

SAVOIR PLUS
UNIVERSITÉS

BERNARD HAUDEVILLE

Économétrie
appliquée



ESTEM



Économétrie appliquée

Économétrie appliquée

ISBN 2-909455-50-5

© 1996 Éditions ESTEM

Toute représentation ou reproduction, intégrale ou partielle, faite sans le consentement de l'auteur, ou de ses ayants droit ou ayants cause, est illicite (loi du 11 mars 1957, alinéa 1^{er} de l'article 40). Cette représentation ou reproduction par quelque procédé que ce soit, constituerait une contrefaçon sanctionnée par les articles 425 et suivant du Code Pénal.

ESTEM Éditions Scientifiques, Techniques et Médicales

5, rue Rousselet, 75007 Paris

Tél. : 33 (1) 42 19 05 11 - Fax : 33 (1) 42 19 05 24

Économétrie appliquée

Bernard Haudeville

*Agrégé de Sciences Économiques
Professeur à l'université d'Orléans*

La diffusion scientifique et technique est un facteur essentiel du développement. Aussi, dès 1988, l'Agence francophone pour l'enseignement supérieur et la recherche (AUPELF-UREF), mandatée par les Sommets francophones pour produire et diffuser revues et livres scientifiques, a créé la collection *Universités francophones*.

Lieu d'expression de la communauté scientifique de langue française, *Universités francophones* vise à instaurer une collaboration entre enseignants et chercheurs francophones en publiant des ouvrages, coédités avec des éditeurs francophones, et largement diffusés dans les pays du Sud, grâce à une politique tarifaire préférentielle.

Quatre séries composent la collection :

- Les usuels : cette série didactique est le cœur de la collection. Elle s'adresse à un public étudiant et vise à constituer une bibliothèque de référence couvrant les principales disciplines enseignées à l'université.
- Actualité scientifique : dans cette série sont publiés les actes de colloques organisés par les réseaux thématiques de recherche de l'UREF.
- Prospectives francophones : s'inscrivent dans cette série des ouvrages de réflexion donnant l'éclairage de la Francophonie sur les grandes questions contemporaines.
- Savoir plus Université : cette nouvelle série, dans laquelle s'inscrit le présent ouvrage, se compose de livres de synthèse qui font un point précis sur des sujets scientifiques d'actualité.

Notre collection, en proposant une approche plurielle et singulière de la science, adaptée aux réalités multiples de la Francophonie, contribue efficacement à promouvoir l'enseignement supérieur et la recherche dans l'espace francophone et le plurilinguisme dans la recherche internationale.

Professeur MICHEL GUILLOU
Directeur général de l'APELF
Recteur de l'UREF

Table des matières

Introduction	1
Chapitre I. L'estimation d'une équation unique	7
Section 1 : L'estimation par les moindres carrés ordinaires dans le cas général	13
Section 2 : La collinéarité des variables explicatives	24
Section 3 : L'autocorrélation des termes d'erreur	31
Section 4 : L'hétéroscédasticité et la méthode des moindres carrés généralisés	50
Section 5 : Problèmes liés à l'instabilité d'une relation au cours du temps ou à l'hétérogénéité d'une population	60
Section 6 : Les retards d'ajustement	70
Chapitre II. L'estimation des modèles à équations simultanées	83
Section 1 : Le modèle comme système d'équation	85
Section 2 : L'estimation des modèles à équations simultanées	108

Chapitre III. L'économétrie sans modèle, les séries temporelles	139
Section 1 : Les propriétés des processus stationnaires	140
Section 2 : Les propriétés des modèles de séries temporelles (Time séries)	147
Section 3 : L'estimation des processus ARMA	164
Chapitre IV. Introduction à la coïntégration	167
Section 1 : Les tests de racine unitaire	168
Section 2 : La coïntégration	171
Section 3 : Propriétés des relations de coïntégration	172
Chapitre V. Application à la consommation	177
Section 1 : La consommation globale	177
Section 2 : L'approche semi-globale	187
Section 3 : L'approche par catégories de produits	190
Chapitre VI. Application aux échanges extérieurs	197
Section 1 : Les importations	197
Section 2 : Les exportations	207
Tables statistiques	215

Introduction

Cet ouvrage est destiné au public des étudiants de second cycle de sciences économiques et de gestion, mais aussi plus largement à tous ceux qui souhaitent acquérir une formation de base en économétrie. Il s'agit donc d'un manuel dont l'ambition est de présenter de façon claire et accessible au plus grand nombre, un ensemble de techniques dont l'usage est devenu courant dans les milieux professionnels. C'est aussi un ouvrage qui privilégie les applications par rapport à la théorie pure. On y trouvera les techniques les plus usuelles illustrées par des exemples d'applications. Mais l'économétrie est une discipline vivante dans laquelle existent une recherche active et un flux régulier de nouveaux développements. Rendre compte de tous ces travaux dans toute leur richesse dépasse largement le cadre de cet ouvrage. Il existe des revues, des manuels et des ouvrages qui leur sont consacrés et auxquels nous renvoyons le lecteur qui souhaitera dépasser le contenu volontairement limité de ce manuel. Susciter, par cette première approche, le désir d'un approfondissement ultérieur ou d'un effort pour suivre l'évolution de la discipline fait partie intégrante des objectifs de cet ouvrage.

L'essentiel de ses développements est tiré d'un cours enseigné sous différentes variantes de contenu et de niveau aux universités d'Orléans, d'Abidjan, de Niamey, de Cotonou, de Butare et dernièrement en maîtrise à l'université du Maine.

L'économétrie est une branche de la science économique consistant à établir des lois ou à vérifier des hypothèses à partir de données chiffrées tirées de la réalité. Les étudiants de première année apprennent que pour différentes raisons, il est impossible d'utiliser en sciences sociales la méthode expérimentale, comme cela se pratique dans les sciences de la nature. Cette limitation poserait des problèmes redoutables concernant la validité des développements de la théorie économique s'il n'existait un substitut de l'expérimentation consistant à en confronter les résultats aux données tirées de l'observation de la réalité. Dans cette démarche essentielle, l'économétrie fournit les méthodes qui permettront de tester les hypothèses ou de quantifier les relations. Elle a par conséquent un caractère

instrumental et constitue le complément indispensable de l'analyse économique. De ce fait il est à peu près impossible de faire de la recherche en sciences économiques sans se trouver devant la nécessité de lire ou de réaliser des travaux d'économétrie à un moment ou à un autre. C'est la raison pour laquelle, dans tous les pays, la formation des économistes suppose l'acquisition de ces techniques.

Mais l'économétrie permet également de quantifier les relations entre variables économiques, ainsi par exemple de chiffrer la réponse de l'investissement en logement à une variation du taux d'intérêt. Pour cela, l'économiste part d'une analyse théorique des déterminants de l'investissement en logement, identifiant les variables explicatives et le type de relation qu'elles sont susceptibles d'avoir avec la variable expliquée. Ces éléments analytiques vont être traduits en une équation mathématique mettant en relation l'investissement en logement et ses principales variables explicatives. En donnant aux variables les valeurs enregistrées dans la réalité, l'économétrie permet à la fois de tester la validité du modèle et d'en chiffrer les paramètres. Elle rend possible la modélisation, c'est-à-dire la représentation de phénomènes économiques, nécessairement complexes, par un ensemble cohérent de relations mathématiques quantifiées. A son tour la modélisation constitue un enrichissement important, sous la forme d'une aide à la décision, en rendant possible la simulation, l'optimisation ou la prévision. La simulation consiste à calculer les conséquences d'un ensemble d'hypothèses représentées par des valeurs des variables explicatives. La prévision étend cet exercice à des périodes futures. L'optimisation adopte le chemin inverse et a pour but de déterminer les valeurs optimales de certaines variables en fonction d'objectifs fixés. La plupart des grands centres de décision, qu'il s'agisse des banques, des agences gouvernementales, des institutions internationales et des grandes entreprises utilisent régulièrement ces techniques. Les constructeurs d'avion utilisent des modèles économétriques pour analyser la demande actuelle et future. Les Ministères des Finances, de l'Économie ou du Plan construisent des modèles macro-économiques pour la prévision et pour l'évaluation des politiques économiques. Les institutions non gouvernementales d'aide au développement construisent et actualisent régulièrement des modèles des grandes régions du monde et des principales économies nationales. Il en résulte que pour travailler *dans* ces institutions, il est indispensable de pouvoir manier les techniques

économétriques souvent à haut niveau. Il en résulte aussi que pour pouvoir travailler *avec* ces institutions, il faut maîtriser ce langage commun.

Il est vrai que depuis quelques années la mode est au libéralisme et à des politiques économiques aussi peu actives que possible. Les marchés sont supposés tout régler dès lors qu'ils ne sont pas entravés. Dans ces conditions, on peut douter de l'utilité des modèles macro-économiques, d'autant plus qu'ils se sont révélés peu performants, pour ne pas dire plus, dans la période récente. De fait, l'un des débouchés traditionnels de la modélisation, qui avait largement contribué à son développement dans le domaine de la macro-économie et exercé un effet d'appel sur l'économétrie, se trouve actuellement quelque peu délaissé. On peut cependant penser que les idées évolueront, en fonction des résultats de l'expérience, vers une plus grande autonomie reconnue à la politique économique. De leur côté les techniques économétriques ne cessent de progresser et la qualité des modèles s'améliore. De nouvelles formes structurelles comme les modèles à correction d'erreur ont permis simultanément d'alléger les équations et d'améliorer leur qualité dans de nombreux cas, en particulier en macro-économie. L'utilisation des modèles de séries temporelles développés par Box et Jenkins ou leur association à des modèles structurels permettent, de leur côté, d'améliorer sensiblement les qualités en prévision. De nouvelles méthodes d'estimation permettent de tenir compte des relations pouvant exister entre les chocs aléatoires subis par plusieurs variables etc. Tout ceci amène à penser que le recours à la modélisation et à l'économétrie devrait connaître de nouveaux développements dans l'avenir. Les applications à l'entreprise et au marché en général, malgré les problèmes particuliers qu'elles posent pourraient bien représenter la partie la plus active de ces futurs développements.

Par ses méthodes comme par la nature des informations qu'elle utilise, l'économétrie fait partie des disciplines quantitatives. On sait que depuis fort longtemps la question du rôle et de la place de ces techniques en sciences économiques fait l'objet de discussions acharnées. Si la possibilité d'utiliser la puissance de l'instrument mathématique en raison du caractère quantifié des variables constitue un avantage indéniable, le caractère réducteur de la formalisation, la nécessité de postuler des comportements simplistes (les seuls que l'on puisse traiter facilement), le recours systématique à l'abstraction sont autant

d'aspects qui prêtent à la critique. Le simple fait que lors d'un des grands débats méthodologiques des années soixante on ait pu forger le concept de "réalisme déplacé" (misplaced concreteness), en analyse économique, montre bien que pour les tenants les plus extrêmes de l'abstraction lorsque la réalité est en contradiction avec la théorie, c'est la réalité qui a tort. Il y aurait beaucoup à dire sur le rapport de la théorie économique à son objet, sur le rôle éclairant ou mystificateur de la formalisation. Force est de constater que cette dernière constitue souvent la seule originalité de nombreux travaux et dissimule plus qu'elle ne supplée l'insuffisance théorique. Un exercice salutaire consiste à reprendre les conclusions et à les confronter directement aux hypothèses de départ. On mesure mieux, alors, ce qu'apportent les étapes intermédiaires du raisonnement.

L'économétrie échappe largement à ces débats malgré son caractère formalisé. En effet, elle s'inscrit dans une démarche inductive qui s'efforce de découvrir à partir de phénomènes particuliers, les lois générales susceptibles de les avoir engendrés. A l'opposé d'une généralisation hâtive de modèles a priori, l'économétrie s'appuie sur la réalité qu'elle permet de mieux déchiffrer. Elle se nourrit de cette confrontation permanente entre les hypothèses théoriques et la réalité telle qu'elle peut être représentée par les données. Dès lors la recherche en économétrie, dont le caractère formalisé est évident, n'a d'autre but que de permettre une meilleure appréhension de la réalité. En perfectionnant les méthodes d'estimation, en précisant leurs conditions de validité, elle offre la possibilité d'obtenir des résultats plus précis et plus fiables. Ainsi les développements récents de la théorie de la cointégration ont permis de mieux cerner les conditions de validité des procédures d'estimation courantes et d'éliminer les corrélations fortuites. L'utilisation généralisée des tests de racine unitaire qui en résulte augmente la fiabilité des résultats numériques. C'est pourquoi même si la matière de ces travaux est souvent d'un accès difficile, leurs conclusions et les procédures qui peuvent en découler sont d'un grand intérêt. La formalisation n'est pas vaine, elle est au contraire le meilleur chemin vers le concret. A partir d'un petit nombre de lois, comme les principales lois de probabilité, l'économétrie a construit un corps de savoirs essentiellement tourné vers les applications.

Comme dans d'autres disciplines, les problèmes qui se posent au niveau de ces applications renvoient, généralement, à la théorie en précisant par exemple la

distribution de probabilité d'un type d'estimateur dans des conditions données. Là encore, la cointégration fournit un bon exemple puisqu'après avoir montré que dans certaines conditions les distributions n'étaient plus standards, les économètres ont calculé les vraies distributions ou leurs approximations.

La connaissance des procédures, si elle est importante, n'est jamais suffisante. Il faut également maîtriser leur conditions d'emploi et leurs limites. Le détour théorique, que l'on a voulu aussi simple que possible, n'est donc pas inutile.

Chapitre I. L'estimation d'une équation unique

Estimer une équation, c'est déterminer la valeur des paramètres de la fonction qui va relier une variable à ses déterminants présumés.

$$y_t = a_1 x_{1t} + \dots + a_l x_{lt} + u_t \quad (1)$$

Les paramètres à estimer sont a_1, \dots, a_l

Si on était dans un univers déterministe, la loi en question serait parfaitement vérifiée, mais ce n'est pas le cas pour plusieurs raisons :

- Les phénomènes économiques sont caractérisés par l'interdépendance entre de nombreux éléments, ce qui entraîne que les variables explicatives susceptibles d'exercer une influence sur la variable expliquée sont très nombreuses (on ne les retient pas toutes). Mais, l'effet des variables qui ont été omises explique qu'il y ait des écarts entre la réalité observée et le résultat de la fonction.
- Les phénomènes économiques sont mis en œuvre par les individus qui n'ont eux mêmes pas un comportement déterminé. C'est donc leur libre arbitre qui fait que parfois ils n'agissent pas comme prévu et donc on n'obtient pas le résultat escompté.

Les lois économiques sont donc seulement vraies en moyenne. Elles ont le caractère de lois statistiques.

On a choisi de traduire cet écart par une variable aléatoire noté u_t appelée terme d'erreur.

Compte tenu de l'existence de ce terme, la loi qui va gouverner y_t n'est plus une loi mathématique mais une loi statistique, et, il faudra se donner une distribution pour caractériser u_t .

L'estimation des coefficients a_1, \dots, a_l , fournit des valeurs qui seront elles-mêmes des variables aléatoires ayant une distribution de probabilité que l'on devra spécifier selon la distribution de u_t .

On dispose d'un ensemble d'observations portant sur la variable expliquée et sur les variables explicatives. Sachant que la relation a un caractère probabiliste, il n'est pas possible de déterminer de façon unique les valeurs vraies des paramètres.

On peut simplement déterminer des estimations et connaissant la distribution de probabilité de ces estimations, remonter vers des intervalles de confiance pour les paramètres.

Les paramètres estimés sont désignés par $\hat{a}_1, \dots, \hat{a}_l$ et, les connaissant, on pourra remonter aux valeurs vraies des paramètres a_1, a_2, \dots, a_l .

Pour que cette opération d'inférence statistique soit possible il faut qu'un certain nombre d'hypothèses soient vérifiées. À défaut, on aura toujours un résultat numérique, mais le résultat est sans valeur car il ne permet pas de porter un jugement sur les vraies valeurs des coefficients recherchés.

Les hypothèses en question portent sur les variables ainsi que sur le terme d'erreur.

Soient $y_1 \dots y_T$ les valeurs observées de y et x_{11} à x_{1T}, \dots, x_{l1} à x_{lT} les valeurs observées des variables explicatives. On peut écrire Y et X sous la forme :

$$Y_{(T,1)} = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix} \quad X_{(T,l)} = \begin{bmatrix} x_{11} & \dots & x_{l1} \\ \vdots & & \vdots \\ x_{1T} & \dots & x_{lT} \end{bmatrix}$$

(variables explicatives)

Les coefficients peuvent aussi s'écrire :

$$A_{(l,1)} = \begin{bmatrix} a_1 \\ \vdots \\ a_l \end{bmatrix} \quad \text{de même :} \quad U_{(T,1)} = \begin{bmatrix} u_1 \\ \vdots \\ u_T \end{bmatrix}$$

On a alors

$$(2) \quad Y = XA + U$$

forme matricielle de l'équation (1)

Le vecteur des estimateurs sera $\hat{A}_{(1,1)} = \begin{bmatrix} \hat{a}_1 \\ \cdot \\ \cdot \\ \cdot \\ \hat{a}_l \end{bmatrix}$

d'où :

$$\hat{Y} = X \hat{A}$$

$$Y - \hat{Y} = \hat{U}$$

U est le terme d'erreur → inconnu

\hat{U} est le résidu → connu.

On distingue 7 hypothèses :

- 1- Les variables sont observées sans erreur. (L'économètre s'interdit de s'immiscer dans la collecte des données).

On peut cependant avoir une petite idée sur la fiabilité des chiffres qu'on utilise. Le degré de précision de la méthode d'estimation est dans une certaine mesure limité par la qualité des données. Il ne sert à rien d'utiliser des méthodes sophistiquées sur des données elles-mêmes douteuses. Leur estimation grossière peut suffire pour donner une idée du phénomène étudié.

- 2- Le terme d'erreur a une espérance mathématique nulle

$$E(u_t) = 0$$

c'est-à-dire qu'on se trompe tantôt dans un sens tantôt de l'autre. L'erreur n'est pas systématique.

- 3- u_t a une variance σ_u^2 qui est constante et finie. C'est ce qu'on appelle l'homoscédasticité c'est-à-dire que la marge d'erreur est la même en début et en fin de période ou bien qu'elle est identique en tout point de l'échantillon.

C'est une hypothèse parfois contestable dans la mesure où on a des raisons de penser que la variance peut augmenter ou diminuer sur l'ensemble des observations.

Cette hypothèse peut être testée et quand on a repéré l'hétéroscédasticité on peut corriger et obtenir néanmoins de bons estimateurs.

4- Les termes d'erreur sont indépendants les uns des autres, dans le temps.

$$\begin{cases} E(u_t u_i) = 0 & \text{si } t \neq i \\ E(u_t u_i) = \sigma_u^2 & \text{si } t = i \end{cases}$$

Cette hypothèse traduit le fait que l'erreur n'est pas systématique. En effet, l'erreur en $t + 1$ est indépendante de l'erreur en t .

Cette hypothèse peut, elle aussi, être testée. Quand H4 n'est pas vérifiée, on parle alors d'auto corrélation, ce qui traduit souvent un problème de spécification. En effet, une spécification défectueuse est de nature à entraîner une erreur systématique :

- Omission d'une variable importante
- Choix erroné de la fonction (linéaire à la place d'exponentielle par exemple).

Les hypothèses sur l'homoscédasticité et l'indépendance peuvent se résumer en écrivant :

$$E(UU') = \sigma_i^2$$

avec U vecteur colonne et U' son transposé, I étant la matrice unité.

$E(UU')$ est la matrice des variances covariances des termes d'erreur. C'est donc une matrice (TXT).

$$E(UU') = \begin{matrix} (T,T) \\ \left[\begin{array}{cc} E(u_1 u_1) \dots & E(u_1 u_T) \\ E(u_2 u_1) & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ E(u_T u_1) & E(u_T u_T) \end{array} \right] \end{matrix}$$

Sous l'hypothèse H4 d'indépendance des termes d'erreur, si 2 termes d'erreur sont indépendants $E(u_i u_j) = 0$ ($i \neq j$). Par conséquent les termes situés en dehors de la première diagonale sont nuls.

Sous l'hypothèse H3 les termes restants sont égaux à σ_u^2

$$\begin{bmatrix} \sigma_u^2 & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & \sigma_u^2 \end{bmatrix} = \sigma_u^2 I_{(T,T)}$$

Donc

$$E(UU') = \sigma_u^2 I$$

- 5- Les variables explicatives $x_1 \dots x_l$ ne sont pas linéairement dépendantes. Sinon, on parle de collinéarité des variables explicatives. On peut la déceler et y remédier le plus souvent.
- 6- Les variables explicatives sont indépendantes du terme d'erreur. Ce n'est pas vérifié dans plusieurs cas :
 - . Dans les modèles à équations simultanées certaines variables explicatives peuvent elles-mêmes être des variables expliquées dans une autre équation et donc être des variables aléatoires.
 - . Quand on travaille sous l'hypothèse d'anticipation rationnelle, puisque les variables anticipées sont supposées résulter de l'application de la théorie économique et sont par conséquent assimilables à des variables endogènes.
 - . Lorsque l'on travaille avec des variables explicatives connues avec erreur (remise en cause de H1).

Dans tous ces cas on utilise la méthode des variables instrumentales.
- 7- Les termes d'erreur suivent une loi normale.

C'est grâce à cette hypothèse que l'inférence statistique peut se réaliser car c'est elle qui va préciser la distribution statistique des estimateurs $\hat{\alpha}$. Elle joue donc un rôle essentiel bien qu'elle soit difficile à vérifier.

Il existe des tests de normalité mais ils ne sont pas applicables sur les termes d'erreurs, qui par définition demeurent inconnus.

Sachant que les termes d'erreur ont une loi de probabilité qui est normale, les estimateurs c'est-à-dire les \hat{a} suivent aussi une loi normale et par conséquent les grandeurs $\frac{\hat{a}_i - E(\hat{a}_i)}{\sigma_{\hat{a}_i}}$ suivent une loi normale centrée réduite. ($\rightarrow N(0,1)$)

Mais, comme on ne connaît pas la variance du terme d'erreur on doit l'estimer. La loi normale est alors remplacée par une loi de Student Fisher :

$$\frac{\hat{a}_i - E(\hat{a}_i)}{\sigma_{\hat{a}_i}} \rightarrow \text{Student}(T-1)$$

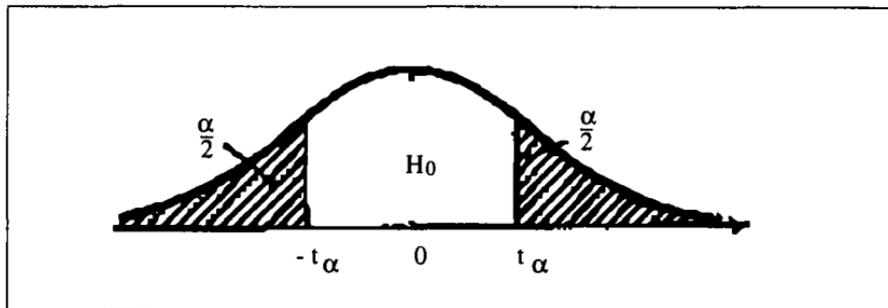
Cette loi permet l'inférence statistique. On peut revenir sur les vraies valeurs a_i , à l'aide d'un test simple

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

Si \hat{a}_i est un estimateur sans biais, $E(\hat{a}_i) = a_i$ et donc sous $H_0 : E(a_i) = 0$ et

$$\frac{\hat{a}_i}{\sigma_{a_i}} \rightarrow \text{Student}(T-1)$$

Soit $t_\alpha =$ Valeur critique de la table de Student



Si $\frac{\hat{a}_i}{\sigma_{\hat{a}_i}}$ est compris entre $-t_\alpha$ et t_α alors on accepte H_0 . Le coefficient a_i n'est donc pas significativement différent de 0. A l'extérieur de l'intervalle on rejette H_0 .

$$\frac{\hat{a}_i}{\sigma_{\hat{a}_i}} > t_\alpha \text{ ou } \frac{\hat{a}_i}{\sigma_{\hat{a}_i}} < -t_\alpha \Rightarrow H_1$$

→ et x_j est bien une variable explicative.

Rappel

La plupart des propriétés qui ont été démontrées en économétrie sont relatives à des variables stationnaires.

SECTION 1 : L'ESTIMATION PAR LES MOINDRES CARRÉS ORDINAIRES DANS LE CAS GÉNÉRAL

Soit le modèle $Y = X A + U$ (1)
(T,1) (T,1)(l,1) (T,1)

\hat{A} est un estimateur de A, donc

$$\hat{Y} = X \hat{A} \quad (2)$$

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

Le principe de la méthode des moindres carrés consiste à trouver les valeurs des estimateurs qui minimisent la somme des carrés des résidus.

A- Les estimateurs simples

On part du calcul de :

$$\text{MIN } \sum \hat{u}_t^2 \text{ c'est - à - dire } \text{MIN } \hat{U}' \hat{U}$$

$$\hat{U} = Y - X \hat{A} \quad \text{d'après (3)}$$

$$\text{Par définition le transposé de } \hat{U} \text{ est } \hat{U}' = (Y - X \hat{A})' = Y' - \hat{A}' X' \quad (4)$$

$$\text{Donc, } \hat{U}' \hat{U} = (Y' - \hat{A}' X') (Y - X \hat{A}) \quad (5)$$

$$\hat{U}' \hat{U} = Y' Y - Y' X \hat{A} - \hat{A}' X' Y + \hat{A}' X' X \hat{A} \quad (6)$$
(l,1) (l,1) (l,1) (l,1)

Tous les éléments du produit sont des scalaires. Sachant qu'un scalaire est égal à son transposé $Y'X\hat{A} = \hat{A}'X'Y$ et le résultat peut s'écrire indifféremment.

$$\hat{U}'\hat{U} = Y'Y - 2Y'X\hat{A} + \hat{A}'X'X\hat{A}$$

ou

$$\hat{U}'\hat{U} = Y'Y - 2\hat{A}'X'Y + \hat{A}'X'X\hat{A} \quad (7)$$

On cherche le minimum de cette expression par rapport à \hat{A} . Cela revient à annuler la dérivée de $\hat{U}'\hat{U}$ par rapport à \hat{A} :

$$\text{MIN}(\hat{U}'\hat{U}) \Rightarrow \frac{\partial \hat{U}'\hat{U}}{\partial \hat{A}} = 0$$

$$\frac{\partial \hat{U}'\hat{U}}{\partial \hat{A}} = -2X'Y + 2X'X\hat{A} = 0$$

d'où $X'Y = X'X\hat{A}$. Si $X'X$ est une matrice régulière, admettant une inverse, \hat{A} s'obtient en multipliant l'équation précédente par $(X'X)^{-1}$:

d'où

$$\hat{A} = (X'X)^{-1} X'Y \quad (8)$$

Remarque

Dans le calcul précédent, on a :

$$\begin{cases} \frac{\partial \hat{A}'X'Y}{\partial \hat{A}} = X'Y \\ \frac{\partial \hat{A}'X'X\hat{A}}{\partial \hat{A}} = 2X'X\hat{A} \end{cases}$$

Ce résultat se vérifie aisément en explicitant les éléments du calcul. Appelons $\alpha_1, \alpha_2, \dots, \alpha_l$ les éléments du produit $X'Y$.

$$(9) \text{ Soit } X'Y = \begin{matrix} (l,1) \\ \left[\begin{array}{c} \alpha_1 \\ \vdots \\ \alpha_l \end{array} \right] \end{matrix}$$

En multipliant $X'Y$ par \hat{A} , il vient :

$$\begin{aligned} \hat{A}'(X'Y) &= (\hat{a}_1 \dots \hat{a}_l) \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_l \end{bmatrix} \\ &= \hat{a}_1 \alpha_1 + \hat{a}_2 \alpha_2 + \dots + \hat{a}_l \alpha_l \quad (10) \end{aligned}$$

La dérivation de cette expression par rapport à \hat{A} revient à dériver par rapport à chacun des éléments de \hat{A} , soit $\hat{a}_1 \dots \hat{a}_l$

$$\frac{\partial \hat{A}'(X'Y)}{\partial \hat{A}} = \begin{bmatrix} \frac{\partial(\hat{a}_1 \alpha_1 + \hat{a}_2 \alpha_2 + \dots + \hat{a}_l \alpha_l)}{\partial \hat{a}_1} = \alpha_1 \\ \vdots \\ \frac{\partial(\hat{a}_1 \alpha_1 + \hat{a}_2 \alpha_2 + \dots + \hat{a}_l \alpha_l)}{\partial \hat{a}_l} = \alpha_l \end{bmatrix}$$

Le résultat est un vecteur (1, 1) composé des termes α_1 à α_l , c'est-à-dire $X'Y$ lui-même.

$$\text{d'où } \frac{\partial \hat{A}' X' Y}{\partial \hat{A}} = X' Y$$

De même, appelons β_{ij} les éléments du produit $X'X$ de format (l,l). Par construction $X'X$ est une matrice symétrique.

$$(11) \quad X'X = \begin{bmatrix} \beta_{11} & \cdot & \cdot & \cdot & \beta_{1l} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \beta_{l1} & \cdot & \cdot & \cdot & \beta_{ll} \end{bmatrix} \text{ avec } \beta_{ij} = \beta_{ji} \quad \forall_i \text{ et } j$$

En multipliant à gauche par \hat{A}' , on obtient un vecteur ligne à l éléments :

$$\hat{A}' X' X = (\hat{a}_1 \beta_{11} + \hat{a}_2 \beta_{21} + \dots + \hat{a}_l \beta_{l1}, \dots, \hat{a}_1 \beta_{1l} + \dots + \hat{a}_2 \beta_{2l} + \dots + \hat{a}_l \beta_{ll}) \quad (12)$$

Puis en multipliant à droite par \hat{A} , il vient

$$(\hat{a}_1 \beta_{11} + \hat{a}_2 \beta_{21} + \dots + \hat{a}_l \beta_{l1}) \hat{a}_1 + \dots + (\hat{a}_1 \beta_{1l} + \hat{a}_2 \beta_{2l} + \dots + \hat{a}_l \beta_{ll}) \hat{a}_l$$

(13) qui est un scalaire.

La dérivée de $\hat{A}' X' X \hat{A}$ par rapport à \hat{A} s'obtient en dérivant par rapport à $\hat{a}_1 \dots \hat{a}_l$

$$\frac{\partial \hat{A}' X' X \hat{A}}{\partial \hat{A}} = \begin{bmatrix} \frac{\partial \hat{A}' X' X \hat{A}}{\partial \hat{a}_1} = 2\beta_{11}\hat{a}_1 + 2\hat{a}_2\beta_{12} + \dots + 2\hat{a}_l\beta_{1l} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial \hat{A}' X' X \hat{A}}{\partial \hat{a}_l} = 2\hat{a}_1\beta_{l1} + 2\hat{a}_2\beta_{l2} + \dots + 2\hat{a}_l\beta_{ll} \end{bmatrix}$$

$$\text{Soit } \frac{\partial \hat{A}' X' X \hat{A}}{\partial \hat{A}} = 2 \begin{bmatrix} \hat{a}_1\beta_{11} + \hat{a}_2\beta_{21} + \dots + \hat{a}_l\beta_{l1} \\ \cdot \\ \cdot \\ \cdot \\ \hat{a}_1\beta_{1l} + \hat{a}_2\beta_{2l} + \dots + \hat{a}_l\beta_{ll} \end{bmatrix}$$

$$D' \text{ où } \frac{\partial \hat{A}' X' X \hat{A}}{\partial \hat{A}} = 2(X' X) \hat{A}$$

Après avoir calculé \hat{A} , nous pouvons examiner les propriétés de cet estimateur.

a) \hat{A} est un estimateur sans biais si $E(U) = 0$ (H2)

On sait que $\hat{A} = (X'X)^{-1} X'Y$

et que $Y = XA + U$

En remplaçant Y par sa valeur :

$$\hat{A} = (X'X)^{-1} X'(XA + U) \quad (14)$$

$$\hat{A} = (X'X)^{-1} X'XA + (X'X)^{-1} X'U \quad (15)$$

$$\hat{A} = A + (X'X)^{-1} X'U \quad (16)$$

$$\text{d'où } E(\hat{A}) = A + (X'X)^{-1} X'E(U)$$

$$\text{donc } E(\hat{A}) = A \text{ si } E(U) = 0$$

Sous H2, \hat{A} est un estimateur sans biais de A .

b) Variance de \hat{A} :

La variance du vecteur \hat{A} est donnée par la matrice des variances covariances obtenue en développant en lignes et en colonnes :

$$V(\hat{A}) = E \left[(\hat{A} - E(\hat{A}))(\hat{A} - E(\hat{A}))' \right]$$

$$V(\hat{A})_{(i,l)} = \begin{bmatrix} V_{(\hat{a}_1)} & \text{cov}_{(\hat{a}_1\hat{a}_2)} & \cdot & \cdot & \cdot & \text{cov}_{(\hat{a}_1\hat{a}_i)} \\ \cdot & \cdot & & & & \cdot \\ \cdot & & V_{(\hat{a}_i)} & & & \cdot \\ \cdot & & & \cdot & & \cdot \\ \cdot & & & & \cdot & \cdot \\ \text{cov}_{(\hat{a}_i\hat{a}_1)} & \cdot & \cdot & \cdot & \cdot & V_{(\hat{a}_i)} \end{bmatrix}$$

On sait que $\hat{A} = (X'X)^{-1} X'Y$ et que $E(\hat{A}) = A$.

$$\text{Donc } \hat{A} - E(\hat{A}) = (X'X)^{-1} X'(XA + U) - A$$

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

$$\text{et } [\hat{A} - E(\hat{A})]' = U'X (X'X)^{-1} \quad (17)$$

$(X'X)^{-1}$ matrice symétrique par construction est égale à sa transposée.

$$\text{D'où } V(\hat{A}) = E [(X'X)^{-1} X'UU' X(X'X)^{-1}] \quad (18)$$

Comme les éléments en X sont des éléments certains, on peut encore écrire :

$$V(\hat{A}) = (X'X)^{-1} X' E(UU') X(X'X)^{-1} \quad (19)$$

Or, sous les hypothèses H3 et H4, on a $E(UU') = \sigma_u^2 I$, σ_u^2 étant un scalaire et I la matrice unité de dimensions (T, T) .

$$\text{Donc } V(\hat{A}) = \sigma_u^2 (X' X)^{-1} X' I X(X' X)^{-1} \quad (20)$$

soit finalement

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

Remarque : La simplification n'est possible que si H3 et H4 sont vérifiées et donc si on peut écrire $E(UU') = \sigma_u^2 I$

On peut montrer que sous les hypothèses qui ont été posées, l'estimateur \hat{A} des moindres carrés ordinaires (MCO) est un estimateur à variance minimale, dans la classe des estimateurs sans biais.

Si maintenant $E(UU') \neq \sigma_u^2 I$, mais $E(UU')$ est connue, par exemple $E(UU') = V$, il est possible de retrouver les propriétés précédentes comme on le verra dans la suite de ce chapitre, à la section 5.

Le calcul de $V(\hat{A})$ fait appel à la variance du terme d'erreur, σ_u^2 .

Celle-ci est inconnue. On peut cependant utiliser un estimateur de σ_u^2 construit à partir de la variance du terme résiduel $\sigma_{\hat{u}}^2$.

$$\hat{\sigma}_u^2 = \frac{\sum \hat{u}_t^2}{T-1} = \frac{\hat{U}' \hat{U}}{T-1} \quad (21)$$

B- Estimations sous contrainte et multiplicateurs de Lagrange :

Les estimateurs qui ont été présentés jusqu'ici présentent la propriété de minimiser la somme des carrés des écarts entre valeurs calculées et valeurs observées de la variable expliquée y_t . On peut leur imposer de satisfaire également d'autres

propriétés. On obtient alors des estimateurs sous contrainte qui minimisent toujours la somme des carrés des écarts, tout en satisfaisant certaines contraintes. Ils sont normalement "moins bons" que les estimateurs sans contrainte et ceci pose la question de savoir si la ou les contraintes sont acceptables compte tenu du modèle et des données utilisées.

Exemple de contraintes : on peut imposer que les rendements d'échelle d'un processus de production soient constants. Dans ce cas si les coefficients à estimer représentent les élasticités de la production par rapport aux facteurs de production, cela se traduira par la contrainte $\sum a_i = 1$.

On supposera dans la suite que les contraintes sont de forme linéaire.

Les restrictions de forme linéaire peuvent être écrites sous forme matricielle telles que : $H\tilde{A}=h$

– \tilde{A} = vecteur des estimateurs sous contrainte. C'est un vecteur colonne à l éléments.

Pour le distinguer du vecteur \hat{A} des MCO on l'écrira \tilde{A}

– H = matrice des restrictions de format $(q \times l)$, "q" étant le nombre de restrictions.

– h est donc un vecteur colonne de format $(q, 1)$.

Exemple : avec 4 coefficients et 2 contraintes

$$l = 4$$

$$q = 2$$

$$\tilde{a}_1 + \tilde{a}_2 = 1$$

$$\alpha_3 \tilde{a}_3 + \alpha_4 \tilde{a}_4 = z$$

On a donc :

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & \alpha_3 & \alpha_4 \end{bmatrix} \begin{bmatrix} \tilde{a}_1 \\ \tilde{a}_2 \\ \tilde{a}_3 \\ \tilde{a}_4 \end{bmatrix} = \begin{bmatrix} 1 \\ z \end{bmatrix}$$

$H \quad \tilde{A} \quad h$

Soit

$$H\hat{A} - H\bar{A} = H(X'X)^{-1}H'\lambda \quad (31)$$

Mais d'après la contrainte, $H\bar{A} = h$

donc

$$H\hat{A} - h = H(X'X)^{-1}H'\lambda$$

soit en multipliant à gauche par $[H(X'X)^{-1}H']^{-1}$

$$\lambda = [H(X'X)^{-1}H']^{-1}[H\hat{A} - h] \quad (32)$$

Cette valeur de λ peut être reportée dans les conditions de premier ordre :

$$-X'Y + X'X\bar{A} + H'[H(X'X)^{-1}H']^{-1}[H\hat{A} - h] = 0 \quad (33)$$

En prémultipliant par $(X'X)^{-1}$, il vient

$$-(X'X)^{-1}X'Y + (X'X)^{-1}X'X\bar{A} + (X'X)^{-1}H[H(X'X)^{-1}H']^{-1}[H\hat{A} - h] = 0 \quad (34)$$

$$-\hat{A} + \bar{A} + (X'X)^{-1}H'[H(X'X)^{-1}H']^{-1}[H\hat{A} - h] = 0 \quad (35)$$

et

$$\bar{A} = \hat{A} - (X'X)^{-1}H'[H(X'X)^{-1}H']^{-1}[H\hat{A} - h] \quad (36)$$

On constate que l'estimateur sous contrainte ne peut pas coïncider avec l'estimateur des MCO. Cela revient à dire que l'imposition de contraintes dégrade la qualité des estimateurs.

On peut tester la validité des contraintes qui ont été imposées sur les estimateurs. C'est le principe des tests de restriction à priori.

λ est un vecteur ayant autant d'éléments que de restrictions qui mesure la perte ou le coût associé aux restrictions à priori.

Il y a un multiplicateur par contrainte dont l'interprétation est identique à celle des multiplicateurs de Lagrange habituels. Il mesure le coût en termes de fonction objectifs, donc ici en termes d'augmentation de la somme des carrés des résidus, de la contrainte à laquelle il est associé.

Il existe toute une série de tests sur la base des propriétés des multiplicateurs de Lagrange. Ces tests ont l'avantage de n'utiliser qu'une seule régression alors que deux seraient normalement nécessaires pour un test de restrictions a priori.

Soient $\sum \hat{u}_t^2$ la somme des carrés des résidus de l'équation sans contrainte et $\sum \hat{w}_t^2$ la somme des carrés des résidus de l'équation avec contrainte. Par définition $\sum \hat{w}_t^2 > \sum \hat{u}_t^2$. Un test de restriction a priori classique passe par la comparaison des variances résiduelles.

$$\text{Le rapport : (37) } F_e = \frac{\frac{\sum \hat{w}_t^2 - \sum \hat{u}_t^2}{q}}{\frac{\sum \hat{u}_t^2}{T-1}} \text{ suit une loi de Fisher } (q, T-1)$$

En posant :

H0 : les contraintes ne modifient pas significativement la variance résiduelle. Le modèle avec contraintes est acceptable.

H1 : Les contraintes modifient significativement la variance résiduelle et ne sont pas compatibles avec les données. Le modèle avec contraintes est rejeté.

Si $F_{\text{empirique}} < F_{\alpha}$ valeur critique de la table de Fisher, on conclut H0.

Si $F_{\text{empirique}} > F_{\alpha}$ on conclut H1.

Mais on sait par ailleurs que la loi de Fisher $F(n, m)$ tend vers la loi du χ^2 centrée lorsque m augmente. Le χ^2 centré (ou χ^2 réduit... ou "augmented χ^2 " en anglais) étant égal à χ_n^2/n .

$$F(n, m) \rightarrow \chi_n^2/n \text{ qd } m \rightarrow \infty. \text{ Soit encore } n F(n, m) \rightarrow \chi_n^2$$

En régressant les résidus de la régression sous H0 sur les contraintes, on obtient un R^2 qui mesure le pouvoir explicatif de la régression, donc ici la part de la variance résiduelle (sous H0) imputable aux contraintes. En utilisant l'expression de F_e donnée dans (37), il vient :

$$F_e = \frac{T-1}{q} R^2 \text{ soit } (T-1)R^2 = q F_e.$$

Pour T grand cette dernière expression peut être approchée par TR^2 à gauche et par χ_q^2 à droite.

Le test de restriction a priori peut donc se ramener au test du χ^2 pratiqué sur le R^2 de la régression auxiliaire.

Soit χ_α^2 la valeur critique de la table, la règle de décision est :

χ^2 empirique $< \chi_\alpha^2$ on conclut H_0 .

χ^2 empirique $> \chi_\alpha^2$ on conclut H_1 .

Les tests sur les multiplicateurs de Lagrange, encore appelés LM tests dans la littérature peuvent être appliqués à un ensemble très diversifié des problèmes en fonction de la nature des contraintes figurant dans l'équation (32). Ce sont en fait des cadres dans lesquels on peut faire rentrer toutes sortes d'hypothèses. On en verra un exemple dans la section 3.

Dans la réalité, les conditions nécessaires pour l'application de la méthode des moindres carrés ordinaires ne seront remplies qu'exceptionnellement. La situation la plus courante est celle où une ou plusieurs hypothèses ne sont pas vérifiées. Il est essentiel, tout d'abord, de s'en apercevoir puis d'appliquer les procédures de correction appropriées qui permettront aux estimateurs des moindres carrés qui sont à la fois parmi les plus simples à calculer et les plus robustes de conserver leurs propriétés.

Les sections qui suivent sont consacrées aux principaux problèmes susceptibles de se poser lorsqu'on cherche à estimer une équation unique. Le problème de la stationnarité des variables, qui est en lui-même un problème essentiel, susceptible de transformer complètement l'approche de l'économétrie sera traité plus en détail aux chapitres 3 et 4.

SECTION 2 : LA COLLINÉARITÉ DES VARIABLES EXPLICATIVES

Dans cette situation, les variables explicatives ne sont pas linéairement indépendantes. Cette collinéarité peut être "stricte", se traduisant par une relation "déterministe" du type : $a_i x_{it} + a_j x_{jt} = k$

Cette collinéarité peut être seulement "approchée", auquel cas la relation n'est plus déterministe mais plus large.

Dans tous les cas c'est l'hypothèse H5 qui est prise en défaut. On commencera par exposer les conséquences de la collinéarité, puis on indiquera les techniques permettant de la mettre en évidence, avant d'indiquer quelques transformations usuelles.

A- Conséquences de la collinéarité

En cas de collinéarité stricte

C'est une situation extrêmement rare et qui correspond généralement à une erreur de conception du modèle ou de définition des variables.

Dans ce cas l'ordinateur affiche le message ERROR car $(X'X)^{-1}$ n'existe pas. En effet, la matrice X aura au moins deux colonnes linéairement dépendantes. $X'X$ aura au moins deux lignes et deux colonnes linéairement dépendantes et sera une matrice singulière ($\det = 0$). Il est donc impossible d'en calculer l'inverse.

En cas de collinéarité approchée

Dans ce cas, la matrice $X'X$ n'est pas singulière. Elle tendra vers la singularité c'est-à-dire que son déterminant tendra vers 0. Ceci va impliquer que les valeurs numériques de $(X'X)^{-1}$ seront très grandes puisque le déterminant tend vers 0.

Il en résulte les conséquences suivantes :

- Les estimateurs tendront à avoir des variances élevées
- On risque d'avoir des estimateurs "instables", c'est-à-dire que si on change légèrement les données, (Exemple : modification de la période), le résultat numérique est susceptible de changer fortement.

B- Comment faire apparaître la présence d'éventuelles collinéarités entre les variables explicatives ?

La méthode la plus simple consiste à calculer le tableau des corrélations partielles entre les variables explicatives.

Comme ce tableau est symétrique, on en donne généralement la partie supérieure.

	x_1	x_2	x_j	x_l
x_1		$R_{x_1 x_2}$...	$R_{x_1 x_l}$
x_2				
.				
x_j			$R_{x_i x_j}$	
.				
x_l				

Ce tableau ne donne pas une réponse claire car le coefficient varie entre -1 et 1. Plus on se rapproche de -1 ou de 1, et plus le risque de collinéarité est fort. Il faut donc se donner des règles de décision en fixant des valeurs critiques.

Ex : Si $R < 0,7 \rightarrow$ pas de problème

Si $R > 0,9 \rightarrow$ présomption de collinéarité.

Mais, le tableau ne garantit pas contre toutes les formes de collinéarité. Il renseigne sur la collinéarité éventuelle entre deux variables mais pas sur la collinéarité entre une variable et un groupe de plusieurs variables.

On ne sait donc rien d'une situation du type :

$$\alpha_i x_{it} = \alpha_j x_{jt} + \alpha_h x_{ht} + k$$

La variable i est en relation avec la somme des variables j et h mais ni avec l'une ni avec l'autre des variables, prises indépendamment.

Il peut donc y avoir des situations où il n'y a pas de corrélation apparente entre variables explicatives mais où il y a quand même collinéarité. Ces situations sont heureusement assez exceptionnelles.

De façon générale, dans toute situation de collinéarité on remarque une incohérence entre les différents tests qui permettent d'apprécier la qualité d'une régression.

Souvent, il y a une bonne qualité d'ensemble (c'est-à-dire un R^2 élevé) associé à des coefficients qui ne sont pas significatifs. La raison est que les variances sont très élevées et donc les tests ne valident pas les coefficients.

Symétriquement, la contradiction éventuelle entre une bonne qualité d'ensemble de l'ajustement et le caractère non significatif de certains coefficients de régression peut constituer un indice révélant la présence de collinéarité.

C- L'estimation en présence de collinéarité

- En cas de collinéarité stricte, c'est souvent lié à un problème de définition de variables ou à l'introduction maladroite de variables muettes (dummies).

Les variables muettes sont des variables qui mesurent l'effet particulier lié à une période. Elles prennent la valeur 0 ou la valeur 1. Lorsque l'observation correspond à la période, elles valent 1. Elles prennent la valeur 0 dans les autres cas.

Exemple

Si on s'intéresse à la consommation d'électricité en données trimestrielles (y_t). On va expliquer cette consommation par le revenu R_t . Mais on sait aussi que la consommation d'électricité est influencée par le climat, la température, la luminosité, la longueur du jour, etc.

Des variables muettes traduisent l'influence des saisons sur la consommation.

A pour l'automne.

H pour l'hiver.

P pour le printemps.

E pour l'été.

Un modèle très simple expliquera la consommation du trimestre par le revenu correspondant et l'effet saisonnier.

$$Y_t = a_0 + a_1 R_t + a_2 A + a_3 H + a_4 P + a_5 E + u_t \quad (1)$$

Influence des saisons = variables muettes

On suppose qu'on prend la période IV 1988 → III 1993.

La première observation correspond à l'Automne

$A = 1$ si Automne ($A = 0$ sinon)

$H = 0$

$P = 0$

$E = 0$ etc.

Donc

t	A	H	P	E	Variables muettes saisonnnières
1	1	0	0	0	
2	0	1	0	0	
3	0	0	1	0	
4	0	0	0	1	
5	1	0	0	0	
6	0	1	0	0	
7	0	0	1	0	
8	0	0	0	1	
9	1	0	0	0	

On remarque qu'à chaque période on a $A + H + E + P = 1 \forall t$. Comme la régression (1) comporte un terme constant, la matrice X contient une colonne de termes égaux à 1. Par conséquent, la somme des quatre colonnes constituées par les variables A , H , P et E est égale à la colonne correspondant au terme constant. Les colonnes de X ne sont plus indépendantes.

On a une relation de collinéarité stricte entre les variables et $(X'X)^{-1}$ ne peut être calculée.

Par conséquent quand on introduit des variables saisonnières sous cette forme, il faut laisser au minimum un degré de liberté.

Autre exemple : on s'intéresse aux performances d'un échantillon d'entreprise et on veut tenir compte de l'effet possible du type d'activité. Les entreprises sont classées dans l'un des 3 secteurs :

- . biens d'équipement;
- . biens de consommation;
- . biens d'investissement,

et une variable muette mesure l'effet du secteur. On retrouve exactement le même problème, si l'équation estimée comporte un terme constant.

Les solutions

1. Dans le cas de collinéarité stricte

- Supprimer l'une des variables "dummies" qui correspond à l'effet le plus faible, ce qui revient à choisir une variable comme référence et transformer les autres de sorte qu'elles mesurent les écarts par rapport à cette variable de référence.

Si on prend A comme référence, on a :

$$y_t = a_0 + a_1 R_t + a_2 (A+H+P+E) + (a_3 - a_2) H + (a_4 - a_2) P + (a_5 - a_2) E \quad (2)$$

et : $y_t = c_0 + a_1 R_t + (a_3 - a_2) H + (a_4 - a_2) P + (a_5 - a_2) E$

On mesure les effets différentiels de H, P, E par rapport à A qui sert de référence.

2. Dans les cas de collinéarité approchée, on dispose de toute une panoplie de solutions possibles qui passent par des transformations de l'équation.

On peut faire la somme de 2 variables (quand ça a un sens)

Autre possibilité : normer l'équation c'est-à-dire diviser toutes les variables par l'une d'elle. Cela permet d'éliminer le trend c'est-à-dire la tendance commune aux différentes variables qui peut être à l'origine de la collinéarité.

Exemple

$$c_t = a_0 + a_1 y_t + a_2 c_{t-1} \quad (3)$$

C'est une vieille formulation de la fonction de consommation dans laquelle le revenu et la consommation des ménages suivent les évolutions

très proches en données chronologiques. On peut la transformer en divisant par y_t ou c_{t-1}

$$\frac{c_t}{y_t} = \frac{a_0}{y_t} + a_1 + a_2 \frac{c_{t-1}}{y_t} \quad (4)$$

La signification des coefficients est celle du modèle du départ⁽¹⁾.

On peut aussi normer par le revenu une fonction d'investissement et la variable expliquée sera un taux d'investissement I_t/Y_t .

En normant par le capital on obtient le taux d'accumulation I_t/K_t .

On peut aussi transformer les variables en taux de croissance ce qui supprime les effets tendanciels. Mais ce n'est possible que si la loi qu'on cherche à estimer peut être assimilée à une loi à élasticité constante :

Exemple :

– Pour les fonctions de commerce extérieur ;

– Pour les fonctions d'emploi.

On peut aussi utiliser certaines propriétés de la fonction :

Soit la fonction de production Cobb-Douglas

$$Q_t = AL^\alpha_t K^\beta_t \quad (5)$$

$$\log Q_t = \log A + \alpha \log L_t + \beta \log K_t \quad (6)$$

Marc Guillaume donne l'exemple suivant⁽²⁾ estimé à partir de la fonction de production linéarisée :

$$\log Q_t = 0,28 \log L_t + 0,52 \log K_t + C \quad R^2 = 0,984 \quad (7)$$

(0,82) (0,46)

Les écarts-types des coefficients sont indiqués entre parenthèses.

C'est un exemple typique de collinéarité : le pouvoir explicatif est bon mais aucune des variables n'est significative.

(1) Remarque : si le terme d'erreur de l'équation de départ est homoscédastique, la transformation proposée est susceptible d'introduire de l'hétéroscédasticité dans (4).

(2) Marc Guillaume "Modèles économiques" Presses universitaires de France 1971 page 99.

Dans le cas particulier ci-dessus il y a une absurdité supplémentaire. Si on somme $\alpha + \beta$ on trouve 0,80 (cas de rendements fortement décroissants), ce qui est inacceptable, compte tenu de ce que l'on sait de l'évolution économique de la période sur laquelle porte l'estimation.

De plus, très souvent pour estimer les fonction de production on se base sur les parts distributives. Si les rendements ne sont pas constants cela ne marche pas car c'est en contradiction avec l'élaboration des données.

En posant par hypothèse que les rendements sont constants, il est possible de transformer l'équation qui devient :

$$\text{Log } \frac{Q_t}{L_t} = (1 - \alpha) \log \frac{K_t}{L_t} + c. \quad (8)$$

$$\text{Les estimations donnent : } \log \frac{Q_t}{L_t} = 0,26 \log \frac{K_t}{L_t} + c \quad (9)$$

(0,08)

$$\begin{cases} \alpha = 0,74 \\ \beta = 0,26 \end{cases}$$

La transformation de la fonction permet donc d'éliminer le problème.

Mais parfois, aucune des transformations précédentes ne donne de bons résultats. On peut alors utiliser une méthode d'estimation par balayage qui va consister à se donner des valeurs plausibles pour l'un des coefficient afin de calculer les autres.

Exemple

$$y_t = a_1 x_{1t} + a_2 x_{2t} + a_3 \quad (10)$$

On se donne a_1 et aux valeurs de a_1 vont être associées celles de a_2 et de a_3 .

On va calculer une nouvelle variable : $y_t - a_1 x_{1t}$

$y_t - a_1 x_{1t} = a_2 x_{2t} + a_3$ et on régresse pour obtenir \hat{a}_2 et \hat{a}_3 .

On fait ce travail pour un ensemble de valeurs de a_1 et on retient celle des valeurs de a_1 qui minimise la somme des carrés des résidus ainsi que les valeurs de a_2 et de a_3 qui lui sont associées.

Mais cette procédure par balayage pose plusieurs problèmes :

- Le choix du paramètre : est-ce a_1 ou a_2 ?
Ça dépend de la signification des variables et de leur importance respective.
- Les différents critères qui caractérisent le résultat de la régression ne sont pas toujours convergents. Il peut arriver que la somme des carrés des résidus diminue tandis qu'un coefficient cesse d'être significatif. Cela suppose que l'on définisse des critères permettant de sélectionner la solution la plus satisfaisante.

SECTION 3 : L'AUTOCORRÉLATION DES TERMES D'ERREUR

C'est une situation dans laquelle les termes d'erreurs ne sont pas indépendants au cours du temps.

On va introduire une hypothèse selon laquelle le terme d'erreur dépend linéairement de sa propre valeur passée. Cette hypothèse se traduit par :

$$u_t = \rho u_{t-1} + v_t \quad (1)$$

ρ = coefficient d'autocorrélation

v_t = nouveau terme d'erreur qui satisfait aux propriétés habituelles
c'est-à-dire

$$\begin{cases} E(v) = 0 \quad \forall t \\ \text{Variance de } (V) = \sigma_v^2 I \end{cases}$$

Le terme d'erreur de l'équation comprend donc une partie systématique et une partie aléatoire.

v_t est une partie purement aléatoire

ρu_{t-1} est une partie systématique

(1) est une relation possible mais n'est pas forcément la seule car :

- Elle est linéaire
- C'est une autocorrélation du premier ordre car elle ne dépend que de la période antérieure.

(autocorrélation d'ordre 2 : $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + v_t$)

Les différentes formes que peut prendre la relation d'autocorrélation peuvent être testées.

L'analyse du signe des résidus peut donner une première information sur l'existence ou non d'une éventuelle auto-corrélation : plus on observera de régularité au niveau des signes d'erreur, plus on vérifiera l'équation (1).

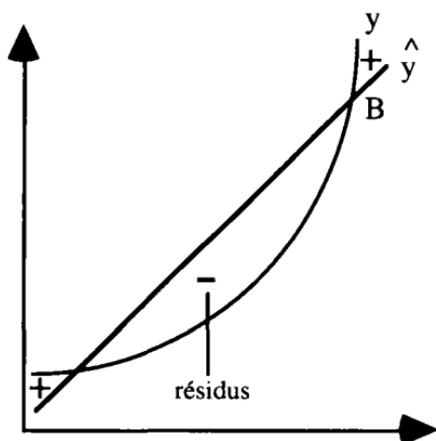
- Si on a une alternance $+/-$, on risque d'avoir une auto-corrélation négative c'est-à-dire $\rho < 0$.

En effet cela traduira le fait qu'à une erreur positive succède une erreur négative et inversement.

- Si on a une succession de résidus du même signe suivie par une succession de termes de signe opposé du type : $+++----++$, une erreur est suivie d'une erreur de même signe, ce sera le signe d'une auto-corrélation positive $\rho > 0$

L'auto-corrélation est un problème statistique qui peut aussi renvoyer à la spécification de l'équation car dans certains cas, elle trouve son origine dans une spécification inadaptée.

Supposons que la vraie fonction soit de type parabolique Y et qu'on l'ait estimé par une fonction linéaire.



La fonction estimée, \hat{y} , va d'abord sous-estimer la variable expliquée, y , puis la surestimer, puis sous-estimer à nouveau. On observera une succession de résidus positifs, puis négatifs et ensuite positifs.

Donc, toute situation d'auto corrélation appelle à une réflexion sur le modèle. (Si il y a un côté systématique dans l'erreur c'est qu'il y a une cause permanente).

- Une distribution non systématique des signes des résidus est compatible avec une absence d'auto-corrélation. Ce n'est évidemment pas un critère très précis et il existe, heureusement, des tests plus performants.

A- Le diagnostic de l'auto-corrélation

On peut tout simplement régresser \hat{u}_t sur \hat{u}_{t-1} et tester le coefficient de régression de l'équation (1) :

$$\hat{u}_t = \rho \hat{u}_{t-1} + v_t$$

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

Ce n'est pas une méthode très performante. On lui préfère le test de Durbin-Watson (DW).

- Test de DURBIN WATSON

$$DW = \frac{2 \sum_{t=1}^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^T (\hat{u}_t)^2}$$

ou

$$DW = \frac{\sum_{t=1}^T (\hat{u}_{t+1} - \hat{u}_t)^2}{\sum_{t=1}^T (\hat{u}_t)^2}$$

Pour un grand nombre d'observation c'est-à-dire quand $T \rightarrow +\infty$,

on montre que :

$$\boxed{DW \rightarrow 2 - 2\rho}$$

En effet :

$$DW = \frac{\sum_1^T \hat{u}_t^2}{\sum_1^T \hat{u}_t^2} + \frac{\sum_1^T \hat{u}_{t-1}^2}{\sum_1^T \hat{u}_t^2} - 2 \frac{\sum_1^T \hat{u}_t \hat{u}_{t-1}}{\sum_1^T \hat{u}_t^2}$$

Les 2 premiers rapports tendent vers 1 car numérateurs et dénominateurs ne diffèrent que par un terme, u_1 ou u_T respectivement.

Le 3^e rapport tend vers $\hat{\rho}$ l'estimateur des MCO dans l'équation de définition (1)

$$\text{Puisque : } \hat{\rho} = \frac{\sum_1^T u_t u_{t-1}}{\sum_1^T u_{t-1}^2}$$

Cette propriété permet d'établir le domaine de variation du test de Durbin WATSON

$$\text{Pour } \rho = 1 \quad DW \rightarrow 0$$

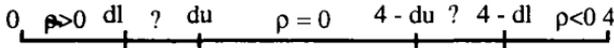
$$\text{Pour } \rho = -1 \quad DW \rightarrow 4$$

Pour $\rho = 0$ $DW \rightarrow 2$ qui est la valeur centrale de ce test.

Plus le test est faible, plus on a de chance d'avoir $\rho > 0$ (c'est-à-dire de l'auto-corrélation positive)

Les valeurs critiques du test de DW tenant compte du nombre d'observation et du nombre de variables explicatives ont été calculées. Le problème est que ces valeurs critiques ne sont pas uniques. Il y a en fait un seuil inférieur et un seuil supérieur. Entre les deux, il y a une zone d'incertitude.

La table de DW est seulement calculée pour les auto-corrélations positives. Les autres se déduiront par symétrie. Il est donc prudent de retracer toutes les situations possibles.



$\left\{ \begin{array}{l} d_l = \text{seuil inférieur} \\ d_u = \text{seuil supérieur} \end{array} \right.$

Remarque : La zone d'incertitude diminue quand le nombre d'observations augmente et elle augmente quand le nombre de variables explicatives augmente.

Dans les tables courantes, elle est maximum pour les valeurs suivantes :

$T = 15$ 5 variables explicatives

$d_l = 0,56$

$d_u = 2,21$

- Il existe une autre table qui a été calculée par THEIL et NAGAR avec un seuil unique (ce qui élimine la zone d'incertitude). On peut donc s'y reporter en cas de doute. Toutefois cette table n'est valable que sous certaines conditions.
- On peut tester des formes plus complexes d'auto-corrélation à l'aide du test de Breush - Godfrey qui appartient à la catégorie des tests sur les multiplicateurs de Lagrange (LM test) présentée dans la section 1. La régression auxiliaire est ici constituée par les résidus de l'équation avec contrainte qui sont régressés sur les variables explicatives et les résidus décalés. Elle permet de tester la significativité globale de l'ensemble des coefficients d'auto-corrélation.

Le modèle sans contrainte est celui qui contient des auto-corrélations. Le modèle avec contraintes est celui qui n'en contient pas, puisque cette absence est équivalente à poser que les coefficients d'auto-corrélation correspondants sont nuls.

Exemple

Soit le modèle général $Y = XA + U$ sur lequel on désire tester une auto-corrélation d'ordre 3.

$$(2) \quad u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + v_t$$

$$\text{et } \rho = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{bmatrix}$$

Le modèle sans contrainte est :

$$(3) \quad Y = XA + Z\rho + V$$

$$\text{où } Z_{(T,3)} = \begin{bmatrix} \hat{u}_{t-1}, \hat{u}_{t-2}, \hat{u}_{t-3} \end{bmatrix}$$

Le modèle avec contrainte est :

$$(4) \quad Y = XA + U$$

Les contraintes sont :

$$\rho_1 = 0 \quad \rho_2 = 0 \quad \rho_3 = 0 \quad \text{ou de façon équivalente } \underset{(3,1)}{\rho} = 0$$

Les hypothèses sont donc :

$H_0 : \rho = 0$ Le modèle contraint est acceptable.

$H_1 : \rho \neq 0$ Le modèle contraint n'est pas acceptable.

Soit \hat{U} le vecteur des résidus du modèle avec contrainte, modèle correspondant à l'équation (4).

La régression auxiliaire consiste à régresser :

$$(5) \quad \hat{U} = XA + Z\rho + V \quad \text{où } Z \text{ est la matrice des résidus décalés de l'équation (4).}$$

Soit R^2 le coefficient de détermination de cette régression. Il mesure le pouvoir explicatif des résidus décalés.

Le test consiste à calculer TR^2 et à comparer avec χ_3^2 pour le seuil de confiance retenu.

Plus les résidus sont corrélés entre eux, plus le R^2 de la régression a des chances d'être élevé et donc plus on a de chance de dépasser les seuils critiques du χ^2 (car plus TR^2 sera élevé).

$$\begin{cases} \text{Si } TR^2 > \chi_\alpha^2 \rightarrow \text{auto-corrélation} \rightarrow H1 \\ \text{Si } TR^2 < \chi_\alpha^2 \rightarrow \text{indépendance} \rightarrow H0 \end{cases}$$

B- Les conséquences de l'auto-corrélation

Les conséquences de l'auto-corrélation sont :

$$- E(UU') \neq \sigma^2 I$$

En effet, les termes d'erreurs n'étant plus indépendants au cours du temps la matrice des variances covariances ne peut plus être une matrice diagonale :

$$E(u_t u_{t'}) \neq 0 \text{ si } t \neq t'$$

On traduit l'auto-corrélation par

$$u_t = \rho u_{t-1} + v_t \quad (1)$$

Cette relation permet de calculer les covariances entre les termes d'erreur :

$$E(u_t u_{t-1}) = E[(\rho u_{t-1} + v_t) u_{t-1}] \quad (6)$$

$$= \rho \sigma_u^2 + E(v_t u_{t-1})$$

$$= \rho \sigma_u^2$$

$$E(u_t u_{t-1}) = \rho \sigma_u^2 \quad (7)$$

$$u_{t-1} = \rho u_{t-2} + v_{t-1}$$

$$u_t = \rho[\rho u_{t-2} + v_{t-1}] + v_t \quad (8)$$

$$u_t = \rho^2 u_{t-2} + \rho v_{t-1} + v_t$$

$$E(u_t u_{t-2}) = E[(\rho^2 u_{t-2} + \rho v_{t-1} + v_t) u_{t-2}] \quad (9)$$

$$= \rho^2 \sigma_u^2 + \rho E(v_{t-1} u_{t-2}) + E(v_t u_{t-2})$$

Donc

$$E(u_t u_{t-2}) = \rho^2 \sigma_u^2$$

soit

$$E(u_t u_{t-i}) = \rho^i \sigma_u^2$$

Remarque : Les termes en v sont postérieurs aux termes en u et ne les contiennent pas donc les deux variables sont indépendantes.

Si il y avait des termes en v antérieurs aux termes en u on n'aurait plus le droit de conclure à l'indépendance car quand on développe u_t selon la formule :

$$u_t = \rho(\rho u_{t-2} + v_{t-1}) + v_t \quad (8)$$

$$u_t = \rho^2 u_{t-2} + \rho v_{t-1} + v_t$$

u_t contient v_{t-1}

On constate que u_{t-1} contient v_{t-2}

u_t se "souvient" des u_{t-i} or dans ces u_{t-i} il y a les v_t de la formule de définition $u_t = \rho u_{t-1} + v_t$

→ les u_t ne sont pas indépendants des v_{t-i} mais les v_t sont indépendants des u_{t-i} .

Exprimons u_t en fonction des $v_t, v_{t-1} \dots$

$$u_t = \rho u_{t-1} + v_t$$

$$u_{t-1} = \rho u_{t-2} + v_{t-1} \quad (10)$$

$$u_{t-2} = \rho u_{t-3} + v_{t-2}$$

$$u_t = \rho (\rho u_{t-2} + v_{t-1}) + v_t$$

$$\begin{aligned} u_t &= \rho (\rho (\rho u_{t-3} + v_{t-2}) + v_{t-1}) + v_t \\ &= \rho^3 u_{t-3} + \rho^2 v_{t-2} + \rho v_{t-1} + v_t \quad (11) \end{aligned}$$

En substituant de proche en proche, on voit qu'il est possible d'exprimer u_t en fonction des termes v_t décalés et des puissances successives de ρ .

$$u_t = \sum \rho^i v_{t-i}$$

Si $E(v_t) = 0 \quad \forall t$ alors $E(u_t) = 0 \quad \forall t$

Les termes d'erreurs restent donc d'espérance nulle.

Conséquences sur la matrice des variances covariances $E(UU')$:

$$E(UU') = \begin{bmatrix} \sigma_{u_1}^2 & \text{cov} u_1 u_2 & \text{cov} u_1 u_T \\ & \cdot & \cdot \\ \text{cov} u_T u_1 & & \sigma_{u_T}^2 \end{bmatrix} \quad (12)$$

La matrice n'est plus diagonale, mais les termes situés sur la diagonale restent égaux entre eux tant que l'hypothèse d'homoscédasticité reste vraie. On peut donc mettre σ_u^2 en facteur puisque tous les éléments de la matrice s'expriment en fonction de σ_u^2 et des puissances de ρ .

$$E(UU') = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ & & \cdot & & \cdot \\ \rho^{T-1} & & & & 1 \end{bmatrix} = \sigma^2 W$$

$$\text{avec } W = \begin{bmatrix} 1 & \rho & \rho^2 \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ & & \cdot & & \cdot \\ \rho^{T-1} & & & & 1 \end{bmatrix}$$

Cette matrice W fait que l'estimation par les MCO n'est plus correcte puisque $E(UU') \neq \sigma_u^2 I$. De plus, il faut distinguer deux cas de figure :

- Équation sans variable endogène décalée (1).
- Équation avec variable endogène décalée (2).

Dans un modèle à une équation, le terme variable endogène désigne la variable expliquée. On donnera dans le chapitre suivant une interprétation plus complète.

Ainsi l'équation $y_t = a_1 x_{1t} + a_2 x_{2t} + u_t$ ne contient pas de variable endogène décalée.

L'équation $y_t = a_1 x_{1t} + a_2 x_{2t} + b_1 y_{t-1} + u_t$ contient y_{t-1} , variable endogène décalée.

• **Équation sans endogène décalée**

Le biais résultant de l'application directe des MCO porte uniquement sur la variance des estimateurs. c'est-à-dire que les estimateurs eux même demeurent sans biais : $E(\hat{A}) = A$

Preuve

La matrice des variances covariance n'intervient pas dans le calcul de ces estimateurs et donc ils n'ont pas à être affectés. Par contre la variance de \hat{A} va être biaisée.

$$v(\hat{A}) = E\left[(X'X)^{-1} X'UU' X(X'X)^{-1}\right]$$

avec $E(UU') = \sigma^2 W$

Il vient :

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

alors que les MCO calculeront :

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

Quel est le sens du biais introduit ?

- En cas d'auto-corrélation positive c'est-à-dire $\rho > 0$

L'ordinateur utilisera la matrice I au lieu de la matrice W

Si $\rho > 0$ alors on va généralement sous-estimer la variance $V(\hat{A})$

Cela implique un risque d'accepter pour valable un coefficient qui n'est pas significatif (On surestime effectivement les tests de Student).

- En cas d'auto-corrélation négatives c'est-à-dire $\rho < 0$.

Les termes non diagonaux de V sont tantôt positifs tantôt négatifs et la conclusion est moins évidente.

Remarque : Le fait que les estimateurs soient sans biais $E(\hat{A}) = A$ ne doit pas être interprété comme un signe d'exactitude mais simplement comme le fait qu'ils ont autant de chance d'être biaisés dans un sens que dans l'autre.

En particulier, quand on estime une équation sans endogène décalé par les MCO et par une procédure de correction de l'auto-corrélation on ne trouve pas les mêmes résultats numériques. Cette différence est normale et provient de la différence dans les méthodes d'estimation.

Donc, même si il n'y a pas de variable endogène décalée en présence d'auto-corrélation il faut quand même utiliser une procédure appropriée. Par contre si il n'y a pas d'auto-corrélation, on n'a pas le droit d'utiliser une méthode d'estimation comportant une procédure de correction.

Quelle relation y a-t-il entre la variance de u et la variance de v ?

$$u_t = v_t + \rho v_{t-1} + \dots + \rho^n v_{t-n}$$

$$E(u_t)^2 = E(v_t^2 + \rho^2 v_{t-1}^2 + \dots + \rho^{2n} v_{t-n}^2) \quad (13)$$

$$E(u_t)^2 = \sigma^2 v(1 + \rho^2 + \dots + \rho^{2n})$$

d'où

$$\sigma_u^2 = \frac{1}{1 - \rho^2} \sigma^2 v$$

Ce qui implique que la variance de u est finie si et seulement si $|\rho| < 1$.

• Équation avec endogène décalée

$$(14) \quad y_t = a_1 x_t + b_1 y_{t-1} + a_2 + u_t$$

$$(1) \quad u_t = \rho u_{t-1} + v_t$$

En décalant d'une période l'équation 14, il vient :

$$(15) \quad y_{t-1} = a_1 x_{t-1} + b_1 y_{t-2} + a_2 + u_{t-1}$$

y_{t-1} variable explicative de (14) contient et n'est pas indépendante de u_{t-1}

Or, d'après (1), u_t le terme d'erreur de l'équation (1) contient lui aussi u_{t-1} .

Donc, dans l'équation (14) on a une variable explicative y_{t-1} qui n'est pas indépendant du terme d'erreur u_t .

y_{t-1} et u_t ont en commun u_{t-1}

L'hypothèse H6 d'indépendance des variables explicatives et du terme d'erreur n'est donc plus satisfaite.

Dans ce second cas on a à la fois un biais sur les variances des estimateurs mais également un biais sur les estimateurs eux-mêmes

Un exemple très simple permet de le vérifier :

Soit le modèle

$$y_t = ay_{t-1} + u_t \quad (15)$$

$$u_t = \rho u_{t-1} + v_t \quad (1)$$

$$\text{d'où } u_t = y_t - ay_{t-1} \rightarrow u_{t-1} = y_{t-1} - ay_{t-2} \quad (16)$$

$$v_t = u_t - \rho u_{t-1}$$

$$\text{donc } v_t = y_t - ay_{t-1} - \rho (y_{t-1} - ay_{t-2}) \quad (17)$$

$$\text{soit } v_t = y_t - (a + \rho) y_{t-1} + (a\rho) y_{t-2} \quad (18)$$

$$\text{et } y_t = (a + \rho) y_{t-1} - (a\rho) y_{t-2} + v_t \quad (19)$$

L'équation (19) traduit l'ensemble des équations (15) et (1).

Dans (19) il y a seulement v_t comme terme d'erreur qui satisfait toutes les propriétés d'un terme d'erreur.

- Estimons l'équation (15) directement par les MCO :

$$\hat{a} = \frac{\sum y_{t-1} y_t}{\sum y_{t-1}^2} \text{ est l'estimateur des moindres carrés ordinaires de l'équation (15).}$$

- Estimons (15) en remplaçant y_t par sa valeur dans (19) :

$$\begin{aligned} \hat{a} &= \frac{\sum [(a + \rho) y_{t-1} - a\rho y_{t-2} + v_t] y_{t-1}}{\sum_i y_{t-1}^2} \\ &= \frac{(a + \rho) \sum y_{t-1}^2 - a\rho \sum y_{t-2} y_{t-1} + \sum v_t y_{t-1}}{\sum_i y_{t-1}^2} \\ &= (a + \rho) - a\rho \frac{\sum y_{t-2} y_{t-1}}{\sum y_{t-1}^2} + \frac{\sum v_t y_{t-1}}{\sum_i y_{t-1}^2} \quad (20) \end{aligned}$$

Donc, on peut déjà constater que $\hat{a} \neq a$.

Le calcul de la limite en probabilité de \hat{a} , $\text{Plim } \hat{a}$, permet de déterminer la valeur vers laquelle \hat{a} tend lorsque le nombre d'observations, T , augmente indéfiniment.

En décalant (15) d'une période

$$y_{t-1} = ay_{t-2} + u_{t-1}$$

on obtient comme valeur de l'estimateur \hat{a} des MCO

$$\hat{a} = \frac{\sum y_{t-1}y_{t-2}}{\sum y_{t-2}^2}$$

qui ne diffère que par un terme du rapport

$$\frac{\sum y_{t-1}y_{t-2}}{\sum y_{t-1}^2}$$

Quand $T \rightarrow +\infty$ cette différence devient négligeable.

d'où

$$\frac{\sum y_{t-1}y_{t-2}}{\sum y_{t-1}^2} = P \lim \hat{a}$$

Quant au dernier terme, son numérateur tend vers 0 puisque v_t et y_{t-1} sont deux variables indépendantes.

Par conséquent, la limite en probabilité de $\hat{a} = a + \rho - a\rho$ $P \lim \hat{a}$

donc

$$P \lim \hat{a} = \frac{a + \rho}{1 + a\rho}$$

Donc, même dans les meilleures conditions, l'estimateur des MCO reste asymptotiquement biaisé. On vérifie que si $\rho = 0$, $P \lim \hat{a} = a$. En l'absence d'auto-corrélation, l'estimateur des moindres carrés ordinaires est sans biais.

- Calculons le sens du biais :

$$P \lim \hat{a} - a = \frac{a + \rho - a(1 + a\rho)}{1 + a\rho}$$

d'où

$$P \lim \hat{a} - a = \frac{\rho(1 - a)}{1 + a\rho}$$

Limitons nous à $|a| < 1$:

- Si $\rho > 0$ alors $\text{Plim } \hat{a} > a \rightarrow$ on surestime a
- Si $\rho < 0$ alors $\text{Plim } \hat{a} < a \rightarrow$ on sous-estime a

Le cas du modèle auto-régressif pur (c'est-à-dire sans vraies variables explicatives) correspond au biais maximum (C'est donc le plus mauvais).

Quand l'équation comporte également de vraies variables explicatives (x_t) le biais a tendance à être plus faible mais il existe toujours.

Par conséquent, les résultats d'une telle estimation ne sont absolument pas fiables.

Mais, facteur aggravant, quand on a ce type de figure (endogène décalée plus auto-corrélation) le test de DW est lui même biaisé. Si on a un problème d'estimation dans (15) alors logiquement on aura aussi un problème dans l'estimation de ρ dans (1). En effet les deux équations sont tout à fait semblables.

Or, le test de DW tend vers $2-2\rho$ ce qui signifie que si $\hat{\rho}$ est biaisé alors le résultat du test est lui même biaisé.

En effectuant le calcul de $\hat{\rho}$ on arrive au résultat suivant :

$$\text{Plim } \hat{\rho} = \frac{a\rho(a+\rho)}{1+a\rho}$$

$$\text{Plim } \hat{\rho} - \rho = \frac{a^2\rho + a\rho^2 - \rho - a\rho^2}{1+a\rho}$$

$$\text{Plim } \hat{\rho} = \frac{\rho(a^2-1)}{1+a\rho}$$

Cas $|a| < 1$:

- Si $\rho < 0$ alors $\text{Plim } \hat{\rho} > \rho$
- Si $\rho > 0$ alors $\text{Plim } \hat{\rho} < \rho \rightarrow$ on sous estime ρ

Soit $|\hat{\rho}| < \rho$ dans tous les cas.

$\hat{\rho}$ est donc biaisé vers la valeur 0 et donc le test de DW est biaisé vers sa valeur centrale qui est 2.

Dans tous les cas de figure, le test de DW a tendance à minimiser l'auto-corrélation (il est optimiste).

Cela amène à remplacer le DW par un autre test en présence de variables endogènes décalées. C'est le test h de DURBIN défini par :

$$h = \hat{\rho} \sqrt{\frac{T}{1 - T \text{var}(\hat{a}_1)}}$$

\hat{a}_1 étant le coefficient de régression de la variable endogène décalée.

La valeur du test de Durbin peut être approchée par :

$$h = \left(\frac{2 - DW}{2} \right) \sqrt{\frac{T}{1 - T \text{var}(\hat{a}_1)}}$$

Remarque : pour pouvoir calculer la valeur de ce test, il faut que :

T variance (\hat{a}_1) < 1

La variable h suit une loi normale centrée réduite

$h \rightarrow N(0,1)$

à partir de laquelle on peut faire un test unilatéral

$H_0 : h = 0$

$H_1 : h > 0$ ou $h < 0$ selon le signe de ρ

Seuil critique $h\alpha = 1,65$ au seuil de 5 %

$|h| < 1,65 \rightarrow H_0$ donc pas d'auto-corrélation

$|h| > 1,65 \rightarrow H_1$ donc auto-corrélation

Si ρ est élevé (forte auto-corrélation) alors on aura une valeur élevée pour h. Les fortes valeurs de h sont généralement associées au rejet de l'indépendance.

Exemple : un ajustement destiné à donner une première idée de la fonction de consommation.

Tunisie (1960 - 1990, données annuelles)

$$c_t = 0,539c_{t-1} + 0,391y_t - 21,296 \quad (21)$$

(0,095)
(0,082)
(8,89)

$$R^2 = 0,989$$

$$DW = 1,92$$

Les tests sont bons et le test de DW semble bon.

Test du h de DW :

$$h = \left(\frac{2 - DW}{2} \right) * \sqrt{\frac{T}{1 - TV(\hat{a}_i)}} = 0,04 \sqrt{\frac{31}{1 - 31 * 0,095^2}}$$

$$h = 0,26$$

$h < h\alpha = 1,65 \rightarrow$ HO Pas d'auto-corrélation.

Dans le cas où $T \sigma_{\hat{a}_i}^2 > 1$ Durbin propose deux étapes :

a- Calcul de \hat{u}_t

b- Régresser \hat{u}_t sur les variables explicatives et \hat{u}_{t-1}

Dans notre exemple, $\hat{u}_t = a_1 c_{t-1} + a_2 y_t + a_3 + \rho \hat{u}_{t-1}$ (22)

On teste ensuite $\frac{\hat{\rho}}{\sigma_{\hat{\rho}}} \rightarrow (T - l - 1)$. Dans l'exemple T - 4 soit 27 degrés de liberté.

Remarque : on pourrait sur l'équation (22) faire le test de Breush-Godfrey et utiliser la table du χ^2 à un degré de liberté.

C- Les procédures de correction

Il existe plusieurs méthodes qui ont comme principe commun de passer par une élimination préalable du terme d'erreur u_t .

$$y_t = a_1 x_{1t} + a_2 + u_t \quad (23)$$

$$u_t = \rho u_{t-1} + v_t \quad (1)$$

Cette élimination peut se faire à partir de (1) :

$$u_t = y_t - a_1 x_{1t} - a_2$$

$$u_t - \rho u_{t-1} = v_t$$

$$y_t - \rho y_{t-1} - a_1(x_{1t} - \rho x_{1t-1}) - a_2(1 - \rho) = v_t \quad (24)$$

$$y_t - \rho y_{t-1} = a_1(x_{1t} - \rho x_{1t-1}) + a_2(1 - \rho) + v_t \quad (25)$$

Dans (25) il ne reste que v_t qui est un bon terme d'erreur, présentant les propriétés requises.

• La méthode de Durbin (2 étapes)

La première étape consiste à éliminer u_t de l'équation et à estimer (26) de façon à ne conserver que v_t comme terme d'erreur.

L'opération conduit à :

$$y_t = \rho y_{t-1} + a_1 x_{1t} - a_1 \rho x_{1t-1} + a_2(1 - \rho) + v_t \quad (26)$$

On estime cette équation ce qui fournit des valeurs pour $\hat{\beta}, \hat{a}_1, \hat{a}_1 \rho, \hat{a}_2$.

Mais les estimateurs sont liés entre eux par une contrainte qui est que le coefficient de x_{1t-1} est égal au produit des coefficients de y_{t-1} et de x_t . Or, dans la procédure d'estimation rien ne permet de tenir compte de cette contrainte. Les estimateurs obtenus ne pourront donc pas être retenus. Ils constituent un ensemble d'estimateurs non cohérents.

On va alors en retenir un seul : $\hat{\beta}$ et connaissant $\hat{\beta}$ il sera possible de calculer \hat{a}_1 et \hat{a}_2 .

Soient les nouvelle variables :

$$\begin{cases} y_t^* = y_t - \hat{\beta} y_{t-1} & (27) \end{cases}$$

$$\begin{cases} x_t^* = x_t - \hat{\beta} x_{t-1} & (28) \end{cases}$$

$$y_t^* = a_1 x_t^* + a_2(1 - \hat{\beta}) + v_t \quad (29)$$

La 2^e étape consiste à régresser (29) par les MCO (on a un bon terme d'erreur) ce qui fournit les valeurs de \hat{a}_1 et \hat{a}_2 .

• Méthode de Cochrane Orcutt

C'est une méthode itérative. Elle consiste à partir de l'équation (25) en se donnant une valeur arbitraire de ρ , et à en déduire des valeurs de a_1 et a_2 .

Connaissant ces valeurs de a_1 et a_2 on revient sur une nouvelle valeur de ρ , etc.

Soit ρ_0 la valeur d'amorçage de la procédure :

Connaissant ρ_0 on forme y_t^0 et x_t^0 :

$$\begin{cases} y_t^0 = y_t - \rho_0 y_{t-1} & (30) \end{cases}$$

$$\begin{cases} x_t^0 = x_t - \rho_0 x_{t-1} & (31) \end{cases}$$

Puis on régresse

$$y_t^0 = a_1 x_t^0 + a_2 (1 - \rho) + v_t \text{ et on obtient des valeurs de } a_1 \text{ et } a_2 \text{ soient } a_1^0 \text{ et } a_2^0$$

Connaissant a_1 et a_2 dans l'équation (1) on peut calculer $\hat{y}_t = a_1^0 x_t + a_2^0$

On peut en déduire $\hat{u}_t = y_t - \hat{y}_t$ et on obtient ρ_1 à partir de (1), puisque

$$\hat{\rho} = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2}.$$

Connaissant ρ_1 on recommence la procédure et on calcule

$$\begin{cases} y_t^1 = y_t - \rho_1 y_{t-1} & (32) \end{cases}$$

$$\begin{cases} x_t^1 = x_t - \rho_1 x_{t-1} & (33) \end{cases}$$

et dans (29) on retrouve a_1^1 et a_2^1

Et dans (1) on en déduit ρ_2 etc.

Sargan a montré que cette procédure est convergente c'est-à-dire qu'au bout d'un certain nombre d'itérations les valeurs de ρ , a_1 et a_2 vont se stabiliser.

Il existe deux types de programme possibles :

- On fixe à priori le nombre d'itérations

- On fixe un critère de divergence sur ρ , par exemple 1 % c'est-à-dire que dès que deux valeurs successives de ρ diffèrent de moins de 1 % on considère que la procédure est stabilisée.

En général la convergence est assez rapide, et, elle sera d'autant plus rapide qu'on aura bien choisi la valeur d'amorçage. Comme le DW est en rapport avec la valeur de ρ il est commode de prendre $\rho_0 = 1 - \frac{DW}{2}$ puisque $DW \rightarrow 2 - 2\rho$.

• **Méthode de Hildreth et Liu** (Méthode d'estimation par balayage)

Elle consiste à estimer une succession d'équation (25) pour des valeurs régulièrement croissante ou décroissante de ρ .

Exemple

$$\rho = 0,1 ; 0,2 ; 0,3 ; \dots ; 0,9$$

On construit pour cela de nouvelles variables sur le modèle de y_t^* et x_t^*

À chacune de ces régressions est associée la somme des carrés des résidus et on peut déterminer quelle est la valeur de ρ qui minimise $\sum (\hat{u}_t)^2$. Puis on va repartir avec un pas plus fin.

Ex. : Si la variance minimale est obtenue pour $\rho = 0,6$ ou $\rho = 0,7$.

$$\rho = 0,61 ; \dots ; 0,69$$

Et ceci jusqu'à obtenir le nombre de décimales souhaité (3 décimales en général).

Les programmes de régression utilisent fréquemment la méthode de Cochran-Orcutt.

Exemple :

L'exemple suivant concerne les importations italiennes en données trimestrielles de 1978.1 à 1992.4.

Un premier ajustement a donné :

$$LIM_t = 1,929 LPB_t - 0,140 LP_t + 0,372 D90 - 12,067 \quad (34)$$

(34,28) (-7,58) (12,31) (-15,68)

$$R^2 = 0,97 \qquad DW = 1,14 \qquad \sum \hat{u}_t^2 = 0,047$$

LIM_t est le logarithme des importations en volume, $L\text{ PIB}$ est le logarithme du PIB réel, LP_t est le logarithme de l'indice du déflateur des importations de produits manufacturés et $D90$ est une dummy pour le 1^{er} trimestre de 1990.

Les valeurs critiques du test de Durbin-Watson sont :

$$dl = 1,48 \quad du = 1,69$$

On est donc dans un cas d'auto-corrélation positive.

L'estimation par la méthode de Cochrane Orcutt donne

$$LIM_t = 1,916LPIB_t - 0,147LP_t + 0,371D90 - 11,899 \quad (35)$$

(22,34) (-4,91) (12,81) (-10,14)

$$R^2 = 0,97 \quad DW = 2,01 \quad \sum \hat{u}_t^2 = 0,033 \quad \hat{\rho} = 0,417$$

Convergence assurée après 4 itérations.

Dans certains cas, il peut arriver que la procédure ne suffise pas à corriger l'auto-corrélation. Cela peut traduire des problèmes de fond concernant les spécifications retenues. À côté d'une réflexion sur la forme de l'équation, il est alors utile d'explorer d'autres formes possibles d'auto-corrélation en particulier d'ordre supérieur à 1. Des corrections, généralisant la méthode de Cochrane Orcutt à des processus d'ordre supérieur, existent dans la plupart des logiciels et permettent de résoudre ce type de difficulté. Le test de Breusch-Godfrey fournit des indications.

SECTION 4 : L'HÉTÉROSCÉDASTICITÉ ET LA MÉTHODE DES MOINDRES CARRÉS GÉNÉRALISÉS

A- Définition

Par définition, il y a hétéroscédasticité lorsque la variance du terme d'erreur (σ_u^2) n'est pas constante sur l'ensemble des observations (au cours du temps ou sur un échantillon).

S'il n'y a pas en plus auto-corrélation, la matrice des variances-covariances des termes d'erreurs $E(UU')$ reste une matrice diagonale

$$E(UU') = \begin{bmatrix} \sigma_{u_1}^2 & & & 0 \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ 0 & & & & \sigma_{u_T}^2 \end{bmatrix} \quad (1)$$

Mais les éléments diagonaux ne sont plus égaux entre eux.

Cette matrice va s'écrire $\sigma_{(T,T)}^2 V \neq \sigma^2 I$ ce qui entraîne la remise en cause des hypothèses 3 et 4.

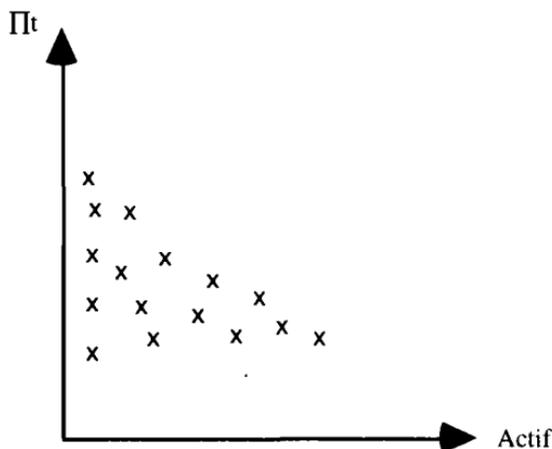
Lorsque l'hétéroscédasticité correspond à la présence d'un élément systématique déterminable dans le terme d'erreur, il est possible de la corriger après l'avoir identifié.

Mais l'hétéroscédasticité peut-être purement aléatoire, ou dûe à des causes inconnues et dans ce cas il est pratiquement impossible d'y remédier.

L'hétéroscédasticité est un problème assez fréquent qui peut être rencontré dans un bon nombre de situations.

Exemple : on s'intéresse au taux de rentabilité d'un ensemble d'entreprises de tailles différentes. On sait par ailleurs que les grandes entreprises ont un taux de rentabilité plus stable que les petites, mais pas nécessairement plus élevé.

Une représentation graphique du taux de profit en fonction de l'actif total de la firme fera apparaître la forme caractéristique du nuage de points :



Supposons maintenant que l'on cherche à évaluer l'influence de la concentration du marché et de la taille de la firme sur sa rentabilité, ce qui a constitué un thème de recherche important dans les années 60 et 70.

On cherche à estimer l'équation :

$$\underset{\substack{\downarrow \\ \text{taux de profit}}}{\Pi_i} = a_0 + a_1 \underset{\substack{\downarrow \\ \text{part de} \\ \text{marché}}}{S_i} + a_2 \underset{\substack{\downarrow \\ \text{valeur des} \\ \text{actifs}}}{A_i} + u_i \quad (2)$$

Dans une telle équation on aura de l'hétéroscédasticité. Le résultat sera différent selon qu'on tient compte ou non de la présence d'hétéroscédasticité. De fait, Hall et Weiss en 1967 ont montré que les résultats des études sur les effets de la concentration étaient substantiellement modifiés lorsqu'on tenait compte de l'hétéroscédasticité⁽¹⁾.

B- Comment faire apparaître la présence d'hétéroscédasticité ?

Il existe différents tests permettant de faire apparaître l'existence éventuelle de l'hétéroscédasticité :

(1) Hall et Weiss "Firm size and profitability". The review of Economics and statistics. Août 1967.

- Le test de GOLDFELD & QUANDT
- Le test de GLEJER

• **Le test de Goldfeld & Quandt**

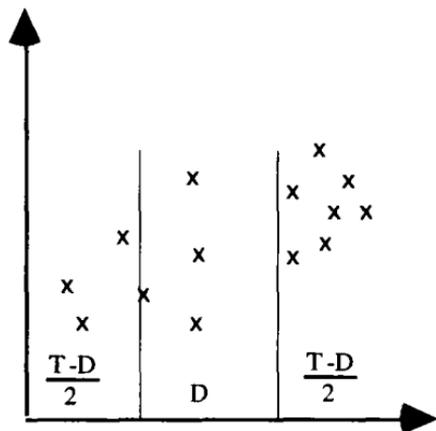
Il part de l'idée que si il y a hétéroscédasticité ayant un caractère systématique, c'est aux deux extrémités du nuage qu'elle sera la plus visible.

Il faut donc exclusivement s'intéresser aux extrémités du nuage. Pour cela on supprime D observations situées au centre de l'échantillon.

Si on a T observations : on ne garde que $\frac{T-D}{2}$ observations aux deux extrémités.

Sur chacune de ces extrémités on va effectuer la régression et on retiendra $\sum \hat{u}_{1i}^2$ et $\sum \hat{u}_{2i}^2$ puis on cherche à vérifier si la variance résiduelle est la même à partir de ces deux régressions.

Pour cela on procède à un test d'analyse de la variance qui s'appuie sur le rapport des deux variances, en plaçant au numérateur la plus forte variance.



Exemple : on suppose que la variance résiduelle est plus forte dans la première sous-période.

$$\sum u_{1t}^2 > \sum u_{2t}^2$$

$$\text{on a } \frac{\frac{\sum u_{1t}^2}{T-D-l}}{\frac{\sum u_{2t}^2}{T-D-l}} \rightarrow \text{Fisher Sned.} \left(\frac{T-D-2l}{2}, \frac{T-D-2l}{2} \right) \quad (3)$$

l = nombre de variables explicatives

Soit $F\alpha$ le seuil critique de la table de Fisher Snedecor, on teste

$$\begin{cases} H0 : \text{homoscédasticité} \\ H1 : \text{hétérosécédasticité} \end{cases}$$

La règle de décision étant la suivante :

- $F < F\alpha \rightarrow H0$
- $F \geq F\alpha \rightarrow H1$

Si les variances sont vraiment différentes, le rapport sera élevé et donc plus la valeur de ce rapport augmente, plus on a de chance de conclure H1.

Remarque : Le test de Goldfeld et Quandt se généralise aisément à 2 sous-périodes de durée quelconque T_1 et T_2 . Dans ce cas :

$$\frac{\frac{\sum \hat{u}_{1t}^2}{T_1-l}}{\frac{\sum \hat{u}_{2t}^2}{T_2-l}} \rightarrow F(T_1-l, T_2-l) \quad (4)$$

• Le test de Glejer

Il consiste à régresser la valeur absolue ou le carré des résidus sur l'ensemble des variables explicatives.

On utilise généralement le carré du résidu \hat{u}_t^2 qui est facile à obtenir après une régression

$$|\hat{u}_i| \text{ ou } \hat{u}_i^2 = a_0 + a_1 x_{1i} + \dots +$$

On s'intéresse au test de Student sur chacun des coefficients de régression.

Si le test de Student classique conclut que tous les coefficients de régression sont non significatifs. $a_j = 0$ ($j = 1 \dots, l$) on conclut à l'homoscédasticité.

Si $a_j \neq 0$ c'est-à-dire qu'il y a au moins un coefficient significatif on conclura à l'hétéroscédasticité et, en même temps on en déduit qu'il y a une des variables explicatives qui explique l'ordre de grandeur des termes d'erreur.

Autrement dit le test de GLEJER donne deux informations :

- La présence d'hétéroscédasticité
- La cause de cette hétéroscédasticité

C- Correction de l'hétéroscédasticité et méthode des Moindres Carrés généralisés

L'hétéroscédasticité ne concerne que la matrice des variances-covariances des termes d'erreur et n'affecte pas les estimateurs des MCO qui demeurent sans biais. Par contre, les variances de ces mêmes estimations et donc les tests de Student seront biaisés. Il est donc nécessaire de corriger l'hétéroscédasticité.

Pour cela on peut utiliser une méthode proposée par AITKEN dans les années 30 connue sous le nom de méthode des Moindres Carrés Généralisés qui permet d'obtenir des estimateurs sans biais \forall la matrice $E(UU')$.

Supposons donc que $E(UU') = \sigma^2 V$, V étant une matrice définie positive c'est-à-dire que pour tout vecteur $Z_{(T,1)}$ alors $Z'VZ > 0$ ($Z'VZ$ est un scalaire).

Cette propriété joue ici le même rôle que la condition qu'une variable soit positive pour admettre des racines carrées réelles.

On peut mettre V sous la forme du produit de deux matrices M et M' : $V = MM'$ de format (T,T) M étant une matrice non singulière admettant une inverse M^{-1} .

V est une matrice symétrique $V = V'$, par construction.

$$\text{De même, si } V = MM' \text{ alors } V^{-1} = M'^{-1} M^{-1}$$

En multipliant à droite et à gauche on retrouve I :

$$M^{-1} V M^{-1} = I$$

Il est possible d'utiliser cette propriété pour corriger l'hétéroscédasticité :

On part du modèle : $Y = XA + U$ (5)

Soient les nouvelles variables construites en prémultipliant par M^{-1} :

$$\begin{cases} Y^* = M^{-1}Y & (6) \end{cases}$$

$$\begin{cases} X^* = M^{-1}X & (7) \end{cases}$$

$$\begin{cases} U^* = M^{-1}U & (8) \end{cases}$$

Sur le nouveau terme d'erreur défini on a :

$$E(U^*U^{*\prime}) = E(M^{-1} U U' M^{-1}) \quad (9)$$

$$= M^{-1} E(UU') M^{-1} \quad (10)$$

$$= M^{-1} \sigma^2 V M^{-1} = \sigma^2 I$$

• Les nouveaux termes d'erreurs satisfont donc toutes les propriétés des MCO.

Vérifions que l'estimateur \hat{A} des MCO sur les données transformées est sans biais :

$$\hat{A} = (X^{*\prime} X^*)^{-1} X^{*\prime} Y^*$$

$$= (X' M^{-1} M^{-1} X)^{-1} X' M^{-1} (M^{-1} XA + M^{-1} U) \quad (11)$$

$$\text{Soit } \hat{A} = A + (X' V^{-1} X)^{-1} X' V^{-1} U \quad (12)$$

donc $E(\hat{A}) = A$ si $E(U) = 0$

L'estimateur est sans biais.

Quant à la variance de l'estimateur sur l'équation transformée,

$$V(\hat{A}) = E \left[(\hat{A} - E(\hat{A}))(\hat{A} - E(\hat{A}))' \right]$$

$$= (X' V^{-1} X)^{-1} X' V^{-1} E(UU') V^{-1} X (X' V^{-1} X)^{-1} \quad (13)$$

$$= \sigma^2 (X' V^{-1} X)^{-1} X' V^{-1} V V^{-1} X (X' V^{-1} X)^{-1} \quad (14)$$

(matrice symétrique)

$$= \sigma^2 (X' V^{-1} X)^{-1}$$

$$= \sigma^2 (X^* X^*)^{-1} \quad \text{car} \quad (X^* = M^{-1} X) \quad (15)$$

$$(X' V^{-1} X = X' M^{-1} M^{-1} X = X^* X^*)$$

• La variance de \hat{A} est donc donnée par les MCO sur les variables transformées

Le problème est cependant de connaître M . Cela suppose d'avoir des informations sur la matrice V . Mais dans la plupart des cas ces informations ne sont pas disponibles. C'est pourquoi la méthode des Moindres Carrés généralisée n'est applicable que dans un nombre limité de cas.

C'est ce qui fait l'intérêt du test de GLEJER puisque ce test peut fournir l'information dont on a besoin.

Si ce test indique par exemple que la variance du taux de profit est inversement proportionnelle aux actifs, dans l'équation (2), on peut corriger l'hétéroscédasticité en utilisant cette information.

$$V = \begin{bmatrix} 1/A_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ & & & & 1/A_K \end{bmatrix} \quad (16)$$

V est donc assez bien spécifiée.

On peut mettre V sous la forme $V = MM'$ d'où :

$$M = \begin{bmatrix} 1/\sqrt{A_1} & & & & \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & 1/\sqrt{A_K} \end{bmatrix} \quad (17)$$

$$M^{-1} = \begin{bmatrix} \sqrt{A_1} & & & & \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \sqrt{A_K} \end{bmatrix} \quad (18)$$

Donc, l'application des Moindres Carrés généralisés revient à multiplier toutes les observations par : $\sqrt{A_i}$

Et, quand les variables sont transformées à faire la régression sur les variables transformées.

Remarque : La méthode des Moindres Carrés Généralisés (MCG) est aussi applicable en cas d'auto-corrélation des termes d'erreurs puisque là aussi on connaît la nature de la matrice V.

$$V = \begin{bmatrix} 1 & \rho & \cdot & \cdot & \cdot & \rho^{T-1} \\ \rho & \cdot & & & & \cdot \\ \cdot & & \cdot & & & \cdot \\ \cdot & & & \cdot & & \cdot \\ \cdot & & & & \cdot & \cdot \\ \rho^{T-1} & \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix} \quad (19)$$

Cette matrice admet un inverse V^{-1}

$$V^{-1} = \begin{bmatrix} 1 & -\rho & 0 & 0 & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & 0 & 0 \\ 0 & 0 & 0 & 0 & -\rho & 1 \end{bmatrix} \quad (20)$$

Qui peut être mis sous la forme $V^{-1} = M^{-1} M^{-1}$

$$\text{avec } M^{-1} = \frac{1}{\sqrt{1-\rho^2}} \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & 0 \\ -\rho & 1 & & \\ 0 & -\rho & 1 & 0 \\ 0 & & 0 & -\rho & 1 \end{bmatrix} \quad (21)$$

L'application des MCG conduit à régresser sur les variables transformées.

$$\begin{cases} Y^* = M^{-1}Y \\ X^* = M^{-1}X \end{cases}$$

Compte tenu de la forme de M^{-1} , appliquer les Moindres Carrés généralisés revient à très peu de choses près à l'application de l'estimation sur l'équation (25) de la section 3.

Dans la plupart des cas où la matrice des variances-covariances est inconnue on ne peut pas appliquer les Moindres Carrés généralisés. C'est pour cela qu'on distingue les Moindres Carrés généralisés et les Moindres Carrés généralisés "faisables".

Exemple : (on laisse de côté la question de la pertinence du modèle qui sera abordée au chapitre 4).

Une régression simple du logarithme de la masse monétaire (M_2) sur le log du PIB, le taux de chômage, le taux d'inflation et le coût d'opportunité sur 1980.1-1993.4 donne :

$$LM_2 = 1,259 LP_{IB} + 0,03 ch - 9,89 Inf + 0,025 CO - 0,928 \quad (22)$$

(6,20) (2,45) (-4,18) (1,84) (-0,65)

$$R^2 = 0,92 \quad DW = 0,46$$

Après correction de l'auto-corrélation, le résultat obtenu est aberrant avec en particulier une élasticité revenu négative.

Le test de Glejer sur cette régression donne :

$$\hat{u}_t^2 = -0,0006 LP_{IB} - 8,4 \cdot 10^{-5} ch - 0,011 inf - 1,710^{-5} CO + 0,005$$

(-1,26) (-2,50) (-1,94) (-0,48) (1,50)

$$R^2 = 0,13$$

Après transformation des variables en utilisant $(ch)^{0,5}$ on obtient :

$$LM_2 = 1,143LPIB + 0,022ch - 10,74inf + 0,022CO - 0,044 \quad (24)$$

(5,69)
(1,75)
(-4,91)
(1,62)
(-0,03)

$$R^2 = 0,99 \quad DW = 0,48$$

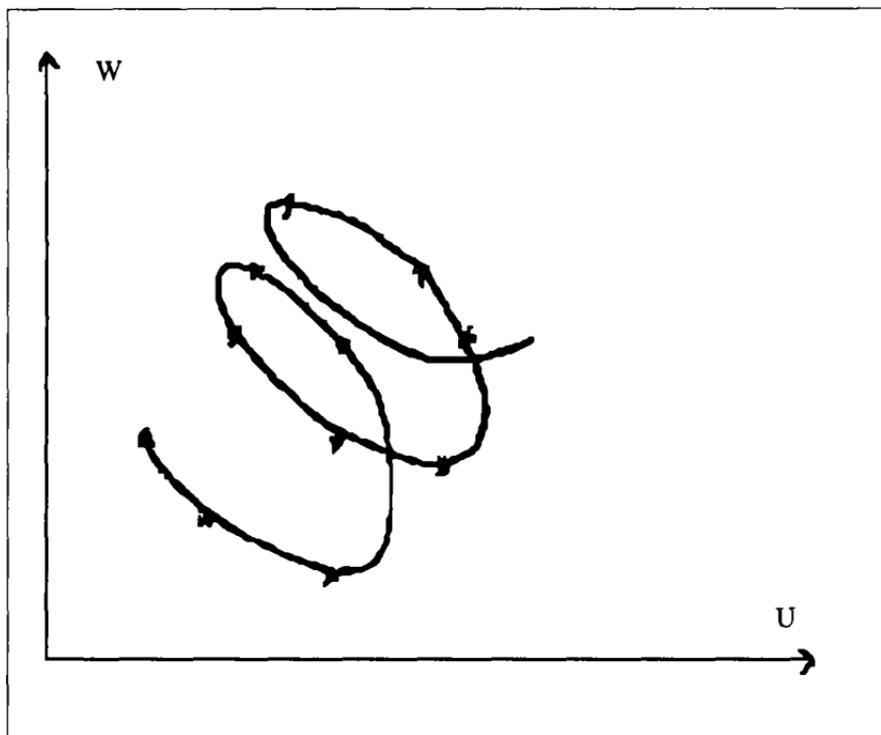
SECTION 5 : PROBLÈMES LIÉS À L'INSTABILITÉ D'UNE RELATION AU COURS DU TEMPS OU À L'HÉTÉROGÉNÉITÉ D'UNE POPULATION

Il arrive souvent qu'on travaille sur des échantillons non homogènes. (Exemple : population de salariés où il y a des hommes et des femmes). De même on peut imaginer que certaines relations ne sont pas stables au cours du temps et qu'elles se déplacent (Exemple : quand le cadre institutionnel change, certaines relations se modifient).

Dans certains cas il est possible de séparer la période ou l'échantillon en deux parties et de travailler séparément. Ce n'est pas toujours possible quand on n'a pas assez d'observations ou quand on s'intéresse justement au déplacement. D'où la nécessité d'estimer sur l'ensemble de la période sachant qu'il y a un problème d'hétérogénéité ou de stabilité. Dans tous les cas, le fait d'ignorer ces phénomènes peut conduire à des absurdités.

Exemple : la courbe de Phillipps sur période longue.

$$\begin{cases} w & \text{taux de croissance des salaires} \\ u & \text{taux de chômage} \end{cases}$$



Sur longue période, les conditions de l'arbitrage inflation-chômage tendent à se dégrader.

Sur chacun des cycles ainsi décrit on peut ajuster des courbes de Phillipps qui auront une élasticité négative mais la courbe d'ajustement sur l'ensemble des points peut très bien avoir une pente positive.

Donc l'estimation sur l'ensemble des observations peut aboutir à des absurdités. Il faut alors vérifier qu'il y a bien stabilité ou homogénéité. À défaut, il faut adopter un modèle un peu différent pour en tenir compte et travailler dans des conditions convenables.

On peut tester la pertinence de l'opération en utilisant des tests de restriction.

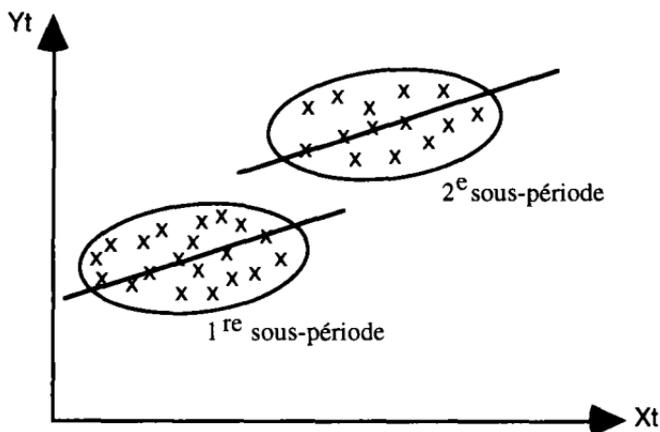
Pour cela, on considère que :

- La stabilité est un cas particulier de l'instabilité ;
- L'homogénéité est un cas particulier de l'hétérogénéité.

A- Comment s'assurer qu'il y a homogénéité ou stabilité ?

Il existe pour cela plusieurs méthodes.

• Premier test



Une première méthode est de chercher si les ajustements effectués sur les deux nuages sont différents. Si le nombre de point est suffisant on peut régresser séparément sur les deux sous-populations ou sous-périodes et à partir de là tester l'égalité des coefficients de régression.

Mais ce test est peu performant car cela revient à construire des intervalles de confiance à ± 2 écarts-types et à regarder s'ils ont une partie commune.

• Deuxième test

On peut régresser sur l'une des sous-population et se demander si cette régression est aussi bonne pour la deuxième sous-population. On calcule des écarts résiduels sur les deux sous-populations et on cherche à savoir si la variance résiduelle de la deuxième sous-population est aussi bonne que celle de la sous-population 1 (test

d'homogénéité). Sinon il y a hétéroscédasticité : la variance résiduelle est plus forte sur la 2^e sous-population que sur la première.

• Troisième test

Le test le plus performant est le test de Chow. Supposons qu'on ait les deux sous-périodes d'observation ayant T_1 et T_2 observations respectivement et l variables explicatives. Ce test demande qu'on fixe à priori la date à partir de laquelle la relation change ou qu'on définisse ce qu'il y a exactement dans les deux sous-populations.

À partir de là 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 variances résiduelles.

On commence par régresser sur la population totale soit T observations.

$$T = T_1 + T_2$$

Puis on refait la même régression sur chacune des 2 sous-populations séparément et on retient la somme des carrés des résidus de chaque régression.

Soit :

$$\text{sur } T_1 \rightarrow \sum \hat{u}_{1t}^2$$

$$\text{sur } T_2 \rightarrow \sum \hat{u}_{2t}^2$$

$$\sum \hat{u}_{1t}^2 + \sum \hat{u}_{2t}^2 \text{ doit être } < \sum \hat{u}_t^2$$

mais dans des proportions qui dépendent de l'amélioration apportée par la prise en compte de l'hétérogénéité des données.

On va calculer

$$\frac{\sum \hat{u}_t^2 - (\sum \hat{u}_{1t}^2 + \sum \hat{u}_{2t}^2)}{l} \quad (1)$$

Nombre de degré de liberté du numérateur = $T - l - (T_1 - l + T_2 - l) = l$

que l'on rapporte à $\frac{\sum \hat{u}_{1t}^2 + \sum \hat{u}_{2t}^2}{T_1 + T_2 - 2l}$

$$\text{soit } \frac{\frac{\sum \hat{u}_1^2 - (\sum \hat{u}_{11}^2 + \sum \hat{u}_{21}^2)}{l}}{\frac{\sum \hat{u}_{11}^2 + \sum \hat{u}_{21}^2}{T_1 + T_2 - 2l}} = F$$

Ce rapport suit une loi de Fisher Snedecor avec respectivement l degrés de liberté pour le numérateur et $(T - 2l)$ degrés de liberté pour le dénominateur. L'ordre numérateur / dénominateur est imposé par le test.

Interprétation

Au numérateur on trouve l'amélioration apportée par la prise en compte de l'hétérogénéité et au dénominateur la plus faible des variances résiduelles.

Soit $F\alpha$ le seuil critique de la table, et F la valeur calculée par le test. On pose :

$H\emptyset$ = homogénéité ou stabilité ;

$H1$ = hétérogénéité ou instabilité.

Si la relation est très instable le gain obtenu est important et on aura une valeur élevée au numérateur.

Donc si $F > F\alpha$ on conclura $H1$

si $F < F\alpha$ on conclura $H\emptyset$

Ce qui fait l'intérêt du test de Chow c'est qu'il est sensible car, au numérateur il porte à la marge sur des variations et non sur les variances elles-mêmes.

Dans cette version, il faut cependant déterminer la partition de la population totale.

Quand il s'agit de phénomènes chronologiques, ce sont des événements extérieurs qui vont donner des indications sur la période charnière. Souvent la période charnière sera celle qui maximisera le test de Chow. Certains logiciels permettent de calculer une succession de tests de Chow pour différentes périodes.

Exemple : la fonction d'importation de l'Italie étudiée dans la section 3 est-elle stable ?

En observant le graphique des résidus, on a remarqué un changement dans la qualité de l'ajustement avant et après le 4^e trimestre 1987.

La régression des importations sur le PIB, les prix et la constante est refaite pour les 2 sous-périodes 1978.1-1987.4 et 1988.1-1992.4.

Les sommes des carrés des résidus sont respectivement :

$$\sum \hat{u}_{1t}^2 = 0,0329$$

$$\sum \hat{u}_{2t}^2 = 0,0423$$

L'ajustement sur la période entière donne $\sum \hat{u}_t^2 = 0,1774$ ⁽¹⁾

Le test de Chow vaut donc :

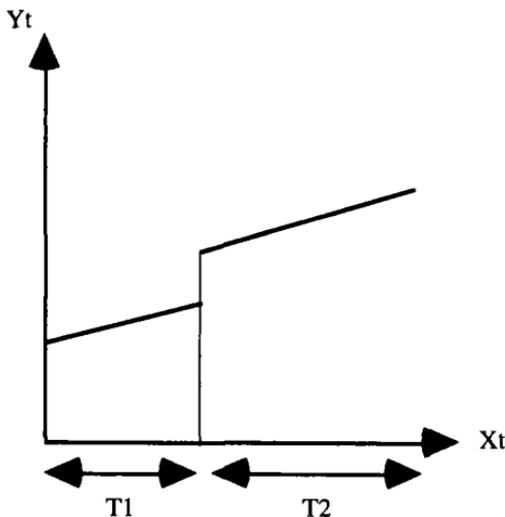
$$\frac{0,1774 - (0,0329 + 0,0423)}{\frac{3}{54}} = 24,4$$

La valeur critique $F_{\alpha}(3,54) = 2,78$.

La fonction est donc instable.

B-La prise en compte d'un éventuel changement

- Premier cas : la fonction se déplace et garde la même forme.



(1) Il s'agit du modèle sans la variable muette D 90, dont la variance résiduelle est nettement supérieure à celle de l'équation présentée à la section 3.

Exemple

Soit une fonction d'importation de la forme.

$$\text{Log Im}_t = a_1 \log y_t + a_2 \log P_{rt} + a_3 \quad (1)$$

On peut rendre compte de ce déplacement en introduisant une variable muette qui servira à caractériser l'une des deux périodes. Tout dépend de ce qu'on choisit comme situation de référence.

Si on considère que la référence est la première période, on caractérisera la seconde période :

$$L = 0 \text{ si } t < T_1 \text{ ou}$$

$$L = 0 \text{ si } t \in T_1$$

$$L = 1 \text{ si } t \geq T_1 \text{ ou}$$

$$L = 1 \text{ si } t \in T_2$$

Ce qui conduit à ajouter une quatrième variable explicative : L dans la régression (1)

$$\text{Log Im}_t = a_1 \log y_t + a_2 \log P_{rt} + a_3 + a_4 L \quad (2)$$

$$(T,4) \left[\begin{array}{cc|c|c} \text{Log } y_1 & \text{Log } P_r & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \text{Log } y_T & & 1 & 1 \end{array} \right]$$

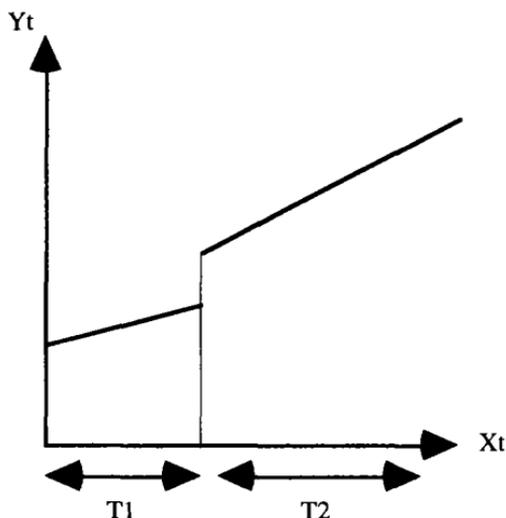
Interprétation des coefficients

a_4 mesure le déplacement de la fonction d'importation qui peut être engendré par un événement en T_2 , par exemple, un changement institutionnel comme la libéralisation du commerce extérieur.

Remarque :

Si on avait estimé l'ensemble par une relation unique tous les coefficients auraient été faux. L'introduction de L rend compte d'un saut unique sur les importations.

• **Second cas : La fonction se déplace et se déforme**



La pente de la fonction est donc différente en T_2 de sa valeur en T_1

Soit la fonction $\text{Log Im}_t = a_1 \log y_t + a_2 \log P_{rt} + a_3$ (1)

On peut introduire sur le coefficient a_1 un effet propre à la deuxième période ce qui va nous conduire à estimer la régression suivante :

$$\text{Log Im}_t = (a_1 + a_4 L) \log y_t + a_2 + \log P_{rt} a_3 + a_5 L \quad (5)$$

a_4 va mesurer la variation de l'élasticité revenu des importations qui est ici imputable à la seconde période.

L'équation modifiée va rendre compte de ce relèvement d'élasticité.

On désigne parfois ce type de fonction sous le nom de modèle "spline"

$$X_{(T,4)} = \left[\begin{array}{cccc} \text{Log } Y_1 & 1 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 0 & 0 \\ \hline \cdot & \cdot & 1 & \text{Log } Y_t \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \text{Log } Y_T & 1 & 1 & \text{Log } Y_T \end{array} \right] \left. \begin{array}{l} \vphantom{X_{(T,4)}} \\ \vphantom{X_{(T,4)}} \end{array} \right\} \begin{array}{l} T_1 \\ T_2 \end{array}$$

Pour savoir si ce modèle est performant, il existe deux méthodes :

- voir si le modèle est validé c'est-à-dire si a_4 est significativement $\neq 0$:

$$\frac{\hat{a}_4}{\hat{\sigma}\hat{a}_4} \rightarrow \text{Student}$$

Et de même pour a_5

- Utiliser un test de restriction à priori. En effet on peut toujours considérer le modèle général comme une version appauvrie du modèle particulier (avec a_4 et a_5). Quand on passe du modèle "Spline" au modèle général on diminue la qualité de l'ajustement. Si la perte en pouvoir explicatif est importante alors le modèle Spline est meilleur que le modèle général c'est-à-dire qu'il y a bien deux périodes séparées. Sinon, si le modèle général est aussi bon que le modèle Spline, la fonction est homogène ou à peu près stable.

Le modèle sans déplacement est un cas particulier. Cela donne la possibilité de tester la validité de l'hypothèse faite sur le déplacement.

On revient par ce moyen au test de restriction à priori..

Soit H = matrice des restrictions ($q,1$) (autant de lignes que de restrictions, autant de colonnes que de coefficients à estimer)

$$HA = h \quad (6)$$

La contrainte peut être que :

- Le terme constant des deux régressions soit le même, ce qui implique que la valeur du coefficient de régression de la variable muette soit nul.

Soit l'équation

$$y_t = a_1 x_t + a_2 + a_3 L \quad (7)$$

On impose $a_3 = 0$

$$A = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \quad H = (0 \ 0 \ 1) \quad h = 0$$

$$(0 \ 0 \ 1) \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = 0 \quad (8)$$

Le modèle dans lequel on impose une restriction est moins bon que le modèle sans restriction. Ce modèle aura donc une $\sum \hat{u}_t^2 >$ à celui du modèle sans restriction.

La différence entre la somme des carrés des résidus avec restriction et sans restriction nous renseignera sur le caractère acceptable ou non de la restriction. Cela servira à effectuer un test sur la validité des restrictions posées :

H_0 $HA = h$ c'est-à-dire que les estimateurs peuvent respecter la contrainte.

H_1 $HA \neq h$ c'est-à-dire que le modèle est non compatible avec la contrainte retenue.

Soit $\sum w_t^2$ la somme des carrés des résidus du modèle avec restriction et $\sum \hat{u}_t^2$ la somme des carrés des résidus du modèle sans restriction.

On a alors

$$\frac{\sum w_t^2 - \hat{\sum u}_t^2}{\frac{\sum \hat{u}_t^2}{T-1}} \rightarrow F(q, T-1)$$

Ce rapport mesure dans quelle proportion la restriction dégrade le modèle.

Soit F_{α} , la valeur critique de la table de Fisher. Si $F < F_{\alpha}$ alors l'imposition d'une contrainte ne dégrade pas beaucoup la qualité de la régression et on conclut donc H_0 .

Si $F > F\alpha$ alors la contrainte diminue de façon importante la qualité du résultat et on conclura alors H_1 .

Remarque : Comme on l'a vu, les tests de restriction à priori sont d'utilisation très variée.

Ils permettent de tester toutes sortes de contraintes de forme linéaire du type $\alpha_1 a_1 + \alpha_2 a_2 = Z$ telles que rendements constants, homogénéité de degré 1 d'une fonction demande par rapport aux prix etc.

Néanmoins une utilisation intéressante des tests de restriction à priori est de servir à l'analyse de variantes d'une même régression. Par rapport à un modèle de base cela servira à évaluer ce qu'apporte une spécification moins générale, plus précise. Cela peut aussi servir à comparer des variantes d'un même modèle explicatif, comme par exemple :

- . variante avec retard / variante sans retard ;
- . variante avec une structure de retard / variante avec une autre structure.

Tous ces problèmes qui relèvent du choix de l'économètre peuvent être éclairés par des tests de restriction à priori.

SECTION 6 : LES RETARDS D'AJUSTEMENT

Les variables économiques ne réagissent pas, le plus souvent, instantanément mais avec un certain délai d'ajustement.

Il faut un délai de perception, puis le temps de faire des choix, de les mettre en application. Tout cela ne peut pas se faire de façon instantanée.

En conséquence, l'explication instantanée d'une variable par une ou plusieurs autres ne rend compte que d'une partie de la réalité et il faut distinguer les réactions de court terme et les effets de long terme.

Avec un modèle du type

$$y_t = a_1 x_t + a_2 x_{t-1} + a_3 x_{t-2} + \dots + a_n x_{t-n} + u_t \quad (1)$$

→ l'effet de court terme est $\frac{\Delta y_t}{\Delta x_t} = a_1$.

→ l'effet de long terme est $\sum_i a_i$.

En effet, si on applique un accroissement de x constant sur toute la période de référence de $i = 0$ à $i = n$ on retrouve $\Delta y = \sum a_i \Delta x$

Un tel modèle est dit à retards échelonnés. C'est donc un modèle dans lequel y dépend des valeurs passées de x . On peut aussi avoir :

$$y_t = a_1 y_{t-1} \dots a_n y_{t-n} + b_1 x_{it} \dots b_n x_{nt} + u_t \quad (2)$$

C'est un modèle auto-régressif, dans lequel la variable expliquée dépend de ses propres valeurs passées et le cas échéant des valeurs présentes et passées d'autres variables explicatives.

Il est commode d'introduire un opérateur qui est l'opérateur de retard désigné par L tel que

$$\begin{cases} Lx_t = x_{t-1} \\ L^i x_t = x_{t-i} \end{cases}$$

Cet opérateur possède les propriétés mathématiques habituelles.

Il peut être utilisé pour réécrire les modèles avec retard.

Ainsi (1) peut s'écrire :

$y_t = A(L) x_t$ c'est-à-dire une suite de retards échelonnés sur la variable x .

La structure générale avec tous les types de retard pourra s'écrire :

$$y_t = a_1 y_{t-1} \dots a_n y_{t-n} + b_0 x_t + b_1 x_{t-1} + \dots + b_k x_{t-k} + u_t \quad (3)$$

$$\text{soit } A(L)y_t = B(L)x_t + u_t \quad (3bis)$$

Un modèle du type (3) ne peut pas dans le cas général être estimé directement. Il y a, en général, différents problèmes économétriques qui rendent l'opération impossible. Au lieu d'avoir une structure simple avec u_t on aura souvent une structure de retard complexe sur u_t et l'estimation par les MCO sera impossible. Même la forme (1) fera apparaître une collinéarité des variables explicatives, dans un bon nombre de situations.

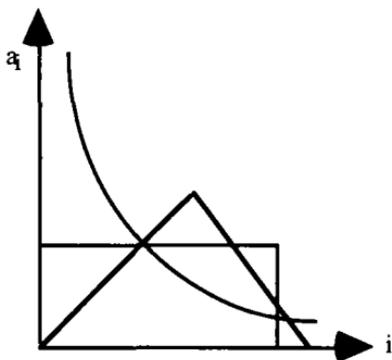
Comme on ne peut pas estimer directement les formes retardées, on recourt à des distributions de retards qui vont permettre une estimation à partir d'un modèle modifié :

Il existe deux types de structures de retards :

- . les distributions à priori;
- . les retards polynomiaux d'Almon. (Polynomial distributed lags. PDL en anglais).

A- Les distributions à priori

Ce sont des distributions que l'on se donne et qui permettent de situer les coefficients retardés les uns par rapports aux autres.



Soit le modèle $y_t = \sum a_i x_{t-i} + u_t$

Plusieurs distributions à priori sont utilisables :

La distribution rectangulaire

On suppose que l'influence de la variable décalée est la même pendant k périodes puis s'estompe.

$a_i = a$ pour $i = 0, \dots, k$

$a_i = 0$ pour $i > k$

$y_t = a [x_t \dots x_{t-k}] + u_t$ (4) et on n'a donc qu'un seul coefficient à estimer.

La distribution en V renversé

Elle permet de traduire une influence croissante jusqu'à un maximum puis décroissante, de la variable explicative.

Supposons que cette influence augmente pendant les deux premières périodes de retard, passe par un maximum pour $i = 2$ et redescende ensuite.

La spécification de ce processus est :

$$y_t = a_0 [x_t + 2x_{t-1} + 3x_{t-2} + 2x_{t-3} + x_{t-4}] + u_t \quad (5)$$

Il y a ici également un seul coefficient à estimer.

Cette formulation simpliste est assez bien adaptée au problème de l'investissement : l'influence des variables décalées (taux d'intérêt, profit) est plus importante que celle de la période contemporaine.

Remarque : Le modèle pourrait aussi être considéré comme une version contrainte d'un modèle plus général et on pourra en tester la validité.

La distribution à poids géométriquement décroissants

$a_i = \lambda^i$ avec $0 < \lambda < 1$ Les coefficients retardés sont fonction des puissances successives d'un coefficient $\lambda < 1$

$$\text{Exemple : } y_t = ax_t + a\lambda x_{t-1} + a\lambda^2 x_{t-2} \dots a\lambda^n x_{t-n} \quad (6)$$

Un tel modèle peut être transformé en modèle auto-régressif par l'intermédiaire de la transformation de KOYCK.

En décalant d'une période l'équation (6) et en la multipliant par λ , il vient :

$$\lambda y_{t-1} = a \lambda x_{t-1} + a \lambda^2 x_{t-2} + \dots a \lambda^{n+1} x_{t-n-1} \quad (7)$$

En soustrayant membre à membre, il vient

$$y_t - \lambda y_{t-1} = ax_t - a \lambda^{n+1} x_{t-n-1} \quad (8)$$

Soit encore

$$y_t = ax_t + \lambda y_{t-1} - a \lambda^{n+1} x_{t-n-1} \quad (9). \text{ Le dernier terme tend vers } 0 \text{ quand } n \text{ augmente.}$$

d'où

$$y_t = ax_t + \lambda y_{t-1} \quad (10)$$

La transformation permet donc de passer d'un modèle à retard échelonné à un modèle auto-régressif.

La transformation de KOYCK pose un problème économétrique puisque si dans l'équation 4 les termes d'erreurs ont les propriétés requises (3-4) ils ne peuvent plus les avoirs dans (10). Dans (10) on a effectivement un terme d'erreur composé ($u_t - \lambda u_{t-1}$)

Il faut distinguer les effets de CT et de LT dans (6)

- A CT $\rightarrow \frac{\Delta y_t}{\Delta x_t} = a$

- A LT : supposons que l'on applique une variation de x égale à Δx sur toute la période considérée.

$$x_{t-i} \text{ devient } x_{t-i} + \Delta x \quad \forall i$$

En conséquence y_t augmente de Δy_t que l'on peut calculer :

$$\begin{aligned} \Delta y_t + y_t &= a(x_t + \Delta x) + a\lambda(x_{t-1} + \Delta x) + \dots + a\lambda^n(x_{t-n} + \Delta x) \\ &= \left[ax_t + a\lambda x_{t-1} + \dots + a\lambda^n x_{t-n} \right] + a\Delta x + a\lambda\Delta x + \dots + a\lambda^n \Delta x \end{aligned}$$

Le terme situé entre crochets n'est autre que y_t lui-même.

D'où il vient : $\Delta y_t = a[\Delta x + \lambda\Delta x + \dots + \lambda^n \Delta x]$ soit en supposant que n est suffisamment grand.

$$\Delta y_t = \left(\frac{a}{1-\lambda} \right) \Delta x$$

Il y a donc bien 2 effets distincts

$$\left\{ \begin{array}{l} \text{effet de court terme } \frac{\Delta y_t}{\Delta x_t} = a \\ \text{effet de long terme } \frac{\Delta y_t}{\Delta x} = \frac{a}{1-\lambda} \end{array} \right.$$

λ étant le coefficient de l'endogène décalée dans (10).

On voit que l'introduction de l'endogène décalée y_{t-1} dans une fonction peut être interprétée comme la traduction d'un mécanisme de retard avec des poids géométriquement décroissants. L'endogène décalée est le signe d'un retard d'ajustement.

Soit le modèle $y_t = a_1 x_{1t} + a_2 x_{2t} + a_3 y_{t-1}$ (11)

Pour faire apparaître ses propriétés de long terme on applique un Δx_1 permanent d'où :

$y_t + \Delta y = a_1 (x_{1t} + \Delta x_1) + a_2 x_{2t} + a_3 (y_{t-1} + \Delta y)$ puisque en régime permanent y_{t-1} augmente de Δy

$y_t + \Delta y = a_1 x_{1t} + a_2 x_{2t} + a_3 y_{t-1} + a_1 \Delta x_1 + a_3 \Delta y$ (12)

d'où $\Delta y = a_1 \Delta x_1 + a_3 \Delta y$

Soit $\Delta y = \frac{a_1}{1 - a_3} \Delta x_1$ (13)

Un mécanisme voisin : l'ajustement partiel

y_t = valeur courante

y_t^* = valeur de long terme (dans certain cas c'est la valeur d'équilibre ou la valeur désirée).

Entre les deux valeurs il y a un mécanisme d'ajustement partiel qui peut s'écrire :

$y_t - y_{t-1} = \lambda (y_t^* - y_{t-1})$ avec $0 < \lambda < 1$ (14)

La valeur de long terme peut se définir par une certaine fonction par exemple :

$y_t^* = a_1 + a_2 x_t$

$y_t - y_{t-1} = \lambda (a_1 + a_2 x_t) - \lambda y_{t-1}$

soit

$y_t = a_1 \lambda + a_2 \lambda x_t + (1 - \lambda) y_{t-1}$

Cette équation traduit l'ajustement du niveau effectif de la variable à ses déterminants, ici x_t .

Si on cherche la réponse de long terme de y_t à x_t on trouve ici a_2 : c'est égal au coefficient de régression x_t sur un moins le coefficient de régression de y_{t-1} :

$$a_2 = \frac{a_2 \lambda}{1 - (1 - \lambda)}$$

– On voit que deux choses assez différentes conduisent finalement à la même spécification :

- un phénomène de retard dans la réponse d'une variable à une autre (équation 6) ;
- un phénomène d'ajustement partiel (équation 14).

Le mécanisme d'ajustement partiel s'écrit souvent sous une forme à élasticité constante :

$$\left[\frac{y_t}{y_{t-1}} \right] = \left[\frac{y_t^*}{y_{t-1}} \right]^\lambda \quad \lambda = \text{coefficient d'élasticité}$$

On peut l'appliquer, par exemple, à l'emploi : dans un premier temps, en inversant une fonction de production, on détermine l'emploi nécessaire.

$$Q : f(K, L)$$

$$\text{d'où } L^* = f^{-1}(Q, K).$$

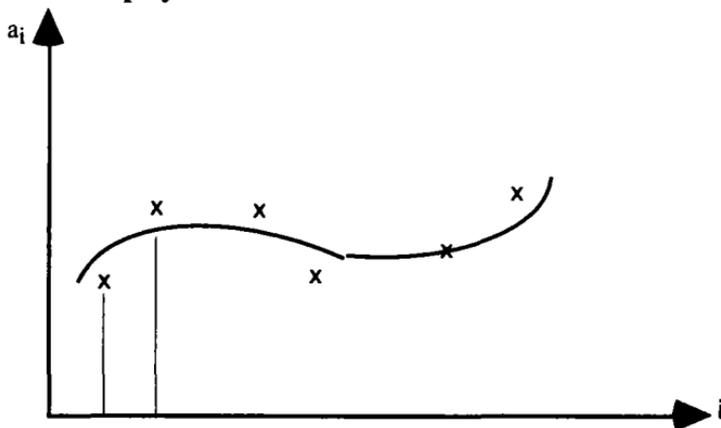
Mais on sait que l'emploi effectif ne s'ajuste pas immédiatement à l'emploi désiré d'où l'appel à un mécanisme d'ajustement. On fera ainsi apparaître des vitesses d'ajustement selon les branches, selon les pays,...

Remarque : si à chaque période on ajuste 20 %, le délai d'ajustement sera

$$\frac{1 - \lambda}{\lambda} \text{ soit 4, dans ce cas.}$$

Toutes ces modalités de retard sont des retards à priori. On préférera généralement utiliser la méthode des retards polynomiaux d'Almon.

B- Les retards polynomiaux d'Almon



Soit le modèle :

$$y_t = a_0 x_t + a_1 x_{t-1} + \dots + a_n x_{t-n} + u_t \quad (15)$$

ou

$$y_t = \sum_i^n a_i x_{t-i} + u_t$$

Comme on ne peut pas directement estimer (15) on utilise un artifice consistant à estimer une fonction polynomiale qui sera suffisamment proche des valeurs vraies des coefficients. Soit F cette fonction polynomiale. On remplacera les vrais valeurs des coefficients par leur valeur dans la fonction. Il faut donc que cette fonction soit une bonne approximation.

Après avoir estimé la fonction on pourra déterminer ainsi n'importe quel coefficient a_i . Cette opération est une application du Théorème de Weirstrass qui établit que n'importe quelle fonction continue et bornée sur un intervalle peut être approchée par un polynôme de degré convenable.

On suppose donc qu'il existe une vraie fonction inconnue qu'on peut approcher par une fonction continue.

Une fonction polynomiale est une fonction de type :

(16) $a_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2 \dots \alpha_r i^r \rightarrow$ polynôme de degré r

Pour connaître les paramètres de F il faut connaître les α_j afin de pouvoir calculer n'importe quelle valeur de a .

Cette opération n'a de sens et d'intérêt que si le nombre de coefficients α est inférieur au nombre de coefficients a car :

- Ce serait illogique d'estimer plus de coefficients que dans la forme de départ..
- Sachant que l'on a $(n+1)$ coefficients a et $(r+1)$ coefficients α , si $r > n$ le polynôme n'est plus défini de façon unique, c'est-à-dire qu'on pourra faire passer plusieurs polynômes par l'ensemble des points qui ont été identifiés.

La méthode consiste à estimer les coefficients α de la relation (16) pour ensuite calculer les a . Pour cela il existe plusieurs méthodes possibles.

La méthode simplifiée d'Almon consiste à réécrire l'équation de départ en fonction des coefficients α et à les estimer par les moindres carrés.

Supposons que l'on ait :

$$\begin{cases} n = 4 \\ r = 2 \end{cases}$$

$$a_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2 \quad (17)$$

$$y_t = a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2} + a_3 x_{t-3} + a_4 x_{t-4} \quad (18)$$

$$\begin{cases} a_0 = \alpha_0 \\ a_1 = \alpha_0 + \alpha_1 + \alpha_2 \\ a_2 = \alpha_0 + 2\alpha_1 + 4\alpha_2 \\ a_3 = \alpha_0 + 3\alpha_1 + 9\alpha_2 \\ a_4 = \alpha_0 + 4\alpha_1 + 16\alpha_2 \end{cases}$$

Remplaçons les a_j par leur valeur en fonction des α :

$$y_t = \alpha_0 x_t + (\alpha_0 + \alpha_1 + \alpha_2) x_{t-1} + (\alpha_0 + 2\alpha_1 + 4\alpha_2) x_{t-2} + (\alpha_0 + 3\alpha_1 + 9\alpha_2) x_{t-3} + (\alpha_0 + 4\alpha_1 + 16\alpha_2) x_{t-4} \quad (19)$$

En mettant en facteurs les α dans (19) :

$$y_t = \alpha_0(x_t + x_{t-1} + x_{t-2} + x_{t-3} + x_{t-4}) \\ + \alpha_1(x_{t-1} + 2x_{t-2} + 3x_{t-3} + 4x_{t-4}) \\ + \alpha_2(x_{t-1} + 4x_{t-2} + 9x_{t-3} + 16x_{t-4}) \quad (20)$$

Dans chacune des parenthèses on retrouve de nouvelles variables soit Z_0, Z_1, Z_2 .
On peut donc écrire :

$$y_t = \alpha_0 Z_0 + \alpha_1 Z_1 + \alpha_2 Z_2 \quad (21)$$

La méthode simplifiée consiste à régresser y sur Z_0, Z_1, Z_2 . pour obtenir des estimateurs $\hat{\alpha}_0, \hat{\alpha}_1$ et $\hat{\alpha}_2$.

Ces estimateurs servent pour calculer les valeurs de a_i .

(Remarque : il existe une autre méthode de résolution, faisant appel aux polynômes d'interpolation de Lagrange.)

Exemple : on s'intéresse aux déterminants de l'emploi en données trimestrielles dans le secteur de l'énergie. Les premiers résultats font apparaître une relation négative entre l'emploi et le capital productif ainsi qu'une influence négative du taux de salaire. Considérant que la réponse de l'emploi aux variations du taux de salaire n'est pas instantanée, on introduit un polynôme de retard sur cette variable. La durée totale des retards est de 8 trimestres et le degré retenu est deux.

$$n = 8$$

$$r = 2$$

Sur la période 1906. I - 1994. IV le résultat obtenu est :

$$LN = -1,330 LK + \sum_1^8 \alpha_i LW_{t-i} + 25,864$$

(8,98) (14,17)

$$R^2 = 0,998$$

$$DW = 1,85$$

$$\hat{\rho} = 0,179$$

$$a_0 = -0,07857 \quad a_1 = -0,05812 \quad a_2 = -0,04060$$

(4,40) (5,40) (7,50)

$$a_3 = -0,02601 \quad a_4 = -0,01435 \quad a_5 = -0,00562$$

(7,92) (3,00) (0,901)

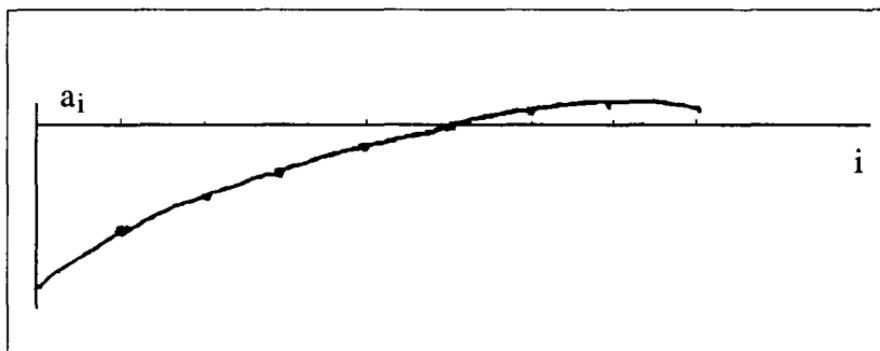
$$a_6 = 0,00018 \quad a_7 = 0,00305 \quad a_8 = 0,00299$$

(0,026) (0,532) (0,847)

$$\Sigma a_i = -0,21706$$

(8,83)

Les variables sont en logarithme. Les chiffres entre parenthèses sont les t de Student.



Dans cette procédure on peut contrôler le choix du degré du polynôme. Il y a deux paramètres importants : n et r :

- n est souvent dicté par la théorie économique ;
- le degré du polynôme : plus il est élevé, plus les calculs sont lourds, plus on augmente le risque de collinéarité dans la méthode ci-dessus puisque chaque degré correspond une variable explicative.

(en passant de $r = 2$ à $r = 3$ on rajoute une variable explicative).

D'un autre côté plus le degré du polynôme est élevé et plus on a de chance de se rapprocher de la vraie distribution. Il y a donc un arbitrage à effectuer entre la lourdeur des calculs et la précision.

Une bonne politique consiste à minimiser le degré du polynôme pour un degré de précision acceptable.

• On peut utiliser les informations à priori disponibles sur la classe de fonction estimée.

– Ex. : Si on estime une fonction d'investissement : la répartition des coefficients retardés à tendance à monter, puis diminuer avant de remonter.

Dans ce cas une fonction du troisième degré autorise deux extremums locaux.

Pour l'investissement : $r = 3$ est un bon point de départ.

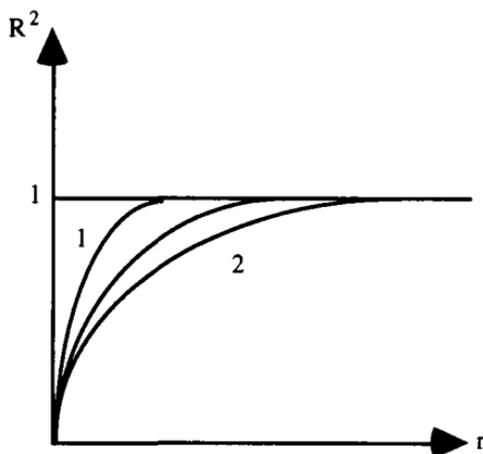
Pour une fonction d'exportation : la répartition des coefficients a la forme d'un V renversé et un polynôme du second degré sera généralement suffisant.

• On peut également suivre le pouvoir explicatif de l'équation (21) en fonction du degré choisi pour le polynôme.

Plus le degré du polynôme est élevé, plus le pouvoir explicatif augmente. Mais cette augmentation peut être très rapide ou au contraire lente.

Dans le premier cas, il semble que l'on ait un bon résultat avec un degré faible.

Dans le second cas, il peut être utile d'explorer des degrés plus élevés.



Dans la pratique on dépasse rarement le troisième degré en essayant de maximiser l'écart entre r et n .

Sur le plan pratique pour obtenir une estimation par les retards polynomiaux d'Almon, les bons logiciels se contentent d'une seule instruction (instruction PDL), qui demande que soient fixés le nombre de retards et le degré du polynôme.

En comparant l'équation de départ et l'équation estimée :

$$y_t = a_0 x_t + a_1 x_{t-1} + \dots + a_n x_{t-n} \quad (15)$$

$$y_t = \alpha_0 z_{1t} + \dots + \alpha_r z_{rt} \quad (22)$$

L'équation (22) peut être considérée comme une version de l'équation (15) avec des restrictions. On peut donc essayer de valider globalement la procédure des retards d'Almon en utilisant la procédure générale de test de restriction.

L'équation (22) est nécessairement moins bonne en terme d'ajustement que l'équation (15). Dans quelle mesure est-elle moins bonne ? C'est à cette question que répond le test de restriction.

Pour cela on calcule

$$F_{(n-r, T-l)} = \frac{\frac{RSS(22) - RSS(15)}{n-r}}{\frac{RSS(15)}{T-l}}$$

Soit F_α le seuil critique, on teste

H_0 : procédure valide

H_1 : procédure non valide

$$\begin{cases} \text{Si } F \text{ calculé} > F_\alpha \rightarrow H_1 \\ \text{Si } F \text{ calculé} < F_\alpha \rightarrow H_0 \end{cases}$$

RSS (22) dépend du degré du polynôme. Il est donc possible de tester différentes variantes de degré.

$(n - r)$ est le nombre de restrictions qui joue le rôle de q dans le test précédent.

• La procédure des retards polynomiaux d'Almon est d'emploi général : pratiquement toute fonction qui présente des retards peut-être traitée par des retards de type polynomiaux. C'est une procédure par défaut qui peut être utilisée quand il n'y a pas d'idées préconçues sur la véritable nature des mécanismes.

Chapitre II : L'estimation des modèles à équations simultanées

Avec les modèles à équations simultanées, de nouvelles catégories de problèmes se posent, à côté de ceux qui ont fait l'objet du chapitre I. Ils conduisent à l'utilisation de méthodes d'estimation spécifiques dont les mérites sont peut-être plus théoriques que réels.

On commencera par présenter les propriétés internes des modèles avant d'aborder la question de leur estimation.

On peut définir un modèle comme un ensemble de variables organisées selon un système cohérent de relations mathématiques.

C'est un ensemble formalisé.

C'est un ensemble cohérent (intéressant en cas d'interactions ou de co-déterminations).

Les modèles sont bien adaptés pour rendre compte des interdépendances multiples entre les variables économiques, c'est une des raisons de leur utilisation. Ils permettent de prendre en compte simultanément un ensemble de phénomènes qui doivent être étudiés conjointement.

Il existe :

- . des modèles théoriques ;
- . des modèles empiriques qui sont ceux qui nous intéressent principalement.

Parmi les modèles empiriques on s'intéressera à ceux dont les paramètres sont estimés économétriquement.

Le champ couvert par les modèles est très large, identique à celui de l'analyse économique elle-même. Il existe donc des modèles macro-économiques et des modèles micro-économiques.

L'avantage de cohérence peut être illustré par l'exemple suivant. Soit le modèle :

$$c_t = f(y_t, c_{t-1}, \dots) \quad (1)$$

$$I_t = g(y_{t-1}, r_t, \dots) \quad (2)$$

$$y_t = c_t + I_t + z_t \quad (3)$$

En travaillant séparément sur (1) et (2), rien n'assure que les valeurs de c_t et I_t sont compatibles entre elles.

Dans le cadre d'un modèle (1) (2) (3) cette compatibilité est assurée immédiatement par la relation (3).

- Historiquement, on a commencé par travailler sur des fonctions séparées ce qui imposait de vérifier la cohérence des résultats.
- Ce n'est qu'ensuite que l'on a utilisé de grands modèles macro-économiques qui assuraient le résultat souhaité, tout en posant d'autres problèmes.

Dans un modèle il y a donc plusieurs choses qui peuvent être classées en :

- . une partie formelle : des variables organisées en équations ;
- . une logique sous jacente : logique interne que traduit le système.

Il faut donc évaluer différemment et selon des critères différents :

- . la qualité des équations ;
- . le caractère plus ou moins adapté à la situation de la logique interne.

Une même situation économique peut être correctement décrite par des modèles d'inspirations différentes.

L'économétrie n'est pas assez sélective pour faire le tri entre les modèles, d'où l'analyse à deux niveaux. Le choix des modèles ne se fera donc qu'avec une analyse théorique.

Ainsi par exemple la situation de l'économie française pendant les années 70-80 pouvait être décrite par des modèles d'inspirations très différentes⁽¹⁾ :

(1) Sans compter un grand nombre d'autres modèles moins connus... dont la liste serait à peu près impossible à établir : autres modèles construits par l'administration, modèles des instituts de recherche publics et privés, modèles universitaires, modèles des grandes institutions internationales OCDE, FMI, modèles construits par les instituts de recherche étrangers etc.

Modèle de David (Banque de France) : Monétariste ;

Modèle STAR (DP) : Néo Cambridgien ;

Modèle DMS (INSEE - DP) : Néo Keynesien.

Les trois modèles impliquaient des politiques différentes, parfois même opposées en terme de dépense publique ou de politique salariale.

Il n'y a pas de relation assez étroite entre un état de la réalité et un type de modélisation.

Il peut arriver qu'un modèle dont les équations sont très bonnes en termes économétriques se révèle inférieur à un modèle dont individuellement les équations sont nettement moins bonnes, parce que la logique théorique du second est plus conforme à la situation qui est décrite. Cette évaluation n'est donc pas simple à faire et ne peut se faire qu'en recourant à des critères variés concernant à la fois la qualité économétrique des équations, mais aussi la logique interne du modèle. Sur le premier point, certaines équations mal spécifiées peuvent tirer vers le bas la qualité d'ensemble d'un modèle. Cela pose le problème du traitement qui peut être fait pour certaines variables difficiles à modéliser, que l'on sait expliquer mais que l'on préférera considérer comme exogènes ou traiter hors modèles.

Ce sont surtout les relations les moins bien établies qui finalement déterminent la performance globale.

SECTION 1 : LE MODÈLE COMME SYSTÈME D'ÉQUATION

La partie visible d'un modèle, ce sont des variables organisées en équations. On s'intéressera d'abord aux variables puis aux équations.

A- Les variables et les équations

Les variables qui figurent dans un modèle sont réparties en deux catégories :

- . les variables endogènes ;
- . les variables exogènes.

• **Les endogènes** sont celles dont la valeur est déterminée par le modèle lui-même. C'est la résolution du système d'équation qui fixe les valeurs des variables endogènes.

Certaines endogènes peuvent intervenir avec retard : ce sont les endogènes décalées.

Quand on résout le modèle, on le résout en t , par conséquent les valeurs des variables endogènes décalées sont déjà fixées et donc les endogènes décalées ne peuvent pas jouer le rôle de vraies variables endogènes. Sur le plan de la logique du système elles sont analogues à la catégorie des variables exogènes. Sur le plan économétrique elles posent des problèmes particuliers en compliquant la structure des termes d'erreur.

• **Les variables exogènes** sont des variables dont la valeur est déterminée en dehors du modèle. Pour certaines d'entre elles, elles sont exogènes par nature car elles ne relèvent pas d'une explication économique (Exemple : la climatique, la démographie, ...).

Mais les plus nombreuses sont des variables qu'on choisit de traiter comme des variables exogènes, c'est-à-dire qu'on ne cherche pas à les expliquer. De ce fait, la frontière entre variables endogènes et variables exogènes traduit le partage entre ce qui est expliqué et ce qui ne l'est pas. C'est un choix que fait le constructeur du modèle.

• La première question à laquelle devra répondre l'économètre est celle du choix des variables et de la catégorie à laquelle elles appartiennent.

Parmi les exogènes on distingue parfois :

- . les variables d'environnement;
 - . les variables de commande.
- Les variables d'environnement décrivent par exemple la situation de l'économie mondiale (prix de l'énergie, taux de change du \$, etc.).
- Les variables de commande sont celles sur lesquelles les autorités peuvent agir et qui décrivent des choix de politique économique.

Cette distinction est utilisée quand on fait des exercices de prévision qu'on appelle des scénarios : on s'efforce dans un premier temps de caractériser différents états possibles de l'environnement : l'intérêt est de les caractériser de façon aussi nette que possible. Sur chacun de ces états on peut alors faire des simulations avec des politiques économique qui sont elles aussi caractérisées.

Puis on évalue le type de politique économique qui est adapté à un type d'environnement.

C'est un exercice qui se pratique à l'occasion des principales décisions de politique économique :

- . loi de finance;
- . préparation des plans à moyen terme.

Parmi les équations on peut distinguer deux catégories :

- 1- les relations de comportement ;
- 2- les conditions d'équilibre et les identités.

Les relations de comportement

Elles traduisent la réaction d'une variable aux variations d'une ou plusieurs autres.

Exemple : les équations (1) et (2)

Dans un modèle macro-économique, la fonction de production, de consommation, la demande de monnaie, etc. sont des relations de comportement.

La valeur des coefficients des relations de comportement résulte de la procédure d'estimation et c'est ce qui les différencie de la catégorie suivante.

Les identités et conditions d'équilibre

Les coefficients des identités et conditions d'équilibre sont, en effet, connus d'avance.

- Soit que cela résulte de la définition des variables (cas des identités).

Exemple

$$K_t = K_{t-1} + Inv_t$$

- Soit qu'il s'agisse d'une condition que l'on impose (cas des conditions d'équilibre). Exemple : équation (3) du modèle précédent.

La différence est que dans certains cas on peut introduire une variable d'écart sur une condition d'équilibre et associer des résultats à cette variable d'écart. Cela a un sens pour une condition d'équilibre, mais n'a aucun sens pour une identité.

Un modèle complet va donc se présenter sous la forme de :

- n variables endogènes ;
- m variables exogènes ou prédéterminées
- organisées dans h relations de comportement ;
- et k identités ou relations d'équilibre.

Pour que le modèle soit dit fermé ou complet il faut qu'on puisse calculer de façon unique la valeur des variables endogènes. Les endogènes vont jouer le rôle des inconnues dans le système d'équations.

- Si toutes les équations sont de forme linéaire et indépendantes on sait que la condition pour qu'il y ait détermination d'une solution unique est que :

$$n = h + k$$

Cette propriété doit être rapprochée du classement des variables : toute augmentation du nombre d'endogènes entraîne automatiquement une augmentation du nombre d'équations c'est-à-dire de la taille du modèle.

Donc pour pouvoir limiter la taille d'un modèle, il faut limiter le nombre d'endogènes. Il y aura donc un certain nombre de variables qui seront traitées comme exogènes.

- Si certaines des relations ne sont pas linéaires on ne peut plus établir strictement la propriété précédente. Il faut raisonner par élimination et dire que si $n \neq h + k$ il est sûr qu'on ne peut pas avoir une solution unique alors que si la propriété est vérifiée elle n'implique pas nécessairement une solution unique (des équilibres multiples sont possibles, ce qui traduit des propriétés particulières du système).

Un modèle dont toutes les équations sont linéaires se résoud facilement par inversion de matrice.

Un modèle comportant des non-linéarités se résoudra en général par itérations : on se donne des valeurs d'amorçage pour certaines inconnues et on regarde si ces valeurs peuvent être solutions, ce qui nécessite des moyens de calculs puissants et ne garantit pas contre l'existence d'autres solutions locales associées à d'autres valeurs d'amorçage.

Quand la propriété n'est pas réalisée, c'est qu'en général il y a un problème de logique interne dans le modèle : si par ex : $n < h + k$ alors en général il n'y aura pas de solution car on aura imposé trop de contraintes sur l'ensemble des endogènes et sauf exception on aura vraisemblablement des mécanismes qui ne seront pas cohérents entre eux. On redonne alors des degrés de liberté au modèle en transformant des exogènes en endogènes : on endogénéise le nombre nécessaire de variables.

On peut illustrer ce problème par un exemple très simple :

$$y_t = a l_t \quad (4)$$

$$c_t = b_1 y_t + b_2 \quad (5)$$

$$y_t = c_t + z_t \quad (6)$$

On suppose deux endogènes : y_t et c_t

Ici $n < h + k$. Il n'y a pas de solution.

Une variable y_t ici est déterminée de façon indépendante par deux mécanismes :

- . l'offre dans l'équation (4);
- . la demande par (5) et (6).

On peut alors redonner un degré de liberté au modèle en considérant l'une des deux variables l_t ou z_t comme endogène.

Si on choisit l_t , y_t sera déterminé par la demande dans (5) et (6) ; et (4) sert uniquement à déterminer l'emploi.

On peut aussi rencontrer la situation inverse : quand il y a trop d'endogènes par rapport au nombre d'équations. Dans ce cas il y a une indétermination : les solutions ne sont plus déterminées de façon unique, il existe une infinité de solutions. Il faut donc réduire le nombre d'endogènes en considérant certaines variables comme exogènes.

Un modèle s'écrit équation par équation et donc il n'est pas évident que $n = h + k$. On ne peut que le vérifier à la fin.

De plus, la variable de gauche n'est pas forcément une endogène (elle le sera souvent cependant).

Pour faire apparaître les propriétés internes d'un modèle on s'intéresse à deux choses :

- L'organisation de la causalité dans le modèle.
- Les propriétés dynamiques qui sont incorporées dans le système d'équation.

C'est la traduction de la logique contenue dans le modèle.

B- La causalité

La causalité porte sur les relations instantanées entre variables.

Il existe différentes approches du phénomène de causalité. Cela est vrai aussi en langage quantitatif.

En économétrie on retient surtout la définition de Granger qui repose sur l'idée que si x et y sont deux processus stationnaires, x est considéré comme cause de y si la prédiction de y qui est faite à partir des valeurs passées de y elle-même et des valeurs passées de x est meilleure que la prédiction de y faite uniquement à partir des valeurs passées de y .

Cette notion a donné naissance aux tests de causalité usuels.

Cela s'applique bien à des modèles à une seule équation.

Exemple : relation entre les prix et la masse monétaire.

Dans le cas des modèles à équation simultanée on retient une autre définition qui est restreinte à l'ordre de détermination des variables endogènes. Ce qui implique que l'analyse de la causalité sera restreinte aux seules endogènes. On retiendra comme définition que y_i est cause de y_j si y_i est déterminée avant et sert à déterminer y_j .

Pour faire apparaître les relations de causalité il faut donc faire apparaître l'ordre dans lequel seront déterminées successivement les différentes variables endogènes, ce qui revient à faire apparaître l'ordre de résolution du modèle. On raisonne toujours dans le cas d'équations supposées linéaires et on utilise les propriétés des systèmes et sous-systèmes d'équations linéaires.

On peut donner une représentation simple de la méthode. Pour cela utilisons un exemple.

Soit le modèle suivant :

$$(7) \quad c_t = a_1 y_t + a_2$$

$$(8) \quad I_t = b_1 \text{Im}_t + b_2 y_{t-1} + b_3$$

$$(9) \quad \text{Im}_t = c_1 x_t - \Delta r_t$$

$$(10) \quad y_t + \text{Im}_t = c_t + I_t + x_t$$

$$(11) \quad y_t = d_1 l_t$$

c_t = consommation

y_t = revenus

Im_t = importation

x_t = exportations

I_t = investissement

l_t = emploi

Δr_t = variation de réserves extérieures

Il y a donc cinq équations. Il faut cinq variables endogènes :

$$c_t, y_t, I_t, \text{Im}_t, l_t$$

Le modèle est donc fermé c'est-à-dire qu'il admet un ensemble de valeurs uniques pour les variables endogènes.

Les exogènes sont : 1, y_{t-1} , x_t , Δr_t

Comment faire apparaître un ordre dans les causalités ?

Définition

Un sous-système d'équations linéaires est dit minimum s'il contient un nombre de variables et d'équations suffisant pour le déterminer.

Un grand système d'équation peut ou non contenir des sous-système plus petits.

Déterminer la causalité revient à hiérarchiser les sous-systèmes par taille croissante.

Pour cela on peut s'aider d'un tableau dans lequel figurent en ligne les équations et en colonnes les variables.

Contient la variable		c_t	y_t	I_t	Im_t	l_t
1 = oui	(7)	1	1	0	0	0
0 = non	(8)	0	0	1	1	0
	(9)	0	0	0	1	0
	(10)	1	1	1	1	0
	(11)	0	1	0	0	1

On peut donc déterminer les sous-systèmes :

(9) ne contient qu' $Im_t \rightarrow$ c'est donc le plus petit sous-système car il n'a besoin d'aucune autre information pour fournir son résultat.

C'est un sous-système d'ordre 0.

On peut réécrire le tableau en permutant les lignes et les colonnes de façon à placer (9) et (8) aux deux premières lignes et Im_t et I_t aux deux premières colonnes.

	Im_t	I_t	y_t	c_t	l_t
(9)	1	0	0	0	0
(8)	1	1	0	0	0
(10)	1	1	1	1	0
(7)	0	0	1	1	0
(11)	0	0	1	0	1

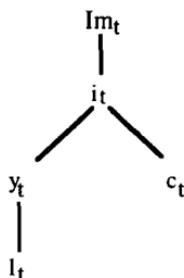
Connaissant la solution de (9) on en déduit la solution de (8).

On a alors un autre sous-système qui est d'ordre 1

(10) et (7) constituent alors un sous-système d'ordre 2.

(11) constitue un système d'ordre 3

L'ordre dans lequel vont être déterminées les 5 variables est donc :



Ici, on a en fait une structure de pré-ordre. (La matrice n'est pas triangulaire). La méthode s'apparente à une diagonalisation de la matrice des (0,1). Quand on a une structure de pré ordre c'est-à-dire que des variables sont déterminées ensemble on ne peut pas diagonaliser la matrice. Si on peut la diagonaliser on obtient un système récursif qui à comme propriété d'être entièrement hiérarchisé sur la totalité des variables.

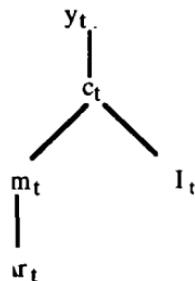
Un ensemble d'équation peut en fait contenir des logiques différentes. Conservons le même système mais considérons les variables différemment. Considérons que les cinq endogènes sont c_t , y_t , I_t , I_m_t , Δr_t (on prend I_t comme exogène : on impose un niveau d'emploi). Le modèle va donc incorporer une logique assez différente.

	c_t	y_t	I_t	I_m_t	Δr_t
(7)	1	1	0	0	0
(8)	0	0	1	1	0
(9)	0	0	0	1	1
(10)	1	1	1	1	0
(11)	0	1	0	0	0

Faisons apparaître un sous-système d'ordre 0 qui est (11), on a :

	y_t	c_t	I_{mt}	i_t	Δr_t	
(11)	1	0	0	0	0	sous-système d'ordre 0
(7)	1	1	0	0	0	sous-système d'ordre 1
(10)	1	1	1	1	0	sous-système d'ordre 2
(8)	0	0	1	1	0	
(9)	0	0	1	0	1	sous-système d'ordre 3

Nouvel arbre de causalité :



• Une même structure formelle peut donc incorporer des logiques différentes. On pourrait encore trouver d'autres logiques en considérant x_t comme endogène par exemple.

À partir de là on peut faire apparaître une typologie assez sommaire des systèmes d'équations :

- Il y a d'abord les systèmes décomposables qui sont des systèmes dans lesquels une partie des variables est déterminée dans une partie des équations et le reste des variables dans les autres équations. Un modèle décomposable se présente en fait comme la juxtaposition de deux sous-modèles indépendants.

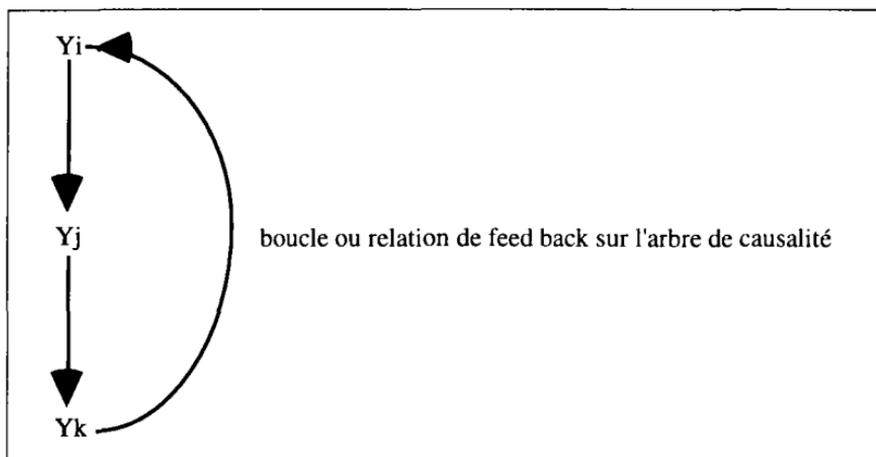
(1)	$y_1 \dots y_i$	$y_j \dots \dots n$
(i)	(0,1)	\emptyset
(j)		
(n)	\emptyset	(0,1)

Le modèle peut donc se décomposer en deux sous-parties indépendantes.

Historiquement cela correspond à un modèle intéressant, le modèle original de Harrod Domar.

Le dépassement de ce type de modèle se fait en établissant des relations entre les deux parties. Dans le modèle Harrod Domar en endogénéisant le taux d'épargne par exemple.

- Les modèles ou systèmes de causalité : ce sont des modèles dans lesquels on peut au moins partiellement faire apparaître l'enchaînement des causes et des effets. Notre exemple correspond à un système de causalité.
- Les systèmes d'interdépendance : ils constituent dans leur totalité le plus petit sous-système admettant une solution unique. Cela revient à dire qu'on ne peut pas hiérarchiser l'ordre de résolution des variables : l'ensemble des endogènes est déterminé simultanément au moment de la résolution du modèle. Sur le plan de la logique, l'interdépendance peut s'interpréter comme l'existence de relation de retour (feed back) dans le schéma de causalité du type :



Il faut malheureusement remarquer que cette technique d'analyse est peu performante car la plupart des modèles ou système de taille importante sont des systèmes d'interdépendance. Plus la taille du modèle augmente plus on a de chance d'avoir un système d'interdépendance.

Une autre faiblesse est que la méthode se bloque rapidement quand le nombre d'équations et de variables augmente. Ce qui est facile à faire apparaître avec un nombre restreint d'équations devient beaucoup plus complexe avec un grand nombre d'équations et de variables.

- Il peut alors être utile d'élaborer des maquettes du modèle de taille réduite pour analyser de façon plus commode la causalité ou les propriétés dynamiques.

C- Les propriétés dynamiques des systèmes d'équation

Par définition on peut dire qu'un modèle est dynamique si dans un environnement donné qui est caractérisé par la stabilité des variables exogènes, un état de l'économie en t entraîne un autre état en $t + 1$ (un état de l'économie étant caractérisé par un ensemble de valeurs des variables endogènes).

Si en $t + 1$ les valeurs sont les mêmes qu'en t on a un état stationnaire.

Soit Y = l'ensemble des variables endogènes.

$Y_t \rightarrow Y_{t+1}$ pour X constant (X = ensemble des exogènes).

Sur le plan de l'écriture (plan formel) un modèle est dynamique s'il contient des variables endogènes décalées.

Tout modèle qui inclut des variables endogènes décalées possède des propriétés dynamiques qui vont se traduire par des trajectoires suivies par les endogènes au cours du temps.

C'est pourquoi les propriétés dynamiques peuvent être analysées et constituent un critère d'évaluation du système d'équation.

Parce qu'on sait que dans la plupart des cas les phénomènes économiques évoluent assez lentement, on demandera le plus souvent au système des propriétés minimum de stabilité.

(Par contre si on veut modéliser un système très instable on privilégiera une structure qui a cette propriété).

On peut donner différentes définitions de la stabilité qui sont plus ou moins pertinentes selon le type de modèle utilisé.

Sachant qu'on a :

Y l'ensemble des variables endogènes

X l'ensemble des variables exogènes

On appelle Y^r et X^r les valeurs de références des variables endogènes et des variables exogènes (valeurs du compte central).

La stabilité statique

Si on applique un écart petit aux valeurs des exogènes $X-X^r < \alpha$ alors on modifie peu les valeurs des endogènes : $Y-Y^r < \beta$

Le système est amorti, il a une forte inertie.

La stabilité dynamique ou asymptotique

On dira qu'un modèle présente cette propriété si après un écart $X-X^r$ petit : $X-X^r < \alpha$, le vecteur des endogènes revient vers le compte central quand le temps passe :

$$Y \rightarrow Y^r \text{ quand } T \rightarrow +\infty$$

Le multiplicateur Keynesien peut servir à illustrer cette propriété : si on applique pendant un nombre limité de période une variation du niveau des dépenses autonomes, lesquelles sont par définition exogènes, alors le revenu, etc. s'éloignent du compte central puis y reviennent progressivement.

La stabilité en trajectoire

Un système est stable en trajectoire si après un écart $X - X^T < \alpha$ les valeurs des variables endogènes rejoignent une trajectoire parallèle à celle du compte central.

Le même multiplicateur Keynesien quand on applique un changement permanent de la dépense autonome : revenu et consommation vont suivre la même évolution mais en dessous ou au dessus du compte central.

Exemple d'analyse des propriétés dynamiques d'un modèle : l'oscillateur de Samuelson.

Comme on sait, il peut être représenté par :

$$c_t = a_1 + a_2 y_{t-1} \quad (12)$$

$$I_t = b_1 + b_2 (y_{t-1} - y_{t-2}) \quad (13)$$

$$y_t = c_t + I_t + G_t \quad (14)$$

En remplaçant (12) et (13) dans (14) on obtient :

$$\begin{aligned} y_t &= a_1 + a_2 y_{t-1} + b_1 + b_2 (y_{t-1} - y_{t-2}) + G_t \\ y_t &= (a_2 + b_2) y_{t-1} - b_2 y_{t-2} + a_1 + b_1 + G_t \\ y_t - (a_2 + b_2) y_{t-1} + b_2 y_{t-2} &= a_1 + b_1 + G_t \end{aligned} \quad (15)$$

C'est donc une équation de récurrence du second ordre avec second membre.

La solution est donnée par $y_t = \lambda^t$ et pour calculer l'équation caractéristique on remplace à gauche dans l'équation sans second membre et on divise par λ^{t-2} :

$$\lambda^t - (a_2 + b_2)\lambda^{t-1} + b_2\lambda^{t-2} = 0 \quad (16)$$

$$\lambda^2 - (a_2 + b_2)\lambda + b_2 = 0 \quad (17)$$

La solution générale est donnée par :

$$y_t = A_1 \lambda_1^t + A_2 \lambda_2^t \quad \text{où } \lambda_1 \text{ et } \lambda_2 \text{ sont solutions de l'équation caractéristique (17).}$$

A_1 et A_2 sont des valeurs calculées en utilisant la partie droite de l'équation (Second membre).

Les propriétés dynamiques de la solution générale dépendent des valeurs de λ .

1- Si λ_1 et λ_2 sont réels, le système engendre des évolutions dynamiques monotones

1a- $|\lambda_1|$ et $|\lambda_2| < 1 \rightarrow$ évolutions monotones amorties (\rightarrow stable)

1b- $|\lambda_1|$ ou $|\lambda_2| > 1 \rightarrow$ évolutions monotones divergentes ou explosives

2- Si λ_1 et λ_2 complexes, alors le système engendre des évolutions cycliques.

2a- modules de λ_1 et $\lambda_2 < 1 \rightarrow$ oscillations amorties

2b- modules de λ_1 et $\lambda_2 > 1 \rightarrow$ oscillations explosives

Soit sur l'exemple de l'oscillateur :

$$\Delta = \text{discriminant} = (a_2 + b_2)^2 - 4b_2$$

Étudions le signe de Δ :

$$\Delta > 0 \text{ si } : (a_2 + b_2)^2 > 4b_2$$

$$\text{Soit } a_2 + b_2 > 2\sqrt{b_2}$$

$$a_2 > 2\sqrt{b_2} - b_2 \quad (18)$$

Calculons quelques points de la fonction (18) : le maximum est atteint par :

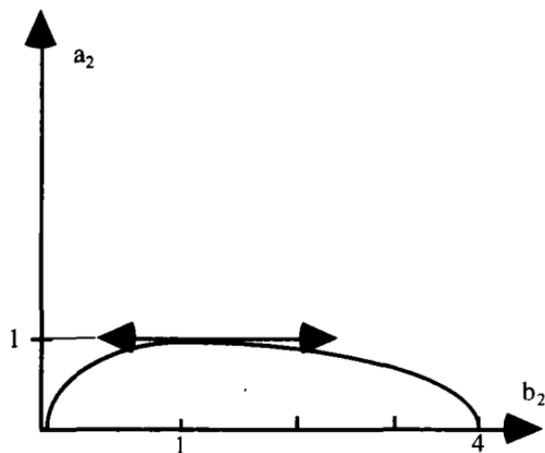
$$\frac{\partial a_2}{\partial b_2} = 2 * \frac{1}{2} b^{-1/2} - 1 = 0$$

$$\text{Soit } \frac{1}{\sqrt{b_2}} - 1 = 0 \text{ d'où } b_2 = 1$$

$$\text{si } \begin{cases} b_2 = 1 & a_2 = 1 \\ b_2 = 0 & a_2 = 0 \\ b_2 = 4 & a_2 = 0 \end{cases}$$

• Graphiquement

Au-dessus de la courbe les solutions sont réelles donc les évolutions sont monotones et en dessous elles sont complexes, donc les évolutions sont cycliques.



• Pour voir si les évolutions ont un caractère amorti ou explosif, on commencera par raisonner dans le cadre de racines complexes. On sait qu'un nombre complexe peut se mettre sous la forme :

$$\rho(\cos \theta + i \sin \theta)$$

ρ = module θ = argument

Quand on élève un nombre complexe à une puissance t on obtient :

$$\rho^t (\cos t\theta + i \sin t\theta)$$

Deux complexes conjugués ont la même partie réelle mais deux parties imaginaires opposées soit : $\rho(\cos \theta + i \sin \theta)$ et $\rho(\cos(-\theta) + i \sin(-\theta))$

Le produit de deux complexes conjugués donne :

$$\rho^2 (\cos^2 \theta - i^2 \sin^2 \theta) = \rho^2$$

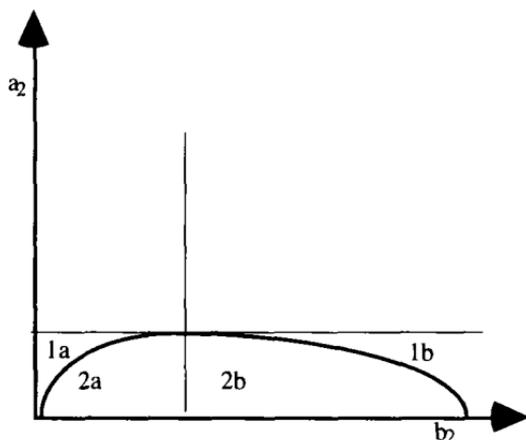
Dans le cas où les racines λ_1 et λ_2 sont complexes, ce sont des complexes conjugués et le produit de ces deux complexes vaut alors ρ^2 c'est-à-dire le carré de leur module commun.

Mais on sait aussi que $\lambda_1 \lambda_2 = b_2$ dans l'équation (17)

Donc si $b_2 > 1$ alors $\rho^2 > 1$ et dans ce cas les oscillations sont divergentes ou explosives.

Si $b_2 < 1$ alors $\rho < 1 \rightarrow$ cas n° 2a oscillations amorties.

On peut représenter graphiquement les différents cas possibles :



Si les solutions sont des nombres réels, on peut utiliser la propriété qu'un nombre réel est un nombre complexe dont la partie imaginaire est nulle, ce qui permet de retrouver la même propriété et la même conclusion.

L'analyse des propriétés dynamiques peut se faire par une méthode plus générale :

• La méthode générale

Elle utilise les propriétés des équations de récurrence matricielles.

Soit $Y_{(n,1)}$ un vecteur de variables endogènes et Y_{-1} ce même vecteur décalé d'une période.

L'équation $Y = AY_{-1}$ (19) constitue une équation de récurrence matricielle du premier ordre, A étant une matrice régulière de format $(n \times n)$.

La solution d'une telle équation est donnée par : $Y_{(n,1)} = r^t V_{(n,1)}$ (20), V étant un vecteur de même format que Y et r étant un scalaire.

L'équation caractéristique de cette équation de récurrence s'écrit :

$$r^t V = A r^{t-1} V \quad (21)$$

$$\text{soit } r V = A V \quad (22)$$

$$\text{ou } (A - rI) V = 0 \quad (22) \text{ bis}$$

Cette équation caractéristique exprime que r est une valeur propre de la matrice A et V est le vecteur propre qui lui est associé.

Si la matrice A est une matrice régulière alors il y aura autant de valeurs propres que de lignes ou de colonnes dans A soit n valeurs propres.

$$|A - rI| = 0 \Rightarrow n \text{ valeurs propres } r_1, \dots, r_n.$$

La solution générale est alors :

$$Y = A_1 r_1^t v_1 + \dots + A_n r_n^t v_n \quad (23)$$

- Les $r_1 \dots r_n$ étant les valeurs propres de A
- Les $A_1 \dots A_n$ étant des coefficients de pondération qui indiquent l'importance respective des différents modes dynamiques indiqués par r_1 à r_n .

Ces coefficients sont calculés par rapport aux conditions de départ : en faisant

$$t = 0 \text{ on obtient } Y_0 = A_1 v_1 + \dots + A_n v_n$$

Il y a ici n équations et n inconnues à calculer, ce qui permet de calculer les valeurs des coefficients $A_1 \dots A_n$

Chaque valeur propre caractérise un mode d'évolution dynamique.

Or, il y a autant de valeurs propres qu'il y a d'équations c'est-à-dire de lignes, et donc autant de valeurs propres qu'il y a de colonnes c'est-à-dire de variables.

La dernière étape consiste alors à associer des propriétés dynamiques c'est-à-dire des valeurs propres à des équations c'est-à-dire des propriétés structurelles du modèle.

- Cela permettra de savoir quel mécanisme est à l'origine de telle ou telle propriété dynamique et ça permet aussi de modifier les propriétés dynamiques du modèle qui ne conviennent pas en agissant sur la cause.

Appliquons cette méthode à l'exemple précédent :

$$(12) \quad c_t = a_1 + a_2 y_{t-1}$$

$$(13) \quad I_t = b_1 + b_2 (y_{t-1} - y_{t-2})$$

$$(14) \quad y_t = c_t + I_t + G_t$$

$$\text{Le vecteur } Y = \begin{bmatrix} c_t \\ I_t \\ y_t \end{bmatrix}$$

La difficulté est la présence de y_{t-2} qui impose de faire un changement de variable.

On considère que (14) est satisfaite en toute période d'où on peut écrire :

$$y_{t-1} = c_{t-1} + I_{t-1} + G_{t-1} \quad (24)$$

On peut définir une nouvelle variable z_t telle que :

$$z_t = c_{t-1} + I_{t-1} + g_{t-1} \quad (25)$$

$$\text{et } z_{t-1} = c_{t-2} + I_{t-2} + G_{t-2} = y_{t-2} \quad (26)$$

$$\text{Posons } Y = \begin{bmatrix} c_t \\ I_t \\ z_t \end{bmatrix}$$

Ce vecteur constitue le nouveau vecteur des variables endogènes.

En remplaçant y_{t-2} par sa valeur tirée de (26), l'équation (13) devient :

$$I_t = b_1 + b_2 c_{t-1} + b_2 I_{t-1} + b_2 G_{t-1} - b_2 z_{t-1} \quad (27)$$

Donc, en remplaçant y_{t-1} et y_{t-2} le système devient :

$$\begin{cases} c_t = a_1 + a_2 c_{t-1} + a_2 I_{t-1} + a_2 G_{t-1} & (28) \\ I_t = b_1 + b_2 c_{t-1} + b_2 I_{t-1} + b_2 G_{t-1} - b_2 z_{t-1} & (27) \\ z_t = c_{t-1} + I_{t-1} + G_{t-1} & (25) \end{cases}$$

que l'on peut mettre sous la forme $Y = AY_{t-1}$

$$Y = \begin{bmatrix} c_t \\ I_t \\ z_t \end{bmatrix} \quad Y-1 = \begin{bmatrix} c_{t-1} \\ I_{t-1} \\ z_{t-1} \end{bmatrix}$$

$$A \begin{bmatrix} a_2 & a_2 & 0 \\ b_2 & b_2 & -b_2 \\ 1 & 1 & 0 \end{bmatrix}$$

(3,3)

Donc :

$$\begin{bmatrix} c_t \\ I_t \\ z_t \end{bmatrix} = \begin{bmatrix} a_2 & a_2 & 0 \\ b_2 & b_2 & -b_2 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} c_{t-1} \\ I_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} a_2 & a_1 \\ b_2 & b_1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} G_{t-1} \\ 1 \end{bmatrix}$$

$Y \quad \quad A \quad \quad Y_{-1} \quad \quad \begin{matrix} (3,2) & (2,1) \\ \text{équivalent de} \\ \text{notre 2}^{\text{e}} \text{ membre} \end{matrix}$

• Calcul des valeurs propres

$$\det(A - rI) = 0$$

$$|A - rI| = \begin{vmatrix} a_2 - r & a_2 & 0 \\ b_2 & b_2 - r & -b_2 \\ 1 & 1 & -r \end{vmatrix} \quad (29)$$

$$\det(A - rI) = -r(a_2 - r)(b_2 - r) - a_2 b_2 + b_2(a_2 - r) + a_2 b_2 r \quad (35)$$

$$= -r(a_2 - r)(b_2 - r) - b_2 r + a_2 b_2 r$$

$$= -r[r^2 - (a_2 + b_2)r + b_2] \quad (30)$$

$$\det(A - rI) = 0 \Leftrightarrow \begin{cases} r = 0 \\ r^2 - (a_2 + b_2)r + b_2 = 0 \end{cases} \quad (31)$$

L'équation (31) est identique à l'équation (17) étudiée précédemment.

Comme le montre cet exemple, dans la plupart des cas la mise sous forme dynamique $Y = AY_{-1}$ ne se fait pas directement. Cela impose des transformations sur le modèle de sorte que l'analyse des propriétés dynamiques se fait surtout sur des versions dérivées c'est-à-dire retravaillées.

Deux exemples permettront d'illustrer l'apport de l'analyse des propriétés dynamiques d'un modèle.

Exemple 1 : modèle de Klein pour l'économie US : (1)

$$c = 16,78 + 0,02\Pi + 0,23\Pi_{-1} + 0,80(w_1 + w_2) \quad (32)$$

$$I = 17,79 + 0,23\Pi + 0,55\Pi_{-1} - 0,15K_{-1} \quad (33)$$

$$w_1 = 1,60 + 0,42(Y + T - w_2) + 0,16(Y + T - w_2)_{-1} + 0,13(t - 1931) \quad (34)$$

$$Y + T = C + I + G \text{ quasi condition d'équilibre.} \quad (35)$$

$$Y = \Pi + w_1 + w_2 \quad (37)$$

$$\Delta K = I \quad (38)$$

Π = profit des entreprises

W_i = salaires versés

K = stock de capital

T = impôts

t = indice du temps

I = investissement

c = consommation

Après transformation ce modèle peut être mis sous la forme :

$$\begin{bmatrix} \Pi \\ Y \\ K \\ C \\ w_1 \\ I \end{bmatrix} = \begin{bmatrix} 0,863 & -0,063 & 0,164 & 0 & 0 & 0 \\ 1,489 & 0,174 & -0,283 & 0 & 0 & 0 \\ 0,746 & -0,015 & 0,816 & 0 & 0 & 0 \\ 0,743 & 0,189 & -0,098 & 0 & 0 & 0 \\ 0,626 & 0,237 & 0,119 & 0 & 0 & 0 \\ 0,746 & -0,015 & -0,184 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Pi_{-1} \\ Y_{-1} \\ K_{-1} \\ C_{-1} \\ w_{1-1} \\ I_{-1} \end{bmatrix} + Z \quad (39)$$

On constate que sous cette forme, le modèle présente la particularité d'avoir trois colonnes nulles, ce qui montre que les trois dernières variables n'interviennent dans aucun processus dynamique.

(1) Cet exemple est tiré de Bridge "Applied Econometrics"

$$\begin{cases} r_1 = -0,26 \\ r_2 = 0,31 \\ r_3 = 1,22 \end{cases}$$

On a donc

- . un mode alternatif (+ - + -);
- . un mode monotone amorti;
- . un mode divergent.

On peut se poser la question de savoir quelles sont les équations qui engendrent ces modes et comment on peut les modifier.

Si on modifie l'équation (46). Cette équation contient des variables décalées qui lui donnent des propriétés dynamiques. On obtient alors :

$$\begin{cases} r_1 = -0,26 \\ r_2 = 0,36 \\ r_3 = 0 \end{cases}$$

On peut aussi modifier l'équation (43). On obtient :

$$\begin{cases} r_1 = -0,27 \\ r_2 = -0,02 \\ r_3 = 1,23 \end{cases}$$

On peut aussi modifier (45). On obtient :

$$\begin{cases} r_1 = 0,06 \\ r_2 = 0,35 \\ r_3 = 1,21 \end{cases}$$

Enfin on peut modifier (42) et on obtient :

$$\begin{cases} r_1 = -0,53 \\ r_2 = 0,29 \\ r_3 = 1,23 \end{cases}$$

À chaque transformation on peut associer une valeur propre qui a été fortement affectée. Le mode d'évolution monotone divergent caractérisé par la valeur propre r_3 est associé à l'équation (46).

Le mode d'évolution monotone convergent caractérisé par r_2 est associé à (43) c'est-à-dire à la fonction de consommation.

Le mode alternatif caractérisé par r_1 est associé à l'équation (45) c'est-à-dire au mécanisme de formation des stocks.

L'équation (42) ne change pas la nature des modes d'évolution et donc cette équation n'est associée à aucune dynamique en particulier. Ce n'est pas surprenant car on avait quatre équations pour 3 racines. Il y avait donc une équation de trop qui semble être ici (42).

Qu'il s'agisse de l'analyse des relations de causalité ou de l'analyse des propriétés dynamiques on se heurte vite à des difficultés pratiques importantes dès lors que la taille des modèles augmente ce qui fait qu'on ne peut pas analyser les modèles en grandeur réelle mais seulement des versions réduites et surtout transformées pour les besoins de la mise en forme (type $Y = A Y_{-1}$).

Les deux exemples qui ont été présentés sont des modèles anciens, mais les versions transformées sont très mauvaises.

Des études ont été faites également sur le modèle STAR pour la France.

Quand il y a des non-linéarités (rapport de variable par exemple...) les méthodes qu'on a vues ne sont plus directement applicables. Au total, l'analyse des propriétés dynamiques, comme du reste celle de la causalité, reste un domaine difficile et finalement assez peu satisfaisant, en dehors de la typologie un peu simpliste des cas de figure possibles.

SECTION 2 : L'ESTIMATION DES MODÈLES À ÉQUATIONS SIMULTANÉES

Comme on l'a vu un modèle se présente comme un ensemble de variables :

. n variables endogènes : $y_1 \dots y_n$;

. m variables exogènes : $x_1 \dots x_m$;

organisées dans un ensemble de $h + k = n$ équations.

Le modèle peut s'écrire en isolant les variables endogènes à gauche :

$$y_{1t} = a_{12}y_{2t} + \dots + a_{1n}y_{nt} + b_{11}x_{1t} + \dots + b_{1m}x_{mt} + u_{1t}$$

.

.

.

$$y_{nt} = a_{n1}y_{1t} + \dots + a_{nn-1}y_{n-1t} + b_{n1}x_{1t} + \dots + b_{nm}x_{mt} + u_{nt}$$

C'est la forme qui vient généralement en premier lors de l'écriture du modèle et qui est censée traduire comment les variables réagissent les unes par rapport aux autres. On l'appelle forme structurelle. On peut aussi isoler à droite les termes d'erreur en respectant le même ordre pour les variables endogènes dans chaque équation.

La forme structurelle s'écrit alors :

$$y_{1t} - a_{12}y_{2t} - \dots - a_{1n}y_{nt} - b_{11}x_{1t} - \dots - b_{1m}x_{mt} = u_{1t}$$

$$-a_{21}y_{1t} + y_{2t} - \dots - a_{2n}y_{nt} - b_{21}x_{1t} - \dots - b_{2m}x_{mt} = u_{2t}$$

.

.

.

$$-a_{n1}y_{1t} \quad \quad \quad + y_{nt} - b_{n1}x_{1t} - \dots - b_{nm}x_{mt} = u_{nt}$$

On peut réécrire le modèle en notations matricielles.

Soit Y = l'ensemble des variables endogènes :

$$Y_{(n,1)} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

Soit X = l'ensemble des variables exogènes :

$$X_{(m,1)} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$$

$$\text{Soit } U_{(n,1)} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$$

Soit A : matrice des coefficients des variables endogènes dans le système d'équations :

$$A_{(n,n)} = \begin{bmatrix} 1 & -a_{12} & \dots & -a_{1n} \\ -a_{21} & 1 & & -a_{2n} \\ \vdots & & \ddots & \vdots \\ -a_{n1} & & & 1 \end{bmatrix}$$

De même

$$B_{(n,m)} = \begin{bmatrix} -b_{11} & \dots & -b_{1m} \\ \vdots & & \vdots \\ -b_{n1} & \dots & -b_{nm} \end{bmatrix}$$

La forme structurelle peut alors se réécrire

$$\boxed{AY + BX = U}$$

Remarque : Dans le chapitre I, Y et X désignaient respectivement les valeurs prises par la variable expliquée et les variables explicatives au cours des T périodes utilisées pour l'estimation.

Ici Y et X représentent l'ensemble des variables endogènes et l'ensemble des variables exogènes du modèle.

A- Le biais des moindres carrés ordinaires

Un premier résultat important est que l'estimation d'un modèle à équations simultanées par les moindres carrés ordinaires conduit à des estimateurs biaisés et non convergents.

On peut le montrer à l'aide d'un exemple très simple en s'inspirant d'une démonstration proposée par Haavelmo.

Soit le modèle suivant :

$$\begin{cases} c_t = ay_t + b + u_t & (1) \\ y_t = c_t + z_t & (2) \end{cases}$$

En remplaçant (2) dans (1) on obtient :

$$(3) \quad c_t = \frac{a}{1-a} z_t + \frac{b}{1-a} + \frac{u_t}{1-a}$$

Si on reporte dans (2) il vient :

$$(4) \quad y_t = \frac{1}{1-a} z_t + \frac{b}{1-a} + \frac{u_t}{1-a}$$

D'après (4), la variable explicative y_t n'est pas indépendante du terme u_t dans (1) et donc l'hypothèse H6 n'est plus vérifiée.

$$E(y_t u_t) = E\left[\left(\frac{1}{1-a} z_t + \frac{b}{1-a} + \frac{u_t}{1-a}\right) u_t\right] = \frac{1}{1-a} \sigma_u^2 \quad (5)$$

L'estimateur des MCO dans (1)

$$\hat{a} = \frac{\sum (c_t - \bar{c})(y_t - \bar{y})}{\sum (y_t - \bar{y})^2} \quad (6) \text{ est biaisé et non convergent.}$$

On peut pour cela remplacer c et y dans :

$$c_t - \bar{c} = \frac{a}{1-a}(z_t - \bar{z}) + \frac{u_t}{1-a} \quad (7)$$

$$y_t - \bar{y} = \frac{1}{1-a}(z_t - \bar{z}) + \frac{u_t}{1-a} \quad (8)$$

Le numérateur de (6) vaut

$$\begin{aligned} & \sum \left[\frac{a}{1-a}(z_t - \bar{z}) + \frac{u_t}{1-a} \right] \left[\frac{1}{1-a}(z_t - \bar{z}) + \frac{u_t}{1-a} \right] \\ &= \frac{a}{(1-a)^2} \sum (z_t - \bar{z})^2 + \frac{1}{(1-a)^2} \sum u_t^2 + \frac{a+1}{(1-a)^2} \sum (z_t - \bar{z})u_t \end{aligned}$$

Le dénominateur vaut

$$\begin{aligned} & \sum \left[\frac{1}{1-a}(z_t - \bar{z}) + \frac{u_t}{1-a} \right]^2 \\ &= \frac{1}{(1-a)^2} \sum (z_t - \bar{z})^2 + \frac{1}{(1-a)^2} \sum u_t^2 + \frac{2}{(1-a)^2} \sum (z_t - \bar{z})u_t \end{aligned}$$

$$\text{Soit } \hat{a} = \frac{a\sigma_z^2 + \sigma_u^2 + (a+1)\text{cov } z_t u_t}{\sigma_z^2 + \sigma_u^2 + 2\text{cov } z_t u_t} \neq a$$

L'estimateur \hat{a} est donc biaisé. De plus il n'est pas convergent c'est-à-dire que le biais demeure même si le nombre d'observation augmente indéfiniment.

La limite en probabilité de \hat{a} est :

$$P\text{lim } \hat{a} = \frac{a\sigma_z^2 + \sigma_u^2}{\sigma_z^2 + \sigma_u^2} \text{ donc ne converge pas vers } a \text{ même pour } T \text{ grand.}$$

La cause du problème est bien la non-vérification de H6.

En effet, si y_t ne contenait pas u_t il n'y aurait pas au numérateur et au dénominateur de termes en σ_u^2 .

• Quel est le sens du biais ?

$$P\text{lim } \hat{a} - a = \frac{a\sigma_z^2 + \sigma_u^2 - a\sigma_z^2 - a\sigma_u^2}{\sigma_u^2 + \sigma_z^2} = \frac{(1-a)\sigma_u^2}{\sigma_z^2 + \sigma_u^2}$$

Si donc : $a > 1$ $1 - a < 0$ $\text{Plim } \hat{a} < a \rightarrow$ on sous-estime
 $a < 1$ $1 - a > 0$ $\text{Plim } \hat{a} > a \rightarrow$ on surestime

Dans l'exemple de Haavelmo :

On surestime la propension à consommer.

Il n'est donc pas correct de traiter un système à équations simultanées par les MCO directs équation par équation.

Il faut donc recourir à des méthodes d'estimation spécifiques. Il en existe plusieurs dont 2 présentent un intérêt particulier.

B- Les moindres carrés indirects

$$\text{Soit : } AY + BX = U$$

Si il existe A^{-1} alors il est possible d'isoler Y :

$$Y = -A^{-1}BX + A^{-1}U$$

On pose

$$C = -A^{-1}B$$

et

$$V = A^{-1}U$$

Alors il vient

$$Y = CX + V$$

Dans cette nouvelle forme les variables endogènes c'est-à-dire les y ne dépendent plus que des variables exogènes, les x et donc il n'y a plus de biais de simultanéité.

La matrice C est de format (n x m) et V est un nouveau vecteur de variables aléatoires de format (n x 1).

Si on développe cette forme réduite, elle s'écrit :

$$\begin{cases} y_{1t} = c_{11}x_{1t} \dots c_{1m}x_{mt} + v_{1t} \\ \vdots \\ y_{nt} = c_{n1}x_{1t} \dots c_{nm}x_{mt} + v_{nt} \end{cases}$$

dans le système : $c_{ij} = \frac{\partial y_i}{\partial x_j}$

Les c_{ij} sont alors les multiplicateurs d'impact des variables exogènes sur les variables endogènes.

La méthode des moindres carrés indirects (MCI) consiste à :

- Dans une 1^{re} étape : estimer les coefficients de la matrice C par les MCO
- Dans une 2^e étape : remonter de C vers la forme structurelle c'est-à-dire vers les coefficients de A et de B.

C'est l'opération d'identification du modèle sous sa forme structurelle.

Comme le montre la relation $C = -A^{-1}B$, les coefficients de C sont des combinaisons des coefficients de A et de B.

$$\begin{bmatrix} 1 & & -a_{1n} \\ -a_{21} & 1 & -a_{2n} \\ & \ddots & \\ -a_{n1} & & 1 \end{bmatrix} = A_{(n,n)} \quad B_{(n,m)} = \begin{bmatrix} b_{11} & & b_{1m} \\ & \ddots & \\ b_{n1} & & b_{nm} \end{bmatrix}$$

Il faut donc dans A calculer $n(n-1)$ coefficients inconnus.

Dans B il y a $(n \times m)$ coefficients inconnus.

Cela fait donc $(n-1)n + nm$ coefficients à calculer à partir des éléments de C qui sont au nombre de $(n \times m)$.

Dans le cas général ce n'est pas possible car il y a trop d'inconnues par rapport au nombre de relations. Le modèle ne peut pas être identifié.

L'identification du modèle n'est possible que si il existe un certain nombre de restrictions à priori : $n(n - 1)$ restrictions à priori qui permettent de réduire le nombre de coefficients à estimer au nombre de relations c'est-à-dire à nm .

Les restrictions à priori sont de nature différentes de celles qui ont été présentées au chapitre I. Ici ce sont toutes les informations qui permettent de réduire le nombre de coefficients à calculer. Ces restrictions à priori peuvent être de plusieurs types.

- 1- Le coefficient est connu d'avance (cas des conditions d'équilibre et des identités) donc il n'y a pas besoin de le calculer.
- 2- Une variable ne figure pas dans l'équation. Le coefficient correspondant est donc nul.
- 3- Les relations possibles entre coefficients : exemple : un coefficient est la somme de deux autres.

De plus, l'identification ne se fait pas sur l'ensemble du modèle mais équation par équation. Le nombre de restrictions doit donc être équitablement réparti sur les équations soit $(n - 1)$ restrictions par équation.

Dans le cas où les restrictions à priori prennent la forme de variables absentes, l'identification n'est possible qu'à condition d'avoir $(m+1)$ variables dans chaque équation. En effet, soit l'équation k . Cette équation a au maximum $n + m$ variables. On lui impose $n - 1$ restriction à priori sous la forme de variables absentes.

Il y a donc $(n + m) - (n - 1) = m + 1$ variables présentes dans l'équation.

On comprend aisément que cette contrainte est inacceptable car beaucoup trop stricte, en macro-économie par exemple.

La plupart des modèles seront donc soit sous-identifiés, soit sur-identifiés.

- Sur-identifiés quand le nombre de restrictions est $> (n - 1)$ par équation.
- Sous-identifiés quand le nombre de restrictions est $< (n - 1)$ par équation.

L'exemple très simple des équations (1) et (2) permet d'illustrer ce problème.

$$\begin{cases} c_t = ay_t + b + u_t & (1) \\ y_t = c_t + z_t & (2) \end{cases}$$

Ce modèle comprend :

- deux endogènes donc $Y = \begin{bmatrix} c_t \\ y_t \end{bmatrix}$;

- deux exogènes donc $X = \begin{bmatrix} z_t \\ 1 \end{bmatrix}$;

- et $U = \begin{bmatrix} u_t \\ 0 \end{bmatrix}$.

En isolant les termes d'erreur à droite, il vient :

$$c_t - ay_t - b = u_t$$

$$-c_t + y_t - z_t = 0$$

$$\text{D'où } A = \begin{bmatrix} 1 & -a \\ -1 & 1 \end{bmatrix} \text{ et } B = \begin{bmatrix} 0 & -b \\ -1 & 0 \end{bmatrix}$$

Le modèle peut donc se réécrire :

$$\begin{bmatrix} 1 & -a \\ -1 & 1 \end{bmatrix} \begin{bmatrix} C_t \\ Y_t \end{bmatrix} + \begin{bmatrix} 0 & -b \\ -1 & 0 \end{bmatrix} \begin{bmatrix} Z_t \\ 1 \end{bmatrix} = \begin{bmatrix} u_t \\ 0 \end{bmatrix} \quad \text{Soit } AY + BX = U$$

Ceci est une autre façon d'écrire la forme structurelle.

Pour la première étape des MCI il faut calculer la forme réduite :

$$A^{-1} = \begin{bmatrix} 1/1-a & a/1-a \\ 1/1-a & 1/1-a \end{bmatrix}$$

$$C = - \begin{bmatrix} 1/1-a & a/1-a \\ 1/1-a & 1/1-a \end{bmatrix} \begin{bmatrix} 0 & -b \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} a/1-a & b/1-a \\ 1/1-a & b/1-a \end{bmatrix}$$

C'est une matrice dont les éléments sont des combinaisons des coefficients de A et B comme indiqué précédemment.

$$V = \begin{bmatrix} 1/1-a & a/1-a \\ 1/1-a & 1/1-a \end{bmatrix} \begin{bmatrix} u_t \\ 0 \end{bmatrix} = \begin{bmatrix} u_t/1-a \\ u_t/1-a \end{bmatrix}$$

La forme réduite est :

$$\begin{bmatrix} c_t \\ y_t \end{bmatrix} = \begin{bmatrix} a/1-a & b/1-a \\ 1/1-a & b/1-a \end{bmatrix} \begin{bmatrix} z_t \\ 1 \end{bmatrix} + \begin{bmatrix} u_t/1-a \\ u_t/1-a \end{bmatrix}$$

On vérifie bien que la partie droite ne contient que des variables exogènes.

Peut-on identifier la forme structurelle, c'est-à-dire connaissant la valeur des coefficients de C, peut-on remonter vers les coefficients de A et de B ?

Il y a quatre relations constituées par les 4 coefficients de C, pour calculer :

2 coefficients dans A

4 coefficients dans B

Le nombre de restrictions à priori sur A et B :

dans A : 1 restriction (le coefficient de c_t est égal à +1 dans l'équation (2)) ;

dans B : 3 restrictions (1 variable absente dans chaque équation et le coefficient de z_t dans (2)).

Nombre de coefficients restant à calculer : $6 - 4 = 2$

Nombre de relations = 4

Le modèle est donc sur-identifié globalement.

• Cela signifie que plusieurs coefficients peuvent être calculés par plusieurs mécanismes indépendants

Appellons c_{ij} les éléments de C :
$$\begin{bmatrix} \hat{c}_{11} & \hat{c}_{12} \\ \hat{c}_{21} & \hat{c}_{22} \end{bmatrix}$$

Chacun des coefficients a et b peut être calculé de plusieurs façons différentes.

Exemple : $a = \frac{\hat{c}_{11}}{\hat{c}_{21}}$

mais $\frac{1}{1-a} = \hat{c}_{21} \quad \frac{1}{\hat{c}_{21}} = 1-a \quad d'où \quad a = 1 - \frac{1}{\hat{c}_{21}}$

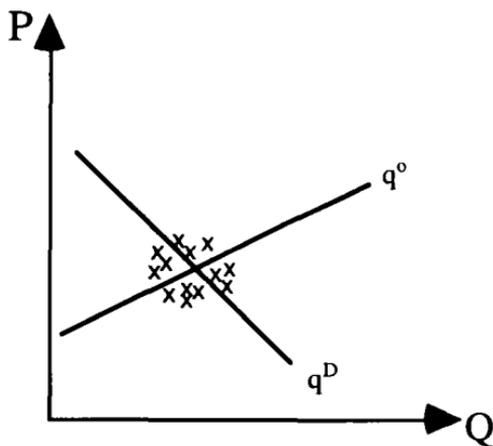
de même $b = \frac{\hat{c}_{12}}{\hat{c}_{21}}$

et $b = \frac{\hat{c}_{22}}{\hat{c}_{21}}$

Il y a donc différentes façons indépendantes de calculer a et b. C'est ce que veut dire la sur-identification : trop d'équations par rapport au nombre d'inconnues.

Cette question de l'identification permet d'expliquer un problème important qui concerne l'économétrie des marchés ou économétrie de l'offre et de la demande.

Supposons qu'on cherche à estimer une fonction de demande et que pour cela on dispose d'un certain nombre d'observations sur le prix et la quantité échangée au cours de plusieurs périodes.

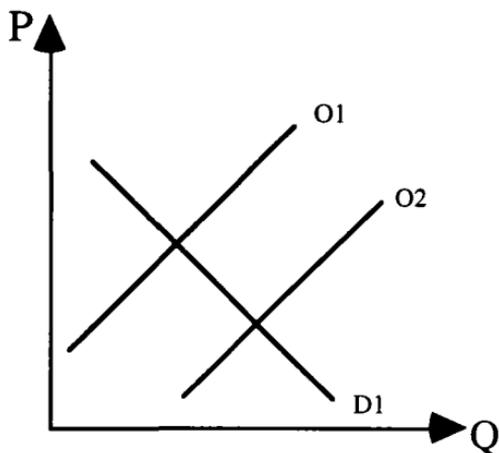


Un premier cas de figure serait celui dans lequel les fonctions d'offre et de demande sont stables. À chaque période la quantité échangée sera proche de l'équilibre.

$$\left\{ \begin{array}{l} q_o = \text{quantité offerte} = a + bP \quad (9) \\ q_D = c + dP \quad (10) \end{array} \right.$$

On obtient un ensemble de points groupés autour de l'intersection des fonctions d'offre et de demande et il est impossible d'estimer ces fonctions.

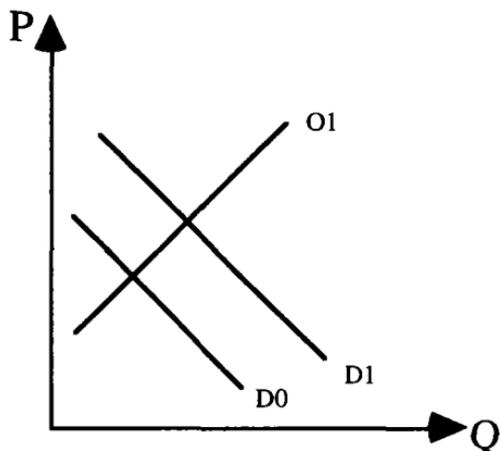
Un deuxième cas est celui où l'une des fonctions se déplace :



La fonction d'offre se déplace et l'intersection décrit la fonction fixe, c'est-à-dire la fonction de demande.

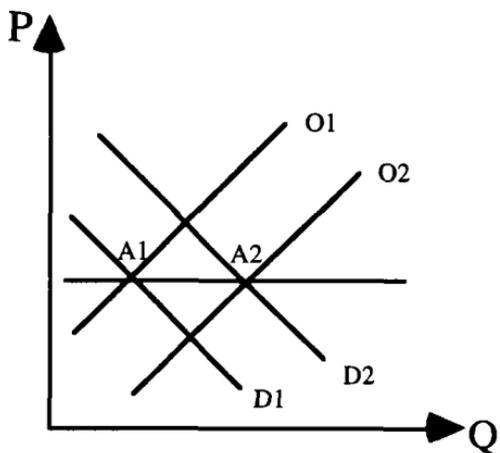
Dans ce cas la régression de la quantité échangée sur le prix d'équilibre donne bien une fonction de demande.

Le cas symétrique est celui où l'offre est fixe et où la demande se déplace :



Le point d'intersection décrit la courbe d'offre et donc la régression de Q sur P donnera la fonction d'offre.

Le cas général est celui dans lequel les 2 fonctions se déplacent.



Le prix d'équilibre passe de $O_1 \cap D_1$ à $O_2 \cap D_2$. Si on régresse Q/P on obtient $A_1 A_2$ qui n'est ni une fonction d'offre, ni une fonction de demande.

Dans le cas général, quand les fonctions d'offre et de demande se déplacent, une régression de Q sur P donne une combinaison ou un Mix de Q_D et Q_0 qu'on peut interpréter comme une forme réduite.

Ceci amène à poser le problème du retour au modèle structurel à partir du modèle réduit.

Le retour possible sur le modèle structurel offre-demande est donc dans une certaine mesure analogue au problème de l'identification.

Cas simple

$Q_0 = a + bp + eR$ pour autoriser les déplacements de la fonction d'offre (11)

$Q_D = c + dp + fR$ (12) pour la fonction de demande.

Le prix observé est le prix d'équilibre et à l'équilibre

$$Q_0 = Q_D = Q_e = \text{quantité d'équilibre}$$

Dans (11) on calcule $R = \frac{[Q_e - a - bP]}{e}$ (13)

Et, en remplaçant R par sa valeur dans (12) on retrouve une relation entre la quantité et le prix :

$$Q_e = c + dP + \frac{f}{e}[Q_e - a - bP]$$

$$Q_e \left[1 - \frac{f}{e} \right] = c - \frac{af}{e} + P \left[d - \frac{bf}{e} \right] \quad (14)$$

$$Q_e = \frac{ec - af}{e - f} + P \frac{[ed - bf]}{e - f} \quad (15)$$

Soit

$$Q_e = \alpha + \beta P \quad (16)$$

On voit bien que (15) apparaît comme une combinaison des coefficients de (11) et (12) qui peut être assimilée à une forme réduite.

Connaissant α et β peut-on retrouver a, b, c, d, e, f ?

En général, il n'est pas possible de remonter à la forme structurelle, comme on peut le voir en comparant (15) et (16).

Par contre, si la fonction de demande est stable alors $f = 0$ et dans ce cas (15) se réduit à $Q_e = c + dp$ ce qui correspond bien à la fonction de demande (10).

Symétriquement si l'offre est stable on a $e = 0$ et on retrouve bien la fonction d'offre $Q_e = a + bp$. (9)

Le problème de l'identification permet de mettre en évidence un certain nombre de difficultés concernant l'estimation des fonctions d'offre et de demande :

- On ne peut pas estimer directement des fonctions d'offre ou des fonctions de demande à partir d'observations sur les prix et sur les quantités échangées.
- Pour que l'on puisse remonter vers des fonctions d'offre ou de demande il faut que ces fonctions soient identifiables, ce qui n'est pas le cas du modèle (11), (12) ici.
- Dans certains cas très particuliers il n'existe pas de comportement d'offre ou de comportement de demande. On peut alors légitimement estimer celle des fonctions qui reste.

Par exemple la demande de dépôts à vue ou réglementés (caisse d'épargne). Il n'y a pas de comportements d'offre c'est-à-dire que les conditions faites par l'institution ne changent pas quelle que soit la quantité déposée. Dans ce cas on peut estimer des fonctions de demande. Mais on n'aurait pas le droit d'estimer dans les mêmes conditions une fonction de demande de dépôts à terme non réglementées puisque là il y a un comportement d'offre de la banque qui fixera les rémunérations en fonction des dépôts qui lui sont proposés.

Autre exemple, lorsqu'on a une offre rationnée, cette fonction pourra être estimée directement.

C- Autre méthode d'estimation : les doubles moindres carrés (DMC) ou TSLS (Two Stages Least Squares)

C'est une méthode simple et performante.

Cette méthode procède en deux étapes :

- Dans un premier temps elle consiste à régresser les variables endogènes figurant en variables explicatives dans la forme structurelle sur toutes les variables exogènes. Ceci permet d'obtenir des estimateurs de ces mêmes variables endogènes $\hat{y}_1 \dots \hat{y}_n$.
- Deuxième étape : on remplace les variables endogènes qui figurent en variables explicatives dans la forme structurelle par leurs valeurs calculées dans l'estimation du modèle équation par équation.

Cette procédure fournit des estimateurs convergents.

Exemple

Soit le modèle suivant :

$$c_t = b_0 + b_1 y_t + b_2 c_{t-1} + u_{1t} \quad (17)$$

$$I_t = a_0 + a_1 y_{t-1} + a_2 y_t + u_{2t} \quad (18)$$

$$y_t = c_t + I_t + g_t \quad (19)$$

Ce modèle contient $\begin{cases} 3 \text{ variables endogènes } c_t; I_t; y_t \\ 4 \text{ variables exogènes } 1; c_{t-1}; y_{t-1}; g_t \end{cases}$

Il y a bien un biais de simultanéité car dans (17) et (18) y_t figure en variable explicative.

L'application des doubles moindres carrés consiste, dans la première étape, à estimer :

$$y_t = \gamma_0 + \gamma_1 c_{t-1} + \gamma_2 y_{t-1} + \gamma_3 g_t \quad (20)$$

ce qui fournit des estimateurs $\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2$ et $\hat{\gamma}_3$ permettant de calculer à chaque période \hat{y}_t d'après (20).

Puis dans la seconde étape, cette valeur calculée \hat{y}_t remplace y_t , ce qui conduit à estimer :

$$c_t = b_0 + b_1 \hat{y}_t + b_2 c_{t-1} + u_{1t} \quad (21)$$

$$I_t = a_0 + a_1 y_{t-1} + a_2 \hat{y}_t + u_{2t} \quad (22)$$

Cette dernière estimation se fait par les moindres carrés.

En remplaçant y_t par \hat{y}_t , on introduit une certaine dose d'imprécision sur la variable explicative mais l'effet qui en résulte sur les estimateurs n'a pas un caractère systématique.

Remarque : La méthode des DMC est une application particulière d'une méthode plus générale qui est la méthode des variables instrumentales, méthode dont le domaine d'application déborde largement le problème de simultanéité.

D- La Méthode des variables instrumentales (M.V.I)

Quand dans une équation l'une au moins des variables explicatives n'est pas indépendante du terme d'erreur, l'hypothèse 6 n'est pas remplie et donc les Moindres Carrés ordinaires ne doivent plus être utilisés.

$$\text{Soit } Y = XA + U \quad (23)$$

$X =$ matrice $[X_1 \dots X_l]$. La matrice des variables est constituée de l colonnes correspondant aux différentes variables explicatives.

Et, il existe au moins X_j tel que $E(X'_j U) \neq 0$

Soit W de format (T, k) un ensemble de K variables observées sur T périodes telles que :

- les variables W soient indépendantes du terme d'erreur U ;
- les variables W sont corrélées avec les variables X ;
- les variables W sont en nombre au moins égal à celui des variables X .

$$E(W'U) = 0$$

$$E(W'X) = Z \quad K > l$$

Cette méthode procède également en deux temps :

• **Première étape**

Elle consiste à régresser les variables X sur W . Ceci permet d'obtenir des estimateurs par les moindres carrés :

$$X_{(i,1)} = W_{(T,k)(i,1)} B + v_i$$

$$\text{et donc } \hat{B}_{(k,1)} = (W' W)^{-1} W' X_i$$

Pour chacune de ces variables X_i on obtient l'estimateur \hat{B}_i correspondant.

Soit au total :

$$\hat{B}_{(k,l)} = \left[\hat{B}_1 \dots \hat{B}_l \right] \quad (24)$$

À l'aide de cette matrice on peut calculer les valeurs de \hat{X} .

$$\hat{X} = W_{(T,k)} \hat{B}_{(k,l)} \quad (25)$$

Cela permet d'obtenir une nouvelle matrice de variables explicatives non plus observées mais calculées.

D'après (24) et (25) on a :

$$\hat{X} = W \hat{B} = W(W' W)^{-1} W' X = MX \quad (26)$$

en posant $M = W(W' W)^{-1} W'$

• **Deuxième étape**

On régresse Y sur \hat{X} (et non plus sur X).

$$Y = \hat{X} C + U \quad (27)$$

L'estimateur \hat{C} des moindres carrés est donné par :

$$\hat{C} = (\hat{X}' \hat{X})^{-1} \hat{X}' Y \quad (28)$$

soit d'après (26) :

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

Par construction la matrice M est idempotente et donc $M'M = M$

En exprimant M en fonction de W :

$$\hat{C} = [X' W(W' W)^{-1} W' X]^{-1} X' W(W' W)^{-1} W' Y. \quad (29-)$$

C'est l'estimateur des variables instrumentales dans le cas général.

Remarque :

Si $K < l$: il y a moins d'instruments que de variables X.

Donc dans le calcul de \hat{X} il faudra calculer l vecteurs à partir d'un nombre de vecteurs indépendants inférieur à l . Les vecteurs obtenus ne sont pas linéairement indépendants et donc $(\hat{X}' \hat{X})$ n'est plus inversible.

La condition $K > l$ renvoie donc à une question de format.

Dans le cas où $K = l$, (l étant le nombre de variables de départ) alors $X' W$ est de format $(l \times T)$. ($T \times l$) soit (l, l) et donc l'inversion peut se faire par blocs à partir de (29) :

$$\hat{C} = (W' X)^{-1} (W' W)(X' W)^{-1} X' W(W' W)^{-1} W' Y \quad (30)$$

Soit

$$\hat{C} = (W' X)^{-1} W' Y \quad (31)$$

Estimateur simple des variables instrumentales.

Dans un programme d'ordinateur il est généralement prévu une procédure d'estimation par les variables instrumentales, I.V., à partir de laquelle l'ordinateur demande la liste des instruments, c'est-à-dire les variables de W sachant que les variables X qui sont exogènes sont leur propre instrument.

L'estimateur des V.I est sans biais car en remplaçant Y par $XA + U$ dans l'équation (29), il vient :

$$\bullet E(\hat{C}) = E \left[\left[X' W (W' W)^{-1} W' X \right]^{-1} X' W (W' W)^{-1} W' (XA + U) \right] \quad (32)$$

$$= A + \left[X' W (W' W)^{-1} W' X \right]^{-1} X' W (W' W)^{-1} W' E(U) \quad (33)$$

$E(U) = 0 \Rightarrow E(\hat{C}) = A$ ce qui établit que l'estimateur est sans biais

$$\bullet \sigma_{(\hat{c})}^2 \sim (\hat{X}' \hat{X})^{-1} \sigma_u^2$$

$$\sigma_u^2 \text{ étant estimé par } \frac{RSS}{T - k}$$

On peut donc calculer la variance des estimateurs par la méthode habituelle sur la dernière régression.

Un domaine d'application important est constitué par l'économétrie des anticipations rationnelles. En effet lorsqu'on utilise une variable anticipée comme variable explicative, celle-ci résulte de l'application de la théorie économique pertinente appliquée aux données disponibles. Les variables résultant d'anticipations rationnelles sont donc, pour cette raison, assimilables à des variables endogènes.

Par exemple, on sait que le taux de change intègre le taux d'inflation anticipé.

$$ch =: f(\dots, f_t^a, \dots) + u_t \quad (34)$$

Mais le taux d'inflation anticipé, dans l'optique des anticipations rationnelles, est lui-même une fonction du taux de croissance de la masse monétaire, du solde des finances publiques, de la pression de la demande, etc.

$$f_t^a = g\left(\dots, \frac{\Delta m_t}{m_t}, \frac{T - G}{y}, \dots\right) + v_t \quad (35)$$

Dans l'équation (34) la variable explicative f_t^a ne peut plus être indépendante du terme d'erreur puisque celui-ci contient v_t d'après (35).

Plus généralement, on peut aussi se poser la question de savoir quelles sont, dans une équation, les variables explicatives qui sont véritablement exogènes et quelles sont celles qui ne le sont pas.

Cette question comme on l'a vu, est le choix de l'économètre dans un modèle à équation simultanée mais peut se discuter dans le cas d'une équation unique. Pour cela on peut utiliser le test d'exogénéité de HAUSMAN - WU.

Soit le modèle standard $Y = XA + U$

$$X_{(T,l)} = [x_1 \dots x_l]$$

On peut avoir des doutes sur l'exogénéité de certaines variables explicatives : supposons qu'il y ait m variables véritablement exogènes et donc $(l - m)$ variables pour lesquelles il y a un doute :

soit

Z_1 : ensemble des $(l - m)$ variables douteuses

Z_2 : ensemble des m autres variables

$$X = \begin{bmatrix} Z_1 & : & Z_2 \\ \vdots & & \vdots \end{bmatrix}$$

Le test consiste à tester :

$$H_0 : E(Z_1' U) = 0$$

$$H_1 : E(Z_1' U) \neq 0$$

On peut réécrire le modèle en faisant apparaître Z_1 et Z_2

$$Y = \underset{(T, l-m)}{Z_1} A_1 + \underset{(T, m)}{Z_2} A_2 + U \quad (36)$$

La régression (36) permet d'obtenir $\sum \hat{u}_t^2 = S_0$

On peut ensuite par la méthode des variables instrumentales calculer des estimateurs des variables douteuses, les \hat{Z}_1 . Ces estimations vont être rajoutées dans l'équation (36), ce qui conduit à estimer

$$Y = Z_1 A_1 + Z_2 A_2 + \hat{Z}_1 B_1 + V \quad (37)$$

On obtient $\sum \hat{v}_t^2 = S_1$ à partir de cette régression avec $S_1 < S_0$.

Le rapport $\frac{S_0 - S_1}{\sigma_u^2}$ où σ_u^2 représente la variance résiduelle dans S_0 , suit une loi de χ_{m-1}^2 (à $m - 1$ degrés de liberté).

Principe du test

Si les variables douteuses (c'est-à-dire les Z_1) sont vraiment exogènes, leur explication par les V.I n'est pas bonne et donc l'introduction de \hat{Z}_1 dans (37) n'a pas beaucoup d'effet sur la somme des carrés des résidus.

Par conséquent S_1 sera peu inférieur à S_0 .

Si par contre les Z_1 sont vraiment endogènes, alors l'introduction des \hat{Z}_i diminuera sensiblement la somme des carrés des résidus d'où la règle de décision suivante :

$$\frac{S_0 - S_1}{\sigma} < X_\alpha^2 \Rightarrow \text{on accepte } H\emptyset$$

$$\frac{S_0 - S_1}{\sigma} > X_\alpha^2 \Rightarrow \text{on rejette } H\emptyset$$

Exemple :

Soit une fonction de consommation élémentaire :

$$c_t = a_0 + a_1 y_t + a_2 c_{t-1} + u_t \quad (38)$$

On peut douter de l'exogénéité de y car le revenu dépend fortement de la consommation. Dans ce cas, les variables explicatives peuvent être réparties entre :

$$z_1 = y_t$$

$$z_2 = c_{t-1}$$

En régressant (38) et on obtient S_0 .

On estime ensuite \hat{Z}_1 c'est-à-dire \hat{y}_t dans ce cas.

Pour cela on peut utiliser comme instruments

- . les dépenses gouvernementales ;
- . les exportations.

$$\text{Soit } y_t = b_0 + b_1 g_t + b_2 x_t + v_t \quad (39)$$

Ce qui permet de calculer \hat{y}_t , qui est rajoutée dans (38)

La nouvelle équation est $c_t = a_0 + a_1 y_t + a_2 c_{t-1} + a_3 \hat{y}_t$ (40) dont la somme des carrés des résidus fournit S_1 .

Ici le χ^2 aura un degré de liberté.

Il existe encore d'autres méthodes qui permettent d'estimer les modèles à équation simultanées. Par exemple on peut estimer l'ensemble des coefficients d'un modèle en une seule étape par la méthode du maximum de vraisemblance.

La méthode est une généralisation de celle qui s'applique dans le cas d'une équation unique :

Dans le modèle à une variable explicative

$$y_t = ax_t + b + u_t \quad (41)$$

$$u_t = y_t - ax_t - b \quad (42)$$

d'après l'hypothèse H7

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

$$\text{et } P(u_1) = \frac{1}{\sigma\sqrt{2\Pi}} e^{-\frac{1}{2} \left(\frac{u_1 - E(u_1)}{\sigma_u} \right)^2}$$

$P(u_1 \dots u_T) = P(u_1) \dots P(u_T)$ puisque ces événements sont indépendants d'après H4. En tenant compte de H2 et H3:

$$P(u_1 \dots u_T) = \left(\frac{1}{\sigma\sqrt{2\Pi}} \right)^T e^{-1/2 \frac{\sum u_i^2}{\sigma_u^2}} \quad (43)$$

N.B. : Maximiser la vraisemblance d'avoir observé $\hat{u}_1 \dots \hat{u}_t$ revient bien à minimiser $\sum u_t^2$, d'après l'équation (43). Ce qui vérifie l'équivalence entre la méthode des moindres carrés et celle du maximum de vraisemblance, dans ce cas.

Cette démarche peut être étendue à un modèle à plusieurs équations.

$$\begin{cases} y_{1t} - a_{12}y_{2t} \dots - a_{1n}y_{nt} - b_{11}x_{1t} \dots - b_{1m}x_{mt} = u_{1t} \\ \vdots \\ y_{nt} - \dots - a_{nn-1}y_{n-1t} - b_{n1}x_{1t} - \dots - b_{nm}x_{mt} = u_{nt} \end{cases} \quad (44)$$

Sur ce modèle on peut calculer une fonction de vraisemblance qui traduira la probabilité d'observer la valeur des résidus des T périodes simultanément.

$$L = P(\hat{u}_{11} \dots \hat{u}_{1T}, \dots, \hat{u}_{n1}, \dots, \hat{u}_{nT})$$

$$L = \prod_{i=1}^n \prod_{t=1}^T P_{it} \text{ sur les T périodes et les n résidus}$$

D'où

$$L = \left(\frac{1}{\sqrt{2\pi}} \right)^{nT} \left(\frac{1}{\prod_i \sigma_{u_i}} \right)^T e^{-\frac{1}{2} \left[\frac{\sum \hat{u}_{1t}^2}{\sigma_{u_1}^2} + \dots + \frac{\sum \hat{u}_{nt}^2}{\sigma_{u_n}^2} \right]} \quad (45)$$

Les estimateurs du maximum de vraisemblance sont les valeurs des coefficients a et b du modèle (44) qui maximisent la fonction (45).

On les obtient en dérivant la fonction de vraisemblance par rapport aux coefficients a et b.

$$\begin{cases} \frac{\partial L}{\partial a_{11}} = 0 \\ \vdots \\ \frac{\partial L}{\partial b_{nm}} = 0 \end{cases}$$

Il est donc possible d'utiliser la méthode du maximum de vraisemblance pour estimer simultanément tous les coefficients d'un modèle.

De toute façon une telle comparaison n'a guère de valeur probante car il est impossible de connaître la vraie valeur du coefficient estimé.

Pour comparer de façon plus pertinente les propriétés des différentes méthodes, on utilise la méthode dite de Monte Carlo dans laquelle on se donne d'avance les valeurs des coefficients recherchés. On crée ensuite des séries de variables expliquées en donnant des valeurs aux variables explicatives. Enfin on ajoute à ces séries un terme d'erreur dont on a spécifié la distribution et dont les valeurs sont obtenues par tirage.

Exemple

$$c_t = a_0 + a_1 y_t$$

. on se donne a_0 et a_1 on tire une série de y

Ceci permet de calculer une chronique certaine de $c_t = a_0 + a_1 y_t$

. On spécifie la distribution de u_t

Exemple

$$u_t \rightarrow N(0; \sigma_u^2)$$

On tire au hasard T valeurs de u_1 à u_T

On calcule $c_t = a_0 + a_1 y_t + u_t$

. On régresse

$$c_t = a_0 + a_1 y_t + u_t$$

Ce qui permet d'obtenir les estimateurs \hat{a}_0 et \hat{a}_1 que l'on compare avec les vraies valeurs de a_0 et a_1 .

Cette comparaison permet d'apprécier les résultats d'une méthode d'estimation.

- En général quand on fait des simulations par cette méthode, on multiplie les séries, par exemple en modifiant la distribution du terme d'erreur ou les valeurs des variables explicatives et on s'intéresse aux propriétés en moyenne des estimateurs.

De même on peut modifier T : la longueur des séries pour analyser leurs propriétés asymptotiques : se rapprochent-ils de leurs vraies valeurs quand on augmente le nombre d'observations ?

Des études par la méthode de Monte Carlo ont été faites et permettent d'avoir un jugement au moins partiel sur l'ensemble des méthodes d'estimation.

Globalement ces études sont assez convergentes. L'une des premières a été réalisée par Basmann :

Étude de BASMANN (1958)⁽¹⁾

Le modèle utilisé est un modèle à trois équations dont l'une est :

$$y_1 - 2y_2 + 1,5 y_3 + 3z_1 - 0,6z_5 + 10 = u_1$$

Dans cette équation on sait d'avance que les valeurs vraies sont :

$$\begin{cases} a_1 = -2 \\ a_2 = 1,5 \\ a_3 = 3 \\ a_4 = -0,6 \end{cases}$$

y = variables endogènes

z = variables exogènes.

Basmann a réalisé 200 expériences et a calculé les valeurs moyennes obtenues pour ces coefficients. Soient α_i les estimateurs obtenus pour le coefficient a_i .

$$\bar{\alpha}_i = \frac{1}{200} \sum_{k=1}^{200} \alpha_{ik}$$

Méthodes d'estimation	α_1	α_2	α_3	α_4
$\bar{\alpha} \left\{ \begin{array}{l} MCO \\ DMC \\ MVIL \end{array} \right.$	- 0,89	1,13	2,33	- 0,59
	- 1,94	1,47	2,95	- 0,65
	- 2,69	1,66	3,52	- 0,60

On vérifie que les estimateurs des MCO sont biaisés et que leurs valeurs sont assez différentes des valeurs vraies des coefficients de régression. Des trois

(1) R.L. Basmann "An Experimental Investigation of Some Small Sample properties of Generalized Classical Linear Estimations of Structural Equations, Some Preliminary Results". G.E. Company ; Hanford Laboratories, 1958.

groupes d'estimateurs, les MCO sont en moyenne les plus éloignés des valeurs vraies des paramètres.

Ensuite Basmann s'est intéressé à la dispersion des coefficients α_i autour de leur moyenne $\bar{\alpha}_i$. Et pour cela il a calculé $\frac{1}{200} \sum (\alpha_i - \bar{\alpha}_i)^2$.

Variance des estimateurs α_i :

	α_1	α_2	α_3	α_4
MCO	1,01	0,05	1,08	0,46
DMC	2,33	0,14	2,10	0,63
MVIL	17,39	1,08	14,97	1,74

Le tableau montre que les estimateurs des MCO sont de loin les mieux groupés, alors que ceux des MVIL ont une variance considérable.

Dans un troisième temps on peut calculer la dispersion autour des valeurs vraies :

$$\frac{1}{200} \sum (\alpha_i - a_i)^2 = \frac{1}{200} \sum (\alpha_i - \bar{\alpha}_i)^2 + (\bar{\alpha}_i - a_i)^2$$

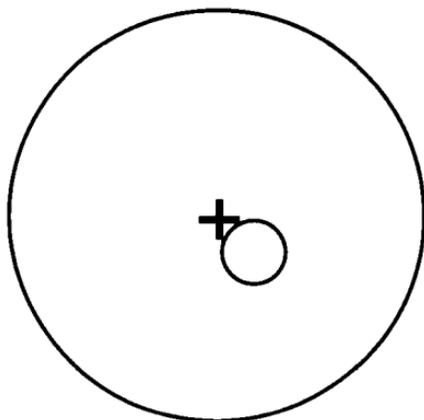
	α_1	α_2	α_3	α_4
MCO	2,24	0,19	1,53	0,46
DMC	2,43	0,14	2,10	0,63
MVIL	17,89	1,11	15,25	1,74

Les estimateurs des MCO ne sont pas tellement plus mauvais que les autres et ceci en raison de leur moindre dispersion. (dans trois cas sur quatre).

Parmi les deux autres méthodes dites exactes la meilleure de très loin est la méthode des DMC. La méthode du MVIL paraît effectivement peu fiable avec une énorme dispersion. Les méthodes correctes sont assez bien centrées sur les valeurs vraies mais avec une dispersion qui peut être importante.

Les estimateurs des MCO sont biaisés mais pas forcément de façon très importante et ils sont bien groupés. Passer des MCO à une méthode appropriée c'est passer d'un biais faible mais certain à un biais important mais pas certain.

On peut en donner une illustration graphique à partir du problème du tir à l'arc.



Si on veut atteindre le centre de la cible. Les méthodes appropriées sont centrées et permettent d'atteindre n'importe quel point de la cible.

La méthode des MCO ne permet d'atteindre que le petit cercle. Il n'est pas centré, mais le résultat n'est jamais très éloigné du centre.

Or, l'estimation d'une équation ou d'un modèle pour un ensemble de données, se fait une fois et non 200 fois. Par conséquent l'estimation s'apparente au tir d'une flèche unique vers la cible et dans ces conditions les propriétés de dispersion ont autant d'importance que les propriétés des valeurs moyennes.

Les résultats de Basmann sont relatifs à un modèle correctement spécifié. Mais la plupart du temps il y a aussi des problèmes de colinéarité, d'autocorrélation, ... Que peut-on dire quand il y a l'un ou l'autre de ces problèmes ?

Les résultats sont un peu différents et pas toujours faciles à interpréter.

En présence de forte colinéarité, les estimateurs des MCO sont meilleurs que ceux des DMC. Alors qu'en présence de faible colinéarité c'est l'inverse.

À partir d'un ensemble d'études, Koutsoyiannis a effectué un classement pour les différents cas suivants⁽¹⁾ :

(1) A. Koutsoyiannis "Theory of Econometrics" 2^e édition, Mac Millan 1981, pp 506-511.

Critère	Omission d'une	Multicollinéarité	Multicollinéarité
	variable		+ omission
Biais	<i>DMC</i>	<i>DMC</i>	<i>DMC</i>
	<i>MVIL</i>	<i>MVIL</i>	<i>MV</i>
	<i>MCO</i>	<i>MV</i>	<i>MCO</i>
	<i>MV</i>	<i>MCO</i>	<i>MVIL</i>
Variance autour de la moyenne	<i>MCO</i>	<i>MCO</i>	<i>DMC</i>
	<i>MV</i>	<i>MV</i>	<i>MCO</i>
	<i>DMC</i>	<i>DMC</i>	<i>MV</i>
	<i>MVIL</i>	<i>MVIL</i>	<i>MVIL</i>
Dispersion autour de la valeur vraie	<i>DMC</i>	<i>DMC</i>	<i>DMC</i>
	<i>MVIL</i>	<i>MV</i>	<i>MCO</i>
	<i>MCO</i>	<i>MCO</i>	<i>MV</i>
	<i>MV</i>	<i>MVIL</i>	<i>MVIL</i>

CONCLUSION

- Il n'est finalement pas complètement absurde d'estimer un modèle à équations multiples équation par équation par les MCO.

En effet, certains, voire la plupart des modèles continuent à être estimés par les MCO bien qu'on sache que c'est incorrect.

- Parmi les méthodes théoriquement correctes la seule qui ait un intérêt pratique est la méthode des DMC. Quand on travaille sur un modèle à équations simultanées il est prudent de faire une estimation par les DMC et de comparer les résultats avec ceux des MCO, au moins pour certaines équations importantes.

Chapitre III : L'économétrie sans modèle, les séries temporelles

Dans les deux premiers chapitres l'estimation se fait à partir de relations structurelles censées traduire de mécanismes économiques. La modélisation consiste à traduire en équations la théorie économique. Dans ce chapitre ces éléments structurels disparaissent. De ce fait, on ne cherche plus à expliquer mais simplement à décrire ou à prévoir. Pour cela on utilise des séries c'est-à-dire des observations d'une variable au cours du temps.

L'utilisation de ces séries se généralise pour tout un ensemble de problèmes pour lesquels la modélisation traditionnelle apparaît peu satisfaisante.

C'est le cas pour des phénomènes complexes dans lesquels il y a de nombreuses actions et réactions simultanées pour lesquelles il est difficile de faire apparaître clairement un enchaînement de causes et d'effets. (Exemple : cours d'une valeur mobilière au jour le jour). Cela peut aussi être le cas de variables extrêmement volatiles. Les modèles explicatifs sont en général incapables de prévoir correctement les points de retournement. Ces modèles sont performants à moyen terme mais peu performants à court terme. Le taux de change, qui subit des fluctuations assez courtes et assez amples, fournit des illustrations de ce type de difficultés.

Une des causes de mauvaise performance des modèles explicatifs vient du fait qu'ils sont beaucoup plus performants pour analyser des tendances que des fluctuations. Une grande partie de leur pouvoir explicatif est tiré du parallélisme des évolutions. Cet argument laisse entendre que certaines corrélations ne correspondent pas véritablement à des relations entre variables, mais peuvent tout simplement résulter d'une évolution semblable des variables sous des influences n'ayant rien à voir avec le problème étudié. C'est la question des corrélations fortuites sur laquelle on reviendra dans le chapitre suivant. Un test intéressant est de comparer le pouvoir explicatif de la même relation macro-économique pour des périodicités annuelle, trimestrielle ou mensuelle.

Dans tout ces cas, quand on a besoin de faire une prévision à court terme et quand les enjeux de cette prévision sont économiquement importants on préfère utiliser des modèles non explicatifs mais assez performants en prévision qui sont donc des modèles de série temporelle.

Ce sont des structures légères qu'on peut facilement réestimer. Le modèle peut donc être mis à jour tous les jours. Le coût de la prévision est relativement faible. Un économètre bien formé avec un bon logiciel peut assurer au quotidien la prévision au jour le jour.

Un des domaines d'application est le domaine des variables financières (taux d'intérêts, taux de change, taux de rendement).

- La méthode consiste à rechercher dans l'histoire de la variable des régularités susceptibles d'aider à prévoir ses valeurs futures.

Remarque : On peut aussi associer les séries temporelles à une équation structurelle : c'est ce qu'on appelle parfois des fonctions de transfert ou des ARMAX et cela permet d'améliorer les qualités en prévision d'un modèle structurel.

Dans ce chapitre, on commencera par présenter une classe particulière de séries qui présentent des propriétés intéressantes. On analysera ensuite les propriétés des différentes catégories de modèles de time séries avant de présenter quelques éléments permettant leur estimation.

SECTION 1 : LES PROPRIÉTÉS DES PROCESSUS STATIONNAIRES

L'utilisation des séries temporelles conduit à rechercher des régularités dans les valeurs passées de la série. Pour que cette démarche ait un sens pour la prévision, il faut que le processus présente une certaine stabilité ou un certain degré d'invariance au cours du temps.

- Si le processus était complètement erratique on pourrait peut être tirer quelques enseignements du passé mais ces éléments n'auraient aucun intérêt prédictif..

C'est cette idée de stabilité ou d'invariance au cours du temps qui est traduite par la notion statistique de stationnarité.

On peut définir la stationnarité d'un processus de série temporelle par les 3 conditions suivantes :

- 1^{re} condition : un processus stochastique de série temporelle est invariant s'il a une moyenne constante au cours du temps :

$$E(y_t) = m \quad \forall t$$

- 2^e condition : ce processus a une variance constante $\sigma_{y_t}^2 = E[(y_t - m)^2]$

$$\text{soit } \sigma_{y_t}^2 = \gamma_0 \quad \forall t$$

- 3^e condition : les covariances entre deux observations de y pour deux périodes différentes ne dépendent que du nombre de décalages.

$$E[(y_t - m)(y_{t-k} - m)] = E[(y_{t+l} - m)(y_{t+l-k} - m)] = \gamma_k$$

$$\text{cov}(y_t, y_{t-k}) = \text{cov}(y_{t+l}, y_{t+l-k}) = \gamma_k$$

On peut retenir l'image d'un processus régulier qui s'inscrit dans une bande horizontale, centrée sur sa moyenne.

À partir de ces propriétés on peut calculer empiriquement et théoriquement des coefficients d'auto-covariance γ_k et des coefficients d'auto-corrélation qu'on désignera par ρ_k coefficient d'auto-corrélation entre y_t et y_{t-k} ou y_{t+k} .

On peut définir des auto covariances qui seront désignées par γ_k

$$\gamma_k = \text{cov}(y_t, y_{t-k}) = \text{cov}(y_t, y_{t+k}).$$

Si le processus est stationnaire cette covariance ne dépend que du nombre de décalages.

$$\text{Covariance particulière : } \gamma_0 = \text{cov}(y_t, y_t) = \sigma_{y_t}^2$$

De même on peut définir des coefficients d'auto-corrélation qu'on désignera par $\rho_k = \rho(y_t, y_{t+k})$

Par application de la formule statistique,

$$\rho_k = \frac{\text{cov}(y_t, y_{t+k})}{\sigma_{y_t} \sigma_{y_{t+k}}}$$

Si le processus est stationnaire la variance de y_t est constante au cours du temps.

Alors

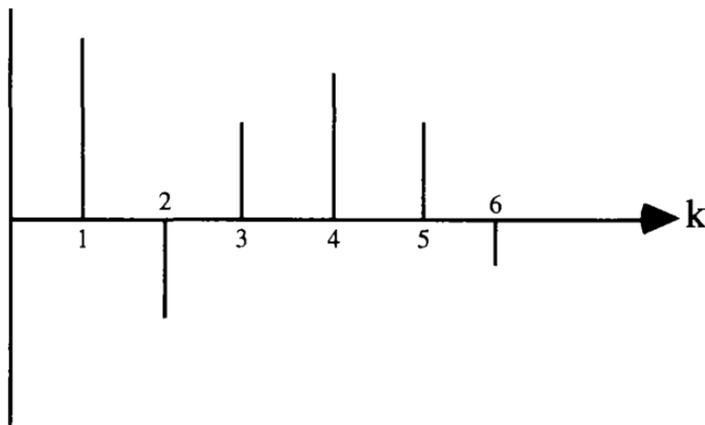
$$\sigma_{y_t} \sigma_{y_{t+k}} = \sigma_{y_t}^2 = \gamma_0 \text{ d'où } \rho_k = \frac{\text{cov}(y_t, y_{t+k})}{\gamma_0} = \frac{\gamma_k}{\gamma_0}$$

Tous les coefficients d'auto-corrélation s'obtiennent facilement à partir des coefficients d'auto-covariance. Il suffit de les rapporter à γ_0 .

Si un processus est stationnaire, son évolution subit une sorte d'amortissement au cours du temps, ce qui se traduit statistiquement par une décroissance rapide des coefficients d'auto-corrélation.

Pour une série quelconque, (par exemple le taux d'intérêt du marché monétaire) on peut toujours calculer des coefficients d'auto-corrélation empiriques.

Ces coefficients peuvent être représentés graphiquement en fonction de k :



Si le processus est stationnaire ces coefficients d'auto-corrélation diminueront assez rapidement.

Un processus non stationnaire peut avoir une résonance de la période t à une période plus éloignée, de telle sorte que des coefficients d'auto-corrélation élevés puissent exister pour des valeurs élevées de k .

Seuls les processus stationnaires sont intéressants car ils sont stables. Il est possible de tester la stationnarité d'un processus c'est-à-dire en fait la décroissance

rapide de ses coefficients d'auto-corrélation. C'est la première opération à faire lorsqu'on étudie une série.

Pour cela il existe différents tests.

Définition

On appelle "bruit blanc" (white noise) un processus stationnaire de moyenne nulle et dont les autocovariances sont nulles à l'exception de γ_0 .

Exemple : $y_t = u_t$ sous les hypothèses faites sur u_t est un bruit blanc.

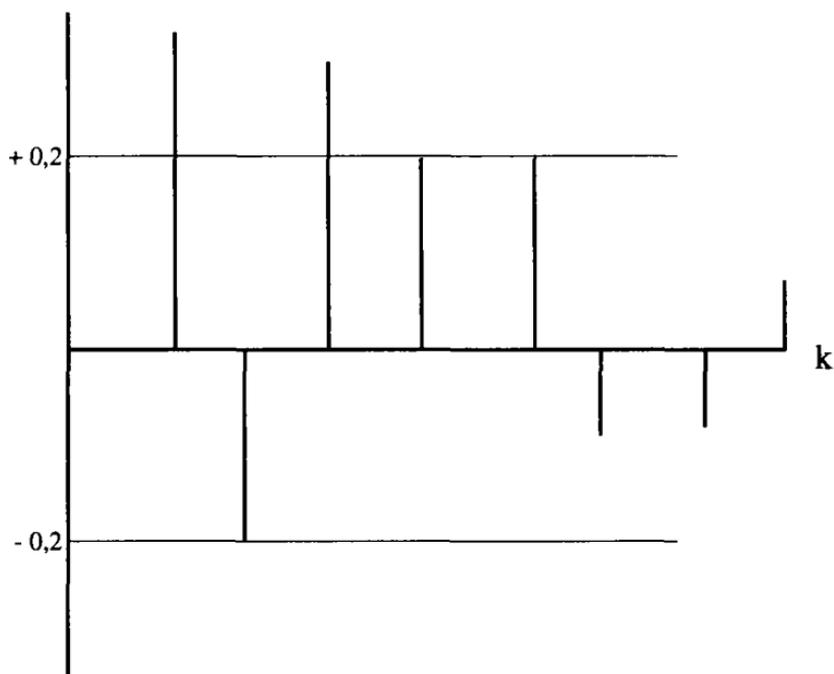
Si le processus étudié est un bruit blanc, Bartlett a montré que les coefficients d'auto-corrélation suivent une loi normale de moyenne nulle et d'écart-type

$$\frac{1}{\sqrt{T}} \cdot \rho \rightarrow N(0, \frac{1}{\sqrt{T}})$$

• Le test de Bartlett consiste donc à vérifier si les coefficients d'auto-corrélation successifs sont significativement différents de 0.

Exemple

Pour $T = 100$, tout coefficient inférieur à $- 0,2$ ou supérieur à $+ 0,2$ sera donc réputé significatif à 95 %.



On utilise aussi deux autres tests qui portent sur un ensemble de coefficients d'auto-corrélation et non sur les coefficients pris individuellement. Ce sont les tests de Box et Pierce et de Ljung et Box.

Test de Box et Pierce

Il consiste à calculer la somme des carrés des k premiers coefficients d'autocorrélation, puis à multiplier cette somme par le nombre d'observations.

$$Q = T \sum_1^k \rho_k^2 \quad \text{Ce produit suit une loi de } \chi^2.$$

$Q \rightarrow \chi^2$ à k - nombre de paramètres estimés (soit $k - p - q$ dans le cas général).

T = nombre de périodes

Le test se formule de la façon suivante :

$$H_0: \sum_k \rho_k^2 = 0$$

$$H_1: \sum_k \rho_k^2 \neq 0$$

Si χ^2 calculé $<$ χ_α^2 on conclut H_0 , c'est-à-dire que les coefficients d'autocorrélation ne sont pas significativement différents de zéro.

Avantage du test

Il permet de ne faire qu'un nombre limité de tests et reste valable quelle que soit la moyenne, à la différence du test de Bartlett.

Mais ce test est peu performant sur de petits échantillons. On le remplace alors par le test de Ljung et Box.

Test de Ljung et Box

$$Q^* = T(T+2) \sum_1^k (T-K)^{-K} \rho_k^2$$

L'interprétation est la même que pour le test de Box et Pierce

$$Q^* \rightarrow \chi_{k-p-q}^2$$

Exemple : on s'intéresse à la série des taux d'inflation trimestriels mesurés par le taux de variation des prix à la consommation, en France de 1985.II à 1993.II.

Sur les 30 premiers coefficients d'auto-corrélation, donc pour $k = 30$, on calcule les tests de stationnarité.

Box et Pierce = 35,75

Ljung et Box = 53,39

Aucun coefficient n'ayant été estimé, ces tests suivent une loi du χ^2 à 30 degrés de liberté. La valeur critique de la table est $\chi^2 = 43,77$.

On constate que les conclusions divergent puisque pour Box et Pierce, la stationnarité est acceptée, alors qu'elle ne l'est pas pour Ljung et Box. Compte tenu du petit nombre d'observations, c'est cette dernière conclusion qui doit être retenue.

Mais, un certain nombre de processus ne sont pas stationnaires. Dans le cas d'un processus non stationnaire il faut essayer de le rendre stationnaire. Pour cela, la méthode la plus courante consiste à appliquer l'opérateur "différence" à la série de départ.

Soit y_t la série de départ.

Alors $w_t = y_t - y_{t-1} = \Delta y_t$

En utilisant l'opérateur retard L ceci est encore égal à $(1-L)y_t$

Sur w_t (la nouvelle série) on peut à nouveau effectuer les tests de stationnarité :

. w_t est stationnaire, il est inutile de continuer.

. w_t n'est pas stationnaire. Il faut continuer la procédure en calculant une nouvelle série :

$$z_t = \Delta w_t = w_t - w_{t-1} = (1-L)w_t = (1-L)^2 y_t$$

De même, on teste la stationnarité de z_t

Si nécessaire on pourra appliquer une troisième fois l'opérateur "différence". Théoriquement, un processus qui est stationnaire après l'application de l'opérateur différence reste stationnaire après toute nouvelle application de l'opérateur c'est-à-dire que si w_t est stationnaire, z_t est stationnaire.

Définition

Un processus qui est stationnaire après une différenciation est dit **intégré d'ordre 1** :

- Si w_t est stationnaire, " y_t est intégré d'ordre 1"
- Si w_t non stationnaire, mais, z_t stationnaire " y_t est intégré d'ordre 2"

Le degré d'intégration d'une série est donc le nombre de fois qu'il faut appliquer l'opérateur "différence" pour rendre le processus stationnaire.

Remarque : L'opérateur retard " L " s'utilise comme un opérateur mathématique ordinaire :

$$\left\{ \begin{aligned} z_t &= (1-L)^2 y_t \\ &= (1-2L+L^2)y_t \\ &= y_t - 2Ly_t + L^2 y_t \\ &= y_t - 2y_{t-1} + y_{t-2} \end{aligned} \right.$$

Tous les processus temporels ne peuvent pas être rendus stationnaires aussi facilement.

Les processus peuvent être représentés par différents modèles qui ont des propriétés particulières.

SECTION 2 : LES PROPRIÉTÉS DES MODÈLES DE SÉRIES TEMPORELLES (TIME SÉRIES)

Il existe différents modèles de séries temporelles :

$$\left\{ \begin{array}{l} \text{AR} \\ \text{MA} \\ \text{ARMA} \\ \text{ARIMA} \end{array} \right.$$

(AR) = modèles auto régressif (Auto Régressive)

(MA) = moyenne mobile (Moving Average)

(ARMA) = combinaison d'un AR et d'un MA

(ARIMA) = processus non stationnaire auto-régressif intégré.

Sous sa forme générale, un processus de série temporelle fait dépendre une variable de ses propres valeurs passées ainsi que des valeurs passées du terme d'erreur.

Soit en utilisant l'opérateur retard :

$$\alpha(L)y_t = a(L)u_t + \delta \quad (1)$$

δ = terme constant, nul si la série est un bruit blanc.

Pour pouvoir estimer économétriquement un tel processus il faut d'abord caractériser sa structure. En particulier déterminer le nombre de termes qui figurent dans les deux polynômes de retard.

En l'absence de cette précision la structure ne peut pas être estimée.

Il y a donc une opération préalable à l'estimation qui consiste à identifier ou fixer la structure à estimer. Et, pour fixer cette structure on peut s'aider des propriétés des processus élémentaires (MA, AR, ARMA). Ces propriétés servent à reconnaître un type de modèle.

A- Le modèle MA (q)

C'est un processus qui a de la mémoire sur le terme d'erreur.

$$y_t = m + u_t - a_1 u_{t-1} - \dots - a_q u_{t-q}. \quad (2)$$

ou de façon équivalente $y_t = a(L)u_t + m$

Sous les hypothèses habituelles concernant les termes d'erreur, ce processus a comme moyenne m : $E(y_t) = m$

Et la variance du terme d'erreur est constante $E(u_t^2) = \sigma_u^2$

Les termes d'erreur sont eux-mêmes indépendants au cours du temps

$$E(u_t u_{t+k}) = 0 \text{ si } k \neq 0$$

Propriétés permettant de l'identifier et de connaître q :

Ce sont les propriétés du coefficient d'auto-corrélation du MA (q). Pour les faire apparaître, il faut passer par les auto-covariances puisque

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

Calcul de γ_0 :

$$\begin{aligned} E(y_t - m)^2 &= E(u_t - a_1 u_{t-1} - \dots - a_q u_{t-q})^2 \\ &= E(u_t^2 + a_1^2 u_{t-1}^2 + \dots + a_q^2 u_{t-q}^2 + \dots - 2a_1 u_t u_{t-1} + \dots \\ &\quad + 2a_q a_{q-1} u_{t-q} u_{t-q-1}) \quad (3) \end{aligned}$$

Or, les termes d'erreurs satisfont aux hypothèses habituelles donc

$E(u_t u_{t+k}) = 0$ par hypothèse (si $k \neq 0$). Ce qui entraîne que tous les doubles produits sont nuls.

Sous l'hypothèse d'homoscédasticité $E(u_t^2) = E(u_{t-k}^2) = \dots = \sigma_u^2$

D'où $\gamma_0 = \sigma_u^2(1 + a_1^2 + \dots + a_q^2)$ (4)

Pour que cette variance soit finie (condition de la stationnarité) il faut que $1 + a_1^2 + \dots + a_q^2 < \infty$.

Calcul de γ_1

$$\gamma_1 = \text{cov}(y_t, y_{t-1})$$

$$\gamma_1 = E[(y_t - m)(y_{t-1} - m)] \quad (5)$$

$$= E[(u_t - a_1 u_{t-1} - \dots - a_q u_{t-q})(u_{t-1} - a_1 u_{t-2} - \dots - a_q u_{t-q-1})] \quad (6)$$

Puisque les termes d'erreurs sont indépendants au cours du temps il ne reste que les termes relatifs au même décalage.

$$\gamma_1 = \sigma^2[-a_1 + a_1 a_2 + \dots + a_q a_{q-1}] \quad (7)$$

Calcul de ρ_1

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{-a_1 + a_1 a_2 + \dots + a_q a_{q-1}}{1 + a_1^2 + \dots + a_q^2} \quad (8)$$

• Soit le processus simple MA (1)

$$y_t = m + u_t - a_1 u_{t-1} \quad (9)$$

• $E(y_t - m)^2 = \text{variance de } y_t$

$$= E[(u_t - a_1 u_{t-1})^2] = \sigma_u^2(1 + a_1^2) = \gamma_0 \quad (10)$$

$$\bullet \gamma_1 = E[(y_t - m)(y_{t-1} - m)] \quad (11)$$

$$\Rightarrow y_{t-1} = m + u_{t-1} - a_1 u_{t-2} \quad (12)$$

$$d' où \gamma_1 = E[(u_t - a_1 u_{t-1})(u_{t-1} - a_1 u_{t-2})] \quad (13)$$

$$soit \gamma_1 = -a_1 \sigma_u^2. \quad (14)$$

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{-a_1}{1 + a_1^2} \quad (15)$$

$$\bullet \gamma_2 = E[(y_t - m)(y_{t-2} - m)] \quad (16)$$

$$y_{t-2} = m + u_{t-2} - a_1 u_{t-3} \quad (17)$$

$$\gamma_2 = E[(u_t - a_1 u_{t-1})(u_{t-2} - a_1 u_{t-3})] = 0 \quad (18)$$

$$or, \rho_2 = \frac{\gamma_2}{\gamma_0} \text{ donc } \rho_2 = 0 \quad (19)$$

Donc, pour un processus MA (1), les coefficients d'auto-corrélation ρ_k sont nuls pour tout $k > 1$.

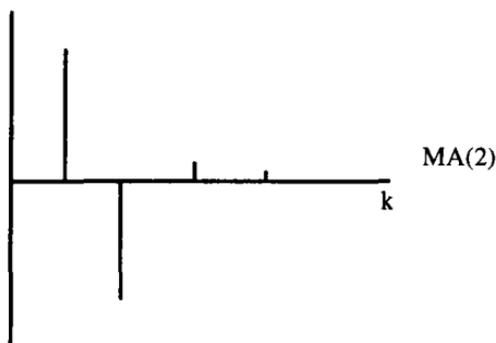
De façon plus générale, pour un processus MA (q), les coefficients d'auto-corrélation sont nuls pour tout décalage $k > q$.

Soit un MA (q) avec $k > q$

$$\gamma_k = E[(u_t - a_1 u_{t-1} \dots a_q u_{t-q})(u_{t-k} - a_1 u_{t-k-1} \dots a_q u_{t-k-q})] \quad (20)$$

Pour $k > q$ il n'y a aucun élément commun entre les deux parenthèses et donc $\gamma_k = 0$ c'est-à-dire $\rho_k = 0$.

Inversement, quand la fonction d'auto-corrélation chute brutalement au bout de q décalages c'est l'indice que le processus pourrait bien être un MA (q).



Un processus MA pose un problème d'estimation car il contient un terme d'erreur composé qui ne peut pas satisfaire les hypothèses habituelles de l'économétrie, même si individuellement les u_t satisfont ces propriétés.

Supposons maintenant qu'on connaisse une fonction inverse de $a(L)$ soit $[a(L)]^{-1}$, alors $y_t [a(L)]^{-1} = u_t$ est un processus qui présente des conditions d'estimation plus favorable car il n'y a plus que u_t comme terme d'erreur. C'est l'opération d'inversion d'un processus MA qui peut être nécessaire pour estimer ce processus.

Exemple

Soit le processus MA (1) supposé être un bruit blanc.

$$y_t = u_t - a_1 u_{t-1} \quad (21)$$

$$\text{soit } y_t = (1 - a_1 L) u_t \quad (22)$$

$$\text{ou } \frac{y_t}{1 - a_1 L} = u_t \quad (23)$$

On fait ainsi apparaître $\frac{1}{1 - a_1 L}$

• Si $a_1 < 1$, $\frac{1}{1 - a_1 L}$ peut s'interpréter comme la somme des termes d'une progression géométrique :

$$\frac{1}{1 - a_1 L} = 1 + a_1 L + a_1^2 L^2 + \dots + a_1^n L^n \quad (L \text{ étant l'opérateur décalage})$$

d'où

$$\frac{y_t}{1 - a_1 L} = y_t (1 + a_1 L + \dots + a_1^n L^n) \quad (24)$$

Avec la définition de l'opérateur retard, cela peut encore s'écrire :

$$y_t + a_1 y_{t-1} + \dots + a_1^n y_{t-n} = u_t \quad (25)$$

Le processus (25) est alors l'inverse du processus (21).

On voit ainsi que l'inverse d'un processus MA est un processus AR de durée infinie.

Ce calcul fait aussi apparaître la condition d'inversibilité d'un MA : cette condition a été exprimée ici sous la forme $a_1 < 1$.

Si on reprend le polynôme de retard $1 - a_1 L$ cette condition a été exprimée sur a_1 : $a_1 < 1$ pour que le processus soit inversible. Mais on peut aussi l'exprimer par $L > 1$.

$$1 - a_1 L = 0 \Rightarrow L = \frac{1}{a_1}$$

L est la racine ici unique, du polynôme de retard $1 - a_1 L$.

La condition d'inversibilité peut aussi s'exprimer par $L > 1$.

Dans un MA (1) la condition d'inversibilité peut s'exprimer alternativement sous la condition $a_1 < 1$ ou bien racine en L du polynôme de retard > 1 .

Cette condition est facilement généralisable et fournit la condition d'inversibilité dans le cas général.

Un processus MA (q) est inversible si les racines du polynôme de retard en L sont situées en dehors du cercle unité. Cette condition rend compte des cas de racines réelles et de racines complexes.

Exemple

$$\text{MA (2)} \quad y_t = u_t - a_1 u_{t-1} - a_2 u_{t-2}$$

$$\text{soit } y_t = u_t (1 - a_1 L - a_2 L^2) \quad \text{Le polynôme de retard est } 1 - a_1 L - a_2 L^2$$

En posant $1 - a_1L - a_2L^2 = 0$, on fait apparaître 2 racines L' et L'' qui peuvent être réelles ou complexes.

Pour que le MA (2) soit inversible il faut que ces racines soient de module > 1 .

Remarque : Cette question a été abordée dans le chapitre II avec les propriétés dynamiques des modèles. La comparaison entre les méthodes suivies aux chapitre II et III permet de mieux situer les deux démarches sur un exemple.

Chapitre II

$$y_t - a_1y_{t-1} - a_2y_{t-2}$$

$$y_t = r^t$$

$$r^2 - a_1r - a_2 = 0$$

polynôme

← caractéristique →

Polynôme de retard en L

$$y_t - a_1y_{t-1} - a_2y_{t-2}$$

$$y_t(1 - a_1L - a_2L^2)$$

$$1 - a_1L - a_2L^2 = 0$$

$$\Delta = a_1^2 + 4a_2$$

$$r' = \frac{a_1 + \sqrt{\Delta}}{2}$$

$$r'' = \frac{a_1 - \sqrt{\Delta}}{2}$$

← racines →

$$\Delta = a_1^2 + 4a_2$$

$$L' = \frac{a_1 + \sqrt{\Delta}}{-2a_2}$$

$$L'' = \frac{a_1 - \sqrt{\Delta}}{-2a_2}$$

On vérifie aisément que $r'L'' = r''L' = 1$. Il est donc logique de retrouver sur les racines en L des conditions de stationnarité qui soient inverses des conditions sur r.

On voit que la stationnarité peut s'aborder de deux façons :

- . de façon empirique;
- . de façon mathématique.

Il se peut aussi que le polynôme de retard admette des solutions où $L = 1$.

On dira alors que "le polynôme en L admet une ou plusieurs racines unitaires". Le processus ne peut alors plus être stationnaire et cela donne une troisième façon d'aborder le problème de la stationnarité : en utilisant des tests de racine unitaire.

B- Propriété du modèle auto régressif AR(p).

Comme précédemment, ces propriétés serviront à reconnaître de tels processus.

Soit le modèle AR(1) :

$$y_t = \alpha_1 y_{t-1} + \delta + u_t \quad (26)$$

La valeur moyenne de ce processus est donnée quand $y_t = m \quad \forall_t$. Dans ce cas :

$$m = \alpha_1 m + \delta \quad (27)$$

d'où

$$m = \frac{\delta}{1 - \alpha_1}$$

Si on pose $\delta = 0$ alors $m = 0$ (Hypothèse de bruit blanc).

$$\gamma_0 = E(y_t^2) = E[(\alpha_1 y_{t-1} + u_t)^2] = E[\alpha_1^2 y_{t-1}^2 + u_t^2 + 2\alpha_1 y_{t-1} u_t] \quad (28)$$

$$E(y_{t-1}^2) = \gamma_0 \quad (\text{cf. hypothèse de stationnarité})$$

$$y_{t-1} \text{ ne peut pas contenir } u_t \text{ d'après (26) et donc } E(y_{t-1} u_t) = 0$$

d'où

$$\gamma_0 = \frac{\sigma_u^2}{1 - \alpha_1^2}$$

$$\gamma_1 = E(y_t y_{t-1})$$

$$= E[(\alpha_1 y_{t-1} + u_t) y_{t-1}]$$

$$= E[\alpha_1 y_{t-1}^2 + u_t y_{t-1}]$$

$$= \alpha_1 \gamma_0$$

$$\gamma_1 = \alpha_1 \gamma_0$$

$$\boxed{\rho_1 = \frac{\gamma_1}{\gamma_0} = \alpha_1 \quad (29)}$$

$$\begin{aligned}\gamma_2 &= E(y_t y_{t-2}) \\ &= E[(\alpha_1 y_{t-1} + u_t) y_{t-2}] \quad (30) \\ &= \alpha_1 \gamma_1\end{aligned}$$

$$\text{Or, } \gamma_1 = \alpha_1 \gamma_0$$

$$\text{donc } \gamma_2 = \alpha_1^2 \gamma_0$$

$$\boxed{\rho_2 = \alpha_1^2}$$

$$\begin{aligned}\gamma_3 &= E(y_t y_{t-3}) \\ &= E[(\alpha_1 y_{t-1} + u_t)] y_{t-3} \quad (31)\end{aligned}$$

$$\text{Soit } \gamma_3 = \alpha_1 \gamma_2$$

$$\text{Or, } \gamma_2 = \alpha_1^2 \gamma_0$$

$$\text{Donc } \gamma_3 = \alpha_1^3 \gamma_0$$

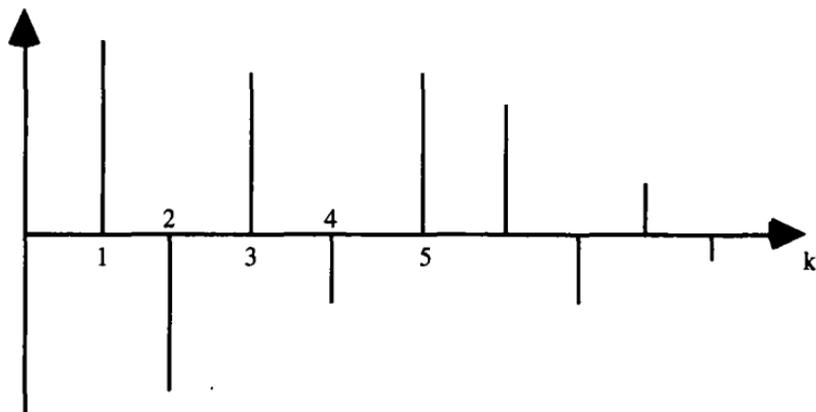
$$\boxed{\rho_3 = \alpha_1^3}$$

Pour un processus AR (1) les coefficients d'auto-corrélation décroissent de façon géométrique à partir du premier.

Ce résultat se généralise pour un AR (p).

- Dans un AR (p) les coefficients d'auto-corrélation sont géométriquement décroissants au-delà de $k = p$.

Symétriquement un processus qui fait apparaître une décroissance géométrique des coefficients d'auto-corrélation au-delà de $k = p$ a de bonnes chances d'être un AR (p).



→ Le processus représenté pourrait être un AR (5).

Un AR(2) s'écrit :

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + u_t \quad (32)$$

(En supposant qu'il s'agit toujours d'un bruit blanc)

$$\left(\text{si } \delta \neq 0 \text{ alors } m = \frac{\delta}{1 - \alpha_1 - \alpha_2} \right) \rightarrow \text{d'où l'intérêt de poser } \delta = 0$$

$$E(y_t^2) = \gamma_0 = E[(\alpha_1 y_{t-1} + \alpha_2 y_{t-2} + u_t)(y_t)] = \alpha_1 \gamma_1 + \alpha_2 \gamma_2 + E(y_t u_t) \quad (33)$$

$$\text{or, } E(y_t u_t) = E[(\alpha_1 y_{t-1} + \alpha_2 y_{t-2} + u_t)(u_t)] = \sigma_u^2 \quad (34)$$

Donc

$$\gamma_0 = \alpha_1 \gamma_1 + \alpha_2 \gamma_2 + \sigma_u^2 \quad (35)$$

$$\gamma_1 = E(y_t y_{t-1})$$

$$= E[(\alpha_1 y_{t-1} + \alpha_2 y_{t-2} + u_t) y_{t-1}] \quad (36)$$

$$= \alpha_1 \gamma_0 + \alpha_2 \gamma_1$$

D'où

$$\gamma_1 = \alpha_1 \gamma_0 + \alpha_2 \gamma_1 \quad (37)$$

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \alpha_1 + \alpha_2 \rho_1 \quad (38)$$

$$\rho_1 = \frac{\alpha_1}{1 - \alpha_2} \quad (39)$$

$$\gamma_2 = E[(\alpha_1 y_{t-1} + \alpha_2 y_{t-2} + u_t)(y_{t-2})] = \alpha_1 \gamma_1 + \alpha_2 \gamma_0$$

soit

$$\gamma_2 = \alpha_1 \gamma_1 + \alpha_2 \gamma_0$$

$$\rho_2 = \alpha_1 \rho_1 + \alpha_2 \quad (40)$$

$$\text{ou } \rho_2 = \frac{\alpha_1^2}{1 - \alpha_2} + \alpha_2 \quad (41)$$

Remarque : Dans un processus AR (2) il y a deux coefficients à calculer (α_1 et α_2). Les équations (39) et (41) constituent un système à deux inconnues (α_1 et α_2). Connaissant les valeurs numériques de ρ_1 et ρ_2 ces équations permettent théoriquement de calculer α_1 et α_2 .

Ces équations sont dites équations de Yule-Walker permettant en effet de calculer les coefficients du modèle à partir des coefficients d'auto-corrélation.

Toutefois pour utiliser ces équations il faut savoir que le processus est un processus AR (2). Sinon il n'y a aucune raison de s'intéresser au système des équations (39) et (41).

$$\gamma_3 = E[(\alpha_1 y_{t-1} + \alpha_2 y_{t-2} + u_t)y_{t-3}] = \alpha_1 \gamma_2 + \alpha_2 \gamma_1 \quad (42)$$

$$\rho_3 = \alpha_1 \rho_2 + \alpha_2 \rho_1$$

Et les coefficients successifs seront obtenus en multipliant par α_1 et α_2 .

$$\gamma_4 = E[(\alpha_1 y_{t-1} + \alpha_2 y_{t-2} + u_t)y_{t-4}] = \alpha_1 \gamma_3 + \alpha_2 \gamma_2$$

$$\rho_4 = \alpha_1 \rho_3 + \alpha_2 \rho_2$$

$$\rho_k = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2} \quad (43)$$

Pour AR (2) au delà de $k = 2$ on retombe donc dans la situation de décroissance géométrique des coefficients d'auto-corrélation.

Ce résultat se généralise pour tout AR(p) pour $k > p$

C- Le processus ARMA (P, Q)

C'est la combinaison d'un AR (p) et d'un MA (q).

$$\alpha(L)y_t = \delta + a(L)u_t \quad (44)$$

Il contient un polynôme de retard sur y et un polynôme de retard sur u avec p et q éléments : $y_t = \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \delta + u_t - a_1 u_{t-1} - \dots - a_q u_{t-q}$ (45)

Comme pour les modèles élémentaires, certaines propriétés permettent de reconnaître un processus ARMA.

Soit un ARMA (1.1). Il s'écrit :

$$y_t = \alpha_1 y_{t-1} + u_t - a_1 u_{t-1} \quad (\text{Hypothèse de bruit blanc}) \quad (46)$$

Condition de stationnarité du processus ARMA.

De façon générale un processus ARMA est stationnaire si sa partie AR est elle-même stationnaire. La partie MA d'un processus ARMA n'est pas susceptible, normalement, d'entraîner des évolutions non stationnaires.

Dans le cas d'un ARMA (1.1) on a :

$$1 - \alpha_1 L_1 = 0 \text{ et la condition de stationnarité est donc :}$$

$$L > 1$$

$$\gamma_0 = E(y_t^2) = E[(\alpha_1 y_{t-1} + u_t - a_1 u_{t-1})^2] \quad (47)$$

$$= E[\alpha_1^2 y_{t-1}^2 + u_t^2 + a_1^2 u_{t-1}^2 + 2\alpha_1 y_{t-1} u_t - 2\alpha_1 a_1 y_{t-1} u_{t-1} - 2a_1 u_t u_{t-1}]$$

Sous les hypothèses habituelles

$$\gamma_0 = \alpha_1^2 \gamma_0 + \sigma_u^2 + a_1^2 \sigma_u^2 - 2\alpha_1 a_1 \sigma_u^2 \quad (48)$$

En effet

$$y_{t-1} = \alpha_1 y_{t-2} + u_{t-1} - a_1 u_{t-2} \text{ et donc}$$

$$E(y_{t-1} u_{t-1}) = \sigma_u^2$$

$$\text{d'où } \gamma_0 = \sigma^2(1 + a_1^2 - 2a_1 \alpha_1) + \alpha_1^2 \gamma_0$$

Soit

$$\gamma_0 = \frac{\sigma_u^2(1+a_1^2-2\alpha_1a_1)}{1-\alpha_1^2} \quad (49)$$

$$\gamma_1 = E[(\alpha_1 y_{t-1} + u_t - a_1 u_{t-1})y_{t-1}] \quad (50)$$

$$\gamma_1 = \alpha_1 \gamma_0 - a_1 \sigma_u^2 \quad (51)$$

$$\gamma_2 = E[(\alpha_1 y_{t-1} + u_t - a_1 u_{t-1})y_{t-2}]$$

$$\text{donc } \gamma_2 = \alpha_1 \gamma_1 \quad (52)$$

$$\gamma_3 = E[(\alpha_1 y_{t-1} + u_t - a_1 u_{t-1})y_{t-3}]$$

$$\gamma_3 = \alpha_1 \gamma_2 \quad (53)$$

ou

$$\gamma_3 = \alpha_1^2 \gamma_1$$

A partir des équations (50) à (53) on calcule aisément les coefficients d'auto-corrélation :

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \alpha_1 - \frac{a_1(1-\alpha_1^2)}{1+a_1^2-2\alpha_1a_1} \quad (54)$$

$$\rho_2 = \alpha_1 \frac{\gamma_1}{\gamma_0} = \alpha_1 \rho_1 \quad (55)$$

$$\rho_3 = \frac{\gamma_3}{\gamma_0} = \alpha_1^2 \rho_1 \quad (56)$$

On constate que dans un processus ARMA (p, q), la partie MA disparaît lorsque $k > q$. À partir de $k = q$, le processus ARMA aura les mêmes propriétés qu'un processus AR simple et donc on retombe sur la décroissance géométrique des coefficients d'auto-corrélation. Il n'est pas toujours facile pour cette raison de distinguer un processus ARMA d'un processus AR de même ordre sur p.

C'est pourquoi il est intéressant d'utiliser aussi un autre indicateur pour identifier les processus ARMA

Exemple : si en $k = 5$ il y a décroissance géométrique cela peut être AR (5) mais aussi un ARMA (5,3). (Le décrochement sur $k = 3$ n'est pas toujours évident).

D- Le modèle ARIMA (p, d, q)

C'est un modèle non stationnaire mais qu'on peut rendre stationnaire en passant aux différences d fois. C'est par conséquent un processus intégré d'ordre d .

Exemple : ARIMA (4, 1, 3)

On le rendra stationnaire en le différenciant une fois.

$$\text{Si } y_t = \text{ARIMA}(4, 1, 3)$$

$$(1 - L)y_t = w_t = \Delta y_t = \text{ARMA}(4,3)$$

$$\text{Par conséquent ARIMA}(p, 0, q) = \text{ARMA}(p,q)$$

Jusqu'ici on s'est appuyé sur la fonction d'auto-corrélation, appelée aussi fonction d'auto-corrélation totale, pour mettre en évidence certaines propriétés des différents modèles étudiés. Mais ce critère peut être insuffisant et on utilise aussi une autre fonction qui est la fonction d'auto-corrélation partielle.

E- Fonction d'auto-corrélation partielle

On s'intéresse uniquement au processus AR (p).

Soit un processus AR (p) qu'on va chercher à identifier. Cela revient à déterminer p .

• Si $p = 1$ alors le processus va s'écrire :

$$y_t = \alpha_1 y_{t-1} + u_t$$

$$\text{donc } \gamma_1 = E[(\alpha_1 y_{t-1} + u_t)(y_{t-1})] = \alpha_1 \gamma_0$$

d'où $\rho_1 = \alpha_1$ comme on l'a vu dans l'équation (29)

Si le processus est un processus AR (1) il n'y a qu'un seul coefficient à calculer, c'est α_1 et il vaut ρ_1 .

• Si le processus est un AR (2) :

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + u_t \text{ et}$$

$$\rho_1 = \frac{\alpha_1}{1 - \alpha_2}$$

$$\rho_2 = \alpha_1 \rho_1 + \alpha_2 \quad (40)$$

Si le processus est un AR(2) les équations (39) et (40) permettent de calculer α_1 et α_2 (Équations de YULE WALKER).

• Si le processus est un AR (3)

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 y_{t-3} + u_t$$

On obtient alors trois équations permettant de calculer $\alpha_1, \alpha_2, \alpha_3$.

Plus généralement pour un AR(P) qui s'écrit :

$$y_t = \alpha_1 y_{t-1} \dots \alpha_p y_{t-p} + u_t$$

$$\gamma_0 = E[y_t (\alpha_1 y_{t-1} \dots \alpha_p y_{t-p} + u_t)] = \alpha_1 \gamma_1 + \alpha_2 \gamma_2 \dots \alpha_p \gamma_p + \sigma_u^2 \quad (57)$$

$$\gamma_1 = E[y_{t-1} (\alpha_1 y_{t-1} \dots \alpha_p y_{t-p} + u_t)] = \alpha_1 \gamma_0 + \alpha_2 \gamma_1 \dots \alpha_p \gamma_{p-1} \quad (58)$$

$$\gamma_2 = E[y_{t-2} (\alpha_1 y_{t-1} \dots \alpha_p y_{t-p} + u_t)] = \alpha_1 \gamma_1 + \alpha_2 \gamma_2 \dots \alpha_p \gamma_{p-2} \quad (59)$$

.

.

.

$$\gamma_k = E[y_{t-k} (\alpha_1 y_{t-1} \dots \alpha_p y_{t-p} + u_t)] = \alpha_1 \gamma_{k-1} + \alpha_2 \gamma_{k-2} \dots \alpha_p \gamma_{p-k} \quad (60)$$

Soient les coefficients d'auto-corrélations $\rho_1 \dots \rho_k$ obtenus en divisant les covariances par γ_0 :

$$\rho_1 = \alpha_1 + \alpha_2 \rho_1 + \dots \alpha_p \rho_{p-1} \quad (61)$$

$$\rho_2 \dots \alpha_1 \rho_1 + \alpha_2 \rho_2 + \dots \alpha_p \rho_{p-2} \quad (62)$$

.

.

.

$$\rho_k = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2} + \dots \alpha_p \rho_{p-k} \quad (63)$$

On appelle coefficient d'auto-corrélation partiels les valeurs de $\alpha_1, \alpha_2, \dots, \alpha_k$ que l'on trouve en posant successivement $p = 1, 2, 3, \dots, k$ dans le système des équations (61) à (63) qui sont les équations de Yule-Walker.

Cette fonction d'auto-corrélation partielle peut aussi être calculée pour les processus ARMA et elle présente certaines propriétés qui peuvent servir à l'identification d'un processus de série temporelle.

C'est un critère qui viendra compléter le critère de l'auto-corrélation totale. Dans la réalité, le choix est facilité par le fait qu'on n'utilise pas de processus trop long, p et $q < 5$, sauf en cas de saisonnalité où des décalages de 4 ou 12 périodes peuvent être nécessaires.

Le tableau ci-dessous, tiré de Cuthbertson, Hall et Taylor⁽¹⁾ donne quelques éléments permettant d'identifier les processus simples

Critères d'identification à partir des propriétés des fonctions d'auto-corrélation totale et partielle

Fonction d'auto-corrélation totale (ρ_k)		Fonction d'auto-corrélation partielle (α_k)
<ul style="list-style-type: none"> • Un pic pour $k = 1$ 	MA (1)	<ul style="list-style-type: none"> • Décroissance exponentielle
<ul style="list-style-type: none"> • Une décroissance exponentielle 	AR (1)	<ul style="list-style-type: none"> • Pic pour $k = 1$
<ul style="list-style-type: none"> • Un pic pour $k = 1$ suivi par une décroissance exponentielle 	ARMA (1,1)	<ul style="list-style-type: none"> • Un pic pour $k = 1$ puis décroissance exponentielle
<ul style="list-style-type: none"> • Deux pics : $k = 1$ et $k = 2$ suivi par décroissance exponentielle 	ARMA (1,2)	<ul style="list-style-type: none"> • Un pic pour $k = 1$ puis décroissance exponentielle
<ul style="list-style-type: none"> • Un pic pour $k = 1$ puis décroissance exponentielle 	ARMA (2,1)	<ul style="list-style-type: none"> • Deux pics pour $k = 1$ et $k = 2$ puis décroissance exponentielle
<ul style="list-style-type: none"> • Deux pics $k = 1$ et $k = 2$ suivis d'une décroissance exponentielle 	ARMA (2,2)	<ul style="list-style-type: none"> • Deux pics pour $k = 1$ et $k = 2$ suivis par décroissance exponentielle

L'appréciation de ces critères suppose une certaine expérience.

(1) H. Cuthbertson, S. Hall, M., Taylor. "Applied Econometric Techniques". Philip Allan 1992.

Au-delà de $k = 2$ l'identification procède à partir des mêmes principes.

Les critères ne sont pas très précis et il existe en général plusieurs modèles possibles avec différentes valeurs de p et de q pour une série chronologique donnée.

Il faut estimer les différents processus possibles et vérifier à posteriori si ces processus conviennent. Autrement dit il y a un contrôle à posteriori possible sur les choix qui ont été faits.

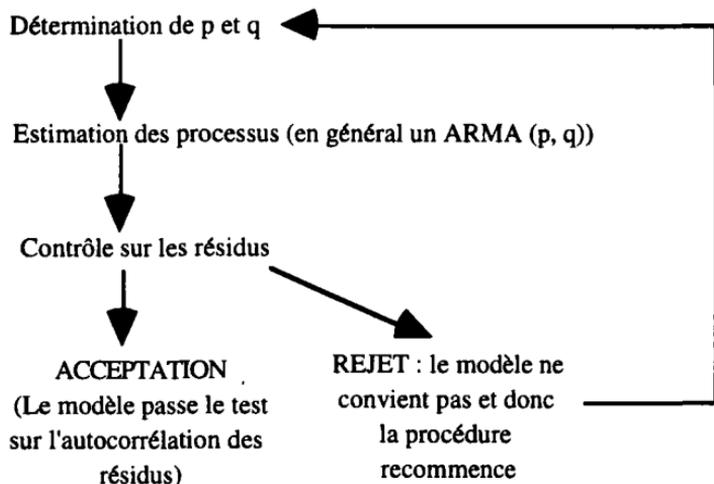
En effet si le processus utilisé est correct, les résidus de la série c'est-à-dire les \hat{u}_t ne contiennent plus aucune autocorrélation significative car tout a été absorbé par le modèle.

Il faut donc tester les coefficients d'auto-corrélation avec les tests habituels (Bartlett, Box et Pierce...) appliqués à la série des résidus.

Si dans la série des résidus on observe des auto-corrélations significatives il faut alors rejeter le modèle. Parmi les modèles possibles, certains ne passeront pas ce test et il restera un nombre assez limité de modèles acceptables.

Toutefois il est exceptionnel qu'il ne reste qu'un seul bon modèle. Pour départager les bons modèles on peut aussi examiner ce que ces modèles donnent sur la période récente.

Box et Jenkins ont proposé la méthode suivante qui résume les différentes étapes de la démarche :



SECTION 3 : L'ESTIMATION DES PROCESSUS ARMA

Un processus ARMA est en général une double distribution

$$\alpha(L)y_t = \delta + a(L)u_t$$

- Quand $q = 0$, il se réduit à un AR (p) et ne pose pas de problème d'estimation particulier.

En effet, sous les hypothèses habituelles un AR (p) a un bon terme d'erreur et peut être estimé directement.

- Si $q \neq 0$: c'est un ARMA (p, q).

Si le terme en u_t possède les propriétés habituelles alors le polynôme de retard en u_t ne peut pas les avoir.

L'estimation d'un ARMA (p, q) suppose qu'on puisse se ramener à une équation dans laquelle on isole à droite u_t . Pour cela on suppose qu'il existe $[a(L)]^{-1}$ le polynôme inverse du polynôme $a(L)$.

L'application de cette transformation sur la partie gauche de l'équation ramène à une formulation du type :

$$[a(L)]^{-1} \alpha(L) y_t = u_t$$

C'est donc sous cette dernière forme qu'on peut estimer un modèle ARMA.

Comme on peut le constater, l'équation ainsi obtenue fera apparaître de nombreuses non-linéarités et ces non-linéarités rendront impossible l'utilisation d'une méthode d'estimation simple comme par exemple, les MCO.

On doit donc recourir à des méthodes itératives donc des méthodes dans lesquelles on se donne des valeurs d'amorçage qui peuvent être tirées des conditions de Yule Walker.

Dans certains cas, ces méthodes d'itération sont convergentes. Il peut arriver qu'elles ne le soient pas.

Il peut arriver aussi que la convergence se fasse sur des solutions locales et qu'en fait le modèle admette d'autres solutions pour des valeurs d'amorçage différentes.

La plupart des logiciels d'économétrie ont des méthodes d'estimation des processus de séries temporelles.

Chapitre IV : Introduction à la cointégration

La plupart des propriétés statistiques des méthodes d'estimation s'appliquent à des variables stationnaires, c'est-à-dire qu'elles ne sont pas valables pour n'importe quel type de données.

Or on applique indifféremment ces méthodes d'estimation à des variables stationnaires et à des variables non stationnaires. D'où la possibilité que ces propriétés statistiques ne soient plus valables pour des variables non stationnaires.

Cet aspect de la méthode économétrique a été ignoré pendant longtemps avant de resurgir au cours des dernières années. Il pose des problèmes redoutables dans la mesure où la plupart des séries représentant des variables économiques sont affectées par une évolution de long terme. C'est par exemple le cas de pratiquement toutes les variables macro-économiques, y compris des variables comme le chômage ou le taux d'inflation sur 20 à 30 ans. C'est aussi vrai pour les données de beaucoup d'entreprises ou d'unités individuelles. En fait, les séries dont on peut logiquement penser a priori qu'elles devraient remplir les conditions de la stationnarité sont peu nombreuses : les taux d'intérêt, certains taux de change (F/dollar mais pas F/Dm par exemple), les prix de certaines matières premières comme l'or ou l'argent.

On s'est intéressé au cas particulier des variables avec racine unitaire. En effet, le modèle avec racine unitaire constitue un cas limite puisqu'il n'est plus stationnaire au sens du chapitre III.

$$(1) y_t = y_{t-1} + u_t$$

est un modèle à racine unitaire ($L = 1$)

$$(1 - L) y_t = u_t$$

$$\rightarrow 1 - L = 0$$

Remarque : Un modèle de ce type s'appelle une "marche au hasard" ("Random Walk")

Ce n'est pas un processus stationnaire :

$$y_{t-1} = y_{t-2} + u_{t-1} \quad (2)$$

$$y_t = y_{t-2} + u_t + u_{t-1} \quad (3)$$

La variance ne peut donc pas être constante, comme on le constate en poursuivant le développement de l'équation (3). Ce n'est donc pas un processus stationnaire.

GRANGER ET NEWBOLD (1974) ont montré à l'aide de simulations empiriques que la distribution du t de Student dans un modèle simple du type $y_t = ax_t + b + u_t$ en présence de racine unitaire ne peut plus être interprétée selon la loi de Student Fisher.

C'est-à-dire que si x_t et y_t peuvent être mises dans la forme d'une AR(1) avec racine unitaire, comme dans l'équation (1), la distribution n'est plus standard.

De plus, entre des séries admettant des racines unitaires il existe des corrélations fortuites ("Spurious Correlation") qui n'ont aucune signification véritable et qui n'ont pas de véritable fondement.

L'économètre peut être amené à conclure à l'existence de relations qui en réalité n'existent pas.

Granger et Newbold ont montré que cela est particulièrement probable lorsque la régression comporte un R^2 élevé et un DW faible.

En 1986, PCB Phillips a confirmé de façon mathématique ces résultats et a également montré qu'en présence de racine unitaire le DW était aussi biaisé vers la valeur zéro. De ce fait la question de savoir s'il était légitime de continuer à travailler sur des séries non stationnaires s'est trouvée posée. En toute hypothèse la première chose à faire consiste à vérifier si cette propriété est remplie. C'est à cela que servent les tests de racine unitaire.

SECTION 1 : LES TESTS DE RACINE UNITAIRE

Pour vérifier la stationnarité des séries, il faut pratiquer des tests de stationnarité ou des tests de racine unitaire. Dans l'équation (4)

$$(4) y_t = ay_{t-1} + u_t$$

on va tester :

$$\begin{cases} H_0: |a| = 1 \\ H_1: |a| < 1 \end{cases}$$

Pour des raisons de commodité on transforme l'équation (4) de façon à pouvoir tester par rapport à la valeur zéro.

$$y_t - y_{t-1} = (a-1)y_{t-1} + u_t \quad (5)$$

$$\Delta y_t = (a-1)y_{t-1} + u_t \quad (6)$$

On teste alors :

$$\begin{cases} H_0: a-1 = 0 \\ H_1: a-1 < 0 \end{cases}$$

Pour ce test on utilise le test de Student mais sachant que la distribution du test de Student sous l'hypothèse H_0 n'est plus standard.

Sous H_0 cette distribution a été tabulée par Dickey et Fuller d'où le nom de test de Dickey-Fuller (1976) qui a été donné au plus courant des tests de racine unitaire.

Il existe une version dans laquelle on introduit dans (6) les accroissements décalés de la variable expliquée

$$\Delta y_t = (a-1)y_{t-1} + \Delta y_{t-1} + \dots + \Delta y_{t-n} + u_t \quad (6)$$

connue sous le nom de Augmented Dickey Fuller (ADF), qui est une version plus générale du test.

Les valeurs critiques du test se lisent dans la table de Dickey-Fuller.

Supposons que l'on ait une équation avec 2 variables explicatives et 25 observations.

$$T = 25$$

$$l = 2 = \text{nombre de variables explicatives}$$

Les valeurs critiques de la table de Dickey-Fuller diffèrent sensiblement de celle de Student.

Seuil unilatéral de Student : - 1,71

$$t > - 1,71 \text{ on conclut } H_0 \text{ c'est-à-dire } a = 1$$

$t < -1,71 \rightarrow H_1$ c'est-à-dire qu'il n'y a pas de racine unitaire

Dans la table de D & F le seuil correspondant vaut - 3.

$t > -3 \rightarrow H_0$ c'est-à-dire racine unitaire

$t < -3 \rightarrow H_1$ c'est-à-dire pas de racine unitaire

Il existe donc une différence substantielle entre les deux tests.

Si la valeur calculée du test est par exemple -2, il existe deux conclusions opposées selon la table utilisée, d'où l'importance de faire le bon choix.

Exemple : on s'intéresse à la variable log PIB pour la France en données trimestrielles sur la période 1980.I-1993.IV.

Le test pratiqué est le test de Dickey-Fuller :

DF = -0,416

Valeur critique à 5 % = -2,914

La variable log PIB n'est donc pas stationnaire.

En prenant les différences premières de log PIB, le test donne le résultat suivant :

DF = -5,964

Valeur critique à 5 % : -2,915

Les différences premières sont stationnaires. La variable log PIB est donc intégrée d'ordre 1.

Le test de racine unitaire (test standard) s'interprète comme un test de stationnarité. En effet, si une série possède au moins une racine unitaire elle ne peut pas être stationnaire tandis qu'une série qui n'a pas de racine unitaire peut être stationnaire.

On peut donc utiliser le test de Dickey-Fuller pour tester la stationnarité d'un processus temporel au sens du chapitre précédent, (ARMA ou ARIMA) au lieu du test de BOX et PIERCE.

Les deux tests posent en fait le même problème.

Il existe une autre façon de tester une racine unitaire à partir du test de Durbin Watson pratiqué sur $y_t = y_{t-1} + u_t$. C'est le test CRDW (Cointégration Régression DW), méthode proposée par Sargan & Bhargava (83). C'est un test moins usuel

que le test de Dickey-Fuller dont les valeurs critiques ont été obtenues par une retabulation de la table de DW.

Le test de Dickey-Fuller se pratique aussi sur les résidus des équations. Quand il est appliqué à des résidus, il porte le nom de test de ENGLE et GRANGER (1987).

SECTION 2 : LA COÏNTEGRATION

Face au problème qui a été rappelé au début de ce chapitre, la théorie de la cointégration permet de préciser les conditions dans lesquelles il est légitime de travailler sur des séries non stationnaires.

Définition

Soit deux séries y_t et x_t toutes les deux intégrées d'ordre d :

$$\begin{cases} x_t \sim I_d \\ y_t \sim I_d \end{cases}$$

Si il existe un vecteur $(\alpha, \beta) \neq 0$ tel que $z_t = \alpha x_t + \beta y_t$

soit intégré d'ordre $d - b$ avec $0 < b < d$, alors on dira que x_t et y_t sont cointégrées d'ordre (d, b) et (α, β) s'appelle vecteur cointégrant.

Par exemple, avec deux séries qui sont intégrées d'ordre 1 si une combinaison linéaire de ces deux séries est par exemple, intégrée d'ordre 0 on dira que ces deux variables sont co intégrées d'ordre $(1, 1)$.

Concrètement si on prend $c_t = a y_t + c_0$,

si une combinaison linéaire des variables c_t et y_t est stationnaire elles sont cointégrées.

Dans le cas le plus usuel où les variables sont intégrées d'ordre 1 avec deux variables il existe deux relations de cointégration qui sont réciproques l'une de l'autre.

$$\begin{cases} y_t = \alpha x_t + u_t & (1) \\ x_t = \alpha' y_t + v_t & (2) \end{cases}$$

Une combinaison particulière de x_t et y_t par rapport à la relation (1) peut être $u_t = y_t - ax_t$. Le vecteur cointégrant est alors $(1, -a)$. On peut tester le fait que $y_t - ax_t - b$ est intégrée d'ordre 0 en pratiquant un test de racine unitaire sur la série des résidus. C'est le test de Engle et Granger.

Si par exemple on travaille sur des variables intégrées d'ordre 1, après régression le test de racine unitaire pratiqué sur les résidus montrera si ceux-ci sont

- . intégrés d'ordre 1 ; il y a présence de racine unitaire et non cointégration.
- . intégrés d'ordre 0 ; il n'y a pas de racine unitaire → les deux variables sont cointégrées.

Ce résultat ou cette définition de la cointégration pour deux variables se généralise à un nombre quelconque de séries, mais le vecteur cointégrant n'est plus unique.

Exemple : En données trimestrielles, les variables suivantes sont intégrées d'ordre 1, en France sur la période 1980.I-1993.IV :

- log PIB
- log M_2
- taux de chômage
- taux d'inflation
- coût d'opportunité mesuré par la différence de rémunération entre les bons du trésor à 5 ans et le taux des livrets A.

Un test de Engle et Granger pratiqué sur une relation entre ces variables donne le résultat suivant :

- . valeur calculée : -2,567
- . valeur critique à : -4,673

Par conséquent les variables ne sont pas cointégrées, en particulier le log de la masse monétaire M_2 et celui du PIB ne suivent pas des évolutions parallèles sur la période.

SECTION 3 : PROPRIÉTÉS DES RELATIONS DE COÏNTEGRATION

Quand des variables sont cointégrées on retrouve certaines propriétés habituelles de la méthode d'estimation par les MCO. En fait les estimateurs sont qualifiés de

super convergents c'est-à-dire qu'ils convergent plus vite que les MCO sur données stationnaires, vers les valeurs vraies des paramètres.

Cela semble impliquer que dans un cas de cointégration on puisse légitimement utiliser le test de Student normal.

En fait, l'interprétation est un peu plus délicate.

Dans le modèle

$$(1) \quad y_t = a_0 + a_1 x_t + u_t$$

$$\text{avec } x_t = x_{t-1} + v_t \quad (2)$$

si les termes d'erreur u_t et v_t sont normaux, de moyenne nulle et indépendants, la distribution des tests de Student et de Fisher est bien la distribution standard.

Lorsqu'il existe des corrélations non nulles entre u_t et v_t , ces distributions ne sont plus standards, mais il est possible de les corriger en spécifiant la relation existant entre u_t et v_t .

Pour cela, on suppose que u_t dépend de v_t pour des décalages de $-k$ à $+k$

$$(3) \quad u_t = \sum_{n=-k}^k \gamma_n v_{t-n} + e_t$$

et que les corrélations sont nulles pour des décalages supérieurs à k , c'est-à-dire si $|n| > k$.

En reportant cette expression dans (1) :

$$y_t = a_0 + a_1 x_t + \sum_{n=-k}^k \gamma_n v_{t-n} + e_t$$

soit puisque d'après (2)

$$\Delta x_t = v_t$$

$$(1') \quad y_t = a_0 + a_1 x_t + \sum_n \gamma_n \Delta x_{t-n} + e_t$$

Le nouveau terme d'erreur, e_t , de la régression (1') a été purgé de toute corrélation avec v_t . En effet si u_t dans (1) n'est pas indépendant de v_t , e_t est bien indépendant de v_t d'après (3).

La correction proposée par Saikhonen⁽¹⁾ (91) consiste à multiplier respectivement les F de Fisher et les t de Student obtenus sur la régression (1') par le rapport $\frac{\sigma_e^2}{\tilde{\lambda}^2}$

pour F ou $\frac{\sigma_e}{\tilde{\lambda}}$ pour t.

σ_e^2 étant obtenu à partir de (1') par l'estimateur habituel $\frac{1}{T-l} \sum \hat{e}_i^2$ où l est le nombre de paramètres estimés dans (1').

$\tilde{\lambda}$ est obtenu à partir d'un modèle AR appliqué à \hat{e}_t .

$$(4) \quad \hat{e}_t = \alpha_1 \hat{e}_{t-1} \dots \alpha_p \hat{e}_{t-p} + \varepsilon_t$$

$$\tilde{\lambda} = \frac{\sigma_\varepsilon}{(1 - \hat{\alpha}_1 \dots - \hat{\alpha}_p)}$$

avec $\sigma_\varepsilon^2 = \frac{1}{T-p} \sum \hat{\varepsilon}_i^2$ d'après (4).

On vérifie que si les termes d'erreur u_t et v_t sont bien indépendants, $\gamma_n = 0 \quad \forall n$ et donc u_t se confond avec e_t qui à son tour est égal à ε_t . Dans ce cas le facteur de correction est égal à 1, ce qui implique bien la possibilité d'utiliser les tests standards de Student ou de Fisher.

Dans le cas très courant où les variables sont intégrées d'ordre 1, le fait qu'elles soient coïntégrées implique que leurs évolutions au cours du temps soient parallèles c'est-à-dire que l'écart entre les deux séries soit stationnaire. C'est ce que traduit la stationnarité des résidus.

(1) Saikhonen, Pentti : "Asymptotically Efficient Estimation of Cointegration Regression" *Econometric Theory* 1991, p 1-21.

On peut montrer qu'il y a alors équivalence entre équations de cointégration et une catégorie particulière de modèle structurel appelé modèle à correction d'erreur (MCE). C'est le théorème de représentation de GRANGER (1983).

Les modèles à correction d'erreur représentent une catégorie de relations particulièrement intéressantes quand on travaille sur des variables intégrées d'ordre 1. Nous reviendrons sur cette question dans les chapitres V et VI.

Les résultats qui viennent d'être exposés permettent de préciser les conditions de validité de l'économétrie.

Les conditions de validité

L'économétrie traditionnelle sur des variables non stationnaires reste légitime à deux conditions principales :

- Ne pas mélanger dans une équation des variables ayant des ordres d'intégration différents.

On rejoint une question souvent abordée de façon intuitive : celle de l'homogénéité d'une équation quand dans une même équation il y a des variables qui sont affectées d'une tendance et des variables stationnaires, il est en général impossible d'obtenir une estimation satisfaisante.

C'est un problème connu depuis longtemps pour des fonctions comme :

$I_t = a_1 y_t + a_2 r_t + a_3$ pour lesquelles il est généralement impossible d'obtenir des estimateurs significatifs de a_2 .

- Dans une équation de nature dynamique quand on introduit des décalages sur les variables il faut respecter le même nombre de décalages sur toutes les variables ou au moins sur les principales variables.

Par conséquent, la théorie de la cointégration valide plutôt la démarche économétrique sous des conditions pas trop restrictives.

Dans un travail d'économétrie, il est donc recommandé de commencer par chercher l'ordre d'intégration des variables (test de racine unitaire puis différences, puis test de racine unitaire etc...).

Ensuite, il est bon de s'assurer que les relations économétriques qui ont été obtenues appartiennent bien à l'espace de cointégration des variables (non unicité

en général de la relation) en déterminant l'ordre d'intégration de la série des résidus et en montrant que l'ordre d'intégration des résidus est inférieur à l'ordre d'intégration des variables.

Le plus souvent on obtient différentes relations possibles pour estimer et on retiendra de préférence celles qui ont aussi le caractère de relation de cointégration c'est-à-dire celles dans lesquelles les variables sont cointégrées.

Moyennant ces précautions on peut considérer que la plupart des propriétés statistiques restent valables.

Chapitre V : Application à la consommation

La consommation est un domaine qui se prête particulièrement bien aux applications de l'économétrie. Les variables qui s'y rattachent, qu'il s'agisse de la consommation elle-même ou du revenu qui constitue la principale variable explicative, font preuve d'une assez grande stabilité et d'une forte inertie. Elles peuvent être traduites par des spécifications relativement simples.

La consommation peut être analysée à trois niveaux qui correspondent aux trois sections de ce chapitre.

SECTION I : LA CONSOMMATION GLOBALE

Le modèle de base, déduit de la théorie générale possède une propension moyenne à consommer décroissante et une propension marginale stable, mais inférieure à l'unité. La traduction la plus simple est donnée par :

$$(1) c_t = a_1 y_t + a_2 + u_t$$

avec $0 < a_1 \leq 1$ et $a_2 > 0$

où c_t représente la consommation par tête à prix constant et y_t le revenu par tête à prix constant.

Les très nombreuses études empiriques faites à partir de ce modèle ont bien confirmé la stabilité de la propension marginale à court terme, mais ont infirmé la décroissance de la propension moyenne. En particulier, une étude de Kuznets sur la consommation aux États-Unis a établi la quasi-stabilité de la propension moyenne à consommer depuis 1865 ⁽¹⁾.

On sait que parmi les très nombreuses reformulations, la plus intéressante est celle proposée par M. Friedman autour du concept de revenu permanent ⁽²⁾.

(1) S. Kuznets "National Product since 1869" NBER 1946.

(2) On trouvera un panorama de ces hypothèses et tentatives d'explication dans l'ouvrage de J.C. Eicher "Consommation et épargne" Ed. Sirey 1961.

Toutefois cette approche pose le problème de la mesure du revenu permanent et de la consommation permanente. M. Friedman fait l'hypothèse que les consommations transitoires sont nulles en moyenne pour une population donnée et que les revenus passés interviennent dans la composition du revenu permanent avec des poids géométriquement décroissants. Il en résulte une spécification à retards échelonnés :

$$(2) \quad c_t = K(1-\lambda) \left[y_t + \lambda y_{t-1} + \dots + \lambda^n y_{t-n} \right] \text{ avec } 0 < \lambda \leq 1$$

L'application de la transformation de Koyck permet de passer à une formulation auto régressive, en supposant que n tend vers l'infini.

$$c_t - \lambda c_{t-1} = K(1-\lambda) \left[y_t - \lambda^{n+1} y_{t-n-1} \right] \text{ d'où}$$

$$(3) \quad c_t = K(1-\lambda) y_t + \lambda c_{t-1}$$

dans laquelle la pension à consommer est :

à court terme : $K(1-\lambda)$

à long terme : K

On peut remarquer que dans cette spécification, l'effet d'une variation transitoire de revenu est parfaitement déterminé, ce qui est en opposition avec la théorie de départ. En effet, celle-ci postule qu'une variation de revenu transitoire affecte la consommation dans une proportion variable. Ce n'est là qu'un exemple du caractère réducteur de la formalisation par laquelle un ensemble de propriétés ou d'hypothèses est traduit par une relation mathématique.

La théorie du revenu permanent conduit à estimer une équation du type :

$$(4) \quad c_t = a_1 y_t + a_2 c_{t-1} + a_3 + u_t$$

avec $a_1 = K(1-\lambda)$

$$a_2 = \lambda$$

Cette spécification, qui s'ajuste généralement bien sur données macroéconomiques brutes, présente de nombreuses faiblesses.

a - L'évolution du revenu et de la consommation suivant des trajectoires à peu près identiques et les comportements faisant preuve d'inertie, les variables explicatives c_{t-1} et y_t seront très souvent colinéaires.

b - Si la fonction de départ, équation (2), comporte des termes d'erreurs indépendants au cours du temps, l'application de la transformation de Koyck rend probable la présence d'autocorrélation dans l'équation (4). De plus le modèle comporte une variable endogène décalée, le test de Durbin Watson n'est pas fiable et les coefficients de régression sont biaisés.

c - La formule de l'équation (4) ne contient que les déterminants principaux. Il convient de la compléter pour lui permettre de prendre en compte l'influence éventuelle de la répartition des revenus ⁽¹⁾, du chômage qui peut exercer un effet de frein ⁽²⁾, des liquidités ou du taux d'intérêt dont l'effet s'exerce principalement sur la demande de biens durables.

d - Les variables y_t et c_t sont généralement intégrées d'ordre 1, parfois d'ordre 2. Il faut s'assurer que les autres variables ont un ordre d'intégration équivalent et vérifier que la relation obtenue appartient bien à l'espace de cointégration des variables. Les travaux anciens ignorent cette précaution.

Les modèles précédents peuvent être utilisés pour une première approche, mais restent peu performants.

Exemple : Estimation d'une fonction de consommation sur le Niger en données annuelles 1962-80. Avec le modèle le plus simple :

$$c_t = 0,88 \text{ PIBRU} - 6,98$$

(46,7) (1,64)

$$R^2 = 0,997 \quad DW = 1,72 \quad \hat{\rho} = 0,475$$

(1) Soit en distinguant les revenus du travail et les revenus du capital dans une approche à la Kaldor qui peut être pertinente pour des économies caractérisées par de fortes inégalités de revenus, soit en séparant les prestations sociales qui sont presque entièrement consommées, des autres revenus dans le cas des économies d'Europe occidentale.

(2) Ce qui montre que la traduction du revenu permanent utilisée est bien imparfaite, car elle n'intègre pas cet élément d'anticipation comme elle le devrait.

PIBRU = PIB moins les revenus de l'uranium. Ceux-ci sont en majorité versés à l'État ou transférés à l'étranger et ne servent pas de support à la consommation des ménages. L'utilisation directe du PIB ou du revenu réel en variable explicative donne des valeurs de la propension marginale à consommer trop faibles, généralement inférieures à 0,7.

Le modèle tiré de la spécification la plus simple de la théorie du revenu permanent donne :

$$c_t = 0,853 \text{ PIBRU} + 0,113 c_{t-1} - 9,621$$

(96,43) (10,45) (19,58)

$$R^2 = 0,998 \quad DW = 1,69$$

Estimation obtenue en normant l'équation par PIB RU.

L'analyse graphique des résidus fait apparaître un écart important pour les années 1973 et 1974 correspondant à 2 années de grande sécheresse au cours desquelles la consommation a moins baissé que le revenu. L'introduction d'une variable muette CL qui vaut 1 pour 1973 et 1974 et 0 ailleurs permet d'effacer ces écarts :

$$c_t = 0,894 \text{ PIBRU} + 0,0811 c_{t-1} + 0,993 \text{ CL} - 11,035$$

(76,4) (6,34) (2,80) (17,8)

$$R^2 = 0,9988 \quad DW = 1,57$$

La propension à consommer à long terme est alors de 0,97, ce qui correspond à un taux d'épargne moyen de l'ordre de 3 %, compatible avec les chiffres enregistrés pour cette variable pendant la période.

Remarque : dans cet exemple, on peut aussi reprendre le problème à l'envers, en supposant que l'on connaît mieux la consommation que le PIB ou le revenu. Le comptable national utilisera les estimations des chiffres de la consommation pour faire son évaluation du revenu ou du PIB en se donnant à priori les coefficients de la fonction... on touche là une limite évidente de la procédure d'estimation dans ce cas particulier.

Cette classe de fonctions s'ajuste bien dans la plupart des cas, en données annuelles ou trimestrielles. Toutefois, leur qualité est sans doute surestimée du fait de la stabilité des évolutions. On constate, en effet, qu'elles sont

incapables de décrire les retournements ou les inflexions importantes. Ce sont, par conséquent tout ou plus, des spécifications exploratoires.

La nécessité de mieux suivre les variations de la consommation a conduit à diverses tentatives de reformulation :

- En reconstruisant la variable revenu permanent directement sur la base de revenus passés et anticipés. C'est par exemple ce qui est fait dans le modèle ICARE de l'IPECODE. Le revenu total est scindé en 5 composantes dont la fraction permanente est estimée à partir des valeurs des 4 trimestres précédents et des trimestres courant ⁽¹⁾.

- En spécifiant une fonction dont la variable expliquée est la propension permanente à consommer. C'est ce qui a été fait dans le modèle PITI de la direction de la prévision ⁽²⁾. On peut dire aujourd'hui que cette solution a l'avantage d'utiliser des variables stationnaires.

Mais la tentative de reformulation à la fois la plus ambitieuse et la plus intéressante est celle qui a été menée principalement en Angleterre à partir de la fin des années 70, par Hendry et Davidson. En travaillant sur la consommation, ils ont construit un nouveau type de spécification, déduit d'une catégorie d'équations désignées dans la littérature sous le nom de modèle à correction d'erreur (MCE) qui présentent des propriétés économétriques particulièrement intéressantes. L'analyse de la consommation a été l'occasion d'un intérêt renouvelé pour le type de modèle.

La démarche de Hendry et Davidson se présente comme une généralisation de la démarche de Friedman avec la théorie du revenu permanent. Comme ce dernier, ils se proposent de rendre compte simultanément de la stabilité de la propension à consommer à long terme et de son instabilité dans le court terme.

(1) "ICARE modèle conjoncturel de l'économie française" Revue de l'IPECODE n°1, Mars 1983 et "ICARE 1980 le modèle ICARE en nouvelle base de comptabilité nationale" Revue de l'IPECODE n° 22, Déc. 1988.

(2) "PITI : éléments pour un modèle à prix implicites trimestriel intégré". Statistiques et études financières n°41, 1980.

Si la propension à long terme est stable, le taux de croissance de la consommation est en longue période égal à celui du revenu :

$$(5) \quad \frac{\Delta C_t}{C_{t-1}} = \frac{\Delta Y_t}{Y_{t-1}} = g$$

Cette condition est par exemple remplie par

$$C_t = K(g)Y_t \text{ soit } \log C_t = \log Y_t + \log K(g)$$

en posant : $c_t = \log C_t$ et $y_t = \log Y_t$

Il vient

$$(6) \quad c_t = y_t + k$$

En passant aux différences

$$(7) \quad c_t - c_{t-1} = y_t - y_{t-1} = g \text{ puisque } c_t - c_{t-1} = \log \frac{C_t}{C_{t-1}} = \log \left(1 + \frac{\Delta C_t}{C_{t-1}}\right)$$

dont l'approximation est $\frac{\Delta C_t}{C_{t-1}} = g$

Le modèle (6) se généralise par l'introduction d'un polynôme de retard sur c_t et d'un autre sur y_t :

$$(8) \quad \alpha(L)c_t = \beta(L)y_t + k$$

Dans la pratique on retient des polynômes du premier degré dont une forme particulière est :

$$(9) \quad c_t = k + \alpha_1 c_{t-1} + \beta_0 y_t + \beta_1 y_{t-1}$$

Cette relation étant vérifiée aussi en $t-1$, il vient :

$$(10) \quad c_t - c_{t-1} = \alpha_1 (c_{t-1} - c_{t-2}) + \beta_0 (y_t - y_{t-1}) + \beta_1 (y_{t-1} - y_{t-2})$$

Dans le cadre d'un régime de croissance équilibrée, le taux de croissance de la consommation est constant et il est égal au taux de croissance du revenu ⁽¹⁾.

(1) On parvient au même résultat en posant que l'équilibre statique de long terme impose que les taux de croissance de C et de Y soient nuls et que les valeurs de Y et C soient constantes, à partir de :

d'où :

$$c_t - c_{t-1} = c_{t-1} - c_{t-2} = y_t - y_{t-1} = y_{t-1} - y_{t-2} = g$$

$$\text{soit } g = \alpha_1 g + \beta_0 g + \beta_1 g$$

$$(11) \quad 1 = \alpha_1 + \beta_0 + \beta_1$$

Ce que l'on peut encore écrire sous la forme :

$$(12) \quad 1 - \alpha_1 = \beta_0 + \beta_1 = \gamma$$

et donc

$$(13) \quad \alpha_1 = 1 - \gamma$$

$$(14) \quad \beta_1 = \gamma - \beta_0$$

Les valeurs obtenues dans (13) et (14) peuvent être reportées dans (9)

$$(15) \quad c_t = k + (1 - \gamma)c_{t-1} + \beta_0 y_t + (\gamma - \beta_0)y_{t-1}$$

$$(16) \quad c_t - c_{t-1} = k + \beta_0(y_t - y_{t-1}) + \gamma(y_{t-1} - c_{t-1})$$

$$(16') \quad c_t - c_{t-1} = k + \beta_0(y_t - y_{t-1}) - \gamma(c_{t-1} - y_{t-1})$$

Cette équation constitue la forme la plus simple d'un modèle à correction d'erreur. Le taux de croissance de la consommation dépend du taux de croissance du revenu et d'un terme de rappel qui est mesuré par l'écart entre $\log c_t$ et $\log y_t$ à la période précédente.

Ce terme peut encore s'interpréter comme $\log \frac{C_{t-1}}{Y_{t-1}}$ et se compare donc à la

valeur de long terme du rapport $\frac{C_t}{Y_t}$. Si la consommation a augmenté moins

vite que le revenu à la période précédente, le terme de rappel accélère la consommation à la période présente. Dans le cas contraire, il freine la consommation.

Un modèle à correction d'erreur peut être complété par diverses variables explicatives.

(9') $c_t - c_{t-1} = K + (\alpha_1 - 1)c_{t-1} + \beta_0(y_t - y_{t-1}) + (\beta_0 + \beta_1)y_{t-1}$ cf Cuthbertson, Hall et Taylor "Applied Économétric Techniques" Philip Allan 1992, p 103.

Un exemple de fonction de consommation de type MCE peut être trouvé dans le modèle DMS 4 ⁽¹⁾ dans lequel ont été ajoutés le taux d'inflation et le taux de chômage anticipé. La formulation est un peu différente de l'équation (16) dans la mesure où elle est résolue par rapport à l'épargne et non par rapport à la consommation.

On passe naturellement de l'une à l'autre par le fait que $E_t = Y_t - C_t$. Nous reviendrons sur les raisons qui peuvent conduire à estimer plutôt une fonction d'épargne dans les modèles multisectoriels, à la fin de ce chapitre.

Le modèle à correction d'erreur présente une propriété remarquable qui a été démontrée par Granger en 1983 ⁽²⁾. Un ensemble de variables co intégrées d'ordre (1,1) peut être mis sous la forme d'un modèle à correction d'erreur dont toutes les variables sont stationnaires et dont les coefficients peuvent être estimés par les méthodes de l'économétrie classique sans risque de corrélations fortuites. Le résultat, connu sous le nom de théorème de représentation de Granger, valide de façon générale la démarche du MCE pour une classe importante de variables.

Les modèles à correction d'erreur peuvent être utilisés pour de nombreuses variables économiques. Il est possible de s'affranchir de l'égalité du taux de croissance de la variable explicative et de la variable expliquée et il est également possible d'appliquer le modèle à plusieurs variables explicatives figurant dans une même équation.

Supposons que la variable X_t , dépendant de Z_t , ait un taux de croissance à long terme égal à ϑ fois celui de Z_t :

$$z_t - z_{t-1} = g \text{ et } x_t - x_{t-1} = \vartheta g$$

L'équation du MCE est :

(1) DMS 4 modèle dynamique multisectoriel - Collections de l'INSEE série C n° 139 - 1987 Chap. IV - p 146 - 150.

(2) Voir aussi Engle et Granger "Co intégration and Error Correction Representation, Estimation and Testing" *Econometrica* 1987.

$$(17) \quad x_t = k + \alpha_1 x_{t-1} + \beta_0 z_t + \beta_1 z_{t-1}$$

La condition de long terme (11) s'écrit désormais

$$(18) \quad \vartheta = \vartheta \alpha_1 + \beta_0 + \beta_1$$

Soit

$$(18) \quad 1 - \alpha_1 = \frac{\beta_0 + \beta_1}{\vartheta} = \gamma$$

$$\text{d'où } \alpha_1 = 1 - \gamma \text{ et } \beta_1 = \gamma \vartheta - \beta_0$$

en remplaçant dans (17) :

$$(19) \quad x_t = k + (1 - \gamma)x_{t-1} + \beta_0 z_t + (\gamma \vartheta - \beta_0)z_{t-1}$$

$$(20) \quad x_t - x_{t-1} = k + \beta_0(z_t - z_{t-1}) - \gamma(x_{t-1} - \vartheta z_{t-1})$$

Cette équation peut être estimée par les MCO sans restrictions sous la forme :

$$(20 \text{ bis}) \quad \Delta x_t = a_0 + a_1 \Delta z_t + a_2 x_{t-1} + a_3 z_{t-1} + u_t$$

Les valeurs des coefficients de (20) sont données par :

$$k = a_0 \quad \beta_0 = a_1 \quad \gamma = -a_2 \quad \vartheta = \frac{-a_3}{a_2}$$

L'équation (20 bis) conduit à régresser sur des variables dont l'ordre d'intégration est différent : Δz_t est $I_{(0)}$ et y_{t-1} comme z_{t-1} sont $I_{(1)}$

Engle et Granger proposent une démarche en 2 étapes permettant d'estimer directement l'équation (20) sans avoir à passer par (20 bis). Dans une première étape, la relation de cointégration, donc de long terme, entre x_t et z_t est estimée.

Elle fournit la valeur du coefficient ϑ , soit $\hat{\vartheta}$, et établit en même temps que $x_t - \hat{\vartheta} z_t$ comme $x_{t-1} - \hat{\vartheta} z_{t-1}$ sont stationnaires par définition de la cointégration.

L'équation (20) peut alors être estimée en utilisant la valeur de $\hat{\vartheta}$:

$\Delta x_t = k + \beta_0 \Delta z_t - \gamma(x_{t-1} - \hat{\vartheta} z_{t-1}) + u_t$ permettant d'obtenir des valeurs de k , β_0 et γ .

Si maintenant Y_t dépend de Z_t et de W_t avec des réponses à long terme, telles que le taux de croissance de X soit égal à ϑ fois celui de Z et ε fois celui de W , le modèle s'écrit :

$$(21) \quad x_t = k + \alpha_1 x_{t-1} + \beta_0 z_t + \beta_1 z_{t-1} + \zeta_0 w_t + \zeta_1 w_{t-1}$$

et l'équation (18) devient :

$$(22) \quad 1 - \alpha_1 = \frac{\beta_0 + \beta_1}{\vartheta} = \frac{\zeta_0 + \zeta_1}{\varepsilon} = \gamma$$

d'où :

$$\alpha_1 = 1 - \gamma$$

$$\beta_1 = \vartheta\gamma - \beta_0$$

$$\zeta_1 = \varepsilon\gamma - \zeta_0$$

Soit en remplaçant dans (21) :

$$(23) \quad x_t = k + (1 - \gamma)x_{t-1} + \beta_0 z_t + (\vartheta\gamma - \beta_0)z_{t-1} + \zeta_0 w_t + (\varepsilon\gamma - \zeta_0)w_{t-1}$$

$$(24) \quad x_t - x_{t-1} = k + \beta_0(z_t - z_{t-1}) + \zeta_0(w_t - w_{t-1}) - \gamma(x_{t-1} - \vartheta z_{t-1} - \varepsilon w_{t-1})$$

Équation qui peut être estimée par les MCO sans contrainte sous la forme

$$(25) \quad \Delta x_t = a_0 + a_1 \Delta z_t + a_2 \Delta w_t + a_3 x_{t-1} + a_4 z_{t-1} + a_5 w_{t-1} + u_t$$

Les coefficients de (23) sont :

$$k = a_0 \quad \beta_0 = a_1 \quad \zeta_0 = a_2 \quad \gamma = -a_3 \quad \vartheta = -\frac{a_4}{a_3} \quad \varepsilon = -\frac{a_5}{a_3}$$

Les réponses à long terme de X , les élasticités à long terme par rapport à Z et W sont donc données par $-\frac{a_4}{a_3}$ et $-\frac{a_5}{a_3}$ respectivement, tandis que les élasticités de court terme sont respectivement a_1 et a_2 .

De tels modèles sont couramment utilisés pour estimer des fonctions d'investissement, d'importations ou d'exportations ⁽¹⁾ et sont intégrés dans

(1) Parmi de nombreux travaux cf : J.P. Urbain "modèles à correction d'erreur et fonctions de demande d'importations agrégées". Économie et Prévision 1990 ; S. Capet et Ph. Gudin

certains modèles macroéconomiques comme METRIC. La version METRICX ⁽²⁾ réestimée en 1988 fait une large place aux fonctions de type MCE. Certains auteurs utilisent aussi l'expression "modèle à correction d'équilibre" qui peut sembler plus approprié que "correction d'erreur" et met bien en évidence l'idée d'évolution de long terme ou de situation d'équilibre de longue période par rapport auxquelles les écarts déclencherait des réactions de correction ⁽³⁾.

SECTION 2 : L'APPROCHE SEMI-GLOBALE

A l'intérieur de la variable consommation, il est parfois utile de distinguer les biens durables et les biens non durables.

Les biens non durables tels que l'alimentation ou les services peuvent être correctement représentés par un modèle du type de ceux qui ont été présentés dans la section précédente. Ils ne posent pas de problèmes particuliers.

Les biens durables, en raison de leur durée de vie précisément et accessoirement de leur valeur unitaire, présentent des caractéristiques particulières.

Comme le notait M.K. Evans dès 1969 ⁽⁴⁾, il est possible de déplacer dans le temps la décision d'achat en la reportant ou au contraire en l'anticipant, ce qui confère à la demande une plus grande variabilité et d'autre part le stock existant de ces mêmes biens exerce un effet de frein sur l'accroissement du parc. De plus, compte tenu de leur mode d'acquisition, qui fait souvent appel à la constitution d'une épargne préalable ou au crédit, la demande pour ce type de biens est beaucoup plus sensible aux variables monétaires et financières que la demande de biens de consommation en général. En somme les biens durables se situent

de Vallerin "Fonctions d'importations et d'exportations : l'apport de la théorie économétrique récente" *Économie et Prévision* 1993 ; J. Glachant et J.F. Nivet "Deux études macroéconomiques de l'investissement". *Économie et Prévision* 1989.

(2) "Le modèle METRICX et l'étude de la variante dévaluation", *Économie et Prévision*, 1988.

(3) C'est l'expression utilisée par Hendry (Equilibrium Correction Model) - cf D. F. Hendry "Dynamic Econometrics" Oxford University Press, 1995.

(4) M.K. Evans "Macro economic Activity", Harper et Row, 1969.

dans une position intermédiaire entre les investissements et la consommation, comme le traduit l'expression "biens d'équipement des ménages" utilisée par l'INSEE.

La modélisation de la consommation de biens durables présente certaines similitudes avec celle de l'investissement, mais s'en distingue par certains aspects tels que la structure des retards de décision et les délais de réalisation.

Les modèles de consommation de biens durables sont à la fois nombreux et divers. C'est depuis longtemps un des domaines principaux d'application de l'économétrie. Faute de pouvoir rendre compte de cette diversité, on se limitera à une présentation des principes utilisés pour la construction de ces fonctions.

Comme pour la FBCF, on peut distinguer un stock désiré et une fonction de réalisation qui décrit l'ajustement à court terme du stock effectif au stock désiré.

Le stock désiré, qui peut aussi être un stock d'équilibre dépend uniquement du revenu des ménages et du prix relatif de ces biens par rapport à l'ensemble de la consommation ou par rapport aux services qu'ils remplacent.

$$(1) S^* = f(y, Pr)$$

Cette fonction peut comporter un mécanisme de retard décrivant l'ajustement du stock désiré aux variations du revenu. L'utilisation du revenu permanent ou du revenu anticipé à moyen terme permet d'éviter ce détour dans certains cas. Un modèle à correction d'équilibre peut également être utilisé pour déterminer le stock désiré. La fonction de réalisation décrit l'ajustement du stock effectif au stock désiré à partir d'une distribution de retards sur les variables qui déterminent la décision d'achat des ménages, c'est-à-dire par exemple la variation de leur revenu, les conditions de crédit ou encore l'évolution anticipée des prix des biens durables.

$$(2) S_t - S_{t-1} \text{ ou } \frac{S_t}{S_{t-1}} = g_1 \left\{ S^*_t - S_{t-1} \text{ ou } \frac{S^*_t}{S_{t-1}} \right\} (+ \text{ ou } \times) g_2(b(L)\Delta y_t, i_t, \rho_t^a)$$

La formulation en $\frac{S_t}{S_{t-1}}$ présente l'avantage de faire apparaître le taux de

croissance du parc, lorsqu'on passe en logarithme. La fonction g_1 peut être un simple mécanisme d'ajustement partiel du type $\lambda(S_t^* - S_{t-1})$ ou $\left[\frac{S_t^*}{S_{t-1}} \right]^\omega$.

La fonction g_2 est additive ou multiplicative selon la spécification retenue.

La combinaison des équations (1) et (2) permet d'exprimer la fonction décrivant l'accroissement du stock de biens durables ou son taux d'accroissement.

La demande de biens durables est égale à l'accroissement du parc auquel s'ajoute la demande de renouvellement. Cette dernière peut être obtenue en appliquant un taux de dépréciation ϑ au stock existant :

$$D_{rt} = \vartheta S_{t-1} \text{ où } \vartheta \equiv \frac{1}{T} \text{ T étant la durée de vie moyenne des biens. Mais cette}$$

solution ne rend pas compte du rythme selon lequel le stock a été constitué. Ainsi un parc automobile peut être très récent et la demande de renouvellement ainsi calculée surestimer grossièrement la demande de renouvellement effective. De même cette méthode ne peut rendre compte des éventuels à-coups dans la constitution du stock, du type crises pétrolières pour un parc automobile.

Il est donc préférable, quand l'information est disponible, d'appliquer une loi de déclassement aux différentes classes d'âge constituant le stock. Pour des biens tels que les automobiles qui font l'objet d'un enregistrement et d'un suivi administratif, cela ne pose guère de problèmes. Pour les autres catégories de biens durables c'est généralement beaucoup plus compliqué puisqu'il faut commencer par construire le stock existant à partir des achats réalisés dans le passé. Rappelons que pour la comptabilité nationale, les biens durables sont réputés consommés dès leur entrée dans le patrimoine des ménages, ce qui justifie qu'il ne soit pas tenu de compte des stocks détenus par les ménages.

Parce qu'elles correspondent à des comportements différents, il est donc intéressant de pouvoir distinguer la consommation de biens durables et celles de biens non durables dans un modèle non sectorialisé. On peut ainsi améliorer la précision avec laquelle l'ensemble de la consommation est décrit.

Dans les grands modèles sectorialisés, la consommation est décrite par produits, comme on va le voir dans la section suivante.

SECTION 3 : L'APPROCHE PAR CATÉGORIES DE PRODUITS

Dans les études détaillées comme dans les grands modèles macroéconomiques, une projection globale de la consommation est insuffisante. D'une part il se produit des déformations continues de la structure de la consommation sous l'influence des élasticités revenus, déformations dont il faut pouvoir rendre compte. D'autre part, l'équilibre offre - demande doit être assuré par catégories de produits et non plus au niveau de l'ensemble des catégories. Dès lors que des catégories ont été identifiées, il n'est plus évident que les excédents d'offre d'une catégorie puissent combler les déficits d'une autre. Or les conséquences des situations d'excédent d'offre ou de pénuries ne sont pas symétriques. Une pénurie est comblée par recours à des importations en économie ouverte, tandis qu'un excédent conduit à une augmentation non désirée des stocks de produits finis, puis à un ajustement du niveau d'activité. On voit que l'équilibre macroéconomique, qui pourrait être réalisé au niveau de l'ensemble des catégories de produits, peut être remis en cause par des désajustements sectoriels.

L'analyse de la consommation par produits se réalise généralement en deux étapes.

On pourrait imaginer de déterminer simultanément toutes les fonctions de demande à partir des conditions de maximisation d'une fonction d'utilité, sous contrainte de revenu. La consommation globale serait alors obtenue par sommation des consommations par produits. Cette démarche est impossible pour plusieurs raisons. Tout d'abord les quantités des différents produits, qui ont été consommées, ne correspondent pas à des quantités demandées, mais à des quantités d'équilibre reflétant le jeu de l'offre et de la demande sur les marchés. Il serait donc incorrect de vouloir en rendre compte uniquement à partir des seules fonctions de demande qui sont celles qui se déduisent de la maximisation sous contrainte de la fonction d'utilité. Il faut rappeler que les fonctions de consommation par produit ne sont pas des fonctions de demande

(on peut se reporter à ce sujet au chapitre II section 2 du présent ouvrage). Une deuxième raison est qu'une telle procédure aurait pour effet d'additionner les erreurs faites sur les consommations par produits au niveau de la consommation globale puisque :

$$c_t = \sum_i c_{it}$$

donc si

$$c_{it} = h_i(y_t, p_{it}, \dots) + u_{it}$$

$$c_t = \sum_i h_i(y_t, p_{it}, \dots) + \sum_i u_{it}$$

Or il n'y a guère de raison à priori pour que les erreurs faites sur les consommations par produits se compensent.

De ce fait la consommation globale serait déterminée avec une assez grande marge d'erreur. Cette marge se retrouverait encore amplifiée au niveau du calcul de l'épargne que l'on obtient par différence avec le revenu.

En effet, compte tenu des proportions respectives de l'épargne et de la consommation dans le revenu, une erreur faible en pourcentage sur la consommation se traduit par une erreur forte en pourcentage sur l'épargne.

Supposons que la consommation représente 90 % du revenu et l'épargne 10 %. Une erreur de 1 % sur la consommation entraîne une erreur de 9 % sur l'épargne.

C'est pourquoi on commence par déterminer l'épargne à l'aide d'une fonction de comportement. Même si cette fonction est de qualité inférieure en terme d'erreur quadratique moyenne rapportée à la variable expliquée ⁽¹⁾, elle reste de meilleure qualité qu'un résultat déterminé par solde de la consommation sur le revenu.

(1) $\sqrt{\frac{1}{T} \sum \left(\frac{y_t - \hat{y}_t}{y_t} \right)^2}$

Ainsi, par exemple, la fonction d'«épargne ajustée» construite dans le modèle DMS 4 a une précision de 3,2 % ⁽¹⁾. Cette fonction est déduite d'un MCE pour la fonction de consommation. Pour obtenir le même degré de précision, il faudrait une fonction de consommation ayant une erreur quadratique moyenne inférieure à 0,5 %.

Une fois estimée l'épargne, on obtient la consommation globale par différence. Il faut ensuite répartir celle-ci entre les différentes catégories de produits. La répartition se fait par un système d'équations inspiré de la microéconomie. Le plus connu de ces systèmes est celui de Houthakker et Taylor. Il conduit à des spécifications dans lesquelles la consommation d'un produit à la période t dépend de la consommation globale, du prix relatif du produit et de la consommation passée du produit :

$$(1) \quad c_{it} = a_0 + a_1 c_t + a_2 p_{rit} + a_3 c_{it-1} + u_t$$

La consommation décalée peut intervenir avec le signe positif, auquel cas elle traduit un effet d'habitude, ou avec le signe négatif traduisant un effet de stock. Un coefficient a_3 positif sera associé aux fonctions de consommation de biens non durables et un coefficient négatif aux fonctions de consommation de biens durables.

Ce type de fonction est compatible avec une spécification linéaire comme dans l'équation (1) ou log linéaire comme (2) ou semi log comme (3) et (4).

$$(2) \quad \log c_{it} = a_0 + a_1 \log c_t + a_2 \log p_{rit} + a_3 \log c_{it-1} + u_t$$

$$(3) \quad c_{it} = a_0 + a_1 \log c_t + a_2 \log p_{rit} + a_3 \log c_{it-1} + u_t$$

$$(4) \quad \log c_{it} = a_0 + a_1 c_t + a_2 p_{rit} + a_3 c_{it-1} + u_t$$

Une catégorie de produit dont le coefficient budgétaire est à peu près stable pourra être représentée par un modèle log linéaire. Par contre la consommation d'une catégorie de produits dont la part régresse dans la consommation globale

(1) DMS 4 modèle dynamique multisectoriel, Collections de l'INSEE série C n° 139 1987, p 148 et suivantes.

comme la consommation alimentaire par exemple à des chances d'être mieux décrite par une fonction de type (3).

Exemple : analyse de la consommation de carburant en France de 1983 I à 1993 II. On commence par vérifier que la consommation de carburant (cc_t) la consommation finale (c_t) et le prix relatif des carburants (Prc_t) exprimés en logarithmes, sont cointégrés.

Test de Dickey - Fuller :

Variable	Lcc_t	Lc_t	$LPrc_t$	ΔLcc_t	ΔLc_t	$\Delta Prct$
Valeurs du test	0,919	14,7	-0,0008	-7,246	-2,059	-5,526
Valeur critique à 5 %	-1,949	-1,949	-1,949	-1,949	-1,949	-1,949
Valeur critique à 1 %	-2,619	-2,619	-2,619	-2,621	-2,621	-2,621

D'où il ressort que les variables sont toutes intégrées d'ordre 1.

Test de Engle et Granger sur Lcc_t , Lc_t et $LPrc_t$

Valeur du test : - 5,80

Valeur critique à 5 % : - 3,954

Valeur critique à 1 % : - 4,662

résultat qui établit que les variables sont bien cointégrées.

Après diverses tentatives on constate que la consommation de carburant décalée n'est jamais significative. D'autre part, la comparaison graphique entre valeurs calculées et valeurs observées montre un léger déplacement après le 2^e trimestre 1987. L'introduction d'une variable muette (D) prenant la valeur 0 jusqu'au 2^e trimestre 1987 et la valeur 1 après se révèle significative. Un test de restriction a priori confirme la pertinence de cette variable.

$$(5) \quad Lcc_t = a_0 + a_1 Lc_t + a_2 LPrc_t + u_t$$

$$(6) \quad Lcc_t = a_0 + a_1 Lc_t + a_2 LPrc_t + a_3 D + u_t$$

$$\sum \hat{w}_t^2 \text{ de l'équation (5)} = 0,0171$$

$$\sum \hat{u}_t^2 \text{ de l'équation (6)} = 0,0143$$

$$\frac{\frac{\sum \hat{w}_t^2 - \sum \hat{u}_t^2}{g}}{\frac{\sum \hat{u}_t^2}{T-l}} = \frac{0,0171 - 0,0143}{\frac{1}{0,0143}} = 6,65$$

La valeur critique de la table de Fisher est :

$$F_{\alpha}(1, 34) = 4,17 \text{ par excès}$$

L'ajustement définitif est :

$$(7) \quad Lcc_t = 1,06 Lc_t + 0,002 D - 0,017 LPr c_t - 4,61 \quad R^2 = 0,968$$

(2,80) (2,53) (3,50) (0,85)

estimation obtenue sous AR(4) pour tenir compte d'un fort mouvement saisonnier.

Incidentement, ce résultat montre que contrairement à ce qui est souvent avancé, l'élasticité de la consommation de carburant est supérieure à l'unité et que la consommation a même eu tendance à augmenter récemment. Une décomposition plus fine montre que cette croissance est essentiellement due au carburant diesel.

Dans un exercice de modélisation de la consommation par produits, rien n'assure que la somme des consommations par produits soit égale à la consommation globale calculée séparément.

Dans le cas général :

$$\sum_i c_{it} \neq y_t - E_t$$

Il y a deux solutions qui permettent d'assurer la compatibilité entre ces deux variables.

La première consiste à laisser libre une catégorie de produits et à la déterminer par solde.

$$c_{jt} = y_t - E_t - \sum_{i \neq j} c_{it}$$

Compte tenu des phénomènes d'accumulation d'erreurs qui ont été évoqués précédemment, il est souhaitable que la catégorie déterminée par solde représente une fraction assez importante de la consommation totale.

C'est pourquoi dans certains exercices de modélisation de la consommation par produits, on solde sur le poste services qui représente près du tiers de la consommation totale.

L'autre solution consiste à répartir l'écart constaté entre les catégories de produits en proportion de leurs valeurs.

$$\text{Si } \frac{y_t - E_t}{\sum_i c_{it}} = m$$

Chaque poste est réévalué par le coefficient m .

Chapitre VI : Application aux échanges extérieurs

Les échanges extérieurs constituent un autre domaine d'application intéressant des instruments de l'économétrie. On sait que les relations économiques avec l'extérieur sont constituées de différentes catégories d'opérations : mouvements de biens et services, mouvements de capitaux, opérations de répartition etc. On s'intéressera exclusivement aux opérations commerciales qui sont les plus simples à modéliser. Certaines opérations peuvent leur être rattachées comme par exemple les crédits à l'exportation ou les investissements directs qui ont tendance à dépendre des flux de marchandises.

La modélisation des échanges commerciaux présente l'avantage de pouvoir se faire avec des spécifications relativement simples, elle aussi, et elle permet d'aborder quelques problèmes intéressants.

On commencera par les importations, puis dans une deuxième section, on étudiera le cas des exportations.

SECTION 1 : LES IMPORTATIONS

Elles dépendent à titre principal de la demande intérieure qui peut être représentée par l'activité économique ou le PIB et de la compétitivité des importations par rapport aux productions nationales pour les produits qui sont substituables à la production intérieure. Pour de nombreux pays engagés dans une démarche d'intégration régionale ou d'ouverture extérieure, il existe une poussée autonome des importations qui traduit cette intégration et l'évolution des spécialisations qui en résulte ⁽¹⁾.

(1) Sur l'évolution des spécialisations internationales cf. H. F. Henner "La spécialisation internationale de l'économie française" Revue Économique 1976 et K. Abdel Rahman "Réexamen de la définition de la mesure des échanges croisés de produits similaires entre les nations", Revue Économique, 1986.

Ces influences peuvent être traduites dans la formule :

$$Im = f(Y, Pr, t)$$

Connue sous le nom de "Formule de Luxembourg" (1).

Dans la mesure où des tensions sur les capacités de production peuvent entraîner des retards de livraison de la part des entreprises intérieures ou conduire à une certaine saturation des capacités de réponse de l'offre intérieure, on peut compléter cette formule par un indicateur des taux d'utilisation de capacités. Lorsque les capacités sont utilisées plus intensément, les importations augmentent.

La réponse des importations aux variations d'activité ou de compétitivité est généralement rapide. En données annuelles, elle se produit entièrement au cours de la période. En données trimestrielles, les retards ne dépassent guère 2 trimestres et il n'est pas nécessaire de construire de structure de retard dans la plupart des cas.

La fonction d'importation est compatible avec une spécification linéaire, ou avec une spécification log linéaire à élasticité constante :

$$Im_t = a_1 Y_t + a_2 Pr_t + a_3 Tu_c + a_4 + u_t$$

a_1 représente donc le contenu en importations d'un accroissement de revenu ou d'activité.

On peut aussi distinguer les différents postes de la demande finale.

Ce qui conduit à estimer des contenus en importations différents pour ces catégories :

$$Im_t = a_1 C_t + a_2 G_t + a_3 X_t + a_4 I_t + a_5 Pr_t + a_6 Tu_c + a_7 + u_t$$

Toutefois, il est quasiment impossible dans cette forme d'éviter les problèmes de colinéarité généralisée.

Avec des élasticités constantes le modèle s'écrit :

(1) Cf B. Bobe et P.H. Derycke "Projections des échanges extérieurs et balances des paiements", *Economica* 1975.

$$(1) \log \text{Im}_t = \alpha_1 \log Y_t + \alpha_2 \log \text{Pr}_t + \alpha_3 \log Tu_c + \alpha_4 + u_t$$

α_1 est l'élasticité revenu des importations

α_2 est l'élasticité prix des importations.

De nombreuses études économétriques ont été réalisées dans le passé à l'aide de la relation (1) ou de formes voisines.

Exemple : importations italiennes en données trimestrielles du chapitre I section 3.

Avec une spécification à élasticité constante, on obtient pour les importations italiennes de produits manufacturés sur la période 1978 I - 1992 II :

$$\log \text{Im}_t = 1,916 \log Y_t - 0,147 \log Pi_t + 0,371 D90 - 11,89$$

(22,34) (4,91) (12,81) (10,14)

$$R^2 = 0,976 \quad DW = 2,02 \quad \hat{\rho} = 0,417$$

Im_t = importations italiennes de produits manufacturés, déflatées

Y_t = PIB déflaté

Pi_t = prix des importations de produits manufacturés

D90 est une variable muette qui vaut 1 pour le premier trimestre de 1990 et permet d'éliminer un écart important inexpliqué pour cette période.

On vérifie que $\log \text{Im}_t$ et $\log Y_t$ sont cointégrés :

Vecteur cointégrant :

Log Im_t : 1

Log Y_t : -1,573

Test de Engle et Granger : - 5,016

Valeur critique à 5 % : - 3,44

Dans la plupart des pays, les élasticités revenu sont bien estimées et varient entre un peu plus de 1 et 1,8 selon les pays et les périodes. Le Japon est un des rares pays de l'OCDE à avoir eu pendant longtemps une élasticité inférieure à 1. L'estimation de l'élasticité prix n'est pas aussi satisfaisante. Il est souvent difficile d'obtenir des coefficients significatifs... et du bon signe. La raison de

ces difficultés a été exposée dans le chapitre 4. En effet l'équation (1) contient des variables qui seront normalement I (1), Im_t et Y_t et une variable qui ne peut pas être I (1) à savoir Pr_t . La logique économique veut que les autorités monétaires veillent à maintenir le niveau de compétitivité des produits nationaux en ajustant le taux de change lorsque les prix intérieurs ont évolué durablement de manière différente des prix des principaux partenaires. Par conséquent Pr_t doit être normalement stationnaire.

La relation (1) contient des variables d'ordre d'intégration différents et ne peut constituer une relation de cointégration, c'est-à-dire une relation de long terme.

Par contre

(2) $\log Im_t = \alpha_1 \log Y_t + \alpha_2$ peut être une telle relation si on suppose que les autres variables n'ont pas d'influence en longue période.

Dans la mesure où $\alpha_1 > 1$ la croissance des importations déplace des productions intérieures. En effet la production intérieure est calculée par solde dans l'équation d'équilibre entre ressources et emploi :

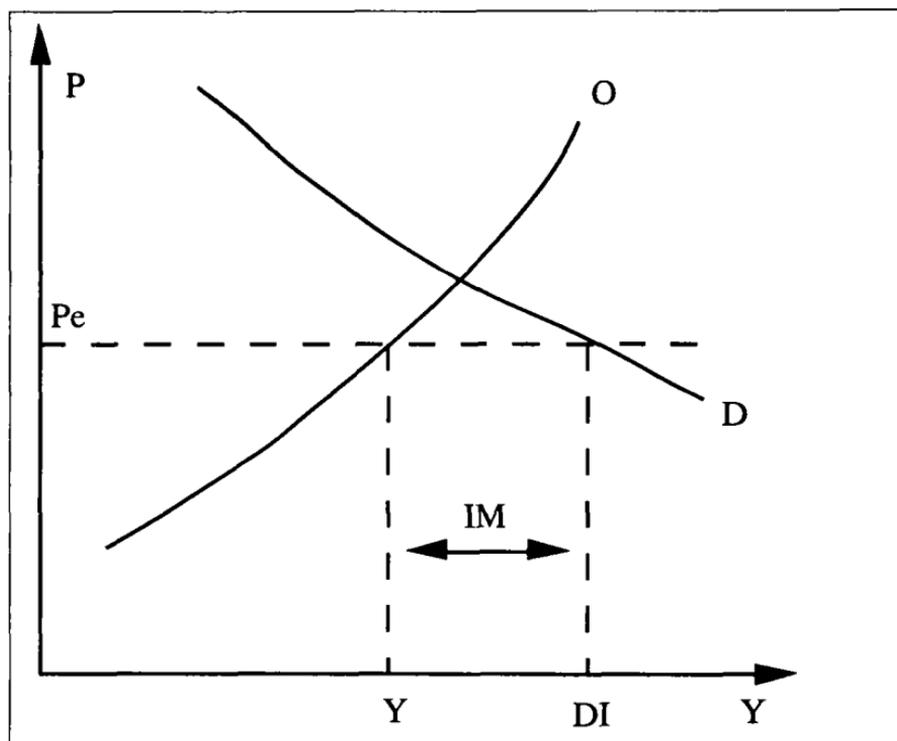
$$(3) Im_t + Y_t = C_t + I_t + G_t + X_t$$

C_t , I_t , X_t et Im_t sont déterminés par des relations de comportement et G_t est exogène.

On peut cependant juger que ce mécanisme est insuffisamment contraignant, dans la mesure où, en particulier, la croissance des importations prend appui sur la croissance du revenu ou de l'activité.

On est alors amené à remplacer la fonction d'importation par une fonction de partage du marché intérieur. Un des premiers modèles à contenir ce type de spécification est le modèle FiFi, construit en France à la fin des années 60 et utilisé pour la préparation du 6^e Plan. Ce modèle inspiré de la théorie des économies concurrentées, elle même inspirée des travaux menés en Norvège par Aukrust, présente une version extrémiste du mécanisme de partage.

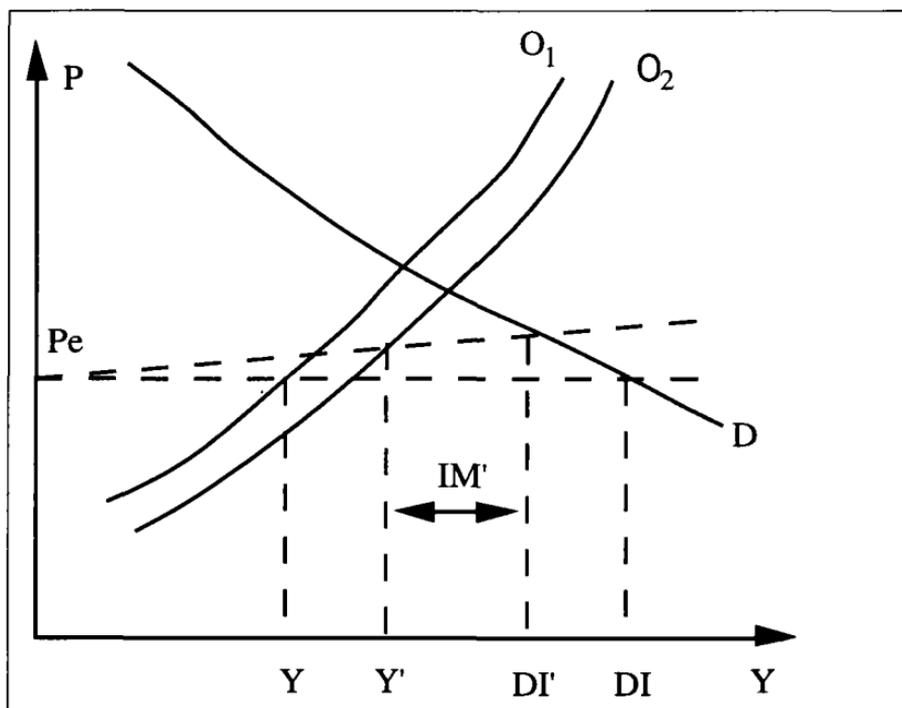
Dans le court terme les prix sont fixés par le marché mondial. La France est un pays non dominant, dans une situation de price taker.



Au niveau de prix P_e , les producteurs nationaux offrent une quantité Y . La différence entre la demande intérieure DI et Y , est comblée par les importations (1).

A moyen terme, les importateurs relèvent leurs marges lorsque leur part de marché augmente, ce qui fait que le niveau de prix n'est plus horizontal et les entreprises se modernisent et réduisent leur coûts, ce qui déplace vers le bas la fonction d'offre de O_1 en O_2 . Les importations s'établissent à $DI' - Y'$.

(1) M. Aglietta, R. Courbis, C. Seibel "Le modèle FiFi" Collection de l'INSEE C22, 1973.



Cette approche consacre la perte quasi totale du contrôle de l'activité économique par les moyens traditionnels de la politique économique qui agissent sur la demande globale. En effet un déplacement vers la droite de la courbe de demande se répercute presque intégralement sur les importations et laisse inchangée la production intérieure et l'emploi. Face aux traditionnels multiplicateurs de demande, il existe donc des multiplicateurs d'offre qui mesurent l'effet du desserrement de la contrainte de compétitivité sur la production et sur l'emploi.

On peut regretter que cette analyse n'ait pas reçu les prolongements qu'elle mérite ⁽¹⁾. Quoiqu'il en soit de nombreux autres travaux ont été faits avec des mécanismes de partage du marché dont la forme générale est :

$$(4) \quad \frac{Im_t}{Y_t} = a(L)Pr_t + f(t) + dTuc_t + u_t$$

$$(5) \quad \log \frac{Im_t}{Y_t} = \alpha(L) \log Pr_t + f(t) + dTuc_t + u_t$$

$a(L)$ et $\alpha(L)$ étant des polynômes de retard et t désignant le temps.

Le partage du marché dépend de la compétitivité prix des importations avec une distribution de retards de l'ordre de 6 trimestres, d'un terme tendanciel et d'un indicateur conjoncturel qui peut être représenté par le taux d'utilisation des capacités.

On remarquera que ces équations ne satisfont pas les conditions posées au chapitre 4.

Dans certains pays, il existe des restrictions aux importations sous la forme de licences ou d'allocations de devises ou de limitations quantitatives. Les importations réalisées sont alors inférieures à la demande d'importation. Il faut dans ce cas distinguer les importations contraintes et les importations potentielles.

Cette distinction est fondamentale lorsque l'on envisage d'assouplir le régime des importations. Se référer aux valeurs des élasticités revenu et prix de la fonction d'importations réalisées conduit à sous-estimer gravement les effets d'une éventuelle libéralisation. Si on connaît la fonction d'utilité des autorités, on peut en déduire le mécanisme de rationnement qu'elles appliquent entre les importations demandées et les importations réalisées. En sens inverse, en

(1) Bien qu'elle ait incontestablement un caractère dérangeant par rapport au mouvement de mondialisation de l'économie. Mais peut-il y avoir un équivalent du "politically correct" en économie ?

connaissant le mécanisme de rationnement et les importations réalisées, il est possible, dans certains cas, de remonter à la fonction de demande d'importations.

Différents modèles théoriques permettent de faire cette opération qui n'est pas sans rappeler l'identification présentée au chapitre 2. Nous présentons ci-dessous un modèle dû à Goldstein et Khan ⁽¹⁾.

La fonction de demande d'importations émane du secteur privé et s'écrit :

$$(6) \quad m_t^d = \alpha_0 + \alpha_1 Y_t + \alpha_2 Pr_t$$

Cette fonction n'est pas entièrement satisfaite faute de moyens de paiement suffisants. Les autorités offrent un montant de devises qui dépend de leurs propres objectifs. Ces objectifs sont à la fois des objectifs de court terme et des objectifs de long terme :

- à court terme, les autorités cherchent à satisfaire le maximum de demandes d'importations et à éviter de trop brusques variations du niveau d'importations.
- à long terme, elles cherchent à maintenir les importations en ligne avec la capacité de paiement, f_t .

Elles désirent également maintenir un certain niveau de réserves en devises, r_t^d

Les réserves en t sont données par :

$$(7) \quad r_t = f_t - m_t + r_{t-1}$$

f_t = flux de recettes provenant des exportations et des entrées de capitaux.

m_t = importations réalisées en t .

Les autorités cherchent à minimiser le coût de déviation par rapport à une fonction de coût quadratique.

$$\min C_t = B_1(m_t - m_t^d)^2 + B_2(m_t - m_{t-1})^2 + B_3(r_t - r_t^d)^2 + B_4(m_t - m_t^*)^2 \quad (8)$$

(1) M. Goldstein et M.S. Khan "The Supply and Demand for Exports : a Simultaneous Approach" The Review of Economics and Statistics, 1978.

où m_t^* représente les importations compatibles avec la capacité de paiement à long terme : $m_t^* = f_t$.

Le coefficient B_i représente le coût associé à un écart par rapport à l'objectif correspondant.

La valeur optimale des importations est celle qui minimise le coût total de déviation mesuré par la fonction C_t . Elle s'obtient en annulant la dérivée par rapport à m_t :

$$\frac{\partial C_t}{\partial m_t} = 0 \text{ sachant que } r_t = f_t - m_t + r_{t-1} \quad (2)$$

$$\frac{\partial C_t}{\partial m_t} = 2B_1 m_t - 2B_1 m_t^d + 2B_2 m_t - 2B_2 m_{t-1} + 2B_3 m_t - 2B_3 f_t + 2B_3 r_t^d - 2B_3 r_{t-1} + 2B_4 m_t - 2B_4 m_t^* = 0 \quad (9)$$

Soit

$$(10) \quad (B_1 + B_2 + B_3 + B_4)m_t = B_1 m_t^d + B_2 m_{t-1} + B_3 f_t - B_3 r_t^d + B_3 r_{t-1} + B_4 m_t^* = 0$$

En normant les coefficients, il vient :

$$b_i = \frac{B_i}{\sum B_i} \text{ et } \sum b_i = 1$$

et en remplaçant dans (10) :

$$(11) \quad m_t = b_1 m_t^d + b_2 m_{t-1} + b_3 f_t + b_3 (r_t^d - r_{t-1}) + b_4 m_t^*$$

$$(12) \quad m_t = (b_3 + b_4) f_t + b_3 (r_t^d - r_{t-1}) + b_2 m_{t-1} + b_1 m_t^d$$

Soit en remplaçant m_t^d par sa valeur dans (6)

$$m_t = (b_3 + b_4) f_t + b_3 (r_t^d - r_{t-1}) + b_2 m_{t-1} + b_1 \alpha_1 Y_t + b_1 \alpha_2 Pr_t + b_1 \alpha_0$$

Ce qui conduit à estimer :

$$(13) \quad m_t = a_0 + a_1 Y_t + a_2 Pr_t + a_3 f_t + a_4 (r_t^d - r_{t-1}) + a_5 m_{t-1} + u_t$$

avec :

$$b_3 + b_4 = a_3$$

$$b_3 = a_4$$

$$b_2 = a_5$$

$$b_1 = 1 - a_3 - a_5$$

$$\alpha_1 = \frac{a_1}{b_1} = \frac{a_1}{1 - a_3 - a_5}$$

$$\alpha_2 = \frac{a_2}{b_1} = \frac{a_2}{1 - a_3 - a_5}$$

Si les variables figurant dans l'équation (13) sont les logarithmes des variables, les coefficients a_1 et a_2 représentent les élasticités apparentes des importations par rapport aux revenus et aux prix, tandis que α_1 et α_2 sont les élasticités réelles de la demande d'importation par rapport aux mêmes variables. Même si les autorités accordent un poids prépondérant au premier objectif et s'efforcent de satisfaire un maximum d'importations, ce résultat montre la possibilité que les "vraies" élasticités soient substantiellement supérieures aux élasticités apparentes qui résulteraient d'une régression directe sur les volumes d'importations réalisés.

Comme indiqué dans le chapitre V, les fonctions d'importations peuvent également être estimées en utilisant une spécification à correction d'erreur.

La formulation la plus générale est :

$$(14) \Delta \log IM_t = k + \beta_0 \Delta \log Y_t + \zeta_0 \Delta \log Pr_t - \gamma (\log IM_{t-1} - \vartheta \log Y_{t-1} - \zeta \log Pr_{t-1})$$

qui peut être estimée par la méthode de Engle et Granger en 2 étapes ⁽¹⁾ sous cette forme ou plus simplement par les MCO à partir de l'équation suivante :

$$(15) \Delta \log IM_t = k + a_1 \Delta \log Y_t + a_2 \Delta \log Pr_t + a_3 \log IM_{t-1} + a_4 \log Y_{t-1} + a_5 \log Pr_{t-1} + u_t$$

avec les propriétés suivantes :

(1) On sait que dans ce cas, la relation de cointégration entre $\log IM_t$, $\log Y_t$ et $\log Pr_t$ n'est pas unique (cf. ch. 4).

Élasticités à court terme des importations :

– par rapport au revenu : $\beta_0 = a_1$

– par rapport aux prix : $\zeta_0 = a_2$

Élasticités à long terme des importations :

– par rapport au revenu : $\vartheta = \frac{-a_4}{a_3}$

– par rapport aux prix : $\varepsilon = \frac{-a_5}{a_3}$

SECTION 2 : LES EXPORTATIONS

Les exportations d'un pays sont par définition les importations du Reste du Monde. Toutefois le Reste du Monde n'est pas, dans la réalité, un ensemble homogène et facilement identifiable comme on le suppose dans les modèles simples à deux pays, mais un ensemble hétérogène de pays à la fois clients et concurrents.

De là il vient que si la spécification d'ensemble des fonctions d'exportations est similaire de celle des fonctions d'importations, la signification des variables qui y figurent est différente.

La variable d'échelle, correspondant au revenu ou à l'activité dans la fonction d'importation, est donnée ici par le revenu ou les importations des pays clients. Les importations présentent l'avantage d'être plus près de la demande adressée par les pays clients au Reste du Monde. Mais parmi les pays clients, tous ne pèsent pas du même poids et par exemple les conséquences d'une augmentation du Revenu ou du Produit Intérieur Brut de l'Allemagne sur les exportations de la France ne sont pas les mêmes que celles d'une augmentation identique du Revenu ou du Produit Intérieur Brut de la Malaisie. C'est pourquoi on construit l'indicateur de demande étrangère en pondérant les Revenus ou les PIB ou les Importations des pays clients exprimés en unité commune, par leur part dans les exportations du pays étudié.

Ainsi pour la France, l'INSEE construit deux Indicateurs :

PIBET = \sum PIB des pays clients de la France pondérés par leur part dans les exportations françaises.

IMPET = \sum Importations des pays clients pondérées par leur part dans les exportations françaises.

Ce dernier indicateur est voisin de ce qu'on désigne parfois en économie internationale sous le nom de "demande adressée".

La compétitivité - prix des exportations peut être mesurée par le rapport des prix à l'exportation sur les prix des produits concurrents, le tout exprimé en une unité monétaire commune. Les différents pays concurrents ne pèsent pas le même poids dans le commerce mondial et les conséquences d'une variation de la compétitivité prix des exportations françaises par rapport aux prix italiens ne sont pas les mêmes que la même variation par rapport aux prix argentins. Il faut donc pondérer les indices de prix des pays concurrents par leurs parts de marché respectives. On peut sur cette base construire un premier indice de compétitivité - prix. Mais les exportations d'un pays ne se réalisent pas sur un vaste marché homogène qui serait constitué par le Reste du Monde. Elles ont lieu sur différents marchés nationaux qui représentent chacun une certaine partie du total et sur ces différents marchés, la concurrence des autres pays est plus ou moins marquée.

Un bon indicateur de compétitivité-prix doit établir pour chaque pays client un indice du prix des exportations par rapport aux prix du pays client et à ceux des principaux concurrents pondérés par leur part du marché dans le pays client. Puis la pondération de ces indices élémentaires par la part des différents marchés dans les exportations totales permet d'obtenir l'indicateur de compétitivité-prix globale.

Supposons que l'on ait identifié m marchés clients désignés par $i = 1 \dots m$ dont la part dans les exportations est e_1 à e_m et qu'il existe n pays concurrents désignés par $\gamma = 1 \dots n$ dont les parts de marchés sont respectivement $e_i \gamma$.

$e_i \gamma$ = part du pays γ dans le marché i .

$$\sum_{\gamma} e_i \gamma = 1 \quad \text{et} \quad \sum_i e_i = 1$$

L'indicateur de compétitivité à l'exportation sur le marché i s'écrit :

$$comp_i = \frac{P}{\prod_{\gamma} (P_{i\gamma})^{\epsilon_{i\gamma}}}$$

P étant l'indice des prix à l'exportations en unité monétaire de référence.

$P_{i\gamma}$: l'indice de prix, en unité monétaire de référence, du pays γ sur le marché i .

L'indicateur global est alors :

$$Comp = \prod_i (comp_i)^{\epsilon_i}$$

Ces variables peuvent être incluses dans une formule de Luxembourg symétrique de celle posée pour les importations :

$$X_t = g(Det, Comp_t, t)$$

ou Det désigne la demande étrangère. La variable temps est souvent omise, comme dans la suite de cette section.

Cette formulation peut être spécifiée de façon linéaire :

$$(1) \quad X_t = b_1 Det + b_2 Comp_t + b_3 + u_t$$

dans ce cas b_1 représente la part du pays dans un accroissement du commerce international.

ou log linéaire :

$$(2) \quad \log X_t = \beta_1 \log Det + \beta_2 \log Comp_t + \beta_3 + u_t$$

β_1 est alors l'élasticité des exportations par rapport à la demande ou par rapport au commerce mondial et β_2 est l'élasticité - prix.

β_1 se compare à la valeur de référence qui est l'unité :

$\beta_1 = 1$ les exportations du pays augmentent au même rythme que la demande étrangère. La part du pays sur ses principaux marchés est stable.

$\beta_1 > 1$ Les exportations augmentent plus vite que les principaux marchés, le pays gagne des parts de marché.

$\beta_1 < 1$ Les exportations augmentent moins vite que la demande étrangère, le pays perd des parts de marché.

Les valeurs de β_1 sont en général légèrement supérieures à 1 pour la plupart des pays développés à l'exception du Royaume-Uni où elle est nettement inférieure. Ces valeurs traduisent la qualité des spécialisations et la capacité des économies à offrir les produits qui font l'objet d'une forte demande sur le marché mondial.

La comparaison des valeurs de long terme des élasticités revenu des importations, α_1 , et des exportations par rapport à la demande étrangère, β_1 , donne des indications sur la capacité d'une économie à soutenir un taux de croissance égal ou supérieur à celui du Reste du Monde.

Un pays pour lequel $\alpha_1 < \beta_1$ dispose de marges de croissance, qui toutes choses égales par ailleurs lui permettent de connaître un taux de croissance durablement supérieur à celui du reste du monde sans mettre en péril son équilibre extérieur.

Le Japon a pendant longtemps bénéficié d'une telle situation.

Au contraire, un pays pour lequel α_1 est nettement supérieur à β_1 viendra buter sur la contrainte extérieure dès que sa croissance interne aura tendance à s'accroître. Un tel pays ne peut suivre le rythme de croissance du reste du monde, qu'en compensant par des gains de compétitivité-prix obtenus par des dépréciations de sa monnaie.

En fait, la symétrie qui existe entre les équations d'importations et d'exportations est plus formelle que réelle. D'une part, il existe des délais de réaction beaucoup plus longs dans le cas des exportations et d'autre part la distinction entre exportations demandées et réalisées s'impose avec davantage d'évidence.

La réaction des exportations est beaucoup plus lente que celle des importations. Ce qui impose presque toujours l'utilisation de retards en données trimestrielles et parfois en données annuelles. Ces retards s'appliquent à la demande et aux prix. Une solution simple consiste à introduire la variable endogène décalée pour traduire la présence d'un mécanisme de retard à poids géométriquement décroissants :

$$(3) \log X_t = \beta_1 \log Det + \beta_2 \log Comp_t + \beta_3 \log X_{t-1} + \beta_4 + u_t$$

ce qui permet de faire apparaître des élasticités de court terme :

CT : élasticité à la demande β_1

élasticité aux prix β_2

et des élasticités de long terme :

$$\text{LT : élasticité par rapport à la demande : } \frac{\beta_1}{1 - \beta_3}$$

$$\text{élasticité prix : } \frac{\beta_2}{1 - \beta_3}$$

Exemple :

L'équation (3) a été utilisée pour estimer les exportations de produits manufacturés du Maroc vers la CEE.

On a vérifié dans un premier temps que toutes les variables sont intégrées d'ordre 1 y compris l'indicateur de compétitivité calculé par rapport à un échantillon de pays concurrents.

L'ajustement réalisé donne sur 1985 - 1 - 1992 - 4

$$LX_t = 0,36 \underset{(2,04)}{LIMPET} + 0,12 \underset{(0,47)}{LPr_t} + 0,30 \underset{(0,96)}{LComp_t} + 0,49 \underset{(2,62)}{LX_{t-1}} - 1,25 \underset{(0,6)}{u_t}$$

$$R^2 = 0,68 \quad DW = 1,89$$

(toutes les variables sont en log)

ou LX_t = indice de volume des exportations désaisonnalisées de produits manufacturés vers la CEE.

$LIMPET$: indice désaisonnalisées des importations des pays de la CEE pondérées par leur part dans les exportations marocaines.

LPr_t : indice de prix relatif du Maroc / CEE

$Lcomp_t$: indice de compétitivité du Maroc par rapport à un échantillon de pays concurrents.

Les élasticités prix ne sont pas significatives.

L'élasticité à la demande CEE est de 0,36 à court terme et de 0,705 à long terme.

On peut également utiliser des retards échelonnés sur Det et des retards polynomiaux sur $Comp_t$.

$$(4) \log X_t = \beta_1 \log Det + \beta_2 \log De_{t-1} + \beta_3 \log De_{t-2} + d(L)Comp_t + \beta_4 + u_t$$

Les exportations qui sont déterminées par l'équation (2) sont les exportations demandées. Elles peuvent différer des exportations réalisées. Dans le cas des importations, l'offre venait du Reste du Monde et pouvait être supposée infiniment élastique. Dans le cas des exportations, l'offre provient d'un seul pays et cette hypothèse ne tient plus.

Il faut donc distinguer l'offre d'exportations et la demande d'exportations. En toute logique ceci impose d'écrire un modèle à plusieurs équations dans lequel les exportations réalisées soit égalisent l'offre et la demande d'exportation, soit se fixent au minimum de l'offre et de la demande.

La fonction d'offre dépend de la production intérieure, de la rentabilité des exportations et des tensions sur les capacités de productions.

$$(5) \log X_t^d = \beta_1 \log Det + \beta_2 \log Comp_t + \beta_3 + u_t$$

$$(6) X_t^o = h(y_t, \frac{Pcx}{P}, Tuc)$$

$$(7) X_t = X_t^d = X_t^o$$

ou (7 bis) $X_t = \min(X_t^d, X_t^o)$

Une solution simple consiste à introduire dans l'équation (2) des variables additionnelles qui traduisent d'éventuelles tensions sur l'offre. L'équation (2) devient alors une équation décrivant les exportations réalisées. Cette transformation est adaptée pour rendre compte d'écarts temporaires résultants d'incapacités ponctuelles, mais conduirait à une sous-estimation de l'élasticité des exportations par rapport à la demande étrangère si les exportations

réalisées étaient de façon permanente limitées par la capacité d'offre. Le recours à une forme structurelle identifiable serait indispensable.

Comme cela a été mentionné dans la section 1, les équations d'exportations posent un problème d'ordre d'intégration des variables. Les variables telles que les exportations ou la demande étrangère seront intégrées d'ordre 1 tandis que la compétitivité doit logiquement être stationnaire. Les mêmes raisons conduisent en effet les autorités monétaires à lisser les écarts importants qui pourraient apparaître au niveau de la compétitivité prix. De ce fait les élasticités prix sont de toute façon délicates à estimer. L'économétrie mesure d'autant mieux que les variations sont plus marquées. Toutefois des travaux récents tendent à infirmer cette hypothèse de stationnarité de la compétitivité et concluent à une intégration d'ordre 1 pour cette variable dans le cas de l'Allemagne, des États-Unis ou de la France.

Les modèles à correction d'erreur peuvent être utilisés pour les fonctions d'exportations, soit que l'on se réfère aux propriétés du théorème de représentation de Granger lorsque toutes les variables sont I (1), soit que l'on introduise une variable stationnaire dans une spécification dont toutes les variables sont stationnaires :

$$(8) \quad \Delta \log X_t = k + \beta_0 \Delta \log Det - \gamma (\log X_{t-1} - \zeta \log De_{t-1}) + u_t$$

dans la mesure où X_t et De_t sont cointégrés (1,1) et si ζ est le paramètre du vecteur de cointégration, il est légitime de rajouter dans cette régression un terme également stationnaire :

$$(9) \quad \Delta \log X_t = k + \beta_0 \Delta \log Det + \beta_2 \log Comp_t - \gamma (\log X_{t-1} - \zeta \log De_{t-1}) + u_t$$

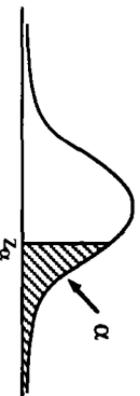
Le terme entre parenthèses n'est autre que l'écart en t-1 par rapport à la relation de long terme. D'où une autre formulation de la fonction d'exportations :

$$(10) \quad \Delta \log X_t = k + \beta_0 \Delta \log Det + \beta_2 \log Comp_t - \gamma \hat{ECM}_{t-1} + u_t$$

qui peut être estimée par les MCO sous cette forme.

Tables statistiques

Loi normale centrée réduite, Z
(loi normale de moyenne 0 et de variance 1)



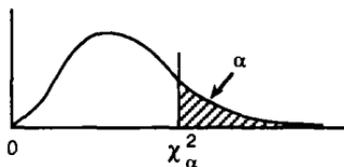
La table donne la valeur de z_α telle que $\alpha = P(Z > z_\alpha)$

α	0,000	0,005	0,010	0,015	0,020	0,025	0,030	0,035	0,040	0,045	0,050	0,055	0,060	0,065	0,070	0,075	0,080	0,085	0,090	0,095
0,0	∞	2,576	2,326	2,170	2,054	1,960	1,881	1,812	1,751	1,695	1,645	1,598	1,555	1,514	1,476	1,440	1,405	1,372	1,341	1,311
0,1	1,282	1,254	1,227	1,200	1,175	1,150	1,126	1,103	1,080	1,058	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,878	0,860
0,2	0,842	0,824	0,806	0,789	0,772	0,755	0,739	0,722	0,706	0,690	0,674	0,659	0,643	0,628	0,613	0,598	0,583	0,568	0,553	0,539
0,3	0,524	0,510	0,496	0,482	0,468	0,454	0,440	0,426	0,412	0,399	0,385	0,372	0,358	0,345	0,332	0,319	0,305	0,292	0,279	0,266
0,4	0,253	0,240	0,228	0,215	0,202	0,189	0,176	0,164	0,151	0,138	0,126	0,113	0,100	0,088	0,075	0,063	0,050	0,038	0,025	0,013
0,5	-0,000	-0,013	-0,025	-0,038	-0,050	-0,063	-0,075	-0,088	-0,100	-0,113	-0,126	-0,138	-0,151	-0,164	-0,176	-0,189	-0,202	-0,215	-0,228	-0,240
0,6	-0,253	-0,266	-0,279	-0,292	-0,305	-0,319	-0,332	-0,345	-0,358	-0,372	-0,385	-0,399	-0,412	-0,426	-0,440	-0,454	-0,468	-0,482	-0,496	-0,510
0,7	-0,524	-0,539	-0,553	-0,568	-0,583	-0,598	-0,613	-0,628	-0,643	-0,659	-0,674	-0,690	-0,706	-0,722	-0,739	-0,755	-0,772	-0,789	-0,806	-0,824
0,8	-0,842	-0,860	-0,878	-0,896	-0,915	-0,935	-0,954	-0,974	-0,994	-1,015	-1,036	-1,058	-1,080	-1,103	-1,126	-1,150	-1,175	-1,200	-1,227	-1,254
0,9	-1,282	-1,311	-1,341	-1,372	-1,405	-1,440	-1,476	-1,514	-1,555	-1,598	-1,645	-1,695	-1,751	-1,812	-1,881	-1,960	-2,054	-2,170	-2,326	-2,576

Exemple : pour $\alpha = 0,025$, la table donne $z_\alpha = 1,960$. C'est-à-dire que $P(Z > 1,960) = 0,025$, ou, de façon équivalente, que $P(Z \leq 1,96) = 0,975$.

Table de la loi de khi-2 (χ^2)

La table donne la valeur χ^2_{α} telle que $\alpha = P(\chi^2 > \chi^2_{\alpha})$

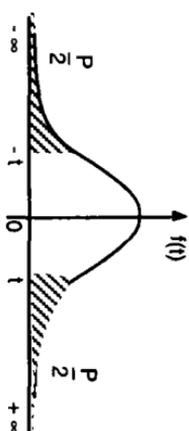


d.d.l.	α										
	0,975	0,95	0,90	0,50	0,30	0,20	0,10	0,05	0,025	0,01	0,001
1	0,00	0,00	0,02	0,46	1,07	1,64	2,71	3,84	5,02	6,63	10,83
2	0,05	0,10	0,21	1,39	2,41	3,22	4,61	5,99	7,38	9,21	13,82
3	0,22	0,35	0,58	2,37	3,66	4,64	6,25	7,81	9,35	11,34	16,27
4	0,48	0,71	1,06	3,36	4,88	5,99	7,78	9,49	11,14	13,28	18,47
5	0,83	1,15	1,61	4,35	6,06	7,29	9,24	11,07	12,83	15,09	20,52
6	1,24	1,64	2,20	5,35	7,23	8,56	10,65	12,59	14,45	16,81	22,46
7	1,69	2,17	2,83	6,35	8,38	9,80	12,02	14,07	16,01	18,48	24,32
8	2,18	2,73	3,49	7,34	9,52	11,03	13,36	15,51	17,53	20,09	26,13
9	2,70	3,33	4,17	8,34	10,66	12,24	14,68	16,92	19,02	21,67	27,88
10	3,25	3,94	4,87	9,34	11,78	13,44	15,99	18,31	20,48	23,21	29,59
11	3,82	4,57	5,58	10,34	12,90	14,63	17,28	19,68	21,92	24,73	31,26
12	4,40	5,23	6,30	11,34	14,01	15,81	18,55	21,03	23,34	26,22	32,91
13	5,01	5,89	7,04	12,34	15,12	16,99	19,81	22,36	24,74	27,69	34,53
14	5,63	6,57	7,79	13,34	16,22	18,15	21,06	23,69	26,12	29,14	36,12
15	6,26	7,26	8,55	14,34	17,32	19,31	22,31	25,00	27,49	30,58	37,70
16	6,91	7,96	9,31	15,34	18,42	20,47	23,54	26,30	28,85	32,00	39,25
17	7,56	8,67	10,09	16,34	19,51	21,62	24,77	27,59	30,19	33,41	40,79
18	8,23	9,39	10,87	17,34	20,60	22,76	25,99	28,87	31,53	34,81	42,31
19	8,91	10,12	11,65	18,34	21,69	23,90	27,20	30,14	32,85	36,19	43,82
20	9,59	10,85	12,44	19,34	22,78	25,04	28,41	31,41	34,17	37,57	45,32
21	10,28	11,59	13,24	20,34	23,86	26,17	29,62	32,67	35,48	38,93	46,80
22	10,98	12,34	14,04	21,34	24,94	27,30	30,81	33,92	36,78	40,29	48,27
23	11,69	13,09	14,85	22,34	26,02	28,43	32,01	35,17	38,08	41,64	49,73
24	12,40	13,85	15,66	23,34	27,10	29,55	33,20	36,42	39,36	42,98	51,18
25	13,12	14,61	16,47	24,34	28,17	30,68	34,38	37,65	40,65	44,31	52,62
26	13,84	15,38	17,29	25,34	29,25	30,80	35,56	38,89	41,92	45,64	54,05
27	14,57	16,15	18,11	26,34	30,32	32,91	36,74	40,11	43,19	46,96	55,48
28	15,31	16,93	18,94	27,34	31,39	34,03	37,92	41,34	44,46	48,28	56,89
29	16,05	17,71	19,77	28,34	32,46	35,14	39,09	42,56	45,72	49,59	58,30
30	16,79	18,49	20,60	29,34	33,53	36,25	40,26	43,77	46,98	50,89	59,70
40	24,43	26,51	29,05	39,34	44,16	47,27	51,80	55,76	59,34	63,69	73,40
60	40,48	43,19	46,46	59,33	65,23	68,97	74,40	79,08	83,30	88,38	99,61
100	74,22	77,93	82,36	99,33	106,91	111,67	118,50	124,34	129,56	135,81	149,45

Quand le nombre de degrés de liberté (ddl) est élevé (en pratique, supérieur à 30), χ^2 suit approximativement une loi normale de moyenne $\mu = \text{ddl}$ et de variance $\sigma^2 = 2\text{ddl}$.

Table de distribution de T (Loi de Student)

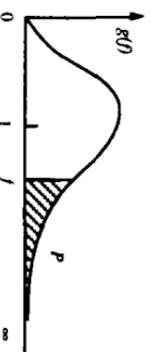
Valeurs de T ayant la probabilité P d'être dépassées en valeur absolue



1	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.001
2	0.142	0.235	0.310	0.372	0.400	0.416	0.426	0.431	0.434	0.436	0.437	0.438	0.439
3	0.137	0.227	0.294	0.348	0.376	0.389	0.396	0.399	0.401	0.402	0.403	0.404	0.405
4	0.134	0.221	0.280	0.326	0.354	0.366	0.371	0.373	0.375	0.376	0.377	0.378	0.379
5	0.132	0.217	0.268	0.314	0.341	0.352	0.356	0.358	0.360	0.361	0.362	0.363	0.364
6	0.131	0.215	0.265	0.310	0.337	0.347	0.351	0.353	0.355	0.356	0.357	0.358	0.359
7	0.130	0.213	0.262	0.306	0.333	0.343	0.346	0.348	0.350	0.351	0.352	0.353	0.354
8	0.130	0.212	0.261	0.304	0.331	0.341	0.344	0.346	0.348	0.349	0.350	0.351	0.352
9	0.129	0.211	0.260	0.302	0.329	0.339	0.342	0.344	0.346	0.347	0.348	0.349	0.350
10	0.129	0.210	0.259	0.301	0.327	0.337	0.340	0.342	0.344	0.345	0.346	0.347	0.348
11	0.129	0.210	0.259	0.300	0.326	0.336	0.339	0.341	0.343	0.344	0.345	0.346	0.347
12	0.128	0.209	0.258	0.299	0.325	0.335	0.338	0.340	0.342	0.343	0.344	0.345	0.346
13	0.128	0.209	0.258	0.298	0.324	0.334	0.337	0.339	0.341	0.342	0.343	0.344	0.345
14	0.128	0.208	0.257	0.297	0.323	0.333	0.336	0.338	0.340	0.341	0.342	0.343	0.344
15	0.128	0.208	0.257	0.297	0.322	0.332	0.335	0.337	0.339	0.340	0.341	0.342	0.343
16	0.128	0.208	0.257	0.296	0.321	0.331	0.334	0.336	0.338	0.339	0.340	0.341	0.342
18	0.127	0.207	0.256	0.295	0.320	0.330	0.333	0.335	0.337	0.338	0.339	0.340	0.341
19	0.127	0.207	0.256	0.294	0.319	0.329	0.332	0.334	0.336	0.337	0.338	0.339	0.340
20	0.127	0.207	0.256	0.294	0.318	0.328	0.331	0.333	0.335	0.336	0.337	0.338	0.339
21	0.127	0.207	0.256	0.293	0.317	0.327	0.330	0.332	0.334	0.335	0.336	0.337	0.338
22	0.127	0.207	0.256	0.293	0.316	0.326	0.329	0.331	0.333	0.334	0.335	0.336	0.337
23	0.127	0.207	0.256	0.292	0.315	0.325	0.328	0.330	0.332	0.333	0.334	0.335	0.336
24	0.127	0.207	0.256	0.292	0.314	0.324	0.327	0.329	0.331	0.332	0.333	0.334	0.335
26	0.127	0.207	0.256	0.291	0.313	0.323	0.326	0.328	0.330	0.331	0.332	0.333	0.334
27	0.127	0.207	0.256	0.291	0.312	0.322	0.325	0.327	0.329	0.330	0.331	0.332	0.333
28	0.127	0.207	0.256	0.290	0.311	0.321	0.324	0.326	0.328	0.329	0.330	0.331	0.332
29	0.127	0.207	0.256	0.290	0.310	0.320	0.323	0.325	0.327	0.328	0.329	0.330	0.331
30	0.127	0.207	0.256	0.289	0.309	0.319	0.322	0.324	0.326	0.327	0.328	0.329	0.330
40	0.126	0.206	0.255	0.288	0.308	0.318	0.321	0.323	0.325	0.326	0.327	0.328	0.329
80	0.126	0.206	0.255	0.287	0.307	0.317	0.320	0.322	0.324	0.325	0.326	0.327	0.328
120	0.126	0.206	0.255	0.286	0.306	0.316	0.319	0.321	0.323	0.324	0.325	0.326	0.327
∞	0.126	0.253	0.385	0.524	0.674	0.842	1.036	1.282	1.645	1.960	2.376	2.576	3.291

Table de distribution de F (Loi de Fisher-Snedecor)

Valeurs de F ayant la probabilité P d'être dépassées ($F = s_1^2 / s_2^2$)

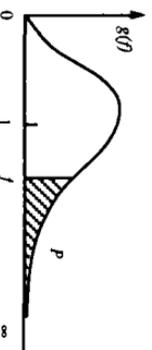


ν_2	$\nu_1 = 1$		$\nu_1 = 2$		$\nu_1 = 3$		$\nu_1 = 4$		$\nu_1 = 5$	
	P = 0.05	P = 0.01								
1	161.4	4 052	199.5	4 990	215.7	5 403	224.6	5 625	230.2	5 764
2	98.48	2 500	120.0	3 000	137.1	3 291	147.6	3 525	153.0	3 704
3	77.13	2 142	95.5	2 380	107.6	2 599	116.3	2 807	121.0	2 979
4	67.57	1 881	83.4	2 127	97.9	2 374	106.4	2 599	111.8	2 776
5	61.26	1 674	75.9	1 927	91.3	2 227	100.0	2 453	106.3	2 639
6	56.99	1 514	70.8	1 781	86.3	2 114	95.9	2 343	102.6	2 547
7	53.9	1 404	67.2	1 666	83.4	2 027	93.0	2 271	100.0	2 488
8	51.72	1 326	64.6	1 571	81.1	1 959	91.0	2 214	98.0	2 437
9	50.04	1 266	62.6	1 496	79.3	1 904	89.4	2 167	96.5	2 392
10	4.96	1 217	61.1	1 437	78.0	1 859	88.4	2 127	95.5	2 352
11	4.84	1 176	59.9	1 390	77.1	1 821	87.6	2 092	94.8	2 317
12	4.75	1 141	58.9	1 352	76.4	1 788	87.0	2 061	94.2	2 285
13	4.67	1 111	58.1	1 320	75.8	1 759	86.5	2 033	93.7	2 255
14	4.60	1 084	57.4	1 292	75.3	1 733	86.1	2 008	93.3	2 227
15	4.54	1 061	56.8	1 268	74.9	1 710	85.7	1 985	93.0	2 201
16	4.49	1 041	56.3	1 247	74.5	1 690	85.4	1 964	92.7	2 176
17	4.45	1 023	55.9	1 228	74.2	1 672	85.1	1 945	92.5	2 152
18	4.41	1 007	55.5	1 211	73.9	1 656	84.9	1 928	92.3	2 129
19	4.38	993	55.2	1 195	73.7	1 641	84.7	1 913	92.1	2 107
20	4.35	980	55.0	1 181	73.5	1 628	84.5	1 900	92.0	2 086
21	4.32	968	54.7	1 168	73.4	1 616	84.4	1 888	91.9	2 066
22	4.30	957	54.6	1 156	73.3	1 605	84.3	1 877	91.8	2 047
23	4.28	946	54.4	1 145	73.2	1 595	84.2	1 867	91.7	2 029
24	4.26	936	54.3	1 134	73.1	1 586	84.1	1 857	91.6	2 012
25	4.24	926	54.2	1 124	73.0	1 577	84.0	1 848	91.5	1 996
26	4.22	917	54.1	1 114	72.9	1 569	83.9	1 839	91.4	1 981
27	4.21	908	54.0	1 104	72.8	1 561	83.8	1 831	91.3	1 967
28	4.20	900	53.9	1 095	72.8	1 553	83.7	1 823	91.2	1 954
29	4.18	892	53.8	1 086	72.7	1 546	83.6	1 816	91.1	1 941
30	4.17	884	53.7	1 078	72.7	1 539	83.5	1 809	91.0	1 929
40	4.10	840	53.3	1 023	72.4	1 500	83.2	1 771	90.7	1 876
50	4.05	800	53.0	973	72.1	1 466	82.9	1 738	90.4	1 832
60	4.01	768	52.8	928	71.8	1 436	82.6	1 709	90.1	1 794
80	3.92	700	52.4	840	71.4	1 380	82.2	1 650	89.7	1 728
100	3.84	664	52.2	799	71.2	1 350	82.0	1 620	89.5	1 696

s_1^2 est la plus grande des deux variances estimées avec ν_1 degrés de liberté.

Table de distribution de F (Loi de Fisher-Snedecor)

Valeurs de F ayant la probabilité P d'être dépassées ($F = s_1^2 / s_2^2$)



v_2	$v_1 = 6$		$v_1 = 8$		$v_1 = 12$		$v_1 = 24$		$v_1 = \infty$	
	P = 0,05	P = 0,01	P = 0,05	P = 0,01	P = 0,05	P = 0,01	P = 0,05	P = 0,01	P = 0,05	P = 0,01
1	234,0	5 859	238,9	5 981	243,9	6 106	249,0	6 234	254,3	6 366
2	19,33	99,33	19,37	99,36	19,41	99,42	19,45	99,46	19,50	99,50
3	8,94	27,91	8,84	27,69	8,74	27,05	8,64	26,60	8,53	26,12
4	6,16	15,21	6,04	14,80	5,91	14,37	5,77	13,93	5,65	13,46
5	4,95	10,67	4,82	10,27	4,68	9,89	4,53	9,47	4,36	9,02
6	4,28	8,47	4,15	8,10	4,00	7,72	3,84	7,31	3,67	6,88
7	3,87	7,19	3,73	6,84	3,57	6,47	3,41	6,07	3,23	5,65
8	3,58	6,37	3,44	6,03	3,28	5,67	3,12	5,28	2,93	4,86
9	3,37	5,80	3,23	5,47	3,07	5,11	2,90	4,73	2,71	4,31
10	3,22	5,39	3,07	5,06	2,91	4,71	2,74	4,33	2,54	3,91
11	3,09	5,07	2,95	4,74	2,79	4,40	2,61	4,02	2,40	3,60
12	3,00	4,82	2,85	4,50	2,69	4,16	2,50	3,78	2,30	3,36
13	2,92	4,62	2,77	4,30	2,60	3,96	2,42	3,59	2,21	3,16
14	2,85	4,46	2,70	4,14	2,50	3,80	2,35	3,45	2,15	3,00
15	2,79	4,33	2,65	4,00	2,48	3,67	2,30	3,32	2,07	2,87
16	2,74	4,20	2,59	3,89	2,43	3,57	2,25	3,22	2,01	2,75
17	2,70	4,10	2,55	3,79	2,38	3,48	2,19	3,16	1,95	2,65
18	2,66	4,01	2,51	3,71	2,34	3,41	2,15	3,08	1,90	2,57
19	2,63	3,94	2,48	3,65	2,31	3,30	2,11	2,97	1,86	2,50
20	2,60	3,87	2,45	3,56	2,28	3,23	2,08	2,86	1,84	2,42
21	2,57	3,81	2,42	3,51	2,25	3,17	2,05	2,80	1,81	2,36
22	2,55	3,76	2,40	3,45	2,23	3,12	2,03	2,75	1,78	2,31
23	2,53	3,71	2,38	3,41	2,20	3,07	2,00	2,70	1,76	2,26
24	2,51	3,67	2,36	3,36	2,18	3,03	1,98	2,66	1,73	2,21
25	2,49	3,63	2,34	3,32	2,16	2,99	1,96	2,62	1,71	2,17
26	2,47	3,59	2,32	3,29	2,15	2,96	1,95	2,58	1,69	2,13
27	2,46	3,56	2,30	3,26	2,13	2,93	1,93	2,55	1,67	2,10
28	2,44	3,53	2,29	3,23	2,12	2,90	1,91	2,52	1,65	2,06
29	2,43	3,50	2,28	3,20	2,10	2,87	1,89	2,49	1,64	2,03
30	2,42	3,47	2,27	3,17	2,09	2,84	1,89	2,47	1,62	2,01
40	2,34	3,29	2,18	2,99	2,00	2,66	1,79	2,29	1,51	1,80
60	2,25	3,12	2,10	2,82	1,92	2,50	1,70	2,12	1,39	1,60
80	2,17	2,96	2,01	2,66	1,83	2,34	1,61	1,95	1,25	1,38
120	2,09	2,80	1,94	2,51	1,75	2,18	1,52	1,79	1,00	1,00

s_1^2 est la plus grande des deux variances estimées avec v_1 degrés de liberté

Tables du d de Durbin-Watson

Risque à 5 %

T	k = 1		k = 2		k = 3		k = 4		k = 5	
	d1	d2								
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
31	1,36	1,50	1,30	1,57	1,23	1,65	1,16	1,74	1,09	1,83
32	1,37	1,50	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82
33	1,38	1,51	1,32	1,58	1,26	1,65	1,19	1,73	1,13	1,81
34	1,39	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,15	1,81
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80
36	1,41	1,52	1,35	1,59	1,29	1,65	1,24	1,73	1,18	1,80
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,80
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,21	1,79
39	1,43	1,54	1,38	1,60	1,33	1,66	1,27	1,72	1,22	1,79
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
55	1,53	1,60	1,49	1,64	1,45	1,68	1,41	1,72	1,38	1,77
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,50	1,70	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,52	1,70	1,49	1,74	1,46	1,77
75	1,60	1,65	1,57	1,68	1,54	1,71	1,51	1,74	1,49	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
85	1,62	1,67	1,60	1,70	1,57	1,72	1,55	1,75	1,52	1,77
90	1,63	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78
95	1,64	1,69	1,62	1,71	1,60	1,73	1,58	1,75	1,56	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,75	1,57	1,78

k est le nombre de variables exogènes (constante exclue).

T est la taille de l'échantillon

Achévé d'imprimer sur les presses d'Offset Services
56, rue Lucien Dupuis - 28500 VERNOUILLET

La collection **Universités francophones**, créée en 1988 à l'initiative de l'Agence francophone pour l'enseignement supérieur et la recherche, propose des ouvrages modernes répondant prioritairement aux besoins des étudiants et chercheurs francophones, et se compose de titres originaux paraissant régulièrement.

Leurs auteurs appartiennent conjointement aux pays du Sud et du Nord et rendent compte des résultats de recherches et des études récentes entreprises en français à travers le monde. Ils permettent à cette collection pluridisciplinaire de couvrir progressivement l'ensemble des enseignements universitaires en français.

Enfin, la vente à un prix préférentiel, des ouvrages destinés aux pays du Sud tient compte des exigences économiques nationales et assure une diffusion adaptée aux pays francophones.

Ainsi, la collection **Universités francophones** constitue une bibliothèque universitaire de référence riche et complète répondant aux besoins de l'étudiant, du chercheur et du professionnel.

Prix 59 FF • Prix préférentiel UREF (Afrique, Asie, Amérique Latine, Moyen-Orient, Haïti) : 10 FF



ISBN 2-909455-50-5