{15.80}{(9.53)}$$

comme n=22, les T suivent une loi de Student à 19 ddl (22-3)

l'écriture, Sous H_0 :

$$P\left[-t_\alpha < \frac{\hat{a}_k}{\widehat{\sigma_{\hat{a}_k}}} < +t_\alpha\right] = 0.95 ; \text{ donne ici } , P\left[-t_\alpha \times \widehat{\sigma_{\hat{a}_k}} < \hat{a}_k < +t_\alpha \times \widehat{\sigma_{\hat{a}_k}}\right] = 0.95$$

Discutons la pertinence de l'investissement :

le t lu sur la table à 95% est 2.093 et on remplace les écarts types par leur valeur, il vient :
 $P(-2.093 \times 0.081 < 0.307 < 2.093 \times 0.081) = 0.95$ ou $P(-0.1695 < 0.307 < 0.1695) = 0.95$

La valeur $\hat{a} = 0.307$ n'appartient pas à l'intervalle, donc rejet de H_0 ; en d'autres termes coefficient a_2 est différent de 0 et l'investissement est une variable pertinente pour expliquer le niveau des importations.

Pour la consommation, on utilise l'approximation :

$|T_c| = |0.192/0.125| < 2$ donc selon l'échantillon, la consommation n'explique pas le niveau des importations.

3°) Test de Durbin & Watson : Problème de l'autocorrélation des résidus :

Il y a autocorrélation des résidus quand le modèle est mal spécifié. On peut préciser que le risque d'autocorrélation des résidus augmente surtout dans les modèles estimés sur des séries chronologiques. Elle sera d'ailleurs d'autant plus grande que la périodicité est courte.

Dans le cas de l'omission d'une variable significative, les résidus ne sont plus distribués de façon aléatoire. Ils comportent l'effet des variables omises.

Lorsqu'on a autocorrélation des aléas, on a tendance à sous-estimer les écart-types des coefficients ; en d'autres termes on surestime le T de Student. Dès lors, on risque de conclure « hâtivement » à la pertinence d'une variable proposée et testée.

Donc, les causes d'un mauvais DW sont de deux types, économique et statistique.

Pour s'assurer du risque d'autocorrélation des erreurs, on procède au test de Durbin & Watson.

a- Principe d'estimation :

Il existe plusieurs procédures d'estimation. On suppose par exemple que les résidus U_t sont corrélés selon un schéma autorégressif de 1^{er} ordre. Soit donc :

$$U_t = \phi U_{t-1} + \varepsilon_t \text{ avec } \varepsilon_t \sim N(0,1) \text{ et } E(\varepsilon_t, \varepsilon_{t'}) = 0 \forall t \neq t'$$

$$\text{On aura donc } U_{t-1} = \phi U_{t-2} + \varepsilon_{t-1}$$

$$\text{D'où : } U_t = \phi^2 U_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t$$

$$\text{enfin } U_t = \sum \phi^k \varepsilon_{t-k}$$

On démontre que dans le cas d'autocorrélation des erreurs, l'estimateur \hat{a} est sans biais et il reste convergent.

Mais, cela ne signifie pas qu'il a la variance minimum. La variance étant forte, l'estimation peut s'éloigner de son espérance mathématique avec une forte probabilité.

Reprenons le modèle initial :

$$Y_t = a_1 X_{1t} + a_2 X_{2t} + \dots + a_k X_{kt} + a_0 + U_t \quad (1)$$

$$Y = Xa + b$$

Soit \hat{a} obtenu par les MCO $\hat{Y} = X\hat{a}$

Pour une période, $\widehat{U}_t = Y - \hat{Y}$. C'est résidus dépendent aussi bien de la suite des erreurs \widehat{U}_t mais également de la suite des X.

Considérons la variable aléatoire \hat{d} définie comme suit :

$$\hat{d} = \frac{\sum_{t=2}^{t=n} (\widehat{U}_t - \widehat{U}_{t-1})^2}{\sum_{t=1}^{t=n} \widehat{U}_t^2} \quad (2)$$

La statistique d de Durbin-Watson varie entre 0 et 4. Ceci peut être confirmé de la manière qui suit. En développant (2), on a :

$$\hat{d} = \frac{\sum \widehat{U}_t^2 + \sum \widehat{U}_{t-1}^2 - 2 \sum \widehat{U}_t \sum \widehat{U}_{t-1}}{\sum \widehat{U}_t^2}$$

Puisque $\sum \widehat{U}_t^2$ et $\sum \widehat{U}_{t-1}^2$ ne diffèrent que d'une observation, ils sont approximativement égaux. Donc, on posant $\sum \widehat{U}_t^2 \approx \sum \widehat{U}_{t-1}^2$, (2) peut s'écrire :

$$\hat{d} \approx 2 \left(1 - \frac{\sum \widehat{U}_t \sum \widehat{U}_{t-1}}{\sum \widehat{U}_t^2} \right)$$

Cette statistique de DW est utilisée pour donner la nature de la corrélation des résidus.

A partir de l'hypothèse $U_t = \phi U_{t-1} + \varepsilon_t$, l'équation $\hat{d} = 2(1 - \hat{\phi})$ permet de statuer sur la valeur de \hat{d} :

Cas 1 : $\hat{\phi} = 0$, cela veut dire qu'il n'y a pas de liaison entre U_t et U_{t-1} ; dans ce cas $\hat{d} = 2$ (ou tend vers 2, si $\hat{\phi}$ est proche de zéro)

Cas 2 : $\hat{\phi} = 1$; cela veut dire qu'il y a une liaison positive entre U_t et U_{t-1} (autocorrélation positive) et $\hat{d} = 0$ [$2(1-1)$].

Cas 3 : $\hat{\phi} = -1$; cela veut dire qu'il y a une liaison négative entre U_t et U_{t-1} (autocorrélation négative) et $\hat{d} = 4$ [$2(1 - (-1)) = 2(1+1)$].

Tableau de décision :

$\hat{d}=0$	d_1	d_2	$4-d_2$	$4-d_2$	4
Autocorrélation positive	Doute	Indépendance	Doute	Autocorrélation négative	

b- Utilisation de la table de Durbin & Watson :

Nous allons tester l'hypothèse $H_0: \phi = 0$ contre l'hypothèse $H_1: \phi \neq 0$ à un seuil α .

Durbin & Watson ont déterminé 2 valeurs d_1 et d_2 fonction de n et de p.

Donc on calcule \hat{d} à partir de sa formule et donner résultats de l'estimation ; ensuite, nous serons face à plusieurs éventualités :

* $\hat{d} < d_1$ on rejette H_0

* $d_1 < \hat{d} < d_2$ doute

$\hat{d} > d_2$ non rejet H_0 , absence d'autocorrélation des résidus.

n	k=2		k=3		k=4	
	d_1	d_2	d_1	d_2	d_1	d_2
15	1.08	1.36	0.95	1.54	0.82	1.75
20	1.20	1.41	1.10	1.54	1.00	1.68
30	1.35	1.49	1.28	1.57	1.21	1.65
50	1.50	1.59	1.46	1.63	1.42	1.67
100	1.65	1.69	1.63	1.72	1.61	1.74

Remarques :

* La statistique de DW appliquée à un modèle contenant des variables endogènes retardées est biaisée vers 2 ce qui laisserait supposer que les erreurs sont moins souvent corrélées dans un processus autorégressif que dans un processus ordinaire.

* Dans le cas d'autocorrélation des aléas, il y a lieu d'opérer sur le modèle initial la transformation de COCHRANE ORCUTT.

Exemple :

$$\text{Log } M_t = 0.469 \text{ Log } C_t + 0.034 \text{ Log } \text{FBCF}_t + 0.471 \text{ Log } X_t - 0.993$$

$$\widehat{DW} = 1.414 \quad n = 15 \text{ et } k = 3 \quad d_1 = 0.82 \text{ et } d_2 = 1.75$$

$0.82 < \widehat{DW} < 1.75$ il y a doute

Bibliographie

➤ Ouvrages

- « **Méthodes statistiques de l'économétrie** » Edmond MALINVAUD - Dunod / 1970
- « **Éléments de mathématiques et statistiques pour économistes** » Tome1 et 2, F Mourji et N Mikou – Wallada
- « **Économétrie** » R Bourbonnais – Dunod- 2011
- « **Introduction à l'économétrie** » C. Labrousse – Dunod / 1976
- « **Méthodes économétriques pour l'analyse économique et l'aide à la décision** » Kamal HADADJ – Walada / 1994
- « **Introduction à l'économétrie** » Dormont B – Montchrestien (1999)
- « **Métodes Econométriques** » Tome 1 et 2, J.Johnston, Economica, 1985
- « **Econométrie** », Jean-François Brun, Jean Louis Combes et Claudio Araujo, Bréal, 2004
- « **The theory and practice of econometrics** » édition Wiley 1985
- Exercices Pedagogiques de statistiques et économetrie** » Claude Mouchot, Economice 1979
- « **Limited dependent and qualitative variables in econometrics** » G.S.Madala , Cambrigde university Press, 1977.
- « **A Guide to Econometrics** » Peter Kenned , Third Edition , 1992
- « **Econométrie appliquée** », sous la direction de Claude Montmarquette, Economica, 1997

➤ Revues à consulter

- **Problèmes économiques**
- **Revue d'Économie du développement**
- **Revue d'économie politique**
- **Economie et prévision**
- **American Economic Review (AER)**
- **Journal of Political Economy**
- **World Bank Economic Review**

➤ Journaux

Il est conseillé de lire

- **Le Monde (du Mardi)**
- **Le Figaro (pages saumon, économiques)**

Voir aussi les sites, comme celui du NBER (National Bureau of Economic Research) ou encore celui de la Direction de la Statistique... ce sont des exemples

ANNEXE 1

Typologie des modèles : les modèles de politique économique

Cette annexe complète l'introduction faite en cours, elle a pour objectif de montrer l'intérêt des travaux de modélisation. Les étudiants(es) comprendront ainsi mieux la portée du cours d'introduction aux méthodes économétriques et de l'apprentissage de la construction des modèles.

La section I de ce chapitre présente les caractéristiques d'un modèle. Elle en définit les constituants (variables, équations, paramètres...).

La section II traite de l'**usage** des modèles de politique économique.

SECTION I - PRESENTATION DES MODELES :

a) Les variables :

On trouve ici les objets de la décision en matière économique : les éléments de la stratégie au niveau microéconomique et de la politique économique, au niveau macroéconomique.

X_i : Quantité de l'**instrument** de stratégie ou de politique économique, numéro (i) utilisée par le décideur.

$i = 1, 2, \dots, m$

Exemple :

MACRO

$i=1$: Dépense publique globale.

$i=2$: Taux des impôts indirects.

$i=3$: Taux de réserves obligatoires.

$i=4$: Taux de salaire dans la fonction publique ou SMIG.

MICRO

$i=1$: Le prix pratiqué par l'entreprise.

$i=2$: Investissement (stock, outillage).

$i=3$: Dépenses publicitaires.

$i=4$: Taux de salaire directs et indirects (transferts)

$i=5$: dépenses en recherches & développement

Y_j : quantité de l'**objectif (j)** visée par le décideur.

$j = 1, 2, \dots, n$.

Exemples :

MACRO

$j=1$: Volume de l'emploi de la main d'œuvre

$j=2$: Solde de la balance commerciale.

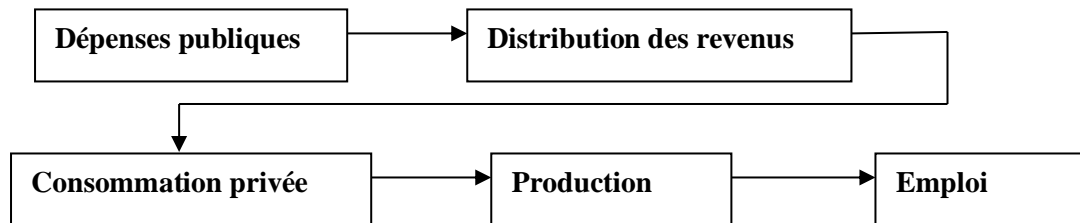
$j=3$: Taux de croissance du PIB.

MICRO

j=1 : Part de marché.
j=2 : Chiffre d'affaire.
j=3 : Productivité.

De manière générale **les instruments** n'agissent pas directement sur **les objectifs**:

Schéma de fonctionnement / Exemple macro:



Dans les modèles qui prennent en compte cet aspect, on trouvera donc des **variables intermédiaires**.
Il s'agit de variables endogènes situées entre l'instrument et l'objectif.

Soit **Z_k** : Quantité de la variable intermédiaire endogène **k**.

On distingue enfin **les variables prédéterminées**. Elles se trouvent également entre l'instrument et l'objectif, mais leur valeur est fixée en dehors du modèle.

Soit **V_k** : Quantité de la variable prédéterminée **k**.

b) Les relations entre les variables :

Dans les modèles de politique économique, on trouve d'abord les relations ordinaires (de comportement), mais on trouve aussi des relations propres à ces modèles.

❖ Les contraintes ou fonctions instrumentales :

Elles relient les objectifs aux instruments. Ceux sont des sortes de fonctions de production d'objectifs (output) à partir des instruments (input).

Exemple :

$Y_j = f_j(x_1, x_2, \dots, x_n)$: fonction de production de j.

$\mu_{j1} = \frac{dY_j}{dX_1}$: efficacité marginale de 1 vis à vis de j.

si l'on tient compte du fait que les instruments n'agissent pas directement, on notera :

$X_1 \longrightarrow Z_k \longrightarrow Y_j$

D'où $Z_k = g(X_1, \dots, X_i, \dots, X_n)$

$Y_j = f_j(Z_1, \dots, Z_k, \dots)$

Exemple :

X_1 : Dépenses publiques.

Z_k : investissement privé.

Y_j : Croissance du PIB.

F_j : Est la fonction de production reliant le PIB et l'investissement privé.

Donc on aura : $Y_j = F_j [\dots\dots\dots g_k (X_1, \dots, X_1 \dots\dots\dots)] \dots\dots\dots$

f : est donc une fonction de fonction (ou fonction composée).

❖ Les contraintes spécifiques sur les objectifs et sur les instruments :

Ce sont des « **planchers** » ou des « **plafonds** » que ne peuvent raisonnablement dépasser les variables de politique économique.

Exemple :

$X \geq X'$: Plancher de production.

$L \geq L'$: Plancher d'emploi de la main d'œuvre.

$P \leq P'$: Plafond des prix.

$G \leq a.Y$: Plafond des dépenses publiques par rapport au revenu national.

$G - F \leq D'$: Plafond du déficit budgétaire (F étant les recettes fiscales).

SECTION II - UTILISATION DES MODELES DE POLITIQUE ECONOMIQUE :

Comme Leur nom l'indique, les modèles de politique économique permettent la mise au point de celle-ci (au niveau macro, comme au niveau microéconomique), or on sait que le décideur agit sur l'activité économique avec une visée plus au moins volontariste.

Ainsi, tantôt ces modèles vont permettre la « **simple** » **prévision**, tantôt, ils seront utilisés pour **la décision**.

a) L'utilisation prévisionnelle :



M . S : Modèle de simulation.

Le schéma sur lequel est fondée cette démarche est **probabiliste**.

A partir des valeurs jugées possibles des **X (instruments)**, et des **V (variables prédéterminées)**, le modèle permet de connaître les valeurs endogènes des variables **Z (intermédiaires)** et **Y (objectifs)**.

Ici, l'optique **du « possible » ou « probable »** est privilégiée. La valeur des objectifs est fournie par le calcul ; elle n'est pas fixée à priori. Cependant on peut introduire des variantes en modifiant la valeur donnée aux **X (instruments)**.

b) L'utilisation décisionnelle :

Ici, les **objectifs sont fixés à priori**. C'est la démarche qui prévaut la planification ou dans les modèles d'optimisation (pour les entreprises).

La démarche se résume comme suit :



ici, Y (Objectifs) et V (Valeurs prédéterminées) sont exogènes.

Le MS permet le calcul des variables endogènes X (instruments) et Z (Variables intermédiaires).

La démarche est plus **volontariste, normative**. Le **souhaitable** est privilégié au **détriment du probable**.

Ce sont les niveaux requis, désirés des objectifs qui déterminent les montants des instruments.

c) La combinaison des deux méthodes s'impose :

En pratique la simple prévision est insuffisante pour parler de politique économique visant des objectifs précis.

D'un autre côté, les objectifs ne peuvent atteindre n'importe quelles valeurs, car la manipulation des instruments admet des limites. Les X ne peuvent prendre toutes les valeurs dictées par Y fixés (cf. les concepts de Plafond et plancher considérés plus haut).

La politique doit non seulement être ambitieuse et cohérente, mais aussi réaliste.

L'utilisation mixte ou alternative du MS permet d'approcher cette réalité.

❖ Première méthode :

Certains objectifs seront traités en variables exogènes, les autres le seront en variables endogènes.

Réciproquement, les instruments les plus rigides (les plus proches de leur limite plafond ou plancher) seront traités en variables exogènes. Les plus souples restent endogènes.

Ainsi, *le modèle est mi-décisionnel, mi-prévisionnel*.

❖ Deuxième méthode :

Les objectifs prioritaires et les instruments souples sont aussi, tantôt endogènes, tantôt exogènes.

Première étape :

Utiliser le modèle dans l'optique décisionnelle.

Il s'agit de fixer Y et de calculer le montant des X, nécessaire pour atteindre le niveau de Y fixé. Si certains niveaux des X semblent irréalistes (forte pression fiscale par exemple), on les réduira à des niveaux plus réalistes.

Deuxième étape :

On utilise le modèle dans une optique prévisionnelle. On recalcule les objectifs.

Par itérations (Vas et vient), on **ira du souhaitable vers le probable** en gagnant à chaque itération un peu du côté du souhaitable en s'approchant le plus près du possible.
Cette méthode est utilisée dans la pratique mais elle présente l'inconvénient d'être lourde.

d) Les modèles d'optimisation :

Il s'agit en fait de modèles décisionnels à objectifs flexibles.

Ici, les objectifs Y et les instruments X sont des variables endogènes. Leur valeur numérique, inconnue au départ, découle de la maximisation (ou minimisation) d'une fonction de préférence sous certaines contraintes.

La démarche se résume comme suit :



Seules les variables prédéterminées V sont exogènes.

A l'aide du Modèle d'Optimisation / MO, on calcule les montants des instruments, des objectifs et des variables intermédiaires.

Il faut en outre disposer d'une part de contraintes (analogues à celle des MS) et d'autre part d'une fonction de préférence ou d'utilité.

Annexe II

Au cours d'Introduction aux méthodes économétriques

**Traitements préalables des séries statistiques:
Compléments de cours**

Notes établies à partir du cours de
Fouzi Mourji

Année universitaire 2019/2020

Sommaire

I - Introduction

II- Comment déflater une série statistique ?

II- 1 L'utilisation des indices des prix

II- 1.1 Les indices simples

II- 1.2 Les indices synthétiques

II- 2 Exemples d'utilisation des indices synthétiques

II- 2.1 Exemple Un

II- 2.2 Exemple Deux

INTRODUCTION

Après la phase de spécification, et avant de procéder à l'estimation des paramètres du modèle spécifié, nous sommes souvent conduits à effectuer des traitements préalables sur les données.

La nature des traitements dépend de plusieurs facteurs; nous retiendrons en particulier la distinction entre données chronologiques et données transversales (en coupe instantanée / données d'enquêtes).

Sur les données transversales, il faudra par exemple définir les ratios en liaison avec la spécification retenue (par exemple : l'effort éducatif appréhendé par la variable «Rdepeduc » rapportant les dépenses d'éducatives à la dépense totale du ménage).

Il faudra aussi tester les risques de colinéarité ; par exemple dans un modèle d'analyse des déterminants de l'accès au crédit : garantie 1 (titre de propriété) et garantie 2 (caution d'un tiers),

Sur les séries chronologiques, il y a la **correction des variations saisonnières (CVS)**, le traitement de la **colinéarité** et la distinction entre « **données en volume** » et « **données en valeur** ». C'est le cas que nous allons considérer ici.

- Lorsque nous « **déflavons** » une série **en valeur** par un indice approprié, nous obtenons une série **en volume**.

Nous parlerons aussi de série exprimée en «dirhams constants » \neq « **dirhams courants** » au niveau des variations, nous aurons respectivement : les variations en terme **nominal** \neq des variations en terme « **réel** ».

- Ce travail consiste à exclure l'effet des prix, dans l'appréciation des variations d'une grandeur.
- En effet l'inflation fausse les conclusions que nous pouvons faire à propos de l'évolution d'une grandeur : par exemple les dépenses des ménages, le chiffre d'affaires d'une entreprise ou d'un commerçant...

Comment procéder pour **déflater** une série statistique ?

Pour observer l'évolution **réelle** d'une grandeur (chiffre d'affaires, dépenses, exportations...), il faut écarter l'effet de la variation des prix.

L'indice le plus connu concerne les prix à la consommation. Il s'agit du taux d'inflation qui mesure l'évolution des prix des biens de consommation ainsi que les prix des services.

Mais il y a beaucoup d'autres indices, comme l'indice des prix des biens d'équipement qui pourrait servir à déflater une série relative à l'investissement, l'indice des prix de gros...

Pour montrer comment utiliser les indices de prix, nous allons recourir à des exemples, en **distinguant le cas d'une grandeur simple et celui d'une grandeur composite**.

1. Travail sur **une grandeur** : Les indices simples

Dans le tableau suivant, nous avons, pour un bien A (des souliers par exemple), les quantités vendues et les prix, sur deux périodes T et T+1

	T		T+1		CA _T	CA _{T+1}
	Px	Qtés	Px	Qtés		
Bien A	10	300	12	290	3000	3480

Pour apprécier / analyser l'évolution du chiffre d'affaires, nous pouvons dire que :

Entre T et T+1, en **terme nominal**, le CA a augmenté de 16% $((3480/3000) \times 100)$; par contre les quantités vendues (volume des ventes) ont baissées de 3.3% $((290/300) - 1) \times 100)$

Nous pouvons aussi mesurer les variations du CA **en terme réel** en écartant l'effet de l'inflation ; ce faisant, nous allons alors déflater, c'est-à-dire «corriger » le **CA en dh courants** par l'indice des prix, et obtenir un **CA en dh constants**.

Prenons comme année de base T :

On rappelle que $IP_{t/0} = (P_t / P_0) \times 100$

Ainsi:

◆ Pour l'**année T** :

Le CA de l'année T va rester le même, puisque l'indice des prix est égal à 100 pour cette année $((10 / 10) \times 100 = 100)$.

◆ Pour **T+1** :

$$IP_{T+1/T} = \frac{12}{10} \times 100 = 120$$

$$CA \text{ en dh constants}_{T+1} = \frac{CA \text{ en dh courants}_{T+1}}{IP_{T+1/T}} \times 100 = \frac{3480}{120} \times 100 = 2900$$

	CA en dh Crt (T)	CA de T+1, en dhCrts	CA de T+1 en dhCst	Δ du CA en dh crt	Δ du CA en dh cst	Δ des qtés produites
Bien A	3000	3480	2900	16%	-3.33%	-3.33%

L'égalité entre la variation du **CA en dh constants** et la variation des **quantités** vendues tient au fait que nous avons éliminé l'effet des prix, entre T et T+1. Mathématiquement parlant, on peut la démontrer par la formule Suivante :

$$\frac{CA \text{ en dh constants}_{T+1}}{CA \text{ en dh constants}_T} = \frac{Q_{T+1} \times P_T}{Q_T \times P_T} \times 100 = \frac{Q_{T+1}}{Q_T} \times 100 = IQ_{T+1/T}$$

- Lorsque nous travaillons sur une seule grandeur (le CA par exemple), raisonner sur les quantités ou sur la série en DH constants revient au même.
- Quand il s'agit de plusieurs grandeurs à la fois (par exemple, le CA réalisé par la vente de **plusieurs biens**, le passage par les **indices synthétiques s'impose**. On ne peut en effet pas ajouter des Kg de tomates avec des litres d'huile ou encore des souliers avec des chemises...

Prenons l'exemple suivant : une épicerie commercialise 3 biens : des conserves (C), de la farine (F) et des bouteilles d'huile (H).

Produits	T ₀			T			Δ du CA en Dh courant
	Prix	Qtés	CA	Px	Qtés	CA	
C	5	20	100	4.5	25	112.5	12.5%
F	12	30	360	18	24	432	20%
H	9	40	360	11	36	396	10%
Total			820			940.5	14.7%

Si nous voulons étudier l'évolution du chiffre d'affaire global de cette épicerie **en termes réels**, nous devons l'exprimer en dh constants. Il faut donc «corriger» le CA en dh courants, par un indice des prix. Or, on ne peut pas recourir à la moyenne arithmétique des trois indices de prix correspondants aux trois biens, du fait que ces trois biens n'entrent pas pour parties égales au niveau des quantités vendues (càd dans la formation du CA global).

Nous allons recourir aux indices synthétiques (cf. le cours de statistique du 1er semestre). On retiendra ici l'indice de Laspeyres

Prenons T0 comme année de base :

Nous savons que l'indice de Laspeyres est la moyenne arithmétique PONDEREE des indices élémentaires. Les coefficients de pondération utilisés sont obtenus à partir du poids de chaque bien dans le CA de l'année de base.

Donc :

$$L_{t/t0} = \sum_{i=1} \alpha_{it0} \cdot \frac{P_{it}}{P_{it0}}$$

$$\text{avec } \alpha_{it0} = \frac{P_{it0} Q_{it0}}{\sum P_{it0} Q_{it0}}$$

Produits	T0			T			P_{it} / P_{it0}	α_{it0}	1x2
	Qtés	Px	CA	Qtés	Px	CA	1	2	
E	20	5	100	25	4.5	112.5	90	0.12	10.8
F	30	12	360	24	18	432	150	0.44	66
H	40	9	360	36	11	396	122	0.44	53.68
Total			820			940.5		1	130.48

D'où : $L_{t/0} (P) = 130.48$

Ainsi, le CA de l'année T en dh Cst (de l'année t0) est égal :

$$(940.5 / 130.48) \times 100 = 720.80$$

Utilisation de la formule simplifiée

- La formule de Laspeyres peut-être simplifiée pour faciliter les calculs:
- On remplace les α_{i0} par leur valeur, ce qui donne :
- **la formule de définition :**

$$L_{t/t0} (P) = \sum_{i=1} \alpha_{it0} \cdot \frac{P_{it}}{P_{it0}}$$

- la formule de calcul

$$L_{t/t0} (P) = \sum_{i=1} \frac{P_{it0} Q_{it0}}{\sum P_{it0} Q_{it0}} \cdot \frac{P_{it}}{P_{it0}}$$

$$= \frac{\sum P_{it} Q_{it0}}{\sum P_{it0} Q_{it0}} \times 100$$

Application :

Numérateur Les $P_{it} Q_{i0}$	$\sum P_{it} Q_{i0}$	Dénominateur $\sum P_{i0} Q_{i0}$
4,5 X 20	90	5 X 20 = 100
18 X 30	540	12 X 30 = 360
11 X 40	440	9 X 40 = 360
	1070	820

$$ILP_{t/t_0}(P) = (1070 / 820) \times 100 = \mathbf{130,48}$$

Le résultat est le même que celui obtenu précédemment en appliquant la formule de définition.

Ainsi le Chiffre d'Affaires relatif à l'année T, et exprimé en dirhams constants de l'année T₀ est : (940,5 / 130,48) x 100 = 720,80

Nous pouvons donc dire que la variation du chiffre d'affaires global en T par rapport à T₀ est de :

* **14,695 %, en terme nominal** (ou encore, c'est la variation du CA sur la base des données exprimées en dirhams courants).

$$((940,5 - 820) / 820) \times 100 = 14,695 \%$$

* **- 12,097 % en terme réel** (ou encore, c'est la variation du CA sur la base des données exprimées en dirhams constants).

$$((720,80 - 820) / 820) \times 100 = - 12,097 \%$$

Exemple Deux :

- Considérons une entreprise de textile qui vend des biens :

		Chemises pour femmes	Pulls pou adultes	Vestes	Chemises pour Hommes	Pulls pour enfants	Pantalons	Shorts
1980	Quantité	216	44	71	193	33	93	45
	Prix	0,55	22	4,5	2,7	11,9	11,6	28
1985	Quantité	221	32	79	225	42	102	45
	Prix	0,8	38	5,8	3,25	12,1	13,6	35
1990	Quantité	180	30	80	230	50	110	38
	Prix	1.2	40	5.8	4	12	14	40

- Les variations des quantités vendues pour chaque bien :

$$\Delta Q_{85/80} = [(Q_{85} - Q_{80})/Q_{80}] \times 100$$

	Chemises pour femme	Pulls pour adultes	Vestes	Chemises pour Hommes	Pulls pour enfants	Pantalons	Shorts
$\Delta Q_{t\acute{e}s\ 85/80}$	+2.31%	-27.27%	+11.27%	+16.6%	+27.27%	+9.68%	0%
$\Delta Q_{t\acute{e}s\ 90/80}$	-16.67%	-31.82%	+12.68%	+19.17%	+51.52%	+18.28%	-15.56%

- Les variations des chiffres d'affaires en valeur par bien :

$$\Delta CA_{85/80} = [(CA_{85} - CA_{80})/CA_{80}] \times 100 \text{ (idem pour 90) ex pour les chemises femmes: } [(221 \times 0.8) - (216 \times 0.55) / (216 \times 0.55)] \times 100 = [(176.8 - 118.8) / 118.8] \times 100 = \mathbf{48.8 \%$$

	Chemises pour femmes	Pulls pour adultes	Vestes	Chemises pour Hommes	Pulls pour enfants	Pantalons	Shorts
$\Delta CA\ 85/80$	+48.8%	+25.6%	+43.4%	+40.3%	+29.4%	+28.5%	+25%

Δ CA 90/80	+81.8%	+23.9%	+45.2%	+76.5%	+52.7%	+42.7%	+20.6%
-------------------	---------------	--------	--------	--------	--------	--------	--------

■ Indices des prix par bien par rapport à 1980 : $I [P] 85/80 = (\text{Prix } 85 / \text{Prix } 80) \times 100$

	Chemises pour femmes	Pulls pour adultes	Vestes	Chemises pour Hommes	Pulls pour enfants	Pantalons	Shorts
Ind Px 85/80	145.45	172.73	128.89	120.37	101.68	117.24	125
Ind Px 90/80	218.18	181.82	128.89	148.15	100.84	120.69	142.86

■ Les chiffres d'affaires en Dh constants de l'année 80 (année de base) :

CA de l'année 1985, exprimés en Dhst de l'année 80 = [CA en Va 85 / (Ind Px 85/80)] x 100 (idem pour 90)

	Chemises pour femme	Pulls pour adultes	Vestes	Chemises pour Hommes	Pulls pour enfants	Pantalons	Shorts
CA de 80 en D de 80	118.8	968	319.5	521.1	392.7	1078.7	1260
CA de 85 en Dh Cst de 80	121.59	703.99	355.5	607.5	499.8	1183.21	1260
CA de 90 en D Cst de 80	99.08	659.99	360	620	595	1276	1063.97

■ La variation du chiffre d'affaires en Dh Cst de l'année 80 (l'année de base) :

Δ CA de l'année 1985 = [(CA de 85 en Dh cst - CA de 80) / CA 80] x 100

	Chemises pour femmes	Pulls pour adultes	Vestes	Chemises pour Hommes	Pulls pour enfants	Pantalons	Shorts
Δ CA 85/80 en termes réels	+2.31%	-27.27%	+11.27%	+16.6%	+27.27%	+9.68%	+0%
Δ CA 90/80 en terme réel	-16.6%	-31.82%	-12.68%	-19.17%	-51.52%	+18.28%	-15.56%

■ Le chiffre d'affaires en Dh constants de l'année 80 (année de base) réalisé par exemple sur les ventes des chemises pour femmes est de 118.8 en 1980, 121.59 en 1985 et 99.08 en 1990. La variation du chiffre d'affaires exprimé en Dh constants, par rapport à 1980 est de + 2.31% en 1985 et de -16.6% en 1990. Il s'agit donc des mêmes variations que les quantités, puisque l'effet de la variation des prix a été écarté. **Ce sont des variations en termes réels.**

Rappel

■ A partir des données en Dh courants, nous calculons les variations en termes nominaux. Ainsi le chiffre d'affaires a connu une hausse importante pour "les chemises destinées aux femmes" (+ **48.8 %** en 1985 et + **81.8 %** en 1990, par rapport à 1980).

En écartant l'effet de l'augmentation des prix, la variation du CA devient : + **2.31%** en 1985 / 80 et - **16.6 %** en 1990/80, ce qui traduit en réalité une perte de part de marché sur ce produit.

Calcul des Variations du **CA global en termes réels**: application de l'indice de Laspeyres

Pour calculer les variations du **CA global** en termes réels, nous devons:

- 1) Calculer un indice **synthétique des prix**, pour l'ensemble des biens (dont nous ne pouvons ajouter les quantités).
- 2) déflater la série du CA global en DH courants par l'indice des prix obtenu
- 3) calculer les variations à l'aide des résultats / données obtenues

1/ Calcul de l'indice synthétique des prix :

- **la formule de définition :**

$$L_{t/80} (P) = \sum_{i=1} \alpha_{i80} \cdot \frac{P_{it}}{P_{i80}}$$

- **la formule de calcul**

$$L_{t/t0} (P) = \sum_{i=1} \frac{P_{i80} Q_{i80}}{\sum P_{i80} Q_{i80}} \cdot \frac{P_{it}}{P_{i80}}$$

$$= \frac{\sum P_{it} Q_{i80}}{\sum P_{i80} Q_{i80}} \times 100$$

Application à 1985

$P_{i85} \cdot q_{i80}$	$\sum P_{i85} \cdot q_{i80}$	$P_{i80} \cdot q_{i80}$	$\sum P_{i80} \cdot q_{i80}$	CA85 enDh courants
0,8*216	172,8	0,55*216	118,8	176,8
38*44	1672	22*44	968	1216
5,8*71	411,8	4,5*71	319,5	458,2
3,25*193	627,25	2,7*193	521,1	731,25
12,1*33	399,3	11,9*33	392,7	508,2
13,6*93	1264,8	11,6*93	1078,8	1387,2
35*45	1575	28*45	1260	1575
	6122,95		4658,9	6052,65

$$L_{85/80} = 6122,95/4658,9 \times 100 = 131,42$$

Les prix ont globalement augmenté de 31,42% en 1985 par rapport à 1980.

2/ Calcul du CA de 1985, exprimé en Dh constants (de l'année 1980)

$$(CA85 \text{ en Dh courants} / L_{85/80} [P]) \times 100 = (6052,65 / 131,42) \times (100) = 4605,577$$

La variation du Ca entre 1980 et 1985 est de :

**** -1,144 % en termes réels ; soit : $((4605,577 - 4658,9) / 4658,9) \times 100$.**

**** 29,9 % en termes nominaux, soit : $((6052,65 - 4658,9) / 4658,9) \times 100$.**

Voir l'annexe III pour d'autres exercices d'application.

ANNEXE III

Exercices complémentaires

Les exercices ci-après portent sur :

- i. les traitements préalables des séries statistiques
- ii. et les régressions / estimation de paramètres et analyse des résultats

I Exercices d'application / les traitements préalables des séries statistiques

Exercice 1 - Analyse de l'évolution des exportations de trois produits

Le tableau suivant donne le niveau des exportations de 3 produits d'un pays pour 3 années 1995, 2000 et 2005

Produits Années	Phosphates		Tomates		Chemises	
	Quantités (en 1000 tonnes)	Prix (DH/T)	Quantités (en 10 ³ Tonnes)	Prix (10 ³ DH/T)	Quantités (milliers d'unités)	Prix (10 ⁶ DH /1000 unités))
1995	250	20	10	50	400	10
2000	290	18	8	78	500	11
2005	220	28	14	52	650	15

- 1) Calculer les taux de variations des quantités exportées pour chaque produit en 2000 par rapport à 1995 et en 2005 par rapport à 1995
- 2) Calculer les recettes d'exportations pour chaque produit, pour les années 1995, 2000 et 2005
- 3) Calculer pour chaque produit les taux de variations des recettes d'exportations en dirhams courants en 2000 par rapport à 1995 et en 2005 par rapport à 1995
- 4) Calculer les indices de prix pour chaque bien, avec 1995 comme année de base
- 5) Calculer les recettes d'exportations, en DH constants de l'année 1995, pour chaque produit et chaque année
- 6) Calculer pour chaque produit les variations en termes réels des recettes d'exportation, en 2000 par rapport à 1995 et en 2005 par rapport à 1995
- 7) Commentez les résultats
- 8) Calculez les recettes globales des exportations et les taux de variations de ces recettes en 2000 par rapport à 1995 et en 2005 par rapport à 1995
- 9) Calculer la valeur globale des recettes des exportations, pour l'année 2005, en dirhams constants de 1995 (année de base).
 - a. Pourquoi est-il nécessaire de passer par un indice synthétique des prix ?
 - b. Calculer l'indice synthétique des prix des exportations
 - c. Déflater la série des exportations
- 10) Calculer les taux de variations en termes réels, des exportations globales de ce pays
- 11) Commentez les résultats

Solution :

- 1) Calcul des taux de variations des quantités exportées pour chaque produit en 2000 par rapport à 1995 et en 2005 par rapport à 1995

	Phosphates	Tomates	Chemises
<u>Δ Qtés 2000/1995</u>	16%	-20%	25%
Δ Qtés 2005/1995	-12%	40%	63%

Par rapport à l'année 1995, les quantités exportées de phosphates ont augmenté de 16 % en 2000 et ont diminué de 12 % en 2005. Celles de tomates ont d'abord baissé en 2000 (-20 %) et ont cru en 2005 (40 %) par rapport à 1995.

2) Calcul des recettes d'exportations pour chaque produit, pour les années 1995, 2000 et 2005

Recettes	Phosphates	Tomates	Chemises
1995	5000	500	4000
2000	5220	624	5500
2005	6160	728	9750

2) Calcul pour chaque produit les taux de variations des recettes d'exportations en dirhams courants en 2000 par rapport à 1995 et en 2005 par rapport à 1995

<u>Δ Rec en Dh Courants</u>	Phosphates	Tomates	Chemises
<u>Δ Rec 2000/1995</u>	4,40%	24,80%	37,50%
Δ Rec 2005/1995	23,20%	45,60%	143,75%

Les recettes tirées des exportations de phosphates ont augmenté de 4,40 % en 2000 par rapport à 1995, la hausse a été de 23,2 % en 2005. En fait ce sont les prix qui expliquent cette dernière progression, puisque les quantités ont diminué de 12 % (voir plus haut). On peut analyser de la même manière les variations des recettes pour les autres produits, comparées à celles de quantités.

3) Calcul des indices de prix pour chaque bien, avec 1995 comme année de base

	Phosphates	Tomates	Chemises
<u>Ind Px 2000/1995</u>	90	156	110
Ind Px 2005/1995	140	104	150

Les prix des phosphates ont diminué de 10 % en 2000 par rapport à 1995 et augmenté de 40 % en 2005. Par rapport à 1995, les prix de la tomate ont très fortement progressé en 2000 (+ 56 %) et faiblement en 2005 (+ 4 %). Les indices montrent qu'en fait les prix ont diminué en 2005 par rapport à 2000 (en réalité 52 DH/T contre 78 DH/T).

4) Calcul des recettes d'exportations, en DH constants de l'année 1995, pour chaque produit et chaque année

<u>Recettes en Dh Cst</u>	Phosphates	Tomates	Chemises
<u>Rec en Dh Cst 1995</u>	5000	500	4000
Rec en Dh Cst 2000/1995	5800	400	5000
Rec en Dh Cst 2005/1995	4400	700	6500

En Dh constants de l'année 1995, les recettes des exportations de phosphates ont atteint 5000 en 1995 (la même valeur qu'en Dh courants), 5800 en 2000 (contre 5220 en DH courants) et 4400 (contre 6160 en DH courants).

- 5) Calcul pour chaque produit les variations en termes réels des recettes d'exportation, en 2000 par rapport à 1995 et en 2005 par rapport à 1995

Δ Rec en Dh cst	Phosphates	Tomates	Chemises
Δ Rec en Dh cst 2000/1995	16%	-20%	25%
Δ Rec en Dh cst 2005/1995	-12%	40%	63%

- 6) Commentaire des résultats

Les variations en termes réels des exportations sont identiques à celles observées pour les quantités. En d'autres termes, travailler sur les quantités physiques ou sur les séries exprimées en Dh constants, revient au même.

- 7) Calcul des recettes globales des exportations et des taux de variations de ces recettes en 2000 par rapport à 1995 et en 2005 par rapport à 1995

Recettes en Dhs Courants	Phosphates	Tomates	Chemises	Recettes Globales	Δ Rec par rapport à 1995
1995	5000	500	4000	9500	-
2000	5220	624	5500	11344	19%
2005	6160	728	9750	16638	75%

Par rapport à 1995, la valeur globale des exportations globales a progressé de 19 % en 2000 et 75 % en 2005. Mais ces variations cachent des variations différentes des quantités exportées de chaque bien ainsi que de leur prix. Il importe d'éliminer l'effet des prix.

- 8) Calcul des recettes globales des exportations en dirhams constants :

- Il est nécessaire de passer par un indice synthétique des prix, car les exportations globales sont composées de produits hétérogènes ; on ne peut additionner les tonnes de phosphates avec les unités de chemises... De plus, les produits n'ont pas une importance égale dans la valeur des exportations globales.
- Calcul de l'indice synthétique des prix des exportations

Application à 2005

On utilise la formule qui simplifie les calculs / voir le cours et le polycopié.

Produits	$\Sigma P_{i2005} \cdot q_{i1995}$	$\Sigma P_{i1995} \cdot q_{i1995}$	Rec 2005 en Dh Courants
Phosphates	7000	5000	6160
Tomates	520	500	728
Chemises	6000	4000	9750
Total	13520	9500	16638

$$L_{2005/1995} [P] = 13520/9500 \cdot 100 = 142,31$$

- Déflater la série des exportations

Calcul des recettes d'exportation de 2005, exprimée en Dh constants (de l'année 1995)

$$(\text{Rec 2005 en Dh courants} / L_{2005/1995} [P]) * 100 = 16638 / 142,31 * 100 = \mathbf{11691,38}$$

9) Calculer les taux de variations en termes réel, des exportations globales de ce pays

La variation des recettes d'exportation entre 2005 et 1995 est de :

* +23% en termes réels (soit : $11691,38/9500$)*100 Ce résultat correspond à la variation du volume des exportations globales.

* On rappelle que le nombre obtenu plus haut (+75%) correspond à la variation en termes nominaux (soit : $16638/ 9500$)*100

10) Commentaire des résultats

De la même manière que pour un produit, il est utile de considérer les variations des quantités, pour avoir une idée précise de la tendance, pour une grandeur composite, l'utilisation d'un indice approprié permet d'obtenir la variation en termes réels.

Exercice 2 :

La fabrication de chaussures, de cartables et d'autres produits en cuir nécessitent 5 inputs.

X1 cuir traité

X2 demi-produits

X3 teinture

X4 produits chimiques

X5 fils et autres fournitures

Les quantités sont exprimées en tonnes

Nous disposons des données sur les prix payés par la branche industrielle des chaussures, cartables et autres produits en cuir et les quantités utilisées de chaque composant, pour les années 2000 et 2005.

	X1	X2	X3	X4	X5
Quantités utilisées en 2000	40	12	18	60	12
Prix unitaires en 2000 (en 10 ⁶ DH)	28	50	11	29	49

Etant donné les modifications au niveau des prix des inputs, des goûts de la clientèle et des capacités des machines, en 2005 les combinaisons productives ont été sensiblement modifiées, pour produire des quantités différentes de chaussures, cartables et autres produits en cuir.

Pour 2005, nous avons :

	X1	X2	X3	X4	X5
Quantités utilisées en 2005	31	15	14	45	9
Prix unitaires en 2005	37	56	19	34	53

- classer les consommations intermédiaires en fonction de leur prix unitaire en 2000 et en 2005
- calculer la valeur de chaque input, en 2000 et en 2005
- calculer les variations de la valeur consommée de chaque input en 2005 par rapport à 2000
- calculer les variation des quantités consommées de chaque input en 2005 par rapport à 2000
- pourquoi ces indications sont-elles insuffisantes, pour appréhender la variation relative des coûts de l'industrie des chaussures, cartables et autres produits en cuir ?
- calculer la variation en termes réels de la **valeur globale** des consommations intermédiaires de cette industrie entre 2000 et 2005 ; pour ce faire :
 - calculer l'indice des prix pour chaque input, en 2005 par rapport à 2000
 - calculer l'indice synthétique des prix (pour les 3 inputs), en 2005 par rapport à 2000

- c. calculer la valeur globale des consommations intermédiaires pour 2005, en dirhams constants de l'année 2000
 - d. Calculer la variation en termes réels de la valeur des consommations intermédiaires
- 7) commentez vos résultats et dire l'intérêt de ce travail

Solution :

1. classement des consommations intermédiaires en fonction de leur prix unitaire en 2000 et en 2005

- a. en 2000

Inputs 2000	Quantités utilisées en 2000	Prix unitaires en 2000
X3	18	11
X1	40	28
X4	60	29
X5	12	49
X2	12	50

- b. en 2005

Inputs 2005	Quantités utilisées en 2005	Prix unitaires en 2005
X3	14	19
X4	45	34
X1	31	37
X5	9	53
X2	15	56

On observe que les prix ont évolué de telle manière que les inputs ont changé de rang par rapport au coût qu'ils représentent. Cela explique en partie les changements des quantités utilisées (consommées) dans le processus de production de la branche industrielle des chaussures etc...

2. Calcul de la valeur de chaque input en 2000 et en 2005

	X1	X2	X3	X4	X5
Val de la CI en 2000	1120	600	198	1740	588
Val de la CI en 2005	1147	840	266	1530	477

Pour produire les chaussures, les cartables et autres produits en cuir, l'industrie considérée a dépensé en 2000 : 1120 millions de DH en cuir traité, 600 en demi-produits etc... En 2005, le valeur consommée en cuir traité a atteint 1147, celle des teintures 266 (contre 198 en 2000) etc...

3. calcul des variations de la valeur consommée de chaque input en 2005 par rapport à 2000

	X1	X2	X3	X4	X5
Δ Val de la CI 2005/2000	2%	40%	34%	-12%	-19%

Etant donné les substitutions opérées pour tenir compte des changements de prix des inputs et des goûts des consommateurs (le niveau de la demande), les valeurs consommées de chaque input ont varié différemment : pour le cuir traité, il y eu une augmentation de 2 %, 40 % pour les demi-produits et 34 % pour les teintures. Par

contre, pour les produits chimiques et le fil et autres fournitures, il y a eu une baisse : - 12 et – 19 % respectivement.

4. calculer la variation de la quantité consommée de chaque produit en 2005 par rapport à 2000

	X1	X2	X3	X4	X5
Δ Quantités 2005/2000	-23%	25%	-22%	-25%	-25%

La quantité consommée en cuir traité a baissé de 23 %, celle des demi-produits a progressé de 25 % etc... On en déduit une baisse globale de la production de l'industrie des chaussures, cartables ...

5. Mais ces indications sont insuffisantes, pour appréhender la variation relative des coûts de l'industrie des chaussures, et au-delà, pour définir une politique optimale de prix. En effet, nous observons que pour certains inputs les prix ont augmenté fortement, pour d'autres faiblement ; les modifications qui ont en résulté des quantités consommées ont induit des changements dans la valeur consommée de chaque input ; mais nous n'avons pas d'idée précise sur le coût global (d'où l'intérêt de la question suivante).
6. Calcul de la variation en termes réels de la **valeur globale** des consommations intermédiaires de l'industrie des chaussures et cartables entre 2000 et 2005
- a. calcul de l'indice des prix pour chaque input, en 2005 par rapport à 2000

	X1	X2	X3	X4	X5
Ind Px 2005/2000	132,14	112,00	172,73	117,24	108,16

- b. calcul de l'indice synthétique des prix (pour les 5 inputs), en 2005 par rapport à 2000

Inputs	Σ P_{i2005}*q_{i2000}	Σ P_{i2000} q_{i2000}	CI 2005 en Dh Courants
X1	1480	1120	1147
X2	672	600	840
X3	342	198	266
X4	2040	1740	1530
X5	636	588	477
Total	5170	4246	4260

$$L_{2005/2000} [P] = 5170/4246*100 = \mathbf{121,76}$$

La valeur globale des consommations intermédiaires a progressé de 21,76 % en 2005 par rapport à 2000. C'est sur cette base que devront ajuster les prix de vente de cette industrie, et sur la base des possibilités de vente et de gains sur chaque produit (chaussures, cartables et autres produits en cuir).

- c. calcul de la valeur globale des consommations intermédiaires de 2005, exprimée en dirhams constants de l'année 2000

$$(\text{Val CI 2005 en Dh courants}/L_{2005/2000} [P]) * 100 = 4260 / 121,76 * 100 = \mathbf{3498,68}$$

- d. Calcul de la variation en termes réels de la valeur globale des consommations intermédiaires entre 2005 et 2000 :

$$* \mathbf{-17,6\%}$$
 en termes réels : $((3498,68 - 4246)/4246)*100$

$$* \mathbf{+0,3\%}$$
 en termes nominaux : $((4260 - 4246)/4246)*100$

7. commentez vos résultats

Pour faire face à l'augmentation des prix des inputs et au changement de la demande, cette industrie a non seulement réduit sa production et modifié son processus de production. Elle a du opéré des changements dans les inputs. Mais de ce fait il devenait difficile d'avoir une idée précise de l'évolution du coût global. Les calculs effectués nous ont permis de connaître la variation des coûts réels de cette industrie (question 6) et la variation de sa consommation globale en inputs (question 7), une fois réalisés les ajustements.

Remarque générale :

Les calculs effectués ici, comme ceux des autres exercices et des exemples présentés dans le cours, montrent qu'il est important de savoir traiter les données, pour passer des séries en valeur (DH courants) à des séries en volume (DH constants). On observe que cela est utiles pour connaître les tendances réelles de grandeurs composites, que celles-ci concernent des chiffres d'affaires (ventes, exportations...) ou coûts de production ; à l'image du panier de la ménagère qui représente la dépense globale du ménage (et pour lequel, on calcule un indice des prix à la consommation, pour mesurer l'inflation).

II) Exercices d'application pour les régressions

Exercice UN :

Le tableau suivant donne pour un pays, la consommation, le revenu et l'indice des prix pour les années 1997 à 2006.

Années	C_t	R_t	Indice des prix (IP)
1997	20	70	100
1998	25	73	121
1999	32	77	145
2000	29	74	169
2001	35	82	187
2002	37	84	199
2003	42	87	229
2004	42	94	270
2005	44	96	306
2006	50	95	369
Moyenne	35,6	83,2	209,5

Etablir le nuage de points représentant la consommation en fonction du temps, qu'en déduisez-vous ?

- 1) Calculer les paramètres du modèle $C_t = a R_t + b$ et commenter les résultats
- 2) Calculer les paramètres du modèle $C_t = \alpha IP_t + \beta$ et commenter les résultats. Sont-ils conformes à nos attentes et pourquoi ?
- 3) Calculer la consommation en DH constants
- 4) Calculer les paramètres de l'équation : $CV_t = \delta IP_t + \lambda$ Commenter et analyser les résultats par rapport aux résultats précédents

Solution :

Le nuage de points montre que la consommation augmente avec le de manière revenu de manière assez continue dans le temps.

L'estimation des deux premiers modèles requiert les calculs suivants.

$(C_t - C_m)$	$(R_t - R_m)$	$(R_t - R_m)^2$
-15,6	-13,2	174,24
-10,6	-10,2	104,04
-3,6	-6,2	38,44
-6,6	-9,2	84,64
-0,6	-1,2	1,44
1,4	0,8	0,64
6,4	3,8	14,44
6,4	10,8	116,64
8,4	12,8	163,84
14,4	11,8	139,24
Somme		837,6

L'indice m signifie qu'il s'agit de la moyenne : C_m = consommation moyenne et R_m = revenu moyen.

$(C_t - C_m) \cdot (R_t - R_m)$	$IP_t - IP_m$	$(IP_t - IP_m)^2$	$(IP_t - IP_m) \cdot (C_t - C_m)$
205,92	-109,5	11990,25	1708,2
108,12	-88,5	7832,25	938,1
22,32	-64,5	4160,25	232,2
60,72	-40,5	1640,25	267,3
0,72	-22,5	506,25	13,5
1,12	-10,5	110,25	-14,7
24,32	19,5	380,25	124,8
69,12	60,5	3660,25	387,2
107,52	96,5	9312,25	810,6
169,92	159,5	25440,25	2296,8
769,8		65032,5	6764

1) L'estimation des paramètres du modèle $C_t = a R_t + b$ donne $\hat{a} = 0.919$ et $\hat{b} = -40.865$

Le fait que le signe du coefficient a soit positif signifie que le revenu agit positivement sur la consommation. Plus précisément, l'augmentation du revenu d'une unité monétaire entraîne un accroissement de la valeur de la consommation de 0,91 unités.

La constante n'a pas toujours de signification dans les modèles économétriques.

2) l'estimation des paramètres de l'équation $C_t = \alpha IP_t + \beta$ donne, après calculs :

$C_t = ,0104 P_t - 13,8$ soit ; $\hat{\alpha} = 0,104$ et $\hat{\beta} = -13,8$

Cela voudrait dire que les prix agissent positivement sur la consommation. Ce résultat n'est pas conforme à l'intuition théorique à l'origine de la spécification du modèle. Normalement α devrait être négatif puisque dans la réalité, une augmentation des prix se traduit par une baisse du niveau de la consommation. Ce résultat « contre intuitif » tient au fait que nous avons travaillé sur des séries en valeur ; nous n'avons pas éliminé l'effet de la variation des prix (voir les questions suivantes).

3) Pour obtenir les niveaux de la consommation en volume, nous déflatons la série de la consommation en DH courants par l'indice des prix. Le résultat figure dans la colonne 1 du tableau ci-après.

C_t (en DH constants) = CV_t	CV_t - CV_m	(CV_t - CV_m) (IP_t - P_m)
20,00	2,10	-229,68
20,66	2,76	-244,14
22,07	4,17	-268,74
17,16	-0,74	30,08
18,72	0,81	-18,32
18,59	0,69	-7,25
18,34	0,44	8,54
15,56	-2,35	-141,99
14,38	-3,52	-340,01
13,55	-4,35	-694,20
17,9		-1905,70

On observe que contrairement à la consommation en valeur, la consommation exprimée en Dh constants ou consommation en volume (Cv) diminue.

4) Pour l'équation $CV_t = \delta IP_t + \lambda$, on trouve : $CV_t = -0,029 P_t + 210,02$ ou encore :

$\delta = -0,029$ $(-1905,70 / 65\,032,5)$ et $\lambda = 23,97$ $(17,9 - (-0,029 \times 209,5))$

On relève bien ici que les prix ont un effet négatif sur le volume de la consommation.

Exercice Deux :

Le tableau suivant donne, pour les années 1996 à 2005, des séries relatives à l'Investissement (Inv), au profit (Pr) et au taux d'intérêt (Ti).

Inv	Pr	Taux d'intérêt (Ti)
125	25	12
132	2	11,2
137	33	9
142	36	8,3
154	38	7,6
161	41	6
170	45	5,2
183	48	5,1
196	50	4,9
208	52	4,7
160,8	37	7,4

1) Soit le modèle simple expliquant l'investissement par le profit : $I_t = a Pr_t + b$; en estimer les paramètres

2) considérer ensuite l'explication des variations de l'investissement par celles du taux d'intérêt : $I_t = \delta T_i + \mu$

Commenter chaque fois les résultats

Solution

1) L'estimation des paramètres de l'équation $I_t = a Pr_t + b$, suppose les calculs suivants :

I_t - I_m	Pr_t - Pr_m	(I_t - I_m)*(Pr_t - Pr_m)	(Pr_t - Pr)²
-35,8	-12	429,6	144
-28,8	-35	1008	1225
-23,8	-4	95,2	16
-18,8	-1	18,8	1
-6,8	1	-6,8	1
0,2	4	0,8	16
9,2	8	73,6	64
22,2	11	244,2	121
35,2	13	457,6	169
47,2	15	708	225
		3029	1982

Ainsi, l'estimation donne $\hat{a} = 1,528$ $(\text{Som } (I_t - I_m) (\text{Prt} - \text{Pr}) / \text{Som } (\text{Prt} - \text{Pr})^2$ ou encore $(3029 / 1982)$

et $b^{\wedge} = 104,255$ $(I \text{ moyen} - \hat{a} \text{ Pr moyen})$

Le profit agit positivement sur l'investissement. Cela est conforme à nos attentes, en effet la réalisation de profit fournit les moyens de financer l'investissement et constitue par la même un signal favorable pour investir.

2) pour la seconde équation $I_t = \delta T_t + \mu$, nous avons :

T_t - T_m	(T_t - T_m)²	(T_t - T_m)(I_t - I_m)
4,6	21,16	-164,68
3,8	14,44	-109,44
1,6	2,56	-38,08
0,9	0,81	-16,92
0,2	0,04	-1,36
-1,4	1,96	-0,28
-2,2	4,84	-20,24
-2,3	5,29	-51,06
-2,5	6,25	-88
-2,7	7,29	-127,44
64,64	-617,5	

Ce qui donne : $\delta = - 9,553$ $(\text{Som } (T_t - T_m) (I_t - I_m) / \text{som } (T_t - T_m)^2$ ou encore $(- 617,5 / 64,64)$

et $\mu = 231,492$

Le signe du coefficient δ correspond à nos attentes, puisque normalement le taux d'intérêt agit négativement sur le niveau de l'investissement.

Exercice 3 : Demande de logements, taux d'intérêt et revenu

Le tableau suivant donne pour les années 1990 à 2004, la série de la demande de logements, de l'évolution du taux d'intérêt et du revenu des ménages.

Années	Demande de logements (D)	Taux d'intérêt (Txi)	Revenu des ménages (R)
1990	40	9	400
1991	45	8	500
1992	50	9	600
1993	55	8	700
1994	60	7	800
1995	70	6	900
1996	65	6	1000
1997	65	8	1100
1998	75	5	1200
1999	75	5	1300
2000	80	5	1400
2001	100	3	1500
2002	90	4	1600
2003	95	3	1700
2004	85	4	1800
Moyenne	70	6	1100

- 1) Quels sont selon vous les déterminants de la demande de logement dans un pays ?
- 2) A partir des données fournies, estimer les paramètres d'un modèle expliquant la demande de logements par le taux d'intérêt et d'un modèle retenant le revenu comme variable explicative.

Solution

- 1) Dans un pays la demande de logements dépend de l'importance de la pression démographique (rythme de croissance de la population, et en particulier de la tranche d'âge des 25 ans et plus) et de l'évolution du revenu disponible. Elle dépend également de l'indice des prix du bâtiment (coûts de construction des logements), des possibilités d'accès au crédit et du coût de celui-ci / taux d'intérêt (cf. le FOGARIM). La demande de logements dépend enfin de facteurs sociologiques.
- 2) L'explication des variations de la demande de logements par celles du taux d'intérêt, suppose l'estimation du modèle suivant : $D_t = a T_{xi_t} + b$, avec D_t la demande de logements et T_{xi} le taux d'intérêt.
Pour évaluer l'effet du revenu, nous pouvons retenir la spécification suivante :
 $D_t = \delta R_t + \lambda$

Pour obtenir les paramètres de liaison contenus dans ces deux équations, il importe d'effectuer les calculs récapitulés dans les tableaux ci-après :

$(D_t - D_m)$	$(T_{xi_t} - T_{xm})$	$(R_t - R_m)$	$(D_t - D_m)^2$	$(T_{xi_t} - T_{xm})^2$	$(R_t - R_m)^2$
-30	3	-700	900	9	490000
-25	2	-600	625	4	360000
-20	3	-500	400	9	250000
-15	2	-400	225	4	160000
-10	1	-300	100	1	90000
0	0	-200	0	0	40000
-5	0	-100	25	0	10000
-5	2	0	25	4	0
5	-1	100	25	1	10000
5	-1	200	25	1	40000
10	-1	300	100	1	90000
30	-3	400	900	9	160000
20	-2	500	400	4	250000
25	-3	600	625	9	360000
15	-2	700	225	4	490000
			4600	60	2800000

$(D_t - D_m) * (T_t - T_m)$	$(D_t - D_m) * (R_t - R_m)$
-90	21000
-50	15000
-60	10000
-30	6000
-10	3000
0	0
0	500
-10	0
-5	500
-5	1000
-10	3000
-90	12000
-40	10000
-75	15000
-30	10500
-505	107500

Les résultats des estimations donnent :

- a. Dans le 1^{er} cas : $\hat{a} = -8,42$ et $\hat{b} = 120,50$. Les deux variables varient en sens inverse. Une baisse du taux d'intérêt entraîne un accroissement de la demande de logement. Cela s'explique par le fait que le taux d'intérêt représente le coût de l'endettement pour l'accès à la propriété.
- b. d'autre part : $\delta = 0,038$ et $\lambda = 27,77$. Le revenu et la demande de logements varient dans le même sens. Une augmentation du revenu encourage les ménages à accéder à la propriété.

Les modèles simples présentés jusqu'ici prennent chaque fois une seule variable explicative. Comme cela est expliqué dans le cours, il est le plus souvent nécessaire de recourir à plusieurs variables, pour raisonner « toutes choses égales par ailleurs ». Tenir compte de toutes les variables sensées agir sur la variable à expliquer, permet d'avoir des estimateurs plus efficaces (nous évitons les « sous-spécifications »).

L'estimation des paramètres d'un modèle linéaire de régression multiple relève cependant de la même logique que celle du MLRS ; seulement au lieu d'appliquer les formules pour \hat{a} et \hat{b} , on recourt au calcul matriciel / voir le cours. Plusieurs logiciels permettent d'effectuer les estimations ; le but des exercices précédents était de familiariser les étudiants avec le travail de modélisation.