

Chapitre I. Introduction générale

I.1. Utilité et définitions de l'économetrie

L'analyse des phénomènes économiques vise essentiellement à mettre en évidence les mécanismes qui régissent ces phénomènes afin de mieux comprendre leur nature et leur fonctionnement, d'une part et de prévoir leur évolution d'autre part.

L'économiste cherche donc, dans sa démarche à caractériser les liens qui unissent les diverses variables intervenant dans l'explication des phénomènes économiques et, si possible, à dégager des lois de comportement sous-jacentes.

Parmi les outils d'analyse quantitative utilisés par les chercheurs et les praticiens dans divers domaines d'application en économie et en gestion, l'économétrie occupe une place de choix.

De nombreuses définitions de l'économétrie ont été proposées par divers auteurs. Elles sont vraisemblablement différentes les unes des autres mais l'idée commune que l'on y retient est que l'économétrie est un outil d'analyse qui intègre les instruments mathématiques et statistiques à l'économie politique.

- Au sens littéral, l'économétrie signifie « mesure de l'économie ». Bien que la mesure soit une part importante de l'économétrie, le domaine de cette discipline est plus vaste et va au-delà d'une simple mesure des faits économiques.
- Dans son acception la plus restreinte, l'économétrie est un ensemble de techniques dans laquelle les outils de la théorie économique, les mathématiques et les déductions statistiques sont appliquées à l'analyse des phénomènes économiques.
- Au sens large, l'économétrie est l'art de construire et d'estimer des modèles empiriques adéquats par rapport aux caractéristiques de la réalité, et conformes avec la théorie économique.

Comme le suggèrent les définitions précédentes, l'économétrie est un mélange de théorie économique, d'économie mathématique, de statistiques économiques et de statistique mathématique.

Il convient de rappeler que :

- la théorie économique avance des propositions et des hypothèses qui sont pour la plupart de nature qualitative ;
- l'économie mathématique traduit la théorie économique sous forme mathématique sans considérer le caractère mesurable ou la vérification empirique de la théorie ;
- la statistique économique est centrée sur la collecte, le traitement et la présentation des données économiques sous la forme de graphique et de tableaux ;

- la statistique mathématique fournir de nombreux outils utilisés en économétrie.

Conçue comme une méthode d'analyse, l'économétrie n'a pas pour objet de se substituer à la pensée économique mais bien de la quantifier sous l'hypothèse que la quantification d'un problème représente un réel progrès scientifique. Elle apporte la rigueur dans le langage et les concepts, la cohérence dans l'analyse des phénomènes interdépendants, la clarté dans la transmission et l'exposé des théories. En outre, elle facilite la confrontation quantitative des résultats obtenus aux données réelles. Elle constitue donc le complément indispensable de l'analyse économique.

Le recours à l'économétrie comme outil d'analyse quantitative nécessite que la théorie économique soit formalisée à l'aide des modèles économiques.

I.2. Notions de modèle en économie

I.2.1. Définition et objectif d'un modèle

Par modèle économique, on entend une représentation simplifiée d'un phénomène sous forme d'équations dont les variables sont des grandeurs économiques. Cette représentation quantitative d'un phénomène économique est fondée sur des hypothèses concernant le comportement des agents impliqués. En effet, à partir d'une réflexion théorique, elle conduit à imaginer des mécanismes d'interaction entre différentes variables économiques.

La qualité du modèle retenu pour expliquer un phénomène économique donné dépendra des variables qu'il contient ainsi que de relations liant ces variables.

Tout modèle permet de comprendre le phénomène étudié, d'en prévoir et, si possible, d'en contrôler l'évolution future. Le choix du modèle, c'est-à-dire des variables utilisées et des relations considérées, dépend de l'importance relative accordée à chacun des aspects explicatifs, prédictifs ou de contrôle ainsi que de la précision souhaitée. Donc, il n'y a pas de modèle unique mais autant des modèles que les problèmes à résoudre. En outre, on ne peut juger un modèle indépendamment de ses objectifs. Il n'est pas non plus juste de dire qu'une représentation « trop simpliste » est nécessairement incorrecte. Pour résoudre un problème donné, la première préoccupation du modélisateur est donc de savoir quelle(s) question(s) spécifique(s) il cherche à solutionner.

Soulignons enfin qu'un modèle n'est donc qu'un ensemble d'équations mathématiques. Si le modèle comporte une seule équation, il prend le nom de *modèle à équation unique* ; s'il a plusieurs équations, on parle de *modèle à équations multiples* ou *modèle à équations simultanées*.

1.2.2. Eléments principaux d'un modèle

Tout modèle comprend des variables des relations et des paramètres dont la nature et le rôle doivent être précisés.

1.2.2.1. Variables

Une *variable économique* est une quantité qui peut prendre n'importe quelle valeur à l'intérieur d'un certain domaine de variation. Ces valeurs numériques peuvent être discrètes ou continues. Très souvent une variable discrète est considérée comme un cas particulier d'une variable continue.

L'usage des variables pose parfois le problème de la capacité à les mesurer. Or, la quantification de la réalité économique peut entraîner des problèmes insolubles. Il arrive souvent que la mesure d'un fait économique soit établie par l'intermédiaire d'une liaison avec un autre fait lui – même plus accessible à la mesure. Ainsi, on peut mesurer le niveau d'activité économique d'un pays par la consommation d'électricité, la quantité de pétrole importé, etc.

Faisons remarquer que la construction des modèles sans la possibilité de disposer des mesures statistiques satisfaisantes est un travail purement formel. D'ailleurs, selon Oskar MORGENSTERN : « il est parfaitement inutile d'élaborer un modèle complexe si l'on ne peut le nourrir qu'avec des données statistiques insuffisantes ou de qualité douteuse ».

Généralement, on classe les variables économiques en :

1.2.2.1.1. Variables endogène et exogène

On appelle « *variable endogène* » (dépendante, expliquée ou régressant) une variable dont la valeur est déterminée au sein du modèle lui – même. Alors que les variables qui sont déterminées en dehors du modèle, c'est-à-dire qui entrent dans le modèle comme des données autonomes sont qualifiées de « *variables exogènes* » (indépendantes, explicatives ou régresseurs). La nature, endogène ou exogène, des variables est rarement une propriété intrinsèque du modèle.

Des variables exogènes d'un modèle économique, on distingue :

- les variables réellement externes au phénomène étudié, et sur lesquelles les agents économiques considérés n'ont aucune maîtrise sur elles ;
- les variables contrôlées par un agent économique spécifique (Ménages, Entreprises, Institutions financières, Etat, Reste du monde), mais dont on se refuse à modéliser le comportement parce qu'on désire maîtriser leur valeur, pour en mesurer les conséquences sur le phénomène étudié.

Ces variables sont parfois appelées « variables de décisions », ou simplement « instruments ».

La manipulation des variables exogènes peut permettre au modélisateur de répondre à des questions du type :

- Que se passe-t-il si par hasard... ?
- Que se passe-t-il si l'Etat ou un autre agent économique décide de ... ?

Le deuxième type de questions peut, par approximations successives, donner la solution du problème inverse : quelle décision prendre pour obtenir tel ou tel résultat ? On peut également combiner les deux approches, en observant les conséquences d'une évolution incontrôlée, puis en la complétant par la recherche de la politique qui permettra d'y faire face.

1.2.2.1.2. Variables de flux et de stock

Les variables économiques peuvent regrouper également en deux grandes catégories : les stocks et les flux.

- Un stock est une quantité mesurée en un point du temps. Il n' a donc pas de dimension temporelle et peut être mesuré à n'importe quel moment donné du temps.
- Un flux est une quantité mesurée par unité de temps. Mesuré entre deux périodes, le flux représente la variation de stock au cours d'une période donnée.

L'exemple le plus cité pour différencier le flux du stock est celui de la baignoire. En effet, la quantité d'eau qu'elle contient est un stock : c'est la quantité d'eau qui se trouve dans la baignoire à tout moment donné. Tandis que la quantité d'eau qui coule du robinet est un flux : c'est la quantité d'eau qui s'ajoute dans la baignoire par unité de temps.

Lors de l'élaboration des modèles économiques, il s'avère souvent utile de déterminer si les variables en cause sont des stocks ou des flux, et quelles sont les éventuelles relations entre eux.

Citons quelques exemples de stocks et des flux liés les uns des autres.

- Le capital fixe ou le capital circulant sont les stocks des biens existant à une date fixe. Alors que l'investissement est une variable de flux, car il représente la différence entre le stock du capital de deux périodes données :

$$I_t = K_t - K_{t-1}$$

1.1

où :

I_t , est le niveau d'investissement au temps t ;

K_t , le stock du capital au temps t ;

K_{t-1} , le stock du capital au temps $t-1$.

- La richesse ou les créances détenues par un agent économique et la masse monétaire dans une économie constituent des stocks. Tandis que son revenu et ses dépenses sont mesurés sur une période, mois ou année, et constituent des variables de flux.
- La dette publique est un stock, alors que le déficit budgétaire est un flux.
- Le nombre de chômeurs est un stock ; le nombre de personnes qui perdent leur emploi est un flux.

1.2.2.1.3. Variable « ex ante » et « ex post »

L'analyse économique se réfère souvent aux valeurs « ex ante » et « ex post » des variables.

- La variable est dite « ex ante » si les valeurs du phénomène considéré sont saisies avant que les réactions économiques sous-jacentes se manifestent (analyse prospective).
- Elle est dite « ex post » si les valeurs liées au phénomène étudié sont captées après la manifestation des réactions économiques (analyse rétrospective).

Signalons ici que la distinction entre la prévision (ex ante) et la réalisation (ex post) permet de mieux interpréter certains équilibres économiques.

1.2.2.1.3. Variable spéciale : le temps

En général, de nombreux facteurs interviennent dans l'explication d'un phénomène économique. Cependant, le manque des données statistiques, la difficulté de saisir quantitativement certaines variables exogènes le concernant limitent leur prise en compte dans le modèle. Pour contourner cette difficulté, on remplace parfois ces différents facteurs par une variable spéciale : le temps.

Le temps intervient donc dans le modèle comme une variable de synthèse. Ainsi, dans l'analyse d'une fonction de demande, il est difficile de saisir les variables qualitatives comme le goût, les habitudes, l'influence des voisins,... tous ces facteurs peuvent apparaître dans la fonction de demande par l'intermédiaire de la variable « temps ».

1.2.2.2. Relations

Les variables intervenant dans un modèle économique sont reliées entre elles par des fonctions mathématiques. La représentation mathématique des relations économiques offre de nombreux avantages, d'une part, elle permet de définir les variables d'une manière exacte et d'établir des hypothèses plus explicites, et d'autre part, elle permet d'obtenir à partir des relations complexes (et parfois même confuses) entre plusieurs variables des conclusions que l'on n'aurait pas pu tirer par une analyse littéraire et / ou graphique.

Les relations entre les variables économiques d'un modèle peuvent être classées en deux grandes catégories : les relations comptables et de définition, et les relations fonctionnelles.

1.2.2.2.1. Relations comptable et de définition

- Les *relations comptables* permettent d'assurer la cohérence du modèle en décrivant les équilibres nécessaires entre les ressources et les emplois pour chaque catégorie d'agent et d'opérations.

Exemple :

Revenu des ménages = consommation + épargne – impôts (des ménages).

- Les *relations de définition ou d'identité* établissent des égalités vraies à partir des définitions des variables considérées.

Exemple :

Recette totale (RT) = quantités vendues (Q) x prix unitaire (P)

- Les *relations d'équilibre expriment* « ex post » la nécessaire égalité des ressources et des emplois, de l'offre et de la demande,... en analyse économique, on dit qu'un système est en équilibre si toutes les variables significatives qu'il contient n'indiquent aucune tendance à se modifier.

Exemples :

Ex.1 : $Y = C + I + X - M$ (Équilibre du revenu sous l'optique dépenses)

Ex. 2 : $Q_d = Q_o$ (Équilibre du marché)

1.2.2.2.2. Relations fonctionnelles

Les équations fonctionnelles constituent véritablement le cœur d'un modèle théorique ou quantitatif. Elles énoncent des propositions que l'on peut tester empiriquement. L'on distingue :

- Relations de comportement** qui décrivent la manière dont se fondent les décisions des agents économiques. Le choix des relations de comportement se fonde à la fois sur des considérations théoriques et/ou sur l'exploitation économique directe des données statistiques.

Exemple : La consommation privée des ménages (C_t) peut être liée à leur revenu disponible (Y_{dt}) par une formule simple :

$$C_t = a + bY_{dt} \quad \text{avec } a \text{ et } b : \text{paramètres réels.}$$

- Relations techniques** schématisent les conditions techniques qui caractérisent un phénomène économique.

Exemple : La fonction de production Cobb – Douglas

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

où : Q est le niveau de production ;

K, facteur capital ;

L, facteur travail ;

A, α , β : paramètres réels.

Cette relation décrit la liaison technique qui existe entre les quantités utilisées des différents facteurs de production et la quantité maximale du bien à produire.

- Relations institutionnelles** décrivent les règles administratives propres à un pays et concernent principalement les problèmes de fiscalité. C'est ainsi que les impôts sont projetés par l'application des

taux d'imposition sur une assiette fiscale qui sert de base à l'imposition :

- l'impôt sur le revenu des ménages dépend du taux d'imposition et du revenu des ménages ;
- l'impôt sur les sociétés est fonction du taux d'imposition et de l'excédent brut d'exploitation ;
- les intérêts versés par l'Etat dépendent du taux d'intérêt et du stock des titres publics émis.

Remarques :

En général, on désigne toutes les trois relations par l'expression « relations fonctionnelles ».

I.2.2.3. Paramètres

Toute équation exprime une relation pondérée entre les variables. Les facteurs de pondération repris dans une équation sont des paramètres. Ils sont considérés comme des non variables ayant pour rôle de relier entre elles les variables d'une équation.

On distingue deux types des paramètres :

- les paramètres fixés a priori par le modélisateur ;
- les paramètres estimés par référence au passé. En effet, partant d'une relation fonctionnelle connue mais qui comporte des paramètres inconnus, on cherche les valeurs de ceux-ci qui fournissent la formulation la plus proche de la réalité passée. On applique pour ce faire la méthode économétrique.

La distinction entre les deux types des paramètres n'est pas très nette, car l'échec des tentatives d'estimation à fournir une valeur cohérente économiquement peut conduire à fixer a priori les valeurs de certains paramètres.

Les paramètres fixés a priori ou estimés sur base des observations passées peuvent prendre la forme des propensions, d'élasticités ou des coefficients techniques.

I.2.2.3.1. Propensions et productivités

a. Propensions

On distingue les propension moyennes et marginales.

- Par « marginal », on entend « additionnel ». Ainsi, la propension marginale à consommer correspondant à la fraction du revenu additionnel qui est consommée. Elle se note :

$$Pmc = \frac{\Delta C}{\Delta Y}$$

où : ΔC représente la variation des quantités consommées ;
 ΔY , la variation du revenu.

Si les variations considérées sont infinitésimales, la formule devient :

$$P_{mc} = \frac{dC}{dY}$$

- Par contre, la propension moyenne à consommer est le ratio de la consommation sur le revenu disponible. Donc :

$$PMC = \frac{C}{Y}$$

b. Productivités

Le concept de productivité se réfère principalement aux activités liées à la production. Ainsi, la productivité moyenne du facteur X (capital, travail, ...) est le rapport de la production (Q) sur la quantité du facteur X dans le processus de production :

$$PM_X = \frac{Q}{X}$$

où : PM_X désigne la productivité moyenne du facteur X ;
 Q, le niveau total de la production ;
 X, le facteur de production considéré.

La productivité marginale du facteur X se définit comme la variation de la production due à l'utilisation d'une unité supplémentaire du facteur X. Donc :

$$Pm_X = \frac{\Delta Q}{\Delta X}$$

où : Pm_X : productivité marginale du facteur X ;
 ΔQ : variation de la production ;
 ΔX : variation du facteur de production X.

Pour des variations infinitésimales, on a :

$$Pm_X = \frac{dQ}{dX}$$

ou, d'une manière générale :

$$Pm_{X_i} = \frac{\delta Q}{\delta X_i}$$

(i = 1, 2, ..., n)

où : Pm_X est la productivité marginale du ième facteur utilisé dans le processus de production ; δ , opérateur de la dérivée partielle.

1.2.2.3.2. Elasticités

Il est parfois intéressant pour un économiste ou un gestionnaire de disposer d'une mesure de la sensibilité d'un phénomène suite à une variation donnée de n'importe quel choc externe. Bien que l'analyse marginale permet aussi de déterminer cette sensibilité, elle présente tout de même le grand défaut d'être influencée par les unités dans lesquelles les grandeurs économiques considérées sont mesurées.

En effet, si les quantités demandées (Q) sont exprimées en unités physiques (litres, kilogrammes, tonnes, ...) et les prix (p) en unités monétaires (Franc, Dollar, Euro,...), alors le rapport $\frac{\Delta Q}{\Delta X}$ aura une dimension du type : Litre/Franc, Kg/Dollar, Tonnes/Euro, etc.

Pour remédier à cet inconvénient, on recourt à une mesure qui n'est pas influencée par les unités des grandeurs utilisées, appelée « élasticité ».

L'élasticité d'une variable Y par rapport à une autre variable X, à laquelle elle est liée par une relation de dépendance quelconque, est le rapport des variations relatives de Y et X :

$$\epsilon_{y/x} = \frac{\frac{\Delta Y}{Y}}{\frac{\Delta X}{X}} = \frac{\Delta Y}{\Delta X} \frac{X}{Y}$$

En considérant des variations infinitésimales, la formule devient :

$$\epsilon_{y/x} = \frac{\delta Y}{\delta X} \frac{X}{Y}$$

ou encore :

$$\epsilon_{y/x} = \frac{d \ln Y}{d \ln X}$$

où ln représente le symbole du logarithme népérien.

1.2.2.3.3. Coefficients techniques

Certains modèles macroéconomiques, notamment les modèles input-output, utilisent des paramètres définis sur base des contraintes techniques. Dans le cas des modèles input-output, par exemple, l'hypothèse fondamentale postule que le rapport des consommations intermédiaires à la production totale de la branche considérée est constant. En vertu de cette hypothèse de constance des coefficients techniques, énoncée par W. LEONTIEF, on peut écrire :

$$a_{ij} = \frac{Q_{ij}}{Q_j}$$

où : les a_{ij} représentent les coefficients technologiques ;

Q_{ij} , la consommation intermédiaire de la branche j en produit i ;

Q_j , la production de la branche j.

1.3. Méthodologie de l'économetrie

Pour analyser un problème économique, l'économetre adopte le plus souvent la méthodologie classique qui domine encore la recherche empirique en Economie et dans d'autres Sciences sociales.

La méthodologie classique propose les étapes suivantes :

- référence à une théorie ;
- spécification du modèle mathématique ;
- spécification du modèle statistique ou économétrique ;
- sélection des variables et des données statistiques ;
- estimation des paramètres du modèle économétrique ;
- validation ou diagnostic du modèle ;
- utilisation du modèle estimé.

1.1.1. Référence à une théorie

Le premier pas décisif lorsqu'on cherche à comprendre ou expliquer empiriquement un phénomène économique est de construire un modèle théorique composé d'une ou de plusieurs relations mathématiques définies sous forme générale au travers d'hypothèses théoriques auxquelles le modèle fait référence. En d'autres termes, la réalisation de travaux économétriques suppose la connaissance préalable des disciplines économiques en jeu, puisqu'elles suggèrent le type de relation à vérifier sur les données réelles observées.

Dans la pratique, l'étape de construction du modèle théorique n'est pas toujours aussi simple.

La question fondamentale est celle de savoir si une théorie économique existante est suffisamment correcte pour être utilisée en tant qu'outil d'analyse. C'est pourquoi, certains modélisateurs ne formulent jamais un modèle théorique de manière définitive, mais préfèrent considérer un ensemble des liaisons en laissant le soin à l'économétrie d'en trancher (économétrie sans théorie).

1.1.2. Spécification du modèle mathématique

La spécification consiste à donner une « forme fonctionnelle » au modèle théorique construit dans la première étape. L'objectif est de vérifier l'adéquation d'un modèle à la réalité observée et de mesurer le taux de réaction des phénomènes expliqués aux variations des phénomènes explicatifs. Pour confronter efficacement le modèle et les données observées, il convient d'exprimer ce dernier sous une forme manipulable.

Cette partie essentielle de la modélisation exige soit une connaissance approfondie de la théorie économique, soit un recours aux résultats économétriques obtenus par d'autres chercheurs ou enfin un « flair intuitif » pour un meilleur choix de relations fonctionnelles.

Les possibilités de spécification sont nombreuses :

1.1.2.1. Choix d'une relation linéaire

Le choix le plus simple est celui d'une relation linéaire. Il se justifie quand on peut raisonnablement supposer que les dérivées partielles de la variable explicative ne sont pas fonction des niveaux atteints par ces variables explicatives. Cette hypothèse signifie que la variation de la variation dépendante, suite à une variation d'une unité de l'une des variables explicatives, est toujours la même quels que soient les niveaux déjà atteints par celles-ci.

Exemple : Considérons le modèle linéaire suivant :

$$Y = \alpha + \beta X$$

avec α et β les paramètres.

La dérivée de Y par rapport à X est donc : $\frac{dY}{dX} = \beta$.

Le paramètre (coefficient) β mesure l'importance de la variation de Y quand X augmente d'une unité. Donc, si X augmente d'une unité, la variable dépendante Y varie de β unités.

1.1.2.2. Choix d'une relation non linéaire

Quoique très commode, la linéarité peut ne pas correspondre de façon adéquate à la relation traitée. Il est parfois irréaliste de supposer que la variation de la variable expliquée est toujours la même, suite à une variation d'une unité de la variable explicative, quels que soient les niveaux déjà atteints par cette dernière. D'où, le recours aux relations formalisées sous la forme d'équations non linéaires.

Exemple : Soit le modèle non linéaire :

$$Y = \ln \alpha + \beta \ln X$$

avec $\frac{dY}{dX} = \frac{\beta}{X}$.

Cette spécification implique une dérivée première de Y par rapport à X qui décroît avec le niveau de X. Autrement dit, au fur et à mesure que X augmente, l'augmentation de Y devient de plus en plus faibles (cf. loi des rendements marginaux non proportionnels).

1.1.3. Spécification du modèle économétrique

Le modèle purement mathématique d'un phénomène économique est d'un intérêt limité pour l'économetre parce qu'il suppose une relation précise ou déterministe entre la variable expliquée et les variables explicatives. Mais les relations entre les variables économiques ne sont pas toujours exactes.

En économétrie, on suppose généralement que les variables économiques sont aléatoires. En d'autres termes, on considère que la valeur observée d'un phénomène économique est en partie due au hasard.

C'est la réalisation d'une variable aléatoire correspondante susceptible de produire d'autres réalisations si l'on répète l'expérience.

Le hasard détermine en partie les réalisations effectivement observées des variables économiques et les résultats auraient pu être différents. Les probabilités d'obtenir telle ou telle valeur effectivement réalisée sont déterminées par les distributions statistiques des variables. Les relations économiques supposées par la théorie économique imposent des liaisons entre ces distributions.

Pour tenir compte des relations imprécises entre les variables, l'économétrie modifie le modèle purement mathématique de la façon suivante :

$$Y = \alpha + \beta X + u$$

où u , appelée perturbation ou erreur, est une variable aléatoire (stochastique) possédant des propriétés probabilistes bien définies. L'équation ci-dessus est un exemple de modèle économétrique.

1.1.4. Sélection et mesure des variables

Le succès de toute analyse économétrique dépend, en dernier ressort, des données représentatives des phénomènes économiques étudiés. Les principaux types des données sont les suivants :

- **Les séries temporelles (séries chronologiques, chroniques ou time series)**

Il s'agit des données observées à des intervalles de temps réguliers (données hebdomadaires, mensuelles, trimestrielles, annuelles, ...). A titre d'exemple, la série des dépenses mensuelles en publicité des sociétés de communication de mars 2000 à janvier 2010 en RDC.

- **Les séries en coupe transversale (coupe instantanée ou cross section)**

Ce sont des données observées au même moment du temps pour un groupe spécifique d'individus. On peut citer l'exemple de consommation des ménages des cadres politiques à Kinshasa en 2009.

- **Les données de panel**

Ce sont des données qui concernent un groupe spécifique d'individus et qui sont observées à des intervalles de temps réguliers. Il s'agit d'un type particulier des séries mixtes dans lesquelles les données représentent les valeurs prises par un même échantillon d'individus sur une période de temps. A titre d'exemple, on peut citer les dépenses de consommation des ménages

des parlementaires nationaux à Kinshasa de 2007 à 2009. Comme on peut le constater, les données de panel ont ainsi une double dimension : individuelle et temporelle.

1.1.5. Choix de décalages temporels

Les variables sur lesquelles s'appuient les modèles économiques ont généralement une dimension temporelle et sont alors connues de manières discrète, le plus souvent avec une périodicité constantes : séries annuelles, mensuelles, ... (séries chronologiques).

Soit le modèle de consommation des ménages kinois :

$$C = a + cY + u$$

où : C représente les dépenses de consommation des ménages kinois ; Y, leur revenu disponible ; u, le terme d'erreur ; a et c les paramètres constants.

Si l'on suppose que le revenu de la date t explique la consommation à la même date, on a alors :

$$C_t = a + cY_t + u_t$$

où t désigne le temps.

Il existe cependant des modèles sans dimension temporelle. Il s'agit le plus souvent des modèles microéconomiques, décrivant des comportements individuels d'entreprises ou des ménages : la dimension sera alors celle des individus dans une population (coupes transversales). Ainsi, la fonction de consommation des ménages des parlementaires nationaux à Kinshasa s'écrit :

$$C_i = a + cY_i + u_i$$

où i désigne l'individu, c'est-à-dire le $i^{\text{ème}}$ parlementaire national.

Dans le cadre des modèles spécifiés en séries temporelles, les relations entre les variables ne sont pas toujours synchrones mais peuvent être décalées dans le temps. D'où, l'on distingue deux types des modèles : les modèles statiques et les modèles dynamiques.

- Les *modèles statiques* ne font intervenir, pour la détermination de l'équilibre associé à une période donnée que les variables de cette période. Ils correspondent à la formulation suivante :

$$F(Y_t, A, X_t, u_t) = 0 \quad 1.2$$

où : Y_t est la variable endogène à la période t ;

X_t , la variable exogène à la période t ;

u_t , le terme d'erreur à la période t ;

A, le vecteur des paramètres constants.

La fonction de consommation des ménages évoquée plus haut : $C_t = a + cY_t + u_t$ constitue une illustration d'un modèle statique où le revenu à la date t explique le niveau de consommation de la même date.

- Au contraire, les *modèles dynamiques* utilisent, pour déterminer l'équilibre d'une période, des variables d'autres périodes.

La justification en pourra être :

- *Théorique* : certains agents seront supposés intégrer dans leur comportement leurs constatations passées.
Ainsi, les consommateurs tiendront comptes de leur niveau de consommation passé pour déterminer la consommation actuelle ;
- *Institutionnel* : l'impôt sur le revenu payé par les ménages sera basé sur leur revenu de la période précédente ;
- *Mécanique* : pour une variable donnée, le passage du niveau instantané au taux de croissance annuel nécessite la prise en compte du niveau précédent.

On constate que chacune de ces justifications suppose, sauf connaissance du futur par les agents, que les influences proviennent des périodes précédentes : on parlera alors d'influences retardées.

La formulation devient donc :

$$F(Y_t, Y_{t-1}, \dots, Y_{t-k}, X_t, X_{t-1}, \dots, X_{t-p}, A, u_t) = 0 \quad 1.3$$

où : Y_t est la variable endogène courante ;

Y_{t-1}, \dots, Y_{t-k} représentent les variables endogènes retardées (décalées ou *laggées*) ;

X_t , la variable exogène courante ;

X_{t-1}, \dots, X_{t-p} : les variables exogènes décalées (retardées ou *laggées*) ;

u_t , le terme d'erreur à la période t ;

A , le vecteur des paramètres constants.

Il convient de souligner ici que, dans les modèles dynamiques, les *variables prédéterminées* comprennent à la fois les variables endogènes décalées, les variables exogènes courantes et les variables exogènes décalées.

Exemple d'un modèle dynamique simple

Soit le modèle de consommation dynamique suivant :

$$C_t = a + cY_t + \phi C_{t-1} + u_t$$

Dans ce modèle, le niveau de consommation à la date t est expliquée non seulement par le revenu de la même date, mais aussi par le niveau de

consommation passée (c'est-à-dire la consommation à la date $t-1$). La variable explicative C_{t-1} est donc la variable endogène décalée. Le coefficient ϕ représente le degré d'inertie de la consommation. Notons que ϕ doit être impérativement inférieur à un ($\phi < 1$). Plus ϕ est proche de 1, plus le degré d'inertie de la consommation est important. Dans ce cas, la consommation passée exerce une forte influence sur la consommation courante, on parle aussi de la persistance.

1.1.6. Validation ou diagnostic du modèle

Cette étape consiste avant tout à déterminer les signes et la grandeur des paramètres théoriquement attendus. Elle permet aussi d'effectuer l'évaluation des résultats des calculs en testant si les estimations des paramètres sont économiquement et statistiquement valides. Elle s'intéresse enfin de la stabilité des coefficients estimés, caractères indispensables pour une meilleure prévision.

1.1.7. Estimation des paramètres du modèle économétrique

Puisque le modèle est spécifié et que les données sont disponibles, la tâche suivante consistera à estimer les paramètres de ce modèle. Cette estimation donne un contenu empirique au modèle.

1.1.8. Applications des modèles

L'utilisation la plus naturelle d'un modèle est de prévoir l'avenir économique. On va distinguer deux types de prévisions : les scénarios et les variantes.

- a) Dans un scénario, on s'intéresse aux résultats dans l'absolu, c'est-à-dire que l'on associe à un ensemble d'hypothèses évaluées dans une évolution future de l'équilibre économique. On cherchera ainsi à obtenir :
 - Soit des prévisions sur la base des hypothèses les plus probables ;
 - Soit une évaluation du champ des possibles ;
 - Soit à déterminer les hypothèses permettant d'atteindre certains résultats économiques.
- b) Dans une variante, on partira d'une simulation de base (dites souvent « compte de référence »), ou d'une simulation sur période historique, et on mesurera la sensibilité de l'équilibre économique à une modification des hypothèses. Il s'agit donc de la comparaison de deux trajectoires économiques.

Les modèles économétriques interviennent également à l'analyse des évolutions passées des phénomènes économiques.

Enfin, les modèles économétriques estimés peuvent être utilisés à des fins de contrôle ou de politique économique. Ainsi, à l'aide d'un « policy mix », fiscal et monétaire, les pouvoirs publics peuvent avoir en main la variable

contrôlée X pour assurer le niveau désiré de la variable objectif Y.

1.2. ROLE DE L'ECONOMETRIE

L'économétrie s'articule autour de deux points suivants : validation de la théorie et outil d'investigation.

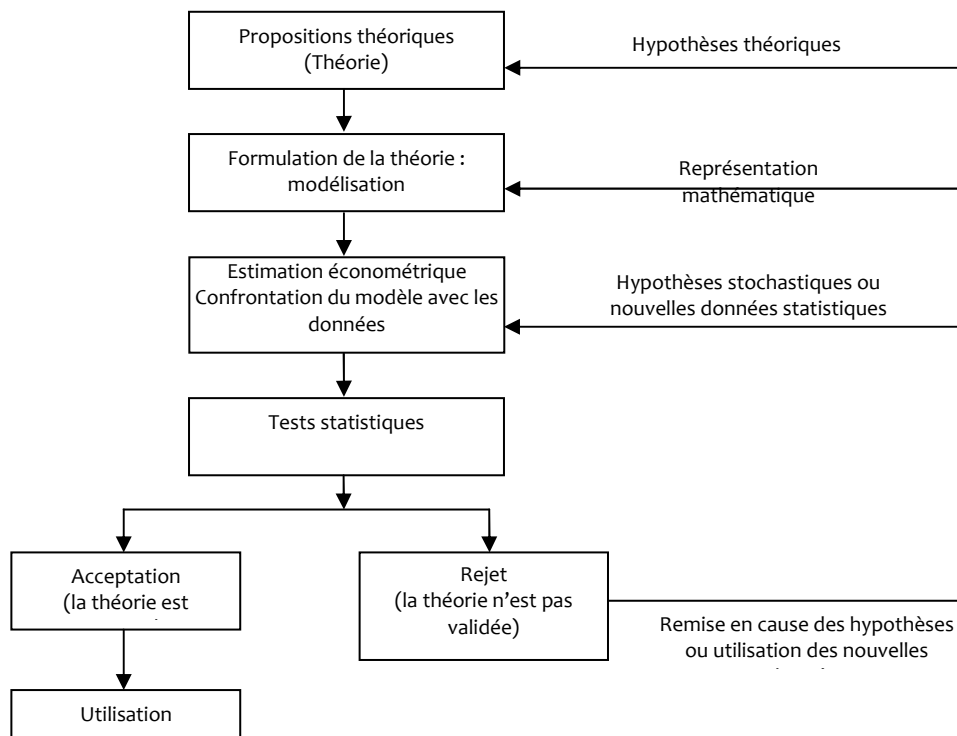
1.2.1. Econométrie comme validation de la théorie

L'économétrie est un outil à la disposition de l'économiste lui permettant d'infirmer ou de confirmer les théories qu'il construit. Le théoricien postule des relations, et l'application des méthodes économétriques fournit des estimations sur la valeur de coefficients ainsi que la précision attendue.

Il serait erroné de penser que les modèles économétrique et les théories économiques forment deux mondes séparés. La modélisation théorique est un préalable obligé. La quantification économétrique s'appuie en général sur les modèles théoriques. En effet, les modèles quantitatifs reprennent et spécifient les propositions théoriques en vue de vérifier leur adéquation aux données statistiques disponibles.

Le schéma ci-après illustre la démarche de validation de la théorie à l'aide de l'économétrie.

Schéma 1 : Processus de modélisation et économétrie



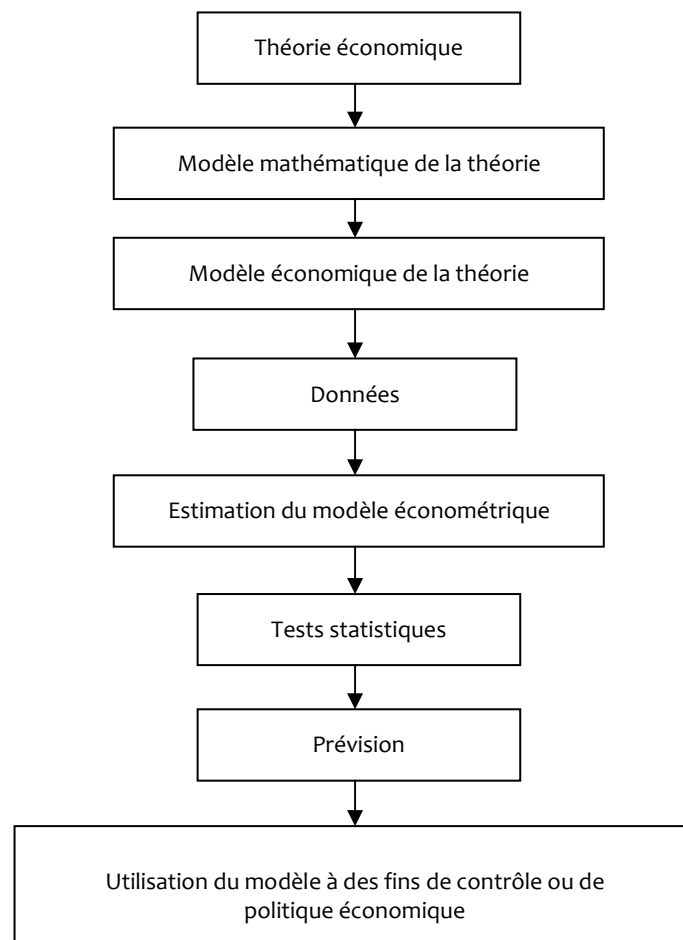
1.2.2. Économétrie comme outil d'investigation

L'économétrie n'est pas seulement un système de validation, mais également un outil d'analyse.

En effet, l'économétrie apporter une aide notamment à la modélisation, à la réflexion théorique ou à l'action économique par :

- la mise en évidence des relations entre les variables qui n'étaient pas a priori évidentes ou pressenties ;
- l'induction statistique ou l'inférence statistique permettant de tester la qualité des paramètres estimés ;
- l'analyse des évolutions passées ;
- la prévision de l'évolution économique ;
- les études en variantes ;
- la simulation qui mesure l'impact de la modification d'une variable sur une autre.

Schéma n°1 : Anatomie de la modélisation économétrique



CHAPITRE II ETUDE DE LA CORRELATION

2.1. PRESENTATION GENERALE

Face aux phénomènes économiques représentés par des grandeurs statistiques quantitatives, on peut chercher à savoir :

- Si les variables considérées sont-elles liées ?
- Comment mesurer cette liaison ?

Lorsque les variables retenues ont une évolution commune, nous disons qu'elles sont «corrélées». La *corrélation simple* mesure le degré de liaison existant entre deux phénomènes représentés par des variables X et Y. Pour mesurer la relation entre trois ou plusieurs variables, on utilise la notion de corrélation multiple.

On parle de la corrélation linéaire, lorsque tous les points du couple de valeurs des deux variables (X,Y) semblent alignés sur une droite, tandis qu'une corrélation est dite non linéaire lorsque les points du couple de valeurs se trouvent sur une même courbe d'allure quelconque.

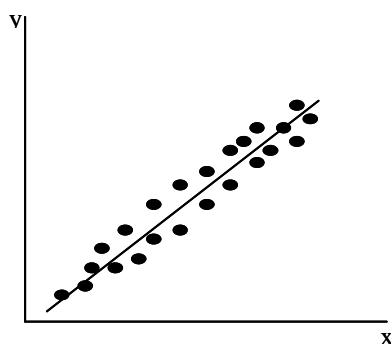
Deux variables X et Y peuvent être :

- en corrélation positive si l'on constate une augmentation (ou diminution, ou constance) simultanée des valeurs des deux variables ;
- en corrélation négative, lorsque les valeurs de l'une augmentent (resp. diminuent), les valeurs de l'autre diminuent (resp. augmentent) ;
- non corrélées, s'il n'y a aucune relation de dépendance entre les valeurs des deux variables.

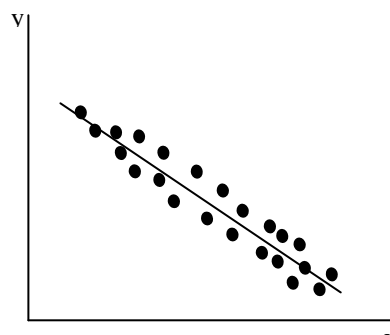
En combinant les critères de linéarité et de corrélation, on peut dégager le tableau et les graphiques suivants.

Tableau 1 - Linéarité et corrélation

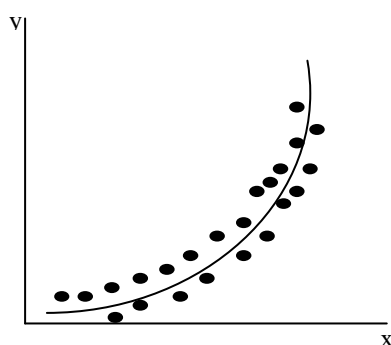
	Corrélation Positive	Corrélation négative	Absence de corrélation
Relation linéaire	Graphe1	Graphe2	Graphe5
Relation non linéaire	Graphe3	Graphe4	Graphe5



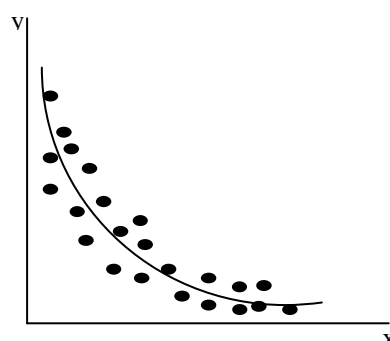
Graphe 1



Graphe 1



Graphe 3



Graphe 3

2.2. MESURE ET LIMITE DU COEFFICIENT DE CORRELATION

2.2.1. Le coefficient de corrélation linéaire

Le coefficient de corrélation linéaire simple entre X et Y, noté $r_{x,y}$, quantifie la liaison entre les deux variables de manière à mettre en évidence le sens de la liaison et son intensité.

Formellement, le coefficient de corrélation linéaire entre les deux variables X et Y, est égal à :

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad 2.1$$

Avec : $\text{Cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n}$: covariance entre X et Y ;

$$\sigma_X = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}} \text{ et } \sigma_Y = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n}} = \text{écart type de X et écart type de Y}$$

;

n : nombre d'observations ;

i : est le numéro de l'observation ou l'année.

Notons que :

- Si i est une date, on parle de données chronologiques ;
- Si i représente un individu statistique (un ménage, une entreprise, ...), on parle de données transversales.

En développant la formule (2.1), il vient :

$$\rho_{X,Y} = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\sqrt{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \sqrt{n \sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2}} \quad 2.2$$

La relation (2.2) peut aussi s'écrire comme :

$$\rho_{X,Y} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad 2.3$$

où $x_i = X_i - \bar{X}$ et $y_i = Y_i - \bar{Y}$ (les écarts par rapport à la moyenne).

Par construction, ce coefficient reste compris entre -1 et 1 :

- proche de 1 , les variables sont corrélées positivement ;
- proche de -1 , les variables sont corrélées négativement ;
- proche de 0 , les variables ne sont pas corrélées.

Dans la pratique, ce coefficient est rarement très proche de l'une de ces trois limites et il est donc difficile de proposer une interprétation fiable à la simple lecture de ce coefficient. Ceci est surtout vrai en économie où les variables sont toutes plus ou moins liées entre elles. De plus, il n'est calculé qu'à partir d'un échantillon (de taille limitée) issu de la population. D'où nécessité d'inférer les résultats obtenus sur la population concernée.

La théorie des tests statistiques nous permet de lever cette indétermination. En effet, concernant la population totale, une hypothèse de nullité du coefficient de corrélation est formulée, et l'on se pose la question de savoir dans quelle mesure cette hypothèse est confirmée ou infirmée par des données observées.

Soit à tester l'hypothèse :

$$\begin{cases} H_0: \Gamma_{X,Y} = 0 \\ H_1: \Gamma_{X,Y} \neq 0 \end{cases}$$

Sous l'hypothèse H_0 , la statistique du test est donc :

$$t_{cal} = \frac{|\rho_{X,Y}|}{\sqrt{\frac{(1 - \rho_{X,Y}^2)}{n - 2}}} \sim t_{\frac{\alpha}{2}; (n-2)} \quad 2.4$$

Si $|t_{\text{cal}}| \geq t_{\frac{\alpha}{2}; (n-2)}$, valeur lue dans une table de Student au seuil $\alpha = 0.05$, (soit $\alpha = 5\%$) à $n - 2$ degrés de liberté, nous rejetons l'hypothèse H_0 . Par conséquent, le coefficient de corrélation est significativement différent de 0 ; dans le cas contraire, l'hypothèse d'un coefficient de corrélation nul est acceptée.

Remarque :

Si le nombre d'observations n est supérieur à 30, on peut approximer la loi de Student par une loi normale, soit $t_{\frac{\alpha}{2}} \approx 1.96$.

2.2.2. Limites de la notion de corrélation

2.2.2.1. La relation testée est linéaire

Les formules (2.1), (2.2) ou (2.3) s'appliquent valablement pour des corrélations linéaires entre variables. Leur utilisation donne des résultats aberrants si les variables considérées ne sont liées par une relation fonctionnelle linéaire. Ainsi, deux variables X et Y auront un coefficient de corrélation égal à zéro ($\rho_{X,Y} = 0$) si les deux variables sont liées entre elles :

- Soit par une fonction circulaire du type : $X^2 + Y^2 = c$ (où c est constante quelconque) ;
- Soit par une fonction non linéaire du type : $Y = \cos X$.
-

Il découle de ces deux cas d'exemple qu'une faible corrélation n'implique pas nécessairement l'absence de relation mais parfois simplement l'absence de relation linéaire.

Pour pallier cette limite, on est éventuellement amené à réaliser des transformations sur certaines variables avant de calculer des corrélations. On parle alors de *linéarisation*.

2.2.2.2. Corrélation n'est pas causalité

L'existence d'une corrélation même élevée n'implique pas nécessairement l'existence d'une relation de cause à effet entre les deux variables considérées. En d'autres termes, un coefficient de corrélation significativement différent de zéro n'implique pas une liaison d'ordre économique ou autre. Nous appelons *corrélation fortuite* ce type de corrélation que rien ne peut expliquer.

Souvent, en effet, les corrélations observées sont dues au fait que les variables étudiées sont toutes deux soumises à des influences communes, qui peuvent en modifier simultanément les valeurs, soit dans le même sens (corrélation positive), soit en sens opposés (corrélation négative).

L'exemple le plus fameux concerne la forte corrélation existante entre le nombre d'entrées à la Fikin et le chiffre d'affaires réalisé par le vendeur de

glaces et autres boissons rafraîchissantes. Cela ne signifie pas qu'il existe une relation entre les deux variables, mais qu'une troisième variable, représentée ici par le temps qu'il fait, c'est-à-dire l'état de la météo du jour, influence conjointement les deux phénomènes.

Exercice résolu

Le tableau ci-après reprend les données statistiques relatives au Produit intérieur brut (PIB) et à la consommation privée de la RDC pour la période allant de 1980 à 1996¹.

Année	Consommation	PIB
1980	545,0	741,1
1981	559,0	748,1
1982	579,0	744,7
1983	591,0	755,2
1984	610,0	797,1
1985	637,6	800,8
1986	656,0	838,6
1987	678,3	861,0
1988	694,4	865,9
1989	701,9	854,1
1990	749,0	798,0
1991	540,8	730,0
1992	502,8	654,3
1993	425,9	565,8
1994	456,7	543,9
1995	464,1	547,7
1996	460,3	541,8

Calculer le coefficient de corrélation simple entre la consommation privée (X)

¹ Notes :

- 1) Consommation des ménages aux prix de 1987 (en milliards de Zaïres);
- 2) PIB aux prix de 1987 (en milliards de Zaïres)
- 3) Taux d'inflation

Source : Iyashi Ile Mbula, Estimation d'un modèle de consommation des ménages. Cas de la RDC de 1980 à 1996, Mémoire de Licence, Faculté d'Administration des Affaires et Sciences Economiques (FASE), Université Protestante au Congo, Octobre 2000.

et le PIB (Y), et tester l'hypothèse de la nullité de ce coefficient au seuil $\alpha = 0,05$.

Solution

Année	Y	X	$Y_i - \bar{Y}$	$X_i - \bar{X}$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})^2$
1980	545	741,1	-34,52	12,39	-427,61	1191,47	153,47
1981	559	748,1	-20,52	19,39	-397,80	420,97	375,90
1982	579	744,7	-0,52	15,99	-8,28	0,27	255,62
1983	591	755,2	11,48	26,49	304,15	131,84	701,63
1984	610	797,1	30,48	68,39	2084,63	929,17	4676,95
1985	637,6	800,8	58,08	72,09	4187,05	3373,56	5196,71
1986	656	838,6	76,48	109,89	8404,51	5849,55	12075,42
1987	678,3	861	98,78	132,29	13067,74	9757,95	17500,18
1988	694,4	865,9	114,88	137,19	15760,51	13197,96	18820,61
1989	701,9	854,1	122,38	125,39	15345,31	14977,44	15722,21
1990	749	798	169,48	69,29	11743,13	28724,27	4800,86
1991	540,8	730	-38,72	1,29	-49,88	1499,06	1,66
1992	502,8	654,3	-76,72	-74,41	5708,70	5885,60	5537,11
1993	425,9	565,8	-153,62	-162,91	25026,12	23598,38	26540,24
1994	456,7	543,9	-122,82	-184,81	22698,15	15084,17	34155,39
1995	464,1	547,7	-115,42	-181,01	20891,95	13321,23	32765,26
1996	460,3	541,8	-119,22	-186,91	22283,18	14212,85	34936,01
Somme	9851,8	12388,1	0	0	166621,57	152155,74	214215,24

$$\text{Moyenne de X} = \frac{12388,1}{17} = 728,7118$$

$$\text{Moyenne de Y} = \frac{9851,8}{17} = 579,5176$$

Le coefficient de corrélation entre la consommation (Y) et le PIB (X) est égal à :

$$\begin{aligned} \rho_{X,Y} &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \\ &= \frac{166621,57}{\sqrt{214215,24} \sqrt{152155,74}} = \frac{166621,57}{(462,8339) \times (390,0715)} = 0,9229. \end{aligned}$$

Donc $\rho_{X,Y} = 92 \%$.

Le t de Student empirique est donc :

$$\begin{aligned}
 t_{\text{cal}} &= \frac{|\rho_{x,y}|}{\sqrt{\frac{(1 - \rho_{x,y}^2)}{n - 2}}} = \frac{0,9229}{\sqrt{\frac{1 - (0,9229)^2}{17 - 2}}} \\
 &= \frac{0,9229}{0,0994} = 9,2831.
 \end{aligned}$$

La valeur lue dans la table de Student au seuil $\alpha = 0,05$ (soit $\alpha = 5 \%$) à 15 degrés de liberté est égale à : $t_{0,025; 15} = 2,131$.

Comme le t calculé est largement supérieur au t de la table de Student, on rejette l'hypothèse nulle H_0 . Le coefficient de corrélation calculé entre la consommation privée et le PIB en RDC est significativement différent de zéro. Donc, il y a une forte liaison positive entre les deux agrégats.

CHAPITRE 3 MODELE DE REGRESSION SIMPLE

Le présent chapitre constitue une introduction aux techniques de régression. C'est pourquoi, nous allons commencer un modèle de régression simple où une variable endogène représentant l'évolution du phénomène considéré est expliquée par une seule variable exogène.

Une telle démarche, appelée analyse de régression simple ou analyse de régression à deux variables, comporte certes un double avantage. Elle permet avant tout de faire ressortir dans le contexte le plus simple possible, un certain nombre de questions fondamentales. Ensuite, elle facilite le développement des concepts et des outils d'analyse qui serviront des matériaux de construction dans les cas plus complexes.

3.1. PRESENTATION DU MODELE

Supposons que, d'après la théorie choisie, le phénomène étudié soit représenté par la relation :

$$Y = f(X) \quad 3.1$$

où Y désigne la variable dépendante (expliquée) ;
X, la variable indépendante (explicative).

La théorie peut aussi nous indiquer certaines contraintes que les paramètres du modèle doivent respecter concernant leur signe, ou concernant leur ordre de grandeur. Dans ce chapitre, nous allons nous limiter à l'étude des modèles de type linéaire sans contrainte sur les paramètres.

Un modèle de régression est dit linéaire lorsque la variable dépendante Y, ou une transformation de Y, peut être exprimée comme une fonction linéaire de X, ou d'une quelconque transformation de X. Ainsi les équations ci-après :

$$Y = \beta_0 + \beta_1 X \quad 3.2$$

$$Y = \alpha + X^\beta \quad 3.3$$

$$Y = \exp \left\{ \beta_0 + \beta_1 \frac{1}{X} \right\} \text{ ou } Y = e^{\beta_0 + \beta_1 \frac{1}{X}} \quad 3.4$$

$$Y = \beta_0 + \beta_1 \frac{1}{X} \quad 3.5$$

$$\frac{1}{Y} = \beta_0 + \beta_1 X \quad 3.6$$

Sont toutes des spécifications linéaires. Elles sont soit linéaires en X et Y (cas de la relation (3.2), soit linéaires après transformation de Y et/ou de X (cas des relations (3.3) à (3.6)).

Par contre, la fonction :

$$(Y - \alpha_1)(X - \alpha_2) = \alpha_3 \quad 3.7$$

Représentant une hyperbole ayant pour asymptotes $Y = \alpha_1$ et $X = \alpha_2$, ne peut être ramenée à une fonction linéaire de transformations de Y et de X .

Exemple :

Soit la fonction de consommation keynésienne : $C = a + cY$
où C : consommation (variable endogène) ;

Y : revenu (variable exogène) ;
 a : consommation autonome ou incompressible ;
 c : propension marginale à consommer,
(a et c sont les paramètres inconnus du modèle ou encore les coefficients de régression).

Cette fonction de consommation peut être spécifiée de deux façons :

- a) En série temporelle : les variables C et Y représentent des grandeurs économiques observées à intervalles de temps réguliers, par exemple la consommation et le revenu de 1990 à 2009 en RDC.
Le modèle s'écrit alors : $C_t = a + cY_t$
($t = 1990, \dots, 2009$)
où C_t : consommation au temps t ;
 Y_t : revenu au temps t .
- b) En coupe transversale : les variables C et Y représentent des grandeurs économiques au même instant mais concernant plusieurs individus, par exemple la consommation et le revenu donnés sur un échantillon de 20 ménages de Kinshasa en 2009.
Le modèle s'écrit alors : $C_i = a + cY_i$
($i = 1, 2, \dots, 20$)
où C_i : est la consommation du ménage i en 2009 ;
 Y_i : le revenu du $i^{\text{ème}}$ ménage en 2009.

3.2. RÔLE DU TERME ALEATOIRE

Le modèle linéaire que nous avons retenu pour l'analyse de la régression simple :

$$Y_t = \beta_0 + \beta_1 X_t \quad 3.8$$

Ne recouvre que partiellement la réalité. En effet, les relations fonctionnelles strictes, ou exactes, fournissent rarement une description adéquate de la dépendance entre les grandeurs économiques. C'est pourquoi nous ajoutons un terme (u) qui synthétise l'ensemble des informations omises dans le modèle mais qui affectent ce dernier.

L'écriture principale du modèle de régression simple devient :

$$Y_t = \beta_0 + \beta_1 X_t + u_t \quad 3.9$$

(avec $t = 1, 2, \dots, n$)

où Y_t : représente la variable endogène au temps t ;

X_t : la variable exogène au temps t ;

u_t : l'erreur ou le terme aléatoire ;

β_0 et β_1 : les paramètres réels non observables.

Plusieurs raisons militent pour la prise en compte du terme d'erreur dans le modèle de régression :

- L'imprécision de la théorie : la théorie, elle existe, peut s'avérer insuffisante pour expliquer le comportement de la variable dépendante.
- L'indisponibilité des données statistiques : le manque des données quantifiables comme la mauvaise qualité des observations peuvent limiter l'utilisation de certaines variables dans le modèle.
- La nature aléatoire du comportement humain : le terme d'erreur peut fort bien refléter cette nature intrinsèquement aléatoire.
- La pauvreté des variables substituées (proxy variables) : le terme d'erreur peut être utilisé pour tenir compte de l'incapacité des variables substituées à remplacer totalement les variables d'origine.
- Le principe de parcimonie : dans le but de rendre le modèle aussi simple que possible tout en maintenant son adéquation avec la théorie, l'économétrie est parfois obligée de ne pas prendre en compte des variables complémentaires.
- Une forme fonctionnelle incorrecte : le terme d'erreur peut être utilisé pour prendre en compte la mauvaise spécification de la relation fonctionnelle entre la variable dépendante et le régresseur (variable explicative).

3.3. NOTION MATRICIELLE DU MODELE

Les n équations qui comportent le modèle linéaire simple (3.9) peuvent s'écrire comme :

$$\begin{cases} Y_1 = \beta_0 + \beta_1 X_1 + u_1 \\ Y_2 = \beta_0 + \beta_1 X_2 + u_2 \\ Y_3 = \beta_0 + \beta_1 X_3 + u_3 \\ \vdots \quad \vdots \quad \vdots \quad \vdots \\ Y_n = \beta_0 + \beta_1 X_n + u_n \end{cases} \quad 3.10$$

Ce système d'équations linéaires peut être réécrit sous la forme :

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{bmatrix} \quad 3.11$$

On obtient la forme matricielle :

$$\begin{matrix} Y & = & X & B & + & U \\ (nx1) & & (nx2) & (2x1) & & (nx1) \end{matrix} \quad 3.12$$

où

$$B_{(2 \times 1)} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}; Y_{(nx1)} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}; U_{(nx1)} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}; X_{(nx2)} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}$$

Faisons remarquer que la première colonne de la matrice X n'est formée que de nombres 1, ce qui est indispensable si on veut tenir compte dans la régression de la constante β_0 .

3.4. SPECIFICATION DES HYPOTHESES

Le modèle linéaire de régression simple s'appuie sur les hypothèses fondamentales suivantes :

- H₁ : Le modèle est linéaire en X_t ou en n'importe quelle transformation de X_t .
- H₂ : Les valeurs de X_t sont observées sans erreur. Donc la variable X_t est non aléatoire. Tandis que Y_t , observée également sans erreur, est aléatoire par l'intermédiaire de u_t .
- H₃ : L'espérance mathématique de l'erreur est nulle.

$$E(u_t) = 0 \quad 3.13$$

Le terme d'erreur pouvant prendre des valeurs négatives et des valeurs positives, l'hypothèse suppose qu'il n'existe pas de biais en faveur des valeurs positives, ni en faveur des valeurs négatives.

La nullité de l'erreur moyenne revient à admettre qu'en moyenne le modèle est correctement spécifié et donc, qu'en moyenne, l'erreur est nulle. Ainsi :

$$\begin{aligned} E(Y_t) &= E(\beta_0 + \beta_1 X_t + u_t) \\ &= \beta_0 + \beta_1 X_t + E(u_t) \\ &= \beta_0 + \beta_1 X_t \end{aligned} \quad 3.14$$

Cette hypothèse implique, sur le plan économique, que le terme d'erreur u_t regroupe l'ensemble des causes aléatoires diverses qui font dévier Y_t de sa valeur théorique ($Y_t = \beta_0 + \beta_1 X_t$) mais dont aucune n'est de poids à déformer la moyenne de u_t vers une valeur systématique différente de zéro.

H_4 : Le terme d'erreur u_t a une variance constante et finie. En effet :

$$\begin{aligned}\text{Var}(u_t) &= E[u_t - E(u_t)]^2 \\ &= E(u_t^2), \text{ car } E(u_t) = 0 \\ &= \sigma_u^2 = \text{constante}, \quad \forall t\end{aligned}\quad 3.15$$

Cette restriction implique que le terme stochastique provient, pour toute la période prise en considération, de la même population ayant la même variabilité. Pour l'économiste, cette restriction l'oblige à accepter que les causes explicatives ignorées dans son modèle influent pratiquement de façon constante sur la variable endogène.

Lorsque cette hypothèse est réalisée, on dit qu'il y a homoscedasticité des erreurs (c'est-à-dire même dispersion ou bien variance égale). Dans le cas contraire où cette hypothèse n'est pas vérifiée, on parle alors de l'hétéroscédasticité des erreurs.

L'hypothèse de l'hétéroscédasticité implique également que la variance de Y_t soit constante et finie. En effet :

$$\begin{aligned}\text{Var}(Y_t) &= E[Y_t - E(Y_t)]^2 \\ &= E[\beta_0 + \beta_1 X_t + u_t - \beta_0 - \beta_1 X_t]^2 \\ &= E(u_t^2) = \sigma_u^2 = \text{constante}\end{aligned}\quad 3.16$$

H_5 : Les erreurs u_t de périodes différentes sont non corrélées, c'est-à-dire indépendantes les une des autres. Donc une erreur à l'instant t n'a pas d'influence sur les erreurs d'autres périodes.

$$\begin{aligned}\text{Cov}(u_i, u_j) &= E\{[u_i - E(u_i)][u_j - E(u_j)]\} \\ &= E(u_i u_j) = 0\end{aligned}\quad 3.17$$

Cette hypothèse est celle de non autocorrélation des erreurs.

H_6 : L'erreur u_t est indépendante de la variable explicative. En effet :

$$\begin{aligned}\text{Cov}(X_t, u_t) &= E\{[X_t - E(X_t)][u_t - E(u_t)]\} \\ &= E\{[X_t - E(X_t)]u_t\} = 0\end{aligned}\quad 3.18$$

Cette hypothèse suppose que chacun des termes d'erreur u_t est non corrélé avec la variable exogène X_t .

H_7 : En vertu du Théorème Central Limite, on suppose que le terme d'erreur u_t suit une loi normale de moyenne 0 et de variance constante σ_u^2 .
Donc :

$$u_t \sim N(0, \sigma_u^2)$$

3.19

Cette hypothèse est justifiée dans la mesure où l'économètre néglige de tenir compte dans son modèle d'un nombre important des variables ou des facteurs explicatifs indépendants.

Remarques

- Les erreurs qui satisfont simultanément les hypothèses d'homoscédasticité et d'absence d'autocorrélation sont parfois appelées erreurs (ou perturbations) sphériques. On parle alors de sphéricité des erreurs. En outre, une série d'erreurs vérifiant les deux hypothèses précitées et celle de la nullité de l'erreur moyenne porte le nom de bruit blanc.
- L'hypothèse de normalité des erreurs n'est pas nécessaire pour estimer les paramètres du modèle par la méthode des moindres carrés. Mais elle est indispensable pour déterminer les estimateurs du maximum de vraisemblance des paramètres du modèle. Elle permet aussi la construction des intervalles de confiance des paramètres et la réalisation des tests d'hypothèses en général.

En fonction des hypothèses évoquées ci-haut, on peut représenter le modèle de régression simple de la façon suivante :

$$Y_t = \beta_0 + \beta_1 X_t + u_t \quad (3.20)$$

(t = 1, 2, ..., n)
avec $E(u_t) = 0$, quelque soit t :

$$E(u_i, u_j) = \begin{cases} 0, & \text{quels que soient i et j, avec } i \neq j \\ \sigma_u^2, & \text{pour tout i et j, avec } i = j \end{cases}$$

$$u_t \sim N(0, \sigma_u^2)$$

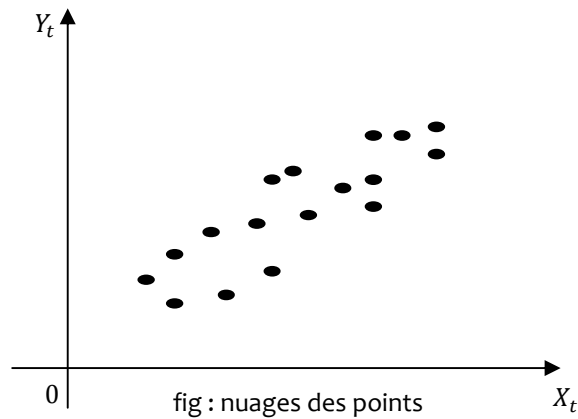
3.5. ESTIMATION PAR LA METHODE DES MOINDRES CARRES

Avant toute chose, rappelons qu'un estimateur est une formule ou une méthode permettant d'estimer des paramètres inconnus de la population à partir d'un échantillon d'observations. Tandis qu'une estimation est la valeur numérique résultant de l'application de cette formule à une réalisation d'un processus aléatoire.

3.5.1. Formulation des estimateurs

L'étape préliminaire dans l'ajustement d'une droite à partir de l'échantillon des données est de représenter le nuage des points dans un diagramme et de s'assurer que celui-ci est à peu près linéaire.

Tracer un graphique des couples des données liant les observations sur Y_t et X_t , en vue d'obtenir un nuage des points :



Remplacer le nuage statistique par une droite de régression ou d'ajustement, qui minimise la somme des carrés des écarts à la droite.

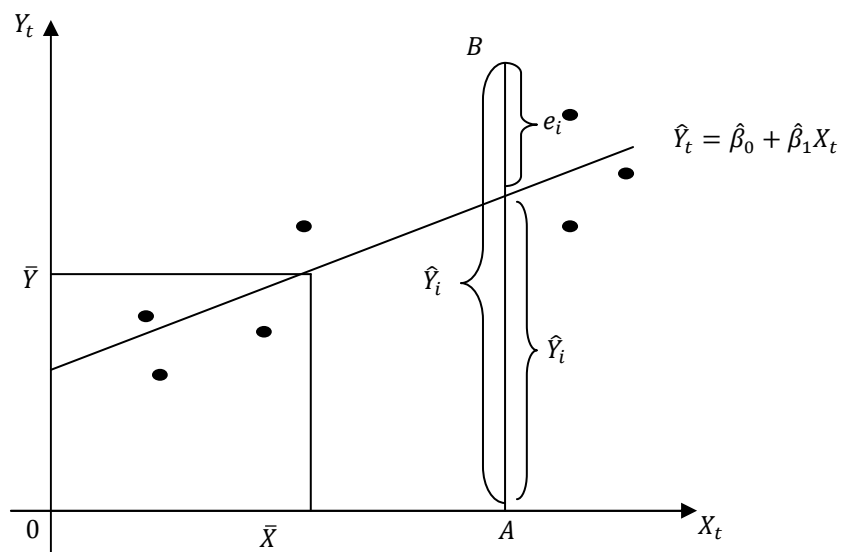


fig. : droite d'ajustement

L'équation de la droite ainsi ajustée s'écrit :

$$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 X_t$$

3.21

avec

$\hat{\beta}_0$, l'estimateur de β ;

$\hat{\beta}_1$, l'estimateur de β_1 ;

\hat{Y}_t , la droite estimée de Y_t .

Notons qu'un estimateur est simplement une formule ou une méthode indiquant comment estimer le paramètre de la population à partir de l'information par l'échantillon disponible porte le nom d'estimateur.

Les écarts à la droite, communément appelés « résidus » sont représentés par :

$$\begin{aligned} e_t &= Y_t - \hat{Y}_t \\ &= Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t \\ &\quad (\text{où } t = 1, 2, \dots, n) \end{aligned}$$

Etant donné qu'en pratique, on a rarement accès à la population entière, mais qu'on dispose habituellement d'un échantillon d'observations issues de cette population, on alors procède à l'estimation de la fonction de régression de la population :

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

Sur la base de la fonction de régression stochastique de l'échantillon :

$$Y_t = \hat{\beta}_0 + \hat{\beta}_1 X_t + e_t$$

Comme la fonction de régression de l'échantillon n'est qu'une approximation de la fonction de régression de la population, il faut élaborer des procédés qui rendent cette approximation aussi faible que possible.

La méthode des moindres carrés ordinaires (OLS) est l'un des procédés largement utilisés dans l'analyse de régression. Le critère d'ajustement par la méthode des moindres carrés ordinaires (Ordinary Least Squares) est donc :

« Choisir $\hat{\beta}_0$ et $\hat{\beta}_1$ afin de minimiser $\sum_{t=1}^n e_t^2$ ».

Formellement, ce critère s'écrit :

$$\begin{aligned} \text{Min} \sum_{t=1}^n e_t^2 &= \text{Min} \sum_{t=1}^n (Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t)^2 \\ &= \text{Min} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 = \text{Min } S(\hat{\beta}_0, \hat{\beta}_1) \end{aligned} \quad 3.22$$

Pour que la SCR soit minimum, il faut que :

$$\frac{\delta(\sum e^2)}{\delta \hat{\beta}_0} = -2 \sum (Y - \hat{\beta}_0 - \hat{\beta}_1 X) = -2 \sum e = 0 \quad 3.23$$

$$\frac{\delta(\sum e^2)}{\delta \hat{\beta}_1} = -2 \sum X(Y - \hat{\beta}_0 - \hat{\beta}_1 X) = -2 \sum Xe = 0 \quad 3.24$$

Après développement, on obtient les équations normales de la régression de Y sur X :

$$\begin{cases} \sum Y = n\hat{\beta}_0 + \hat{\beta}_1 \sum X & 3.25 \end{cases}$$

$$\begin{cases} \sum XY = \hat{\beta}_0 \sum X + \hat{\beta}_1 \sum X^2 & 3.26 \end{cases}$$

De l'équation (3.25), on tire l'estimateur de β_0 :

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad 3.27$$

En remplaçant $\hat{\beta}_0$ dans la deuxième équation normale (3.26), on obtient :

$$\hat{\beta}_1 = \frac{\sum_t (X_t - \bar{X})(Y_t - \bar{Y})}{\sum_t (X_t - \bar{X})^2} \quad 3.28$$

La relation (3.28) peut s'écrire aussi comme :

$$\hat{\beta}_1 = \frac{\sum_t X_t Y_t - n\bar{X}\bar{Y}}{\sum_t X_t^2 - n\bar{X}^2} \quad 3.29$$

ou simplement :

$$\hat{\beta}_1 = \frac{\sum_t x_t y_t}{\sum_t x_t^2} \quad 3.30$$

avec $x_t = X_t - \bar{X}$ et $y_t = Y_t - \bar{Y}$

3.5.2. Estimateur des variables centrées

On peut simplifier les calculs en exprimant les observations comme des écarts par rapport à leur moyenne.

Ainsi :

$$\begin{aligned} y_t &= Y_t - \bar{Y} ; \quad x_t = X_t - \bar{X} \\ \hat{y}_t &= \hat{Y}_t - \bar{Y} ; \quad e_t = y_t - \hat{y}_t = \hat{Y}_t - Y_t \end{aligned}$$

Le critère d'ajustement par la méthode des moindres carrés devient :

$$\begin{aligned} \text{Min } S &= \text{Min} \sum_{t=1}^n e_t^2 = \text{Min} \sum_{t=1}^n (y_t - \hat{y}_t)^2 \\ &= \text{Min} \sum_{t=1}^n e_t^2 = \text{Min} \sum_{t=1}^n (y_t - \hat{\beta}_1 x_t)^2 \\ &= \text{Min } f(\hat{\beta}_1) \end{aligned} \quad 3.31$$

Minimiser cette fonction implique :

$$\frac{\delta S}{\delta \hat{\beta}_1} = -2 \sum_{t=1}^n x_t (y_t - \hat{\beta}_1 x_t) = 0 \quad 3.32$$

D'où :

$$\begin{aligned} \sum_{t=1}^n x_t y_t &= \hat{\beta}_1 \sum_{t=1}^n x_t^2 \\ \Rightarrow \hat{\beta}_1 &= \frac{\sum_{t=1}^n x_t y_t}{\sum_{t=1}^n x_t^2} \end{aligned} \quad 3.33$$

La relation (3.33) peut s'écrire aussi comme :

$$\hat{\beta}_1 = \frac{\sum_{t=1}^n x_t y_t}{\sum_{t=1}^n x_t^2} = \frac{\sum_{t=1}^n x_t Y_t}{\sum_{t=1}^n x_t^2} \quad 3.34$$

car :

$$\begin{aligned} \sum_{t=1}^n x_t Y_t &= \sum_{t=1}^n x_t (y_t - \bar{Y}) \\ &= \sum_{t=1}^n x_t y_t - \bar{Y} \underbrace{\sum_{t=1}^n x_t}_{=0} = \sum_{t=1}^n x_t y_t \end{aligned}$$

ou encore :

$$\sum_{t=1}^n x_t Y_t = \sum_{t=1}^n y_t X_t$$

3.5.3. Estimation d'un modèle sans terme constant

La théorie économique postule parfois des relations dans lesquelles le terme constant $\beta_0 = 0$. C'est le cas des fonctions des recettes où la quantité vendue nulle entraîne une recette nulle.

La spécification d'un tel modèle se réduit à :

$$Y_t = \beta_1 X_t + u_t \quad 3.35$$

et l'estimation de β est alors donné par la relation suivante :

$$\hat{\beta}_1 = \frac{\sum_{t=1}^n X_t Y_t}{\sum_{t=1}^n X_t^2} \quad 3.36$$

Nous remarquons qu'il s'agit de l'application de la relation (3.36) dans

laquelle \bar{X} et \bar{Y} sont nulles.

Remarques

- a) L'écriture du modèle n'est pas neutre. En effet, la spécification : $Y_t = \beta_0 + \beta_1 X_t + u_t$ n'est pas équivalente à : $X_t = \beta_0^* + \beta_1^* Y_t + u_t^*$. Dans le premier modèle, X_t est la cause de Y_t , alors que dans la deuxième spécification, c'est Y_t qui est la cause de X_t .

Notons que :

$$\hat{\beta}_0 \cdot \beta_0^* = \rho^2 \quad 3.37$$

où ρ est le coefficient de corrélation entre X et Y

- b) Le modèle de régression simple peut s'écrire sous deux formes selon qu'il s'agit du modèle théorique spécifié par l'économiste ou du modèle estimé à partir d'un échantillon d'observations :

- Modèle spécifié par l'économiste

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

où u_t est l'erreur inconnue.

- Modèle estimé à partir d'un échantillon

$$\begin{aligned} Y_t &= \hat{Y}_t + e_t \\ &= \hat{\beta}_0 + \hat{\beta}_1 X_t + e_t \end{aligned}$$

où e_t est le résidu.

- c) Le coefficient $\hat{\beta}_1$ représente la pente de la droite des moindres carrés ou encore la propension marginale. Ainsi, l'impact d'une variation de X_t se mesure directement sur Y_t au travers du coefficient $\hat{\beta}_1$:

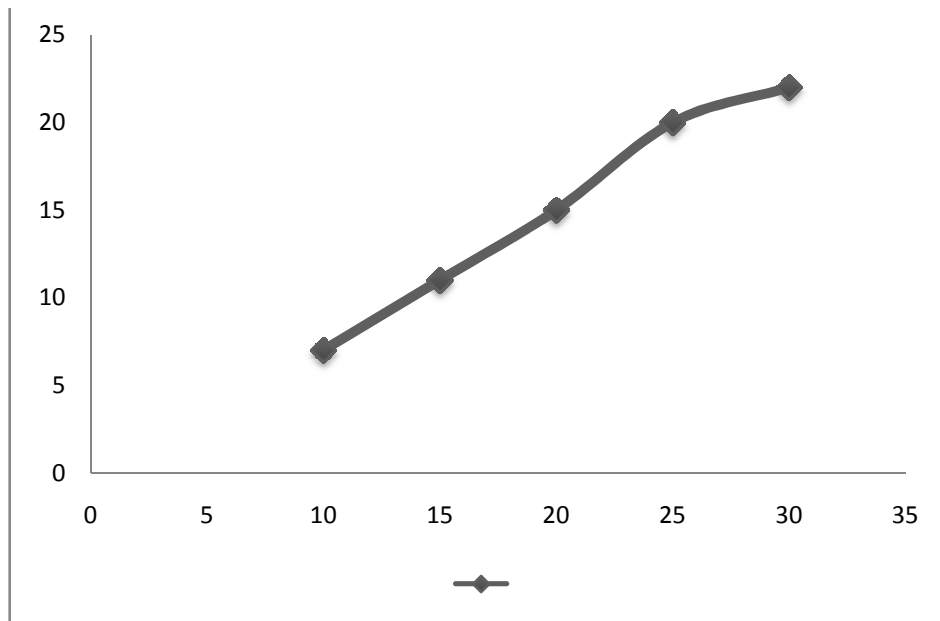
$$\Delta Y_t = \hat{\beta}_1 \Delta X_t$$

Exemple 1

A partir des données du tableau ci-après, on demande de calculer les estimations de $\hat{\beta}_0$ et $\hat{\beta}_1$.

Année	X_t	Y_t
1995	10	7
1996	15	11
1997	20	15
1998	25	20
1999	30	22

a) Représentation graphique du nuage des points

b) Calcul des estimateurs avec X_t et Y_t

Année	X _t	Y _t	X _t Y _t	X _t ²
1995	10	7	70	100
1996	15	11	165	225
1997	20	15	300	400
1998	25	20	500	625
1999	30	22	660	900
Σ	100	75	1695	2250

$$\bar{X}_t = \frac{\sum_{t=1}^5 X_t}{5} = \frac{100}{5} = 20$$

$$\bar{Y}_t = \frac{\sum Y_t}{5} = \frac{75}{5} = 15$$

1. En utilisant les équations normales, on trouve :

$$\begin{cases} \sum Y_t = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_t \\ \sum X_t Y_t = \hat{\beta}_0 \sum X_t + \hat{\beta}_1 \sum X_t^2 \end{cases}$$

$$\Rightarrow \begin{cases} 75 = 5\hat{\beta}_0 + 100\hat{\beta}_1 \\ 1695 = 100\hat{\beta}_0 + 2250\hat{\beta}_1 \end{cases}$$

$$\begin{cases} 15 = \hat{\beta}_0 + 20\hat{\beta}_1 \\ 339 = 20\hat{\beta}_0 + 450\hat{\beta}_1 \end{cases} \Rightarrow \hat{\beta}_1 = 0.78 \quad \text{et} \quad \hat{\beta}_0 = -0.6$$

La droite de régression est donc :

$$\boxed{\hat{Y}_t = -0.6 + 0.78X_t}$$

2. A l'aide des estimateurs $\hat{\beta}_0$ et $\hat{\beta}_1$, on calcule directement :

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum X_t Y_t - \bar{Y} \sum X_t}{\sum X_t^2 - \bar{X} \sum X_t} \\ &= \frac{100 - (15)(100)}{2250 - (20)(100)} = \frac{195}{250} = 0.78 \end{aligned}$$

et

$$\begin{aligned} \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X} \\ &= 15 - (0.78)(20) = -0.6 \end{aligned}$$

$$\Rightarrow \boxed{\hat{Y}_t = -0.6 + 0.78X_t}$$

c) Estimation avec des variables centrées

	X_t	Y_t	$x_t = X_t - \bar{X}$	$y_t = Y_t - \bar{Y}$	$x_t y_t$	x_t^2
	10	7	-10	-8	80	100
	15	11	-5	-4	20	25
	20	15	0	0	0	0
	25	20	5	5	25	25
	30	22	10	7	70	100
\sum	100	75	0	0	195	250

$$\hat{\beta}_1 = \frac{\sum x_t y_t}{\sum x_t^2} = \frac{195}{250} = 0.78$$

La droite d'ajustement sur base des variables centrées est donc :

$$\boxed{\hat{y}_t = 0.78x_t}$$

De cette équation, on peut déduire la droite d'ajustement exprimée en données brutes. En effet :

$$\begin{aligned} \hat{y}_t &= 0.78x_t \\ (\hat{Y}_t - \bar{Y}) &= 0.78 (X_t - \bar{X}) \end{aligned}$$

$$\hat{Y}_t - 15 = 0.78 (X_t - 20)$$

$$\Rightarrow \boxed{\hat{Y}_t = -0.6 + 0.78X_t}$$

ou bien, connaissant la valeur de $\hat{\beta}_1$, on peut tout simplement appliquer la formule de $\hat{\beta}_0$:

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ &= 15 - (0.78)(20) = -0.6\end{aligned}$$

d) Calcul des résidus

Les estimateurs des erreurs (u_t) , appelés résidus et notés (e_t), s'obtiennent à partir de la relation ci-après : $e_t = Y_t - \hat{Y}_t = y_t - \hat{y}_t$.

X_t	Y_t	x_t	y_t	$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 X_t$	$\hat{y}_t = \hat{\beta}_1 x_t$	e_t
10	7	-10	-8	7.2	-7.8	-0.2
15	11	-5	-4	11.2	-3.9	-0.1
20	15	0	0	15.0	0	0
25	20	5	5	18.9	3.9	1.1
30	22	10	7	22.8	7.8	-0.8
Σ						0

e) Les droites d'ajustements

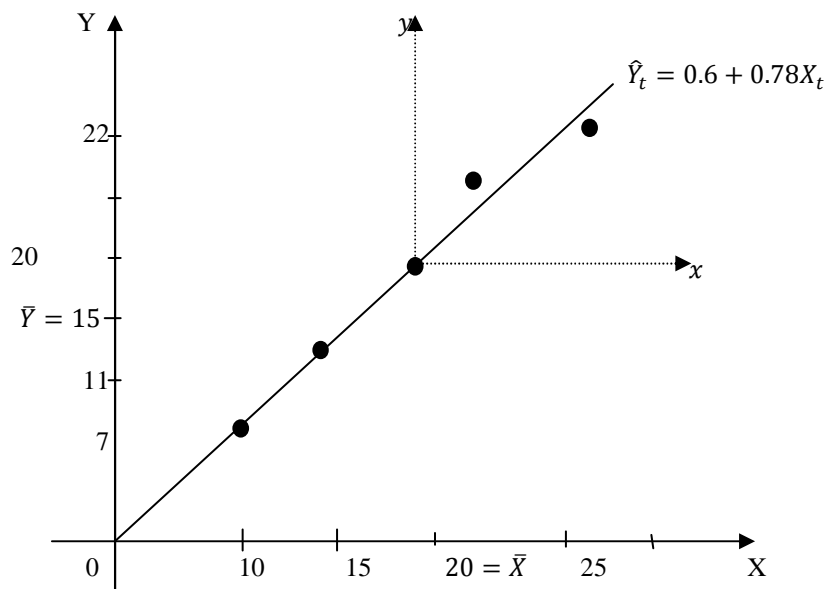


fig. : droite

Exemple 2

Soit le modèle suivant : $Y_t = \beta_0 + \beta_1 X_t + u_t$

Nous disposons des données du tableau ci-dessous :

X_t	Y_t
3	8
5	13
2	7
8	19
10	22
8	15

- Estimer les paramètres du modèle ;
- Calculer le coefficient de détermination (R^2).

Solution

	X	Y	XY	X^2	\hat{Y}	$e = Y - \hat{Y}$	Xe
	3	8	24	9	8.54	-0.54	-1.62
	5	13	65	25	12.18	0.82	4.10
	2	7	14	4	6.72	0.28	0.56
	8	19	152	64	17.64	1.36	10.88
	10	22	220	100	21.28	0.72	7.20
	8	15	120	64	17.64	-2.64	-21.12
Σ	36	84	595	266		0	0

$$\bar{X} = \frac{\sum_t X_t}{n} = \frac{36}{6} = 6$$

$$\bar{Y} = \frac{\sum_t Y_t}{n} = \frac{84}{6} = 14$$

a. Utilisation des équations normales

$$\begin{cases} \sum_t Y_t = n\hat{\beta}_0 + \hat{\beta}_1 \sum_t X_t \\ \sum_t Y_t X_t = \hat{\beta}_0 \sum_t X_t + \hat{\beta}_1 \sum_t X_t^2 \end{cases}$$

$$\begin{cases} 84 = 6\hat{\beta}_0 + 36\hat{\beta}_1 \\ 595 = 36\hat{\beta}_0 + 266\hat{\beta}_1 \end{cases}$$

$$\Rightarrow \hat{\beta}_1 = 1.82 \quad \text{et} \quad \hat{\beta}_0 = 3.08$$

D'où :

$$\hat{Y}_t = 3.08 + 1.82X_t$$

b. Utilisation des données brutes

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_t X_t Y_t - n\bar{X}\bar{Y}}{\sum_t X_t^2 - n\bar{X}^2} \\ &= \frac{595 - 6(6)(14)}{266 - 6(6)^2} = \frac{91}{50} = 1.82 \end{aligned}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 14 - (1.82)(6) = 3.08$$

D'où : $\hat{Y}_t = 3.08 + 1.82X_t$

c. Utilisation des données centrées

	X	Y	x	y	xy	x ²	\hat{y}	e	xe	y ²
	3	8	-3	-6	18	9	-5.46	-0.54	1.62	36
	5	13	-1	-1	1	1	-1.82	0.82	-0.82	1
	2	7	-4	-7	28	16	-7.28	0.28	-1.12	49
	8	19	2	5	10	4	3.64	1.36	2.72	25
	10	22	4	8	32	16	7.28	0.72	2.88	64
	8	15	2	1	2	4	3.63	-2.64	-5.28	1
\sum	36	84	0	0	91	50		0	0	176

$$\hat{\beta}_1 = \frac{\sum_t x_t y_t}{\sum_t x_t^2} = \frac{91}{50} = 1.82$$

D'où : $\hat{y}_t = 1.82x_t$

Etant donné que : $\hat{y}_t = \hat{Y}_t - \bar{Y}$ et $x_t = X_t - \bar{X}$, alors :

$$\hat{Y}_t - \bar{Y} = 1.82(X_t - \bar{X})$$

$$\hat{Y}_t - 14 = 1.82(X_t - 6)$$

$$\hat{Y}_t - 14 = 1.82X_t - 10.92$$

$$\hat{Y}_t = 3.08 + 1.82X_t$$

d. Les coefficients de détermination

- La sommation des carrés expliquée est égale à :

$$\begin{aligned} \text{SCE} &= \hat{\beta}_1 \sum_t y_t x_t \\ &= 1.82(91) = 165.62 \end{aligned}$$

- La somme des carrés totale :

$$\text{SCT} = \sum_t y_t^2 = 176$$

- La somme des carrés résiduelle :

$$\text{SCR} = \text{SCT} - \text{SCE} = 176 - 165.62 = 10.38$$

Le coefficient de détermination est donc :

$$R^2 = \frac{\text{SCE}}{\text{SCT}} = \frac{165.62}{176} = 0.9410$$

ou encore :

$$R^2 = 1 - \frac{\text{SCR}}{\text{SCT}} = 1 - \frac{10.38}{176} = 0.9410$$

$$\boxed{R^2 = 94\%}$$

3.6. DECOMPOSITION DE LA SOMME DES CARRES ET ANALYSE DE LA VARIANCE

La variation totale Y (lorsque X varie peut être exprimée comme la somme de deux termes) : la variation expliquée par régression, et la variation non expliquée par la régression.

Soit :

$$\begin{aligned} y_t &= \hat{y}_t + e_t \\ &= \hat{\beta} x_t + e_t \end{aligned} \tag{3.38}$$

En prenant la somme de carré de cette relation on obtient :

$$\begin{aligned} \sum y_t^2 &= \sum \hat{y}_t^2 + \sum e_t^2 + 2 \sum \hat{y}_t e_t \\ &= \hat{\beta}^2 \sum x_t^2 + \sum e_t^2 + 2\hat{\beta} \sum x_t e_t \end{aligned} \tag{3.39}$$

Comme :

$$\begin{aligned}
 \sum x_t e_t &= \sum x_t (y_t - \hat{y}_t) \\
 &= \sum x_t y_t - \hat{\beta} \sum x_t^2 \\
 &= \sum x_t y_t - \sum x_t y_t = 0 \\
 (\text{car } \hat{\beta} &= \frac{\sum x_t y_t}{\sum x_t^2} \rightarrow \hat{\beta} \sum x_t^2 = \sum x_t y_t), \text{ alors :} \\
 \sum y_t^2 &= \sum \hat{y}_t^2 + \sum e_t^2 \text{ ou } \sum y_t^2 = \hat{\beta}^2 \sum x_t^2 + \sum e_t^2
 \end{aligned}
 \tag{3.40}$$

La relation (3.40) peut s'écrire aussi sous la forme :

$$SCT = SCE + SCR \tag{3.41}$$

où : $SCT = \sum y_t^2$ est la somme des carrés totale ou somme de carrés des écarts par rapport à sa moyenne de la variable dépendante.

$SCR = \sum e_t^2$ est la somme de carrés résiduelle, ou non expliquée.

$SCE = \sum \hat{y}_t^2$ = somme des carrés expliquée.

La relation (3.41) montre que la variabilité totale de Y est égale à la variabilité expliquée de Y et de la variabilité résiduelle.

Cette décomposition de la variance de Y permet d'apprécier la qualité de l'ajustement d'un modèle. En effet, plus la variance expliquée est proche de la variance totale, meilleur est l'ajustement du nuage des points par la droite des moindres carrés.

La qualité de l'ajustement d'un modèle peut être mesurée à partir du rapport :

$$\begin{aligned}
 R^2 &= \frac{SCE}{SCT} = \frac{\sum \hat{y}_t^2}{\sum y_t^2} \\
 &= \frac{\sum (Y_t - \bar{Y})^2}{\sum (Y_t - \bar{Y})^2}
 \end{aligned}
 \tag{4.42}$$

ou bien :

$$\begin{aligned}
 R^2 &= 1 - \frac{SCR}{SCT} \\
 &= 1 - \frac{\sum e_t^2}{\sum y_t^2}
 \end{aligned}
 \tag{4.43}$$

Le coefficient de détermination, R^2 , peut être interprété comme la part de la variation de Y attribuable à la régression linéaire sur X. Le domaine de variation du coefficient de détermination (R^2) est compris entre 0 et 1 : $0 \leq R^2 \leq 1$.

où R^2 est appelé le coefficient de détermination.

Ainsi, R^2 peut être interprété comme la part de variation de Y attribuable à la régression linéaire sur X. des relations (4.42) et (4.43), on aura toujours :

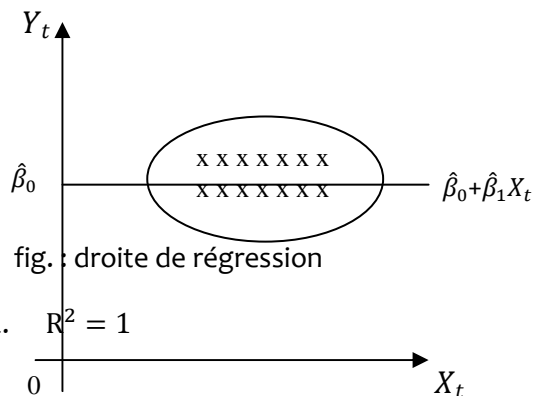
$$0 \leq R^2 \leq 1.$$

Considérons les deux cas limites :

1. $R^2 = 0$

Dans ce cas : $SCR = SCT$ et $R^2 = 1 - \frac{SCR}{SCT}$.

Il n'y a donc pas d'association linéaire entre X et Y ; la droite des moindres carrés est horizontale.



2. $R^2 = 1$

Dans ce cas : $SCE = SCT$ et $R^2 = \frac{SCE}{SCT} = 1$

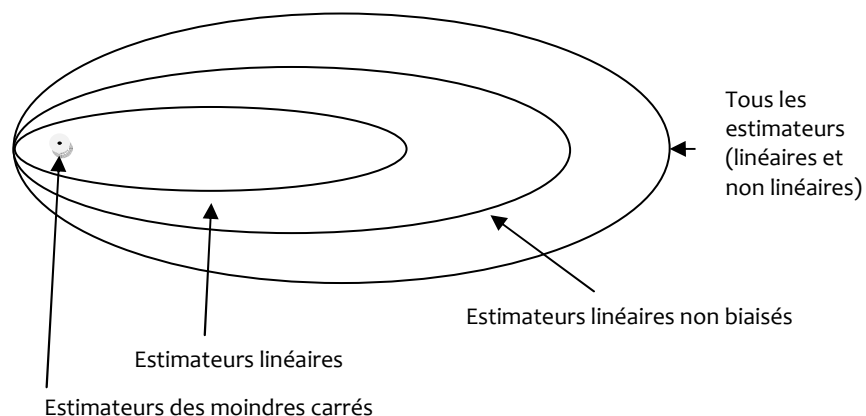
Tous les résidus sont égaux et comme ils sont de moyenne nulle, ils sont égaux à zéro. Les points (X_i, Y_i) se trouvent sur une droite ; l'association linéaire entre X et Y est parfaite.

*

3.7. PROPRIETES DES ESTIMATEURS DES MOINDRES CARRES

La méthode des moindres carrés n'est qu'une méthode possible d'estimation des paramètres d'un modèle parmi tant d'autres méthodes, mais, elle a l'avantage de donner des estimations qui sont les plus proches possibles des vraies valeurs de la population.

D'ailleurs, le théorème de Gauss-Markov montre que les estimateurs des moindres carrés sont BLEU (Best Linear Unbiased Estimator), c'est-à-dire que dans la classe des estimateurs linéaires non biaisés, $\hat{\beta}_0$ et $\hat{\beta}_1$ sont (à variance minimum) les meilleurs, en ce sens qu'ils ont la variance la plus faible.



Comme souligné avant, les estimateurs des moindres carrés, β_0 et β_1 , possèdent donc trois propriétés :

- Ils sont linéaires ;
- non biaisés ;
- et à variance minimale.

3.7.1. Estimateurs linéaires

Les estimateurs des moindres carrés, $\hat{\beta}_0$ et $\hat{\beta}_1$, sont des fonctions linéaires en Y .

$$a) \quad \hat{\beta}_1 = \frac{\sum_t x_t y_t}{\sum_t x_t^2} = \frac{\sum_t x_t (Y_t - \bar{Y})}{\sum_t x_t^2} = \frac{\sum_t x_t Y_t}{\sum_t x_t^2},$$

$$\text{car } \sum_t x_t \bar{Y} = \bar{Y} \sum_t x_t$$

De ce qui précède, $\frac{x_t}{\sum_t x_t^2}$ joue le rôle de pondération de l'effet Y_t par la cause X_t .

Cette pondération a les caractéristiques suivantes :

$$1. \quad \sum_t w_t^2 = 0$$

Démonstration :

$$\sum_t w_t = \frac{\sum_t x_t}{\sum_t x_t^2} = 0 \quad \text{car} \quad \sum_t x_t = 0.$$

$$2. \quad \sum_t w_t^2 = \frac{1}{\sum_t x_t^2}$$

Démonstration :

$$\sum_t w_t^2 = \sum_t \left[\frac{x_t^2}{(\sum_t x_t^2)^2} \right] = \frac{\sum_t x_t^2}{(\sum_t x_t^2)^2} = \frac{1}{\sum_t x_t^2}.$$

$$3. \quad \sum_t w_t x_t = \sum_t w_t X_t = 1$$

Démonstration :

$$\begin{aligned} \sum_t w_t x_t &= \sum_t w_t (X_t - \bar{X}) \\ &= \sum_t \left[\frac{x_t X_t - x_t \bar{X}}{\sum_t x_t^2} \right] = \frac{\sum_t x_t X_t}{\sum_t x_t^2} = \frac{\sum_t x_t X_t}{\sum_t x_t X_t} = 1. \end{aligned}$$

En tenant compte de la pondération ainsi définie, on peut déduire :

$$\hat{\beta}_1 \frac{\sum_t x_t y_t}{\sum_t x_t^2} = \frac{\sum_t x_t Y_t}{\sum_t x_t^2} = \sum_t w_t Y_t \quad 3.44$$

Donc $\hat{\beta}_1$ est une combinaison linéaire des valeurs de Y .

$$\begin{aligned}
b) \quad \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X} \\
&= \frac{1}{n} \sum_t Y_t - \bar{X} \sum_t w_t Y_t \\
&= \sum_t \left[\frac{1}{n} - \bar{X} w_t \right] Y_t = \sum_t Q_t Y_t \\
&\text{où } Q_t = \frac{1}{n} - \bar{X} w_t
\end{aligned} \tag{3.45}$$

Par conséquent, l'estimateur de α est aussi une combinaison linéaire des variables aléatoires Y .

3.7.2. Estimateurs non biaisés

Les estimateurs des moindres carrés, $\hat{\beta}_0$ et $\hat{\beta}_1$, sont non biaisés. Cette assertion implique que :

$$E(\hat{\beta}_1) = \beta_1 \quad \text{et} \quad E(\hat{\beta}_0) = \beta_0$$

a) $\hat{\beta}_1$ estimateur non biaisé

De la relation (1), nous savons que : $\hat{\beta}_1 = \sum_t w_t Y_t$.

En remplaçant dans $\hat{\beta}_1$ la variable aléatoire Y_t par sa valeur, on obtient :

$$\begin{aligned}
\hat{\beta}_1 &= \sum_t w_t (\beta_0 + \beta_1 X_t + \mu_t) \\
&= \beta_0 \sum_t w_t + \beta_1 \sum_t w_t X_t + \sum_t w_t \mu_t \\
&= \beta_1 + \sum_t w_t \mu_t
\end{aligned} \tag{3.46}$$

$$\text{Car } \sum_t w_t = 0 \text{ et } \sum_t w_t X_t = 1$$

D'où :

$$\begin{aligned}
E(\hat{\beta}_1) &= \beta_1 + \sum_t w_t E(\mu_t) \\
&\Rightarrow E(\hat{\beta}_1) = \beta_1
\end{aligned} \tag{3.47}$$

Etant donné que $E(\mu_t) = 0$

b) $\hat{\beta}_0$ estimateur non biaisé

Comme $\beta_1 \hat{\beta}_0 = \sum_t \left(\frac{1}{n} - \bar{X} w_t \right) Y_t$, on peut aussi écrire :

$$\begin{aligned}
\hat{\beta}_0 &= \sum_t \left(\frac{1}{n} - \bar{X} w_t \right) (\beta_0 + \beta_1 X_t + \mu_t) \\
&= \sum_t \left(\frac{\beta_0}{n} - \beta_0 \bar{X} w_t + \beta_1 \frac{X_t}{n} - \beta_1 \bar{X} w_t X_t \right) + \sum_t \left(\frac{1}{n} - \bar{X} w_t \right) \mu_t \\
&= \beta_0 + \beta_1 \bar{X} - \beta_1 \bar{X} + \sum_t \left(\frac{1}{n} - \bar{X} w_t \right) \mu_t
\end{aligned}$$

Car

$$\sum_t \beta_1 \bar{X} w_t X_t = \beta_1 \bar{X} \sum_t w_t X_t = \beta_1 \bar{X}$$

Et

$$\sum_t \beta_0 \bar{X} w_t = \beta_0 \bar{X} \sum_t w_t$$

Par conséquent, on trouve :

$$\hat{\beta}_0 = \beta_0 + \sum_t \left(\frac{1}{n} - \bar{X} w_t \right) \mu_t \quad 3.48$$

Et

$$\begin{aligned} E(\hat{\beta}_0) &= \beta_0 + \sum_t \left(\frac{1}{n} - \bar{X} w_t \right) E(\mu_t) \\ \Rightarrow E(\hat{\beta}_0) &= \beta_0 \end{aligned} \quad 3.49$$

3.7.3. Estimateurs convergents

Pour qu'ils soient convergents, les estimateurs doivent vérifier la condition suivante :

$$\lim \text{Var}(\hat{\theta}) \rightarrow 0 \text{ lorsque } n \rightarrow \infty \quad 3.50$$

Où $\text{Var}(\hat{\theta})$ est la variance du paramètre θ ; n étant la taille de l'échantillon.

a) Variance de $\hat{\beta}_1$

De la relation (3.46), nous savons que :

$$\hat{\beta}_1 - \beta_1 = \sum_t w_t \mu_t$$

Ainsi :

$$\begin{aligned} \text{Var}(\hat{\beta}_1) &= E[\hat{\beta}_1 - \beta_1]^2 \\ &= E[(\sum_t w_t \mu_t)^2] \\ &= E[\sum_t w_t^2 \mu_t^2 + 2 \sum_{t < t'} w_t w_{t'} \mu_t \mu_{t'}] \\ &= E \sum_t w_t^2 E(\mu_t^2) + 2 \sum_{t < t'} w_t w_{t'} E(\mu_t \mu_{t'}) \end{aligned} \quad 3.51$$

Or : $E(\mu_t^2) = \sigma_u^2, \forall t$
 $E(\mu_t \mu_{t'}) = 0, \forall t \neq t'$

$$\text{D'où : } \text{Var}(\hat{\beta}_1) = \sigma_u^2 \sum_t W_t^2$$

$$\boxed{(\hat{\beta}_1) = \sigma_{\hat{\beta}_1}^2 = \frac{\sigma_u^2}{\sum_t X_t^2}} \quad 3.52$$

L'estimateur $\hat{\beta}_1$ est convergent, car lorsque $n \rightarrow \infty$ alors $\sum_t x_t^2$ tend également vers ∞ et $\text{Var}(\hat{\beta}_1)$ tend vers 0. Donc cet estimateur est d'autant plus précis que $\sum_t x_t^2$ est élevé, c'est-à-dire lorsque :

- Le nombre d'observations est important ;
- Et/ou les valeurs de la variable explicative sont très dispersées autour de la moyenne

b) Variance de $\hat{\beta}_0$

$$\text{Var}(\hat{\beta}_0) = E[(\hat{\beta}_0 - \beta_0)^2]$$

$$\text{Or, nous savons que : } \hat{\beta}_0 - \beta_0 = \sum_t \left(\frac{1}{n} - \bar{X}w_t \right) \mu_t$$

D'où :

$$\begin{aligned} \text{Var}(\hat{\alpha}) &= E \left[\sum_t \left(\frac{1}{n} - \bar{X}w_t \right) \mu_t \right]^2 \\ &= E \left[\sum_t \left(\frac{1}{n} - \bar{X}w_t \right)^2 \mu_t^2 \right] + E \underbrace{[\text{somme des produits croisés}]}_{=0} \\ &= E \left[\sum_t \left(\frac{1}{n} - \bar{X}w_t \right)^2 \mu_t^2 \right] \\ &= \sum_t \left(\frac{1}{n} - \bar{X}w_t \right)^2 E(\mu_t^2) \\ &= \sigma_\mu^2 \sum_t \left(\frac{1}{n} - \bar{X}w_t \right)^2 \\ &= \sigma_\mu^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_t x_t^2} \right) \\ \Rightarrow \text{Var}(\hat{\alpha}) &= \sigma_{\hat{\alpha}}^2 = \sigma_\mu^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_t x_t^2} \right) \end{aligned} \quad 3.53$$

ou encore :

$$\text{Var}(\hat{\alpha}) = \sigma_{\hat{\alpha}}^2 = \sigma_\mu^2 \frac{\sum_t X_t^2}{n \sum_t x_t^2} \quad 3.54$$

Nous constatons également que $\hat{\alpha}$ est un estimateur qui converge en probabilité vers α , $\lim \text{Var}(\hat{\alpha}) \rightarrow 0$ lorsque $n \rightarrow +\infty$

c) Covariance de $\hat{\beta}_0$ et $\hat{\beta}_1$

$$\text{Cov}(\hat{\alpha}, \hat{\beta}) = E[(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1)]$$

Or, nous savons que :

$$\begin{aligned}
\hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X} \\
&= (\beta_0 + \beta_1 \bar{X} + \bar{\mu}) - \hat{\beta}_1 \bar{X} \\
&= \beta_0 - \bar{X}(\hat{\beta}_1 - \beta_1) + \bar{\mu}
\end{aligned}$$

$$\begin{aligned}
\text{D'où : } \hat{\beta}_0 - \beta_0 &= \bar{\mu} - \bar{X}(\hat{\beta}_1 - \beta_1) \\
\text{avec } \bar{\mu} &= \frac{1}{n} \sum_t \mu_t
\end{aligned}$$

En substituant $(\hat{\beta}_0 - \beta_0)$ dans l'expression de la covariance, nous trouvons :

$$\begin{aligned}
\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) &= E[(\hat{\beta}_1 - \beta_1)\bar{\mu} - (\hat{\beta}_1 - \beta_1)^2 \bar{X}] \\
&= -\bar{X} E(\hat{\beta}_1 - \beta_1)^2 = -\bar{X} \frac{\sigma_\mu^2}{\sum_t x_t^2}
\end{aligned}$$

car

$$\begin{aligned}
E[(\hat{\beta}_1 - \beta_1)\bar{\mu}] &= E\left[\bar{\mu} \sum_t w_t \mu_t\right] \\
&= \frac{1}{n} E[(\mu_1 + \mu_2 + \dots + \mu_n)(w_1 \mu_1 + w_2 \mu_2 + \dots + w_n \mu_n)] \\
&= \frac{1}{n} E[(w_1 \mu_1^2 + w_2 \mu_2^2 + \dots + w_n \mu_n^2) \\
&\quad + (w_1 \mu_1 \mu_2 + w_2 \mu_1 \mu_3 + \dots + w_n \mu_n \mu_{n-1})] \\
&= \frac{1}{n} \sigma_\mu^2 \sum_t w_t = 0
\end{aligned}$$

$$\text{Donc : } \boxed{\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = -\bar{X} \frac{\sigma_\mu^2}{\sum_t x_t^2}} \quad 3.55$$

Cette relation montre que les estimations de α et β_1 sont corrélées.

La matrice des variances et covariances de $\hat{\beta}_0$ et $\hat{\beta}_1$ est donc :

$$\Omega(\hat{\beta}_0, \hat{\beta}_1) = \begin{bmatrix} \text{Var}(\hat{\beta}_0) & \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) \\ \text{Cov}(\hat{\beta}_1, \hat{\beta}_0) & \text{Var}(\hat{\beta}_1) \end{bmatrix}$$

En utilisant les résultats obtenus ci-haut, on obtient :

$$\Omega(\hat{\beta}_0, \hat{\beta}_1) = \sigma_\mu^2 \begin{bmatrix} \frac{\sum_t x_t^2}{n \sum_t x_t^2} & -\frac{\bar{X}}{\sum_t x_t^2} \\ -\frac{\bar{X}}{\sum_t x_t^2} & \frac{1}{\sum_t x_t^2} \end{bmatrix} \quad 3.56$$

Il convient de noter que : $\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = \text{Cov}(\hat{\beta}_1, \hat{\beta}_0)$

3.7.4. Le théorème de Gauss-Markov

Les estimateurs des moindres carrés s'avèrent être des combinaisons linéaires de la variable Y et donc des combinaisons linéaires de la variable stochastique u . Comme, ils sont aussi sans biais, ils appartiennent à la classe des estimateurs linéaires sans biais. Leur grande importance dans la théorie et la pratique des statistiques provient de ce que leurs variances est la plus faible, dans cette classe.

En effet, si on considère un estimateur linéaire sans biais quelconque de β :

$$b^* = \sum c_t Y_t \quad 3.57$$

On peut montrer que :

$$\text{Var}(b^*) = \text{Var}(\hat{\beta}_1) + \delta_u^2 \sum (c_t - w_t)^2 \quad 3.58$$

Puisque $\sum (c_t - w_t)^2 \geq 0$, alors $\text{Var}(b^*) \geq \text{Var}(\hat{\beta}_1)$. Il n'y a égalité que si $c_t - w_t$ pour tout t , c'est-à-dire que si $b^* = \hat{\beta}_1$.

Donc l'estimateur des moindres carrés possède la variance la plus petite dans la classe des estimateurs linéaires sans biais. On dit que c'est un estimateur **BLUE** (de l'anglais Best Linear Unbiased Estimator).

3.7.5. Compléments : meilleurs estimateurs

Par meilleur estimateur, il faut comprendre celui qui est le plus efficace (efficient), c'est-à-dire celui qui a la variance minimale.

On démontre facilement que les estimateurs des moindres carrés, $\hat{\beta}_0$ et $\hat{\beta}_1$, répondent à ce critère d'efficacité.

Considérons un estimateur général de β_1 , soit :

$$\hat{\beta}_1 = \sum c_t y_t \quad 3.59$$

Où c_t est la pondération qui doit rendre la variance de $\hat{\beta}_1$ minimale tout en respectant le critère non biaisé de $\hat{\beta}_1$, c'est-à-dire $E(\hat{\beta}_1) = \beta_1$.

En remplaçant la variable aléatoire Y_t par sa valeur, on obtient :

$$\hat{\beta}_1 = \beta_0 \sum c_t + \beta_1 \sum c_t X_t + \sum c_t u_t$$

et

$$E(\hat{\beta}_1) = \beta_1 \sum c_t + \beta_1 \sum c_t X_t \quad 3.60$$

Pour que $E(\hat{\beta}_1) = \beta_1$, il faut que :

$$\sum c_t = 0 \quad \text{et} \quad \sum c_t X_t = 1 \quad 3.61$$

Ces deux questions forment les contraintes de la fonction objectif qui vise à minimiser la variance de $\hat{\beta}_1$.

La variance de $\hat{\beta}_1$ est égale à :

$$\begin{aligned} \text{Var}(\hat{\beta}_1) &= E[\hat{\beta}_1 - \beta_1] = E\left[\left(\sum_t C_t u_t\right)^2\right] \\ &= (\sum_t C_t^2 u_t^2) + E(\underbrace{\text{somme des produits croisés}}_{=0}) \\ &= \sigma_u^2 \sum_t C_t^2 \end{aligned} \quad 3.62$$

Formons la fonction de Lagrange afin de déterminer la valeur de c_t^2 qui rend minimale la variance de $\hat{\beta}_1$ sous contrainte des équations de la relation (3.62) :

$$L = \sum_t c_t^2 - 2\lambda \sum_t c_t - 2\gamma(\sum_t c_t X_t - 1) \quad 3.63$$

où λ et γ sont les multiplicateurs de Lagrange.

Les conditions du premier ordre de l'optimisation sous contraintes permettent de trouver :

$$\frac{\partial L}{\partial C_t} = 2C_t - 2\lambda - 2\gamma X_t = 0 \quad 3.64$$

$$\frac{\partial L}{\partial \lambda} = -2 \sum_t c_t = 0 \quad 3.65$$

$$\frac{\partial L}{\partial C_t} = (2 \sum_t c_t X_t - 1) = 0 \quad 3.66$$

De (3.54) et (3.65), on tire :

$$\begin{aligned} \sum_t (\lambda + \gamma X_t) &= 0 \\ n\lambda &= -\gamma \sum_t X_t \\ \lambda &= -\gamma \bar{X} \end{aligned} \quad 3.67$$

et

$$\begin{aligned} c_t &= -\gamma \bar{X} + \gamma X_t \\ &= \gamma(\bar{X} + X_t) = \gamma x_t \end{aligned} \quad 3.68$$

En substituant (3.68) dans (3.66), il vient :

$$\begin{aligned} \gamma \sum_t x_t X_t &= 1 \\ \gamma &= \frac{1}{\sum_t x_t X_t} \\ \Rightarrow \gamma &= \frac{1}{\sum_t x_t^2} \end{aligned} \quad 3.69$$

$$\text{car } \sum_t x_t X_t = \sum_t x_t^2$$

Et la relation (3.68) devient :

$$c_t = \frac{x_t}{\sum_t x_t^2} \quad 3.70$$

Par conséquent, la variance de $\hat{\beta}_1$ atteint son minimum lorsque $c_t = \frac{x_t}{\sum_t x_t^2}$

3.7.6. Estimateur de la variance résiduelle (σ_u^2)

Les variances et covariances des estimateurs des moindres carrés $\hat{\beta}_0$ et $\hat{\beta}_1$ contiennent le paramètre inconnu δ_u^2 (variance de la perturbation ; variance résiduelle ou variance du terme aléatoire u). Pour rendre opérationnels ces estimateurs, il faut nécessairement remplacer le paramètre inconnu par son estimateur calculé à partir des valeurs observées des résidus de la régression : e_t . En effet, la variance résiduelle, σ_u^2 , ne peut être estimée à partir d'un échantillon des valeurs de μ , car celle-ci dépend des paramètres inconnus β_0 et β_1 qui sont inobservables. Mais, on peut en faire une estimation sur la base des résidus.

Etant donné :

$$e_t = y_t - \hat{y}_t \quad \text{avec} \quad \begin{cases} y_t = \beta_0 x_t + (\mu_t - \bar{\mu}) \\ \text{et } \hat{y}_t = \hat{\beta}_1 x_t \end{cases}$$

On obtient :

$$e_t = -(\hat{\beta}_1 - \beta_1)x_t + (\mu_t - \bar{\mu}) \quad 3.71$$

L'espérance mathématique de $\sum_t e_t^2$ est égale à :

$$E\left(\sum_t e_t^2\right) = E\left[\underbrace{(\hat{\beta}_1 - \beta_1)^2 \sum_t x_t^2}_1 + \underbrace{\sum_t (\mu_t - \bar{\mu})^2}_2 - 2(\hat{\beta}_1 - \beta_1) \underbrace{\sum_t x_t (\mu_t - \bar{\mu})}_3\right]$$

$$1) E(\hat{\beta}_1 - \beta_1)^2 \sum_t x_t^2 = \sigma_\mu^2$$

$$2) E\left[\sum_t (\mu_t - \bar{\mu})^2\right] = E\left[\sum_t \mu_t^2 - \frac{(\sum_t \mu_t)^2}{n}\right]$$

$$= n\sigma_\mu^2 - \frac{n\sigma_\mu^2}{n} = (n-1)\sigma_\mu^2$$

$$3) -2E\left[(\hat{\beta}_1 - \beta_1) \sum_t x_t (\mu_t - \bar{\mu})\right] = -2E(\hat{\beta}_1 - \beta_1) \sum_t x_t \mu_t$$

$$= -2 \frac{E(\sum_t x_t \mu_t)^2}{\sum_t x_t}$$

$$\left(\text{Notons que } \hat{\beta}_1 - \beta_1 = \sum_t w_t \mu_t = \frac{\sum_t x_t \mu_t}{\sum_t x_t} \right)$$

$$= -2\sigma_\mu^2 \frac{\sum_t x_t^2}{\sum_t x_t^2} = -2\sigma_\mu^2$$

Il en résulte que :

$$E \left[\sum_t e_t^2 \right] = \sigma_\mu^2 + (n-1)\sigma_\mu^2 - 2\sigma_\mu^2$$

$$= (n-1)\sigma_\mu^2 \quad 3.72$$

La relation (3.72) peut s'écrire aussi :

$$\sigma_\mu^2 = E \left[\frac{\sum_t e_t^2}{n-2} \right]$$

D'où :

$$\hat{\sigma}_\mu^2 = \frac{\sum_t e_t^2}{n-2} \quad 3.73$$

Avec $\hat{\sigma}_\mu^2$, estimateur non biaisé de σ_μ^2 .

Soulignons ici que le dénominateur de l'estimateur de la variance résiduelle représente le nombre de degré de liberté² : (n - 2)

Exemple

Soit le tableau des données suivant :

t	X _t	Y _t
1	2	4
2	3	7
3	1	3
4	5	9
5	9	17

a) Calcul de $\hat{\beta}_1$, $\hat{\beta}_1$ et \hat{y}

- En utilisant les données brutes

² La notion de degré de liberté correspond au nombre de valeurs restant réellement à disposition après une procédure d'estimation statistique. Si un échantillon comprend n valeurs, il faut lui retirer le nombre des paramètres estimés.

	X_t	Y_t	$X_t Y_t$	X_t^2	\hat{y}_t	e_t	$X_t e_t$
	2	4	8	4	4.50	-0.50	-1
	3	7	21	9	6.25	0.75	2.25
	1	3	3	1	2.75	0.25	0.25
	5	9	45	25	9.75	-0.75	-3.75
	9	17	153	81	16.75	0.25	2.25
Σ	20	40	230	120	40	0	0

$$\bar{X} = \frac{\Sigma X_t}{n} = \frac{20}{5} = 4 ; \quad \bar{Y} = \frac{\Sigma Y_t}{n} = \frac{40}{5} = 8$$

- En utilisant les écarts par rapport à leurs moyennes

	x_t	y_t	$x_t y_t$	x_t^2	y_t^2	\hat{y}_t	e_t	$x_t e_t$
	-2	-4	8	4	16	-3.50	-0.50	1.00
	-1	-1	1	1	1	-1.75	0.75	-0.75
	-3	-5	15	9	25	-5.25	0.25	-0.75
	1	1	1	1	1	1.75	0.75	-0.75
	5	9	45	25	81	8.75	0.25	1.25
Σ	0	0	70	40	124	0	0	0

Les équations normales peuvent s'écrire sous la forme :

$$\begin{cases} 40 = 5\hat{\beta}_0 + 20\hat{\beta}_1 \\ 230 = 20\hat{\beta}_0 + 120\hat{\beta}_1 \end{cases}$$

D'où on déduit la solution :

$$\hat{\beta}_0 = 1 ; \quad \hat{\beta}_1 = 1.75 \text{ et } \boxed{\hat{Y}_t = 1 + 1.75X_t}$$

Les mêmes estimations peuvent être obtenues en utilisant les données présentées sous leur forme d'écart par rapport à la moyenne. Ainsi :

$$\hat{\beta}_1 = \frac{\Sigma x_t y_t}{\Sigma x_t^2} = \frac{70}{40} 1.75$$

et

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 8 - (1.75)(4) = 1$$

Par conséquent :

$$\hat{y}_t = 1.75x_t \text{ et } \hat{Y}_t = 1 + 1.75X_t$$

b) Décomposition de la somme des carrés

- La somme des carrés expliquée est donnée par :

$$\begin{aligned} \text{SCR} + \sum \hat{y}_t^2 &= \hat{\beta}_1 \sum x_t y_t \\ &= (1.75)(70) = 122.5 \end{aligned}$$

- La somme des carrés totales

$$\text{SCT} = \sum y_t^2 = 124$$

- La somme des carrés résiduelle

$$\text{SCR} = \text{SCT} - \text{SCE} = 124 - 122.5 = 1.5$$

- La part de la variation de Y expliquée par la régression linéaire est donc

$$R^2 = \frac{\text{SCE}}{\text{SCT}} = \frac{122.5}{124} = 0.9879$$

c) Calcul des variances des estimateurs $\hat{\beta}_0$ et $\hat{\beta}_1$

- Variance résiduelle

$$\begin{aligned} \hat{\sigma}_\mu^2 &= \frac{\sum e_t^2}{n-2} = \frac{\text{SCR}}{n-2} \\ &= \frac{1.5}{(5-2)} = \frac{1.5}{3} = 0.5 \end{aligned}$$

- Variance de $\hat{\beta}_1$

$$\hat{\sigma}_{\hat{\beta}_1}^2 = \frac{\hat{\sigma}_\mu^2}{\sum x_t^2} = \frac{0.5}{40} = 0.0125$$

et

$$\hat{\sigma}_{\hat{\beta}_1} = \sqrt{0.0125} = 0.111803398$$

- Variance de $\hat{\beta}_0$

$$\hat{\sigma}_{\hat{\beta}_0}^2 = \hat{\sigma}_\mu^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x_t^2} \right) = 0.5 \left(\frac{1}{5} + \frac{16}{40} \right) = 0.3$$

et

$$\hat{\sigma}_{\hat{\beta}_0} = \sqrt{0.3} = 0.5477$$

3.8. INTRODUCTION STATISTIQUE DANS LE MODELE DES MOINDRES CARRES ORDINAIRES

L'utilisation de la méthode des moindres carrés ordinaires nous a permis d'estimer les paramètres β_0 , β_1 et σ_u^2 .

Compte tenu des hypothèses fondamentales, nous avons montré que ces paramètres satisfaisaient des propriétés statistiques souhaitables (théorème) de Gauss – Markov : estimateurs BLUES).

Pour apprécier la validité du modèle estimé et interpréter avec exactitude les résultats de l'estimation, c'est-à-dire pour effectuer les tests d'hypothèses et calculer les intervalles de confiance des paramètres, l'on doit considérer une hypothèse supplémentaire sur la forme de la loi de probabilité des erreurs (μ_t). Cette hypothèse suppose que chaque μ_t soit une distribution normale avec moyenne nulle et variance σ_u^2 . Donc :

$$\mu_t \sim N(0, \sigma_u^2) \quad 3.74$$

La normalité des erreurs, non indispensable pour l'obtention des estimateurs BLUE, permet de déterminer la distribution statistique des valeurs estimées $\hat{\beta}_0$ et $\hat{\beta}_1$.

3.8.1. Propriétés des estimateurs des moindres carrés ordinaires selon l'hypothèse de normalité

Sous l'hypothèse que μ_t soit une distribution normale, les estimateurs des moindres carrés ordinaires ont les propriétés :

- Ils sont biaisés
- Ils ont une variance minimale. Associé à la première, ils sont non biaisés à variance minimale, où qu'ils sont des estimateurs performants.
- Ils sont robustes : lorsque la taille de l'échantillon croît indéfiniment, ils convergent vers leurs valeurs effectives dans la population.

a) $\hat{\beta}_0$ (fonction linéaire de μ_t) est distribué normalement avec :

$$\text{Moyenne : } E(\hat{\beta}_0) = \beta_0 \quad 3.75$$

$$\text{Var}(\hat{\beta}_0) : \sigma_{\hat{\beta}_0}^2 = \frac{\sum x_t^2}{n \sum x_t^2} \cdot \sigma_u^2 \quad 3.76$$

Ou mieux :

$$\hat{\beta}_0 \sim N(\beta_0, \sigma_u^2) \quad 3.77$$

Compte tenu de l'hypothèse de normalité, la variable Z, qui se définit comme :

$$Z = \frac{\hat{\beta}_0 - \beta_0}{\alpha_{\hat{\beta}_0}} \quad 3.78$$

Suit la distribution normale standard, c'est-à-dire une distribution ayant une moyenne nulle et une variance égale à 1. Donc :

$$Z \sim N(0, 1) \quad 3.79$$

b) $\hat{\beta}_1$ (étant une fonction linéaire de u_t) a une distribution normale avec :

$$\text{- Moyenne : } E(\hat{\beta}_1) = \beta_1 \quad 3.80$$

$$\text{- Var}(\hat{\beta}_1) : \sigma_{\hat{\beta}_1}^2 = \frac{\sigma_u^2}{n \sum x_t^2} \cdot \sigma_u^2 \quad 3.81$$

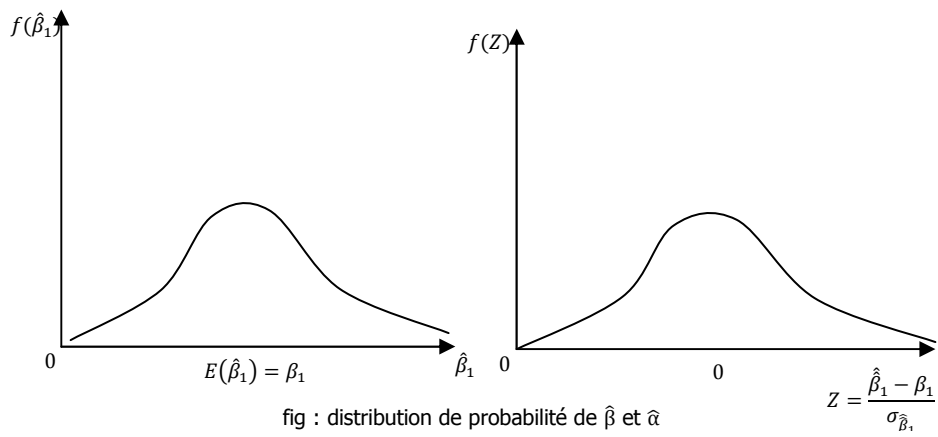
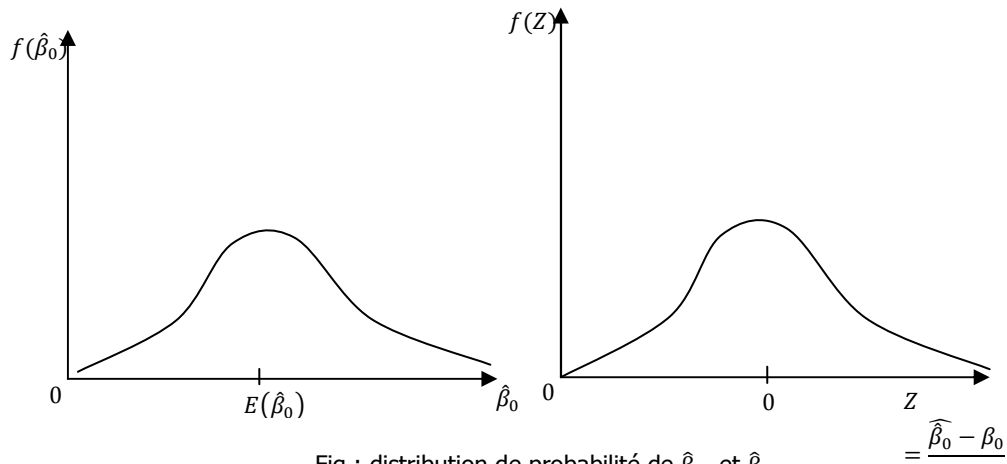
Ou encore :

$$\hat{\beta}_1 \sim N(\beta_1, \sigma_{\hat{\beta}_1}^2) \quad 3.82$$

La variable Z est donc :

$$Z = \frac{\hat{\beta}_1 - \beta_1}{\alpha_{\hat{\beta}_1}} \quad 3.83$$

Les distributions de probabilité de β_0 et de $\hat{\beta}_1$ peuvent être représentées graphiquement comme ci – après :



- c) $(n - 2) \frac{\hat{\sigma}_u^2}{\sigma_u^2} = \frac{\sum e_t^2}{\sigma_u^2}$ est distribué comme la distribution de χ^2 (chi – carré ou chi deux) avec $(n - 2)$ degrés de liberté :

$$(n - 2) \frac{\hat{\sigma}_u^2}{\sigma_u^2} \sim \chi_{(n-2)}^2 \quad 3.84$$

- d) Y_t , fonction linéaire de u_t , possède lui – même une distribution normale avec la moyenne et la variance données par :

$$E(Y_t) = \beta_0 + \beta_1 X_t \quad 3.85$$

$$\text{Var}(Y_t) = \sigma_u^2 \quad 3.86$$

L'on peut écrire simplement :

$$Y_t \sim N(\beta_0 + \beta_1 X_t, \sigma_u^2), \quad 3.87$$

3.8.2. Les intervalles de confiance

3.8.2.1. L'intervalle de confiance β_1

Il s'agit ici de trouver un intervalle I tel que :

$$P[\beta_1 \in I] \quad 3.88$$

Où I est un intervalle de confiance pour le paramètre β_1 au niveau de confiance $(1 - \varepsilon)$

$\varepsilon (0 < \varepsilon < 1)$, seuil de confiance au niveau de signification, peut être interprété comme le risque que l'intervalle de confiance I ne contienne pas la vraie valeur du paramètre.

Pour ce faire, considérons la distribution de probabilité de $\hat{\beta}_1$:

$$\hat{\beta}_1 \sim N(\beta, \sigma_{\hat{\beta}_1}^2) \quad 3.89$$

D'où :

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} = \frac{\hat{\beta}_1 - \beta_1}{\sigma_u} \cdot \sqrt{\sum x_t^2} \sim N(0, 1) \quad 3.90$$

(suit une loi normale centrée réduite)

Etant donné que la variance résiduelle σ_u^2 doit être remplacée par son estimateur $\hat{\sigma}_u^2$, la loi normale centrée réduite devient une loi de Student. En effet, le rapport d'une loi normale centrée réduite et de la racine carrée d'une loi χ^2 divisée par le nombre de ses degrés de liberté permet de trouver :

$$\begin{aligned} t &= \frac{(\hat{\beta}_1 - \beta_1) \sqrt{\sum x_t^2}}{\sigma_u} / \frac{\sqrt{\sum x_t^2} / \sigma_u}{\sqrt{(n-2)}} \\ &= \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_u / \sqrt{\sum x_t^2}} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}} \sim t_{(n-2)} \end{aligned} \quad 3.91$$

Le ratio t suit une loi de Student à $(n - 2)$ de degrés de liberté.

A partir de $\hat{\beta}_1$ du paramètre β_1 , du ratio de Student $t_{\hat{\beta}_1}$ et d'un seuil de signification ε ou d'un niveau de confiance $(1 - \varepsilon)$, on définit l'intervalle de confiance pour β_1 de telle sorte que :

$$P\left[-t_{\frac{\varepsilon}{2}} \leq t_{\hat{\beta}_1} \leq t_{\frac{\varepsilon}{2}}\right] = 1 - \varepsilon \quad 3.92$$

C'est-à-dire la probabilité que $t_{\hat{\beta}_1}$ soit comprise entre $-t_{\frac{\varepsilon}{2}}$ et $t_{\frac{\varepsilon}{2}}$ est de $(1 - \varepsilon)$ degrés de liberté.

En remplaçant le ratio de Student $t_{\hat{\beta}_1}$ par sa valeur dans la relation (19), on obtient :

$$P\left[-t_{\frac{\varepsilon}{2}} \leq \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}} \leq t_{\frac{\varepsilon}{2}}\right] \quad 3.93$$

Cette inéquation simultanée peut s'écrire aussi de façon équivalente comme

suit :

$$P\left(\hat{\beta}_1 - t_{\frac{\varepsilon}{2}} \cdot \hat{\sigma}_{\hat{\beta}_1} \leq \beta_1 \leq \hat{\beta}_1 + t_{\frac{\varepsilon}{2}} \cdot \hat{\sigma}_{\hat{\beta}_1}\right) \quad 3.94$$

De façon plus précise, l'intervalle de confiance pour β_1 de $100(1 - \varepsilon)\%$ est donc :

$$\hat{\beta}_1 \pm t_{\frac{\varepsilon}{2}} \cdot \hat{\sigma}_{\hat{\beta}_1} \quad 3.95$$

Cette inéquation simultanée peut s'écrire de façon équivalente comme suit :

$$P\left(\hat{\beta}_1 - t_{\frac{\varepsilon}{2}} \cdot \hat{\sigma}_{\hat{\beta}_1} \leq \beta_1 \leq \hat{\beta}_1 + t_{\frac{\varepsilon}{2}} \cdot \hat{\sigma}_{\hat{\beta}_1}\right) \cdot 1 - \varepsilon \quad 3.96$$

Cette relation signifie que la probabilité que la vraie valeur inconnue de β_1 soit comprise dans l'intervalle ainsi défini par la relation (3.98) est égale à $(1 - \varepsilon)\%$.

De façon plus concise, l'intervalle de confiance pour β_1 de $100(1 - \varepsilon)\%$ est donc :

$$\boxed{\hat{\beta}_1 \pm t_{\frac{\varepsilon}{2}} \cdot \hat{\sigma}_{\hat{\beta}_1}} \quad 3.97$$

3.8.2.2. Intervalle de confiance de β_0

Un développement similaire permet de calculer l'intervalle de confiance du paramètre inconnu β_0 .

Comme :

$$\begin{aligned} t_{\hat{\beta}_0} &= \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\hat{\beta}_0} \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum x_t^2}}} \\ &= \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\hat{\beta}_0}} \sim t_{(n-2)} \end{aligned} \quad 3.98$$

On déduit :

$$P\left(\hat{\beta}_0 - t_{\frac{\varepsilon}{2}} \cdot \hat{\sigma}_{\hat{\beta}_0} \leq \beta_0 \leq \hat{\beta}_0 + t_{\frac{\varepsilon}{2}} \cdot \hat{\sigma}_{\hat{\beta}_0}\right) \cdot 1 - \varepsilon \quad 3.99$$

Et l'intervalle de confiance de niveau $(1 - \varepsilon)$ à $(n - 2)$ degré de liberté pour α est donné par :

$$\boxed{\hat{\beta}_0 \pm t_{\frac{\varepsilon}{2}} \cdot \hat{\sigma}_{\hat{\beta}_0}} \quad 3.100$$

où $t_{\frac{\varepsilon}{2}}$ est la valeur critique de t pour le seuil de signification $\frac{\varepsilon}{2}$ et $(n - 2)$ degré de liberté.

3.8.2.3. Intervalle de confiance σ_u^2

Etant donné que :

$$\chi^2 = (n-2) \frac{\hat{\sigma}_u^2}{\sigma_u^2} = \frac{\sum e_e^2}{\sigma_u^2} \sim \chi_{(n-2)}^2 \quad 3.101$$

(suit une loi χ^2 à $(n-2)$ degrés de liberté), on peut utiliser cette distribution χ^2 pour bâtir l'intervalle de confiance pour σ_u^2 :

où $\chi_{\frac{\varepsilon}{2}}^2$ et $\chi_{1-\frac{\varepsilon}{2}}^2$ sont des valeurs critiques à lire dans la table du χ^2 à $(n-2)$ degrés de liberté.

En substituant de χ^2 et réaménageant les termes, on obtient :

$$P \left[(n-2) \frac{\sigma_u^2}{\chi_{\frac{\varepsilon}{2}}^2} \leq \sigma_u^2 \leq (n-2) \frac{\sigma_u^2}{\chi_{1-\frac{\varepsilon}{2}}^2} \right] \quad 3.102$$

Remarques

Le seuil de signification le plus communément utilisé est $\varepsilon = 0.05$.

3.8.3. tests de significativité des paramètres estimés

3.8.3.1. Notion

Les paramètres calculés sont des estimations des paramètres de la population. L'on doit alors s'assurer de leur fiabilité statistique, c'est-à-dire de leur proximité aux vraies valeurs mais inconnues des paramètres de la population en procédant aux tests.

Les tests de significativité des paramètres estimés, appelés aussi couramment tests d'hypothèses, ont pour but de vérifier une hypothèse émise sur une population à partir de l'information contenue dans un échantillon. L'hypothèse testée sera en fonction des valeurs possibles du paramètre inconnue.

Pour ce faire, il faut d'abord préciser qu'elle est l'hypothèse soumise au test, 'est à dire l'hypothèse à vérifier. On l'appellera "hypothèse nulle" et elle sera notée H_0 . En outre, il faut définir également "l'hypothèse alternative", notée H_1 , qui est l'hypothèse qu'on est prêt à accepter si H_0 est rejetée. On dira que l'on teste H_0 contre H_1 .

Une hypothèse alternative peut être simple ou composite.

- Une hypothèse statistique est dite simple si elle fixe de façon précise la valeur du paramètre à tester.

Ex.1 : $H_0 : \theta = \theta_0$
contre
 $H_1 : \theta \neq \theta_0$

Ex.2 : $H_0 : \theta = \theta_0$
contre
 $H_1 : \theta = \theta_0$

- Une hypothèse est composite quand elle représente un intervalle des valeurs possibles pour θ .

Ex.1 :

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases}$$

Ex.2 :

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$$

Si l'hypothèse est simple, on utilise le test bilatéral qui répartit le niveau de signification du test en deux parties égales à chacune des extrémités de la courbe de densité de probabilité du paramètre.

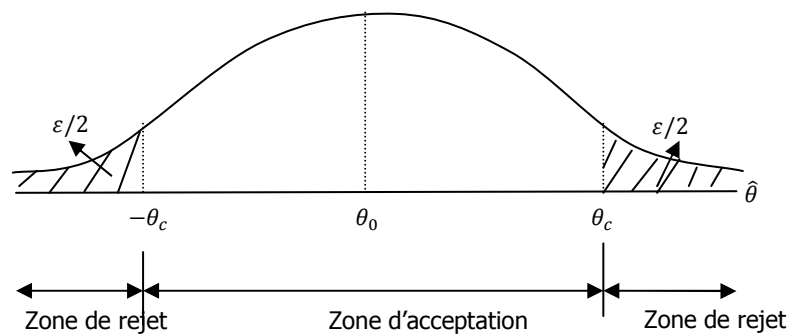


fig : test bilatéral

Dans le cas d'hypothèse composite, on fait appel au test unilatéral à gauche lorsque $H_1 : \theta < \theta_0$ ou au test unilatéral à droite lorsque $H_1 : \theta > \theta_0$.

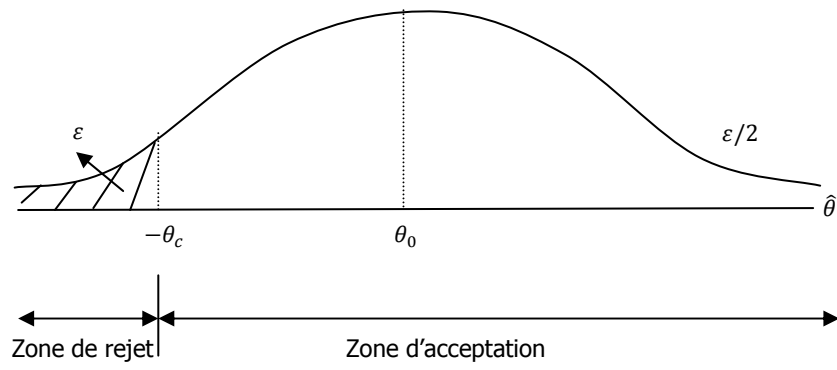


fig : test unilatéral à gauche

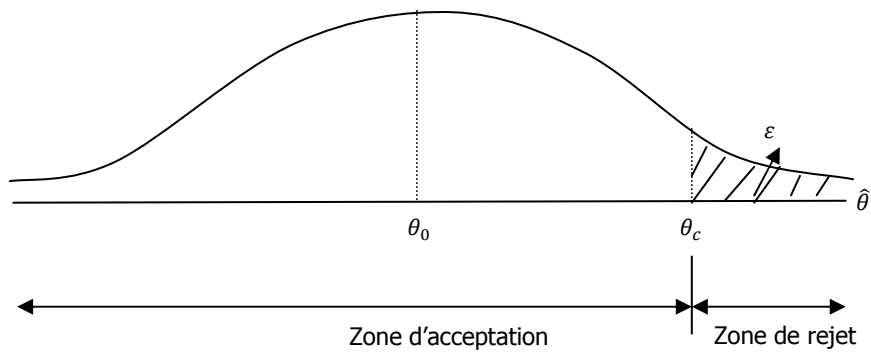


fig : test unilatéral à droite

3.8.3.2. Test de signification du coefficient β_1

Pour tester l'hypothèse de base selon laquelle le paramètre inconnu β_1 du modèle linéaire simple a une valeur donnée, β_1^* soit :

$$\begin{cases} H_0 : \beta_1 = \beta_1^* \\ H_1 : \beta_1 \neq \beta_1^* \end{cases}$$

On calcule l'indicateur t_c (t – calculé) en substituant β_1 par β_1^* dans la relation (3.97) et nous obtenons :

$$t_c = \frac{\hat{\beta}_1 - \beta_1^*}{\hat{\sigma}_{\hat{\beta}_1}} \sim t_{(n-1)} \quad 3.103$$

On compare t_c à $t_{(n-1)}$ obtenu par lecture dans la table t de student pour risque ε en fonction du nombre de degrés de liberté $(n - 2)$.

- Si $|t_c| > t_{(n-2); \frac{\varepsilon}{2}}$, on rejette l'hypothèse H_0 ;
- Si $|t_c| \leq t_{(n-2); \frac{\varepsilon}{2}}$, on accepte l'hypothèse H_0 ;

Le tableau ci – après reprend les règles de décision des différents types de test selon les hypothèses retenues :

Type de test	H_0 Hypothèse nulle	H_1 Hypothèse alternative	Règle de décision
			Rejet de l'hypothèse H_0
1. Test bilatéral	$\beta = \beta_1^*$	$\beta \neq \beta_1^*$	$ t_c > t_{(n-2)}$
2. Test unilatéral à gauche	$\beta \geq \beta_1^*$	$\beta < \beta_1^*$	$t_c < -t_{(n-2)}$
3. Test unilatéral à droite	$\beta \leq \beta_1^*$	$\beta > \beta_1^*$	$t_c > t_{(n-2)}$

Note : $(n - 2)$ est le nombre des degrés de liberté pour le modèle à deux variables ;

t_{ε} , valeur critique de t au seuil de signification ε ;
 $t_{\frac{\varepsilon}{2}}$, valeur critique de t au seuil de signification $\frac{\varepsilon}{2}$.

La région critique est l'ensemble des valeurs de l'observation qui conduisent à rejeter H_0 et donc à accepter H_1 . Le point critique est celui duquel on rejette H_0 .

Faisons remarquer que l'hypothèse la plus couramment testée dans les études empiriques est la suivante :

$$H_0 : \beta_1 = 0 \quad 3.104$$

Le but de cette hypothèse est de tester si la variable endogène Y est reliée à la variable exogène X . si l'hypothèse H_0 est vraie, alors la variable X ne joue aucun rôle dans la détermination de la variable endogène Y , c'est-à-dire elle

ne contribue pas dans l'explication du phénomène étudié.

3.8.3.3. Test de signification global du modèle

Le test d'hypothèse $H_0 : \beta_1 = 0$ peut se ramener à un problème d'analyse de la variance. Mais, cette approche sera particulièrement utile lorsqu'on abordera la régression multiple. Pour l'instant, on la décrit dans le cadre d'un modèle de régression simple à deux variables.

Comme :

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma_u} \sim N(0,1) \quad 3.105$$

Alors :

$$\frac{(\hat{\beta}_1 - \beta_1)^2 \sum x_t^2}{\sigma_u^2} \sim \chi_{(n-2)}^2 \quad 3.106$$

et

$$\frac{\sum x_t^2}{\sigma_u^2} \sim \chi_{(n-2)}^2 \quad 3.107$$

(la loi χ^2 : somme des carrés des variables normales)

En utilisant le fait que le rapport entre deux χ^2 indépendants, divisés par leur nombre de degrés de liberté, suit une loi *F de Fisher – Snedecor*, on trouve :

$$F_c = \frac{(\hat{\beta}_1 - \beta_1)^2 \sum x_t^2}{\sum x_t^2 / (n-2)} \sim F_{(1, n-2)} \quad 3.108$$

Donc la statistique F_c est distribuée selon une loi *F* de Fisher-Snedecor à 1 et $n - 2$ degrés de liberté.

Si $\beta_1 = 0$, la statistique F_c devient :

$$F_c = \frac{\hat{\beta}_1^2 \sum x_t^2}{\sum e_t^2 / (n-2)} \sim F_{(1, n-2)} \quad 3.109$$

En considérant la décomposition de la somme des carrés :

$$SCT = SCE + SCR$$

où

$$SCT = \sum y_t^2 ;$$

$$SCE = \sum \hat{y}_t^2 = \hat{\beta}_1^2 \sum x_t^2 ;$$

$$SCR = \sum e_t^2$$

La relation (3.109) peut s'écrire aussi comme :

$$F_c = \frac{SCE/1}{SCR/(n-2)} \sim F_{(1,n-2)} \quad 3.110$$

On peut enfin écrire cette formule de F_c en fonction du coefficient de détermination :

$$F_c = \frac{R^2/1}{(1-R^2)/(n-2)} \sim F_{(1,n-2)} \quad 3.111$$

sachant que : $SCE = R^2 SCT = R^2 \sum y_t^2$ et $SCR = (1 - R^2) SCT = (1 - R^2) \sum y_t^2$

Les données nécessaires pour faire ce test sont d'habitude classées dans un tableau d'analyse de la variance (ANOVA selon la dénomination anglaise "Analysis Of Variance").

Généralement, pour utiliser cette approche, on commence par construire un tableau d'analyse de la variance (ANOVA selon la dénomination anglaise "Analysis Of Variance").

Source de variation	Somme des carrés	Degrés de liberté	Carrés moyens	F calculé
Variation expliquée	$SCE = \sum \hat{y}_t^2$ $= \hat{\beta}_1^2 \sum \hat{x}_t^2 = \hat{\beta}_1^2 \sum x_t y_t$	1	SCE/1	$\frac{SCE/1}{SCR/(n-2)}$
Variation résiduelle	$SCR = \sum e_t^2$	n-2	SCR/(n-2)	
Variation totale	$SCT = \sum y_t^2$	n-1		

La statistique F_c définie ci-dessus apparait ainsi comme le rapport entre la somme des carrés expliquée par la variable x et la somme des carrés des résidus, chacune d'entre elles étant divisée par son nombre des degrés de liberté. La somme des carrés des résidus peut être considérée comme une mesure du "bruit" existant dans le système, et par conséquent l'effet de x n'est décelé que s'il est plus grand que celui qui est provoqué par ce bruit. On peut donc juger si x est une variable significative en examinant le F de l'échantillon.

La formulation des hypothèses est la suivante :

$$H_0 : \beta_1 = 0$$

contre

H_1 : négation de H_0

Intuitivement, le nombre de degrés de liberté représente le nombre des valeurs que l'on peut choisir arbitrairement. Ainsi, dans le tableau ANOVA, nous pouvons choisir arbitrairement $n-1$ valeurs de y , la $n - i\grave{e}me$ étant déterminée par la condition $\sum y = 0$. De même, nous pouvons choisir arbitrairement $n-2$ valeurs de e (résidu), l'ajustement par les moindres carrés imposant deux conditions que les résidus (e) doivent satisfaire : $\sum e = \sum xe = 0$.

La règle de décision pour un niveau spécifié ε du test est la suivante :

- si $F_c > F_{\varepsilon; (1, n-2)}$, on rejette H_0 ; la variable x est significative ;
- si $F_c \leq F_{\varepsilon; (1, n-2)}$, on accepte H_0 et la variable n'est pas significative.

Remarques

Dans le modèle de régression simple, le test F n'est rien d'autre que le test appliqué à β_1 pour juger si la variable exogène x joue un rôle significatif ou non dans l'explication de y .

On dispose ainsi de deux façons de tester si x est significatif ou non, l'une basée sur le t de Student, l'autre basée sur le F de Fisher. En fait, ces tests sont identiques puisque :

$$t_{(n)}^2 = F_{(1, n)}$$

Donc, on peut utiliser indifféremment chacune de ces statistiques et les comparer à la loi appropriée (t ou F) pour tester si x est significatif. Ces deux approches donnent d'ailleurs la même réponse à la question : " X joue-t-il un rôle significatif dans la détermination de Y ?

3.8.4. Test de normalité

Pour définir la loi de probabilité de coefficients estimés par la méthode des moindres et effectuer les tests de signification sur ces coefficients, il convient de vérifier la normalité des erreurs. Le test de Jarque et Bera, fondé sur la notion de Skewness (asymétrie) et de Kurtosis (aplatissement), permet de vérifier la normalité d'une distribution statistique. Basé sur les résidus de la régression, le test de Jarque et Bera est un test pour les grands échantillons.

3.8.4.1. Les tests du Skewness et du Kurtosis

Le coefficient du Skewness est égal à :

$$S = \frac{\mu_3}{\mu_2^{3/2}}$$

3.112

Et le coefficient de Kurtosis :

$$K = \frac{\mu_4}{\mu_2^2} \quad 3.113$$

où $\mu_k = \frac{1}{n} \sum (x_t - \bar{x})^k$ représente le moment centré d'ordre k .

Si la distribution est normale et le nombre d'observations grand ($n > 30$), alors :

$$S \sim N\left(0; \sqrt{\frac{6}{n}}\right) \quad 3.114$$

et

$$K \sim N\left(3; \sqrt{\frac{24}{n}}\right) \quad 3.115$$

On construit ainsi les statistiques :

$v_1 = \frac{|S-0|}{\sqrt{\frac{6}{n}}}$ et $v_2 = \frac{|K-3|}{\sqrt{\frac{24}{n}}}$ que l'on compare à 1,96 valeur de la loi normale au seuil de 5%).

Si les hypothèses $H_0: v_1 = 0$ (symétrie) et $H_0: v_2 = 0$ (aplatissement normal) sont vérifiées, alors $v_1 < 1,96$ et $v_2 < 1,96$; dans ce cas contraire, l'hypothèse de normalité est rejetée.

3.8.4.2. Le test de JARQUE et BERA

Il s'agit d'un test de normalité qui synthétise les tests de Skewness et de Kurtosis. En effet, si S et K obéissent à des lois normales, alors la statistique :

$$JB = \frac{n}{6} S^2 + \frac{n}{24} (K - 3)^2 \quad 3.116$$

suit une loi du khi-carré à deux degrés de liberté $X_{(2)}^2$.

La formulation de l'hypothèse est donc :

H_0 : les résidus sont normalement distribués ;

Contre H_0 : les résidus ne sont pas normalement distribués.

Et la règle de décision pour un seuil spécifié ε est la suivante :

- Si $JB < X_{1-\varepsilon; (2)}^2$ où accepte l'hypothèse H_0 de normalité des résidus au seuil ε ;
- Si $JB \geq X_{1-\varepsilon; (2)}^2$ où rejette l'hypothèse H_0 , donc la distribution des résidus n'est pas normale.

3.9. LA PREVISION DANS LE MODELE DE REGRESSION SIMPLE

L'un des objectifs essentiels de l'économétrie est de fournir des prévisions. Lorsque les coefficients du modèle ont été estimés, il est possible de calculer les prévisions à un horizon h .

Soit le modèle estimé sur la prévision $t = 1, 2, \dots, n$:

$$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 X_t \quad 3.117$$

Si la valeur de la variable explicative X_t est supposée connue en la période $t + h$, c'est-à-dire X_{t+h} , la prévision ponctuelle est donnée par :

$$\hat{Y}_{t+h} = \hat{\beta}_0 + \hat{\beta}_1 X_{t+h} \quad 3.118$$

La vraie valeur prise par y lors de la période sur laquelle porte la prévision sera :

$$Y_{t+h} = \beta_0 + \beta_1 X_{t+h} + u_{t+h} \quad 3.119$$

Où u_{t+h} est le terme d'erreur lors de la période $t + h$.

Ainsi, l'erreur de prévision peut être définie par :

$$\begin{aligned} e_{t+h} &= Y_{t+h} - \hat{Y}_{t+h} \\ &= u_{t+h} - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1) X_{t+h} \end{aligned} \quad 3.120$$

En se référant aux hypothèses fondamentales du modèle, on a :

1. Espérance mathématique de l'erreur de prévision

$$E(e_{t+h}) = 0 \quad 3.121$$

Car $E(u_{t+h}) = 0$ et que $\hat{\beta}_1$ et $\hat{\beta}_0$ sont respectivement les estimateurs non biaisés de β_0 et β_1 . Est une prévision non biaisée de Y_{t+h} .

2. Variance de l'erreur de prévision

$$\begin{aligned} \text{Var}(e_{t+h}) &= E[(Y_{t+h} - \hat{Y}_{t+h}) - E(Y_{t+h} - \hat{Y}_{t+h})]^2 \\ &= E[(Y_{t+h} - \hat{Y}_{t+h})]^2 = E(e_{t+h})^2 \\ &= \text{Var}(e_{t+h}) + \text{Var}(\hat{\beta}_0) + X_{t+h}^2 \text{Var}(\hat{\beta}_1) + 2X_{t+h} \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) \\ &= \sigma_u^2 \left[1 + \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x_t^2} \right) + \frac{X_{t+h}^2}{\sum x_t^2} - \frac{2X_{t+h}\bar{X}}{\sum x_t^2} \right] \\ &= \sigma_u^2 \left[1 + \frac{1}{n} + \frac{(X_{t+h} - \bar{X})^2}{\sum x_t^2} \right] \end{aligned} \quad 3.122$$

On remarque que la variance de l'erreur de prévision est donc minimale lorsque $X_{t+h} = \bar{X}$ et elle augmente de manière non linéaire au fur et à

mesure que X_{t+h} s'éloigne de \bar{X} .

$$\hat{Y}_{t+h} \pm t_{(n-2); \frac{\varepsilon}{2}} \cdot \sigma_u \sqrt{1 + \frac{1}{n} + \frac{(X_{t+h} - \bar{X})^2}{\sum x_t^2}} \quad 3.123$$

Ou encore :

$$(\hat{\beta}_0 + \hat{\beta}_1 X_{t+h}) \pm t_{(n-2); \frac{\varepsilon}{2}} \cdot \sigma_u \sqrt{1 + \frac{1}{n} + \frac{(X_{t+h} - \bar{X})^2}{\sum x_t^2}} \quad 3.124$$

Comme l'erreur de prévision e_{t+h} suit une loi normale, on peut donc, de la manière habituelle, trouver pour e_{t+h} un intervalle de confiance au seuil de signification ε :

Ou encore

$$E(Y_{t+h}) = \hat{\beta}_0 + \hat{\beta}_1 X_{t+h} \quad 3.125$$

Il arrive parfois que l'intérêt de l'analyse porte non pas sur la prévision de la valeur effective de Y_{t+h} mais plutôt et son espérant mathématique :

Dans cas, l'erreur de prévision devient :

$$\begin{aligned} e_{t+h} &= E(Y_{t+h}) - \hat{Y}_{t+h} \\ &= (\beta_0 + \beta_1 X_{t+h}) - (\hat{\beta}_0 + \hat{\beta}_1 X_{t+h}) \\ &= (\beta_0 - \hat{\beta}_0)(\beta_1 - \hat{\beta}_1) X_{t+h} \end{aligned} \quad 3.126$$

En conséquence :

$$E(e_{t+h}) = 0 \quad 3.127$$

$$\text{Var}(e_{t+h}) = \sigma_u^2 \cdot \left[\frac{1}{n} + \frac{(X_{t+h} - \bar{X})^2}{\sum x_t^2} \right] \quad 3.128$$

Et l'intervalle de confiance pour $E(Y_{n+h})$ au seuil de signification ε est donné par :

$$\hat{Y}_{t+h} \pm t_{(n-2); \frac{\varepsilon}{2}} \cdot \sqrt{\frac{1}{n} + \frac{(X_{t+h} - \bar{X})^2}{\sum x_t^2}} \quad 3.129$$

Exemple 1

On dispose, pour un secteur industriel donné et sur une période de onze années, de la série du nombre des salariés (Y) et du chiffre d'affaires (X) du

secteur.

Année	Nombre de salariés (en milliers)	Chiffre d'affaires (en millions de CDF)
1987	294	624
1988	271	661
1989	314	728
1990	356	782
1991	383	819
1992	369	869
1993	402	938
1994	422	1023
1995	475	1136
1996	475	1227
1997	486	1317

On veut tester la réalité d'une relation linéaire entre Y et X , soit :

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

(où $t = 1, 2, \dots, 11$)

- donner les estimations $\hat{\beta}_0$ et $\hat{\beta}_1$ de β_0 et β_1 obtenues par la méthode des moindres carrés ordinaires.
- Calculer l'estimation de la variance résiduelle ($\hat{\sigma}_u$), les estimations des variances des estimateurs $\hat{\beta}_0$ et $\hat{\beta}_1$, et le coefficient de détermination de la régression R^2 .
- On pose le problème de test :

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$$

A quelle question ce test répond – t – il ? Peut – on dire que β_1 est significativement différent de zéro, au risque de 0.05 ?

- On suppose que, connu de façon exogène, le chiffre d'affaires en 2000 est de 1772 millions. Sur la base de la relation linéaire estimée à partir des données fournies, à combien des salariés peut – on s'attendre en 2000 ? A posteriori, on dénombre, en 1970, 514000 salariés. Que pouvez – vous déduire de cette observation.

Année	Y_t	X_t	$y_t = Y_t - \bar{Y}$	$x_t = X_t - \bar{X}$	$x_t y_t$	x_t^2	\hat{Y}_t	\hat{y}_t	$e_t = y_t - \hat{y}_t$	e_t^2	y_t^2
1987	294	624	-92.0909	-296.3636	27292.3966	87831.4047	293.5458	-92.5455	0.4545	0.2066	8480.7355
1988	271	661	-115.0909	-259.3636	29850.3966	67269.4957	305.0998	-80.9915	-34.0994	1162.7709	13245.9173
1989	314	728	-72.0909	-192.3636	13867.6694	37003.7685	326.0219	-60.0694	-12.0215	144.5169	5197.0992
1990	356	782	-30.0909	-138.3636	4163.4876	19144.4958	342.8845	43.2968	13.1159	172.0269	905.4628
1991	383	819	-3.0909	-101.3636	313.3058	10274.5867	354.4385	-31.6525	28.5619	815.7829	9.5537
1992	369	869	-17.0909	-51.3636	877.8512	2638.2231	370.0520	-16.0393	-1.0516	1.1058	292.0992
1993	402	938	15.9091	17.6364	280.5785	311.0413	391.5986	5.5073	10.4018	108.1971	253.0992
1994	422	1023	35.9091	102.6364	3685.5786	10534.2232	418.1416	32.0503	3.8588	14.8906	1289.4628
1995	475	1136	88.9091	215.6364	19172.0331	46499.0415	453.4281	67.3368	21.5723	465.3651	7904.8265
1996	475	1227	88.9091	306.6364	27262.7604	94025.8597	481.8446	95.7533	-6.8442	46.8437	7904.8265
1997	486	1317	99.9091	396.6364	39627.5786	157320.4050	509.9489	123.8576	-23.9485	573.5329	9981.8265
	4247	10124	0	0	166393.6360	532852.5450			0	3505.2395	55464.9091

$$n = 11$$

$$\bar{Y} = \frac{\sum Y_t}{n} = \frac{4247}{11} = 386.090909$$

$$\bar{X} = \frac{\sum x_t}{n} = \frac{10124}{11} = 920.363636$$

- I. Calcul des estimations de $\hat{\beta}_0$ et $\hat{\beta}_1$

$$\hat{\beta}_1 = \frac{\sum x_t y_t}{\sum x_t^2} = \frac{166393;6360}{532852;5450} = 0.31227$$

$$\begin{aligned}\hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X} = 386.0909 - (0.31227)(920.3636) \\ &= 98.68935\end{aligned}$$

La droite de régression de Y sur X est :

$$\boxed{Y_t = 98.68935 + 0.31227x_t}$$

- II. La somme des carrés des résidus :

$$\begin{aligned}\sum e_t^2 &= \sum y_t^2 - \sum \hat{y}_t^2 \\ \sum_{SCR} e_t^2 &= \sum y_t^2 - \hat{\beta}_1^2 \sum x_t^2 \\ SCR &= SCT - SCE\end{aligned}$$

$$SCT = \sum y_t^2 = 55467.9091$$

$$\begin{aligned}SCE &= \sum \hat{y}_t^2 = \hat{\beta}_1^2 \sum x_t^2 \\ &= (0.31227)^2 (532852.5450) \\ &= 51959.8120\end{aligned}$$

$$\begin{aligned}SCR &= SCT - SCE = 55467.9091 - 51959.8120 \\ &= 3505.0971\end{aligned}$$

$$\text{Donc } \boxed{\sum e_t^2 = 3505.0971}$$

- La variable résiduelle estimée :

$$\hat{\sigma}_u^2 = \frac{\sum e_t^2}{n - 2} = \frac{3505.0971}{11 - 2} = 389.4552$$

- La variance estimée de $\hat{\beta}_0$

$$\begin{aligned}\hat{\sigma}_{\hat{\beta}_0}^2 &= \hat{\sigma}_u^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum x_t^2} \right] \\ &= 389.4552 \left[\frac{1}{11} + \frac{(920.3636)^2}{532852.5450} \right] = 654.5172\end{aligned}$$

Et l'écart - type de $\hat{\beta}_0$:

$$\hat{\sigma}_u = \sqrt{654.5172} = 25.5835$$

- La variance estimée de $\hat{\beta}_0$

$$\hat{\sigma}_{\beta_1}^2 = \frac{\hat{\sigma}_u^2}{\sum x_t^2} = \frac{389.4552}{532852.5450} = 0.0007309$$

et son écart – type :

$$\hat{\sigma}_{\beta_1} = \sqrt{0.0007309} = 0.027035$$

- Le coefficient de détermination

$$R^2 = \frac{SCE}{SCT} = \frac{51959.8120}{55467.9091} = 0.93673$$

3. Tests d'hypothèses et intervalles de confiance

a) Intervalle de confiance à 95%

- Pour β_0

$$\begin{aligned} IC &= \hat{\beta}_0 \pm t_{\epsilon/2; (n-2)} \cdot \hat{\sigma}_{\hat{\beta}_0} = \hat{\beta}_0 \pm t_{0.025} \cdot \hat{\sigma}_{\hat{\beta}_0} \\ &= 98.68935 \pm (2.2620)(25.5835) \\ &= 98.68935 \pm 57.86988 \end{aligned}$$

$$\Rightarrow IC = [40.81947 ; 156.55923]$$

Notons que : $t_{0.025; (9)} \cdot \hat{\sigma}_{\hat{\beta}_0}$ (voir table de Student)

- Pour β_1

$$\begin{aligned} IC &= \hat{\beta}_1 \pm t_{0.025; (9)} \cdot \hat{\sigma}_{\hat{\beta}_1} \\ &= 0.31227 \pm (2.2620)(0.027035) \\ &= [0.251117 ; 0.373423] \end{aligned}$$

b) Test d'hypothèses

- Tester $\beta_1 = 0$ revient à tester l'influence significative de X dans l'explication de Y.
Sous l'hypothèse nulle : $H_0 : \beta_1 = 0$, la statistique t est donc :

$$t_{calculé} = \frac{|\hat{\beta}_1|}{\hat{\sigma}_{\hat{\beta}_1}} = \frac{0.31227}{0.027035} = 11.55058$$

Etant donné que :

$t_{calculé} = 11.55$ est supérieur au $t_{table} = 2.2620$, nous rejetons l'hypothèse nulle. Par conséquent, le coefficient de régression est fortement significatif.

- La constante α est aussi significativement différente de zéro puisque :

$$\frac{|\hat{\beta}_0|}{\hat{\sigma}_{\hat{\beta}_0}} = \frac{98.68935}{25.5835} = 3.857539 > 2.2620$$

- c) Le Test D'analyse de la Variance (test F)

La statistique F est définie par :

$$\begin{aligned} F_{\text{calculé}} &= \frac{SCE/1}{SCR/(n-2)} \\ &= \frac{51959,8120}{3505,0971/9} = \frac{51959,8120}{389,45523} \\ &= 122,41665 \end{aligned}$$

Alors que $F_{0,95;(1;9)} = 5,12$. Donc, nous rejetons $H_0 : \beta_1 = 0$

Le test d'analyse de la variance n'est ici rien d'autre que le appliqué à β_1 pour juger si X joue un rôle significatif ou non : c'est le même test sous une autre forme.

4. Préviation

En supposant que le chiffre d'affaire en 2000 est de 1772 millions de FC, le nombre des salariés prévu pour cette période est donc :

$$\begin{aligned} IC &= \hat{Y}_{2000} \pm t_{0.025} \hat{\sigma}_u \sqrt{1 + \frac{1}{n} + \frac{X_0 - \bar{X}}{\sum x_t^2}} \\ &= 652 \pm (2.262)(19.73462) \sqrt{1 + \frac{1}{11} + \frac{(1772 - 920)^2}{532852.5450}} \\ &= 652 \pm 70 \end{aligned}$$

$$\text{Donc } 582 < Y_{2000} < 722$$

Ex – post, on observe $Y_{2000} = 514$ (en millions). Il semble que l'on doive douter de l'utilisation en prévision du modèle ainsi ajusté.

Exemple 2.

Soit la fonction de demande keynésienne :

$$C = a_0 + a_1 y_d$$

Où C = consommation ;

y_d = revenu disponible ;

a_0 = consommation autonome ou disponible ;

a_1 = propension marginale à consommer.

Les données statistiques sur le niveau de consommation et du revenu disponible pour la période allant de 1993 à 2002 sont reproduites dans le

tableau ci – après :

Année	Consommation	Revenu disponible
1993	7.4	8.0
1994	8.2	9.0
1995	8.8	9.5
1996	8.7	9.5
1997	8.8	9.8
1998	9.6	11.0
1999	10.6	12.0
2000	11.2	13.0
2001	12.8	15.0
2002	13.9	16.0

t	Y_t	X_t	y_t	x_t	$x_t y_t$	x_t^2	y_t^2	\hat{y}_t	e_t	e_t^2	\hat{y}_t^2	X_t^2
1993	7.4	8.0	-2.6	-3.28	8.5280	10.7584	6.76	-2.57109	-0.02891	0.00084	6.61050	64.00
1994	8.2	9.0	-1.8	-2.28	4.1040	5.1984	3.24	-1.78722	-0.01278	0.00016	3.19416	81.00
1995	8.8	9.5	-1.2	-1.78	2.1360	3.1684	1.44	-1.30529	0.19529	0.03814	1.94683	90.25
1996	8.7	9.5	-1.3	-1.78	2.3140	3.1684	1.69	-1.30529	0.09529	0.00908	1.94683	90.25
1997	8.8	9.8	-1.2	-1.48	1.7760	2.1904	1.44	-1.16013	-0.03987	0.00159	1.34590	96.04
1998	9.6	11.0	-0.4	-0.28	0.1120	0.0784	0.16	-0.21948	-0.18052	0.03259	0.04817	121.00
1999	10.6	12.0	0.6	0.72	0.4320	0.5184	0.36	0.56439	0.03561	0.00127	0.13854	144.00
2000	11.2	13.0	1.2	1.72	2.0640	2.9584	1.44	1.34826	-0.14826	0.02198	1.81781	169.00
2001	12.8	15.0	2.8	3.72	10.4160	13.8384	7.84	2.91600	-0.11600	0.01346	8.50306	225.00
2002	13.9	16.0	3.9	4.72	18.4080	22.2784	15.21	3.69987	0.20013	0.04005	13.68904	256.00
Σ	100.0	112.8	0	0	50.2900	64.1560	39.58	0.00002	-0.00002	0.15916	39.42084	1336.54

$$\bar{X} = \frac{\sum_t x_t}{n} = \frac{112.8}{10} = 11.28$$

$$\bar{Y} = \frac{\sum_t y_t}{n} = \frac{100.0}{10} = 10.0$$

III. Calcul des estimations de $\hat{\beta}_0$ et $\hat{\beta}_1$

$$\hat{\beta}_1 = \frac{\sum_{t=1}^n x_t y_t}{\sum_{t=1}^n x_t^2} = \frac{50.29}{64.1560} = 0.78387$$

$$\begin{aligned}\hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X} = 10.0 - (0.78387)(11;28) \\ &= 10.0 - 8.84205 = 1.15795\end{aligned}$$

Donc :

$$\hat{y}_t = 0.78387x_t \quad (1)$$

Ou bien :

$$\hat{Y}_t = 1.15795 + 0.78387x_t \quad (2)$$

IV. Décomposition de la somme des carrées

$$SCT = \sum_{t=1}^n y_t^2 = 39.58$$

$$SCE = \hat{\beta}_1^2 \sum_{t=1}^n x_t^2 = (0.78387)^2 (64.1560) = 39.4208$$

$$SCR = \sum_{t=1}^n e_t^2 = SCT - SCE = 39.58 - 39.4208 = 0.1592$$

V. Le coefficient de détermination

$$R^2 = \frac{SCE}{SCT} = \frac{39.4208}{39.58} = 0.99598$$

La part de la variation de la consommation expliquée par la régression linéaire est de 99%.

VI. Calcul des variances

1. Estimation de la variance de l'erreur

$$\hat{\sigma}_u^2 = \frac{\sum e_t^2}{n-2} = \frac{0.1592}{(10-2)} = 0.0199$$

et son écart – type est égal à :

$$\hat{\sigma}_u = \sqrt{0.0199} = 0.1411$$

2. Variance estimée de $\hat{\beta}_1$

$$\hat{\sigma}_{\hat{\beta}_1}^2 = \frac{\hat{\sigma}_u^2}{\sum_{t=1}^n x_t^2} = \frac{0.0199}{64.1560} = 0.00031$$

et $\hat{\sigma}_{\hat{\beta}_1} = 0.01761$

3. Variance estimée de $\hat{\beta}_0$

$$\hat{\sigma}_{\hat{\beta}_0}^2 = \hat{\sigma}_u^2 \cdot \frac{\sum_{t=1}^n X_t^2}{n \cdot \sum_{t=1}^n x_t^2} = 0.0199 \frac{(1336.54)}{(10)(64.1560)} = 0.04146$$

ou encore :

$$\begin{aligned} \hat{\sigma}_{\hat{\beta}_0}^2 &= \hat{\sigma}_u^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum x_t^2} \right] = 0.0199 \left[\frac{1}{10} + \frac{(11.28)^2}{(64.1560)} \right] \\ &= (0.0199)(2.0833) \\ &= 0.04146 \end{aligned}$$

Son écart – type est donc :

$$\hat{\sigma}_{\hat{\beta}_0} = \sqrt{0.04146} = 0.20362$$

VII. Intervalles de confiance

1. Intervalle de confiance à 95% pour α

$$\beta_0 \in \hat{\beta}_0 \pm \hat{\sigma}_{\hat{\beta}_0} \cdot t_{\frac{\alpha}{2}; (n-2)}$$

$$\beta_0 \in 1.15795 \pm (0.20362)(2.3060)$$

$$\in 1.15795 \pm 0.46955$$

$$\text{Donc IC} = [0.6884 ; 1.6273]$$

2. Intervalle de confiance à 95% pour β_1

$$\beta_1 \in \hat{\beta}_1 \pm \hat{\sigma}_{\hat{\beta}_1} \cdot t_{0.025; 8}$$

$$\beta_1 \in 0.78387 \pm (0.01761)(2.3060)$$

$$\in 0.78387 \pm 0.04069$$

$$\text{Donc IC} = [0.74318 ; 0.82456]$$

3. Intervalle de confiance à 95% pour σ_u^2

$$\sigma_u^2 \in \left[\frac{(n-2)\hat{\sigma}_u^2}{\chi_{0.975}^2; (n-2)} ; \frac{(n-2)\hat{\sigma}_u^2}{\chi_{0.025}^2; (n-2)} \right]$$

D'après la table du χ^2 , avec 8 degrés de liberté, on trouve : $\chi_{0.025; 8}^2 = 2.18$ et $\chi_{0.975; (n-2)}^2 = 17.53$. Ainsi, un intervalle de confiance à 95% pour σ_u^2

sera donné par les valeurs comprises entre :
 $IC = [0.00908 ; 0.7303]$

Remarques

Notons ici que $\chi^2_{0.975}$ est un nombre dont la probabilité que le χ^2 lui soit supérieur est égale à 0.025. de même, la probabilité pour que le χ^2 dépasse $\chi^2_{0.025}$ est égale à 0.975.

VIII. Tests d'hypothèses

1. Pour β_0

$$\begin{cases} H_0 : \beta_0 = 0 \\ H_1 : \beta_0 \neq 0 \end{cases}$$

Sous l'hypothèse H_0 , le ratio de Student (le t calculé) est donné par :

$$t_{\alpha}^* = \frac{|\hat{\beta}_0|}{\hat{\sigma}_{\hat{\beta}_0}} = \frac{1.15795}{0.20362} = 5.69$$

La table de Student donne :

$$t_{0.025;8} = 2.3060$$

Décision : $t_{\hat{\beta}_0}^* > t_{table}$, il y a donc rejet de l'hypothèse nulle. La constante α du modèle est significativement différente de zéro.

2. Pour β_1

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$$

Sous l'hypothèse H_0 , le ratio de Student est égal à :

$$t_{\beta}^* = \frac{|\hat{\beta}_1|}{\hat{\sigma}_{\hat{\beta}_1}} = \frac{0.78387}{0.01761} = 44.5128$$

Décision : $t_{\hat{\beta}_1}^* > t_{table}$, rejet de l'hypothèse nulle.

Donc la propension marginale à consommer est significativement différent de zéro. Ce résultat implique que la variable « revenu disponible » est bien explicativement de la variable « consommation ».

IX. Test de Fisher

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$$

Sous l'hypothèse nulle H_0 , la statistique F de l'échantillon est donc :

$$F^* = \frac{SCE/1}{SCR/(n-2)} = \frac{39.4208}{(0.1592)/8} = 1980.94$$

Alors que $F_{0.95 ; (1,8)} = 5.32$

Donc, nous rejetons $H_0 : \beta_1 = 0$. La régression est globalement significative.

X. Prédiction

En 2003 et 2004, on prévoit respectivement 16.8 et 17.0 de revenu par habitant. Déterminer la prévision de consommation pour ces deux années ainsi que l'intervalle de prévision ou seuil de signification de 5%.

1. Prédiction ponctuelle

Les prévisions ponctuelles sont calculées par l'utilisation du modèle estimé.

$$\begin{aligned} \hat{Y}_{2003} &= 1.15795 + 0.78387 X_{2003} \\ &= 1.15795 + (0.78387)(16.8) = 14.3 \end{aligned}$$

$$\begin{aligned} \hat{Y}_{2004} &= 1.15795 + 0.78387 X_{2004} \\ &= 1.15795 + (0.78387)(17) = 14.5 \end{aligned}$$

2. Intervalles de prévision

L'intervalle de prévision peut alors être calculé à partir de la formule ci-dessous :

$$Y_{2003} = \hat{Y}_{2003} \pm t_{0.025 ; 8} \cdot \hat{\sigma}_u \sqrt{1 + \frac{1}{n} + \frac{(X_{2003} - \bar{X})^2}{\sum_{t=1}^n x_t^2}}$$

CHAPITRE 4

MODELE DE REGRESSION MULTIPLE

4.1. Exposé du problème

Soient X_1, X_2, \dots, X_k : k variables indépendantes et non aléatoires et Y une variable aléatoire dépendant de X_1, X_2, \dots, X_k . On dit qu'on est en présence d'un modèle linéaire général quand on a la relation :

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_i X_{it} + \dots + \beta_k X_{kt} + u_t \quad (4.1)$$

où u_t est une variable aléatoire centrée de variance σ_u^2 , indépendante des variables X_i et $\beta_1, \beta_2, \dots, \beta_k$ des paramètres réels mais inconnus. On dit que la variable à expliquer Y est une variable endogène et que les variables explicatives X_1, X_2, \dots, X_k sont des variables exogènes. La variable aléatoire u est l'écart ou l'erreur.

On désire estimer les paramètres $\beta_1, \beta_2, \dots, \beta_k$ et la variance σ_u^2 de la variable aléatoire u à partir d'un ensemble de n observations indépendantes :

$$\begin{cases} Y_t \text{ où } t = 1, 2, \dots, n \\ \text{et} \\ X_{it} \text{ où } i = 1, 2, \dots, k \end{cases}$$

4.2. Notation matricielle du modèle

En posant :

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}; \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}; U = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}; \begin{bmatrix} X_{11} & \dots & X_{k1} \\ \vdots & \ddots & \vdots \\ X_{1n} & \dots & X_{kn} \end{bmatrix}$$

On obtient la forme matricielle :

$$\underbrace{Y}_{(n \times 1)} = \underbrace{X}_{(n \times k)} \underbrace{\beta}_{(k \times 1)} + \underbrace{U}_{(n \times 1)} \quad (4.2)$$

L'introduction d'un terme constant s'obtient en donnant à la valeur X_{it} la valeur 1 pour tout $t = 1, \dots, n$.

Exemple : Considérons le modèle linéaire simple (Fonction de consommation keynésienne) :

$$C_t = \beta_1 + \beta_2 Y_{dt} + u_t$$

($t = 1, 2, \dots, n$)

où : C_t = consommation au temps t ;

Y_{dt} = revenu disponible au temps t .

Ce modèle peut aussi s'écrire sous forme de matrice comme ci-après :

$$\underbrace{C}_{(n \times 1)} = \underbrace{Y_d}_{(n \times 2)} \underbrace{\beta}_{(2 \times 1)} + \underbrace{U}_{(n \times 1)}$$

$$C = \begin{bmatrix} C_1 \\ \vdots \\ C_n \end{bmatrix}; \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}, U = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \quad Y_d = \begin{bmatrix} 1 & Y_{d1} \\ \vdots & \vdots \\ 1 & Y_{dn} \end{bmatrix}$$

L'écriture sous forme matricielle rend plus aisée la manipulation du modèle linéaire général, c'est pourquoi nous l'adoptons par la suite.

4.3. Estimation et propriétés des estimateurs

4.3.1. Hypothèses d'application de la méthode des moindres carrés ordinaires.

Deux catégories d'hypothèses doivent être faites pour résoudre le problème des moindres carrés dans le cas d'un modèle linéaire général. Il s'agit des hypothèses stochastiques (liées à l'erreur u), d'une part, et des hypothèses structurelles, de l'autre.

A. Hypothèses stochastiques

- H_1 : X est une matrice non aléatoire, c'est-à-dire les valeurs de X sont observées sans erreur.
- H_2 : L'espérance mathématique de l'erreur est nulle.

$$E(U) = 0 \quad (4.3)$$

Cette hypothèse peut s'écrire aussi comme ci-après :

$$E(U) = \begin{bmatrix} E(u_1) \\ E(u_2) \\ E(u_3) \\ \vdots \\ E(u_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (4.4)$$

De la relation (4.4), on déduit :

$$\begin{aligned} E(Y) &= E(X\beta + U) \\ &= E(X\beta) + E(U) = X\beta. \end{aligned}$$

$$\begin{aligned} - \quad H_3 : \quad \text{Var}(U) &= \{[U - E(U)][U - E(U)]'\} = E[UU'] = \sigma_u^2 I = \Omega_u \\ (4.5) \end{aligned}$$

où Ω_u est la matrice des variantes et covariances de l'erreur U .
L'hypothèse H_3 s'écrit en détail comme ci-dessous:

$$\begin{aligned} E(UU') &= E \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} [u_1 \quad u_2 \quad \cdots \quad u_n] \\ &= \begin{bmatrix} E(u_1^2) & E(u_1 u_2) & \cdots & \cdots & E(u_1 u_n) \\ E(u_2 u_1) & E(u_2^2) & \cdots & \cdots & E(u_2 u_n) \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ E(u_n u_1) & E(u_n u_2) & \cdots & \cdots & E(u_n^2) \end{bmatrix} \end{aligned}$$

$$= \sigma_u^2 \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & \dots & \dots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \dots & 1 \end{bmatrix} = \sigma_u^2 I = \underbrace{\Omega_{yy}}_{(n \times n)} \quad (4.6)$$

En principe, on fait ici deux hypothèses :

- Les lois de probabilité suivies par les erreurs U ont la même variance.

Donc $E(u_t^2) = \sigma_u^2$, $\forall t$ (variance constante et finie). C'est ce qu'on appelle l'*homoscédasticité*, c'est-à-dire que la dispersion ou variance du terme d'erreur est la même du début à la fin de la période. Lorsque cette hypothèse n'est pas vérifiée, on dit qu'il y a *hétéroscédasticité* (dispersion ou variance inégale).

- Les erreurs, prises deux à deux, ne sont pas corrélées.

$$E(u_i u_j) = 0, \quad \forall i \neq j$$

(absence d'autocorrélation ou absence de corrélation sérielle).

Si tel n'est pas le cas, on dit que les erreurs sont *autocorrélées*.

- H_4 : Le vecteur de perturbations U suit une loi normale à n dimensions.
Donc :

$$U \sim N(0, \sigma_u^2 I) \quad (4.7)$$

Cette hypothèse de normalité des erreurs est indispensable, comme nous l'avons souligné plus haut, si on veut utiliser le principe d'estimation du maximum de vraisemblance mais aussi pour procéder aux tests d'inférences.

- H_5 : Covariance nulle entre U et X , c'est-à-dire :

$$\text{Cov}(u_i, X_{ij}) = E(u_i X_{ij}) = 0 \quad (4.8)$$

B. Hypothèses structurelles

- H_5 : absence de colinéarité entre les variables explicatives; cela implique que la matrice $(X'X)$ est non singulière et que la matrice $(XX')^{-1}$ existe. Donc le rang de la matrice des variables explicatives doit être égal à k : $\rho(X) = k$.
- H_6 : le nombre d'observations doit être supérieur au nombre de paramètres à estimer ($n > k$).

4.3.2. Estimateurs des moindres carrés ordinaires

Soit le modèle sous forme matricielle à k variables explicatives et n observations :

$$Y = X\beta + U \quad (4.9)$$

Le vecteur des résidus (e) s'obtient en remplaçant le vecteur inconnu des paramètres β par son estimateur $\hat{\beta}$ dans l'équation (4.9). D'où :

$$e = Y - X\hat{\beta} = Y - \hat{Y} \quad (4.10)$$

Pour déterminer l'estimateur de β , on applique le principe de la méthode des moindres carrés qui consiste à minimiser la somme des carrés des résidus : ee' , soit :

$$\text{Min } e'e = \text{Min } \sum_{t=1}^n e_t^2 = \text{Min } \sum_{t=1}^n [Y_t - (\hat{\beta}_1 + \hat{\beta}_2 X_{2t} + \dots + \hat{\beta}_k X_{kt})]^2 \quad (4.11)$$

En écriture matricielle, on sait que :

$$\begin{aligned} \sum_{t=1}^n e_t^2 &= e'e = (Y - \hat{Y})'(Y - \hat{Y}) \\ &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) \\ &= Y'Y - Y'X\hat{\beta} - \hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta} \end{aligned}$$

Etant donné que $Y'X\hat{\beta} = \hat{\beta}'X'Y$ (car les deux produits sont des scalaires), on peut alors écrire :

$$\text{SCR} = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta} \quad (4.12)$$

Pour minimiser cette expression par rapport à $\hat{\beta}$ on annule sa dérivée par rapport à $\hat{\beta}$:

$$\frac{\delta \text{SCR}}{\delta \hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0 \quad (4.13)$$

La résolution de la relation (4.13) permet de trouver les équations normales sous leur forme matricielle :

$$X'Y = X'X\hat{\beta} \quad (4.14)$$

La relation (4.14) peut s'écrire aussi comme ci-après :

$$\underbrace{\begin{bmatrix} n & \sum X_{2t} & \sum X_{3t} & \dots & \sum X_{kt} \\ \sum X_{2t} & \sum X_{2t}^2 & \sum X_{2t}X_{3t} & \dots & \sum X_{2t}X_{kt} \\ \sum X_{3t} & \sum X_{3t}X_{2t} & \sum X_{3t}^2 & \dots & \sum X_{3t}X_{kt} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum X_{kt} & \sum X_{kt}X_{2t} & \sum X_{kt}X_{3t} & \dots & \sum X_{kt}^2 \end{bmatrix}}_{X'X} \underbrace{\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}}_{\hat{\beta}} = \underbrace{\begin{bmatrix} \sum Y_t \\ \sum X_{2t}Y_t \\ \sum X_{3t}Y_t \\ \vdots \\ \sum X_{kt}Y_t \end{bmatrix}}_{X'Y} \quad (4.15)$$

En multipliant la relation (4.14) par $(X'X)^{-1}$, on obtient l'expression de $\hat{\beta}$.

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (4.16)$$

La résolution de cette équation n'est possible que si la matrice $(X'X)$ est inversible.

Exemples

Ex. 1 : Pour un modèle linéaire simple de type : $Y_t = \beta_1 + \beta_2 X_{2t} + u_t$, la relation (3.15) devient :

$$\begin{bmatrix} n & \sum X_{2t} \\ \sum X_{2t} & \sum X_{2t}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \sum Y_t \\ \sum X_{2t}Y_t \end{bmatrix}$$

La résolution donne le système d'équations ci-après :

$$\begin{cases} \sum_{t=1}^n Y_t = n\hat{\beta}_1 + \hat{\beta}_2 \sum_{t=1}^n X_{2t} \\ \sum_{t=1}^n X_t Y_t = \hat{\beta}_1 \sum_{t=1}^n X_{2t} + \hat{\beta}_2 \sum_{t=1}^n X_{2t}^2 \end{cases}$$

Ex. 2 : Pour un modèle linéaire multiple ayant deux variables explicatives :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

où X_{2t} et X_{3t} sont les variables explicatives.

Les équations normales pour ce modèle se présentent comme suit :

$$\begin{bmatrix} n & \sum X_{2t} & \sum X_{3t} \\ \sum X_{2t} & \sum X_{2t}^2 & \sum X_{2t}X_{3t} \\ \sum X_{3t} & \sum X_{3t}X_{2t} & \sum X_{3t}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} \sum Y_t \\ \sum X_{2t}Y_t \\ \sum X_{3t}Y_t \end{bmatrix}$$

ou encore après la résolution matricielle :

$$\begin{cases} \sum_{t=1}^n Y_t = n\hat{\beta}_1 + \hat{\beta}_2 \sum_{t=1}^n X_{2t} + \hat{\beta}_3 \sum_{t=1}^n X_{3t} \\ \sum_{t=1}^n X_{2t}Y_t = \hat{\beta}_1 \sum_{t=1}^n X_{2t} + \hat{\beta}_2 \sum_{t=1}^n X_{2t}^2 + \hat{\beta}_3 \sum_{t=1}^n X_{2t}X_{3t} \\ \sum_{t=1}^n X_{3t}Y_t = \hat{\beta}_1 \sum_{t=1}^n X_{3t} + \hat{\beta}_2 \sum_{t=1}^n X_{3t}X_{2t} + \hat{\beta}_3 \sum_{t=1}^n X_{3t}^2 \end{cases}$$

4.3.3. Propriétés des estimateurs des moindres carrés ordinaires

Les propriétés des estimateurs des moindres carrés ordinaires sont, dans le modèle de régression multiple, analogues à celles du modèle de régression simple. En effet, les estimateurs des moindres carrés ordinaires $\hat{\beta} = (X'X)^{-1}X'Y$ sont non seulement linéaires et sans biais, mais aussi ont une variance minimale dans la classe de tous les estimateurs non biaisés. Ils sont donc « BLUE » (Best Linear Unbiased Estimator). En d'autres termes, ils satisfont le théorème de Gauss-Markov.

A. Estimateur linéaire

L'estimateur des moindres carrés ordinaires est une fonction linéaire de Y . En effet, soit :

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

Posons : $\gamma = (X'X)^{-1}X'$, on peut alors écrire :

$$\hat{\beta} = \gamma Y \quad (4.16)$$

Donc, les composantes de $\hat{\beta}$ sont des fonctions linéaires des composantes Y_t de Y .

B. Estimateur sans biais

L'espérance de l'estimateur $\hat{\beta}$ est donc :

$$\begin{aligned} E(\hat{\beta}) &= E(\gamma Y) \\ &= \gamma E(X\beta + U) = \gamma[X\beta + E(U)] \\ &= \gamma X\beta, \end{aligned}$$

(car $E(U) = 0$).

Par conséquent :
$$E(\hat{\beta}) = (X'X)^{-1}X'X\beta = (X'X)^{-1}(X'X)\beta = \beta$$

$\Rightarrow E(\hat{\beta}) = \beta$. Donc, $\hat{\beta}$ est un estimateur sans biais de β .

C. Calcul de la matrice des variances-covariances de $\hat{\beta}$

La matrice des variances-covariances de $\hat{\beta}$ est donnée par :

$$\begin{aligned} \Omega_{\hat{\beta}} &= \text{Var}(\hat{\beta}) = \text{Var}(\gamma Y) = \gamma \text{Var}(Y) \gamma' \\ &= \gamma \text{Var}(X\beta + U) \gamma' = \gamma \text{Var}(U) \gamma' \\ &= \gamma \cdot \sigma_u^2 I \cdot \gamma' = (X'X)^{-1} X' \cdot \sigma_u^2 I \cdot X (X'X)^{-1} \\ &= \sigma_u^2 (X'X)^{-1} (X'X) (X'X)^{-1} \end{aligned}$$

D'où :

$$\Omega_{\hat{\beta}} = \sigma_u^2 (X'X)^{-1} \quad (4.18)$$

Puisque la variance résiduelle σ_u^2 est inconnue, on l'estimera à partir des valeurs observées des résidus afin de rendre opérationnelle la relation (4.18). Donc :

$$\hat{\sigma}_u^2 = \frac{ee'}{n-k} \quad (4.19)$$

où le facteur $(n-k)$ représente le nombre des degrés de liberté.

Ainsi, la matrice des variances-covariances estimée devient :

$$\hat{\Omega}_{\hat{\beta}} = \hat{\sigma}_u^2 (X'X)^{-1} \quad (4.20)$$

Cette détermination de la matrice des variances-covariances de $\hat{\beta}$ permet de montrer que cet estimateur est meilleur, c'est-à-dire de variance minimale.

D. Estimateur de variance minimale

Montrons que, parmi la classe des estimateurs linéaires sans biais, l'estimateur $\hat{\beta}$ des moindres carrés ordinaires est celui qui possède la variance la plus faible :

$$\text{Var}(\hat{\beta}) \leq \text{Var}(\tilde{\beta}) \quad (4.21)$$

où $\tilde{\beta}$ représente tout autre estimateur sans biais de β .

En effet, si $\tilde{\beta}$, fonction linéaire des Y_t , est un estimateur sans biais de β , on doit avoir :

$$\tilde{\beta} = MY \quad \text{et} \quad E(\tilde{\beta}) = \beta$$

Comme $Y = X\beta + U$, on a $\tilde{\beta} = MX\beta + MU$.

Donc la matrice M doit vérifier : $MX\beta = \beta$ ou $MX = I$, avec I est la matrice identité d'ordre k .

Soit $\tilde{\beta} = MY$, estimateur sans biais différent de $\hat{\beta} = (X'X)^{-1}X'Y$.

Posons : $M = (X'X)^{-1}X' + N$.

Comme $MX = I$, alors :

$$MX = [(X'X)^{-1}X' + N]X = (X'X)^{-1}X'X + NX = I + NX = I$$

$$\text{où } \underbrace{N}_{(k \times n)} \underbrace{X}_{(n \times k)} = \underbrace{0}_{(k \times k)}.$$

Montrons que les variances des composantes $\tilde{\beta}_1, \tilde{\beta}_2, \dots, \tilde{\beta}_k$ du vecteur $\tilde{\beta}$ sont supérieures aux composantes $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ de $\hat{\beta}$, c'est-à-dire les termes de la diagonale principale de la matrice des covariances de $\text{Var}(\tilde{\beta}) = E\{[\tilde{\beta} - E(\tilde{\beta})][\tilde{\beta} - E(\tilde{\beta})]'\}$ sont supérieurs à ceux de $\text{Var}(\hat{\beta}) = \sigma_u^2(X'X)^{-1}$. En effet, on a :

$$\tilde{\beta} - E(\tilde{\beta}) = MX\beta + MU - \beta = \beta + MU - \beta = MU,$$

car $MU = I$.

Il en résulte que :

$$\begin{aligned} \text{Var}(\tilde{\beta}) &= E\{[MU][MU]'\} = E(MUU'M') \\ &= ME(UU')M' = M \cdot \sigma_u^2 I \cdot M' = \sigma_u^2 MM'. \end{aligned}$$

Comme $M = (X'X)^{-1}X' + N$, sa transposée est égale à : $M' = X(X'X)^{-1} + N'$.

$$\begin{aligned} \text{D'où : } \text{Var}(\tilde{\beta}) &= \sigma_u^2 MM' \\ &= \sigma_u^2 [(X'X)^{-1}(X'X)^{-1} + (X'X)^{-1}X'N' + NX(X'X)^{-1} + NN'] \\ &= \sigma_u^2 [(X'X)^{-1} + NN'], \end{aligned}$$

car $NX = 0$ et $X'N' = 0$.

Par conséquent :

$$\begin{aligned} \text{Var}(\tilde{\beta}) &= \sigma_u^2(X'X)^{-1} + \sigma_u^2(NN') \\ &= \text{Var}(\hat{\beta}) + \sigma_u^2(NN') \end{aligned}$$

et

$$\text{Var}(\tilde{\beta}_i) = \text{Var}(\hat{\beta}_i) + \sigma_u^2 \sum_{i=1}^n n_i^2$$

(où $\sigma_u^2 \sum_{i=1}^n n_i^2 \geq 0$).

$$\text{D'où : } \text{Var}(\tilde{\beta}_i) \geq \text{Var}(\hat{\beta}_i), \quad \forall i = 1, 2, \dots, k \quad (4.22)$$

Il s'ensuit que, parmi la classe des estimateurs linéaires sans biais, l'estimateur des moindres carrés ordinaires $\hat{\beta}$ est celui qui possède la variance la plus faible.

4.4. Analyse de la variance et qualité de l'ajustement

4.4.1. Equation d'analyse de la variance

L'équation fondamentale d'analyse de la variance est :

$$\sum_t (Y_t - \bar{Y})^2 = \sum_t (\hat{Y}_t - \bar{Y})^2 + \sum_t (Y_t - \hat{Y}_t)^2 \quad (4.23)$$

ou en écriture matricielle :

$$(Y'Y - n\bar{Y}^2) = (\hat{\beta}'X'Y - n\bar{Y}^2) + e'e \quad (4.24)$$

Sous forme de variables centrées :

$$\sum_t y_t^2 = \sum_t \hat{y}_t^2 + \sum_t e_t^2 \quad (4.25)$$

ou encore :

$$y'y = \hat{\beta}x'y + e'e \quad (4.26)$$

Cette équation est l'équation d'analyse de la variance. Au facteur $\left(\frac{1}{n}\right)$ près, la variance empirique de Y, appelée la variance totale (SCT), est égale à la somme de la variance expliquée par le modèle (SCE) et de la variance résiduelle (SCR).

$$\begin{cases} \text{SCT} = \text{Variabilité totale} = \sum_t (Y_t - \bar{Y})^2 \\ \text{SCE} = \text{Variabilité expliquée} = \sum_t (\hat{Y}_t - \bar{Y})^2 \\ \text{SCR} = \text{Variabilité des résidus} = \sum_t (Y_t - \hat{Y}_t)^2 \end{cases}$$

A partir de cette équation, la qualité de l'ajustement peut être mesurée par le coefficient de détermination.

4.4.2. Qualité d'un ajustement

L'équation d'analyse de la variance permet de juger de la qualité de l'ajustement d'un modèle. En effet, plus la variance expliquée est «proche» de la variance totale, meilleur est l'ajustement global du modèle.

Le coefficient de détermination, noté R^2 , est défini à partir de la décomposition de la variance de Y :

$$R^2 = \frac{\text{Variance expliquée}}{\text{Variance totale}} = \frac{\text{SCE}}{\text{SCT}} = 1 - \frac{\text{SCR}}{\text{SCT}} \quad (4.25)$$

Dans le cas de variables centrées :

$$R^2 = \frac{\hat{\beta}x'y}{y'y} = \frac{\sum_t \hat{y}_t^2}{\sum_t y_t^2} \quad (4.26)$$

ou

$$R^2 = 1 - \frac{e'e}{y'y} = 1 - \frac{\sum_t e_t^2}{\sum_t y_t^2} \quad (4.27)$$

Dans le cas de variables non centrées :

$$\begin{aligned} R^2 &= \frac{\hat{\beta}X'Y - n\bar{Y}^2}{Y'Y - n\bar{Y}^2} \\ &= 1 - \frac{e'e}{Y'Y - n\bar{Y}^2} \end{aligned} \quad (4.28)$$

Par construction, cette statistique est comprise dans l'intervalle $[0, 1]$. Donc : $0 \leq R^2 \leq 1$.

Le coefficient de détermination R^2 est un indicateur de qualité de l'ajustement linéaire entre les variables explicatives et la variable expliquée. Il mesure la part de la variance totale expliquée par les variables explicatives et permet donc de juger de la qualité de l'ajustement du modèle. Une valeur proche de 1 indique que la qualité de l'ajustement est bonne dans la mesure où la part de la variance de Y expliquée par le modèle est élevée. Cependant, il doit être utilisé avec précaution. Un R^2 élevé ne doit en aucun cas être interprété comme une

mesure du degré d'explication de la variable dépendante par les variables explicatives, mais seulement comme une forte association entre ces variables. En outre, le R^2 a tendance à croître avec le nombre de variables explicatives retenues dans le modèle, indépendamment du pouvoir explicatif de ces variables. Il ya donc un biais dans l'évaluation du coefficient de détermination dû au fait que l'on a tenu compte explicitement ni du nombre de variables explicatives, ni du nombre d'observations. Pour y remédier, on a défini un coefficient de détermination ajusté, connu aussi sous le nom de R^2 corrigé, noté \bar{R}^2 , qui tient compte du nombre de variables explicatives présentes dans le modèle :

$$\bar{R}^2 = 1 - \frac{e'e/(n-k)}{y'y/(n-1)} \quad (4.29)$$

ou encore :

$$\begin{aligned} \bar{R}^2 &= 1 \\ &- \frac{(n-1)}{(n-k)}(1 - R^2) \end{aligned} \quad (4.30)$$

ou enfin :

$$\begin{aligned} \bar{R}^2 &= \left(\frac{n-1}{n-k} \right) R^2 \\ &- \left(\frac{k-1}{n-k} \right) \end{aligned} \quad (4.31)$$

Notons que :

- a) $\bar{R}^2 < R^2$, donc lorsque le nombre de variables explicatives augmente, le coefficient de détermination ajusté \bar{R}^2 croît moins que le R^2 ;
- b) $\bar{R}^2 = R^2$, si le nombre d'observations n est grand ;
- c) \bar{R}^2 peut prendre la valeur négative bien que R^2 soit exclusivement positif. Dans ce cas, la valeur de \bar{R}^2 doit être considérée comme nulle.

Sachant que la même série d'observations relatives à un phénomène donné peut avoir plusieurs représentations, la statistique \bar{R}^2 peut être utilisée pour retenir le modèle dont le pouvoir explicatif est élevé, c'est-à-dire le modèle dont le coefficient de détermination ajusté est le plus élevé. On peut également utiliser deux autres critères d'information pour le choix d'un modèle ayant un pouvoir explicatif élevé. Il s'agit de :

- critère de SCHWARTZ

$$SC = \ln\left(\frac{SCR}{n}\right) + \frac{k}{n} \ln(n) \quad (4.32)$$

- critère d'information d'AKAIKE

$$CIA = \ln\left(\frac{SCR}{n}\right) + \frac{2k}{n} \quad (4.33)$$

Ces critères visent à retenir, parmi les modèles à comparer, celui qui minimise le critère d'information.

Il convient de noter que le coefficient de détermination R^2 n'est utilisable que dans un modèle avec terme constant. L'utilisation de cette statistique dans un modèle de régression passant par l'origine (sans terme constant) peut aussi donner un résultat négatif : $R^2 < 0$.

Enfin, pour les régressions ne présentant pas de valeur en ordonnée à l'origine, on calcule parfois un coefficient de détermination spécifique, nommé R^2 brut. Il se définit comme suit :

$$R^2 \text{ brut} = \frac{(\sum X_t Y_t)^2}{\sum X_t^2 \sum Y_t^2} \quad (4.34)$$

(Il s'agit donc des sommes des carrés brutes, non corrigées de la moyenne). Bien que ce R^2 brut satisfait la relation $0 \leq R^2 \text{ brut} \leq 1$, il n'est pas directement comparable à la valeur du R^2 conventionnel.

4.5. Les tests statistiques

Soit le modèle linéaire général suivant :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_i X_{it} + \dots + \beta_k X_{kt} + u_t$$

Examinons quelques problèmes statistiques qui découlent de l'écriture de ce modèle.

- i. Peut-on comparer les paramètres β_i ($i = 1, \dots, k$) à des valeurs fixées a priori ($\beta_i = \beta$).
En particulier une question intéressante est " $\beta_i = 0$ ", ce qui revient à supposer que la variable endogène Y ne dépend pas de la variable explicative X_i (test de Student).
- ii. Le modèle linéaire est-il globalement significatif (test de Fisher) ?
- iii. Le modèle considéré est-il stable sur la totalité de la période (test de stabilité, ou de changement structurel) ?

4.5.1. Hypothèse de normalité

Comme il a été souligné dans le cadre d'un modèle de régression linéaire simple, l'hypothèse de normalité des termes d'erreur joue un rôle essentiel car elle va préciser la distribution statistique des estimateurs. En plus, c'est grâce à cette hypothèse que l'inférence statistique peut se réaliser.

En supposant que les erreurs suivent une loi normale centrée et de variance σ_u^2 alors :

$$Y = X\beta + U \sim N(X\beta, \sigma_u^2 I) \quad \text{et} \quad \hat{\beta} = (X'X)^{-1}X'Y \sim N(\beta, \sigma_u^2 (X'X)^{-1})$$

L'hypothèse de normalité des erreurs implique aussi que :

$$\frac{\sum_t e^2}{\sigma_u^2} = (n - k) \frac{\hat{\sigma}_u^2}{\sigma_u^2} \sim \chi^2_{(n-k)}.$$

4.5.2. Test de Student

A. Comparaison d'un paramètre β_i à une valeur fixée β_i^*

Le test d'hypothèse est le suivant :

$$\begin{aligned} H_0: \beta_i &= \beta_i^* \\ H_1: \beta_i &\neq \beta_i^* \end{aligned}$$

Comme $\text{Var}(\hat{\beta}) = \sigma_u^2 (X'X)^{-1}$, on a :

$$\text{Var}(\hat{\beta}_i) = \sigma_u^2 v_{ii}; \quad \hat{\beta}_i \sim N(\beta_i, \sigma_u^2 v_{ii}) \quad \text{et} \quad \frac{\hat{\beta}_i - \beta_i}{\sigma_u \sqrt{v_{ii}}} = \frac{\hat{\beta}_i - \beta_i}{\sigma_{\hat{\beta}_i}^2} \sim N(0, 1)$$

où v_{ii} , $i = 1, 2, \dots, k$ est l'élément (i, i) de la matrice $(X'X)^{-1}$.

La variance des erreurs n'étant pas connue, on peut diviser la variable centrée réduite de $\hat{\beta}_i$ par la racine carrée de la statistique $\frac{\hat{\sigma}_u^2}{\sigma_u^2}$ afin d'obtenir la valeur de la statistique t de $\hat{\beta}_i$:

$$t_{\hat{\beta}_i} = \frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_u \sqrt{v_{ii}}} = \frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_{\hat{\beta}_i}} \quad (4.,.,.,.)$$

La statistique $t_{\hat{\beta}_i}$ est distribuée suivant une loi de Student à $(n - k)$ degrés de liberté sous l'hypothèse H_0 .

En vertu de la relation (4.,.,.), il est possible de construire un intervalle de confiance au seuil de signification α pour β_i ($i = 1, 2, \dots, k$), soit :

$$\hat{\beta}_i \pm t_{\frac{\alpha}{2}; (n-k)} \cdot \hat{\sigma}_u \sqrt{v_{ii}} \quad \text{ou bien} \quad \hat{\beta}_i \pm t_{\frac{\alpha}{2}; (n-k)} \cdot \hat{\sigma}_{\hat{\beta}_i}$$

L'intervalle de confiance (...) peut également s'écrire sous la forme :

$$P \left[\hat{\beta}_i - t_{\frac{\alpha}{2}; (n-k)} \cdot \hat{\sigma}_u \sqrt{v_{ii}} < \beta_i < \hat{\beta}_i + t_{\frac{\alpha}{2}; (n-k)} \cdot \hat{\sigma}_u \sqrt{v_{ii}} \right] = 100(1 - \alpha)\%$$

Le seuil de risque α est généralement fixé à 5 % ou 10 %.

Il est alors possible de tester l'hypothèse nulle selon laquelle le coefficient β_i est égal à une valeur donnée β_i^* , soit :

$$H_0: \beta_i = \beta_i^*$$

$$H_1: \beta_i \neq \beta_i^*$$

Si l'hypothèse nulle est vraie, alors :

$$t_{\hat{\beta}_i} = \frac{\hat{\beta}_i - \beta_i^*}{\hat{\sigma}_u \sqrt{v_{ii}}} = \frac{\hat{\beta}_i - \beta_i^*}{\hat{\sigma}_{\hat{\beta}_i}}$$

La règle de décision est la suivante :

- Si $|t_{\hat{\beta}_i}| > t_{\frac{\alpha}{2}}$: on rejette l'hypothèse H_0 au seuil statistique de $100\alpha\%$. Le coefficient β_i peut être considéré significativement différent de β_i^* .
- Si $|t_{\hat{\beta}_i}| \leq t_{\frac{\alpha}{2}}$: on accepte l'hypothèse H_0 au seuil statistique de $100\alpha\%$. Le coefficient β_i peut être considéré significativement égal à β_i^* .

B. Comparaison d'un paramètre β_i à la valeur $\beta_i = 0$

Le test le plus utilisé est celui qui consiste à tester l'hypothèse nulle :

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

Il s'agit donc d'un test de significativité de coefficient. Sous l'hypothèse nulle, la variable X_i n'est pas significative, c'est-à-dire qu'elle ne joue aucun rôle dans la détermination de la variable expliquée Y .

Sous l'hypothèse nulle, la statistique t s'écrit :

$$t_{\hat{\beta}_i} = \frac{\hat{\beta}_i}{\hat{\sigma}_{\hat{\beta}_i}}$$

La règle de décision est donc :

- Si $|t_{\hat{\beta}_i}| \leq t_{\frac{\alpha}{2}}$: on accepte l'hypothèse H_0 au seuil statistique de $100\alpha\%$, donc le coefficient $\beta_i = 0$. La variable X_i n'est pas significative et n'a aucune influence sur Y .
- Si $|t_{\hat{\beta}_i}| > t_{\frac{\alpha}{2}}$: on rejette l'hypothèse H_0 au seuil statistique de $100\alpha\%$, donc le coefficient $\beta_i \neq 0$. La variable X_i n'est pas significative et n'a aucune influence sur Y .

Les décisions associées étant d'exclure ou de conserver la variable X_i dans le modèle selon qu'on accepte ou rejette l'hypothèse H_0 .

4.5.3. Tests de Fisher

A) Tests d'hypothèse linéaire

Une hypothèse linéaire (ou restriction linéaire) est un ensemble de une ou plusieurs conditions du premier degré portant sur les coefficients d'un modèle de régression.

Soit le modèle linéaire multiple suivant :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

Les hypothèses linéaires suivantes peuvent faire l'objet de test :

- $\beta_2 = 0$ (une condition ou contrainte);
- $\beta_2 = \beta_3$ (une condition ou contrainte);
- $\beta_2 = -1$ et $\beta_3 = 0$ (deux conditions ou contraintes);
- $\beta_2 + \beta_3 = 1$ et $\beta_1 = 0$ (deux conditions ou contraintes);
- $\beta_1 = \beta_2 = \beta_3 = 0$ (trois conditions ou contraintes).

De ce qui précède, on distingue deux types de tests de contraintes linéaires sur les coefficients du modèle : les tests d'une seule contrainte et les tests de plusieurs contraintes. Les premiers types de tests utilisent une statistique distribuée selon une loi de Student, tandis que les seconds utilisent une statistique distribuée selon une loi de Fisher (dite aussi loi de Fisher-Snedecor).

B) Test d'une contrainte

Soit le modèle de régression multiple suivant :

$$Y = X\beta + U$$

où le terme d'erreur U est indépendant de X et distribué normalement.

De façon générale, une contrainte linéaire sur les paramètres du modèle de régression peut s'écrire sous la forme :

$$\begin{cases} H_0: r_1\beta_1 + r_2\beta_2 + \dots + r_k\beta_k = q \\ H_1: r_1\beta_1 + r_2\beta_2 + \dots + r_k\beta_k \neq q \end{cases}$$

ou de manière succincte :

$$\begin{cases} H_0: r'\beta = q \\ H_1: r'\beta \neq q \end{cases}$$

(avec q une contrainte et $r' = [r_1 \ r_2 \ \dots \ r_k]$).

Exemple

Supposons le modèle de régression suivant : $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$, l'on veut tester l'hypothèse : $\beta_1 + 0.4\beta_2 = 1$. Cette hypothèse peut s'écrire sous la forme : $r'\beta = q$ comme suit :

$$\underbrace{\begin{bmatrix} 0 & 1 & 0.4 \end{bmatrix}}_{r'} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}}_{\beta} = \underbrace{\begin{bmatrix} 1 \end{bmatrix}}_q$$

Pour tester cette hypothèse, on estime le modèle par la méthode des moindres carrés ordinaires et on calcule la statistique $r'\hat{\beta}$. Ce procédé porte le nom de *régression non contrainte*. Sous hypothèse H_0 , son espérance et sa variance sont respectivement :

$$E(r'\hat{\beta}) = r'E(\hat{\beta}) = r'\beta = q$$

et

$$\text{Var}(r'\hat{\beta}) = r'\text{Var}(\hat{\beta})r = \sigma_u^2 r'(X'X)^{-1}r$$

La statistique $r'\hat{\beta}$, sous hypothèse nulle, est donnée par l'expression suivante :

$$t_{\text{cal}} = \frac{r'\hat{\beta} - q}{\hat{\sigma}_u \sqrt{r'(X'X)^{-1}r}}$$

La statistique t_{cal} est distribuée selon une loi de Student à $(n-k)$ degrés de liberté au seuil de signification α choisi.

Si La statistique t_{cal} est supérieure à la valeur critique de t au niveau de signification choisi pour un nombre de degrés de liberté donné, on peut rejeter l'hypothèse nulle ; sinon, on ne peut la rejeter. De manière alternative, si la valeur p (p – value) du test t associée est assez basse, c'est-à-dire si elle est inférieure au niveau de signification choisi, on peut rejeter l'hypothèse nulle.

C) Test de plusieurs contraintes

Soit le modèle de régression multiple suivant :

$$Y = X\beta + U$$

où le terme d'erreur U est indépendant de X et a une distribution normale.

On peut définir un ensemble de m contraintes (ou restrictions) linéaires sur les paramètres du modèle de régression par l'expression générale ci-après :

$$R\beta = q$$

où R est matrice de format $(m \times k)$; β , vecteur colonne des paramètres du modèle est de dimension $(k \times 1)$; q , un vecteur colonne de dimension $(m \times 1)$.

Exemple

Supposons le modèle de régression suivant :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t.$$

Les deux contraintes linéaires suivantes : $\beta_3 + \beta_4 = 1$ et $\beta_2 = 0$ sur ce modèle peuvent s'écrire aussi comme :

$$\underbrace{\begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}}_R \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}}_{\beta} = \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_q$$

Le test sur plusieurs contraintes consiste donc à tester l'hypothèse :

$$\begin{cases} H_0: R\beta = q \\ H_1: R\beta \neq q \end{cases}$$

Pour réaliser ce test, on estime avec la méthode des moindres carrés ordinaires le modèle sans contraintes afin de calculer $R\hat{\beta}$. Etant fonction linéaire de l'estimateur $\hat{\beta}$, la statistique est également distribuée selon une loi normale.

Sous l'hypothèse H_0 , son espérance et sa variance sont données respectivement par :

$$E(R\hat{\beta}) = R E(\hat{\beta}) = R\beta = q$$

et

$$\text{Var}(R\hat{\beta}) = R \text{Var}(\hat{\beta}) R' = \sigma_u^2 R (X'X)^{-1} R'$$

Par conséquent :

$$R\hat{\beta} \sim N[q, \sigma_u^2 R (X'X)^{-1} R']$$

Sous l'hypothèse H_0 que ces contraintes sont vraies, c'est-à-dire $R\beta = q$, la statistique :

$$F_{\text{cal}} = \frac{(R\hat{\beta} - R\beta)' [\hat{\sigma}_u^2 R (X'X)^{-1} R']^{-1} (R\hat{\beta} - R\beta)}{m} \sim F_{(m, n-k)}$$

Cette statistique peut s'écrire aussi comme suit :

$$F_{\text{cal}} = \frac{\{(R\hat{\beta} - R\beta)' [R (X'X)^{-1} R']^{-1} (R\hat{\beta} - R\beta)\} / m}{\text{SCR} / (n - k)} \sim F_{(m, n-k)}$$

où $\hat{\beta}$ est le vecteur des coefficients estimés sur le modèle non contraint.

De manière équivalente, on peut également trouver une valeur identique en utilisant un test qui nécessite les estimations du modèle sans contraintes et avec contraintes.

Les étapes de calcul sont les suivantes :

- Commencer par régresser le modèle initial sans aucune restriction, et garder la somme des carrés des résidus ainsi obtenus (SCR_{NC}) ;
- Ensuite régresser par la méthode des moindres carrés le modèle transformé (appelé aussi modèle contraint) obtenu en imposant les restrictions au modèle initial, et calculer la somme des carrés des résidus (SCR_C) ;
- Enfin, calculer l'une des deux statistiques équivalentes suivantes :

$$F_{\text{cal}} = \frac{(\text{SCR}_C - \text{SCR}_{\text{NC}}) / m}{\text{SCR}_{\text{NC}} / (n - k)} \sim F_{(m, n-k)}$$

ou

$$F_{\text{cal}} = \frac{(R_{\text{NC}}^2 - R_{\text{C}}^2)/m}{(1 - R_{\text{NC}}^2)/(n - k)} \sim F_{(m, n-k)}$$

avec : R_{NC}^2 représente le coefficient de détermination du modèle non contraint (modèle initial) ;

R_{C}^2 , le coefficient de détermination du modèle contraint (modèle transformé).

- Si $F_{\text{cal}} > F_{(m, n-k)}$, on rejette l'hypothèse nulle.
- Si $F_{\text{cal}} \leq F_{(m, n-k)}$ on accepte l'hypothèse H_0 .

De même, si la valeur p (p -value) du F_{cal} est suffisamment faible, on peut rejeter H_0 .

Une autre façon de procéder consiste à utiliser le ratio de vraisemblance (connu comme test LR: likelihood ratio) des modèles contraint et non contraint. Si la contrainte est vraie, on aura : $L_{\text{C}} < L_{\text{NC}}$, où L_{NC} est la fonction de vraisemblance du modèle non contraint et L_{C} , la fonction de vraisemblance du modèle contraint.

Ce test se ramène aussi à un test du χ^2 par le calcul de la statistique :

$\text{LR} = -2 \ln \lambda \sim \chi^2$ à m degrés de liberté (où $\lambda = \frac{L_{\text{C}}}{L_{\text{NC}}}$ et m étant le nombre des contraintes).

Si LR est supérieur à χ_m^2 lu dans la table statistique au seuil α choisi et à m degrés de liberté, on rejette l'hypothèse H_0 , les restrictions ne sont pas vérifiées.

Exemple : Soit le modèle de demande de thé de Ceylan aux Etats-Unis:

$$\ln Q_t = \beta_1 + \beta_2 \ln P_{\text{Ct}} + \beta_3 \ln P_{\text{It}} + \beta_4 \ln P_{\text{Bt}} + \beta_5 \ln R_t + u_t$$

où Q représente les importations de thé de Ceylan, P_{C} le prix du thé de Ceylan, P_{I} le prix du thé d'Inde, P_{B} le prix du café du Brésil, R le revenu national et u le terme d'erreur.

On veut tester l'hypothèse : $\beta_2 = -1$ et $\beta_3 = 0$, c'est-à-dire l'hypothèse d'élasticité unitaire des quantités au prix du thé de Ceylan ($\varepsilon_{Q/P_{\text{C}}}$) et conjointement d'absence d'influence du prix du thé d'Inde ($\varepsilon_{Q/P_{\text{I}}}$).

- La régression par les moindres carrés ordinaires du modèle non contraint donne les résultats suivants :

$$\ln Q_t = 2,8370 + -1,4810 \ln P_{\text{Ct}} + 1,1810 \ln P_{\text{It}} + 0,1860 \ln P_{\text{Bt}} + 0,2570 \ln R_t$$

(2,0000) (0,9870) (0,6900) (0,1340) (0,3700)

avec $n = 22$ et $SCR_{NC} = 0,4277$

- Sous l'hypothèse $H_0 : \beta_2 = -1$ et $\beta_3 = 0$, le modèle transformé (modèle contraint) s'écrit :

$$\ln Q_t = \beta_1 - \ln P_{Ct} + \beta_4 \ln P_{Bt} + \beta_5 \ln R_t + u_t$$

ou encore :

$$\ln Q_t + \ln P_{Ct} = \beta_1 + \beta_4 \ln P_{Bt} + \beta_5 \ln R_t + u_t$$

L'application de la méthode des moindres carrés sur le modèle contraint donne les résultats suivants :

$$\ln Q_t + \ln P_{Ct} = -0,7380 + 0,1990 \ln P_{Bt} + 0,2610 \ln R_t$$

(0,6900) (0,1340) (0,3700)

avec $n = 22$ et $SCR_C = 0,6788$

- La statistique du test est donc :

$$F_{cal} = \frac{(0,6788 - 0,4277)/2}{0,4277/(22 - 5)} = 4,99$$

Au seuil de 5 %, la valeur de la loi de Fisher $F_{(2,17)}$ lue dans la table est égale à 3,59. On a donc $F_{cal} = 4,99 > 3,59$, ce qui signifie que l'on rejette au risque 5% l'hypothèse linéaire envisagée.

D) Test de nullité d'un coefficient de régression particulier

La nullité d'un coefficient de régression est une hypothèse linéaire particulière où le nombre de restrictions $m = 1$. En effet, ce cas correspond à :

$$R = [0 \quad \dots \quad 0 \quad 1 \quad 0 \quad \dots \quad 0] \text{ et } q = [0]$$

La matrice R contient un seul 1 à la $i^{\text{ème}}$ place et q est nul, ce qui revient bien à tester la significativité du coefficient β_i . Si l'on pose $q = [\beta_i^*]$, on retrouve le test de significativité de l'égalité de β_i à une certaine valeur β_i^* .

Il convient de faire remarquer que le test de Fisher d'une telle hypothèse linéaire est sans intérêt car il est mathématiquement équivalent au test de Student de significativité d'un coefficient.

E) Test de signification globale d'une régression multiple

Ce test est aussi un cas particulier du test d'hypothèses linéaires portant sur la significativité de l'ensemble des coefficients des variables explicatives de la régression. Il s'intéresse particulièrement à l'évaluation globale du modèle dans le cadre d'une régression multiple, c'est-à-dire si l'ensemble des variables explicatives a une influence sur la variable à expliquer Y . Un tel test est parfois appelé *test de significativité de la régression* ou encore *test de la relation d'ensemble* ou plus communément *test de Fisher*. Il correspond au cas suivant :

$$\underbrace{\begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}}_R \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}}_{\beta} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_q$$

On teste donc :

$$\begin{cases} H_0 : \beta_2 = \beta_3 = \dots = \beta_k = 0 \\ H_1 : \exists \text{ au moins un des coefficients non nuls.} \end{cases}$$

Ce test ne considère pas le terme constant β_1 , car un modèle dans lequel seul le terme constant est significatif n'a aucun sens économique.

Si l'hypothèse nulle est vraie, alors :

$$\begin{aligned} F_{\text{cal}} &= \frac{(\hat{y}_t - \bar{y})^2 / (k - 1)}{\sum e_t^2 / (n - k)} \\ &= \frac{\text{SCE} / (k - 1)}{\text{SCR} / (n - k)} \\ &= \frac{R^2 / (k - 1)}{(1 - R^2) / (n - k)}. \end{aligned}$$

Exemple

Soit la fonction de production Cobb-Douglas :

$$\ln Q_t = \beta_1 + \beta_2 \ln K_t + \beta_3 \ln L_t + u_t$$

où Q_t représente le Produit National Brut au temps t ;

K_t , le facteur capital à la période t ;

L_t , le facteur travail à la période t ;

t , le temps (de 1929 à 1961).

En effectuant la régression par les moindres carrés ordinaires sur la période globale on obtient :

$$\widehat{\ln Q_t} = -3,87 + 0,41 \ln K_t + 1,41 \ln L_t$$

(0,255) (0,050) (0,089)

$$n = 33$$

$$R^2 = 0,9937$$

(•) = écarts types estimés des estimateurs des coefficients de régression.

- 1) Interpréter les résultats de la régression par la méthode des moindres carrés ;
- 2) Tester au seuil de 5 % les hypothèses linéaires suivantes :
 - a) $\beta_2 = 0.4$;
 - b) $\beta_3 = 0$.

Solution

- 1) Interprétation

Les coefficients du modèle de production Cobb Douglas représentent respectivement l'élasticité du Produit National Brut au facteur capital et l'élasticité du Produit National Brut au facteur travail. En effet :

$$\beta_2 = \frac{\partial \ln Q}{\partial \ln K} = \frac{\Delta Q/Q}{\Delta K/K} = \varepsilon_{Q/K};$$

$$\beta_3 = \frac{\partial \ln Q}{\partial \ln L} = \frac{\Delta Q/Q}{\Delta L/L} = \varepsilon_{Q/L};$$

β_2 et β_3 sont des rapports de dérivées logarithmiques.

L'élasticité du produit national brut au facteur travail est égale à $\beta_2 = 1,41$.

L'élasticité du produit national brut au facteur capital est égale à $\beta_3 = 0,41$.

Si le facteur travail augmente de 10 %, le Produit National Brut augmente de 14,1 % et si le facteur capital augmente de 10 %, le Produit National Brut augmente de 4,1%.

L'on peut également déterminer les rendements d'échelle car ces derniers sont :

- **décroissants** si $\beta_2 + \beta_3 < 1$ (le Produit National Brut augmente dans une proportion moindre que les facteurs de production);
- **constants** si $\beta_2 + \beta_3 = 1$ (le Produit National Brut augmente dans une proportion identique aux facteurs de production);
- **croissants** si $\beta_2 + \beta_3 > 1$ (le Produit National Brut augmente plus vite que les facteurs de production).

La valeur de $R^2 = 0,99$ signifie qu'environ 99 % de la variation du logarithme du Produit National Brut sont expliqués par les logarithmes des quantités de travail et de capital.

2) Tests statistiques

a) Test d'hypothèse de significativité du coefficient β_3

$$H_0: \beta_3 = 0$$

$$H_1: \beta_3 \neq 0$$

Le facteur travail a-t-il une influence significative sur le Produit National Brut ?

La valeur du ratio de Student permet de répondre à cette question. En effet :

$$t_{\text{calculé}} = \left| \frac{\hat{\beta}_3}{\hat{\sigma}_{\beta_3}} \right| = \left| \frac{1,41}{0,089} \right| = 15,84.$$

Au seuil de 5 %, la lecture de la table de Student donne $t_{(0,025; 30)} = 2,042$.

Le ratio de Student est supérieur à $t_{(0,025; 30)}$, le coefficient β_3 est significativement différent de 0 ; le facteur travail a une influence significative sur le produit national brut.

b) Test d'hypothèse sur β_2

$$H_0: \beta_2 = 0,40$$

$$H_1: \beta_2 \neq 0,40$$

Sous hypothèse nulle, alors :

$$t_{\text{calculé}} = \left| \frac{\hat{\beta}_2 - 0,40}{\hat{\sigma}_{\beta_2}} \right| = \left| \frac{0,41 - 0,40}{0,050} \right| = 0,20.$$

La valeur de t_{cal} est inférieure à $t_{(0,025; 30)} = 2,042$: on accepte l'hypothèse nulle H_0 que $\beta_2 = 0,40$.

c) Test de significativité globale du modèle

$$H_0: \beta_2 = \beta_3 = 0$$

$$H_1: \exists \text{ au moins un coefficient non nul}$$

Sous hypothèse nulle, la statistique Fisher empirique est égal à :

$$F_{\text{cal}} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{0,9937/2}{(1-0,9937)/30} = 2365,95.$$

Au seuil de 5 %, la valeur critique fournie par la table de Fisher pour 2 (c'est-à-dire $k-1 = 3-1$) degrés de liberté au numérateur et 30 (c'est-à-dire $n-k = 33-3$) degrés de liberté au dénominateur est $F_{(2;30)} = 3,32$. Puisque $F_{\text{cal}} > F_{(2;30)}$, nous rejetons l'hypothèse de nullité de tous les coefficients, la régression est globalement significative.

4.6. Tests de stabilité

Les tests de stabilité des paramètres estimés du modèle de régression, appelés aussi tests de robustesse du modèle estimé, visent à vérifier si les valeurs des coefficients estimés sont stables sur l'ensemble de la période d'étude considérée. On cherche donc à vérifier s'il n'y a pas eu de changements structurels entre la variable expliquée et les variables explicatives. Notons que l'instabilité des paramètres peut avoir plusieurs sources attribuables soit aux forces externes, aux calamités naturelles ou aux variations de politique économique.

Il existe diverses méthodes pour détecter l'instabilité des coefficients estimés d'un modèle de régression. Nous allons exposer deux types de tests de stabilité des coefficients :

- le test de Chow ;
- les tests Cusum et Cusum Carré de Brown, Durbin et Evans.

4.6.1. Test de Chow

Le test de Chow, appelé aussi test de changement structurel, permet d'examiner si les coefficients d'une régression sont stables par rapport aux observations utilisées.

Sur des séries temporelles, on compare les estimations effectuées sur deux (ou plusieurs) sous ensembles d'observations qui correspondent à un découpage en périodes de l'échantillon initial. On parle dans ce cas de *test de stabilité temporelle* de la régression.

Sur les données en coupe transversale, on peut comparer les résultats obtenus par exemple sur des individus, des pays, des régions, des secteurs industriels différents. Concernant les individus, on peut s'intéresser à des résultats par classe d'âge, par sexe, etc. Dans ce cas, le test de Chow est utilisé afin de déterminer si des groupes d'individus sont homogènes ou pas. On parle alors de *test d'homogénéité des comportements*.

Considérons le modèle linéaire général suivant, pour $t=1,2,\dots,n$:

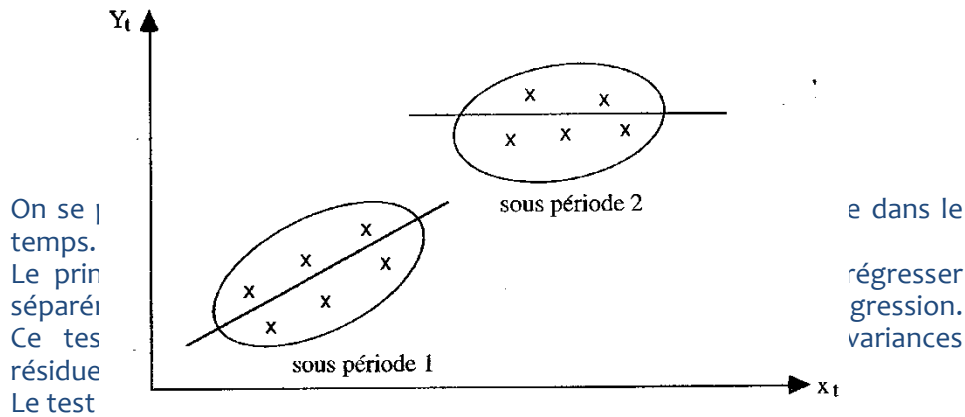
$$\underset{(n \times 1)}{\underline{Y}} = \underset{(n \times k)}{\underline{X}} \underset{(k \times 1)}{\underline{\beta}} + \underset{(n \times 1)}{\underline{U}} \quad (4.1)$$

Supposons que l'on divise l'échantillon d'observations en deux sous-périodes distinctes (n_1 et n_2 ; $n_1 + n_2 = n$) et que l'on estime les deux modèles suivants :

$$\underset{(n_1 \times 1)}{\underline{Y}^1} = \underset{(n_1 \times k)}{\underline{X}^1} \underset{(k \times 1)}{\underline{\beta}^1} + \underset{(n_1 \times 1)}{\underline{U}^1} \quad (4.2)$$

et

$$\underbrace{Y^2}_{(n_2 \times 1)} = \underbrace{X^2}_{(n_2 \times k)} \underbrace{\beta^2}_{(k \times 1)} + \underbrace{U^2}_{(n_2 \times 1)} \quad (4.3)$$



$$H_0 : \begin{cases} \beta_0^1 = \beta_0^2 = \beta_0 \\ \beta_1^1 = \beta_1^2 = \beta_1 \\ \vdots \\ \beta_k^1 = \beta_k^2 = \beta_k \end{cases} \quad (4.4)$$

De manière équivalente, ce test peut s'écrire comme suit :

$$\begin{aligned} H_0 : SCR &= SCR_1 + SCR_2 \quad (\text{stabilité}) \\ H_1 : SCR &\neq SCR_1 + SCR_2 \quad (\text{instabilité}) \end{aligned} \quad (4.5)$$

Les étapes de test de Chow sont les suivantes :

- Estimer le modèle (4.1) et calculer la somme des carrés des résidus (SCR) correspondante, avec $(n - K)$ degrés de liberté.
- Estimer ensuite le modèle (4.2) et calculer la somme des carrés des résidus (SCR_1) correspondante, avec $(n_1 - K)$ degrés de liberté.
- Estimer enfin le modèle (4.3) et calculer la somme des carrés des résidus (SCR_2) correspondante, avec $(n_2 - K)$ degrés de liberté.
- Calculer $SCR_3 = SCR_1 + SCR_2$, $(n_1 - K) + (n_2 - K) = (n - 2K)$ degrés de liberté.
- Calculer la statistique de test :

$$F_{cal} = \frac{[SCR - (SCR_1 + SCR_2)]/K}{(SCR_1 + SCR_2)/(n - 2K)} \quad (4.6)$$

où $K = k + 1$ est le nombre des coefficients du modèle.

Cette statistique suit, sous l'hypothèse nulle d'absence de changement structurel, une loi de Fisher avec respectivement K degrés de liberté pour le numérateur et $(n - 2K)$ degrés de liberté pour le dénominateur.

La règle de décision est alors la suivante :

- Si $F_{cal} < F_{(K, n-2K)}$, on accepte l'hypothèse nulle de stabilité des coefficients. Il n'y a donc pas de changement structurel.

- Si $F_{cal} < F_{(K,n-2K)}$, on rejette l'hypothèse nulle de stabilité des coefficients. Il y a eu donc changement structurel.

Remarques :

- Le test de Chow peut être facilement généralisé à la présence de plus d'un changement structurel.
- Ce test nécessite nécessairement que soit connue la date à laquelle se produit la rupture structurelle.
- Enfin, ce test s'applique de manière identique sous les données en coupe transversale (cross section) pour tester l'homogénéité du comportement de groupes d'individus.

Exemple

Soit la fonction d'importations ci-après :

$$\ln M_t = \beta_1 + \beta_2 \ln \text{PIB}_t + u_t$$

où :

- M désigne les importations ;
- PIB est le Produit intérieur brut.

L'estimation de cette relation sur la période 1962-1995 donne les résultats suivants :

$$\begin{aligned} \widehat{\ln M_t} &= 0,65 + 0,75 \ln \text{PIB}_t \\ R^2 &= 0,87 ; \quad \text{SCR} = 0,134 ; \quad n = 34 \end{aligned}$$

Cette fonction d'importations estimée est-elle stable sur la totalité de la période considérée au seuil de 5% ?

Solution

Pour appliquer le test de Chow, on estime deux nouvelles régressions : une régression sur la période 1960-1978 et une régression sur la période 1979-1995. Les résultats de ces régressions sont donnés ci-après :

- Sur la période 1962-1978, soit $t=1, \dots, 17$:

$$\begin{aligned} \widehat{\ln M_t} &= 1,34 + 0,65 \ln \text{PIB}_t \\ R_1^2 &= 0,46 ; \quad \text{SCR}_1 = 0,102 ; \quad n_1 = 17 \end{aligned}$$
- Sur la période 1979-1995, soit $t=18, \dots, 34$:

$$\begin{aligned} \widehat{\ln M_t} &= 2,15 + 0,54 \ln \text{PIB}_t \\ R_2^2 &= 0,84 ; \quad \text{SCR}_2 = 0,014 ; \quad n_2 = 17 \end{aligned}$$

La statistique du test de Chow est donc :

$$\begin{aligned} F_{cal} &= \frac{[\text{SCR} - (\text{SCR}_1 + \text{SCR}_2)]/k}{(\text{SCR}_1 + \text{SCR}_2)/(n - 2k)} \\ &= \frac{0,134 - (0,102 + 0,014)/2}{(0,102 + 0,014)/30} = 2,30 \end{aligned}$$

La lecture de la table de Fisher au seuil de 5 % donne : $F_{(2,30)} = 3,32$. On constate que valeur calculée de la statistique de test est inférieure à la valeur critique. Par conséquent, l'hypothèse nulle de stabilité des coefficients du modèle est acceptée au seuil de 5 %. La fonction d'importations est donc stable sur la période considérée.

La valeur critique $F_{théorique}$ est $F_{0,05}(2,30)=3,32$. Donc : $F_{cal} < F_{théorique}$, on accepte l'hypothèse de stabilité. Cette la fonction d'importations estimée est stable.

4.6.2. Tests de stabilité temporelle basés sur les résidus récurrents

L'intérêt de ces tests réside dans le fait qu'il permet d'étudier la stabilité d'une régression sans définir a priori la date de rupture sur les coefficients. Contrairement au test de Chow, les tests proposés par Brown, Durbin et Evans sont basés sur l'utilisation des résidus récurrents.

Soit le modèle de régression linéaire suivant :

$$Y = X\beta + U$$

ou encore :

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_i X_{it} + \dots + \beta_k X_{kt} + u_t \\ (t = 1, 2, \dots, n)$$

Pour calculer les résidus récurrents du modèle, on doit estimer par la méthode des moindres carrés ordinaires les paramètres β avec un nombre d'observations qui varie de p à n . Et le $r^{\text{ème}}$ résidu récurrent, e_r (où $r = p + 1, \dots, n$), appelé aussi l'erreur de prévision ex-post sur l'observation y_r , est donné par l'expression :

$$e_r = y_r - \hat{y}_r = y_r - x_r' \hat{\beta}_{r-1}$$

où : $\hat{\beta}_{r-1} = (X_{r-1}' X_{r-1})^{-1} X_{r-1}' Y_{r-1}$, estimateur des moindres carrés ordinaires lorsque la taille de l'échantillon est égale à $(r - 1)$;

$x_r' = [1 \quad X_{1r} \quad X_{2r} \quad \dots \quad X_{kr}]$ est le vecteur des variables explicatives plus la constante pour la $r^{\text{ème}}$ observation ;

Y_r , le sous vecteur des $(r - 1)$ premiers éléments de Y .

Sous l'hypothèse de stabilité $y_r = x_r' \beta + u_r$ et l'erreur de prévision devient :

$$e_r = y_r - \hat{y}_r = x_r' \beta + u_r - x_r' \hat{\beta}_{r-1} \\ = u_r - x_r' (\hat{\beta}_{r-1} - \beta)$$

Or $E(\hat{\beta}_{r-1}) = \beta$ et $\text{Var}(\hat{\beta}_{r-1}) = E[(\hat{\beta}_{r-1} - \beta)(\hat{\beta}_{r-1} - \beta)'] = \sigma_u^2 (X_{r-1}' X_{r-1})^{-1}$, d'où :

$$\text{Var}(e_r) = \sigma_{e_r}^2 = \sigma_u^2 [1 + x_r' (X_{r-1}' X_{r-1})^{-1} x_r].$$

Les résidus récurrents, notés w_r , sont définis comme les erreurs de prévision normalisées :

$$w_r = \frac{e_r}{\sqrt{1 + x_r' (X_{r-1}' X_{r-1})^{-1} x_r}}$$

avec $w_r \sim N(0, \sigma_u^2)$.

A partir de ces résidus récurrents, Brown, Durbin et Evans (1975) ont proposé les tests Cusum (Cumulated Sum of residuals) et Cusum carré (Cumulated Sum of Squared residuals : Cusumsq) qui permettent de tester la stabilité des coefficients estimés dans un modèle de régression.

Il s'agit de tester l'hypothèse nulle que le vecteur β est identique à chaque période conditionnellement à l'hypothèse $\sigma_{u_1}^2 = \dots = \sigma_{u_i}^2 = \dots = \sigma_{u_n}^2 = \sigma_u^2$, soit :

$$H_0: \beta_1 = \beta_2 = \dots = \beta_n = \beta \\ H_1: \beta_t \neq \beta \text{ ou } \sigma_{u_n}^2 = \sigma_u^2$$

où les coefficients $\beta_t, t = 1, 2, \dots, n$ sont les vecteurs des coefficients de régression pour la période t et les $\sigma_{u_t}^2$ désignent les variances des erreurs de la même période.

A. Test Cusum

Le test Cusum est basé sur la somme cumulée des résidus récurrents w_t définie par :

$$w_t = \sum_{j=r+1}^t \frac{w_j}{\hat{\sigma}_w}$$

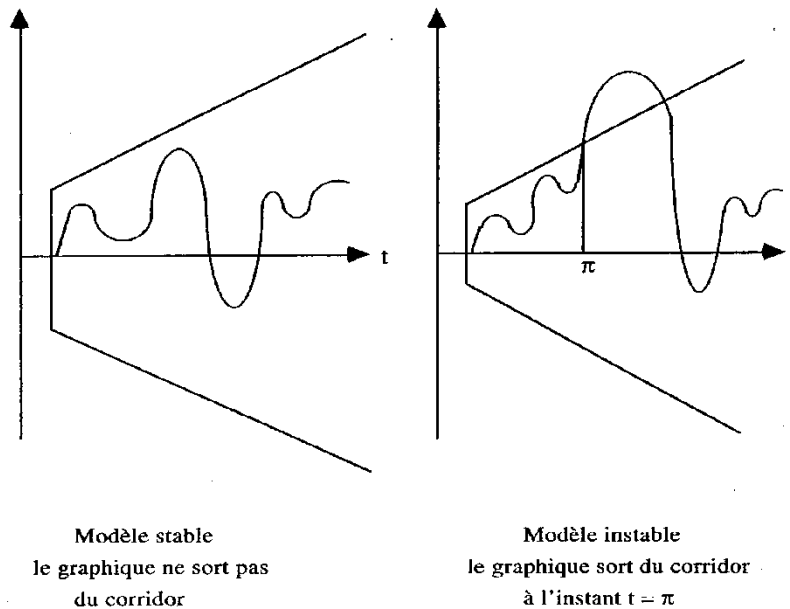
où : $t = 1, 2, \dots, n$;

$$\bar{w} = \frac{1}{n-r-1} \sum_{j=r+1}^n w_j \quad \text{et} \quad \hat{\sigma}_w^2 = \frac{1}{n-r-1} \sum_{j=r+1}^n (w_j - \bar{w})^2.$$

Sous l'hypothèse nulle de stabilité, w_t est de moyenne nulle et de variance approximativement égale au nombre total des résidus (car les termes sont indépendants et ont chacun une variance égale à 1).

Le test est effectué en représentant graphiquement la série w_t par rapport à t et la région de confiance donnée par l'intervalle de confiance $[-C_\alpha, C_\alpha]$. Pour $t = r$, $C_\alpha = a(n-r)^{1/2}$ et pour $t = n$, $C_\alpha = 3a(n-r)^{1/2}$. La région de confiance du test Cusum pour les deux points $t = r+1, \dots, n-1$ est représentée par l'espace entre les deux droites reliant ces points.

Si w_t reste dans cet intervalle quel que soit $t = r+1, \dots, n$, alors l'hypothèse de stabilité des paramètres est retenue. Sinon, si à une date donnée, w_t va au-delà d'une des droites, alors on considère qu'à cette date il y a une rupture et que les paramètres du modèle estimé ne sont pas stables.



Le test Cusum est généralement utilisé pour détecter d'éventuels mouvements systématiques dans la valeur des coefficients reflétant une possible instabilité

structurelle. Si une rupture est constatée, on rejette la spécification choisie sur l'ensemble de la période.

B. Test Cusum Carré

Le test de Cusum Carré est similaire au précédent. Il est utilisé pour détecter des mouvements aléatoires c'est-à-dire des mouvements ne provenant pas forcément d'une modification structurelle des coefficients. Ce test consiste à représenter graphiquement les sommes cumulées des résidus récurrents S_t :

$$S_t = \frac{\sum_{j=r+1}^t w_j^2}{\sum_{j=r+1}^n w_j^2}$$

($t = r + 1, \dots, n$)

Comme dans le test Cusum, la région critique du test Cusum carré peut être représentée graphiquement. Les deux bornes sont données par $E(S_t) \pm c_0$, où les valeurs de c_0 sont données dans la table de Durbin pour diverses tailles d'échantillon et divers seuils de signification. Si, quelque soit $t = r + 1, \dots, n$, S_t se situe dans le corridor défini par les deux bornes alors les paramètres du modèle sont supposés stables. Sinon, si S_t sort des bornes données par les deux droites à une période $t = \pi$, on conclut à l'existence d'une rupture aléatoire qui traduit l'instabilité des coefficients de la régression pour cette date.

Exemple d'illustration

La figure ci-dessous correspond à l'application du test Cusum pour un seuil statistique de 5 %. On constate que la série de la somme cumulée des résidus récurrents reste à l'intérieur de l'intervalle formé des droites, suggérant l'absence d'instabilité de la relation sur la période considérée.

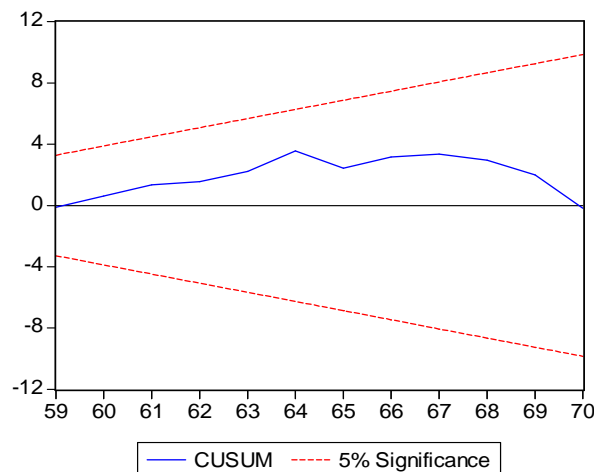


Fig. : Test Cusum

La figure suivante correspond à l'application du test Cusum carré au seuil de signification de 5 %. Comme dans la figure précédente, ce graphique met en évidence que la somme cumulée des résidus récurrents au carré est pratiquement à l'intérieur du corridor confirmant ainsi la stabilité des paramètres sur la période.

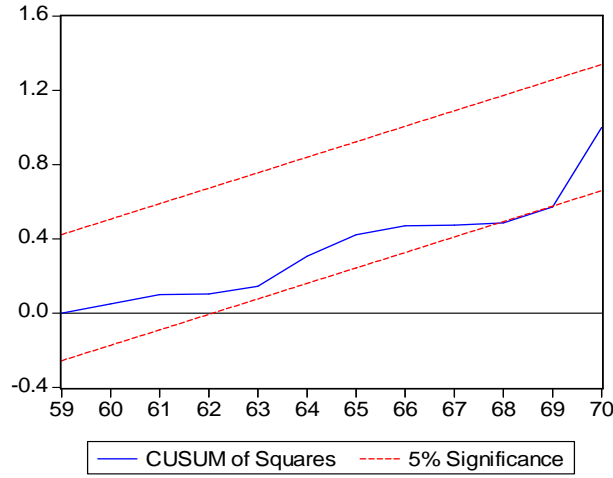


Fig. : Test Cusum carré

4.7. Prédiction

Le problème consiste à déterminer, à partir d'un modèle dont les paramètres ont été estimés, quelle valeur attribuée à la variable expliquée lorsque nous connaissons les valeurs des variables exogènes.

Si le modèle général estimé est comme ci-dessous :

$$\hat{Y}_t = \hat{\beta}_1 + \hat{\beta}_2 X_{2t} + \dots + \hat{\beta}_k X_{kt}$$

Pour un vecteur X donné des valeurs prises par les variables explicatives lors de la période sur laquelle porte la prédiction, la prédiction ponctuelle pour la période $t+h$ est la suivante :

$$\hat{Y}_{t+h} = \hat{\beta}_1 + \hat{\beta}_2 X_{2,t+h} + \dots + \hat{\beta}_k X_{k,t+h}$$

Sous forme matricielle, l'équation ci-dessus peut s'écrire comme suit :

$$\hat{Y}_{t+h} = [1 \quad X_{1,t+h} \quad X_{2,t+h} \quad \dots \quad X_{k,t+h}] \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = R\hat{\beta}$$

où R est le vecteur des valeurs prévues des variables explicatives.

La vraie valeur de Y pendant de la période de prédiction est donc :

$$Y_{t+h} = R\beta + u_{t+h},$$

avec u_{t+h} désignant la vraie valeur prise par le terme d'erreur pendant cette période.

L'erreur de prédiction e_{t+h} est donnée par l'écart entre la valeur observée et la valeur prévue :

$$\begin{aligned} e_{t+h} &= Y_{t+h} - \hat{Y}_{t+h} \\ &= R(\beta - \hat{\beta}) + u_{t+h} \end{aligned}$$

Notons que e_{t+h} est une variable aléatoire car la réalisation Y_{t+h} n'est pas connue au moment de la prédiction. Les moments de cette variable aléatoire

sont :

a) L'espérance mathématique de l'erreur de prévision est égale à :
 $E(e_{t+h}) = 0$,
 car $E(\hat{\beta}) = \beta$ et $E(u_{t+h}) = 0$.

b) La variance de l'erreur de prévision est donc :

$$\begin{aligned} \text{Var}(e_{t+h}) &= \sigma_{e_{t+h}}^2 \\ &= E\{[-R(\hat{\beta} - \beta) + u_{t+h}][-R(\hat{\beta} - \beta) + u_{t+h}]'\} \\ &= \sigma_u^2 R(X'X)^{-1} R' + \sigma_u^2 \\ &= \sigma_u^2 [R(X'X)^{-1} R' + 1]. \end{aligned}$$

L'erreur de prévision e_{t+h} est une variable aléatoire distribuée selon une loi normale :

$$e_{t+h} \sim N(0, \sigma_{e_{t+h}}^2)$$

par conséquent :

$$\frac{e_{t+h}}{\sqrt{\sigma_u^2 [R(X'X)^{-1} R' + 1]}} \sim N(0, 1)$$

et, en remplaçant la variance théorique de l'erreur σ_u^2 par la variance empirique $\hat{\sigma}_u^2$, on obtient la statistique :

$$\begin{aligned} t_{\text{cal}} &= \frac{e_{t+h}}{\sqrt{\hat{\sigma}_u^2 [R(X'X)^{-1} R' + 1]}} = \frac{Y_{t+h} - \hat{Y}_{t+h}}{\sqrt{\hat{\sigma}_u^2 [R(X'X)^{-1} R' + 1]}} \\ &= \frac{Y_{t+h} - \hat{Y}_{t+h}}{\sqrt{\hat{\sigma}_{e_{t+h}}^2}} = \frac{Y_{t+h} - \hat{Y}_{t+h}}{\hat{\sigma}_{e_{t+h}}} \sim t_{(n-k)}. \end{aligned}$$

On peut ainsi établir l'intervalle de confiance d'une prédiction de Y_{t+h} , pour un ensemble d'observations donné :

$$P\left[\hat{Y}_{t+h} - \hat{\sigma}_{e_{t+h}} \cdot t_{\frac{\alpha}{2}, (n-k)} \leq Y_{t+h} \leq \hat{Y}_{t+h} + \hat{\sigma}_{e_{t+h}} \cdot t_{\frac{\alpha}{2}, (n-k)}\right] = 1 - \alpha.$$

L'intervalle de prévision au seuil de signification α est alors :

$$\hat{Y}_{t+h} \pm \hat{\sigma}_{e_{t+h}} \cdot t_{\frac{\alpha}{2}, (n-k)}$$

ou encore :

$$\hat{Y}_{t+h} - t_{\frac{\alpha}{2}, (n-k)} \sqrt{\hat{\sigma}_u^2 [R(X'X)^{-1} R' + 1]}$$

ou enfin :

$$\hat{Y}_{t+h} - t_{\frac{\alpha}{2}, (n-k)} \hat{\sigma}_u \sqrt{[R(X'X)^{-1} R' + 1]}.$$

Pour évaluer la qualité prédictive d'un modèle, plusieurs méthodes sont proposées. L'une des méthodes consiste à comparer les valeurs observées et les valeurs estimées, c'est-à-dire s'intéresser à la série des résidus. La série des valeurs estimées doit bien reproduire l'évolution de la série observée et, en particulier, les points de retournement.

On peut également, pour plus de précision, utiliser des méthodes plus formelles. Il s'agit de calculer différentes statistiques permettant d'évaluer l'exactitude des prévisions :

1. L'erreur moyenne (mean error)

$$ME = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)$$

2. Erreur absolue moyenne (Mean Absolute Error)

$$MAE = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t|$$

3. Racine de l'écart-type moyen (Root Mean Square Error)

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2}$$

4. Statistique U de Theil

$$U = \sqrt{\frac{\sum_{t=2}^n (\Delta Y_t - \Delta \hat{Y}_t)^2}{\sum_{t=2}^n (\Delta Y_t)^2}}$$

$$\text{où } \Delta Y_t = Y_t - Y_{t-1} \text{ et } \Delta \hat{Y}_t = \hat{Y}_t - \hat{Y}_{t-1}.$$

Les trois premières mesures (ME, MAE et RMSE) sont utilisées lorsqu'on compare le pouvoir de prévision de plusieurs modèles estimés. Le modèle pour lequel les trois statistiques sont les plus faibles est considéré avoir une meilleure capacité prédictive. Mais, notons que ces statistiques présentent un problème évident de normalisation. En effet, la multiplication des valeurs de la variable dépendante par n'importe quel facteur multiplie également la mesure par ce scalaire.

La statistique U de Theil n'est pas influencée par ce facteur de normalisation. Elle permet de comparer les prévisions du modèle à des prévisions naïves. Lorsque $\hat{Y}_t = Y_{t-1}$ alors $\Delta \hat{Y}_t = 0$ et, par conséquent, la statistique de Theil s'égale à 1. Ainsi, une valeur de la statistique de Theil inférieure à 1 indique que le modèle permet de réaliser des prévisions meilleures que des prévisions naïves.

Exemple illustratif

L'estimation du modèle de Cobb-Douglas pour l'économie mexicaine de 1955-1970 a donné les résultats suivants :

$$\ln \widehat{Q}_t = -3,2116 + 0,7931 \ln K_t + 0,5811 \ln L_t$$

(0,3933) (0,0625) (0,1234)

$$n = 16;$$

$$R^2 = 0,9987; \bar{R}^2 = 0,9984;$$

$$\hat{\sigma}_u = 0,012209 \text{ (écart-type estimé de l'erreur);}$$

$$F_{\text{calculé}} = 4808$$

(•) = écarts types estimés des estimateurs des coefficients de régression.

Notons que :

$$X'X = \begin{bmatrix} 16 & 200,57189 & 147,78782 \\ 200,57189 & 2515,53683 & 1853,23835 \\ 147,78782 & 1853,23835 & 1365,39069 \end{bmatrix}$$

$$\Rightarrow (X'X)^{-1} = \begin{bmatrix} 1037,7331 & 141,69943475 & -304,650632 \\ 141,699434 & 26,2374792 & -51,9493389 \\ -304,650632 & -51,9493389 & 102,128953 \end{bmatrix}$$

- 1) Les prévisions ponctuelles sont données par :

- Pour l'année 1971 :

$$\begin{aligned}\hat{M}_{t+1} &= \hat{M}_{1971} = -3,211626 + (0,793127) * (13,09124) + (0,581095) \\ &\quad * (9,498372) \\ &= 12,6908384 \\ - \text{ Pour l'année 1972 :} \\ \hat{M}_{t+2} &= \hat{M}_{1972} \\ &= -3,211626 + (0,793127) * (13,16265) + (0,581095) \\ &\quad * (9,527921) \\ &= 12,7646463 \\ - \text{ Pour l'année 1973 :} \\ \hat{M}_{t+3} &= \hat{M}_{1973} \\ &= -3,211626 + (0,793127) * (13,23842) + (0,581095) \\ &\quad * (9,675583) \\ &= 12,9105408 \\ - \text{ Pour l'année 1974 :} \\ \hat{M}_{t+4} &= \hat{M}_{1974} \\ &= -3,211626 + (0,793127) * (13,32093) + (0,581095) \\ &\quad * (9,557753) \\ &= 12,9075196\end{aligned}$$

2) Les prévisions par intervalle au seuil de 5 % :

$$\begin{aligned}- \text{ Pour l'année 1971 :} \\ \hat{M}_{t+h} \pm t_{\frac{\alpha}{2}, (n-k)} \hat{\sigma}_u \sqrt{[R(X'X)^{-1}R' + 1]} \\ \hat{M}_{1971} \pm t_{0,025; (16-3)} \hat{\sigma}_u \sqrt{[R(X'X)^{-1}R' + 1]}\end{aligned}$$

Comme :

$$t_{0,025; (16-3)} = 2,160 ;$$

$$\hat{\sigma}_u = 0,012209 ;$$

$$R_{1971} = [1 \quad 13,09 \quad 9,50]$$

$$R(X'X)^{-1}R' = [1 \quad 13,09124 \quad 9,498372] \begin{bmatrix} 1037,7331 & 141,69943475 & -304,650632 \\ 141,699434 & 26,2374792 & -51,9493389 \\ -304,650632 & -51,9493389 & 102,128953 \end{bmatrix} \begin{bmatrix} 1 \\ 13,09124 \\ 9,498372 \end{bmatrix}$$

$$= 0,34011116,$$

alors, l'intervalle de prévision à 95 % pour M_{1971} est donné par :

$$12,6908384 \pm (2,160) * (0,012209) * (0,34011116)$$

$$= 12,6908384 \pm (2,16) * (0,00415242)$$

$$= 12,6908384 \pm 0,00896922$$

$$\Rightarrow [12,6818691 ; 12,6998076]$$

- Pour l'année 1972 :

$$\begin{aligned}\hat{M}_{1972} \pm t_{0,025; (16-3)} \hat{\sigma}_u \sqrt{[R(X'X)^{-1}R' + 1]} \\ R(X'X)^{-1}R' = [1 \quad 13,16265 \quad 9,527921] \begin{bmatrix} 1037,7331 & 141,69943475 & -304,650632 \\ 141,699434 & 26,2374792 & -51,9493389 \\ -304,650632 & -51,9493389 & 102,128953 \end{bmatrix} \begin{bmatrix} 1 \\ 13,16265 \\ 9,527921 \end{bmatrix} \\ = 0,43236024\end{aligned}$$

D'où :

$$12,7646463 \pm (2,16) * (0,00527869) = 12,7646463 \pm 0,01140196$$

$$\Rightarrow [12,7532444 ; 12,7760483]$$

- Pour l'année 1973 :

$$\begin{aligned}\hat{M}_{1973} \pm t_{0,025; (16-3)} \hat{\sigma}_u \sqrt{[R(X'X)^{-1}R' + 1]} \\ R(X'X)^{-1}R' = [1 \quad 13,23842 \quad 9,675583] \begin{bmatrix} 1037,7331 & 141,69943475 & -304,650632 \\ 141,699434 & 26,2374792 & -51,9493389 \\ -304,650632 & -51,9493389 & 102,128953 \end{bmatrix} \begin{bmatrix} 1 \\ 13,23842 \\ 9,675583 \end{bmatrix}\end{aligned}$$

$$= 1,26375373$$

D'où :

$$12,9105408 \pm (2,16) * (0,01542917) = 12,9105408 \pm 0,03332701$$

$$\Rightarrow [12,8772138 ; 12,9438678]$$

- Pour l'année 1974 :

$$\hat{M}_{1974} \pm t_{0,025 ; (16-3)} \hat{\sigma}_u \sqrt{[R(X'X)^{-1}R' + 1]}$$

$$R(X'X)^{-1}R' = \begin{bmatrix} 1 & 13,32093 & 9,557753 \end{bmatrix} \begin{bmatrix} 1037,7331 & 141,69943475 & -304,650632 \\ 141,699434 & 26,2374792 & -51,9493389 \\ -304,650632 & -51,9493389 & 102,128953 \end{bmatrix} \begin{bmatrix} 1 \\ 13,32093 \\ 9,557753 \end{bmatrix}$$

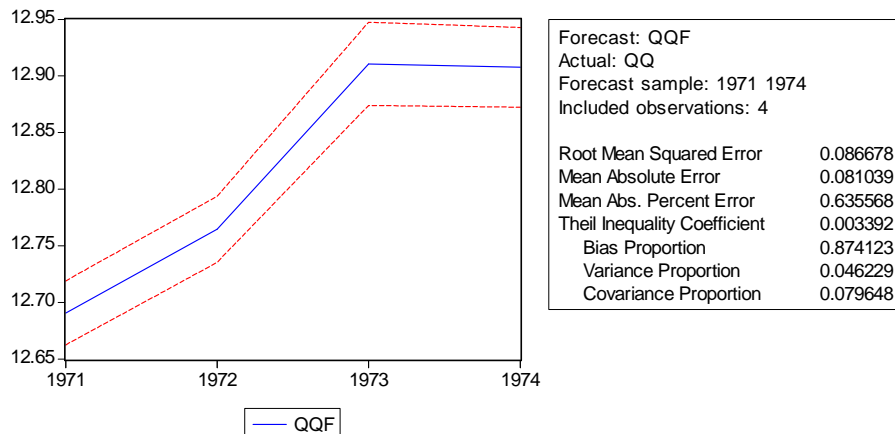
$$= 1,07860217$$

D'où :

$$12,9075196 \pm (2,16) * (0,01316865) = 12,9075196 \pm 0,02844429$$

$$\Rightarrow [12,8790754 ; 12,9359639].$$

La qualité prédictive du modèle peut être évaluée à partir des éléments ci-dessous fournis par le logiciel d'économétrie EvIEWS 5.1. Le graphique retrace l'évolution des valeurs ponctuelles prévues pour l'horizon considéré ainsi que les intervalles de prédiction associés. Cette évaluation graphique est complétée par une série de mesures conçues pour apprécier les prévisions a posteriori. Au vu des résultats ci-dessous, on constate que la valeur de la statistique de Theil ($U=0,003392$) est largement inférieure à 1 indiquant un pouvoir de prévision élevé du modèle estimé.



CHAPITRE 4

MODELE DE REGRESSION MULTIPLE

4.8. Exposé du problème

Soient X_1, X_2, \dots, X_k : k variables indépendantes et non aléatoires et Y une variable aléatoire dépendant de X_1, X_2, \dots, X_k . On dit qu'on est en présence d'un modèle linéaire général quand on a la relation :

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_i X_{it} + \dots + \beta_k X_{kt} + u_t \quad (4.1)$$

où u_t est une variable aléatoire centrée de variance σ_u^2 , indépendante des variables X_i et $\beta_1, \beta_2, \dots, \beta_k$ des paramètres réels mais inconnus. On dit que la variable à expliquer Y est une variable endogène et que les variables explicatives X_1, X_2, \dots, X_k sont des variables exogènes. La variable aléatoire u est l'écart ou l'erreur.

On désire estimer les paramètres $\beta_1, \beta_2, \dots, \beta_k$ et la variance σ_u^2 de la variable aléatoire u à partir d'un ensemble de n observations indépendantes :

$$\begin{cases} Y_t \text{ où } t = 1, 2, \dots, n \\ \text{et} \\ X_{it} \text{ où } i = 1, 2, \dots, k \end{cases}$$

4.9. Notation matricielle du modèle

En posant :

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}; \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}; U = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}; \begin{bmatrix} X_{11} & \dots & X_{k1} \\ \vdots & \ddots & \vdots \\ X_{1n} & \dots & X_{kn} \end{bmatrix}$$

On obtient la forme matricielle :

$$\underbrace{Y}_{(n \times 1)} = \underbrace{X}_{(n \times k)} \underbrace{\beta}_{(k \times 1)} + \underbrace{U}_{(n \times 1)} \quad (4.2)$$

L'introduction d'un terme constant s'obtient en donnant à la valeur X_{it} la valeur 1 pour tout $t = 1, \dots, n$.

Exemple : Considérons le modèle linéaire simple (Fonction de consommation keynésienne) :

$$C_t = \beta_1 + \beta_2 Y_{dt} + u_t$$

$$(t = 1, 2, \dots, n)$$

où : C_t = consommation au temps t ;

Y_{dt} = revenu disponible au temps t .

Ce modèle peut aussi s'écrire sous forme de matrice comme ci-après :

$$\underbrace{C}_{(n \times 1)} = \underbrace{Y_d}_{(n \times 2)} \underbrace{\beta}_{(2 \times 1)} + \underbrace{U}_{(n \times 1)}$$

$$C = \begin{bmatrix} C_1 \\ \vdots \\ C_n \end{bmatrix}; \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}, U = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \quad Y_d = \begin{bmatrix} 1 & Y_{d1} \\ \vdots & \vdots \\ 1 & Y_{dn} \end{bmatrix}$$

L'écriture sous forme matricielle rend plus aisée la manipulation du modèle linéaire général, c'est pourquoi nous l'adoptons par la suite.

4.10. Estimation et propriétés des estimateurs

4.10.1. Hypothèses d'application de la méthode des moindres carrés ordinaires.

Deux catégories d'hypothèses doivent être faites pour résoudre le problème des moindres carrés dans le cas d'un modèle linéaire général. Il s'agit des hypothèses stochastiques (liées à l'erreur u), d'une part, et des hypothèses structurelles, de l'autre.

C. Hypothèses stochastiques

- H_1 : X est une matrice non aléatoire, c'est-à-dire les valeurs de X sont observées sans erreur.
- H_2 : L'espérance mathématique de l'erreur est nulle.

$$E(U) = 0 \quad (4.3)$$

Cette hypothèse peut s'écrire aussi comme ci-après :

$$E(U) = \begin{bmatrix} E(u_1) \\ E(u_2) \\ E(u_3) \\ \vdots \\ E(u_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (4.4)$$

De la relation (4.4), on déduit :

$$\begin{aligned} E(Y) &= E(X\beta + U) \\ &= E(X\beta) + E(U) = X\beta. \end{aligned}$$

$$\begin{aligned} - \quad H_3 : \quad \text{Var}(U) &= \{[U - E(U)][U - E(U)]'\} = E[UU'] = \sigma_u^2 I = \Omega_u \\ (4.5) \end{aligned}$$

où Ω_u est la matrice des variantes et covariances de l'erreur U .

L'hypothèse H_3 s'écrit en détail comme ci-dessous:

$$E(UU') = E \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \begin{bmatrix} u_1 & u_2 & \cdots & \cdots & u_n \end{bmatrix}$$

$$\begin{aligned}
&= \begin{bmatrix} E(u_1^2) & E(u_1 u_2) & \cdots & \cdots & E(u_1 u_n) \\ E(u_2 u_1) & E(u_2^2) & \cdots & \cdots & E(u_2 u_n) \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ E(u_n u_1) & E(u_n u_2) & \cdots & \cdots & E(u_n^2) \end{bmatrix} \\
&= \sigma_u^2 \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & 1 \end{bmatrix} = \sigma_u^2 I = \underbrace{\Omega_u}_{(n \times n)}
\end{aligned}$$

(4.6)

En principe, on fait ici deux hypothèses :

- Les lois de probabilité suivies par les erreurs U ont la même variance.

Donc $E(u_t^2) = \sigma_u^2$, $\forall t$ (variance constante et finie). C'est ce qu'on appelle l'*homoscédasticité*, c'est-à-dire que la dispersion ou variance du terme d'erreur est la même du début à la fin de la période. Lorsque cette hypothèse n'est pas vérifiée, on dit qu'il y a *hétéroscédasticité* (dispersion ou variance inégale).

- Les erreurs, prises deux à deux, ne sont pas corrélées.

$$E(u_i u_j) = 0, \quad \forall i \neq j$$

(absence d'autocorrélation ou absence de corrélation sérielle).

Si tel n'est pas le cas, on dit que les erreurs sont *autocorrélées*.

- H_4 : Le vecteur de perturbations U suit une loi normale à n dimensions.
Donc :

$$U \sim N(0, \sigma_u^2 I) \quad (4.7)$$

Cette hypothèse de normalité des erreurs est indispensable, comme nous l'avons souligné plus haut, si on veut utiliser le principe d'estimation du maximum de vraisemblance mais aussi pour procéder aux tests d'inférences.

- H_5 : Covariance nulle entre U et X , c'est-à-dire :

$$\text{Cov}(u_i, X_{ij}) = E(u_i X_{ij}) = 0 \quad (4.8)$$

D. Hypothèses structurelles

- H_5 : absence de colinéarité entre les variables explicatives ; cela implique que la matrice $(X'X)$ est non singulière et que la matrice $(XX')^{-1}$ existe. Donc le rang de la matrice des variables explicatives doit être égal à k : $\rho(X) = k$.
- H_6 : le nombre d'observations doit être supérieur au nombre de paramètres à estimer ($n > k$).

4.10.2. Estimateurs des moindres carrés ordinaires

Soit le modèle sous forme matricielle à k variables explicatives et n observations :

$$Y = X\beta + U \quad (4.9)$$

Le vecteur des résidus (e) s'obtient en remplaçant le vecteur inconnu des paramètres β par son estimateur $\hat{\beta}$ dans l'équation (4.9). D'où :

$$e = Y - X\hat{\beta} = Y - \hat{Y} \quad (4.10)$$

Pour déterminer l'estimateur de β , on applique le principe de la méthode des moindres carrés qui consiste à minimiser la somme des carrés des résidus : ee' , soit :

$$\text{Min } e'e = \text{Min } \sum_{t=1}^n e_t^2 = \text{Min } \sum_{t=1}^n [Y_t - (\hat{\beta}_1 + \hat{\beta}_2 X_{2t} + \dots + \hat{\beta}_k X_{kt})^2] \quad (4.11)$$

En écriture matricielle, on sait que :

$$\begin{aligned} \sum_{t=1}^n e_t^2 &= e'e = (Y - \hat{Y})'(Y - \hat{Y}) \\ &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) \\ &= Y'Y - Y'X\hat{\beta} - \hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta} \end{aligned}$$

Etant donné que $Y'X\hat{\beta} = \hat{\beta}'X'Y$ (car les deux produits sont des scalaires), on peut alors écrire :

$$\text{SCR} = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta} \quad (4.12)$$

Pour minimiser cette expression par rapport à $\hat{\beta}$ on annule sa dérivée par rapport à $\hat{\beta}$:

$$\frac{\delta \text{SCR}}{\delta \hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0 \quad (4.13)$$

La résolution de la relation (4.13) permet de trouver les équations normales sous leur forme matricielle :

$$X'Y = X'X\hat{\beta} \quad (4.14)$$

La relation (4.14) peut s'écrire aussi comme ci-après :

$$\underbrace{\begin{bmatrix} n & \sum X_{2t} & \sum X_{3t} & \dots & \sum X_{kt} \\ \sum X_{2t} & \sum X_{2t}^2 & \sum X_{2t}X_{3t} & \dots & \sum X_{2t}X_{kt} \\ \sum X_{3t} & \sum X_{3t}X_{2t} & \sum X_{3t}^2 & \dots & \sum X_{3t}X_{kt} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum X_{kt} & \sum X_{kt}X_{2t} & \sum X_{kt}X_{3t} & \dots & \sum X_{kt}^2 \end{bmatrix}}_{X'X} \underbrace{\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}}_{\hat{\beta}} = \underbrace{\begin{bmatrix} \sum Y_t \\ \sum X_{2t}Y_t \\ \sum X_{3t}Y_t \\ \vdots \\ \sum X_{kt}Y_t \end{bmatrix}}_{X'Y} \quad (4.15)$$

En multipliant la relation (4.14) par $(X'X)^{-1}$, on obtient l'expression de $\hat{\beta}$.

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (4.16)$$

La résolution de cette équation n'est possible que si la matrice $(X'X)$ est inversible.

Exemples

Ex. 1 : Pour un modèle linéaire simple de type : $Y_t = \beta_1 + \beta_2 X_{2t} + u_t$, la relation (3.15) devient :

$$\begin{bmatrix} n & \sum X_{2t} \\ \sum X_{2t} & \sum X_{2t}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \sum Y_t \\ \sum X_{2t} Y_t \end{bmatrix}$$

Sa résolution donne le système d'équations ci-après :

$$\begin{cases} \sum_{t=1}^n Y_t = n\hat{\beta}_1 + \hat{\beta}_2 \sum_{t=1}^n X_{2t} \\ \sum_{t=1}^n X_{2t} Y_t = \hat{\beta}_1 \sum_{t=1}^n X_{2t} + \hat{\beta}_2 \sum_{t=1}^n X_{2t}^2 \end{cases}$$

Ex. 2 : Pour un modèle linéaire multiple ayant deux variables explicatives :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

où X_{2t} et X_{3t} sont les variables explicatives.

Les équations normales pour ce modèle se présentent comme suit :

$$\begin{bmatrix} n & \sum X_{2t} & \sum X_{3t} \\ \sum X_{2t} & \sum X_{2t}^2 & \sum X_{2t} X_{3t} \\ \sum X_{3t} & \sum X_{3t} X_{2t} & \sum X_{3t}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} \sum Y_t \\ \sum X_{2t} Y_t \\ \sum X_{3t} Y_t \end{bmatrix}$$

ou encore après la résolution matricielle :

$$\begin{cases} \sum_{t=1}^n Y_t = n\hat{\beta}_1 + \hat{\beta}_2 \sum_{t=1}^n X_{2t} + \hat{\beta}_3 \sum_{t=1}^n X_{3t} \\ \sum_{t=1}^n X_{2t} Y_t = \hat{\beta}_1 \sum_{t=1}^n X_{2t} + \hat{\beta}_2 \sum_{t=1}^n X_{2t}^2 + \hat{\beta}_3 \sum_{t=1}^n X_{2t} X_{3t} \\ \sum_{t=1}^n X_{3t} Y_t = \hat{\beta}_1 \sum_{t=1}^n X_{3t} + \hat{\beta}_2 \sum_{t=1}^n X_{3t} X_{2t} + \hat{\beta}_3 \sum_{t=1}^n X_{3t}^2 \end{cases}$$

4.10.3. Propriétés des estimateurs des moindres carrés ordinaires

Les propriétés des estimateurs des moindres carrés ordinaires sont, dans le modèle de régression multiple, analogues à celles du modèle de régression simple. En effet, les estimateurs des moindres carrés ordinaires $\hat{\beta} = (X'X)^{-1}X'Y$ sont non seulement linéaires et sans biais, mais aussi ont une variance

minimale dans la classe de tous les estimateurs non biaisés. Ils sont donc « BLUE » (Best Linear Unbiased Estimator). En d'autres termes, ils satisfont le théorème de Gauss-Markov.

E. Estimateur linéaire

L'estimateur des moindres carrés ordinaires est une fonction linéaire de Y . En effet, soit :

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

Posons : $\gamma = (X'X)^{-1}X'$, on peut alors écrire :

$$\hat{\beta} = \gamma Y \quad (4.16)$$

Donc, les composantes de $\hat{\beta}$ sont des fonctions linéaires des composantes Y_t de Y .

F. Estimateur sans biais

L'espérance de l'estimateur $\hat{\beta}$ est donc :

$$\begin{aligned} E(\hat{\beta}) &= E(\gamma Y) \\ &= \gamma E(X\beta + U) = \gamma[X\beta + E(U)] \\ &= \gamma X\beta, \\ &(\text{car } E(U) = 0). \end{aligned}$$

$$\text{Par conséquent : } E(\hat{\beta}) = (X'X)^{-1}X'X\beta = (X'X)^{-1}(X'X)\beta = \beta \quad (4.17)$$

$\Rightarrow E(\hat{\beta}) = \beta$. Donc, $\hat{\beta}$ est un estimateur sans biais de β .

G. Calcul de la matrice des variances-covariances de $\hat{\beta}$

La matrice des variances-covariances de $\hat{\beta}$ est donnée par :

$$\begin{aligned} \Omega_{\hat{\beta}} &= \text{Var}(\hat{\beta}) = \text{Var}(\gamma Y) = \gamma \text{Var}(Y) \gamma' \\ &= \gamma \text{Var}(X\beta + U) \gamma' = \gamma \text{Var}(U) \gamma' \\ &= \gamma \cdot \sigma_u^2 I \cdot \gamma' = (X'X)^{-1}X' \cdot \sigma_u^2 I \cdot X(X'X)^{-1} \\ &= \sigma_u^2 (X'X)^{-1}(X'X)(X'X)^{-1} \end{aligned}$$

D'où :

$$\Omega_{\hat{\beta}} = \sigma_u^2 (X'X)^{-1} \quad (4.18)$$

Puisque la variance résiduelle σ_u^2 est inconnue, on l'estimera à partir des valeurs observées des résidus afin de rendre opérationnelle la relation (4.18). Donc :

$$\hat{\sigma}_u^2 = \frac{ee'}{n-k} \quad (4.19)$$

où le facteur $(n-k)$ représente le nombre des degrés de liberté.

Ainsi, la matrice des variances-covariances estimée devient :

$$\hat{\Omega}_{\hat{\beta}} = \hat{\sigma}_u^2 (X'X)^{-1} \quad (4.20)$$

Cette détermination de la matrice des variances-covariances de $\hat{\beta}$ permet de montrer que cet estimateur est meilleur, c'est-à-dire de variance minimale.

H. Estimateur de variance minimale

Montrons que, parmi la classe des estimateurs linéaires sans biais, l'estimateur $\hat{\beta}$ des moindres carrés ordinaires est celui qui possède la variance la plus faible :

$$\text{Var}(\hat{\beta}) \leq \text{Var}(\tilde{\beta}) \quad (4.21)$$

où $\tilde{\beta}$ représente tout autre estimateur sans biais de β .

En effet, si $\tilde{\beta}$, fonction linéaire des Y_t , est un estimateur sans biais de β , on doit avoir :

$$\tilde{\beta} = MY \quad \text{et} \quad E(\tilde{\beta}) = \beta$$

Comme $Y = X\beta + U$, on a $\tilde{\beta} = MX\beta + MU$.

Donc la matrice M doit vérifier : $MX\beta = \beta$ ou $MX = I$, avec I est la matrice identité d'ordre k .

Soit $\tilde{\beta} = MY$, estimateur sans biais différent de $\hat{\beta} = (X'X)^{-1}X'Y$.

Posons : $M = (X'X)^{-1}X' + N$.

Comme $MX = I$, alors :

$$MX = [(X'X)^{-1}X' + N]X = (X'X)^{-1}X'X + NX = I + NX = I$$

$$\text{où } \underbrace{N}_{(k \times n)} \underbrace{X}_{(n \times k)} = \underbrace{0}_{(k \times k)}.$$

Montrons que les variances des composantes $\tilde{\beta}_1, \tilde{\beta}_2, \dots, \tilde{\beta}_k$ du vecteur $\tilde{\beta}$ sont supérieures aux composantes $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ de $\hat{\beta}$, c'est-à-dire les termes de la diagonale principale de la matrice des covariances de $\text{Var}(\tilde{\beta}) = E\{[\tilde{\beta} - E(\tilde{\beta})][\tilde{\beta} - E(\tilde{\beta})]'\}$ sont supérieurs à ceux de $\text{Var}(\hat{\beta}) = \sigma_u^2(X'X)^{-1}$. En effet, on a :

$$\tilde{\beta} - E(\tilde{\beta}) = MX\beta + MU - \beta = \beta + MU - \beta = MU,$$

$$\text{car } MU = I.$$

Il en résulte que :

$$\begin{aligned} \text{Var}(\tilde{\beta}) &= E\{[MU][MU]'\} = E(MUU'M') \\ &= ME(UU')M' = M \cdot \sigma_u^2 I \cdot M' = \sigma_u^2 MM'. \end{aligned}$$

Comme $M = (X'X)^{-1}X' + N$, sa transposée est égale à : $M' = X(X'X)^{-1} + N'$.

$$\text{D'où : } \text{Var}(\tilde{\beta}) = \sigma_u^2 MM'$$

$$= \sigma_u^2 [(X'X)^{-1}(X'X)^{-1} + (X'X)^{-1}X'N' + NX(X'X)^{-1} + NN']$$

$$= \sigma_u^2 [(X'X)^{-1} + NN'],$$

$$\text{car } NX = 0 \quad \text{et} \quad X'N'.$$

Par conséquent :

$$\begin{aligned} \text{Var}(\tilde{\beta}) &= \sigma_u^2(X'X)^{-1} + \sigma_u^2(NN') \\ &= \text{Var}(\hat{\beta}) + \sigma_u^2(NN') \end{aligned}$$

et

$$\text{Var}(\tilde{\beta}_i) = \text{Var}(\hat{\beta}_i) + \sigma_u^2 \sum_{i=1}^n n_i^2$$

$$(\text{où } \sigma_u^2 \sum_{i=1}^n n_i^2 \geq 0).$$

$$\text{D'où : } \text{Var}(\tilde{\beta}_i) \leq \text{Var}(\hat{\beta}_i), \quad \forall i = 1, 2, \dots, k \quad (4.22)$$

Il s'ensuit que, parmi la classe des estimateurs linéaires sans biais, l'estimateur des moindres carrés ordinaires $\hat{\beta}$ est celui qui possède la variance la plus faible.

4.11. Analyse de la variance et qualité de l'ajustement

4.11.1. Equation d'analyse de la variance

L'équation fondamentale d'analyse de la variance est :

$$\sum_t (Y_t - \bar{Y})^2 = \sum_t (\hat{Y}_t - \bar{Y})^2 + \sum_t (Y_t - \hat{Y}_t)^2 \quad (4.23)$$

ou en écriture matricielle :

$$(Y'Y - n\bar{Y}^2) = (\hat{\beta}X'Y - n\bar{Y}^2) + e'e \quad (4.24)$$

Sous forme de variables centrées :

$$\sum_t y_t^2 = \sum_t \hat{y}_t^2 + \sum_t e_t^2 \quad (4.25)$$

ou encore :

$$y'y = \hat{\beta}x'y + e'e \quad (4.26)$$

Cette équation est l'équation d'analyse de la variance. Au facteur $\left(\frac{1}{n}\right)$ près, la variance empirique de Y, appelée la variance totale (SCT), est égale à la somme de la variance expliquée par le modèle (SCE) et de la variance résiduelle (SCR).

$$\begin{cases} \text{SCT} = \text{Variabilité totale} = \sum_t (Y_t - \bar{Y})^2 \\ \text{SCE} = \text{Variabilité expliquée} = \sum_t (\hat{Y}_t - \bar{Y})^2 \\ \text{SCR} = \text{Variabilité des résidus} = \sum_t (Y_t - \hat{Y}_t)^2 \end{cases}$$

A partir de cette équation, la qualité de l'ajustement peut être mesurée par le coefficient de détermination.

4.11.2. Qualité d'un ajustement

L'équation d'analyse de la variance permet de juger de la qualité de l'ajustement d'un modèle. En effet, plus la variance expliquée est «proche» de la variance totale, meilleur est l'ajustement global du modèle.

Le coefficient de détermination, noté R^2 , est défini à partir de la décomposition de la variance de Y :

$$R^2 = \frac{\text{Variance expliquée}}{\text{Variance totale}} = \frac{\text{SCE}}{\text{SCT}} = 1 - \frac{\text{SCR}}{\text{SCT}} \quad (4.25)$$

Dans le cas de variables centrées :

$$R^2 = \frac{\hat{\beta}x'y}{y'y} = \frac{\sum_t \hat{y}_t^2}{\sum_t y_t^2} \quad (4.26)$$

ou

$$R^2 = 1 - \frac{e'e}{y'y} = 1 - \frac{\sum_t e_t^2}{\sum_t y_t^2} \quad (4.27)$$

Dans le cas de variables non centrées :

$$R^2 = \frac{\hat{\beta}X'Y - n\bar{Y}^2}{Y'Y - n\bar{Y}^2} = 1 - \frac{e'e}{Y'Y - n\bar{Y}^2} \quad (4.28)$$

Par construction, cette statistique est comprise dans l'intervalle $[0, 1]$. Donc : $0 \leq R^2 \leq 1$.

Le coefficient de détermination R^2 est un indicateur de qualité de l'ajustement linéaire entre les variables explicatives et la variable expliquée. Il mesure la part de la variance totale expliquée par les variables explicatives et permet donc de juger de la qualité de l'ajustement du modèle. Une valeur proche de 1 indique que la qualité de l'ajustement est bonne dans la mesure où la part de la variance de Y expliquée par le modèle est élevée. Cependant, il doit être utilisé avec précaution. Un R^2 élevé ne doit en aucun cas être interprété comme une mesure du degré d'explication de la variable dépendante par les variables explicatives, mais seulement comme une forte association entre ces variables. En outre, le R^2 a tendance à croître avec le nombre de variables explicatives retenues dans le modèle, indépendamment du pouvoir explicatif de ces variables. Il ya donc un biais dans l'évaluation du coefficient de détermination dû au fait que l'on a tenu compte explicitement ni du nombre de variables explicatives, ni du nombre d'observations. Pour y remédier, on a défini un coefficient de détermination ajusté, connu aussi sous le nom de R^2 corrigé, noté \bar{R}^2 , qui tient compte du nombre de variables explicatives présentes dans le modèle :

$$\bar{R}^2 = 1 - \frac{e'e/(n-k)}{y'y/(n-1)} \quad (4.29)$$

ou encore :

$$\begin{aligned} \bar{R}^2 &= 1 \\ &- \frac{(n-1)}{(n-k)} (1 - R^2) \end{aligned} \quad (4.30)$$

ou enfin :

$$\begin{aligned} \bar{R}^2 &= \left(\frac{n-1}{n-k} \right) R^2 \\ &- \left(\frac{k-1}{n-k} \right) \end{aligned} \quad (4.31)$$

Notons que :

- d) $\bar{R}^2 < R^2$, donc lorsque le nombre de variables explicatives augmente, le coefficient de détermination ajusté \bar{R}^2 croît moins que le R^2 ;
- e) $\bar{R}^2 = R^2$, si le nombre d'observations n est grand ;
- f) \bar{R}^2 peut prendre la valeur négative bien que R^2 soit exclusivement positif. Dans ce cas, la valeur de \bar{R}^2 doit être considérée comme nulle.

Sachant que la même série d'observations relatives à un phénomène donné peut avoir plusieurs représentations, la statistique \bar{R}^2 peut être utilisée pour retenir le modèle dont le pouvoir explicatif est élevé, c'est-à-dire le modèle dont le coefficient de détermination ajusté est le plus élevé. On peut également utiliser deux autres critères d'information pour le choix d'un modèle ayant un pouvoir explicatif élevé. Il s'agit de :

- critère de SCHWARTZ

SC

$$\begin{aligned} &= \ln \left(\frac{SCR}{n} \right) \\ &+ \frac{k}{n} \ln(n) \end{aligned} \quad (4.32)$$

$$\begin{aligned}
 & - \text{critère d'information d'AKAIKE} \\
 & \text{CIA} \\
 & = \ln \left(\frac{\text{SCR}}{n} \right) \\
 & + \frac{2k}{n}
 \end{aligned} \tag{4.33}$$

Ces critères visent à retenir, parmi les modèles à comparer, celui qui minimise le critère d'information.

Il convient de noter que le coefficient de détermination R^2 n'est utilisable que dans un modèle avec terme constant. L'utilisation de cette statistique dans un modèle de régression passant par l'origine (sans terme constant) peut aussi donner un résultat négatif : $R^2 < 0$.

Enfin, pour les régressions ne présentant pas de valeur en ordonnée à l'origine, on calcule parfois un coefficient de détermination spécifique, nommé R^2 brut. Il se définit comme suit :

$$R^2 \text{ brut} = \frac{(\sum X_t Y_t)^2}{\sum X_t^2 \sum Y_t^2} \tag{4.34}$$

(Il s'agit donc des sommes des carrés brutes, non corrigées de la moyenne).

Bien que ce R^2 brut satisfait la relation $0 \leq R^2 \text{ brut} \leq 1$, il n'est pas directement comparable à la valeur du R^2 conventionnel.

4.12. Les tests statistiques

Soit le modèle linéaire général suivant :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_i X_{it} + \dots + \beta_k X_{kt} + u_t$$

Examinons quelques problèmes statistiques qui découlent de l'écriture de ce modèle.

- iv. Peut-on comparer les paramètres β_i ($i = 1, \dots, k$) à des valeurs fixées a priori ($\beta_i = \beta$).
En particulier une question intéressante est " $\beta_i = 0$ ", ce qui revient à supposer que la variable endogène Y ne dépend pas de la variable explicative X_i (test de Student).
- v. Le modèle linéaire est-il globalement significatif (test de Fisher) ?
- vi. Le modèle considéré est-il stable sur la totalité de la période (test de stabilité, ou de changement structurel) ?

4.12.1. Hypothèse de normalité

Comme il a été souligné dans le cadre d'un modèle de régression linéaire simple, l'hypothèse de normalité des termes d'erreur joue un rôle essentiel car elle va préciser la distribution statistique des estimateurs. En plus, c'est grâce à cette hypothèse que l'inférence statistique peut se réaliser.

En supposant que les erreurs suivent une loi normale centrée et de variance σ_u^2 alors :

$$Y = X\beta + U \sim N(X\beta, \sigma_u^2 I) \quad \text{et} \quad \hat{\beta} = (X'X)^{-1}X'Y \sim N(\beta, \sigma_u^2 (X'X)^{-1})$$

L'hypothèse de normalité des erreurs implique aussi que :

$$\frac{\sum_t e^2}{\sigma_u^2} = (n - k) \frac{\hat{\sigma}_u^2}{\sigma_u^2} \sim \chi^2_{(n-k)}.$$

4.12.2. Test de Student

C. Comparaison d'un paramètre β_i à une valeur fixée β_i^*

Le test d'hypothèse est le suivant :

$$H_0: \beta_i = \beta_i^*$$

$$H_1: \beta_i \neq \beta_i^*$$

Comme $\text{Var}(\hat{\beta}) = \sigma_u^2 (X'X)^{-1}$, on a :

$$\text{Var}(\hat{\beta}_i) = \sigma_u^2 v_{ii}; \quad \hat{\beta}_i \sim N(\beta_i, \sigma_u^2 v_{ii}) \quad \text{et} \quad \frac{\hat{\beta}_i - \beta_i}{\sigma_u \sqrt{v_{ii}}} = \frac{\hat{\beta}_i - \beta_i}{\sigma_{\hat{\beta}_i}^2} \sim N(0, 1)$$

où v_{ii} , $i = 1, 2, \dots, k$ est l'élément (i, i) de la matrice $(X'X)^{-1}$.

La variance des erreurs n'étant pas connue, on peut diviser la variable centrée réduite de $\hat{\beta}_i$ par la racine carrée de la statistique $\frac{\hat{\sigma}_u^2}{\sigma_u^2}$ afin d'obtenir la valeur de la statistique t de $\hat{\beta}_i$:

$$t_{\hat{\beta}_i} = \frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_u \sqrt{v_{ii}}} = \frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_{\hat{\beta}_i}} \quad (4.)$$

La statistique $t_{\hat{\beta}_i}$ est distribuée suivant une loi de Student à $(n - k)$ degrés de liberté sous l'hypothèse H_0 .

En vertu de la relation (4.), il est possible de construire un intervalle de confiance au seuil de signification α pour β_i ($i = 1, 2, \dots, k$), soit :

$$\hat{\beta}_i \pm t_{\frac{\alpha}{2}, (n-k)} \cdot \hat{\sigma}_u \sqrt{v_{ii}} \quad \text{ou bien} \quad \hat{\beta}_i \pm t_{\frac{\alpha}{2}, (n-k)} \cdot \hat{\sigma}_{\hat{\beta}_i}$$

L'intervalle de confiance (...) peut également s'écrire sous la forme :

$$P \left[\hat{\beta}_i - t_{\frac{\alpha}{2}, (n-k)} \cdot \hat{\sigma}_u \sqrt{v_{ii}} < \beta_i < \hat{\beta}_i + t_{\frac{\alpha}{2}, (n-k)} \cdot \hat{\sigma}_u \sqrt{v_{ii}} \right] = 100(1 - \alpha)\%$$

Le seuil de risque α est généralement fixé à 5 % ou 10 %.

Il est alors possible de tester l'hypothèse nulle selon laquelle le coefficient β_i est égal à une valeur donnée β_i^* , soit :

$$H_0: \beta_i = \beta_i^*$$

$$H_1: \beta_i \neq \beta_i^*$$

Si l'hypothèse nulle est vraie, alors :

$$t_{\hat{\beta}_i} = \frac{\hat{\beta}_i - \beta_i^*}{\hat{\sigma}_u \sqrt{v_{ii}}} = \frac{\hat{\beta}_i - \beta_i^*}{\hat{\sigma}_{\hat{\beta}_i}}$$

La règle de décision est la suivante :

- Si $|t_{\hat{\beta}_i}| > t_{\frac{\alpha}{2}}$: on rejette l'hypothèse H_0 au seuil statistique de 100 α %. Le coefficient β_i peut être considéré significativement différent de β_i^* .
- Si $|t_{\hat{\beta}_i}| \leq t_{\frac{\alpha}{2}}$: on accepte l'hypothèse H_0 au seuil statistique de 100 α %. Le coefficient β_i peut être considéré significativement égal à β_i^* .

D. Comparaison d'un paramètre β_i à la valeur $\beta_i = 0$

Le test le plus utilisé est celui qui consiste à tester l'hypothèse nulle :

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

Il s'agit donc d'un test de significativité de coefficient. Sous l'hypothèse nulle, la variable X_i n'est pas significative, c'est-à-dire qu'elle ne joue aucun rôle dans la détermination de la variable expliquée Y .

Sous l'hypothèse nulle, la statistique t s'écrit :

$$t_{\hat{\beta}_i} = \frac{\hat{\beta}_i}{\hat{\sigma}_{\hat{\beta}_i}}$$

La règle de décision est donc :

- Si $|t_{\hat{\beta}_i}| \leq t_{\frac{\alpha}{2}}$: on accepte l'hypothèse H_0 au seuil statistique de $100\alpha\%$, donc le coefficient $\beta_i = 0$. La variable X_i n'est pas significative et n'a aucune influence sur Y .
- Si $|t_{\hat{\beta}_i}| > t_{\frac{\alpha}{2}}$: on rejette l'hypothèse H_0 au seuil statistique de $100\alpha\%$, donc le coefficient $\beta_i \neq 0$. La variable X_i n'est pas significative et n'a aucune influence sur Y .

Les décisions associées étant d'exclure ou de conserver la variable X_i dans le modèle selon qu'on accepte ou rejette l'hypothèse H_0 .

4.12.3. Tests de Fisher

F) Tests d'hypothèse linéaire

Une hypothèse linéaire (ou restriction linéaire) est un ensemble de une ou plusieurs conditions du premier degré portant sur les coefficients d'un modèle de régression.

Soit le modèle linéaire multiple suivant :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

Les hypothèses linéaires suivantes peuvent faire l'objet de test :

$$\beta_2 = 0 \quad (\text{une condition ou contrainte});$$

$$\beta_2 = \beta_3 \quad (\text{une condition ou contrainte});$$

$$\beta_2 = -1 \quad \text{et} \quad \beta_3 = 0 \quad (\text{deux conditions ou contraintes});$$

$$\beta_2 + \beta_3 = 1 \quad \text{et} \quad \beta_1 = 0 \quad (\text{deux conditions ou contraintes});$$

$$\beta_1 = \beta_2 = \beta_3 = 0 \quad (\text{trois conditions ou contraintes}).$$

De ce qui précède, on distingue deux types de tests de contraintes linéaires sur les coefficients du modèle : les tests d'une seule contrainte et les tests de plusieurs contraintes. Les premiers types de tests utilisent une statistique distribuée selon une loi de Student, tandis que les seconds utilisent une statistique distribuée selon une loi de Fisher (dite aussi loi de Fisher-Snedecor).

G) Test d'une contrainte

Soit le modèle de régression multiple suivant :

$$Y = X\beta + U$$

où le terme d'erreur U est indépendant de X et distribué normalement.

De façon générale, une contrainte linéaire sur les paramètres du modèle de régression peut s'écrire sous la forme :

$$\begin{cases} H_0: r_1\beta_1 + r_2\beta_2 + \dots + r_k\beta_k = q \\ H_1: r_1\beta_1 + r_2\beta_2 + \dots + r_k\beta_k \neq q \end{cases}$$

ou de manière succincte :

$$\begin{cases} H_0: r'\beta = q \\ H_1: r'\beta \neq q \end{cases}$$

(avec q une contrainte et $r' = [r_1 \quad r_2 \quad \dots \quad r_k]$).

Exemple

Supposons le modèle de régression suivant : $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$, l'on veut tester l'hypothèse : $\beta_1 + 0.4\beta_2 = 1$. Cette hypothèse peut s'écrire sous la forme : $r'\beta = q$ comme suit :

$$\underbrace{\begin{bmatrix} 0 & 1 & 0.4 \end{bmatrix}}_{r'} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}}_{\beta} = \underbrace{\begin{bmatrix} 1 \end{bmatrix}}_q$$

Pour tester cette hypothèse, on estime le modèle par la méthode des moindres carrés ordinaires et on calcule la statistique $r'\hat{\beta}$. Ce procédé porte le nom de *régression non contrainte*. Sous hypothèse H_0 , son espérance et sa variance sont respectivement :

$$E(r'\hat{\beta}) = r'E(\hat{\beta}) = r'\beta = q$$

et

$$\text{Var}(r'\hat{\beta}) = r'\text{Var}(\hat{\beta})r = \sigma_u^2 r'(X'X)^{-1}r$$

La statistique $r'\hat{\beta}$, sous hypothèse nulle, est donnée par l'expression suivante :

$$t_{\text{cal}} = \frac{r'\hat{\beta} - q}{\hat{\sigma}_u \sqrt{r'(X'X)^{-1}r}}$$

La statistique t_{cal} est distribuée selon une loi de Student à $(n-k)$ degrés de liberté au seuil de signification α choisi.

Si La statistique t_{cal} est supérieure à la valeur critique de t au niveau de signification choisi pour un nombre de degrés de liberté donné, on peut rejeter l'hypothèse nulle ; sinon, on ne peut la rejeter. De manière alternative, si la valeur p (p – value) du test t associée est assez basse, c'est-à-dire si elle est inférieure au niveau de signification choisi, on peut rejeter l'hypothèse nulle.

H) Test de plusieurs contraintes

Soit le modèle de régression multiple suivant :

$$Y = X\beta + U$$

où le terme d'erreur U est indépendant de X et a une distribution normale.

On peut définir un ensemble de m contraintes (ou restrictions) linéaires sur les paramètres du modèle de régression par l'expression générale ci-après :

$$R\beta = q$$

où R est matrice de format $(m \times k)$; β , vecteur colonne des paramètres du modèle est de dimension $(k \times 1)$; q , un vecteur colonne de dimension $(m \times 1)$.

Exemple

Supposons le modèle de régression suivant :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t.$$

Les deux contraintes linéaires suivantes : $\beta_3 + \beta_4 = 1$ et $\beta_2 = 0$ sur ce modèle peuvent s'écrire aussi comme :

$$\underbrace{\begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}}_R \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}}_{\beta} = \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_q.$$

Le test sur plusieurs contraintes consiste donc à tester l'hypothèse :

$$\begin{cases} H_0: R\beta = q \\ H_1: R\beta \neq q \end{cases}$$

Pour réaliser ce test, on estime avec la méthode des moindres carrés ordinaires le modèle sans contraintes afin de calculer $R\hat{\beta}$. Etant fonction linéaire de l'estimateur $\hat{\beta}$, la statistique est également distribuée selon une loi normale.

Sous l'hypothèse H_0 , son espérance et sa variance sont données respectivement par :

$$E(R\hat{\beta}) = R E(\hat{\beta}) = R\beta = q$$

et

$$\text{Var}(R\hat{\beta}) = R \text{Var}(\hat{\beta}) R' = \sigma_u^2 R (X'X)^{-1} R'$$

Par conséquent :

$$R\hat{\beta} \sim N[q, \sigma_u^2 R (X'X)^{-1} R']$$

Sous l'hypothèse H_0 que ces contraintes sont vraies, c'est-à-dire $R\beta = q$, la statistique :

$$F_{\text{cal}} = \frac{(R\hat{\beta} - R\beta)' [\sigma_u^2 R (X'X)^{-1} R']^{-1} (R\hat{\beta} - R\beta)}{m} \sim F_{(m, n-k)}$$

Cette statistique peut s'écrire aussi comme suit :

$$F_{\text{cal}} = \frac{\{(R\hat{\beta} - R\beta)' [R (X'X)^{-1} R']^{-1} (R\hat{\beta} - R\beta)\} / m}{\text{SCR} / (n - k)} \sim F_{(m, n-k)}$$

où $\hat{\beta}$ est le vecteur des coefficients estimés sur le modèle non contraint.

De manière équivalente, on peut également trouver une valeur identique en utilisant un test qui nécessite les estimations du modèle sans contraintes et avec contraintes.

Les étapes de calcul sont les suivantes :

- Commencer par régresser le modèle initial sans aucune restriction, et garder la somme des carrés des résidus ainsi obtenus (SCR_{NC}) ;

- Ensuite régresser par la méthode des moindres carrés le modèle transformé (appelé aussi modèle contraint) obtenu en imposant les restrictions au modèle initial, et calculer la somme des carrés des résidus (SCR_C) ;
- Enfin, calculer l'une des deux statistiques équivalentes suivantes :

$$F_{cal} = \frac{(SCR_C - SCR_{NC})/m}{SCR_{NC}/(n-k)} \sim F_{(m, n-k)}$$

ou

$$F_{cal} = \frac{(R_{NC}^2 - R_C^2)/m}{(1 - R_{NC}^2)/(n-k)} \sim F_{(m, n-k)}$$

avec : R_{NC}^2 représente le coefficient de détermination du modèle non contraint (modèle initial) ;

R_C^2 , le coefficient de détermination du modèle contraint (modèle transformé).

- Si $F_{cal} > F_{(m, n-k)}$, on rejette l'hypothèse nulle.
- Si $F_{cal} \leq F_{(m, n-k)}$ on accepte l'hypothèse H_0 .

De même, si la valeur p (p -value) du F_{cal} est suffisamment faible, on peut rejeter H_0 .

Une autre façon de procéder consiste à utiliser le ratio de vraisemblance (connu comme test LR: likelihood ratio) des modèles contraint et non contraint. Si la contrainte est vraie, on aura : $L_C < L_{NC}$, où L_{NC} est la fonction de vraisemblance du modèle non contraint et L_C , la fonction de vraisemblance du modèle contraint.

Ce test se ramène aussi à un test du χ^2 par le calcul de la statistique :

$LR = -2 \ln \lambda \sim \chi^2$ à m degrés de liberté (où $\lambda = \frac{L_C}{L_{NC}}$ et m étant le nombre des contraintes).

Si LR est supérieur à χ_m^2 lu dans la table statistique au seuil α choisi et à m degrés de liberté, on rejette l'hypothèse H_0 , les restrictions ne sont pas vérifiées.

Exemple : Soit le modèle de demande de thé de Ceylan aux Etats-Unis:

$$\ln Q_t = \beta_1 + \beta_2 \ln P_{Ct} + \beta_3 \ln P_{It} + \beta_4 \ln P_{Bt} + \beta_5 \ln R_t + u_t$$

où Q représente les importations de thé de Ceylan, P_C le prix du thé de Ceylan, P_I le prix du thé d'Inde, P_B le prix du café du Brésil, R le revenu national et u le terme d'erreur.

On veut tester l'hypothèse : $\beta_2 = -1$ et $\beta_3 = 0$, c'est-à-dire l'hypothèse d'élasticité unitaire des quantités au prix du thé de Ceylan (ε_{Q/P_C}) et conjointement d'absence d'influence du prix du thé d'Inde (ε_{Q/P_I}).

- La régression par les moindres carrés ordinaires du modèle non contraint donne les résultats suivants :

$$\ln Q_t = 2,8370 + -1,4810 \ln P_{Ct} + 1,1810 \ln P_{It} + 0,1860 \ln P_{Bt} + 0,2570 \ln R_t$$

(2,0000) (0,9870) (0,6900) (0,1340)

(0,3700)

avec $n = 22$ et $SCR_{NC} = 0,4277$

- Sous l'hypothèse $H_0 : \beta_2 = -1$ et $\beta_3 = 0$, le modèle transformé (modèle contraint) s'écrit :

$$\ln Q_t = \beta_1 - \ln P_{Ct} + \beta_4 \ln P_{Bt} + \beta_5 \ln R_t + u_t$$

ou encore :

$$\ln Q_t + \ln P_{Ct} = \beta_1 + \beta_4 \ln P_{Bt} + \beta_5 \ln R_t + u_t$$

L'application de la méthode des moindres carrés sur le modèle contraint donne les résultats suivants :

$$\ln Q_t + \ln P_{Ct} = -0,7380 + 0,1990 \ln P_{Bt} + 0,2610 \ln R_t$$

(0,6900) (0,1340) (0,3700)

avec $n = 22$ et $SCR_C = 0,6788$

- La statistique du test est donc :

$$F_{cal} = \frac{(0,6788 - 0,4277)/2}{0,4277/(22 - 5)} = 4,99$$

Au seuil de 5 %, la valeur de la loi de Fisher $F_{(2,17)}$ lue dans la table est égale à 3,59. On a donc $F_{cal} = 4,99 > 3,59$, ce qui signifie que l'on rejette au risque 5% l'hypothèse linéaire envisagée.

I) Test de nullité d'un coefficient de régression particulier

La nullité d'un coefficient de régression est une hypothèse linéaire particulière où le nombre de restrictions $m = 1$. En effet, ce cas correspond à :

$$R = [0 \quad \dots \quad 0 \quad 1 \quad 0 \quad \dots \quad 0] \text{ et } q = [0]$$

La matrice R contient un seul 1 à la $i^{\text{ème}}$ place et q est nul, ce qui revient bien à tester la significativité du coefficient β_i . Si l'on pose $q = [\beta_i^*]$, on retrouve le test de significativité de l'égalité de β_i à une certaine valeur β_i^* .

Il convient de faire remarquer que le test de Fisher d'une telle hypothèse linéaire est sans intérêt car il est mathématiquement équivalent au test de Student de significativité d'un coefficient.

J) Test de signification globale d'une régression multiple

Ce test est aussi un cas particulier du test d'hypothèses linéaires portant sur la significativité de l'ensemble des coefficients des variables explicatives de la régression. Il s'intéresse particulièrement à l'évaluation globale du modèle dans le cadre d'une régression multiple, c'est-à-dire si l'ensemble des variables

explicatives a une influence sur la variable à expliquer Y. Un tel test est parfois appelé *test de significativité de la régression* ou encore *test de la relation d'ensemble* ou plus communément *test de Fisher*. Il correspond au cas suivant :

$$\underbrace{\begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}}_{\mathbf{R}} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}}_{\boldsymbol{\beta}} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{q}}$$

On teste donc :

$$\begin{cases} H_0 : \beta_2 = \beta_3 = \dots = \beta_k = 0 \\ H_1 : \exists \text{ au moins un des coefficients non nuls.} \end{cases}$$

Ce test ne considère pas le terme constant β_1 , car un modèle dans lequel seul le terme constant est significatif n'a aucun sens économique.

Si l'hypothèse nulle est vraie, alors :

$$\begin{aligned} F_{\text{cal}} &= \frac{(\hat{Y}_t - \bar{y})^2 / (k - 1)}{\sum e_t^2 / (n - k)} \\ &= \frac{\text{SCE} / (k - 1)}{\text{SCR} / (n - k)} \\ &= \frac{R^2 / (k - 1)}{(1 - R^2) / (n - k)} \end{aligned}$$

Exemple

Soit la fonction de production Cobb-Douglas :

$$\ln Q_t = \beta_1 + \beta_2 \ln K_t + \beta_3 \ln L_t + u_t$$

où Q_t représente le Produit National Brut au temps t ;

K_t , le facteur capital à la période t ;

L_t , le facteur travail à la période t ;

t , le temps (de 1929 à 1961).

En effectuant la régression par les moindres carrés ordinaires sur la période globale on obtient :

$$\begin{array}{ccc} \ln \widehat{Q}_t = & -3,87 & + 0,41 \ln K_t + 1,41 \ln L_t \\ & (0,255) & (0,050) \quad (0,089) \end{array}$$

$$n = 33$$

$$R^2 = 0,9937$$

(•) = écarts types estimés des estimateurs des coefficients de régression.

3) Interpréter les résultats de la régression par la méthode des moindres carrés ;

4) Tester au seuil de 5 % les hypothèses linéaires suivantes :

a) $\beta_2 = 0.4$;

b) $\beta_3 = 0$.

Solution

3) Interprétation

Les coefficients du modèle de production Cobb Douglas représentent respectivement l'élasticité du Produit National Brut au facteur capital et l'élasticité du Produit National Brut au facteur travail. En effet :

$$\beta_2 = \frac{\partial \ln Q}{\partial \ln K} = \frac{\Delta Q/Q}{\Delta K/K} = \varepsilon_{Q/K};$$

$$\beta_3 = \frac{\partial \ln Q}{\partial \ln L} = \frac{\Delta Q/Q}{\Delta L/L} = \varepsilon_{Q/L};$$

β_2 et β_3 sont des rapports de dérivées logarithmiques.

L'élasticité du produit national brut au facteur travail est égale à $\beta_2 = 1,41$.

L'élasticité du produit national brut au facteur capital est égale à $\beta_3 = 0,41$.

Si le facteur travail augmente de 10 %, le Produit National Brut augmente de 14,1 % et si le facteur capital augmente de 10 %, le Produit National Brut augmente de 4,1%.

L'on peut également déterminer les rendements d'échelle car ces derniers sont :

- **décroissants** si $\beta_2 + \beta_3 < 1$ (le Produit National Brut augmente dans une proportion moindre que les facteurs de production) ;
- **constants** si $\beta_2 + \beta_3 = 1$ (le Produit National Brut augmente dans une proportion identique aux facteurs de production) ;
- **croissants** si $\beta_2 + \beta_3 > 1$ (le Produit National Brut augmente plus vite que les facteurs de production).

La valeur de $R^2 = 0,99$ signifie qu'environ 99 % de la variation du logarithme du Produit National Brut sont expliqués par les logarithmes des quantités de travail et de capital.

4) Tests statistiques

d) Test d'hypothèse de significativité du coefficient β_3

$$H_0 : \beta_3 = 0$$

$$H_1 : \beta_3 \neq 0$$

Le facteur travail a-t-il une influence significative sur le Produit National Brut ?

La valeur du ratio de Student permet de répondre à cette question. En effet :

$$t_{\text{calculé}} = \left| \frac{\hat{\beta}_3}{\hat{\sigma}_{\beta_3}} \right| = \left| \frac{1,41}{0,089} \right| = 15,84.$$

Au seuil de 5 %, la lecture de la table de Student donne $t_{(0,025; 30)} = 2,042$.

Le ratio de Student est supérieur à $t_{(0,025; 30)}$, le coefficient β_3 est significativement différent de 0 ; le facteur travail a une influence significative sur le produit national brut.

e) Test d'hypothèse sur β_2

$$H_0 : \beta_2 = 0,40$$

$$H_1 : \beta_2 \neq 0,40$$

Sous hypothèse nulle, alors :

$$t_{\text{calculé}} = \left| \frac{\hat{\beta}_2 - 0,40}{\hat{\sigma}_{\beta_2}} \right| = \left| \frac{0,41 - 0,40}{0,050} \right| = 0,20.$$

La valeur de t_{cal} est inférieure à $t_{(0,025; 30)} = 2,042$: on accepte l'hypothèse nulle H_0 que $\beta_2 = 0,40$.

f) Test de significativité globale du modèle

$$H_0 : \beta_2 = \beta_3 = 0$$

$$H_1 : \exists \text{ au moins un coefficient non nul}$$

Sous hypothèse nulle, la statistique Fisher empirique est égal à :

$$F_{\text{cal}} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{0,9937/2}{(1-0,9937)/30} = 2365,95.$$

Au seuil de 5 %, la valeur critique fournie par la table de Fisher pour 2 (c'est-à-dire $k-1 = 3-1$) degrés de liberté au numérateur et 30 (c'est-à-dire $n-k = 33-3$) degrés de liberté au dénominateur est $F_{(2;30)} = 3,32$.

Puisque $F_{\text{cal}} > F_{(2;30)}$, nous rejetons l'hypothèse de nullité de tous les coefficients, la régression est globalement significative.

4.13. Tests de stabilité

Les tests de stabilité des paramètres estimés du modèle de régression, appelés aussi tests de robustesse du modèle estimé, visent à vérifier si les valeurs des coefficients estimés sont stables sur l'ensemble de la période d'étude considérée. On cherche donc à vérifier s'il n'y a pas eu de changements structurels entre la variable expliquée et les variables explicatives. Notons que l'instabilité des paramètres peut avoir plusieurs sources attribuables soit aux forces externes, aux calamités naturelles ou aux variations de politique économique.

Il existe diverses méthodes pour détecter l'instabilité des coefficients estimés d'un modèle de régression. Nous allons exposer deux types de tests de stabilité des coefficients :

- le test de Chow ;
- les tests Cusum et Cusum Carré de Brown, Durbin et Evans.

4.13.1. Test de Chow

Le test de Chow, appelé aussi test de changement structurel, permet d'examiner si les coefficients d'une régression sont stables par rapport aux observations utilisées.

Sur des séries temporelles, on compare les estimations effectuées sur deux (ou plusieurs) sous ensembles d'observations qui correspondent à un découpage en périodes de l'échantillon initial. On parle dans ce cas de *test de stabilité temporelle* de la régression.

Sur les données en coupe transversale, on peut comparer les résultats obtenus par exemple sur des individus, des pays, des régions, des secteurs industriels différents. Concernant les individus, on peut s'intéresser à des résultats par classe d'âge, par sexe, etc. Dans ce cas, le test de Chow est utilisé afin de déterminer si des groupes d'individus sont homogènes ou pas. On parle alors de *test d'homogénéité des comportements*.

Considérons le modèle linéaire général suivant, pour $t=1,2,\dots,n$:

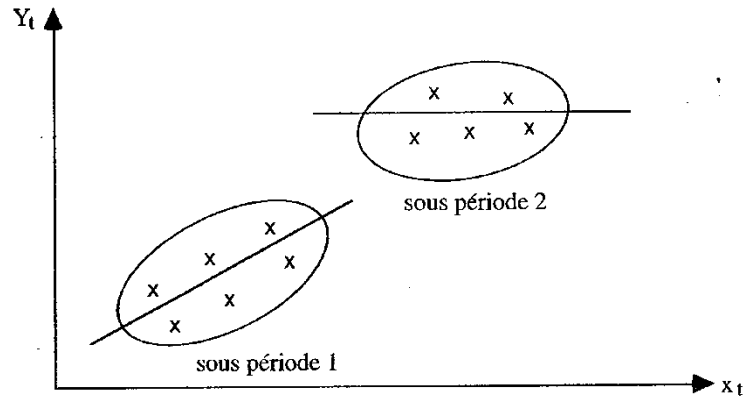
$$\underbrace{Y}_{(n \times 1)} = \underbrace{X}_{(n \times k)} \underbrace{\beta}_{(k \times 1)} + \underbrace{U}_{(n \times 1)} \quad (4.1)$$

Supposons que l'on divise l'échantillon d'observations en deux sous-périodes distinctes (n_1 et n_2 ; $n_1 + n_2 = n$) et que l'on estime les deux modèles suivants :

$$\underbrace{Y^1}_{(n_1 \times 1)} = \underbrace{X^1}_{(n_1 \times k)} \underbrace{\beta^1}_{(k \times 1)} + \underbrace{U^1}_{(n_1 \times 1)} \quad (4.2)$$

et

$$\underbrace{Y^2}_{(n_2 \times 1)} = \underbrace{X^2}_{(n_2 \times k)} \underbrace{\beta^2}_{(k \times 1)} + \underbrace{U^2}_{(n_2 \times 1)} \quad (4.3)$$



On se pose le problème de la stabilité des coefficients du modèle dans le temps.

Le principe du test est de voir dans quelle mesure le fait de régresser séparément sur les deux sous périodes améliore le résultat de la régression. Ce test portera sur les sommes des carrés des résidus (variances résiduelles).

Le test de Chow consiste à tester l'hypothèse nulle :

$$H_0 : \begin{cases} \beta_0^1 = \beta_0^2 = \beta_0 \\ \beta_1^1 = \beta_1^2 = \beta_1 \\ \vdots \\ \beta_k^1 = \beta_k^2 = \beta_k \end{cases} \quad (4.4)$$

De manière équivalente, ce test peut s'écrire comme suit :

$$\begin{aligned} H_0 : SCR &= SCR_1 + SCR_2 \quad (\text{stabilité}) \\ H_1 : SCR &\neq SCR_1 + SCR_2 \quad (\text{instabilité}) \end{aligned} \quad (4.5)$$

Les étapes de test de Chow sont les suivantes :

- Estimer le modèle (4.1) et calculer la somme des carrés des résidus (SCR) correspondante, avec $(n - K)$ degrés de liberté.
- Estimer ensuite le modèle (4.2) et calculer la somme des carrés des résidus (SCR_1) correspondante, avec $(n_1 - K)$ degrés de liberté.
- Estimer enfin le modèle (4.3) et calculer la somme des carrés des résidus (SCR_2) correspondante, avec $(n_2 - K)$ degrés de liberté.
- Calculer $SCR_3 = SCR_1 + SCR_2$, $(n_1 - K) + (n_2 - K) = (n - 2K)$ degrés de liberté.
- Calculer la statistique de test :

$$F_{cal} = \frac{[SCR - (SCR_1 + SCR_2)]/K}{(SCR_1 + SCR_2)/(n - 2K)} \quad (4.6)$$

où $K = k + 1$ est le nombre des coefficients du modèle.

Cette statistique suit, sous l'hypothèse nulle d'absence de changement structurel, une loi de Fisher avec respectivement K degrés de liberté pour le numérateur et $(n - 2K)$ degrés de liberté pour le dénominateur.

La règle de décision est alors la suivante :

- Si $F_{\text{cal}} < F_{(K, n-2K)}$, on accepte l'hypothèse nulle de stabilité des coefficients. Il n'y a donc pas de changement structurel.
- Si $F_{\text{cal}} > F_{(K, n-2K)}$, on rejette l'hypothèse nulle de stabilité des coefficients. Il y a eu donc changement structurel.

Remarques :

- Le test de Chow peut être facilement généralisé à la présence de plus d'un changement structurel.
- Ce test nécessite nécessairement que soit connue la date à laquelle se produit la rupture structurelle.
- Enfin, ce test s'applique de manière identique sous les données en coupe transversale (cross section) pour tester l'homogénéité du comportement de groupes d'individus.

Exemple

Soit la fonction d'importations ci-après :

$$\ln M_t = \beta_1 + \beta_2 \ln \text{PIB}_t + u_t$$

où :

- M désigne les importations ;
- PIB est le Produit intérieur brut.

L'estimation de cette relation sur la période 1962-1995 donne les résultats suivants :

$$\begin{aligned} \widehat{\ln M_t} &= 0,65 + 0,75 \ln \text{PIB}_t \\ R^2 &= 0,87 ; \quad \text{SCR} = 0,134 ; \quad n = 34 \end{aligned}$$

Cette fonction d'importations estimée est-elle stable sur la totalité de la période considérée au seuil de 5%?

Solution

Pour appliquer le test de Chow, on estime deux nouvelles régressions : une régression sur la période 1960-1978 et une régression sur la période 1979-1995. Les résultats de ces régressions sont donnés ci-après :

- Sur la période 1962-1978, soit $t=1, \dots, 17$:

$$\begin{aligned} \widehat{\ln M_t} &= 1,34 + 0,65 \ln \text{PIB}_t \\ R_1^2 &= 0,46 ; \quad \text{SCR}_1 = 0,102 ; \quad n_1 = 17 \end{aligned}$$
- Sur la période 1979-1995, soit $t=18, \dots, 34$:

$$\begin{aligned} \widehat{\ln M_t} &= 2,15 + 0,54 \ln \text{PIB}_t \\ R_2^2 &= 0,84 ; \quad \text{SCR}_2 = 0,014 ; \quad n_2 = 17 \end{aligned}$$

La statistique du test de Chow est donc :

$$F_{cal} = \frac{[SCR - (SCR_1 + SCR_2)]/k}{(SCR_1 + SCR_2)/(n - 2k)} = \frac{0,134 - (0,102 + 0,014)/2}{(0,102 + 0,014)/30} = 2,30$$

La lecture de la table de Fisher au seuil de 5 % donne : $F_{(2;30)} = 3,32$. On constate que valeur calculée de la statistique de test est inférieure à la valeur critique. Par conséquent, l'hypothèse nulle de stabilité des coefficients du modèle est acceptée au seuil de 5 %. La fonction d'importations est donc stable sur la période considérée.

La valeur critique $F_{théorique}$ est $F_{0,05}(2,30) = 3,32$. Donc : $F_{cal} < F_{théorique}$, on accepte l'hypothèse de stabilité. Cette la fonction d'importations estimée est stable.

4.13.2. Tests de stabilité temporelle basés sur les résidus récurrents

L'intérêt de ces tests réside dans le fait qu'il permet d'étudier la stabilité d'une régression sans définir a priori la date de rupture sur les coefficients. Contrairement au test de Chow, les tests proposés par Brown, Durbin et Evans sont basés sur l'utilisation des résidus récurrents.

Soit le modèle de régression linéaire suivant :

$$Y = X\beta + U$$

ou encore :

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_i X_{it} + \dots + \beta_k X_{kt} + u_t \quad (t = 1, 2, \dots, n)$$

Pour calculer les résidus récurrents du modèle, on doit estimer par la méthode des moindres carrés ordinaires les paramètres β avec un nombre d'observations qui varie de p à n . Et le $r^{ème}$ résidu récurrent, e_r (où $r = p + 1, \dots, n$), appelé aussi l'erreur de prévision ex-post sur l'observation y_r , est donné par l'expression :

$$e_r = y_r - \hat{y}_r = y_r - x_r' \hat{\beta}_{r-1}$$

où : $\hat{\beta}_{r-1} = (X_{r-1}' X_{r-1})^{-1} X_{r-1}' Y_{r-1}$, estimateur des moindres carrés ordinaires lorsque la taille de l'échantillon est égale à $(r - 1)$;

$x_r' = [1 \quad X_{1r} \quad X_{2r} \quad \dots \quad X_{kr}]$ est le vecteur des variables explicatives plus la constante pour la $r^{ème}$ observation ;

Y_r , le sous vecteur des $(r - 1)$ premiers éléments de Y .

Sous l'hypothèse de stabilité $y_r = x_r' \beta + u_r$ et l'erreur de prévision devient :

$$e_r = y_r - \hat{y}_r = x_r' \beta + u_r - x_r' \hat{\beta}_{r-1} = u_r - x_r' (\hat{\beta}_{r-1} - \beta)$$

Or $E(\hat{\beta}_{r-1}) = \beta$ et $Var(\hat{\beta}_{r-1}) = E[(\hat{\beta}_{r-1} - \beta)(\hat{\beta}_{r-1} - \beta)'] = \sigma_u^2 (X_{r-1}' X_{r-1})^{-1}$, d'où :

$$Var(e_r) = \sigma_{e_r}^2 = \sigma_u^2 [1 + x_r' (X_{r-1}' X_{r-1})^{-1} x_r].$$

Les résidus récurrents, notés w_r , sont définis comme les erreurs de prévision normalisées :

$$w_r = \frac{e_r}{\sqrt{1 + \mathbf{x}_r'(X_{r-1}'X_{r-1})^{-1}\mathbf{x}_r}}$$

avec $w_r \sim N(0, \sigma_u^2)$.

A partir de ces résidus récurrents, Brown, Durbin et Evans (1975) ont proposé les tests Cusum (Cumulated Sum of residuals) et Cusum carré (Cumulated Sum of Squared residuals : Cusumsq) qui permettent de tester la stabilité des coefficients estimés dans un modèle de régression.

Il s'agit de tester l'hypothèse nulle que le vecteur β est identique à chaque période conditionnellement à l'hypothèse $\sigma_{u_1}^2 = \dots = \sigma_{u_i}^2 = \dots = \sigma_{u_n}^2 = \sigma_u^2$, soit :

$$H_0: \beta_1 = \beta_2 = \dots = \beta_n = \beta$$

$$H_1: \beta_t \neq \beta \text{ ou } \sigma_{u_n}^2 \neq \sigma_u^2$$

où les coefficients $\beta_t, t = 1, 2, \dots, n$ sont les vecteurs des coefficients de régression pour la période t et les $\sigma_{u_t}^2$ désignent les variances des erreurs de la même période.

C. Test Cusum

Le test Cusum est basé sur la somme cumulée des résidus récurrents w_r définie par :

$$W_t = \sum_{j=r+1}^t \frac{w_j}{\hat{\sigma}_w}$$

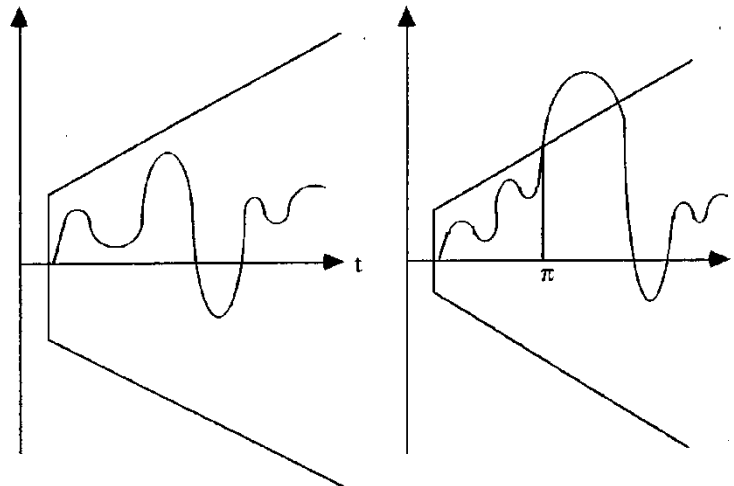
où : $t = 1, 2, \dots, n$;

$$\bar{w} = \frac{1}{n-r-1} \sum_{j=r+1}^n w_j \quad \text{et} \quad \hat{\sigma}_w^2 = \frac{1}{n-r-1} \sum_{j=r+1}^n (w_j - \bar{w})^2.$$

Sous l'hypothèse nulle de stabilité, W_t est de moyenne nulle et de variance approximativement égale au nombre total des résidus (car les termes sont indépendants et ont chacun une variance égale à 1).

Le test est effectué en représentant graphiquement la série W_t par rapport à t et la région de confiance donnée par l'intervalle de confiance $[-C_\alpha, C_\alpha]$. Pour $t = r$, $C_\alpha = a(n-r)^{1/2}$ et pour $t = n$, $C_\alpha = 3a(n-r)^{1/2}$. La région de confiance du test Cusum pour les deux points $t = r+1, \dots, n-1$ est représentée par l'espace entre les deux droites reliant ces points.

Si W_t reste dans cet intervalle quel que soit $t = r+1, \dots, n$, alors l'hypothèse de stabilité des paramètres est retenue. Sinon, si à une date donnée, W_t va au-delà d'une des droites, alors on considère qu'à cette date il y a une rupture et que les paramètres du modèle estimé ne sont pas stables.



Le test Cusum est généralement utilisé pour détecter d'éventuels mouvements systématiques dans la valeur des coefficients reflétant une possible instabilité structurelle. Si une rupture est constatée, on rejette la spécification choisie sur l'ensemble de la période.

D. Test Cusum Carré

Le test de Cusum Carré est similaire au précédent. Il est utilisé pour détecter des mouvements aléatoires c'est-à-dire des mouvements ne provenant pas forcément d'une modification structurelle des coefficients. Ce test consiste à représenter graphiquement les sommes cumulées des résidus récurrents S_t :

$$S_t = \frac{\sum_{j=r+1}^t w_j^2}{\sum_{j=r+1}^n w_j^2}$$

($t = r + 1, \dots, n$)

Comme dans le test Cusum, la région critique du test Cusum carré peut être représentée graphiquement. Les deux bornes sont données par $E(S_t) \pm c_0$, où les valeurs de c_0 sont données dans la table de Durbin pour diverses tailles d'échantillon et divers seuils de signification. Si, quelque soit $t = r + 1, \dots, n$, S_t se situe dans le corridor défini par les deux bornes alors les paramètres du modèle sont supposés stables. Sinon, si S_t sort des bornes données par les deux droites à une période $t = \pi$, on conclut à l'existence d'une rupture aléatoire qui traduit l'instabilité des coefficients de la régression pour cette date.

Exemple d'illustration

La figure ci-dessous correspond à l'application du test Cusum pour un seuil statistique de 5 %. On constate que la série de la somme cumulée des résidus récurrents reste à l'intérieur de l'intervalle formé des droites, suggérant l'absence d'instabilité de la relation sur la période considérée.

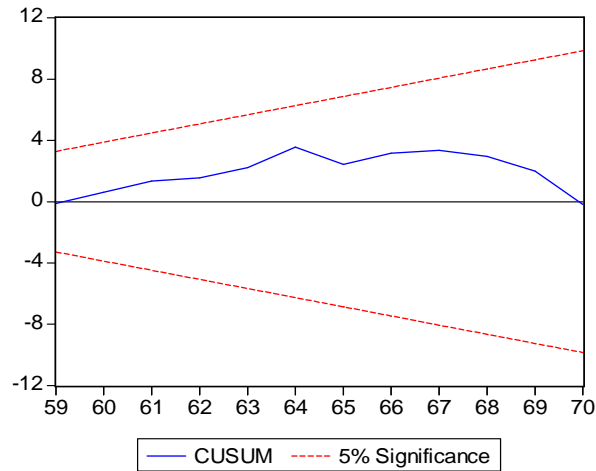


Fig. : Test Cusum

La figure suivante correspond à l'application du test Cusum carré au seuil de signification de 5 %. Comme dans la figure précédente, ce graphique met en évidence que la somme cumulée des résidus récurrents au carré est pratiquement à l'intérieur du corridor confirmant ainsi la stabilité des paramètres sur la période.

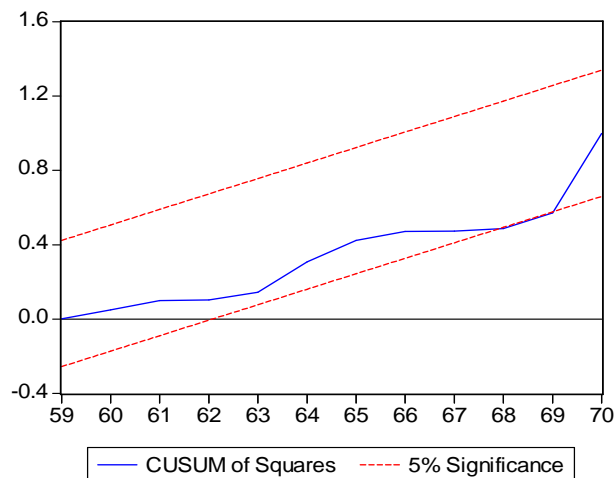


Fig. : Test Cusum carré

4.14. Prédiction

Le problème consiste à déterminer, à partir d'un modèle dont les paramètres ont été estimés, quelle valeur attribuée à la variable expliquée lorsque nous connaissons les valeurs des variables exogènes.

Si le modèle général estimé est comme ci-dessous :

$$\hat{Y}_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt}$$

Pour un vecteur X donné des valeurs prises par les variables explicatives lors de la période sur laquelle porte la prévision, la prévision ponctuelle pour la période $t+h$ est la suivante :

$$\hat{Y}_{t+h} = \hat{\beta}_1 + \hat{\beta}_2 X_{2,t+h} + \dots + \hat{\beta}_k X_{k,t+h}$$

Sous forme matricielle, l'équation ci-dessus peut s'écrire comme suit :

$$\hat{Y}_{t+h} = [1 \quad X_{1,t+h} \quad X_{2,t+h} \quad \dots \quad X_{k,t+h}] \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = R\hat{\beta}$$

où R est le vecteur des valeurs prévues des variables explicatives.
La vraie valeur de Y pendant de la période de prévision est donc :

$$Y_{t+h} = R\beta + u_{t+h},$$

avec u_{t+h} désignant la vraie valeur prise par le terme d'erreur pendant cette période.

L'erreur de prévision e_{t+h} est donnée par l'écart entre la valeur observée et la valeur prévue :

$$\begin{aligned} e_{t+h} &= Y_{t+h} - \hat{Y}_{t+h} \\ &= R(\hat{\beta} - \beta) + u_{t+h} \end{aligned}$$

Notons que e_{t+h} est une variable aléatoire car la réalisation Y_{t+h} n'est pas connue au moment de la prévision. Les moments de cette variable aléatoire sont :

c) L'espérance mathématique de l'erreur de prévision est égale à :
 $E(e_{t+h}) = 0$,
 car $E(\hat{\beta}) = \beta$ et $E(u_{t+h}) = 0$.

d) La variance de l'erreur de prévision est donc :

$$\begin{aligned} \text{Var}(e_{t+h}) &= \sigma_{e_{t+h}}^2 \\ &= E\{[-R(\hat{\beta} - \beta) + u_{t+h}][-R(\hat{\beta} - \beta) + u_{t+h}]'\} \\ &= \sigma_u^2 R(X'X)^{-1}R' + \sigma_u^2 \\ &= \sigma_u^2 [R(X'X)^{-1}R' + 1]. \end{aligned}$$

L'erreur de prévision e_{t+h} est une variable aléatoire distribuée selon une loi normale :

$$e_{t+h} \sim N(0, \sigma_{e_{t+h}}^2)$$

par conséquent :

$$\frac{e_{t+h}}{\sqrt{\sigma_u^2 [R(X'X)^{-1}R' + 1]}} \sim N(0, 1)$$

et, en remplaçant la variance théorique de l'erreur σ_u^2 par la variance empirique $\hat{\sigma}_u^2$, on obtient la statistique :

$$t_{\text{cal}} = \frac{e_{t+h}}{\sqrt{\hat{\sigma}_u^2 [R(X'X)^{-1}R' + 1]}} = \frac{Y_{t+h} - \hat{Y}_{t+h}}{\sqrt{\hat{\sigma}_u^2 [R(X'X)^{-1}R' + 1]}}$$

$$= \frac{Y_{t+h} - \hat{Y}_{t+h}}{\sqrt{\hat{\sigma}_{e_{t+h}}^2}} = \frac{Y_{t+h} - \hat{Y}_{t+h}}{\hat{\sigma}_{e_{t+h}}} \sim t_{(n-k)}.$$

On peut ainsi établir l'intervalle de confiance d'une prédiction de Y_{t+h} , pour un ensemble d'observations donné :

$$P\left[\hat{Y}_{t+h} - \hat{\sigma}_{e_{t+h}} \cdot t_{\frac{\alpha}{2}; (n-k)} \leq Y_{t+h} \leq \hat{Y}_{t+h} + \hat{\sigma}_{e_{t+h}} \cdot t_{\frac{\alpha}{2}; (n-k)}\right] = 1 - \alpha.$$

L'intervalle de prévision au seuil de signification α est alors :

$$\hat{Y}_{t+h} \pm \hat{\sigma}_{e_{t+h}} \cdot t_{\frac{\alpha}{2}; (n-k)}$$

ou encore :

$$\hat{Y}_{t+h} - t_{\frac{\alpha}{2}; (n-k)} \sqrt{\hat{\sigma}_u^2 [R(X'X)^{-1}R' + 1]}$$

ou enfin :

$$\hat{Y}_{t+h} - t_{\frac{\alpha}{2}; (n-k)} \hat{\sigma}_u \sqrt{[R(X'X)^{-1}R' + 1]}.$$

Pour évaluer la qualité prédictive d'un modèle, plusieurs méthodes sont proposées. L'une des méthodes consiste à comparer les valeurs observées et les valeurs estimées, c'est-à-dire s'intéresser à la série des résidus. La série des valeurs estimées doit bien reproduire l'évolution de la série observée et, en particulier, les points de retournement.

On peut également, pour plus de précision, utiliser des méthodes plus formelles. Il s'agit de calculer différentes statistiques permettant d'évaluer l'exactitude des prévisions :

1. L'erreur moyenne (mean error)
$$ME = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)$$
2. Erreur absolue moyenne (Mean Absolute Error)
$$MAE = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t|$$
3. Racine de l'écart-type moyen (Root Mean Square Error)
$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2}$$
4. Statistique U de Theil
$$U = \sqrt{\frac{\sum_{t=2}^n (\Delta Y_t - \Delta \hat{Y}_t)^2}{\sum_{t=2}^n (\Delta Y_t)^2}}$$

où $\Delta Y_t = Y_t - Y_{t-1}$ et $\Delta \hat{Y}_t = \hat{Y}_t - \hat{Y}_{t-1}$.

Les trois premières mesures (ME, MAE et RMSE) sont utilisées lorsqu'on compare le pouvoir de prévision de plusieurs modèles estimés. Le modèle pour lequel les trois statistiques sont les plus faibles est considéré avoir une meilleure capacité prédictive. Mais, notons que ces statistiques présentent un problème évident de normalisation. En effet, la multiplication des valeurs de la variable dépendante par n'importe quel facteur multiplie également la mesure par ce scalaire.

La statistique U de Theil n'est pas influencée par ce facteur de normalisation. Elle permet de comparer les prévisions du modèle à des prévisions naïves. Lorsque $\hat{Y}_t = Y_{t-1}$ alors $\Delta Y_t = 0$ et, par conséquent, la statistique de Theil s'égale à 1. Ainsi, une valeur de la statistique de Theil

inférieure à 1 indique que le modèle permet de réaliser des prévisions meilleures que des prévisions naïves.

Exemple illustratif

L'estimation du modèle de Cobb-Douglas pour l'économie mexicaine de 1955-1970 a donné les résultats suivants :

$$\ln \widehat{Q}_t = -3,2116 + 0,7931 \ln K_t + 0,5811 \ln L_t$$

(0,3933) (0,0625) (0,1234)

$$n = 16 ;$$

$$R^2 = 0,9987 ; \bar{R}^2 = 0,9984 ;$$

$$\hat{\sigma}_u = 0,012209 \text{ (écart-type estimé de l'erreur)} ;$$

$$F_{\text{calculé}} = 4808$$

(•) = écarts types estimés des estimateurs des coefficients de régression.

Notons que :

$$X'X = \begin{bmatrix} 16 & 200,57189 & 147,78782 \\ 200,57189 & 2515,53683 & 1853,23835 \\ 147,78782 & 1853,23835 & 1365,39069 \end{bmatrix}$$

$$\Rightarrow (X'X)^{-1} = \begin{bmatrix} 1037,7331 & 141,69943475 & -304,650632 \\ 141,699434 & 26,2374792 & -51,9493389 \\ -304,650632 & -51,9493389 & 102,128953 \end{bmatrix}$$

3) Les prévisions ponctuelles sont données par :

- Pour l'année 1971 :

$$\hat{M}_{t+1} = \hat{M}_{1971} = -3,211626 + (0,793127) * (13,09124) + (0,581095) * (9,498372) \\ = 12,6908384$$

- Pour l'année 1972 :

$$\hat{M}_{t+2} = \hat{M}_{1972} = -3,211626 + (0,793127) * (13,16265) + (0,581095) * (9,527921) \\ = 12,7646463$$

- Pour l'année 1973 :

$$\hat{M}_{t+3} = \hat{M}_{1973} = -3,211626 + (0,793127) * (13,23842) + (0,581095) * (9,675583) \\ = 12,9105408$$

- Pour l'année 1974 :

$$\hat{M}_{t+4} = \hat{M}_{1974} = -3,211626 + (0,793127) * (13,32093) + (0,581095) * (9,557753) \\ = 12,9075196$$

4) Les prévisions par intervalle au seuil de 5 % :

- Pour l'année 1971 :

$$\hat{M}_{t+h} \pm t_{\frac{\alpha}{2}, (n-k)} \hat{\sigma}_u \sqrt{R(X'X)^{-1}R' + 1}.$$

$$\hat{M}_{1971} \pm t_{0,025; (16-3)} \hat{\sigma}_u \sqrt{R(X'X)^{-1}R' + 1}$$

Comme :

$$t_{0,025; (16-3)} = 2,160 ;$$

$$\hat{\sigma}_u = 0,012209 ;$$

$$R_{1971} = [1 \quad 13,09 \quad 9,50]$$

$$R(X'X)^{-1}R' = [1 \quad 13,09124 \quad 9,498372] \begin{bmatrix} 1037,7331 & 141,69943475 & -304,650632 \\ 141,699434 & 26,2374792 & -51,9493389 \\ -304,650632 & -51,9493389 & 102,128953 \end{bmatrix} \begin{bmatrix} 1 \\ 13,09124 \\ 9,498372 \end{bmatrix}$$

$$= 0,34011116,$$

alors, l'intervalle de prévision à 95 % pour M_{1971} est donné par :

$$\begin{aligned} & 12,6908384 \pm (2,160) * (0,012209) * (0,34011116) \\ & = 12,6908384 \pm (2,16) * (0,00415242) \\ & = 12,6908384 \pm 0,00896922 \\ & \Rightarrow [12,6818691 ; 12,6998076] \end{aligned}$$

- Pour l'année 1972 :

$$\begin{aligned} & \hat{M}_{1972} \pm t_{0,025 ; (16-3)} \hat{\sigma}_u \sqrt{[R(X'X)^{-1}R' + 1]} \\ & R(X'X)^{-1}R' = [1 \quad 13,16265 \quad 9,527921] \begin{bmatrix} 1037,7331 & 141,69943475 & -304,650632 \\ 141,699434 & 26,2374792 & -51,9493389 \\ -304,650632 & -51,9493389 & 102,128953 \end{bmatrix} \begin{bmatrix} 1 \\ 13,16265 \\ 9,527921 \end{bmatrix} \\ & = 0,43236024 \end{aligned}$$

D'où :

$$\begin{aligned} & 12,7646463 \pm (2,16) * (0,00527869) = 12,7646463 \pm 0,01140196 \\ & \Rightarrow [12,7532444 ; 12,7760483] \end{aligned}$$

- Pour l'année 1973 :

$$\begin{aligned} & \hat{M}_{1973} \pm t_{0,025 ; (16-3)} \hat{\sigma}_u \sqrt{[R(X'X)^{-1}R' + 1]} \\ & R(X'X)^{-1}R' = [1 \quad 13,23842 \quad 9,675583] \begin{bmatrix} 1037,7331 & 141,69943475 & -304,650632 \\ 141,699434 & 26,2374792 & -51,9493389 \\ -304,650632 & -51,9493389 & 102,128953 \end{bmatrix} \begin{bmatrix} 1 \\ 13,23842 \\ 9,675583 \end{bmatrix} \\ & = 1,26375373 \end{aligned}$$

D'où :

$$\begin{aligned} & 12,9105408 \pm (2,16) * (0,01542917) = 12,9105408 \pm 0,03332701 \\ & \Rightarrow [12,8772138 ; 12,9438678] \end{aligned}$$

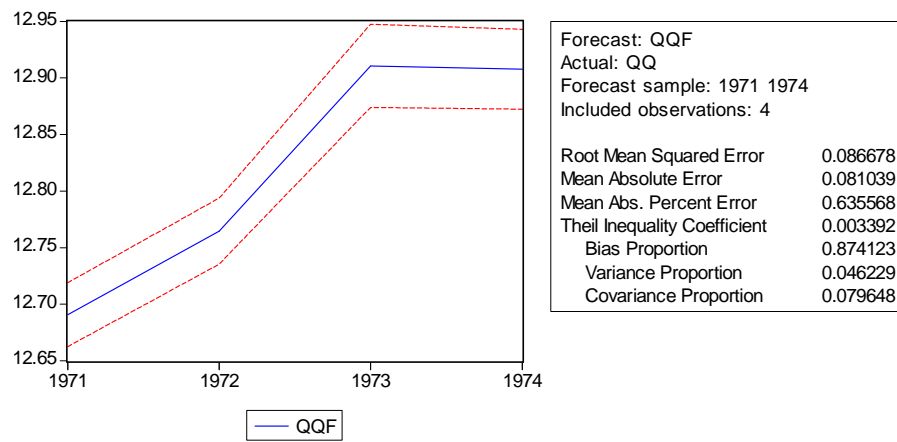
- Pour l'année 1974 :

$$\begin{aligned} & \hat{M}_{1974} \pm t_{0,025 ; (16-3)} \hat{\sigma}_u \sqrt{[R(X'X)^{-1}R' + 1]} \\ & R(X'X)^{-1}R' = [1 \quad 13,32093 \quad 9,557753] \begin{bmatrix} 1037,7331 & 141,69943475 & -304,650632 \\ 141,699434 & 26,2374792 & -51,9493389 \\ -304,650632 & -51,9493389 & 102,128953 \end{bmatrix} \begin{bmatrix} 1 \\ 13,32093 \\ 9,557753 \end{bmatrix} \\ & = 1,07860217 \end{aligned}$$

D'où :

$$\begin{aligned} & 12,9075196 \pm (2,16) * (0,01316865) = 12,9075196 \pm 0,02844429 \\ & \Rightarrow [12,8790754 ; 12,9359639]. \end{aligned}$$

La qualité prédictive du modèle peut être évaluée à partir des éléments ci-dessous fournis par le logiciel d'économétrie Eviews 5.1. Le graphique retrace l'évolution des valeurs ponctuelles prévues pour l'horizon considéré ainsi que les intervalles de prédiction associés. Cette évaluation graphique est complétée par une série de mesures conçues pour apprécier les prévisions a posteriori. Au vu des résultats ci-dessous, on constate que la valeur de la statistique de Theil ($U=0,003392$) est largement inférieure à 1 indiquant un pouvoir de prévision élevé du modèle estimé.



CHAPITRE 5 MODELES DE REGRESSION NON LINEAIRES

1.1. Introduction

On entend par modèle linéaire un modèle dans lequel la variable endogène Y , ou une quelconque transformation de Y , s'exprime en fonction des paramètres à estimer par une relation du premier degré.

Ainsi :

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + u_t \quad (5.1)$$

Est un modèle linéaire.

Par contre, les modèles ci – après :

$$Y_t = \beta_0 X_{1t}^{\beta_1} X_{2t}^{\beta_2} \dots X_{kt}^{\beta_k} \quad (5.2)$$

$$Y_t = \left\{ \beta_0 + \beta_1 \frac{1}{X_t} \right\} \quad (5.3)$$

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_t^2 + \dots + \beta_k X_{kt} \quad (5.4)$$

sont des modèles dont la spécification n'est pas linéaire mais qui sont linéaires par rapport à leurs paramètres. Ces modèles sont aussi considérés comme linéaires.

Enfin, les modèle du type :

$$Y_t = \beta_0 + \frac{1}{X_t - \beta_2} \quad (5.5)$$

$$Y_t = \alpha_1 e^{\beta_1 X_t} + \alpha_2 e^{\beta_2 X_{2t}} \quad (5.6)$$

Dans lesquels les paramètres à estimer sont $\alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2$ ne sont pas des modèles linéaires.

Si la méthode des moindres carrés ordinaires appliquée aux modèles linéaires par rapport à leurs paramètres conduisent à des estimateurs linéaires centrés à variance minimale (BLUE), il n'en est pas de même pour les modèles non linéarisables. Pour ces derniers, il faut nécessairement recourir aux méthodes d'estimation des paramètres non linéaires disponibles dans certains logiciels d'économétrie notamment RATS et EvIEWS.

1.2. Linéarisation des modèles non linéaires

Examinons le cas des modèles dont la spécification n'est pas linéaire, mais qui sont linéaires par rapport aux paramètres.

Le modèle log – linéaire

Les formes analogiques suivantes sont très intéressantes en Economie, plus particulièrement en raison de l'élasticité constante qu'on y dégage :

$$Y_t = \beta_0 X^{\beta_1} e^{u_t} \quad (5.7)$$

En appliquant une transformation logarithmique sur les deux membres de la relation (1), il vient :

$$\ln Y_t = \ln \beta_0 + \beta_1 \ln X_t + u_t \quad (5.8)$$

Ce modèle est linéaire dans les paramètres β_0 et β_1 , dans les logarithmes des variables X et Y , et peut être estimé par une régression des moindres carrés ordinaires.

A cause de cette linéarité, de tels modèles sont appelés des modèles log – log, double – log ou log – linéaire.

La relation (2) peut s'écrire simplement comme ci – dessous :

$$Y_t^* = \beta_0^* + \beta_1 X_t^* + u_t \quad (5.9)$$

où $Y_t^* = \ln Y_t$

$X_t^* = \ln X_t$

$\beta_0^* = \ln \beta_0$

Il convient de souligner ici que les estimateurs des moindres carrés ordinaires $\hat{\beta}_0$ et $\hat{\beta}_1$ obtenus sont les meilleurs estimateurs linéaires sans biais (BLUE) de β_0 et β_1 .

L'un des traits intéressants du modèle log – log est que le coefficient de la pente (β_1) mesure l'élasticité de Y par rapport à X . en effet :

$$\epsilon_{y/x} = \frac{\frac{dY}{Y}}{\frac{dX}{X}} = \frac{d \ln Y}{d \ln X} = \beta_1 \quad (5.10)$$

Cette élasticité est indépendante de la variation de X . tandis que l'élasticité de Y par rapport à X obtenue à partir de la fonction linéaire du type $Y_t = \beta_0 + \beta_1 X_t + u_t$ est égal à :

$$\epsilon_{y/x} = \beta_1 \frac{X_t}{Y_t} \quad (5.11)$$

Dépendante des variables X et Y .

Signalons en outre que le modèle log – linéaire suppose que le coefficient d'élasticité entre Y et X demeure constant ; d'où l'application alternative de modèle à élasticité constante.

L'expression analytique généralisée du modèle log – linéaire est donnée par :

$$Y_t = \beta_0 X_{1t}^{\beta_1} X_{2t}^{\beta_2} \dots X_{kt}^{\beta_k} e^{u_t} \quad (5.12)$$

Sous sa forme logarithmique, cette relation devient :

$$\ln Y_t = \ln \beta_0 + \beta_1 \ln X_{1t} + \beta_2 \ln X_{2t} + \dots + \beta_k \ln X_{kt} + u_t \quad (5.13)$$

Et les paramètres associés aux variables explicatives représentent les élasticités de Y par rapport à X_j ($j = 1, 2, \dots, k$) respectivement. Donc :

$$\epsilon_{y/x_j} = \frac{\delta \ln Y_t}{\delta \ln X_{jt}} = \beta_j \quad (j = 1, 2, \dots, k) \quad (5.14)$$

Les modèles semi – logarithmiques : les modèles log –ln et lin – log

1. Le modèle log – lin

Pour connaître le taux de croissance de certaines variables telles que la population, le PIB, l'offre de monnaie, l'emploi, ... on peut utiliser le modèle de régression suivant :

$$Y_t = Y_0(1 + r)^t \quad (5.15)$$

Notons que la relation (5.15) n'est autre que la formule des intérêts composés (où r est le taux de croissance de Y) ;
En utilisant le logarithme naturel de (5.15), on obtient :

$$\ln Y_t = \ln Y_0 + t \ln(1 + r) \quad (5.16)$$

Cette relation peut s'écrire aussi comme :

$$\begin{aligned} \ln Y_t &= \beta_0 + \beta_1 t \\ \text{où } \beta_0 &= \ln Y_0 ; \\ \beta_1 &= \ln(1 + r), \end{aligned} \quad (5.17)$$

ou simplement

$$Y_t^* = \beta_0 + \beta_1 t \quad (5.18)$$

En ajoutant le terme d'erreur, on obtient :

$$Y_t^* = \beta_0 + \beta_1 t + u_t \quad (5.19)$$

Ce modèle est identique à n'importe quel autre modèle linéaire en ce que les paramètres β_0 et β_1 sont linéaires. Les modèles du type (5.17) sont appelés des modèles semi – log parce qu'une seule variable (la variable dépendante) apparaît sous forme logarithmique. Un modèle dans lequel la variable dépendante est logarithmique est connu comme modèle log – lin.

Remarquons que le coefficient de la pente mesure la variation constante proportionnelle ou relative de Y , consécutive à un changement de la valeur absolue du régresseur X (ici, la variable t), c'est – à – dire :

$$\begin{aligned} \beta_1 &= \frac{\text{Variation relative de la variable dépendante}}{\text{Variation absolue du régresseur}} \\ &= \frac{d \ln Y_t}{d t} \end{aligned}$$

Le coefficient β_1 est appelé la semi – élasticité de Y par rapport à t .

Les modèles du trend linéaire

Au lieu d'estimer le modèle log – lin, on peut utiliser le modèle suivant :

$$Y_t^* = \beta_0 + \beta_1 t + u_t \quad (5.20)$$

Un modèle est appelé modèle de trend linéaire et la variable temps (t) est

nommée le trend. Si le coefficient β_1 de la relation (5.20) est positif, le trend de Y est croissant ; un coefficient négatif indique un trend de Y décroissant.

Exemple

Considérons les données sur la consommation des services. Les résultats de la régression sont les suivants :

$$\begin{aligned} \text{a) } \ln \widehat{DS}_t &= 7.7890 + 0.00743 t \\ &\quad (3387.62) \quad (44.28) \\ R^2 &= 0.9894 \end{aligned}$$

Etant donné qu'il s'agit des données trimestrielles, la dépense de service a augmenté trimestriellement à un taux de 0.743%. Ceci correspond à une croissance annuelle de 2.94% (c'est-à-dire 0.743×4). Puisque $7.7890 = \ln DS$ au début de la période étudiée, alors son anti log permet d'obtenir 2413.9 millions de FC comme leur initiale de la dépense c'est-à-dire au temps 0.

$$\begin{aligned} \text{b) } \ln \widehat{DS}_t &= 2405 + 19.6920 t \\ &\quad (322.99) \quad (36.25) \\ R^2 &= 0.9843 \end{aligned}$$

Sur la période sous étude, la dépense en services a augmenté, en moyenne, au taux absolu (et non négatif) d'environ 20 millions de FC par trimestre. Il y a donc un trend croissant de la consommation de services.

2. Le modèle lin – log

A la différence du modèle de croissance log – lin dans lequel on cherche le pourcentage de croissance de Y correspondant à une variation absolue de X, nous allons nous intéresser à la variation absolue de Y consécutive à un changement de X en pourcentage. D'où le modèle suivant :

$$Y_t = \beta_0 + \beta_1 \ln X_t + u_t \quad (5.21)$$

Un tel modèle est appelé modèle lin – log.

Le coefficient β_1 peut être interprété comme suit :

$$\begin{aligned} \beta_1 &= \frac{\text{Variation de Y}}{\text{Variation de } \ln X} \\ &= \frac{\text{Variation de Y}}{\text{Variation relative de X}} \end{aligned} \quad (5.22)$$

L'équation (9) peut s'écrire d'une autre manière :

$$\Delta Y = \beta_1 \left(\frac{\Delta X}{X} \right) \quad (5.23)$$

Cette équation signifie que la variation absolue de Y est égale à la pente multipliée par la variation relative de X.

Une application intéressante est celle des modèles appelés modèles d'Engel, « la dépense totale consacrée à la nourriture trend à croître selon une

progression arithmétique lorsque la dépense totale augmente en progression géométrique ».

Exemple

La liaison entre les dépenses d'alimentation et la total des dépenses peut être régressée par le modèle lin – log, car les dépenses d'alimentation progressent plus lentement que le total des dépenses. Les résultats d'un modèle lin – log sont les suivants :

$$\widehat{\text{DépAlim}}_t = -11283.91 + 257.27 \ln \text{DépTot}_t$$

(4.38) (5.66)

$$R^2 = 0.3769$$

Le coefficient de la pente ($\hat{\beta}_1$) signifie qu'une croissance de 1% de la dépense alimentaire totale se traduit en moyenne par une hausse de 2.57 des dépenses de nourriture (on a divisé) le coefficient estimé de la pente par 100).

Les modèles réciproques

Les modèles réciproques s'avèrent utiles dans l'analyse économique. Ces modèles se caractérisent par l'existence d'une asymptote pour au moins l'une des deux variables, et se présentent généralement comme ci – après :

$$(Y_t - \alpha_1)(X_t - \alpha_2) = \alpha_3 \quad (5.24)$$

Cette relation peut être représentée par les graphiques suivants :

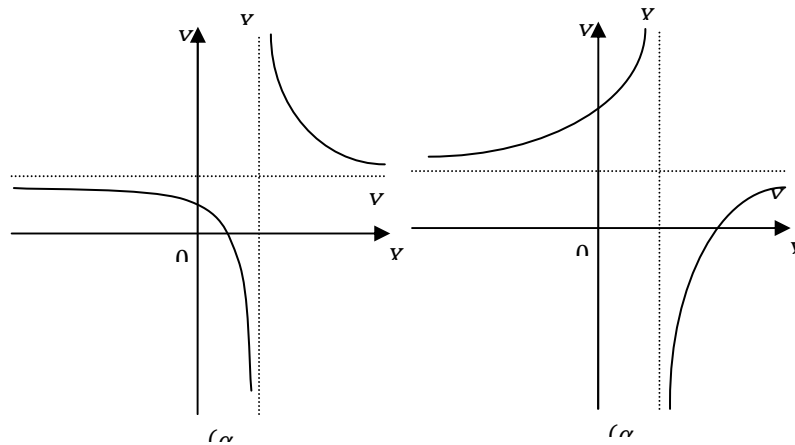


fig : Hyperboles rectangulaires

La relation (11) peut s'écrire aussi sous la forme :

$$Y_t = \alpha_1 + \frac{\alpha_3}{X_t - \alpha_2} \quad (5.25)$$

En ajoutant le terme d'erreur, le modèle (5.25) devient :

$$Y_t = \alpha_1 + \frac{\alpha_3}{X_t - \alpha_2} + u_t \quad (5.26)$$

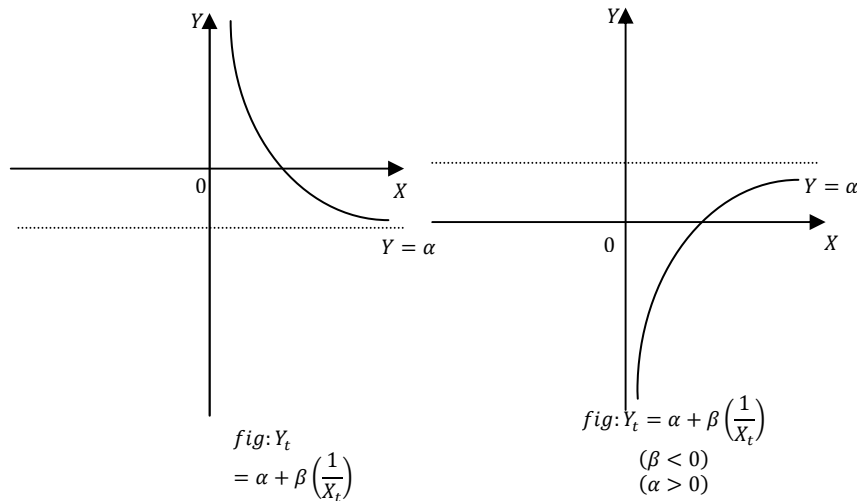
Le modèle (5.26) est non linéaire dans les paramètres (α_2, α_3). Dans ce cas, il n'existe pas de transformation qui permette de linéariser le modèle, et donc d'utiliser les méthodes des moindres carrés ordinaires pour estimer les paramètres du modèle. Cependant, il y a deux cas particuliers où le modèle (5.26) peut être linéarisé.

a) Si $\alpha_2 = 0$, le modèle (5.26) devient :

$$Y_t = \alpha + \beta \left(\frac{1}{X_t} \right) + u_t \quad (5.27)$$

où $\alpha = \alpha_1$ et $\beta = \alpha_3$

Graphiquement, la relation (5.27) se présente comme suit :



Bien que le modèle (5.27) ne soit pas linéaire par rapport à la variable X car elle se présente de manière inverse (ou réciproque), il est linéaire par rapport à α et β . C'est donc un modèle de régression linéaire.

b) Si $\alpha_1 = 0$, le modèle (5.25) prend la forme suivante :

$$\left(\frac{1}{Y_t} \right) = \alpha + \beta X_t \quad (5.28)$$

Avec $\alpha = \frac{-\alpha_2}{\alpha_3}$ et $\beta = \frac{1}{\alpha_3}$

Graphiquement, nous avons :

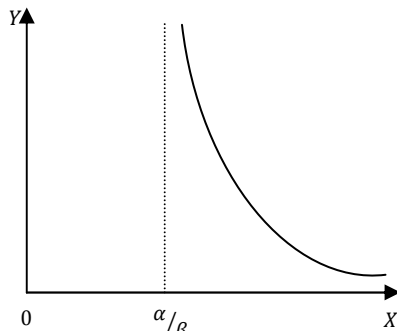


fig : modèle $\frac{1}{Y_t}$
 $= \alpha + \beta X_t$

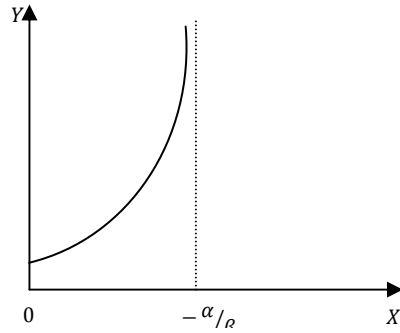


fig : modèle $\frac{1}{Y_t}$
 $= \alpha + \beta X_t$

Pour linéariser les deux modèles non linéaires en variable (5.27) et (5.28), il suffit seulement de poser :

$$X_t^* = \frac{1}{X_t} \quad \text{et} \quad Y_t^* = \frac{1}{Y_t}$$

D'où :

$$Y_t = \alpha + \beta X_t^* + u_t \quad (5.29)$$

et

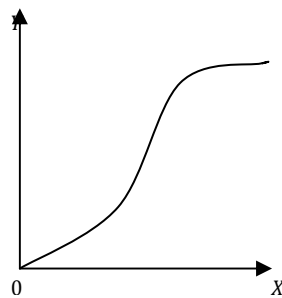
$$Y_t^* = \alpha + \beta X_t + u_t \quad (5.30)$$

Le modèle log – hyperbole ou modèle logarithmique réciproque

La forme générale des modèles log – hyperbole est :

$$\ln Y_t = \beta_0 - \beta_1 \left(\frac{1}{X_t} \right) + u_t \quad (5.31)$$

Sa représentation graphique est donc celle – ci :



Initialement, Y croît à un taux croissant (courbe est convexe) ; par la suite, Y augmente à un taux décroissant (la courbe devient concave). Un tel modèle est apte à représenter une fonction de production de court terme. En effet, une hausse du facteur variable (le travail) se traduit par une croissance de la production ayant la forme de la figure ci – haut.

Le modèle (5.31) peut s'écrire aussi comme suit :

$$Y_t^* = \beta_0 + \beta_1^* X_t^* + u_t \quad (5.32)$$

$$\begin{aligned} \text{où : } Y_t^* &= \ln Y_t ; \\ X_t^* &= \ln \frac{1}{X_t} ; \\ \beta_1^* &= -\beta_1 \end{aligned}$$

Le modèle polynomial

Une fonction polynomiale de degré k est de la forme générale :

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_t^2 + \dots + \beta_k X_t^k \quad (5.33)$$

Une telle fonction est linéaire dans les paramètres.

Remarquons que dans les modèles de régression polynomiaux, il n'y a qu'une seule droite variable explicative dans le membre de droite, mais elle figure avec des puissances différentes.

Pour $k = 3$, sa représentation graphique est donc :

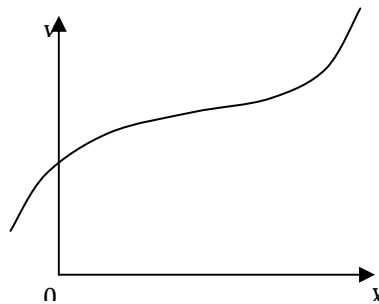


fig : modèle $Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_t^2 + \beta_3 X_t^3$

La transformation linéaire des fonctions polynomiales consiste à réécrire le modèle (5.33) comme ci – après :

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + u_t \quad (5.34)$$

où : $X_{1t} = X_t$; $X_{2t} = X_t^2$; ... ; $X_{kt} = X_t^k$

Quelques applications courantes de ce type de modèle sont :

- L'estimation d'une tendance (trend) pour une série chronologique accusant, par exemple, deux points de retournement :

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + u_t \quad (5.35)$$

où t représente le temps

- L'estimation d'une fonction de coût total :

$$CT_t = \beta_0 + \beta_1 Q_t + \beta_2 Q_t^2 + u_t \quad (5.36)$$

où : CT_t = coût total à la période t ;
 Q_t = quantité produite au temps t ;
 u_t = le terme d'erreur à la période t

Remarquons que dans les modèles de régression polynomiaux, il n'y a qu'une seule variable explicative dans le membre de droite, mais figure avec des puissances différentes.

Applications

1. Les modèles de courbe de vie du produit

Retenons la première génération de ces modèles. Ils tentent de déterminer l'évolution probable des ventes connaissant le seuil de saturation. Il s'agit du modèle de GOMPERTZ et du modèle logistique.

A. Le modèle de GOMPERTZ

Ce modèle est défini par la formulation suivante :

$$Y_t = e^{br^t + a} \quad (5.37)$$

ou bien, sous sa forme logarithmique :

$$\ln Y_t = br^t + a \quad (5.38)$$

$$(b < 0 \text{ et } 0 < r < 1)$$

Le coefficient r définit la vitesse de diffusion ;

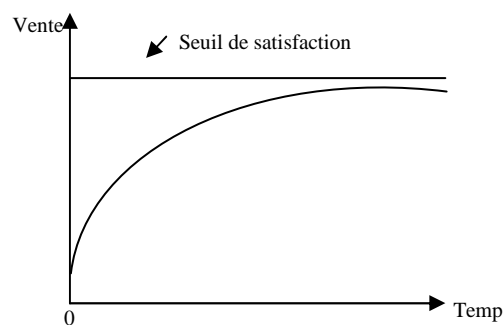


Fig : courbe de GOMPERTZ

Il est possible d'estimer ce modèle lorsqu'on connaît ou lorsqu'on postule la valeur de saturation

Linéarisation du modèle avec seuil de saturation estimé :

$$\ln Y_t = a + br^t$$

$$\begin{aligned}
\ln Y_t - a &= br^t \\
\ln(\ln Y_t - a) &= \ln b + t \ln r \\
Y_t^* &= a_0 + a_1 t \\
\text{où : } Y_t^* &= \ln(\ln Y_t - a) \\
a_0 &= \ln b \\
a_1 &= \ln r
\end{aligned} \tag{5.39}$$

B. Le modèle logistique

Il s'exprime comme suit :

$$Y_t = \frac{Y_{\max}}{1 + br^t} \quad (0 < r < 1) \tag{5.40}$$

où : Y_{\max} représente le seuil de saturation
 r , la vitesse de diffusion

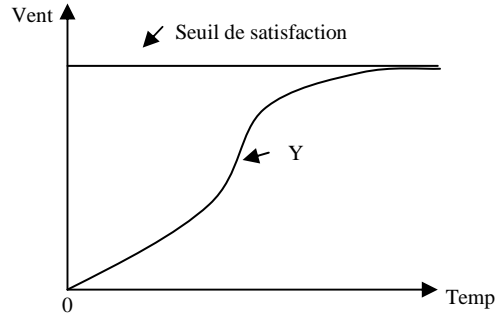


Fig : courbe logistique

La linéarisation du modèle avec seuil de saturation connu est donc :

$$\frac{Y_{\max}}{Y_t} = 1 + br^t \tag{5.41}$$

$$\left(\frac{Y_{\max}}{Y_t} \right) - 1 = br^t$$

$$\ln \left[\frac{Y_{\max}}{Y_t} - 1 \right] = \ln b + t \ln r$$

$$\boxed{Y_t^* = a_0 + a_1 t} \tag{5.42}$$

où :

$$Y_t^* = \ln \left[\frac{Y_{\max}}{Y_t} - 1 \right]$$

$$a_0 = \ln b$$

$$a_1 = \ln r$$

2. Calculs de tendance et de taux de croissance

Beaucoup des variables économiques augmentent ou diminuent avec le temps. Si la tendance observée est linéaire, elle peut être modélisée de la façon suivante :

$$Y_t = \alpha + \beta T + u_t \quad (5.43)$$

où T indique le temps

La variable T peut être spécifiée indifféremment sous les différentes formes. Par exemple, si on dispose d'observations annuelles d'une variable pour les années allant de 1970 à 2008, on peut prendre T sous les formes suivantes :

T = 1970, 1971, 1972, ..., 2007, 2008

T = 1, 2, 3, ..., 39

T = -19, -18, -17, ..., 18, 19

Dans chaque cas, il faut définir l'origine des temps et l'unité de mesure utilisée. Dans les trois cas, l'unité de temps est l'année. Le troisième cas est intéressant pour les calculs à petite échelle puisque dans ce cas, la moyenne de T est nulle, de sorte que les coefficients des moindres carrés relatifs au modèle (5.43) se simplifient pour devenir :

$$\hat{\alpha} = \bar{Y} \text{ et } \hat{\beta} = \frac{\sum TY}{\sum T^2} \quad (5.44)$$

Beaucoup des logiciels fournissent une tendance (trend), à utiliser dans les régressions, correspondant à la deuxième spécification pour T ci – dessus.

- Courbes à taux de croissance constant

En prenant les différences premières de la relation (5.43), on obtient :

$$\Delta Y_t = \beta + (u_t - u_{t-1}) \quad (5.45)$$

Si on néglige les perturbations, il découle de (5.43) que la série augmente (ou diminue) d'un montant constant à chaque période. Ainsi, pour des séries croissantes ($\beta > 0$), ceci implique un taux de croissance qui diminue, alors que les séries décroissantes ($\beta < 0$), le taux de décroissance de Y augmente au cours du temps.

Donc, pour les suites avec un taux de croissance constant, aussi bien positif que négatif, la relation (5.43) est une modélisation inappropriée. La spécification appropriée doit exprimer le logarithme de la suite comme une fonction linéaire du temps. D'où, une série croissant à un taux constant g est telle que :

$$Y_t = Y_0(1 + g)^t \quad (5.46)$$

Si on considère les transformations logarithmiques, la relation (5.46) devient :

$$\ln Y_t = \ln Y_0 + t \ln(1 + g) \quad (5.47)$$

ou encore :

$$\ln Y_t = \alpha + \beta t \quad (5.48)$$

avec $\alpha = \ln Y_0$ et $\beta = (1 + g)$

En effet, si on pense qu'une suite croît à un taux constant, on exprime son logarithme en fonction du temps. Si le nuage des points obtenu graphiquement semble être linéaire, on estime alors la relation (5.48) par la méthode des moindres carrés, en faisant une régression de Y en fonction du temps. Le coefficient associé au temps fournit une estimation \hat{g} du taux de croissance, car :

$$\hat{\beta} = \ln(1 + \hat{g}) \quad (5.49)$$

ou bien :

$$\hat{g} = e^{\hat{\beta}} - 1 \quad (5.50)$$

Le coefficient β de la relation (5.49) représente le taux de croissance continu $\left(\beta = \frac{d \ln Y_t}{dt}\right)$, alors que g représente le taux de croissance discret.

Une suite qui croît continûment à un taux de croissance est telle que :

$$Y_t = Y_0 e^{\beta t} \quad (5.51)$$

ou encore :

$$\ln Y_t = \alpha + \beta t \quad (5.52)$$

avec $\alpha = \ln Y_0$

En considérant les différences premières de la relation (5.49), on a :

$$\Delta \ln Y_t = \beta = \ln(1 + g) \approx g \quad (5.53)$$

Donc, en prenant les différences premières des logarithmes, on obtient le taux de croissance continu qui, à son tour, est une approximation du taux de croissance discret.

Chapitre VI. Autocorrélation

6.1. LA NATURE DU PROBLEME

Pour estimer les paramètres inconnus du modèle linéaire par la méthode des moindres carrés ordinaires, les hypothèses classiques suivantes ont été énumérées :

- H_1 : le modèle est linéaire en X ;
- H_2 : les valeurs X sont observées sans erreur ;
- H_3 : $E(\mu_t) = 0$;
- H_4 : $E(\mu_t^2) = \sigma_\mu^2$;
- H_5 : $E(\mu_t \mu_{t-s}) = 0, \forall s > 0$;
- H_6 : $\text{Cov}(X_t, \mu_t) = 0$.

Dans un modèle linéaire général, la spécification de la matrice des variances – covariances de l'erreur est :

$$\Omega_U = E[UU'] = \begin{bmatrix} \sigma_\mu^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_\mu^2 & & \dots & 0 \\ 0 & 0 & \sigma_\mu^2 & \dots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \dots & \sigma_\mu^2 \end{bmatrix}$$

Les erreurs ne sont plus indépendantes (violation de H_5), lorsque l'hypothèse H_5 n'est plus vérifiée, on dit qu'il y a présence d'autocorrélation des erreurs. En d'autres termes, la perturbation se produisant pour l'observation X_t est liée à la perturbation correspondant à l'observation X_{t-s} . Donc, l'existence de l'autocorrélation implique que l'effet total d'une perturbation n'est pas instantané mais est aussi senti dans les périodes futures.

L'autocorrélation des erreurs peut provenir d'un des facteurs suivants :

- Phénomènes d'inertie ;
- Omission des variables explicatives importantes ;
- Lissage, interpolation ou extrapolation des données ;
- Mauvaise spécification (de la forme) du modèle : forme fonctionnelle incorrecte.

L'autocorrélation est particulièrement fréquente lorsqu'on travaille avec des

séries chronologiques. En effet, même des événements aléatoires ont souvent des incidences qui se prolongent dans le temps. Ainsi une grève des ouvriers ou une panne de machine aura une influence sur la production de plusieurs mois. Donc, il est parfois peu naturel d'admettre l'indépendance de l'erreur d'une période à l'autre.

6.2. PROPRIETES DES ESTIMATEURS DES MOINDRES CARRES ORDINAIRES EN PRESENCE D'AUTOCORRELATION

Soit le modèle linéaire général :

$$Y = X\beta + U$$

où X est une matrice non aléatoire ;

$$E(U) = 0$$

$$E(UU') = \Omega_U = \sigma_u^2 \Psi + \sigma_u^2 I$$

1. Les estimateurs des moindres carrés ordinaires restent non biaisés.

$$\text{En effet : } \hat{\beta} = \beta + (X'X)^{-1}X'U$$

et

$$E(\hat{\beta}) = \beta \quad 6.1$$

2. Les estimateurs des ainsi obtenus ne sont plus à variance minimale. La matrice des variances – covariances des estimateurs devient :

$$\begin{aligned} \Omega_{\hat{\beta}} &= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] \\ &= (X'X)^{-1}X'E(UU')X(X'X)^{-1} \\ &= (X'X)^{-1}X'\Omega_U X(X'X)^{-1} \end{aligned} \quad 6.2$$

expression différente de $\sigma_u^2(X'X)^{-1}$ précédemment obtenue.

Conséquences

En présence d'autocorrélation des erreurs, les estimateurs des moindres carrés ordinaires sont encore linéaires et non biaisés, mais ils ne sont plus performants, c'est-à-dire ils n'ont plus une variance minimale. Par conséquent, la matrice de variance – covariance n'est plus estimée de manière convergente et les tests de significativité (tests de t et F) ne sont plus opérants. Ils risquent d'aboutir à des conclusions erronées sur la signification des coefficients de régression estimés.

6.3. SCHEMA AUTOREGRESSIF DU PREMIER ORDRE DES ERREURS

L'autocorrélation des erreurs peut prendre différentes formes. Dans le cadre de ce cours, nous allons nous limiter au schéma autorégressif du premier ordre des résidus.

Soit le modèle linéaire général ci-après :

$$Y = X\beta + U \quad 6.3$$

ou encore :

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + u_t \quad 6.4$$

avec :

$$u_t = \rho u_{t-1} + v_t \quad 6.5$$

Le terme aléatoire u_t est un processus autorégressif d'ordre 1, noté généralement AR(1), où :

$$\begin{cases} E(v_t) = 0 \\ E(v_t^2) = \sigma_v^2 \\ E(v_t v_{t-s}) = 0, \forall s \neq 0 \\ |\rho| < 1 : \text{condition de stabilité} \end{cases} \quad 6.6$$

En procédant par substitution successive de (6.5), on obtient :

$$\begin{aligned} u_t &= \rho u_{t-1} + v_t = \rho(\rho u_{t-2} + v_{t-1}) + v_t \\ &= \rho^2 u_{t-2} + \rho v_{t-1} + v_t \\ &= \rho^2(\rho u_{t-3} + v_{t-2}) + \rho v_{t-1} + v_t \\ &= \rho^3 u_{t-3} + \rho^2 v_{t-2} + \rho v_{t-1} + v_t \\ &= \rho^i u_{t-i} + v_t + \rho v_{t-1} + \rho^2 v_{t-2} + \dots + \rho^{i-1} v_{t-i+1} \end{aligned} \quad 6.7$$

En supposant que $|\rho| < 1$ et $i \rightarrow \infty$, la relation (5.7) devient :

$$u_t = u_t + \rho v_{t-1} + \rho^2 v_{t-2} + \dots + \sum_{s=0}^{\infty} \rho^s v_{t-s} \quad 6.8$$

De la relation (5.8), on déduit les propriétés de u_t suivantes :

$$\begin{aligned} 1. \quad E(u_t) &= \sum_{s=0}^{\infty} (\rho^s v_{t-s}) \\ &= \sum_{s=0}^{\infty} \rho^s E(v_{t-s}) = 0, \quad \forall t \\ \text{car } E(v_{t-s}) &= 0. \end{aligned} \quad 6.9$$

2. La matrice des variances – covariances

- En élevant au carré l'expression (6.8) et en prenant son espérance mathématique, on obtient :

$$\begin{aligned} \text{Var}(u_t) &= E(u_t^2) \\ &= E(v_t + \rho v_{t-1} + \rho^2 v_{t-2} + \dots)^2 \\ &= E[v_t^2 + \rho^2 v_{t-1}^2 + \rho^4 v_{t-2}^2 + \dots] \\ &= E[v_t^2] + \rho^2 E[v_{t-1}^2] + \rho^4 E[v_{t-2}^2] + \dots \\ &= \sigma_v^2 (1 + \rho^2 + \rho^4 + \dots) \\ &= \frac{\sigma_v^2}{1 + \rho^2} = \sigma_u^2 \end{aligned} \quad 6.10$$

- En utilisant les propriétés de v_t , on démontre aussi que :

$$\begin{aligned} \text{Cov}(u_t, u_{t-1}) &= E(u_t u_{t-1}) \\ &= E[(\rho u_{t-1} + v_t) u_{t-1}] \\ &= \rho \sigma_u^2 \end{aligned}$$

On arrive également à ce résultat en procédant comme ci-après :

$$\text{Cov}(u_t, u_{t-1}) = E(u_t u_{t-1})$$

$$\begin{aligned}
&= E[(v_t + \rho v_{t-1} + \rho^2 v_{t-2} + \dots)(v_{t-1} + \rho v_{t-2} + \rho^2 v_{t-3} + \dots)] \\
&= E[\rho v_{t-1}^2 + \rho^3 v_{t-2}^2 + \rho^4 v_{t-3}^2 \dots] \\
&= \sigma_v^2 \rho (1 + \rho^2 + \rho^4 + \dots) = \frac{\rho \sigma_v^2}{1 - \rho^2} = \rho \sigma_u^2
\end{aligned}$$

$$\begin{aligned}
\text{Cov}(u_t, u_{t-2}) &= E(u_t u_{t-2}) \\
&= E[(\rho u_{t-1} + v_t) u_{t-2}] \\
&= E[(\rho^2 u_{t-2} + \rho v_{t-1} + v_t) u_{t-1}] \\
&= \rho^2 \sigma_u^2
\end{aligned}$$

et en général :

$$\begin{aligned}
\text{Cov}(u_t, u_{t-s}) &= E(u_t u_{t-s}) = \rho^s \sigma_u^2 \\
&= \frac{\rho^s \sigma_v^2}{1 - \rho^2}
\end{aligned} \tag{6.11}$$

Faisons remarquer que le coefficient de corrélation entre u_t et u_{t-s} (c'est-à-dire le coefficient de corrélation de décalage s) est donné par le rapport de la covariance et écarts types ainsi :

$$\begin{aligned}
\frac{E[u_t u_{t-1}]}{\{E[u_t^2]E[u_{t-1}^2]\}^{1/2}} &= \frac{\rho^s \sigma_u^2}{\sigma_u^2} \\
&= \frac{\rho^s \sigma_v^2 / (1 - \rho^2)}{\sigma_v^2 / (1 - \rho^2)} = \rho^s
\end{aligned} \tag{6.12}$$

Si $s = 1$, on obtient :

$$\rho = \frac{E[u_t u_{t-1}]}{\sqrt{E[u_t^2]E[u_{t-1}^2]}}$$

Notons également que plus grand sera le décalage entre les erreurs, et plus petit sera leur coefficient de corrélation.

De ce qui précède, la matrice des variances - covariances de l'erreur est :

$$\text{donc : } \Omega_U = E[UU'] = E \begin{bmatrix} \mu_1^2 & \mu_1 \mu_2 & \mu_1 \mu_3 & \dots & \mu_1 \mu_n \\ \mu_2 \mu_1 & \mu_2^2 & \mu_2 \mu_3 & \dots & \mu_2 \mu_n \\ \mu_3 \mu_1 & \mu_3 \mu_2 & \mu_3^2 & \dots & \mu_3 \mu_n \\ . & . & . & & . \\ . & . & . & & . \\ . & . & . & & . \\ \mu_n \mu_1 & \mu_n \mu_2 & \mu_n \mu_3 & \dots & \mu_n^2 \end{bmatrix}$$

$$= \frac{\sigma_v^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix} \quad 6.13$$

La matrice des variances et covariances peut s'écrire aussi comme :

$$\Omega_U = \sigma_v^2 \Psi \quad 6.14$$

où

$$\Psi = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & \rho & \dots & \rho^{n-1} \\ \rho & 1 & \dots & \rho^{n-2} \\ \rho^2 & \rho & \dots & \rho^{n-3} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \rho^{n-1} & \rho^{n-2} & \dots & 1 \end{bmatrix}$$

6.4. ESTIMATEURS DES MOINDRES CARRES GENERALISES

Lorsque la matrice de covariance des résidus ne répond plus aux hypothèses classiques, il faut alors estimer les paramètres du modèle par la méthode de Gauss-Aitken ou la méthode généralisée des moindres carrés (GLS = Generalized Least Squares).

Soit le modèle ci-après :

$$\begin{aligned} Y &= X\beta + U \\ \text{avec } E(U) &= 0 \\ E(UU') &= \sigma_u^2 \Psi \end{aligned} \quad 6.15$$

Prémultiplions les deux membres de la relation (6.15) par une matrice de transformation non singulière P de format (n x n).

$$\text{Soit : } PY = PX\beta + PU \quad 6.16$$

Cette relation peut s'écrire simplement comme :

$$\begin{aligned} Y^* &= X^* + U^* \\ \text{où : } Y^* &= PY ; \\ X^* &= PX ; \\ U^* &= PU. \end{aligned} \quad 6.17$$

Sur le terme aléatoire PU, on fait les hypothèses suivantes :

$$E(U^*) = E(PU) = PE(U) = 0 \quad 6.18$$

$$\begin{aligned} E(U^*U^*) &= E(PUU'P') \\ &= PE(UU')P' \\ &= \sigma_u^2 P\phi P' \\ &= P\Omega_u P' = \sigma_u^2 I \end{aligned} \quad 6.19$$

La relation (5.19) suppose donc que :

$$P\Psi P' = I \quad (5.20)$$

En effet, il est possible de trouver une matrice P qui vérifie la relation (5.20).

$$\text{Supposons que : } \Psi = T T' \quad (5.21)$$

Si T est une matrice non singulière, on a :

$$T^{-1} \Psi (T')^{-1} = I \quad (5.22)$$

De la relation (5.20) et (5.22), on peut déduire que :

$$P = T^{-1} \quad (5.23)$$

Il en résulte immédiatement que :

$$\begin{aligned} T^{-1} \Psi (T')^{-1} &= T^{-1} \Psi (T - 1)' \\ &= P \Psi P' = I \end{aligned} \quad (5.24)$$

$$\text{et } \Psi^{-1} = (T')^{-1} T^{-1} = (T')^{-1} T^{-1} = P'P \quad (5.25)$$

Appliquons maintenant la méthode des moindres carrés ordinaires aux variables transformées $PY = P^*$ et $PX = X^*$ du modèle. Cette méthode consiste à minimiser la somme des carrés des résidus :

$$\begin{aligned} \text{Min } S &= \text{Min } (Y^* - X^*\beta) (Y^* - X^*\beta)' \\ &= \text{Min } (Y^* - X\beta)' P'P (Y - X\beta) \\ &= \text{Min } (Y - X\beta)' \varphi^{-1} (Y - X\beta) \end{aligned} \quad (5.26)$$

En dérivant cette relation par rapport au vecteur de β et en l'égalisant à zéro, on trouve :

$$\begin{aligned} \frac{\delta S}{\delta \beta} &= -2X'P'PY + 2X'P'PX\hat{\beta} = 0 \\ &= -X'P'PY + X'P'PX\hat{\beta} = 0 \\ &= X'\Psi^{-1}Y + X'\Psi^{-1}X\hat{\beta} = 0 \end{aligned} \quad (5.27)$$

De la relation (13), on déduit :

$$\begin{aligned} \hat{\beta} &= (X^{**'}X^*)^{-1} X^{**'}Y^* \\ &= [(PX)'PX]^{-1} (PX)'PY \\ &= (X'P'PX)^{-1} X'P'PY \\ &= (X'\Psi^{-1}X)^{-1} X'\Psi^{-1}Y \end{aligned} \quad (5.28)$$

L'estimateur $\hat{\beta}$ est estimateur des moindres carrés généralisés de Gauss-Aitken. L'estimateur $\hat{\beta}$ obtenu par la méthode de GLS est non biaisé :

$$\begin{aligned} E(\hat{\beta}) &= E[(X'\Psi^{-1}X)^{-1}X'\Psi^{-1}Y] \\ &= E[(X'\Psi^{-1}X)^{-1}X'\Psi^{-1}(X\beta + U)] \\ &= (X'\Psi^{-1}X)^{-1}X'\Psi^{-1}X\beta + (X'\Psi^{-1}X)^{-1}X'\Psi^{-1}E(U) \\ &= \beta \end{aligned} \quad (5.29)$$

Donc l'estimateur $\hat{\beta}$ des moindres carrés généralisés est biaisé.

La matrice des variances – covariances de $\hat{\beta}$ est donnée par l'expression ci-après :

$$\begin{aligned} E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] &= E[(X'\Psi^{-1}X)^{-1}X'\Psi^{-1}UU'\Psi^{-1}X(X'\Psi^{-1}X)^{-1}] \\ &= (X'\Psi^{-1}X)^{-1}X'\Psi^{-1}E(UU')\Psi^{-1}X(X'\Psi^{-1}X)^{-1} \\ &= \sigma_v^2(X'\Psi^{-1}X)^{-1} \end{aligned} \quad (5.30)$$

Cette matrice diffère largement de la matrice des variances – covariances obtenue en utilisant directement la méthode des moindres carrés ordinaires sur les données non transformées (brutes). En effet :

$$\sigma_v^2 (X' \Psi^{-1} X)^{-1} \sigma_v^2 (X' X)^{-1} X' \Psi X (X' X)^{-1} \quad (5.31)$$

IV.5. Procédures d'estimation en cas d'autocorrélation des erreurs

1° Principes généraux

Le modèle linéaire général sous l'hypothèse d'une autocorrélation des erreurs d'ordre 1 s'écrit :

$$Y = X \beta + U \quad (5.32)$$

avec

$$u_t = \rho u_{t-1} + v_t$$

$$\begin{cases} |\rho| < 1 \\ E(v_t) = 0 \\ E(v_t^2) = \sigma_v^2 \\ E(v_t v_{t-s}) = 0, \forall s \neq 0 \end{cases}$$

Sous ses conditions, on sait que :

$$\Omega_U = \frac{\sigma_v^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix} = \sigma_v^2 \Psi \quad (5.33)$$

L'estimateur des moindres carrés généralisés est égal à :

$$\beta^* = (X' \Psi^{-1} X)^{-1} X' \Psi^{-1} Y \quad (5.34)$$

et sa variance est donc :

$$\Omega_{\beta} = \sigma_u^2 (X' \Psi^{-1} X)^{-1} \quad (5.35)$$

$$\text{où } \Psi^{-1} = \begin{bmatrix} 1 & -\rho & 0 & \dots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & \dots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{bmatrix}$$

Etant donné que $PP' = \Psi^{-1}$, l'estimateur des moindres carrés généralisés peut être déterminé par :

$$\begin{aligned} \beta^* &= (X' P' P X)^{-1} X' P' P Y \\ &= [(PX)' (PX)]^{-1} (PX)' (PY) \\ &= (X^{*'} X^*)^{-1} X^{*'} Y^* \end{aligned} \quad (5.36)$$

Par conséquent, il faut trouver la matrice de transformation P telle que le modèle linéaire général : $PY = PX\beta + PU$ ait ses erreurs indépendantes et homoscedastiques. La matrice P telle que $P'P = \Psi^{-1}$ est donnée par :

$$P = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{bmatrix} \quad (5.37)$$

Ainsi, on peut substituer à la méthode des moindres carrés généralisés, la méthode des moindres carrés, la méthode des moindres carrés ordinaires au modèle linéaire général $PY = PX\beta + PU$ qui n'est autre que le modèle initial où les variables sont transformées. En effet :

$$Y^* = PY = \begin{bmatrix} \sqrt{1-\rho^2} y_1 \\ y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n - \rho y_{n-1} \end{bmatrix} \quad (5.38)$$

et

$$X^* = \begin{bmatrix} \sqrt{1-\rho^2} & \sqrt{1-\rho^2}x_{21} & \dots & \sqrt{1-\rho^2}x_{k1} \\ 1-\rho & x_{22}-\rho x_{21} & \dots & x_{k2}-\rho x_{k1} \\ 1-\rho & x_{23}-\rho x_{22} & \dots & x_{k3}-\rho x_{k2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 1-\rho & x_{2n}-\rho x_{2,n-1} & \dots & x_{kn}-\rho x_{k,n-1} \end{bmatrix} \quad (5.39)$$

Ces transformations sont appelées indistinctement différences partielles, quasi différences ou pseudo différences.

2° Exemple de la transformation en quasi différences

Soit le modèle à deux variables explicatives :

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t \quad (5.40)$$

avec

$$U_t = \rho u_{t-1} + v_t$$

$$|\rho| < 1$$

$$E(v_t) = 0$$

$$E(v_t^2) = \sigma_v^2$$

$$E(v_t v_{t-s}) = 0, \quad \forall s \neq 0$$

Le modèle peut s'écrire en $t-1$:

$$Y_{t-1} = \beta_0 + \beta_1 X_{1,t-1} + \beta_2 X_{2,t-1} + u_{t-1} \quad (5.41)$$

Le calcul de (5.40) - ρ (5.41) donne :

$$Y_t - \rho Y_{t-1} = \beta_0(1-\rho) + \beta_1(X_{1t} - \rho X_{1,t-1}) + \beta_2(X_{2t} - \rho X_{2,t-1}) + (u_t - \rho u_{t-1}) \quad (5.42)$$

Cette relation peut s'écrire simplement comme :

$$Y_t^* = \beta_0 + \beta_1 X_{1t}^* + \beta_2 X_{2t}^* + v_t \quad (5.43)$$

Le terme aléatoire u_t répond aux hypothèses d'application de la méthode des moindres carrés ordinaires, nous pouvons donc utiliser les procédures d'estimation des moindres carrés ordinaires sur les variables transformées. Les estimations des coefficients calculées à partir du modèle en différence s'interprètent directement comme étant les coefficients du modèle initial, sauf le terme constant qui est égal à :

$$b_0 = \beta_0(1-\rho) \Rightarrow \beta_0 = \frac{b_0}{1-\rho} \quad (5.44)$$

3° Procédures d'estimation de ρ

Examinons quelques méthodes pratiques d'estimation de ρ lorsque :

$$u_t = \rho u_{t-1} + v_t \quad (5.45)$$

a. Estimation de ρ à partir des résidus de la régression sur le modèle initial.

Etape 1 : estimation de ρ par régression directe de e_t sur e_{t-1} :

$$\hat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=2}^n e_{t-1}^2} \quad (5.46)$$

Remarquons que, pour de grands échantillons :

$$\sum e_t^2 \approx \sum e_{t-1}^2$$

$$\hat{\rho} \approx \rho = \frac{\sum e_t e_{t-1}}{\sqrt{\sum e_t^2} \sqrt{\sum e_{t-1}^2}} \quad (5.47)$$

(où ρ est le coefficient de corrélation).

Etape 2 : transformation des variables et régression sur les quasi – différences :

$$Y_t - \hat{\rho} Y_{t-1} = b_0 + \beta_1 (X_{1t} - \hat{\rho} X_{1,t-1}) + \beta_2 (X_{2t} - \hat{\rho} X_{2,t-1}) + \dots + \beta_k (X_{kt} - \hat{\rho} X_{k,t-1}) + v_t \quad (5.48)$$

Les paramètres estimés par OLS sont donc :

$$\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \dots, \hat{\beta}_k \text{ et le terme constant } \hat{\beta}_0 = \frac{\hat{b}_0}{(1-\hat{\rho})}.$$

b. Procédure itérative (méthode de Cochrane – Orcutt)

Etape 1 : initialisation de ρ

Par une technique d'estimation directe par régression, on fixe une première valeur à ρ , soit :

$$\hat{\rho} = \hat{\rho}_0$$

(on peut aussi donner une valeur a priori de ρ , soit $\hat{\rho}_0 = 0$)

Etape 2 : régression sur les quasi – différences

$$Y_t - \hat{\rho}_0 Y_{t-1} = b_0 + \beta_1 (X_{1t} - \hat{\rho}_0 X_{1,t-1}) + \beta_2 (X_{2t} - \hat{\rho}_0 X_{2,t-1}) + \dots + \beta_k (X_{kt} - \hat{\rho}_0 X_{k,t-1}) + v_t \quad (5.49)$$

Les paramètres estimés sont alors :

$$\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \dots, \hat{\beta}_k \text{ et le terme constant } \hat{\beta}_0 = \frac{\hat{b}_0}{(1-\hat{\rho})}.$$

Etape 3 : Réestimation de ρ

A partir des nouveaux résidus d'estimation (e_t^1), nous recalculons une nouvelle valeur de ρ , soit ρ_1 :

$$e_t^1 = Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_{1t} - \hat{\beta}_2 X_{2t} - \dots - \hat{\beta}_k X_{kt}$$

et

$$\hat{\rho}_1 = \frac{\sum_t e_t^1 e_{t-1}^1}{\sum_t (e_{t-1}^1)^2}$$

Etape 4 : régression sur les quasi – différences :

$$Y_t - \hat{\rho}_1 Y_{t-1} = b_0 + \beta_1 (X_{1t} - \hat{\rho}_1 X_{1,t-1}) + \dots + \beta_k (X_{kt} - \hat{\rho}_1 X_{k,t-1}) + v_t$$

Puis nous calculons un nouveau résidu (e_t^1) à partir de la nouvelle estimation des coefficients, cce qui permet d'obtenir un nouveau $\hat{\rho}_2$.

Et ainsi de suite jusqu'à la stabilité des coefficients β estimés (en général, après 3 ou 4 itérations). Donc, on arrête l'opération de calcul lorsqu'il y a convergence des estimateurs $\hat{\beta}$, c'est-à-dire lorsque les valeurs trouvées après deux itérations successives ne diffèrent plus que d'un nombre donné aussi petit que l'on veut.

c. Méthode du balayage (Procédure de HILDRETH – LU)

Supposons que ρ est positif et qu'il varie dans un intervalle $[0, 1]$.

Etape : régression, pour toutes les valeurs successives de $\rho = \{0.1; 0.2; 0.3; \dots; 1\}$ sur l'intervalle $[0, 1]$ avec un pas fixé généralement à 0.1, l'équation au quasi – différences :

$$Y_t - \rho_i Y_{t-1} = b_0 + \beta_1 (X_{1t} - \rho_i X_{1,t-1}) + \dots + \beta_k (X_{kt} - \rho_i X_{k,t-1}) + v_t$$

Etape 2 : retenir la valeur de ρ qui minimise la somme des carrés des résidus ($\sum e_t^1$) et déduire les valeurs des estimateurs $\hat{\beta}$ correspondantes.

Faisons remarquer que l'on peut affiner la valeur estimée de ρ à partir de la méthode de balayage en prenant un pas plus fin (par exemple : 0.01). Cette technique est optimale selon le critère des moindres carrés puisque ρ minimise la somme des carrés des résidus.

Remarques

1. Les différentes procédures d'estimation de ρ évoquées ici ne sont valables que dans le cas restrictif où la liaison des erreurs est effectivement autorégressive d'ordre 1, c'est-à-dire :
 $u_t = \rho u_{t-1} + v_t$.
2. Elles ne peuvent donner des résultats valables que pour un échantillon d'observations suffisamment grand.
3. Elles ne peuvent s'appliquer correctement sur un modèle initial qui contient une ou plusieurs variables endogènes retardées (décalées).
4. Tester l'autocorrélation des résidus n'a aucun sens sur les données en coupes transversales. Car, il n'y a pas d'ordonancement naturel des observations.

VI.6. Détection de l'autocorrélation des erreurs

En raison l'importance des conséquences de l'autocorrélation des erreurs sur l'efficacité des estimateurs, il s'avère essentiel de pouvoir déceler ce biais. Et, la détection d'une éventuelle dépendance des erreurs ne peut s'effectuer qu'à partir de l'analyse des résidus e_t car les erreurs u_t sont inconnues et restent inconnues.

1. Examen visuel des résidus

L'analyse graphique des résidus donne une indication sur la présence ou l'absence ou l'absence d'auto corrélation des erreurs. En effet, lorsque :

- Les résidus sont pendant plusieurs périodes consécutives soit positifs, soit négatifs (c'est-à-dire lorsqu'ils sont successivement de même), alors il y a autocorrélation positive.

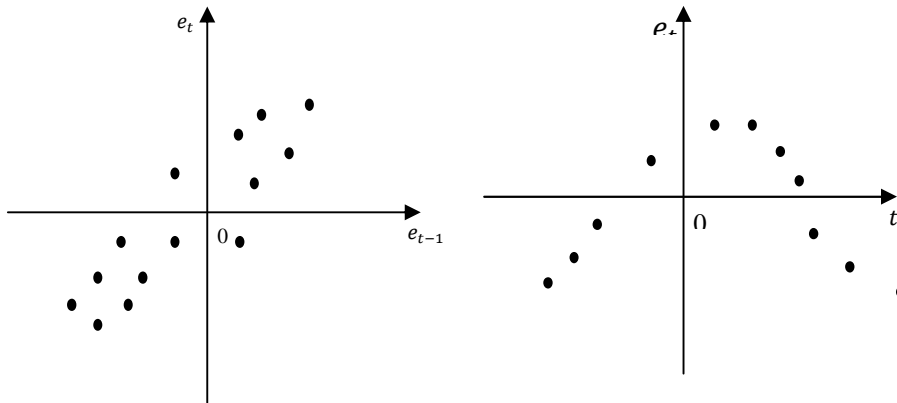


fig : autocorrélation positive

- Les résidus sont alternés (les résidus successifs sont généralement de signe contraire) : autocorrélation négative

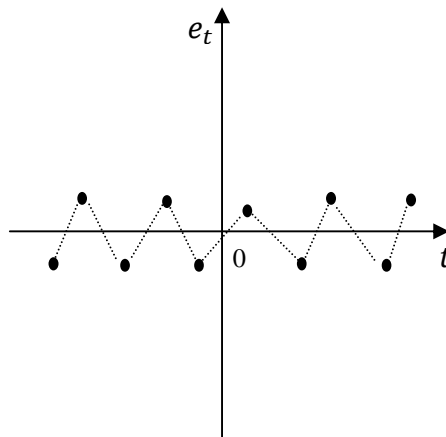


fig. : autocorrélation négative absence

- Les résidus évoluent indépendamment les uns des autres, alors il y a absence d'autocorrélation.

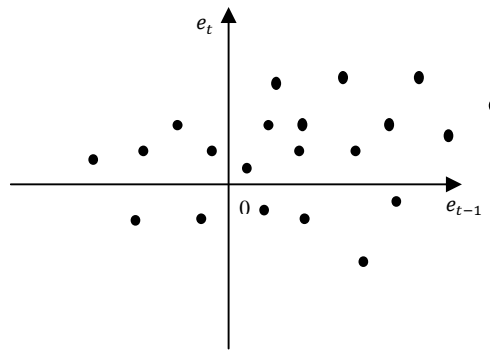


Fig. : absence d'autocorrélation

Soulignons que l'analyse graphique est parfois délicate d'interprétation car profil des résidus ne présente pas toujours des caractéristiques évidentes.

2. Test de Durbin Watson

La test de Durbin Watson (DW) permet de détecter une autocorrélation des erreurs d'ordre 1 de la forme :

$$u_t = \rho u_{t-1} + v_t \quad (5.50)$$

où $v_t \sim N(0, \sigma_v^2)$

Le test d'hypothèse est le suivant :

$H_0 : \rho = 0$ (il y a indépendance)

contre

$H_1 : \rho \neq 0$ (il n'y a pas indépendance).

Pour tester l'hypothèse nulle H_0 , nous calculons la statistique de Durbin et Watson :

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad (5.51)$$

où les e_t sont les résidus des moindres carrés ordinaires (résidus de l'estimation du modèle initial).

Notons ici la statistique DW est étroitement liée au coefficient du premier ordre des résidus (ρ) en effet, la statistique de Durbin Watson (DW ou simplement d) peut d'écrire :

$$\begin{aligned} d &= \frac{\sum_{t=2}^n e_t^2 + \sum_{t=2}^n e_{t-1}^2 - 2 \sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2} \\ &= \frac{2 \sum_{t=1}^n e_t^2 - 2 \sum_{t=2}^n e_t e_{t-1} - e_1^2 - e_n^2}{\sum_{t=1}^n e_t^2} \\ &= 2 - 2\gamma_1 \hat{\rho} - \gamma_2 \end{aligned}$$

où

$$\gamma_1 = \frac{\sum_{t=2}^n e_{t-1}^2}{\sum_{t=1}^n e_t^2}$$

$$\gamma_2 = \frac{e_1^2 + e_n^2}{\sum_{t=1}^n e_t^2}$$

$$\hat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=2}^n e_{t-1}^2}$$

Si le nombre d'observations (n) augmente : $\gamma_1 \approx 1$ et $\gamma_2 \approx 0$. Ainsi :

$$d \approx 2 - 2\hat{\rho} = 2(1 - \hat{\rho}) \text{ et } \hat{\rho} \approx 1 - \frac{d}{2}$$

On peut donc approximer la valeur de $\hat{\rho}$ à partir de la statistique d de Durbin Watson.

De par sa construction, la statistique d varie entre 0 et 4 :

- d = 2 Lorsque $\hat{\rho} = 0$ (absence d'autocorrélation) ;
- d = 0 lorsque $\hat{\rho} = 1$ (autocorrélation positive) ;
- d = 4 lorsque $\hat{\rho} = -1$ (autocorrélation négative)

Afin de tester l'hypothèse H_0 d'indépendance des erreurs :

$H_0 : \rho = 0$ (absence d'autocorrélation)

contre

$H_1 : \rho \neq 0$ (présence d'autocorrélation).

Durbin et Watson ont tabulé les valeurs critiques de la statistique d au seuil de 1% et 5% respectivement en fonction de la taille de l'échantillon (n) et du nombre des variables explicatives du modèle (k).

La lecture de la table statistique permet de déterminer deux valeurs d_L et d_U comprises entre 0 et 2 qui délimite l'espace entre 0 et 4 selon le schéma ci-après :

0	Autocorrélation positive	Doute	Indépendance des erreurs	Doute	Doute Autocorrélation négative	0
---	--------------------------	-------	--------------------------	-------	--------------------------------	---

La procédure du test est donc :

1. Si $0 \leq d < d_L$, on rejette l'hypothèse d'indépendance des erreurs et on admet une autocorrélation positive
2. Si $d_L \leq d \leq d_U$, on se trouve dans la zone d'indétermination de la table : il y a donc doute. Le test ne permet pas de conclure. Mais, dans doute la pratique, on considère le doute comme une présomption favorable d'autocorrélation.
3. Si $d_U < d < 4 - d_U$, on accepte l'hypothèse d'indépendance des erreurs. Il y a absence d'autocorrélation.
4. Si $4 - d_U \leq d \leq 4 - d_L$, il y a doute.
5. Si $4 - d_L < d \leq 4$, on rejette l'hypothèse d'indépendance des erreurs au profit d'une autocorrélation négative.

Remarques

Le test de Durbin Watson n'est qu'un test présomptif d'indépendance des erreurs du fait qu'il utilise les résidus pour sa construction. En outre, il ne test qu'une autocorrélation d'ordre 1 [AR(1)] des erreurs.

L'utilisation de ce test est conditionnée par les hypothèses suivantes :

- Le modèle de régression à tester ne peut pas comprendre des variables prédéterminées (variables endogènes et/ou exogènes) décalées ;
- Il est nécessaire d'avoir un terme constant dans le modèle de régression. Pour des modèles sans terme constant, il existe actuellement des tables statistiques appropriées ;
- La taille de l'échantillon doit excéder 15, car on trouve généralement des tables de Durbin Watson qui commencent par $n = 15$ aux seuils de signification de 1% et 5% respectivement. Mais, il existe actuellement des tables commençant par $n = 6$.

3. Test de Durbin pour une régression comprenant des valeurs retardées de la variable dépendante.

Comme nous l'avons fait remarquer, la procédure du test de Durbin et Watson suppose que la matrice X soit non aléatoire. Dans le cas d'un modèle autorégressif d'ordre 1 du type :

$$Y_t = b_1 Y_{t-1} + \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_k X_{kt} + u_t \quad (a)$$

avec

$$u_t = \rho u_{t-1} + v_t$$

Un test approprié d'autocorrélation des erreurs doit alors être utilisé.

La statistique utilisée est suivante :

$$h = \hat{\rho} \sqrt{\frac{n}{1 - n\hat{\sigma}_{b_1}^2}} = \left(1 - \frac{d}{2}\right) \sqrt{\frac{n}{1 - n\hat{\sigma}_{b_1}^2}}$$

où : n est la taille de d'échantillon ;

$\hat{\sigma}_{b_1}^2$, la variance estimée du coefficient Y_{t-1} dans la régression par la méthode des moindres carrés ordinaires ;

$\hat{\rho}$, le coefficient d'autocorrélation estimé des résidus $\left(\hat{\rho} = 1 - \frac{DW}{2}\right)$.

La statistique « h » est distribuée de manière asymptotique comme une variable normale centrée réduite. Le test d'hypothèse est le suivant :

$H_0 : h = 0$ (Indépendance des erreurs) ;

$H_1 : h \neq 0$ (non indépendances des erreurs).

Si $|h| \leq t_{\frac{\alpha}{2}}$, nous acceptons l'hypothèse H_0 d'indépendance des erreurs

($t_{\frac{\alpha}{2}}$ est la valeur issue de la loi normale pour un test bilatéral au seuil de α).

Notons que si $n\hat{\sigma}_{b_1}^2 \geq 1$, la statistique « h » ne peut être calculée. Dans tel cas, Durbin suggère une procédure asymptotiquement équivalente :

1. Estimer le modèle (a) par la méthode des moindres carrés ordinaires et calculer les résidus e ;
2. Régresser par la méthode des moindres carrés ordinaires :

$$e_t \text{ en } e_{t-1}, Y_{t-1}, X_{1t}, X_{2t}, \dots, X_{ht}$$
3. Si le coefficient de e_{t-1} dans cette régression est significativement différent de zéro, on regrette l'hypothèse base ($H_0 : \rho = 0$).

4. Test de BREUSCH – GODFREY :

Ce test, fondé soit sur un test de Fisher de nullité des coefficients, soit sur test de multiplicateur de Lagrange (LM test), permet de tester une autocorrélation d'un ordre supérieur à 1 et reste valide en présence de la variable dépendante décalée en tant que variable explicative. L'idée générale de ce test réside dans la recherche d'une relation signification entre le résidu et ce même résidu décalé.

Soit le modèle linéaire général ci-après :

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + u_t$$

où les erreurs sont autocorrélée d'ordre p donc :

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + v_t$$

En substituant u_t par sa valeur dans le modèle, on obtient :

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + v_t$$

Ce test est mené en trois étapes :

- Estimation par la méthode des moindres carrés ordinaires du modèle et calcul des résidus e_t ($t = 1, 2, \dots, n$) ;
- Estimation par la méthode des moindres carrés ordinaires de l'équation intermédiaire :

$$e_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \rho_1 e_{t-1} + \dots + \rho_p e_{t-p} + v_t$$

- Test d'hypothèse sur l'équation intermédiaire. L'hypothèse H_0 d'absence d'autocorrélation des erreurs à tester est :

$$H_0: \rho_1 = \rho_2 = \dots = \rho_p = 0.$$

Si on rejette l'hypothèse nulle, alors il existe un risque d'autocorrélation des erreurs à l'ordre p .

Pour mener ce test, nous avons deux possibilités :

- Soit effectuer un test de Fisher classique de nullité des coefficients ρ_i (où $i = 1, 2, \dots, p$) ;
- Soit utiliser la statistique LM qui est distribuée comme un χ^2 à P degrés de liberté, si la statistique $LM = n \times R^2$ (où n est le nombre d'observations et R^2 , le coefficient de détermination issu de l'estimation de l'équation intermédiaire) est supérieure à $\chi^2_{(p)}$ lu dans la table au seuil α , on rejette l'hypothèse d'indépendance des erreurs. Donc : $LM > \chi^2_{(p)} \Rightarrow R_{H_0}$.

5. Tests de Box-Pierce et Ljung-Box

L'absence d'autocorrélation des erreurs implique que :

$\rho_1 = \rho_2 = \dots = \rho_k = 0$, soit les hypothèses :

$H_0: \rho_1 = \rho_2 = \dots = \rho_k = 0$ (indépendance des erreurs) ;

H_1 : il existe au moins un ρ_i significativement différent de 0 (non indépendance des erreurs).

- Pour effectuer ce test, on recourt à la statistique Q (due à Box-Pierce) qui est donnée par

$$Q = n \sum_{k=1}^h \hat{\rho}_k^2$$

où

n = Nombre d'observations ;

h = nombre de retards (lags) ;

$$\text{et } \hat{\rho}_k = \frac{\sum_{t=k+1}^n e_t e_{t-k}}{\sum_{t=1}^n e_t^2}.$$

La statistique Q de Box-Pierce est distribuée de manière asymptotique comme un χ^2 à h degrés de liberté. Nous rejetons l'hypothèse nulle, d'absence d'autocorrélation si Q est supérieure au χ^2 lu dans la table au seuil ε et h degrés de liberté.

- b. On peut également utiliser une autre statique dont propriétés asymptotiques sont meilleures. Il s'agit de la statistique Q' de Ljung et Box :

$$Q' = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k}$$

qui est aussi distribuée selon un χ^2 à h degrés de liberté et dont les règles de décisions sont identiques au test de Box-Pierce.

VI. 7. Prévisions lorsque les erreurs sont autocorrélées

Soit le modèle linéaire général :

$$Y = X\beta + U \quad (1)$$

On se propose de déterminer T_0 observations futures de la variable endogène Y , représentées par le vecteur Y_0 de format $(T_0 \times 1)$:

$$Y_0 = X_0\beta + U_0 \quad (2)$$

où X_0 , matrice des variable explicative de format $(T_0 \times k)$;

U_0 , vecteur de perturbations admettant les mêmes propriétés que U .

Etant donné que les erreurs sont autocorrélées et qu'elles suivent un processus autorégressif d'ordre 1, on déduit :

$$E(U_0 U_0') = \sigma_v^2 \Psi_0 \quad (3)$$

où Ψ_0 , matrice carrée de dimension $(T_0 \times T_0)$, est définie de la même façon que Ψ .

Notons, en outre, que les éléments de U et U_0 sont autocorrélés. Par conséquent :

$$\begin{aligned} E(U_0 U) &= E \begin{bmatrix} u_{n+1} \\ u_{n+2} \\ \vdots \\ u_{n+T_0} \end{bmatrix} [u_1 \quad u_2 \quad u_3 \quad \dots \quad u_n] \\ &= E \begin{bmatrix} u_{n+1}u_1 & u_{n+1}u_2 & \dots & u_{n+1}u_n \\ u_{n+2}u_1 & u_{n+2}u_2 & \dots & u_{n+2}u_n \\ \vdots & \vdots & \ddots & \vdots \\ u_{n+T_0}u_1 & u_{n+T_0}u_2 & \dots & u_{n+T_0}u_n \end{bmatrix} \end{aligned}$$

$$= \frac{\sigma_v^2}{1-\rho^2} \begin{bmatrix} \rho^n & \rho^{n+1} & \dots & \dots & \rho \\ \rho^{n+1} & \rho^n & \dots & \dots & \rho^2 \\ \rho^{n+2} & \rho^{n+1} & \dots & \dots & \rho^3 \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ \rho^{n+T_0-1} & \rho^{n+T_0-2} & \dots & \dots & \rho^{T_0} \end{bmatrix}$$

Enfin, comme le vecteur de perturbation U est inconnu, on peut le remplacer par le vecteur des résidus obtenu par la méthode des moindres carrés généralisés :

$$e = Y - X\beta^* \quad (5)$$

où $\beta^* = (X'\Psi^{-1}X)^{-1}X'\Psi^{-1}Y$

La meilleure prévision de U_0 est donnée par :

$$e_0 = \begin{bmatrix} \rho \\ \rho^2 \\ \rho^3 \\ \vdots \\ \vdots \\ \rho^{T_0} \end{bmatrix} (Y_n - X_n'\beta^*)$$

$$= W(Y_n - X_n'\beta^*) \quad (6)$$

$$\text{avec } W = \begin{bmatrix} \rho \\ \rho^2 \\ \vdots \\ \vdots \\ \rho^{T_0} \end{bmatrix}$$

et $X_n' = (1 \ X_{2n} \ X_{3n} \dots X_{kn})$.

En effet, le premier élément de U_0 est :

$$u_{n+1} = \rho u_n + v_{n+1} \quad (7)$$

où $E(v_{n+1}) = 0$

et v_{n+1} est non corrélée avec (v_1, v_2, \dots, v_n) .

Comme u_n est connu, la prévision de u_{n+1} sera :

$$e_{n+1} = \rho u_n \quad (8)$$

De même :

$$u_{n+2} = \rho u_{n+1} + v_{n+2} \\ = \rho^2 u_n + \rho v_{n+1} + \rho v_{n+2} \quad (9)$$

Et la prévision de u_{n+2} est donc :

$$e_{n+2} = \rho^2 u_n \quad (10)$$

En continuant ce processus et en substituant le terme inconnu, u_n , par son estimateur : $e_n = Y_n - X'_n \beta^*$, on trouve :

$$e_0 = W(Y_n - X'_n \beta^*)$$

Etant donné que $X_0 \beta^*$ est la prévision de $X_0 \beta$, et en le combinant avec la prévision de la perturbation (6), on obtient :

$$\hat{Y}_0 = X_0 \beta^* + W(Y_n - X'_n \beta^*) \quad (11)$$

Ainsi, la prévision de Y pour un horizon prévisionnel h (où $h=1, 2, \dots, T_0$) est donc :

$$\hat{Y}_{n+h} = X'_{n+h} \beta^* + \rho^h (Y_n - X'_n \beta^*) \quad (12)$$

La relation (12) montre également que la contribution du second membre diminue si l'horizon de prévision h devient grand.

Exemple :

Soit le modèle :

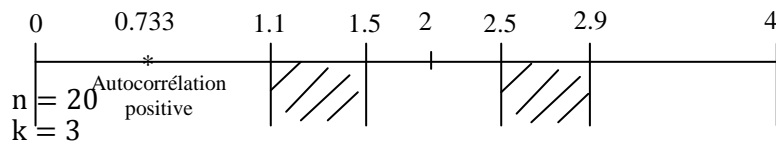
$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

$n = 20$ (nombre d'observations)

1. Application de la méthode des moindres carrés ordinaires a donné les résultats suivants :

$$\hat{\beta} = (X'X)^{-1}X'Y = \begin{bmatrix} 20.00 & 416.74 & 420.03 \\ 416.74 & 8946.93 & 8925.56 \\ 420.03 & 8925.56 & 3190.06 \end{bmatrix}^{-1} \begin{bmatrix} 1065.23 \\ 22353.79 \\ 22665.64 \end{bmatrix} = \begin{bmatrix} 35.3460 \\ 0.1060 \\ 0.7450 \end{bmatrix}$$

$$DW = 0.733$$



Au seuil de $\alpha\% = 5\%$, $d_L = 1.1000$ et $d_U = 1.5370$

Le test de DW suggère l'existence d'une autocorrélation positive des erreurs. La matrice des variances et covariances de $\hat{\beta}$ est donc :

$$\hat{\Omega}_{\hat{\beta}} = \hat{\sigma}_e (X'X)^{-1} = \begin{bmatrix} 38.8360 & -1.0380 & -0.4920 \\ - & 0.0950 & -0.0450 \\ - & - & 0.068 \end{bmatrix}$$

$$\text{Où } \hat{\sigma}_e = \frac{e'e}{17} = 17.2130$$

2. Estimation de ρ

Pour estimer ρ , on a utilisé la statistique DW.

D'où :

$$\begin{aligned}\hat{\rho} &= 1 - \frac{DW}{2} \\ &= 1 - \frac{1}{2}(0.7330) = 0.63346\end{aligned}$$

3. Estimateur des moindres carrés généralisés

$$\begin{aligned}\beta^* &= (X'P'PX)^{-1}X'P'PY \\ &= (X^{*'}X^*)^{-1}X^{*'}Y^* \\ &= \begin{bmatrix} 3.1510 & 66.0340 & 66.2760 \\ - & 1455.4390 & 1429.5970 \\ - & - & 1730.5950 \end{bmatrix}^{-1} \begin{bmatrix} 168.1080 \\ 3588.2590 \\ 3859.6380 \end{bmatrix} = \begin{bmatrix} 25.8200 \\ 0.3950 \\ 0.9150 \end{bmatrix}\end{aligned}$$

$$\begin{aligned}\hat{\Omega}_{\beta^*} &= \hat{\sigma}_v^2 (X'P'PX)^{-1} \\ &= \begin{bmatrix} 64.4250 & -2.6490 & -0.2790 \\ - & 0.1440 & -0.0170 \\ - & - & 0.0310 \end{bmatrix}\end{aligned}$$

où

$$\hat{\sigma}_v^2 = \frac{(Y^* - X^*\beta^*)'(Y^* - X^*\beta^*)}{n - k} = 3.6170$$

4. Prévision de Y_{n+1} étant donné $X'_{n+1} = [1 \ 20 \ 20]$
 $\hat{Y}_{n+1} = X'_{n+1}\beta^* + \hat{\rho}(Y_n - X'_n\beta^*)$

$$\begin{aligned}&= [1 \ 20 \ 20] \begin{bmatrix} 25.8200 \\ 0.3950 \\ 0.9150 \end{bmatrix} + 0.63346 \left[61.75 - [1 \ 28.73 \ 25.66] \begin{bmatrix} 25.8200 \\ 0.3950 \\ 0.9150 \end{bmatrix} \right] \\ &= 52.02 + 0.83 = 52.85\end{aligned}$$

Notons que : $X_{2n} = 28.73$ et $X_{3n} = 25.66$

Chapitre VII. HETEROSCEDASTICITE

6.1 Présentation du problème

Soit le modèle linéaire général :

$$Y = X\beta + U \quad (7.1)$$

Pour lequel l'hypothèse de la constance de la variance des erreurs n'est pas vérifiée. Dans ce cas, les erreurs sont de dispersion variable d'une observation (période) à l'autre, c'est-à-dire :

$$\text{Var}(U_i) = E(U_i^2) = \sigma_i^2 \quad (7.2)$$

où : $i = 1, 2, \dots, n$

Et la matrice des variances – covariances des erreurs prend la forme ci – après :

$$\begin{aligned} \Omega_U = U(UU') &= \begin{bmatrix} E(u_1^2) & E(u_1 u_2) & \dots & E(u_1 u_n) \\ E(u_2 u_1) & E(u_2^2) & \dots & E(u_2 u_n) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ E(u_n u_1) & E(u_n u_2) & \dots & E(u_n^2) \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix} = \sigma^2 \Psi = V \end{aligned} \quad (7.3)$$

La relation (6.3) peut s'écrire aussi comme :

$$V = \Omega_U = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2) \quad (7.4)$$

où $\sigma_i^2 = \sigma^2 w_i$
 $i = 1, 2, \dots, n$

Généralement, on considère que :

$$\text{tr}(\Psi) = \sum_{i=1}^n w_i = n \quad (7.5)$$

Si tel est le cas, la régression classique avec l'hypothèse des erreurs homoscedastiques peut être considérée comme un cas spécial où $w_i = 1$ pour tout i .

Les causes de l'hétéroscédasticité sont multiples. Elle peut arriver lorsque :

- Les observations représentent des moyennes calculées sur des échantillons de taille différentes ;
- Une même valeur de la variable endogène ne répète pour des valeurs différentes d'une variable explicative ;
- Lorsque les erreurs sont liées aux valeurs prises par une variable explicative. Ceci se retrouve souvent lorsque le modèle est spécifié en coupes transversales (cross – section data).

Remarque :

Ce problème se rencontre plus fréquemment pour les modèles en coupe transversale (ou instantanée) contrairement au problème de l'autocorrélation qui est plutôt spécifique aux modèles en séries temporelles.

6.2 Conséquences de l'hétéroscédasticité

Les conséquences de l'hétéroscédasticité sont identiques à celles de l'autocorrélation des erreurs :

- L'estimateur de β obtenu par la méthode des moindres carrés ordinaires :

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (7.6)$$

est non biaisé ;

- Mais sa variance n'est plus minimale :

$$\Omega_{\hat{\beta}} = \sigma^2(X'X)^{-1}(X'\Psi X)(X'X)^{-1} \quad (7.7)$$

Entraînant ainsi des modifications très sensibles des niveaux de signification des différents tests statistiques.

6.3 Estimation lorsqu'il y a hétéroscédasticité

$$\text{Soit : } Y = X\beta + U \quad (7.8)$$

un modèle hétéroscédastique pour lequel la matrice des variances – covariances :

$$\Omega_U = V = \sigma^2\Psi \neq \sigma_U^2 I \quad (7.9)$$

La matrice V est symétrique et définie positive, on peut donc la diagonaliser et ses valeurs propres sont positives.

On a alors :

$$\Psi = PDP' \quad (7.10)$$

où P est une matrice orthogonale ($P' = P^{-1}$) ;

D est une matrice diagonale à éléments positifs

Donc :

$$\begin{aligned}\Psi &= PD^{1/2}D^{1/2}P' \\ &= \left(PD^{\frac{1}{2}}\right)\left(PD^{\frac{1}{2}}\right)' \\ &= HH'\end{aligned}\quad (7.11)$$

Trouvons une transformation du modèle (7.1) telle que l'erreur dans le modèle transformé puisse vérifier les hypothèses fondamentales, c'est-à-dire cherchons λ telle que :

$$\lambda Y = \lambda X\beta + \lambda U \quad (7.12)$$

avec : $V(\lambda U) = \sigma^2 I$

En supposant que $\lambda = H^{-1}$, la matrice des variances – covariances devient :

$$\begin{aligned}V(H^{-1}U) &= E[(H^{-1}U)(H^{-1}U)'] \\ &= E[(H^{-1})UU'(H^{-1})'] \\ &= (H^{-1})E(UU')(H^{-1})' \\ &= H^{-1}\sigma^2\Psi(H^{-1})'\end{aligned}$$

Comme $\Psi = HH'$, on a alors :

$$V(H^{-1}U) = H^{-1}\sigma^2 HH'(H^{-1})' = \sigma^2 I \quad (7.13)$$

Le modèle (7.12) devient :

$$H^{-1}Y = H^{-1}X\beta + H^{-1}U \quad (7.14)$$

En appliquant la méthode des moindres carrés au modèle transformé (7.14), obtient :

$$\begin{aligned}\hat{\beta} &= [(H^{-1}X)'(H^{-1}X)]^{-1}(H^{-1}X)'H^{-1}Y \\ &= [X'(H^{-1})'(H^{-1})X]^{-1}X'(H^{-1})'(H^{-1})Y \\ &= (X'\Psi X)^{-1}X'\Psi^{-1}Y\end{aligned}\quad (7.15)$$

$$\text{avec } \Omega_{\hat{\beta}} = \sigma^2(X'\Psi X)^{-1} \quad (7.16)$$

L'estimation (7.15) est appelé estimateur des moindres carrés généralisés (GLS = Generalized Least Squares) ou estimateur de AITKEN.

Faisons remarquer que si $V = \sigma^2 I$, on retrouve l'estimateur de Gauss – Markov :

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (7.17)$$

L'estimateur de AITKEN généralise donc l'estimateur de Gauss – Markov.

Remarques :

Il n'existe pas une méthodologie unique de correction de l'hétéroscédasticité. La règle générale consiste à transformer les données de

variables du modèle afin de se ramener à un modèle à variances constantes, c'est-à-dire à un modèle homoscédastique.

Ainsi, pour le modèle hétéroscédastique suivant :

$$Y_i = \beta_0 + \beta_1 X_i + u_i \quad (7.18)$$

où : $i = 1, 2, \dots, n$

Les transformations éventuelles à opérer pour corriger l'hétéroscédasticité consiste en :

$$\frac{Y_i}{n_i} = \frac{\beta_0}{n_i} + \beta_1 \frac{X_i}{n_i} + \frac{u_i}{n_i} \quad (7.19)$$

si $E(u_i^2) = \frac{\sigma_u^2}{n_i}$

$$\frac{Y_i}{X_i} = \frac{\beta_0}{X_i} + \beta_1 + \frac{u_i}{X_i} \quad (7.20)$$

si $E(u_i^2) = \sigma_u^2 X_i^2$

$$\frac{Y_i}{\sqrt{f(X_i)}} = \frac{\beta_0}{\sqrt{f(X_i)}} + \frac{\beta_1}{\sqrt{f(X_i)}} + \frac{u_i}{\sqrt{f(X_i)}} \quad (7.21)$$

si $E(u_i^2) = \sigma_u^2 f(X_i)$

Exemple

Soit à estimer la fonction de consommation ci – après :

$$C_t = aR_t + b + u_t \quad (7.22)$$

où $t = 1, 2, \dots, n$

En supposant que $\text{Var}(u_t) = a^2 R_t^2$. Le modèle étant hétéroscédastique, pour l'estimer, il faut procéder à la transformation des toutes les variables du modèle (7.22). D'où :

$$Y_t = \frac{C_t}{R_t}, \quad X_t = \frac{1}{R_t} \quad \text{et} \quad \varepsilon_t = \frac{u_t}{R_t}$$

Le modèle (6.22) devient alors :

$$Y_t = a + bX_t + \varepsilon_t \quad (7.23)$$

avec :

$$\begin{aligned} \text{Var}(u_t) &= \text{Var}\left(\frac{u_t}{R_t}\right) \\ &= \frac{1}{R_t^2} \text{var}(u_t) = \frac{\sigma^2 R_t^2}{R_t^2} = \sigma^2 \end{aligned} \quad (7.24)$$

On peut donc estimer le modèle (7.23) par la méthode des moindres carrés ordinaires :

$$\hat{a} = \bar{Y} - \hat{b}\bar{X} \quad (7.25)$$

et

$$\hat{b} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} \quad (7.26)$$

\hat{a} et \hat{b} sont les estimateurs des moindres carrés généralisés du modèle de départ (7.22); ils correspondent aux estimateurs des moindres carrés ordinaires du modèle transformé (7.23).

6.4 Tests de détection de l'hétéroscédasticité

Examinons quelques tests permettant de détecter une hétéroscédasticité.

1. Test de Park

L'approche de Park propose d'établir une relation fonctionnelle quelconque entre la variance du terme d'erreur et la variable explicative X_i . La forme fonctionnelle utilisée par l'auteur est la suivante :

$$\sigma_i^2 = \gamma^2 X_i^\beta e^{v_i} \quad (7.27)$$

ou bien :

$$\ln \sigma_i^2 = \ln \gamma^2 + \beta \ln X_i + v_i \quad (7.28)$$

où v_i est le terme d'erreur stochastique.

Comme σ_i^2 n'est généralement pas connue, Park suggéra d'utiliser les carrés des résidus, e_i^2 , comme substitut et posa la régression suivante :

$$\begin{aligned} \ln e_i^2 &= \ln \gamma^2 + \beta \ln X_i + v_i \\ &= \alpha + \beta \ln X_i + v_i \end{aligned} \quad (7.29)$$

- Si $\hat{\beta}$ est statistiquement significatif, on conclut par la présence d'hétéroscédasticité.
- S'il n'est pas significatif, on peut admettre l'hypothèse d'homoscédasticité.

Les étapes du test de Park sont les suivantes :

- Exécuter la régression du modèle original par la méthode des moindres carrés ordinaires sans tenir compte de l'hétéroscédasticité et tirer les résidus e_i .
- Faire la régression du modèle (7.29) et procéder au test de β .

Remarques :

La forme fonctionnelle particulière choisie par Park n'est qu'une suggestion. Une autre forme peut révéler des relations significatives. On peut utiliser e_i^2 au lieu de $\ln e_i^2$ comme variable explicative.

2. Test de Goldfeld – Quandt

Ce test est applicable si l'une des variables explicatives est la cause de l'hétéroscédasticité et le nombre d'observations est important. On suppose que l'écart – type de l'erreur augmente proportionnellement avec X_i qui en est la cause.

Les étapes du test sont les suivantes :

Etape 1 : Ordonner les observations en ordre croissant en fonction des

valeurs de X_i .

Etape 2 : Dériver l'échantillon en deux groupes en omettant C observations situées au centre de l'échantillon.

Etape 3 : Appliquer séparément la méthode des moindres carrés ordinaires sur les deux sous – échantillons pour obtenir respectivement SCR_1 (la somme des carrés des résidus de la régression sur les plus petites valeurs de X_i) et SCR_2 (somme des carrés des résidus de la régression sur les plus grandes valeurs de X_i).

Etape 4 : Calculer, sous l'hypothèse nulle H_0 d'homoscédasticité :

$$\begin{cases} H_0 : \sigma_i^2 = \sigma^2, & \forall i \\ H_1 : \sigma_i^2 \neq \sigma^2, & \text{pour au moins un } i. \end{cases}$$

$$\text{La statistique (le ratio)} \quad F = \frac{SCR_2/ddl_2}{SCR_1/ddl_1} \quad (7.30)$$

où $ddl_1 = n_1 - k$, degrés de liberté du processus sous – échantillon.

La relation (6.30) suit une loi de Fisher à ddl_1 et ddl_2 degrés de liberté.

Si $F_{\text{calculé}} > F_{\text{table}}$, au niveau choisi de signification, l'hypothèse d'homoscédasticité H_0 est rejetée, le modèle est donc hétéroscédastique.

3. Test de Glejser

Le test de Glejser permet non seulement de détecter une éventuelle hétéroscédasticité, mais aussi d'identifier la forme que revêt cette hétéroscédasticité. Ce test est fondé sur la relation entre les résidus de l'estimation du modèle de base par la méthode des moindres carrés ordinaires et la variable explicative, supposée être la cause de l'hétéroscédasticité.

Les étapes du test de Glejser sont les suivantes :

Etape 1 : Régression par la méthode des moindres carrés de base pour trouver le vecteur des résidus e .

Etape 2 : Régression de la valeur absolue des résidus $|e_j|$ sur X_j . La régression porte sur différentes formes de la relation, notamment :

$$- |e_j| = \beta_0 + \beta_1 X_j + v_j \quad (7.31)$$

avec une hétéroscédasticité du type $\hat{\sigma}_{uj}^2 = k^2 X_j^2$
(où k est une constante et v_j est le terme d'erreur)

$$- |e_j| = \beta_0 + \beta_1 X_j^{1/2} + v_j \quad (7.32)$$

avec $\hat{\sigma}_{uj}^2 = k^2 X_j$

$$- |e_j| = \beta_0 + \beta_1 X_j^{-1} + v_j \quad (7.33)$$

$$\begin{aligned}
 & \text{avec } \hat{\sigma}_{uj}^2 = k^2 X_j^{-2} \\
 - \quad |e_j| &= \beta_0 + \beta_1 X_j^{-1/2} + v_j \\
 & \text{avec } \hat{\sigma}_{uj}^2 = k^2 X_j^{-1}
 \end{aligned}
 \tag{7.34}$$

Etape 3 : Tester l'hypothèse de nullité de β_1 en comparant la statistique t empirique avec le t de la table. L'hypothèse nulle $H_0 : \beta_1 = 0$ d'homoscédasticité sera rejetée si $t_{\text{calculé}} > t_{\text{table}}$.

4. Test d'hétéroscédasticité de White

A. Présentation du test

Ce test ne nécessite pas que l'on spécifie les variables dont on pense qu'elles sont à l'origine de l'homoscédasticité. Il est fondé sur l'existence d'une relation entre carré des résidus de la régression d'origine et les variables explicatives en niveau, leurs valeurs au carré et en termes croisés :

$$\begin{aligned}
 e_t^2 &= \beta_0 + \beta_1 X_{2t} + \dots + \beta_p X_{pt} + \alpha_1 X_{2t}^2 + \dots + \alpha_p X_{pt}^2 + \gamma_1 X_{1t} X_{2t} \\
 &+ \dots + \gamma_m X_{p-t,t} X_{pt} + v_t
 \end{aligned}
 \tag{7.35}$$

Si au moins un des coefficients de régression est significatif, on rejette l'hypothèse nulle d'homoscédasticité en faveur de l'hypothèse alternative d'hétéroscédasticité.

Pour détecter la présence de l'hétéroscédasticité, on utilise :

- Soit le test de Fisher classique de nullité des coefficients :

$$H_0 : \beta_1 = \alpha_1 = \beta_2 = \alpha_2 = \dots = 0 \tag{7.36}$$

Si l'on rejette l'hypothèse nulle, il existe un risque d'homoscédasticité ;

- Soit le test LM (du multiplicateur de Lagrange). Sous l'hypothèse de base d'homoscédasticité :

$$LM = nR^2 \sim \chi_{(p)}^2 \tag{7.37}$$

où $p = 2k$ est le nombre des variables dans la régression auxiliaire (7.35) ; R^2 , le coefficient de détermination de la régression auxiliaire.

En conséquence, si $nR^2 < \chi_{(p)}^2$, on accepte l'hypothèse d'homoscédasticité. Si $nR^2 \geq \chi_{(p)}^2$, on conclut en faveur de l'hypothèse alternative d'hétéroscédasticité.

B. Les inconvénients du test de White

- Si on rejette l'homoscédasticité, on a aucune indication quant à la forme prise par l'hétéroscédasticité et, par conséquent, pas de guide pour déterminer un estimateur des moindres carrés généralisés.
- Le nombre de degrés de liberté dans le LM test χ^2 peut devenir assez important, ce qui tend à réduire le pouvoir du test.

Remarque

Le test de White peut également être utilisé comme test de mauvaise spécification d'un modèle. En effet, sous l'hypothèse nulle, le test de White suppose que les erreurs sont non seulement homoscédastiques, mais aussi non corrélées avec les régresseurs et que la spécification linéaire du modèle est correcte. Si l'une de ces conditions est violée, la statistique de test sera supérieure à la valeur critique. Au contraire, si la valeur de la statistique est inférieure à la valeur critique, ceci témoigne du fait qu'aucune de ces trois conditions n'est violée.

5. Test de ARCH

Les modèles de type ARCH (Autoregressive Conditional Heteroscedasticity) ou (Hétéroscédasticité Autorégressive conditionnelle), initiés par R. Engle, permettent de modéliser des séries chronologiques (time series) qui présentent de larges variations, ou volatilité (ou variance ou variabilité), ce qui suggère que la variance des séries chronologiques se modifie dans le temps.

En effet, traditionnellement, les économistes ont été vigilants sur la possibilité des perturbations hétéroscédastiques dans les analyses des séries en coupes transversales (cross section) et des perturbations autocorrélées dans les études des séries chronologiques. Dans le cas des séries en coupes transversales, on fait généralement l'hypothèse de l'existence d'homoscédasticité, tandis que toutes les corrélations des résidus pris deux à deux sont supposées être nulles lorsqu'il s'agit des séries temporelles. Mais, actuellement, il est établi que l'hétéroscédasticité peut aussi être présente dans le contexte des séries chronologiques. Les études des séries temporelles financières telles que les prix des actions, le taux de change, le taux d'inflation, etc., présentent souvent une *volatilité groupée*, c'est-à-dire des périodes dans lesquelles leurs prix observent d'amples oscillations pour une période longue suivies par des périodes de calme relatif. D'où l'idée selon laquelle le passé proche peut fournir des informations sur la variance conditionnelle d'observations. Ainsi :

$$\text{Var}(u_t) = \sigma_{u_t}^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \dots + \alpha_p u_{t-p}^2 \quad (7.38)$$

S'il n'y a pas d'autocorrélation dans la variance d'erreur, on a :

$$H_0 = \alpha_1 = \alpha_2 = \dots = \alpha_p = 0 \quad (7.39)$$

auquel cas $\text{Var}(u_t)=0$ et l'on n'obtient pas l'effet ARCH.

Puisqu'on n'observe pas directement σ_t^2 , Engle a démontré que faire la régression suivante peut facilement tester l'hypothèse nulle précédente :

$$e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \alpha_2 e_{t-2}^2 + \dots + \alpha_p e_{t-p}^2 \quad (7.40)$$

où e_t est le résidu estimé de la régression du modèle de régression original.

Le test d'hétéroscédasticité autorégressive conditionnelle est une conséquence de (7.38). Ce test est fondé soit sur un test de Fisher classique, soit sur le test du multiplicateur de Lagrange (LM).

De manière pratique, on procède de la manière suivante :

Etape 1 : faire l'estimation par la méthode des moindres carrés ordinaires de la régression de Y sur X pour obtenir les résidus (e_t).

Etape 2 : faire la régression par la méthode des moindres carrés ordinaires :

$$e_t^2 = \beta_0 + \sum_{i=1}^p \beta_i e_{t-i}^2 + v_t \quad (7.41)$$

Etape 3 : tester si $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$ sont conjointement significatifs, l'hypothèse de perturbations homoscédastiques conditionnelles est rejetée en faveur de perturbations ARCH.

Etape 3' : ou bien calculer la statistique $LM = nR^2$ à partir de la régression auxiliaire (6.40). La règle de décision est alors :

- $LM \leq \chi_{(p)}^2$, on accepte l'hypothèse nulle d'homoscédasticité ;
- $LM > \chi_{(p)}^2$, on rejette l'hypothèse nulle en faveur de l'hypothèse alternative d'hétéroscédasticité.

6. Test de BREUSCH – PAGAN-GODFREY

Il s'agit d'un test très général dans la mesure où il couvre un grand nombre de cas d'hétéroscédasticité.

Soit le modèle linéaire à k variables :

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i \quad (7.42)$$

La variance de l'erreur, σ_i^2 , peut s'écrire :

$$\sigma_i^2 = E(u_i^2) = f(Z_i' \alpha) \quad (7.43)$$

où : Z_i est un vecteur ($p \times 1$) des variables dont on pense qu'elles peuvent être responsables de l'hétéroscédasticité (quelques-unes ou la totalité des variables X peuvent représenter les variables Z) ;

α , un vecteur des coefficients inconnus de dimension ($p \times 1$) ;

$f(\cdot)$, une fonction non spécifiée supposée continue et au moins deux fois différentiables.

En supposant que :

$$\sigma_i^2 = \alpha_1 + \alpha_2 Z_{2i} + \dots + \alpha_p Z_{pi} \quad (7.44)$$

l'hypothèse de base d'homoscédasticité devient alors :

$$H_0 : \alpha_2 = \alpha_3 = \dots = \alpha_p = 0 \quad (7.45)$$

de sorte que la variance des erreurs $\sigma_i^2 = \alpha_1$ est une constante au cours du temps.

L'hypothèse alternative d'hétéroscédasticité correspond alors au cas où α contient des éléments non nuls.

Afin de mettre en œuvre le test, la procédure à suivre est la suivante :

- Estimer le modèle original par la méthode des moindres carrés ordinaires (OLS) afin d'obtenir les résidus e_i et de calculer :

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^n e_i^2}{n} \quad \text{et} \quad v_i = \frac{e_i^2}{\tilde{\sigma}^2}$$

(Notons que $\tilde{\sigma}^2 \neq \hat{\sigma}^2 = \frac{\sum_{i=1}^n e_i^2}{n-k}$).

- Faire la régression de v_i sur Z_i par la méthode des moindres carrés ordinaires et calculer la somme des carrés expliquée (SCE) :

$$v_i = \alpha_1 + \alpha_2 Z_{2i} + \dots + \alpha_p Z_{pi} \quad (7.46)$$

- Sous l'hypothèse nulle d'homoscédasticité H_0 , la statistique $Q = \frac{SCE}{2}$ suit asymptotiquement une loi de χ^2 à $(p-1)$ degrés de liberté. Donc :

$$\frac{SCE}{2} \sim \chi_{(p-1)}^2 \quad (7.47)$$

On rejette l'homoscédasticité si $\frac{SCE}{2}$ dépasse la valeur donnée par la table de χ^2 .

Soulignons qu'une procédure plus simple mais asymptotiquement équivalente est de faire la régression de e_t^2 sur Z_t . Alors le $LM = nR^2$ de cette régression suit asymptotiquement une loi de $\chi_{(p-1)}^2$ sous l'hypothèse de base.

Remarque :

Ce test requiert la connaissance des variables Z à l'origine de l'hétéroscédasticité, bien qu'il ne soit pas nécessaire de connaître la forme exacte de leur action. Une telle information risque de ne pas être aisément disponible. En pratique, les variables candidates peuvent être une ou plusieurs des variables explicatives présentes dans la matrice X .

7. Test d'égalité des variances

Ce test s'applique aux données constituées en m groupes de n observations. Pour tester l'égalité des variances, on fait l'hypothèse :

$$H_0 : \hat{\sigma}_1^2 = \hat{\sigma}_2^2 = \dots = \hat{\sigma}_m^2 \quad (7.48)$$

Les étapes de calcul sont :

- *Etape 1* : Calcul de la variance empirique de chaque groupe

$$\hat{\sigma}_i^2 = \frac{\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}{n_i - 1} \quad (7.49)$$

- *Etape 2* : Calcul de la variance totale

$$\hat{\sigma}_T^2 = \frac{\sum_{i=1}^m (n_i - 1) \hat{\sigma}_i^2}{\sum_{i=1}^m (n_i - 1)} = \frac{\sum_{i=1}^m v_i \hat{\sigma}_i^2}{v} \quad (7.50)$$

$$\text{Avec } v_i = n_i - 1 ; v = \sum_{i=1}^m v_i = \sum_{i=1}^m (n_i - 1)$$

- *Etape 3* : Calculer χ^2 et tester la quantité :

$$Q' = v \ln \hat{\sigma}_T^2 - \sum_{i=1}^m v_i \ln \hat{\sigma}_i^2 \sim \chi_{(m-1)}^2 \quad (7.51)$$

Cette statistique peut être améliorée en divisant Q' par une constante d'échelle. D'où :

$$Q = \frac{Q'}{C} \sim \chi^2_{(m-1)} \quad (7.52)$$

avec $C = 1 + \frac{1}{3(m-1)} \left[\sum_{i=1}^m \frac{1}{v_i} - \frac{1}{v} \right]$.

Décision : Si $Q > \chi^2_{0.95; (m-1)}$: l'hypothèse nulle H_0 est rejetée, le modèle est donc hétéroscédastique.

8. Test de KOENKER-BASSET

Comme la plupart des tests que nous avons évoqués ci-haut, le test de KB est basé sur le carré des résidus, e_i^2 ; mais , au lieu d'être régressé sur une variable explicative (ou plusieurs), le carré des résidus est régressé sur le carré des valeurs estimées de la variable dépendante.

Soit le modèle d'origine est :

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i \quad (7.53)$$

Les étapes du test sont les suivantes :

- Estimer le modèle (6.53) afin d'obtenir les résidus e_i ;
- Procéder à l'estimation de :

$$e_i^2 = \alpha_1 + \alpha_2 (\hat{Y}_i)^2 + v_i \quad (7.54)$$

où les \hat{Y}_i sont les valeurs estimées du modèle (7.53).

L'hypothèse nulle est que : $H_0 : \alpha_2 = 0$.

Si ceci n'est pas rejeté, on peut conclure à l'absence d'hétéroscédasticité.

Remarques :

- Si le modèle (7.53) est double log, alors le carré des résidus est régressé sur $(\log \hat{Y})^2$.
- Le test de KB s'applique s'il y a un ou plusieurs régresseurs.

Présentation du problème

L'une des hypothèses fondamentales du modèle de régression multiple est que le rang de la matrice X est maximum. Cette hypothèse signifie que les variables explicatives incluses dans le modèle sont linéairement indépendantes, ou orthogonales.

Cependant, il est assez fréquent en pratique que les variables explicatives soient plus ou moins liées entre elles. Dans de tels cas, on parle de la présence de multicollinéarité, c'est-à-dire l'existence d'une forte corrélation entre les variables explicatives qui rend difficile la mise en évidence de leur impact individuel.

On distingue deux types de multicollinéarité : la multicollinéarité parfaite et la multicollinéarité imparfaite.

a) On parle de *multicollinéarité parfaite* ou *exacte* lorsqu'une variable explicative est les résultats d'une combinaison linéaire de plusieurs autres variables explicatives.

b)

Dans un modèle comprenant k variables explicatives, la multicollinéarité se manifeste s'il existe une relation telle que :

$$\lambda_1 X_{1t} + \lambda_2 X_{2t} + \dots + \lambda_k X_{kt} = 0 \quad (1.1)$$

où $\lambda_1, \lambda_2, \dots, \lambda_k$ sont des constantes non toutes nulles.

Si la multicollinéarité est parfaite, il est théoriquement impossible d'inverser la matrice $(X'X)$ et, par conséquent, de calculer les coefficients du modèle.

c) Les cas de parfaite multicollinéarité sont rares. En pratique, les variables explicatives présentes une multicollinéarité forte, mais pas parfaite. On parle alors de *multicollinéarité imparfaite*, de *quasi multicollinéarité* ou simplement de *multicollinéarité*.

On est en présence de multicollinéarité imparfaite si, dans un modèle à k variables explicatives, on peut définir la relation ci-après :

$$\lambda_1 X_{1t} + \lambda_2 X_{2t} + \dots + \lambda_k X_{kt} + u_t = 0 \quad (1.2)$$

où : $\lambda_1, \lambda_2, \dots, \lambda_k$ sont des constantes non toutes nulles ;

u_t , le terme d'erreur.

Dans ce cas, quels que soient les résultats, les coefficients de régression peuvent être estimés.

Sources de multicollinéarité

La multicollinéarité peut provenir des facteurs suivants :

- La méthode de collectes des données ;
- Les contraintes sur le modèle ou dans la population échantillonnée ;
- La spécification du modèle ;
- Un modèle surdéterminé. Ceci se produit lorsqu'un modèle a plus de variables explicatives que le nombre d'observations ;

- Les variables explicatives présentes dans le modèle partagent un trend commun, c'est-à-dire quand elles croissent et décroissent toutes dans le temps.

Conséquences de la multicollinéarité

Les principaux symptômes de la présence de multicollinéarité sont les suivants :

- Les variances et covariances ont tendance à augmenter, et les résultats de l'estimation perdent donc en précision. En effet, du fait de la valeur élevée des variances des coefficients estimés, les t statistiques sont faibles alors que le R^2 et le F test sont élevés.
- Les coefficients estimés sont fortement sensibles à de petites variations des données. Une petite modification de l'échantillon ou du nombre d'observations peut engendrer des changements notables dans les coefficients estimés.
- Il est difficile, voire impossible, de distinguer les effets de différentes variables explicatives sur la variable expliquée (effet de masque).

Détection de la multicollinéarité

Le soupçon de multicollinéarité peut être confirmé et son importance évaluée en utilisant l'une des techniques de détection ci-après :

Corrélation entre les variables explicatives

Cette méthode consiste à calculer les coefficients de corrélation linéaire entre les variables explicatives. Si celles-ci sont fortement corrélées, il y a présomption de multicollinéarité. Mais, le problème avec cette technique est que la présence de fortes corrélations n'est pas obligatoire pour avoir un phénomène de multicollinéarité. Un tel phénomène peut exister même si les coefficients de corrélation sont relativement faibles (inférieurs à 0.5 par exemple).

Test de Klein

La méthode proposée par Klein (1962) consiste simplement à comparer le coefficient de détermination (R^2) du modèle et les coefficients de détermination entre les variables explicatives considérées deux à deux (R_{ij}^2). Il y a présomption de multicollinéarité lorsque la plupart des R_{ij}^2 sont supérieurs à R^2 . La faiblesse de cette méthode vient du fait que la colinéarité peut impliquer plus de deux variables explicatives.

Test de Theil

Theil (1971) propose de calculer dans un modèle à k variables explicatives la statistique suivante :

$$THEIL = R^2 - \sum_{h=2}^k (R^2 - R_h^2) \quad (1.3)$$

où : R^2 est le coefficient de détermination du modèle ;

R_h^2 , le coefficient de détermination de la régression où la $h^{\text{ième}}$ variable explicative est exclue (à l'exception de la constante).

La statistique de Theil mesure donc la différence entre la contribution globale des k variables explicatives et la somme de leurs contributions individuelles. En l'absence de multicolinéarité, la statistique de Theil doit être nulle. Une valeur élevée est donc le signe de présomption de multicolinéarité.

Test de Farrar et Glauber

La technique proposée par Farrar et Glauber (1971) consiste à étudier si le déterminant de la matrice des coefficients de corrélation entre les variables explicatives est ou non proche de zéro. Si tel est le cas, il y a présomption de multicolinéarité. En effet, si les variables explicatives sont parfaitement corrélées, le déterminant de la matrice des coefficients de corrélation sera nul. Les auteurs ont suggéré de procéder à un test du Khi-carré visant à tester l'hypothèse nulle selon laquelle le déterminant de la matrice des coefficients de corrélation est égal à 1, signifiant que les variables sont orthogonales, contre l'hypothèse alternative selon laquelle le déterminant est inférieur à 1, indiquant que les variables explicatives sont indépendantes.

La première étape du test consiste à calculer le déterminant de la matrice des coefficients de corrélation entre les variables explicatives.

$$D = \begin{vmatrix} 1 & r_{x_1 r_{x_2}} & r_{x_1 r_{x_3}} & \cdots & r_{x_1 r_{x_k}} \\ r_{x_2 r_{x_1}} & 1 & r_{x_2 r_{x_3}} & \cdots & r_{x_2 r_{x_k}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{x_k r_{x_1}} & r_{x_k r_{x_2}} & r_{x_k r_{x_3}} & \cdots & 1 \end{vmatrix} \quad (1.4)$$

La deuxième étape consiste à effectuer un test du χ^2 , en posant l'hypothèse nulle suivante :

$$\begin{cases} H_0: D = 1 & \text{(les variables explicatives sont corrélées)} \\ H_1: D < 1 & \text{(les variables explicatives sont indépendantes)} \end{cases}$$

Sous l'hypothèse nulle, la statistique de test est donnée par la relation suivante :

$$FG = - \left[n - 1 - \frac{1}{6}(2K + 5) \right] \cdot \ln D \sim \chi^2_{\alpha; \frac{1}{2}K(K-1)} \quad (1.5)$$

où n est la taille de l'échantillon, K le nombre de variables explicatives (terme constant inclus) et D le déterminant de la matrice des coefficients de corrélation des k variables explicatives.

Les règles de décision sont alors les suivantes :

- Si $FG \geq \chi^2$ lu dans la table à $\frac{1}{2}K(K - 1)$ degrés de liberté et au seuil de α choisi, alors l'hypothèse H_0 est rejetée. Il y a donc présomption de multicolinéarité.
- Si $FG < \chi^2$, on accepte l'hypothèse nulle. L'hypothèse d'orthogonalité n'est pas rejetée. Donc, les variables explicatives sont indépendantes.

Remèdes à la multicolinéarité

Parmi les moyens permettant d'apporter des solutions au problème de multicolinéarité, on peut citer :

- Remplacer les variables explicatives par un nombre plus faible de combinaisons linéaires de ces variables ;
- Combiner les séries en coupes transversales et les séries temporelles, donc opter pour l'estimation en panel ;
- Transformer les variables en considérant soit les différences premières, soit les ratios ;
- Utiliser la « ridge regression » qui consiste à remplacer dans la formule de l'estimateur des moindres carrés la matrice (XX') par la matrice $(XX' + rI)$ avec r un scalaire choisi arbitrairement et I la matrice identité ;
- Augmenter la taille de l'échantillon. Mais cette technique n'est efficace que si l'ajout d'observations diffère significativement de celles figurant déjà dans le modèle. Dans le cas contraire, il y a risque de reconduction de la multicolinéarité ;
- Eliminer certaines variables explicatives pour remédier à la multicolinéarité. Mais cette pratique doit être évitée, car l'abandon a priori d'une variable peut entraîner un biais de spécification ou une erreur de spécification.

Remarque :

Lorsqu'il est impossible de remédier à la multicolinéarité, il faut alors privilégier les tests d'hypothèses jointes (F test) qui permettent d'évaluer la significativité globale des variables colinéaires.

I. Modèles à équations simultanée

Dans les modèles ne comportant qu'une seule équation, une hypothèse implicite supposait que la relation de cause à effet entre la variable dépendante (Y) et les variables explicatives (X) était unidirectionnelle : les variables explicatives étaient la cause et la variable dépendante l'effet. Or, nombre des théories économiques sont basées sur des modèles à plusieurs équations, c'est-à-dire des systèmes d'équations. Ces équations n'étant pas indépendantes les une des autres, l'interaction des différentes variables peut avoir des conséquences importantes au niveau de l'estimation de chacune des équations et du système d'équations dans son ensemble. Donc, on ne peut pas, sauf cas particulier, utiliser efficacement la méthode des moindres carrés ordinaires (MCO) équation par équation lorsqu'on cherche à tester des théories économiques décrivant un système complet d'équations pour

faire des prévisions simultanées sur un ensemble des variables liées.

I.1 : présentation du modèle à équations simultanées

I.1.1 : Exemples introductifs

A. Modèle de demande et d'offre

$$\left\{ \begin{array}{l} Q_{dt} = \alpha_0 + \alpha_1 P_t + \alpha_2 y_t + u_{dt} \end{array} \right. \quad (1)$$

$$\left\{ \begin{array}{l} Q_{0t} = \beta_0 + \beta_1 P_t + u_{0t} \end{array} \right. \quad (2)$$

$$\left\{ \begin{array}{l} Q_{dt} = Q_{0t} = Q_t \end{array} \right. \quad (3)$$

L'équation (1) est l'équation de demande avec Q_{dt} désignant la quantité demandée à la période courante, P_t le prix du bien à la période t et y_t , le revenu à la période t .

La relation (2) est l'équation d'offre où Q_{0t} désigne la quantité offerte du bien considéré à la période t , les termes d'erreurs sont représentés respectivement par u_{dt} et u_{0t} pour l'équation de demande et l'équation d'offre.

La relation (3) est une identité. Elle représente la condition d'équilibre du marché. Elle ne comporte pas de terme d'erreur.

B. Modèle macroéconomique de Klein

C.

Soit le modèle de Klein portant sur l'économie américaine sur la période 1920-1941 :

$$\left\{ \begin{array}{l} C_t = \alpha_0 + \alpha_1 \Pi_t + \alpha_2 \Pi_{t-1} + \alpha_3 (w_{1t} + w_{2t}) + u_{1t} \end{array} \right. \quad (4)$$

$$\left\{ \begin{array}{l} I_t = \beta_0 + \beta_1 \Pi_t + \beta_2 \Pi_{t-1} + \beta_3 K_{t-1} + u_{2t} \end{array} \right. \quad (5)$$

$$\left\{ \begin{array}{l} w_{1t} = \gamma_0 + \gamma_1 y_t + \gamma_2 y_{t-1} + \gamma_3 t + u_{3t} \end{array} \right. \quad (6)$$

$$\left\{ \begin{array}{l} Y_t = C_t + I_t + G_t \end{array} \right. \quad (7)$$

$$\left\{ \begin{array}{l} \Pi_t = Y_t + W_{1t} - W_{2t} - T_t \end{array} \right. \quad (8)$$

$$\left\{ \begin{array}{l} I_t = K_t - K_{t-1} \end{array} \right. \quad (9)$$

Où

C : désigne la consommation ;

Π : les profits ;

W_1 : les salaires versés par l'industrie (salaires du secteur privé) ;

W_2 : les salaires versés par l'administration (salaires du secteur public) ;

I : l'investissement net ;

K : le stock de capital ;

Y : le revenu national ;

G : les dépenses publiques ;
 T : les impôts ;
 T : le trend temporel ;
 U : le terme d'erreur.

L'équation (4) est l'équation de consommation, la relation (5) est l'équation d'investissement et la relation (6) exprime la demande du travail. Ces équations sont des équations de comportement. Les trois dernières équations sont des identités. La relation (7) définit le revenu comme la somme de la consommation privée, de l'investissement des entreprises et des dépenses publiques. Selon l'équation (8), le revenu national est égal à la somme des profits, des salaires et des impôts. Enfin, l'équation (9) définit l'investissement comme la variation stock de capital.

Le modèle de Klein est ainsi composé de 6 équations. Les variables endogènes du système sont : C_t , I_t , w_{1t} , Π_t et K_t . Quant aux variables prédéterminées, n distingue : d'une part, les variables endogènes décalées (Π_{t-1} , K_{t-1} , y_{t-1}) et, d'autre part, les variables exogènes courantes (W_{2t} , T_t , G_t et t).

I.1.2 : forme générale des modèles à équations simultanées

En général, on représente la forme structurelle du modèle à équations simultanées comme suit :

$$\beta_{i1}Y_{1t} + \beta_{i2}Y_{2t} + \dots + \beta_{iG}Y_{Gt} + \gamma_{i1}X_{1t} + \gamma_{i2}X_{2t} + \dots + \gamma_{iK}X_{Kt} = u_{it} \quad (10)$$

Pour tout $i = 1, 2, \dots, G$ et pour tout $t = 1, 2, \dots, n$; où les Y sont les variables endogènes et les X sont les variables prédéterminées.

De façon détaillée, la relation (10) peut s'écrire comme ci-après :

$$\begin{cases} \beta_{11}Y_{1t} + \beta_{12}Y_{2t} + \dots + \beta_{1G}Y_{Gt} + \gamma_{11}X_{1t} + \gamma_{12}X_{2t} + \dots + \gamma_{1K}X_{Kt} = u_{1t} \\ \beta_{21}Y_{1t} + \beta_{22}Y_{2t} + \dots + \beta_{2G}Y_{Gt} + \gamma_{21}X_{1t} + \gamma_{22}X_{2t} + \dots + \gamma_{2K}X_{Kt} = u_{2t} \\ \beta_{G1}Y_{1t} + \beta_{G2}Y_{2t} + \dots + \beta_{GG}Y_{Gt} + \gamma_{G1}X_{1t} + \gamma_{G2}X_{2t} + \dots + \gamma_{GK}X_{Kt} = u_{Gt} \end{cases} \quad (11)$$

Où $t = 1, 2, \dots, n$

Les équations de ce système, issues de la théorie économique, sont appelées des *équations structurelles ou de comportement*, car elles décrivent soit la structure d'une économie ou le comportement d'un agent économique.

Le système (11) comporte G équations et G variables endogènes (Y_1, Y_2, \dots, Y_G). il comprend K variables prédéterminées (exogènes courantes et décalées plus les endogènes décalées). Le modèle comporte aussi G termes d'erreurs ($u_{11}, u_{2t}, \dots, u_{Gt}$).

Dans chaque équation du système (11), une des variables endogènes a son coefficient associé égal à 1. Il s'agit de la variable dépendante. On parle donc de *normalisation*. En outre, les équations dans lesquelles tous les coefficients sont égaux à 1 et qui ne comporte pas de termes d'erreur sont les *identités*. Enfin, notons que l'une des variables peut prendre la valeur 1 afin de tenir compte du terme constant dans chacune des équations. Par conséquent, la relation (10) devient :

$$Y_{it} = -\gamma_{i1} + \sum_{l \neq i} (-\beta_{il}) Y_{lt} + \sum_{j=2}^K (-\gamma_{ij}) X_{jt} + u_{it} \quad (12)$$

Avec $i = 1, 2, \dots, G$; $l = 1, 2, \dots, G$

Sous sa forme matricielle, le système d'équation (11) s'écrit :

$$\underbrace{B}_{(G \times G)} \underbrace{Y}_{(G \times i)} + \underbrace{\Gamma}_{(G \times k)} \underbrace{X}_{(k \times i)} = \underbrace{U}_{(G \times i)} \quad (13)$$

Avec :

$$B = \begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1G} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2G} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \beta_{G1} & \beta_{G2} & \dots & \beta_{GG} \end{bmatrix} ; \quad Y = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \cdot \\ \cdot \\ \cdot \\ y_{Gt} \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1K} \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2K} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \gamma_{G1} & \gamma_{G2} & \dots & \gamma_{GG} \end{bmatrix} ; \quad X = \begin{bmatrix} x_{1t} \\ x_{2t} \\ \cdot \\ \cdot \\ \cdot \\ x_{Kt} \end{bmatrix} ; \quad U = \begin{bmatrix} u_{1t} \\ u_{2t} \\ \cdot \\ \cdot \\ \cdot \\ u_{Gt} \end{bmatrix}$$

Si la matrice B est non singulière, elle est inversible et il est possible de dériver la forme réduite du modèle (13) permettant ainsi d'exprimer la matrice Y en fonction de la matrice X :

$$Y = -B^{-1}\Gamma X + B^{-1}U \quad (14)$$

La condition selon laquelle la matrice B est non singulière est appelée condition de complétude.

La relation (15) peut s'écrire aussi :

$$Y = AX + E \quad (15)$$

Où $A = -B^{-1}\Gamma$ et $E = B^{-1}U$

La relation (15) représente le système d'équation (13) sous sa forme réduite. Celle-ci permet d'exprimer chaque variable endogène en fonction des variables prédéterminées et de termes d'erreur.

Sous cette forme, on peut, sous certaines conditions à préciser plus loin, estimer la matrice A. mais il faudra, pour obtenir des estimateurs ayant une signification économique, revenir aux matrices B et C. donc, le problème qui se pose est de déterminer les paramètres de la forme structurelle à partir des estimations obtenues sur la forme réduite. C'est le problème de l'identification.

Lorsque ces conditions sont réalisées, on peut alors estimer chaque équation du modèle de la forme réduite par les MCO. Les estimateurs obtenus par les MCO sont sans biais et à variance minimale.

On obtient ainsi une matrice \hat{A} dont tous les éléments sont des estimations sans biais des paramètres de la forme réduite et qui converge en probabilité vers les vraies valeurs de ces paramètres.

En effet, dans ces équations, les variables endogènes sont exprimées en fonction de variables prédéterminées en fonction des variables prédéterminées, supposées non corrélées avec les termes d'erreurs.

2. Test de simultanéité de Hausman

Le problème de la simultanéité se pose parce que quelques régresseurs sont endogènes et, donc, susceptibles d'être corrélés avec les termes d'erreur. S'il y a simultanéité, les estimateurs des MCO sont biaisés et non convergents, et les méthodes alternatives (doubles moindres carrés, par exemple) peuvent fournir des estimateurs cohérents et performants. En l'absence de simultanéité, les estimateurs MCO produisent des résultats cohérents et performants. L'application des méthodes alternatives lorsqu'il n'y a pas simultanéité fournit des estimateurs non cohérents mais performants (à savoir avec une variance minimale).

Pour vérifier le problème de simultanéité le test d'erreur de spécification d'Hausman. Examinons les étapes du test d'Hausman à l'aide d'un modèle simple.

Soit le modèle à équations simultanées suivant :

$$\text{Fonction de demande : } Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + \alpha_3 R_t + u_{1t} \quad (1)$$

$$\text{Fonction d'offre : } Q_t = \beta_0 + \beta_1 P_t + u_{2t} \quad (2)$$

Où P : le prix ;

Q : la quantité ;

I : le revenu

R : la richesse ;

U : termes de l'erreur.

Supposons que I et R soient exogènes et que P et Q soient endogènes.

Considérons la fonction d'offre (2). S'il n'y a pas de problème de simultanéité (c'est-à-dire si P et Q sont mutuellement indépendants), P_t et u_{2t} devraient être non corrélés. D'autre part, s'il y a simultanéité, P_t et u_{2t} seront corrélés.

La démarche consiste, avant tout, à obtenir la forme réduite du système d'équations structurelles :

$$\begin{cases} P_t = \Pi_0 + \Pi_1 I_t + \Pi_2 R_t + v_t \end{cases} \quad (3)$$

$$\begin{cases} Q_t = \Pi_3 + \Pi_4 I_t + \Pi_5 R_t + w_t \end{cases} \quad (4)$$

Où v_t et w_t sont les termes d'erreur de forme réduite.

Ensuite, il faut régresser l'équation (3) par les MCO pour obtenir \hat{P}_t et \hat{v}_t , car

$$P_t = \hat{P}_t + \hat{v}_t;$$

L'étape suivante consiste à régresser l'équation structurelle (2) en remplaçant P_t par \hat{P}_t et \hat{v}_t . Donc, effectuer la régression de :

$$Q_t = \gamma_0 + \gamma_1 \hat{P}_t + \gamma_2 \hat{v}_t + u_{2t} \quad (5)$$

Enfin, faire un test t sur le coefficient de \hat{v}_t . S'il est significatif, ne pas rejeter l'hypothèse de simultanéité ; dans le cas inverse, le rejeter.

Il convient de remarquer ici que :

- Certains auteurs suggèrent de régresser, dans le cas présent, Q_t sur P_t car le régresseur est P_t et non \hat{P}_t ;
- S'il y a plus d'un régresseur endogène, il faut utiliser le test F.
-

Autre exemple

L'étude du comportement des dépenses publiques à l'aide du modèle à équations simultanées suivant :

$$\begin{cases} \text{DEP}_t = \beta_0 + \beta_1 \text{SUB}_t + \beta_2 \text{IMP}_t + \beta_3 \text{POP}_t + u_{1t} \end{cases} \quad (1)$$

$$\begin{cases} \text{SUB}_t = \delta_0 + \delta_1 \text{DEP}_t + \delta_2 \text{PS}_t + u_{2t} \end{cases} \quad (2)$$

Où

DEP : dépenses publiques ;

SUB : niveau des subventions ;

IMP : impôts ;

POP : population ;

PS : population des enfants scolarisés dans le primaire et le secondaire ;

Dans ce modèle, les variables IMP, POP et PS sont considérées comme exogènes.

Le test d'Hausman implique donc les étapes suivantes :

Etape 1 : Régresser SUB sur IMP, POP et PS (c'est-à-dire la régression de forme réduite) ; et calculer le résidu estimé de cette régression \hat{w}_{1t}

Etape 2 : Régresser ensuite l'équation structurelle DEP_t sur SUB_t , IMP_t , POP_t et \hat{w}_{1t} .

$$\text{Soit : } \text{DEP}_t = \gamma_0 + \gamma_1 \text{SUB}_t + \gamma_2 \text{IMP}_t + \gamma_3 \text{POP}_t + \gamma_4 \hat{w}_{1t} + u_{1t} \quad (3)$$

Ou encore :

$$\text{DEP}_t = \alpha_0 + \alpha_1 \text{SUB}_t + \alpha_2 \text{IMP}_t + \alpha_3 \text{POP}_t + \alpha_4 \hat{w}_{1t} + u_{1t} \quad (4)$$

Etape 3 : faire le test t sur le coefficient de \hat{w}_{1t} .

Exemple 3 : soit un modèle d'équations simultanées ayant 3 équations et 3 variables endogènes : Y_1 , Y_2 et Y_3

Admettons que la première équation du modèle soit :

$$Y_{1t} = \beta_0 + \beta_1 Y_{2t} + \beta_2 Y_{3t} + \beta_3 Y_{4t} \quad (1)$$

Les étapes de calcul :

- Régresser les équations de forme réduite Y_{2t} et Y_{3t} . a partir de ces régressions, on obtient : \hat{Y}_{2t} , \hat{Y}_{3t} , \hat{w}_{2t} et \hat{w}_{3t}
- Estimer l'équation (1) modifiée par les MCO :

$$Y_{1t} = \alpha_0 + \alpha_1 \hat{Y}_{2t} + \alpha_2 \hat{Y}_{3t} + \alpha_3 X_{1t} + \gamma_1 \hat{W}_{2t} + \gamma_2 \hat{W}_{3t} + u_{1t} \quad (2)$$

- En utilisant le test F, on teste l'hypothèse que $\gamma_1 = \gamma_2 = 0$. Si l'hypothèse est rejetée, on conclut qu'il y a simultanéité.

I.2. Problème de l'identification des équations

I.2.1. Notion

Le problème de l'identification consiste à savoir s'il est possible de déduire les coefficients de la forme structurelle à partir des estimations numériques des paramètres de la forme réduite.

- s'il est impossible d'obtenir des estimations des paramètres de la forme structurelle à partir des coefficients de la forme réduite, le modèle est dit non identifié ou sous-identifié. un modèle est sous-identifié si une équation du modèle est sous identifiable. cela signifie que le nombre d'équations est plus faible que le nombre des paramètres à identifier dans la forme structurelle et il alors impossible se résoudre le système.
- si des valeurs numériques des paramètres structurels peuvent être obtenues à partir des coefficients de la forme réduite, on dit que le modèle est identifié. une équation identifiée peut être soit exactement identifié ou sur-identifiée.

On dit qu'une équation est exactement (ou pleinement ou strictement ou juste) identifiée si des valeurs numériques uniques des paramètres structures peuvent être obtenues. Le modèle est exactement identifié si toutes ses équations sont strictement identifiables.

- elle est dite sur-identifiée si plusieurs valeurs numériques peuvent correspondre aux paramètres structurels. le modèle est sur-identifié si les équations sont sur-identifiables.

I.2.2. Restrictions sur les coefficients

s'il est généralement admis que le problème de l'identification peut être examiné à partir de la forme réduite du modèle à équations simultanées, la connaissance de la forme réduite ne permet pas toujours de déterminer la forme structurelle du modèle. en effet, à une forme structurelle, sans contrainte sur les paramètres, correspond une forme réduite, mais la réciproque n'est pas vraie. si la structure générale :

$$BY_t + \Gamma X_t = U_t \quad (1)$$

est multipliée par une matrice F non singulière :

$$(FB)Y_t + (F\Gamma)X_t = FU_t \quad (2)$$

alors la forme réduite du modèle n'est modifiée :

$$\begin{aligned}
 Y_t &= B^{-1}F\Gamma X_t + B^{-1}F^{-1}FU_t \\
 &= -B^{-1}\Gamma X_t + B^{-1}U_t
 \end{aligned}
 \tag{3}$$

Donc, si l'on ne dispose d'aucune information supplémentaire liée à la théorie économique, l'identification est impossible. La théorie économique conduit à imposer des contraintes sur les paramètres de la forme structurelle du modèle qui réduisent le nombre de paramètres à estimer.

Il y a une restriction sur un coefficient de la forme structurelle, chaque fois qu'un paramètre est contraint, par l'écriture du modèle, à être égal à une valeur déterminée. L'on retient généralement deux types de restrictions :

A. les restrictions d'exclusion

Le fait ne pas introduire une variable dans une des équations du système est considéré comme une relation d'exclusion. Cela revient en effet à affecter la variable en question d'un coefficient nul. En d'autres termes, cela équivaut à placer des zéros dans les éléments de la matrice B et/ou Γ . Une telle procédure permet de réduire le nombre des paramètres inconnus et fournit ainsi une aide à l'identification.

B. Les restrictions linéaires

Certaines spécifications des modèles imposent que des variables soient affectées d'un coefficient identique ou qu'une combinaison linéaire des coefficients prennent une valeur précise. En imposant de telles restrictions sur les paramètres, on facilite la procédure d'estimations en diminuant le nombre des paramètres inconnus.

I.2.3. Règles de l'identification

On détermine l'identification d'une équation, on applique les *conditions d'ordre et de rang de l'identification*. Ces conditions cherchent à déterminer si le nombre des contraintes sur les paramètres du système d'équations structurelles est suffisant pour permettre de déduire les coefficients de la forme structurelle à partir des estimations numériques des paramètres de la forme réduite.

Soit le modèle à équations simultanées :

$$\underbrace{B}_{(G \times G)} \underbrace{Y}_{(G \times i)} + \underbrace{\Gamma}_{(G \times k)} \underbrace{X}_{(k \times i)} = \underbrace{U}_{(G \times i)} \tag{1}$$

Où G : nombre de variables endogènes dans le modèle ;

g : nombre de variables endogènes dans une équation donnée ;

K : nombre de variables prédéterminées dans le modèle y compris la constante (intercept) ;

k : nombre de variables prédéterminées dans une équation donnée.

A. Condition d'ordre de l'identification

Dans un modèle à G équations simultanées, pour identifier une équation, le nombre de variables prédéterminées exclues de l'équation ne doit pas être inférieur au nombre de variables endogènes incluses dans cette équation moins 1, c'est-à-dire :

$$K - k \geq g - 1 \quad (2)$$

- Si $K - k = g - 1$, l'équation est exactement identifiée ;
- Si $K - k > g - 1$, elle est sur-identifiée ;
- Si $K - k < g - 1$, elle est sous-identifiée.

Notons que la condition d'ordre de l'identification est une condition nécessaire mais non satisfaisante pour identifier une équation ce qui signifie que, même si elle est satisfaisante, il peut se faire qu'une équation ne soit pas identifiée.

Exemples

Ex.1 : soit le modèle à équations simultanées suivant :

$$\begin{cases} Q_t = \alpha_0 + \alpha_1 P_t + u_{1t} \end{cases} \quad (1)$$

$$\begin{cases} Q_t = \beta_0 + \beta_1 P_t + u_{2t} \end{cases} \quad (2)$$

Où Q_t : quantité ; P_t : prix et u : le terme d'erreur

Le modèle comporte deux variables endogènes : Q_t et P_t , et aucune variable prédéterminées.

Equation	Nombre des variables prédéterminées exclues ($K - k$)	Nombre des variables endogènes incluses moins 1 ($g - 1$)	Décision
(1)	0	$2 - 1 = 1$	Sous-id
(2)	0	$2 - 1 = 1$	Sous-id

Aucune équation du modèle n'est identifiée :

Ex.2 : soit le modèle :

$$\begin{cases} Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 R_t + u_{1t} \end{cases} \quad (1)$$

$$\begin{cases} Q_t = \beta_0 + \beta_1 P_t + u_{2t} \end{cases} \quad (2)$$

Où Q_t et P_t : variables endogènes ;

R_t : variable prédéterminée.

Equation	($K - k$)	($g - 1$)	Décision
(1)	$1 - 1 = 0$	$2 - 1 = 1$	Sous-id
(2)	$2 - 1 = 1$	$2 - 1 = 1$	Sous-id

Ex.3 : soit le modèle :

$$\begin{cases} Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 R_t + u_{1t} & (1) \\ Q_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t} & (2) \end{cases}$$

Où Q_t et P_t : variables endogènes ;

R_t et P_{t-1} : variables prédéterminées.

Equation	(K - k)	(g - 1)	Décision
(1)	2 - 1 = 1	2 - 1 = 1	Ex-id
(2)	2 - 1 = 1	2 - 1 = 1	Ex-id

Chaque équation est identifiée par les conditions d'ordre. Dans son ensemble, le modèle est donc exactement identifié.

Ex.4 : soit le modèle :

$$\begin{cases} Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 R_t + \alpha_3 S_t + u_{1t} & (1) \\ Q_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t} & (2) \end{cases}$$

Où Q_t et P_t : variables endogènes ;

R_t , P_{t-1} et S_t : variables prédéterminées.

Equation	(K - k)	(g - 1)	Décision
(1)	2 - 1 = 1	2 - 1 = 1	Ex-id
(2)	2 - 1 = 1	2 - 1 = 1	Ex-id

Il s'ensuit qu'une estimation unique de tous les paramètres du modèle est impossible. En effet, la forme réduite du modèle est donc :

$$\begin{cases} P_t = \Pi_{10} + \Pi_{11} R_t + \Pi_{12} S_t + \Pi_{13} P_{t-1} + u_{1t} \\ Q_t = \Pi_{20} + \Pi_{21} R_t + \Pi_{22} S_t + \Pi_{23} P_{t-1} + u_{2t} \end{cases}$$

Où

$$\Pi_{10} = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} ; \Pi_{11} = -\frac{\alpha_2}{\alpha_1 - \beta_1} ;$$

$$\Pi_{12} = -\frac{\alpha_3}{\alpha_1 - \beta_1} ; \Pi_{13} = \frac{\beta_2}{\alpha_1 - \beta_1} ;$$

$$\Pi_{20} = \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1} ; \Pi_{21} = -\frac{\alpha_2 \beta_1}{\alpha_1 - \beta_1} ;$$

$$\Pi_{22} = \frac{\alpha_3 \beta_1}{\alpha_1 - \beta_1} ; \Pi_{23} = -\frac{\alpha_1 \beta_1}{\alpha_1 - \beta_1} ;$$

$$v_{1t} = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1} \text{ et } v_{2t} = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}$$

A partir des coefficients de forme réduite ci-haut, on obtient deux estimations du coefficient β_1 :

$$\beta_1 = \frac{\Pi_{22}}{\Pi_{12}} \text{ et } \beta_1 = \frac{\Pi_{21}}{\Pi_{11}}$$

Puisque rien ne garantit que ces deux estimations seront identiques et que le paramètre β_1 apparaît aux dénominateurs de tous les coefficients de la forme réduite, l'ambiguïté dans l'estimation de β_1 sera transmise aux autres estimations.

B. Condition de rang d'identification

Contrairement à la condition d'ordre et l'identification, la condition de rang est suffisante. Cette condition assure qu'il existe une correspondance un à un entre les coefficients structurels et ceux de forme réduite. En conséquence, il faut à la fois une condition nécessaire et suffisante pour l'identification.

Les étapes à suivre pour appliquer la condition de rang d'identification sont :

- a) Ecrire le système d'équations simultanées sous forme de tableau des coefficients des variables endogènes et prédéterminées.
- b) Supprimer les coefficients de la ligne sur laquelle l'équation considérée apparaît ;
- c) Supprimer aussi les colonnes des éléments non nuls de cette ligne (de l'équation considérée) ;
- d) A partir des coefficients restant du tableau, construire toutes les matrices possibles, telles que P d'ordre $(G - 1)$ et calculer les déterminants correspondants. Si au moins un des déterminants de P est non nul, l'équation considérée est exactement sur-identifiée. Si les déterminants possibles d'ordre $(G - 1)$ sont nuls, alors le rang de la matrice P est inférieur à $(G - 1)$ et l'équation considérée n'est pas identifiée.

Les conditions d'ordre et de rang de l'identification conduit aux principaux généraux suivants pour identifier une équation structurelle dans un système de G équations simultanées :

1. Si $K - k > g - 1$ et si le rang de la matrice P d'ordre $(G - 1)$ est maximum, alors l'équation considérée est sur-identifiée ;
2. Si $K - k = g - 1$ et si le rang de la matrice P d'ordre $(G - 1)$ est maximum, alors l'équation considérée est exactement identifiée ;
3. Si $K - k \geq g - 1$ et si le rang de la matrice P d'ordre $(G - 1)$ est maximum, alors l'équation considérée est sou-identifiée ;
4. Si $K - k < g - 1$ si l'équation structurelle n'est pas identifiée. Dans ce cas, le rang de la matrice P est inférieur à $(G - 1)$.

Exemple de la condition de rang

Soit le modèle à équations simultanée suivantes :

$$\begin{cases} Y_{1t} = \beta_{10} + \beta_{12}Y_{2t} + \beta_{13}Y_{3t} + \gamma_{11}X_{1t} + u_{1t} & (1) \\ Y_{2t} = \beta_{20} + \beta_{23}Y_{3t} + \gamma_{21}X_{1t} + \gamma_{22}X_{2t} + u_{2t} & (2) \\ Y_{3t} = \beta_{30} + \beta_{31}Y_{1t} + \gamma_{31}X_{1t} + \gamma_{32}X_{2t} + u_{3t} & (3) \\ Y_{4t} = \beta_{40} + \beta_{41}Y_{1t} + \beta_{42}Y_{2t} + \gamma_{42}X_{3t} + u_{4t} & (4) \end{cases}$$

Sous sa forme générale, le modèle à équations simultanées peut s'écrire comme suit :

$$BY + \Gamma X = U \quad (5)$$

D'où :

$$\begin{cases} Y_{1t} - \beta_{10} - \beta_{12}Y_{2t} - \beta_{13}Y_{3t} - \gamma_{11}X_{1t} = u_{1t} & (1)' \\ Y_{2t} - \beta_{20} - \beta_{23}Y_{3t} - \gamma_{21}X_{1t} - \gamma_{22}X_{2t} = u_{2t} & (2)' \\ Y_{3t} - \beta_{30} - \beta_{31}Y_{1t} - \gamma_{31}X_{1t} - \gamma_{32}X_{2t} = u_{3t} & (3)' \\ Y_{4t} - \beta_{40} - \beta_{41}Y_{1t} - \beta_{42}Y_{2t} - \gamma_{42}X_{3t} = u_{4t} & (4)' \end{cases}$$

Pour faciliter l'identification, écrivons ce système dans le tableau des coefficients des variables :

Equation n°	1	Y_1	Y_2	Y_3	Y_4	X_1	X_2	X_3
(1)	$-\beta_{10}$	1	$-\beta_{12}$	$-\beta_{13}$	0	$-\gamma_{11}$	0	0
(2)	$-\beta_{20}$	0	1	$-\beta_{23}$	0	$-\gamma_{21}$	$-\gamma_{22}$	0
(3)	$-\beta_{30}$	$-\beta_{31}$	0	1	0	$-\gamma_{31}$	$-\gamma_{32}$	0
(4)	$-\beta_{40}$	$-\beta_{41}$	$-\beta_{42}$	0	1	0	0	$-\gamma_{43}$

- Vérifions la condition d'ordre de l'identification :

Equation n°	$K - k$	$g - 1$	Décision
(1)	$3 - 1 = 2$	$3 - 1 = 2$	Exactement identifiée
(2)	$3 - 2 = 1$	$2 - 1 = 1$	Exactement identifiée
(3)	$3 - 2 = 1$	$2 - 1 = 1$	Exactement identifiée
(4)	$3 - 1 = 2$	$3 - 1 = 2$	Exactement identifiée

La condition d'ordre de l'identification montre que chaque équation du système est exactement identifiée.

- Appliquons la condition de rang

Considérons la première équation qui exclut Y_4 , X_2 et X_3 . Pour que cette

équation soit identifiée, il faut obtenir au moins un déterminant non nul d'ordre (3×3) à partir des coefficients des variables exclues de cette équation mais présentes dans d'autres.

La matrice P associée à la première équation est donc :

$$P = \begin{vmatrix} 0 & -\gamma_{22} & 0 \\ 0 & -\gamma_{32} & 0 \\ 1 & 0 & -\gamma_{43} \end{vmatrix} \quad (4)$$

Et son déterminant :

$$|P| = \begin{vmatrix} 0 & -\gamma_{22} & 0 \\ 0 & -\gamma_{32} & 0 \\ 1 & 0 & -\gamma_{43} \end{vmatrix} = 0 \quad (5)$$

Puisque le déterminant est nul, le rang de la matrice P est inférieur à 3. En conséquence, l'équation (1) ne satisfait pas la condition de rang et n'est pas identifiée (elle est sous-identifiée).

On peut vérifier que, d'après la condition de rang, les équations (2) et (3) sont aussi non identifiées, mais que l'équation (4) l'est.

En conclusion, la condition de rang indique si l'équation considérée est ou non identifiée, alors que la condition d'ordre si elle est exactement identifiée ou sur-identifiée.

Conditions d'identification à des restrictions linéaires sur les paramètres

Pour établir les conditions d'identification évoquées ci-haut, on avait considéré uniquement des relations d'exclusion. Si l'on a en plus des restrictions linéaires sur les paramètres, la condition d'ordre devient :

$$(K - k) + r \geq g - 1$$

Où P désigne le nombre de restrictions autres que celles d'exclusion. Si par exemple, pour une raison quelconque deux coefficients doivent être égaux, on a une restriction et on pose $r = 1$.

- Si $(K - k) + r > g - 1$, alors l'équation est sur-identifiée ;
- Si $(K - k) + r = g - 1$, alors l'équation est juste identifiée ;

- Si $(K - k) + r < g - 1$, alors l'équation est sous-identifiée ;

Les conditions nécessaires et suffisantes de l'identification deviennent alors :

- Si $(K - k) + r < g - 1$, ou si la condition de rang n'est pas vérifiée, l'équation est sous-identifiée ;
- Si $(K - k) + r = g - 1$, et que la matrice de rang est vérifiée, l'équation est exactement identifiée ;
- Si $(K - k) + r > g - 1$, et que la condition de rang est vérifiée, l'équation est sur-identifiée.

Méthodes d'estimation des modèles à équations simultanées

Les méthodes d'estimation des modèles à équations simultanées sont fonction de critères d'identifications des modèles. Seules les modèles

exactes identifiés ou sur-identifiés sont estimables.

Si le modèle est sous-identifié, il n'y a pas d'estimation possible. Dans ce cas, il faut réexaminer le modèle en vue d'en proposer un nouveau pour une meilleure spécification. En effet, la sous-identification signifie qu'il y a moins d'équation que les paramètres à identifier dans la forme structurelle.

Dans le cas d'un modèle exactement identifié ou sur identifié, on peut adopter deux approches à l'estimation des équations structurelles, à savoir les méthodes à équations unique, appelées aussi *méthodes à information limitée*. Et les méthodes des systèmes, nommées aussi *méthodes à information complète*.

Dans les méthodes à équation unique, on estime chaque équation du système d'équation du système d'équations simultanées individuellement tout en négligeant l'information contenue dans les autres équations du système; d'où le nom de *méthodes à information limitée*. Dans cette catégorie figurent les méthodes des *moindres carrés indirects*, des doubles moindres carrés et méthodes du maximum de vraisemblance à information limitée.

Au contraire, dans les ms simultanément toutes les méthodes des systèmes, on estime simultanément toutes les équations du système en utilisant toute l'information contenue dans l'ensemble des équations du système (c'est-à-dire toutes les restrictions sur ces équations par omission ou absence de certaines variables); d'où le nom de *méthodes à information complète*. Dans cette catégorie figurent les méthodes des triples moindres carrés, du maximum de vraisemblance à information complète ou encore la méthode de moments généralisés de systèmes d'équations.

Pour maintenir le caractère simultané des équations du modèle, on devrait idéalement utiliser les méthodes des systèmes. Cependant, en pratique, de telles méthodes sont rarement utilisées pour essentiellement trois raisons :

- la lourdeur des calculs ;
 - l'existence de solutions non linéaires. s'il
 - la sensibilité aux erreurs de spécification. s'il existe une erreur de spécification (due à une mauvaise forme fonctionnelle ou l'exclusion des variables pertinentes) dans une ou plusieurs équations du système, cette erreur est transmise dans le reste du système.
 - en conséquence, les méthodes à équations unique sont souvent utilisées.
 -
1. Méthodes des moindres carrés indirects

La méthode des moindres carrés indirects (MCI) consiste à appliquer la méthode des moindres carrés ordinaires aux équations justes identifiables (exactement identifiées).

Les étapes indispensables pour l'application de cette méthode sont :

- a) transformer les équations structurelles en équation de forme réduite

$$By + \Gamma X = U \quad (1)$$

$$y = -B^{-1}\Gamma X + B^{-1}U \quad (2)$$

$$\text{Où } \Pi = -B^{-1}\Gamma \text{ et } V = B^{-1}U$$

- b) Estimer les paramètres par la méthode des moindres carrés de chaque équation de la forme réduite ;
- c) obtenir les estimations des coefficients structurels (\hat{B} et $\hat{\Gamma}$) à partir des coefficients de la forme réduites ($\hat{\Pi}$).

Comme le montre cette procédure en trois étapes, le nom de moindres carrés indirects provient du fait que les coefficients structurels (brut principale de l'analyse) sont indirectement obtenus des estimations par les moindres carrés ordinaires des coefficients de forme réduite.

Le seul inconvénient de cette méthode est le retour, à partir de la matrice Π , aux matrices des coefficients structurels B et Γ . Les calculs sont souvent fastidieux, à moins de considérer de modèles très simples.

Exemple

Soit le modèle d'offre et de demande suivant :

Fonction de demande : $Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 X_t + u_{1t}$

(1)

Fonction d'offre : $Q_t = \beta_0 + \beta_1 P_t + u_{2t}$

(2)

où Q = quantité ;

P = pris ;

X = revenu ou dépenses de consommation ;

P et Q : variables endogènes ;

X : variable prédéterminée.

- les équations de forme réduite correspondant aux équations structurelles du modèle considéré sont :

$$\begin{cases} P_t = \Pi_{10} + \Pi_{11}X_t + v_{1t} \end{cases} \quad (3)$$

$$\begin{cases} Q_t = \Pi_{20} + \Pi_{21}X_t + v_{2t} \end{cases} \quad (4)$$

où

$$\Pi_{10} = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} ; \Pi_{11} = \frac{\alpha_2}{\alpha_1 - \beta_1} ; v_{1t} = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1} ;$$

$$\Pi_{20} = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} ; \Pi_{21} = \frac{\alpha_2\beta_1}{\alpha_1 - \beta_1} ; v_{1t} = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1} ;$$

- appliquons les moindres carrés ordinaires aux équations de forme réduite, prises individuellement. d'où

- la deuxième étape consiste à remplacer les variables endogènes figurant à droite des équations structurelles par leurs valeurs estimées dans le modèle (1).

$$\begin{cases} y_{1t} = b_{12}\hat{y}_{2t} + \dots + b_{1G}\hat{y}_{Gt} + c_{11}X_{1t} + \dots + c_{1K}X_{Kt} + u_{1t} \\ y_{2t} = b_{22}\hat{y}_{1t} + \dots + b_{2G}\hat{y}_{Gt} + c_{21}X_{1t} + \dots + c_{2K}X_{Kt} + u_{2t} \\ \vdots \\ y_{Gt} = b_{G1}\hat{y}_{1t} + \dots + b_{GG}\hat{y}_{Gt} + c_{G1}X_{1t} + \dots + c_{GK}X_{Kt} + u_{Gt} \end{cases} \quad (3)$$

Les estimateurs ainsi obtenus sont cohérents : ils convergent vers leurs vraies valeurs à mesure que la taille de l'échantillon croît indéfiniment.

Remarques

- les MDC peuvent s'appliquer à une équation individuelle du système sans prendre directement en compte aucune autre équation de ce système. d'où, pour résoudre des modèles économétriques comportant un grand nombre d'équations, les MDC offrent une méthode économique.
- à la différence des MCI, qui fournissent de nombreuses estimations des paramètres dans les équations sur-identifiées, les MDC ne procurent qu'une estimation par paramètre.
- ils sont facilement applicables car on a seulement besoin de connaître le nombre total des variables exogènes dans le système sans être informé sur d'autres variables du système.
- bien que spécialement connue pour traiter les équations sur-identifiées, la méthode peut aussi s'appliquer à celles qui sont exactement identifiées. mais, dans ce dernier cas, les MCI et les DMC donnent des estimations identiques.

Sous forme matricielle, ce modèle peut s'écrire :

$$\begin{pmatrix} 1 & 0 & 0 \\ -\beta_{21} & 1 & 0 \\ -\beta_{31} & -\beta_{32} & 1 \end{pmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} + \begin{bmatrix} -\gamma_{10} & -\gamma_{11} & -\gamma_{12} \\ -\gamma_{20} & -\gamma_{21} & -\gamma_{22} \\ -\gamma_{30} & -\gamma_{33} & -\gamma_{32} \end{bmatrix} \begin{bmatrix} 1 \\ X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix} \quad (3)$$

Lorsqu'un modèle à équations simultanées répond au critère de récursivité, il est recommandé d'appliquer la méthode des MCO équation par équation. En effet, dans ce cas particulier, il y a indépendance entre les variables endogènes et les erreurs. Par exemple, dans la deuxième équation, y_{1t} dépend de u_{1t} mais pas de u_{2t} ;

Cas particulier : les modèles récursifs

Un système d'équations est appelé récursif si chaque variable endogène peut être déterminée de manière séquentielle :

$$\begin{cases} y_{1t} = f(X_{1t}, X_{2t}, \dots + X_{Kt}; u_{1t}) \\ y_{2t} = f(X_{1t}, X_{2t}, \dots + X_{Kt}; y_{1t} + u_{2t}) \\ y_{3t} = f(X_{1t}, X_{2t}, \dots + X_{Kt}; y_{1t}, y_{2t} + u_{3t}) \\ \vdots \\ y_{Gt} = f(X_{1t}, \dots, X_{Kt}; y_{1t}, \dots, y_{G-1,t}; y_{Gt}; u_{1t}) \end{cases} \quad (1)$$

Le modèle récurrents sont aussi appelés *système triangulaire* car les coefficients des variables endogènes forment un triangle au sein de la matrice B. la variable y_i est alors expliquée par des variables de rang inférieur. Ainsi, la première équation ne contient aucune variable endogène. La deuxième équation a pour variable explicative endogène, la variable de la première équation, et ainsi de suite.

Considérons l'illustration d'un modèle récurrent :

$$\begin{cases} y_{1t} = \gamma_{10} + \gamma_{11}X_{1t} + \gamma_{12}X_{2t} + u_{1t} \\ y_{2t} = \gamma_{20} + \beta_{21}y_{1t} + \gamma_{21}X_{1t} + \gamma_{22}X_{2t} + u_{2t} \\ y_{3t} = \gamma_{30} + \beta_{31}y_{1t} + \beta_{32}y_{2t} + \gamma_{31}X_{1t} + \gamma_{32}X_{2t} + u_{3t} \end{cases} \quad (2)$$