

Économétrie

Cours et exercices corrigés

Régis Bourbonnais

9^e édition

DUNOD

Le pictogramme qui figure ci-contre mérite une explication. Son objet est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, particulièrement dans le domaine de l'édition technique et universitaire, le développement massif du photocopillage.

Le Code de la propriété intellectuelle du 1^{er} juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique s'est généralisée dans les établissements

d'enseignement supérieur, provoquant une baisse brutale des achats de livres et de revues, au point que la possibilité même pour

les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée. Nous rappelons donc que toute reproduction, partielle ou totale, de la présente publication est interdite sans autorisation de l'auteur, de son éditeur ou du Centre français d'exploitation du droit de copie (CFC, 20, rue des Grands-Augustins, 75006 Paris).



© Dunod, 2015

5 rue Laromiguière, 75005 Paris

www.dunod.com

ISBN 978-2-10-072151-1

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2° et 3° a), d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

Table des matières

Avant-propos	IX
1. Qu'est-ce que l'économétrie ?	1
I. La notion de modèle	1
A. Définition	1
B. La construction des modèles en économétrie	2
II. Le rôle de l'économétrie	5
A. L'économétrie comme validation de la théorie	5
B. L'économétrie comme outil d'investigation	5
III. La théorie de la corrélation	6
A. Présentation générale	6
B. Mesure et limite du coefficient de corrélation	8
2. Le modèle de régression simple	13
I. Présentation du modèle	13
A. Exemple introductif	13
B. Rôle du terme aléatoire	14
C. Conséquences du terme aléatoire	16
II. Estimation des paramètres	17
A. Modèle et hypothèses	17
B. Formulation des estimateurs	18
C. Les différentes écritures du modèle : erreur et résidu	21
D. Propriétés des estimateurs	22
III. Conséquences des hypothèses : construction des tests	24
A. Hypothèse de normalité des erreurs	24
B. Conséquences de l'hypothèse de normalité des erreurs	25
C. Test bilatéral, test unilatéral et probabilité critique d'un test	27
IV. Équation et tableau d'analyse de la variance	33
A. Équation d'analyse de la variance	33
B. Tableau d'analyse de la variance	34
V. La prévision dans le modèle de régression simple	39

3. Le modèle de régression multiple	47
I. Le modèle linéaire général	47
A. Présentation	47
B. Forme matricielle	48
II. Estimation et propriétés des estimateurs	49
A. Estimation des coefficients de régression	49
B. Hypothèses et propriétés des estimateurs	51
C. Équation d'analyse de la variance et qualité d'un ajustement	54
III. Les tests statistiques	59
A. Le rôle des hypothèses	59
B. Construction des tests	60
C. Tests sur les résidus : valeur anormale, effet de levier et point d'influence	62
IV. L'analyse de la variance	67
A. Construction du tableau d'analyse de la variance et test de signification globale d'une régression	67
B. Autres tests à partir du tableau d'analyse de la variance	68
C. Généralisation des tests par analyse de la variance	73
V. L'utilisation de variables indicatrices	75
A. Constitution et finalités des variables indicatrices	75
B. Exemples d'utilisation	76
VI. La prévision à l'aide du modèle linéaire général et la régression récursive	81
A. Prédiction conditionnelle	81
B. Fiabilité de la prévision et intervalle de prévision	82
C. Les tests de stabilité par la régression récursive	84
D. Le test de spécification de Ramsey	86
VII. Exercices récapitulatifs	90
Annexe	102
A) Interprétation géométrique de la méthode des moindres carrés	102
B) Résolution de l'exercice 1 par des logiciels informatiques de régression multiple	103
C) Estimation de la variance de l'erreur	105
4. Multicolinéarité et sélection du modèle optimal	107
I. Corrélation partielle	107
A. Exemple introductif	107
B. Généralisation de la notion de corrélation partielle	108
II. Relation entre coefficients de corrélation simple, partielle et multiple	112

III. Multicolinéarité : conséquences et détection	114
A. Conséquences de la multicolinéarité	114
B. Tests de détection d'une multicolinéarité	115
C. Comment remédier à la multicolinéarité ?	118
IV. Sélection du modèle optimal	119
5. Problèmes particuliers : la violation des hypothèses	125
I. L'autocorrélation des erreurs	125
A. Présentation du problème	125
B. L'estimateur des Moindres Carrés Généralisés (MCG)	126
C. Les causes et la détection de l'autocorrélation des erreurs	127
D. Les procédures d'estimation en cas d'autocorrélation des erreurs	134
II. L'hétéroscédasticité	142
A. Présentation du problème	142
B. Correction de l'hétéroscédasticité	144
C. Tests de détection de l'hétéroscédasticité	147
D. Autre test d'hétéroscédasticité : le test ARCH	153
III. Modèles à erreurs sur les variables	154
A. Conséquences lorsque les variables sont entachées d'erreurs	154
B. La méthode des variables instrumentales	155
C. Le test d'exogénéité d'Hausman	156
D. La méthode des moments généralisée	157
6. Les modèles non linéaires	165
I. Les différents types de modèles non linéaires	165
A. Les fonctions de type exponentiel	165
B. Les modèles de diffusion	168
II. Méthodes d'estimation des modèles non linéaires	170
A. Initiation aux méthodes d'estimation non linéaires	170
B. Exemples d'application	172
7. Les modèles à décalages temporels	177
I. Les modèles linéaires autorégressifs	177
A. Formulation générale	177
B. Test d'autocorrélation et méthodes d'estimation	178
II. Les modèles à retards échelonnés	183
A. Formulation générale	183
B. Détermination du nombre de retards	184

C. Distribution finie des retards	188
D. Distribution infinie des retards	192
III. Deux exemples de modèles dynamiques	198
A. Le modèle d'ajustement partiel	198
B. Le modèle d'anticipations adaptatives	199
8. Introduction aux modèles à équations simultanées	217
I. Équations structurelles et équations réduites	218
A. Exemple introductif	218
B. Le modèle général	220
C. Cas particulier : les modèles récurrents	221
II. Le problème de l'identification	221
A. Restrictions sur les coefficients	221
B. Conditions d'identification	222
III. Les méthodes d'estimation	223
A. Les moindres carrés indirects	223
B. Les doubles moindres carrés	223
C. Autres méthodes d'estimation	224
<i>Annexe</i>	236
<i>Identification : les conditions de rang</i>	236
9. Éléments d'analyse des séries temporelles	239
I. Stationnarité	239
A. Définition et propriétés	239
B. Fonctions d'autocorrélation simple et partielle	240
C. Tests de « bruit blanc » et de stationnarité	241
II. La non-stationnarité et les tests de racine unitaire	245
A. La non-stationnarité : les processus TS et DS	245
B. Les tests de racine unitaire et la stratégie séquentielle de test	248
III. Les modèles ARIMA	256
A. Typologie des modèles AR, MA et ARMA	256
B. L'extension aux processus ARIMA et SARIMA	259
IV. La méthode de Box et Jenkins	260
A. Recherche de la représentation adéquate : l'identification	260
B. Estimation des paramètres	261
C. Tests d'adéquation du modèle et prévision	262
10. La modélisation VAR	275
I. Représentation d'un modèle VAR	276
A. Exemple introductif	276

B. La représentation générale	277
C. La représentation ARMAX	278
II. Estimation des paramètres	279
A. Méthode d'estimation	279
B. Détermination du nombre de retards	279
C. Prévision	280
III. Dynamique d'un modèle VAR	284
A. Représentation VMA d'un processus VAR	284
B. Analyse et orthogonalisation des « chocs »	285
C. Décomposition de la variance	288
D. Choix de l'ordre de décomposition	288
IV. La causalité	292
A. Causalité au sens de Granger	292
B. Causalité au sens de Sims	293
11. La cointégration et le modèle à correction d'erreur	297
I. Exemples introductifs	297
II. Le concept de cointégration	299
A. Propriétés de l'ordre d'intégration d'une série	299
B. Conditions de cointégration	301
C. Le modèle à correction d'erreur (ECM)	301
III. Cointégration entre deux variables	302
A. Test de cointégration entre deux variables	303
B. Estimation du modèle à correction d'erreur	303
IV. Généralisation à k variables	306
A. La cointégration entre k variables	306
B. Estimation du modèle à correction d'erreur	307
C. Le modèle à correction d'erreur vectoriel	308
D. Tests de relation de cointégration	310
E. Test d'exogénéité faible	313
F. Synthèse de la procédure d'estimation	314
12. Introduction à l'économétrie des variables qualitatives	319
I. Les problèmes et les conséquences de la spécification binaire	320
II. Les modèles de choix binaires	322
A. Le modèle linéaire sur variable latente	322
B. Les modèles Probit et Logit	323
C. Interprétation des résultats et tests statistiques	325
III. Les modèles à choix multiples	330
A. Les modèles Probit et Logit ordonnés	331

B. Le modèle de choix multiples non ordonné : le Logit multinomial	335
IV. Les modèles à variable dépendante limitée : le modèle Tobit	337
A. Le modèle Tobit simple : modèle de régression tronqué ou censuré	338
B. Estimation et interprétation des résultats	340
13. Introduction à l'économétrie des données de panel	345
I. Présentation des modèles à données de panel	346
A. Spécificités des données de panel	346
B. La méthode SUR	347
C. Le modèle linéaire simple	348
II. Les tests d'homogénéité	349
A. Procédure séquentielle de tests	349
B. Construction des tests	350
III. Spécifications et estimations des modèles à effets individuels	355
A. Le modèle à effets fixes individuels	355
B. Le modèle à effets aléatoires	357
C. Effets fixes ou effets aléatoires ? Le test d'Hausman	358
Liste des exercices	363
Tables statistiques	367
Bibliographie	375
Index	379

Avant-propos


Cette neuvième édition est enrichie de nouveaux exercices et des développements les plus récents de l'économétrie. Ce livre couvre tous les champs de l'économétrie : régression simple et multiple, violation des hypothèses (hétéroscédasticité, autocorrélation des erreurs, variables explicatives aléatoires), modèle à décalage, analyse des séries temporelles, tests de racine unitaire, équations multiples, VAR, cointégration, VECM, économétrie des variables qualitatives et des données de panel...

Sur l'ensemble de ces thèmes, ce livre vous propose un cours, des exercices corrigés, et une présentation des logiciels d'économétrie les plus répandus. Souhaitons qu'il corresponde à votre attente.

En effet, nous avons voulu, par une alternance systématique de cours et d'exercices, répondre à un besoin pédagogique qui est de mettre rapidement en pratique les connaissances théoriques et ainsi, d'utiliser de manière opérationnelle les acquis du cours ; les exercices sont repérés grâce à un bandeau grisé. De surcroît, le recours à des logiciels¹, lors de la résolution des exercices, permet une découverte de ces outils et donne une dimension pratique que recherchent l'étudiant et le praticien.

Afin que le lecteur puisse lui-même refaire les exercices, les données utilisées (sous format Excel, ASCII, RATS et Eviews) ainsi que les programmes de traitement « Batch » de Eviews ou de RATS sont disponibles gratuitement par téléchargement sur le serveur web :

<http://regisbourbonnais.dauphine.fr>

Pour chaque exercice faisant appel à un fichier de données, le nom du fichier est cité en tête de l'exercice et repéré par l'icône suivante : 

Nous avons voulu faire de ce manuel un livre d'apprentissage facilement accessible ; c'est pourquoi les démonstrations les plus complexes font l'objet de renvois à une bibliographie plus spécialisée. Cependant, il convient de préciser que l'économétrie fait appel à des notions d'algèbre linéaire et d'induction statistique qu'il est souhaitable de connaître.

1. Trois logiciels sont utilisés : EXCEL (© Microsoft), RATS (© Var Econometrics version 3 et Estima version 4), Eviews (© Quantitative Micro Software). Nous recommandons aussi particulièrement le logiciel GRETL (<http://gretl.sourceforge.net>) qui est un logiciel d'économétrie gratuit, complet et très facile d'apprentissage.

Dans le terme « économétrie » figure la racine du mot « économie » car son utilisation est surtout destinée à des fins de traitement de données économiques ; cependant, d'autres domaines tels que la finance, la recherche agronomique, la médecine, etc., font maintenant le plus souvent appel à ces techniques.

Ce livre s'adresse en premier lieu aux étudiants (sciences économiques, gestion, écoles de commerce et d'ingénieurs, etc.) dont la formation requiert une connaissance de l'économétrie. Gageons qu'il sera un support de cours indispensable et un allié précieux pour préparer les séances de travaux dirigés.

N'oublions pas cependant le praticien de l'économétrie (économiste d'entreprise, chercheur, etc.) qui, confronté à des problèmes d'estimation statistique, trouvera dans ce livre les réponses pratiques aux différentes questions qu'il peut se poser.

Enfin, j'exprime toute ma gratitude à toutes les personnes – collègues et étudiants – qui ont eu la gentillesse de me faire des commentaires et dont les conseils et suggestions contribuent à la qualité pédagogique de ce livre. Je reste, bien entendu, le seul responsable des erreurs qui subsisteraient¹.

1. Les lecteurs souhaitant faire des commentaires ou des remarques peuvent me contacter : Régis Bourbonnais, université de Paris-Dauphine, place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16, E-mail : regis.bourbonnais@dauphine.fr

1. Qu'est-ce que l'économétrie ?

Ce premier chapitre est consacré à la présentation de l'économétrie et à sa liaison avec la théorie économique. Nous abordons tout d'abord la notion de modèle ainsi que les différentes étapes de la modélisation. L'apport de l'économétrie en tant qu'outil de validation est étudié en II. Enfin, la théorie de la corrélation – fondement de l'économétrie – fait l'objet de la section III.

I. La notion de modèle

A. Définition

Il est délicat de fournir une définition unique de la notion de modèle¹. Dans le cadre de l'économétrie, nous pouvons considérer qu'un modèle consiste en une *présentation formalisée d'un phénomène* sous forme d'équations dont les variables sont des grandeurs économiques. L'objectif du modèle est de représenter les traits les plus marquants d'une réalité qu'il cherche à styliser. Le modèle est donc l'outil que le modélisateur utilise lorsqu'il cherche à comprendre et à expliquer des phénomènes. Pour ce faire, il émet des hypothèses et explicite des relations.

1. La notion de modèle est relative au point de vue auquel nous nous plaçons : la physique, l'épistémologie...

► *Pourquoi des modèles ?*

- *Nombreux sont ceux – sociologues, économistes ou physiciens – qui fondent leurs analyses ou leurs jugements sur des raisonnements construits et élaborés. Ces constructions réfèrent implicitement à des modèles ; alors pourquoi ne pas expliciter clairement les hypothèses et les relations au sein d'un modèle ?*

Le modèle est donc une présentation schématique et partielle d'une réalité naturellement plus complexe. Toute la difficulté de la modélisation consiste à ne retenir que la ou les représentations intéressantes pour le problème que le modélisateur cherche à expliciter. Ce choix dépend de la nature du problème, du type de décision ou de l'étude à effectuer. La même réalité peut ainsi être formalisée de diverses manières en fonction des objectifs.

B. La construction des modèles en économétrie

Dans les sciences sociales, et particulièrement en économie, les phénomènes étudiés concernent le plus souvent des comportements afin de mieux comprendre la nature et le fonctionnement des systèmes économiques. L'objectif du modélisateur est, dans le cadre de l'économétrie et au travers d'une mesure statistique, de permettre aux agents économiques (ménages, entreprises, État...) d'intervenir de manière plus efficace. La construction d'un modèle comporte un certain nombre d'étapes qui sont toutes importantes. En effet, en cas de faiblesse d'un des « maillons », le modèle peut se trouver invalidé pour cause d'hypothèses manquantes, de données non représentatives ou observées avec des erreurs, etc. Examinons les différentes étapes à suivre lors de la construction d'un modèle, ceci à partir de l'exemple du modèle keynésien simplifié.

1) Référence à une théorie

Une théorie s'exprime au travers d'hypothèses auxquelles le modèle fait référence. Dans la théorie keynésienne, quatre propositions sont fondamentales :

1. la consommation et le revenu sont liés ;
2. le niveau d'investissement privé et le taux d'intérêt sont également liés ;
3. il existe un investissement autonome public ;
4. enfin, le produit national est égal à la consommation plus l'investissement privé et public.

2) Formalisation des relations et choix de la forme des fonctions

À partir des propositions précédentes, nous pouvons construire des relations :

1. la consommation est fonction du revenu : $C = f(Y)$ avec $f' > 0$;
2. l'investissement privé dépend du taux d'intérêt : $I = g(r)$ avec $g' < 0$;
3. il existe un investissement autonome public : \bar{I} ;
4. enfin, le produit national (ou le revenu national) est égal à la consommation plus l'investissement : $Y \equiv C + I + \bar{I}$.

À ce stade, nous n'avons postulé aucune forme particulière en ce qui concerne les fonctions f et g . Ainsi, bien que des considérations d'ordre théorique nous renseignent sur le signe des dérivées, il existe une multitude de fonctions de formes très différentes et ayant des signes de dérivées identiques, par exemple $C = a_0 + a_1 Y$ et $C = a_0 Y^{a_1}$. Cependant ces deux relations ne reflètent pas le même comportement ; une augmentation du revenu provoque un accroissement proportionnel pour la première relation, alors que, dans la seconde, l'effet s'estompe avec l'augmentation du revenu (si $0 < a_1 < 1$). Nous appelons « forme fonctionnelle » ce choix (arbitraire ou fondé) de spécification précise du modèle. Dans notre exemple, le modèle explicité s'écrit :

$$\begin{aligned}C &= a_0 + a_1 Y && \text{avec } a_0 > 0 \text{ et } 0 < a_1 < 1 \\&&& a_1 = \text{propension marginale à consommer} \\&&& \text{et } a_0 = \text{consommation incompressible ;} \\I &= b_0 + b_1 r && \text{avec } b_0 > 0 \text{ et } b_1 < 0 ; \\Y &\equiv C + I + \bar{I}\end{aligned}$$

Les deux premières équations reflètent des relations de comportements alors que la troisième est une identité (aucun paramètre n'est à estimer).

3) Sélection et mesure des variables

Le modèle étant spécifié, il convient de collecter les variables représentatives des phénomènes économiques. Ce choix n'est pas neutre et peut conduire à des résultats différents, les questions qu'il convient de se poser sont par exemple :

- *Faut-il raisonner en euros constants ou en euros courants ?*
- *Les données sont-elles brutes ou CVS¹ ?*
- *Quel taux d'intérêt faut-il retenir (taux au jour le jour, taux directeur de la Banque Centrale Européenne,...) ? etc.*

1. Corrigées des Variations Saisonnières.

Nous distinguons plusieurs types de données selon que le modèle est spécifié en :

- *série temporelle* : c'est le cas le plus fréquent en économétrie, il s'agit de variables observées à intervalles de temps réguliers (la consommation annuelle, totale France, exprimée en euros courants sur 20 ans) ;
- *coupe instantanée* : les données sont observées au même instant et concernent les valeurs prises par la variable pour un groupe d'individus¹ spécifiques (consommation observée des agriculteurs pour une année donnée) ;
- *panel* : la variable représente les valeurs prises par un échantillon d'individus à intervalles réguliers (la consommation d'un échantillon de ménages de la région parisienne sur 20 ans) ;
- *cohorte* : très proches des données de panel, les données de cohorte se distinguent de la précédente par la constance de l'échantillon, les individus sondés sont les mêmes d'une période sur l'autre.

4) Décalages temporels

Dans le cadre de modèle spécifié en séries temporelles, les relations entre les variables ne sont pas toujours synchrones mais peuvent être décalées dans le temps. Nous pouvons concevoir que la consommation de l'année t est expliquée par le revenu de l'année $t - 1$ et non celui de l'année t . Pour lever cette ambiguïté, il est d'usage d'écrire le modèle en le spécifiant à l'aide d'un indice de temps : $C_t = a_0 + a_1 Y_{t-1}$. La variable Y_{t-1} est appelée « variable endogène retardée ».

► On appelle « variable exogène » une variable dont les valeurs sont prédéterminées, et « variable endogène » une variable dont les valeurs dépendent des variables exogènes.

5) Validation du modèle

La dernière étape est celle de la validation² du modèle :

- Les relations spécifiées sont-elles valides ?
- Peut-on estimer avec suffisamment de précision les coefficients ?
- Le modèle est-il vérifié sur la totalité de la période ?
- Les coefficients sont-ils stables ? Etc.

À toutes ces questions, les techniques économétriques s'efforcent d'apporter des réponses.

1. Le terme d'individu est employé au sens statistique, c'est-à-dire comme un élément d'une population : une personne, une parcelle de terre...

2. Validation, c'est-à-dire en conformité avec les données disponibles.

II. Le rôle de l'économétrie

A. L'économétrie comme validation de la théorie

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

Une question se pose alors : pourquoi estimer ces relations, et les tester statistiquement ? Plusieurs raisons incitent à cette démarche : tout d'abord cela force l'individu à établir clairement et à estimer les interrelations sous-jacentes. Ensuite, la confiance aveugle dans l'intuition peut mener à l'ignorance de liaisons importantes ou à leur mauvaise utilisation. De plus, des relations marginales mais néanmoins explicatives, qui ne sont qu'un élément d'un modèle global, doivent être testées et validées afin de les mettre à leur véritable place. Enfin, il est nécessaire de fournir, en même temps que l'estimation des relations, une mesure de la confiance que l'économiste peut avoir en celles-ci, c'est-à-dire la précision que l'on peut en attendre. Là encore, l'utilisation de méthodes purement qualitatives exclut toute mesure quantitative de la fiabilité d'une relation.

B. L'économétrie comme outil d'investigation

L'économétrie n'est pas seulement un système de validation, mais également un outil d'analyse. Nous pouvons citer quelques domaines où l'économétrie apporte une aide à la modélisation, à la réflexion théorique ou à l'action économique par :

- la mise en évidence de relations entre des variables économiques qui n'étaient pas *a priori* évidentes ou pressenties ;
- l'induction statistique ou l'inférence statistique consiste à inférer, à partir des caractéristiques d'un échantillon, les caractéristiques d'une population. Elle permet de déterminer des intervalles de confiance pour des paramètres du modèle ou de tester si un paramètre est significativement¹ inférieur, supérieur ou simplement différent d'une valeur fixée ;

1. Au sens statistique, c'est-à-dire avec un seuil (risque d'erreur à ne pas dépasser, souvent 5 %).

- la simulation qui mesure l’impact de la modification de la valeur d’une variable sur une autre ($\Delta C_t = a_1 \Delta Y_t$) ;
- la prévision¹, par l’utilisation de modèles économétriques, qui est utilisée par les pouvoirs publics ou l’entreprise afin d’anticiper et éventuellement de réagir à l’environnement économique.

Dans cet ouvrage, nous nous efforcerons de montrer, à l’aide d’exemples, les différentes facettes de l’utilisation des techniques économétriques dans des contextes et pour des objectifs différents.

III. La théorie de la corrélation

A. Présentation générale

Lorsque deux phénomènes ont une évolution commune, nous disons qu’ils sont « corrélés ». La corrélation simple mesure le degré de liaison existant entre ces deux phénomènes représentés par des variables. Si nous cherchons une relation entre trois variables ou plus, nous ferons appel alors à la notion de corrélation multiple.

Nous pouvons distinguer la corrélation linéaire, lorsque tous les points du couple de valeurs (x, y) des deux variables semblent alignés sur une droite, de la corrélation non linéaire lorsque le couple de valeurs se trouve sur une même courbe d’allure quelconque.

Deux variables peuvent être :

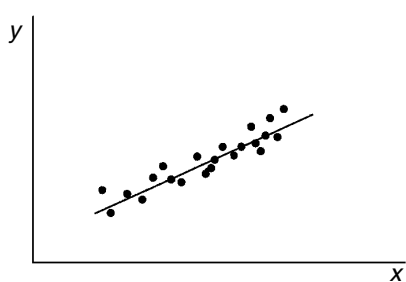
- en corrélation positive ; on constate alors une augmentation (ou diminution, ou constance) simultanée des valeurs des deux variables ;
- en corrélation négative, lorsque les valeurs de l’une augmentent, les valeurs de l’autre diminuent ;
- non corrélées, il n’y a aucune relation entre les variations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

Le tableau 1, en croisant les critères de linéarité et de corrélation, renvoie à une représentation graphique.

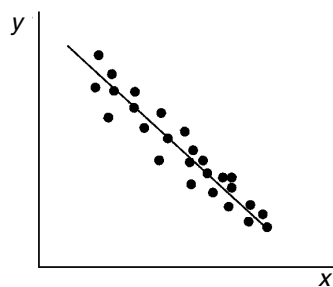
1. Pour découvrir l’utilisation de l’économétrie à des fins de prévision de ventes, voir Bourbonnais R. et Usunier J. C. (2013).

Tableau 1 – Linéarité et corrélation

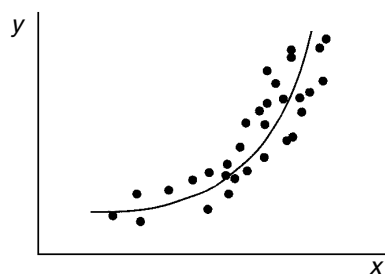
	Corrélation positive	Corrélation négative	Absence de corrélation
Relation linéaire	Graphe 1	Graphe 2	Graphe 5
Relation non linéaire	Graphe 3	Graphe 4	Graphe 5



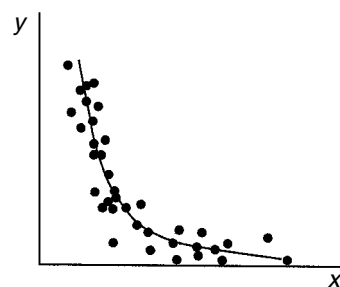
Graphe 1



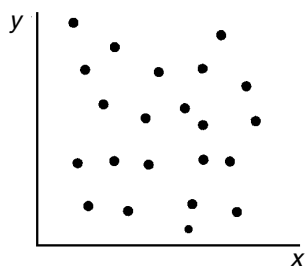
Graphe 2



Graphe 3



Graphe 4



Graphe 5

B. Mesure et limite du coefficient de corrélation

1) Le coefficient de corrélation linéaire

La représentation graphique ne donne qu'une « impression » de la corrélation entre deux variables sans donner une idée précise de l'intensité de la liaison, c'est pourquoi nous calculons une statistique appelée *coefficient de corrélation linéaire simple*, noté $r_{x,y}$. Il est égal à :

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

avec :

$\text{Cov}(x,y)$ = covariance entre x et y ;

σ_x et σ_y = écart type de x et écart type de y ;

n = nombre d'observations.

En développant la formule [1], il vient :

$$r_{x,y} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \sqrt{n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i\right)^2}} \quad [2]$$

On peut démontrer que, par construction ce coefficient reste compris entre -1 et 1 :

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

Dans la pratique, ce coefficient est rarement très proche de l'une de ces trois bornes et il est donc difficile de proposer une interprétation fiable à la simple lecture de ce coefficient. Ceci est surtout vrai en économie où les variables sont toutes plus au moins liées entre elles. De plus, il n'est calculé qu'à partir d'un échantillon d'observations et non pas sur l'ensemble des valeurs. On appelle $\rho_{x,y}$

ce coefficient empirique qui est une estimation du coefficient vrai $r_{x,y}$. La théorie des tests statistiques nous permet de lever cette indétermination.

Soit à tester l'hypothèse $H_0 : r_{x,y} = 0$, contre l'hypothèse $H_1 : r_{x,y} \neq 0$.

Sous l'hypothèse H_0 , nous pouvons démontrer que $\frac{\rho_{x,y}}{\sqrt{\frac{(1 - \rho_{x,y}^2)}{n - 2}}}$ suit une loi

de Student à $n - 2$ degrés de liberté¹. Nous calculons alors une statistique, appelé le t de Student empirique :

$$t^* = \frac{|\rho_{x,y}|}{\sqrt{\frac{(1 - \rho_{x,y}^2)}{n - 2}}} \quad [3]$$

Si $t^* > t_{n-2}^{\alpha/2}$ valeur lue dans une table de Student² au seuil $\alpha = 0,05$ (5 %) à $n - 2$ degrés de liberté³, nous rejetons l'hypothèse H_0 , le coefficient de corrélation est donc significativement différent de 0 ; dans le cas contraire, l'hypothèse d'un coefficient de corrélation nul est acceptée. La loi de Student étant symétrique, nous calculons la valeur absolue du t empirique et nous procédons au test par comparaison avec la valeur lue directement dans la table.

1. La notion de degrés de liberté est explicitée au chapitre 2.

2. Les lois de probabilité sont en fin d'ouvrage.

3. Si le nombre d'observations n est supérieur à 30, on peut approximer la loi de Student par une loi normale, soit $t^{\alpha/2} \approx 1,96$.

Exercice n° 1

📁 fichier C1EX1

Calcul d'un coefficient de corrélation

Un agronome s'intéresse à la liaison pouvant exister entre le rendement de maïs x (en quintal) d'une parcelle de terre et la quantité d'engrais y (en kilo). Il relève 10 couples de données consignés dans le tableau 2

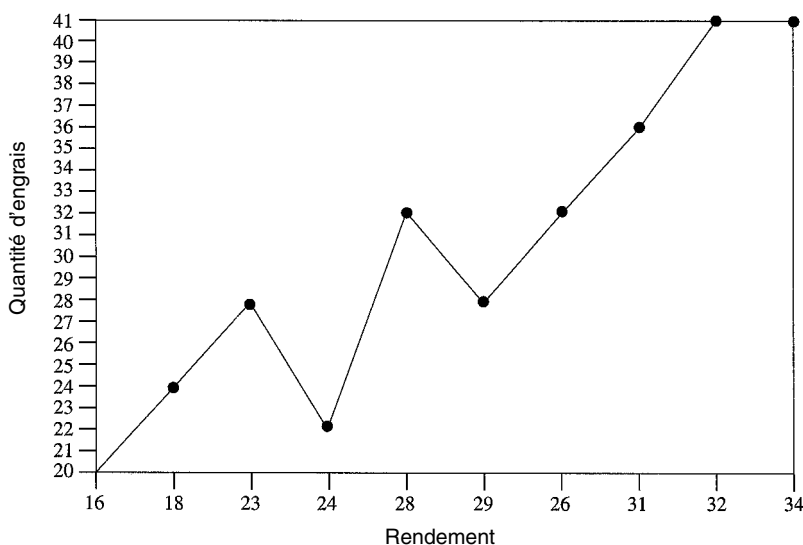
Tableau 2 – Rendement de maïs et quantité d'engrais

Rendement x	16	18	23	24	28	29	26	31	32	34
Engrais y	20	24	28	22	32	28	32	36	41	41

- 1) Tracer le nuage de points et le commenter.
- 2) Calculer le coefficient de corrélation simple et tester sa signification par rapport à 0 pour un seuil $\alpha = 0,05$.

Solution

- 1) Le nuage de points (graphique 6) indique que les couples de valeurs sont approximativement alignés : les deux variables semblent corrélées positivement.



Graphique 6 – Nuage du couple de valeurs :
rendement-quantité d'engrais

2) Afin d'appliquer la formule [2], nous dressons le tableau de calcul 3.

Tableau 3 – Calcul d'un coefficient de corrélation

	x	y	x^2	y^2	xy
	16	20	256	400	320
	18	24	324	576	432
	23	28	529	784	644
	24	22	576	484	528
	28	32	784	1 024	896
	29	28	841	784	812
	26	32	676	1 024	832
	31	36	961	1 296	1 116
	32	41	1 024	1 681	1 312
	34	41	1 156	1 681	1 394
Somme	261	304	7 127	9 734	8 286

$$\rho_{x,y} = \frac{(10)(8\,286) - (261)(304)}{\sqrt{(10)(7\,127) - 261^2} \sqrt{(10)(9\,734) - 304^2}} = \frac{3\,516}{(56,11)(70,17)}$$

soit $\rho_{x,y} = 0,89$ et $\rho_{x,y}^2 = 0,79$

Le t de Student empirique (d'après [3]) est égal à :

$$t^* = \frac{|\rho_{x,y}|}{\sqrt{\frac{(1 - \rho_{x,y}^2)}{n - 2}}} = \frac{0,89}{0,1620} = 5,49 > t_8^{0,025} = 2,306$$

le coefficient de corrélation entre x et y est significativement différent de 0.

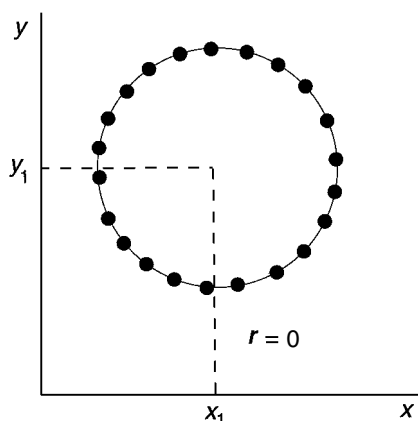
2) Limites de la notion de corrélation

a) La relation testée est linéaire

L'application de la formule [1] ou [2] ne permet de déterminer que des corrélations linéaires entre variables. Un coefficient de corrélation nul indique que la covariance entre la variable x et la variable y est égale à 0. C'est ainsi que deux variables en totale dépendance peuvent avoir un coefficient de corrélation nul, comme l'illustre l'exemple suivant : l'équation d'un cercle nous est donnée par $(x - x_1)^2 + (y - y_1)^2 = R^2$, les variables x et y sont bien liées entre elles fonctionnellement (graphique 7) et pourtant leur covariance est nulle et donc leur coefficient de corrélation égal à 0.

Pour pallier cette limite, il convient éventuellement de transformer les variables, préalablement au calcul du coefficient de corrélation, afin de linéariser

leur relation, par exemple au moyen d'une transformation de type logarithmique.



Graphique 7 – Relation fonctionnelle n'est pas corrélation linéaire

b) Corrélacion n'est pas causalité

Le fait d'avoir un coefficient de corrélation élevé entre deux variables ne signifie pas qu'il existe un autre lien que statistique. En d'autres termes, une covariance significativement différente de 0 n'implique pas une liaison d'ordre économique, physique ou autre. Nous appelons *corrélacion fortuite* ce type de corrélation que rien ne peut expliquer.

L'exemple le plus fameux concerne la forte corrélation existante entre le nombre de taches solaires observées et le taux de criminalité aux États-Unis. Cela ne signifie pas qu'il existe une relation entre les deux variables, mais qu'une troisième variable, l'évolution de long terme (la tendance) ici, explique conjointement les deux phénomènes. La théorie de la cointégration traite de ce problème (cf. chapitre 11).

2. Le modèle de régression simple

Nous commençons notre étude par le modèle le plus simple : une variable endogène est expliquée par une variable exogène. Après avoir étudié les conséquences probabilistes de l'erreur d'observation, nous présentons en I. les formules de base permettant d'estimer les paramètres du modèle. Les hypothèses stochastiques et leurs conséquences sont étudiées au paragraphe II. En III. et IV., la qualité de l'estimation d'un modèle est examinée à l'aide des premiers tests statistiques (Student, Fisher). Enfin, en V., le modèle de régression simple est étudié en tant qu'outil de prévision avec le degré de confiance que nous pouvons en attendre.

I. Présentation du modèle

A. Exemple introductif

Soit la fonction de consommation keynésienne :

$$C = a_0 + a_1 Y$$

où :

C = consommation,

Y = revenu,

a_1 = propension marginale à consommer,

a_0 = consommation autonome ou incompressible.

1) Vocabulaire

- La variable consommation est appelée « variable à expliquer » ou « variable endogène ».
- La variable revenu est appelée « variable explicative » ou « variable exogène » (c'est le revenu qui explique la consommation).
- a_1 et a_0 sont les paramètres du modèle ou encore les coefficients de régression.

2) Spécification

Nous pouvons distinguer deux types de spécifications :

- Les modèles en série temporelle, les variables représentent des phénomènes observés à intervalles de temps réguliers, par exemple la consommation et le revenu annuel sur 20 ans pour un pays donné. Le modèle s'écrit alors :

$$C_t = a_0 + a_1 Y_t \quad t = 1, \dots, 20$$

où :

C_t = consommation au temps t ,

Y_t = revenu au temps t .

- Les modèles en coupe instantanée, les variables représentent des phénomènes observés au même instant mais concernant plusieurs individus, par exemple la consommation et le revenu observés sur un échantillon de 20 pays. Le modèle s'écrit alors :

$$C_i = a_0 + a_1 Y_i \quad i = 1, \dots, 20$$

où :

C_i = consommation du pays i pour une année donnée,

Y_i = revenu du pays i pour une année donnée.

B. Rôle du terme aléatoire

Le modèle tel qu'il vient d'être spécifié n'est qu'une caricature de la réalité. En effet ne retenir que le revenu pour expliquer la consommation est à l'évidence même insuffisant ; il existe une multitude d'autres facteurs susceptibles d'expliquer la consommation. C'est pourquoi nous ajoutons un terme (ε_t) qui synthétise l'ensemble de ces informations non explicitées dans le modèle : $C_t = a_0 + a_1 Y_t + \varepsilon_t$ si le modèle est spécifié en série temporelle ($C_i = a_0 + a_1 Y_i + \varepsilon_i$ si le modèle est spécifié en coupe instantanée), où ε_t représente l'erreur de spécification du modèle, c'est-à-dire l'ensemble des phénomènes explicatifs de la consommation non liés au revenu. Le terme ε_t mesure la

différence entre les valeurs réellement observées de C_t et les valeurs qui auraient été observées si la relation spécifiée avait été rigoureusement exacte. Le terme ε_t regroupe donc trois erreurs :

- une erreur de spécification, c'est-à-dire le fait que la seule variable explicative n'est pas suffisante pour rendre compte de la totalité du phénomène expliqué ;
- une erreur de mesure, les données ne représentent pas exactement le phénomène ;
- une erreur de fluctuation d'échantillonnage, d'un échantillon à l'autre les observations, et donc les estimations, sont légèrement différentes.

Exercice n° 1

↓ fichier C2EX1

Génération d'une consommation aléatoire

Le tableau 1 présente le revenu moyen par habitant sur 10 ans exprimé en dollars pour un pays.

Tableau 1 – Évolution du revenu moyen par habitant en dollars

Année	Revenu
1	8 000
2	9 000
3	9 500
4	9 500
5	9 800
6	11 000
7	12 000
8	13 000
9	15 000
10	16 000

Sachant que la propension marginale à consommer est de 0,8 et que la consommation incompressible est 1 000, on demande :

- 1) de calculer la consommation théorique sur les 10 ans ;
- 2) considérant que notre erreur d'observation suit une loi normale de moyenne 0 et de variance 20 000, de générer cette variable aléatoire et de calculer une consommation observée tenant compte de cette erreur.

Solution

Les calculs des questions 1) et 2) sont présentés dans le tableau 2.

La consommation théorique (colonne 3) est calculée par application directe de la formule : $C_t = 1\,000 + 0,8 Y_t$.

La génération de la variable aléatoire $\varepsilon_t (\varepsilon_t \rightarrow N(0; 20\,000))$ ne pose pas de difficulté particulière ; bien entendu il en existe une infinité, un exemple en est présenté en colonne 4.

La consommation « observée » (colonne 5) est donc égale à $C_t = 1\,000 + 0,8 Y_t + \varepsilon_t$, soit la somme de la colonne 3 et de la colonne 4.

Tableau 2 – Calcul de la consommation observée


(1) Année	(2) Revenu disponible	(3) Consommation théorique	(4) Aléa ε_t	(5) Consommation observée
1	8 000	7 400	– 10,01	7 389,99
2	9 000	8 200	– 30,35	8 169,65
3	9 500	8 600	231,71	8 831,71
4	9 500	8 600	52,84	8 652,84
5	9 800	8 840	– 51,92	8 788,08
6	11 000	9 800	– 183,79	9 616,21
7	12 000	10 600	– 6,55	10 593,45
8	13 000	11 400	– 213,89	11 186,11
9	15 000	13 000	– 241,91	12 758,09
10	16 000	13 800	69,62	13 869,62
Moyenne :			– 38,42	
Écart type :			137,24	

Nous observons que la moyenne de ε_t , $\bar{\varepsilon} = -38,42$ et la variance de ε_t , $\text{Var}(\varepsilon_t) = 18\,834,81$ sont légèrement différentes des valeurs théoriques. Cela est la conséquence du tirage particulier d'un échantillon de taille assez faible (dix observations).

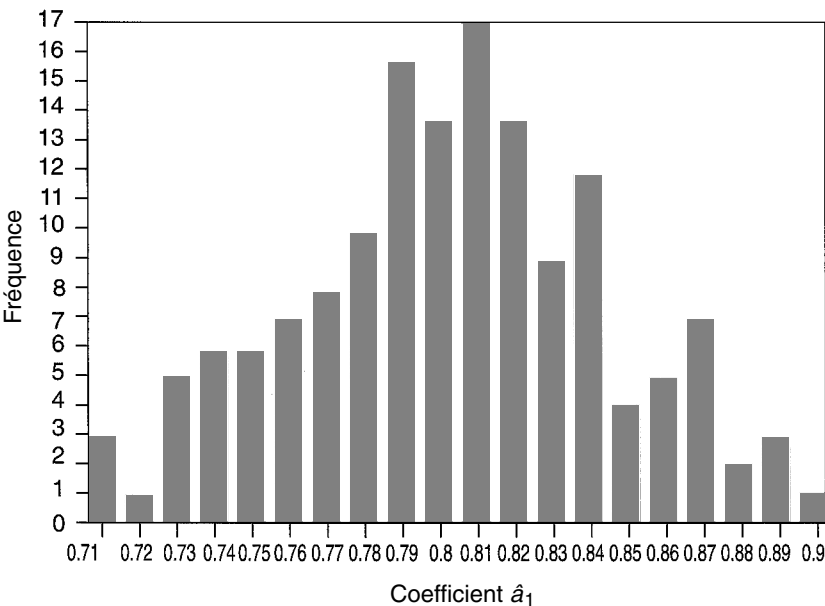
C. Conséquences du terme aléatoire

Dans l'exercice précédent, les valeurs vraies a_0 et a_1 sont parfaitement connues, cependant, dans la réalité, nous ne connaissons pas ces valeurs mais seulement les deux séries d'observations C_t et R_t . Les estimateurs¹ de a_0 et a_1 , notés respectivement \hat{a}_0 et \hat{a}_1 , sont des variables aléatoires, qui suivent les mêmes lois de probabilité, celle de ε_t , puisqu'ils sont fonctions de la variable aléatoire ε_t . Les caractéristiques de moyenne et d'écart type de ces coefficients permettent de construire des tests de validité du modèle estimé.

1. Il ne faut pas confondre : estimateur \hat{a} de a et estimation de \hat{a} qui est la valeur particulière de l'estimateur pour un échantillon.

Si à l'aide d'un programme informatique ( C2EX1.PRG) nous construisons 150 échantillons de valeurs différentes pour la variable aléatoire $\varepsilon_t \rightarrow N(0; 20\,000)$, nous allons alors trouver 150 estimations de a_1 légèrement différentes entre elles dont la distribution aura la forme de l'histogramme du graphique 1.

Cette distribution a pour moyenne 0,801 et écart type 0,032, nous pouvons observer qu'elle est à peu près symétrique par rapport à la moyenne et qu'elle a la forme d'une courbe en « cloche » : tous ces éléments suggèrent bien une distribution normale de : $a_1 \rightarrow N(0,801; 0,032)$. Cela est la conséquence directe de la normalité des erreurs.



Graphique 1 – Histogramme de la distribution de 150 \hat{a}_1

II. Estimation des paramètres

A. Modèle et hypothèses

Soit le modèle suivant :

$$y_t = a_0 + a_1 x_t + \varepsilon_t \quad \text{pour} \quad t = 1, \dots, n$$

avec :

y_t = variable à expliquer au temps t ;

x_t = variable explicative au temps t ;

a_0, a_1 = paramètres du modèle ;

ε_t = erreur de spécification (différence entre le modèle vrai et le modèle spécifié), cette erreur est inconnue et restera inconnue ;

n = nombre d'observations.

Hypothèses

- H1 : le modèle est linéaire en x_t (ou en n'importe quelle transformation de x_t).
- H2 : les valeurs x_t sont observées sans erreur (x_t non aléatoire).
- H3 : $E(\varepsilon_t) = 0$, l'espérance mathématique de l'erreur est nulle : en moyenne le modèle est bien spécifié et donc l'erreur moyenne est nulle.
- H4 : $E(\varepsilon_t^2) = \sigma_\varepsilon^2$, la variance de l'erreur est constante¹ : le risque de l'amplitude de l'erreur est le même quelle que soit la période.
- H5 : $E(\varepsilon_t \varepsilon_{t'}) = 0$ si $t \neq t'$, les erreurs sont non corrélées (ou encore indépendantes) : une erreur à l'instant t n'a pas d'influence sur les erreurs suivantes.
- H6 : $\text{Cov}(x_t, \varepsilon_t) = 0$, l'erreur est indépendante de la variable explicative.

B. Formulation des estimateurs

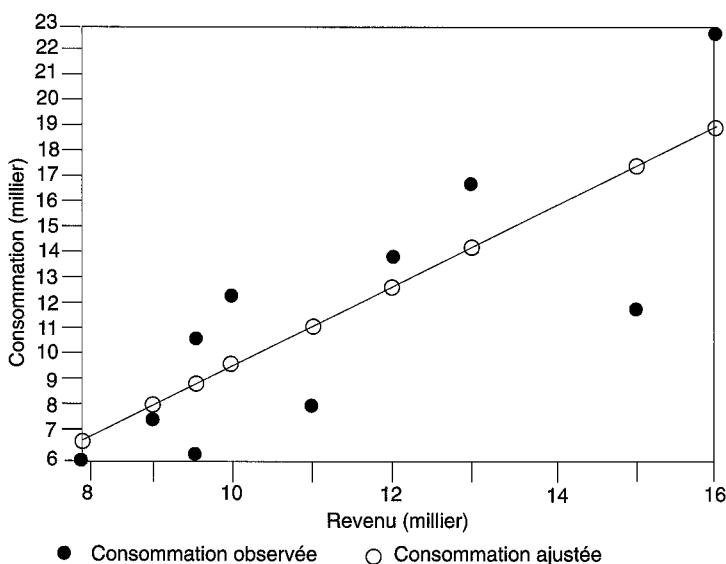
En traçant un graphique (2) des couples de données liant le revenu et la consommation observée, nous obtenons un nuage de points que nous pouvons ajuster à l'aide d'une droite.

L'estimateur des coefficients a_0 et a_1 est obtenu en minimisant la distance au carré entre chaque observation et la droite, d'où le nom d'estimateur des moindres carrés ordinaires (MCO).

La résolution analytique est la suivante :

$$\text{Min} \sum_{t=1}^{t=n} \varepsilon_t^2 = \text{Min} \sum_{t=1}^{t=n} (y_t - a_0 - a_1 x_t)^2 = \text{Min } S$$

1. Cette hypothèse s'appelle l'hypothèse d'homoscédasticité ; dans le cas où cette hypothèse n'est pas vérifiée, on parle alors de modèle hétéroscédastique.



Graphique 2 – Ajustement d'un nuage de points par une droite

En opérant par dérivation par rapport à a_0 et a_1 afin de trouver le minimum¹ de cette fonction, on obtient les résultats suivants :

$$\frac{\delta S}{\delta a_0} = -2 \sum_t (y_t - \hat{a}_0 - \hat{a}_1 x_t) = 0 \quad \text{et} \quad \frac{\delta S}{\delta a_1} = -2 \sum_t x_t (y_t - \hat{a}_0 - \hat{a}_1 x_t) = 0$$

Sommant par rapport à t , il vient :

$$\begin{aligned} \sum_t x_t y_t - \hat{a}_0 \sum_t x_t - \hat{a}_1 \sum_t x_t^2 &= 0 \\ \sum_t y_t - n \hat{a}_0 - \hat{a}_1 \sum_t x_t &= 0 \end{aligned}$$

qu'on appelle les équations normales et qui impliquent que :

$$\hat{a}_1 = \frac{\sum_{t=1}^{t=n} (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{t=n} (x_t - \bar{x})^2} = \frac{\sum_{t=1}^{t=n} x_t y_t - n \bar{x} \bar{y}}{\sum_{t=1}^{t=n} x_t^2 - n \bar{x}^2} \quad [1]$$

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x} \quad [2]$$

1. Nous considérons les conditions du deuxième ordre comme vérifiées car la fonction est convexe.

Nous pouvons faire deux remarques :

– l'écriture du modèle n'est pas neutre ;

La spécification : $y_t = a_0 + a_1 x_t + \varepsilon_t$ n'est pas équivalente à : $x_t = a'_0 + a'_1 y_t + \varepsilon'_t$. Dans le premier modèle, x_t est la cause de y_t , alors que dans la deuxième spécification, c'est y_t qui est la cause de x_t .

Nous remarquons que : $\hat{a}_1 \times \hat{a}'_1 = \rho^2$
(ρ = coefficient de corrélation entre x et y) ;

– le coefficient a_1 représente la pente de la droite ou encore une propension marginale.

Nous avons : $\Delta y_t = \hat{a}_1 \Delta x_t$. L'impact d'une variation de x_t se mesure directement sur y_t au travers du coefficient \hat{a}_1 .

Cas particulier : modèle sans terme constant

La théorie économique postule parfois des relations dans lesquelles $a_0 = 0$: c'est le cas par exemple pour une fonction de production de produit industriel où le facteur de production (unique) nul entraîne une production nulle. L'estimation de a_1 est alors donnée par la formule suivante :

$$\hat{a}_1 = \frac{\sum_{t=1}^{t=n} x_t y_t}{\sum_{t=1}^{t=n} x_t^2} \quad [3]$$

Nous remarquons qu'il s'agit de l'application de la formule [1] dans laquelle \bar{x} et \bar{y} sont nulles. Dans le cas de variables centrées¹, c'est donc cette formule [3] qu'il convient d'employer car le terme constant est nul.

Exercice n° 2

 **fichier C2EX2**

Estimation des coefficients de régression

À partir des données du tableau 2 de l'exercice 1, on demande de calculer les estimations de \hat{a}_0 et \hat{a}_1 .

Solution

Le tableau 3 présente les calculs à effectuer.

1. Les données sont centrées lorsque les observations sont centrées sur leur moyenne : $(x_t - \bar{x})$, la somme des données centrées est donc par construction nulle.

Tableau 3 – Calcul d'un coefficient de régression

(1) t	(2) y_t	(3) x_t	(4) $y_t - \bar{y}$	(5) $x_t - \bar{x}$	(6) (5)* (5)	(7) (4)* (5)
1	7 389,99	8 000	- 2 595,59	- 3 280	10 758 400	8 513 518
2	8 169,65	9 000	- 1 815,93	- 2 280	5 198 400	4 140 300
3	8 831,71	9 500	- 1 153,87	- 1 780	3 168 400	2 053 879
4	8 652,84	9 500	- 1 332,74	- 1 780	3 168 400	2 372 268
5	8 788,08	9 800	- 1 197,50	- 1 480	2 190 400	1 772 292
6	9 616,21	11 000	- 369,37	- 280	78 400	103 422
7	10 593,45	12 000	607,88	720	518 400	437 670
8	11 186,11	13 000	1 200,54	1 720	2 958 400	2 064 920
9	12 758,09	15 000	2 772,52	3 720	13 838 400	10 313 755
10	13 869,62	16 000	3 884,05	4 720	22 278 400	18 332 692
Somme	99 855,75	112 800	0	0	64 156 000	50 104 729
Moyenne	9 985,57	11 280	0	0	6 415 600	5 010 472

$$\hat{a}_1 = \frac{\sum_{t=1}^{t=n} (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{t=n} (x_t - \bar{x})^2} = \frac{50\,104\,729}{64\,156\,000} = 0,78$$

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x} = 9\,985,57 - 0,78 \times 11\,280 = 1\,176,08$$

Ces estimations sont à comparer aux valeurs vraies (respectivement 0,8 et 1 000), les différences importantes en ce qui concerne surtout le terme constant sont imputables à l'aléa d'observation qui « perturbe » l'estimation des coefficients.

C. Les différentes écritures du modèle : erreur et résidu

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

- Modèle théorique spécifié par l'économiste avec ε_t l'erreur inconnue :

$$y_t = a_0 + a_1 x_t + \varepsilon_t$$

- Modèle estimé à partir d'un échantillon d'observations :

$$y_t = \hat{a}_0 + \hat{a}_1 x_t + e_t = \hat{y}_t + e_t \quad e_t = \text{résidu}$$

(Bien noter les « chapeaux » sur les a .)

Le résidu observé e_t est donc la différence entre les valeurs observées de la variable à expliquer et les valeurs ajustées à l'aide des estimations des coefficients du modèle ;

ou encore :

$$\hat{y}_t = \hat{a}_0 + \hat{a}_1 x_t$$

D. Propriétés des estimateurs

$$y_t = a_0 + a_1 x_t + \varepsilon_t \quad [4]$$

$$\bar{y} = a_0 + a_1 \bar{x} + \bar{\varepsilon} \quad [5]$$

$$[4] - [5] = y_t - \bar{y} = a_1(x_t - \bar{x}) + (\varepsilon_t - \bar{\varepsilon})$$

En remplaçant dans la formule [1] de l'estimateur \hat{a}_1 , $y_t - \bar{y}$ par son expression, on obtient :

$$\hat{a}_1 = a_1 + \frac{\sum_{t=1}^{t=n} (x_t - \bar{x})(\varepsilon_t - \bar{\varepsilon})}{\sum_{t=1}^{t=n} (x_t - \bar{x})^2} = a_1 + \frac{\sum_{t=1}^{t=n} (x_t - \bar{x})\varepsilon_t}{\sum_{t=1}^{t=n} (x_t - \bar{x})^2} \quad [6]$$

$$\text{car } \bar{\varepsilon} \sum_{t=1}^n (x_t - \bar{x}) = \bar{\varepsilon} \sum_{t=1}^n x_t - \bar{\varepsilon} n \bar{x} = \bar{\varepsilon} n \bar{x} - \bar{\varepsilon} n \bar{x} = 0 \quad \text{avec } \bar{x} = \frac{\sum_{t=1}^n x_t}{n}$$

- Les estimateurs sont-ils sans biais ?

(Rappel : un estimateur est sans biais si $E(\hat{a}) = a$.)

$$\hat{a}_1 = a_1 + \frac{\sum_{t=1}^{t=n} (x_t - \bar{x})\varepsilon_t}{\sum_{t=1}^{t=n} (x_t - \bar{x})^2} \quad [7]$$

$$\text{d'où :} \quad E(\hat{a}_1) = E(a_1) + \frac{\sum_{t=1}^{t=n} (x_t - \bar{x})E(\varepsilon_t)}{\sum_{t=1}^{t=n} (x_t - \bar{x})^2}$$

Soit $E(\hat{a}_1) = a_1$ car $E(\varepsilon_t) = 0$

De même on démontre que $E(\hat{a}_0) = a_0$.

$$\begin{aligned}\bar{y} &= \hat{a}_0 + \hat{a}_1 \bar{x} \\ \bar{y} &= a_0 + a_1 \bar{x} + \bar{\varepsilon} \quad \rightarrow \quad \hat{a}_0 = a_0 + \bar{\varepsilon} - (\hat{a}_1 - a_1) \bar{x} \\ E(\hat{a}_0) &= a_0 + E(\bar{\varepsilon}) - E((\hat{a}_1 - a_1) \bar{x}) = a_0 \\ \text{car } E(\hat{a}_1 - a_1) &= 0 \quad \text{et } E(\bar{\varepsilon}) = 0\end{aligned}$$

car $E(\bar{\varepsilon})$ est par hypothèse nulle, $E\left(\sum_{t=1}^n \varepsilon_t\right) = 0$

► *Les estimateurs sont sans biais.*

• Les estimateurs sont-ils convergents ?

Puisque les estimateurs sont sans biais, il suffit pour qu'ils soient convergents que :

$\lim V(\hat{a}_1) \rightarrow 0$ lorsque $n \rightarrow \infty$ où $V(\hat{a}_1)$ est la variance de \hat{a}_1 .

$$V(\hat{a}_1) = E\{\hat{a}_1 - E(\hat{a}_1)\}^2 = E(\hat{a}_1 - a_1)^2 = E \left[\left(\frac{\sum_{t=1}^{t=n} (x_t - \bar{x}) \varepsilon_t}{\sum_{t=1}^{t=n} (x_t - \bar{x})^2} \right)^2 \right] \quad (\text{d'après [7]})$$

$$V(\hat{a}_1) = E \left[\left(\sum_t w_t \varepsilon_t \right)^2 \right] = E \left[\sum_t w_t^2 \varepsilon_t^2 + 2 \sum_{t < t'} w_t w_{t'} \varepsilon_t \varepsilon_{t'} \right]$$

avec $w_t = \frac{(x_t - \bar{x})}{\sum_{t=1}^{t=n} (x_t - \bar{x})^2}$

$$V(\hat{a}_1) = \sum_t w_t^2 E(\varepsilon_t^2) + 2 \sum_{t < t'} w_t w_{t'} E(\varepsilon_t \varepsilon_{t'})$$

Or, d'après les hypothèses :

– H4 $E(\varepsilon_t^2) = \sigma_\varepsilon^2$;

– H5 $E(\varepsilon_t \varepsilon_{t'}) = 0$ si $t \neq t'$.

$$\boxed{V(\hat{a}_1) = \sum_t w_t^2 \sigma_\varepsilon^2 = \frac{\sigma_\varepsilon^2}{\sum_t (x_t - \bar{x})^2}} \quad [8]$$

Lorsque $n \rightarrow \infty$ alors $\sum_t (x_t - \bar{x})^2$ tend également vers ∞ , d'où $V(\hat{a}_1)$ tend vers 0 puisque σ_ε^2 est constant.

Nous pouvons observer sur l'expression de la variance de \hat{a}_1 que cet estimateur est d'autant plus précis que $\sum_t (x_t - \bar{x})^2$ est élevé, c'est-à-dire lorsque :

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

Une démonstration analogue pour \hat{a}_0 conduit aux résultats suivant :

$$V(\hat{a}_0) = \sigma_\varepsilon^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_t (x_t - \bar{x})^2} \right) \quad [9]$$

$\lim V(\hat{a}_0) \rightarrow 0$ lorsque $n \rightarrow \infty$.

Nous remarquons que :

$$V(\hat{a}_0) = \frac{\sigma_\varepsilon^2}{n} + \bar{x}^2 V(\hat{a}_1) \quad \text{Cov}(\hat{a}_0, \hat{a}_1) = -\bar{x} V(\hat{a}_1)$$

► Les estimateurs sont convergents.

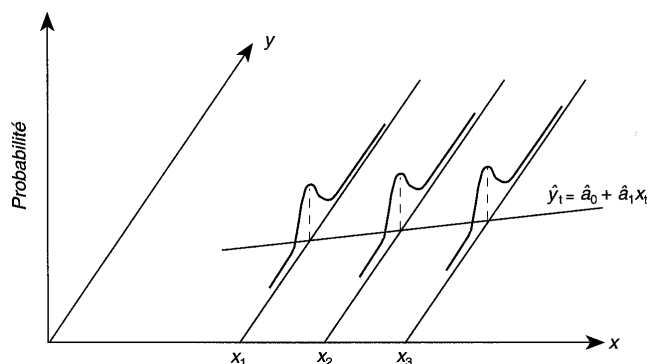
III. Conséquences des hypothèses : construction des tests

A. Hypothèse de normalité des erreurs

Nous allons maintenant introduire de nouveau l'hypothèse qui est celle de la normalité des erreurs (cf. exercice 1). Cette hypothèse n'est pas indispensable afin d'obtenir des estimateurs convergents mais elle va nous permettre de construire des tests statistiques¹ concernant la validité du modèle estimé.

Soit $\varepsilon_t \rightarrow N(0, \sigma_\varepsilon^2)$

Le graphique 3 illustre cette hypothèse sur la distribution des valeurs projetées sur la droite des moindres carrés.



Graphique 3 – Les erreurs suivent une loi normale

1. En effet, cette hypothèse permet de définir la loi de probabilité des estimateurs.

B. Conséquences de l'hypothèse de normalité des erreurs

En préliminaire, cherchons un estimateur de la variance de l'erreur.

Le résidu est donné par :

$$e_t = y_t - \hat{y}_t = y_t - \hat{a}_1 x_t - \hat{a}_0$$

Sans rien changer à l'expression précédente, nous pouvons écrire :

$$e_t = y_t - \hat{a}_1 x_t - \hat{a}_0 - \hat{a}_1 \bar{x} + \hat{a}_1 \bar{x}$$

Ou encore : $e_t = y_t - \hat{a}_1 \bar{x} - \hat{a}_0 - \hat{a}_1 x_t + \hat{a}_1 \bar{x}$

Or $\bar{y} = \hat{a}_1 \bar{x} + \hat{a}_0$

D'où : $e_t = y_t - \bar{y} - \hat{a}_1 (x_t - \bar{x})$

En remplaçant y_t et \bar{y} par leurs expressions [4] et [5], il vient :

$$e_t = (a_1 - \hat{a}_1)(x_t - \bar{x}) + (\varepsilon_t - \bar{\varepsilon})$$

Nous obtenons en élevant ce terme au carré et effectuant la somme sur les n observations :

$$\sum_t e_t^2 = (a_1 - \hat{a}_1)^2 \sum_t (x_t - \bar{x})^2 + \sum_t (\varepsilon_t - \bar{\varepsilon})^2 + 2(a_1 - \hat{a}_1) \sum_t (x_t - \bar{x})(\varepsilon_t - \bar{\varepsilon})$$

Or d'après l'expression [6], nous avons :

$$\sum_t (x_t - \bar{x})(\varepsilon_t - \bar{\varepsilon}) = -(a_1 - \hat{a}_1) \sum_t (x_t - \bar{x})^2$$

que nous remplaçons dans l'expression précédente et après simplification :

$$\sum_t e_t^2 = \sum_t (\varepsilon_t - \bar{\varepsilon})^2 - (a_1 - \hat{a}_1)^2 \sum_t (x_t - \bar{x})^2$$

En prenant l'espérance mathématique de cette dernière expression :

$$E\left[\sum_t e_t^2\right] = E\left[\sum_t (\varepsilon_t - \bar{\varepsilon})^2\right] - E[(a_1 - \hat{a}_1)^2] \sum_t (x_t - \bar{x})^2$$

Examinons les deux membres de cette équation.

$$\begin{aligned} \bullet \text{ a) } E\left[\sum_t (\varepsilon_t - \bar{\varepsilon})^2\right] &= E\left[\sum_t (\varepsilon_t^2 - 2\bar{\varepsilon}\varepsilon_t + \bar{\varepsilon}^2)\right] = E\left[\sum_t \varepsilon_t^2 - 2\bar{\varepsilon} \sum_t \varepsilon_t + \sum_t \bar{\varepsilon}^2\right] \\ &= E\left[\sum_t \varepsilon_t^2 - 2\bar{\varepsilon}n\bar{\varepsilon} + n\bar{\varepsilon}^2\right] = E\left[\sum_t \varepsilon_t^2 - 2n\bar{\varepsilon}^2 + n\bar{\varepsilon}^2\right] \\ &= E\left[\sum_t \varepsilon_t^2 - n\bar{\varepsilon}^2\right] = E\left[\sum_t \varepsilon_t^2 - \bar{\varepsilon} \sum_t \varepsilon_t\right] \\ &= E\left[\sum_t \varepsilon_t^2 - \frac{\left[\sum_t \varepsilon_t\right]^2}{n}\right] = \left[\sum_t E(\varepsilon_t^2) - \frac{1}{n}E\left(\sum_t \varepsilon_t\right)^2\right] \end{aligned}$$

Nous savons que : $E[\varepsilon_i^2] = \sigma_\varepsilon^2$

$$E\left[\sum_t (\varepsilon_t - \bar{\varepsilon})^2\right] = n\sigma_\varepsilon^2 - \frac{1}{n}E(\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_n)^2$$

Or, d'après l'hypothèse d'indépendance des erreurs, les doubles produits sont donc tous nuls. Nous obtenons alors :

$$E\left[\sum_t (\varepsilon_t - \bar{\varepsilon})^2\right] = n\sigma_\varepsilon^2 - \frac{1}{n}n\sigma_\varepsilon^2 = n\sigma_\varepsilon^2 - \sigma_\varepsilon^2$$

• b) $E[(a_1 - \hat{a}_1)^2] \sum_t (x_t - \bar{x})^2 = \sigma_\varepsilon^2$

En effet $E[(a_1 - \hat{a}_1)^2] = \frac{\sigma_\varepsilon^2}{\sum_t (x_t - \bar{x})^2}$

D'où :

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

L'estimateur de la variance de l'erreur (σ_ε^2) noté $\hat{\sigma}_\varepsilon^2$ est donc égal à :

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n-2} \sum_t e_t^2 \quad [10]$$

Ce qui nous permet de définir, en remplaçant la variance des erreurs par son estimateur dans les expressions [8] et [9], les estimateurs empiriques¹ de la variance de chacun des coefficients.

$$\begin{aligned} \hat{\sigma}_{\hat{a}_1}^2 &= \frac{\hat{\sigma}_\varepsilon^2}{\sum_t (x_t - \bar{x})^2} \\ \hat{\sigma}_{\hat{a}_0}^2 &= \hat{\sigma}_\varepsilon^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_t (x_t - \bar{x})^2} \right) \end{aligned} \quad [11]$$

L'hypothèse de normalité des erreurs implique que :

$$\frac{\hat{a}_1 - a_1}{\sigma_{\hat{a}_1}} \quad \text{et} \quad \frac{\hat{a}_0 - a_0}{\sigma_{\hat{a}_0}}$$

suivent une loi normale centrée réduite $N(0, 1)$.

1. Bien noter les chapeaux qui différencient les estimateurs théoriques (inconnus) des estimations empiriques (calculables).

$$\frac{\sum_{t=1}^n e_t^2}{\sigma_\varepsilon^2} = (n-2) \frac{\widehat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}$$

suit une loi du χ^2 (chi-deux) à $n-2$ degrés de liberté¹ (somme au carré de $n-2$ variables aléatoires *indépendantes* normales centrées réduites).

$$\text{Nous pouvons remarquer que } (n-2) \frac{\widehat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} = (n-2) \frac{\widehat{\sigma}_a^2}{\sigma_a^2} = \frac{\sum_{t=1}^n e_t^2}{\sigma_a^2 \sum_t (x_t - \bar{x})^2}$$

suit donc aussi une loi du χ^2 (chi-deux) à $n-2$ degrés de liberté.

Il en résulte que :

$$\Rightarrow \frac{\widehat{a}_1 - a_1}{\widehat{\sigma}_{\widehat{a}_1}} \text{ et } \frac{\widehat{a}_0 - a_0}{\widehat{\sigma}_{\widehat{a}_0}} \quad (\text{l'écart type théorique est remplacé par l'écart type empirique})$$

suivent une loi de Student à $n-2$ degrés de liberté.

$$\text{En effet } \frac{\widehat{a}_1 - a_1}{\widehat{\sigma}_{\widehat{a}_1}} = \frac{\frac{\widehat{a}_1 - a_1}{\sigma_{\widehat{a}_1}}}{\sqrt{(n-2) \frac{\widehat{\sigma}_{\widehat{a}_1}^2}{\sigma_{\widehat{a}_1}^2} \frac{1}{(n-2)}}} \text{ est le rapport d'une loi normale}$$

centrée réduite à la racine carrée d'un chi-deux divisé par son degré de liberté.

Il est donc possible maintenant de mettre en place des tests statistiques afin d'apporter des réponses à des problèmes tels que :

- comparaison d'un coefficient de régression par rapport à une valeur fixée ;
- comparaison de deux coefficients de régression provenant de deux échantillons différents ;
- détermination d'un intervalle de confiance pour un coefficient.

C. Test bilatéral, test unilatéral et probabilité critique d'un test

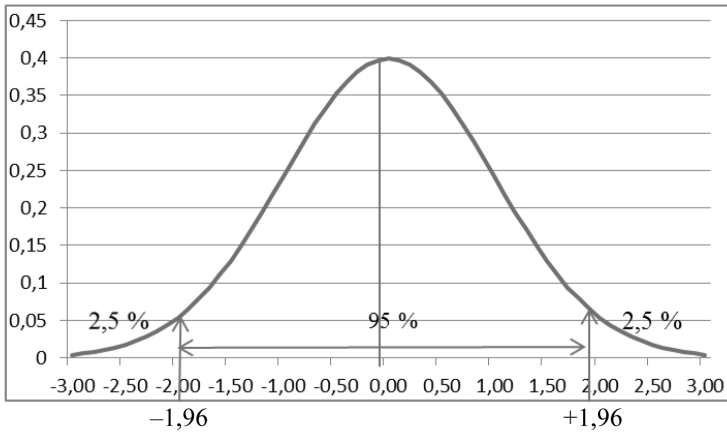
1) Test bilatéral

Soit à tester, à un seuil de 5 %, l'hypothèse $H_0 : a_1 = 0$ contre l'hypothèse $H_1 : a_1 \neq 0$.

Nous savons que $\frac{\widehat{a}_1 - a_1}{\widehat{\sigma}_{\widehat{a}_1}}$ suit une loi de Student à $n-2$ degrés de liberté.

1. La notion de degré de liberté correspond au nombre de valeurs restant réellement à disposition après une procédure d'estimation statistique. Si un échantillon comprend 10 observations et qu'on dispose en plus de la moyenne de cet échantillon, on ne peut choisir librement les valeurs que pour 9 de ces observations, la dixième se déduisant de la valeur de la moyenne. Dans le cas présent, le modèle de régression simple, le nombre de degrés de liberté est donc de $n-2$ car nous avons estimé deux paramètres a_0 et a_1 .

Sous H_0 ($a_1 = 0$) le ratio appelé ratio de Student $\frac{\hat{a}_1 - 0}{\hat{\sigma}_{\hat{a}_1}}$ suit donc une loi de Student à $n - 2$ degrés de liberté. Le test d'hypothèses bilatéral consiste donc à comparer le ratio de Student empirique $t^* = \frac{|\hat{a}_1|}{\hat{\sigma}_{\hat{a}_1}}$ à la valeur du t de Student lue dans la table à $n - 2$ degrés de liberté¹ et pour un seuil de probabilité égal à 5 %, soit si $n - 2 > 30$, $t_{\infty}^{0,05} = 1,96$ (la table 2. de Student en fin du livre est tabulée pour les tests bilatéraux). Si $t^* > t_{\infty}^{0,05} = 1,96$, nous rejetons l'hypothèse H_0 (cf. graphique 4), le coefficient théorique et inconnu a_1 est significativement différent de 0.



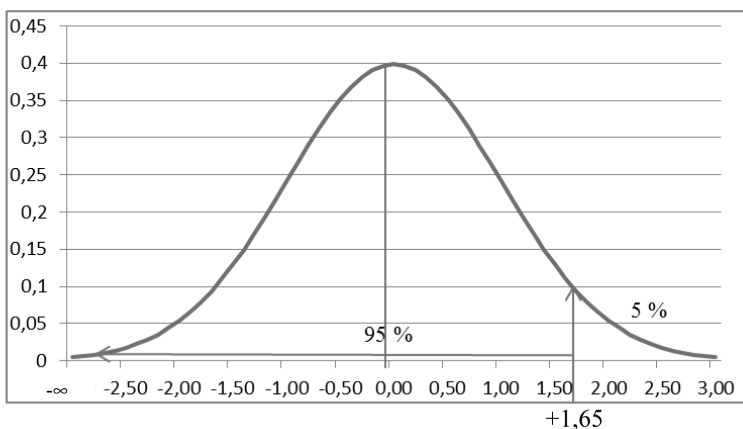
Graphique 4 – Test bilatéral à 5 %

2) Test unilatéral

Soit à tester, à un seuil de 5 %, l'hypothèse $H_0 : a_1 = 0$ contre l'hypothèse $H_1 : a_1 > 0$ ou $a_1 < 0$ selon que le coefficient estimé soit positif ou négatif.

Le test d'hypothèses unilatéral consiste donc à comparer le ratio de Student empirique $t^* = \frac{|\hat{a}_1|}{\hat{\sigma}_{\hat{a}_1}}$ à la valeur du t de Student lue dans la table à $n - 2$ degrés de liberté et pour un seuil de probabilité égal à 5 %, soit si $n - 2 > 30$, $t_{\infty}^{0,05} = 1,65$. Si $t^* > t_{\infty}^{0,05} = 1,65$ nous rejetons l'hypothèse H_0 (cf. graphique 5), le coefficient théorique et inconnu a_1 est significativement différent de 0. Attention, la table de Student en fin du livre est tabulée pour les tests bilatéraux, il faut donc lire à $10 \% = 2 \times 0,05$.

1. Si le degré de liberté est supérieur à 30, la loi de Student peut être approximée par une loi normale.



Graphique 5 – Test unilatéral à 5 % ($H1 : a_1 > 0$)

Remarque : si nous rejetons l’hypothèse H_0 pour un test bilatéral, alors nous rejetons forcément (pour un même seuil de probabilité) l’hypothèse H_0 pour un test unilatéral.

3) Probabilité critique d’un test

La probabilité critique – risque de rejeter à tort l’hypothèse H_0 – ou encore *risque de première espèce* est donnée par la valeur de la probabilité α^c telle que :

$$t_{n-2}^{\alpha^c} = t^* = \frac{\hat{a}_1}{\hat{\sigma}_{\hat{a}_1}}, \text{ par la lecture sur une table de Student à } n - 2 \text{ degrés de liberté.}$$

La probabilité critique d’un test unilatéral peut directement se déduire de la probabilité critique d’un test bilatéral par une simple division par 2 $\rightarrow \alpha_{\text{unilatéral}}^c = \alpha_{\text{bilatéral}}^c / 2$ (opération inverse de celle de la lecture de la table).

Exemple : Pour $t^* = 2,53$ et $n - 2 = 20$, soit à déterminer α^c tel que $t_{20}^{\alpha^c} = 2,53$. Par lecture de la table de Student à 20 degrés de liberté, nous trouvons $\alpha^c = 2\%$ pour un test bilatéral et donc 1% pour un test unilatéral.

Exercice n° 3

↓ fichier C2EX2

Test de coefficient et intervalle de confiance

En reprenant les résultats de l’exercice 2, on demande de répondre aux questions suivantes.

- 1) La propension marginale à consommer est-elle significativement différente de 0 ?

- 2) Quel est l'intervalle de confiance au seuil (ou niveau) de 95 % pour la propension marginale à consommer ?

Solution

- 1) La propension marginale à consommer est-elle significativement différente de 0 ?

Cette question est très importante en économétrie. En effet, dans le cas d'une réponse négative – le coefficient n'est pas significativement différent de 0 – la variable explicative Revenu ne sera pas considérée comme étant explicative de la consommation puisque son coefficient de pondération est nul.

Il peut paraître étonnant de tester la différence par rapport à zéro et non pas seulement la positivité ou la négativité du coefficient de régression. En effet, il est commode de ne s'interroger que sur la contribution de la variable explicative, qu'elle soit positive ou négative.

Ce problème peut être formulé à l'aide de la théorie des tests à partir des deux hypothèses suivantes :

$$H_0 : a_1 = 0$$

$$H_1 : a_1 \neq 0$$

Si nous rejetons l'hypothèse H_0 , à un seuil α ¹ fixé, alors la propension marginale à consommer est considérée comme étant significativement différente de 0. Le seuil le plus communément employé est $\alpha = 0,05$, soit un risque de rejeter à tort H_0 de 5 %.

Nous savons que :

$$\frac{\hat{a}_1 - a_1}{\hat{\sigma}_{\hat{a}_1}} \quad \text{suit une loi de Student à } n - 2 \text{ degrés de liberté.}$$

Sous l'hypothèse H_0 , cette relation devient :

$$\frac{\hat{a}_1 - 0}{\hat{\sigma}_{\hat{a}_1}} = \frac{\hat{a}_1}{\hat{\sigma}_{\hat{a}_1}} = t_{\hat{a}_1}^* \rightarrow \text{loi de Student à } n - 2 \text{ degrés de liberté.}$$

$t_{\hat{a}_1}^*$ est appelé le *ratio de Student*.

Nous avons calculé \hat{a}_1 lors de l'exercice 2, il convient donc de calculer $\hat{\sigma}_{\hat{a}_1}$; or, d'après [11],

$$\hat{\sigma}_{\hat{a}_1}^2 = \frac{\hat{\sigma}_\varepsilon^2}{\sum_t (x_t - \bar{x})^2}, \quad \text{nous connaissons } \sum_t (x_t - \bar{x})^2 = 64\,156\,000$$

(colonne 6 du tableau de calcul 3). L'estimateur de la variance de l'erreur nous est donné par [10] :

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_t e_t^2}{n - 2}, \quad \text{où } e_t \text{ est le résidu de l'estimation } (e_t = y_t - \hat{y}_t)$$

-
1. Le seuil α est aussi appelé risque de première espèce : c'est la probabilité de rejeter l'hypothèse H_0 bien qu'elle soit vraie.

a) Calcul de \hat{y}_t et de e_t

La série ajustée \hat{y}_t est calculée par application des estimations \hat{a}_0 et \hat{a}_1 :

$$\hat{y}_t = \hat{a}_0 + \hat{a}_1 x_t$$

$$\text{soit } \hat{y}_1 = \hat{a}_0 + \hat{a}_1 x_1 \rightarrow e_1 = y_1 - \hat{y}_1$$

$$\hat{y}_1 = 1\,176,08 + 0,78 \times 8\,000 = 7\,423,95^1 \rightarrow e_1 = 7\,389,99 - 7\,423,95 = -33,96$$

$$\hat{y}_2 = 1\,176,08 + 0,78 \times 9\,000 = 8\,204,93 \rightarrow e_2 = 8\,169,65 - 8\,204,93 = -35,28$$

Les résultats sont consignés dans le tableau 4. Nous remarquons bien que $\sum_t e_t = 0$ (propriété de la méthode des moindres carrés).

Tableau 4 – Calcul du résidu d'estimation

\hat{y}_t	e_t	e_t^2
7 423,95	- 33,96	1 153,38
8 204,93	- 35,28	1 244,98
8 595,43	236,28	55 830,26
8 595,43	57,41	3 296,40
8 829,72	- 41,64	1 733,93
9 766,90	- 150,69	22 707,42
10 547,88	45,57	2 076,39
11 328,87	- 142,76	20 379,08
12 890,83	- 132,74	17 620,12
13 671,81	197,81	39 127,38
Somme	0,00	165 169,3
Moyenne	0,00	16 516,93

b) Calcul de l'estimation de la variance de l'erreur et de l'écart type du coefficient de régression.

L'estimation de la variance de l'erreur est donc égale à :

$$\hat{\sigma}_e^2 = \frac{\sum_t e_t^2}{n - 2} = \frac{165\,169,3}{10 - 2} = 20\,646,16$$

ce qui nous permet de calculer la variance estimée de \hat{a}_1 :

$$\hat{\sigma}_{\hat{a}_1}^2 = \frac{\hat{\sigma}_e^2}{\sum_t (x_t - \bar{x})^2} = \frac{20\,646,16}{64\,156\,000} = 0,000\,321\,8$$

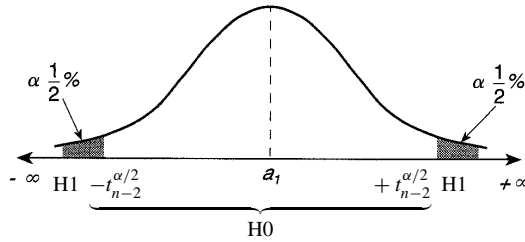
soit $\hat{\sigma}_{\hat{a}_1} = 0,017\,9$

c) Calcul du ratio de Student et règle de décision.

Nous savons que $\frac{\hat{a}_1}{\hat{\sigma}_{\hat{a}_1}} = t_{\hat{a}_1}^*$ est distribué selon une loi de Student à $n - 2$

degrés de liberté, la distribution d'échantillonnage sous H_0 est donc représentée par le graphe 6.

1. Les petites différences constatées sont dues à des problèmes d'arrondi.



Graphique 6 – Distribution d'échantillonnage sous l'hypothèse H_0

La règle de décision pour un seuil α est alors la suivante :

On calcule le ratio empirique de Student $t_{a1}^* = \frac{\hat{a}_1}{\hat{\sigma}_{\hat{a}_1}}$ (rapport du coefficient sur son écart type)

– Si t_{a1}^* est inférieur à $-t_{n-2}^{\alpha/2}$ ou supérieur à $+t_{n-2}^{\alpha/2}$ alors on rejette l'hypothèse H_0 (nous sommes dans la zone hachurée H_1), le coefficient a_1 est alors significativement différent de 0 (on accepte $a_1 \neq 0$) ; la variable explicative R_i est donc contributive à l'explication de la variable C_i .

– Si t_{a1}^* est compris dans l'intervalle $\pm t_{n-2}^{\alpha/2}$, alors nous ne sommes pas en mesure de rejeter l'hypothèse H_0 (donc on l'accepte), le coefficient a_1 n'est pas significativement différent de 0 (on accepte $a_1 = 0$) ; la variable explicative R_i n'est donc pas explicative de la variable C_i .

Il est plus simple de profiter de la symétrie de la loi de Student et donc de calculer la valeur absolue du ratio de Student et de la comparer directement à la valeur lue dans la table.

La règle de décision pour un seuil $\alpha = 0,05$ est alors la suivante :

– si $t_{a1}^* = \frac{|\hat{a}_1|}{\hat{\sigma}_{\hat{a}_1}} > t_{n-2}^{0,025} \rightarrow$ on rejette l'hypothèse H_0 , le coefficient a_1 est alors significativement différent de 0 (on accepte $a_1 \neq 0$) ; la variable explicative R_i est donc contributive à l'explication de la variable C_i ;

– si $t_{a1}^* = \frac{|\hat{a}_1|}{\hat{\sigma}_{\hat{a}_1}} \leq t_{n-2}^{0,025} \rightarrow$ on accepte l'hypothèse H_0 , le coefficient a_1 n'est donc pas significativement différent de 0 (on accepte $a_1 = 0$) ; la variable explicative R_i n'est donc pas contributive à l'explication de C_i .

Nous voyons l'importance que revêt ce test dans l'investigation économétrique ; en effet, il permet de tester la pertinence d'une variable explicative qui figure dans un modèle et sa contribution à l'explication du phénomène que l'on cherche à modéliser.

Dans notre exemple, nous calculons le ratio de Student :

$$t_{a1}^* = \frac{|\hat{a}_1|}{\hat{\sigma}_{\hat{a}_1}} = \frac{0,78}{0,0179} = 43,57 > t_8^{0,025} = 2,306^1 \rightarrow a_1 \neq 0$$

La propension marginale à consommer est donc significativement différente de 0, la variable Revenu est bien explicative de la variable Consommation.

2) Détermination d'un intervalle de confiance, au seuil de 95 %, pour la propension marginale à consommer.

1. Les tables statistiques sont en fin d'ouvrage. Dans le cas d'un test unilatéral, il convient d'utiliser une table correspondant à ce type de test, c'est-à-dire dont seule la partie de gauche est hachurée. Dans le cas d'un test bilatéral, ce sont les tables dont les deux parties extrêmes sont hachurées. Bien entendu, cela n'est vrai que pour les lois de probabilité symétriques.

Nous savons que :

$\frac{\hat{a}_1 - a_1}{\hat{\sigma}_{\hat{a}_1}}$ suit une loi de Student à $n - 2$ degrés de liberté, soit $\frac{\hat{a}_1 - a_1}{\hat{\sigma}_{\hat{a}_1}} = t_{n-2}^\alpha$, l'intervalle de confiance nous est donné (cf. graphe 4) par :

$$\frac{\hat{a}_1 - a_1}{\hat{\sigma}_{\hat{a}_1}} = \pm t_{n-2}^{\alpha/2} \rightarrow a_1 = \hat{a}_1 \pm \hat{\sigma}_{\hat{a}_1} \times t_{n-2}^{\alpha/2}$$

Application numérique pour un seuil $\alpha = 0,95$:

$$a_1 = 0,78 \pm 2,306 \times 0,0179$$

Nous avons donc un risque de 5 % que le véritable coefficient a_1 se trouve à l'extérieur de l'intervalle $[0,74 ; 0,82]$; nous constatons que 0 ne figure pas dans cet intervalle de confiance, ce qui est bien entendu cohérent avec la question précédente.

IV. Équation et tableau d'analyse de la variance

A. Équation d'analyse de la variance

Démontrons les deux relations suivantes :

– $\sum_t e_t = 0$: la somme des résidus est nulle (la droite de régression passe par le point moyen¹).

$$y_t = \hat{a}_0 + \hat{a}_1 x_t + e_t \rightarrow \sum_t y_t = \sum_t \hat{a}_0 + \hat{a}_1 \sum_t x_t + \sum_t e_t \rightarrow$$

$$\sum_t y_t - n\hat{a}_0 - \hat{a}_1 \sum_t x_t = \sum_t e_t \text{ or } \hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}, \text{ en divisant par } n \text{ il apparaît que } \sum_t e_t = 0.$$

– $\sum_t y_t = \sum_t \hat{y}_t$, il y a égalité entre la moyenne de la série à expliquer et la moyenne de la série ajustée.

$$y_t - \hat{y}_t = e_t \rightarrow \sum_t y_t - \sum_t \hat{y}_t = \sum_t e_t = 0 \rightarrow \bar{y} = \bar{\hat{y}}$$

De ces deux relations nous pouvons en déduire l'équation fondamentale d'analyse de la variance :

$$\boxed{\begin{array}{lcl} \sum_t (y_t - \bar{y})^2 & = & \sum_t (\hat{y}_t - \bar{\hat{y}})^2 + \sum_t e_t^2 \\ SCT & = & SCE + SCR \end{array}} \quad [12]$$

1. Cela n'est vrai que pour les modèles comportant un terme constant, ou bien, pour les modèles sans terme constant, si les données sont centrées sur leur moyenne.

La variabilité totale (SCT) est égale à la variabilité expliquée (SCE) + la variabilité des résidus (SCR).

Cette équation va nous permettre de juger de la qualité de l’ajustement d’un modèle. En effet, plus la variance expliquée est proche de la variance totale, meilleur est l’ajustement du nuage de points par la droite des moindres carrés. Il est d’usage de calculer le rapport :

$$R^2 = \frac{\sum_t (\hat{y}_t - \bar{y})^2}{\sum_t (y_t - \bar{y})^2} = 1 - \frac{\sum_t e_t^2}{\sum_t (y_t - \bar{y})^2}$$

[13]

R^2 est appelé le coefficient de détermination, et R le coefficient de corrélation multiple (dans le cas particulier du modèle de régression à une seule variable explicative, il est égal au coefficient de corrélation linéaire simple entre x et y).

B. Tableau d’analyse de la variance

Le tableau 5 présente l’analyse de la variance pour un modèle de régression simple.

Tableau 5 – Analyse de la variance par une régression simple

Source de variation	Somme des carrés	Degré de liberté	Carrés moyens
x	$SCE = \sum_t (\hat{y}_t - \bar{y})^2$	1	$SCE/1$
Résidu	$SCR = \sum_t e_t^2$	$n - 2$	$SCR/(n - 2)$
Total	$SCT = \sum_t (y_t - \bar{y})^2$	$n - 1$	

Les degrés de liberté correspondent au nombre de valeurs que nous pouvons choisir arbitrairement (par exemple, pour la variabilité totale, connaissant $n - 1$ valeurs, nous pourrons en déduire la n -ième, puisque nous connaissons la moyenne \bar{y}).

Le test $H_0 : a_1 = 0$ est équivalent au test d’hypothèse¹ $H_0 : SCE = 0$ (la variable explicative x_t ne contribue pas à l’explication du modèle).

Soit le test d’hypothèses $H_0 : SCE = 0$ contre l’hypothèse $H_1 : SCE \neq 0$.

La statistique² de ce test est donnée par :

1. Cela n’est vrai que dans le cas du modèle de régression simple.
2. Nous comparons la somme des carrés expliqués SCE à la somme des carrés des résidus SCR qui est représentative de la somme des carrés théoriquement la plus faible.

$$F^* = \frac{\frac{SCE}{ddl_{SCE}}}{\frac{SCR}{ddl_{SCR}}} = \frac{\frac{\sum_t (\hat{y}_t - \bar{y})^2}{1}}{\frac{\sum_t e_t^2}{(n-2)}} \quad [14]$$

Ou encore :

$$F^* = \frac{\frac{SCE}{ddl_{SCE}}}{\frac{SCR}{ddl_{SCR}}} = \frac{\frac{\sum_t (\hat{y}_t - \bar{y})^2}{1}}{\frac{\sum_t e_t^2}{(n-2)}} = \frac{\frac{R^2}{1}}{\frac{(1-R^2)}{(n-2)}} \quad [15]$$

La statistique F^* est le rapport de la somme des carrés expliqués par x_t sur la somme des carrés des résidus, chacune de ces sommes étant divisée par son degré de liberté respectif. Ainsi, si la variance expliquée est significativement supérieure à la variance résiduelle, la variable x_t est considérée comme étant une variable réellement explicative.

F^* suit une statistique de Fisher à 1 et $n - 2$ degrés de liberté. Si $F^* > F_{1;n-2}^\alpha$ nous rejetons au seuil α l'hypothèse H_0 d'égalité des variances, la variable x_t est significative ; dans le cas contraire, nous acceptons l'hypothèse d'égalité des variances, la variable x_t n'est pas explicative de la variable y_t .

En effet, $\frac{(\hat{a}_1 - a_1)^2}{\sigma_\varepsilon^2 / \sum_t (x_t - \bar{x})^2}$ suit une loi du χ^2 (chi-deux) à un degré de liberté (carré d'une variable aléatoire normale centrée réduite) et $\frac{\sum_{t=1}^n e_t^2}{\sigma_\varepsilon^2} = (n-2) \frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} = (n-2) \frac{\hat{\sigma}_{\hat{a}_1}^2}{\sigma_{\hat{a}_1}^2}$ suit une loi du χ^2 (chi-deux) à $n - 2$ degrés de liberté (somme au carré de $n - 2$ variables aléatoires *indépendantes* normales centrées réduites).

$$\text{En effectuant le rapport des deux chi-deux on obtient : } F^* = \frac{(\hat{a}_1 - a_1)^2 \sum_t (x_t - \bar{x})^2}{\frac{\sum_t e_t^2}{(n-2)}},$$

$$\text{soit sous } H_0 (a_1 = 0) \quad F^* = \frac{\hat{a}_1^2 \sum_t (x_t - \bar{x})^2}{\frac{\sum_t e_t^2}{(n-2)}} = \frac{\frac{SCE}{1}}{\frac{SCR}{(n-2)}} \text{ suit une loi de Fisher à 1}$$

et $n - 2$ degrés de liberté (rapport de chi-deux divisés par leurs degrés de liberté).

En effet, nous avons $SCE = \sum_t (\hat{y}_t - \bar{y})^2 = \hat{a}_1^2 \sum_t (x_t - \bar{x})^2$ car

$$\hat{y}_t - \bar{y} = \hat{a}_0 + \hat{a}_1 x_t - \hat{a}_0 - \hat{a}_1 \bar{x} = \hat{a}_1 (x_t - \bar{x}).$$

Nous remarquons :

$$F^* = (t_{\hat{a}_1}^*)^2 = \left(\frac{\hat{a}_1}{\hat{\sigma}_{\hat{a}_1}} \right)^2 = \frac{\hat{a}_1^2}{\hat{\sigma}_\varepsilon^2 / \sum_t (x_t - \bar{x})^2} = \frac{\hat{a}_1^2 \sum_t (x_t - \bar{x})^2}{\sum_t e_i^2 / (n - 2)}$$

Exercice n° 4

Équivalence des tests et analyse de la variance

Un agronome cherche à estimer la relation liant la production de maïs y_i au taux de bauxite x_i se trouvant dans la terre en formalisant la relation :

$$y_i = a_0 + a_1 x_i + \varepsilon_i$$

À partir d'une étude statistique portant sur 85 parcelles de terre, un économètre lui fournit les résultats suivants :

$$y_i = 132,80 - 1,1 x_i + e_i \quad i = 1, \dots, 85$$

$$(4,3) \quad (10,2)$$

$$(\cdot) = \text{ratio de Student}$$

$$\sum_i e_i^2 = 6\,234,32$$

- 1) Montrer que tester l'hypothèse $H_0 : a_1 = 0$ revient à tester l'hypothèse $r = 0$, où r est le coefficient de corrélation linéaire simple entre y_i et x_i ; le calculer.
- 2) Construire le tableau d'analyse de la variance et vérifier les résultats obtenus en 1) à partir du test de Fisher.
- 3) Le coefficient a_1 est-il significativement inférieur à -1 ?

Solution

- 1) Les observations ne sont pas datées, le modèle est spécifié en coupe instantanée.

Pour tester l'hypothèse $H_0 : a_1 = 0$, nous comparons le ratio de Student empirique $t^* = 10,2$ à la valeur critique $t_{83}^{0,05} = 1,96$ ¹.

1. Le degré de liberté de la loi de Student étant supérieur à 30, il est licite de l'approximer par une loi normale, 1,96 est la valeur de la loi normale à un seuil de 0,05 (test bilatéral).

Puisque t^* est largement supérieur à 1,96, nous rejetons l'hypothèse H_0 , a_1 est donc significativement différent de 0. Le taux de bauxite est un facteur explicatif (négatif) de la production de maïs.

L'écart type du coefficient \hat{a}_1 est égal à : $\hat{\sigma}_{\hat{a}_1} = \hat{a}_1 / t^* = 0,107$.

Montrons l'équivalence de ce test avec $H_0 : r = 0$.

Le coefficient de corrélation linéaire simple est égal à :

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \rightarrow r^2 = \frac{\left[\sum_i (x_i - \bar{x})(y_i - \bar{y}) \right]^2}{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}$$

Or, d'après [1], nous savons que :

$$\hat{a}_1 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} \rightarrow r^2 = \frac{\hat{a}_1 \sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (y_i - \bar{y})^2} = \frac{SCE}{SCT} = R^2$$

En effet :

$$\begin{aligned} \hat{a}_1 \sum_i (x_i - \bar{x})(y_i - \bar{y}) &= \hat{a}_1 \times \hat{a}_1 \sum_i (x_i - \bar{x})^2 = \hat{a}_1^2 \sum_i (x_i - \bar{x})^2 \\ &= \sum_i (\hat{a}_1 x_i - \hat{a}_1 \bar{x})^2 = \sum_i (\hat{y}_i - \hat{a}_0 - \bar{y} + \hat{a}_0)^2 = \sum_i (\hat{y}_i - \bar{y})^2 = SCE \end{aligned}$$

Pour le modèle de régression simple, nous avons égalité entre le coefficient de détermination et le carré du coefficient de corrélation.

Nous avons la relation :

$$F^* = \frac{R^2}{(1 - R^2)/(n - 2)} = \frac{r^2}{(1 - r^2)/(n - 2)} = (t^*)^2$$

donc on en déduit :

$$t^* = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \text{ qui suit une loi de Student à } n - 2 \text{ degrés de liberté.}$$

Ceci permet de tester si la relation entre y_i et x_i est significative, ou encore si le coefficient r est significativement différent de 0. Le calcul du coefficient de corrélation ne pose alors plus de problème :

$$t^* = 10,2 = \frac{r\sqrt{83}}{\sqrt{1-r^2}} \rightarrow r^2 = 0,556 \rightarrow |r| = 0,745^1$$

1. Nous savons que le coefficient r est en réalité négatif puisque le coefficient de régression \hat{a}_1 est lui-même négatif.

2) Pour construire le tableau d'analyse de la variance, il nous faut connaître :

$$SCE = \sum_i (\hat{y}_i - \bar{y})^2$$

$$SCR = \sum_i e_i^2$$

$$SCT = \sum_i (y_i - \bar{y})^2$$

Or, d'après [13], nous avons : $R^2 = 1 - \frac{\sum_i e_i^2}{\sum_i (y_i - \bar{y})^2} = 1 - \frac{SCR}{SCT}$

$R^2 = r^2 = 0,556$, la connaissance de $SCR = \sum_i e_i^2 = 6\,234,32$ permet de déterminer $SCT = 14\,041,26$ ainsi que, d'après [12], $SCE = 7\,806,94$. Nous pouvons maintenant construire le tableau 6 d'analyse de la variance.

Tableau 6 – Analyse de la variance

Source de variation	Somme des carrés	Degré de liberté	Carrés moyens
x_i	$SCE = 7\,806,94$	1	7 806,94
Résidu	$SCR = 6\,234,32$	85 – 2	75,11
Total	$SCT = 14\,041,26$	85 – 1	

$$F^* = \frac{SCE/1}{SCR/(n-2)} = \frac{7\,806,94}{75,11} = 103,94 > F_{1,83}^{0,05} = 3,96$$

Nous remarquons que $F^* = (t^*)^2$.

Dans le modèle de régression simple, il y a équivalence à tester :

H0	H1
$a_1 = 0$	$a_1 \neq 0$
$r_{x,y} = 0$	$r_{x,y} \neq 0$
$SCE = 0$	$SCE \neq 0$

Le premier test porte sur la pente de la droite de régression, le deuxième test sur le coefficient de corrélation entre x et y et, enfin, le troisième a pour but de juger si la somme des carrés expliqués est significative, ces trois tests néanmoins répondent à la même interrogation.

3) Le coefficient a_1 est-il significativement inférieur à -1 ?

La formulation des hypothèses est la suivante :

$$H0 : a_1 = -1$$

$$H1 : a_1 < -1$$

Sous l'hypothèse H_0 , nous avons :

$$\frac{|\hat{a}_1 - a_1|}{\hat{\sigma}_{\hat{a}_1}} = \frac{|\hat{a}_1 - (-1)|}{\hat{\sigma}_{\hat{a}_1}} = \frac{|-1,1 + 1|}{0,1078} = 0,92 < t^{0,05} = 1,65^1$$

Nous acceptons l'hypothèse H_0 , a_1 n'est pas significativement différent de -1 . Nous pouvons aussi répondre à cette question en vérifiant que la valeur -1 est comprise dans l'intervalle de confiance :

$$a_1 = \hat{a}_1 \pm 1,96 \hat{\sigma}_{\hat{a}_1} \rightarrow a_1 \in [-1,31 ; -0,89]$$

(Attention : $t^{0,05} = 1,96$ car le test est maintenant bilatéral.)

V. La prévision dans le modèle de régression simple

Lorsque les coefficients du modèle ont été estimés, il est possible de calculer une prévision à un horizon h .

Soit le modèle estimé sur la période $t = 1, \dots, n$: $y_t = \hat{a}_0 + \hat{a}_1 x_t + e_t$, si la valeur de la variable explicative x_t est connue en $n+1$ (x_{n+1}), la prévision est donnée par : $\hat{y}_{n+1} = \hat{a}_0 + \hat{a}_1 x_{n+1}$.

Montrons que cette prévision est sans biais.

L'erreur de prévision est égale à : $e_{n+1} = y_{n+1} - \hat{y}_{n+1}$

que l'on peut écrire : $e_{n+1} = (a_0 + a_1 x_{n+1} + \varepsilon_{n+1}) - (\hat{a}_0 + \hat{a}_1 x_{n+1})$

soit : $e_{n+1} = (a_0 - \hat{a}_0) + (a_1 - \hat{a}_1)x_{n+1} + \varepsilon_{n+1}$. En se référant aux hypothèses du modèle, on a :

$$E(e_{n+1}) = 0$$

Une démonstration analogue permet d'obtenir $E(e_{n+h}) = 0$.

La prévision sans biais est donc obtenue par l'application directe du modèle de régression estimé. Cependant, dans la pratique, il n'est que de peu d'utilité de connaître la prévision si nous ne savons pas quel degré de confiance nous pouvons lui accorder. Nous allons donc calculer la variance de l'erreur de prévision qui nous permet de déterminer un intervalle de

1. Valeur de la loi normale au seuil de 5 % ; en effet il s'agit d'un test unilatéral.

confiance¹ bornant la prévision. La variance de l'erreur de prévision est donnée par :

$$V(e_{n+1}) = V((\widehat{a}_0 - a_0) + (\widehat{a}_1 - a_1)x_{n+1} + \varepsilon_{n+1})$$

Puisque la variable x_{n+1} est certaine et l'erreur ε_{n+1} est non autocorrélée avec les ε_t cette expression peut s'écrire :

$$V(e_{n+1}) = V(\widehat{a}_0) + x_{n+1}^2 V(\widehat{a}_1) + 2x_{n+1} \text{Cov}(\widehat{a}_0, \widehat{a}_1) + V(\varepsilon_{n+1})$$

En remplaçant les variances et la covariance des coefficients par leurs expressions d'après [9] et connaissant $V(\varepsilon_{n+1}) = \sigma_\varepsilon^2$, nous obtenons :

$$V(e_{n+1}) = \left(\frac{\sigma_\varepsilon^2}{n} + \bar{x}^2 V(\widehat{a}_1) + x_{n+1}^2 V(\widehat{a}_1) - 2x_{n+1} \bar{x} V(\widehat{a}_1) + \sigma_\varepsilon^2 \right)$$

ou encore :

$$V(e_{n+1}) = V(y_{n+1} - \widehat{y}_{n+1}) = \widehat{\sigma}_\varepsilon^2 \left[\frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_t (x_t - \bar{x})^2} + 1 \right] \quad [16]$$

Nous pouvons observer que, dans cette formule, la variance de l'erreur de prévision est fonction de l'écart quadratique entre la variable exogène prévue et la moyenne de cette même variable : plus la valeur prévue s'éloigne de cette moyenne, plus le risque d'erreur est important. De même, nous remarquons que la variance de l'erreur de prévision est une fonction inverse de la variabilité de la série explicative.

L'hypothèse de normalité de ε_t permet alors de déterminer un intervalle à $(1 - \alpha) \%$ pour la prévision :

$$e_{n+1} = y_{n+1} - \widehat{y}_{n+1} \rightarrow N \left(0, \sigma_\varepsilon^2 \left[\frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_t (x_t - \bar{x})^2} + 1 \right] \right)$$

Soit $\frac{\widehat{a}_0 + \widehat{a}_1 x_{n+1} - y_{n+1}}{\widehat{\sigma}_\varepsilon \sqrt{\frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_t (x_t - \bar{x})^2} + 1}} \rightarrow t_{n-2}$ (Student à $n - 2$ d.d.l.)

$$\boxed{y_{n+1} = \widehat{y}_{n+1} \pm t_{n-2}^{\alpha/2} \widehat{\sigma}_\varepsilon \sqrt{\frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_t (x_t - \bar{x})^2} + 1}} \quad [17]$$

1. Il est plus juste de parler d'intervalle de prédiction.

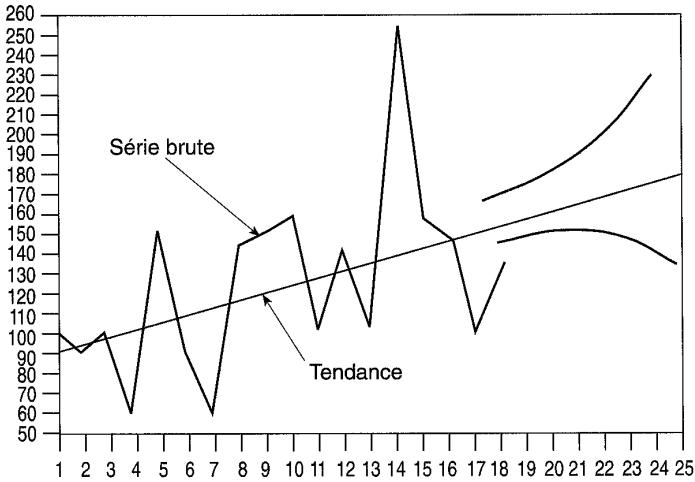
Cas particulier

Lorsque nous utilisons le modèle de régression simple pour calculer une droite de tendance (moindres carrés sur le temps), le modèle est spécifié ainsi :

$$T_t = \hat{a}_0 + \hat{a}_1 t + e_t \quad \text{pour } t = 1, \dots, n$$

Pour calculer la prévision à l'horizon h , nous employons la formule d'extrapolation :

$\hat{T}_{n+h} = \hat{a}_0 + \hat{a}_1(n+h)$, et l'intervalle de prédiction se trouve alors sur deux branches d'hyperbole¹ $(n+h-\bar{t})^2$ illustré par le graphique 7.



Graphique 7 – Intervalle de la prévision par extrapolation de tendance

Exercice n° 5

↓ fichier C2EX2

Prévision dans un modèle de régression simple

Nous reprenons le modèle consommation-revenu estimé lors de l'exercice 2 :

$$y_t = 1\,176,08 + 0,78 x_t + e_t$$

(0,21) (43,53)

$$n = 10$$

(·) = t de Student

- 1) Calculer le coefficient de détermination et effectuer le test de Fisher permettant de déterminer si la régression est globalement significative.

1. C'est l'effet « trompette », le lecteur devinera pourquoi...

- 2) Quelle est la conséquence sur la consommation de l'augmentation du revenu de 8 % ?
- 3) Pour les années 11 et 12, on prévoit respectivement 16 800 et 17 000 dollars de revenu par habitant. Déterminer la prévision de consommation pour ces deux années ainsi que l'intervalle de prédiction au seuil de 95 %.

Solution

- 1) Pour calculer le coefficient de détermination, nous nous référons à la relation [15] :

$$F^* = \frac{R^2}{(1 - R^2)/(n - 2)} = \frac{r^2}{(1 - r^2)/(n - 2)} = (r^*)^2 = 43,5^2$$

Soit : $r^2 = 0,99$, nous pouvons alors calculer le Fisher empirique :

$$F^* = 1895,3 > F_{1,8}^{0,05} = 5,32, \text{ la variable explicative est significative.}$$

- 2) Augmentation de 8 % du revenu.

Nous avons :

$$\Delta \hat{y}_t = \hat{a}_1 \Delta x_t \text{ soit } \Delta \hat{y}_t = 0,78 \times \Delta x_t = 0,78 \times 0,08 = 0,0624$$

La consommation augmente de 6,24 %, soit un peu moins que le revenu.

- 3) Les prévisions sont calculées par l'utilisation du modèle estimé.

$$\begin{aligned} \hat{y}_{11} &= 1\,176,08 + 0,78 x_{11} = 1\,176,08 + 0,78 \times 16\,800 \\ &= 14\,280,08 \end{aligned}$$

L'intervalle de prévision [17] peut alors être calculé :

$$y_{11} = \hat{y}_{11} \pm t_{n-2}^{\alpha/2} \hat{\sigma}_\varepsilon \sqrt{\frac{1}{n} + \frac{(x_{11} - \bar{x})^2}{\sum_t (x_t - \bar{x})^2} + 1}$$

avec :

$$\begin{aligned} n &= 10 \\ \hat{\sigma}_\varepsilon &= 143,69 \text{ (d'après l'exercice 3)} \\ \sum_t (x_t - \bar{x})^2 &= 64\,156\,000 \text{ (d'après l'exercice 3)} \\ \bar{x} &= 11\,280 \text{ (d'après l'exercice 2)} \\ t_{n-2}^{\alpha/2} &= 2,306 \\ x_{11} &= 16\,800 \\ y_{11} &= 14\,280,08 \pm 2,306 \times 180,32 \\ IC &= [13\,864,24 ; 14\,695,91] \end{aligned}$$

La réalisation a 95 % de chance de se trouver à l'intérieur de cet intervalle¹.

De même, pour l'année 12, nous obtenons :

$$\begin{aligned} \hat{y}_{12} &= 1\,176,08 + 0,78 x_{12} = 1\,176,08 + 0,78 \times 17\,000 \\ &= 14\,436,08 \\ y_{12} &= 14\,436,08 \pm 2,306 \times 182,32 \\ IC &= [14\,015,65 ; 14\,856,51] \end{aligned}$$

1. La distribution étant normale (loi continue), il est clair que la probabilité d'apparition de chacune des valeurs de l'intervalle n'est pas identique. La valeur la plus probable se trouve au centre de l'intervalle.

Exercice n° 6

Comparaison de coefficients de régression

Un économiste spécialisé en économie du travail s'intéresse à la relation liant la rémunération et la durée des études (théorie du capital humain). Pour ce faire, il dispose d'un échantillon de 40 hommes et 25 femmes ayant le même âge, dont il relève la rémunération annuelle (y_i), exprimée en milliers d'euros, et le nombre d'années d'études (x_i).

Les estimations économétriques conduisent aux résultats suivants :

Pour les hommes :

$$y_i = 18,60 + 1,8x_i + e_i \quad i = 1, \dots, 40 \quad n_1 = 40$$
$$(9,3) \quad (5,2)$$

(\cdot) = ratio de Student

$$R^2 = 0,42$$

Pour les femmes :

$$y_i = 14,50 + 0,7x_i + e_i \quad i = 1, \dots, 25 \quad n_2 = 25$$
$$(12,8) \quad (2,5)$$

(\cdot) = ratio de Student

$$R^2 = 0,22$$

- 1) L'influence de la durée des études sur la rémunération vous semble-t-elle significative ?
- 2) Existe-t-il une différence significative de l'impact de la durée des études sur la rémunération des hommes et des femmes ?

Solution

- 1) Pour répondre à cette première question, nous pouvons analyser soit les ratios de Student, soit le coefficient de détermination.

Le ratio de Student empirique de la variable « années d'études » est égal à :

Pour les hommes :

$$t_{\hat{a}_h}^* = \frac{\hat{a}_h}{\hat{\sigma}_{\hat{a}_h}} = 5,2 > t_{38}^{0,05} = 1,96$$

Pour les femmes :

$$t_{\hat{a}_f}^* = \frac{\hat{a}_f}{\hat{\sigma}_{\hat{a}_f}} = 2,5 > t_{23}^{0,05} = 2,06$$

Les écarts types sont :

$$\hat{\sigma}_{\hat{a}_h} = 0,34 \quad \text{et} \quad \hat{\sigma}_{\hat{a}_f} = 0,28$$

Les deux coefficients de régression sont donc significativement différents de 0. Il est à noter que, pour les femmes, le coefficient de pondération des années d'études est plus

faible et moins significatif que celui des hommes. Les probabilités critiques¹ sont, pour les hommes, de $\alpha^c = 0,0001$ et de $\alpha^c = 0,02$ pour les femmes (nous considérons le coefficient comme non significativement différent de 0 au seuil de 2 %).

Le test de Fisher mené sur les coefficients de détermination conduit aux mêmes résultats.

- 2) Ce problème se ramène à un test de différence de moyennes de variables aléatoires normales indépendantes et de variances inégales.

$$H0 : a_h = a_f \quad H0 : d = a_h - a_f = 0$$

$$H1 : a_h \neq a_f \quad H1 : d = a_h - a_f \neq 0$$

Le rapport $\frac{(\hat{a}_h - \hat{a}_f) - (a_h - a_f)}{\hat{\sigma}_{\hat{a}_h - \hat{a}_f}}$ suit une loi de Student à $n_1 + n_2 - 4$ degrés de liberté.

Avec $\hat{\sigma}_d^2 = \hat{\sigma}_{a_f}^2 + \hat{\sigma}_{a_h}^2$ et $\hat{d} = \hat{a}_h - \hat{a}_f$ et sous l'hypothèse $H0$, le rapport s'écrit :

$$\frac{\hat{d}}{\hat{\sigma}_d} = t^* = \frac{(1,8 - 0,7)}{\sqrt{0,34^2 + 0,28^2}} = 2,49 > t_{61}^{0,05} = 1,96$$

Rappel : $\text{Var}(a \times x_1 + b \times x_2) = a^2 \text{Var}(x_1) + b^2 \text{Var}(x_2) + 2ab \text{Cov}(x_1, x_2)$ où x_1 et x_2 sont deux variables aléatoires et a et b sont deux scalaires. Ici $\text{Cov}(\hat{a}_h, \hat{a}_f) = 0$ car les deux régressions sont indépendantes. D'où $\text{Var}(\hat{a}_f - \hat{a}_h) = \text{Var}(\hat{a}_f) + \text{Var}(\hat{a}_h)$.

Nous rejetons l'hypothèse $H0$, il existe une différence significative des coefficients de régression : la durée des études des femmes a moins d'impact sur la rémunération que la durée des études des hommes (sur cet échantillon...)

Nous pouvons déterminer la probabilité critique de ce test (probabilité à partir de laquelle nous sommes amenés à accepter l'hypothèse $H0$). La lecture de la table de Student à 61 degrés de liberté (sur la table ∞) indique une probabilité comprise entre 0,01 et 0,02 (la valeur exacte déterminée par la fonction Excel « *loi.student* » est 0,0154). Le risque de se tromper en rejetant l'hypothèse $H0$ est donc de 1,54 % ; compte tenu de ce très faible risque, nous rejetons l'hypothèse $H0$. Evidemment, plus la probabilité critique est faible plus nous sommes confortés dans notre décision.

Exercice n° 7

Apprendre à manipuler les formules

Soit les résultats d'une estimation économétrique :

$$y_t = 1,251 x_t - 32,95 + e_t$$

$$n = 20$$

$$R^2 = 0,23$$

$$\hat{\sigma}_e = 10,66$$

1. Cette notion est importante : il s'agit de déterminer le seuil de probabilité à partir duquel on accepte l'hypothèse $H0$. Plus ce seuil est faible, moins le risque de se tromper est important.

- 1) A partir des informations connues, on demande de retrouver les statistiques suivantes : la somme des carrés des résidus (SCR), la somme des carrés totaux (SCT), la somme des carrés expliqués (SCE), la valeur de la statistique du Fisher empirique (F^*) et l'écart type du coefficient $\hat{a}_1(\hat{\sigma}_{\hat{a}_1})$.
- 2) Le coefficient de la variable x est-il significativement supérieur à 1 ?

Solution

1) On sait que $\hat{\sigma}_\varepsilon = \sqrt{\frac{SCR}{n-2}} = 10,66 \Rightarrow SCR = (10,66)^2 \times 18 = 2\,045,44$

Nous pouvons calculer SCE et SCT à l'aide du coefficient de détermination.

$$R^2 = 0,23 = 1 - SCR/SCT \\ \Rightarrow SCT = SCR/(1 - R^2) = 2045,44/(1 - 0,23) = 2\,656,42$$

Or $SCT = SCE + SCR \Rightarrow SCE = 610,98$

Nous pouvons calculer maintenant :

$$F^* = \frac{R^2}{(1 - R^2)/(n - 2)} = \frac{SCE}{SCR/(n - 2)} = 5,40 \Rightarrow t^* = \sqrt{F^*} = 2,32$$

(dans le cas d'un modèle de régression simple $t^{*2} = F^*$). Nous pouvons en déduire

l'écart type du coefficient : $\hat{\sigma}_{\hat{a}_1} = \frac{\hat{a}_1}{t^*} = \frac{1,251}{2,32} = 0,54$.

- 2) On pose le test d'hypothèses :
 $H_0 : a_1 = 1$ contre l'hypothèse $H_1 : a_1 > 1$
 Sous H_0 , nous pouvons écrire :

$$t_{\hat{a}_1}^* = \frac{\hat{a}_1 - a_1}{\hat{\sigma}_{\hat{a}_1}} = \frac{1,25 - 1}{0,54} = 0,46 < t_{18}^{0,05} = 1,734^1 \Rightarrow \text{nous sommes donc dans la zone de l'acceptation de } H_0, \text{ le coefficient } a_1 \text{ n'est pas significativement supérieur à 1.}$$

Exercice n° 8

Apprendre à utiliser les formules

À partir d'un échantillon de 190 observations, on étudie la relation entre la variable à expliquer y_i et la variable explicative x_i .

À l'aide des informations fournies ci-dessous reconstituez les huit valeurs manquantes signalées par VM1, ...VM8.

1. Attention, comme le test est unilatéral et que la table de Student de cet ouvrage est tabulée directement pour $\alpha/2$ (cas le plus général d'un test bilatéral), il convient donc ici de lire sur la table à un seuil de $0,10 = 2 \times 0,05$.

Dependent Variable: Y				
Method: Least Squares				
Sample: 1 190				
Included observations: 190				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4364.928	VM1	-16.61141	0.0000
X	VM4	VM3	VM2	0.0000
R-squared	VM5	Mean dependent var		VM6
S.E. of regression	322.8850	S.D. dependent var		VM8
Sum squared resid	VM7	F-statistic		778.9623

On donne $\sigma_x = 3,447$ et $\bar{x} = 38,416$

Solution

$$\text{VM1 : directement par } \hat{\sigma}_{\hat{a}_0} = \frac{\hat{a}_0}{t_{\hat{a}_0}^*} = 262,76$$

$$\text{VM2 : } t_{\hat{a}_1}^* = \sqrt{F^*} = \sqrt{778,96} = 27,91$$

$$\text{VM3 : } \hat{\sigma}_{\hat{a}_1} = \frac{\hat{\sigma}_\varepsilon}{\sqrt{\sum (x_i - \bar{x})^2}}, \text{ or } \hat{\sigma}_\varepsilon = 322,88 ;$$

$$\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \text{ d'où } \sum (x_i - \bar{x})^2 = (n-1) \times \sigma_x^2 = 189 \times (3,447)^2 = 2245,66$$

$$\hat{\sigma}_{\hat{a}_1} = \frac{\hat{\sigma}_\varepsilon}{\sqrt{\sum (x_i - \bar{x})^2}} = \frac{322,88}{\sqrt{2245,66}} = 6,81$$

$$\text{d'où VM4} = \text{VM3} \times \text{VM2} = 190,15$$

$$\begin{aligned} \text{VM5 : on sait que } F^* &= \frac{R^2/1}{(1-R^2)/(n-2)} = \frac{R^2}{(1-R^2)/188} \\ &= 778,96 \rightarrow R^2 = \frac{778,96}{778,96 + 188} = 0,80 \end{aligned}$$

$$\text{VM6 : } \bar{y} = \hat{a}_0 + \hat{a}_1 \times \bar{x} = -4364,928 + 190,15 \times 38,416 = 2939,8$$

$$\text{VM7 : } \text{SCR} = \sum e_i^2, \text{ or } \hat{\sigma}_\varepsilon^2 = \frac{\sum e_i^2}{n-2}$$

$$\text{d'où } \sum e_i^2 = (n-2) \times \hat{\sigma}_\varepsilon^2 = 188 \times (322,88)^2 = 19\,599\,280.$$

$$\text{VM8 : } R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} \rightarrow SCT = \frac{SCR}{1-R^2} = 98\,986\,262$$

$$\sigma_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{SCT}{n-1}} = 723,7$$

1. Nous prenons ici la formule de l'estimateur de l'écart type calculé à partir d'un échantillon, donc nous divisons par $n-1$.

3. Le modèle de régression multiple

Le modèle linéaire général est une extension du modèle de régression simple abordé au chapitre précédent. Après avoir présenté le modèle linéaire général (I), nous envisageons une procédure d'estimation des paramètres en étudiant les propriétés statistiques des estimateurs (II). Les différents tests d'hypothèses concernant les coefficients du modèle sont proposés en (III), et la section (IV) est consacrée à l'analyse de la variance ainsi qu'aux tests s'y rattachant.

En (V) nous présentons une classe particulière de séries explicatives : les variables indicatrices, puis la prévision à l'aide du modèle linéaire général fait l'objet de la section (VI).

Enfin nous terminons ce chapitre par quelques exercices récapitulatifs.

I. Le modèle linéaire général

A. Présentation

Lors du chapitre précédent, nous avons considéré qu'une variable endogène est expliquée à l'aide d'une seule variable exogène. Cependant, il est extrêmement rare qu'un phénomène économique ou social puisse être appréhendé par une seule variable. Le modèle linéaire général est une généralisation du modèle de régression simple dans lequel figurent plusieurs variables explicatives :

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \dots + a_k x_{kt} + \varepsilon_t \text{ pour } t = 1, \dots, n$$

avec :

y_t = variable à expliquer à la date t ;

x_{1t} = variable explicative 1 à la date t ;

x_{2t} = variable explicative 2 à la date t ;

...

x_{kt} = variable explicative k à la date t ;

a_0, a_1, \dots, a_k = paramètres du modèle ;

ε_t = erreur de spécification (différence entre le modèle vrai et le modèle spécifié), *cette erreur est inconnue et restera inconnue* ;

n = nombre d'observations.

B. Forme matricielle

L'écriture précédente du modèle est d'un maniement peu pratique. Afin d'en alléger l'écriture et de faciliter l'expression de certains résultats, on a habituellement recours aux notations matricielles. En écrivant le modèle, observation par observation, nous obtenons :

$$y_1 = a_0 + a_1 x_{11} + a_2 x_{21} + \dots + a_k x_{k1} + \varepsilon_1$$

$$y_2 = a_0 + a_1 x_{12} + a_2 x_{22} + \dots + a_k x_{k2} + \varepsilon_2$$

...

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \dots + a_k x_{kt} + \varepsilon_t$$

...

$$y_n = a_0 + a_1 x_{1n} + a_2 x_{2n} + \dots + a_k x_{kn} + \varepsilon_n$$

Soit, sous forme matricielle :

$$\begin{matrix} Y \\ (n,1) \end{matrix} = \begin{matrix} X \\ (n,k+1) \end{matrix} \begin{matrix} a \\ (k+1,1) \end{matrix} + \begin{matrix} \varepsilon \\ (n,1) \end{matrix}$$

avec :

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_t \\ \vdots \\ y_n \end{pmatrix} ; X = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{1t} & x_{2t} & \dots & x_{kt} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{pmatrix} ; a = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix} ; \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_t \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Nous remarquons la première colonne de la matrice X , composée de 1, qui correspond au coefficient a_0 (coefficient du terme constant).

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

L'écriture sous forme matricielle rend plus aisée la manipulation du modèle linéaire général, c'est pourquoi nous l'adoptons par la suite.

II. Estimation et propriétés des estimateurs

A. Estimation des coefficients de régression

Soit le modèle sous forme matricielle à k variables explicatives et n observations :

$$Y = Xa + \varepsilon \quad [1]$$

Afin d'estimer le vecteur a composé des coefficients $a_0, a_1 \dots a_k$, nous appliquons la méthode des Moindres Carrés Ordinaires (MCO) qui consiste à minimiser la somme des carrés des erreurs, soit :

$$\text{Min} \sum_{i=1}^n \varepsilon_i^2 = \text{Min} \varepsilon' \varepsilon = \text{Min} (Y - Xa)'(Y - Xa) = \text{Min} S \quad [2]$$

avec ε' transposé¹ du vecteur ε .

Pour minimiser cette fonction par rapport à a , nous différencions² S par rapport à a :

$$\frac{\partial S}{\partial a} = -2 X'Y + 2 X'X \hat{a} = 0 \rightarrow \boxed{\hat{a} = (X'X)^{-1} X'Y} \quad [3]$$

Cette solution est réalisable³ si la matrice carrée $X'X$ de dimension $(k+1, k+1)$ est inversible. La matrice $X'X$ est la matrice des produits croisés des variables explicatives ; en cas de colinéarité parfaite entre deux variables explicatives, la matrice $X'X$ est singulière et la méthode des MCO défaille.

On appelle *équations normales* les équations issues de la relation :

$$X'X\hat{a} = X'Y$$

1. Nous désignerons par le signe ' les transposés de vecteur ou de matrice.

2. $S = (Y - Xa)'(Y - Xa) = Y'Y - Y'Xa - a'X'Y + a'X'Xa$
 $= Y'Y - 2a'X'Y + a'X'Xa$.

3. Les conditions de second ordre sont vérifiées du fait que $X'X$ est une matrice définie semi-positive.

Soit, sous forme matricielle :

$$\begin{pmatrix} n & \sum x_{1t} & \sum x_{2t} & \dots & \sum x_{kt} \\ \sum x_{1t} & \sum x_{1t}^2 & \sum x_{1t} x_{2t} & \dots & \sum x_{1t} x_{kt} \\ \sum x_{2t} & \sum x_{2t} x_{1t} & \sum x_{2t}^2 & \dots & \sum x_{2t} x_{kt} \\ \vdots & \dots & \dots & \dots & \dots \\ \sum x_{kt} & \sum x_{kt} x_{1t} & \sum x_{kt} x_{2t} & \dots & \sum x_{kt}^2 \end{pmatrix} \begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{a}_2 \\ \dots \\ \hat{a}_k \end{pmatrix} = \begin{pmatrix} \sum y_t \\ \sum x_{1t} y_t \\ \sum x_{2t} y_t \\ \dots \\ \sum x_{kt} y_t \end{pmatrix}$$

Le modèle estimé s'écrit :

$$y_t = \hat{a}_0 + \hat{a}_1 x_{1t} + \hat{a}_2 x_{2t} + \dots + \hat{a}_k x_{kt} + e_t$$

avec $e_t = y_t - \hat{y}_t$ où e_t est le résidu, c'est-à-dire l'écart entre la valeur observée de la variable à expliquer et sa valeur estimée (ajustée).

Attention : il convient de bien distinguer entre l'erreur de spécification du modèle (noté ε_t) qui est et restera inconnue et le résidu (e_t) qui, lui, est connu.

1) Cas particulier

Si nous raisonnons sur des données centrées, l'estimateur de a peut s'écrire en fonction des matrices des variances et covariances empiriques :

$$\begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \dots \\ \hat{a}_k \end{pmatrix} = \begin{pmatrix} \text{Var}(x_1) & \text{Cov}(x_1, x_2) & \text{Cov}(x_1, x_3) & \dots & \text{Cov}(x_1, x_k) \\ \text{Cov}(x_2, x_1) & \text{Var}(x_2) & \text{Cov}(x_2, x_3) & \dots & \text{Cov}(x_2, x_k) \\ \text{Cov}(x_3, x_1) & \text{Cov}(x_3, x_2) & \text{Var}(x_3) & \dots & \text{Cov}(x_3, x_k) \\ \dots & \dots & \dots & \dots & \dots \\ \text{Cov}(x_k, x_1) & \text{Cov}(x_k, x_2) & \text{Cov}(x_k, x_3) & \dots & \text{Var}(x_k) \end{pmatrix}^{-1} \times \begin{pmatrix} \text{Cov}(x_1, y) \\ \text{Cov}(x_2, y) \\ \text{Cov}(x_3, y) \\ \dots \\ \text{Cov}(x_k, y) \end{pmatrix}$$

avec $\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}_1 - \hat{a}_2 \bar{x}_2 - \dots - \hat{a}_k \bar{x}_k$.

Que sont des données centrées sur la moyenne ? Soit x_t une variable connue sur n observations et \bar{x} sa moyenne, nous pouvons calculer une nouvelle variable ($X_t = x_t - \bar{x}$) dont la somme est par construction nulle : $\sum_{t=1}^n (x_t - \bar{x}) = \sum_{t=1}^n X_t = 0$.

Nous avons donc $\bar{X} = 0$.

2) Effet de la variation d'une seule des variables explicatives

Soit le modèle estimé : $y_t = \hat{a}_0 + \hat{a}_1 x_{1t} + \hat{a}_2 x_{2t} + \dots + \hat{a}_k x_{kt} + e_t$.

Si la variable x_2 passe de la valeur x_{2t} à $(x_{2t} + \Delta x_{2t})$, toutes choses étant égales par ailleurs (les $k - 1$ autres variables restant constantes), alors la variable à expliquer varie de $\hat{a}_2 \Delta x_2 : \Delta \hat{y}_t = \hat{a}_2 \Delta x_{2t}$.

Les coefficients s'interprètent donc directement en terme de propension marginale.

B. Hypothèses et propriétés des estimateurs

Par construction, le modèle est linéaire en X (ou sur ces coefficients) et nous distinguons les hypothèses stochastiques (liées à l'erreur ε) des hypothèses structurelles.

1) Hypothèses stochastiques

- H1 : les valeurs $x_{i,t}$ sont observées sans erreur.
- H2 : $E(\varepsilon_t) = 0$, l'espérance mathématique de l'erreur est nulle.
- H3 : $E(\varepsilon_t^2) = \sigma_\varepsilon^2$, la variance de l'erreur est constante ($\forall t$) (homoscédasticité).
- H4 : $E(\varepsilon_t \varepsilon_{t'}) = 0$ si $t \neq t'$, les erreurs sont non corrélées (ou encore indépendantes).
- H5 : $\text{Cov}(x_{it}, \varepsilon_t) = 0$, l'erreur est indépendante des variables explicatives.

2) Hypothèses structurelles

- H6 : absence de colinéarité entre les variables explicatives, cela implique que la matrice $(X' X)$ est régulière et que la matrice inverse $(X' X)^{-1}$ existe.
- H7 : $(X' X)/n$ tend vers une matrice finie non singulière.
- H8 : $n > k + 1$, le nombre d'observations est supérieur¹ au nombre des séries explicatives.

3) Propriétés des estimateurs

Considérons les propriétés de l'estimateur [3].

Le modèle sous forme matricielle peut s'écrire, comme pour le modèle de régression simple, de différentes manières :

1. Dans le cas d'une égalité, nous aurions un système de n équations à n inconnues, donc parfaitement déterminé.

$$\left. \begin{array}{l} Y = Xa + \varepsilon \\ Y = X\hat{a} + e \\ \hat{Y} = X\hat{a} \end{array} \right\} \rightarrow e = Y - \hat{Y} \quad (e = \text{résidu})$$

Nous obtenons :

$$\begin{aligned} \hat{a} &= (X'X)^{-1} X'Y = (X'X)^{-1} X'(Xa + \varepsilon) \\ \hat{a} &= (X'X)^{-1} X'(Xa) + (X'X)^{-1} X'\varepsilon \\ \hat{a} &= a + (X'X)^{-1} X'\varepsilon \end{aligned} \quad [4]$$

d'où $E(\hat{a}) = a + (X'X)^{-1} X'E(\varepsilon) = a$ car $E(\varepsilon) = 0$

L'estimateur est donc sans biais :

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

Calculons maintenant la matrice des variances et covariances des coefficients de régression $\Omega_{\hat{a}}$.

$$\Omega_{\hat{a}} = E\{(\hat{a} - a)(\hat{a} - a)'\}$$

Or, d'après [4], $(\hat{a} - a) = (X'X)^{-1} X'\varepsilon$ et $(\hat{a} - a)' = \varepsilon'X(X'X)^{-1}$ puisque $(X'X)^{-1}$ est symétrique¹.

$$(\hat{a} - a)(\hat{a} - a)' = (X'X)^{-1} X'\varepsilon\varepsilon'X(X'X)^{-1} \text{ d'où}$$

$\Omega_{\hat{a}} = (X'X)^{-1} X'E(\varepsilon\varepsilon')X(X'X)^{-1}$ avec $E(\varepsilon\varepsilon') = \Omega_{\varepsilon} = \sigma_{\varepsilon}^2 I =$ matrice des variances et covariances de l'erreur ε .

En effet, d'après les hypothèses H3 et H4 nous avons :

$$\begin{aligned} \Omega_{\varepsilon} = E(\varepsilon\varepsilon') &= \begin{bmatrix} E(\varepsilon_1\varepsilon_1) & E(\varepsilon_1\varepsilon_2) & \dots & E(\varepsilon_1\varepsilon_n) \\ E(\varepsilon_2\varepsilon_1) & E(\varepsilon_2\varepsilon_2) & \dots & E(\varepsilon_2\varepsilon_n) \\ \dots & \dots & \dots & \dots \\ E(\varepsilon_n\varepsilon_1) & E(\varepsilon_n\varepsilon_2) & \dots & E(\varepsilon_n\varepsilon_n) \end{bmatrix} = \\ &= \begin{bmatrix} \sigma_{\varepsilon}^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_{\varepsilon}^2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \sigma_{\varepsilon}^2 \end{bmatrix} \end{aligned}$$

Soit : $\Omega_{\hat{a}} = \sigma_{\varepsilon}^2 (X'X)^{-1} X'X(X'X)^{-1}$

$$\Omega_{\hat{a}} = \sigma_{\varepsilon}^2 (X'X)^{-1}$$

[5]

1. Nous rappelons quelques règles du calcul matriciel : $(ABC)' = C'B'A'$ et si A est symétrique $A' = A$ et donc $[(X'X)^{-1}]' = (X'X)^{-1}$.

$$\Omega_{\hat{a}} = \frac{\sigma_{\varepsilon}^2}{n} \left(\frac{X'X}{n} \right)^{-1} \Rightarrow \text{Lim } \Omega_{\hat{a}} = 0 \text{ si } n \rightarrow \infty \text{ (d'après les hypothèses H3}$$

et H7). L'estimateur est donc convergent.

Théorème de Gauss-Markov : L'estimateur [3] des moindres carrés est qualifié de BLUE (*Best Linear Unbiased Estimator*), car il s'agit du meilleur estimateur linéaire sans biais (au sens qu'il fournit les variances les plus faibles pour les estimateurs).

Il est à noter que l'estimateur du maximum de vraisemblance des paramètres fournit des résultats identiques à ceux de l'estimateur des MCO si l'hypothèse de normalité des erreurs est vérifiée.

Après un calcul matriciel¹, il apparaît que nous pouvons estimer sans biais σ_{ε}^2 par :

$$\hat{\sigma}_{\varepsilon}^2 = \frac{e' e}{n - k - 1} \quad [6]$$

En remplaçant la variance de l'erreur par son estimateur, nous obtenons :

$$\hat{\Omega}_{\hat{a}} = \hat{\sigma}_{\varepsilon}^2 (X' X)^{-1} \quad [7]$$

4) Le théorème de Frisch, Waugh et Lovell (FWL)

Supposons que les variables explicatives soient séparées en deux groupes représentés par les matrices X_1 et X_2 . Le modèle s'écrit : $Y = X_1 a_1 + X_2 a_2 + \varepsilon$.

Le théorème de FWL nous dit que l'estimateur des MCO du paramètre a_2 et des résidus sont les mêmes que ceux du modèle : $M_1 Y = M_1 X_2 a_2 + v$ avec $M_1 = I - X_1 (X_1' X_1)^{-1} X_1'$.

La matrice M_1 est symétrique et idempotente ($M_1' M_1 = M_1$), de plus $M_1 X_1 = 0$.

Ce théorème permet de bien comprendre le problème de la spécification d'un modèle : si une variable Y est effectivement expliquée par le groupe de variables X_1 et X_2 , il faut impérativement faire figurer dans le modèle ces deux groupes de variables même si seul le groupe de variables X_1 nous intéresse.

1. Voir démonstration en annexe C.

C. Équation d'analyse de la variance et qualité d'un ajustement

Comme pour le modèle de régression simple, nous avons :

$$\begin{aligned} a) \sum_t y_t &= \sum_t \hat{y}_t \rightarrow \bar{y} = \bar{\hat{y}} \\ b) \sum_t e_t &= 0 \end{aligned}$$

De ces deux relations, nous en déduisons l'équation fondamentale d'analyse de la variance :

$$\boxed{\begin{array}{lcl} \sum_t (y_t - \bar{y})^2 & = & \sum_t (\hat{y}_t - \bar{\hat{y}})^2 + \sum_t e_t^2 \\ SCT & = & SCE + SCR \end{array}} \quad [8]$$

La variabilité totale (SCT) est égale à la variabilité expliquée (SCE) + la variabilité des résidus (SCR).

Cette équation va nous permettre de juger de la qualité de l'ajustement d'un modèle ; en effet, plus la variance expliquée est « proche » de la variance totale, meilleur est l'ajustement global du modèle. C'est pourquoi nous calculons le rapport *SCE* sur *SCT* :

$$\boxed{R^2 = \frac{\sum_t (\hat{y}_t - \bar{\hat{y}})^2}{\sum_t (y_t - \bar{y})^2} = 1 - \frac{\sum_t e_t^2}{\sum_t (y_t - \bar{y})^2}} \quad [9]$$

R^2 est appelé le coefficient de détermination, et R le coefficient de corrélation multiple. R^2 mesure la proportion de la variance de Y expliquée par la régression de Y sur X .

Dans le cas de données centrées (moyenne nulle) et seulement dans ce cas, le coefficient de détermination est égal à :

$$\boxed{R^2 = \frac{\hat{Y}' \hat{Y}}{Y' Y} = 1 - \frac{e' e}{Y' Y}} \quad [10]$$

Cette qualité de l'ajustement et l'appréciation que l'on a du R^2 doivent être tempérées par le degré de liberté de l'estimation. En effet, lorsque le degré de

liberté est faible¹, il convient de corriger le R^2 afin de tenir compte du relativement faible nombre d'observations comparé au nombre de facteurs explicatifs par le calcul d'un R^2 « corrigé » noté \overline{R}^2 :

$$\overline{R}^2 = 1 - \frac{n-1}{n-k-1}(1-R^2) \quad [11]$$

On a $\overline{R}^2 < R^2$ et si n est grand $\overline{R}^2 \simeq R^2$.

Nous verrons au chapitre 4 section IV deux critères (Akaike et Schwarz) permettant d'arbitrer, lors de l'introduction d'une ou plusieurs variables explicatives, entre la perte de degrés de liberté et l'apport d'information.

Exercice n° 1

↓ fichier C3EX1

Mise sous forme matricielle d'un modèle et calcul des coefficients de régression

Soit le modèle à trois variables explicatives :

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + a_3 x_{3t} + \varepsilon_t$$

Nous disposons des données du tableau 1.

- 1) Mettre le modèle sous forme matricielle en spécifiant bien les dimensions de chacune des matrices.
- 2) Estimer les paramètres du modèle.
- 3) Calculer l'estimation de la variance de l'erreur ainsi que les écarts types de chacun des coefficients.
- 4) Calculer le R^2 et le \overline{R}^2 corrigé.

1. Dans le cas d'un modèle où le nombre d'observations est égal au nombre de variables explicatives (degré de liberté = 0), le $R^2 = 1$. Cependant, le pouvoir explicatif de ce modèle est nul.

Tableau 1 – Valeurs observées de y , x_1 , x_2 et x_3

t	y	x_1	x_2	x_3
1	12	2	45	121
2	14	1	43	132
3	10	3	43	154
4	16	6	47	145
5	14	7	42	129
6	19	8	41	156
7	21	8	32	132
8	19	5	33	147
9	21	5	41	128
10	16	8	38	163
11	19	4	32	161
12	21	9	31	172
13	25	12	35	174
14	21	7	29	180

Solution

1) Forme matricielle

Nous disposons de 14 observations et trois variables explicatives, le modèle peut donc s'écrire :

$$Y = Xa + \varepsilon \quad \text{avec :}$$

$$Y = \begin{pmatrix} 12 \\ 14 \\ 10 \\ \vdots \\ 21 \end{pmatrix}; \quad X = \begin{pmatrix} 1 & 2 & 45 & 121 \\ 1 & 1 & 43 & 132 \\ 1 & 3 & 43 & 154 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 7 & 29 & 180 \end{pmatrix}; \quad a = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix}; \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_t \\ \vdots \\ \varepsilon_{14} \end{pmatrix}$$

Dimensions :

(14,1)

(14,4)

(4,1)

(14,1)

2) Estimation des paramètres

Nous savons d'après [3] que $\hat{a} = (X'X)^{-1}X'Y$.

Calcul de $X'X$ et de $(X'X)^{-1}$

$$\begin{matrix} X' & X \end{matrix}$$

$$\begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 2 & 1 & 3 & \dots & 7 \\ 45 & 43 & 43 & \dots & 29 \\ 121 & 132 & 154 & \dots & 180 \end{pmatrix} \begin{pmatrix} 1 & 2 & 45 & 121 \\ 1 & 1 & 43 & 132 \\ 1 & 3 & 43 & 154 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 7 & 29 & 180 \end{pmatrix} =$$

$$= \begin{pmatrix} 14 & 85 & 532 & 2\,094 \\ 85 & 631 & 3\,126 & 13\,132 \\ 532 & 3\,126 & 20\,666 & 78\,683 \\ 2\,094 & 13\,132 & 78\,683 & 317\,950 \end{pmatrix}$$

$$(X' X)^{-1} = \begin{pmatrix} 20,16864 & 0,015065 & -0,23145 & -0,07617 \\ 0,015065 & 0,013204 & 0,001194 & -0,00094 \\ -0,23145 & 0,001194 & 0,003635 & 0,000575 \\ -0,07617 & -0,00094 & 0,000575 & 0,000401 \end{pmatrix}$$

Calcul de $X' Y$

$$\begin{matrix} & X' & & Y \\ \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 2 & 1 & 3 & \dots & 7 \\ 45 & 43 & 43 & \dots & 29 \\ 121 & 132 & 154 & \dots & 180 \end{pmatrix} & \begin{pmatrix} 12 \\ 14 \\ 10 \\ \vdots \\ 21 \end{pmatrix} & = & \begin{pmatrix} 248 \\ 1\,622 \\ 9\,202 \\ 37\,592 \end{pmatrix} \end{matrix}$$

Calcul de \hat{a}

$$\begin{matrix} & (X' X)^{-1} & & X' Y \\ \begin{pmatrix} 20,16864 & 0,015065 & -0,23145 & -0,07617 \\ 0,015065 & 0,013204 & 0,001194 & -0,00094 \\ -0,23145 & 0,001194 & 0,003635 & 0,000575 \\ -0,07617 & -0,00094 & 0,000575 & 0,000401 \end{pmatrix} & \begin{pmatrix} 248 \\ 1\,622 \\ 9\,202 \\ 37\,592 \end{pmatrix} & = & \hat{a} \\ & = & \begin{pmatrix} 32,89132 \\ 0,801900 \\ -0,38136 \\ -0,03713 \end{pmatrix} \end{matrix}$$

Soit $\hat{a}_0 = 32,89$; $\hat{a}_1 = 0,80$; $\hat{a}_2 = -0,38$; $\hat{a}_3 = -0,03$

Les calculs que nous venons de développer sont longs et fastidieux et mettent en évidence l'intérêt incontestable d'utiliser un ordinateur.

3) Calcul de $\hat{\sigma}_\varepsilon^2$ et de $\hat{\sigma}_a^2$

D'après [6] $\hat{\sigma}_\varepsilon^2 = \frac{e' e}{n - k - 1}$, nous devons donc calculer le résidu e .

$$e = Y - \hat{Y} = Y - X\hat{a}$$

Soit $e_t = y_t - (\hat{a}_0 + \hat{a}_1 x_{1t} + \hat{a}_2 x_{2t} + \hat{a}_3 x_{3t})$

$$e_t = y_t - 32,89 - 0,80 x_{1t} + 0,38 x_{2t} + 0,03 x_{3t}$$

Par exemple pour e_1 :

$$e_1 = y_1 - 32,89 - 0,80 x_{11} + 0,38 x_{21} + 0,03 x_{31}$$

$$e_1 = 12 - 32,89 - 0,80 \times 2 + 0,38 \times 45 + 0,03 \times 121 = -0,84$$

Le tableau 2 présente l'ensemble des résultats.

Par construction, la somme des résidus est bien nulle.

$$\hat{\sigma}_\varepsilon^2 = \frac{e' e}{n - k - 1} = \frac{\sum_{t=1}^{t=14} e_t^2}{14 - 3 - 1} = \frac{67,45}{10} = 6,745$$

Tableau 2 – Calcul du résidu

t	y_t	\hat{y}_t	e_t	e_t^2
1	12	12,84	- 0,84	0,71
2	14	12,39	1,61	2,58
3	10	13,18	- 3,18	10,11
4	16	13,39	1,61	2,58
5	14	17,70	- 3,70	13,67
6	19	17,88	1,12	1,26
7	21	22,20	- 1,20	1,44
8	19	18,86	0,14	0,02
9	21	16,51	4,49	20,14
10	16	18,76	- 2,76	7,63
11	19	17,92	1,08	1,17
12	21	21,90	- 0,90	0,81
13	25	22,71	2,29	5,27
14	21	20,76	0,24	0,06
Somme			0	67,45

La matrice des variances et covariances estimées des coefficients nous est donnée par [7], soit :

$$\hat{\Omega}_{\hat{a}} = \hat{\sigma}_\varepsilon^2 (X' X)^{-1}$$

$$\hat{\Omega}_{\hat{a}} = 6,745 \times \begin{pmatrix} 20,16864 & 0,015065 & -0,23145 & -0,07617 \\ 0,015065 & 0,013204 & 0,001194 & -0,00094 \\ -0,23145 & 0,001194 & 0,003635 & 0,000575 \\ -0,07617 & -0,00094 & -0,000575 & 0,000401 \end{pmatrix}$$

Les variances des coefficients de régression se trouvent sur la première diagonale :

$$\begin{aligned} \hat{\sigma}_{\hat{a}_0}^2 &= 6,745 \times 20,17 = 136,04 \rightarrow \hat{\sigma}_{\hat{a}_0} = 11,66 \\ \hat{\sigma}_{\hat{a}_1}^2 &= 6,745 \times 0,013 = 0,087 \rightarrow \hat{\sigma}_{\hat{a}_1} = 0,29 \\ \hat{\sigma}_{\hat{a}_2}^2 &= 6,745 \times 0,0036 = 0,024 \rightarrow \hat{\sigma}_{\hat{a}_2} = 0,15 \\ \hat{\sigma}_{\hat{a}_3}^2 &= 6,745 \times 0,0004 = 0,0026 \rightarrow \hat{\sigma}_{\hat{a}_3} = 0,05 \end{aligned}$$

4) Le calcul du R^2 est effectué à partir de la formule [9].

Nous connaissons $e' e = 67,45$, il convient de calculer $\sum_t (y_t - \bar{y})^2 = 226,86$.

$$R^2 = 1 - \frac{\sum_t e_t^2}{\sum_t (y_t - \bar{y})^2} = 1 - \frac{67,45}{226,86} = 0,702$$

Le \bar{R}^2 corrigé est donné par [11] :

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1} (1 - R^2) = 1 - \frac{14-1}{14-4} (1 - 0,702) = 0,613$$

Nous observons la baisse du coefficient de détermination lorsque nous le corrigeons par le degré de liberté.

III. Les tests statistiques

A. Le rôle des hypothèses

L'hypothèse de normalité des erreurs implique¹ que :

$$\sum_{i=1}^n \frac{e_i^2}{\sigma_\varepsilon^2} = (n - k - 1) \frac{\widehat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} = (n - k - 1) \frac{\widehat{\sigma}_{a_i}^2}{\sigma_{a_i}^2} \quad [12] \text{ suit une loi du } \chi^2 \text{ (chi-deux) à}$$

$n - k - 1$ degrés de liberté (somme au carré de $n - k - 1$ variables aléatoires *indépendantes* normales centrées réduites).

Il en résulte que :

$\Rightarrow \frac{\widehat{a}_i - a_i}{\widehat{\sigma}_{\widehat{a}_i}}$ (l'écart type théorique est remplacé par l'écart type empirique) suit une loi de Student à $n - k - 1$ degrés de liberté.

$$\text{En effet, } \frac{\widehat{a}_i - a_i}{\widehat{\sigma}_{\widehat{a}_i}} = \frac{\frac{\widehat{a}_i - a_i}{\sigma_{\widehat{a}_i}}}{\sqrt{(n - k - 1) \frac{\widehat{\sigma}_{\widehat{a}_i}^2}{\sigma_{\widehat{a}_i}^2} \frac{1}{(n - k - 1)}}} \text{ est le rapport d'une loi nor-}$$

male centrée réduite à la racine carrée d'un chi-deux divisé par son degré de liberté.

$(\widehat{a} - a)' \Omega_a^{-1} (\widehat{a} - a)$ suit une loi du χ^2 (chi-deux) à $k + 1$ degrés de liberté (somme au carré de $k + 1$ variables aléatoires normales centrées réduites, les $k + 1$ coefficients).

Si on remplace la matrice des variances covariances théoriques des coefficients, par son estimateur $\widehat{\Omega}_a = \widehat{\sigma}_\varepsilon^2 (X'X)^{-1}$, la loi de probabilité de $\frac{1}{k + 1} (\widehat{a} - a)' \widehat{\Omega}_a^{-1} (\widehat{a} - a)$ est alors un Fisher à $k + 1$ et $n - k - 1$ degrés de liberté.

1. L'ensemble de ces résultats sont une généralisation à k variables explicatives des résultats présentés au chapitre 2, paragraphe III.B.

$$\text{En effet, } F = \frac{\frac{1}{k+1}(\widehat{a} - a)'[\sigma_\varepsilon^2(X'X)^{-1}]^{-1}(\widehat{a} - a)}{(n-k-1)\frac{\widehat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}\frac{1}{(n-k-1)}} \text{ est le rapport de deux}$$

chi-deux divisés par leurs degrés de liberté (caractéristique d'une loi de Fisher à $k+1$ et $n-k-1$ degrés de liberté).

B. Construction des tests

À partir des différentes relations définies ci-dessus, nous pouvons mettre en place un certain nombre de tests statistiques que nous allons expliciter.

1) Comparaison d'un paramètre a_i à une valeur fixée \bar{a}

Le test d'hypothèses est le suivant :

$$H0 : a_i = \bar{a}$$

$$H1 : a_i \neq \bar{a}$$

Nous savons que :

$\frac{\widehat{a}_i - a_i}{\widehat{\sigma}_{\widehat{a}_i}}$ suit une loi de Student à $n - k - 1$ degrés de liberté.

Sous l'hypothèse $H0$, cette relation devient¹ :

$$\frac{|\widehat{a}_i - \bar{a}|}{\widehat{\sigma}_{\widehat{a}_i}} = t_{\widehat{a}_i}^* \rightarrow \text{loi de Student } n - k - 1 \text{ degrés de liberté} \quad [13]$$

Si $t_{\widehat{a}_i}^* > t_{n-k-1}^{\alpha/2}$ alors nous rejetons l'hypothèse $H0$, a_i est significativement différent de \bar{a} (au seuil de α).

Si $t_{\widehat{a}_i}^* \leq t_{n-k-1}^{\alpha/2}$ alors nous acceptons l'hypothèse $H0$, a_i n'est pas significativement différent de \bar{a} (au seuil de α).

Cas particulier : test par rapport à une valeur particulière $\bar{a} = 0$.

Si nous désirons savoir si une variable explicative figurant dans un modèle est réellement – significativement – contributive pour expliquer la variable endogène, il convient de tester si son coefficient de régression est significativement différent de 0 pour un seuil choisi, en général $\alpha = 5 \%$.

1. Comme il s'agit d'un test bilatéral, on calcule la valeur absolue du t empirique.

La relation [13], sous $H_0 (a_i = 0)$, devient :

$$\left| \frac{\widehat{a}_i}{\widehat{\sigma}_{\widehat{a}_i}} \right| = t_{\widehat{a}_i}^* \rightarrow \text{loi de Student à } n - k - 1 \text{ degrés de liberté} \quad [14]$$

$t_{\widehat{a}_i}^*$ est appelé le *ratio de Student*, les règles de décision citées plus haut s'appliquent alors.

Ce test est très important ; en effet, si dans un modèle estimé, un des coefficients (hormis le terme constant) n'est pas significativement différent de 0, il convient d'éliminer cette variable¹ et de ré-estimer les coefficients du modèle. La cause de cette non-significativité, est due :

- soit à une absence de corrélation avec la variable à expliquer,
- soit à une colinéarité trop élevée avec une des variables explicatives.

2) Comparaison d'un ensemble de paramètres à un ensemble de valeurs fixées

Nous cherchons à tester *simultanément* l'égalité d'un sous-ensemble de coefficients de régression à des valeurs fixées.

$$H_0 : a_q = \bar{a}_q$$

$$H_1 : a_q \neq \bar{a}_q$$

q étant le nombre de coefficients retenus, c'est-à-dire la dimension de chacun des vecteurs a_q .

Nous avons démontré que $\frac{1}{k+1}(\widehat{a} - a)' \widehat{\Omega}_a^{-1} (\widehat{a} - a)$ suit une loi de Fisher à $k+1$ et $n-k-1$ degrés de liberté ; pour un sous-ensemble de paramètres q , l'expression $\frac{1}{q}(\widehat{a}_q - a_q)' \widehat{\Omega}_{\widehat{a}_q, q}^{-1} (\widehat{a}_q - a_q)$ suit alors une loi de Fisher à q et $n-k-1$ degrés de liberté.

Pour accepter H_0 , il suffit que :

$$\boxed{\frac{1}{q}(\widehat{a}_q - \bar{a}_q)' \widehat{\Omega}_{\widehat{a}_q, q}^{-1} (\widehat{a}_q - \bar{a}_q) \leq F^\alpha(q, n-k-1)} \quad [15]$$

$F^\alpha(q, n-k-1) =$ loi de Fisher au seuil α à q et $n-k-1$ degrés de liberté.

1. En effet, elle n'apporte aucune contribution et dégrade l'estimation des autres variables.

3) Intervalle de confiance de la variance de l'erreur

L'intervalle de confiance de la variance de l'erreur permet de déterminer une fourchette de variation de l'amplitude de l'erreur. Pour un intervalle à $(1 - \alpha)\%$, il est donné par :

$$IC = \left[\frac{(n - k - 1) \hat{\sigma}_\varepsilon^2}{\chi_1^2} ; \frac{(n - k - 1) \hat{\sigma}_\varepsilon^2}{\chi_2^2} \right] \quad [16]$$

Avec χ_1^2 à $n - k - 1$ degrés de liberté et $\alpha/2$ de probabilité¹ d'être dépassée et χ_2^2 à $n - k - 1$ degrés de liberté et $(1 - \alpha/2)$ de probabilité d'être dépassée.

C. Tests sur les résidus : valeur anormale, effet de levier et point d'influence

Nous n'aborderons pas ici les tests sur les résidus remettant en cause les hypothèses d'homoscédasticité ou d'indépendance, qui sont traités au chapitre 5, mais trois notions qui sont proches et néanmoins différentes dans leur conséquence sur l'estimation des paramètres du modèle et leurs précisions.

1) La matrice *HAT*

La matrice « *HAT*² », notée *H*, joue un rôle essentiel dans la détection de l'effet de levier.

Nous calculons la matrice « *HAT* » $H = X(X'X)^{-1}X'$.

Les éléments de la première diagonale de cette matrice *H* sont appelés les leviers, qui déterminent l'influence de l'observation *i* sur les estimations obtenues par la régression.

Le levier est situé sur la première diagonale de cette matrice soit $h_i = x_i(X'X)^{-1}x_i'$

Deux propriétés : $0 \leq h_i \leq 1$ et $\sum_{i=1}^n h_i = k + 1$, la somme des éléments de la première diagonale de la matrice *H* est égale au nombre de paramètres estimés du modèle.

-
1. Attention, la loi du chi-deux n'est pas symétrique, il convient donc de lire sur la table les deux probabilités $1 - \alpha/2$ et $\alpha/2$.
 2. Cette matrice permet de passer du vecteur *Y* au vecteur \hat{Y} d'où le nom de matrice *HAT* (chapeau en anglais), en effet nous avons $\hat{Y} = X\hat{a} = X(X'X)^{-1}X'Y = HY$.

Si chaque observation pèse le même poids, alors les valeurs des h_i doivent être proches de $\frac{k+1}{n}$.

Le levier d'une observation i est donc anormalement élevé si : $h_i > 2\frac{k+1}{n}$, l'observation est alors considérée comme un point de levier (*leverage point*) ou point d'influence.

2) Point de levier et valeur anormale

Une observation exerce un effet de levier si elle est éloignée des autres en termes de combinaison des variables explicatives ; c'est-à-dire que les valeurs prises par les variables explicatives, pour cette observation de la variable à expliquer, sont inattendues. Par exemple, un pays dont la population est faible mais le PIB élevé, chaque facteur explicatif pris individuellement n'est pas surprenant, mais la survenance de deux valeurs à la fois pour un pays est insolite.

Le point d'influence est une observation qui contribue très fortement au pouvoir explicatif du modèle (sans cette valeur la régression peut être non significative !), l'observation pèse de manière exagérée dans la régression, au point que les résultats obtenus sont très différents selon que nous l'intégrons ou non dans la régression.

La valeur prise par la variable explicative est anormale si le résidu de cette observation est beaucoup plus élevé que les autres résidus, pour identifier une valeur anormale nous pouvons calculer le résidu standardisé (ou encore appelé le résidu studentisé). Une autre approche permettant de détecter des valeurs anormales à l'aide de variables indicatrices sera présentée dans ce chapitre à la section V.

3) Résidu standardisé (ou studentisé)

Les résidus standardisés¹ notés e_i^S permettent de détecter des valeurs anormales. Le résidu e_i est divisé par son écart type estimé pondéré par le levier² :

$e_i^S = \frac{e_i}{\hat{\sigma}_e \sqrt{1 - h_i}}$ suit une loi de Student à $n - k - 1$ degrés de liberté, avec

$\hat{\sigma}_e = \sqrt{\frac{\sum_i e_i^2}{n - k - 1}}$. Si, par exemple, les résidus standardisés e_i^S sont compris

dans l'intervalle $\pm t_{n-k-1}^{0,025}$, on ne suspecte pas de valeurs anormales pour un seuil de confiance 95 %.

1. Dans cette méthode de calcul nous parlons de résidus studentisés internes car le calcul porte sur l'ensemble des observations.

2. Certains auteurs divisent le résidu par son écart type $e_i^S = \frac{e_i}{\hat{\sigma}_e}$ sans intégrer la valeur du levier.

Il est aussi possible de calculer des résidus studentisés externes, ils sont calculés de la même manière mais en excluant du calcul l'observation i .

Exercice n° 2

 **fichier C3EX1**

Tests statistiques sur les coefficients et la variance de l'erreur : calcul des leviers et des résidus studentisés

En reprenant les données du tableau 1 et les résultats de l'exercice 1, on demande de répondre aux questions suivantes :

- 1) Les variables explicatives sont-elles significativement contributives pour expliquer la variable endogène ?
- 2) Le coefficient a_1 est-il significativement inférieur à 1 ?
- 3) Les coefficients a_1 et a_2 sont-ils simultanément et significativement différents de 1 et $-0,5$?
- 4) Quel est l'intervalle de confiance pour la variance de l'erreur ?
- 5) Calculer les leviers et les résidus standardisés, existe-t-il des valeurs aberrantes ?
(Les seuils choisis seront de 5 %.)

Solution

- 1) Il convient de calculer les trois ratios de Student et de les comparer à la valeur lue dans la table pour un seuil de 5 % (la table de Student en fin d'ouvrage indique directement les valeurs de α pour un test bilatéral).

D'après [14], nous obtenons :

$$\frac{\hat{a}_1}{\hat{\sigma}_{\hat{a}_1}} = \frac{0,80}{0,29} = t_{\hat{a}_1}^* = 2,75 > t_{10}^{0,05} = 2,228 \rightarrow a_1 \neq 0, \text{ la variable explicative } x_1 \text{ est}$$

contributive à l'explication de y ; de même :

$$\frac{\hat{a}_2}{\hat{\sigma}_{\hat{a}_2}} = \left| \frac{-0,38}{0,15} \right| = t_{\hat{a}_2}^* = 2,53 > t_{10}^{0,05} = 2,228 \rightarrow a_2 \neq 0$$

$$\frac{\hat{a}_3}{\hat{\sigma}_{\hat{a}_3}} = \left| \frac{-0,03}{0,05} \right| = t_{\hat{a}_3}^* = 0,60 < t_{10}^{0,05} = 2,228 \rightarrow a_3 = 0$$

la variable x_2 est explicative de y alors que la variable x_3 n'est pas contributive à l'explication de y , il convient donc de la retirer de ce modèle et de procéder à une nouvelle estimation.

Nous aurions pu tout aussi bien répondre à cette question en calculant les intervalles de confiance de chacun des coefficients :

$$IC_{a_1} = \hat{a}_1 \pm t_{n-k-1}^{0,05} \cdot \hat{\sigma}_{\hat{a}_1} = 0,80 \pm 2,228 \times 0,29 = [0,14 ; 1,45]$$

De même nous obtenons :

$$IC_{a_2} = [-0,71 ; -0,04] \quad \text{et} \quad IC_{a_3} = [-0,14 ; 0,08]$$

La valeur 0 n'appartient pas à l'intervalle de confiance à 95 % de a_1 et a_2 , donc ces deux coefficients sont significativement différents de 0 ; en revanche, 0 appartient à l'intervalle de confiance de a_3 , ce coefficient n'est pas significativement différent de 0.

2) Nous posons le test d'hypothèses suivant :

$$H_0 : a_1 = 1$$

$$H_1 : a_1 < 1$$

Sous H_0 , la relation [13] s'écrit :

$$\frac{\hat{a}_1 - a_1}{\hat{\sigma}_{\hat{a}_1}} = \frac{0,80 - 1}{0,29} = -0,68 > -t_{10}^{0,05} = -1,81^1 \Rightarrow \text{acceptation de } H_0$$

Nous sommes bien dans la zone d'acceptation de H_0 (cf. graphique 4 du chapitre 2).

Par souci de simplification, nous pouvons procéder au test de Student en profitant de la symétrie de cette loi, soit à calculer :

$$\frac{|\hat{a}_1 - a_1|}{\hat{\sigma}_{\hat{a}_1}} = \frac{|0,80 - 1|}{0,29} = 0,68 < t_{10}^{0,05} = 1,81 \Rightarrow \text{acceptation de } H_0$$

Le fait de raisonner sur la valeur absolue du numérateur entraîne une lecture directe de la table et ainsi une construction et interprétation immédiate du test.

3) Le test d'hypothèses est le suivant :

$$H_0 : \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -0,5 \end{pmatrix}$$

$$H_1 : \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \neq \begin{pmatrix} 1 \\ -0,5 \end{pmatrix}$$

Examinons les différents éléments de la relation [15] sous H_0 :

$$\frac{1}{q} (\hat{a}_q - a_q)' \hat{\Omega}_{\hat{a}_q}^{-1} (\hat{a}_q - a_q)$$

Nous avons : $q = 2$, $\hat{a}_q = \begin{pmatrix} 0,80 \\ -0,38 \end{pmatrix}$ et $a_q = \begin{pmatrix} 1 \\ -0,5 \end{pmatrix}$. La matrice des variances

et covariances des coefficients a été calculée lors de l'exercice 1, nous ne retenons que la sous-matrice de dimension 2×2 correspondant aux deux coefficients de régression faisant l'objet du test.

$$\hat{\Omega}_{\hat{a}_q} = 6,745 \cdot \begin{pmatrix} 0,013204 & 0,001194 \\ 0,001194 & 0,003635 \end{pmatrix} \rightarrow \hat{\Omega}_{\hat{a}_q}^{-1} = \begin{pmatrix} 11,57140 & -3,80213 \\ -3,80213 & 42,03506 \end{pmatrix}$$

1. Attention, le test est unilatéral, il convient de lire sur la table de Student tabulée pour $\alpha/2$ à un seuil de $0,10 = 2 \times 0,05$.

$$F^* = \frac{1}{2} (0,80 - 1; -0,38 + 0,5) \begin{pmatrix} 11,57140 & -3,80213 \\ -3,80213 & 42,03506 \end{pmatrix} \times \begin{pmatrix} 0,80 - 1 \\ -0,38 + 0,5 \end{pmatrix}$$

$F^* = 0,612$ est à comparer à $F^\alpha(q, n - k - 1) = F_{2,10}^{0,05} = 4,10$, le F^* empirique est inférieur au F lu dans la table, on accepte l'hypothèse H_0 . Les données ne sont pas incompatibles avec la possibilité que les coefficients a_1 et a_2 soient simultanément et respectivement égaux à 1 et $-0,5$.

- 4) L'intervalle de confiance de la variance de l'erreur à un seuil $(1 - \alpha)\% = 95\%$ ($\alpha = 0,05$) est calculé à partir de la formule [16] pour 10 degrés de liberté :

$$IC = \left[\frac{(n - k - 1) \hat{\sigma}_\varepsilon^2}{\chi_{0,025}^2}; \frac{(n - k - 1) \hat{\sigma}_\varepsilon^2}{\chi_{0,975}^2} \right] = \left[\frac{10 \times 6,745}{20,48}; \frac{10 \times 6,745}{3,25} \right]$$

Soit $3,30 \leq \sigma_\varepsilon^2 \leq 20,75$. La variance vraie (mais inconnue) σ_ε^2 de l'erreur a 95 % de chance de se situer à l'intérieur de cet intervalle.


- 5) Le calcul de h_i et des résidus standardisés sont effectués à l'aide du programme  C3EX2.PR.G. Les résultats sont présentés sur le tableau 3.

Tableau 3 – Valeur des leviers h_i et des résidus standardisés e_i^S

	Résidus		
i	e_i	h_i	e_i^S
1	-0,8408	0,2790	-0,3813
2	1,6068	0,2966	0,7377
3	-3,1800	0,3091	-1,4732
4	1,6055	0,3248	0,7523
5	-3,6973	0,2609	-1,6559
6	1,1220	0,1825	0,4778
7	-1,2015	0,5327	-0,6768
8	0,1426	0,2025	0,0615
9	4,4880	0,1804	1,9088
10	-2,7622	0,1442	-1,1497
11	1,0830	0,3066	0,5008
12	-0,8994	0,2115	-0,3900
13	2,2946	0,4086	1,1489
14	0,2387	0,3605	0,1149

Le seuil du levier est égal à $2 \frac{k+1}{n} = 2 \frac{4}{14} = 0,57$, aucune valeur n'est supérieure à 0,57, nous ne détectons pas de point de levier (ou de point d'influence).

Les résidus studentisés sont tous dans l'intervalle $\pm t_{10}^{0,025} = \pm 2,228$, nous ne détectons pas de valeur anormale.

IV. L'analyse de la variance

A. Construction du tableau d'analyse de la variance et test de signification globale d'une régression

Dans cette section, nous allons nous interroger sur la signification globale du modèle de régression, c'est-à-dire si l'ensemble des variables explicatives a une influence sur la variable à expliquer. Ce test peut être formulé de la manière suivante : existe-t-il au moins une variable explicative significative ? Soit le test d'hypothèses :

$H_0 : a_1 = a_2 = \dots = a_k = 0$ (tous les coefficients sont nuls¹)

H_1 : il existe au moins un des coefficients non nul

Nous ne testons pas le cas où le terme constant a_0 est nul, car seules nous intéressent les variables explicatives. Un modèle dans lequel seul le terme constant est significatif n'a aucun sens économique.

Le cas où l'hypothèse H_0 est acceptée signifie qu'il n'existe aucune relation linéaire significative entre la variable à expliquer et les variables explicatives (ou encore que la Somme des Carrés Expliqués n'est pas significativement différente de 0).

Nous reprenons l'équation fondamentale [8] d'analyse de la variance :

$$\sum_i (y_i - \bar{y})^2 = \sum_i (\hat{y}_i - \bar{y})^2 + \sum_i e_i^2$$

La régression est jugée significative si la variabilité expliquée est significativement différente de 0. Le tableau 4 présente le tableau d'analyse de la variance² permettant d'effectuer le test de Fisher.

$$F^* = \frac{\sum_i (\hat{y}_i - \bar{y})^2 / k}{\sum_i e_i^2 / (n - k - 1)} = \frac{R^2 / k}{(1 - R^2) / (n - k - 1)}$$

(d'après [9])

[17]

1. Nous remarquons que nous pouvons répondre à cette question par le test d'un sous-ensemble de coefficients [15], le test ici présenté conduit évidemment à des résultats identiques.

2. Voir chapitre 2, paragraphe 4, pour la construction de ce tableau.

Tableau 4 – Analyse de la variance pour une régression multiple

Source de variation	Somme des carrés	Degré de liberté	Carrés moyens
x_1, x_2, \dots, x_k	$SCE = \sum_t (\hat{y}_t - \bar{y})^2$	k	SCE/k
Résidu	$SCR = \sum_t e_t^2$	$n - k - 1$	$SCR/(n - k - 1)$
Total	$SCT = \sum_t (y_t - \bar{y})^2$	$n - 1$	

L'hypothèse de normalité des erreurs implique que sous l'hypothèse H_0 , F^* suit une loi de Fisher (rapport de deux chi-deux). Nous comparons donc ce F^* calculé au F théorique à k et $(n - k - 1)$ degrés de liberté : si $F^* > F$ nous rejetons l'hypothèse H_0 , le modèle est globalement explicatif.

Dans la pratique, ce test est effectué immédiatement grâce à la connaissance du coefficient de détermination R^2 (seulement si le modèle comporte un terme constant) qui permet de calculer le Fisher empirique (calculé).

B. Autres tests à partir du tableau d'analyse de la variance

À partir d'un exercice, nous allons présenter quatre tests usuels se référant à l'analyse de la variance, la généralisation de ces tests ne posera par la suite aucune difficulté au lecteur.

1) Introduction d'une ou de plusieurs variables explicatives supplémentaires

L'ajout d'un bloc supplémentaire de variables explicatives améliore-t-il significativement la qualité de l'ajustement ?

2) Stabilité des coefficients du modèle dans le temps (test de CHOW)

Peut-on considérer le modèle comme étant stable sur la totalité de la période, ou bien doit-on considérer deux sous-périodes distinctes d'estimation (changement structurel du modèle) ? La spécification du modèle est la même, mais les valeurs estimées des coefficients pour les deux échantillons sont différentes.

3) Test de restrictions et de contraintes sur les coefficients

Les contraintes souhaitées ou envisagées sur les coefficients sont-elles justifiées et validées par l'estimation économétrique ?

4) Augmentation de la taille de l'échantillon servant à estimer le modèle

Lorsque la taille de l'échantillon augmente (le nombre d'observations à disposition est plus important), le modèle reste-t-il stable ? Ce test se ramène au test de Chow de stabilité des coefficients sur deux sous-périodes. L'estimation sur la sous-période 1 est effectuée à partir de l'échantillon initial et l'estimation de la sous-période 2 à partir des nouvelles observations.

Nous voyons l'intérêt pratique de ces tests et l'apport pour l'économiste des réponses à ces questions.

Exercice n° 3

↓ fichier C3EX1

Tests à partir de l'analyse de la variance

En reprenant les données de l'exercice 1 (tableau 1), dont nous rappelons les résultats de l'estimation du modèle¹ :

$$\begin{aligned}y_t &= 32,89 + 0,80 x_{1t} - 0,38 x_{2t} - 0,03 x_{3t} + e_t \\ &\quad (11,66) \quad (0,29) \quad (0,15) \quad (0,05) \\ R^2 &= 0,702 \\ n &= 14 \\ (\cdot) &= \text{écart type des coefficients}\end{aligned}$$

on demande de tester les hypothèses suivantes.

- 1) L'ajout des variables explicatives x_2 et x_3 améliore-t-il significativement la qualité de l'estimation par rapport à x_1 seul ?
- 2) Peut-on considérer le modèle (à trois variables explicatives) comme stable sur l'ensemble de la période, ou doit-on procéder à deux estimations, l'une de la période 1 à 7, et l'autre de la période 8 à 14 ?
- 3) Un économiste suggère que dans ce modèle $a_1 = 1$ et $a_2 = a_3$, qu'en pensez-vous ?

Solution²

Nous pouvons tout d'abord appliquer le test de Fisher [17] afin de tester la signification globale de la régression à trois variables x_1 , x_2 et x_3 .

$$F^* = \frac{R^2/k}{(1 - R^2)/(n - k - 1)} = \frac{0,702/3}{(1 - 0,702)/10} = 7,878 > F_{3,10}^{0,05} = 3,71$$

1. Le lecteur notera la présentation « standard » des résultats d'estimation d'un modèle. Les informations listées ici doivent impérativement figurer. À noter que le t de Student est souvent indiqué à la place de l'écart type des coefficients afin de pouvoir, sans aucun calcul, procéder aux tests de significativité des coefficients.
2. Les calculs sont effectués à partir d'un logiciel, il peut apparaître de légères différences entre les calculs manuels et les résultats, imputables au fait que le logiciel tient compte d'un nombre élevé de décimales.

Nous rejetons l'hypothèse H_0 de nullité de tous les coefficients, la régression est globalement significative.

1) Test d'ajout de variables

Étape 1 : calcul de la variabilité totale, expliquée et résiduelle sur le modèle complet. Les résultats calculés précédemment lors de l'exercice 1 nous ont donné :

$$SCT = \sum_i (y_i - \bar{y})^2 = 226,86; \quad SCE = \sum_i (\hat{y}_i - \bar{y})^2 = 159,41; \quad SCR = e' e = 67,45$$

Étape 2 : calcul de la variabilité totale, expliquée et résiduelle sur le modèle à une seule variable explicative x_1 .

Le modèle estimé est le suivant :

$$\begin{aligned} y_t &= 1,011x_{1,t} + 11,57 + e_t \\ &\quad (0,281) \\ n &= 14 \\ R^2 &= 0,52 \\ (.) &= \text{Ecart type} \\ \hat{\sigma}_e &= 3,0165 \end{aligned}$$

Nous calculons d'abord

$$SCR^1 = e' e = \text{d.d.l.} \times \hat{\sigma}_e^2 = 12 \times 3,0165^2 = 109,20$$

puis à partir du coefficient de détermination R^2 , nous déduisons :

$$SCT^1 = 226,86 \text{ et } SCE^1 = 117,65$$

Le test d'hypothèses est le suivant :

$$\begin{aligned} H_0 : a_2 = a_3 &= 0 \\ H_1 : \text{il existe au moins un des deux coefficients non nul.} \end{aligned}$$

Ce test se ramène à un test par analyse de la variance : le fait d'ajouter des variables explicatives dans un modèle entraîne automatiquement une augmentation¹ de SCE (et donc une diminution de SCR) ; on souhaite donc tester que la différence entre SCE et SCE^1 soit significativement positive (ou bien que la différence entre SCR^1 et SCR soit significativement positive, il s'agit du même test). On compare donc la différence par rapport à la somme des carrés la plus faible, soit ici SCR . Le tableau 5 d'analyse de la variance permet de procéder au test de Fisher.

Étape 3 : tableau d'analyse de la variance.

Tableau 5 – Tableau d'analyse de la variance pour tester l'ajout d'un bloc de variables explicatives

Source de variation	Somme des carrés	Degré de liberté	Carrés moyens
x_1	$SCE^1 = 117,65$	1	117,65
x_1, x_2, x_3	$SCE = 159,41$	3	53,14
Résidu	$SCR = 67,45$	10	6,74
Total	$SCT = 226,85$	13	

1. Sauf si la ou les variables ajoutées sont orthogonales à la variable à expliquer, SCE reste alors identique. Ce cas est évidemment rare.

Étape 4 : calcul du Fisher empirique.

$$F^* = \frac{(SCE - SCE^1)/(k - k')}{SCR/(n - k - 1)} = \frac{41,67/(3 - 1)}{67,45/10} = 3,09 < F_{2,10}^{0,05} = 4,10$$

Ou encore : $F^* : \frac{(SCR^1 - SCR)/(k - k')}{SCR/(n - k - 1)} = \frac{(109,2 - 67,45)/2}{67,45/10} = 3,09$

Avec k = nombre de variables explicatives du modèle complet et k' = nombre de variables explicatives du modèle sans l'ajout du bloc de variables. Nous acceptons l'hypothèse H_0 , il n'y a donc pas de différence significative entre les deux variances expliquées, l'ajout des variables explicatives x_2 et x_3 n'améliore pas de manière significative – au seuil de 5 % – le pouvoir explicatif du modèle.

2) Le modèle est-il stable sur la totalité de la période ?

Soit le modèle estimé sur une seule période :

$$y_t = \hat{a}_1 x_{1t} + \hat{a}_2 x_{2t} + \hat{a}_3 x_{3t} + \hat{a}_0 + e_t \text{ pour } t = 1, \dots, 14$$

ou le modèle estimé sur deux sous-périodes :

$$y_t = \hat{a}_1^1 x_{1t} + \hat{a}_2^1 x_{2t} + \hat{a}_3^1 x_{3t} + \hat{a}_0^1 + e_t \text{ pour } t = 1, \dots, 7$$

$$y_t = \hat{a}_1^2 x_{1t} + \hat{a}_2^2 x_{2t} + \hat{a}_3^2 x_{3t} + \hat{a}_0^2 + e_t \text{ pour } t = 8, \dots, 14$$

Le test d'hypothèses jointes est alors le suivant :

$$H_0 : \begin{pmatrix} a_1 = a_1^1 = a_1^2 \\ a_2 = a_2^1 = a_2^2 \\ a_3 = a_3^1 = a_3^2 \\ a_0 = a_0^1 = a_0^2 \end{pmatrix}$$

Ce test de stabilité des coefficients (test de Chow) se ramène à la question suivante : existe-t-il une différence significative entre la somme des carrés des résidus (SCR) de l'ensemble de la période et l'addition de la somme des carrés des résidus calculée à partir des deux sous-périodes ($SCR^1 + SCR^2$) ?

En effet, dans le cas d'une réponse négative, cela signifie que le fait de scinder en deux échantillons n'améliore pas la qualité du modèle, donc qu'il est stable sur la totalité de la période.

Les étapes sont alors les suivantes.

Étape 1 : estimation du modèle sur chacune des deux sous-périodes¹ et calcul des sommes des carrés de résidus.

sous-période 1 : données de 1 à 7

$$y_t = 0,774x_{1,t} - 0,293x_{2,t} - 0,012x_{3,t} + 25,27 + e_t$$

(0,53) (0,31) (0,10)

$$n = 7$$

$$R^2 = 0,692$$

$$(.) = \text{Ecart type}$$

$$\hat{\sigma}_e = 3,01759$$

1. Les deux sous-périodes peuvent être de longueur inégale, cependant elles doivent impérativement recouvrir la totalité des observations de la période.

Nous pouvons en déduire comme précédemment :

$$SCT^1 = 88,85; SCE^1 = 61,54; SCR^1 = 27,31$$

sous-période 2 : données de 8 à 14

$$y_t = 1,228x_{1,t} - 0,620x_{2,t} - 0,184x_{3,t} + 62,63 + e_t$$

(0,69) (0,52) (0,15)

$$n = 7$$

$$R^2 = 0,543$$

$$(.) = \text{Ecart type}$$

$$\hat{\sigma}_e = 2,6281$$

$$D'où SCT^2 = 45,43; SCE^2 = 24,70; SCR^2 = 20,73.$$

Étape 2 : calcul du Fisher empirique.

En prenant au dénominateur la plus faible des sommes des carrés (soit $SCR^1 + SCR^2$), le Fisher empirique est égal à :

$$F^* = \frac{[SCR - (SCR^1 + SCR^2)]/ddl_n}{(SCR^1 + SCR^2)/ddl_d}$$

$$\text{avec } ddl_n = (n - k - 1) - [(n_1 - k - 1) + (n_2 - k - 1)] = k + 1 = 4$$

$$\text{car } n = n_1 + n_2$$

$$ddl_d = (n_1 - k - 1) + (n_2 - k - 1) = n - 2(k + 1) = 6$$

d'où

$$F^* = \frac{[(67,45 - (27,31 + 20,73))]/4}{(27,31 + 20,73)/6} = \frac{4,852}{8,00} = 0,606 < F_{4;6}^{0,05} = 4,53$$

L'hypothèse H_0 est acceptée, les coefficients sont significativement stables sur l'ensemble de la période.

Attention, en cas d'hétéroscédasticité (cf. chapitre 5), le test de Chow est biaisé dans le sens d'une surestimation du seuil de rejet du test, nous rejetons trop souvent l'hypothèse H_0 .

3) Test de $a_1 = 1$ et $a_2 = a_3$

Si cette hypothèse est vérifiée, le modèle :

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + a_3 x_{3t} + \varepsilon_t$$

peut s'écrire :

$$y_t = a_0 + 1 x_{1t} + a_2 x_{2t} + a_2 x_{3t} + \varepsilon_t$$

ou encore :

$$\begin{aligned} y_t - x_{1t} &= a_0 + a_2(x_{2t} + x_{3t}) + \varepsilon_t \\ z_t &= a_0 + a_2 v_t + \varepsilon_t \end{aligned}$$

Il convient de constituer la nouvelle variable à expliquer, z_t , et la nouvelle variable explicative v_t , puis d'effectuer la régression de z_t sur v_t .

Le tableau 6 présente ces nouvelles variables.

L'estimation des deux $(k' + 1)$ coefficients du modèle conduit aux résultats suivants :

$$\begin{aligned}
 z_t &= -0,0111v_t + 13,74 + e_t \\
 &\quad (0,051) \\
 n &= 14 \\
 R^2 &= 0,0389 \\
 (.) &= \text{Ecart type} \\
 \hat{\sigma}_\varepsilon &= 3,0109
 \end{aligned}$$

Tableau 6 – Variables transformées sous l'hypothèse de vérification des contraintes

t	$z_t = y_t - x_{1t}$	$v_t = x_{2t} + x_{3t}$
1	10	166
2	13	175
3	7	197
4	10	192
5	7	171
6	11	197
7	13	164
8	14	180
9	16	169
10	8	201
11	15	193
12	12	203
13	13	209
14	14	209

Nous pouvons en déduire :

$$SCT^1 = 109,21 ; SCE^1 = 0,425 ; SCR^1 = 108,78$$

L'hypothèse à tester est donc :

H0 : les restrictions sont toutes vérifiées ($SCR^1 = SCR$).

H1 : il existe au moins une restriction non vérifiée ($SCR^1 \neq SCR$).

Le Fisher empirique est donné par :

$$F^* = \frac{(SCR^1 - SCR)/ddl_n}{SCR/(n - k - 1)} = \frac{(108,78 - 67,45)/2}{67,45/10} = 3,06 < F_{2,10}^{0,05} = 4,10$$

avec $ddl_n = (n - k' - 1) - (n - k - 1) = k - k' = 2$.

L'hypothèse H0 est acceptée, les contraintes envisagées sur les coefficients sont compatibles avec les données.

C. Généralisation des tests par analyse de la variance

Nous pouvons remarquer que tous ces tests par analyse de la variance se ramènent à un test unique, tel que les éléments du vecteur des coefficients α vérifient un ensemble de q contraintes linéaires :

$$H_0 : Ra = r$$

$$H_1 : Ra \neq r$$

Ainsi le Fisher empirique est donné par :

$$F^* = \frac{\{(R\hat{a} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{a} - r)\}/q}{SCR/(n - k - 1)}$$

où \hat{a} est le vecteur des coefficients estimés sur le modèle non contraint.

On rejette H_0 si le F^* est supérieur au F lu à q et $n - k$ degrés de liberté.

Quelques exemples d'utilisation pour un modèle à k variables explicatives :

– Test sur un coefficient de régression. Hypothèse $H_0 : a_3 = 0,5$.

Soit à tester l'égalité du 3^e coefficient par rapport à 0,5 ; la contrainte s'écrit $Ra = r$ avec $R = (0\ 0\ 0\ 1\ \dots\ 0\ 0\ 0)$ et $r = (0\ 0\ 0\ 0,5\ \dots\ 0\ 0\ 0)$. Le premier élément des vecteurs correspond au terme constant a_0 . Ce test peut aussi être mené par un classique test de Student.

– Test d'égalité de coefficients. Hypothèse $H_0 : a_1 = a_3 \rightarrow a_1 - a_3 = 0$.

La contrainte s'écrit $Ra = r$ avec $R = (0\ 1\ 0\ -1\ \dots\ 0\ 0\ 0)$

et $r = (0\ 0\ 0\ 0\ \dots\ 0\ 0\ 0)$.

– Test de significativité globale de la régression.

Hypothèse $H_0 : a_1 = a_2 = a_3 = \dots = a_k = 0$.

Ce test de Fisher (cf. équation [17]) est équivalent à l'écriture contrainte $Ra = r$ avec $R = (0\ 1\ 1\ 1\ \dots\ 1\ 1)$ et $r = (0\ 0\ 0\ 0\ \dots\ 0\ 0\ 0)$.

De manière équivalente, nous pouvons utiliser la statistique :

$F^* = \frac{(SCR_c - SCR)/q}{SCR/(n - k - 1)}$ où SCR_c est la somme des carrés des résidus du modèle contraint.

Une autre manière de procéder consiste à comparer le ratio de vraisemblance¹ du modèle contraint et non contraint : si la contrainte est valide, nous devons avoir $L_c < L_{nc}$ où L_{nc} est la fonction de vraisemblance du modèle non contraint et L_c est la fonction de vraisemblance du modèle contraint. Soit $L_c/L_{nc} < 1$, sous forme logarithmique $\ln(L_c) - \ln(L_{nc}) < 0$ ou encore $l_c - l_{nc} < 0$, la différence entre les logarithmes des fonctions doit être significativement négative. On démontre que ce test se ramène à un test du χ^2 par calcul de la statistique $LR = -2(l_c - l_{nc})$ qui suit un χ^2 à r degrés de liberté (r étant le nombre de contraintes). Ainsi, si LR est supérieur au χ^2 lu dans la table au seuil α choisi et à r degrés de liberté, on rejette l'hypothèse H_0 , les restrictions ne sont pas vérifiées.

Enfin, nous pouvons citer le test du multiplicateur de Lagrange (« LM test ») fondé sur l'estimation d'une équation intermédiaire et de la valeur d'une statistique $LM = n \times R^2$ (n = nombre d'observations et R^2 = coefficient de détermi-

1. Dans un modèle de régression classique, la maximisation de la fonction de vraisemblance fournit des estimations identiques à celle de la méthode des MCO.

nation issu de l'estimation de l'équation intermédiaire) qui suit un χ^2 à r degrés de liberté (r étant le nombre de contraintes) ; nous verrons des applications de cette statistique au chapitre 5 concernant les tests de détection de l'autocorrélation des erreurs et de l'hétéroscédasticité.

V. L'utilisation de variables indicatrices¹

A. Constitution et finalités des variables indicatrices

Une variable indicatrice est une variable explicative particulière qui n'est composée que de 0 ou de 1. Cette variable est utilisée lorsque, dans un modèle, nous désirons intégrer un facteur explicatif binaire : « le phénomène a lieu ou n'a pas lieu » pour corriger, par exemple, d'une valeur anormale ; ou bien lorsque le facteur explicatif est qualitatif : « le genre d'un individu, homme ou femme ». Il s'agit donc d'incorporer une ou des variables explicatives supplémentaires au modèle spécifié initialement et d'appliquer les méthodes classiques d'estimation.

Le modèle de régression diffère selon l'apparition du phénomène par les valeurs d'un ou plusieurs coefficients alors que les autres paramètres sont identiques. En cas de modification structurelle d'un coefficient de régression, la variable muette affecte alors le coefficient de la ou des variables explicatives considérées.

Par exemple, soit le modèle à deux variables explicatives x_{1t} et x_{2t} :

$$y_t = a_0 + a_1x_{1t} + a_2x_{2t} + b_0D_t + b_1D_tx_{1t} + b_2D_tx_{2t} + \varepsilon_t$$

Si le phénomène existe, $D_t = 1$ et $D_t = 0$ sinon.

Si $D_t = 0$, le modèle s'écrit : $y_t = a_0 + a_1x_{1t} + a_2x_{2t} + \varepsilon_t$

Si $D_t = 1$, le modèle s'écrit : $y_t = (a_0 + b_0) + (a_1 + b_1)x_{1t} + (a_2 + b_2)x_{2t} + \varepsilon_t$

Si $b_1 = b_2 = 0$, le modèle ne diffère que par la valeur du terme constant.

Domaine d'utilisation des variables indicatrices.

Le domaine d'utilisation des variables indicatrices est très vaste, nous pouvons citer : la correction des valeurs anormales (cf. exercice n° 4), la modification structurelle (0 pour la période avant le changement structurel, 1 après le changement structurel), l'intégration de la saisonnalité (cf. exercice n° 6), la caractérisation d'un individu (genre, situation matrimoniale...), l'intégration de facteurs qualitatifs (appartenance d'un pays à la zone euro, promotion non quantifiable...), etc.

1. Les termes de variables indicatrices, de variables auxiliaires ou de variables muettes sont indifféremment employés en français. Le terme anglo-saxon *dummy* est le plus couramment utilisé.

B. Exemples d'utilisation

1) Correction de valeurs anormales

Les séries statistiques sont parfois affectées de valeurs anormales liées à la survenance d'événements exceptionnels : grève, guerre, aberration climatique, etc. Deux problèmes se posent alors : détecter la valeur anormale et la corriger afin qu'elle ne perturbe pas l'estimation statistique des autres variables.

Exercice n° 4

Détection et correction de valeurs anormales par variable indicatrice

Un modèle de production de service du secteur du tourisme est spécifié de la manière suivante :

$$QPS_t = a_0 + a_1 VA_t + a_2 POP_t + \varepsilon_t$$

avec :

QPS_t = production du secteur tourisme pour l'année t ;

VA_t = valeur ajoutée du secteur tourisme pour l'année t ;

POP_t = population pour l'année t .

L'économetre chargé de l'estimation de ce modèle sur 18 ans s'interroge sur la perturbation entraînée par l'effet d'une guerre pour l'année 16. Pour répondre à cette question, il intègre à son modèle de base une variable indicatrice D_t tel que :

$$D_t = 0 \quad \text{pour } t = 1 \text{ à } 15 \text{ et } t = 17 \text{ à } 18$$

$$D_t = 1 \quad \text{pour } t = 16$$

L'estimation du modèle économétrique est la suivante :

$$QPS_t = 2\,340,4 + 23,5 VA_t + 0,3 POP_t - 120,56 D_t + e_t$$

$$(4,5) \quad (2,2) \quad (2,9) \quad (5,8)$$

$$n = 18$$

$$R^2 = 0,65$$

$$(\cdot) = t \text{ de Student}$$

L'effet « guerre » a-t-il une influence significative sur la production du service du secteur du tourisme ?

Solution

La variable indicatrice D_t a un ratio de Student de $t^* = 5,8 > t_{14}^{0,05} = 2,14$, le coefficient de régression de cette variable est significativement différent de 0, la production de service pour l'année 16 est donc anormalement basse ($-120,56$). Cette baisse est, sans doute imputable à l'effet de la guerre.

• Généralisation

Dans le cas d'un phénomène se produisant de manière sporadique, la variable indicatrice prend la valeur 1 pour la ou les périodes que l'on désire corriger et 0 pour les autres.

Nous remarquons que nous pouvons procéder au test de Chow (stabilité du modèle sur l'ensemble de la période) en recourant à une variable indicatrice prenant la valeur 1 pour la première sous-période et la valeur 0 pour la deuxième sous-période. Le test de Student portant sur le coefficient de la variable indicatrice permet alors de se déterminer sur un modèle à un régime ou un modèle à deux régimes.

2) Variable qualitative

Il peut s'avérer important dans certaines spécifications de modèle de tenir compte de l'effet, sur la variable endogène, de variables qualitatives : être titulaire d'un diplôme, genre d'un individu, appartenance politique, etc. L'utilisation d'une variable indicatrice permet de segmenter les individus en deux groupes et de déterminer si le critère de segmentation est réellement discriminant.

Attention, dans le cas de variables qualitatives à plusieurs modalités, par exemple la couleur des yeux (bleu, vert, marron, autres), ou bien la situation familiale (célibataire, marié, divorcé, veuf, autres), etc. Il convient alors de coder autant de variables indicatrices que de modalités moins une. En reprenant l'exemple de la couleur des yeux nous définissons trois variables indicatrices : bleu (= 1 si l'individu a les yeux bleus, 0 sinon), vert (= 1 si l'individu a les yeux verts, 0 sinon), marron (= 1 si l'individu a les yeux marrons, 0 sinon), le cas des autres individus n'appartenant pas à l'une des trois premières catégories est implicitement contenu dans le terme constant. Une erreur à ne pas commettre consiste à créer une seule variable explicative codée, par exemple, de la manière suivante : bleu = 1, vert = 2, marron = 3, ...

Au chapitre 12 nous traitons du cas particulier des variables qualitatives qui figurent en tant que variables à expliquer.

Exercice n° 5

Intégration d'une variable qualitative

Afin de déterminer les facteurs explicatifs de la réussite de la licence en sciences économiques, on spécifie le modèle suivant :

$$NL = a_0 + a_1 ND + a_2 DS + \varepsilon$$

où :

NL = note moyenne obtenue en licence,

ND = note moyenne obtenue en fin de deuxième année,

DS = variable indicatrice de genre (1 pour les hommes et 0 pour les femmes).

L'estimation à partir d'un échantillon de 60 étudiants conduit aux résultats suivants :

$$NL = 8,5 + 0,3 ND - 1,2 DS + e$$

(4,5) (7,1) (2,3)

$$n = 60$$

$$R^2 = 0,72$$

$$(\cdot) = t \text{ de Student}$$

Le fait d'être homme ou femme a-t-il une influence sur la note obtenue en licence de sciences économiques ?

Solution

La variable indicatrice DS a un ratio de Student de

$$t^* = 2,3 > t_{57}^{0,05} = 1,96$$

le coefficient de régression a_2 est significativement différent de 0, le facteur sexe est donc bien un facteur discriminant de la note obtenue en licence. La probabilité critique, associée au risque de première espèce (risque de rejeter l'hypothèse H_0 à tort), est égale à 0,025. Nous avons donc 2,5 % de risque de nous tromper en rejetant l'hypothèse H_0 . Le risque d'erreur est très faible, nous la rejetons.

Il est à noter que le coefficient négatif indique qu'il est « pénalisant » d'être un homme ($DS = 1$) et qu'en moyenne, sur notre échantillon, les hommes ont une note inférieure de 1,2 point à celle des femmes.

Toutefois le facteur explicatif le plus important reste bien la note obtenue en fin de deuxième année ($t^* = 7,1$).

3) Analyse de saisonnalité

Exercice n° 6

↓ fichier C3EX6

Étude de saisonnalité par variables indicatrices

Une entreprise cherche à appréhender une relation entre ses ventes et ses dépenses publicitaires. Le directeur du marketing dispose des données (tableau 7) de ventes et de dépenses publicitaires sur 5 ans par trimestre.

1) Ce directeur du marketing commence par estimer la relation :

$$V_t = a_0 + a_1 \text{ Pub}_t + \varepsilon_t$$

Commenter les résultats obtenus.

2) Tracer le graphique de la série des ventes, que pouvez-vous en conclure ?

3) Spécifier et estimer le modèle adéquat.

Tableau 7 – Ventes et dépenses publicitaires

Années		T_1	T_2	T_3	T_4
1	Ventes	164	198	85	179
	Pub.	34	36	32	29
2	Ventes	168	201	98	197
	Pub.	45	67	76	75
3	Ventes	197	209	100	216
	Pub.	75	78	72	75
4	Ventes	223	245	119	260
	Pub.	78	81	84	83
5	Ventes	298	309	124	267
	Pub.	89	82	81	83

Solution

1) L'estimation du modèle de régression simple

$$V_t = a_0 + a_1 \text{ Pub}_t + \varepsilon_t$$

conduit aux résultats suivants :

$$V_t = 104,89 + 1,29 \text{ Pub}_t + e_t$$

(1,85)

$$n = 20$$

$$R^2 = 0,16$$

$$(\cdot) = t \text{ de Student}$$

La publicité a-t-elle un effet significatif sur les ventes ? La valeur du ratio du Student empirique permet de répondre à cette question.

$t^* = 1,85 < t_{18}^{0,05} = 2,10 \rightarrow$ le coefficient a_1 n'est pas significativement différent de 0, la publicité n'a pas, *a priori*, d'impact sur les ventes.

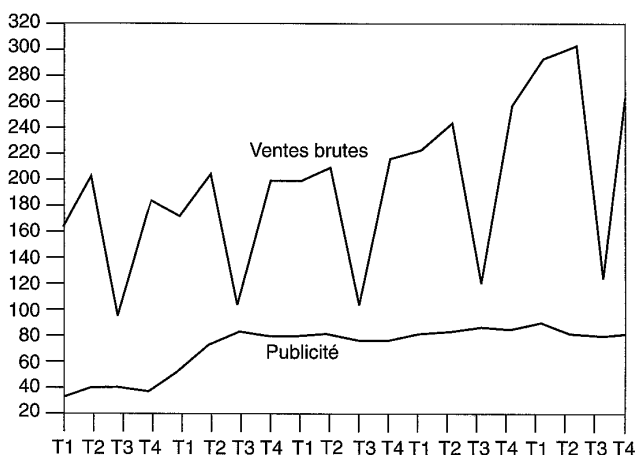
2) Le graphique 1 de la série des ventes et des dépenses publicitaires indique une série des ventes fortement saisonnière avec un creux très affirmé au troisième trimestre alors que la variable publicité ne semble pas affectée de variations saisonnières.

La relation entre les ventes et la publicité ne peut être déterminée puisque le mouvement saisonnier vient occulter l'estimation économétrique¹. Il convient donc d'intégrer ce mouvement saisonnier à l'aide de variables muettes.

3) Le modèle tenant compte de ce mouvement saisonnier s'écrit :

$$V_t = a_0 + a_1 \text{ Pub}_t + a_2 D_{1t} + a_3 D_{2t} + a_4 D_{3t} + \varepsilon_t$$

avec D_{1t} = variable indicatrice du trimestre 1 : cette variable est composée de 0, sauf pour les premiers trimestres pour lesquels la valeur 1 figure ; D_{2t} = variable indicatrice du trimestre 2 : cette variable est composée de 0, sauf pour les deuxièmes trimestres pour lesquels la valeur 1 figure ; D_{3t} = variable indicatrice du trimestre 3 : cette variable est composée de 0, sauf pour les troisièmes trimestres pour lesquels la valeur 1 figure.



Graphique 1 – Série des ventes et de dépenses publicitaires

La variable indicatrice du quatrième trimestre ne doit pas être incorporée au modèle. En effet, nous aurions dans le cas de son existence

$$D_{1t} + D_{2t} + D_{3t} + D_{4t} = U$$

U étant le vecteur unité qui serait par construction colinéaire au vecteur unité du terme constant, ce qui rendrait la matrice $X'X$ singulière (non inversible).

La valeur du coefficient saisonnier du quatrième trimestre est donc implicitement incorporée à la valeur du terme constant.

Nous pourrions, dans une solution alternative, utiliser les quatre variables indicatrices et procéder à une régression sans terme constant. Les données utilisées pour l'estimation sont présentées dans le tableau 8.

L'estimation du modèle est la suivante :

$$V_t = 129,10 + 1,37 \text{ Pub}_t - 7,21 D_{1t} + 8,87 D_{2t} - 118,6 D_{3t} + e_t$$

(3,97) (0,38) (0,47) (6,25)

1. Dans un modèle économétrique, les variables doivent être non saisonnières ou Corrigées des Variations Saisonnières (CVS).

$$n = 20$$

$$R^2 = 0,83$$

$$(\cdot) = t \text{ de Student}$$

Les dépenses publicitaires, dont le ratio de Student est égal 3,97, sont maintenant explicatives¹ des ventes ; la variable indicatrice D_{3t} est la seule variable muette significative ($t^* = 6,25$), ce qui implique que la saisonnalité des ventes est liée essentiellement au creux du troisième trimestre.

*Tableau 8 – Variables indicatrices
pour une désaisonnalisation trimestrielle*

	V_t	Pub_t	D_{1t}	D_{2t}	D_{3t}
T_1	164	34	1	0	0
T_2	198	36	0	1	0
T_3	85	32	0	0	1
T_4	179	29	0	0	0
T_1	168	45	1	0	0
T_2	201	67	0	1	0
T_3	98	76	0	0	1
T_4	197	75	0	0	0
T_1	197	75	1	0	0
T_2	209	78	0	1	0
T_3	100	72	0	0	1
T_4	216	75	0	0	0
T_1	223	78	1	0	0
T_2	245	81	0	1	0
T_3	119	84	0	0	1
T_4	260	83	0	0	0
T_1	298	89	1	0	0
T_2	309	82	0	1	0
T_3	124	81	0	0	1
T_4	267	83	0	0	0

VI. La prévision à l'aide du modèle linéaire général et la régression récursive

A. Prédiction² conditionnelle

Le problème consiste à déterminer quelle valeur doit être attribuée à la variable endogène lorsque nous connaissons les valeurs des variables exogènes.

1. Ceci est à noter : dans un modèle, l'absence d'une ou de plusieurs variables explicatives importantes peut entraîner une mauvaise estimation des variables figurant déjà dans le modèle.
2. Il ne s'agit pas toujours de prévoir une valeur dans le futur, mais dans le cadre de séries temporelles de simuler le passé ou bien dans les modèles en coupe instantanée de prédire des valeurs.

Le modèle général estimé est le suivant :

$$y_t = \widehat{a}_0 + \widehat{a}_1 x_{1t} + \widehat{a}_2 x_{2t} + \dots + \widehat{a}_k x_{kt} + e_t$$

La prévision pour la donnée $t + h$ (respectivement $i + h$ pour les modèles en coupe instantanée) est la suivante :

$$\widehat{y}_{t+h} = \widehat{a}_0 + \widehat{a}_1 x_{1t+h} + \widehat{a}_2 x_{2t+h} + \dots + \widehat{a}_k x_{kt+h}$$

L'erreur de prévision est donnée par :

$$e_{t+h} = y_{t+h} - \widehat{y}_{t+h}$$

Considérant que les hypothèses du modèle linéaire général sont vérifiées, la prévision \widehat{y}_{t+h} est sans biais.

Nous avons postulé que nous *connaissons sans erreur* les valeurs en $t + h$ des variables explicatives. Si pour un modèle en coupe instantanée cette hypothèse est réaliste, il n'en est pas de même dans les modèles en séries temporelles pour lesquelles les valeurs futures des variables explicatives sont estimées avec imprécision et donc introduisent un élément supplémentaire d'incertitude dans le calcul de la prévision. Ce problème est exclu du champ de notre analyse, puisque nous traitons de prédiction conditionnelle.

B. Fiabilité de la prévision et intervalle de prévision

L'erreur de prévision calculée en t à l'horizon h peut s'écrire aussi :

$$e_{t+h} = y_{t+h} - \widehat{y}_{t+h} = X'_{t+h}(a - \widehat{a}) + \varepsilon_{t+h}$$

Calculons la variance de cette expression : $V(e_{t+h}) = V(X'_{t+h}(a - \widehat{a}) + \varepsilon_{t+h})$

$X'_{t+h}(a - \widehat{a})$ est une combinaison linéaire des y_i et puisque ε_{t+h} est sans corrélation avec les y_i , nous avons $\text{Cov}(X'_{t+h}(a - \widehat{a}), \varepsilon_{t+h}) = 0$.

On a alors : $V(e_{t+h}) = V(X'_{t+h}(a - \widehat{a})) + V(\varepsilon_{t+h})$

$$V(X'_{t+h}(a - \widehat{a})) = X'_{t+h} V(a - \widehat{a}) X_{t+h} = \sigma_\varepsilon^2 [X'_{t+h} (X'X)^{-1} X_{t+h}]$$

(puisque $V(\widehat{a}) = \sigma_\varepsilon^2 (X'X)^{-1}$)

On connaît : $V(\varepsilon_{t+h}) = \sigma_\varepsilon^2$

La variance de l'erreur de prévision est donc égale à :

$$\sigma_{et+h}^2 = \sigma_\varepsilon^2 [X'_{t+h} (X'X)^{-1} X_{t+h} + 1] \quad [18]$$

Avec $X_{t+h} = \begin{bmatrix} 1 \\ x_{1t+h} \\ x_{2t+h} \\ \dots \\ x_{kt+h} \end{bmatrix}$ vecteur des valeurs prévues des variables explicatives.

Or l'erreur de prévision ($e_{t+h} = y_{t+h} - \hat{y}_{t+h}$) est distribuée suivant une loi normale $N(0, \sigma_{e_{t+h}}^2)$, en remplaçant la variance théorique σ_ε^2 par la variance empirique $\hat{\sigma}_\varepsilon^2$ dans l'expression [18], nous pouvons en déduire que :

$\frac{\hat{y}_{t+h} - y_{t+h}}{\hat{\sigma}_\varepsilon [X'_{t+h} (X' X)^{-1} X_{t+h} + 1]^{1/2}}$ suit une loi de Student à $n - k - 1$ degrés de liberté.

Nous remarquons que, comme pour le modèle de régression simple, la variance de l'erreur de prévision est d'autant plus faible lorsque :

- la variance résiduelle est faible ;
- les valeurs prévues des variables explicatives se rapprochent de leur moyennes.

L'intervalle au seuil $(1 - \alpha)$ de la prévision est alors :

$$y_{t+h} = \hat{y}_{t+h} \pm t_{n-k-1}^{\alpha/2} \sqrt{\hat{\sigma}_\varepsilon^2 [X'_{t+h} (X' X)^{-1} X_{t+h} + 1]} \quad [19]$$

Exercice n° 7

↓ fichier C3EX1

Prévision à partir du modèle linéaire

En reprenant les données de l'exercice 1 (tableau 1), on demande :

- 1) d'estimer le modèle à deux variables explicatives :

$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \varepsilon_t$ (puisque nous avons montré que la variable x_3 n'est pas significative) ;

- 2) de calculer une prévision et son intervalle à 95 % pour les périodes 15 et 16, sachant que :

$$x_{1\,15} = 3; x_{1\,16} = 6 \text{ et } x_{2\,15} = 24; x_{2\,16} = 38$$

Solution

- 1) L'estimation du modèle à deux variables explicatives conduit aux résultats suivants :

$$y_t = 25,84 + 0,715x_{1,t} - 0,328x_{2,t} + e_t$$

(0,26) (0,13)

$$n = 14$$

$$R^2 = 0,687$$

$$(.) = \text{Ecart type}$$

$$\hat{\sigma}_\varepsilon = 2,538$$

Nous remarquons les t de Student supérieurs à 2,201, les coefficients a_1 et a_2 sont significativement différents de 0.

- 2) La prévision pour la période 15 est calculée à partir du modèle estimé :

$$\hat{y}_{15} = 25,84 + 0,71 x_{15} - 0,33 x_{215} = 25,84 + 0,71 \times 3 - 0,33 \times 24$$

$$\hat{y}_{15} = 20,25$$

De même, pour la période 16, on obtient :

$$\hat{y}_{16} = 25,84 + 0,71 x_{16} - 0,33 x_{216} = 25,84 + 0,71 \times 6 - 0,33 \times 38$$

$$\hat{y}_{16} = 17,26$$

Les écarts types de l'erreur de prévision sont donnés par [18] :

$\hat{\sigma}_{e15}^2 = \hat{\sigma}_e^2 [X'_{15} (X' X)^{-1} X_{15} + 1]$. Nous devons calculer $(X' X)^{-1}$, les autres éléments étant connus.

$$\begin{aligned} X' X &= \begin{bmatrix} 14 & 85 & 532 \\ 85 & 631 & 3126 \\ 532 & 3126 & 20666 \end{bmatrix} \rightarrow (X' X)^{-1} = \\ &= \begin{bmatrix} 5,707687 & -0,16341 & -0,12221 \\ -0,16341 & 0,011001 & 0,002542 \\ -0,12221 & 0,002542 & 0,002809 \end{bmatrix} \\ \hat{\sigma}_{e15}^2 &= (2,538)^2 \left[(1 \ 3 \ 24) \begin{bmatrix} 5,707687 & -0,16341 & -0,1222 \\ -0,16341 & 0,011001 & 0,002542 \\ -0,12221 & 0,002542 & 0,002809 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 24 \end{bmatrix} + 1 \right] \\ \hat{\sigma}_{e15}^2 &= 6,44 \cdot [0,94 + 1] = 12,53 \rightarrow \hat{\sigma}_{e15} = 3,54 \end{aligned}$$

De même, nous pouvons déterminer :

$$\hat{\sigma}_{e16}^2 = 6,44 \cdot [0,071 + 1] = 6,90 \rightarrow \hat{\sigma}_{e16} = 2,62$$

Les intervalles de prévision peuvent être calculés par [19] :

$$\begin{aligned} y_{t+h} &= \hat{y}_{t+h} \pm t_{n-k-1}^{\alpha/2} \sqrt{\hat{\sigma}_e^2 [X'_{t+h} (X' X)^{-1} X_{t+h} + 1]} \\ y_{15} &= \hat{y}_{15} \pm t_{14-2-1}^{0,025} \cdot \hat{\sigma}_{e15} = 20,05 \pm 2,201 \times 3,54 \end{aligned}$$

$IC_{15}^{0,05} = [12,26; 27,84]$, la prévision pour la période 15 a 95 % de chances de se situer dans cet intervalle et la prévision de la période 16 :

$$IC_{16}^{0,05} = [11,49; 23,03]$$

Les intervalles de la prévision semblent assez larges, cependant il convient de souligner que la distribution de probabilité suit une loi de Student et qu'à ce titre la valeur la plus probable demeure la valeur centrale (la prévision estimée) et que la probabilité d'apparition diminue lorsque l'on s'éloigne de cette valeur centrale.

Nous remarquons que l'écart type de l'erreur de prévision de la période 16 est inférieur à celui de la période 15. Cela est la conséquence des valeurs des variables explicatives pour la période 16 que nous avons choisies sensiblement égales à leur moyenne.

C. Les tests de stabilité par la régression récursive

On appelle régression récursive la procédure qui consiste à estimer une succession de régressions en augmentant progressivement le nombre d'observations.

On estime d'abord le modèle avec les $k + 2$ premières observations (modèle à un degré de liberté) puis on incorpore l'observation suivante et on re-estime le modèle, et ainsi de suite jusqu'à n (le nombre d'observations total disponible).

Un simple examen graphique de l'évolution des coefficients du modèle – assortis de leurs intervalles de confiance à \pm deux écarts types – permet éventuellement de déterminer des changements structurels.

– Les tests du « *CUSUM* » sont fondés sur la dynamique de l'erreur de prévision.

Ces tests permettent de détecter les instabilités structurelles des équations de régression au cours du temps. Contrairement au test de Chow, ils ne nécessitent pas de connaître *a priori* la date de modification structurelle. L'idée générale de ces tests est d'étudier l'évolution au cours du temps de l'erreur de prévision normalisée, on appelle résidu récursif cette succession d'erreur de prévision calculée en $t - 1$ pour t .

Le résidu récursif w_t au temps t est donc défini comme le rapport :

- de l'écart entre la prévision calculée en $t - 1$ pour t et la réalisation en t ,
- et une variable d'échelle¹ calculée pour cette même période t ,

soit :

$$w_t = \frac{e_t = y_t - \hat{y}_t}{S_e} = \frac{y_t - x'_t \hat{a}_{t-1}}{\sqrt{(1 + x'_t (X'_{t-1} X_{t-1})^{-1} x_t)}}$$

avec $t = K + 2, K + 3, K + 4, \dots, n$ ($K = k + 1$ est le nombre total de paramètres estimés du modèle). Ce résidu récursif suit donc une loi normale $N(0, \sigma^2)$.

Soit les deux tests :

– le *CUSUM* (CUMulative SUM) fondé sur la somme cumulée des résidus récursifs,

– le *CUSUM SQ* fondé sur la somme cumulée du carré des résidus récursifs.

À partir du résidu récursif w_t , on calcule la statistique *CUSUM* :

$$W_t = \frac{n - K}{SCR} \sum_{j=K+2}^t w_j$$

avec $t = K + 2, \dots, n$ ($K = k + 1$ est le nombre total de paramètres estimés du modèle) et *SCR* la somme des carrés des résidus calculée à partir du modèle à n observations.

Si les coefficients sont stables au cours du temps, alors les résidus récursifs doivent rester dans l'intervalle défini par les deux droites : $[K, \pm \alpha \sqrt{n - K}]$ et $[n, \pm 3\alpha \sqrt{n - K}]$ avec $\alpha = 1,143 - 0,948$ et $0,850$ respectivement pour des seuils de confiance de 1 %, 5 % et 10 %. Dans le cas contraire, le modèle est réputé instable.

1. On remarque qu'il s'agit de la formule [18], hormis la variance résiduelle.

- La statistique du *CUSUM SQ* est donnée par le carré du résidu récursif :

$$S_t = \frac{\sum_{j=K+2}^t w_j^2}{\sum_{j=K+2}^n w_j^2} \quad \text{avec } t = K + 2, \dots, n \text{ et } 0 \leq S_t \leq 1 ;$$

ce test permet de détecter des modifications aléatoires (ponctuelles) dans le comportement du modèle. Si les coefficients sont stables au cours du temps, alors les résidus récursifs carrés doivent rester dans l'intervalle défini par :

$$\left[\pm C \frac{(t - K)}{n - K} \right]$$

où C est la statistique de Kolmogorov-Smirnov¹.

D. Le test de spécification de Ramsey

Le test de Ramsey, aussi appelé le test de RESET (*Regression Error Specification Test*), porte sur la pertinence de la forme fonctionnelle du modèle, telle que :

- une relation fonctionnelle non adaptée (passage aux logarithmes, fonctions inverses...) entre la variable à expliquer et les variables explicatives ;
- l'absence d'une variable explicative dans le modèle ;
- la corrélation entre la variable explicative et le terme d'erreur ;
- ...

Plutôt que d'estimer des spécifications alternatives (par exemple linéaire ou non linéaire), le test porte sur la significativité d'un ou des coefficients d'une équation intermédiaire dans laquelle figure la série à expliquer ajustée et élevée à la puissance 2, 3, 4... Le test RESET est mené en trois étapes :

- 1) Estimation par les MCO du modèle initial et calcul de la série ajustée :

$$\hat{y}_t = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2 + \dots + \hat{a}_k x_k$$

- 2) Estimation par les MCO de l'équation intermédiaire :

$$y_t = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k + \phi_2 \hat{y}_t^2 + \phi_3 \hat{y}_t^3 + \dots + \phi_h \hat{y}_t^h + v_t$$

1. Pour la table, cf. Johnston J. et Dinardo J., pages 379-380, 1999.

3) Test de l'hypothèse $H_0 : \phi_2 = \phi_3 = \dots = \phi_h$ par un classique test de Fisher d'un sous-ensemble de coefficients (ou si $h = 2$, par un test de Student portant sur le coefficient unique, $H_0 : \phi_2 = 0$).

Si l'hypothèse H_0 est acceptée, le modèle est bien linéaire et il n'existe pas de problème de spécification.

Exercice n° 8

↓ fichier C3EX1

Tests de stabilité structurelle, Test de spécification de Ramsey

Soit les données de l'exercice 1 (tableau 1) et le modèle :

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \varepsilon_t$$

On demande :

- 1) À partir d'une régression récursive de tester un éventuel changement structurel en examinant la stabilité des coefficients et en procédant aux tests CUSUM ;
- 2) De procéder au test de spécifications de Ramsey.

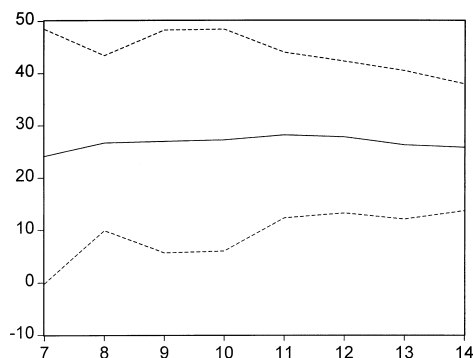
Solution

1) Stabilité des coefficients

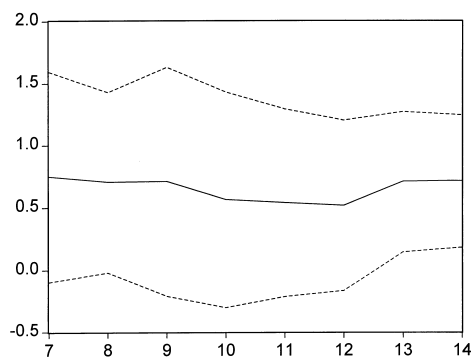
Nous procédons à l'estimation des coefficients a_0 , a_1 et a_2 avec leur intervalle de confiance à \pm deux écarts types, sur 4, 5, ..., 14 observations, soit 11 coefficients différents.

Le programme Eviews (C3EX8) est en téléchargement.

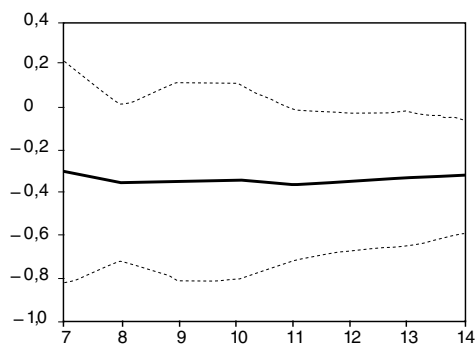
Les graphiques 2, 3 et 4 illustrent l'évolution des trois coefficients à partir de la septième observation ; en effet compte tenu du degré de liberté très faible, l'estimation de l'écart type des coefficients (et donc l'intervalle de confiance) pour les trois premiers calculs n'est pas significatif.



Graphique 2 – Évolution de a_0 et de son intervalle de confiance à \pm deux écarts types



Graphique 3 – Évolution de a_1 et de son intervalle de confiance à \pm deux écarts types

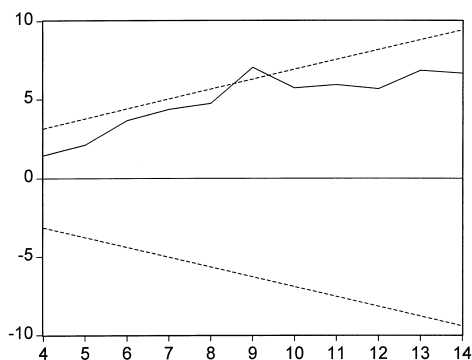


Graphique 4 – Évolution de a_2 et de son intervalle de confiance à \pm deux écarts types

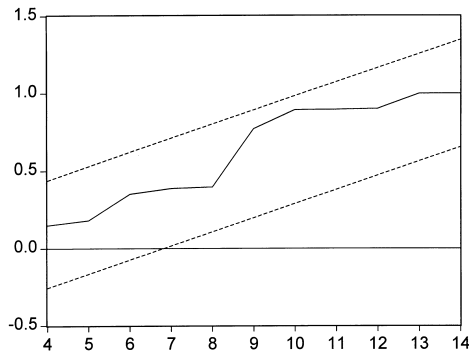
Nous constatons que les trois coefficients restent toujours dans leur intervalle de confiance, nous rejetons l'hypothèse d'un changement structurel.

Tests *CUSUM*

Les tests *CUSUM* et *CUSUM SQ* peuvent être effectués directement sur Eviews :



Graphique 5 – Test *CUSUM* pour un seuil de confiance de $\pm 5\%$



Graphique 6 – Test CUSUM SQ pour un seuil de confiance de $\pm 5\%$

Hormis à la période 9 du test *CUSUM* pour laquelle nous constatons un – très léger – franchissement de l’intervalle de confiance, les statistiques *CUSUM* et *CUSUM SQ* restent dans leur intervalle ; là encore nous rejetons donc l’hypothèse d’un changement structurel. Ces résultats sont en conformité avec ceux de l’exercice n°3 (test de Chow).

2) Le test de Ramsey est conduit de la manière suivante :

- Calculer la série à expliquer ajustée : $\hat{y}_t = 25,84 + 0,71x_{1t} - 0,33x_{2t}$
- Élever au carré la série à expliquer : $\hat{y}_t^2 = (\hat{y}_t)^2$, nous ne testons qu’une spécification non linéaire de type quadratique.
- Estimer le modèle en ajoutant comme variable explicative supplémentaire \hat{y}_t^2 :

$$y_t = b_0 + b_1x_{1t} + b_2x_{2t} + \phi_2\hat{y}_t^2 + \varepsilon_t$$

– Effectuer le test d’hypothèses $H_0 : \phi_2 = 0$ ($H_1 : \phi_2 \neq 0$). La probabilité critique du coefficient ϕ_2 est égale à 0,827. Nous acceptons l’hypothèse H_0 , le modèle est donc correctement spécifié.

Le test de Ramsey peut être effectué directement sur Eviews :

Ramsey RESET Test :				
F-statistic	0.049866	Probability	0.827792	
Test Equation :				
Dependent Variable : Y				
Included observations : 14				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	16.40322	42.74223	0.383771	0.7092
X1	0.326961	1.759415	0.185835	0.8563
X2	− 0.150146	0.809298	− 0.185526	0.8565
FITTED^2 = \hat{y}_t^2	0.015484	0.069342	0.223308	0.8278

Remarque : La *F*-statistique est égale à $F^* = 0,0498 = (t^*)^2 = (0,223)^2$.

VII. Exercices récapitulatifs

Exercice n° 9

Estimation d'un modèle et tests de validité à partir de la connaissance de quelques éléments statistiques

Sur $n = 100$ observations et pour trois séries (y , x_1 et x_2), nous avons les résultats numériques suivants :

$$V(y) = 1\,000; r_{y,x_1}^2 = 0,75; r_{x_1,x_2}^2 = 0,45; r_{y,x_2}^2 = 0,85; \bar{y} = 12$$

1) Nous avons effectué la régression : $\hat{y} = 10 x_1 - 6$.

Le coefficient de x_1 est-il significativement différent de 0 ?

2) La régression de y sur x_2 a donné $\hat{y} = 4 x_2 + 8$, le coefficient de x_2 est-il significativement différent de 0 ?

3) Calculer les coefficients du modèle : $y = a_0 + a_1 x_1 + a_2 x_2 + \varepsilon$, et le coefficient de corrélation multiple.

4) Les coefficients a_1 et a_2 sont-ils significativement différents de 0 ? La régression est-elle globalement significative ?

Solution

1) La variance estimée du coefficient de régression est donnée par :

$$\hat{\sigma}_a^2 = \frac{\hat{\sigma}_\varepsilon^2}{\sum_t (x_{1t} - \bar{x}_1)^2}$$

Nous savons que dans le cadre de la régression simple, il y a égalité entre corrélation simple et corrélation multiple, soit :

$$\begin{aligned} r_{y,x_1}^2 &= \frac{\text{Cov}(y, x_1)^2}{V(y) \cdot V(x_1)} = \frac{\left[\sum_t (y_t - \bar{y})(x_{1t} - \bar{x}_1) \right]^2}{\sum_t (y_t - \bar{y})^2 \sum_t (x_{1t} - \bar{x}_1)^2} \\ &= R^2 = 1 - \frac{\sum_t e_t^2}{\sum_t (y_t - \bar{y})^2} = 0,75 \end{aligned}$$

Or :

$$V(y) = 1\,000 = \frac{\sum_t (y_t - \bar{y})^2}{n} = \frac{\sum_t (y_t - \bar{y})^2}{100} \rightarrow \sum_t (y_t - \bar{y})^2 = 100\,000$$

Nous utilisons ici la formule de la variance d'une population (division par n) et non la formule de la variance d'un échantillon (division par $n - 1$).

D'où :

$$\sum_t e_t^2 = (1 - 0,75) \cdot 100\,000 = 25\,000 \rightarrow \hat{\sigma}_\varepsilon^2 = 25\,000 / (n - 2) = 255,1$$

Pour calculer la variance de x_1 , nous utilisons les relations :

$$\hat{a} = \frac{\sum_t (y_t - \bar{y})(x_{1t} - \bar{x}_1)}{\sum_t (x_{1t} - \bar{x}_1)^2} = 10$$

et

$$r_{y,x_1}^2 = \frac{\left[\sum_t (y_t - \bar{y})(x_{1t} - \bar{x}_1) \right]^2}{\sum_t (y_t - \bar{y})^2 \sum_t (x_{1t} - \bar{x}_1)^2} = 0,75$$

Connaissant $\sum_t (y_t - \bar{y})^2 = 100\,000$, nous trouvons :

$$\sum_t (x_{1t} - \bar{x}_1)^2 = 750 \text{ et } \sum_t (y_t - \bar{y})(x_{1t} - \bar{x}_1) = 7\,500$$

Soit : $\text{Var}(x_1) = 7,5$ et $\text{Cov}(y, x_1) = 75$

$$\hat{\sigma}_a^2 = \frac{\hat{\sigma}_\varepsilon^2}{\sum_t (x_{1t} - \bar{x}_1)^2} = \frac{255,1}{750} = 0,34 \rightarrow t^* = \frac{10}{0,58} = 17,14 > t^{0,05} = 1,96$$

Le coefficient de régression de y sur x_1 est significativement différent de 0.

2) Un calcul analogue pour l'écart type du coefficient de régression de y sur x_2 conduit aux résultats suivants :

$$\sum_t e_t^2 = (1 - 0,85) \cdot 100\,000 = 15\,000 \rightarrow \hat{\sigma}_\varepsilon^2 = 15\,000/98 = 153,06$$

$$\sum_t (x_{2t} - \bar{x}_2)^2 = 5\,312,5 \text{ et } \sum_t (y_t - \bar{y})(x_{2t} - \bar{x}_2) = 21\,250$$

Soit : $\text{Var}(x_2) = 53,12$ et $\text{Cov}(y, x_2) = 212,50$

$$\hat{\sigma}_a^2 = \frac{\hat{\sigma}_\varepsilon^2}{\sum_t (x_{2t} - \bar{x}_2)^2} = \frac{153,06}{5\,312,5} = 0,03 \rightarrow t^* = \frac{4}{0,17} = 23,5 > t^{0,05} = 1,96$$

Le coefficient de régression de y sur x_2 est significativement différent de 0.

3) Calculer les coefficients du modèle :

$$y = a_0 + a_1 x_1 + a_2 x_2 + \varepsilon$$

Nous avons la relation d'après les équations normales (cf. section II, p. 50) :

$$\begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} = \begin{bmatrix} \text{Var}(x_1) & \text{Cov}(x_1, x_2) \\ \text{Cov}(x_1, x_2) & \text{Var}(x_2) \end{bmatrix}^{-1} \begin{bmatrix} \text{Cov}(y, x_1) \\ \text{Cov}(y, x_2) \end{bmatrix}$$

Nous connaissons tous les éléments de ce calcul sauf $\text{Cov}(x_1, x_2)$, or :

$$r_{x_1, x_2}^2 = \frac{[\text{Cov}(x_1, x_2)]^2}{V(x_1) \cdot V(x_2)} = \frac{\left[\sum_t (x_{1t} - \bar{x}_1)(x_{2t} - \bar{x}_2) \right]^2}{\sum_t (x_{1t} - \bar{x}_1)^2 \sum_t (x_{2t} - \bar{x}_2)^2} = 0,45$$

$\text{Var}(x_1) = 7,5$ et $\text{Var}(x_2) = 53,12 \rightarrow \text{Cov}(x_1, x_2) = 13,389$, en effet la covariance entre x_1 et x_2 est positive puisque les covariances entre y et x_1 , et y et x_2 sont positives.

$$\begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} = \begin{bmatrix} 7,5 & 13,39 \\ 13,39 & 53,12 \end{bmatrix}^{-1} \begin{bmatrix} 75 \\ 212,5 \end{bmatrix} = \begin{bmatrix} 0,242 & -0,061 \\ -0,061 & 0,034 \end{bmatrix} \begin{bmatrix} 75 \\ 212,5 \end{bmatrix} = \begin{bmatrix} 5,19 \\ 2,69 \end{bmatrix}$$

Le terme constant est donné par $\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}_1 - \hat{a}_2 \bar{x}_2$.

Or $\bar{y} = 12$; $\hat{a}_1 = 5,19$; $\hat{a}_2 = 2,69$, de plus $\bar{x}_1 = 1,8$ car $\bar{y} = 10 \bar{x}_1 - 6$ (modèle 1) et $\bar{x}_2 = 1$ car $\bar{y} = 4 \bar{x}_2 + 8$ (modèle 2) $\rightarrow \hat{a}_0 = -0,032$.

Le coefficient de détermination est égal, si l'on raisonne sur des données centrées, à :

$$R^2 = \frac{\hat{Y}' \hat{Y}}{Y' Y} = \frac{\hat{a}' X' Y}{Y' Y} = \frac{(5,19 \ 2,69) \begin{bmatrix} 7 \ 500 \\ 21 \ 250 \end{bmatrix}}{\sum_t (y_t - \bar{y})^2} = \frac{96 \ 087,5}{100 \ 000} = 0,96$$

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

Le coefficient de corrélation multiple est égal à $R = 0,98$

4) Analyse des résultats et tests statistiques.

• Significativité des coefficients

Pour calculer les ratios de Student, il convient d'abord de connaître les écarts types de chacun des coefficients.

$$\hat{\sigma}_{\hat{a}_i}^2 = \hat{\sigma}_\varepsilon^2 (X' X)^{-1} = \hat{\sigma}_\varepsilon^2 \times \frac{1}{n} \times \begin{bmatrix} \text{Var}(x_1) & \text{Cov}(x_1, x_2) \\ \text{Cov}(x_1, x_2) & \text{Var}(x_2) \end{bmatrix}^{-1}$$

Or :

$$\hat{\sigma}_\varepsilon^2 = \frac{SCR}{n - k - 1} = \frac{(1 - R^2) SCT}{n - k - 1} = \frac{(1 - R^2) \sum_t (y_t - \bar{y})^2}{n - k - 1} = \frac{(1 - 0,96) 100 \ 000}{100 - 2 - 1}$$

$$\hat{\sigma}_\varepsilon^2 = 41,23$$

$$\hat{\sigma}_{\hat{a}_1}^2 = 41,23 \times \frac{0,242}{100} = 0,097 \rightarrow \hat{\sigma}_{\hat{a}_1} = 0,316 \rightarrow t^* = \frac{5,19}{0,316} = 16,4 > t^{0,05} = 1,96$$

$$\hat{\sigma}_{\hat{a}_2}^2 = 41,23 \times \frac{0,034}{100} = 0,014 \rightarrow \hat{\sigma}_{\hat{a}_2} = 0,118 \rightarrow t^* = \frac{2,69}{0,118} = 22,8 > t^{0,05} = 1,96$$

Les coefficients a_1 et a_2 sont significativement différents de 0. Nous pouvons observer des différences par rapport aux valeurs estimées précédemment, sur les coefficients et sur leurs écarts types. Cela est la conséquence de la colinéarité¹ entre x_1 et x_2 ($r_{x_1, x_2} = 0,67$).

- Analyse de la régression dans son ensemble

Nous calculons la statistique de Fisher :

$$F^* = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{0,96/2}{(1-0,96)/97} = 1\,164 > F_{2,97}^{0,05} = 3,10$$

La régression est globalement significative.

Exercice n° 10

Test d'une combinaison linéaire de coefficients de régression (test de contrainte linéaire)

Pendant 23 ans nous avons relevé, sur une parcelle de terre située en Auvergne, les rendements de la culture de blé (y), la température moyenne (x_1) et le niveau des précipitations (x_2).

L'ajustement de ce modèle conduit aux résultats suivants :

$$y = 0,510 x_1 - 0,350 x_2 + 27,3 + e$$

$$n = 23$$

$$R^2 = 0,937$$

$$(X'X)^{-1} = \begin{bmatrix} 0,0009 & -0,08 & -0,3 \\ -0,08 & 0,0025 & 0,02 \\ -0,3 & 0,02 & 0,2 \end{bmatrix}$$

et
$$SCT = \sum_t (y_t - \bar{y})^2 = 317,46$$

On demande de tester les deux hypothèses suivantes :

- 1) Existe-t-il une influence d'au moins un des facteurs ?
- 2) Le coefficient de la température est-il significativement deux fois plus élevé que celui des précipitations ?

Solution

- 1) Nous procédons au test de Fisher :

$$H_0 : a_1 = a_2 = 0$$

1. Les différents aspects de la colinéarité sont étudiés au chapitre 4.

H1 : il existe au moins un des coefficients non nul

$$F^* = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{0,937/2}{(1-0,937)/20} = 148,731 > F_{2,20}^{0,05} = 3,49$$

Nous rejetons l'hypothèse $a_1 = a_2 = 0$, il existe au moins un des coefficients non nul.

2) Le test d'hypothèses est le suivant¹ :

$$H0 : a_1 - 2 a_2 = 0$$

$$H1 : a_1 - 2 a_2 \neq 0$$

Sous H0, le rapport $\frac{(\hat{a}_1 - 2 \hat{a}_2) - 0}{\hat{\sigma}_{\hat{a}_1 - 2 \hat{a}_2}}$ suit une loi de Student à $n - k - 1$ degrés de liberté.

Il convient tout d'abord de calculer la variance de $(\hat{a}_1 - 2 \hat{a}_2)$:

On rappelle que $\text{Var}(a x + b y) = a^2 \text{Var}(x) + b^2 \text{Var}(y) + 2 a b \text{Cov}(x, y)$ (a et b sont deux nombres réels ; x et y sont deux variables aléatoires).

$$\text{Var}(\hat{a}_1 - 2 \hat{a}_2) = \text{Var}(\hat{a}_1) + \text{Var}(-2 \hat{a}_2) + 2 \text{Cov}(\hat{a}_1, -2 \hat{a}_2)$$

$$\text{Var}(\hat{a}_1 - 2 \hat{a}_2) = \text{Var}(\hat{a}_1) + (-2)^2 \text{Var}(\hat{a}_2) + 2 \times -2 \text{Cov}(\hat{a}_1, \hat{a}_2)$$

De plus, la variance résiduelle est donnée par :

$$\sum_t e_t^2 = (1 - R^2) \cdot SCT = (1 - 0,937) \cdot 317,46 \rightarrow \hat{\sigma}_\varepsilon^2 = 20/(n - 3) = 1$$

$$\text{D'où} \quad \hat{\sigma}_{\hat{a}_1 - 2 \hat{a}_2}^2 = \hat{\sigma}_{\hat{a}_1}^2 + 4 \hat{\sigma}_{\hat{a}_2}^2 - 4 \text{Cov}(\hat{a}_1, \hat{a}_2)$$

$$\hat{\sigma}_{\hat{a}_1 - 2 \hat{a}_2}^2 = 0,0009 + 4 \times 0,0025 - 4(-0,08) = 0,33$$

$$\frac{(\hat{a}_1 - 2 \hat{a}_2) - 0}{\hat{\sigma}_{\hat{a}_1 - 2 \hat{a}_2}} = \frac{(0,51 + 2 \times 0,35)}{0,57} = 2,10 > t_{20}^{0,05} = 2,08$$

Nous rejetons (de justesse !), au seuil de 5 %, l'hypothèse H0. Cependant, pour un seuil de 4 %, l'hypothèse H0 est acceptée.

Remarque : Nous pouvons aussi procéder à ce test en calculant la statistique $\left(\frac{\hat{a}_1 - 2 \hat{a}_2}{\hat{\sigma}_{\hat{a}_1 - 2 \hat{a}_2}} \right)^2$ qui suit un χ^2 à 1 degré de liberté ; c'est ainsi que ce test de restriction linéaire sur les coefficients du modèle peut être généralisé à des restrictions non linéaires (test de Wald). Par analogie, prenons par exemple une contrainte du type $a_1 \times a_2 = 1$, la statistique $\left(\frac{\hat{a}_1 \times \hat{a}_2}{\hat{\sigma}_{\hat{a}_1 \times \hat{a}_2}} \right)^2$ suit un χ^2 à 1 degré de liberté. Cependant, la variance de $(\hat{a}_1 \times \hat{a}_2)$ n'est plus directement calculable, elle est alors estimée par une approximation de Taylor. Sans procéder à deux estimations différentes, l'une sur le modèle non contraint et l'autre sur le modèle contraint (cf. exercice 3), le test de Wald permet, donc en une seule étape, de tester des restrictions sur les valeurs des coefficients. Les logiciels d'économétrie proposent directement ce test.

1. Nous pouvons répondre à cette question par un test d'analyse de la variance (section IV).

Exercice n° 11

Reconstitution d'informations manquantes

Un économètre recherche une relation entre le taux de mortalité (TX), les dépenses de santé (DS), le pourcentage de plus de 60 ans dans la population (POP) et la densité de médecins (DM).

Une estimation sur 10 pays effectuée à partir du logiciel RATS conduit aux résultats suivants (les données ont été préalablement centrées sur leur moyenne) :

DEPENDENT VARIABLE	6	TX				
TOTAL OBSERVATIONS	10	SKIPPED/MISSING		0		
USABLE OBSERVATIONS	10	DEGREES OF FREEDOM		????		
R**2	????	RBAR**2		.9973919		
SSR	.33774258	SEE		????		
NO.	LABEL	VAR	LAG	COEFFICIENT	STAND. ERROR	T-STATISTIC
***	*****	***	***	*****	*****	*****
1	DS	7	0	????	.1285115E-01	8.062011
2	POP	8	0	-.1881785	????	????
3	DM	9	0	.3185637	.1126177	2.828718

$$(X' X)^{-1} = \begin{pmatrix} 0,0034229 & & \\ -0,00074442 & 0,1031686 & \\ -0,018265 & -0,125482 & ??? \end{pmatrix}$$

- 1) On demande de calculer les valeurs manquantes notées par : ???.
- 2) Les coefficients des dépenses de santé et de la densité médicale sont-ils respectivement et simultanément différents à 0,10 et 0,30 ?

Solution

- 1) Les résultats complets sont les suivants :

DEPENDENT VARIABLE	6	TX				
TOTAL OBSERVATIONS	10	SKIPPED/MISSING		0		
USABLE OBSERVATIONS	10	DEGREES OF FREEDOM		7		
R**2	.99797152	RBAR**2		.99739195		
SSR	.33774258	SEE		.21965641		
NO.	LABEL	VAR	LAG	COEFFICIENT	STAND. ERROR	T-STATISTIC
***	*****	***	***	*****	*****	*****
1	DS	7	0	.1036061	.1285115E-01	8.062011
2	POP	8	0	-.1881785	.7055335E-01	-2.667180
3	DM	9	0	.3185637	.1126177	2.828718

$$(X' X)^{-1} = \begin{pmatrix} 0,0034229 & & \\ -0,00074442 & 0,1031686 & \\ -0,018265 & -0,125482 & 0,26286 \end{pmatrix}$$

2) Le test d'hypothèses est le suivant :

$$H_0 : \begin{pmatrix} a_1 \\ a_3 \end{pmatrix} = \begin{pmatrix} 0,1 \\ 0,3 \end{pmatrix}$$

$$H_1 : \begin{pmatrix} a_1 \\ a_3 \end{pmatrix} \neq \begin{pmatrix} 0,1 \\ 0,3 \end{pmatrix}$$

La relation [15] sous H_0 s'écrit :

$$\frac{1}{q} (\hat{a}_q - a_q)' \hat{\Omega}_{\hat{a}q}^{-1} (\hat{a}_q - a_q)$$

Nous avons $q = 2$, $\hat{a}_q = \begin{pmatrix} 0,103 \\ 0,318 \end{pmatrix}$ et $a_q = \begin{pmatrix} 0,1 \\ 0,3 \end{pmatrix}$. La matrice des variances covariances des coefficients est égale à :

$$\hat{\Omega}_{\hat{a}} = 0,0482 \begin{pmatrix} 0,0034229 & & \\ -0,00074442 & 0,1031686 & \\ -0,018265 & -0,125482 & 0,26286 \end{pmatrix}$$

d'où :

$$\begin{aligned} \hat{\Omega}_{\hat{a}q} &= 0,0482 \begin{pmatrix} 0,0034229 & -0,018265 \\ -0,018265 & 0,26286 \end{pmatrix} \rightarrow \\ &\rightarrow \hat{\Omega}_{\hat{a}q}^{-1} = \begin{pmatrix} 9632,962 & 669,3578 \\ 669,3578 & 125,4386 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} F^* &= \frac{1}{2} (0,103 - 0,1 \quad 0,318 - 0,3) \times \\ &\times \begin{pmatrix} 9632,962 & 669,3578 \\ 669,3578 & 125,4386 \end{pmatrix} \begin{pmatrix} 0,103 - 0,1 \\ 0,318 - 0,3 \end{pmatrix} \end{aligned}$$

$F^* = 0,172$ est à comparer à $F^\alpha(q, n - k - 1) = F_{2,7}^{0,05} = 4,74$, le F^* empirique est inférieur au F lu dans la table, on accepte l'hypothèse H_0 .

Les coefficients a_1 et a_3 sont simultanément et respectivement égaux à 0,1 et 0,3.

Exercice n° 12

Modèle à deux régimes : estimation par variable muette ou par analyse de la variance

Nous reprenons l'exercice 6 du chapitre 2 concernant la relation liant la rémunération et la durée des études. Nous disposons d'un échantillon de 40 hommes et 25 femmes ayant le même âge dont nous connaissons la rémunération annuelle (y_i) et le nombre d'année d'études (x_i).

Pour étudier l'influence du genre sur la rémunération, nous avons comparé les coefficients de régression.

1) Quel type de variable peut-on introduire pour étudier cet effet ?

2) Nous avons procédé à trois estimations :

sur l'échantillon global et nous obtenons : SCR ;

sur l'échantillon des 40 hommes : SCR_h ;

sur l'échantillon des 25 femmes : SCR_f .

Comment construire le test d'homogénéité ?

Solution

1) Une variable muette prenant la valeur 1 pour les hommes et 0 pour les femmes permet d'estimer l'effet du genre sur la rémunération. Si le coefficient de cette variable indicatrice est significativement différent de 0, alors le genre est un facteur discriminant.

2) Il s'agit d'une application du test de Chow à un modèle en coupe instantanée (test d'homogénéité de comportement).

Le Fisher empirique est égal à :

$$F^* = \frac{[SCR - (SCR_h + SCR_f)]/ddl_n}{(SCR_h + SCR_f)/ddl_d}$$

avec $ddl_n = (n - k - 1) - [(n_1 - k - 1) + (n_2 - k - 1)] = 2$

$ddl_d = (n_1 - k - 1) + (n_2 - k - 1) = n - 2(k + 1) = 65 - 4 = 61$

Si $F^* < F^{0,05} = 3,15$, l'hypothèse H_0 est acceptée, les coefficients sont stables sur l'ensemble de l'échantillon ; dans le cas contraire, il existe une différence significative entre la rémunération des hommes et des femmes (c'est le cas ici, cf. les résultats de l'exercice 6, chapitre 2).

Nous avons donc développé trois méthodes différentes d'analyse pour répondre à la même question.

Exercice n° 13

↓ fichier C3EX13

Estimation d'une fonction de production : impact de la privatisation

Soit une fonction de production de type Cobb-Douglas :

$\text{Log } Q = \text{Log } a_0 + a_1 \text{Log } K + a_2 \text{Log } L + \varepsilon_t$

Q = production

K = facteur capital

L = facteur travail

ε_t = terme d'erreur répondant aux hypothèses classiques.

Nous disposons de données concernant 50 entreprises du secteur des télécommunications en Europe pour une année donnée. Le tableau 9 présente un extrait des données.

Tableau 9 – Extrait des données de production, facteur travail, facteur capital et variable indicatrice

Entreprise	Q	K	L	DUM
1	5 192 511	11 523	4 610	1
2	8 717 373	22 476,2	6 948	1
...
49	3 701 601	10 147,4	4 760	0
50	440 380	835	248	0

1) Écrire le modèle sous sa forme originale, c'est-à-dire non linéaire. Comment s'interprètent les coefficients a_1 et a_2 de ce modèle ?

2) On désire mesurer l'homogénéité des comportements entre les entreprises publiques et les entreprises privées, pour ce faire nous intégrons une variable muette (nommée DUM) qui prend les valeurs 0 pour les entreprises publiques et 1 pour les entreprises privées.

Soit à estimer le modèle suivant :

$$\begin{aligned} \text{Log } Q = & \text{Log } a_0 + a_1 \text{Log } K + a_2 \text{Log } L + b_0 DUM + b_1 DUM \times \text{Log } K \\ & + b_2 DUM \times \text{Log } L + \varepsilon \end{aligned}$$

Existe-t-il une différence des fonctions de production entre les entreprises publiques et privées ? Si oui, pouvez-vous la quantifier ?

3) Proposer et expliciter une autre méthode afin de tester l'homogénéité des comportements des entreprises.

4) Nous proposons maintenant d'estimer une fonction de production de type « Translog » en négligeant les spécificités « public/privé » :

$$\begin{aligned} \text{Log } Q = & \text{Log } a_0 + a_1 \text{Log } K + a_2 \text{Log } L + a_3 \times 0,5 \times (\text{Log } K)^2 \\ & + a_4 \times 0,5 \times (\text{Log } L)^2 + a_5 \text{Log } L \times \text{Log } K + \varepsilon \end{aligned}$$

La spécification Translog vous semble-t-elle mieux adaptée que la spécification Cobb-Douglas ? Vous justifierez votre réponse par un test approprié.

Solution

1) le modèle sous sa forme initiale s'écrit : $Q = a_0 K^{a_1} L^{a_2} e^\varepsilon$

Les coefficients s'interprètent comme étant des élasticités :

a_1 = élasticité de la production au facteur capital

a_2 = élasticité de la production au facteur travail

Il s'agit d'un modèle linéaire sur ses coefficients, c'est-à-dire non linéaire sous sa forme initiale mais linéarisable par une simple transformation logarithmique, le modèle est Log-linéaire.

2) Les résultats d'estimation sont les suivants :

Dependent Variable : LOG(Q)				
Included observations : 50				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.477194	0.287830	22.50358	0.0000
LOG(L)	0.230860	0.143231	1.611803	0.1142
LOG(K)	0.747993	0.143545	5.210878	0.0000
DUM	- 1.018567	0.457088	- 2.228384	0.0310
DUM*LOG(L)	0.212946	0.241356	0.882290	0.3824
DUM*LOG(K)	- 0.091142	0.225917	- 0.403431	0.6886

Les variables LOG(L), DUM*LOG(L) et DUM*LOG(K) ne sont pas significatives (probabilités critiques supérieures à 5 %) nous commençons par retirer du modèle la moins significative DUM*LOG(K). La nouvelle estimation est la suivante.

Dependent Variable : LOG(Q)				
Included observations : 50				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.517159	0.267719	24.34327	0.0000
LOG(L)	0.266781	0.111144	2.400321	0.0206
LOG(K)	0.711198	0.109808	6.476720	0.0000
DUM	- 1.072324	0.433148	- 2.475654	0.0171
DUM*LOG(L)	0.118229	0.055440	2.132570	0.0385

Les variables sont maintenant toutes significatives, nous constatons que les variables indicatrices DUM sont explicatives : il existe donc une différence significative entre les fonctions de production des entreprises publiques et privées.

Pour les entreprises publiques, la fonction de production s'écrit ($DUM = 0$) :

$$\text{Log } Q = 6,52 + 0,266 \text{ Log } L + 0,711 \text{ Log } K + e$$

Pour les entreprises privées, la fonction de production s'écrit ($DUM = 1$) :

$$\begin{aligned} \text{Log } Q = 6,52 + 0,266 \text{ Log } L + 0,711 \text{ Log } K - 1,07 DUM \\ + 0,11 DUM \times \text{Log } L + e \end{aligned}$$

$$\text{Log } Q = 5,44 + 0,384 \text{ Log } L + 0,711 \text{ Log } K + e$$

Nous remarquons que le facteur travail contribue plus à la production pour les entreprises du secteur privé que pour les entreprises du secteur public.

3) Nous pouvons procéder au test de Chow en segmentant notre échantillon entre les entreprises du secteur privé et les entreprises du secteur public.

Soit à calculer la statistique :
$$F^* = \frac{(SCR - (SCR^{Privé} + SCR^{Publique})) / (k + 1)}{(SCR^{Privé} + SCR^{Publique}) / (n - 2(k + 1))}$$

qui suit alors une loi de Fisher à $(k + 1)$ et $n - 2 \times (k + 1)$ degrés de liberté.

SCR = somme des carrés des résidus calculée sur la totalité de l'échantillon des 50 entreprises.

$SCR^{Privé}$ = somme des carrés des résidus calculée sur les 21 entreprises privées.

$SCR^{Publique}$ = somme des carrés des résidus calculée sur les 29 entreprises publiques.

$k = 2$ et $n = 50$.

Soit
$$F^* = \frac{(1,853 - (0,795 + 0,629)) / 3}{(0,795 + 0,629) / 44} \approx 4,42 > F_{3;44}^{0,05} \approx 2,84$$
. Nous rejetons

l'hypothèse H_0 , le modèle est instable, il existe une différence significative entre les fonctions de production des entreprises publiques et des entreprises privées.

4) Les résultats d'estimation de la fonction « Tranlog » sont les suivants :

Dependent Variable : LOG(Q)				
Included observations : 50				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.404574	1.957160	2.761437	0.0084
LOG(K)	1.686188	1.201063	1.403913	0.1674
LOG(L)	- 0.612883	1.071805	- 0.571823	0.5704
0.5*LOG(K)^2	0.031900	0.939072	0.033970	0.9731
0.5*LOG(L)^2	0.289379	1.031212	0.280620	0.7803
LOG(L)*LOG(K)	- 0.156205	0.978803	- 0.159588	0.8739

La fonction Translog se ramène à une fonction Cobb-Douglas si $a_3 = a_4 = a_5 = 0$, soit à tester l'hypothèse $H_0 : a_3 = a_4 = a_5 = 0$. Si nous rejetons l'hypothèse H_0 (au moins un coefficient non nul), cela signifie que la spécification Translog est mieux adaptée que la spécification Cobb-Douglas. Il s'agit d'un test de Fisher de contrainte (test de Wald) d'un sous-ensemble de coefficients, le résultat est le suivant :

Wald Test :			
Equation : TRANSLOG			
Test Statistic	Value	df	Probability
F-statistic	0.401096	(3.44)	0.7529

La probabilité critique du test est de 75 %, nous acceptons donc l'hypothèse H_0 , la spécification Translog n'est donc pas mieux adaptée que la spécification Cobb-Douglas.

Exercice n° 14

Fonction économétrique d'investissement

Une estimation économétrique de la part des emprunts dans les investissements d'une importante entreprise pétrolière donne :

$$\frac{\text{EMP}}{\text{INV}} = -0,6645 \times \text{TIR} + 179,07 \times \frac{\text{INVO}}{\text{INVI}} + 36,31 + e$$

(1,26) (4,69)

$R^2 = 0,65$; $n = 17$ (nombre d'années de la période d'estimation)

Les chiffres entre parenthèses sous les coefficients sont les t de Student avec :

EMP = emprunts totaux

INV = investissements totaux

TIR = taux d'intérêt réel

INVO = investissements de la société dans les forages offshore

INVI = investissements industriels

e = résidu d'estimation

- 1) Commentez les résultats obtenus, les coefficients des variables explicatives ont-ils le signe attendu ?
- 2) Quel est l'effet d'un passage de 13 % à 15 % de INVO/INVI sur la variable endogène ?
- 3) Le coefficient de la variable TIR n'est pas significativement différent de zéro. Quelle explication économique peut-on donner à ce résultat ?
- 4) On estime de nouveau ce modèle en retirant la variable TIR, les résultats sont les suivants :

$$\frac{\text{EMP}}{\text{INV}} = 162,15 \times \frac{\text{INVO}}{\text{INVI}} + 41,28 + e$$

(7,20)

$R^2 = 0,57$

Pourquoi a-t-on retiré la variable explicative TIR (donnez deux raisons possibles de la non-significativité statistique de cette variable) ?

Le coefficient de détermination a diminué, est ce fortuit ?

Les coefficients du modèle sont légèrement différents de la première estimation, pourquoi ?

- 5) L'estimation économétrique de l'équation de la question 4 permet-elle de donner une approximation du taux d'autofinancement de cette entreprise ? Donner le cas échéant une valeur lorsque INVO/INVI = 13%.

Solution

1) Le taux d'emprunt dépend :

- négativement du taux d'intérêt, ce qui est conforme à l'intuition économique ; mais le coefficient de cette variable n'est pas significativement différent de 0 ($t^* = 1,26 < t_{14}^{0,05} = 2,14$) ;

- positivement de l'investissement dans les forages *offshore*, ce qui est logique car l'investissement dans les forages « *offshore* » est très important ; le coefficient est significativement différent de 0 ($t^* = 4,69 > t_{14}^{0,05} = 2,14$).

2) Par application du modèle estimé, il vient :

$$\Delta \frac{\text{EMP}}{\text{INV}} = 179,07 \times (0,15 - 0,13) = 3,58$$

3) La compagnie pétrolière doit pratiquer l'autofinancement, donc il n'existe pas de sensibilité au taux d'intérêt.

4) La variable explicative TIR n'est pas significative car elle n'est peut être pas assez corrélée avec la variable à expliquer ou bien colinéaire avec l'autre variable explicative.

Le coefficient de détermination a diminué car le fait de retirer une variable explicative même non significative fait augmenter mécaniquement la Somme des Carrés des Résidus.

Les coefficients du modèle sont légèrement différents car la colinéarité entre les variables explicatives entraîne une modification de la valeur estimée des coefficients.

5) Par application du modèle estimé il vient :

$$\frac{\text{EMP}}{\text{INV}} = 162,15 \times 0,13 + 41,28 = 62,35$$

$$\Rightarrow \text{Autofinancement} \approx 100 - 62,35 = 37,65 \%$$

Annexe

A. Interprétation géométrique de la méthode des moindres carrés

Soit les vecteurs Y, X_1, X_2, \dots, X_k composés de n éléments que nous pouvons donc représenter dans R^n .

Nous cherchons à minimiser :

$$\text{Min} \sum_{i=1}^n \varepsilon_i^2 = \text{Min} \varepsilon' \varepsilon = \text{Min} (Y - Xa)' (Y - Xa) = \text{Min} S$$

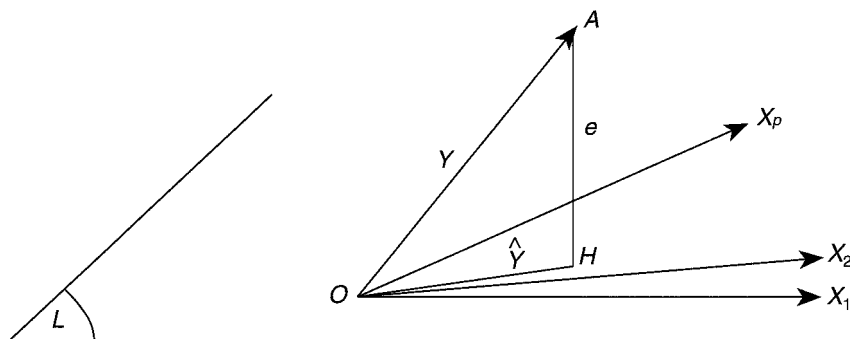
Avec $X = (X_1, X_2, \dots, X_k)$.

Le vecteur $Xa = (X_1, X_2, \dots, X_k) \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix}$ appartient au sous-espace vectoriel

(L) de R^n engendré par X_1, X_2, \dots, X_k .

La quantité S est minimale lorsque le vecteur $e = Y - Xa$ est orthogonal au sous-espace (L) .

Le graphique 7 illustre ce propos.



Graphique 7 – Interprétation géométrique de la régression multiple

B. Résolution de l'exercice 1 par des logiciels informatiques de régression multiple

• Logiciel RATS

```
Dependent Variable Y - Estimation by Least Squares
Usable Observations      14      Degrees of Freedom      10
Centered R**2      0.702687      R Bar **2      0.613493
Uncentered R**2      0.985401      T x R**2      13.796
Mean of Dependent Variable      17.714285714
Std Error of Dependent Variable      4.177385480
Standard Error of Estimate      2.597068848
Sum of Squared Residuals      67.447666031
Regression F(3,10)      7.8782
Significance Level of F      0.00545230
Durbin-Watson Statistic      3.186886
```

Variable	Coeff	Std Error	T-Stat	Signif
1. Constant	.89132428	11.66331015	2.820068	0.01815860
2. X1	0.80190069	0.29843584	2.687012	0.02281643
3. X2	-0.38136236	0.15658069	-2.435564	0.03511440
4. X3	-0.03713244	0.05202312	-0.713768	0.49169355

Nous retrouvons bien l'ensemble des résultats de l'exercice 1.

• Logiciel Eviews

=====				
LS // Dependent Variable is Y				
Sample: 1 14				
Included observations: 14				
=====				
Variable	Coefficient	Std. Error	T-Statistic	Prob.
=====				
C	32.89132	11.66331	2.820068	0.0182
X1	0.801901	0.298436	2.687012	0.0228
X2	-0.381362	0.156581	-2.435564	0.0351
X3	-0.037132	0.052023	-0.713768	0.4917
=====				
R-squared	0.702687	Mean dependent var	17.71429	
Adjusted R-squared	0.613493	S.D. dependent var	4.177385	
S.E. of regression	2.597069	Akaike info criter	2.143723	
Sum squared resid	67.44767	Schwartz criterion	2.326311	
Log likelihood	-30.87120	F-statistic	7.878181	
Durbin-Watson stat	3.186886	Prob(F-statistic)	0.005452	
=====				

• Tableau EXCEL

– par la fonction DROITEREG

-0.03713	-0.38136	0.801901	32.89132
0.052023	0.156581	0.298436	11.66331
0.702687	2.597069	#N/A	#N/A
7.878181	10	#N/A	#N/A
159.4095	67.44767	#N/A	#N/A

Interprétation et correspondance des informations issues du tableau EXCEL

\hat{a}_3	\hat{a}_2	\hat{a}_1	\hat{a}_0
$\hat{\sigma}_{\hat{a}_3}$	$\hat{\sigma}_{\hat{a}_2}$	$\hat{\sigma}_{\hat{a}_1}$	$\hat{\sigma}_{\hat{a}_0}$
R^2	$\hat{\sigma}_e$	#N/A	#N/A
F^*	ddl	#N/A	#N/A
SCE	SCR	#N/A	#N/A

Par l'utilitaire d'analyse « Régression Linéaire » (Macro-complémentaire)

RAPPORT DÉTAILLÉ	
Statistiques de la régression	
Coefficient de détermination multiple	0,838
Coefficient de détermination R^2	0,702
Coefficient de détermination R^2 (ajusté)	0,613
Erreur-type	2,597
Observations	14

ANALYSE DE VARIANCE				
	Degré de liberté	Somme des carrés	Moyenne des carrés	F
Régression	3	159,409	53,1368	7,8786
Résidus	10	67,447	6,74476	
Total	13	226,857		

	Coefficients	Erreur-type	Statistique <i>t</i>	Probabilité
Constante	32,891	11,663	2,820	0,0181
X1	0,80190	0,298	2,687	0,0228
X2	-0,3813	0,156	-2,435	0,0351
X3	-0,0371	0,05202	-0,713	0,4916

C. Estimation de la variance de l'erreur

La variance de l'erreur est inconnue, nous allons donc l'estimer à partir de la somme des carrés des résidus.

Soit : $e = Y - X\hat{a} = Y - X(X'X)^{-1}X'Y = PY$ avec $P = I - X(X'X)^{-1}X'$.

Cette matrice P est symétrique et idempotente ($P^2 = P$). Elle est aussi telle que $PX = 0$ et $Pe = e$.

$e = PY = P(Xa + \varepsilon) = P\varepsilon$ puisque $PX = 0$.

D'où l'espérance de l'erreur est donnée par : $E(e'e) = E(\varepsilon'P'\varepsilon) = E(\varepsilon'P\varepsilon)$ car ($P^2 = P$).

En utilisant le fait que la trace¹ d'un scalaire est un scalaire, il vient :

$$\begin{aligned}
 E(\varepsilon'P\varepsilon) &= E[\text{Trace}(\varepsilon'P\varepsilon)] = E[\text{Trace}(\varepsilon\varepsilon'P)] \\
 &= \sigma_\varepsilon^2 \text{Trace}(P) = \sigma_\varepsilon^2 \text{Trace}(I - X(X'X)^{-1}X') \\
 &= \sigma_\varepsilon^2 \text{Trace}(I) - \sigma_\varepsilon^2 \text{Trace}[X(X'X)^{-1}X'] \\
 &= \sigma_\varepsilon^2 \text{Trace}(I) - \sigma_\varepsilon^2 \text{Trace}[(X'X)^{-1}(X'X)] = \sigma_\varepsilon^2(n - (k + 1)) = \sigma_\varepsilon^2(n - k - 1) \cdot
 \end{aligned}$$

$\text{Trace}(I) = n$ et $\text{Trace}[(X'X)^{-1}(X'X)] = k + 1$ car $(X'X)$ est de dimension $(k + 1, k + 1)$.

D'où l'estimateur de la variance de l'erreur (appelée aussi la variance résiduelle) s'écrit donc :

$$\hat{\sigma}_\varepsilon^2 = \frac{e'e}{n - k - 1} = \frac{\sum_t e_t^2}{n - k - 1}$$

1. La trace d'une matrice est égale à la somme des valeurs de la première diagonale.

4. Multicolinéarité et sélection du modèle optimal

Dans ce chapitre, nous examinons comment gérer l'abondance de l'information. En effet, il est fréquent, lorsque l'économiste spécifie un modèle, qu'il hésite à intégrer telle ou telle variable explicative. La question essentielle qu'il se pose est la suivante : comment déterminer le mix optimal des variables explicatives ? Formulée en terme statistique, cette question se résume à trouver les variables explicatives qui maximisent leur coefficient de corrélation avec la série à expliquer, tout en étant les moins corrélées entre elles.

Nous présentons, en I., tout d'abord la notion de corrélation partielle qui permet de déterminer l'apport relatif de chaque série explicative à l'explication de la série endogène.

Les relations entre les coefficients de corrélation simple, partielle et multiple sont examinées en II.

Les conséquences de la multicolinéarité sur l'estimation des coefficients de régression font l'objet du III.

Enfin, les techniques de sélection de variables explicatives sont exposées et illustrées à partir d'exemples en IV.

I. Corrélation partielle

A. Exemple introductif

Un marchand de glaces, situé près de la tour Eiffel, cherche à calculer le coefficient de corrélation entre ses ventes (x_1) et le nombre de touristes visitant ce monument (x_2). Ces deux variables sont influencées par le climat : la consom-

mation de glaces est plus importante lorsqu'il fait chaud et les touristes sont peu enclins à visiter un monument extérieur en cas de froid ou de pluie, on appelle x_3 cette variable climatique.

Nous pouvons penser que la corrélation entre x_1 et x_2 est positive, cependant un calcul de coefficient de corrélation simple n'est pas révélateur du degré de liaison réelle entre ces deux variables ; en effet, la variable climat influence la vente des glaces et la fréquentation des touristes. En d'autres termes, le coefficient de corrélation simple calculé ainsi intègre l'apport de la variabilité des conditions climatiques sans pouvoir isoler l'influence relative du nombre de touristes.

B. Généralisation de la notion de corrélation partielle

Le coefficient de corrélation partielle mesure la liaison entre deux variables lorsque l'influence d'une ou des autres variables explicatives est retirée.

Soit y une variable à expliquer, et x_1 , x_2 et x_3 des variables explicatives :

r^2_{y, x_1} , r^2_{y, x_2} et r^2_{y, x_3} mesurent respectivement la variance de y expliquée par la variable x_1 seule, x_2 seule et x_3 seule.

Nous pouvons calculer six coefficients de corrélation partielle du premier ordre :

$$r^2_{yx1.x2} ; r^2_{yx1.x3} ; r^2_{yx2.x1} ; r^2_{yx2.x3} ; r^2_{yx3.x1} ; r^2_{yx3.x2} ;$$

ainsi que trois coefficients de corrélation partielle du deuxième ordre :

$$r^2_{yx1.x2 \ x3} ; r^2_{yx2.x1 \ x3} ; r^2_{yx3.x1 \ x2}.$$

Cette notion de corrélation partielle est très importante car elle permet de juger de la pertinence d'inclure une variable explicative dans un modèle.

Plus le coefficient de corrélation partielle d'une variable est élevé, plus la contribution de cette variable est importante à l'explication globale du modèle.

Le coefficient de corrélation partielle peut se calculer de deux manières à partir :

1) Du coefficient de corrélation simple

- Entre le résidu de la régression de la variable à expliquer sur le sous-ensemble des $k - 1$ autres variables explicatives,
- et le résidu de la régression de la variable explicative x_i sur les $k - 1$ variables explicatives.

Exemple de calcul de $r^2_{yx3.x1 \ x2}$

Étape 1 : calcul de résidu e_1 de la régression de y sur x_1 et x_2

$$e_1 = y - (\hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2)$$

Étape 2 : calcul de résidu e_2 de la régression de x_3 sur x_1 et x_2

$$e_2 = x_3 - (\hat{a}'_0 + \hat{a}'_1 x_1 + \hat{a}'_2 x_2)$$

Étape 3 : calcul du coefficient de corrélation simple entre e_1 et e_2

$$r^2_{yx3.x1\ x2} = r^2_{e1,e2}$$

2) Du t de Student

Dans un modèle à k variables explicatives, il existe une relation entre le coefficient de corrélation partielle et le t de Student¹ :

$$r^2_{yxi.(autres\ variables)} = \frac{t_i^2}{t_i^2 + (n - k - 1)} \quad [1]$$

Attention : cette relation n'est vérifiée que pour un coefficient de corrélation partielle d'ordre $k - 1$.

Exercice n° 1

↓ fichier C4EX1

Calcul des coefficients de corrélation partielle du premier ordre et du deuxième ordre

Une entreprise cherche à déterminer quelles sont les influences respectives de la publicité (x_1), de la promotion auprès des distributeurs (x_2), et de la promotion auprès des consommateurs (x_3) exprimées en milliers d'euros sur les ventes en tonnes (y).

Pour ce faire, elle dispose, sur 10 ans, des séries statistiques du tableau 1.

Tableau 1 – Ventes, publicité et promotions

Année	y	x_1	x_2	x_3
1	49,0	35,0	53,0	200,0
2	40,0	35,0	53,0	212,0
3	41,0	38,0	50,0	211,0
4	46,0	40,0	64,0	212,0
5	52,0	40,0	70,0	203,0
6	59,0	42,0	68,0	194,0
7	53,0	44,0	59,0	194,0
8	61,0	46,0	73,0	188,0
9	55,0	50,0	59,0	196,0
10	64,0	50,0	71,0	190,0

1. Greene W.H., pages 234-235, 2000.

Pour répondre à cette question, on demande de calculer les coefficients de corrélation partielle du premier ordre et du deuxième ordre et de commenter les résultats obtenus.

Solution

Nous traitons complètement le calcul de $r^2_{yx1.x2}$ et $r^2_{yx3.x1\ x2}$, les résultats des autres coefficients sont fournis afin que le lecteur puisse vérifier ses propres calculs.

a) Calcul de $r^2_{yx1.x2}$ par corrélation entre les résidus

Étape 1 : régression de y sur x_2

DEPENDENT VARIABLE y						
R**2	.64684945					
NO.	LABEL	VAR	LAG	COEFFICIENT	STAND. ERROR	T-STATISTIC
***	*****	***	***	*****	*****	*****
1	CONSTANT	0	0	3.580952	12.75204	.2808141
2	x2	3	0	.7809524	.2040130	3.827954

Résidu $e_1 = y - (0,7809\ x_2 + 3,5809)$

Étape 2 : régression de x_1 sur x_2

DEPENDENT VARIABLE x1						
R**2	.33862434					
NO.	LABEL	VAR	LAG	COEFFICIENT	STAND. ERROR	T-STATISTIC
***	*****	***	***	*****	*****	*****
1	CONSTANT	0	0	18.38095	11.76556	1.562268
2	x2	3	0	.3809524	.1882308	2.023858

Résidu $e_2 = x_1 - (0,3809\ x_2 + 18,3809)$

Étape 3 : calcul du coefficient de corrélation entre e_1 et e_2

	e_1	e_2
	—	—
1	4,02857	−3,57143
2	−4,97143	−3,57143
3	−1,62857	0,57142
4	−7,56190	−2,76190
5	−6,24762	−5,04762
6	2,31429	−2,28571
7	3,34286	3,14286
8	0,40952	−0,19047
9	5,34286	9,14286
10	4,97143	4,57143

$$r_{e1,e2} = 0,6798 \rightarrow r^2_{yx1.x2} = r^2_{e1,e2} = 0,4621$$

La corrélation entre les ventes et la publicité est donc égale à 0,46, lorsque l'influence de la promotion auprès des distributeurs est retirée.

b) Calcul de $r^2_{yx3.x1\ x2}$ par corrélation entre les résidus

Étape 1 : régression de y sur x_1 et x_2

DEPENDENT VARIABLE y						
R**2 .81005118						
NO.	LABEL	VAR	LAG	COEFFICIENT	STAND. ERROR	T-STATISTIC
***	*****	***	***	*****	*****	*****
1	CONSTANT	0	0	-9.962133	11.42177	-.8722058
2	x1	2	0	.7368000	.3004391	2.452410
3	x2	3	0	.5002667	.1968836	2.543510

$$\text{Résidu } e_1 = y - (0,7368 x_1 + 0,50 x_2 - 9,9621)$$

Étape 2 : régression de x_3 sur x_1 et x_2

DEPENDENT VARIABLE x3						
R**2 .58912000						
NO.	LABEL	VAR	LAG	COEFFICIENT	STAND. ERROR	T-STATISTIC
***	*****	***	***	*****	*****	*****
1	CONSTANT	0	0	258.8320	18.87600	13.71223
2	x1	2	0	-.9520000	.4965159	-1.917361
3	x2	3	0	-.3040000	.3250459	-.9352524

$$\text{Résidu } e_2 = x_3 - (0,952 x_1 - 0,304 x_2 + 258,932)$$

Étape 3 : calcul du coefficient de corrélation entre e_1 et e_2

	e_1	e_2
	—	—
1	6,66000	-9,4000
2	-2,34000	2,6000
3	-2,04960	3,5440
4	-5,52693	10,7040
5	-2,52853	3,5280
6	3,99840	-4,1760
7	1,02720	-5,0080
8	0,54986	-4,8480
9	-1,39360	2,7040
10	1,60320	0,3520

$$r_{e_1, e_2} = -0,921 \rightarrow r_{yx3, x1 \ x2}^2 = r_{e_1, e_2}^2 = 0,848$$

La corrélation entre les ventes et la promotion « consommateur » est donc égale à 0,84, lorsque les influences de la publicité et de la promotion « distributeur » sont retirées.

c) Calcul de $r_{yx3, x1 \ x2}^2$ à partir du t Student

Étape 1 : régression de y sur x_1 , x_2 et x_3

DEPENDENT VARIABLE 1 y						
R**2 .97119009						
NO.	LABEL	VAR	LAG	COEFFICIENT	STAND. ERROR	T-STATISTIC
***	*****	***	***	*****	*****	*****
1	CONSTANT	0	0	134.2903	25.36038	5.295282
2	x1	2	0	.2062305	.1560786	1.321325
3	x2	3	0	.3308411	.8775302E-01	3.770140
4	x3	4	0	-.5573209	.9620561E-01	-5.793019

Étape 2 : calcul du coefficient de corrélation partielle

D'après [1], nous avons :

$$r_{yx3.x1\ x2}^2 = \frac{t_i^2}{t_i^2 + (n - k - 1)} = \frac{5,79^2}{5,79^2 + (10 - 4)} = 0,848$$

d) Autres résultats

$$\begin{array}{lll} r_{yx1.x2}^2 = 0,461 & r_{yx1.x3}^2 = 0,212 & r_{yx2.x1}^2 = 0,480 \\ r_{yx2.x3}^2 = 0,698 & r_{yx3.x1}^2 = 0,734 & r_{yx3.x2}^2 = 0,895 \\ r_{yx1.x2\ x3}^2 = 0,225 & r_{yx2.x1\ x3}^2 = 0,703 & r_{yx3.x1\ x2}^2 = 0,848 \end{array}$$

À la lecture de ces chiffres, l'apport intrinsèque de la variable promotion « consommateur » semble le plus important pour expliquer les ventes : $r_{yx3.x1}^2 = 0,734$; $r_{yx3.x2}^2 = 0,895$ et $r_{yx3.x1\ x2}^2 = 0,848$ sont les trois coefficients de corrélation partielle les plus élevés.

II. Relation entre coefficients de corrélation simple, partielle et multiple

Dans le cas d'un modèle à une seule variable explicative x_1 , la somme des carrés des résidus est égale à :

$$SCR = \sum_t e_t^2 = \sum_t (y_t - \bar{y})^2 (1 - R_{y.x1}^2) = \sum_t (y_t - \bar{y})^2 (1 - r_{yx1}^2)$$

$R_{y.x1}^2$ = coefficient de détermination de la régression de y sur x_1 .

Supposons maintenant un modèle à deux variables explicatives :

$$y = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2 + e$$

La somme des carrés des résidus, après avoir retiré l'influence de x_1 et x_2 , est égale à :

$$SCR = \sum_t e_t^2 = \sum_t (y_t - \bar{y})^2 (1 - R_{y.x1x2}^2)$$

$R_{y.x1\ x2}^2$ = coefficient de détermination de la régression de y sur x_1 et x_2 .

Or $r_{yx2.x1}^2$ est la proportion du résidu expliquée par la variable x_2 seule, cette dernière expression peut donc s'écrire :

$$\sum_t (y_t - \bar{y})^2 (1 - R_{y.x1x2}^2) = (1 - r_{yx2.x1}^2) \sum_t (y_t - \bar{y})^2 (1 - r_{yx1}^2) \quad [2]$$

D'où

$$1 - R_{y.x1\ x2}^2 = (1 - r_{yx2.x1}^2)(1 - r_{yx1}^2) \quad [3]$$

Nous avons ainsi la décomposition de l'influence relative de chacune des variables : l'apport de x_2 sur y lorsque l'influence de x_1 est retirée et l'apport de x_1 sur y (certains auteurs emploient le terme de rendement pour qualifier ces apports relatifs).

Cette formule [3] peut être généralisée pour un modèle à trois variables explicatives :

$$1 - R_{y.x_1 x_2 x_3}^2 = (1 - r_{yx1}^2)(1 - r_{yx2.x1}^2)(1 - r_{yx3.x1 x_2}^2) \quad [4]$$

Pour un modèle à quatre variables explicatives, nous obtenons :

$$1 - R_{y.x_1 x_2 x_3 x_4}^2 = (1 - r_{yx1}^2)(1 - r_{yx2.x1}^2)(1 - r_{yx3.x1 x_2}^2)(1 - r_{yx4.x1 x_2 x_3}^2)$$

Il est à noter que les indices peuvent permuter, ainsi, dans l'ordre, 3, 4, 1, 2 ; la relation précédente devient :

$$1 - R_{y.x_1 x_2 x_3 x_4}^2 = (1 - r_{yx3}^2)(1 - r_{yx4.x3}^2)(1 - r_{yx1.x3 x_4}^2)(1 - r_{yx2.x1 x_3 x_4}^2)$$

Exercice n° 2

Relation entre coefficients de corrélation simple, partielle et multiple

Un économètre estime un modèle à trois variables explicatives :

$$\hat{y} = 348,4 + 56,3 x_1 - 9,5 x_2 + 234,8 x_3$$

(4,5) (8,3) (2,1) (4,6)

$$R^2 = 0,76$$

$$n = 65$$

$$(.) = t \text{ de Student}$$

Le coefficient de corrélation linéaire entre y et x_1 est connu, $r_{yx1}^2 = 0,52$.

On demande de calculer le coefficient de corrélation partielle $r_{yx2.x1}^2$.

Solution

Le coefficient $r_{yx3.x1 x_2}^2$ est égal à : $\frac{t_{\hat{a}_3}^2}{t_{\hat{a}_3}^2 + (n - k - 1)} = \frac{4,6^2}{4,6^2 + (65 - 4)}$

$$r_{yx3.x1 x_2}^2 = 0,257$$

Nous pouvons alors utiliser la relation suivante :

$$1 - R_{y.x_1 x_2 x_3}^2 = (1 - r_{yx1}^2)(1 - r_{yx2.x1}^2)(1 - r_{yx3.x1 x_2}^2)$$

$$1 - 0,76 = (1 - 0,52)(1 - r_{yx2.x1}^2)(1 - 0,257)$$

d'où $r_{yx2.x1}^2 = 0,324$

III. **M**ulticolinéarité :

conséquences et détection

Le terme de multicolinéarité est employé dans le cas d'un modèle incorporant des séries explicatives qui sont liées entre elles.

À l'opposé, pour des séries explicatives de covariance nulle ($\text{Cov}(x_1, x_2) = 0$), nous dirons qu'elles sont orthogonales. Si, pour des études théoriques, nous pouvons supposer que deux séries statistiques sont orthogonales, dans la pratique, lorsque l'économiste modélise des phénomènes économiques, les séries explicatives sont toujours plus ou moins liées entre elles. Nous allons donc examiner les conséquences de la multicolinéarité.

A. Conséquences de la multicolinéarité

Nous pouvons citer trois effets principaux :

- a) *augmentation de la variance estimée de certains coefficients* lorsque la colinéarité entre les variables explicatives augmente (le t de Student diminue) ;
- b) *instabilité des estimations des coefficients* des moindres carrés, des faibles fluctuations concernant les données entraînent des fortes variations des valeurs estimées des coefficients ;
- c) *en cas de multicolinéarité parfaite*, la matrice $X'X$ est singulière (le déterminant est nul), *l'estimation des coefficients est alors impossible et leur variance est infinie*.

Exercice n° 3

Conséquences de la multicolinéarité

Soit un modèle à deux variables explicatives dont les matrices $X'X$ et $X'Y$ sont calculées à partir d'un échantillon d'observations de variables centrées.

$$X'X = \begin{bmatrix} 200 & 150 \\ 150 & 113 \end{bmatrix} \quad X'Y = \begin{bmatrix} 350 \\ 263 \end{bmatrix}$$

La suppression d'une observation modifie les résultats de la manière suivante :

$$X'X = \begin{bmatrix} 199 & 149 \\ 149 & 112 \end{bmatrix} \quad X'Y = \begin{bmatrix} 347,5 \\ 261,5 \end{bmatrix}$$

On demande :

- 1) de calculer les coefficients de régression du modèle dans les deux cas ;
- 2) puis de calculer le coefficient de corrélation entre x_1 et x_2 ;
- 3) de commenter les résultats.

Solution

1) Calcul des coefficients de régression

- a) Estimation des coefficients à partir des données de base. Les équations normales s'écrivent :

$$\begin{aligned} 200 \hat{a}_1 + 150 \hat{a}_2 &= 350 & \hat{a}_1 &= 1 \\ &\Rightarrow & & \\ 150 \hat{a}_1 + 113 \hat{a}_2 &= 263 & \hat{a}_2 &= 1 \end{aligned}$$

- b) Estimation des coefficients après suppression d'une observation. Les équations normales s'écrivent :

$$\begin{aligned} 199 \hat{a}_1 + 149 \hat{a}_2 &= 347,5 & \hat{a}_1 &= -0,5 \\ &\Rightarrow & & \\ 149 \hat{a}_1 + 112 \hat{a}_2 &= 261,5 & \hat{a}_2 &= 3 \end{aligned}$$

2) Calcul du coefficient de corrélation

$$\rho_{x_1, x_2}^2 = \frac{[\text{Cov}(x_1, x_2)]^2}{\text{Var}(x_1) \times \text{Var}(x_2)} = 0,995 \longrightarrow \rho_{x_1, x_2} = \sqrt{0,995} = 0,997$$

- 3) **Commentaires** : une faible modification d'une des observations ou du nombre d'observations entraîne une profonde modification des valeurs estimées des coefficients, cela est la conséquence directe de la très forte colinéarité entre x_1 et x_2 (0,997).

Afin d'éviter ce problème, il convient alors de détecter une éventuelle multicollinéarité lors de l'estimation d'un modèle.

B. Tests de détection d'une multicollinéarité

1) Test de Klein¹

Le test de Klein est fondé sur la comparaison du coefficient de détermination R_y^2 calculé sur le modèle à k variables :

$$y = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2 + \dots + \hat{a}_k x_k + e$$

1. Klein L.R., page 101, 1962.

et les coefficients de corrélation simple $r_{xi, xj}^2$ entre les variables explicatives pour $i \neq j$.

Si $R_y^2 < r_{xi, xj}^2$, il y a présomption de multicolinéarité.

Il ne s'agit pas d'un test statistique au sens test d'hypothèses mais simplement d'un critère de présomption de multicolinéarité.

2) Test de Farrar et Glauber¹

La première étape consiste à calculer le déterminant de la matrice des coefficients de corrélation entre les variables explicatives.

$$D = \begin{vmatrix} 1 & r_{x1\ x2} & r_{x1\ x3} & \dots & r_{x1\ xk} \\ r_{x2\ x1} & 1 & r_{x2\ x3} & \dots & r_{x2\ xk} \\ \dots & \dots & \dots & \dots & \dots \\ r_{xk\ x1} & r_{xk\ x2} & r_{xk\ x3} & \dots & 1 \end{vmatrix}$$

Lorsque la valeur du déterminant D tend vers zéro, le risque de multicolinéarité est important.

Par exemple, pour un modèle à deux variables explicatives, si les deux séries sont parfaitement corrélées, le déterminant D s'écrit :

$$D = \begin{vmatrix} 1 & r_{x1\ x2} \\ r_{x2\ x1} & 1 \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} = 0$$

Dans le cas opposé où les séries explicatives sont orthogonales, le déterminant devient :

$$D = \begin{vmatrix} 1 & r_{x1\ x2} \\ r_{x2\ x1} & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1$$

La deuxième étape consiste à effectuer un test du χ^2 , en posant les hypothèses suivantes :

$H_0 : D = 1$ (les séries sont orthogonales) ;

$H_1 : D < 1$ (les séries sont dépendantes).

La valeur empirique du χ^2 calculée à partir de l'échantillon est égale à :

$$*\chi^2 = - \left[n - 1 - \frac{1}{6} (2K + 5) \right] \cdot \text{Ln } D$$

où n est la taille de l'échantillon, K le nombre de variables explicatives (terme constant inclus, $K = k + 1$) et Ln le logarithme népérien.

Si $*\chi^2 \geq \chi^2$ lu dans la table à $\frac{1}{2}K(K-1)$ degrés de liberté et au seuil α choisi, alors l'hypothèse H_0 est rejetée, il y a donc présomption de multicolinéarité.

Si $*\chi^2 < \chi^2$, alors nous acceptons l'hypothèse d'orthogonalité.

1. Farrar D.E. et Glauber R.R., 1967.

Exercice n° 4

↓ fichier C4EX4

Tests de multicolinéarité

Un économiste cherche à expliquer la variable y à l'aide de quatre séries explicatives x_1 , x_2 , x_3 et x_4 . Il désire auparavant tester une éventuelle multicolinéarité entre ces quatre séries ; pour ce faire, il dispose des données du tableau 2.

Tableau 2 – Ces séries statistiques sont-elles colinéaires ?

y	x_1	x_2	x_3	x_4
8,40	82,90	17,10	92,00	94,00
9,60	88,00	21,30	93,00	96,00
10,40	99,90	25,10	96,00	97,00
11,40	105,30	29,00	94,00	97,00
12,20	117,70	34,00	100,00	100,00
14,20	131,00	40,00	101,00	101,00
15,80	148,20	44,00	105,00	104,00
17,90	161,80	49,00	112,00	109,00
19,30	174,20	51,00	112,00	111,00
20,80	184,70	53,00	112,00	111,00

On demande donc d'appliquer les tests suivants :

- 1) test de Klein,
- 2) test de Farrar-Glauber.

Solution

1) Test de Klein

a) Régression : $y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4 + \varepsilon$.

Les résultats de l'estimation sont les suivants :

$$y = -13,53 + 0,096 x_1 + 0,015 x_2 - 0,199 x_3 + 0,34 x_4 + e$$

$$(1,80) \quad (3,66) \quad (0,30) \quad (2,20) \quad (2,27)$$

$$n = 10$$

$$R^2 = 0,998$$

(.) = t de Student

b) Calcul des coefficients de corrélation simple entre les variables explicatives :

$$r_{x_1, x_2}^2 = 0,976 \quad r_{x_1, x_3}^2 = 0,960 \quad r_{x_1, x_4}^2 = 0,974$$

$$r_{x_2, x_3}^2 = 0,938 \quad r_{x_2, x_4}^2 = 0,938 \quad r_{x_3, x_4}^2 = 0,982$$

À la lecture de ces coefficients il ne semble pas ressortir des risques graves de multicolinéarité puisque tous les coefficients de corrélation simple sont inférieurs au coefficient de détermination. Toutefois, nous observons qu'ils sont tous très élevés.

2) Test de Farrar-Glauber

Calcul du déterminant

$$D = \begin{vmatrix} 1 & r_{x1\ x2} & r_{x1\ x3} & r_{x1\ x4} \\ r_{x2\ x1} & 1 & r_{x2\ x3} & r_{x2\ x4} \\ r_{x3\ x1} & r_{x3\ x2} & 1 & r_{x3\ x4} \\ r_{x4\ x1} & r_{x4\ x2} & r_{x4\ x3} & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0,988 & 0,980 & 0,987 \\ 0,988 & 1 & 0,969 & 0,969 \\ 0,980 & 0,969 & 1 & 0,991 \\ 0,987 & 0,969 & 0,991 & 1 \end{vmatrix} = 0,92198 \times 10^{-5}$$

La valeur empirique du $^*\chi^2$ calculée à partir de l'échantillon est :

$$^*\chi^2 = - \left[n - 1 - \frac{1}{6} (2K + 5) \right] \times \text{Ln } D$$

$$^*\chi^2 = - \left[10 - 1 - \frac{1}{6} (2 \times 5 + 5) \right] \times -11,59 = 75,33$$

Cette valeur est à comparer à la valeur lue dans la table :

χ^2 à $\frac{1}{2} K (K - 1) = 10$ degrés de liberté et pour un seuil $\alpha = 0,05$, soit :

$$\chi^2 = 18,31$$

Puisque $^*\chi^2 > \chi^2$, nous rejetons l'hypothèse H_0 , il y a présomption de multicollinéarité.

Ces deux tests conduisent donc à des résultats différents, cependant le test de Farrar-Glauber, dont le fondement théorique est plus affirmé, semble devoir être privilégié.

C. Comment remédier à la multicollinéarité ?

Nous ne ferons que citer les différentes techniques permettant d'apporter des solutions au problème de multicollinéarité, le lecteur intéressé par les développements méthodologiques peut se reporter à Judge¹.

- *Augmenter la taille de l'échantillon* : cette technique n'est efficace que si l'ajout d'observations diffère significativement de celles figurant déjà dans le modèle, sinon il y aura reconduction de la multicollinéarité.
- *La « Ridge Regression »* est une réponse purement numérique, il s'agit de transformer la matrice $X'X$ en $(X'X + cI)$ où c est une constante choisie arbitrairement qui, en augmentant les valeurs de la première diagonale, réduit les effets « numériques » de la multicollinéarité.

1. Judge G.G. *et al*, pages 874-882, 1988.

Face à ces artifices de calcul, la seule parade vraiment efficace consiste, lors de la spécification de modèle, à éliminer les séries explicatives susceptibles de représenter les mêmes phénomènes et donc d'être corrélées entre elles, ceci afin d'éviter *l'effet de masque*.

Au paragraphe suivant, nous présentons des méthodes permettant de déterminer le « mix optimal » de variables explicatives.

IV. Sélection du modèle optimal

Dans la pratique, l'économètre est souvent confronté au choix de plusieurs variables explicatives x_1, x_2, \dots, x_k candidates pour expliquer la variable y . Des procédures statistiques permettent de déterminer quelles variables retirer ou quelles variables ajouter dans un modèle. Cette démarche exclut tout raisonnement économique car elle permet d'aboutir à des modèles économétriques qui sont souvent bons sur le plan statistique mais dont l'interprétation économique s'avère nulle, voire aberrante. Ces techniques automatiques de sélection de variables explicatives sont donc à manier avec prudence.

Au problème de la sélection des variables explicatives s'ajoute celui du choix à effectuer entre plusieurs modèles concurrents, c'est-à-dire des modèles dont les variables sont toutes significatives mais qui ne sont pas les mêmes.

La question se pose alors de sélectionner le meilleur modèle.

Le critère de maximisation du R^2 (coefficient de détermination) consiste à retenir le modèle dont le R^2 est le plus élevé. Ce critère présente l'inconvénient de ne pas arbitrer entre la perte de degrés de liberté du modèle et l'ajustement qui en résulte. C'est pourquoi on préfère utiliser les critères de Akaike¹ ou de Schwarz² afin de comparer des modèles impliquant un nombre différents de variables explicatives.

Nous retenons le modèle qui minimise la fonction de Akaike (*Akaike Information Criterion*) :

$$AIC = \text{Ln} \left(\frac{SCR}{n} \right) + \frac{2k}{n}$$

ou de Schwarz (*Schwarz Criterion*) :

$$SC = \text{Ln} \left(\frac{SCR}{n} \right) + \frac{k \text{Ln}(n)}{n}$$

1. Akaike H., 1973 ; 1974.

2. Schwarz, 1978.

Avec :

Ln = Logarithme népérien,

SCR = somme des carrés des résidus du modèle,

n = nombre d'observations,

k = nombre de variables explicatives.

Nous allons examiner cinq méthodes qui vont nous permettre de retenir le meilleur modèle, celui qui est composé des variables qui sont :

- les plus corrélées avec la variable à expliquer ;
- les moins corrélées entre elles.

Toutes les régressions possibles

Il s'agit de la méthode la plus simple dans son exposé : nous estimons toutes les combinaisons de régressions possibles ($2^k - 1$ possibilités, k = nombre de variables explicatives candidates) et le modèle retenu est celui dont le critère de Akaike ou de Schwarz est minimum pour un modèle comportant des variables explicatives toutes significatives. La limite d'utilisation de cette méthode est liée au nombre initial de variables explicatives candidates, par exemple si $k = 10$, le nombre de combinaisons possibles est de 1023.

L'élimination progressive (« Backward Elimination »)

Cette procédure consiste, sur le modèle complet à k variables explicatives, à éliminer de proche en proche (c'est-à-dire en réestimant l'équation après chaque élimination) les variables explicatives dont les t de Student sont en dessous du seuil critique. Cette procédure n'est utilisable que si la première équation peut être effectivement estimée, ce qui n'est pas toujours le cas. En effet, lorsque le modèle comporte un nombre important de variables explicatives, le risque de colinéarité entre ces variables est élevé et la matrice $X'X$ peut être singulière.

La sélection progressive (« Forward Regression »)

Dans cette procédure, on sélectionne, dans une première étape, la variable explicative dont le coefficient de corrélation simple est le plus élevé avec la variable y , soit x_i cette variable. La deuxième étape consiste alors à calculer les coefficients de corrélation partielle $r_{yx_j \cdot x_i}^2$ pour $j \neq i$ et à retenir la variable explicative ayant le coefficient le plus élevé. La sélection s'arrête lorsque les t de Student des variables explicatives sont inférieurs au seuil critique.

La régression pas à pas (« Stepwise Regression »)

Cette procédure est identique à la précédente, sauf qu'après avoir incorporé une nouvelle variable explicative, nous examinons les t de Student de chacune des variables explicatives préalablement sélectionnées et nous éliminons du modèle celle(s) dont le t du Student est inférieur au seuil critique.

La régression par étage (« Stagewise Regression »)

C'est un processus de sélection de variables explicatives permettant de minimiser les intercorrélations entre les séries explicatives par étude du résidu.

- 1^{re} étape

La variable explicative dont le coefficient de corrélation simple est le plus élevé avec la variable y est sélectionnée, soit x_i cette variable.

- 2^e étape

Calcul du résidu de la régression de y sur x_i :

$$e_1 = y - \hat{a}_0 - \hat{a}_1 x_i$$

Les coefficients de corrélation simple entre le résidu e_1 et les variables explicatives sont calculés, nous retenons la variable explicative dont le coefficient est le plus élevé. Soit x_j cette variable explicative.

- 3^e étape

Nous calculons un nouveau résidu :

$$e_2 = y - \hat{a}_0 - \hat{a}_1 x_i - \hat{a}_2 x_j$$

Les coefficients de corrélation simple entre le résidu e_2 et les variables explicatives sont calculés, nous retenons la variable explicative dont le coefficient est le plus élevé, ce qui permet de dégager un nouveau résidu.

La procédure est arrêtée lorsque les coefficients de corrélation ne sont plus significativement différents de 0.

Exercice n° 5

 **fichier C4EX4**

Méthodes de sélection de variables explicatives

En utilisant les cinq procédures décrites précédemment et en reprenant les données du tableau 2, on demande de sélectionner, parmi la (ou les) variable(s) explicative(s) candidate(s), celle(s) dont le pouvoir explicatif est le plus important.

Solution

a) Toutes les régressions possibles

Nous devons estimer $(2^k - 1 = 2^4 - 1)$ 15 modèles.

Tableau 3 – Résultats des 15 régressions (les coefficients non significativement différents de 0 sont en caractères gras, italiques)

M1	Séries Coefs. <i>t</i> de Student	X1 0,118 41,9				R^2 0,995	AIC^1 – 2,185	SC – 2,125
M2	Séries Coefs. <i>t</i> de Student	X2 0,327 15,3				0,967	– 0,199	– 0,138
M3	Séries Coefs. <i>t</i> de Student	X3 0,516 12,5				0,952	0,185	0,245
M4	Séries Coefs. <i>t</i> de Student	X4 0,663 18,6				0,978	– 0,583	– 0,523
M5	Séries Coefs. <i>t</i> de Student	X1 0,132 6,91	X2 – 0,039 0,72			0,996	– 2,058	– 1,967
M6	Séries Coefs. <i>t</i> de Student	X1 0,126 8,42	X3 – 0,036 0,53			0,996	– 2,026	– 1,936
M7	Séries Coefs. <i>t</i> de Student	X1 0,102 5,60	X4 0,090 0,86			0,996	– 2,087	– 1,996
M8	Séries Coefs. <i>t</i> de Student	X2 0,209 2,55	X3 0,194 1,48			0,975	– 0,273	– 0,183
M9	Séries Coefs. <i>t</i> de Student	X2 0,137 2,42	X4 0,395 3,45			0,988	– 0,993	– 0,903
M10	Séries Coefs. <i>t</i> de Student	X3 – 0,166 0,73	X4 0,871 3,04			0,979	– 0,457	– 0,366
M11	Séries Coefs. <i>t</i> de Student	X1 0,139 5,58	X2 – 0,038 0,66	X3 – 0,035 0,49		0,996	– 1,898	– 1,778
M12	Séries Coefs. <i>t</i> de Student	X1 0,115 3,55	X2 – 0,028 0,48	X4 0,074 0,64		0,996	– 1,926	– 1,805
M13	Séries Coefs. <i>t</i> de Student	X1 0,104 7,46	X3 – 0,188 2,47	X4 0,319 2,62		0,998	– 2,59	– 2,469
M14	Séries Coefs. <i>t</i> de Student	X2 0,164 3,32	X3 – 0,300 1,99	X4 0,720 3,81		0,993	– 1,304	– 1,183
M15	Séries Coefs. <i>t</i> de Student	X1 0,097 3,66	X2 0,015 0,30	X3 – 0,199 2,20	X4 0,340 2,27	0,998	– 2,408	– 1,804

1. Nous utilisons les formules de la section IV afin de calculer les critères d'information de Akaike et de Schwarz, le logiciel EvIEWS utilise des formules légèrement différentes.

Parmi ces 15 équations, laquelle est la meilleure ?

- Il faut éliminer celles dont un ou plusieurs ratios de Student (sauf en ce qui concerne le terme constant dont nous acceptons qu'il soit non significativement différent de 0) sont inférieurs au seuil critique (environ 2,30) : M5, M6, M7, M8, M10, M11, M12, M14, M15.
- Parmi les 6 équations restantes, on sélectionne celle dont le critère de Akaike et de Schwarz est le plus faible¹, d'où M13 :

$$\text{choix : } x_1 \ x_3 \ x_4$$

À partir de cet exemple, nous prenons la dimension du travail à mener² et les limites, donc, de cette méthode.

b) L'élimination progressive

- Estimation du modèle complet

À partir du modèle estimé avec 4 variables explicatives, nous retirons la variable dont le ratio de Student est le plus faible, soit x_2 .

- Estimation à 3 variables

Le modèle estimé est le modèle [M13], toutes les variables sont significatives ($t^* > 2,30$), la procédure de sélection s'arrête :

$$\text{choix : } x_1 \ x_3 \ x_4$$

c) La sélection progressive

- Calcul des coefficients de corrélation entre y et x_1, x_2, x_3, x_4

$$\rho_{yx1} = 0,997 ; \quad \rho_{yx2} = 0,983 ; \quad \rho_{yx3} = 0,975 ; \quad \rho_{yx4} = 0,988$$

La première variable sélectionnée est x_1 .

- Estimation des 3 modèles à deux variables explicatives : x_1 et x_2 , x_1 et x_3 , x_1 et x_4 .

Modèles : [M5], [M6] et [M7]. Dans ces trois modèles la nouvelle variable explicative n'est jamais significative, la procédure est terminée :

$$\text{choix } x_1$$

d) La régression pas à pas

- Calcul des coefficients de corrélation entre y et x_1, x_2, x_3, x_4 . La première variable sélectionnée est x_1 .
- Estimation des 3 modèles à deux variables explicatives : x_1 et x_2 , x_1 et x_3 , x_1 et x_4 , dans chacun de ces modèles au moins une des variables n'est pas significative, la procédure est donc terminée et nous sommes ramenés au cas précédent :

$$\text{choix } x_1$$

1. Nous constatons qu'il s'agit aussi de l'équation dont le R^2 est le plus élevé.

2. Et pourtant nous n'avons que quatre variables explicatives candidates.

e) *La régression par étage*

- Calcul des coefficients de corrélation entre y et x_1, x_2, x_3, x_4 . La première variable sélectionnée est x_1 .
- Régression et calcul de $e_1 = y + 1,24 - 0,11 x_1$ (modèle [M1])

	e_1
	—
1	-0,123762
2	0,475231
3	-0,127120
4	0,236519
5	-0,424754
6	0,007913
7	-0,419014
8	0,078300
9	0,017027
10	0,279659

- Calcul des coefficients de corrélation entre e_1 et x_1, x_2, x_3, x_4

$$\rho_{e_1 x_1} = 0,000 ; \quad \rho_{e_1 x_2} = -0,040 ; \quad \rho_{e_1 x_3} = -0,039 ; \quad \rho_{e_1 x_4} = 0,048$$

Le coefficient de corrélation entre e_1 et x_1 est par construction nul puisque l'information relative à x_1 n'est plus contenue dans le résidu (ou encore les vecteurs e_1 et x_1 sont orthogonaux).

Les autres coefficients de corrélation ne sont pas significativement différents de 0, la procédure de sélection est donc terminée :

choix x_1

Commentaires :

Seule la procédure de l'élimination progressive ne fournit pas les mêmes résultats de sélection que les quatre autres procédures. Il apparaît au vu de ces résultats que si l'élimination progressive ne maximise pas le R^2 , elle permet, en revanche, de sélectionner une information plus abondante en terme de variables explicatives.

Dans la pratique, la procédure la plus utilisée est la régression pas à pas, c'est cette dernière qui est disponible dans la plupart des logiciels d'économétrie.

5. Problèmes particuliers : la violation des hypothèses

Ce chapitre est consacré aux problèmes particuliers liés au non-respect des hypothèses. Nous nous attachons particulièrement à deux formes classiques :

- en I., l'autocorrélation des erreurs ;
- en II., l'hétéroscédasticité.

L'étude de ces deux phénomènes nous permet de définir un nouvel estimateur : l'estimateur des moindres carrés généralisés, utilisé lorsque la matrice des variances et covariances de l'erreur ne répond plus aux hypothèses classiques, telles que nous les avons supposées réalisées jusqu'à maintenant.

Pour chacun de ces deux cas, nous présentons les méthodes d'investigation : tests de détection, conséquences et procédures d'estimation.

Enfin, la section III. est consacrée au cas où les variables sont observées avec erreur.

I. L'autocorrélation des erreurs

A. Présentation du problème

Jusqu'à maintenant, nous avons considéré que, lors de l'estimation des paramètres du modèle, les hypothèses suivantes sont toutes respectées.

- H1 : le modèle est linéaire en x_t .
- H2 : les valeurs x_t sont observées sans erreur.
- H3 : $E(\varepsilon_t) = 0$.
- H4 : $E(\varepsilon_t^2) = \sigma_\varepsilon^2$.
- H5 : $E(\varepsilon_t \varepsilon_{t'}) = 0$ si $t \neq t'$.
- H6 : $\text{Cov}(x_t, \varepsilon_t) = 0$.

La spécification de la matrice des variances-covariances de l'erreur est :

$$\Omega_\varepsilon = E(\varepsilon \varepsilon') = \begin{bmatrix} E(\varepsilon_1 \varepsilon_1) & E(\varepsilon_1 \varepsilon_2) & \dots & E(\varepsilon_1 \varepsilon_n) \\ E(\varepsilon_2 \varepsilon_1) & E(\varepsilon_2 \varepsilon_2) & \dots & E(\varepsilon_2 \varepsilon_n) \\ \dots & \dots & \dots & \dots \\ E(\varepsilon_n \varepsilon_1) & E(\varepsilon_n \varepsilon_2) & \dots & E(\varepsilon_n \varepsilon_n) \end{bmatrix} = \begin{bmatrix} \sigma_\varepsilon^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_\varepsilon^2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \sigma_\varepsilon^2 \end{bmatrix}$$

Lorsque l'hypothèse H5 n'est plus vérifiée, la matrice $E(\varepsilon \varepsilon') = \Omega_\varepsilon \neq \sigma_\varepsilon^2 I$ n'a plus cette forme particulière (elle n'est plus composée de 0 à l'extérieur de la première diagonale, puisque $\text{Cov}(\varepsilon_t, \varepsilon_{t'}) \neq 0$) et les estimateurs obtenus par la méthode des MCO sont sans biais mais ne sont plus à variance minimale, en effet :

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

C'est-à-dire que \hat{a} est un estimateur dont la première diagonale de la matrice des variances et covariances est supérieure à celle de $\sigma_\varepsilon^2 (X'X)^{-1}$.

Nous sommes donc amenés à nous poser trois questions :

- Comment déterminer un nouvel estimateur pour a ?
- Comment détecter une éventuelle autocorrélation des erreurs ?
- Quelles méthodes d'estimation doit-on utiliser ?

B. L'estimateur des Moindres Carrés Généralisés (MCG)

Considérons le modèle linéaire général

$$\begin{matrix} Y & = & X & a & + & \varepsilon \\ (n, 1) & & (n, k+1) & (k+1, 1) & & (n, 1) \end{matrix}$$

dans lequel $E(\varepsilon \varepsilon') = \Omega_\varepsilon \neq \sigma_\varepsilon^2 I$ (Ω_ε est de dimension n, n).

Nous désirons déterminer un estimateur de a qui ait les mêmes propriétés que l'estimateur des MCO : sans biais, fonction linéaire de Y et à variance minimale. Il est démontré que cet estimateur est donné par :

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

$$\Omega_{\hat{a}} = (X' \Omega_\varepsilon^{-1} X)^{-1} \quad [2]$$

Cet estimateur est appelé estimateur des Moindres Carrés Généralisés (MCG) ou encore estimateur de Aitken.

Remarque : lorsque les hypothèses classiques sont satisfaites, nous retrouvons l'estimateur des MCO :

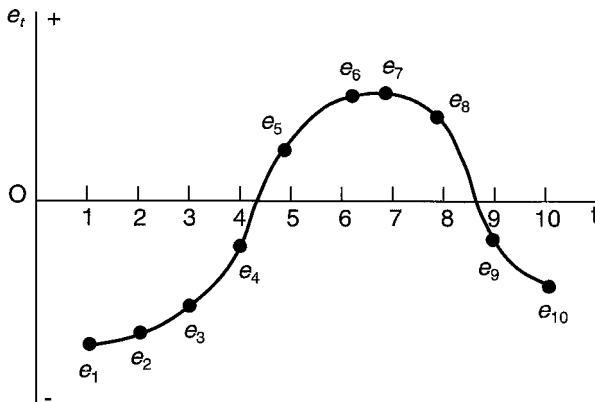
$$\begin{aligned} \hat{a} &= (X' \Omega_\varepsilon^{-1} X)^{-1} (X' \Omega_\varepsilon^{-1} Y) = \\ &= \left(X' \left(\frac{1}{\sigma_\varepsilon^2} I \right) X \right)^{-1} \left(X' \left(\frac{1}{\sigma_\varepsilon^2} I \right) Y \right) = (X' X)^{-1} (X' Y) \end{aligned}$$

Dans la pratique, nous ne connaissons pas la matrice Ω_ε , les formules ci-dessus ne sont pas utilisables, sauf dans des cas exceptionnels. Il convient donc de présenter des procédures d'estimation opérationnelles.

C. Les causes et la détection de l'autocorrélation des erreurs

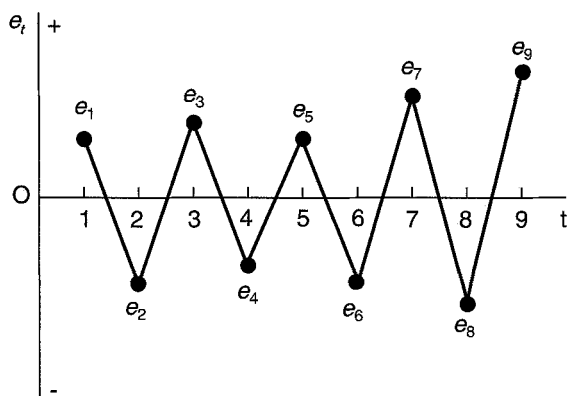
1) Définition et causes

Nous sommes en présence d'une autocorrélation des erreurs lorsque les erreurs sont liées par un processus de reproduction¹. Nous pouvons distinguer l'autocorrélation positive (graphique 1) qui est caractérisée par des successions de résidus de même signe, de l'autocorrélation négative (graphique 2) qui est caractérisée par une alternance positive et négative des résidus.



Graphique 1 – Autocorrélation positive

1. Nous pouvons parler de processus à « mémoire », par comparaison à des processus purement aléatoires.



Graphique 2 – Autocorrélation négative

L'autocorrélation des erreurs peut être observée pour plusieurs raisons :

- *l'absence d'une variable explicative importante* dont l'explication résiduelle permettrait de « blanchir »¹ les erreurs ;
- *une mauvaise spécification du modèle*, les relations entre la variable à expliquer et les variables explicatives ne sont pas linéaires et s'expriment sous une autre forme que celle du modèle estimé (logarithmes, différences premières, etc.) ;
- *un lissage par moyenne mobile ou une interpolation des données* crée une autocorrélation artificielle des erreurs due à l'usage de ces deux opérateurs.

L'autocorrélation des erreurs se rencontre essentiellement dans les modèles en série temporelle où l'influence d'une erreur – due à une mauvaise spécification – d'une période sur l'autre est plausible. Dans le cas de modèle spécifié en coupe instantanée, nous ne pouvons concevoir une autocorrélation des erreurs que si les observations ont été préalablement triées en fonction croissante (ou décroissante) de la variable à expliquer. En effet, la collecte des données s'effectue de manière aléatoire et la succession des valeurs de la variable à expliquer n'a donc aucune raison de générer une autocorrélation des erreurs.

2) Détection

La détection d'une éventuelle dépendance des erreurs ne peut s'effectuer qu'à partir de l'analyse des résidus, en effet eux seuls sont connus.

a) Examen visuel des résidus

L'analyse graphique des résidus permet le plus souvent de détecter un processus de reproduction des erreurs lorsque :

1. Ce terme de « blanchir » vient de la notion de « bruit blanc » (« White Noise »), qui qualifie des processus purement aléatoires.

- les résidus sont pendant plusieurs périodes consécutives soit positifs, soit négatifs : autocorrélation positive (graphique 1) ;
- les résidus sont alternés : autocorrélation négative (graphique 2).

Cependant, le plus souvent, l'analyse graphique est délicate d'interprétation car le dessin des résidus ne présente pas des caractéristiques toujours évidentes.

b) Test de Durbin et Watson

Le test de Durbin et Watson (DW) permet de détecter une autocorrélation des erreurs d'ordre 1 selon la forme :

$$\varepsilon_t = \rho \varepsilon_{t-1} + v_t \text{ avec } v_t \rightarrow N(0, \sigma_v^2)$$

Le test d'hypothèses est le suivant :

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

(ou $\rho < 0$ ou $\rho > 0$, attention le seuil n'est alors pas le même).

Pour tester l'hypothèse nulle H_0 , nous calculons la statistique de Durbin et Watson :

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad [3]$$

où e_t sont les résidus de l'estimation du modèle.

De par sa construction, cette statistique varie entre 0 et 4 et nous avons $DW = 2$ lorsque $\hat{\rho} = 0$ ($\hat{\rho}$ est le ρ observé). Afin de tester l'hypothèse H_0 , Durbin et Watson ont tabulé les valeurs critiques de DW au seuil de 5 % en fonction de la taille de l'échantillon n et du nombre de variables explicatives (k). La lecture de la table permet de déterminer deux valeurs d_1 et d_2 comprises entre 0 et 2 qui délimitent l'espace entre 0 et 4 selon le schéma 1 :

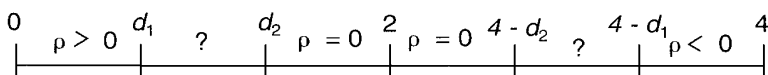


Schéma 1 – Interprétation du test de Durbin et Watson

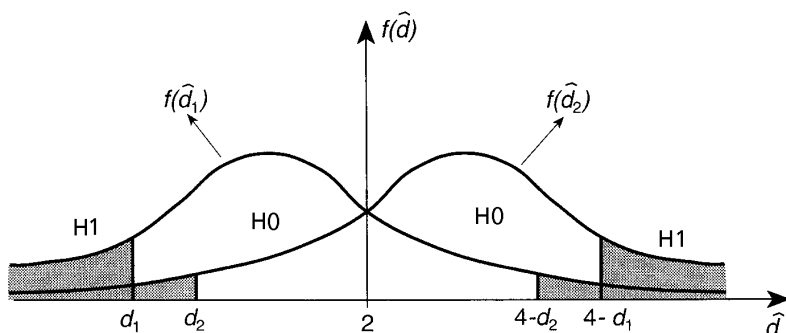
Selon la position du DW empirique dans cet espace, nous pouvons conclure :

- $d_2 < DW < 4 - d_2$, on accepte l'hypothèse $H_0 \rightarrow \rho = 0$;
- $0 < DW < d_1$, on rejette l'hypothèse $H_0 \rightarrow \rho > 0$;

- $4 - d_1 < DW < 4$, on rejette l'hypothèse $H_0 \rightarrow \rho < 0$;
- $d_1 < DW < d_2$ ou $4 - d_2 < DW < 4 - d_1$;

nous sommes dans une zone d'indétermination, ou zone de doute, c'est-à-dire que nous ne pouvons pas conclure¹ dans un sens comme dans l'autre.

Le graphique 3 illustre ces règles de décision.



Graphique 3 – Densités de probabilité de d_1 et d_2

Conditions d'utilisation :

- le modèle doit comporter impérativement un terme constant² ;
- la variable à expliquer ne doit pas figurer parmi les variables explicatives (en tant que variable retardée), il faut alors recourir à la statistique h de Durbin (cf. chapitre 7, I. B.) ou au test de Breusch-Godfrey (cf. ci-dessous) ;
- pour les modèles en coupe instantanée, les observations doivent être ordonnées en fonction des valeurs croissantes ou décroissantes de la variable à expliquer ou d'une variable explicative soupçonnée être la cause de l'autocorrélation ;
- le nombre d'observations doit être supérieur ou égal à 15.

Le test de Durbin et Watson est un test présomptif d'indépendance des erreurs du fait qu'il utilise les résidus ; de plus, il ne teste qu'une autocorrélation d'ordre 1³.

c) Test de Breusch-Godfrey⁴

Ce test, fondé sur un test de Fisher de nullité de coefficients ou de Multiplicateur de Lagrange (« *LM test* », cf. chapitre 3), permet de tester une autocorrélation

1. Dans la pratique, on accepte l'hypothèse la plus désastreuse ! Cependant il serait licite d'accepter l'hypothèse d'indépendance.
2. Les tables de DW sont généralement tabulées pour des modèles comportant un terme constant, cependant il existe des tables pour des modèles sans terme constant.
3. Cela signifie que le test de Durbin et Watson ne détecte pas une autocorrélation d'ordre supérieur à 1, par exemple entre t et $t + 4$ pour une série trimestrielle qui ne serait pas désaisonnalisée.
4. Breusch (1978) et Godfrey (1978).

d'un ordre supérieur à 1 et reste valide en présence de la variable dépendante décalée en tant que variable explicative. L'idée générale de ce test réside dans la recherche d'une relation significative entre le résidu et ce même résidu décalé.

Une autocorrélation des erreurs d'un ordre p s'écrit :

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \dots + \rho_p \varepsilon_{t-p} + v_t$$

Soit le modèle général à erreurs autocorrélées d'ordre p :

$$y_t = a_1 x_{1t} + a_2 x_{2t} + \dots + a_k x_{kt} + a_0 + \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \dots + \rho_p \varepsilon_{t-p} + v_t$$

Ce test est mené en trois étapes :

- Estimation par les MCO du modèle et calcul du résidu e_t , puisque les erreurs sont inconnues, le test porte sur les résidus.
- Estimation par les MCO de l'équation intermédiaire :

$$e_t = a_1 x_{1t} + a_2 x_{2t} + \dots + a_k x_{kt} + a_0 + \rho_1 e_{t-1} + \rho_2 e_{t-2} + \dots + \rho_p e_{t-p} + v_t$$

Soit n le nombre d'observations disponibles (attention chaque décalage entraîne la perte d'une observation) pour estimer les paramètres du modèle et R^2 le coefficient de détermination. Certains auteurs préconisent, afin de ne pas perdre d'observations, de mettre à 0 les premières valeurs du résidu décalé. La différence n'est perceptible que pour des petits échantillons.

- Test d'hypothèses sur l'équation intermédiaire.

L'hypothèse H_0 d'absence d'autocorrélation des erreurs à tester est :

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_p = 0$$

Si on refuse l'hypothèse nulle, alors il existe un risque d'autocorrélation des erreurs à l'ordre p .

Pour mener ce test, nous avons deux possibilités : soit effectuer un test de Fisher classique de nullité des coefficients ρ_i (cf. chapitre 3), soit recourir à la statistique LM qui est distribuée comme un χ^2 à p degrés de liberté ; si $n \times R^2 > \chi^2(p)$ lu dans la table au seuil α , on rejette l'hypothèse d'indépendance des erreurs.

d) Tests de Box-Pierce et Ljung-Box

Ces tests présentés au chapitre 9 (section I. C.) permettent de tester un ensemble de coefficients d'autocorrélation des résidus d'un ordre égal ou supérieur à 1.

Exercice n° 1

 **fichier C5EX1**

Tests d'indépendance des erreurs

Soit le modèle à trois variables explicatives :

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + a_3 x_{3t} + \varepsilon_t$$

Nous disposons des observations annuelles (tableau 1) sur 20 ans :

Tableau 1 – Données annuelles (extrait)

Année	y	x_1	x_2	x_3
1	87,40	98,60	99,10	108,5
2	97,60	101,20	99,10	110,1
...
19	110,70	105,30	93,00	108,5
20	127,10	107,60	106,60	111,3

Nous nous proposons de déceler une éventuelle autocorrélation d'ordre 1 des erreurs, pour cela on demande :

- 1) d'estimer les coefficients du modèle ;
- 2) d'effectuer l'analyse graphique des résidus ;
- 3) de calculer la statistique de Durbin et Watson et d'effectuer le test ;
- 4) enfin, d'effectuer le test de Breusch-Godfrey.

Solution

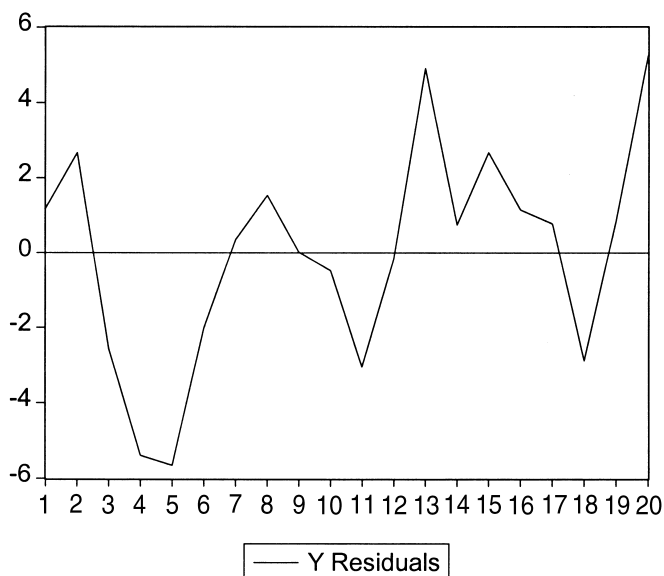
1) Les résultats de l'estimation à partir de Eviews sont les suivants :

Dependent Variable : Y
Included observations : 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-242.7951	26.79995	-9.059538	0.0000
X1	3.897408	0.400325	9.735616	0.0000
X2	0.404365	0.061351	6.590977	0.0000
X3	-0.878886	0.240216	-3.658727	0.0021
R-squared	0.938895	Mean dependent var		97.53500
Adjusted R-squared	0.927438	S.D. dependent var		11.83048
S.E. of regression	3.186826	Akaike info criterion		5.332784
Sum squared resid	162.4937	Schwarz criterion		5.531931
Log likelihood	-49.32784	F-statistic		81.94784
Durbin-Watson stat	1.053794	Prob(F-statistic)		0.000000

Les trois coefficients des variables explicatives sont significativement différents de 0 ($t_{16}^{0,05} = 2,12$), le Fisher empirique est supérieur au Fisher lu ($F_{3,16}^{0,05} = 3,24$), le modèle semble satisfaisant.

2) Eviews fournit directement le graphique des résidus.



L'analyse de ce graphique révèle des résidus qui semblent cycliques, ceci est symptomatique d'une autocorrélation positive des résidus.

3) Les conditions d'utilisation du test de Durbin et Watson sont bien respectées : le modèle est spécifié en série temporelle, le nombre d'observations ($n = 20$) est supérieur à 15 et, enfin, le modèle estimé comporte un terme constant.

Le calcul de la statistique à partir des résidus est alors :

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} = \frac{171,23}{162,49} = 1,053$$

Cette valeur est à comparer à celles lues dans la table de Durbin et Watson à $n = 20$ et $k = 3$, soit $d_1 = 1,00$ et $d_2 = 1,68$. La valeur DW se situe dans la zone de doute ($d_1 < DW < d_2 \rightarrow$ incertitude), cependant à proximité immédiate de la zone de rejet de H_0 , nous pouvons plutôt conclure à une autocorrélation positive des résidus, donc à une présomption de dépendance des erreurs.

4) Les données étant annuelles, nous testons une autocorrélation d'ordre 2, un ordre supérieur à deux ans ne semble pas justifié. Nous estimons donc le modèle :

$$e_t = a_1 x_{1t} + a_2 x_{2t} + a_3 x_{3t} + a_0 + \rho_1 e_{t-1} + \rho_2 e_{t-2} + v_t$$

$$\text{Soit : } \hat{e}_t = -17,26 + 0,5287x_{1t} - 0,0369x_{2t} - 0,3149x_{3t} + 0,589 e_{t-1} - 0,497e_{t-2}$$

avec : $R^2 = 0,33$ et $n = 18$ (car nous avons perdu deux observations du fait du décalage).

$LM = nR^2 = 18 \times 0,33 = 5,94 < \chi_{0,05}^2(2) = 5,99$, nous sommes à la limite d'acceptation de l'hypothèse H_0 pour un seuil de 5 %.

Nous testons alors un ordre 1, soit l'estimation suivante :

$$\hat{e}_t = -11,82 + 0,087 x_{1t} - 0,0158 x_{2t} - 0,044 x_{3t} + 0,507 e_{t-1}$$

avec : $R^2 = 0,196$ et $n = 19$, d'où $LM = 19 \times 0,196 = 3,72 < \chi_{0,05}^2(1) = 3,84$.

Nous sommes, là encore, à la limite d'acceptation (probabilité critique $\alpha^c = 5,37$ %) de l'hypothèse H_0 , compte tenu de toutes ces informations (test de Durbin-Watson, test de Breusch-Godfrey), nous sommes enclins à accepter l'hypothèse d'autocorrélation des erreurs.

Les estimateurs que nous avons trouvés sont sans biais mais non à variance minimale, il convient donc de déterminer une procédure adéquate d'estimation.

Il est à noter que tous les commentaires de la question 1 concernant la validité statistique du modèle sont maintenant caducs.

D. Les procédures d'estimation en cas d'autocorrélation des erreurs

1) Principes généraux

Si nous retenons l'hypothèse d'une autocorrélation des erreurs d'ordre 1, le modèle linéaire s'écrit :

$$Y = Xa + \varepsilon$$

avec

$$\varepsilon_t = \rho \varepsilon_{t-1} + v_t \quad |\rho| < 1 \quad [4]$$

(processus autorégressif d'ordre 1 : $AR(1)$)¹ où $v_t \rightarrow N(0; \sigma_v^2)$ et $E(v_t v_{t'}) = 0$ pour $t \neq t'$.

En procédant par substitution successive de [4], nous obtenons :

$$\begin{aligned} \varepsilon_t &= \rho(\rho \varepsilon_{t-2} + v_{t-1}) + v_t = \rho^2 \varepsilon_{t-2} + (\rho v_{t-1} + v_t) \\ \varepsilon_t &= v_t + \rho v_{t-1} + \rho^2 v_{t-2} + \rho^3 v_{t-3} + \dots \end{aligned} \quad [5]$$

Ce processus tend vers 0 puisque $|\rho| < 1$.

Nous allons étudier les propriétés de ε_t .

1. On note les processus autorégressifs par : $AR(p)$, où p est l'ordre du processus autorégressif. Le chapitre 9 est consacré à l'étude générale de ce type de processus.

$$E(\varepsilon_t) = E(\rho\varepsilon_{t-1} + v_t) = \rho E(\varepsilon_{t-1}) + E(v_t) = 0$$

$$E(\varepsilon_t^2) = E(\rho\varepsilon_{t-1} + v_t)^2 = \rho^2 E(\varepsilon_{t-1}^2) + E(v_t^2) \text{ car } v_t \text{ est ind pendant de } \varepsilon_t.$$

La variance de ε_t  tant homosc dastique, nous avons : $E(\varepsilon_t^2) = E(\varepsilon_{t-1}^2) = \sigma_\varepsilon^2$.

$$\text{Soit : } \sigma_\varepsilon^2(1 - \rho^2) = \sigma_v^2 \longrightarrow \sigma_\varepsilon^2 = \frac{\sigma_v^2}{(1 - \rho^2)} \quad [6]$$

$$\text{Cov}(\varepsilon_t, \varepsilon_{t+1}) = E(\varepsilon_t, \varepsilon_{t+1}) = E(\varepsilon_t, (\rho\varepsilon_t + v_{t+1})) = \rho\sigma_\varepsilon^2$$

$$\text{Cov}(\varepsilon_t, \varepsilon_{t+2}) = E(\varepsilon_t, \varepsilon_{t+2}) = E(\varepsilon_t, (\rho\varepsilon_{t+1} + v_{t+2})) = \rho E(\varepsilon_t, \varepsilon_{t+1}) = \rho^2\sigma_\varepsilon^2$$

...

$$\text{Cov}(\varepsilon_t, \varepsilon_{t+i}) = E(\varepsilon_t, \varepsilon_{t+i}) = \rho^i\sigma_\varepsilon^2 = \frac{\rho^i\sigma_v^2}{(1 - \rho^2)}$$

La matrice des variances-covariances de l'erreur dans ce cas est alors :

$$\Omega_\varepsilon = E(\varepsilon\varepsilon') = \frac{\sigma_v^2}{1 - \rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix} \text{ avec } \rho \neq 1$$

L'estimateur des MCG est  gal d'apr s [1] :

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

o 

$$\Omega_\varepsilon^{-1} = \frac{1}{\sigma_v^2} \begin{bmatrix} 1 & -\rho & 0 & 0 & \dots & \dots & 0 \\ -\rho & 1 + \rho^2 & -\rho & 0 & \dots & \dots & 0 \\ 0 & -\rho & 1 + \rho^2 & -\rho & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & 1 + \rho^2 & -\rho & \dots \\ 0 & \dots & \dots & \dots & 0 & -\rho & 1 \end{bmatrix}$$

Cependant, nous ne connaissons ni le terme ρ , ni la variance σ_v^2 ; nous allons donc chercher une transformation matricielle M telle que le mod le $MY = MXa + M\varepsilon$ ait ses erreurs ind pendantes et homosc dastiques.

Soit :

$$E(M\varepsilon(M\varepsilon)') = E(M\varepsilon\varepsilon' M') = ME(\varepsilon\varepsilon') M' = M\Omega_\varepsilon M' = \sigma_v^2 I$$

Dans ce cas, on peut d terminer l'estimateur BLUE de a par la m thode des MCO :

$$\hat{a} = ((MX)' MX)^{-1} (MX)' MY = (X' M' MX)^{-1} X' M' MY \quad [7]$$

En identifiant terme à terme [1] et [7] on obtient :

$M' M = \lambda \Omega_\varepsilon^{-1} = \sigma_v^2 \Omega_\varepsilon^{-1}$ puisque si l'égalité [1] = [7] est vérifiée pour $M' M = \Omega_\varepsilon^{-1}$, elle l'est aussi pour $\lambda \Omega_\varepsilon^{-1}$ quel que soit λ .

La matrice

$$M_{(n-1, n)} = \begin{bmatrix} -\rho & 1 & 0 & 0 & \dots & 0 \\ 0 & -\rho & 1 & 0 & \dots & 0 \\ 0 & 0 & -\rho & 1 & & 0 \\ \dots & & & & & \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{bmatrix}$$

remplit cette condition puisque :

$$M' M = \begin{bmatrix} \rho^2 & -\rho & 0 & 0 & \dots & 0 \\ -\rho & 1 + \rho^2 & -\rho & 0 & \dots & 0 \\ 0 & -\rho & 1 + \rho^2 & -\rho & \dots & 0 \\ \dots & & & & & \\ 0 & \dots & & \dots & 1 + \rho^2 & -\rho \\ 0 & \dots & & 0 & -\rho & 1 \end{bmatrix}$$

et $\sigma_v^2 \Omega_\varepsilon^{-1}$ sont identiques au premier terme près ρ^2 au lieu de 1.

Ainsi, nous pouvons substituer à la méthode des MCG, la méthode des MCO (si n est suffisamment grand) appliquée au modèle linéaire $MY = MXa + M\varepsilon$ qui n'est autre que le modèle initial dans lequel les variables sont transformées par un passage aux quasi-différences premières¹. En effet :

$$MY = \begin{bmatrix} y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ \dots \\ y_n - \rho y_{n-1} \end{bmatrix} \quad \text{et} \quad MX_k = \begin{bmatrix} x_{k2} - \rho x_{k1} \\ x_{k3} - \rho x_{k2} \\ \dots \\ x_{kn} - \rho x_{k,n-1} \end{bmatrix}$$

Cas particulier : lorsqu'il y a autocorrélation parfaite ($\rho = 1$), nous sommes amenés à appliquer la méthode des MCO sur le modèle en différences premières.

Nous allons montrer à partir d'un exemple le bien-fondé de la transformation en quasi-différences.

Soit le modèle à deux variables explicatives :

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \varepsilon_t \quad [8]$$

avec $\varepsilon_t = \rho \varepsilon_{t-1} + v_t$ avec $|\rho| < 1$ ($v_t \rightarrow N(0; \sigma_v^2)$ et $E(v_t v_{t'}) = 0$ pour $t \neq t'$).

Le modèle [8] peut s'écrire en $t - 1$:

$$y_{t-1} = a_0 + a_1 x_{1,t-1} + a_2 x_{2,t-1} + \varepsilon_{t-1} \quad [9]$$

Le calcul de [8] - ρ [9] nous donne :

$$\begin{aligned} y_t - \rho y_{t-1} &= a_0(1 - \rho) + a_1(x_{1t} - \rho x_{1,t-1}) + a_2(x_{2t} - \rho x_{2,t-1}) + (\varepsilon_t - \rho \varepsilon_{t-1}) \\ dy_t &= b_0 + a_1 dx_{1t} + a_2 dx_{2t} + v_t \end{aligned}$$

1. Cette transformation entraîne la perte d'une observation, toutefois on peut remplacer la première observation par $\sqrt{1 - \rho^2} y_1$ et $\sqrt{1 - \rho^2} x_{k1}$.

Le terme aléatoire v_t répond aux hypothèses d'application de la méthode des MCO, nous pouvons donc utiliser les procédures d'estimation des MCO sur les variables transformées. Les estimations des coefficients calculées à partir du modèle en différence s'interprètent directement comme étant les coefficients du modèle initial, sauf en ce qui concerne le terme constant :

$$a_0 = b_0 / (1 - \rho)$$

Si nous connaissons la technique d'estimation permettant de lever l'autocorrélation d'ordre 1, nous pouvons l'appliquer de manière opérationnelle, il convient d'estimer le paramètre ρ . C'est l'objet des procédures d'estimations suivantes.

2) Procédures d'estimation de ρ

Les procédures suivantes ne sont valides que si : $\varepsilon_t = \rho \varepsilon_{t-1} + v_t$.

a) Estimation directe de ρ à partir des résidus de la régression sur le modèle initial

Étape 1 : *estimation de ρ de deux manières.*

- par régression directe de e_t sur e_{t-1}

$$\hat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_t e_t^2} \quad (\bar{e} = 0) \quad [10]$$

- ou à partir de la statistique de Durbin et Watson

$$\hat{\rho} \simeq 1 - \frac{DW}{2} \quad [11]$$

En effet, le numérateur de la statistique de Durbin et Watson est égal à :

$$\sum_{t=2}^n (e_t - e_{t-1})^2 = \sum_{t=2}^n e_t^2 + \sum_{t=2}^n e_{t-1}^2 - 2 \sum_{t=2}^n e_t e_{t-1}$$

En divisant par le dénominateur de cette même statistique et sachant que

$$\rho_{e_t e_{t-1}} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2}, \text{ nous avons } DW = 2 - 2\rho_{e_t e_{t-1}} - \frac{e_1^2 + e_n^2}{\sum_{t=1}^n e_t^2},$$

si n est grand, $\frac{e_1^2 + e_n^2}{\sum_{t=1}^n e_t^2}$ tend vers 0.

Étape 2 : transformation des variables et régressions¹ sur les quasi-différences :

$$y_t - \hat{\rho}y_{t-1} = b_0 + a_1(x_{1t} - \hat{\rho}x_{1t-1}) + \dots + a_k(x_{kt} - \hat{\rho}x_{kt-1}) + v_t$$

Les paramètres estimés par MCO sont alors : $\hat{a}_1, \dots, \hat{a}_k$ et $\hat{a}_0 = \hat{b}_0/(1 - \hat{\rho})$.

b) Estimation itérative du vecteur a et de ρ (méthode de Cochrane-Orcutt)

Étape 1 : initialisation de ρ .

Par une technique d'estimation directe par régression [10], on fixe une première valeur à ρ soit :

$$\hat{\rho}_0 = \hat{\rho}$$

Étape 2 : régression sur les quasi-différences :

$$y_t - \hat{\rho}_0 y_{t-1} = b_0 + a_1(x_{1t} - \hat{\rho}_0 x_{1t-1}) + \dots + a_k(x_{kt} - \hat{\rho}_0 x_{kt-1}) + v_t$$

Les paramètres estimés sont alors : $\hat{a}_1, \dots, \hat{a}_k$ et $\hat{a}_0 = \hat{b}_0/(1 - \hat{\rho}_0)$.

Étape 3 : réestimation de ρ .

À partir des nouveaux résidus d'estimation ($e_t^{(1)}$) nous recalculons une nouvelle valeur de ρ soit $\hat{\rho}_1$:

$$e_t^{(1)} = y_t - \hat{a}_0 - \hat{a}_1 x_{1t} - \dots - \hat{a}_k x_{kt}$$

$$\hat{\rho}_1 = \frac{\sum_{t=2}^n e_t^{(1)} e_{t-1}^{(1)}}{\sum_t (e_t^{(1)})^2}$$

Étape 4 : régression sur les quasi-différences :

$$y_t - \hat{\rho}_1 y_{t-1} = b_0 + a_1(x_{1t} - \hat{\rho}_1 x_{1t-1}) + \dots + a_k(x_{kt} - \hat{\rho}_1 x_{kt-1}) + v_t$$

Puis nous calculons un nouveau résidu ($e_t^{(2)}$) à partir de la nouvelle estimation des coefficients, ce qui nous permet d'obtenir un $\hat{\rho}_2$.

Et ainsi de suite jusqu'à la stabilité des coefficients a_k estimés (en général 3 ou 4 itérations).

c) Méthode du « balayage » (« Hildreth-Lu »)

Étape 1 : détermination du type d'autocorrélation.

À partir de la statistique de Durbin et Watson, on détermine une autocorrélation positive ou négative ($\rho < 0$ ou $\rho > 0$).

Étape 2 : régression pour l'intervalle des valeurs possibles de ρ .

Par exemple, on sait que $\rho \in [0; 1]$, nous régressons pour toutes les valeurs successives de $\rho = \{0,1; 0,2; \dots 0,9; 1\}$ sur l'intervalle $[0; 1]$ avec un pas fixé ici égal à 0,1, l'équation :

$$y_t - \hat{\rho}_i y_{t-1} = b_0 + a_1(x_{1t} - \hat{\rho}_i x_{1t-1}) + \dots + a_k(x_{kt} - \hat{\rho}_i x_{kt-1}) + v_{it}$$

1. Nous perdrons une observation lors de la transformation en différences.

et nous retenons la valeur de ρ qui minimise la somme des carrés des résidus $\left(\sum_t \widehat{v}_{it}^2\right)$.

Il est possible d'affiner la valeur estimée de ρ en réemployant la même procédure sur un intervalle restreint et avec un pas plus fin (par exemple 0,01).

Il est à noter que cette technique est optimale selon le critère des moindres carrés puisque l'on retient le ρ qui minimise la somme des carrés des résidus.

Exercice n° 2

 **fichier C5EX1**

Procédures d'estimation en cas d'autocorrélation des erreurs d'ordre 1

En reprenant le modèle de l'exercice 1 (et les données du tableau 1) dans lequel nous avons détecté une autocorrélation des erreurs, on demande d'en corriger les effets à l'aide des quatre procédures présentées. On comparera les résultats.

Solution

La solution de cet exercice est réalisée à l'aide du programme Eviews  C5EX2.PRG

a) Estimation directe de ρ à partir des résidus de la régression sur le modèle initial

Étape 1 : régression, calcul des résidus et estimation de ρ .

Les résultats de la régression ainsi que le calcul des résidus ont déjà été présentés lors de la résolution de l'exercice 1.

• **Première possibilité :** par régression de e_t sur e_{t-1}

Résultats

Dependent Variable : e				
Included observations : 19 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.065400	0.643073	0.101700	0.9202
$e(-1)$	0.464845	0.241301	1.926413	0.0709

$$\widehat{\rho}_1 = 0,46$$

Étape 2 : transformations des variables et régression sur les quasi-différences

• Calcul des quasi-différences :

$$\begin{aligned} \text{genr dy} &= y - \text{rau1} * y(-1) \\ \text{genr dx1} &= x1 - \text{rau1} * x1(-1) \\ \text{genr dx2} &= x2 - \text{rau1} * x2(-1) \\ \text{genr dx3} &= x3 - \text{rau1} * x3(-1) \end{aligned}$$

- Régression

Dependent Variable : DY				
Included observations : 19 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	- 135.5132	18.38346	- 7.371473	0.0000
DX1	3.593675	0.583072	6.163342	0.0000
DX2	0.439019	0.071365	6.151765	0.0000
DX3	- 0.514596	0.337097	- 1.526550	0.1477

Calcul du terme constant :

$$\widehat{a}_0 = \widehat{b}_0 / (1 - \widehat{\rho}_1) = -135,51 / (1 - 0,46) = -253,22$$

- *Deuxième possibilité* : estimation de ρ à partir de la statistique de Durbin et Watson

$$\widehat{\rho}_2 = 1 - \frac{DW}{2} = 1 - \frac{1,053}{2} = 0,473$$

Étape 2 : transformations des variables et régression sur les quasi-différences.

Le calcul est identique au précédent avec la nouvelle valeur de ρ .

Dependent Variable : DY				
Included observations : 19 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	- 133.4095	18.14245	- 7.353445	0.0000
DX1	3.580221	0.583673	6.133954	0.0000
DX2	0.440490	0.071400	6.169361	0.0000
DX3	- 0.503033	0.336991	- 1.492721	0.1562

Calcul du terme constant :

$$\widehat{a}_0 = \widehat{b}_0 / (1 - \widehat{\rho}_2) = -133,40 / (1 - 0,47) = -253,20$$

- b)** Estimation itérative du vecteur a et de ρ (méthode de Cochrane-Orcutt)

• Estimation d'un modèle autorégressif d'ordre 1

LS Y C X1 X2 X3 AR(1)

Dependent Variable : Y				
Included observations : 19 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	- 250.3442	38.62520	- 6.481369	0.0000
X1	3.142385	0.573681	5.477582	0.0001
X2	0.497558	0.069854	7.122783	0.0000
X3	- 0.144028	0.306106	- 0.470517	0.6452
AR(1)	0.811373	0.336991	5.048921	0.0002
Durbin-Watson stat = 2.149				

$$\hat{\rho}_3 = 0,81$$

Il est à noter que le logiciel calcule directement la valeur de ρ (assortie de son écart type) ainsi que l'estimation du terme constant du modèle initial.

c) Méthode du « balayage » (« Hildreth-Lu »)

Les résultats sont identiques à ceux de la méthode de Cochrane-Orcutt.

Tableau 2 – Récapitulation des résultats

MÉTHODES				
Paramètres	<i>a</i> Régression	<i>a</i> DW	<i>b</i> C.O.	<i>c</i> H.L.
$\hat{\rho}$	0,464	0,473	0,81	0,81
(<i>t</i>)	(1,92)		(5,04)	(5,04)
\hat{a}_0	- 253,22	- 253,20	- 250,3	- 250,3
(<i>t</i>)			(6,49)	(6,49)
\hat{a}_1	3,59	3,58	3,14	3,14
(<i>t</i>)	(6,16)	(6,13)	(5,48)	(5,48)
\hat{a}_2	0,44	0,44	0,49	0,49
(<i>t</i>)	(6,15)	(6,16)	(7,12)	(7,12)
\hat{a}_3	- 0,51	- 0,50	- 0,14	- 0,14
(<i>t</i>)	(1,52)	(1,49)	(0,47)	(0,47)
R^2	0,91	0,91	0,96	0,96

Les coefficients de régression en italique ne sont pas significativement différents de 0 (pour un seuil de 5 %).

À la lecture des résultats (tableau 2), nous constatons que :

- le coefficient de régression a_3 n'est jamais significativement différent de 0, alors que, dans le modèle initial, à erreurs autocorrélées, tous les coefficients sont significatifs ;

- les deux méthodes d'estimation directe des résidus aboutissent à des résultats quasiment identiques, il subsiste néanmoins une autocorrélation des erreurs ;
- les deux méthodes itératives (C.O. et H.L.) donnent des résultats identiques, mais assez différents des deux méthodes précédentes.

Laquelle choisir ? Cette question est sans réponse définitive, la plus répandue est certainement la méthode de Cochrane-Orcutt. Les deux procédures d'estimation directe sont plus un artifice de calcul qu'une méthode proprement dite ; la technique du « balayage » aboutit à des résultats similaires à celle de Cochrane-Orcutt.

Nous sommes dans le domaine de l'estimation statistique donc avec des marges d'erreurs qui peuvent parfois dépasser le raffinement de telle ou telle méthode, c'est pourquoi nous pouvons recommander d'utiliser la technique la plus simple à mettre en œuvre donc celle directement utilisée par le logiciel habituel.

Il convient de plus de noter que ces méthodes particulières d'estimation constituent une réponse technique à un problème le plus souvent de spécifications économiques. Nous recommandons, donc, plutôt que d'appliquer immédiatement ces procédures, de réfléchir à une spécification plus adaptée du modèle.

II. L'hétéroscédasticité

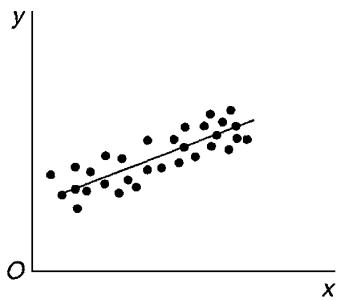
A. Présentation du problème

Soit le modèle $Y = Xa + \varepsilon$ pour lequel l'hypothèse H4 n'est pas vérifiée, la matrice des erreurs est alors :

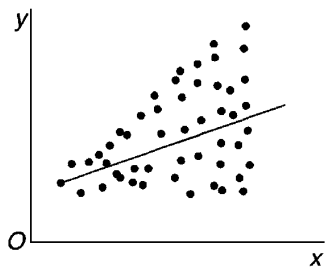
$$\Omega_\varepsilon = E(\varepsilon\varepsilon') = \begin{bmatrix} E(\varepsilon_1 \varepsilon_1) & E(\varepsilon_1 \varepsilon_2) & \dots & E(\varepsilon_1 \varepsilon_n) \\ E(\varepsilon_2 \varepsilon_1) & E(\varepsilon_2 \varepsilon_2) & \dots & E(\varepsilon_2 \varepsilon_n) \\ \dots & \dots & \dots & \dots \\ E(\varepsilon_n \varepsilon_1) & E(\varepsilon_n \varepsilon_2) & \dots & E(\varepsilon_n \varepsilon_n) \end{bmatrix} = \begin{bmatrix} \sigma_{\varepsilon_1}^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_{\varepsilon_2}^2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \sigma_{\varepsilon_n}^2 \end{bmatrix}$$

Les variances des erreurs ne sont plus constantes sur la première diagonale. Ce problème se rencontre plus fréquemment pour les modèles spécifiés en

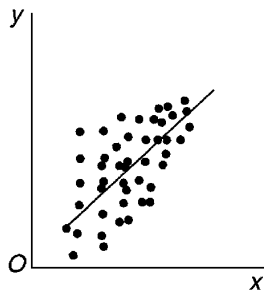
coupe instantanée ou bien lorsque les observations sont représentatives de moyennes. La variance de l'erreur est alors liée aux valeurs de la variable explicative comme l'illustrent les graphiques 4 à 6.



Graphique 4 – Modèle homoscédastique



Graphique 5 – La variance de l'erreur croît avec la variable explicative



Graphique 6 – La variance de l'erreur décroît avec la variable explicative

Les conséquences de l'hétéroscédasticité sont identiques à celles de l'auto-corrélation des erreurs, c'est-à-dire :

- estimateur sans biais ;
- l'estimateur de MCO n'est plus à variance minimale.

Les causes de l'hétéroscédasticité sont multiples :

- comme nous l'avons déjà cité, lorsque les observations représentent des moyennes calculées sur des échantillons de taille différente ;
- un autre cas résulte de la répétition d'une même valeur de la variable à expliquer pour des valeurs différentes d'une variable explicative, par exemple lors de regroupements en tranches (de salaires, d'effectifs...) ;
- lorsque les erreurs sont liées aux valeurs prises par une variable explicative, dans un modèle en coupe instantanée la variance de la consommation croît, par exemple, avec le revenu disponible, etc.

Dans le cas d'une hétéroscédasticité révélée, le test de Chow (cf. chapitre 3, section B.) est biaisé, nous risquons alors de rejeter à tort¹ l'hypothèse H_0 de la stabilité des coefficients du modèle. Si, lors d'un test de Chow, nous sommes amenés à accepter l'hypothèse H_0 , le problème ne se pose donc pas ; en revanche, dans le cas du rejet de H_0 , il convient préalablement de corriger l'hétéroscédasticité et d'effectuer le test de Chow sur le modèle transformé.

B. Correction de l'hétéroscédasticité

L'estimateur BLUE du modèle hétéroscédastique est alors celui des MCG :

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

avec

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

Il n'existe pas une méthodologie unique de correction, à la différence de la correction de l'autocorrélation d'ordre 1 des erreurs, mais des méthodes que l'on applique en fonction de la cause présumée de l'hétéroscédasticité. La règle générale consiste à déterminer une transformation² concernant les données – de la variable à expliquer et des variables explicatives – afin de se ramener à un modèle à variances constantes (homoscédastique). Les exercices 3 et 4 illustrent les principales méthodes.

Exercice n° 3

↓ fichier C5EX3

Correction de l'hétéroscédasticité lorsque les données représentent des moyennes

Afin de tester l'efficacité d'un nouveau fortifiant sur la croissance des poulets, un éleveur a retenu 40 poulets répartis en cinq groupes de tailles inégales et leur a administré des doses différentes de ce fortifiant. Les résultats sont consignés sur le tableau 3.

1. C'est-à-dire trop souvent.

2. Cette transformation consiste à déterminer un vecteur de pondération pour chaque observation, d'où le nom de régression pondérée.

Tableau 3 – Relation entre le poids des poulets
et la dose de fortifiant

Groupe i	Nombre de poulets n_i	Poids moyen des poulets \bar{Y}_i	Dose moyenne administrée \bar{X}_i
1	12	1,7	5,8
2	8	1,9	6,4
3	6	1,2	4,8
4	9	2,3	6,9
5	5	1,8	6,2

On demande d'estimer les coefficients de la relation :

$$Y_i = a_0 + a_1 X_i + \varepsilon_i$$

relation entre le poids de chacun des poulets Y_i et la dose individuelle reçue X_i et de tester si le coefficient a_1 est significativement différent de 0.

Solution

• Estimation des coefficients

Nous ne connaissons que la relation entre les moyennes, soit : $\bar{Y}_i = a_0 + a_1 \bar{X}_i + \bar{\varepsilon}_i$. Les variables représentent des moyennes, donc la variance de l'erreur $\bar{\varepsilon}_i$ est égale à : $\text{Var}(\bar{\varepsilon}_i) = \frac{\sigma_\varepsilon^2}{n_i}$.

$$\begin{aligned} \text{En effet, } \text{Var}(\bar{\varepsilon}_i) &= \text{Var}\left(\frac{\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_n}{n_i}\right) = \frac{1}{n_i^2}(\text{Var}(\varepsilon_1) + \text{Var}(\varepsilon_2) + \dots + \text{Var}(\varepsilon_n)) \\ &= \frac{1}{n_i^2}(n_i \times \sigma_\varepsilon^2) = \frac{\sigma_\varepsilon^2}{n_i} \text{ (les covariances } \text{Cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j \text{).} \end{aligned}$$

La matrice des variances et covariances des erreurs est donc la forme :

$$\text{Var}(\bar{\varepsilon}_i) = \Omega_\varepsilon = \sigma_\varepsilon^2 V = \sigma_\varepsilon^2 \begin{bmatrix} \frac{1}{n_1} & & & & \\ & \frac{1}{n_2} & & & \\ & & \frac{1}{n_3} & & \\ & & & \frac{1}{n_4} & \\ & & & & \frac{1}{n_5} \end{bmatrix}$$

Puisque le modèle est hétéroscédastique, nous allons utiliser l'estimateur des MCG :

$$\begin{aligned} \hat{a} &= (X' \Omega_\varepsilon^{-1} X)^{-1} (X' \Omega_\varepsilon^{-1} Y) = (X' V^{-1} X)^{-1} (X' V^{-1} Y) \\ V^{-1} &= \begin{bmatrix} n_1 & & & & \\ & n_2 & & & \\ & & n_3 & & \\ & & & n_4 & \\ & & & & n_5 \end{bmatrix} \quad \text{d'où } (X' V^{-1} X) = \begin{bmatrix} \sum n_i & \sum n_i \bar{X}_i \\ \sum n_i \bar{X}_i & \sum n_i \bar{X}_i^2 \end{bmatrix} \end{aligned}$$

et

$$(X' V^{-1} Y) = \left[\begin{array}{c} \sum n_i \bar{Y}_i \\ \sum n_i \bar{X}_i \bar{Y}_i \end{array} \right]$$

Le tableau de calcul 4 permet d'obtenir les résultats dont nous avons besoin.

Tableau 4 – Exemple de calcul en cas d'hétéroscédasticité

\bar{Y}_i	\bar{X}_i	n_i	$n_i \bar{Y}_i$	$n_i \bar{X}_i$	$n_i \bar{X}_i^2$	$n_i \bar{Y}_i^2$	$n_i \bar{X}_i \bar{Y}_i$
1,7	5,8	12	20,4	69,6	403,68	34,68	118,32
1,9	6,4	8	15,2	51,2	327,68	28,88	97,28
1,2	4,8	6	7,2	28,8	138,24	8,64	34,56
2,3	6,9	9	20,7	62,1	428,49	47,61	142,83
1,8	6,2	5	9,0	31,0	192,20	16,20	55,8
Total		40	72,5	242,7	1 490,29	136,01	448,79

$$(X' V^{-1} X)^{-1} = \left[\begin{array}{cc} 40 & 242,7 \\ 242,7 & 1\,490,29 \end{array} \right]^{-1} = \left[\begin{array}{cc} 2,1040 & -0,3426 \\ -0,3426 & 0,05647 \end{array} \right]$$

et

$$(X' V^{-1} Y) = \left[\begin{array}{c} 72,5 \\ 448,79 \end{array} \right] \text{ soit } \hat{a} = \left[\begin{array}{c} -1,23 \\ 0,502 \end{array} \right]$$

Nous pouvons remarquer que la matrice M telle que $V^{-1} = M' M$ est égale à :

$$M = \left[\begin{array}{ccccc} \sqrt{n_1} & & & & \\ & \sqrt{n_2} & & & \\ & & \sqrt{n_3} & & \\ & & & \sqrt{n_4} & \\ & & & & \sqrt{n_5} \end{array} \right]$$

L'estimateur des MCG est alors égal à :

$$\hat{a} = [(MX)'(MX)]^{-1}[(MX)'(MY)] = (X' M' MX)^{-1}(X' M' MY)$$

Nous aurions donc trouvé les mêmes résultats en employant la technique de la régression pondérée, c'est-à-dire par multiplication du vecteur des observations de la variable à expliquer (Y) et du vecteur de la variable explicative (X) par la matrice M (dont la forme générale est $\sqrt{n_i}$). La méthode des MCO appliquée aux données transformées (y compris le terme constant) suivantes conduit donc à des résultats identiques :

$$\begin{array}{lllll} Y_i : & 1,7 \times \sqrt{12} & 1,9 \times \sqrt{8} & 1,2 \times \sqrt{6} & 2,3 \times \sqrt{9} & 1,8 \times \sqrt{5} \\ X_i : & 5,8 \times \sqrt{12} & 6,4 \times \sqrt{8} & 4,8 \times \sqrt{6} & 6,9 \times \sqrt{9} & 6,2 \times \sqrt{5} \\ U_i : & 1 \times \sqrt{12} & 1 \times \sqrt{8} & 1 \times \sqrt{6} & 1 \times \sqrt{9} & 1 \times \sqrt{5} \end{array}$$

- Nous remarquons que : $\text{Var}(\sqrt{n_i} \varepsilon_i) = \sigma_\varepsilon^2$, l'hétéroscédasticité est donc levée.
- Estimation de la variance de l'erreur

Un estimateur sans biais de σ_ε^2 peut être déduit de l'application des MCO à la relation précédente :

$$\begin{aligned} \hat{\sigma}_\varepsilon^2 &= \frac{e'e}{n-k-1} = \frac{(MY - MX\hat{a})'(MY - MX\hat{a})}{n-k-1} \\ &= \frac{(Y - X\hat{a})'M'M(Y - X\hat{a})}{n-k-1} \end{aligned}$$

$$\hat{\sigma}_\varepsilon^2 = \frac{(Y - X\hat{a})'V^{-1}(Y - X\hat{a})}{n - k - 1}$$

Nous connaissons :

$$Y - X\hat{a} = \begin{pmatrix} 0,0161 \\ -0,0885 \\ 0,0238 \\ 0,0575 \\ -0,0869 \end{pmatrix} ; V^{-1} = \begin{pmatrix} 12 & & & & \\ & 8 & & & \\ & & 6 & & \\ & & & 9 & \\ & & & & 5 \end{pmatrix}$$

L'estimation de la variance de l'erreur est donnée par :

$$\hat{\sigma}_\varepsilon^2 = \frac{(Y - X\hat{a})'V^{-1}(Y - X\hat{a})}{n - k - 1} = \frac{0,1369}{5 - 1 - 1} = 0,0445$$

(n est le nombre d'observations servant à estimer les paramètres, ici $n = 5$).

- Estimation de la variance des coefficients

La matrice des variances et covariances des coefficients est :

$$\hat{\Omega}_{\hat{a}} = \hat{\sigma}_\varepsilon^2 (X'V^{-1}X)^{-1} = 0,0445 \begin{bmatrix} 2,1040 & -0,3426 \\ -0,3426 & 0,05647 \end{bmatrix}$$

soit

$$\hat{\sigma}_{a_0}^2 = 0,0445 \times 2,1040 = 0,094 \quad \text{et} \quad \hat{\sigma}_{a_1}^2 = 0,0445 \times 0,0565 = 0,00251$$

Le coefficient a_1 est-il significativement différent de 0 ?

$t^* = 0,502/0,05 = 10,04 > t_3^{0,05} = 3,18$, nous rejetons l'hypothèse H_0 , le fortifiant a une influence positive sur le poids des poulets.

C. Tests de détection de l'hétéroscédasticité

Nous allons présenter, à partir d'un exercice, différents tests permettant de détecter une hétéroscédasticité.

Exercice n° 4

 **fichier C5EX4**

Tests de détection d'une hétéroscédasticité et correction

Un directeur de la production d'une unité de construction automobile désire déterminer une relation entre le nombre de défauts constatés (Y_i) et le temps de vérification (X_i) d'une automobile, selon le modèle suivant :

$$Y_i = a_0 + a_1 X_i + u_i$$

Pour ce faire, il procède à un test sur 30 véhicules qu'il regroupe en 6 classes de 5 voitures en demandant à chaque chef d'atelier de passer un nombre d'heures de vérification fixé.

Les résultats sont consignés dans le tableau 5.

Tableau 5 – Relation entre temps de vérification
et nombre de défauts

Nombre de défauts Y_i					Temps passés en heures X_i
4	5	6	8	8	4
6	11	13	15	17	3,5
9	13	14	15	21	2
6	13	16	23	26	1,5
11	15	17	22	34	1
7	21	23	28	38	0,5

- On demande de procéder aux tests de détection d’hétéroscédasticité suivants :
- 1) test d’égalité des variances ;
 - 2) test de Goldfeld-Quandt ;
 - 3) test de Gleisjer ;
 - 4) test de White ;
 - 5) puis, en cas d’hétéroscédasticité, d’en corriger les effets.

Solution

1) Test d’égalité des variances

Les données sont constituées de $m = 6$ groupes d’observations. Nous allons tester l’égalité des variances¹, soit l’hypothèse :

$$H0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_6^2$$

Étape 1 : calcul de la variance empirique pour chaque groupe.

$$\widehat{\sigma}_i^2 = \frac{\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}{n_i - 1} = \frac{\sum_{j=1}^5 (Y_{ij} - \bar{Y}_i)^2}{4}$$

Étape 2 : calcul de la variance totale.

$$\widehat{\sigma}_T^2 = \frac{\sum_{i=1}^m (n_i - 1) \widehat{\sigma}_i^2}{\sum_{i=1}^m (n_i - 1)} = \frac{\sum_{i=1}^m v_i \widehat{\sigma}_i^2}{v}$$

avec $v_i = n_i - 1$ et $v = \sum_{i=1}^m v_i = \sum_{i=1}^m (n_i - 1)$

Étape 3 : calcul du χ^2 empirique et test.

1. Kendall M.G. et Stuart A., pages 234-236, 1961.

La quantité $Q' = v \ln \hat{\sigma}_T^2 - \sum_{i=1}^m v_i \ln \hat{\sigma}_i^2$ suit une loi du χ^2 à $m - 1$ degrés de

liberté (\ln = logarithme népérien). Cependant, l'estimation peut être améliorée en divisant Q' par une constante d'échelle C :

$$C = 1 + \frac{1}{3(m-1)} \left(\sum_{i=1}^m \frac{1}{v_i} - \frac{1}{v} \right) \quad \text{ainsi} \quad Q = \frac{Q'}{C} \rightarrow \chi^2(m-1)$$

Si $Q > \chi_{0,95}^2(m-1)$, l'hypothèse H_0 est rejetée, le modèle est hétéroscédastique.

Le tableau 6 illustre les étapes de calcul.

Tableau 6 – Étapes de calcul pour un test d'égalité de variances

i	Y_i					X_i	$\hat{\sigma}_i^2$	$v_i \ln(\hat{\sigma}_i^2)$
1	4	5	6	8	8	4	3,2	4,65
2	6	11	13	15	17	3,5	17,8	11,51
3	9	13	14	15	21	2	18,8	11,73
4	6	13	16	23	26	1,5	63,7	16,61
5	11	15	17	22	34	1	78,7	17,46
6	7	21	23	28	38	0,5	127,3	19,38

$$\hat{\sigma}_T^2 = 51,58333$$

$Q' = 13,26$; $C = 1,097$; $Q = 12,09 > \chi_{0,95}^2(m-1) = 11,07$, le modèle est donc hétéroscédastique.

2) Test de Goldfeld-Quandt¹

Ce test n'est valable que si l'une des variables est la cause de l'hétéroscédasticité et que le nombre d'observations est important. Il se compose de trois étapes.

Étape 1 : ordonner les observations en fonction des valeurs croissantes ou décroissantes de la variable à expliquer ou bien de la variable explicative soupçonnée être la source de l'hétéroscédasticité.

Cette opération est déjà effectuée puisque le nombre de défauts constatés est classé (tableau 7) en fonction du nombre d'heures décroissant de vérification.

Tableau 7 – Classement des observations en fonction du temps de vérification (extrait des données)

j	Y_j	X_j
1	4	4
2	5	4
...
29	28	0,5
30	38	0,5

1. Goldfeld S.M. et Quandt R.E., pages 539-547, 1965.

Étape 2 : omettre C observations centrales.

Nous choisissons arbitrairement C observations situées au centre de l'échantillon. Ces C observations sont exclues de l'analyse. La valeur de C doit être approximativement égale au quart du nombre d'observations totales.

$$C = \text{partie entière de } (30/4) = 8$$

Étape 3 : régressions sur les deux sous-échantillons et test.

Premier échantillon : $j = 1, 11$

$$Y_j = 16,93 - 2,13 X_j + e_j$$

(2,31)

$$n = 11$$

$$R^2 = 0,08$$

$$SCR_1 = \sum_j e_j^2 = 164,66$$

$$(\cdot) = t \text{ de Student}$$

$$ddl_1 = n - 2 = 9$$

Deuxième échantillon : $j = 20, 30$

$$Y_j = 9,84 - 1,32 X_j + e_j$$

(9,23)

$$n = 11$$

$$R^2 = 0,002$$

$$SCR_2 = \sum_j e_j^2 = 872,02$$

$$(\cdot) = t \text{ de Student}$$

$$ddl_2 = n - 2 = 9$$

Sous l'hypothèse H_0 d'homoscédasticité, le rapport¹ :

$$F^* = \frac{SCR_2/ddl_2}{SCR_1/ddl_1} = \frac{872,02/9}{164,66/9} = 5,29 \text{ suit une loi de Fisher à } ddl_2 \text{ et } ddl_1 \text{ degrés de liberté.}$$

$F^* = 5,29 > F_{9,9}^{0,05} = 3,18$, l'hypothèse d'homoscédasticité H_0 est rejetée, le modèle est donc hétéroscédastique.

3) Test de Gleisjer²

Le test de Gleisjer permet non seulement de déceler une éventuelle hétéroscédasticité, mais aussi d'identifier la forme que revêt cette hétéroscédasticité. Ce test est fondé sur la relation entre le résidu de l'estimation par les MCO effectuée sur le modèle de base et la variable explicative supposée être la cause de l'hétéroscédasticité.

Étape 1 : régression par les MCO de Y_j en X_j ($j = 1, 30$)

$$Y_j = 24,09 - 4,12 X_j + e_j$$

(4,19)

$$n = 30$$

$$R^2 = 0,38$$

Le vecteur des résidus e_j est alors connu.

1. La valeur la plus élevée des deux SCR se situe toujours au numérateur.

2. Gleisjer H., pages 316-323, 1969.

Étape 2 : régression de la valeur absolue $|e_j|$ des résidus sur X_j .

Glejser suggère de tester différentes formes de relation, par exemple :

Forme générale

Estimation ($J = 1,30$)

$$\text{I} \quad |e_j| = a_0 + a_1 X_j + v_j$$

$$|e_j| = 8,09 - 1,46 X_j + \hat{v}_j$$

$$n = 30 \quad (2,55)$$

Hétéroscédasticité de type :

$$R^2 = 0,19$$

$$\hat{\sigma}_{e_j}^2 = k^2 X_j^2 \quad (k \text{ étant une constante quelconque})$$

En effet, si nous avons une relation entre la valeur absolue du résidu (e_j) et la variable X_j , cela signifie que nous avons de fait une relation entre le résidu au carré et la variable X_j au carré.

$$\text{II} \quad |e_j| = a_0 + a_1 X_j^{1/2} + v_j$$

$$|e_j| = 10,7 - 4,15 X_j^{1/2} + \hat{v}_j$$

$$(2,60)$$

$$\hat{\sigma}_{e_j}^2 = k^2 X_j$$

$$n = 30$$

$$R^2 = 0,19$$

$$\text{III} \quad |e_j| = a_0 + a_1 X_j^{-1} + v_j$$

$$|e_j| = 2,79 + 2,86 X_j^{-1} + \hat{v}_j$$

$$(2,30)$$

$$\hat{\sigma}_{e_j}^2 = k^2 X_j^{-2}$$

$$n = 30$$

$$R^2 = 0,16$$

L'hypothèse d'homoscédasticité est rejetée si le coefficient a_1 d'une des spécifications est significativement différent de 0. Les trois t de Student empiriques sont supérieurs à 1,96, l'hétéroscédasticité est donc détectée. La forme retenue est la II ; en effet, le coefficient est affecté du ratio de Student le plus élevé (toutefois, la forme I est très proche).

4) Test de White¹

Le test de White est très proche du précédent, il est fondé sur une relation significative entre le carré du résidu et une ou plusieurs variables explicatives en niveau et au carré au sein d'une même équation de régression² :

$$e_t^2 = a_1 x_{1t} + b_1 x_{1t}^2 + a_2 x_{2t} + b_2 x_{2t}^2 + \dots + a_k x_{kt} + b_k x_{kt}^2 + a_0 + v_t$$

Soit n le nombre d'observations disponibles pour estimer les paramètres du modèle et R^2 le coefficient de détermination. Si l'un de ces coefficients de régression est significativement différent de 0, alors on accepte l'hypothèse d'hétéroscédasticité. Nous pouvons procéder à ce test soit à l'aide d'un test de Fisher classique de nullité de coefficients :

$$H_0 : a_1 = b_1 = a_2 = b_2 = \dots = a_k = b_k = 0$$

Si on refuse l'hypothèse nulle, alors il existe un risque d'hétéroscédasticité.

Soit recourir à la statistique LM qui est distribuée comme un χ^2 à $p = 2k$ degrés de liberté (autant que de coefficients que nous estimons, hormis le terme constant), si $n \times R^2 > \chi^2(p)$ lu dans la table au seuil α , on rejette l'hypothèse d'homoscédasticité des erreurs.

1. White, 1980.

2. En cas de doute sur une influence croisée de deux variables comme étant la source de l'hétéroscédasticité, on peut ajouter au modèle à estimer des termes croisés, par exemple : $x_{1t} \times x_{2t}$.

Soit ici à estimer le modèle : $e_j^2 = a_1 x_j + b_1 x_j^2 + a_0 + v_j$

$$e_j^2 = -78,58 x_j + 11,98 x_j^2 + 136,02 + \hat{v}_j$$

avec $n = 30$; $R^2 = 0,226$; $F^* = 3,956$

– Test de Fisher $F^* = 3,956 > F_{2; 27}^{0,05} = 3,35$.

– Test LM $n R^2 = 30 \times 0,22 = 6,78 > \chi_{0,05}^2(2) = 5,99$.

Nous sommes, dans les deux cas, amenés à rejeter l'hypothèse H_0 pour un seuil de 5 %. Le modèle est donc hétéroscédastique.

5) Correction de l'hétéroscédasticité

Les quatre tests sont concordants : le modèle est hétéroscédastique, il convient donc d'en corriger les effets.

Supposons, par exemple, que l'on retienne la forme I : $\hat{\sigma}_{ej}^2 = k^2 X_j^2$; l'application de la régression pondérée par le facteur $1/X_j$ conduit à un modèle homoscdastique :

$$\frac{Y_j}{X_j} = \frac{a_0}{X_j} + a_1 + \frac{e_j}{X_j} \quad \text{d'où} \quad E\left(\frac{e_j}{X_j}\right)^2 = \frac{1}{X_j^2} \hat{\sigma}_{ej}^2 = k^2$$

Or, le test de Gleisjer a mis en évidence une relation (II) du type : $\hat{\sigma}_{ej}^2 = k^2 X_j$. Pour lever l'hétéroscédasticité, dans ce cas, nous employons la régression pondérée sur les données brutes divisées par $\sqrt{X_j}$. En effet :

$$\frac{Y_j}{\sqrt{X_j}} = \frac{a_0}{\sqrt{X_j}} + a_1 \frac{X_j}{\sqrt{X_j}} + \frac{e_j}{\sqrt{X_j}} \quad \text{d'où} \quad E\left(\frac{e_j}{\sqrt{X_j}}\right)^2 = \frac{1}{X_j} \hat{\sigma}_{ej}^2 = k^2$$

En général, lorsque nous avons détecté une hétéroscédasticité de type : $\hat{\sigma}_{ej}^2 = k^2 f(X_j)$, il convient de diviser les données par $\sqrt{f(X_j)}$ afin de se ramener à un modèle homoscdastique.

Dans le cadre de notre exercice, les données sont alors transformées (tableau 8).

Tableau 8 – Transformation pour une régression pondérée
(extrait des données)

$Z_j = \frac{Y_j}{\sqrt{X_j}}$	$X_{1j} = \frac{1}{\sqrt{X_j}}$	$X_{2j} = \frac{X_j}{\sqrt{X_j}}$
2,00	0,50	2,00
2,50	0,50	2,00
...
39,60	1,41	0,71
53,74	1,41	0,71

Le modèle estimé par les MCO est le suivant :

$Z_j = b_1 X_{1j} + b_2 X_{2j} + \varepsilon_j$, ε_j répond aux hypothèses classiques. Il est à noter l'absence du terme constant. Les résultats fournis par Eviews sont :

```
SMPL      1 - 30
30 Observations
LS // Dependent Variable is Z
```

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
X1	24.960028	2.5120465	9.9361327	0.000
X2	-4.5376024	1.5410077	-2.9445683	0.006

```
=====
R-squared      0.642533      Mean of dependent var    14.42033
Adjusted R-squared 0.629766      S.D. of dependent var    12.44159
S.E. of regression 7.570315      Sum of squared resid    1604.671
Durbin-Watson stat 1.712764      F-statistic              50.32894
Log likelihood   -102.2603
=====
```

Les coefficients du modèle initial sont $\hat{a}_0 = \hat{b}_1 = 24,96$ et $\hat{a}_1 = \hat{b}_2 = -4,53$. Le modèle estimé¹ est donc :

$$Y_j = 24,96 - 4,53 X_j + e_j$$

(2,94)

$$n = 30$$

(.) = *t* de Student

Il existe bien une influence significative du temps de vérification sur le nombre de défauts constatés, chaque heure de vérification permet de supprimer en moyenne 4,5 défauts.

D. Autre test d'hétéroscédasticité : le test ARCH

Les modèles de type ARCH² (« AutoRegressive Conditional Heteroscedasticity ») permettent de modéliser des chroniques (la plupart du temps financières³) qui ont une volatilité (ou variance ou variabilité) instantanée qui dépend du passé. Il est ainsi possible d'élaborer une prévision dynamique de la chronique en termes de moyenne et de variance. Le test est fondé soit sur un test de Fisher classique, soit sur le test du multiplicateur de Lagrange (*LM*).

De manière pratique, on procède de la manière suivante :

– 1^{re} étape : calcul de e_j , le résidu du modèle de régression ;

1. Ce modèle est à comparer à celui estimé sur les données brutes (premier modèle du test de Gleisjer).

Le coefficient de détermination n'est pas présenté, car, dans tout modèle estimé sans terme constant, il n'est pas interprétable.

2. Engle, 1982.

3. En effet, les séries financières sont particulièrement concernées par les modèles ARCH, car on constate des périodes de forte spéculation (variabilité élevée) suivie de périodes d'accalmie (variabilité faible).

- 2^e étape : calcul des e_t^2 ;
- 3^e étape : régression autorégressive des résidus sur p retards (résidu décalé)

où seuls les retards significatifs sont conservés, $e_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i e_{t-i}^2$;

Soit à tester l'hypothèse $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_p = 0$.

- 4^e étape : calcul de la statistique du multiplicateur Lagrange, $LM = n \times R^2$ avec :
 n = nombre d'observations servant au calcul de la régression de l'étape 3,
 R^2 = coefficient de détermination de l'étape 3.

Si $LM > \chi^2(p)$ à p degrés de liberté lu dans la table à un seuil α fixé (en général 0,05), on rejette H_0 ; on considère que le processus est justifiable d'un modèle ARCH(p).

C'est le test de significativité des coefficients α_i de la régression e_t^2 sur e_{t-p}^2 qui permet de déterminer l'ordre p du processus ARCH sachant qu'un processus ARCH d'ordre 3 semble un maximum. Une autre approche consiste à calculer le corrélogramme des résidus aux carrés issus du modèle initial. Si des termes de ce corrélogramme sont significativement différents de 0, alors on peut conclure à une spécification de type ARCH.

III. **Modèles à erreurs sur les variables**

A. **Conséquences lorsque les variables sont entachées d'erreurs**

Quand nous avons étudié le modèle linéaire, nous avons admis que la variable endogène et les variables exogènes étaient observables sans erreur. Dans la pratique, cette hypothèse est rarement vérifiée ; cependant, nous pouvons admettre, généralement, que l'erreur de mesure des observations est faible par rapport à l'erreur de spécification.

Toutefois, dans certains modèles, les variables économiques retenues peuvent être entachées d'une erreur de mesure relativement importante. C'est le cas, par exemple, lorsque les données proviennent, non pas, d'une mesure directe, mais de données d'enquêtes par sondage. Dans ce cas, il convient de distinguer les variables vraies (et inconnues) : $y^*, x_1^*, x_2^*, \dots, x_k^*$ des valeurs observées y, x_1, x_2, \dots, x_k et d'étudier les conséquences concernant les propriétés de l'estimateur obtenu par les MCO.

Soit le modèle $Y^* = X^*a + \varepsilon$ avec ε qui satisfait aux hypothèses habituelles.

Posons :

$$X = X^* + \mu \quad \text{et} \quad Y = Y^* + v$$

avec : $E(\mu) = 0$; $E(v) = 0$; $E(X^*\mu) = 0$; $E(Y^*v) = 0$; $E(X^*v) = 0$; $E(Y^*\mu) = 0$.

On a alors :

$$E(\varepsilon'\mu) = E\{(Y^* - X^*a)'\mu\} = E(Y^*\mu) - a'E(X^*\mu) = 0$$

$$E(\varepsilon'v) = E\{(Y^* - X^*a)'v\} = E(Y^*v) - a'E(X^*v) = 0$$

Nous avons donc indépendance entre les erreurs sur les variables μ et v , et l'erreur de spécification du modèle ε .

La relation entre les variables observées X et Y est la suivante :

$$Y^* = Y - v = (X - \mu)a + \varepsilon \rightarrow Y = Xa + v - \mu a + \varepsilon = Xa + \eta$$

avec

$$\eta = v - \mu a + \varepsilon$$

Les propriétés stochastiques de η sont :

$$E(\eta) = E(v - \mu a + \varepsilon) = E(v) - E(\mu)a + E(\varepsilon) = 0$$

$$E(X^*\eta) = E(X^*v) - E(X^*\mu)a + E(X^*\varepsilon) = 0$$

$$\begin{aligned} E(X'\eta) &= E\{(X^* + \mu)'\eta\} = E(\mu'\eta) \\ &= E(\mu'v) - E(\mu'\mu)a + E(\mu'\varepsilon) \\ &= -E(\mu'\mu)a \neq 0 \end{aligned}$$

L'hypothèse H6 du modèle général n'est donc pas vérifiée puisque η et X sont corrélés, la méthode des MCO fournit des estimateurs biaisés négativement.

B. La méthode des variables instrumentales

Lorsqu'on se trouve en présence d'un modèle à erreurs sur les variables $Y = Xa + \eta$, l'hypothèse H6 est en défaut et l'estimateur \hat{a} ne converge pas asymptotiquement vers a . Les autres hypothèses sont réputées vérifiées.

Le but de la technique des variables instrumentales est de déterminer k variables z_1, z_2, \dots, z_k telles que :

$$E(Z'\eta) = 0 \quad \text{et} \quad Z = (z_1, z_2, \dots, z_k)$$

$$\text{Cov}(Z'X) \neq 0$$

c'est-à-dire qu'aucune combinaison linéaire des variables z_k n'est orthogonale aux variables x_1, x_2, \dots, x_k ou encore que les variables Z et X soient corrélées.

Nous avons alors :

$$E(Z'Y) = E\{Z'(Xa + \eta)\} = E(Z'X)a + E(Z'\eta) = E(Z'X)a$$

soit

$$\hat{a} = (Z'X)^{-1}Z'Y \quad [12]$$

On démontre¹ que \hat{a} est un estimateur convergent de a , la variance de l'estimateur \hat{a} est d'autant plus faible que la corrélation entre $Z'X$ est forte, la matrice des variances et covariances des coefficients est égale à :

$$\hat{\Omega}_{\hat{a}} = \hat{\sigma}_{\varepsilon}^2 (Z'X)^{-1} (Z'Z) (X'Z)^{-1} \quad [13]$$

La difficulté de mise en œuvre de cette méthode réside dans la sélection des variables instrumentales² « miracles » Z qui doivent être non corrélées avec η et fortement corrélées avec X . Dans certains cas, nous pouvons simplement retenir, comme variable instrumentale, la variable exogène décalée d'une période.

C. Le test d'exogénéité d'Hausman

Le test d'exogénéité d'Hausman (1978) permet de détecter une éventuelle corrélation entre le terme d'erreur ε_t et une ou des variables explicatives x_{it} . Dans cette hypothèse, nous ne pouvons plus utiliser l'estimateur des MCO qui est non convergent, il faut alors recourir à la méthode des Variables Instrumentales (VI) ou à la Méthode des Moments Généralisée (GMM).

Soit le test d'hypothèses, $H_0 : \text{Cov}(x_t, \varepsilon_t) = 0$ (la variable x_t est exogène) contre l'hypothèse d'endogénéité $H_1 : \text{Cov}(x_t, \varepsilon_t) \neq 0$.

Sous l'hypothèse H_0 les estimateurs des MCO et des VI sont convergent alors que sous l'hypothèse H_1 la covariance est non nulle et l'estimateur des MCO est biaisé et non convergent. Ce test peut être mené de deux manières, soit un test de différence entre l'estimateur des VI et des MCO, soit un test à partir d'une régression augmentée.

1) Test de différence

Nous calculons la statistique :

$$H = (\hat{a}_{VI} - \hat{a}_{MCO})' [\text{Var}(\hat{a}_{VI}) - \text{Var}(\hat{a}_{MCO})]^{-1} (\hat{a}_{VI} - \hat{a}_{MCO}).$$

La statistique H est distribuée selon un chi-deux à k degrés de liberté. Si $H < \chi^2(k)$ pour un seuil α % fixé, nous acceptons l'hypothèse H_0 , l'estimateur MCO est non biaisé.

1. Judge G.G. *et al.*, pages 577-579, 1988.

2. Les anglo-saxons emploient le terme de « proxy variable », que l'on peut traduire littéralement par « variable par délégation ».

2) Régression augmentée

La procédure proposée par Hausman est en quatre étapes :

- Estimation d'un modèle par les MCO avec pour variable à expliquer la variable dont nous désirons tester l'exogénéité et comme variables explicatives le ou les instruments, le plus souvent les variables explicatives décalées d'une période.
- Estimation de la ou des variables ajustées \hat{x}_{it} à partir de la ou des régressions précédentes.
- Estimation du modèle augmenté (modèle initial dans lequel nous rajoutons la ou les variables explicatives ajustées \hat{x}_{it}).
- Test de significativité par rapport à 0 du ou des coefficients de la ou des variables explicatives ajustées. Si ce ou ces coefficients ne sont pas significativement de 0 (test de Student ou de Fisher), alors nous retenons l'hypothèse $H_0 : \text{Cov}(x_i, \varepsilon_i) = 0$.

D. La méthode des moments généralisée

La Méthode des Moments Généralisée, GMM (*Generalized Method of Moments*), est utilisée lorsque la ou les variables explicatives sont supposées exogènes ($\text{Cov}(x_i, \varepsilon_i) \neq 0$) et que, de plus, la matrice des variances covariances des erreurs est quelconque ($E(\varepsilon_i, \varepsilon_i') \neq \sigma^2 I$). L'estimateur des GMM combine alors la méthode des moindres carrés généralisés avec celle des variables instrumentales. L'estimateur des GMM est donné par :

$$\hat{a} = (X'Z(Z'\hat{\Omega}Z)^{-1}Z'X)^{-1}X'Z(Z'\hat{\Omega}Z)^{-1}Z'y \quad [14]$$

avec :

y = la variable à expliquer

X = les variables explicatives

Z = les instruments

$\hat{\Omega}$ = la matrice des variances covariances des résidus estimés dans une première étape par la méthode des variables instrumentales

Il est à noter que dans le cas où les hypothèses classiques sont vérifiées ($E(\varepsilon_i, \varepsilon_i') = \sigma^2 I$), l'estimateur des GMM (expression [14]) se ramène à l'estimateur VI (expression [13]).

Exercice n° 5

↓ fichier C5EX5

Test d'exogénéité d'Hausman, modèle à erreurs sur les variables : la technique des variables instrumentales

Un agronome désire estimer la relation entre le rendement de blé (y_i) et la quantité utilisée d'engrais (x_i^*). Pour ce faire, il ne dispose que de la quantité d'engrais (x_i) déclarée par l'agriculteur, qui est donc entachée d'une erreur. Cependant, il pense que la variable de dépense effective en achat d'engrais (z_i) est indépendante de l'erreur d'observation de la quantité d'engrais déclarée et bien corrélée avec la consommation réelle d'engrais.

On dispose des vingt observations présentées au tableau 9.

Tableau 9 – Rendement de blé, quantité d'engrais utilisée et dépense en achat d'engrais

Observation	y_i	x_i	z_i
1	15,30	17,30	3,00
2	19,91	21,91	7,00
3	20,94	22,96	5,40
...
18	25,83	29,43	22,20
19	25,15	28,95	24,60
20	25,06	28,86	24,60

On demande :

- 1) de tester une éventuelle endogénéité de la variable x_i à l'aide du test d'Hausman ;
- 2) d'estimer la relation entre y_i et x_i^* par une méthode adaptée.

Solution

- 1) *Test de différence* : Nous calculons la statistique d'Hausman :

$$H = (\widehat{a}_{VI} - \widehat{a}_{MCO})' [\text{Var}(\widehat{a}_{VI}) - \text{Var}(\widehat{a}_{MCO})]^{-1} (\widehat{a}_{VI} - \widehat{a}_{MCO})$$

Avec :

$$\begin{aligned} \widehat{a}_{VI} &= \begin{bmatrix} 0,795 \\ 2,153 \end{bmatrix}; \widehat{a}_{MCO} = \begin{bmatrix} 0,822 \\ 1,471 \end{bmatrix}; \text{Var}(\widehat{a}_{VI}) = \begin{bmatrix} 0,0004489 & -0,01106 \\ -0,01106 & 0,27613 \end{bmatrix}; \\ \text{Var}(\widehat{a}_{MCO}) &= \begin{bmatrix} 0,0003486 & -0,00859 \\ -0,00859 & 0,214867 \end{bmatrix}; (\widehat{a}_{VI} - \widehat{a}_{MCO}) = \begin{bmatrix} -0,0276 \\ 0,6817 \end{bmatrix}; \\ [\text{Var}(\widehat{a}_{VI}) - \text{Var}(\widehat{a}_{MCO})] &= \begin{bmatrix} 0,0001 & -0,0024 \\ -0,0024 & 0,0612 \end{bmatrix} \end{aligned}$$

$H = [-0,0276 \quad 0,6817] \begin{bmatrix} 1568341,7 & 63248,1 \\ 63248,1 & 2566,99 \end{bmatrix} \begin{bmatrix} -0,0276 \\ 0,6817 \end{bmatrix} = 7,63 > \chi^2(2)$
 pour un seuil de 5 % = 5,99. Nous rejetons l'hypothèse H_0 , l'estimateur des MCO est biaisé, il convient d'utiliser l'estimateur des VI.

Régression augmentée

Nous procédons au test d'Hausman en quatre étapes.

- Estimation par les MCO de la régression de x_i sur l'instrument z_i :

Dependent Variable : X				
Method : Least Squares				
Included observations : 20				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	18.37253	0.622495	29.51436	0.0000
Z	0.440680	0.039908	11.04238	0.0000

- Calcul de la série ajustée : $\hat{x}_i = 18,37 + 0,44z_i$
- Estimation du modèle augmenté avec ($XF = \hat{x}_i$) :

Dependent Variable : Y				
Method : Least Squares				
Included observations : 20				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
X	1,010328	0,022324	45,25680	0.0000
XF	- 0,215107	0,023915	- 8,994504	0.0000
C	2,153561	0,212721	10,12386	0.0000

- Le coefficient de la variable XF est significativement différent de 0, nous ne sommes pas en mesure d'accepter l'hypothèse H_0 , nous avons donc $\text{Cov}(x_i, \varepsilon_i) \neq 0$.

2) La régression de y_i sur x_i conduit aux résultats suivants :

$$\hat{y}_i = 1,47 + 0,82 X_i \quad (41)$$

$$R^2 = 0,99$$

$$n = 20$$

$$(\cdot) = t \text{ de Student}$$

Cependant la méthode des MCO n'est pas applicable car $E(x_i \eta) \neq 0$. En revanche, nous savons par hypothèses que $E(z_i \eta) = 0$ et que $\text{Cov}(x_i^* z_i) \neq 0$.

L'estimateur \hat{a} est donné par [12], soit :

$$\hat{a} = \begin{pmatrix} 2 & 1 \end{pmatrix} = \begin{pmatrix} Z' & X \end{pmatrix}^{-1} \begin{pmatrix} Z' & Y \end{pmatrix} = \begin{pmatrix} (20,2) & (20,2) \end{pmatrix} \begin{pmatrix} (2,20) & (20,1) \end{pmatrix}$$

La matrice Z est composée de 1 pour la première colonne et des valeurs de z_i pour la deuxième colonne. De même, la matrice X est composée de 1 pour la première colonne et des valeurs de x_i pour la deuxième colonne.

Nous obtenons :

$$Z'X = \begin{bmatrix} 20,00 & 492,78 \\ 284,40 & 7\,369,53 \end{bmatrix} ; \quad (Z'X)^{-1} = \begin{bmatrix} 1,02 & -0,07 \\ -0,04 & 0,00 \end{bmatrix} ;$$

$$Z'Y = \begin{bmatrix} 434,94 \\ 6472,88 \end{bmatrix}$$

soit
$$\hat{a} = \begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \end{bmatrix} = \begin{bmatrix} 2,15 \\ 0,795 \end{bmatrix}$$

Ces coefficients sont à comparer à ceux trouvés lors de l'estimation du modèle par les MCO.

Nous pouvons calculer l'estimation de la variance de l'erreur :

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_i e_i^2}{n-2} = \frac{1,29}{18} = 0,071$$

La matrice des variances et covariances de \hat{a} est alors d'après [13] :

$$\hat{\Omega}_{\hat{a}} = \hat{\sigma}_\varepsilon^2 (Z'X)^{-1} (Z'Z)(X'Z)^{-1}$$

$$Z'Z = \begin{bmatrix} 20,00 & 284,40 \\ 284,40 & 4\,866,08 \end{bmatrix} ; \quad \text{soit} \quad \hat{\Omega}_{\hat{a}} = 0,071 \begin{bmatrix} 3,85 & -0,15 \\ -0,15 & 0,00 \end{bmatrix}$$

$$\rightarrow \hat{\sigma}_{\hat{a}1} = 0,0211$$

Le modèle estimé à partir de la méthode des variables instrumentales est donc le suivant :

$$\hat{y}_i = 2,15 + 0,795 X_i$$

(37)

$$n = 20$$

(.) = t de Student

Exercice n° 6

 **fichier C5EX6**

Prévision de part de marché

Une entreprise de la grande consommation cherche à prévoir la part de marché d'un produit apéritif (PM_t) en fonction de sa présence dans l'univers de vente ¹ (DN_t) et de son indice de prix par rapport aux concurrents (IP_t). Les données portent sur 41 semaines.

1. La DN ou distribution numérique représente le nombre de fois où le produit est présent dans la distribution GMS (Grande et Moyennes Surfaces). Si $DN = 100$, cela signifie que le produit est présent dans tous les magasins, si $DN = 80$, le produit est présent dans 80 % des magasins. Bien sûr, plus la DN augmente et plus le produit est présent dans les linéaires et donc la probabilité de vendre le produit s'accroît.

Le chef de produit estime un premier modèle dont les résultats sont présentés ci-dessous :

$$\text{Log}(PM_t) = 2,91 + 1,03 \text{Log}(DN_t) + e_t$$

(48,5) (6,83)

$$n = 41$$

$$R^2 = 0,54$$

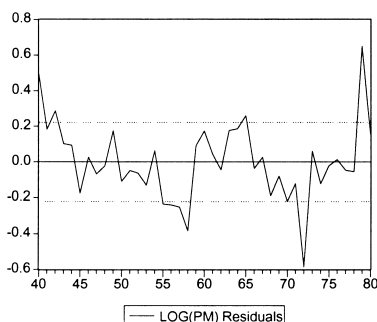
$$DW = 1,29$$

$$(.) = t \text{ de Student}$$

$$\text{Log} = \text{Logarithme népérien}$$

1) Que représente le coefficient a_1 du modèle ? Est-il significativement de 0 ?

2) Le graphique des résidus est le suivant qu'en pensez-vous ?



Existe-t-il une autocorrélation des erreurs ?

3) Le chef de produit ajoute l'indice des prix à son équation de régression et obtient le résultat suivant :

$$\text{Log}(PM_t) = 25,22 + 1,28 \text{Log}(DN_t) - 4,71 \text{Log}(IP_t) + e_t$$

(6,7) (10,9) (-5,9)

$$n = 41$$

$$R^2 = 0,76$$

$$DW = 1,83$$

Commentez d'un point de vue statistique et économique les résultats obtenus.

4) Connaissant pour la semaine 42 les valeurs de $DN_{42} = 0,60$ et $IP_{42} = 100$, on vous demande de calculer une prévision de part de marché pour la semaine 42 assortie de son intervalle à 95 %.

On donne la matrice $(X'X)^{-1}$

$$(X'X)^{-1} = \begin{bmatrix} 543,63 & 6,38 & -114,92 \\ 6,38 & 0,54 & -1,32 \\ -114,92 & -1,32 & 24,29 \end{bmatrix}$$

Solution

1) Le coefficient a_1 du modèle représente une élasticité car le modèle est sous la forme Log-Log. Ce coefficient est significativement de 0 car la valeur empirique du t de Student est largement supérieure au t lu dans la table ($t^{0,05} = 1,96$).

2) Le graphique laisse supposer une autocorrélation des erreurs. Puisque les conditions d'application du test de Durbin et Watson sont vérifiées, nous pouvons interpréter

cette statistique : $DW = 1,29 < 1,44$ (valeur lue dans la table). Il existe une présomption d'autocorrélation des erreurs d'ordre 1.

3) Commentaires des résultats

Commentaires statistiques :

- la statistique de Fisher indique que le modèle est globalement significatif,
- les t de Student ont des probabilités critiques inférieures à 0,05, les coefficients sont donc tous significatifs,
- la statistique de $DW = 1,82$ ne laisse plus présager d'une autocorrélation des erreurs.

Nous constatons que le fait d'avoir ajouter au modèle une variable explicative corrige l'autocorrélation des erreurs. Le modèle est validé sur le plan statistique.

Commentaires économiques :

- la variable DN agit positivement, plus l'entreprise est présente dans l'univers de vente plus la part de marché augmente (10 % d'augmentation entraîne 12,5 % d'augmentation de part de marché),
- la variable IP agit négativement, il s'agit d'un effet prix classique (10 % d'augmentation entraîne 47 % de baisse de part de marché),

Les coefficients des variables ont bien le signe attendu.

4) On calcule le logarithme népérien : $LN(DV) = -0,51$; $LN(IP) = 4,605$

La prévision en semaine 42 est donnée par :

$$\text{Log}(\widehat{PM}) = 25,22 + 1,285 \times -0,51 - 4,716 \times 4,605 = 2,84$$

D'où $PM = e^{2,84} = 17,18$. La part de marché prévue est de 17,18 %

L'écart type de l'erreur de prévision est égal à :

$$\hat{\sigma}_{e_{t+h}}^2 = \hat{\sigma}_\varepsilon^2 (1 + X'_{t+h} (X'X)^{-1} X_{t+h}) = (0,16132)^2 (1 + 0,27) = 0,033.$$

$$\text{avec } X_{t+h} = \begin{pmatrix} 1 \\ -0,51 \\ 4,605 \end{pmatrix}$$

L'intervalle de prévision à 95 % de $\text{Ln}(PM)$ est donné par : $2,84 \pm 1,96 \times \sqrt{0,033}$.

Soit $IC = [2,483; 3,196]$

Nous passons aux exponentiels afin d'obtenir l'intervalle de confiance de la part de marché : $[11,95; 24,43]$.

Exercice n° 7

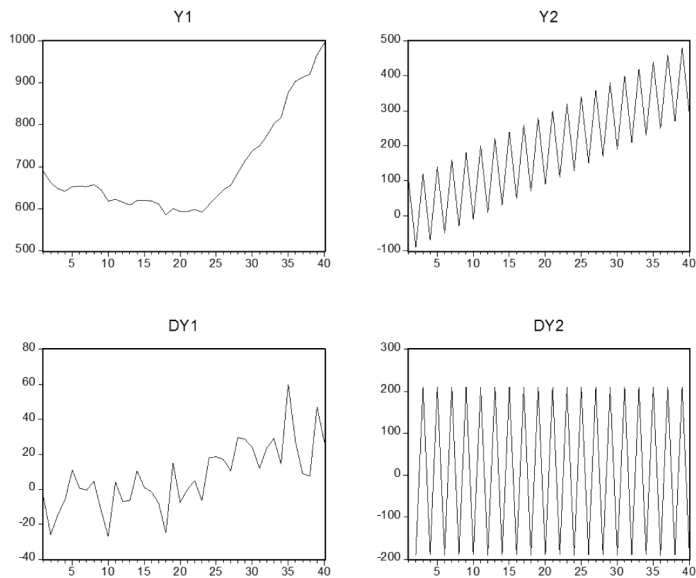
📁 fichier C5EX7

Un peu de réflexion économétrique...

Soit trois variables construites artificiellement connues sur $n = 40$ périodes.

- $Y1$ et $Y1$ en différences premières : $DY1 = Y1 - Y1(-1)$
- $Y2$ et $Y2$ en différences premières : $DY2 = Y2 - Y2(-1)$
- une variable $TENDANCE = 1, 2, \dots, n$

Les graphiques 7 montrent les variations des variables Y1, Y2, DY1, DY2 en fonction du temps.



Graphique 7 – Évolution des variables Y1, Y2, DY1, DY2 en fonction du temps

En vous aidant des graphiques, de la statistique de Durbin-Watson et du coefficient de détermination, on demande de marier chacune des 4 régressions proposées (Résultats du modèle) avec le couple « Variable dépendante/Variable explicative ».

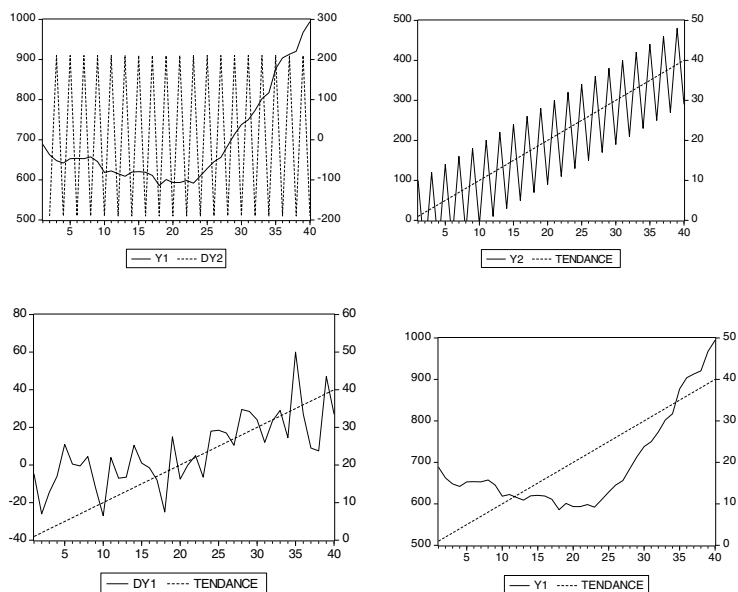
Résultats du modèle			Variables	
	Durbin et Watson	R ²	Dépendante	Explicative
Modèle 1	1,97	0,52	Y1	TENDANCE
Modèle 2	3,91	0,55	DY1	TENDANCE
Modèle 3	0,03	0	Y2	TENDANCE
Modèle 4	0,053	0,52	Y1	DY2

Solution

On procède par déduction et élimination.

- Modèle 3 : Y1/DY2 car coefficient de détermination quasi nul et très forte autocorrélation des erreurs.
- Modèle 2 : Y2/TENDANCE car coefficient de détermination élevé et très forte autocorrélation négative des erreurs.
- Modèle 1 : DY1/TENDANCE car coefficient de détermination élevé et absence d'autocorrélation.
- Modèle 4 : Y1/TENDANCE car forte autocorrélation positive des erreurs et coefficient de détermination élevé.

Les graphiques 8 suivants illustrent les autocorrélations et les coefficients de détermination.



Graphique 8 – Illustrations des modèles : autocorrélations et les coefficients de détermination

6. Les modèles non linéaires

Nous avons examiné le cas du modèle linéaire, cependant, très souvent, la théorie économique doit être formalisée à l'aide de relations non linéaires.

Dans ce chapitre, la part consacrée au cours est très réduite. En effet, l'étude des modèles non linéaires se prête bien à l'exercice pratique. Nous examinons tout d'abord en I., les différents types de modèles non linéaires : ceux qui sont linéaires sur leurs paramètres et des modèles un peu plus complexes (modèle de diffusion, par exemple) qui font l'objet d'une étude particulière. Puis, les méthodes d'estimation des paramètres des modèles non linéarisables sont présentées en II.

I. Les différents types de modèles non linéaires

A. Les fonctions de type exponentiel

Nous allons examiner le cas de modèles dont la spécification n'est pas linéaire, mais qui sont linéaires par rapport à leurs paramètres.

Exercice n° 1

 **fichier C6EX1**

Estimation d'une fonction de production de type Cobb-Douglas

Nous avons relevé pour 25 entreprises : la production Q , le facteur capital K et le facteur travail L ; ces variables sont exprimées en millions d'euros.

Une fonction de production de type Cobb-Douglas est de la forme :

$$Q = \alpha_0 K^{\alpha_1} L^{\alpha_2}$$

On demande :

- 1) de donner une interprétation économique des coefficients α_1 et α_2 ;
- 2) d'estimer les paramètres du modèle et de commenter les résultats ;
- 3) de calculer le risque α de première espèce concernant l'hypothèse des rendements d'échelle constants.

Les données collectées sont présentées dans le tableau 1 suivant :

*Tableau 1 – Données de production pour 25 entreprises
(extrait des données)*

K	L	Q
8	23	106,00
9	14	81,08
...
3	10	43,21
6	71	121,24

Solution

- 1) Les coefficients α_1 et α_2 représentent respectivement l'élasticité de la production au facteur capital et l'élasticité de la production au facteur travail.

• Démonstration :

Nous avons une relation du type : $Q = \alpha_0 K^{\alpha_1} L^{\alpha_2}$.

Une transformation logarithmique conduit à :

$$\text{Log}(Q) = \text{Log}(\alpha_0) + \alpha_1 \text{Log}(K) + \alpha_2 \text{Log}(L)$$

Or l'élasticité de y à x se définit comme étant un rapport de variation relative :

$$e = \frac{\Delta y/y}{\Delta x/x} = \frac{\partial \text{Log}(y)}{\partial \text{Log}(x)} = \text{rapport de dérivées logarithmiques.}$$

Les coefficients α_1 et α_2 sont donc des élasticités.

► Dans tout modèle spécifié sous forme Log-Log, les coefficients s'interprètent directement en termes d'élasticités.

- 2) Estimation des coefficients du modèle Log-linéaire

Le modèle à estimer est donc le suivant :

$$\begin{array}{rclclclclcl} \text{Log}(Q) & = & \text{Log}(\alpha_0) & + & \alpha_1 & \text{Log}(K) & + & \alpha_2 & \text{Log}(L) & + \varepsilon \\ LQ & = & a_0 & + & \alpha_1 & LK & + & \alpha_2 & LL & + \varepsilon \end{array}$$

où les variables LQ , LK et LL (tableau 2) sont les variables Q , K et L transformées.

Les résultats de l'estimation sont les suivants :

25 Observations
 LS // Dependent Variable is LQ

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	1.0838457	0.0568868	19.052660	0.000
LK	0.6389324	0.0355138	17.991123	0.000
LL	0.2543517	0.0273723	9.2922871	0.000
R-squared	0.938754	Mean of dependent var	1.901200	
Adjusted R-squared	0.933186	S.D. of dependent var	0.157647	
S.E. of regression	0.040749	Sum of squared resid	0.036531	
Durbin-Watson stat	2.515340	F-statistic	168.6042	
Log likelihood	46.13244			

Les t^* de Student calculés sont tous supérieurs au t lu ($t_{22}^{0,05} = 2,07$), les coefficients sont donc significativement différents de 0 au seuil de 5 %.

Tableau 2 – Données transformées en logarithmiques décimaux
 (extrait des données)

$LK =$ Log(k)	$LL =$ Log(L)	$LQ =$ Log(Q)
0,90	1,36	2,03
0,95	1,15	1,91
...
0,48	1,00	1,64
0,78	1,85	2,08

Attention, la statistique de Durbin et Watson ne s’interprète pas puisque le modèle est spécifié en coupe instantanée.

L’estimation statistique est donc satisfaisante.

L’élasticité de la production au capital est égale à : $\alpha_1 = 0,64$.

L’élasticité de la production au travail est égale à : $\alpha_2 = 0,25$.

(Si le facteur capital augmente de 10 %, la production augmente de 6,4 % et si le fac-
 teur travail augmente de 10 %, la production augmente de 2,5 %.)

Le coefficient α_0 est égal à : $\alpha_0 = 10^{1,084} = 12,13$.

La fonction de production estimée s’écrit : $\widehat{Q} = 12,13 K^{0,64} L^{0,25}$.

3) Les rendements d'échelle sont-ils constants ?

Les rendements d'échelle sont :

- décroissants si $\alpha_1 + \alpha_2 < 1$ (la production augmente dans une proportion moindre que les facteurs de production) ;
- constants si $\alpha_1 + \alpha_2 = 1$ (la production augmente dans une proportion identique aux facteurs de production) ;
- croissants si $\alpha_1 + \alpha_2 > 1$ (la production augmente plus vite que les facteurs de production).

Or nous avons $\hat{\alpha}_1 + \hat{\alpha}_2 = 0,89$, pour quelle probabilité critique cette valeur est-elle significativement inférieure à 1 ?

On teste donc les hypothèses suivantes :

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

$$H_1 : \alpha_1 + \alpha_2 < 1$$

Le t calculé de Student est donc égal, sous H_0 , à :

$$t^* = \frac{\hat{\alpha}_1 + \hat{\alpha}_2 - 1}{\hat{\sigma}_{\hat{\alpha}_1 + \hat{\alpha}_2}}$$

Or la matrice des variances et covariances des coefficients est fournie par Eviews :

Covariance Matrix			
C, C	0.003236	C, LK	-0.001303
C, LL	-0.001430	LK, LK	0.001261
LK, LL	0.000314	LL, LL	0.000749

$$\text{Soit } \hat{\sigma}_{\hat{\alpha}_1 + \hat{\alpha}_2}^2 = \hat{\sigma}_{\hat{\alpha}_1}^2 + \hat{\sigma}_{\hat{\alpha}_2}^2 + 2 \text{ cov}(\hat{\alpha}_1, \hat{\alpha}_2)$$

$$= 0,000749 + 0,001261 + 2 \times 0,000314 = 0,002638$$

$$t^* = \frac{\hat{\alpha}_1 + \hat{\alpha}_2 - 1}{\hat{\sigma}_{\hat{\alpha}_1 + \hat{\alpha}_2}} = \frac{0,64 + 0,25 - 1}{0,05136} = -2,14 \simeq t_{22}^{0,020}$$

Nous avons 2 % de risques de rejeter l'hypothèse H_0 à tort, nous pouvons donc considérer les rendements d'échelle comme significativement décroissants.

B. Les modèles de diffusion

La courbe de vie du produit est une notion utilisée en marketing : les ventes évoluent en fonction du temps ; le démarrage est relativement lent lorsque le produit est nouveau, puis elles connaissent une croissance rapide. La courbe passe ensuite par un point d'inflexion à partir duquel le rythme de croissance des ventes diminue.

Nous allons examiner deux modèles classiquement utilisés.

1) Le modèle logistique (ou courbe de Verhulst ou courbe de Pearl)

Ce modèle est défini par la formulation suivante :

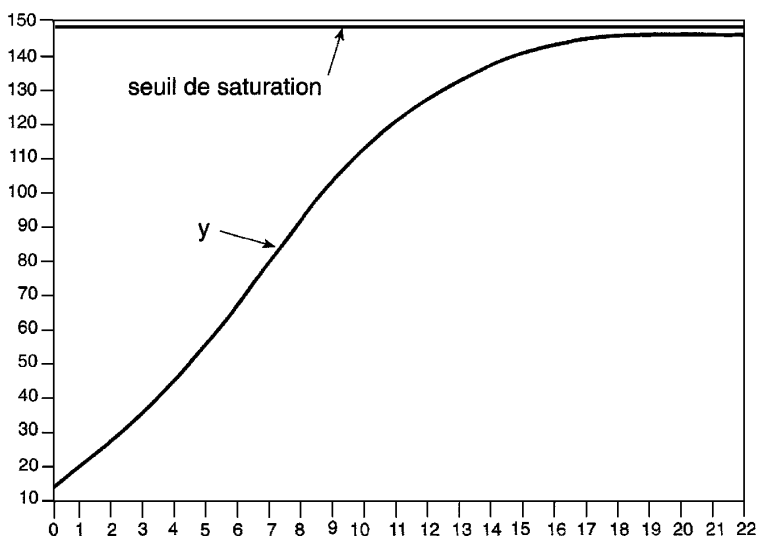
$$y_t = \frac{y_{\max}}{1 + br^t}$$

où b et r sont les deux paramètres du modèle ($0 < r < 1$), r définit la vitesse de diffusion (plus r est proche de 0, plus vite nous atteindrons l'asymptote) ; b est une caractéristique d'ordonnée à l'origine ; y_{\max} = seuil de saturation.

Si $t \rightarrow \infty$ alors $y_t \rightarrow y_{\max}$, et si $t \rightarrow -\infty$ alors $y_t \rightarrow 0$.

Le point d'inflexion est fixe et est atteint pour $y_t = y_{\max}/2$.

Le graphique 1 illustre l'allure générale de la courbe logistique.



Graphique 1 – La courbe logistique ($y_{\max} = 148,31$, $b = 10$, $r = 0,7$)

2) Le modèle de Gompertz

Ce modèle est défini par la formulation suivante :

$$y_t = e^{br^t + a} \rightarrow \ln(y_t) = br^t + a$$

(e = base du logarithme népérien).

Avec $y_{\max} = e^a$ = seuil de saturation, b et r sont les deux paramètres du modèle ($b < 0$ et $0 < r < 1$).

Si $t \rightarrow \infty$ alors $y_t \rightarrow y_{\max}$, et si $t \rightarrow -\infty$ alors $y_t \rightarrow 0$.

3) Les méthodes d'estimation

La difficulté d'estimation de ce type de modèles¹ vient du fait qu'ils ne sont pas linéaires. L'emploi du modèle de régression est donc impossible tel quel. L'estimation des paramètres d'un tel modèle doit être effectuée par des méthodes de régression non linéaire qui nécessitent des procédures numériques itératives assez complexes, objet du paragraphe suivant.

II. Méthodes d'estimation des modèles non linéaires

Nous avons jusqu'à maintenant procédé à des artifices de calcul afin de se ramener à des modèles linéaires. Cependant, pour certaines spécifications de modèle il s'avère impossible, même sous certaines hypothèses, d'utiliser les MCO. Nous allons présenter, très succinctement², les méthodes d'estimation non linéaires en essayant d'en montrer l'inspiration générale et les limites. Des exemples d'utilisation sont présentés, ensuite, à partir des logiciels RATS et Eviews.

A. Initiation aux méthodes d'estimation non linéaires

L'inspiration générale des techniques d'estimation non linéaires sont les mêmes quel que soit le type d'algorithme. Il s'agit de méthodes itératives dans lesquelles l'équation non linéaire est linéarisée à l'aide d'un développement limité de Taylor (à partir d'un jeu de valeurs de coefficients initiaux). Les MCO sont alors utilisés sur cette équation linéaire afin d'estimer des nouveaux coefficients. Ces nouveaux coefficients permettent, à l'aide d'un nouveau développement limité, de procéder à une nouvelle linéarisation. La procédure est arrêtée lorsque les coefficients sont relativement stables d'une itération à l'autre.

Pour ce type de méthode soit efficace, il convient que les valeurs initiales ne soient pas trop éloignées des valeurs optimales. Dans le cas contraire, il n'y a pas convergence et la procédure est défailante.

-
1. Pour un approfondissement des modèles de diffusion cf. Bourbonnais R. et Usunier J. C. Chapitre 7, (2013).
 2. Pour un développement très complet, le lecteur peut se référer à Greene W.H., chapitre 10, 2000.

Algorithme de Gauss-Newton

Soit le modèle non linéaire : $y_t = f(X, a) + \varepsilon_t$ où X est la matrice des observations des variables explicatives (de dimension $n, k + 1$) et a est le vecteur (de dimension $k + 1$) des paramètres à estimer.

Sous les hypothèses classiques concernant ε_t , l'estimateur des moindres carrés est la valeur de a qui minimise la somme des carrés des résidus :

$$S(a) = e'e = [y_t - f(X, a)]'[y_t - f(X, a)]$$

Nous avons $k + 1$ conditions du premier ordre ($\partial S / \partial a = 0$), soit :

$$\frac{\partial S}{\partial a} = -2 \frac{\partial f(X, a)}{\partial a} [y_t - f(X, a)] = 0$$

$$\text{Avec } \frac{\partial f(X, a)}{\partial a} = Z(a) = \begin{bmatrix} \frac{\partial f(x_1, a)}{\partial a_0} & \cdots & \frac{\partial f(x_1, a)}{\partial a_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(x_n, a)}{\partial a_0} & \cdots & \frac{\partial f(x_n, a)}{\partial a_k} \end{bmatrix}$$

Soit $Z(a^1)$ cette matrice calculée pour les valeurs particulières de $a (a = a^1)$. Puis, à l'aide d'un développement limité de Taylor au voisinage de a^1 , nous pouvons approximer la t -ième observation

$$f(x_t, a) \simeq f(x_t, a^1) + \left[\frac{\partial f(x_t, a)}{\partial a_0} \bigg|_{a=a^1} \cdots \frac{\partial f(x_t, a)}{\partial a_k} \bigg|_{a=a^1} \right] (a - a^1)$$

Ou encore, en notation matricielle :

$$f(X, a) \simeq f(X, a^1) + Z(a^1)(a - a^1)$$

$$\text{Soit } y = f(X, a^1) + Z(a^1)(a - a^1) + \varepsilon$$

$$\text{ou encore : } y = f(X, a^1) + Z(a^1)a - Z(a^1)a^1 + \varepsilon$$

En posant $\bar{y}(a^1) = y - f(X, a^1) + Z(a^1)a^1$, nous pouvons alors construire un pseudo-modèle linéaire qui est égal à :

$$\bar{y}(a^1) = Z(a^1)a + \varepsilon$$

L'estimateur des MCO de ce modèle linéaire est donné par :

$$\begin{aligned} a^2 &= [Z(a^1)'Z(a^1)]^{-1}Z(a^1)'\bar{y}(a^1) \\ &= a^1 + [Z(a^1)'Z(a^1)]^{-1}Z(a^1)'[y - f(X, a^1)] \end{aligned}$$

Ce qui donne $k + 1$ nouvelles valeurs pour le vecteur $a = a^2$. Le processus est arrêté à la p -ième itération, lorsque nous observons une relative stabilité des coefficients ainsi estimés : $\hat{a} = a^p \simeq a^{p-1}$.

Cette méthode assez complexe connaît quelque raffinement par l'utilisation d'un développement limité de Taylor du second ordre (*algorithme de Newton-Raphson*).

Les logiciels d'économétrie (Eviews, RATS, ...) ainsi que les tableurs – à l'aide de la fonction *SOLVEUR* – permettent, très simplement, d'utiliser ces méthodes d'estimation des modèles non linéaires.

Néanmoins, il convient d'insister sur le fait qu'elles ne sont efficaces que si nous sommes capables de fournir des valeurs initiales compatibles avec la spécification du modèle et les données.

B. Exemples d'application

Exercice n° 2

↓ fichier C6EX2

Estimation des paramètres d'un modèle de diffusion Logistique et de Gompertz

Le tableau 3 présente les ventes cumulées (en milliers d'unités) d'un CD audio d'un artiste à succès sur 19 semaines.

Tableau 3 – Extrait des ventes cumulées sur 19 semaines (milliers d'unités)

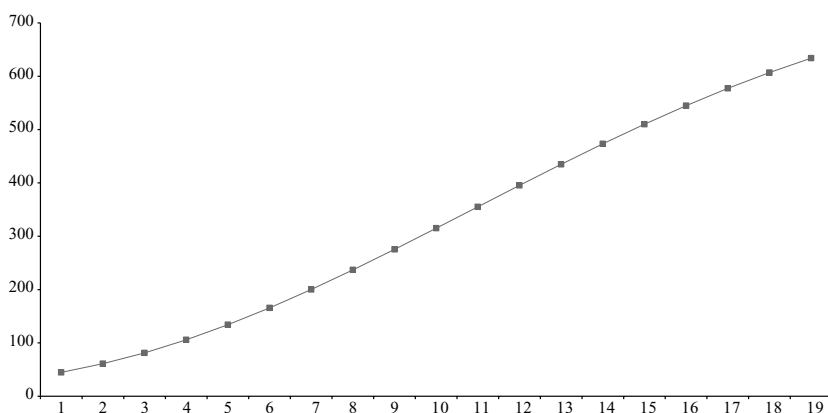
Semaines	Ventes
1	44,7
2	61,0
...	...
18	607,0
19	633,9

On demande :

- 1) D'estimer les paramètres d'un modèle de diffusion de type Logistique.
- 2) D'estimer les paramètres d'un modèle de diffusion de type Gompertz.

Solution

Le graphique 2 présente l'évolution des ventes cumulées en fonction des semaines. Ce graphique peut suggérer une évolution selon un modèle de diffusion, nous serions alors à proximité du point d'inflexion.



Graphique 2 – Ventes cumulées

1) Estimation d'un modèle de diffusion de type Logistique (Eviews).

Valeurs initiales compatibles

PARAM 1 600. 2 5. 3 0.5

equation NLS ventes=C(1)/(1+c(2)*c(3)^@trend)

Résultats (Eviews) :

Dependent Variable: VENTES				
Sample: 1 19				
Included observations: 19				
Convergence achieved after 8 iterations				
VENTES=C(1)/(1+C(2)*C(3)^TEND)				
	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	718.0033	11.87762	60.45008	0.0000
C(2)	14.77212	0.536144	27.55254	0.0000
C(3)	0.784909	0.004580	171.3614	0.0000

R-squared 0.999009

Le modèle estimé s'écrit : $y_t = \frac{y_{\max}}{1 + br^t} = \frac{718}{1 + 14,77 \times 0,785^t}$

2) Estimation d'un modèle de diffusion de type Gompertz (RATS).

* Régression non linéaire

nonlin b r a

← Définition des paramètres à estimer

frml gompertz = exp(b*r**(t)+a)

← Formulation du modèle

```
*Initialisation des paramètres
eval a = 5.
eval b = -3.5
eval r = 0.5
nlls(frml=gompertz,trace) taux    ← Estimation des coefficients
```

Résultats

```
Dependent Variable TAUX - Estimation by Nonlinear Least Squares
Iterations Taken      15
Usable Observations   19      Degrees of Freedom      16
Centered R**2         0.999966    R Bar **2         0.999962
Uncentered R**2       0.999991    T x R**2         19.000
Variable              Coeff      Std Error      T-Stat      Signif
*****
1.  B                  -3.464637364    0.010317751    -335.79385    0.00000000
2.  R                  0.889725397    0.000737070    1207.11145    0.00000000
3.  A                  6.832158116    0.006144379    1111.93637    0.00000000
```

Soit le modèle : $y_i = e^{br^i+a} = e^{(-3,46 \times 0,88^i + 6,83)}$; le seuil de saturation, selon un modèle de Gompertz, est donc de $e^{6,83} = 927$.

Exercice n° 3

↓ fichier C6EX1

Estimation des paramètres d'une fonction de production de type CES

Les fonctions de production à l'élasticité de substitution constante, appelées CES¹ (Constant Elasticity of Substitution), sont de la forme :

$$Q = \beta_1 [\beta_2 K^{\beta_3} + (1 - \beta_2) L^{\beta_3}]^{\beta_4} \text{ avec } \beta_3 \text{ et } \beta_4 < 0$$

où $1/(1 - \beta_3)$ représente l'élasticité de substitution².

Nous pouvons aussi l'écrire de la manière suivante :

$$Q = \beta_1 [\beta_2 K^{\beta_3} + (1 - \beta_2) L^{\beta_3}]^{\nu/\beta_3}$$

où ν représente alors les rendements d'échelle.

En reprenant les données du tableau 1, on demande :

1. La fonction de production à l'élasticité de substitution constante (CES) a été introduite par Arrow K.J. *et al.*, 1961.
2. L'élasticité de substitution mesure l'incidence de la variation relative des prix des facteurs sur la combinaison productive (le rapport K/L).

- 1) de transformer cette fonction de production par un passage aux logarithmes ;
- 2) d'estimer les paramètres du modèle ;
- 3) de comparer les résultats avec ceux obtenus à l'exercice 1.

Solution

1) Transformation du modèle

$$Q = \beta_1 [\beta_2 K^{\beta_3} + (1 - \beta_2)] L^{\beta_3}]^{\beta_4}$$

$$\text{Log}(Q) = \text{Log}(\beta_1) + \beta_4 \text{Log} [\beta_2 K^{\beta_3} + (1 - \beta_2)] L^{\beta_3}]$$

Nous observons que cette fonction de production n'est pas linéarisable (contrairement à la fonction de production de type Cobb-Douglas). Il convient donc d'utiliser une technique d'estimation des modèles non linéaires.

2) Estimation des paramètres

Les instructions du logiciel Eviews sont les suivantes :

PARAM 1 12. 2 -1. 3 0.5 4 -0.5 ← Initialisation des 4 paramètres

NLS Q = C(1)*(C(3)*K^C(4))+(1-C(3))*L^C(4))^C(2) ← Équation

Résultats

```

NLS // Dependent Variable is Q
SMPL range: 1 - 25
Number of observations: 25
Q=C(1)*(C(3)*K^C(4))+(1.-C(3))*L^C(4))^C(2)
Convergence achieved after 13 iterations

```

	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C(1)	11.213534	1.4270360	7.8579193	0.0000
C(3)	0.4049360	0.1434756	2.8223333	0.0102
C(4)	-0.5970243	0.3020160	-1.9767972	0.0613
C(2)	-1.3853333	0.7651230	-1.8106021	0.0845
R-squared	0.953721	Mean of dependent var	84.56080	
Adjusted R-squared	0.947110	S.D. of dependent var	29.70348	
S.E. of regression	6.831156	Sum of squared resid	979.9586	
Log likelihood	-81.33139	F-statistic	144.2573	
Durbin-Watson stat	2.555634	Prob(F-statistic)	0.000000	

La fonction de production estimée s'écrit :

$$\begin{aligned}
 Q &= \beta_1 [\beta_2 K^{\beta_3} + (1 - \beta_2) L^{\beta_3}]^{\beta_4} \\
 &= 11,21 [0,404 K^{-0,597} + 0,594 L^{-0,597}]^{-1,38}
 \end{aligned}$$

3) Comparaison des résultats entre une fonction de production Cobb-Douglas et une fonction de production de type CES

Lors de l'exercice 1, nous avons estimé les rendements d'échelle de la fonction Cobb-Douglas ($\hat{\alpha}_1 + \hat{\alpha}_2 = 0,89$) qui sont significativement décroissants.

Pour la fonction de type CES, nous avons :

$$\hat{v} = \hat{\beta}_3 \times \hat{\beta}_4 = -1,38 \times -0,597 = 0,82 < 1$$

Les rendements d'échelle sont décroissants, et l'élasticité de substitution constante est égale à :

$$1/(1 - \hat{\beta}_3) = 0,42$$

7. Les modèles à décalages temporels

Jusqu'à maintenant, nous avons spécifié des modèles où toutes les variables sont exprimées à la même période. Cependant, la théorie économique postule couramment – non pas des effets synchrones – mais des effets retardés.

Ces modèles à décalages peuvent inclure comme variables retardées aussi bien la variable endogène que les variables exogènes. Nous étudierons, tout d'abord en I., les modèles autorégressifs dont la variable retardée est la variable endogène. Puis, en II., une autre classe de modèle est envisagée : les modèles à retards échelonnés. Dans cette spécification, les variables exogènes apparaissent à plusieurs décalages.

Enfin, en III., nous présentons deux exemples de modèles dynamiques.

I. Les modèles linéaires autorégressifs

A. Formulation générale

Dans ce type de modèles temporels, la variable endogène y_t dépend :

- de k variables exogènes $x_{1t}, x_{2t}, \dots, x_{kt}$ à l'époque t ;
- des valeurs que prend la variable y_t pendant les périodes précédentes, $y_{t-1}, y_{t-2}, \dots, y_{t-h}$.

Soit la formulation :

$$y_t = b_1 y_{t-1} + b_2 y_{t-2} + \dots + b_h y_{t-h} + a_0 + a_1 x_{1t} + a_2 x_{2t} + \dots + a_k x_{kt} + \varepsilon_t$$

ou encore : $y_t = \sum_{j=1}^h b_j y_{t-j} + \frac{X}{(n, k+1)} \frac{a}{(k+1, 1)} + \varepsilon_t$

Dans ce modèle, l'hypothèse H5, d'indépendance entre les variables explicatives et l'erreur, n'est plus satisfaite car les variables $y_{t-1}, y_{t-2}, \dots, y_{t-h}$ qui dépendent de $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-h}$ sont aléatoires puisque y_{t+1} est fonction de y_t qui dépend de ε_t , soit : $E(y_{t+1} \varepsilon_t) \neq 0$.

Nous remarquons que si les variables exogènes x_{it} et les erreurs ε_t sont fixées, les variables endogènes sont solutions de l'équation de récurrence :

$$y_t = b_1 y_{t-1} + b_2 y_{t-2} + \dots + b_h y_{t-h} + S_t$$

avec $S_t = Xa + \varepsilon_t$.

Il est démontré que la solution générale de cette équation de récurrence est explosive si une des h racines de son équation caractéristique¹ a son module supérieur à 1.

Or nous devons considérer les processus explosifs – que ce soit d'une façon monotone ou oscillatoire – comme des phénomènes rares en économie (les variables sont limitées généralement dans leur croissance), c'est pourquoi il convient d'ajouter l'hypothèse de stabilité du processus.

Cas particulier : le modèle autorégressif d'ordre 1

Le modèle général spécifié ci-dessus est rarement utilisé, le plus souvent nous nous limitons à des processus autorégressifs d'ordre 1², de la forme :

$$y_t = b_1 y_{t-1} + a_0 + a_1 x_{1t} + a_2 x_{2t} + \dots + a_k x_{kt} + \varepsilon_t \quad [1]$$

Ce modèle est stable si $|b_1| < 1$ et explosif si $|b_1| > 1$.

B. Test d'autocorrélation et méthodes d'estimation

La méthode d'estimation adéquate dépend d'une éventuelle dépendance des erreurs ; or, dans le cas d'un modèle autorégressif, le test de Durbin et Watson a une puissance réduite et est biaisé. C'est pourquoi il convient d'utiliser une autre statistique, celle du « h » de Durbin.

1) Test d'autocorrélation des erreurs

Dans le cas d'un modèle autorégressif d'ordre 1, le test classique de Durbin et Watson sous-estime le risque d'autocorrélation, un nouveau test d'autocorrélation des erreurs doit alors être utilisé³. La statistique utilisée est la suivante :

1. L'équation caractéristique s'écrit $r^h - b_1 r^{h-1} - b_2 r^{h-2} - \dots - b_h = 0$.

2. La variable y_t dépend néanmoins des valeurs passées $y_{t-1}, y_{t-2}, \dots, y_0$.

3. Durbin J., 1970.

$$h = \hat{\rho} \sqrt{\frac{n}{1 - n\hat{\sigma}_{b1}^2}} \quad [2]$$

avec $\hat{\rho} = 1 - dw/2$ (dw est la statistique de Durbin et Watson calculée sur la modèle [1]) ;

n = nombre d'observations ;

$\hat{\sigma}_{b1}^2$ = variance estimée du coefficient b_1 à partir du modèle [1].

Cette statistique « h » est distribuée de manière asymptotique comme une variable normale centrée réduite. Ainsi, il y a équivalence entre les deux tests d'hypothèses suivants :

$$H0 : \rho = 0 \quad H0 : h = 0$$

$$H1 : \rho \neq 0 \quad H1 : h \neq 0$$

Si $|h| \leq t^{\alpha/2}$, nous acceptons l'hypothèse $H0$ d'indépendance des erreurs.

($t^{\alpha/2}$ = valeur issue de la loi normale pour un test bilatéral au seuil de α .)

Nous remarquons que si $n\hat{\sigma}_{b1}^2 \geq 1$, la statistique « h » ne peut pas être calculée ; dans ce cas, nous pouvons utiliser la statistique de Durbin et Watson en incluant la zone de doute dans la zone d'autocorrélation des erreurs.

2) Estimation si les erreurs ne sont pas corrélées

Dans le cas d'absence d'autocorrélation des erreurs, les estimateurs des MCO appliqués au modèle [1] convergent asymptotiquement vers les valeurs vraies des paramètres et ont une variance minimale parmi tous les estimateurs convergents.

Cependant, pour les petits échantillons, lors de l'estimation d'un modèle autorégressif d'ordre h , les résultats asymptotiques sont alors très approximatifs, car le nombre de périodes d'estimation est de $n - h$. De plus, les problèmes de colinéarité entre les variables explicatives décalées interdisent pratiquement d'utiliser les MCO.

En résumé, pour l'estimation des modèles autorégressifs à erreurs indépendantes, l'application de la méthode classique des MCO est licite si le nombre d'observations est suffisant (souvent dans la pratique $n > 15$).

3) Estimation en cas d'autocorrélation des erreurs

Le modèle s'écrit alors :

$$y_t = b_1 y_{t-1} + a_0 + a_1 x_{1t} + a_2 x_{2t} + \dots + a_k x_{kt} + \varepsilon_t \quad [3]$$

avec $\varepsilon_t = \rho \varepsilon_{t-1} + v_t$

Il s'agit d'un modèle à autocorrélation des erreurs, nous avons montré lors du chapitre 5 que la transformation en quasi-différences premières peut lever l'autocorrélation des erreurs d'ordre 1.

Nous pouvons citer différentes méthodes d'estimation.

a) Régression sur les différences premières

On procède à deux régressions par la méthode des MCO : la première sur le modèle spécifié avec des variables non transformées et la seconde sur le modèle spécifié avec des variables transformées en différences premières.

Si les deux résultats d'estimation sont proches, nous pouvons conclure que les paramètres du modèle sont suffisamment bien déterminés sans rechercher la liaison éventuelle des erreurs.

b) Correction de l'autocorrélation des erreurs par régression sur les quasi-différences premières

Par une des méthodes présentées au chapitre 5, nous procédons à la correction de l'autocorrélation des erreurs : estimation directe de ρ , Hildreth-Lu, Cochrane-Orcutt.

c) D'autres méthodes

Nous pouvons aussi citer :

- la méthode des variables instrumentales (cf. chapitre 5), où nous retenons comme variable instrumentale la variable y_{t-2} ;
- la méthode du maximum de vraisemblance.

La méthode des variables instrumentales résout le problème de convergence des estimateurs lorsque nous sommes en présence d'erreurs à moyenne mobile de type :

$$\varepsilon_t = v_t - \theta_1 v_{t-1} - \theta_2 v_{t-2} - \dots - \theta_p v_{t-p}$$

Dans le cas d'erreurs autocorrélées, nous devons utiliser une des méthodes d'estimation robuste présentées au chapitre 5.

4) Prédiction dans un modèle autorégressif

La prédiction de la variable endogène y , à l'horizon h , est donnée par :

$$\hat{y}_{n+h} = \hat{b}_1 \hat{y}_{n+h-1} + \hat{a}_0 + \hat{a}_1 x_{1n+h} + \hat{a}_2 x_{2n+h} + \dots + \hat{a}_k x_{kn+h}$$

L'erreur qui affecte les valeurs successives de \hat{y}_{n+h-1} se cumule d'une période sur l'autre, et ainsi la variance de l'erreur de prédiction peut devenir très importante.

L'utilisation à des fins de prédiction du modèle autorégressif doit donc être limitée à quelques périodes.

Exercice n° 1

↓ fichier C7EX1

Estimation des paramètres d'un modèle autorégressif à erreurs liées

Un économètre désire tester la relation entre les prix officiels (PO) de la tonne de café et les prix réellement appliqués à l'exportation (PE) par les pays producteurs. Il se propose d'estimer la relation :

$$PO_t = a_0 + a_1 PO_{t-1} + a_2 PE_t + \varepsilon_t$$

dans laquelle le prix officiel est fonction de manière instantanée du prix observé et s'ajuste avec un retard d'un an au prix officiel.

Il dispose des données, sur 16 ans, présentées au tableau 1.

*Tableau 1 – Prix officiel et prix pratiqué
à l'exportation de la tonne de café (en dollars)*

t	PO	PE
1	455,0	615,0
2	500,0	665,0
3	555,0	725,0
4	611,0	795,0
5	672,0	870,0
6	748,5	970,0
7	846,0	1 095,0
8	954,5	1 235,0
9	1 090,0	1 415,0
10	1 243,5	1 615,0
11	1 390,0	1 795,0
12	1 559,0	2 015,0
13	1 781,0	2 315,0
14	2 046,5	2 660,0
15	2 311,0	2 990,0
16	2 551,0	3 280,0

On demande :

- 1) d'estimer la relation et de tester une éventuelle autocorrélation des erreurs ;
- 2) d'en corriger les effets, le cas échéant.

Solution

1) Les résultats de l'estimation sont les suivants :

Dependent Variable : PO

Included observations : 15 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7.078673	1.851077	-3.824084	0.0024
PO(-1)	0.224016	0.037090	6.039727	0.0001
PE	0.622340	0.025520	24.38636	0.0000
R-squared	0.999980	Mean dependent var	1257.267	
Adjusted R-squared	0.999976	S.D. dependent var	663.5723	
S.E. of regression	3.236240	Akaike info criterion	5.363558	
Sum squared resid	125.6790	Schwarz criterion	5.505168	
Log likelihood	-37.22668	F-statistic	294296.0	
Durbin-Watson stat	0.625901	Prob(F-statistic)	0.000000	

L'équation est estimé sur 15 périodes ($n - 1$) car dans le modèle figure une variable retardée d'une période.

La statistique de Durbin et Watson laisse augurer d'une autocorrélation des erreurs, ce qui est confirmé par le « h » de Durbin [2] :

$$h = \hat{\rho} \sqrt{\frac{n}{1 - n\hat{\sigma}_{a1}^2}} = 0,687 \sqrt{\frac{15}{1 - 15 \times 0,0013}} = 2,68 > t^{0,05} = 1,96$$

Nous rejetons l'hypothèse H_0 , soit $h \neq 0 \rightarrow \rho \neq 0$.

2) Nous allons tout d'abord comparer les estimations obtenues à partir du modèle brut à celles obtenues à partir du modèle en différences premières.

Nous procédons à l'estimation du modèle en différences premières sur 14 ans car nous perdons de nouveau une observation lors du calcul des différences premières de PO_{t-1} .

Les résultats sont les suivants :

Dependent Variable : D(PO)

Included observations : 14 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.891649	1.298314	2.227234	0.0478
D(PO(-1))	0.240680	0.024114	9.981022	0.0000
D(PE)	0.598017	0.018553	32.23277	0.0000
R-squared	0.999299	Mean dependent var	146.5000	
Adjusted R-squared	0.999172	S.D. dependent var	76.37030	
S.E. of regression	2.197504	Akaike info criterion	4.599930	
Sum squared resid	53.11925	Schwarz criterion	4.736871	
Log likelihood	-29.19951	F-statistic	7845.105	
Durbin-Watson stat	1.781510	Prob(F-statistic)	0.000000	

Nous observons que les différences entre les coefficients de régression sont assez faibles pour a_1 (0,224 et 0,240) comme pour a_2 (0,622 et 0,598); nous pouvons considérer les résultats obtenus à la première régression comme valides.

Cependant, à titre pédagogique, nous allons utiliser la méthode de correction d'auto-corrélation des erreurs.

Le modèle estimé, selon la méthode de Hildreth-Lu, est alors :

$$PO_t = -2,765 + 0,234 \times PO_{t-1} + 0,6135 \times PE_t + e_t$$

(1,67) (15,5) (57,9)

$$e_t = 0,418 e_{t-1} + v_t$$

$$R^2 = 0,99; n = 14; DW = 2,08; (.) = t \text{ de Student}$$

Les prix officiels de la période t sont donc plus influencés par les prix réellement pratiqués en t que par les prix officiels de l'année précédente.

II. Les modèles à retards échelonnés

A. Formulation générale

Dans certaines spécifications de modèles économétriques temporels, nous pouvons postuler que la variable endogène dépend des valeurs prises par une variable exogène à des époques antérieures, tel que :

$$y_t = b_0 + a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2} + \dots + a_h x_{t-h} + \varepsilon_t = \sum_{j=0}^h a_j x_{t-j} + b_0 + \varepsilon_t \quad [6]$$

En général, l'effet de la variable explicative s'estompe avec le temps :

$$a_0 > a_1 > a_2 > \dots > a_h$$

Ce modèle [6] peut se simplifier dans son écriture en utilisant un opérateur décalage D défini par : $Dx_t = x_{t-1}$ et en général : $D^i x_t = x_{t-i}$, soit :

$$y_t = \sum_{j=0}^h a_j x_{t-j} + b_0 + \varepsilon_t = \sum_{j=0}^h a_j D^j x_t + b_0 + \varepsilon_t = \left[\sum_{j=0}^h a_j D^j \right] x_t + b_0 + \varepsilon_t$$

$$y_t = A(D)x_t + b_0 + \varepsilon_t \quad [7]$$

où $A(D)$ est un polynôme de degré h , tel que :

$$A(D) = a_0 + a_1 D^1 + a_2 D^2 + \dots + a_h D^h$$

Le nombre de retards, h , peut être fini ou infini. Cependant, la somme des coefficients a_j tend vers une limite finie, sinon y_t serait un processus explosif.

Le polynôme $A(1) = a_0 + a_1 + a_2 + \dots + a_h$ permet de mesurer, à long terme, l'impact de la modification de la variable explicative x_t d'une quantité Δx sur la valeur de y_t . En effet, les coefficients a_j représentent les multiplicateurs instantanés et leur somme le multiplicateur cumulé.

Retard moyen

La notion de retard moyen permet d'appréhender le laps de temps qui correspond à la valeur moyenne des coefficients a_j . Il est égal à la moyenne pondérée des coefficients, soit :

$$\overline{D} = \frac{\sum_{j=0}^h j a_j}{\sum_{j=0}^h a_j} = \frac{a_1 + 2 a_2 + \dots + h a_h}{a_0 + a_1 + a_2 + \dots + a_h} = \frac{A'(1)}{A(1)} \quad [8]$$

L'estimation des paramètres du modèle [6] soulève deux types de difficultés :

- en pratique, la valeur h qui mesure l'importance du nombre de retards est rarement connue, même si nous en connaissons l'ordre de grandeur ;
- le second problème résulte de la colinéarité entre les variables exogènes décalées. En effet, lorsque le nombre de retards est important, la colinéarité entre les variables explicatives décalées risque d'entraîner une imprécision dans l'estimation des coefficients. C'est pourquoi, le plus souvent, nous émettons des hypothèses sur la forme de retards.

Ces deux points sont abordés ci-dessous.

B. Détermination du nombre de retards

Lorsque la valeur h du nombre de retards du modèle [6] est inconnue, il existe des critères statistiques permettant de la déterminer¹.

1) Test de Fisher

Le test le plus naturel est celui de Fisher dans lequel nous testons l'hypothèse de la nullité des coefficients de régression pour les retards supérieurs à h .

La formulation des hypothèses est la suivante, lorsque l'on teste, d'une manière descendante, une valeur de h comprise entre 0 et M : $0 < h < M$.

1. Il s'agit de déterminer quelle est la période maximum d'influence de la série explicative. Il peut très bien arriver, c'est même souvent le cas, que des coefficients de rang inférieur au décalage h ne soit pas significativement différents de 0.

$$\begin{array}{ll}
H0^1 : h = M - 1 \rightarrow a_M = 0 & H1^1 : h = M \rightarrow a_M \neq 0 \\
H0^2 : h = M - 2 \rightarrow a_{M-1} = 0 & H1^2 : h = M - 1 \rightarrow a_{M-1} \neq 0 \\
\vdots & \vdots \\
H0^i : h = M - i \rightarrow a_{M-i+1} = 0 & H1^i : h = M - i + 1 \rightarrow a_{M-i+1} \neq 0
\end{array}$$

Chacune de ces hypothèses fait l'objet du classique test de Fisher, soit :

$$F_1^* = \frac{(SCR_{M-i} - SCR_{M-i+1})/1}{SCR_{M-i+1}/(n - M + i - 3)} \quad [9]$$

que l'on compare au F lu dans la table à 1 et $(n - M + i - 3)$ degrés¹ de liberté.

Dès que, pour un seuil donné, le F empirique est supérieur au F lu, nous rejetons alors l'hypothèse $H0^i$ et la procédure est terminée. La valeur du retard est égale à $M - i + 1 : h = M - i + 1$.

Afin de pouvoir procéder à ce test dans ses conditions d'application, la Somme des Carrés Totaux doit rester constante d'une estimation à l'autre. Cela oblige donc à estimer les différents modèles avec un nombre d'observations identique correspondant donc au nombre d'observations réellement disponible pour le décalage le plus important, chaque décalage entraînant la perte d'une donnée.

2) Critère de Akaike (AIC)²

Une autre méthode consiste à retenir comme valeur de h celle qui minimise la fonction de Akaike qui est donnée par :

$$AIC(h) = \text{Ln}\left(\frac{SCR_h}{n}\right) + \frac{2h}{n} \quad [10]$$

avec SCR_h = Sommes des Carrés des Résidus pour le modèle à h retards

n = nombre d'observations disponible
(chaque retard entraîne la perte d'une observation)

Ln = logarithme népérien

3) Critère de Schwarz (SC)³

Enfin, une méthode très proche de la précédente consiste à retenir la valeur de h qui minimise la fonction de Schwarz :

$$SC(h) = \text{Ln}\left(\frac{SCR_h}{n}\right) + \frac{h \text{ Ln } n}{n} \quad [11]$$

1. Le degré de liberté du dénominateur est égal au nombre d'observations (n) moins le nombre de paramètres à estimer dans le modèle [6]. Soit pour $h = M - i + 1$ et en tenant compte du terme constant et du coefficient de régression pour la variable x à 0 retard :

$$n - (M - i + 1) - 1 - 1 = n - M + i - 3.$$

2. Akaike H., 1973 ; 1974.

3. Schwarz G., 1978.

Exercice n° 2

↓ fichier C7EX2

Détermination du nombre de retards dans un modèle à retards échelonnés

La théorie économique postule que les dépenses d'investissement (notées y_t) peuvent être expliquées par les profits passés (notés x_t). Le modèle prend la forme d'un modèle à retards échelonnés tel le modèle [6].

Nous disposons d'un extrait des données trimestrielles du tableau 2 concernant l'industrie chimique française.

N'étant pas certains de la spécification exacte du modèle, nous désirons tout d'abord rechercher le nombre de décalages trimestriels qui semblent avoir un effet sur les dépenses d'investissement. Puis après avoir déterminé le nombre de retards, nous calculerons le délai moyen.

Tableau 2 – Dépenses d'investissement et profit
(en millions d'euros)

Trimestre	y_t	x_t
1	2 072	1 660
2	2 077	1 926
...
43	2 840	3 449
44	2 837	3 764

Solution

Nous procédons aux estimations des modèles sur 34 trimestres, puisque le décalage maximum testé est de 10, on élimine les 10 premières observations.

Nous calculons les trois critères : Akaike, Schwarz et Fisher à l'aide d'un programme Eviews en téléchargement (C7EX2).

Les résultats sont consignés dans le tableau 3.

Tableau 3 – Résultats de la recherche du nombre de décalages optimal

Décalage	Akaike	Schwarz	Fisher
0	11,96	11,96	16,82
1	11,50	11,54	18,91
2	11,04	11,12	19,18
3	10,55	10,68	10,89
4	10,25	10,42	8,59
5	9,99	10,21	6,86
6	9,84	10,10	1,09
7	9,88	10,19	0,13
8	9,96	10,31	0,77
9	10,03	10,43	0,25
10	10,10	10,55	–

Nous observons immédiatement que les minima des deux critères ($AIC(h)$ et $SC(h)$) sont situés sur la ligne 6, correspondant donc à 6 décalages. La première valeur du Fisher empirique qui est significative – en partant du retard 10 – est celle de la ligne 5. Elle est égale à 6,86, ce qui vient confirmer une influence de la variable explicative allant jusqu’à un décalage de 6 périodes.

Nous allons détailler le calcul des critères¹ de Akaike et Schwarz pour le retard 6. Le résultat de l’estimation pour 6 retards est le suivant :

- Variable à expliquer : Y
- $n = 38$

Variable	Coefficient	Écart type	t de Student	Prob.
C	501,5414	154,8486	3,238915	0,0029
X	– 0,011389	0,081532	– 0,139687	0,8898
X(– 1)	0,061265	0,124906	0,490487	0,6274
X(– 2)	0,227569	0,119635	1,902194	0,0668
X(– 3)	0,167932	0,112997	1,486158	0,1477
X(– 4)	0,118734	0,127454	0,931580	0,3590
X(– 5)	0,000169	0,136907	0,001235	0,9990
X(– 6)	0,237174	0,084065	2,821310	0,0084
$R^2 = 0,92$	0,921953	Somme des carrés des résidus =		

D’après [10], nous avons :

$$AIC(6) = \text{Ln} \left[\frac{521116}{38} \right] + \frac{2 \times 6}{38} = 9,84$$

Et d’après [11] :

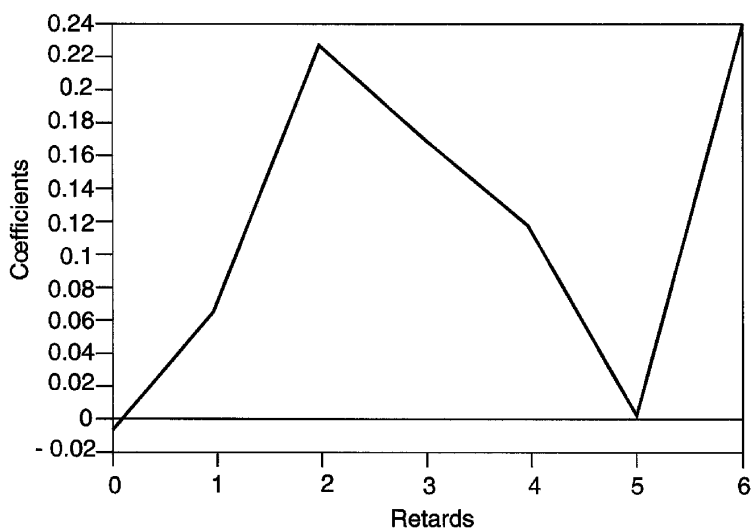
$$SC(6) = \text{Ln} \left[\frac{521116}{38} \right] + \frac{6 \times \text{Ln}(38)}{38} = 10,10$$

Nous remarquons que la probabilité critique du coefficient de la variable X(–6) est largement inférieure à 0,05 ; le coefficient est donc significativement différent de 0. Ceci est évidemment conforme aux résultats précédents.

Les trois critères sont (heureusement) cohérents, le modèle à retards échelonnés comporte 6 retards : l’investissement des entreprises de ce secteur est fonction des profits réalisés sur les six derniers trimestres, soit un an et demi. Il convient de noter que seul le coefficient du sixième retard est significativement différent de 0.

Le graphique 1 illustre la structure de pondération des retards pour le modèle ainsi estimé.

1. Nous ne revenons pas sur le test de Fisher qui a été déjà abondamment illustré.



Graphique 1 – Structure de la pondération des retards

Le retard moyen est égal d'après [8] :

$$\bar{D} = \frac{\sum_{j=0}^6 j\hat{a}_j}{\sum_{j=0}^6 \hat{a}_j} = \frac{0,0612 + 2 \times 0,227 + 3 \times 0,168 + \dots + 6 \times 0,237}{-0,011 + 0,227 + 0,168 + \dots + 0,237} = 3,64$$

Le délai moyen de réaction est 3,64 trimestres, soit presque une année.

C. D Distribution finie des retards

Considérant que nous avons déterminé la longueur des retards h , il convient d'estimer les coefficients a_i du modèle. Cependant, l'utilisation des MCO peut se révéler impossible du fait de la multicolinéarité entre les variables explicatives décalées. C'est pourquoi, dans ce contexte, nous allons présenter une méthode permettant, moyennant des hypothèses sur la forme des coefficients, de réduire le nombre de paramètres à estimer.

Fonction polynomiale des retards (retards d'Almon¹)

Nous pouvons imaginer une infinité de formes concernant la distribution des coefficients a_i (graphique 2). La méthode des retards d'Almon est très utilisée, car elle permet de dégager des profils de retards s'adaptant à des représentations différentes.

1. Almon S., 1965.

Cette technique consiste à imposer aux coefficients d'appartenir à un même polynôme de degré q (graphique 3), tel que :

$$a_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2 + \dots + \alpha_q i^q = \sum_{j=0}^q \alpha_j i^j \tag{12}$$

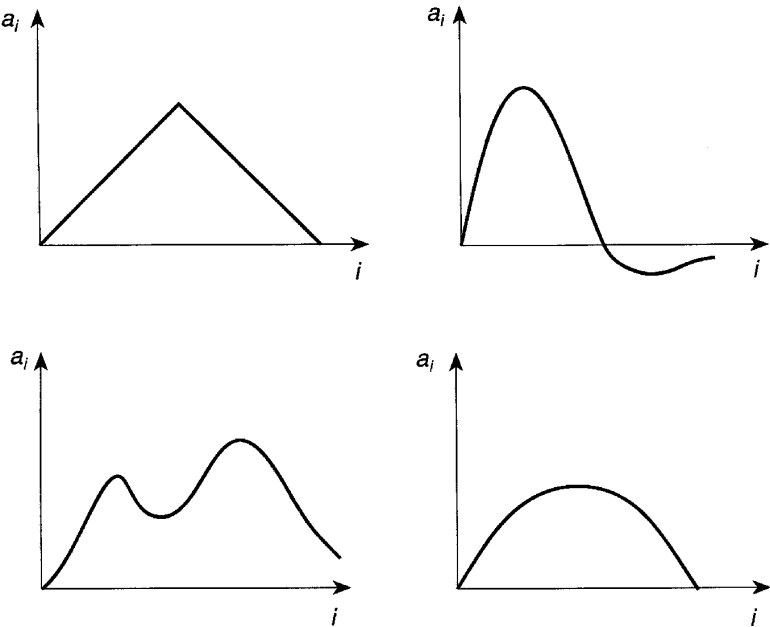
Par exemple, pour un polynôme de degré 2 ($q = 2$), nous avons la séquence des coefficients :

$$\begin{aligned} a_0 &= \alpha_0 \\ a_1 &= \alpha_0 + \alpha_1 + \alpha_2 \\ a_2 &= \alpha_0 + 2\alpha_1 + 4\alpha_2 \\ a_3 &= \alpha_0 + 3\alpha_1 + 9\alpha_2 \end{aligned}$$

ou encore

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 2^2 \\ 1 & 3 & 3^2 \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix}$$

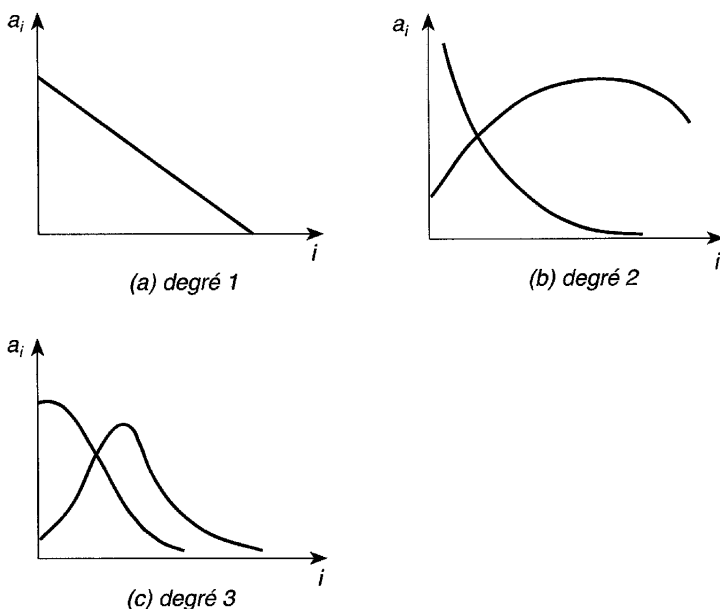
$a = H \alpha$



Graphique 2 – Exemples de structures de retards

Nous pouvons généraliser la formule précédente pour h retards et un polynôme de degré q .

$$\begin{matrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_h \end{bmatrix} & = & \begin{bmatrix} 1 & 0 & 0 & \cdot & \cdot & 0 \\ 1 & 1 & 1 & \cdot & \cdot & 1 \\ 1 & 2 & 2^2 & \cdot & \cdot & 2^q \\ 1 & 3 & 3^2 & \cdot & \cdot & 3^q \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & h & h^2 & \cdot & \cdot & h^q \end{bmatrix} & \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_q \end{bmatrix} \\ (h+1, 1) & & (h+1, q+1) & & (q+1, 1) \end{matrix} \tag{13}$$



Graphique 3 – Exemples de retards polynomiaux

Le modèle initial $Y = Xa + \varepsilon$ peut alors s'écrire $Y = XH\alpha + \varepsilon = Z\alpha + \varepsilon$ dans lequel $XH = Z$ est la matrice des observations des « nouvelles » variables explicatives.

L'estimation par la méthode des MCO des $q + 1$ coefficients α , $(\hat{\alpha}_0, \hat{\alpha}_1, \dots, \hat{\alpha}_q)$, permet d'obtenir, à partir des relations [12], les estimations des $h + 1$ paramètres de a $(\hat{a}_0, \hat{a}_1, \dots, \hat{a}_h)$.

Dans cette méthode, nous sommes confrontés à un double choix :

- la valeur h du nombre de retards (nous avons la réponse, cf. 2),
- le degré du polynôme.

Le degré du polynôme q , $q < h$, peut être déterminé à l'aide d'un test portant sur la significativité, par rapport à 0, du coefficient (α_q) de la dernière « nouvelle » variable explicative. En partant d'une valeur $q = h - 1$, on teste la significativité du coefficient α du terme le plus élevé à l'aide d'un t de Student et on réduit le degré du polynôme jusqu'à ce que ce coefficient soit significatif.

Les logiciels d'économétrie traitant des estimateurs d'Almon offrent la possibilité d'introduire des contraintes supplémentaires concernant les valeurs des coefficients a_{-1} et/ou a_{h+1} afin qu'elles soient nulles. Cela se justifie par le fait que a_{-1} est le coefficient théorique de la variable x_{t+1} qui est sans influence sur y_t .

Exercice n° 3

↓ fichier C7EX2

Estimation des coefficients à l'aide d'un polynôme (méthode d'Almon)

En reprenant le modèle – Investissement/Profit – de l'exercice 2, on demande d'estimer les coefficients par la méthode des retards d'Almon.

Solution

Le corrigé de cet exercice est réalisé à l'aide de Eviews, qui permet le recours direct à la technique du polynôme d'Almon.

L'instruction est la suivante : *LS Y C PDL (X, 6, 4, 1)*

PDL : Polynomial Distributed Lag

X = série explicative

$h = 6$: nombre de décalages (résultat de l'exercice 2) ;

$q = 4$: degré du polynôme (après un essai avec $q = 5$, le coefficient de la variable PDL5 n'est pas significativement différent de 0),

1 = indicateur de la contrainte : $a_{-1} = 0$ (cet indicateur vaut 0 si l'on ne désire aucune contrainte, 2 pour la contrainte $a_{h+1} = 0$ et 3 pour les deux contraintes simultanées).

Résultats

Dependent Variable : Y

Method: Least Squares

Included observations : 38 after adjusting endpoints

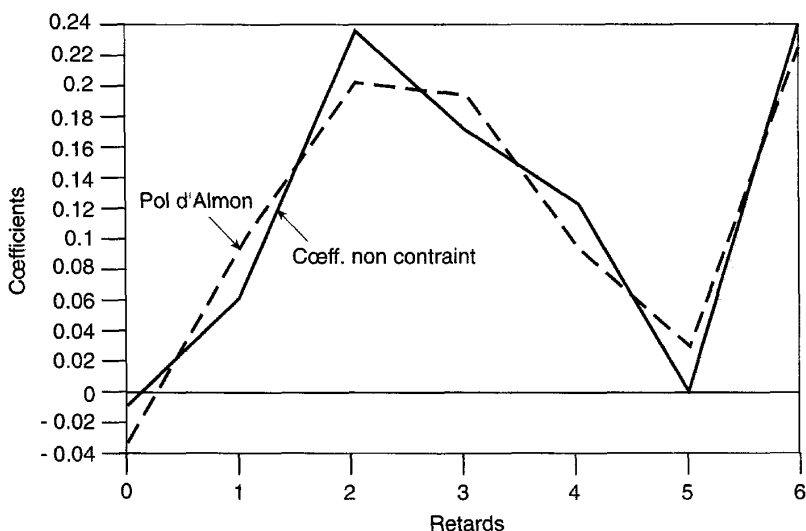
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	511.6833	143.4181	3.567774	0.0011
PDL01	-0.193452	0.120131	-1.610334	0.1169
PDL02	0.215794	0.103537	2.084228	0.0450
PDL03	-0.055823	0.025884	-2.156616	0.0384
PDL04	0.004228	0.001956	2.160957	0.0381
R-squared	0.921681	Mean dependent var	2567.553	
Adjusted R-squared	0.912187	S.D. dependent var	424.8027	
S.E. of regression	125.8826	Akaike info criterion	12.63066	
Sum squared resid	522932.5	Schwarz criterion	12.84613	
Log likelihood	-234.9825	F-statistic	97.08789	
Durbin-Watson stat	0.583162	Prob(F-statistic)	0.000000	

Lag Distribution of X		Lag	Coef	S.E.	T-Stat
*	:	0	-0.02925	0.04183	-0.69931
:	:	1	0.09733	0.01879	5.18007
:	:	2	0.19701	0.03741	5.26622
:	:	3	0.18851	0.02940	6.41132
:	:	4	0.09200	0.03462	2.65719
:	*	5	0.02912	0.05538	0.52584
:	:	6	0.22298	0.06084	3.66474
0		Sum	0.79770	0.05409	14.7484

Les coefficients du polynôme (α_i) sont représentés par les variables PDLi, ils sont (presque) tous significativement différents de 0. La présence d'une contrainte entraîne la perte d'un degré de polynôme.

Le calcul des coefficients et la structure de pondération apparaissent directement après, les coefficients (a_i) sont tous significativement différents de 0, sauf ceux des variables x_t et x_{t-5} .

Le graphique 4 présente les coefficients calculés lors de l'exercice 2 et les coefficients ajustés à l'aide du polynôme estimé.



Graphique 4 – Ajustement des coefficients à l'aide d'un polynôme de degrés 4

D. Distribution infinie des retards

Dans ce type de modèle, l'effet de la variable exogène n'est plus limité dans le temps, mais a un effet illimité, bien que, naturellement, cet effet s'estompe pour les périodes anciennes.

$$y_t = \sum_{j=0}^{\infty} a_j x_{t-j} + b_0 + \varepsilon_t = \sum_{j=0}^{\infty} a_j D^j x_t + b_0 + \varepsilon_t = \left[\sum_{j=0}^{\infty} a_j D^j \right] x_t + b_0 + \varepsilon_t \quad [14]$$

Afin de se ramener à un nombre fini de paramètres à estimer, nous devons postuler une forme particulière que peut prendre la succession des coefficients $a_0, a_1, a_2 \dots$

Nous pouvons admettre plusieurs types de spécifications, deux modèles particuliers – les plus utilisés – font l'objet d'une présentation :

- une décroissance géométrique des effets de la variable exogène avec le temps ;
- une croissance suivie d'une décroissance.

1) Modèle de Koyck (progression géométrique)

Le modèle de Koyck¹ postule une décroissance géométrique de la structure des retards (graphique 5) telle que, dans le modèle [14], les coefficients soient liés de la manière suivante :

$$a_1 = \lambda a_0$$

$$a_2 = \lambda^2 a_0$$

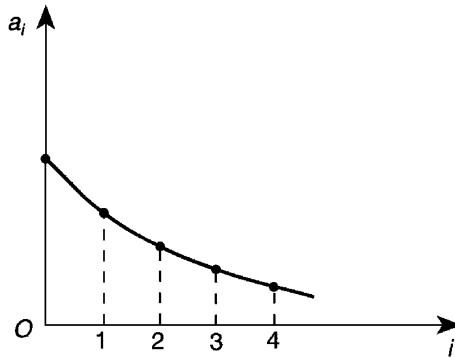
et en général $a_i = \lambda^i a_0$ avec $0 < \lambda < 1$

Soit

$$y_t = b_0 + a_0 x_t + \lambda a_0 x_{t-1} + \lambda^2 a_0 x_{t-2} + \dots + \lambda^i a_0 x_{t-i} + \dots + \varepsilon_t$$

ou encore :

$$y_t = b_0 + a_0 (x_t + \lambda x_{t-1} + \lambda^2 x_{t-2} + \dots + \lambda^i x_{t-i} + \dots) + \varepsilon_t$$



Graphique 5 – Modèle de Koyck :
les coefficients suivent une progression géométrique

La fonction $A(D)$ associée est alors :

$$A(D) = a_0 + \lambda a_0 D^1 + \lambda^2 a_0 D^2 + \dots$$

Le modèle $y_t = A(D)x_t + b_0 + \varepsilon_t$ peut alors s'écrire :

$$B(D)y_t = B(D)A(D)x_t + B(D)b_0 + B(D)\varepsilon_t \text{ avec } B(D) = A(D)^{-1}.$$

$$\text{Or } A(D) = a_0(1 + \lambda D^1 + \lambda^2 D^2 + \dots) = a_0(1 - \lambda D)^{-1}$$

(somme d'une progression géométrique), nous avons donc $B(D) = (1 - \lambda D)/a_0$.

1. Koyck L. M., 1954.

Soit $(1 - \lambda D)y_t = a_0x_t + (1 - \lambda)b_0 + (1 - \lambda D)\varepsilon_t$ ou encore :

$$y_t = \lambda y_{t-1} + a_0x_t + (1 - \lambda)b_0 + \varepsilon_t - \lambda\varepsilon_{t-1} \quad [15]$$

qui est un modèle autorégressif à erreurs liées dont la procédure d'estimation a été exposée en I.

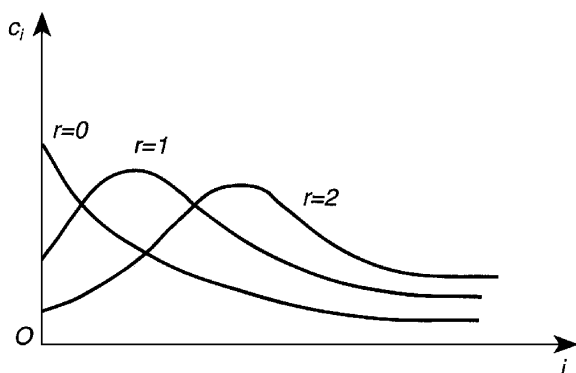
Cette transformation – qui consiste à passer d'un modèle à retards échelonnés à un modèle autorégressif – est habituellement appelée « transformation de Koyck ». Nous pouvons noter que nous passons d'un modèle à retards échelonnés, difficile à estimer par l'abondance des paramètres, à un modèle autorégressif, simple dans sa spécification.

2) Modèle de Solow¹ (distribution de Pascal)

Les coefficients sont distribués selon :

$a_i = (1 - \lambda)^{r+1} C_{r+i}^i \lambda^i$ où C_{r+i}^i est le coefficient du binôme de Newton, r et λ sont deux paramètres avec $0 < \lambda < 1$ et $r \in N$.

Pour $r = 0$, nous retrouvons la distribution géométrique de Koyck, le graphique 6 illustre la distribution des coefficients des retards.



Graphique 6 – Distribution de Pascal des coefficients pour différentes valeurs de r

Le modèle général s'écrit :

$$y_t = b_0 + \sum_{i=0}^{i=\infty} (1 - \lambda)^{r+1} C_{r+i}^i \lambda^i x_{t-i} + \varepsilon_t \text{ ou encore}$$

$$y_t = A(D)x_t + b_0 + \varepsilon_t \quad [16]$$

Une démonstration analogue à la précédente permet d'écrire le modèle [16]

$$B(D)y_t = B(D)A(D)x_t + B(D)b_0 + B(D)\varepsilon_t \text{ avec } B(D) = A(D)^{-1}$$

1. Solow R., 1960.

avec :

- pour $r = 0 \rightarrow B(D) = (1 - \lambda D)/a_0$, soit le modèle précédent
- pour $r = 1 \rightarrow B(D) = (1 - \lambda D)^2/a_0$

Soit :

$$y_t = 2\lambda y_{t-1} - \lambda^2 y_{t-2} + a_0 x_t + (1 - 2\lambda + \lambda^2)b_0 + v_t \quad [17]$$

avec $v_t = (1 - 2\lambda D + \lambda^2 D^2)\varepsilon_t = \varepsilon_t - 2\lambda\varepsilon_{t-1} + \lambda^2\varepsilon_{t-2}$

- pour $r = 2 \rightarrow B(D) = (1 - \lambda D)^3/a_0$

Soit :

$$y_t = 3\lambda y_{t-1} - 3\lambda^2 y_{t-2} + \lambda^3 y_{t-3} + a_0 x_t + (1 - 3\lambda + 3\lambda^2 - \lambda^3)b_0 + v_t \quad [18]$$

avec $v_t = (1 - 3\lambda D + 3\lambda^2 D^2 - \lambda^3 D^3)\varepsilon_t = \varepsilon_t - 3\lambda\varepsilon_{t-1} + 3\lambda^2\varepsilon_{t-2} - \lambda^3\varepsilon_{t-3}$

Afin de déterminer les valeurs du paramètre r , Maddala et Rao (1971) suggèrent l'utilisation d'une procédure de balayage dont la fonction objectif à maximiser est le coefficient de détermination corrigé (\bar{R}^2)

Exercice n° 4

 **fichier C7EX2**

Estimation des coefficients d'un modèle selon une spécification des retards de Koyck et une distribution de Pascal

Nous reprenons les données du tableau 2 concernant la relation Investissement/Profit, on demande :

- 1) d'estimer les paramètres du modèle si on suppose que les coefficients suivent une spécification de Koyck ;
- 2) de rechercher l'ordre d'une distribution de Pascal et d'estimer les paramètres du modèle pour l'ordre adéquat ;
- 3) de comparer les résultats obtenus par les différentes méthodes employées lors des exercices 2, 3 et 4 ;
- 4) enfin, de calculer l'élasticité de long terme des dépenses d'investissement au profit.

Solution

1) On suppose que les coefficients du modèle à retards échelonnés suivent une progression géométrique. D'après [15], l'estimation des paramètres du modèle s'effectue sous la forme d'un modèle autorégressif à autocorrélation des erreurs.

Les résultats de l'estimation du modèle sont les suivants :

$$\begin{aligned}
y_t &= -222,23 + 0,907y_{t-1} + 0,179x_t + e_t \\
&\quad (-2,10) \quad (25,24) \quad (7,59) \\
n &= 43 \\
R^2 &= 0,95 \\
(.) &= t \text{ de Student} \\
DW &= 0,97
\end{aligned}$$

Nous avons utilisé directement la méthode des MCO, car la méthode de correction d'autocorrélation des erreurs envisagées en I donne des résultats quasiment identiques.

Nous obtenons :

$$\hat{\lambda} = 0,907 ; \hat{a}_0 = 0,179 ; \hat{b}_0 = -222,23/(1 - 0,907) = -2\,400,23$$

Le modèle peut donc s'écrire :

$$y_t = -2\,400,23 + 0,179x_t + 0,907 \times 0,179x_{t-1} + 0,907^2 \times 0,179x_{t-2} + \dots + e_t$$

ou encore : $y_t = 0,907y_{t-1} + 0,179x_t - 222,23 + e_t$

2) Nous supposons maintenant que les coefficients suivent une distribution de Pascal.

$$\begin{aligned}
y_t &= 40,78 + 1,42y_{t-1} - 0,54y_{t-2} + 0,096x_t + e_t \\
&\quad (0,40) \quad (13,59) \quad (-5,10) \quad (3,83) \\
n &= 42 \\
R^2 &= 0,97 \\
(.) &= t \text{ de Student} \\
DW &= 2,22
\end{aligned}$$

Il s'agit donc d'estimer des modèles autorégressifs d'ordre 1, 2, 3, etc. Afin de déterminer l'ordre, nous avons procédé à toutes les estimations de ces modèles. À partir d'un modèle autorégressif d'ordre 3, les coefficients ne sont plus significativement différents de 0. Le modèle autorégressif d'ordre 1 ayant déjà été estimé, nous avons procédé à l'estimation du modèle d'ordre 2.

D'après [17], nous obtenons en procédant par identification :

$$\begin{aligned}
\hat{\lambda} &\simeq \sqrt{0,541} \simeq 1,426/2 \simeq 0,72 ; \hat{a}_0 = 0,096 ; \\
\hat{b}_0 &= 40,78/(1 - 2\hat{\lambda} + \hat{\lambda}^2) = 40,78/0,0784 = 520,15
\end{aligned}$$

Le modèle estimé est alors :

$$y_t = 1,426y_{t-1} - 0,541y_{t-2} + 0,096x_t + 40,78 + e_t$$

$$\text{ou encore : } y_t = 520,15 + \sum_{i=0}^{i=\infty} (1 - 0,72)^2 C_{1+i}^i 0,72^i x_{t-i} + e_t$$

3) Comparaison des méthodes

Nous pouvons construire le tableau 4 dans lequel nous indiquons les valeurs des coefficients associées à chaque retard (limités à 6 pour les modèles à retards infinis), la somme des retards et le délai moyen, ceci pour les quatre méthodes employées.

Tableau 4 – Comparaison des méthodes de calcul des coefficients

Retard i	Non contraint ¹	Almon $q = 4$	Koyck	Pascal $r = 1$
0	- 0,01138	- 0,0292	0,1795	0,0784
1	0,06126	0,09733	0,1623	0,0169
2	0,22756	0,19701	0,1472	0,2438
3	0,16793	0,18851	0,1335	0,2926
4	0,11893	0,0920	0,1211	0,3160
5	0,00016	0,02912	0,1098	0,3185
6	0,23717	0,22298	0,0996	0,3058
Somme	0,0814	0,7977	0,9530	1,7246
Délai moyen	3,6421	3,6459	2,1026	3,6104

À la lecture de ces résultats, nous constatons la diversité des profils des coefficients estimés. En effet, nous avons tenté, à des fins pédagogiques, de faire correspondre les coefficients à des structures bien déterminées qui représentent des spécifications de modèles différentes les unes des autres. Dans la pratique, cette démarche est absurde car il convient de tester une seule forme de spécification :

- soit celle que la réflexion économique nous conduit à retenir ;
- soit celle dont le profil du modèle non contraint semble la plus proche.

4) Calcul de l'élasticité de long terme des dépenses d'investissement au profit.

En considérant le modèle à distribution géométrique des retards où la variable exogène et la variable endogène sont sous forme logarithmique :

$$\text{Log } y_t = b_0 + a_0 \text{Log } x_t + \lambda a_0 \text{Log } x_{t-1} + \lambda^2 a_0 \text{Log } x_{t-2} + \dots + \varepsilon_t$$

Le modèle, sous forme réduite, s'exprime par :

$$\text{Log } y_t = \lambda \text{Log } y_{t-1} + a_0 \text{Log } x_t + c_0 + v_t$$

L'élasticité de long terme est égal à λ^2 : $e_{LT} = a_0 / (1 - \lambda)$, l'élasticité de court terme est donnée par : $e_{CT} = a_0$, et le paramètre d'ajustement est λ .

L'estimation du modèle Dépenses d'Investissement/Profit sous forme logarithmique conduit aux résultats suivants :

1. Modèle non contraint = estimation libre des coefficients.
2. L'élasticité de long terme est la somme des élasticités (les coefficients) aux différents décalages, soit une somme d'une progression géométrique de premier terme a_0 et de raison λ .

$$\text{Log } y_t = 0,904 \text{ Log } y_{t-1} + 0,184 \text{ Log } x_t - 0,699 + v_t$$

$$(27,3) \qquad (8,48) \qquad (-2,36)$$

$$n = 43$$

$$R^2 = 0,95$$

(·) = *t* de Student

$$\text{Soit : } e_{LT} = \hat{a}_0 / (1 - \hat{\lambda}) = 0,184 / (1 - 0,904) = 1,91.$$

L'élasticité de long terme est donc égale à 1,91, nous sommes dans la zone des rendements croissants. Lorsque le profit augmente par exemple de 10 %, l'investissement à long terme augmente de 19,1 %.

III. Deux exemples de modèles dynamiques

Les exemples d'application des modèles à décalages sont très nombreux dans la littérature. Nous allons en présenter deux particulièrement intéressants.

A. Le modèle d'ajustement partiel

Le plus souvent, en économie, la rigidité des institutions ou des structures empêche les variables à expliquer d'atteindre le niveau désiré en une seule période. Examinons deux exemples :

- Afin de faire face à une augmentation de la demande, une entreprise cherche à se doter de moyens de production supplémentaires. Cependant cet investissement ne peut pas être réalisé immédiatement et demande un certain temps d'ajustement.
- Lors de l'augmentation des prix du pétrole, les consommateurs et les institutions réagissent avec retard à cette modification des prix. Il en résulte qu'à court terme le comportement des consommateurs ne reflète pas l'ajustement qui s'est produit à long terme.

La formulation de ce modèle est la suivante :

il est nécessaire de distinguer entre la valeur désirée, de la variable à expliquer (y_t^D), et la valeur vraie de cette même variable (y_t). Le niveau désiré de la variable à expliquer est fonction de la variable explicative x_t tel que :

$$y_t^D = a_0 + a_1 x_t + \varepsilon_t \quad [19]$$

Une modification de la valeur de x_t entraîne une modification de la valeur désirée de la variable à expliquer. Cependant, malgré une spécification simple, nous ne pouvons pas appliquer les MCO puisque nous n'avons aucune mesure

de y'_D . En revanche, nous connaissons les valeurs de y_t et nous pouvons spécifier une relation entre y_t^D et y_t :

$$y_t - y_{t-1} = \lambda(y_t^D - y_{t-1}) \text{ avec } 0 \leq \lambda \leq 1 \quad [20]$$

λ est appelé le coefficient d'ajustement.

La relation [20] représente l'ajustement progressif qui s'opère entre la valeur désirée et la valeur observée.

Ce modèle peut se ramener de manière simple à un modèle de Koyck.

En substituant l'équation [19] dans [20], on obtient :

$$\begin{aligned} y_t - y_{t-1} &= \lambda(a_0 + a_1 x_t + \varepsilon_t - y_{t-1}) \quad \text{ou encore} \\ y_t &= (1 - \lambda)y_{t-1} + \lambda a_0 + \lambda a_1 x_t + \lambda \varepsilon_t \end{aligned} \quad [21]$$

$y_t = b_1 y_{t-1} + c_0 + c_1 x_t + v_t$ dont on peut estimer les paramètres par la méthode des MCO, avec $b_1 = (1 - \lambda)$; $c_0 = \lambda a_0$; $c_1 = \lambda a_1$.

B. Le modèle d'anticipations adaptatives

Dans cette spécification, les valeurs de la variable à expliquer y_t sont fonctions des valeurs, non pas observées d'une variable explicative, mais des valeurs attendues, telles que :

$$y_t = a_0 + a_1 x_t^P + \varepsilon_t \quad [22]$$

où x_t^P est la valeur prévue de la variable explicative x_t .

Par exemple, la production d'une entreprise est fonction des valeurs prévues des ventes.

Comme pour le modèle précédent, nous sommes confrontés à l'absence de mesure de la variable x_t^P . Pour lever cette difficulté, nous devons poser une hypothèse concernant la formation de la variable x_t^P : celle des anticipations adaptatives.

Cette relation s'écrit :

$$x_t^P - x_{t-1}^P = \lambda(x_t - x_{t-1}^P) \text{ avec } 0 \leq \lambda \leq 1 \quad [23]$$

où λ est le coefficient d'anticipation.

Nous remarquons que si $\lambda = 0$, l'anticipation reste constante quelles que soient les réalisations de x_t , le modèle ne réagit pas aux nouvelles valeurs réalisées. Si $\lambda = 1$, au contraire, les valeurs anticipées sont toujours égales aux valeurs courantes. Dans la pratique, ce coefficient est donc compris entre ces deux valeurs en fonction du degré de réactivité de l'anticipation.

L'équation [23] peut aussi s'écrire :

$$x_t^P = \lambda x_t + (1 - \lambda)x_{t-1}^P \text{ en développant cette formule il vient :}$$

$$x_t^P = \lambda x_t + \lambda(1 - \lambda)x_{t-1} + \lambda(1 - \lambda)^2 x_{t-2} + \lambda(1 - \lambda)^3 x_{t-3} + \dots$$

$$x_t^P = \lambda \sum_{i=0}^{i=\infty} (1 - \lambda)^i x_{t-i} \quad [24]$$

En remplaçant x_t^P par son expression dans [22] :

$$y_t = a_0 + a_1 \lambda \sum_{i=0}^{i=\infty} (1 - \lambda)^i x_{t-i} + \varepsilon_t$$

qui est un modèle à retards échelonnés¹ dont une transformation de Koyck permet de mettre sous une forme autorégressive :

$$y_t = \lambda a_0 + \lambda a_1 x_t + (1 - \lambda)y_{t-1} + [\varepsilon_t - (1 - \lambda)\varepsilon_{t-1}]$$

Nous pouvons estimer ce modèle autorégressif à autocorrélation des erreurs, et ainsi, en déduire une estimation des paramètres λ , a_0 , a_1 .

Il est à noter que ces deux modèles peuvent être combinés, bien qu'alors cette spécification semble un peu abondante et d'interprétation délicate.

Exercice n° 5

↓ fichier C7EX2

Estimation d'un modèle d'ajustement partiel et d'anticipations adaptatives

À partir des données du tableau 2, on désire estimer les paramètres d'un modèle d'ajustement partiel et les paramètres d'un modèle d'anticipations adaptatives.

Solution

Nous estimons le modèle autorégressif d'ordre 1 avec autocorrélation des erreurs (méthode Hildreth-Lu) :

$$y_t = -103,07 + 0,882 \times y_{t-1} + 0,1555 \times x_t + e_t$$

$$(0,47) \quad (11,04) \quad (4,41)$$

$$e_t = 0,5489 e_{t-1} + v_t$$

$R^2 = 0,96$; $n = 42$; $DW = 1,83$; $(\cdot) = t$ de Student

Soit : $\hat{\lambda} = 1 - 0,882 = 0,1178$; $\hat{a}_0 = -103,07/0,1178 \approx -874,9$;
 $\hat{a}_1 = 0,1555/0,1178 \approx 1,32$

– Modèle d'ajustement partiel :

$$y_t^D = -874,9 + 1,32 x_t + e_t \quad \text{et} \quad y_t - y_{t-1} = 0,118(y_t^D - y_{t-1})$$

– Modèle d'anticipations adaptatives :

$$y_t = -874,9 + 1,32 x_t^P + e_t \quad \text{et} \quad x_t^P - x_{t-1}^P = 0,118(x_t - x_{t-1}^P)$$

1. Nous remarquons que la valeur de la variable à expliquer y dépend d'une pondération géométriquement décroissante de la variable explicative x . Cette structure de pondération est le fondement de la méthode de prévision appelée « lissage exponentiel ».

Exercice n° 6

Modélisation du prix d'une matière première

Nous nous intéressons à modéliser l'évolution du prix d'une matière première y_t selon un modèle autorégressif d'ordre 1 et une variable exogène x_t – la demande – à partir d'une équation de la forme : $y_t = a_1 y_{t-1} + a_2 x_t + \varepsilon_t$ [E1] avec ε_t répondant aux hypothèses classiques. Les données ont été préalablement centrées sur leur moyenne.

L'estimation de l'équation [E1] sur 22 périodes donne :

$$y_t = 0,757y_{t-1} + 0,256x_t + e_t \quad [E2]$$

(0,025) (0,02)

$$R^2 = 0,996 ; DW = 2,17 ; n = 22 ; (.) = \text{écart type}$$

- 1) Le coefficient de détermination a-t-il une signification dans cette équation ?
- 2) Le coefficient a_2 est-il significativement différent de zéro ?
- 3) Les erreurs du modèle sont-elles autocorrélées à l'ordre 1 ?
- 4) Quel est, à long terme, l'influence d'une augmentation de la demande x_t de 1 unité sur le prix y_t ?
- 5) En effectuant une nouvelle régression de y_t sur y_{t-1} seule, on obtient :

$$y_t = 0,615y_{t-1} + e_t \quad [E3]$$

(0,044)

$$R^2 = 0,981 ; DW = 1,64 ; n = 22$$

Est-il normal ou fortuit que le R^2 soit plus élevé dans l'équation [E2] que dans l'équation [E3] ?

- 6) En ôtant la variable x_t dans la régression [E3] le coefficient de y_{t-1} a été modifié, pourquoi ?
- 7) Que signifierait un modèle du type $y_t = a y_{t-1} + \varepsilon_t$ dans lequel a serait supérieur à 1 ?
- 8) L'estimation selon une même spécification mais pour une autre matière première donne :

$$y_t = 0,695y_{t-1} + 0,336x_t + e_t \quad [E4]$$

(0,053) (0,046)

$$R^2 = 0,984 ; DW = 2,27 ; n = 22$$

Les coefficients a_1 sont-ils significativement différents entre la régression [E2] avec ceux de l'équation [E4] ?

Solution

- 1) Bien que le modèle ne possède pas de terme constant, le R^2 peut s'interpréter car les données ont été préalablement centrées sur leurs moyennes.

2) Le coefficient a_2 est significativement différent de zéro car $t^* = 0,256/0,02 = 12,8 > t_{19}^{0,05} = 2,09$.

3) Nous sommes dans les conditions d'application de la statistique de Durbin et Watson car bien que le modèle ne possède pas de terme constant, le DW peut s'interpréter car les données ont été préalablement centrées. Mais il existe un biais de sous-estimation dû au modèle autorégressif. Il faut calculer le h de Durbin :

$$h = \rho \sqrt{\frac{n}{1 - n\hat{\sigma}_{\hat{a}_1}^2}} \text{ avec } \rho = 1 - \frac{DW}{2} = 2,17/2 = -0,085$$

n = nombre d'observations

$\hat{\sigma}_{\hat{a}_1}^2$ = variance estimée du coefficient \hat{a}_1 de la variable décalée

Cette statistique h est distribuée de manière asymptotique comme une variable normale centrée réduite. Ainsi, il y a équivalence entre les deux tests d'hypothèses suivants :

$$H_0 : \rho = 0 \quad H_0 : h = 0$$

$$H_1 : \rho \neq 0 \quad H_1 : h \neq 0$$

Si $|h| < t^{\alpha/2}$ nous acceptons l'hypothèse H_0 d'indépendance des erreurs.

($t^{0,05} = 1,96$ valeur issue de la loi normale pour un test bilatéral au seuil de 5%).

$|h| = -0,085 \sqrt{\frac{22}{1 - 22 \times 0,025^2}} = 0,40 < t^{0,05} = 1,96$, nous acceptons l'hypothèse d'indépendance des erreurs.

4) À la période 1 : $\Delta y_t = 0,757 \Delta y_{t-1} + 0,256 \Delta x_t = 0,757 \times 0 + 0,256 \times 1 = 0,256$

À la période 2 : $\Delta y_t = 0,757 \times 0,256 = 0,194$

À la période 3 : $\Delta y_t = 0,757 \times 0,194 = 0,146$

Il s'agit d'une progression géométrique de premier terme $U_0 = 0,256$ et de raison $q = 0,757 < 1$.

La somme tend vers une limite finie et est égale à : $U_0 / (1 - q) = 0,256 / (1 - 0,757) = 1,053$.

Le prix augmentera de 1,053 unité à long terme.

5) Il est normal que le R^2 soit plus élevé dans l'équation [E2] que dans l'équation [E3] car une variable explicative a été retirée du modèle ce qui entraîne une diminution de la somme des carrés des résidus.

6) À cause de la covariance existant entre les variable y_{t-1} et x_t .

7) Un modèle du type $y_t = a y_{t-1} + \varepsilon_t$ dans lequel le coefficient a serait supérieur à 1 conduirait à une solution explosive (croissance infinie) ce qui impossible.

8) Tests de comparaison de coefficient

$$H_0 : a_{1,MP1} = a_{1,MP2}$$

$$H_1 : a_{1,MP1} \neq a_{1,MP2}$$

$$t^* = \frac{(\hat{a}_{1,MP1} - \hat{a}_{1,MP2}) - 0}{\hat{\sigma}_{\hat{a}_{1,MP1} - \hat{a}_{1,MP2}}} = \frac{0,757 - 0,695}{\sqrt{0,025^2 + 0,053^2}} = 1,05 \text{ suit une Student à}$$

$n_1 + n_2 - 6 = 38$ degrés de liberté, la covariance entre $a_{1,MP1}$ et $a_{1,MP2}$ est nulle car les deux régressions sont indépendantes (les deux échantillons sont différents), $t^* = 1,05 < t_{\infty}^{0,05} = 1,96$ donc non rejet de H_0 .

Exercice n° 7

Un modèle d'évaluation de la politique marketing

Nous intéressons à mesurer l'impact de la publicité et des promotions sur les ventes d'un produit. Les impacts sont mesurés au moyen d'estimations en coupe instantanée menées sur des échantillons représentatifs de produit de grande consommation à forte pression marketing de deux secteurs : les produits cosmétiques et les détergents.

Les chiffres entre parenthèses sous les coefficients de régression sont les écarts types des coefficients.

Partie A

Le premier modèle testé est le suivant :

$$V_{it} = a_0 + a_1 Pu_{it} + a_2 Pr_{it} + \varepsilon_{it} \quad [E1]$$

V_{it} = Vente en volume du produit i pour l'année t

Pu_{it} = Dépenses publicitaires en euros pour le produit i pour l'année t

Pr_{it} = Dépenses promotionnelles en euros pour le produit i pour l'année t

Les résultats obtenus pour une année ($t = 1$) pour le secteur des cosmétiques est le suivant :

$$V_i = 812,86 + 29,94Pu_i + 21,31Pr_i + e_i \quad i = 1, \dots, 20$$

(10,73) (8,23)

$$R^2 = 0,868 ; DW = 1,68 ; \text{Somme des Carrés des Résidus} = SCR_C = 291$$

- 1) Interpréter d'un point de vue statistique ce résultat d'estimation.
- 2) Théoriquement si l'entreprise respecte une politique optimale d'allocation de son budget publi-promotionnel (*marketing mix*), les effets positifs d'un euro dépensé en publicité ou en promotion sont égaux ($a_1 = a_2$). Sachant que $Cov(\hat{a}_1, \hat{a}_2) = 6,75$, effectuez le test permettant de savoir si cette politique optimale est vérifiée.
- 3) Une interrogation concernant les études menées sur des échantillons d'entreprises réside en l'existence de comportements homogènes.

Nous procédons donc à deux nouvelles estimations.

Les résultats obtenus pour la même année ($t = 1$) pour le secteur des détergents est le suivant :

$$V_i = 627,32 + 14,27Pu_i + 17,87Pr_i + e_i \quad i = 1, \dots, 20$$

(3,12) (5,57)

$$R^2 = 0,41 ; DW = 1,68 ; SCR_D = 249$$

Les résultats obtenus pour la même année ($t = 1$) pour le secteur des détergents et des cosmétiques sont les suivants :

$$V_i = 785 + 16,56Pu_i + 18,91Pr_i + e_i \quad i = 1, \dots, 40$$
$$SCR_T = 926$$

Tester l'homogénéité des comportements dans ces deux secteurs. Qu'en concluez-vous ?

4) Dans le même esprit nous disposons pour le secteur des cosmétiques d'une estimation pour l'année suivante ($t = 2$), nous allons donc chercher à tester la stabilité temporelle pour ce secteur.

Sachant que l'estimation du modèle en $t = 2$ indique $SCR_2 = 316$ ($i = 1, \dots, 20$) et qu'une estimation pour $t = 1$ et $t = 2$ indique $SCR_T = 1054$, qu'en concluez-vous ?

Partie B

Un des problèmes de ce type de modélisation consiste en un biais dû à des variables omises, en particulier des variables caractéristiques du produit.

1) Supposons que ces variables soient additives et peuvent être synthétisées globalement par une seule notée F_i , cette variable est constante dans le temps car il s'agit d'une caractéristique permanente du produit. Cependant cette variable n'est pas mesurable.

Le modèle s'écrit donc maintenant :

$$V_{it} = a_{0t} + a_{1t}Pu_{it} + a_{2t}Pr_{it} + F_{it} + \varepsilon_{it} \quad [E2]$$

Où t désigne l'année d'estimation et i le produit.

Sachant que nous disposons des données pour 20 produits *pour deux années consécutives*, comment estimer les paramètres du modèle intégrant cet effet individuel sans qu'il apparaisse effectivement ?

2) Si cet effet individuel n'est pas additif mais multiplicatif nous considérons alors qu'il peut être fonction du budget marketing M_{it} du produit i pour l'année t . On appelle $k_t = \frac{V_t}{M_t}$, le rapport moyen entre les ventes totales et le budget total marketing dans l'échantillon pour tous les produits pour l'année t .

Sous l'hypothèse que les effets individuels sont constants dans le temps on peut définir une variable $\left(\frac{\bar{V}}{\bar{M}}\right)$ telle que $\left(\frac{\bar{V}}{\bar{M}}\right)_{i(t-1)} = \frac{V_{i(t-1)}}{M_{i(t-1)}} - k_{t-1}$

Interpréter le sens de cette variable économique. En quoi son introduction dans l'équation [E1] permet-elle de mesurer les effets individuels ?

3) On a obtenu pour les secteurs des cosmétiques et des détergents les résultats pour l'année $t = 1$.

• Cosmétique :

$$V_i = 582 + 25,19Pu_i + 23,81Pr_i - 0,97\left(\frac{\bar{V}}{\bar{M}}\right)_{i(t-1)} + e_i \quad i = 1, \dots, 20$$

(9,3) (5,7) (0,37)

$$R^2 = 0,967$$

• Détergent :

$$V_i = 328 + 17,78Pu_i + 18,74Pr_i - 0,49\left(\frac{\bar{V}}{\bar{M}}\right)_{i(t-1)} + e_i \quad i = 1, \dots, 20$$

(5,7) (8,2) (0,21)

$$R^2 = 0,967$$

L'augmentation des coefficients de détermination R^2 pour chaque équation est-elle suffisante pour affirmer que la relation estimée est de meilleure qualité que l'équation [E1] ? Sinon comment jugez-vous de l'amélioration ou la détérioration des résultats.

4) Les coefficients des variables $\left(\frac{\bar{V}}{M}\right)_{i(t-1)}$ pour le modèle cosmétique et détergent sont-ils significativement différents ?

Partie C

Un des biais envisageables dans l'équation [E1] est que les ventes d'un produit soient fonction d'un événement imprévu. Dans cette situation, cela peut créer un biais sur les ventes du produit qui diffère de ce qu'elles seraient dans des conditions normales. Pour effectuer des prévisions correctes, on peut redresser ce biais en introduisant comme variable supplémentaire dans [E1] les ventes de la période précédente :

$$V_{it} = a_0 + a_1 Pu_{it} + a_2 Pr_{it} + a_3 V_{it-1} + \varepsilon_{it} \quad [E3]$$

On a obtenu pour les secteurs des cosmétiques les résultats pour l'année $t = 1$.

$$V_{it} = -1,52 + 2,52 Pu_{it} + 6,36 Pr_{it} + 0,75 V_{it-1} + e_{it} \quad i = 1, \dots, 20 \quad [E4]$$

(1,1) (1,7) (0,3)

$$R^2 = 0,979$$

1) Interpréter d'un point de vue statistique et économique ces résultats, est-il envisageable dans l'équation [E3] de tester une éventuelle autocorrélation des erreurs ?

2) Pour ce modèle cosmétique nous avons effectué la régression suivante :

$$e_{it}^2 = 2,1 Pu_{it} + 0,18 Pr_{it} + 5,8 + u_{it} \quad i = 1, \dots, 20$$

(0,2) (0,5)

e_{it} = Résidu de la régression [E4]

Pourquoi procéder à cette régression ? Qu'en concluez-vous et que proposez-vous ?

Solution

A.1) Soit les tests suivants à mener.

- Test de significativité globale de la régression : $H_0 : a_1 = a_2 = 0$

Soit le test de Fisher : $F^* = \frac{R^2/k}{(1 - R^2)/(n - k - 1)} = \frac{0,868/2}{(1 - 0,868)/(20 - 2 - 1)}$
 $= 55,89 > F_{2,17}^{0,05} = 3,59$, nous rejetons l'hypothèse H_0 , il existe au moins une variables Pu_i ou/et Pr_i d'explicative de V_i

- Tester $H_0 : a_1 = 0$ contre l'hypothèse $H_1 : a_1 \neq 0$

Le test d'hypothèses bilatéral consiste donc à comparer le ratio de Student empirique $t_{a1}^* = \frac{\hat{a}_1}{\hat{\sigma}_{\hat{a}_1}} = 2,79$ à la valeur du t de Student lue dans la table à $20 - 2 - 1 = 17$ degrés de liberté et pour un seuil de probabilité égal à 5%, soit $t_{17}^{0,05} = 2,11$. Puisque $t_{a1}^* = 2,79 > t_{17}^{0,05} = 2,11$, nous refusons l'hypothèse H_0 , le coefficient a_1 est donc significativement différent de 0.

- Soit à tester $H_0 : a_2 = 0$ contre l'hypothèse $H_1 : a_2 \neq 0$

$$t_{\hat{a}_2}^* = \frac{\hat{a}_2}{\hat{\sigma}_{\hat{a}_2}} = 2,59 > t_{17}^{0,05} = 2,11 \text{ nous refusons l'hypothèse } H_0, \text{ le coefficient } a_2 \text{ est}$$

donc significativement différent de 0.

• Le modèle étant spécifié en coupe instantanée et n'ayant aucune information sur un éventuel tri des données selon un critère, la statistique de Durbin et Watson n'est pas interprétable.

A.2) Le test d'hypothèse est : $H_0 : a_1 = a_2$

Sous l'hypothèse H_0 $t^* = \frac{|(\hat{a}_1 - \hat{a}_2) - 0|}{\hat{\sigma}_{\hat{a}_1 - \hat{a}_2}}$ suit une loi de Student à 17 degrés de liberté.

Il convient de calculer : $\hat{\sigma}_{\hat{a}_1 - \hat{a}_2}$

$$\begin{aligned} \text{Var}(\hat{a}_1 - \hat{a}_2) &= \text{Var}(\hat{a}_1) + \text{Var}(\hat{a}_2) - 2 \times \text{Cov}(\hat{a}_1, \hat{a}_2) \\ &= 10,73^2 + 8,23^2 - 2 \times 6,75 = 169,36 \end{aligned}$$

$$\text{D'où : } t^* = \frac{|29,94 - 21,31|}{\sqrt{169,36}} = 0,66 < t_{17}^{0,05} \approx 2,11. \text{ Nous acceptons l'hypothèse}$$

H_0 , la politique optimale est donc respectée.

A.3) Soit à tester (test de Chow) l'hypothèse $H_0 : SCR_T = SCR_D + SCR_C \rightarrow H_0 : SCR_T - (SCR_D + SCR_C) = 0$

La statistique du Fisher empirique du test de Chow est donnée par :

$F^* = \frac{(SCR_T - (SCR_D + SCR_C))/(k+1)}{(SCR_D + SCR_C)/(n-2(k+1))}$ qui suit alors une loi de Fisher à $(k+1)$ et $n-2(k+1)$ degrés de liberté.

$$\text{Soit, } F^* = \frac{(926 - (291 + 249))/3}{(291 + 249)/(34)} \approx 8,10 > F_{3;34}^{0,05} \approx 2,92. \text{ Nous refusons l'hypo-}$$

thèse H_0 , les comportements sont hétérogènes entre les deux secteurs.

A.4) Il s'agit toujours du test de Chow mais mené maintenant en données temporelles.

$$\text{Soit, } F^* = \frac{(1\,054 - (291 + 316))/3}{(291 + 316)/(34)} \approx 8,34 > F_{3;34}^{0,05} \approx 2,92. \text{ Nous rejetons l'hy-}$$

pothèse H_0 , les comportements sont différents, sur les deux années étudiées, pour le secteur des cosmétiques.

B.1) Nous avons :

$$V_{it} = a_{0t} + a_{1t}Pu_{it} + a_{2t}Pr_{it} + F_{it} + \varepsilon_{it}$$

$$V_{it-1} = a_{0t-1} + a_{1t-1}Pu_{it-1} + a_{2t-1}Pr_{it-1} + F_{it-1} + \varepsilon_{it-1}$$

Puisque les effets sont constants nous avons : $F_{it} = F_{it-1}$

Nous pouvons alors écrire :

$$\begin{aligned} V_{it} - V_{it-1} &= (a_{0t} - a_{0t-1}) + a_{1t}Pu_{it} - a_{1t-1}Pu_{it-1} + a_{2t}Pr_{it} - a_{2t-1}Pr_{it-1} \\ &\quad + \varepsilon_{it} - \varepsilon_{it-1} \end{aligned}$$

L'estimation de ce modèle permet de lever le biais du aux caractéristiques de produit.

B.2) La variable $\left(\frac{\bar{V}}{M}\right)_{i(t-1)} = \frac{V_{i(t-1)}}{M_{i(t-1)}} - k_{t-1} = \frac{V_{i(t-1)}}{M_{i(t-1)}} - \frac{V_{(t-1)}}{M_{(t-1)}}$ mesure les déviations individuelles du produit i par rapport au ratio des ventes totales sur les dépenses marketing totales pour tous les produits de l'année antérieure. Ainsi cette variable capte ce qui est spécifique du produit par rapport au comportement moyen tendanciel. Si le coefficient de cette variable n'est pas significativement différent de 0, cela signifie que les produits ont tous le même comportement et donc qu'il n'existe pas d'effet individuel.

B.3) Le coefficient de détermination est insuffisant pour évaluer la pertinence d'ajout d'une variable explicative supplémentaire car même une variable non significative contribue à augmenter la somme des carrées expliquées. En revanche il convient d'effectuer un test de significativité par rapport à 0 sur la variable nouvellement introduite.

Soit pour le secteur cosmétique : $t_{\hat{a}_3}^* = \frac{|0,97|}{0,37} = 2,62$ et pour le secteur Détergent : $t_{\hat{a}_3}^* = \frac{|0,49|}{0,21} = 2,33$, les deux ratios sont supérieurs à $t_{17}^{0,05} = 2,11$. L'ajout de la variable $\left(\frac{\bar{V}}{M}\right)_{i(t-1)}$ contribue significativement à l'explication des ventes, il existe des effets individuels par produit. De plus, les coefficients des variables Publicité et Promotion sont significativement différents de 0 et leurs valeurs se rapprochent, ce qui correspond mieux à l'intuition théorique.

B.4) Le test d'hypothèse est : $H_0 : a_3^C = a_3^D$

Sous l'hypothèse H_0 , $t^* = \frac{|(\hat{a}_3^C - \hat{a}_3^D) - 0|}{\hat{\sigma}_{\hat{a}_3^C - \hat{a}_3^D}}$ suit une loi de Student à $n_1 + n_2 - 2(k + 1) = 32$ degrés de liberté.

Il convient de calculer : $\hat{\sigma}_{\hat{a}_3^C - \hat{a}_3^D}$

$$\begin{aligned}\text{Var}(\hat{a}_3^C - \hat{a}_3^D) &= \text{Var}(\hat{a}_3^C) + \text{Var}(\hat{a}_3^D) - 2 \times \text{Cov}(\hat{a}_3^C, \hat{a}_3^D) \\ &= 0,37^2 + 0,21^2 - 2 \times 0 = 0,181\end{aligned}$$

Les régressions étant effectuées sur des échantillons indépendants la covariance est nulle.

D'où : $t^* = \frac{|-0,97 + 0,49|}{\sqrt{0,181}} = 1,13 < t_{\infty}^{0,05} \approx 1,96$. Nous acceptons l'hypothèse H_0 ,

les deux coefficients ne sont pas significativement différents entre eux.

C.1) Les coefficients du modèle sont tous significativement différents de 0, les t de Student empiriques des trois coefficients sont supérieurs à $t_{17}^{0,05} = 2,11$ nous refusons l'hypothèse H_0 de nullité des coefficients. Nous ne pouvons pas interpréter la statistique de Durbin et Watson car le modèle est estimé en coupe instantanée pour l'année t et nous ne savons pas si les données ont été préalablement triées.

C.2) L'estimation de ce modèle permet de tester une éventuelle hétéroscédasticité des résidus (test de White), le coefficient de la variable Pu_{it} est significativement différent de 0, il existe donc une hétéroscédasticité liée à la variable Publicité.

Il convient donc de corriger le modèle de l'hétéroscédasticité en procédant à une régression pondérée, dont le facteur de pondération est : $\frac{1}{\sqrt{Pu_{it}}}$.

Exercice n° 8

↓ fichier C7EX8

Une comparaison des courbes de Phillips française et américaine¹

Soit les données annuelles françaises et américaines concernant les taux d'inflation et de chômage sur 34 ans. Le but de cet exercice est d'essayer de valider empiriquement les courbes de Phillips. Une démarche progressive est proposée. Dans un premier temps, vous devrez estimer la courbe qui relie le taux d'inflation au taux de chômage, puis celle qui relie l'inflation et le chômage retardée d'une période. Dans un deuxième temps, nous prendrons mieux en compte la dynamique qui relie l'inflation (niveau général des prix) à l'inflation salariale afin de valider empiriquement les mécanismes invoqués par Phillips. Enfin, dans un dernier temps, les anticipations d'inflation formées de manière adaptative seront prises en compte. Il est à noter que les séries peuvent suivre une dynamique dite « non-stationnaire », des procédures d'estimation abordées au chapitre 9 permettent alors d'en tenir compte.

chofra = Chômage France

chousa = Chômage États-Unis

inffra = Taux d'inflation France

infusa = Taux d'inflation États-Unis

Première partie : Courbes inflation-chômage

1) Visualiser successivement les deux courbes de chaque pays inflation-chômage dans le plan (chômage en abscisse et inflation en ordonnée). Constaté la grande instabilité de celles-ci.

2) Estimer économétriquement les deux équations inflation-chômage (respectivement France et États-Unis) qui relient le niveau d'inflation, variable expliquée (nommés respectivement inffra et infusa) à une constante et au niveau retardé du chômage, variable explicative (nommés respectivement chofra et chousa).

Constaté que cela ne modifie pas notablement les résultats de prendre comme variable explicative le taux de chômage contemporain et le taux de chômage retardé. Dorénavant, nous considérerons le taux de chômage retardé.

3) Examiner les résidus des deux équations (chômage retardé). Vous semblent-ils liés par une dynamique cyclique ? (présomption d'autocorrélation des résidus). Existe-t-il des résidus très éloignés des autres ? (présomption de valeurs aberrantes).

4) Effectuer les tests de Durbin-Watson. Confirmer vos résultats en utilisant les tests d'autocorrélation des résidus de Breusch-Godfrey.

1. Merci à Jean-François Jacques, professeur à l'Université de Paris-Est-Marne-la-Vallée, auteur de ce cas.

Deuxième partie : courbes de Phillips originelles (sans anticipation d'inflation) avec autocorrélation des résidus et choc négatif d'offre (année 11 et année 17)

5) Construire une variable indicatrice appelée *DUM* qui vaut 0 à chaque date à l'exception des périodes 11 et 17 pour lesquelles elle vaut 1.

6) Soit les trois équations suivantes :

- Équation de Phillips originelle qui relie le taux de croissance des salaires nominaux (\hat{w}_t) au taux de chômage (u_t) : $\hat{w}_t = \lambda(\bar{u} - u_{t-1})$ avec \bar{u} = constante.
- Équation qui relie inversement le taux de croissance de la production (\hat{y}_t) et celui du chômage (pseudo-loi d'Okun) : $\hat{y}_t = -\frac{1}{\alpha}\hat{u}_t$ avec $\hat{u}_t = (u_t - u_{t-1})$.
- Équation de quasi-offre ou ce qui identifie formellement équation de fixation des prix (\hat{p}_t) en taux de croissance en concurrence imparfaite : $\hat{p}_t = c \hat{y}_t + \hat{w}_t$.

Montrer qu'il est possible, à partir de ces équations, d'écrire une dynamique qui relie le taux d'inflation \hat{p}_t à une constante, au niveau du chômage u_t et au niveau du chômage retardé u_{t-1} .

Estimer économétriquement cette équation. Est-ce que cela améliore les résultats de la première partie ?

7) Estimer de nouveau le modèle de la question 6 en introduisant dans la liste des variables explicatives la variable *DUM* et en corrigeant de l'autocorrélation des résidus (cf. chapitre 5).

Troisième partie : courbes de Phillips augmentées des anticipations d'inflation

8) On cherche maintenant à estimer un modèle intégrant des anticipations d'inflation :

$$\hat{p}_t = a_0 + a_1 \hat{p}_{t-1} + a_2 u_t + a_3 u_{t-1} + \varepsilon_t$$

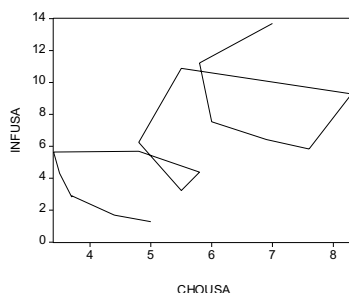
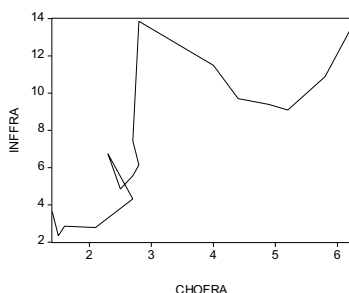
Estimer cette relation inflation-chômage en ignorant l'autocorrélation des résidus mais en conservant la variable *DUM*. En faisant un test du h de Durbin (présence de la variable endogène retardée, cf. chapitre 7, I. B.) confirmer qu'il n'existe plus d'autocorrélation des résidus.

9) Réaliser sur ce nouveau modèle des tests de Chow à la période 21 pour la France et à la période 18 pour les États-Unis (date de début des politiques de désinflation). Attention : la variable *DUM* doit être retirée de la régression.

10) Construire les séries pour la France et les États-Unis (*phifra* et *phiusa*) $\phi_i = p_i - \hat{a}_1 p_{i-1} - \hat{a}_4 dum$, les coefficients de régression proviennent du résultat de la question 8. Puis regarder l'allure de la courbe dans le plan chômage en abscisse et ϕ en ordonné. Que constatez-vous lorsque que l'on supprime les déplacements verticaux de long terme de la courbe de Phillips dus aux anticipations et aux chocs d'offre ? Une politique de désinflation est-elle plus coûteuse, en termes de chômage, en France ou aux États-Unis ?

Solution

1) À la lecture des représentations graphiques (graphique 7) de la relation inflation-chômage pour la France et pour les États-Unis, nous pouvons penser qu'il est très difficile de trouver une relation significative, pour les deux pays, entre le taux de chômage et le taux d'inflation à une date donnée car les deux courbes sont très instables.



Graphique 7 – Relation inflation-chômage pour la France et pour les États-Unis

2) Estimations économétriques.

Les résultats obtenus pour la France sont les suivants :

Dependent Variable: INFFRA

Included observations: 33 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.918367	1.304888	6.834582	0.0000
CHOFR(-1)	-0.442073	0.173685	-2.545262	0.0161

DW = 0,23

On remarque que la probabilité critique de la variable chofra(-1) est inférieure à 5 %, donc le taux de chômage retardé est une variable significative et dont le coefficient est négatif, ce qui est conforme à l'intuition économique.

Estimons maintenant la relation entre l'inflation et le taux de chômage contemporain.

Dependent Variable: INFFRA

Included observations: 34

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.273400	1.322230	6.257158	0.0000
CHOFR	-0.345105	0.171588	-2.011250	0.0528

Dans ce cas, nous remarquons que le taux de chômage contemporain n'est pas aussi significatif. En effet, la probabilité critique est légèrement supérieure à 5 % mais reste quand même acceptable.

Pour les États-Unis :

Dependent Variable: INFUSA

Included observations: 33 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.718162	2.147358	2.662883	0.0122
CHOUA(-1)	-0.093347	0.343655	-0.271629	0.7877

DW = 0,42

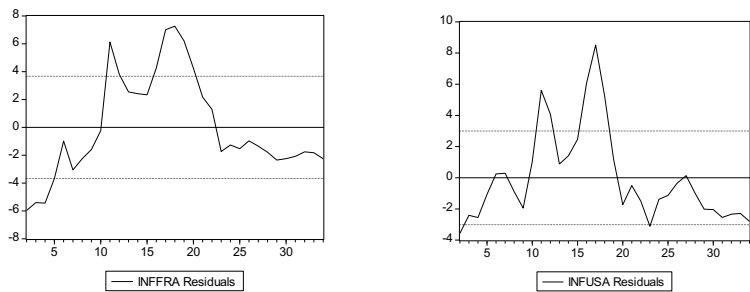
Dependent Variable: INFUSA
Included observations: 34

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.329524	2.082007	1.118884	0.2715
CHOUA	0.449518	0.335099	1.341446	0.1892

On remarque que ni le taux de chômage retardé, ni le taux de chômage contemporain ne sont significatifs car leurs probabilités critiques sont supérieures à 5 %.

Nous pouvons donc conclure que le fait de prendre le chômage retardé ou le chômage contemporain ne change pas significativement les résultats. Nous prendrons donc dans la suite le chômage retardé afin de rester homogène par rapport à la France.

3) La représentation graphique (graphique 8) des résidus de la relation inflation-chômage retardé pour la France et pour les États-Unis laisse supposer une présence d'autocorrélation positive des résidus car il semblerait que les résidus soient liés par une dynamique cyclique. En ce qui concerne les points aberrants, il apparaît des résidus éloignés pour les périodes 10 et 18 pour la France alors que pour les États-Unis, les résidus sont élevés pour les périodes 10 et 17.



Graphique 8 – Représentation des résidus pour la France et pour les États-Unis

4) Tests d'autocorrélation des erreurs.

Pour la France, il est possible d'appliquer le test de Durbin et Watson à la relation entre l'inflation et le taux de chômage retardé car le nombre d'observations est supérieur à 15, la variable à expliquer ne figure pas parmi les variables explicatives et le modèle est spécifié en séries temporelles. La statistique de Durbin et Watson est égale à 0,23. Dans la table de Durbin et Watson, pour 33 observations et une variable explicative, on trouve que $d_1 = 1,38$ et $d_2 = 1,51$. La statistique de Durbin et Watson dans ce cas est inférieure à d_1 , on peut donc conclure à une autocorrélation positive des erreurs.

Il est également possible de faire le test de Breusch-Godfrey.

Breusch-Godfrey : Serial Correlation LM Test:			
F-statistic	73.24807	Probability	0.000000

On rejette l'hypothèse d'absence d'autocorrélation car la probabilité critique du Fisher est très inférieure à 5 %.

Pour les États-Unis, la statistique de Durbin et Watson est égale à 0,42. Or d_1 et d_2 ont les mêmes valeurs que pour la France, donc on peut conclure à une autocorrélation des erreurs.

Breusch-Godfrey : Serial Correlation LM Test:			
F-statistic	48.13823	Probability	0.000000

On rejette l'hypothèse d'absence d'autocorrélation car la probabilité critique du Fisher est inférieure à 5 %. On peut donc conclure à une autocorrélation positive des erreurs.

5) Construire une variable appelé *DUM* qui vaut 0 à chaque date à l'exception des périodes 11 et 17 pour lesquelles elle vaut 1.

6) Estimation de la courbe de Philips sans autocorrélation des résidus.

Soit le système suivant :

$$\hat{w}_t = \lambda(\tilde{u} - u_{t-1}) \quad (1)$$

$$\hat{y}_t = -\alpha^{-1}\hat{u}_t \quad (2)$$

$$\hat{p}_t = c\hat{y}_t + \hat{w} \quad (3)$$

En remplaçant dans l'expression (3) \hat{y}_t et \hat{w}_t par les relations (2) et (3), on trouve que :

$$\hat{p}_t = c(-\alpha^{-1}\hat{u}_t) + \lambda(\tilde{u} - u_{t-1})$$

$$\hat{p}_t = -\alpha^{-1}c(u_t - u_{t-1}) + \lambda(\tilde{u} - u_{t-1}) \quad \text{car } \hat{u}_t = (u_t - u_{t-1})$$

$$\text{D'où } \hat{p}_t = -\alpha c u_t - (\lambda + \alpha^{-1}c)u_{t-1} + \lambda\tilde{u}$$

Nous pouvons alors estimer cette équation.

Pour la France :

Dependent Variable: INFFRA

Included observations: 33 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.195813	1.341090	6.111305	0.0000
CHOFRA	1.978700	1.187570	1.666176	0.1061
CHOFRA(-1)	-2.412074	1.194353	-2.019565	0.0524
Durbin-Watson stat	0.436616			

On remarque que les deux variables ne sont pas significatives car leurs probabilités critiques sont supérieures à 5 %.

Pour les États-Unis :

Dependent Variable: INFUSA

Included observations: 33 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.123759	2.073971	1.988340	0.0560
CHOUA	1.363228	0.533217	2.556610	0.0159
CHOUA(-1)	-1.192817	0.533991	-2.233777	0.0331
Durbin-Watson stat	0.683192			

En ce qui concerne les États-Unis, ce modèle est plus adapté que les deux précédents car les deux variables (chômage et chômage retardé) sont significatives.

7) Estimation d'un nouveau le modèle en introduisant la variable *DUM* et en prenant en compte l'autocorrélation des résidus.

Pour la France :

Dependent Variable: INFFRA

Included observations: 32 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	15.14060	6.269402	2.414999	0.0228
CHOFRA	-0.254229	0.522127	-0.486910	0.6303
CHOFRA(-1)	-0.788179	0.495101	-1.591958	0.1230
DUM	2.674737	0.797477	3.354000	0.0024
AR(1)	0.866577	0.081842	10.58842	0.0000
Durbin-Watson stat	1.791241			

On constate dans ce modèle que les chômages retardés et contemporains ne sont pas significatifs.

Pour les États-Unis :

Dependent Variable: INFUSA

Included observations: 32 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	12.66870	3.475067	3.645597	0.0011
CHOUSA	-0.221311	0.292094	-0.757670	0.4552
CHOUSA(-1)	-0.963667	0.296656	-3.248427	0.0031
DUM	2.293875	0.767918	2.987137	0.0059
AR(1)	0.895121	0.086378	10.36288	0.0000
Durbin-Watson stat	1.552390			

On remarque que dans ce cas, le chômage contemporain n'est plus significatif, alors que le chômage contemporain l'était dans le modèle précédent, non corrigé des autocorrélations des résidus et des points aberrants.

8) On cherche maintenant à estimer le modèle suivant :

$$\hat{p}_t = a_0 + a_1 \hat{p}_{t-1} + a_2 u_t + a_3 u_{t-1} + \varepsilon_t.$$

Il s'agit d'un modèle autorégressif d'ordre 1.

Pour la France :

Dependent Variable: INFFRA

Included observations: 33 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.663308	0.625929	2.657345	0.0129
INFFRA(-1)	0.831080	0.063433	13.10170	0.0000
CHOFRA	-0.360813	0.439505	-0.820953	0.4186
CHOFRA(-1)	0.225252	0.448785	0.501916	0.6197
DUM	4.995335	0.898138	5.561881	0.0000
Durbin-Watson stat	2.195938			

Nous remarquons que les variables choffra et choffra(-1) ne sont pas significatives ; en revanche, le terme autorégressif est bien significativement différent de 0.

Nous effectuons le test h de Durbin car nous avons un modèle autorégressif (présence de la variable à expliquer comme variable explicative retardée).

Soit le test d'hypothèses $H_0 : h = 0$ contre $H_1 : h \neq 0$.

$$h = |\rho| \sqrt{\frac{n}{1 - n\hat{\sigma}_{\hat{a}1}^2}} \text{ avec } \rho = 1 - \frac{DW}{2} \text{ donc } \rho = -0,098$$

$$h = |0,98| \sqrt{\frac{33}{1 - 33 \times 0,0634^2}} = 0,60 < t^{0,05} = 1,96. \text{ Il n'existe donc plus d'auto-corrélation des erreurs.}$$

Pour les États-Unis :

Dependent Variable: INFUSA

Included observations: 33 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.704137	0.966380	3.833004	0.0007
INFUSA(-1)	0.916993	0.115893	7.912443	0.0000
CHOUSA	-1.025115	0.356171	-2.878159	0.0076
CHOUSA(-1)	0.442910	0.295441	1.499151	0.1450
DUM	4.630472	1.005135	4.606815	0.0001
Durbin-Watson stat	2.047469			

Dans ce modèle, seul le taux de chômage décalé n'est pas significatif. On peut également effectuer le test du h de Durbin, $h = 0,182$ qui est très inférieur à 1,96. Il n'existe donc plus d'autocorrélation des erreurs.

9) Tests de Chow¹.

Pour la France :

Le test de Chow est effectué à la période 20 comme période charnière. Comme pour la deuxième période, la variable *DUM* est égale à 0 (colinéarité avec le terme constant), nous la retirons pour les deux périodes.

Nous obtenons : *SCR* = 83,25 ; *SCR1* = 70,03 ; *SCR2* = 5,056.

Chow Breakpoint Test: 21			
F-statistic	0.678912	Probability	0.613020

Pour les États-Unis :

Nous obtenons : *SCR* = 80,59 ; *SCR1* = 52,04 ; *SCR2* = 7,93.

Chow Breakpoint Test: 18			
F-statistic	2.148648	Probability	0.104452

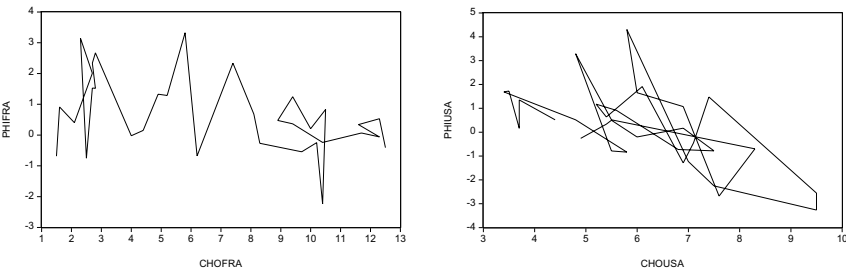
Nous acceptons l'hypothèse que le fait de scinder en deux sous périodes n'améliore pas le modèle pour la France et les États-Unis.

10) Nous construisons les deux séries $\phi_i = p_i - \hat{a}_1 p_{i-1} - \hat{a}_4 dum$ pour la France et les États-Unis :

$$\phi_{ifra_t} = \text{inf} fra_t - 0,83 \text{inf} fra_{t-1} - 4,99 dum_t$$

$$\phi_{iusa_t} = \text{inf} usa_t - 0,91 \text{inf} usa_{t-1} - 4,63 dum_t$$

À la lecture de la représentation graphique (graphique 9) pour les États-Unis et la France de la relation entre le chômage et ϕ , nous remarquons que la tendance de la courbe de la France a une pente moins importante que celle des États-Unis, ce qui nous permet d'émettre l'hypothèse qu'une politique de désinflation a moins d'effets négatifs aux États-Unis qu'en France. Nous pouvons donc en conclure qu'une politique de désinflation est plus coûteuse en termes de chômage en France qu'aux États-Unis.



Graphique 9 – Relation entre le chômage et ϕ pour la France et pour les États-Unis

1. À titre d'exercice, le lecteur vérifiera les résultats proposés par le logiciel, cf. chapitre 3, IV. B.

8. Introduction aux modèles à équations simultanées

Ce chapitre est une introduction¹ aux modèles dont la spécification nécessite l'écriture d'équations multiples reliées entre elles au travers de variables figurant dans plusieurs équations.

Nous ne pouvons pas, sauf cas particulier, utiliser la méthode des MCO équation par équation comme si chacune de ces équations était indépendante les unes des autres.

En I., nous envisageons la présentation matricielle d'un modèle à équations simultanées et à son écriture sous forme réduite (les variables endogènes sont exprimées en fonction des seules variables exogènes).

Le problème de l'identification, c'est-à-dire la possibilité d'estimation du modèle sous forme réduite, est étudié en II. Enfin, en III., nous présentons les méthodes d'estimation spécifiques des modèles à équations simultanées.

Les exercices d'application se trouvent exceptionnellement tous en fin de chapitre, la partie théorique étant relativement réduite.

1. Dans le cadre de ce manuel, nous limitons volontairement les développements théoriques qui peuvent se révéler complexes dans le domaine de l'économétrie en général, et des équations simultanées en particulier. Le lecteur désirant approfondir ce chapitre peut se référer à Greene W.H. chapitre 16, 2000.

I. Équations structurelles et équations réduites

Lorsque nous sommes en présence d'un modèle linéaire à équations multiples, il arrive fréquemment qu'une variable endogène d'une équation apparaisse en tant que variable explicative d'une autre équation. Ce double statut de certaines variables entraîne un biais dans les estimations des coefficients lorsque nous employons les MCO, équation par équation.

Nous allons donc chercher à transformer le modèle initial en un modèle où les variables endogènes ne sont exprimées qu'en fonction des variables exogènes.

A. Exemple introductif

Considérons le modèle macro-économique à trois équations :

$$C_t = a_0 + a_1 Y_t + \varepsilon_{1t} \quad [\text{E1}]$$

$$I_t = b_0 + b_1 Y_{t-1} + \varepsilon_{2t} \quad [\text{E2}]$$

$$Y_t = C_t + I_t \quad [\text{E3}]$$

avec :

C_t = consommation totale pour l'année t ;

I_t = investissement total pour l'année t ;

Y_t = revenu total pour l'année t .

Ce système d'équations multiples, spécifié par l'économiste, qui traduit directement les relations entre les variables, s'appelle : *le système d'équations structurelles*.

Ce modèle comporte trois équations dont une identité [E3]. En effet, dans la relation [E3], il n'y a aucun coefficient à estimer et par conséquent pas de terme aléatoire. L'équation [E1] est une fonction de consommation et l'équation [E2] est relative à l'investissement.

Ce système contient trois variables endogènes : C_t , I_t , Y_t et une variable exogène Y_{t-1} . Nous remarquons, par exemple, que la variable Y_t apparaît comme variable explicative en [E1], ce qui est contraire à son statut de variable endogène.

Pour lever ce problème, nous allons exprimer les trois variables endogènes (C_t, I_t, Y_t) en fonction de la seule variable exogène (Y_{t-1}).

En substituant [E3] dans [E1], nous obtenons :

$$C_t = a_0 + a_1(C_t + I_t) + \varepsilon_{1t} = a_0 + a_1C_t + a_1I_t + \varepsilon_{1t}$$

En remplaçant I_t par son expression [E2] et en faisant passer a_1C_t à gauche de l'équation, nous pouvons écrire :

$$(1 - a_1)C_t = a_0 + a_1b_0 + a_1b_1Y_{t-1} + a_1\varepsilon_{2t} + \varepsilon_{1t}$$

ou encore :

$$C_t = \frac{a_0 + a_1b_0}{1 - a_1} + \frac{a_1b_1}{1 - a_1} Y_{t-1} + \frac{a_1\varepsilon_{2t} + \varepsilon_{1t}}{1 - a_1}$$

Il en résulte que :

$$Y_t = C_t + I_t = \frac{a_0 + a_1b_0}{1 - a_1} + b_0 + \left[\frac{a_1b_1}{1 - a_1} + b_1 \right] Y_{t-1} + \frac{a_1\varepsilon_{2t} + \varepsilon_{1t}}{1 - a_1} + \varepsilon_{2t}$$

$$Y_t = \frac{a_0 + b_0}{1 - a_1} + \frac{b_1}{1 - a_1} Y_{t-1} + \frac{\varepsilon_{2t} + \varepsilon_{1t}}{1 - a_1}$$

Les équations structurelles sont alors équivalentes aux *équations réduites* (les variables endogènes sont exprimées en fonction des seules variables exogènes) :

$$C_t = \frac{a_0 + a_1b_0}{1 - a_1} + \frac{a_1b_1}{1 - a_1} Y_{t-1} + \frac{a_1\varepsilon_{2t} + \varepsilon_{1t}}{1 - a_1} \quad [\text{E4}]$$

$$Y_t = \frac{a_0 + b_0}{1 - a_1} + \frac{b_1}{1 - a_1} Y_{t-1} + \frac{\varepsilon_{2t} + \varepsilon_{1t}}{1 - a_1} \quad [\text{E5}]$$

$$I_t = b_0 + b_1Y_{t-1} + \varepsilon_{2t} \quad [\text{E6}]$$

L'équation [E5] indique que la variable Y_t est fonction de ε_{1t} et par conséquent $E(Y_t | \varepsilon_{1t}) \neq 0$. Il en résulte que, dans l'équation [E1], l'hypothèse d'indépendance entre la « variable explicative » Y_t et l'erreur ε_{1t} n'est pas respectée et l'application des MCO sur le modèle [E1] conduit à des estimateurs biaisés et non convergents.

En revanche, l'utilisation des MCO sur les équations réduites est licite puisque la variable Y_{t-1} est indépendante de ε_{1t} et ε_{2t} .

Il est noter que la forme réduite permet de mesurer l'effet total, direct et indirect, d'une modification de la variable exogène Y_{t-1} sur les variables endogènes. Par exemple, le schéma suivant illustre la cascade des causalités :

[E2] [E3] [E1]

$Y_{t-1} \rightarrow I_t \rightarrow Y_t \rightarrow C_t$ ce qui est résumé dans [E4].

B. Le modèle général

Soit le système général de g équations structurelles liant g variables endogènes à k variables prédéterminées :

$$\begin{aligned} b_{11}y_{1t} + b_{12}y_{2t} + \dots + b_{1g}y_{gt} + c_{11}x_{1t} + c_{12}x_{2t} + \dots + c_{1k}x_{kt} &= \varepsilon_{1t} \\ b_{21}y_{1t} + b_{22}y_{2t} + \dots + b_{2g}y_{gt} + c_{21}x_{1t} + c_{22}x_{2t} + \dots + c_{2k}x_{kt} &= \varepsilon_{2t} \\ \dots & \\ b_{g1}y_{1t} + b_{g2}y_{2t} + \dots + b_{gg}y_{gt} + c_{g1}x_{1t} + c_{g2}x_{2t} + \dots + c_{gk}x_{kt} &= \varepsilon_{gt} \end{aligned}$$

Soit, sous forme matricielle :

$$\begin{matrix} B & Y & + & C & X & = & \varepsilon \\ (g, g) & (g, 1) & & (g, k) & (k, 1) & & (g, 1) \end{matrix} \quad [1]$$

Bien entendu, dans chaque équation, quelques coefficients sont nuls et la variable dont le coefficient est égal à 1 est la variable dépendante. Si une équation a tous ses coefficients, non nuls, égaux à 1 et ne comporte pas de terme aléatoire, cela signifie qu'il s'agit d'une identité (aucun coefficient n'est à estimer).

Si la matrice B est régulière, nous passons de la forme structurelle à la forme réduite en exprimant Y en fonction de X , soit :

$$Y = -B^{-1}CX + B^{-1}\varepsilon \quad [2]$$

et nous pouvons appliquer les MCO. En effet, les erreurs $B^{-1}\varepsilon$ sont indépendantes de X .

Si, sur le plan de la présentation, la formalisation est simple, sa mise en œuvre pratique est plus complexe. En effet, la connaissance des $g \times k$ éléments de la matrice $(-B^{-1}C)$ ne permet pas de déterminer – d'identifier – la matrice B (composée de $g \times g$ éléments) ainsi que la matrice C (composée de $g \times k$ éléments). Nous sommes alors en présence de $g \times k$ équations à $(g \times g) + (g \times k)$ inconnues qui, sans restrictions supplémentaires, s'avèrent impossibles à résoudre, il s'agit du problème de l'identification.

Nous allons donc présenter en II. des règles simples permettant, à partir de restrictions sur les paramètres, de déterminer des conditions d'identification.

À titre d'illustration le modèle introductif peut s'écrire sous forme matricielle :

$$Y = \begin{bmatrix} C_t \\ I_t \\ Y_t \end{bmatrix}; \quad B = \begin{bmatrix} 1 & 0 & -a_1 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix}; \quad X = \begin{bmatrix} U \\ Y_{t-1} \end{bmatrix}; \quad C = \begin{bmatrix} -a_0 & 0 \\ -b_0 & -b_1 \\ 0 & 0 \end{bmatrix};$$

$$\varepsilon = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ 0 \end{bmatrix} \quad (U \text{ est le vecteur unité.})$$

C. Cas particulier : les modèles récurrents

Un système d'équations est appelé « récurrent » si chaque variable endogène peut être déterminée de manière séquentielle :

$$\begin{aligned}y_{1t} &= f(x_{1t}, x_{2t} \dots x_{kt}; \varepsilon_{1t}) \\y_{2t} &= f(x_{1t}, x_{2t} \dots x_{kt}; y_{1t}; \varepsilon_{2t}) \\y_{3t} &= f(x_{1t}, x_{2t} \dots x_{kt}; y_{1t}, y_{2t}; \varepsilon_{3t})\end{aligned}$$

(Les variables aléatoires (ε_{it}) sont, par hypothèse, indépendantes.)

Nous observons que la première équation ne contient aucune variable endogène. La deuxième équation a pour variable « explicative » endogène, la variable de la première équation, et ainsi de suite.

Lorsqu'un modèle à équations simultanées répond au critère de récursivité, il est licite d'appliquer la méthode des MCO équation par équation. En effet, dans ce cas particulier, il y a indépendance entre les variables endogènes et les erreurs. Par exemple, dans la deuxième équation, y_{1t} dépend de ε_{1t} mais pas de ε_{2t} .

Les modèles récurrents sont aussi appelés « système triangulaire » car les coefficients des variables endogènes forment un triangle au sein de la matrice B .

II. Le problème de l'identification

A. Restrictions sur les coefficients

Il y a une restriction sur un coefficient de la forme structurelle, chaque fois qu'un paramètre est contraint – par l'écriture du modèle – à être égal à une valeur déterminée. Nous distinguons deux types de restrictions.

1) Restrictions d'exclusion

Nous pouvons considérer que chaque fois qu'une variable endogène ou exogène n'apparaît pas dans une équation structurelle, cela revient à l'affecter d'un coefficient nul. Par exemple, dans notre modèle introductif, la variable I_t ne figure pas dans l'équation [E1], son coefficient est donc nul : dans la matrice B , l'élément de la première ligne et de la deuxième colonne est égal à 0.

2) Restrictions linéaires

Certaines spécifications de modèle imposent que des variables soient affectées d'un coefficient identique, il s'agit là encore de restrictions *a priori* sur les paramètres du modèle.

B. Conditions d'identification

Les conditions d'identification se déterminent équation par équation. Nous pouvons distinguer trois cas d'identifications :

- le modèle est sous-identifié si une équation du modèle est sous-identifiable (il y a moins d'équations que de paramètres à identifier dans la forme structurelle, le système est donc impossible à résoudre) ;
- le modèle est juste identifié si toutes les équations sont justes identifiables ;
- le modèle est sur-identifié si les équations du modèle sont soit justes identifiables, soit sur-identifiables.

Si le modèle est sous-identifié, il n'existe aucune possibilité d'estimation des paramètres du modèle, celui-ci doit être respécifié.

Les conditions d'identification peuvent faire l'objet d'un développement complexe, nous nous bornons ici à édicter des règles simples qui sont, dans la pratique, appliquées en premier lieu.

Soit :

g = nombre de variables endogènes du modèle (ou encore nombre d'équations du modèle) ;

k = nombre de variables exogènes du modèle ;

g' = nombre de variables endogènes figurant dans une équation ;

k' = nombre de variables exogènes figurant dans une équation.

Lorsque les restrictions ne sont que des restrictions d'exclusion, les conditions – nécessaires – d'identifiabilité s'énoncent ainsi :

► $g - 1 > g - g' + k - k' \rightarrow$ l'équation est sous-identifiée ;

► $g - 1 = g - g' + k - k' \rightarrow$ l'équation est juste identifiée ;

► $g - 1 < g - g' + k - k' \rightarrow$ l'équation est sur-identifiée.

Ce qui peut se résumer ainsi : pour qu'une équation ne soit pas sous-identifiée, le nombre de variables exclues de l'équation doit être au moins égal au nombre d'équations du modèle moins 1.

Lorsque nous avons r restrictions, autres que celles d'exclusion, concernant les paramètres d'une équation (égalité de deux coefficients, par exemple), les conditions précédentes deviennent :

► $g - 1 > g - g' + k - k' + r \rightarrow$ l'équation est sous-identifiée ;

► $g - 1 = g - g' + k - k' + r \rightarrow$ l'équation est juste identifiée ;

► $g - 1 < g - g' + k - k' + r \rightarrow$ l'équation est sur-identifiée.

Ces conditions – nécessaires – sont appelées *conditions d'ordre* d'identifiabilité.

Il convient de vérifier des conditions suffisantes, qualifiées de *conditions de rang* qui, dans la pratique, se révèlent difficiles, voire parfois impossibles à mettre en œuvre. En annexe, ces conditions sont exposées.

III. Les méthodes d'estimation

Les méthodes d'estimation que nous pouvons utiliser dans le cadre des équations simultanées sont fonction du critère d'identifiabilité du modèle.

- Si le modèle est sous-identifiable ; pas d'estimation possible.

Dans le cas d'un modèle juste ou sur-identifiable, nous pouvons distinguer la méthode à employer équation par équation, selon le critère d'identifiabilité :

- Si l'équation est juste identifiée : les moindres carrés indirects ou les doubles moindres carrés.
- Si l'équation est sur-identifiée : les doubles moindres carrés.

A. Les moindres carrés indirects

La méthode des Moindres Carrés Indirects (MCI) consiste à appliquer la méthode des MCO aux équations justes identifiées du modèle sous formes réduite ; elle se compose de trois étapes :

- mise sous forme réduite du modèle structurel ;
- estimation par les MCO des paramètres de chacune des équations ;
- détermination des coefficients des équations structurelles à partir des relations algébriques entre coefficients réduits et structurels (puisque le modèle est juste identifiable, la solution est unique).

L'estimateur des MCI de la forme réduite est BLUE. Cependant, l'estimateur des coefficients de la forme structurelle, obtenu à partir de l'estimateur optimal des MCI, est biaisé pour les petits échantillons. Les propriétés asymptotiques font tendre ce biais vers 0, lorsque la taille de l'échantillon augmente.

Cet estimateur des MCI est rarement utilisé du fait de la difficulté, pour les modèles importants, de déterminer la forme réduite du modèle. Nous lui préférons l'estimateur des doubles moindres carrés, plus simple à mettre en œuvre et qui fournit les mêmes résultats que les MCI pour les équations justes identifiées.

B. Les doubles moindres carrés

La procédure d'estimation des Doubles Moindres Carrés (DMC) est la plus utilisée en pratique. Elle s'applique pour tous les modèles justes ou sur-identifiables.

Cette méthode des DMC est fondée, comme son nom l'indique, sur l'application en deux étapes des MCO.

Soit le modèle à équations simultanées à g variables endogènes et k variables exogènes :

$$\begin{aligned} b_{11}y_{1t} + b_{12}y_{2t} + \dots + b_{1g}y_{gt} + c_{11}x_{1t} + c_{12}x_{2t} + \dots + c_{1k}x_{kt} &= \varepsilon_{1t} \\ b_{21}y_{1t} + b_{22}y_{2t} + \dots + b_{2g}y_{gt} + c_{21}x_{1t} + c_{22}x_{2t} + \dots + c_{2k}x_{kt} &= \varepsilon_{2t} \\ \dots & \\ b_{g1}y_{1t} + b_{g2}y_{2t} + \dots + b_{gg}y_{gt} + c_{g1}x_{1t} + c_{g2}x_{2t} + \dots + c_{gk}x_{kt} &= \varepsilon_{gt} \end{aligned}$$

La première étape consiste à effectuer une régression de chacune des variables endogènes sur toutes les variables exogènes :

$$\begin{aligned} y_{1t} &= \alpha_{11}x_{1t} + \alpha_{12}x_{2t} + \dots + \alpha_{1k}x_{kt} + u_{1t} \\ y_{2t} &= \alpha_{21}x_{1t} + \alpha_{22}x_{2t} + \dots + \alpha_{2k}x_{kt} + u_{2t} \\ \dots & \\ y_{gt} &= \alpha_{g1}x_{1t} + \alpha_{g2}x_{2t} + \dots + \alpha_{gk}x_{kt} + u_{gt} \end{aligned}$$

Puis, dans une deuxième étape, il convient de remplacer les variables endogènes figurant à droite des équations structurelles par leurs valeurs ajustées à l'aide des modèles estimés :

$$\begin{aligned} y_{1t} &= \beta_{12}\widehat{y}_{2t} + \dots + \beta_{1g}\widehat{y}_{gt} + c_{11}x_{1t} + c_{12}x_{2t} + \dots + c_{1k}x_{kt} + \varepsilon_{1t} \\ y_{2t} &= \beta_{21}\widehat{y}_{1t} + \dots + \beta_{2g}\widehat{y}_{gt} + c_{21}x_{1t} + c_{22}x_{2t} + \dots + c_{2k}x_{kt} + \varepsilon_{2t} \\ \dots & \\ y_{gt} &= \beta_{g1}\widehat{y}_{1t} + \beta_{g2}\widehat{y}_{2t} + \dots + c_{g1}x_{1t} + c_{g2}x_{2t} + \dots + c_{gk}x_{kt} + \varepsilon_{gt} \end{aligned}$$

Les propriétés de l'estimateur des DMC sont identiques, de manière asymptotique, à celles d'un estimateur classique ; c'est-à-dire que pour les petits échantillons les estimations des paramètres peuvent être biaisées.

Il est à noter que l'estimateur des DMC peut s'interpréter comme étant un estimateur de la méthode des Variables Instrumentales¹, les variables exogènes des autres équations étant les instruments.

C. Autres méthodes d'estimation

Nous pouvons citer trois méthodes :

- La méthode des triples moindres carrés est adaptée lorsque les erreurs sont autocorrélées et/ou hétéroscédastiques, elle consiste à déterminer l'estimateur des doubles moindres carrés puis à calculer l'estimateur des MCG.
- La méthode du maximum de vraisemblance à information complète (MVIC) consiste à maximiser la fonction Log-vraisemblance associée au modèle.

1. Greene W.H., pages 681-684, 2000.

- La méthode des moments généralisée est utilisée en cas de présence d'hétéroscédasticité.

Exercice n° 1

↓ fichier C8EX1

Comparaison des résultats par moindres carrés ordinaires, moindres carrés indirects et doubles moindres carrés

Soit le modèle suivant sous forme structurelle :

$$Y_{1t} = a(Y_{2t} + X_{1t}) + \varepsilon_{1t} \quad [E1]$$

$$Y_{2t} = bY_{1t} + cY_{1t-1} + \varepsilon_{2t} \quad [E2]$$

avec :

Y_{1t} = Produit National Brut pour l'année t ;

Y_{2t} = consommation des ménages pour l'année t ;

X_{1t} = demande finale autre que consommation des ménages pour l'année t .

Nous disposons des données du tableau 1 dans lequel les variables ont été préalablement centrées.

Tableau 1 – Variables macro-économiques centrées

t	Y_{1t}	Y_{2t}	X_{1t}	Y_{1t-1}
1	- 30,0	- 14,0	- 26,0	20,0
2	- 4,0	10,0	5,0	- 30,0
3	- 19,0	- 19,0	- 36,0	- 4,0
4	- 6,0	- 11,0	- 6,0	- 19,0
5	- 9,0	6,0	- 13,0	- 6,0
6	11,0	- 12,0	25,0	- 9,0
7	9,0	10,0	5,0	11,0
8	19,0	11,0	32,0	9,0
9	29,0	19,0	14,0	19,0

- 1) Préciser les conditions d'identifiabilité du modèle.
- 2) Estimer les équations de la forme structurelle par les MCO et démontrer que les estimateurs ne sont pas BLUE.
- 3) Estimer les coefficients par les MCI et les DMC.
- 4) Comparer les résultats obtenus.

Solution

- 1) Nous nous référons aux conditions d'identification développées en II.

Le modèle comporte deux variables endogènes, Y_{1t} et Y_{2t} soit $g = 2$, et deux variables exogènes, X_{1t} et Y_{1t-1} soit $k = 2$.

La première équation [E1] comporte une restriction d'exclusion et une restriction de contrainte (les coefficients de Y_2 et X_1 sont égaux). Soit, en appliquant les conditions d'identifiabilité, les variables figurant dans l'équation donnent : $g' = 2$, $k' = 1$ et la contrainte d'égalité des coefficients : $r = 1$.

Soit : $g - g' + k - k' + r = 2 - 2 + 2 - 1 + 1 = 2 > g - 1 = 1$, l'équation [E1] est sur-identifiée.

Pour l'équation [E2], nous avons $g' = 2$, $k' = 1$ et $r = 0$, soit

$$g - g' + k - k' + r = 2 - 2 + 2 - 1 + 0 = 1 = g - 1,$$

l'équation [E2] est juste identifiée.

Nous pouvons donc estimer les coefficients du modèle.

- 2) Les résultats de l'estimation par les MCO sont les suivants après avoir créé la nouvelle variable explicative : somme de Y_2 et X_1 . Il est à noter que le modèle ne comporte pas de terme constant puisque les données sont centrées.

[E3]

$$Y_{1t} = 0,516(Y_{2t} + X_{1t}) + e_{1t}$$

$$(5,89)$$

$$R^2 = 0,81$$

$$n = 9$$

$$(\cdot) = t \text{ de Student}$$

[E4]

$$Y_{2t} = 0,518 Y_{1t} + 0,040 Y_{1t-1} + e_{2t}$$

$$(2,58) \quad (0,18)$$

$$R^2 = 0,50$$

$$n = 9$$

Ces estimateurs ne sont pas BLUE. En effet, si on explicite Y_{1t} en fonction de X_{1t} et de Y_{1t-1} , on obtient :

$$Y_{1t} = \frac{a}{1-ab} X_{1t} + \frac{ac}{1-ab} Y_{1t-1} + \frac{\varepsilon_{2t}}{1-ab}$$

Or, dans l'équation [E2], on régresse Y_{2t} sur Y_{1t} qui dépend aléatoirement de ε_{2t} , ce qui est en contradiction avec l'hypothèse d'indépendance entre la variable explicative et l'erreur. Les estimateurs ne sont donc pas BLUE.

- 3) Pour appliquer la méthode des MCI, il convient d'abord de mettre le modèle sous forme réduite, soit :

$$Y_{1t} = \frac{a}{1-ab} X_{1t} + \frac{ac}{1-ab} Y_{1t-1} + \frac{\varepsilon_{2t}}{1-ab} + \frac{\varepsilon_{1t}}{1-ab} \quad [E5]$$

$$Y_{2t} = \frac{ab}{1-ab} X_{1t} + \frac{c}{1-ab} Y_{1t-1} + \frac{\varepsilon_{2t} + b\varepsilon_{1t}}{1-ab} \quad [E6]$$

ou encore :

$$Y_{1t} = \alpha_1 X_{1t} + \alpha_2 Y_{1t-1} + v_{1t}$$

$$Y_{2t} = \alpha_3 X_{1t} + \alpha_4 Y_{1t-1} + v_{2t}$$

Les résultats d'estimation par les MCO sont les suivants :

[E7]

$$Y_{1t} = 0,717 X_{1t} + 0,190 Y_{1t-1} + e_{1t}$$

$$(4,93) \quad (1,0)$$

$$R^2 = 0,78$$

$$n = 9$$

$$(\cdot) = t \text{ de Student}$$

[E8]

$$Y_{2t} = 0,355 X_{1t} + 0,139 Y_{1t-1} + e_{2t}$$

$$(1,92) \quad (1,0)$$

$$R^2 = 0,36$$

$$n = 9$$

Or le modèle sous forme matricielle s'écrit :

$$\begin{bmatrix} 1 & -a \\ -b & 1 \end{bmatrix} \times \begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} + \begin{bmatrix} -a & 0 \\ 0 & -c \end{bmatrix} \times \begin{bmatrix} X_{1t} \\ Y_{1t-1} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

$$B \quad \times \quad Y \quad + \quad C \quad \times \quad X \quad = \quad \varepsilon$$

ou encore $Y = -B^{-1}CX + B^{-1}\varepsilon = AX + v$ qui est le modèle sous forme réduite avec $A = -B^{-1}C$ (A est la matrice des coefficients estimés de la forme réduite).

Nous avons donc $BA = -C$, soit :

$$\begin{bmatrix} 1 & -a \\ -b & 1 \end{bmatrix} \begin{bmatrix} 0,717 & 0,190 \\ 0,355 & 0,139 \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & c \end{bmatrix}$$

ce qui donne :

$$0,717 - a \times 0,355 = a$$

$$0,190 - a \times 0,139 = 0$$

$$-b \times 0,717 + 0,355 = 0$$

$$-b \times 0,190 + 0,139 = c$$

Nous pouvons, à partir des deux dernières relations, estimer $\hat{b} = 0,495$ et $\hat{c} = 0,045$ (puisque l'équation [E2] est juste identifiée), mais, en revanche, le coefficient a est indéterminé (l'équation [E1] est sur-identifiée).

Nous appliquons maintenant la méthode des DMC afin d'estimer tous les coefficients de ce modèle.

Nous allons développer le calcul pour la première équation.

Étape 1 : régresser la variable endogène Y_{2t} sur toutes les variables exogènes (X_{1t} et Y_{1t-1}).

Le résultat est déjà connu, puisqu'il s'agit de l'équation [E8].

Étape 2 : calculer la série ajustée : \hat{Y}_{2t}

soit :

$$\hat{Y}_{2t} = 0,355 X_{1t} + 0,139 Y_{1t-1}$$

Étape 3 : la variable endogène Y_{2t} est remplacée par sa valeur ajustée \hat{Y}_{2t} dans l'équation [E1] : $Y_{1t} = a(\hat{Y}_{2t} + X_{1t}) + \varepsilon_{1t}$.

Le résultat d'estimation par les MCO est le suivant :

$$Y_{1t} = 0,535(\hat{Y}_{2t} + X_{1t}) + e_{1t}$$

(5,22)

$$R^2 = 0,77$$

$$n = 9$$

$$(\cdot) = t \text{ de Student}$$

Cette procédure des DMC est donc un peu lourde à mettre en œuvre, c'est pourquoi les logiciels permettent en une seule instruction d'utiliser cette méthode. Par exemple, pour estimer la deuxième équation de notre modèle, l'instruction de Eviews est la suivante :

TSLS Y2 Y1(-1) Y1 @ X1 Y1(- 1)

où TSLS est l'abréviation de Two-Stage Least Squares. La variable à expliquer (Y2) figure suivie des variables exogènes – Y1(- 1), la seule dans cette équation – et, enfin, on cite, la variable endogène que l'on désire remplacer par sa valeur ajustée à l'aide des séries exogènes (elles sont séparées de la variable endogène par le signe @).

Le résultat, fourni par Eviews, est le suivant :

Dependent Variable : Y2
Method: Two-Stage Least Squares
Included observations: 9
Instrument list: X1 Y1D

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Y1D	0.045084	0.220336	0.204614	0.8437
Y1	0.495344	0.228271	2.169985	0.0666
R-squared	0.503610	Mean dependent var		0.000000
Adjusted R-squared	0.432697	S.D. dependent var		13.87444
S.E. of regression	10.45016	Sum squared resid		764.4403
F-statistic	5.176628	Durban-Watson stat		3.011822
Prob(F-statistic)	0.057037			

Nous remarquons que l'estimation des coefficients est exactement la même que par la méthode des MCI. En effet, nous rappelons que la méthode des MCI fournit des résultats strictement identiques à la méthode des DMC. C'est pourquoi, dans la pratique, seule cette dernière est utilisée car elle se révèle d'un maniement bien plus aisé.

4) Nous pouvons au sein du tableau 2 comparer l'ensemble des résultats obtenus.

Tableau 2 – Résultats d'estimation par les MCO et DMC

	MCO	DMC
<i>a</i>	0,516 (5,89)	0,535 (5,29)
<i>b</i>	0,518 (2,58)	0,495 (2,17)
<i>c</i>	0,040 (0,18)	0,045 (0,20)

À la lecture de ces résultats, nous constatons des différences assez faibles entre les estimations des MCO et celles des DMC. En effet, les propriétés des estimateurs des DMC ne sont qu'asymptotiques, or le nombre d'observations dans cet exemple est limité à 9.

Exercice n° 2

Le problème de l'identification lorsque la structure d'un modèle change

On considère le système d'équations :

$$Y_{1t} + aY_{2t} + b = \varepsilon_{1t} \quad [\text{E1}]$$

$$cY_{1t} + Y_{2t} + dZ_t + e = \varepsilon_{2t} \quad [\text{E2}]$$

dans lequel Y_{1t} et Y_{2t} sont des variables endogènes et Z_t une variable exogène.

- 1) Mettre ce système sous forme matricielle et étudier les conditions d'identifiabilité.
- 2) Vérifier les conditions d'identifiabilité à partir du calcul matriciel.
- 3) Répondre aux mêmes questions avec $e = 0$.

Solution

1) Le modèle sous forme matricielle s'écrit :

$$\begin{bmatrix} 1 & a \\ c & 1 \end{bmatrix} \begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} + \begin{bmatrix} 0 & b \\ d & e \end{bmatrix} \begin{bmatrix} Z_t \\ U_t \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \text{ soit } BY + CX = \varepsilon$$

avec $U =$ vecteur unité.

Nous avons $g = 2$ variables endogènes : Y_{1t} , Y_{2t} et $k = 2$ variables exogènes Z_t et U . Examinons les conditions d'identifiabilité :

- pour l'équation [E1] : $g' = 2$, $k' = 1$ et $r = 0$ soit $g - g' + k - k' + r = 1 = g - 1$, l'équation [E1] est juste identifiée ;
- pour l'équation [E2] : $g' = 2$, $k' = 2$ et $r = 0$ soit $g - g' + k - k' + r = 0 < g - 1$, l'équation [E2] est sous-identifiée.

2) Par le calcul matriciel, nous allons exprimer les variables endogènes par les variables exogènes : $Y = -B^{-1}CX + B^{-1}\varepsilon$.

Posons $A = -B^{-1}C$

$$B^{-1} = \frac{1}{1-ac} \begin{bmatrix} 1 & -a \\ -c & 1 \end{bmatrix} \rightarrow A = \frac{-1}{1-ac} \begin{bmatrix} -ad & b - ae \\ d & -bc + e \end{bmatrix}$$

Cette matrice A correspond à la matrice des coefficients estimés (notés $\hat{\alpha}_i$) de la forme réduite.

Nous avons donc la relation $BA = -C$:

$$\begin{bmatrix} 1 & a \\ c & 1 \end{bmatrix} \times \begin{bmatrix} \hat{\alpha}_1 & \hat{\alpha}_2 \\ \hat{\alpha}_3 & \hat{\alpha}_4 \end{bmatrix} = - \begin{bmatrix} 0 & b \\ d & e \end{bmatrix}$$

soit en explicitant cette relation matricielle :

$$\hat{\alpha}_1 + a\hat{\alpha}_3 = 0 \quad [E3]$$

$$c\hat{\alpha}_1 + \hat{\alpha}_3 = -d \quad [E5]$$

$$\hat{\alpha}_2 + a\hat{\alpha}_4 = -b \quad [E4]$$

$$c\hat{\alpha}_2 + \hat{\alpha}_4 = -e \quad [E6]$$

Les équations [E3] et [E4] constituent deux équations à deux inconnues. Les coefficients a et b sont donc parfaitement identifiables dès lors que l'on connaît les coefficients de la forme réduite. Ces deux coefficients a et b appartiennent à l'équation [E1] qui est juste identifiée, ce qui est une vérification des règles énoncées.

En revanche, les équations [E5] et [E6] forment deux équations à trois inconnues, soit un système indéterminé. L'équation [E3] est bien sous-identifiée.

3) Si $e = 0$, l'équation [E1] reste juste identifiée et l'équation [E2] devient, dès lors, juste identifiée. Matriciellement, le système s'écrit :

$$\begin{bmatrix} 1 & a \\ c & 1 \end{bmatrix} \times \begin{bmatrix} \hat{\alpha}_1 & \hat{\alpha}_2 \\ \hat{\alpha}_3 & \hat{\alpha}_4 \end{bmatrix} = - \begin{bmatrix} 0 & b \\ d & 0 \end{bmatrix}$$

Soit, en explicitant cette relation matricielle :

$$\hat{\alpha}_1 + a\hat{\alpha}_3 = 0 \quad [E7]$$

$$c\hat{\alpha}_1 + \hat{\alpha}_3 = -d \quad [E9]$$

$$\hat{\alpha}_2 + a\hat{\alpha}_4 = -b \quad [E8]$$

$$c\hat{\alpha}_2 + \hat{\alpha}_4 = 0 \quad [E10]$$

Soit quatre équations à quatre inconnues, le système est maintenant parfaitement déterminé.

Exercice n° 3

↓ fichier C8EX3

Le modèle de Klein

Ce modèle¹ macro-économique fournit un excellent exemple pédagogique des problèmes qui peuvent se poser dans le cadre des équations simultanées.

Soit le modèle suivant :

$$Cons_t = a_0 + a_1 P_t + a_2 P_{t-1} + a_3 (W_t + W'_t) + \varepsilon_{1t} \quad [E1]$$

$$I_t = b_0 + b_1 P_t + b_2 P_{t-1} + b_3 K_{t-1} + \varepsilon_{2t} \quad [E2]$$

$$W_t = c_0 + c_1 X_t + c_2 X_{t-1} + c_3 t + \varepsilon_{3t} \quad [E3]$$

$$X_t = Cons_t + I_t + G_t \quad [E4]$$

$$P_t = X_t - W_t - Tax_t \quad [E5]$$

$$K_t = I_t + K_{t-1} \quad [E6]$$

avec :

$Cons_t$ = consommation pour l'année t ;

P_t = profits de l'année t ;

W_t et W'_t = salaires de l'année t versés par l'industrie et l'administration ;

I_t = investissements de l'année t ;

1. Klein L. R., 1950.

K_t = stock de capital à la fin de l'année t ;
 X_t = production industrielle de l'année t ;
 G_t = dépenses publiques de l'année t ;
 Tax_t = impôts sur les profits de l'année t .

- 1) Distinguer les variables endogènes et exogènes.
- 2) Estimer les équations par la méthode des MCO.
- 3) Examiner les conditions d'identifiabilité de ce modèle.
- 4) Estimer ce modèle à l'aide d'une méthode adéquate.

Les données – exprimées dans des unités convenablement choisies – sont présentées sur le tableau 3.

Tableau 3 – Données du modèle de Klein (1950) (extrait des données)

	$Const_t$	G_t	I_t	K_{t-1}	P_t	Tax_t	W_t	W'_t	X_t
1920	39,8	2,4	2,7	180,1	12,7	3,4	28,8	2,2	44,9
1921	41,9	3,9	-0,2	182,8	12,4	7,7	25,5	2,7	45,6
...
1940	65,0	7,4	3,3	201,2	21,1	9,6	45,0	8,0	75,7
1941	69,7	13,8	4,9	204,5	23,5	11,6	53,3	8,5	88,4

(NB : ce tableau indique directement la variable K_{t-1})

Solution

- 1) Ce modèle comporte six équations et onze variables économiques (y compris le vecteur et la tendance). Les trois dernières équations sont des identités (aucun paramètre n'est à estimer).

Les variables exogènes sont les variables non expliquées par une relation :

- les variables retardées : P_{t-1} , X_{t-1} , K_{t-1} et
- les variables « véritablement exogènes » : la tendance t , les dépenses gouvernementales G_t , les salaires des administrations W'_t , les impôts sur les profits Tax_t et le vecteur unité U . Soit, $k = 8$.

Les variables endogènes sont celles qui sont déterminées par une relation de comportement : $Const_t$, I_t , W_t . En ce qui concerne les identités, on choisit comme variables endogènes, compte tenu de leur signification, X_t , P_t et K_t . Soit $g = 6$.

Nous remarquons que, dans ce choix de variables endogènes et exogènes, la réflexion économique joue un grand rôle. Néanmoins, il subsiste une part d'arbitraire représentative du schéma de pensée de l'économiste.

- 2) Les estimations par les MCO sont les suivantes :

$$Const_t = 16,23 + 0,193 P_t + 0,089 P_{t-1} + 0,796(W_t + W'_t) + e_{1t} \quad [E7]$$

(12,4) (2,11) (0,99) (19,9)

$$R^2 = 0,98 \quad n = 21 \quad DW = 1,36 \quad (-) = t \text{ de Student}$$

$$I_t = 10,12 + 0,479 P_t + 0,333 P_{t-1} - 0,111 K_{t-1} + e_{2t} \quad [E8]$$

(1,85) (4,93) (3,30) (4,189)

$$R^2 = 0,93 \quad n = 21 \quad DW = 1,81 \quad (-) = t \text{ de Student}$$

$$W_t = -0,066 + 0,439 X_t + 0,146 X_{t-1} + 0,13 t + e_{3t} \quad [E9]$$

(2,02) (13,5) (3,90) (4,90)

$R^2 = 0,98$ $n = 21$ $DW = 1,95$ $(\cdot) = t$ de Student

Les estimateurs des MCO sont biaisés, cette méthode d'estimation est donc inappropriée pour ce type de modèle. Nous devons étudier les conditions d'identification de ce modèle et utiliser, le cas échéant, la méthode des DMC.

- 3) Les conditions d'identification – conditions d'ordre – s'étudient équation par équation.

Le modèle comporte $g = 6$ variables endogènes et $k = 8$ variables exogènes (cf. la réponse à la question 1).

Pour l'équation [E1], nous avons $g' = 3$, variables endogènes présentes et $k' = 3$, variables exogènes présentes, soit $(g - g' + k - k' = 8)$ restrictions d'exclusion. De plus, l'égalité entre le coefficient de W_t et W'_t introduit une restriction supplémentaire $r = 1$.

$g - g' + k - k' + r = 9 > g - 1 = 5 \rightarrow$ l'équation [E1] est sur-identifiée.

Pour l'équation [E2] : $g' = 2$, $k' = 3$ et $r = 0$ soit :

$g - g' + k - k' + r = 9 > g - 1 = 5$, l'équation [E2] est sur-identifiée.

Pour l'équation [E3] : $g' = 2$, $k' = 3$ et $r = 0$ soit :

$g - g' + k - k' + r = 9 > g - 1 = 5$, l'équation [E3] est sur-identifiée.

Les équations étant toutes sur-identifiées, il est licite d'estimer les coefficients de ce modèle à l'aide de la méthode des DMC.

- 4) Nous avons appliqué la méthode des DMC pour chacune des trois équations à estimer de trois manières différentes.

• Équation [E1] en décomposant la méthode des DMC.

Les variables endogènes P_t et W_t apparaissent à droite de l'équation ; il convient donc de les remplacer par leurs valeurs en fonction des 8 variables exogènes.

Étape 1 : estimation du modèle pour P_t et calcul de la variable ajustée PA_t .

Dependent Variable : P

Sample (adjusted): 1921 1941

Included observations: 21 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	46.55153	23.15829	2.010145	0.0656
P(-1)	0.802501	0.518856	1.546674	0.1459
K	- 0.216103	0.119113	- 1.814267	0.0928
X(-1)	0.022000	0.282164	0.077969	0.9390
WP	- 0.079608	2.533823	- 0.031418	0.9754
TREND	0.319405	0.778129	0.410478	0.6881
G	0.439016	0.391143	1.122394	0.2820
TAX	- 0.923098	0.433759	- 2.128133	0.0530

Étape 2 : estimation pour W_t et calcul de la variable ajustée WA_t .

Dependent Variable : W

Sample (adjusted): 1921 1941

Included observations: 21 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	34.87254	18.61034	1.873826	0.0836
P(-1)	0.871921	0.416960	2.091137	0.0567
K	-0.122952	0.095721	-1.284477	0.2214
X(-1)	0.095329	0.226751	0.420412	0.6811
WP	-0.443726	2.036218	-0.217917	0.8309
TREND	0.713584	0.625316	1.141158	0.2744
G	0.866220	0.314328	2.755783	0.0164
TAX	-0.60415	0.348575	-1.733206	0.1067

Étape 3 : estimation du modèle final en remplaçant dans l'équation [1] P_t par PA_t et W_t par WA_t .

Dependent Variable : CONS

Sample (adjusted): 1921 1941

Included observations: 21 after adjusting

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	16.55476	2.571080	6.438834	0.0000
PA	0.017302	0.229797	0.075294	0.9409
P(-1)	0.216234	0.208810	1.035554	0.3149
WA + WP	0.810183	0.078351	10.34044	0.0000

• Équation [E2] par application directe de la méthode des DMC à l'aide de Eviews, l'instruction unique est la suivante (la liste des instruments – les variables exogènes – figurent après @ :

TSLS I C P(-1) K P @ C P(-1) K X(-1) WP TREND G TAX

Dependent Variable : I

Method: Two-Stage Least Squares

Sample (adjusted): 1921 1941

Included observations: 21 after adjusting endpoints

Instrument list: C P(-1) K X(-1) WP TREND G TAX

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	20.27821	8.383247	2.418897	0.0271
P(-1)	0.615943	0.180926	3.404398	0.0034
K	-0.157788	0.040152	-3.929752	0.0011
P	0.150222	0.192534	0.780238	0.4460

• *Équation [E3]* par application directe de la méthode des DMC à l'aide de RATS, les instructions sont les suivantes :

```
INSTRUMENTS CONSTANT TREND WP TAX G P{1} K X{1}
LINREG(INST) W
# CONSTANT TREND X{1} X
```

Dependent Variable W – Estimation by Instrumental Variables

Annual Data From 1921: 01 To 1941: 01

Usable observations 21 Degrees of Freedom 17

Centered R**2 0.987414 R Bar **2 0.985193

Variable	Coeff	Std Error	T-Stat	Signif
1. Constant	-0.064451360	1.147038221	-0.05619	0.95584607
2. TREND	0.130395687	0.032388389	4.02600	0.00087642
3. X{1}	0.146673821	0.043163948	3.39806	0.00342209
4. X	0.438859065	0.039602662	11.08155	0.00000000

Nous pouvons aussi appliquer la méthode des triples moindres carrés. Le système (SYS01) s'écrit :

$$\text{cons} = C(1) + C(2) \times P + C(3) \times P(-1) + C(4) \times (W + WP)$$

$$I = C(5) + C(6) \times P + C(7) \times P(-1) + C(8) \times K$$

$$W = C(9) + C(10) \times X + C(11) \times X(-1) + C(12) \times \text{@trend}(1919)$$

$$\text{inst C @trend}(1919) \text{ WP TAX G P}(-1) \text{ K X}(-1)$$

Les résultats (SYS01.3SLS) sont les suivants :

System : SYS01

Estimation Method : Three-Stage Least Squares

Sample : 1921 1941

Included observations : 21

Total system (balanced) observations 63

	Coefficient	Std Error	t-Statistic	Prob.
C(1)	16.44079	1.304549	12.60267	0.0000
C(2)	0.124891	0.108129	1.155014	0.2535
C(3)	0.163144	0.100438	1.624322	0.1105
C(4)	0.790081	0.037938	20.82563	0.0000
C(5)	28.17784	6.793769	4.147601	0.0001
C(6)	-0.013079	0.161896	-0.080786	0.9359
C(7)	0.755724	0.152933	4.941532	0.0000
C(8)	-0.194848	0.032531	-5.98974	0.0000
C(9)	0.001128	1.009957	0.001117	0.9991
C(10)	0.400492	0.031813	12.58878	0.0000
C(11)	0.181291	0.034159	5.307308	0.0000
C(12)	0.149674	0.027935	5.357897	0.0000

Equation : $\text{CONS} = C(1) + C(2) \times P + C(3) \times P(-1) + C(4) \times (W + WP)$

Equation : $I = C(5) + C(6) \times P + C(7) \times P(-1) + C(8) \times K$

Equation : $W = C(9) + C(10) \times X + C(11) \times X(-1) + C(12) \times \text{@TREND}(1919)$

Instruments : C @TREND(1919) WP TAX G P(-1) K X(-1)

Enfin la méthode du maximum de vraisemblance à information complète (SYS01.fiml) permet d'obtenir les résultats suivants :

System : SYS01

Estimation Method : Full Information Maximum Likelihood (Marquardt)

Sample : 1921 1941

Included observations : 21

Total system (balanced) observations 63

	Coefficient	Std Error	z-Statistic	Prob.
C(1)	15.83050	4.095952	3.864913	0.0001
C(2)	0.300548	0.408091	0.736472	0.4614
C(3)	0.042528	0.164388	0.258707	0.7959
C(4)	0.780851	0.078507	9.946226	0.0000
C(5)	15.70774	14.35232	1.094439	0.2738
C(6)	0.382189	0.341696	1.118507	0.2634
C(7)	0.409626	0.248385	1.649155	0.0991
C(8)	-0.137683	0.071215	-1.933357	0.0532
C(9)	-0.136911	4.099237	-0.033399	0.9734
C(10)	0.370767	0.129208	2.869534	0.0041
C(11)	0.207299	0.090622	2.287518	0.0222
C(12)	0.184379	0.101556	1.815549	0.0694

Equation : $CONS = C(1) + C(2) \times P + C(3) \times P(-1) + C(4) \times (W + WP)$

Equation : $I = C(5) + C(6) \times P + C(7) \times P(-1) + C(8) \times K$

Equation : $W = C(9) + C(10) \times X + C(11) \times X(-1) + C(12) \times @TREND(1919)$

Annexe

Identification : les conditions de rang

Soit le modèle sous forme matricielle :

$$\begin{matrix} B & Y & + & C & X & = & \varepsilon \\ (g,g) & (g,1) & & (g,k) & (k,1) & & (g,1) \end{matrix}$$

On appelle P la matrice des structures qui est formée par la juxtaposition des matrices B et C , soit $P = [BC]$, P est donc de dimension $(g, g + k)$.

Soit ϕ_i , une matrice des restriction afférente à l'équation i construite telle que : $P_i \phi_{ih} = 0$, i est la i -ième ligne de la matrice P et h la h -ième colonne de la matrice ϕ_i .

Soit $\mu_i = \text{rang}[P\phi_i]$ (rang de la matrice $P\phi_i$), et g le nombre de variables endogènes du modèle, la condition de rang est alors la suivante :

- si $\mu_i < g - 1 \rightarrow$ l'équation i est sous-identifiée ;
- si $\mu_i = g - 1 \rightarrow$ l'équation i est juste identifiée ;
- si $\mu_i > g - 1 \rightarrow$ l'équation i est sur-identifiée.

Cette condition nécessaire et suffisante d'identification est, bien entendu, plus lourde à mettre en œuvre, c'est pourquoi le plus souvent on se satisfait des conditions d'ordre.

À partir du modèle de l'exercice 1, nous allons construire ces matrices et appliquer les conditions de rang.

Ce modèle sous forme matricielle s'écrit :

$$\begin{matrix} \begin{bmatrix} 1 & -a \\ -b & 1 \end{bmatrix} & \times & \begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} & + & \begin{bmatrix} -a & 0 \\ 0 & -c \end{bmatrix} & \times & \begin{bmatrix} X_{1t} \\ Y_{1t-1} \end{bmatrix} & = & \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \\ B & & Y & & C & & X & & \varepsilon \end{matrix}$$

$$\text{Soit } P = [BC] = \begin{bmatrix} 1 & -a & -a & 0 \\ -b & 1 & 0 & c \end{bmatrix}.$$

La matrice ϕ_1 , matrice des restrictions de la première équation, est donc construite de la manière suivante :

- autant de lignes que de variables endogènes et exogènes, chaque ligne est représentative d'une variable (ligne 1 = Y_{1t} , ligne 2 = Y_{2t} , ligne 3 = X_{1t} , ligne 4 = Y_{1t-1}) ;
- une colonne par restriction d'exclusion (variable endogène ou exogène manquante) et par restriction linéaire sur les coefficients ;
- pour les relations d'exclusion, les colonnes sont composées de 0, sauf pour les variables dont le coefficient est nul où l'on met la valeur 1 ;

– pour les restrictions de contraintes sur les coefficients, on explicite la relation entre les coefficients. Dans notre modèle, nous avons $c_{11} - c_{12} = 0$ (Y_2 et X_1 ont le même coefficient).

Nous pouvons alors récrire la condition : $P_i \phi_{ih} = 0$.

Soit, pour l'équation 1, nous avons une seule variable manquante et une restriction sur le coefficient des variables Y_2 et X_1 qui doit être identique, soit deux colonnes :

$$\phi_i = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & -1 \\ 1 & \leftarrow 0 \end{bmatrix} \begin{array}{l} \leftarrow \\ \leftarrow \\ \text{—————} \end{array} \begin{array}{l} \text{les variables } Y_2 \text{ et } X_1 \text{ ont le même coefficient} \\ \text{la variable } Y_{1t-1} \text{ est absente} \end{array}$$

Pour ϕ_2 , nous avons une seule restriction : la variable X_1 est absente de l'équation 2, soit :

$$\phi_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \leftarrow \text{la variable } X_1 \text{ est absente de l'équation 2}$$

$$P\phi_1 = \begin{bmatrix} 1 & -a & -a & 0 \\ -b & 1 & 0 & -c \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ -c & 1 \end{bmatrix}$$

rang de $P\phi_1 = \mu_1 = g - 1$.

L'équation 1 est juste identifiée.

$$P\phi_2 = \begin{bmatrix} 1 & -a & -a & 0 \\ -b & 1 & 0 & -c \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -a \\ 0 \end{bmatrix} \rightarrow \rightarrow \text{rang de } P\phi_2 = \mu_2 = 1 = g - 1$$

L'équation 2 est juste identifiée (sauf si $a = 0$).

Nous remarquons que la condition de rang vient modifier la condition d'ordre de la première équation, puisque d'un statut de sur-identifiée, elle devient juste identifiée. Toutefois, cela ne remet aucunement en cause les résultats trouvés lors de l'exercice 1.

9. Éléments d'analyse des séries temporelles

Ce chapitre est consacré à une présentation sommaire des techniques d'analyse des séries chronologiques. Ce thème, à lui seul, peut faire l'objet de longs développements et de nombreux ouvrages¹ y sont intégralement consacrés. Nous étudions en I. les caractéristiques statistiques – en terme de stationnarité – des séries temporelles en présentant les différents tests (Dickey-Fuller, corrélogramme, etc.) s'y rapportant. Puis en II., nous présentons différentes classes de modèles (AR, MA, ARMA) en étudiant leurs propriétés. Enfin, la méthode de Box et Jenkins qui systématise une démarche d'analyse des séries temporelles fait l'objet de la section III.

I. Stationnarité

A. Définition et propriétés

Avant le traitement d'une série chronologique, il convient d'en étudier les caractéristiques stochastiques. Si ces caractéristiques – c'est-à-dire son espérance et sa variance – se trouvent modifiées dans le temps, la série chronologique est considérée comme non stationnaire ; dans le cas d'un processus stochastique invariant, la série temporelle est alors stationnaire. De manière formalisée, le processus stochastique y_t est stationnaire si :

1. Pour un approfondissement de ce chapitre, se référer au livre de Bourbonnais R. et Terraza M., Dunod, 2010.

$E(y_t) = E(y_{t+m}) = \mu \quad \forall t \text{ et } \forall m$, la moyenne est constante et indépendante du temps ;

$\text{var}(y_t) < \infty \quad \forall t$, la variance est finie et indépendante du temps ;

$\text{cov}(y_t, y_{t+k}) = E[(y_t - \mu)(y_{t+k} - \mu)] = \gamma_k$, la covariance est indépendante du temps.

Il apparaît, à partir de ces propriétés, qu'un processus de bruit blanc¹ ε_t dans lequel les ε_t sont indépendants et de même loi $N(0, \sigma_\varepsilon^2)$ est stationnaire.

Une série chronologique est donc stationnaire si elle est la réalisation d'un processus stationnaire². Ceci implique que la série ne comporte ni tendance, ni saisonnalité et plus généralement aucun facteur n'évoluant avec le temps.

B. Fonctions d'autocorrélation simple et partielle

La *fonction d'autocorrélation* (FAC) est la fonction notée ρ_k qui mesure la corrélation de la série avec elle-même décalée de k périodes, comme l'illustre le tableau 1.

Sa formulation est la suivante :

$$\rho_k = \frac{\text{cov}(y_t, y_{t-k})}{\sigma_{y_t} \sigma_{y_{t-k}}} = \frac{\sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sqrt{\sum_{t=k+1}^n (y_t - \bar{y})^2} \sqrt{\sum_{t=k+1}^n (y_{t-k} - \bar{y})^2}} \quad [1]$$

avec \bar{y} moyenne de la série calculée sur $n - k$ périodes, n = nombre d'observations.

Nous pouvons en déduire que :

$$\rho_0 = 1 \quad \text{et} \quad \rho_k = \rho_{-k}$$

Cette formule [1] est malaisée à manier puisqu'elle exige de recalculer pour chaque terme ρ_k les moyennes et les variances, c'est pourquoi on lui préfère la fonction d'autocorrélation d'échantillonnage :

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2} \quad [2]$$

avec \bar{y} moyenne de la série calculée sur n périodes.

1. Un processus de bruit blanc est une suite de variables aléatoires de même distribution et mutuellement indépendantes. Ce terme est emprunté à la physique faisant référence au spectre de la lumière blanche.
2. Si les conditions énoncées ci-dessous sont vérifiées, on dit que la série est stationnaire de second ordre, ou encore faiblement stationnaire. Dans le cas où, aux conditions définies, s'ajoute la constante de la distribution de probabilité, on parle de série strictement stationnaire.

Tableau 1 – Exemple de calcul d'une fonction d'autocorrélation

k	0	1	2	3	4
t	y_t	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-4}
1	123				
2	130	123			
3	125	130	123		
4	138	125	130	123	
5	145	138	125	130	123
6	142	145	138	125	130
7	141	142	145	138	125
8	146	141	142	145	138
9	147	146	141	142	145
10	157	147	146	141	142
11	150	157	147	146	141
12	145	150	157	147	146

Moyenne y_t	140,75	142,36	143,60	145,67	146,63
Moyenne y_{t-k}	140,75	140,36	139,40	137,44	136,25
Variance y_t	95,02	72,41	62,84	24,11	22,23
Variance y_{t-k}	95,02	101,87	101,84	74,91	71,44
ρ_k	1	0,77	0,62	0,59	0,55

Lorsque le nombre d'observations n est suffisamment grand, les deux formules [1] et [2] donnent des résultats très proches.

La *fonction d'autocorrélation partielle* (FAP) s'apparente à la notion de corrélation partielle. Nous avons défini au chapitre 4 le coefficient de corrélation partielle comme étant le calcul de l'influence de x_1 sur x_2 en éliminant les influences des autres variables x_3, x_4, \dots, x_k .

Par analogie, nous pouvons définir l'autocorrélation partielle de retard k comme le coefficient de corrélation partielle entre y_t et y_{t-k} , c'est-à-dire comme étant la corrélation entre y_t et y_{t-k} l'influence des autres variables décalées de k périodes ($y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$) ayant été retirée.

Afin d'éviter par la suite toutes ambiguïtés entre les deux fonctions d'autocorrélation, nous appelons fonction d'autocorrélation simple, la fonction d'autocorrélation.

C. Tests de « bruit blanc » et de stationnarité

Nous ne pouvons identifier clairement les caractéristiques stochastiques d'une série chronologique que si elle est stationnaire. Cette étude de stationnarité

s'effectue essentiellement à partir de l'étude des fonctions d'autocorrélation (ou de leur représentation graphique appelée « corrélogramme »). Une série chronologique est stationnaire si elle ne comporte ni tendance ni saisonnalité. Nous allons donc, à partir de l'étude du corrélogramme d'une série, essayer de montrer de quelle manière nous pouvons mettre en évidence ces deux composantes.

Nous pouvons distinguer différents types de séries stationnaires :

- à mémoire, c'est-à-dire dont on peut modéliser, par une loi de reproduction, le processus ;
- identiquement et indépendamment distribuée notée *i.i.d.* ou appelée Bruit Blanc (« *White Noise* ») ;
- normalement (selon une loi normale) et indépendamment distribuée notée *n.i.d.* ou appelée Bruit Blanc gaussien.

1) Analyse des fonctions d'autocorrélation

Lorsque nous étudions la fonction d'autocorrélation d'une série chronologique, la question qui se pose est de savoir quels sont les termes ρ_k qui sont significativement différents de 0.

En effet, par exemple, si aucun terme n'est significativement différent de 0, on peut en conclure que le processus étudié est sans mémoire et donc qu'à ce titre il n'est affecté ni de tendance ni de saisonnalité. Ou encore si une série mensuelle présente une valeur élevée pour ρ_{12} (corrélation entre y_t et y_{t-12}), la série étudiée est certainement affectée d'un mouvement saisonnier¹. Nous verrons à partir d'exemples différents cas de figures.

Le test d'hypothèses pour un terme ρ_k est le suivant :

$$H_0 : \rho_k = 0$$

$$H_1 : \rho_k \neq 0$$

Nous pouvons utiliser le test d'hypothèses d'un coefficient de corrélation, présenté au chapitre 1, fondé sur la comparaison d'un t de Student empirique et théorique. Toutefois, Quenouille² a démontré que pour un échantillon de taille importante ($n > 30$), le coefficient ρ_k tend de manière asymptotique vers une loi normale de moyenne 0 et d'écart type $1/\sqrt{n}$.

L'intervalle de confiance du coefficient ρ_k est alors donné par :

$$\rho_k = 0 \pm t^{\alpha/2} \frac{1}{\sqrt{n}}$$

n = nombre d'observations.

1. Il s'agit même d'un test de détection de saisonnalité.

2. Quenouille M. H., 1949.

Si le coefficient calculé $\hat{\rho}_k$ est à l'extérieur de cet intervalle de confiance, il est significativement différent de 0 au seuil α (en général $\alpha = 0,05$ et $t^{\alpha/2} = 1,96$). La plupart des logiciels fournissent, avec le corrélogramme, l'intervalle de confiance, ce qui autorise une interprétation instantanée.

Nous devons souligner une limite des tests à 5 %. En effet, lorsqu'une fonction d'autocorrélation est calculée pour un nombre important de retards, nous pouvons nous attendre à ce que quelques-uns soient, de manière fortuite, significativement différents de 0. Si h est le nombre de retards, le nombre possible de faux rejets est alors de $0,05 \times h$, pour un seuil de confiance de 5 %.

Dans le cas où le corrélogramme ne laisse apparaître aucune décroissance de ses termes (absence de « *cut off* »), nous pouvons en conclure que la série n'est pas stationnaire en tendance.

2) Statistiques de Box-Pierce et Ljung-Box

Le test de Box-Pierce permet d'identifier les processus sans mémoire (suite de variables aléatoires indépendantes entre elles). Nous devons donc identifier $\text{cov}(y_t, y_{t-k}) = 0$ ou encore $\rho_k = 0 \forall k$.

Un processus de bruit blanc implique que $\rho_1 = \rho_2 = \dots = \rho_h = 0$, soit les hypothèses :

$H_0 : \rho_1 = \rho_2 = \dots = \rho_h = 0$

$H_1 : \text{il existe au moins un } \rho_i \text{ significativement différent de 0.}$

Pour effectuer ce test, on recourt à la statistique Q (due à Box-Pierce¹) qui est donnée par :

$$Q = n \sum_{k=1}^h \hat{\rho}_k^2$$

h = nombre de retards, $\hat{\rho}_k$ = autocorrélation empirique d'ordre k , n = nombre d'observations.

La statistique Q est distribuée de manière asymptotique comme un χ^2 (chi-deux) à h degrés de liberté. Nous rejetons donc l'hypothèse de bruit blanc, au seuil α , si la statistique Q est supérieure au χ^2 lu dans la table au seuil $(1 - \alpha)$ et h degrés de liberté.

Nous pouvons utiliser aussi une autre statistique, dont les propriétés asymptotiques sont meilleures, dérivée de la première qui est le Q' de Ljung et Box² :

$Q' = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k}$ qui est aussi distribuée selon un χ^2 à h degrés de liberté et dont les règles de décisions sont identiques au précédent. Ces tests sont appelés par les anglo-saxons : « *portmanteau test* » soit littéralement test « *fourre-tout* ».

1. Box G E. P. et Pierce D. A., 1970.

2. Ljung G. M. et Box G E. P., 1978.

3) Tests de normalité

Pour calculer des intervalles de confiance prévisionnels et aussi pour effectuer les tests de Student sur les paramètres, il convient de vérifier la normalité des erreurs. Le test de Jarque et Bera (1984), fondé sur la notion de Skewness (asymétrie) et de Kurtosis (aplatissement), permet de vérifier la normalité d'une distribution statistique.

a) Les tests du Skewness et du Kurtosis

Soit $\mu_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k$ le moment centré d'ordre k , le coefficient de Skewness

$(\beta_1^{1/2})$ est égal à : $\beta_1^{1/2} = \frac{\mu_3}{\mu_2^{3/2}}$ et le coefficient de Kurtosis : $\beta_2 = \frac{\mu_4}{\mu_2^2}$.

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

$$\beta_1^{1/2} \rightarrow N\left(0; \sqrt{\frac{6}{n}}\right) \text{ et } \beta_2 \rightarrow N\left(3; \sqrt{\frac{24}{n}}\right)$$

On construit alors les statistiques :

$$v_1 = \frac{|\beta_1^{1/2} - 0|}{\sqrt{\frac{6}{n}}} \text{ et } v_2 = \frac{|\beta_2 - 3|}{\sqrt{\frac{24}{n}}} \text{ que l'on compare à 1,96 (valeur de la loi nor-}$$

male au seuil de 5 %).

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

b) Le test de Jarque et Bera

Il s'agit d'un test qui synthétise les résultats précédents ; si $\beta_1^{1/2}$ et β_2 obéissent à des lois normales alors la quantité $s : s = \frac{n}{6} \beta_1 + \frac{n}{24} (\beta_2 - 3)^2$ suit un χ^2 à deux degrés de liberté.

Donc si $s > \chi_{1-\alpha}^2(2)$, on rejette l'hypothèse H_0 de normalité des résidus au seuil α .

Ces tests de normalité servent également dans le cas où il y a hétéroscédasticité. En effet, l'hétéroscédasticité se manifeste sur le graphe de la distribution par des queues de probabilité plus épaisses (distribution leptokurtique) que les queues de la loi normale.

4) Tests d'homoscédasticité

Un processus de bruit blanc doit être homoscédastique, les tests d'hétéroscédasticité du chapitre 5, Section II peuvent être utilisés.

- Le test de Goldfeld-Quandt a pour but de comparer la somme des carrés des résidus d'estimation après avoir scindé les résidus en deux sous-échantillons.
- Un autre test consiste à étudier la distribution des carrés des résidus. L'analyse des termes du corrélogramme des résidus au carré permet de tester l'existence d'une hétéroscédacité. Si certaines valeurs de la FAC (tests de Box-Pierce ou Ljung-Box) sont significativement différentes de 0, nous pouvons conclure à la présence d'une hétéroscédacité.

II. La non-stationnarité et les tests de racine unitaire

A. La non-stationnarité : les processus TS et DS

Les chroniques économiques sont rarement des réalisations de processus aléatoires stationnaires. Pour analyser la non-stationnarité, deux types de processus sont distingués :

- les processus TS (*Trend Stationary*) qui représentent une non-stationnarité de type déterministe¹ ;
- les processus DS (*Differency Stationary*) pour les processus non stationnaires aléatoires.

a) Les processus TS

Un processus TS s'écrit : $x_t = f_t + \varepsilon_t$ où f_t est une fonction polynômiale du temps, linéaire ou non linéaire, et ε_t un processus stationnaire. Le processus TS le plus simple (et le plus répandu) est représenté par une fonction polynômiale de degré 1. Le processus TS porte alors le nom de linéaire et s'écrit :

$$x_t = a_0 + a_1 t + \varepsilon_t.$$

Si ε_t est un bruit blanc (gaussien ou non), les caractéristiques de ce processus sont alors :

$$\begin{aligned} E[x_t] &= a_0 + a_1 t + E[\varepsilon_t] = a_0 + a_1 t \\ V[x_t] &= 0 + V[\varepsilon_t] = \sigma_\varepsilon^2 \\ \text{cov}[x_t, x_{t'}] &= 0 \text{ pour } t \neq t' \end{aligned}$$

Ce processus TS est non stationnaire car $E[x_t]$ dépend du temps. Connaissant \hat{a}_0 et \hat{a}_1 , le processus x_t peut être stationnarisé en retranchant, de la valeur de x_t en t , la valeur estimée $\hat{a}_0 + \hat{a}_1 t$. Dans ce type de modélisation, l'ef-

1. Par définition, un processus est aléatoire d'où l'ambiguïté du terme de processus déterministe.

est produit par un choc (ou par plusieurs chocs aléatoires) à un instant t est transitoire. Le modèle étant déterministe, la chronique retrouve son mouvement de long terme qui est ici la droite de tendance. Il est possible de généraliser cet exemple à des fonctions polynômiales de degré quelconque.

b) Les processus DS

Les processus DS sont des processus que l'on peut rendre stationnaires par l'utilisation d'un filtre aux différences : $(1 - D)^d x_t = \beta + \varepsilon_t$ où ε_t est un processus stationnaire, β une constante réelle, D l'opérateur décalage et d l'ordre du filtre aux différences.

Ces processus sont souvent représentés en utilisant le filtre aux différences premières ($d = 1$). Le processus est dit alors processus du premier ordre. Il s'écrit :

$$(1 - D)x_t = \beta + \varepsilon_t \Leftrightarrow x_t = x_{t-1} + \beta + \varepsilon_t$$

L'introduction de la constante β dans le processus DS permet de définir deux processus différents :

- $\beta = 0$: le processus DS est dit sans dérive.

Il s'écrit : $x_t = x_{t-1} + \varepsilon_t$.

Comme ε_t est un bruit blanc, ce processus DS porte le nom de modèle de marche au hasard ou de marche aléatoire (*Random Walk Model*). Il est très fréquemment utilisé pour analyser l'efficacité des marchés financiers.

Pour étudier les caractéristiques de ce modèle, écrivons-le sous sa forme développée :

$$x_t = x_{t-1} + \varepsilon_t$$

$$x_{t-1} = x_{t-2} + \varepsilon_{t-1} \Rightarrow x_t = x_{t-2} + \varepsilon_{t-1} + \varepsilon_t$$

$$x_{t-2} = x_{t-3} + \varepsilon_{t-2} \Rightarrow x_t = x_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t$$

etc.

Si le premier terme de la chronique est x_0 , le modèle s'écrit alors :

$$x_t = x_0 + \sum_{i=1}^t \varepsilon_i$$

Les caractéristiques de ce processus sont (en supposant x_0 certain) :

$$E[x_t] = x_0$$

$$V[x_t] = t\sigma_\varepsilon^2$$

$$\text{cov}[x_t, x_{t'}] = \sigma_\varepsilon^2 \times \text{Min}(t, t') \text{ si } t \neq t'$$

Par exemple :

$$\begin{aligned} \text{cov}(x_4, x_2) &= E[(x_0 + \varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \varepsilon_4)(x_0 + \varepsilon_1 + \varepsilon_2)] \\ &= E(\varepsilon_1 \times \varepsilon_1) + E(\varepsilon_2 \times \varepsilon_2) = 2\sigma_\varepsilon^2. \end{aligned}$$

En effet tous les produits $E(\varepsilon_t \times \varepsilon_{t'}) = 0$ si $t \neq t'$.

Ce processus est non stationnaire en variance puisqu'elle dépend du temps. Cette non stationnarité est dite aléatoire ou stochastique.

Pour stationnariser la marche aléatoire, il suffit d'appliquer au processus le filtre aux différences premières : $x_t = x_{t-1} + \varepsilon_t \Leftrightarrow (1 - D)x_t = \varepsilon_t$.

- $\beta \neq 0$: le processus porte alors le nom de processus DS avec dérive.

Il s'écrit : $x_t = x_{t-1} + \beta + \varepsilon_t$.

Comme précédemment, on peut rechercher sa forme équivalente développée :

$$x_t = x_{t-1} + \beta + \varepsilon_t$$

$$x_{t-1} = x_{t-2} + \beta + \varepsilon_{t-1} \Rightarrow x_t = x_{t-2} + 2\beta + \varepsilon_{t-1} + \varepsilon_t$$

$$x_{t-2} = x_{t-3} + \beta + \varepsilon_{t-2} \Rightarrow x_t = x_{t-3} + 3\beta + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t$$

etc.

Si on suppose la valeur d'origine x_0 connue et déterministe, on a alors :

$$x_t = x_0 + \beta t + \sum_{i=1}^t \varepsilon_i$$

On peut analyser les caractéristiques de ce processus :

$$E[x_t] = x_0 + \beta t$$

$$V[x_t] = t\sigma_\varepsilon^2$$

$$\text{cov}[x_t, x_{t'}] = \sigma_\varepsilon^2 \times \text{Min}(t, t') \text{ si } t \neq t'$$

Le processus est non stationnaire de par son espérance et sa variance. L'espérance étant de la même forme que celle d'un processus TS, on reconnaît dans ce processus une non stationnarité déterministe et aléatoire à la fois.

La stationnarisation de ce processus est réalisée en utilisant le filtre aux différences premières :

$$x_t = x_{t-1} + \beta + \varepsilon_t \Leftrightarrow (1 - B)x_t = \beta + \varepsilon_t$$

Soit en employant la forme développée :

$$x_t = x_0 + \beta t + \sum_{i=1}^t \varepsilon_i \text{ et en calculant : } x_{t-1} = x_0 + \beta(t-1) + \sum_{i=1}^{t-1} \varepsilon_i$$

On a : $x_t - x_{t-1} = (1 - B)x_t = \beta + \varepsilon_t$.

Dans les processus de type DS, un choc à un instant donné se répercute à l'infini sur les valeurs futures de la série ; l'effet du choc est donc permanent et va en décroissant.

En résumé, pour stationnariser un processus TS, la bonne méthode est celle des moindres carrés ordinaires ; pour un processus DS, il faut employer le filtre aux différences. Le choix d'un processus DS ou TS comme structure de la chronique n'est donc pas neutre.

c) Conséquences d'une mauvaise stationnarisation du processus

Pour un processus TS, la bonne méthode de stationnarisation est celle des moindres carrés ordinaires. Supposons que l'on applique au processus TS du premier ordre un filtre aux différences premières. *A priori*, comme le degré du polynôme est 1, ce filtre peut être considéré comme correct puisqu'un filtre aux différences d'ordre d élimine un polynôme de même degré. Cependant, on démontre que l'application du filtre aux différences a créé une perturbation artificielle.

Pour un processus DS, la bonne méthode de stationnarisation est le filtre aux différences premières. Supposons que l'on applique la méthode des moindres carrés ordinaires (régression sur le temps) sur les observations d'un échantillon du processus, les paramètres de la tendance sont estimés et par conséquent le résidu de la régression doit être un bruit blanc. Nelson et Kang montrent à partir de simulations, que l'élimination d'une tendance linéaire sur un processus de marche aléatoire crée artificiellement une forte autocorrélation des résidus pour les premiers retards.

Sur le plan économétrique, il est donc primordial d'identifier clairement le processus sous-jacent et d'employer la méthode adéquate de stationnarisation. Sinon le risque de créer des « bruits parasites » artificiels est très élevé.

Les conséquences sont également importantes sur le plan économique. Considérons, par exemple, le PIB d'un pays comme la France en valeur réelle. Si ce PIB est DS plutôt que TS, il est alors nécessaire de remettre en cause la décomposition traditionnelle (tendance et cycle) et sa justification théorique : l'indépendance des schémas explicatifs. Si le PIB est en effet DS, la croissance et le cycle sont liés et ne peuvent être en conséquence étudiés de façon séparés. Or, d'après les travaux de Nelson et Plosser (1982) sur des chroniques macro-économiques américaines, la variabilité constatée de la composante conjoncturelle serait due à une structure DS. Comme jusqu'à présent, l'analyse de cette composante s'effectue à partir du résidu d'une régression entre le PIB et une tendance déterministe, cette analyse surestime l'amplitude du cycle et sous-estime l'importance de la tendance. Sur ce constat, Beveridge S. et Nelson C.R. (1981) proposent une décomposition des processus selon une tendance stochastique (permanente) qui obéit à une marche aléatoire avec ou sans dérive et une composante stationnaire (transitoire). Par la suite Harvey A.C. (1988) utilise les modèles structurels à composantes inobservables (modèle tendance plus cycle et tendance-cycle) représentées sous forme d'un modèle espace d'états estimé par le filtre de Kalman.

B. Les tests de racine unitaire et la stratégie séquentielle de test

Les tests de racine unitaire « *Unit Root Test* » permettent non seulement de détecter l'existence d'une non-stationnarité mais aussi de déterminer de quelle non-stationnarité il s'agit (processus TS ou DS) et donc la bonne méthode pour stationnariser la série.

a) Tests de racines unitaires : tests de Dickey-Fuller (1979)

Les tests de Dickey-Fuller (DF) permettent de mettre en évidence le caractère stationnaire ou non d'une chronique par la détermination d'une tendance déterministe ou stochastique.

Les modèles servant de base à la construction de ces tests sont au nombre de trois. Le principe des tests est simple : si l'hypothèse $H_0 : \phi_1 = 1$ est retenue dans l'un de ces trois modèles, le processus est alors non stationnaire.

- | | |
|---|--------------------------------------|
| [1] $x_t = \phi_1 x_{t-1} + \varepsilon_t$ | Modèle autorégressif d'ordre 1. |
| [2] $x_t = \phi_1 x_{t-1} + \beta + \varepsilon_t$ | Modèle autorégressif avec constante. |
| [3] $x_t = \phi_1 x_{t-1} + bt + c + \varepsilon_t$ | Modèle autorégressif avec tendance. |

Si l'hypothèse H_0 est vérifiée, la chronique x_t n'est pas stationnaire quel que soit le modèle retenu.

Dans le dernier modèle [3], si on accepte $H_1 : \phi_1 < 1$ et si le coefficient b est significativement différent de 0, alors le processus est un processus TS ; on peut le rendre stationnaire en calculant les résidus par rapport à la tendance estimée par les moindres carrés ordinaires.

Sous H_0 , les règles habituelles de l'inférence statistique ne peuvent pas être appliquées pour tester cette hypothèse, en particulier la distribution de Student du paramètre ϕ_1 ; Dickey et Fuller ont donc étudié la distribution asymptotique de l'estimateur $\hat{\phi}_1$ sous l'hypothèse H_0 . À l'aide de simulations de Monte-Carlo, ils ont tabulé les valeurs critiques pour des échantillons de tailles différentes. Ces tables sont des tables¹ analogues aux tables du t de Student. Les auteurs ont choisi de tester la valeur $(\hat{\phi}_1 - 1)$ au lieu de $\hat{\phi}_1$ pour des raisons purement statistiques. Cela n'est pas gênant pour le test. En effet, $x_t = \phi_1 x_{t-1} + \varepsilon_t$ s'écrit aussi :

$$\begin{aligned}x_t - x_{t-1} &= \phi_1 x_{t-1} - x_{t-1} + \varepsilon_t \\ \Delta x_t &= (\phi_1 - 1)x_{t-1} + \varepsilon_t\end{aligned}$$

Il est donc équivalent de tester comme hypothèse $H_0 : \phi_1 = 1$ ou $\phi_1 - 1 = 0$.

Les principes généraux du test sont les suivants.

On estime par les moindres carrés ordinaires le paramètre ϕ_1 noté $\hat{\phi}_1$ pour les modèles [1], [2] et [3]. L'estimation des coefficients et des écarts types du modèle par les moindres carrés ordinaires fournit $t_{\hat{\phi}_1}$ qui est analogue à la statistique de Student (rapport du coefficient sur son écart type). Si $t_{\hat{\phi}_1} \geq t_{\text{tabulé}}$, alors on accepte l'hypothèse H_0 ; il existe une racine unité, le processus n'est donc pas stationnaire.

Remarque : les principaux logiciels d'analyse de séries temporelles calculent automatiquement les valeurs critiques $t_{\hat{\phi}_1}$.

b) Les tests de Dickey et Fuller Augmentés

Dans les modèles précédents, utilisés pour les tests de Dickey-Fuller simples, le processus ε_t est, par hypothèse, un bruit blanc. Or il n'y a aucune raison pour

1. Cf. table 7 en fin d'ouvrage.

que, *a priori*, l'erreur soit non corrélée ; on appelle tests de Dickey-Fuller Augmentés (ADF, 1981) la prise en compte de cette hypothèse.

Les tests ADF sont fondés, sous l'hypothèse alternative $|\phi_1| < 1$, sur l'estimation par les MCO des trois modèles :

$$\text{Modèle [4]} : \Delta x_t = \rho x_{t-1} - \sum_{j=2}^p \phi_j \Delta x_{t-j+1} + \varepsilon_t$$

$$\text{Modèle [5]} : \Delta x_t = \rho x_{t-1} - \sum_{j=2}^p \phi_j \Delta x_{t-j+1} + c + \varepsilon_t$$

$$\text{Modèle [6]} : \Delta x_t = \rho x_{t-1} - \sum_{j=2}^p \phi_j \Delta x_{t-j+1} + c + bt + \varepsilon_t$$

avec $\varepsilon_t \rightarrow i.i.d.$

Le test se déroule de manière similaire aux tests DF simples, seules les tables statistiques diffèrent. La valeur de p peut être déterminée selon les critères de Akaike ou de Schwarz, ou encore, en partant d'une valeur suffisamment importante de p , on estime un modèle à $p-1$ retards, puis à $p-2$ retards, jusqu'à ce que le coefficient du $p^{\text{ième}}$ retard soit significatif.

c) Le test de Phillips et Perron (1988)

Ce test est construit sur une correction non paramétrique des statistiques de Dickey-Fuller pour prendre en compte des erreurs hétéroscédastiques. Il se déroule en quatre étapes :

1) Estimation par les moindres carrés ordinaires des trois modèles de base des tests de Dickey-Fuller et calcul des statistiques associées, soit e_t le résidu estimé.

2) Estimation de la variance dite de court terme $\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n e_t^2$

3) Estimation d'un facteur correctif s_t^2 (appelé variance de long terme) établi à partir de la structure des covariances des résidus des modèles précédemment estimés de telle sorte que les transformations réalisées conduisent à des distributions identiques à celles du Dickey-Fuller standard :

$$s_t^2 = \frac{1}{n} \sum_{i=1}^n e_i^2 + 2 \sum_{i=1}^l \left(1 - \frac{i}{l+1}\right) \frac{1}{n} \sum_{t=i+1}^n e_t e_{t-i}.$$

Pour estimer cette variance de long terme, il est nécessaire de définir un nombre de retards l (troncature de Newey-West) estimé en fonction du nombre d'observations n , $l \approx 4(n/100)^{2/9}$.

4) Calcul de la statistique de *PP* : $t_{\phi_1}^* = \sqrt{k} \times \frac{(\hat{\phi}_1 - 1)}{\hat{\sigma}_{\phi_1}} + \frac{n(k-1)\hat{\sigma}_{\phi_1}}{\sqrt{k}}$

avec $k = \frac{\hat{\sigma}^2}{s_t^2}$ (qui est égal à 1 – de manière asymptotique – si e_t est un bruit blanc). Cette statistique est à comparer aux valeurs critiques de la table de MacKinnon.

d) Stratégie de tests

Nous constatons que pour réaliser un test de racine unitaire, le résultat n'est pas identique selon l'utilisation de l'un des trois modèles comme processus générateur de la chronique de départ. Les conclusions auxquelles on parvient sont donc différentes et peuvent entraîner des transformations erronées. C'est la raison pour laquelle Dickey et Fuller, et à leur suite d'autres auteurs, ont élaboré des stratégies de tests¹. Nous présentons un exemple simplifié (schéma 1) d'une stratégie de tests. Les valeurs critiques des t_c et t_b permettant de tester la nullité des coefficients c et b des modèles [2] et [3] sont données en fin d'ouvrage (table 7).

e) Le test KPSS (1992)

Kwiatkowski *et al.* (1992) propose d'utiliser un test du multiplicateur de Lagrange (LM) fondé sur l'hypothèse nulle de stationnarité. Après estimation des modèles [2] ou [3], on calcule la somme partielle des résidus : $S_t = \sum_{i=1}^t e_i$ et on estime la variance de long terme (s_t^2) comme pour le test de Phillips et Perron.

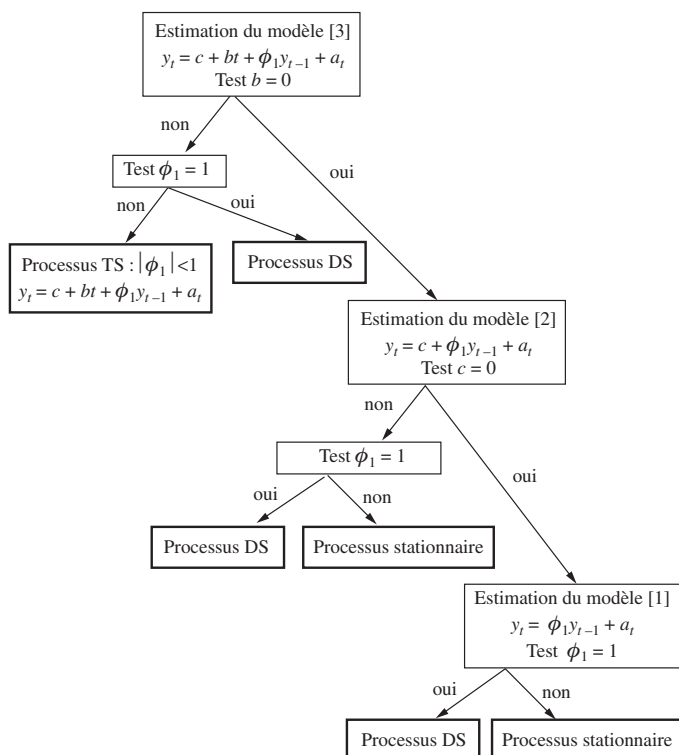


Schéma 1 – Stratégie simplifiée des tests de racine unitaire

1. Cf. Bourbonnais R. et Terraza M. Chapitre 5, 2010.

La statistique est alors $LM = \frac{1}{s_t^2} \frac{\sum_{t=1}^n S_t^2}{n^2}$. On rejette l'hypothèse de stationnarité si cette statistique est supérieure aux valeurs critiques lues dans une table élaborée par les auteurs.

Il est à noter que les logiciels RATS et Eviews permettent directement l'utilisation de ces tests.

Exercice n° 1

↓ fichier C9EX1

Exemple d'application des tests de racine unitaire au CAC40

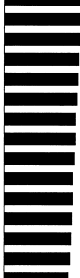
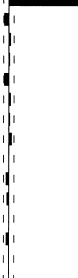
On demande d'appliquer les tests de non-stationnarité à l'indice CAC40 (indice représentatif de l'évolution des cours de bourse) sur une période de 1160 jours.

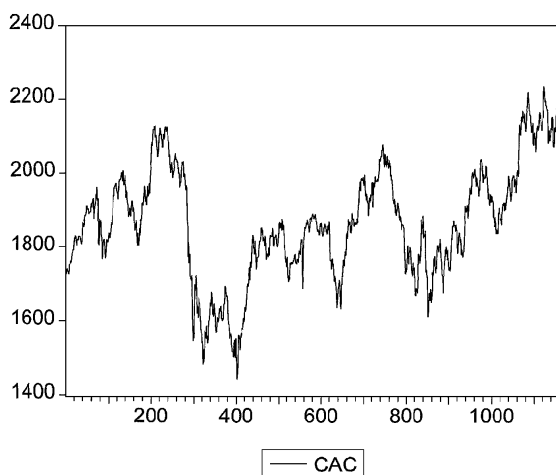
Solution

Nous allons étudier le CAC40 sur 1 160 observations quotidiennes. Le graphique 1 illustre l'évolution de l'indice CAC40.

Les fonction d'autocorrélation simple et partielle, pour $h = 15$ retards, sont obtenues directement par une fonction :

Sample: 1 1160
Included observations: 1160

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.989	0.989	1136.6	0.000
		2 0.976	-0.049	2245.7	0.000
		3 0.964	0.020	3329.0	0.000
		4 0.953	0.015	4387.8	0.000
		5 0.941	-0.055	5420.2	0.000
		6 0.929	0.024	6427.5	0.000
		7 0.917	0.009	7411.2	0.000
		8 0.907	0.035	8373.4	0.000
		9 0.896	-0.005	9314.4	0.000
		10 0.885	-0.039	10233.	0.000
		11 0.874	-0.018	11128.	0.000
		12 0.862	0.004	12002.	0.000
		13 0.850	-0.040	12852.	0.000
		14 0.838	0.003	13679.	0.000
		15 0.827	0.000	14483.	0.000



Graphique 1 – Évolution de l'indice boursier CAC40

EvIEWS fournit les résultats des fonctions d'autocorrélation simple (colonne AC) et partielle (colonne PAC), avec les corrélogrammes respectifs. Les bornes de l'intervalle de confiance sont stylisées par des traits pointillés horizontaux ; chaque terme qui sort de cet intervalle est donc significativement différent de 0 au seuil de 5 %. Nous nous apercevons que tous les termes du corrélogramme simple sont extérieurs à l'intervalle de confiance. Le processus n'est pas un bruit blanc (il semble même caractéristique d'un processus non stationnaire). La statistique Q de Ljung-Box (la seule calculée par EvIEWS) confirme ce fait : $Q\text{-Stat} = 14483$ (au retard $k = 15$) $> \chi^2_{0,05;15} = 25$, on refuse l'hypothèse de nullité des coefficients ρ_k (la probabilité critique de ce test est indiquée $\alpha_c = 0,000 < 0,05$, donc on refuse H_0). Le processus CAC40 n'est pas un bruit blanc. A partir des tests de Dickey-Fuller nous allons examiner si le processus est non stationnaire.

Le test de Dickey-Fuller simple consiste à estimer¹ les trois modèles :

$$[1] : CAC_t - CAC_{t-1} = DCAC_t = (\phi_1 - 1) CAC_{t-1}$$

$$[2] : CAC_t - CAC_{t-1} = DCAC_t = (\phi_1 - 1) CAC_{t-1} + c$$

$$[3] : CAC_t - CAC_{t-1} = DCAC_t = (\phi_1 - 1) CAC_{t-1} + c + bt$$

L'estimation par les MCO des paramètres du modèle [3] donne :

$$DCAC_t = -0,008447 CAC_{t-1} + 14,94 + 0,002103 t$$

$$(-2,053)$$

$$(2,0)$$

$$(1,11)$$

$$n = 1\,159 - (\cdot) = t \text{ empirique.}$$

Le coefficient de la droite de tendance n'est pas significativement différent de 0 ($t^* = 1,11$), on rejette l'hypothèse d'un processus TS et $t_{\hat{\phi}_1} = -2,053 > t_{\text{tabulé}} = -3,41$ (ou $n(\hat{\phi}_1 - 1) = 1\,159 \times -0,008447 = -9,79 > -21,8$) ; on accepte l'hypothèse H_0 (valeurs critiques au seuil de 5 %) ; le processus n'est pas stationnaire.

L'estimation par les MCO des paramètres du modèle [2] donne :

$$DCAC_t = -0,00709 CAC_{t-1} + 13,63$$

$$(-1,805)$$

$$(1,85)$$

$$n = 1\,159 - (\cdot) = t \text{ empirique.}$$

1. Ces tests très brièvement présentés ici doivent faire l'objet d'une stratégie afin de tester les cas des hypothèses jointes. Cf. Bourbonnais R. et Terraza M., 2010.

Le terme constant n'est pas significativement différent de 0 ($t^* = 1,85$), on rejette l'hypothèse d'un processus DS avec dérive et $t_{\hat{\phi}_1} = -1,805 > t_{\text{tabulé}} = -2,86$; on accepte l'hypothèse H_0 ; le processus n'est pas stationnaire.

Enfin l'estimation par les MCO des paramètres du modèle [1] donne :

$$\text{DCAC}_t = 0,000174 \text{ CAC}_{t-1} \\ (0,536)$$

$$n = 1\,159 - (\cdot) = t \text{ empirique.}$$

$t_{\hat{\phi}_1} = 0,536 > t_{\text{tabulé}} = -1,95$, on accepte l'hypothèse H_0 ; le processus n'est pas stationnaire.

Le test DFA (avec quatre retards) conduit à des résultats similaires :

Hypothèse H_0 : CAC possède une racine unitaire		
Nombre de retards (minimum du critère de Akaike) = 4		
Test de Dickey-Fuller Augmenté	t Statistique	Probabilité critique
Modèle [4]	0,46	0,81
Modèle [5]	-2,03	0,27
Modèle [6]	-2,28	0,44

Les probabilités critiques sont toutes supérieures à 0,05, nous ne rejetons pas l'hypothèse H_0 ; nous pouvons donc conclure que le processus CAC40 possède une racine unitaire et n'est donc pas stationnaire.

Nous procédons ensuite au test de Phillips-Perron avec une troncature $l = 6$.

Hypothèse H_0 : CAC possède une racine unitaire		
Troncature $l = 6$		
Test de Phillips-Perron	t Statistique ajusté	Probabilité critique
Modèle [1]	0,50	0,82
Modèle [2]	-1,94	0,31
Modèle [3]	-2,20	0,49

Les probabilités critiques sont toutes supérieures à 0,05, nous ne rejetons pas l'hypothèse H_0 ; le processus CAC40 possède une racine unitaire.

Enfin, nous procédons aux tests KPSS.

Hypothèse H_0 : CAC ne possède pas une racine unitaire		
Troncature $l = 6$		
Test de Kwiatkowski-Phillips-Schmidt-Shin	LM Statistique	Valeur critique à 5%
Modèle [2]	2,58	0,46
Modèle [3]	1,31	0,14

La statistique LM est supérieure à la valeur critique (pour un seuil de 5%) pour les deux spécifications, nous rejetons l'hypothèse H_0 , le processus CAC40 possède donc une racine unitaire.

Tous les résultats sont convergents, nous pouvons donc conclure que le processus CAC40 n'est pas stationnaire.

S'agit-il d'un processus de marche au hasard sans dérive ? Nous allons le vérifier en calculant le corrélogramme de la série filtrée par les différences premières ceci afin de stationnariser le processus :

$$\text{DCAC}_t = \text{CAC}_t - \text{CAC}_{t-1}.$$

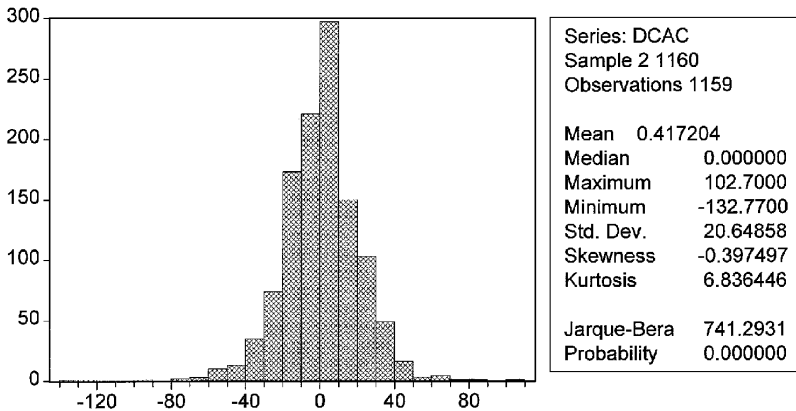
Correlogram of DCAC

Sample: 1 1160 Included observations: 1159						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.062	0.062	4.4625	0.035
		2	-0.010	-0.014	4.5772	0.101
		3	-0.039	-0.038	6.3597	0.095
		4	0.059	0.064	10.409	0.034
		5	-0.015	-0.024	10.665	0.058
		6	-0.020	-0.018	11.145	0.084
		7	-0.017	-0.010	11.493	0.119
		8	-0.008	-0.012	11.572	0.171
		9	0.044	0.046	13.813	0.129
		10	0.021	0.016	14.347	0.158
		11	0.018	0.016	14.712	0.196
		12	0.040	0.043	16.558	0.167
		13	-0.008	-0.018	16.634	0.217
		14	-0.051	-0.049	19.652	0.141
		15	-0.009	0.001	19.741	0.182

La statistique Q a une probabilité critique de 0,182 (pour $k = 15$) largement supérieure à 0,05 ; nous acceptons l'hypothèse H_0 de nullité des coefficients du correlogramme. Le correlogramme de la série CAC40 filtrée par les différences premières est caractéristique d'un bruit blanc. La série CAC40 est donc bien un processus DS sans dérive.

S'agit-il d'un bruit blanc gaussien ?

L'histogramme de la distribution et les valeurs empiriques des Skewness, Kurtosis et de la statistique de Jarque-Bera sont donnés par :



Les tests sont effectués à partir des valeurs :

$$v_1 = \frac{|\beta_1^{1/2} - 0|}{\sqrt{\frac{6}{n}}} = \frac{|-0,39 - 0|}{\sqrt{\frac{6}{1159}}} = 5,42 > 1,96$$

et

$$v_2 = \frac{|\beta_2 - 3|}{\sqrt{\frac{24}{n}}} = \frac{|6,83 - 3|}{\sqrt{\frac{24}{1159}}} = 26,61 > 1,96.$$

Nous rejetons l'hypothèse de normalité en ce qui concerne la symétrie et l'aplatissement de la distribution, ce qui est confirmé par la statistique de Jarque-Bera :

$JB = 741,3 > \chi_{0,05}^2(2) = 5,99$. Le processus CAC40 en différences est donc un bruit blanc non gaussien.

III. Les modèles ARIMA

Nous allons présenter une famille de processus aléatoires qui sont censés recouvrir une gamme très large d'évolution possible de séries chronologiques : les processus autorégressifs et les processus de moyenne mobile.

A. Typologie des modèles AR, MA et ARMA

1) Modèle AR (Auto Régressif)

a) Formulation

Dans le processus autorégressif d'ordre p , l'observation présente y_t est générée par une moyenne pondérée des observations passées jusqu'à la p -ième période sous la forme suivante :

$$AR(1) : y_t = \theta_1 y_{t-1} + \varepsilon_t$$

$$AR(2) : y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \varepsilon_t$$

...

$$AR(p) : y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + \varepsilon_t \quad [4]$$

où $\theta_1, \theta_2, \dots, \theta_p$ sont des paramètres à estimer positifs ou négatifs, ε_t est un aléa gaussien.

Nous pouvons ajouter à ce processus une constante qui ne modifie en rien les propriétés stochastiques. L'équation [4] peut aussi s'écrire à l'aide de l'opérateur décalage D :

$$(1 - \theta_1 D - \theta_2 D^2 - \dots - \theta_p D^p) y_t = \varepsilon_t$$

b) Caractéristiques des corrélogrammes

Il est démontré que le corrélogramme simple d'un processus $AR(p)$ est caractérisé par une décroissance géométrique de ses termes de type :

$$\rho_k = \rho^k$$

Le corrélogramme partiel a ses seuls p premiers termes différents de 0.

2) Modèle MA (Moving Average : Moyenne Mobile)

a) Formulation

Dans le processus de moyenne mobile d'ordre q , chaque observation y_t est générée par une moyenne pondérée d'aléas jusqu'à la q -ième période.

$$\text{MA}(1) : y_t = \varepsilon_t - \alpha_1 \varepsilon_{t-1}$$

$$\text{MA}(2) : y_t = \varepsilon_t - \alpha_1 \varepsilon_{t-1} - \alpha_2 \varepsilon_{t-2}$$

...

$$\text{MA}(q) : y_t = \varepsilon_t - \alpha_1 \varepsilon_{t-1} - \alpha_2 \varepsilon_{t-2} - \dots - \alpha_q \varepsilon_{t-q} \quad [5]$$

où $\alpha_1, \alpha_2, \dots, \alpha_q$ sont des paramètres pouvant être positifs ou négatifs et ε_t est un aléa gaussien.

L'équation [5] peut aussi s'écrire :

$$(1 - \alpha_1 D - \alpha_2 D^2 - \dots - \alpha_q D^q) \varepsilon_t = y_t.$$

Dans ce processus, tout comme dans le modèle auto-régressif AR, les aléas sont supposés être engendrés par un processus de type bruit blanc. Nous pouvons interpréter le modèle MA comme étant représentatif d'une série chronologique fluctuant autour de sa moyenne de manière aléatoire, d'où le terme de moyenne mobile car celle-ci, en lissant la série, gomme le bruit créé par l'aléa.

Il est à noter qu'il y a équivalence entre un processus MA(1) et un processus AR d'ordre p infini :

$$\text{MA}(1) = \text{AR}(\infty).$$

b) Caractéristiques des corrélogrammes

Le corrélogramme simple d'un processus MA(q) est de la forme générale :

$$\rho_k = \frac{\sum_{i=0}^{i=q-k} \alpha_i \alpha_{i+k}}{\sum_{i=0}^{i=q} \alpha_i^2} \text{ pour } k = 0, 1, \dots, q \text{ et } \rho_k = 0 \text{ pour } k > q.$$

C'est-à-dire que seuls les q premiers termes du corrélogramme simple sont significativement différents de 0.

Le corrélogramme partiel est caractérisé par une décroissance géométrique des retards.

3) Modèle ARMA (mélange de processus AR et MA)

a) Formulation

Les modèles ARMA sont donc représentatifs d'un processus généré par une combinaison des valeurs passées et des erreurs passées. Ils sont définis par l'équation :

$$\begin{aligned} \text{ARMA}(p,q) : (1 - \theta_1 D - \theta_2 D^2 - \dots - \theta_p D^p) y_t &= \\ &= (1 - \alpha_1 D - \alpha_2 D^2 - \dots - \alpha_q D^q) \varepsilon_t \end{aligned}$$

Nous avons :

$$\text{ARMA}(1,0) = \text{AR}(1); \text{ARMA}(0,1) = \text{MA}(1).$$

Dans le cas d'un processus $\text{ARMA}(p, q)$ avec constante :

$$\begin{aligned} y_t = \mu + \theta_1 x_{t-1} + \theta_2 x_{t-2} + \dots + \theta_p x_{t-p} + \varepsilon_t - \alpha_1 \varepsilon_{t-1} \\ - \alpha_2 \varepsilon_{t-2} - \dots - \alpha_q \varepsilon_{t-q} \end{aligned}$$

L'espérance du processus est donnée¹ par $E(x_t) = \frac{\mu}{(1 - \theta_1 - \theta_2 - \dots - \theta_p)}$

Donc connaissant l'espérance du processus (Eviews calcule directement l'espérance du processus), la constante du processus ARMA est déterminée par :

$$\mu = E(x_t) \times (1 - \theta_1 - \theta_2 - \dots - \theta_p)$$

b) Caractéristiques des corrélogrammes

Les corrélogrammes simples et partiels sont, par voie de conséquence, un mélange des deux corrélogrammes des processus AR et MA purs. Il s'avère ainsi plus délicat d'identifier ces processus à partir de l'étude des fonctions d'autocorrélation empiriques.

Le tableau 2 synthétise les caractéristiques, en termes de corrélogrammes, des processus AR, MA et ARMA.

Tableau 2 – Résumé des propriétés des fonctions d'autocorrélation simple et partielle

Processus	Fonction autocorrélation simple	Fonction autocorrélation partielle
AR(1)	Décroissance exponentielle ($\theta_1 > 0$) ou sinusoidale amortie ($\theta_1 < 0$)	Pic significatif pour le premier retard: Positif si $\theta_1 > 0$ et négatif si $\theta_1 < 0$, les autres coefficients nuls pour des retards > 1
AR(2)	Décroissance exponentielle ou sinusoidale selon les signes de θ_1 et θ_2	Pics significatifs pour le premier et second retards, les autres coefficients sont nuls pour des retards > 2
AR(p)	Décroissance exponentielle et/ou sinusoidale	Pics significatifs pour les p premiers retards, les autres coefficients sont nuls pour des retards $> p$

1. Il suffit de remplacer chaque terme $E(x_t) = E(x_{t-1}) = E(x_{t-2}) = \dots = E(x_{t-p})$ car la série est stationnaire.

<i>Processus</i>	<i>Fonction autocorrélation simple</i>	<i>Fonction autocorrélation partielle</i>
MA(1)	Pic significatif pour le premier retard : positif si $\alpha_1 < 0$ et négatif si $\alpha_1 > 0$. Les autres coefficients sont nuls pour des retards > 1	Décroissance exponentielle ($\alpha_1 > 0$) ou sinusoïdale amortie ($\alpha_1 < 0$)
MA(2)	Pics significatifs pour le premier et second retards. Les autres coefficients sont nuls pour des retards > 2	Décroissance exponentielle ou sinusoïdale selon les signes de α_1 et α_2
MA(q)	Pics significatifs pour les q premiers retards. Les autres coefficients nuls pour des retards $> q$	Décroissance exponentielle et/ou sinusoïdale
ARMA(1, 1)	Décroissance géométrique à partir du premier retard, le signe est déterminé par $\theta_1 - \alpha_1$	Décroissance exponentielle ($\alpha_1 > 0$) ou sinusoïdale amortie ($\alpha_1 < 0$)
ARMA(p, q)	Décroissance exponentielle ou sinusoïdale amortie tronquée après ($q - p$) retards	Décroissance exponentielle ou sinusoïdale amortie tronquée après $p - q$ retards

4) Conditions d'utilisation

Les modèles AR, MA, ARMA ne sont représentatifs que de chroniques :

- stationnaires en tendance ;
- corrigées des variations saisonnières.

B. L'extension aux processus ARIMA et SARIMA

Les tests de Dickey-Fuller et Dickey-Fuller Augmenté envisagés précédemment permettent de déterminer si la série est stationnaire et dans le cas d'une non-stationnarité de quel type il s'agit : TS et DS.

Si la série étudiée est de type TS, il convient de la stationnariser par régression sur le temps et le résidu d'estimation est alors étudié selon la méthodologie de Box-Jenkins. Ceci permet de déterminer les ordres p et q des parties AR et MA du résidu. Le modèle est toujours dans ce cas un ARMA(p, q).

Si la série étudiée est de type DS, il convient de la stationnariser par passage aux différences selon l'ordre d'intégration $I = d$ (c'est-à-dire le nombre de fois qu'il faut différencier la série pour la rendre stationnaire). La série différenciée est alors étudiée selon la méthodologie de Box-Jenkins qui permet de déterminer les ordres p et q des parties AR et MA. On note ce type de modèle ARIMA(p, d, q).

Les modèles SARIMA permettent d'intégrer un ordre de différenciation lié à une saisonnalité généralisée par la transformation : $(1 - D^s)y_t = y_t - y_{t-s}$ où s correspond à la périodicité des données ($s = 4$ pour une série trimestrielle, $s = 12$ pour une série mensuelle).

Exercice n° 2

Génération de processus ARMA et analyse des corrélogrammes

On demande de générer sur 200 observations, de manière artificielle, les processus suivants et d'en étudier les corrélogrammes.

1) MA(1) : $y_t = 2 + (1 + 0,8 D)\varepsilon_t$

2) AR(1) : $(1 - 0,9 D)y_t = 2 + \varepsilon_t$

3) MA(2) : $y_t = 2 + (1 + 0,6 D - 0,3 D^2)\varepsilon_t$

4) AR(2) : $(1 - 0,9 D + 0,7 D^2)y_t = 2 + \varepsilon_t$

5) ARMA(1, 1) : $(1 - 0,9 D)y_t = 2 + (1 + 0,8 D)\varepsilon_t$

avec $\varepsilon_t \rightarrow N(0; 1)$.

Solution

Le programme Eviews ([C9EX2.PRG](#)) permettant de générer les processus est disponible en téléchargement.

IV. La méthode de Box et Jenkins

La partie autorégressive d'un processus, notée AR, est constituée par une combinaison linéaire finie des valeurs passées du processus. La partie moyenne mobile, notée MA, est constituée d'une combinaison linéaire finie en t des valeurs passées d'un bruit blanc. Wold (1954) montre que les modèles ARMA permettent de représenter la plupart des processus stationnaires. L'approche de Box et Jenkins (1976) consiste en une méthodologie d'étude systématique des séries chronologiques à partir de leurs caractéristiques afin de déterminer, dans la famille des modèles ARIMA, le plus adapté à représenter le phénomène étudié. Trois étapes principales sont définies.

A. Recherche de la représentation adéquate : l'identification¹

La phase d'identification est la plus importante et la plus difficile : elle consiste à déterminer le modèle adéquat dans la famille des modèles ARIMA. Elle est

1. Attention : il ne faut pas confondre l'emploi de ce terme avec son utilisation dans le cadre des modèles à équations simultanées.

fondée sur l'étude des corrélogrammes simple et partiel. Nous pouvons essayer d'édicter quelques règles simples facilitant la recherche des paramètres p , d , q du modèle ARIMA.

1) Désaisonnalisation

Dans le cas d'une série affectée d'un mouvement saisonnier, il convient de la retirer préalablement à tout traitement statistique. Cette saisonnalité est ajoutée à la série prévue à la fin du traitement afin d'obtenir une prévision en terme brut.

2) Recherche de la stationnarité en terme de tendance

Si l'étude du corrélogramme simple et les tests statistiques s'y rapportant (statistique Q) présagent d'une série affectée d'une tendance, il convient d'en étudier les caractéristiques selon les tests de Dickey-Fuller. La méthode d'élimination de la tendance est fonction du processus DS ou TS sous-jacent à la chronique étudiée.

Après stationnarisation, nous pouvons identifier les valeurs des paramètres p , q du modèle ARMA.

- Si le corrélogramme simple n'a que ses q premiers termes ($q = 3$ maximum) différents de 0 et que les termes du corrélogramme partiel diminuent lentement, nous pouvons pronostiquer un MA(q).
- Si le corrélogramme partiel n'a que ses p premiers termes ($p = 3$ maximum) différents de 0 et que les termes du corrélogramme simple diminuent lentement, cela caractérise un AR(p).
- Si les fonctions d'autocorrélation simple et partiel ne paraissent pas tronquées, il s'agit alors d'un processus de type ARMA, dont les paramètres dépendent de la forme particulière des corrélogrammes.

B. Estimation des paramètres

Les méthodes d'estimation diffèrent selon le type de processus diagnostiqué. Dans le cas d'un modèle AR(p), nous pouvons appliquer une méthode des moindres carrés (cf. chapitre 7) ou bien nous pouvons utiliser les relations existantes entre les autocorrélations et les coefficients du modèle (équations de Yule-Walker).

L'estimation des paramètres d'un modèle MA(q) s'avère plus complexe. Box et Jenkins suggèrent d'utiliser une procédure itérative de type balayage que nous pouvons illustrer de la manière suivante.

Supposons le processus :

$$(1 - \theta_1 D - \theta_2 D^2)y_t = (1 - \alpha_1 D - \alpha_2 D^2)\varepsilon_t$$

que nous pouvons écrire :

$$y_t = \frac{1}{1 - \theta_1 D - \theta_2 D^2} (1 - \alpha_1 D - \alpha_2 D^2) \varepsilon_t$$

Nous posons $v_t = \frac{1}{1 - \theta_1 D - \theta_2 D^2} \varepsilon_t$,

il vient alors :

$$v_t - \theta_1 v_{t-1} - \theta_2 v_{t-2} = \varepsilon_t \quad [6]$$

Ce qui nous donne :

$$y_t = v_t - \alpha_1 v_{t-1} - \alpha_2 v_{t-2} \text{ où } v_t = y_t + \alpha_1 v_{t-1} + \alpha_2 v_{t-2} \quad [7]$$

Nous pouvons alors initialiser la procédure de balayage en partant de deux intervalles de valeurs plausibles pour $(\hat{\alpha}_1; \hat{\alpha}_2)$ et d'un pas d'incrémentation.

Puis, pour chaque couple de valeurs $(\hat{\alpha}_1; \hat{\alpha}_2)$, nous posons : $\hat{v}_0 = 0$ et $\hat{v}_1 = 0$, et nous calculons les valeurs estimées de \hat{v}_t à partir de la relation [7] :

$$\begin{aligned} \hat{v}_2 &= y_2 \\ \hat{v}_3 &= y_3 + \hat{\alpha}_1 \hat{v}_2 \\ \hat{v}_4 &= y_4 + \hat{\alpha}_1 \hat{v}_3 + \hat{\alpha}_2 \hat{v}_2, \text{ etc.} \end{aligned}$$

Après calcul de toutes les valeurs de \hat{v}_t , nous estimons les paramètres θ_1 et θ_2 par la méthode des moindres carrés appliquée à l'équation [6] :

$$v_t = \theta_1 v_{t-1} + \theta_2 v_{t-2} + \varepsilon_t \quad [8]$$

Nous retenons les valeurs $\hat{\theta}_1$, $\hat{\theta}_2$ et $\hat{\alpha}_1$, $\hat{\alpha}_2$ qui rendent minimum la somme des carrés des résidus issue de la régression de l'équation [8].

Cette méthode d'estimation n'est valide que si le nombre de paramètres à estimer n'est pas trop important.

Nous pouvons mentionner des méthodes d'estimation fondées sur une maximisation de fonctions de vraisemblance recourant alors à des procédures itératives de régression non linéaire, telles que celles envisagées au chapitre 6.

C. Tests d'adéquation du modèle et prévision

Les paramètres du modèle étant estimés (on vérifie la convergence de la procédure itérative d'estimation), nous examinons les résultats d'estimation.

- *Les coefficients du modèle* doivent être significativement différents de 0 (le test du t de Student s'applique de manière classique). Si un coefficient n'est pas significativement différent de 0, il convient d'envisager une nouvelle spécification éliminant l'ordre du modèle AR ou MA non valide.
- *L'analyse des résidus* : si les résidus obéissent à un bruit blanc, il ne doit pas exister d'autocorrélation dans la série et les résidus doivent être homoscedastiques. Les tests suivants peuvent être utilisés.

- le test de Durbin Watson, bien qu’il ne permette de détecter que des autocorrélations d’ordre 1 ;
- les tests de Box et Pierce et de Ljung et Box (*cf.* I, C de ce chapitre) qui permettent de tester l’ensemble des termes de la fonction d’autocorrélation. Si le résidu est à mémoire, cela signifie que la spécification du modèle est incomplète et qu’il convient d’ajouter au moins un ordre au processus ;
- le test ARCH d’hétéroscédasticité (*cf.* chapitre 5, II, D) effectué à partir de la fonction d’autocorrélation du résidu au carré.

La phase de validation du modèle est très importante et nécessite le plus souvent un retour à la phase d’identification.

Lorsque le modèle est validé, la prévision peut alors être calculée à un horizon de quelques périodes, limitées car la variance de l’erreur de prévision croît très vite avec l’horizon.

Nous pouvons résumer les différentes étapes de la méthodologie de Box et Jenkins à partir du schéma 2.

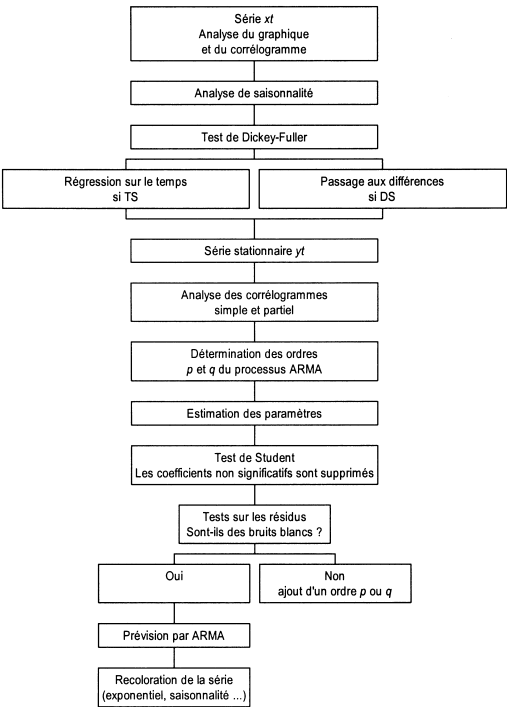


Schéma 2 – Les étapes de la méthodologie de Box et Jenkins

Exercice n° 3

↓ fichier C9EX3

Analyse par la méthodologie de Box et Jenkins et prévision de trois processus aléatoires

Soit les trois processus x_1 , x_2 et x_3 connus sur 220 périodes, on demande de :

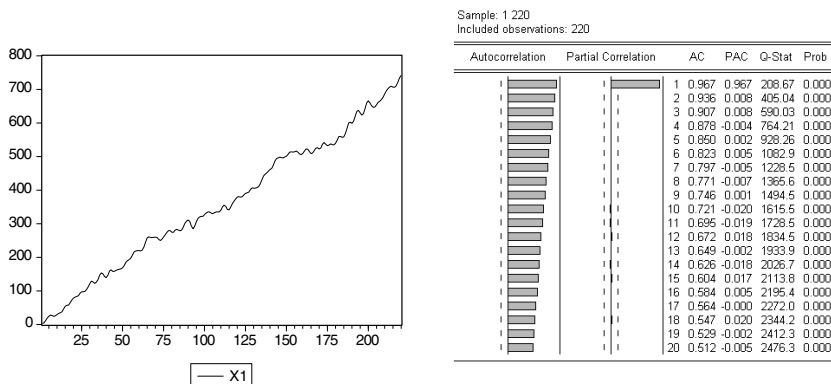
- analyser leurs propriétés stochastiques ;
- déterminer, le cas échéant, l'ordre du processus ARMA(p , q) ;
- effectuer une prévision, par une méthode adaptée, à un horizon de 3 périodes.

Solution

Les résultats sont fournis sans les calculs intermédiaires.

1) Analyse du processus x_1

Le graphique 2 présente l'évolution du processus x_1 et son corrélogramme, à leur lecture le processus est non stationnaire.



Graphique 2 – Évolution et fonction d'autocorrélation simple du processus x_1

a) Stratégie simplifiée de test : tests de Phillips-Perron (fenêtre de Newey-West = 4)

Estimation du modèle [3]

Null Hypothesis: X1 has a unit root				
Exogenous: Constant, Linear Trend				
Bandwidth: 4 (Fixed using Bartlett kernel)				
Phillips-Perron test statistic			Adj. t-Stat – 2.148148	Prob. 0.5157
Dependent Variable: D(X1)				
Included observations: 219 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
X1(-1)	– 0.039046	0.021398	– 1.824772	0.0694
C	3.725580	0.811468	4.591160	0.0000
@TREND(1)	0.124280	0.066711	1.862974	0.0638

Nous acceptons l'hypothèse H_0 de racine unitaire, nous ne pouvons donc pas interpréter les probabilités critiques calculées directement selon les lois de Student. Nous devons nous référer à la table statistique 7 en fin d'ouvrage : $t_b^* = 1,86 < t_b^{0,05} = 2,79$ (valeur lue à 5%, modèle [3]), acceptation de l'hypothèse $b = 0$. Selon la stratégie séquentielle de test nous estimons le modèle [2].

Estimation du modèle [2]

Null Hypothesis: X1 has a unit root				
Exogenous: Constant				
Bandwidth: 4 (Fixed using Bartlett kernel)				
Phillips-Perron test statistic			Adj. t-Stat	Prob.
Dependent Variable: D(X1)			0.307267	0.9783
Included observations: 219 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
X1(-1)	0.000673	0.001830	0.367843	0.7133
C	3.128542	0.749736	4.172860	0.0000

Nous acceptons l'hypothèse H_0 de racine unitaire, nous ne pouvons pas interpréter les probabilités critiques calculées directement selon les lois de Student : $t_c^* = 4,17 > t_c^{0,05} = 2,84$ (valeur lue à 5%, modèle [2]), nous refusons l'hypothèse $c = 0$. Il s'agit donc d'un processus DS avec dérive : $x_{1t} = x_{1t-1} + c + \varepsilon_t$. Nous le stationarisons par le passage aux différences premières.

b) Analyse des fonctions d'autocorrélation simple et partielle sur la série stationnaire

Sample: 1 220
Included observations: 219

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.555	0.555	68.446	0.000
		2 -0.187	-0.715	76.209	0.000
		3 -0.481	0.278	128.09	0.000
		4 -0.266	-0.158	144.04	0.000
		5 0.119	0.212	147.22	0.000
		6 0.278	-0.183	164.81	0.000
		7 0.125	0.087	168.36	0.000
		8 -0.094	-0.040	170.38	0.000
		9 -0.132	0.109	174.39	0.000
		10 0.015	0.026	174.44	0.000
		11 0.144	0.003	179.28	0.000
		12 0.134	0.095	183.50	0.000
		13 -0.010	-0.172	183.53	0.000
		14 -0.136	0.159	187.92	0.000
		15 -0.133	-0.230	192.11	0.000
		16 -0.050	0.120	192.71	0.000
		17 -0.008	-0.301	192.72	0.000
		18 -0.028	0.212	192.92	0.000
		19 -0.047	-0.276	193.45	0.000
		20 0.025	0.492	193.59	0.000

Il ne s'agit pas d'une marche au hasard (les probabilités critiques de la Q -Stat sont toutes très largement inférieures à 0,05), le processus est à mémoire, il existe donc une représentation dans la classe des processus ARMA.

c) Recherche des ordres p et q de la représentation ARMA

Compte tenu de la forme des corrélogrammes simple et partiel nous sélectionnons un modèle ARMA(2, 1). Nous constatons que la constante est significativement différente de 0, et des essais d'autres représentations concurrentes (ARMA(1, 1)) indiquent des valeurs des critères d'information (AIC ou SC) supérieures au modèle retenu.

Dependent Variable: D(X1)				
Included observations: 217 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.327748	0.346054	9.616272	0.0000
AR(1)	0.760134	0.066149	11.49128	0.0000
AR(2)	- 0.604867	0.062249	- 9.716905	0.0000
MA(1)	0.500890	0.073470	6.817644	0.0000

d) Validation de la représentation

- Tests de Student sur les coefficients :

Les coefficients sont tous significativement différents de 0 (probabilités critiques inférieures à 0,05)

- Analyse des résidus

Corrélogramme du résidu

Sample: 4 220
Included observations: 217
Q-statistic probabilities adjusted for 3 ARMA term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.024 -0.024 0.1229			
		2 -0.009 -0.010 0.1409			
		3 0.047 0.046 0.6220			
		4 -0.096 -0.054 1.3265 0.249			
		5 0.007 0.005 1.3361 0.513			
		6 0.044 0.042 1.7618 0.619			
		7 0.030 0.037 1.9858 0.738			
		8 -0.050 -0.051 2.5434 0.770			
		9 0.085 0.082 4.2128 0.648			
		10 0.004 0.009 4.2170 0.754			
		11 0.018 0.027 4.2891 0.830			
		12 0.094 0.081 6.3491 0.705			
		13 -0.043 -0.033 6.7773 0.746			
		14 -0.080 -0.082 8.2758 0.688			
		15 -0.058 -0.072 9.0667 0.697			
		16 -0.025 -0.026 9.2119 0.757			
		17 -0.063 -0.061 10.161 0.750			
		18 0.004 -0.018 10.164 0.808			
		19 -0.093 -0.103 12.219 0.729			
		20 0.037 0.051 12.554 0.766			

Corrélogramme du résidu au carré

Sample: 4 220
Included observations: 217
Q-statistic probabilities adjusted for 3 ARMA term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.028 0.028 0.1705			
		2 -0.084 -0.085 1.7200			
		3 -0.087 -0.083 3.4030			
		4 -0.033 -0.036 3.6521 0.056			
		5 0.027 0.014 3.8097 0.149			
		6 0.009 -0.004 3.8298 0.280			
		7 0.031 0.029 4.0432 0.400			
		8 0.093 0.096 6.0034 0.306			
		9 0.003 0.007 6.0060 0.423			
		10 0.030 0.051 6.2059 0.516			
		11 -0.015 0.002 6.2574 0.618			
		12 -0.009 0.004 6.2759 0.712			
		13 0.004 0.006 6.2806 0.791			
		14 -0.025 -0.028 6.4299 0.843			
		15 0.043 0.037 6.6725 0.866			
		16 -0.049 -0.067 7.4476 0.878			
		17 -0.108 -0.108 10.112 0.754			
		18 -0.017 -0.028 10.185 0.808			
		19 -0.000 -0.025 10.185 0.857			
		20 0.024 -0.004 10.328 0.889			

Le corrélogramme du résidu indique qu'il s'agit d'un processus sans mémoire, le corrélogramme du résidu au carré (test ARCH, cf. chapitre 5) n'indique aucun terme de significativement différent de 0 ; l'homoscédasticité des résidus est donc vérifiée. Les résidus sont donc un processus de bruit blanc.

Les résidus sont-ils gaussiens ? La statistique de Jarque-Bera ($JB = 2,06$) indique une probabilité critique de 0,35, nous acceptons l'hypothèse H_0 de normalité des résidus.

La représentation est validée, x_{1t} est un processus ARIMA(2, 1, 1) avec constante.

e) Prévision

La valeur de la constante est égale à : $\mu = E(x_{1t}) \times (1 - \theta_1 - \theta_2)$

La prévision est donnée par :

$$\begin{aligned}\Delta x_{1t} &= \theta_1 \times \Delta x_{1t-1} + \theta_2 \times \Delta x_{1t-2} + E(x_{1t}) \times (1 - \theta_1 - \theta_2) + \varepsilon_t + \alpha_1 \varepsilon_{t-1} \\ \Delta x_{1t} &= 0,76 \times \Delta x_{1t-1} - 0,60 \times \Delta x_{1t-2} + 3,33 \times (1 - 0,76 + 0,60) + \varepsilon_t + 0,50 \times \varepsilon_{t-1} \\ x_{1t} &= x_{1t-1} + \Delta x_{1t} \\ \Delta x_{1,221} &= 0,76 \times \Delta x_{1,220} - 0,60 \times \Delta x_{1,219} + 2,81 + \varepsilon_{221} + 0,50 \times \varepsilon_{220} \\ \Delta x_{1,221} &= 0,76 \times 7,76 - 0,60 \times 12,39 + 2,81 + 0 + 0,50 \times 0,87 = 1,65\end{aligned}$$

Jusqu'à l'observation $t = 220$ nous prenons $\varepsilon_t = \text{résidu}$, pour $t > 220$, $E(\varepsilon_t) = 0$.

$$\begin{aligned}x_{1,221} &= x_{1,220} + \Delta x_{1,221} = 741,90 + 1,65 = 743,55 \\ \Delta x_{1,222} &= 0,76 \times \Delta x_{1,221} - 0,60 \times \Delta x_{1,220} + 2,81 + \varepsilon_{222} + 0,50 \times \varepsilon_{221} \\ \Delta x_{1,222} &= 0,76 \times 1,65 - 0,60 \times 7,76 + 2,81 + 0 + 0 = -0,63 \\ x_{1,222} &= x_{1,221} + \Delta x_{1,222} = 743,55 - 0,63 = 742,92 \\ \Delta x_{1,223} &= 0,76 \times \Delta x_{1,222} - 0,60 \times \Delta x_{1,221} + 2,81 + \varepsilon_{223} + 0,50 \times \varepsilon_{222} \\ \Delta x_{1,223} &= 0,76 \times -0,63 - 0,60 \times 1,65 + 2,81 + 0 + 0 = 1,34 \\ x_{1,223} &= x_{1,222} + \Delta x_{1,223} = 742,92 + 1,34 = 744,26\end{aligned}$$

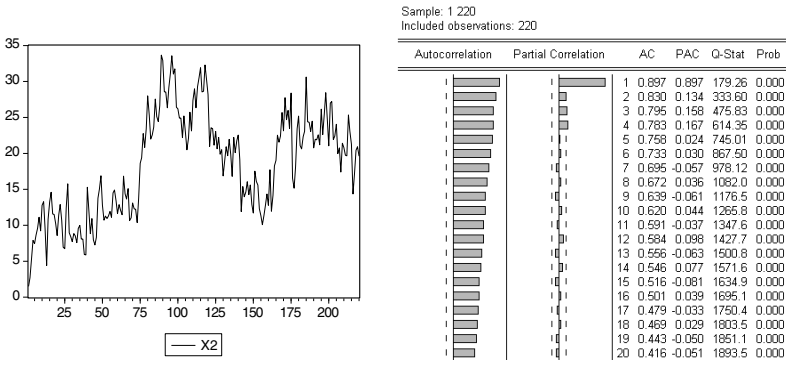
Les résultats sont présentés au tableau 3.

Tableau 3 – Calcul des prévisions du processus ARIMA(2, 1, 1)

t	x_{1t}	Δx_{1t}	$\varepsilon_t = \text{résidus}$
218	721,75	10,2	4,07
219	734,14	12,39	2,28
220	741,90	7,76	0,87
221	743,55	1,65	0
222	742,92	- 0,63	0
223	744,26	1,34	0

2) Analyse du processus x_2

Le graphique 3 présente l'évolution du processus x_2 et son corrélogramme, à leur lecture le processus semble non stationnaire.



Graphique 3 – Évolution et fonction d'autocorrélation simple du processus x_2

a) *Stratégie simplifiée de test : tests de Dickey-Fuller*

Doit-on utiliser le test de Dickey-Fuller simple ou le test de Dickey-Fuller Augmenté ?

Préalablement nous déterminons la valeur de p , si $p = 0$ nous utilisons le test de Dickey-Fuller. À partir du modèle [2] (modèle avec constante) nous sélectionnons l'ordre p du modèle à partir des critères d'information d'Akaike et de Schwarz. Les deux critères sont convergents¹ $p = 3$.

Null Hypothesis: X2 has a unit root
Exogenous: Constant
Lag Length: 3 (Automatic based on AIC, MAXLAG=8)
Null Hypothesis: X2 has a unit root
Exogenous: Constant
Lag Length: 3 (Automatic based on SIC, MAXLAG=8)

Nous appliquons donc la stratégie séquentielle des tests à partir des tests de Dickey-Fuller Augmenté avec $p = 3$.

Estimation du modèle [3]

Null Hypothesis: X2 has a unit root				
Exogenous: Constant, Linear Trend				
Lag Length: 3 (Fixed)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			- 2.253814	0.4570
Test critical values:	1% level		- 4.001108	
	5% level		- 3.430766	
	10% level		- 3.138998	
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(X2)				
Sample (adjusted): 5 220				
Included observations: 216 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
X2(- 1)	- 0.080344	0.035648	- 2.253814	0.0252
D(X2(- 1))	- 0.229159	0.070487	- 3.251071	0.0013
D(X2(- 2))	- 0.209633	0.069319	- 3.024181	0.0028
D(X2(- 3))	- 0.168372	0.067849	- 2.481571	0.0139
C	1.314041	0.575187	2.284546	0.0233
@TREND(1)	0.002458	0.003842	0.639940	0.5229

Nous acceptons l'hypothèse H_0 de racine unitaire, nous ne pouvons pas interpréter les probabilités critiques calculées directement selon les lois de Student. Nous devons

1. En cas de divergence des critères nous retenons la valeur de p minimum des deux critères.









































donc nous référer à la table statistique 7 en fin d'ouvrage : $t_b^* = 0,64 < t_b^{0,05} = 3,12$ (valeur lue à 5 %, modèle [3]), acceptation de l'hypothèse $b = 0$. Nous estimons le modèle [2].

Estimation du modèle [2]

Null Hypothesis: X2 has a unit root				
Exogenous: Constant				
Lag Length: 3 (Fixed)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			– 2.306282	0.1709
Test critical values:	1% level		– 3.460596	
	5% level		– 2.874741	
	10% level		– 2.573883	
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(X2)				
Sample (adjusted): 5 220				
Included observations: 216 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
X2(– 1)	– 0.067264	0.029165	– 2.306282	0.0221
D(X2(– 1))	– 0.239180	0.068630	– 3.485076	0.0006
D(X2(– 2))	– 0.217876	0.068017	– 3.203270	0.0016
D(X2(– 3))	– 0.174967	0.066968	– 2.612710	0.0096
C	1.347763	0.571966	2.356368	0.0194

Nous acceptons l'hypothèse H0 de racine unitaire, nous ne pouvons pas interpréter les probabilités critiques calculées directement selon les lois de Student : $t_c^* = 2,35 < t_c^{0,05} = 3,39$ (valeur lue à 5 %, modèle [2]), nous acceptons l'hypothèse $c = 0$. Il s'agit donc d'un processus DS sans dérive : $x_{2t} = x_{2t-1} + \varepsilon_t$. Nous le stationnarisons par le passage aux différences premières.

b) Analyse des fonctions d'autocorrélation simple et partielle sur la série stationnaire

Sample: 1 220					
Included observations: 219					
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.195	-0.195	8.4614	0.004
		2 -0.151	-0.197	13.568	0.001
		3 -0.101	-0.189	15.860	0.001
		4 0.046	-0.061	16.344	0.003
		5 0.018	-0.043	16.414	0.006
		6 0.084	0.067	18.010	0.006
		7 -0.072	-0.035	19.203	0.008
		8 0.026	0.037	19.357	0.013
		9 -0.066	-0.054	20.346	0.016
		10 0.042	0.010	20.764	0.023
		11 -0.141	-0.167	25.368	0.008
		12 0.068	-0.026	26.484	0.009
		13 -0.043	-0.095	26.925	0.013
		14 0.133	0.081	31.085	0.005
		15 -0.066	-0.019	32.116	0.006
		16 0.014	0.033	32.165	0.010
		17 -0.071	-0.031	33.379	0.010
		18 0.061	0.017	34.289	0.012
		19 -0.020	-0.016	34.386	0.017
		20 0.011	-0.042	34.414	0.023

Il ne s'agit pas d'une marche au hasard (les probabilités critiques de la Q -Stat sont toutes très largement inférieures à 0,05), le processus est à mémoire, il existe donc une représentation dans la classe des processus ARMA.

c) Recherche des ordres p et q de la représentation ARMA

Compte tenu de la forme des corrélogrammes simple et partiel nous sélectionnons un modèle ARMA(1, 1). Nous constatons que la constante n'est pas significativement différente de 0, et des essais d'autres représentations concurrentes (ARMA(2, 0) ou ARMA(0, 2)) indiquent des valeurs des critères d'information (AIC ou SC) supérieures au modèle retenu.

Dependent Variable: D(X2)				
Sample (adjusted): 3 220				
Included observations: 218 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.437628	0.127029	3.445091	0.0007
MA(1)	- 0.743275	0.094329	- 7.879625	0.0000

d) Validation de la représentation

- Tests de Student sur les coefficients

Les coefficients sont tous significativement différents de 0 (probabilités critiques inférieures à 0,05).

- Analyse des résidus

Corrélogramme du résidu

Sample: 3 220
Included observations: 218
Q-statistic probabilities adjusted for 2 ARMA term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.014 0.014 0.0447			
		2 -0.051 -0.052 0.6303			
		3 -0.046 -0.044 1.0970 0.295			
		4 0.068 0.067 2.1379 0.343			
		5 0.050 0.044 2.7006 0.440			
		6 0.092 0.096 4.6062 0.330			
		7 -0.048 -0.041 5.1389 0.399			
		8 0.010 0.020 5.1629 0.523			
		9 -0.067 -0.072 6.1881 0.518			
		10 0.006 -0.009 6.1973 0.625			
		11 -0.141 -0.153 10.763 0.291			
		12 0.032 0.026 11.028 0.355			
		13 -0.030 -0.033 11.235 0.424			
		14 0.117 0.121 14.456 0.273			
		15 -0.043 -0.016 14.892 0.314			
		16 0.009 0.033 14.910 0.384			
		17 -0.058 -0.030 15.711 0.402			
		18 0.049 0.020 16.265 0.433			
		19 -0.009 -0.019 16.306 0.502			
		20 0.024 -0.017 16.450 0.561			

Corrélogramme du résidu au carré

Sample: 3 220
Included observations: 218
Q-statistic probabilities adjusted for 2 ARMA term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.019 -0.019 0.0785			
		2 -0.056 -0.056 0.7779			
		3 -0.015 -0.017 0.8290 0.363			
		4 -0.072 -0.076 1.9862 0.370			
		5 -0.065 0.061 2.9377 0.401			
		6 -0.116 -0.125 6.0054 0.199			
		7 -0.004 -0.002 6.0093 0.305			
		8 -0.094 -0.116 8.0142 0.237			
		9 -0.119 0.128 11.231 0.126			
		10 -0.098 0.064 13.495 0.096			
		11 -0.059 -0.029 14.314 0.112			
		12 -0.028 0.009 14.494 0.152			
		13 -0.028 -0.001 14.677 0.198			
		14 -0.103 0.085 17.192 0.143			
		15 0.007 0.019 17.205 0.190			
		16 -0.000 0.034 17.205 0.245			
		17 -0.089 -0.085 19.109 0.209			
		18 -0.088 -0.067 20.981 0.179			
		19 0.001 -0.058 20.981 0.227			
		20 -0.032 -0.012 21.224 0.268			

Le corrélogramme du résidu indique qu'il s'agit d'un processus sans mémoire, le corrélogramme du résidu au carré n'indique aucun terme de significativement différent de 0 ; les résidus sont donc homoscedastiques. Les résidus sont donc un bruit blanc.

Les résidus sont-ils gaussiens ? La statistique de Jarque-Bera ($JB = 5,88$) indique une probabilité critique de 0,052, nous acceptons (de justesse !) l'hypothèse de normalité des résidus.

La représentation est validée, x_{2t} est un processus ARIMA(1, 1, 1) sans constante.

e) *Prévision*

La prévision est donnée ($E(\varepsilon_t) = 0$ pour $t > 220$) par :

$$\Delta x_{2t} = \theta_1 \times \Delta x_{2t-1} + \alpha_1 \varepsilon_{t-1}$$
$$\Delta x_{2t} = 0,437 \times \Delta x_{2t-1} - 0,743 \varepsilon_{t-1}$$
$$x_{2t} = x_{2t-1} + \Delta x_{2t}$$
$$\Delta x_{2,221} = 0,437 \times \Delta x_{2,220} - 0,743 \times \varepsilon_{220}$$
$$\Delta x_{2,221} = 0,437 \times -1,318 - 0,743 \times -0,889 = 0,084$$
$$x_{2,221} = x_{2,220} + \Delta x_{2,221} = 19,61 + 0,084 = 19,69$$
$$\Delta x_{2,222} = 0,437 \times \Delta x_{2,221} - 0,743 \times \varepsilon_{221}$$
$$\Delta x_{2,222} = 0,437 \times 0,084 - 0,743 \times 0 = 0,037$$
$$x_{2,222} = x_{2,221} + \Delta x_{2,222} = 19,69 + 0,037 = 19,73$$
$$\Delta x_{2,223} = 0,437 \times \Delta x_{2,222} - 0,743 \times \varepsilon_{222}$$
$$\Delta x_{2,223} = 0,437 \times 0,037 - 0,743 \times 0 = 0,016$$
$$x_{2,223} = x_{2,222} + \Delta x_{2,223} = 19,73 + 0,016 = 19,75$$

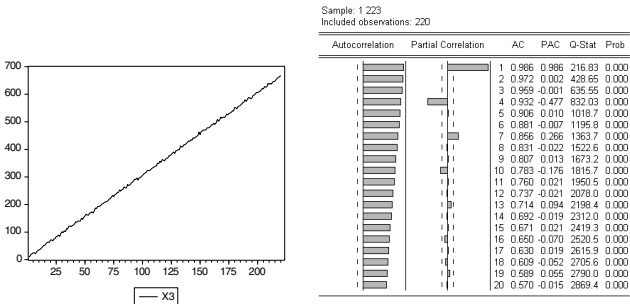
Les résultats sont présentés au tableau 4.

Tableau 4 – Calcul des prévisions du processus ARIMA(1, 1, 1)

t	x_{2t}	Δx_{2t}	$\varepsilon_t = \text{résidus}$
218	20,438	3,328	2,464
219	20,930	0,492	0,867
220	19,611	- 1,318	- 0,889
221	19,695	0,084	0
222	19,732	0,037	0
223	19,748	0,016	0

3) *Analyse du processus x_3*

Le graphique 4 présente l'évolution du processus x_3 et son corrélogramme, à leur lecture le processus est clairement non stationnaire.



Graphique 4 – Évolution et fonction d'autocorrélation simple du processus x_3

a) Stratégie simplifiée de test : tests de Phillips-Perron (choix automatique de la fenêtre de Newey-West)

Estimation du modèle [3]

Null Hypothesis: X3 has a unit root				
Exogenous: Constant, Linear Trend				
Bandwidth: 1 (Newey-West using Bartlett kernel)				
			Adj. t-Stat	Prob.*
Phillips-Perron test statistic			- 15.03107	0.0000
Test critical values:	1% level		- 4.000511	
	5% level		- 3.430477	
	10% level		- 3.138828	
Phillips-Perron Test Equation				
Dependent Variable: D(X3)				
Sample (adjusted): 2 220				
Included observations: 219 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
X3(- 1)	- 1.023098	0.068065	- 15.03108	0.0000
C	10.33495	0.639662	16.15689	0.0000
@TREND(1)	3.065841	0.203980	15.03013	0.0000

Nous refusons l'hypothèse H_0 de racine unitaire, nous pouvons donc interpréter directement les probabilités critiques calculées selon les lois de Student, refus aussi de l'hypothèse $b = 0$.

Le processus x_3 est un TS (*Trend Stationnary*), $x_{3t} = a_0 + a_1t + \varepsilon_t$ avec $t = 1, 2, \dots, n$. La bonne méthode de stationnarisation consiste donc à effectuer une régression sur le temps.

b) Modélisation

Soit la régression de x_3 sur une tendance déterministe $t = 1, 2, \dots, n$

Dependent Variable: X3				
Method: Least Squares				
Included observations: 220				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.171985	0.409466	17.51547	0.0000
TENDANCE	2.996632	0.003213	932.7294	0.0000

c) Validation de la représentation

- Tests de Student sur les coefficients

Les coefficients sont tous significativement différents de 0 (probabilités critiques inférieures à 0,05).

• Analyse des résidus

Corrélogramme du résidu

Sample: 1 220
Included observations: 220

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob		
1	1	1	1	0.023	0.023	0.1187	0.730
1	1	2	0.008	0.008	0.1334	0.935	
1	1	3	0.035	0.035	0.4059	0.939	
1	1	4	-0.040	-0.038	0.7625	0.943	
1	1	5	-0.024	-0.027	0.8957	0.971	
1	1	6	-0.096	-0.098	2.9818	0.811	
1	1	7	0.011	0.009	3.0085	0.884	
1	1	8	-0.048	-0.047	3.5457	0.896	
1	1	9	0.035	0.038	3.8266	0.922	
1	1	10	-0.018	-0.026	3.9039	0.952	
1	1	11	0.062	0.061	4.7971	0.941	
1	1	12	0.061	0.049	5.6712	0.932	
1	1	13	0.122	0.132	9.2099	0.757	
1	1	14	-0.126	-0.139	12.965	0.529	
1	1	15	0.006	0.010	12.972	0.604	
1	1	16	0.075	0.067	14.330	0.574	
1	1	17	-0.085	-0.045	16.087	0.518	
1	1	18	-0.018	-0.025	16.166	0.581	
1	1	19	-0.020	0.001	16.261	0.640	
1	1	20	0.018	0.000	16.337	0.695	

Corrélogramme du résidu au carré

Sample: 1 220
Included observations: 220

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	1	1 - 0.086	-0.086	1.6666	0.197
1	1	2 0.006	-0.001	1.6748	0.433
1	1	3 0.045	0.046	2.1362	0.545
1	1	4 -0.061	-0.053	2.9703	0.563
1	1	5 -0.079	-0.090	4.3945	0.494
1	1	6 0.021	0.006	4.4993	0.609
1	1	7 -0.089	-0.082	6.3173	0.503
1	1	8 -0.109	-0.124	9.0763	0.336
1	1	9 0.011	-0.022	9.1030	0.428
1	1	10 -0.067	-0.069	10.147	0.428
1	1	11 -0.024	-0.041	10.281	0.505
1	1	12 -0.005	-0.043	10.288	0.591
1	1	13 -0.037	-0.061	10.613	0.643
1	1	14 0.028	0.002	10.795	0.702
1	1	15 0.010	-0.026	10.817	0.765
1	1	16 0.123	0.105	14.459	0.565
1	1	17 0.024	0.024	14.600	0.624
1	1	18 -0.101	-0.129	17.042	0.520
1	1	19 -0.030	-0.074	17.261	0.572
1	1	20 -0.050	-0.073	17.867	0.596

Le corrélogramme du résidu indique qu'il s'agit d'une marche au hasard, le corrélogramme du résidu au carré n'indique aucun terme de significativement différent de 0 ; les résidus sont donc homoscédastiques. Les résidus sont donc un bruit blanc.

Les résidus sont-ils gaussiens ? La statistique de Jarque-Bera ($JB = 1,249$) indique une probabilité critique de 0,54, nous acceptons l'hypothèse de normalité des résidus.

La représentation est validée x_{3t} est un processus TS.

d) Prévision

La prévision est donnée par :

$$x_{3t} = a_0 + a_1 t + \varepsilon_t \text{ avec } (E(\varepsilon_t) = 0) \forall t$$

$$x_{3t} = 7,17 + 2,996 \times t$$

$$x_{3,221} = 7,17 + 2,996 \times 221 = 669,428$$

$$x_{3,222} = 7,17 + 2,996 \times 222 = 672,424$$

$$x_{3,223} = 7,17 + 2,996 \times 223 = 675,421$$

10. La modélisation VAR

La modélisation économétrique classique à plusieurs équations structurelles (cf. chapitre 8) a connu beaucoup de critiques (Granger [1969] et Sims [1980]) et de défaillances face à un environnement économique très perturbé. Les prévisions élaborées à l'aide de ces modèles se sont révélées très médiocres. Les critiques principales formulées à l'encontre de ces modèles structurels concerne la simultanéité des relations et la notion de variable exogène. La représentation¹ VAR (Vector AutoRegressive) – généralisation des modèles autorégressifs (AR) au cas multivarié (cf. chapitre 9) – apporte une réponse statistique à l'ensemble de ces critiques. Dans cette représentation, les variables sélectionnées² en fonction du problème étudié ont toutes, *a priori*, le même statut et on s'intéresse alors à des relations purement statistiques.

Nous présentons tout d'abord en section I. la spécification générale d'un modèle VAR, puis nous abordons les méthodes d'estimation et de prévision en section II. La dynamique d'un modèle VAR et l'analyse des chocs sont traitées dans la section III. ; enfin, la section IV est consacrée à la recherche du sens de la causalité entre les variables.

-
1. Le terme de représentation convient mieux que le terme de modélisation, car l'économiste ne spécifie pas des relations économiques en tant que telles.
 2. En général celles d'un modèle théorique.

I. Représentation d'un modèle VAR

A. Exemple introductif

Soit une représentation VAR dans laquelle on considère deux variables y_{1t} et y_{2t} .

Chacune de ces variables est fonction de ses propres valeurs passées et de celles de l'autre.

Par exemple le modèle VAR d'un ordre $p = 4$, s'écrit :

$$\begin{aligned}y_{1t} &= a_1 + \sum_{i=1}^4 b_{1i} y_{1t-i} + \sum_{i=1}^4 c_{1i} y_{2t-i} - d_1 y_{2t} + \varepsilon_{1t} \\y_{2t} &= a_2 + \sum_{i=1}^4 b_{2i} y_{1t-i} + \sum_{i=1}^4 c_{2i} y_{2t-i} - d_2 y_{1t} + \varepsilon_{2t}\end{aligned}$$

Les variables y_{1t} et y_{2t} sont considérées comme étant stationnaires (*cf.* chapitre 9), les perturbations ε_{1t} et ε_{2t} (les innovations ou les chocs) sont des bruits blancs de variances constantes $\sigma_{\varepsilon_1}^2$ et $\sigma_{\varepsilon_2}^2$ et non autocorrélées. Nous pouvons immédiatement constater l'abondance de paramètres à estimer (ici 20 coefficients) et les problèmes de perte de degrés de liberté qui en résultent. À la lecture de ce modèle, il apparaît qu'il n'est pas sous forme réduite : en effet, y_{1t} a un effet immédiat sur y_{2t} et réciproquement y_{2t} a un effet immédiat sur y_{1t} . Ce système initial est appelé *forme structurelle* de la représentation VAR. Sous forme matricielle, ce modèle devient :

$$BY_t = A_0 + \sum_{i=1}^4 A_i Y_{t-i} + \varepsilon_t \quad [1]$$

avec :

$$B = \begin{bmatrix} 1 & d_1 \\ d_2 & 1 \end{bmatrix} \quad Y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} \quad A_0 = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \quad A_i = \begin{bmatrix} b_{1i} & c_{1i} \\ b_{2i} & c_{2i} \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}.$$

Pour obtenir un modèle VAR sous forme standard, on multiplie de part et d'autre dans [1] par B^{-1} . Le modèle sous forme standard s'écrit :

$$\begin{aligned}y_{1t} &= a_1^0 + \sum_{i=1}^4 a_{1i}^1 y_{1t-i} + \sum_{i=1}^4 a_{1i}^2 y_{2t-i} + v_{1t} \\y_{2t} &= a_2^0 + \sum_{i=1}^4 a_{2i}^1 y_{1t-i} + \sum_{i=1}^4 a_{2i}^2 y_{2t-i} + v_{2t}\end{aligned}$$

Dans cette spécification, les erreurs v_{1t} et v_{2t} sont fonction des innovations ε_{1t} et ε_{2t} ; en effet, puisque $v = B^{-1}\varepsilon$, on obtient :

$$v_{1t} = (\varepsilon_{1t} - d_1 \varepsilon_{2t}) / (1 - d_1 d_2) \text{ et } v_{2t} = (\varepsilon_{2t} - d_2 \varepsilon_{1t}) / (1 - d_1 d_2)$$

On démontre que :

$$E(v_{1t}) = 0; E(v_{2t}) = 0; E(v_{1t}v_{1t-i}) = 0; E(v_{2t}v_{2t-i}) = 0$$

Les erreurs ont une espérance nulle et sont non autocorrélées.

$$E(v_{1t}^2) = (\sigma_{\varepsilon 1}^2 + d_1^2 \sigma_{\varepsilon 2}^2)/(1 - d_1 d_2)^2; E(v_{2t}^2) = (\sigma_{\varepsilon 2}^2 + d_2^2 \sigma_{\varepsilon 1}^2)/(1 - d_1 d_2)^2$$

La variance de l'erreur est constante (indépendante du temps).

$$E(v_{1t}v_{2t}) = -(d_2 \sigma_{\varepsilon 1}^2 + d_1 \sigma_{\varepsilon 2}^2)/(1 - d_1 d_2)^2$$

Si $d_1 = d_2 = 0$, les variables y_{1t} et y_{2t} n'ont pas d'influence synchrone l'une sur l'autre, les erreurs v_{1t} et v_{2t} sont non corrélées. Dans le cas contraire, les erreurs v_{1t} et v_{2t} sont corrélées et donc une variation de l'une de ces erreurs à un instant donné a un impact sur l'autre.

B. La représentation générale

La généralisation de la représentation VAR à k variables et p décalages (notée VAR(p)) s'écrit sous forme matricielle :

$$Y_t = A_0 + A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + v_t$$

$$\text{avec } Y_t = \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{k,t} \end{bmatrix}; A_{i \neq 0} = \begin{bmatrix} a_{1i}^1 & a_{1i}^2 & \dots & a_{1i}^k \\ a_{2i}^1 & a_{2i}^2 & & a_{2i}^k \\ \vdots & & & \\ a_{ki}^1 & a_{ki}^2 & & a_{ki}^k \end{bmatrix}; A_0 = \begin{bmatrix} a_1^0 \\ a_2^0 \\ \vdots \\ a_k^0 \end{bmatrix}; v_t = \begin{bmatrix} v_{1t} \\ v_{2t} \\ \vdots \\ v_{kt} \end{bmatrix}$$

On note : $\sum_v = E(v_t v_t')$, la matrice de dimension (k, k) des variances covariances des erreurs. Cette matrice est bien sûr inconnue.

Cette représentation peut s'écrire à l'aide de l'opérateur retard :

$$(I - A_1 D - A_2 D^2 - \dots - A_p D^p) Y_t = A_0 + v_t \text{ ou encore } A(D) Y_t = A_0 + v_t$$

Condition de stationnarité

Un modèle VAR est stationnaire, s'il satisfait les trois conditions classiques :

- $E(Y_t) = \mu \forall t$;
- $\text{Var}(Y_t) < \infty$;
- $\text{Cov}(Y_t, Y_{t+k}) = E[(Y_t - \mu)(Y_{t+k} - \mu)] = \Gamma_k \forall t$.

On démontre¹ qu'un processus VAR(p) est stationnaire si le polynôme défini à partir du déterminant : $\det(I - A_1 z - A_2 z^2 - \dots - A_p z^p) = 0$ a ses racines à l'extérieur du cercle unité du plan complexe.

1. Cf. Hamilton J. D., page 259, 1994.

Exercice n° 1

Recherche des conditions de stationnarité d'un modèle VAR

Soit le processus :

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} + \begin{bmatrix} 0,2 & 0,7 \\ 0,3 & 0,4 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

On demande d'étudier les conditions de stationnarité.

Solution

Le déterminant est égal à :

$$\det \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0,2 & 0,7 \\ 0,3 & 0,4 \end{bmatrix} z \right) = 1 - 0,6z - 0,13z^2 \Rightarrow \begin{cases} z_1 = 1,30 \\ z_2 = -5,91 \end{cases}$$

Les deux racines sont en valeur absolue supérieures à 1, le processus est donc stationnaire.

C. La représentation ARMAX

La représentation précédente peut être généralisée¹, par analogie avec les processus ARMA(p, q), à un modèle dont les erreurs sont autocorrélées d'ordre q .

$$Y_t = A_0 + A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + v_t + B_1 v_{t-1} + B_2 v_{t-2} + \dots + B_q v_{t-q}$$

Il s'agit d'un processus ARMA multivarié noté : ARMAX ou parfois VARMA.

Les conditions de stationnarité sont analogues à celles d'un processus ARMA univarié : un processus VAR est toujours inversible, il est stationnaire lorsque les racines de son polynôme sont à l'extérieur du cercle unité du plan complexe.

- Un processus VMA est toujours stationnaire. Il est inversible si les racines de son polynôme retard sont à l'extérieur du cercle unité du plan complexe.
- Les conditions de stationnarité et d'inversibilité d'un ARMAX sont données par la partie VAR et la partie VMA de l'ARMAX.

1. Il s'agit d'une application au cas multivarié du théorème de décomposition de Wold (1954).

II. Estimation des paramètres

Les paramètres du processus VAR ne peuvent être estimés que sur des séries chronologiques stationnaires. Ainsi, après étude des caractéristiques des séries, soit les séries sont stationnarisées par différence, préalablement à l'estimation des paramètres dans le cas d'une tendance stochastique, soit il est possible d'ajouter une composante tendance à la spécification VAR, dans le cas d'une tendance déterministe.

A. Méthode d'estimation

Dans le cas d'un processus VAR, chacune des équations peut être estimée par les MCO, indépendamment les unes des autres (ou par une méthode de maximum de vraisemblance).

Soit le modèle $\text{VAR}(p)$ estimé :

$$Y_t = \hat{A}_0 + \hat{A}_1 Y_{t-1} + \hat{A}_2 Y_{t-2} + \dots + \hat{A}_p Y_{t-p} + e$$

e étant le vecteur de dimension $(k, 1)$ des résidus d'estimation $e_{1t}, e_{2t}, \dots, e_{kt}$.

Et on note : \sum_e la matrice des variances covariances estimées des résidus du modèle.

Les coefficients du processus VAR ne peuvent être estimés qu'à partir de séries stationnaires. Ainsi, après étude des caractéristiques des chroniques, soit les séries sont stationnarisées par différence, préalablement à l'estimation des paramètres dans le cas d'une tendance stochastique, soit il est possible d'ajouter une composante tendance à la spécification VAR, dans le cas d'une tendance déterministe.

De même, nous pouvons ajouter à la spécification VAR des variables binaires afin de corriger un mouvement saisonnier ou une période anormale.

B. Détermination du nombre de retards

Pour déterminer le nombre de retards d'un modèle à retards échelonnés (*cf.* chapitre 7), nous avons présenté les critères de Akaike et de Schwarz. Dans le cas de la représentation VAR, ces critères peuvent être utilisés pour déterminer l'ordre p du modèle. La procédure de sélection de l'ordre de la représentation consiste à estimer tous les modèles VAR pour un ordre allant de 0 à h (h étant le retard maximum admissible par la théorie économique ou par les données

disponibles). Les fonctions $AIC(p)$ et $SC(p)$ sont calculées de la manière suivante :

$$AIC(p) = \text{Ln} [\det |\sum_e|] + \frac{2k^2 p}{n}$$

$$SC(p) = \text{Ln} [\det |\sum_e|] + \frac{k^2 p \text{Ln}(n)}{n}$$

avec : k = nombre de variables du système ; n = nombre d'observations ; p = nombre de retards ; \sum_e = matrice des variances covariances des résidus du modèle.

Le retard p qui minimise les critères AIC ou SC est retenu.

C. Prévision

Les coefficients du modèle étant estimés, la prévision peut être calculée en n à l'horizon d'une période, par exemple pour un VAR(1), de la manière suivante :

$$\hat{Y}_n(1) = \hat{A}_0 + \hat{A}_1 Y_n$$

À l'horizon de 2 périodes, la prévision est :

$$\hat{Y}_n(2) = \hat{A}_0 + \hat{A}_1 \hat{Y}_n(1) = \hat{A}_0 + \hat{A}_1 \hat{A}_0 + \hat{A}_1^2 Y_n$$

À l'horizon de 3 périodes, la prévision s'écrit :

$$\hat{Y}_n(3) = \hat{A}_0 + \hat{A}_1 \hat{Y}_n(2) = (I + \hat{A}_1 + \hat{A}_1^2) \hat{A}_0 + \hat{A}_1^3 Y_n$$

etc.

Lorsque h tend vers l'infini, nous constatons que la prévision tend vers une valeur constante (état stationnaire) puisque $\hat{A}_i^i \rightarrow 0$ si $i \rightarrow \infty$.

L'espérance de la matrice de l'erreur de prévision est nulle, sa variance est donnée par :

$\sum_e(h) = M_0 \sum_e M'_0 + M_1 \sum_e M'_1 + \dots + M_{h-1} \sum_e M'_{h-1}$ où M_i est calculé par la formule de récurrence suivante :

$$M_i = \sum_{j=1}^{\min(p,i)} \hat{A}_j M_{i-j} \quad i = 1, 2, \dots \text{ et } M_0 = I$$

Ainsi, il vient :

$$M_1 = \hat{A}_1 ; M_2 = \hat{A}_1 M_1 + \hat{A}_2 M_0 = \hat{A}_1^2 + \hat{A}_2 ;$$

$$M_3 = \hat{A}_1 M_2 + \hat{A}_2 M_1 + \hat{A}_3 M_0 = \hat{A}_1^3 + \hat{A}_1 \hat{A}_2 + \hat{A}_2 \hat{A}_1 + \hat{A}_3$$

etc.

La variance de l'erreur de prévision pour chacune des prévisions des k variables ($\hat{\sigma}_n^2(h)$) se lit sur la première diagonale de la matrice $\sum_e(h)$. L'intervalle de prévision au seuil de $(1 - \alpha/2)$ est donné par : $\hat{Y}_n(h) \pm t^{\alpha/2} \times \hat{\sigma}_n(h)$ avec $t^{\alpha/2}$ valeur de la loi normale.

Exercice n° 2

↓ fichier C10EX2

Spécification, estimation et prévision d'un modèle VAR

Nous cherchons à modéliser, sous la forme VAR, la demande (y_{1t}) et les prix (y_{2t}) d'une matière première. Nous disposons des données trimestrielles CVS et en différences premières sur 18 ans (tableau 1).

Tableau 1 – Demande et prix CVS et en différences premières (extrait des données)

Date	y_{1t}	y_{2t}
0001Q1	-146,4	-75,6
0001Q2	19,2	1,8
...
0018Q3	19,2	-16,2
0018Q4	50,4	3,6

On demande :

- 1) de rechercher l'ordre du modèle VAR,
- 2) d'estimer les paramètres du modèle,
- 3) de calculer une prévision pour l'année 19 avec son intervalle de confiance à 95 %.

Solution

1) Rechercher l'ordre du modèle VAR

Nous allons utiliser les critères de Akaike et de Schwarz pour des décalages h allant de 0 à 4. Nous devons donc estimer quatre modèles différents et retenir celui dont les critères AIC et SC sont les plus faibles.

Estimation du modèle pour $h = 1$

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} a_1^0 \\ a_2^0 \end{bmatrix} + \begin{bmatrix} a_{11}^1 & a_{12}^1 \\ a_{21}^1 & a_{22}^1 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix}$$

Ou encore :

$$y_{1t} = a_1^0 + a_{11}^1 y_{1t-1} + a_{12}^1 y_{2t-1} + v_{1t}$$

$$y_{2t} = a_2^0 + a_{21}^1 y_{1t-1} + a_{22}^1 y_{2t-1} + v_{2t}$$

Pour estimer les paramètres de ce modèle, nous pouvons utiliser les MCO équation par équation.

$$y_{1t} = 0,00676 \times y_{1t-1} - 0,6125 \times y_{2t-1} + 17,129$$

$$y_{2t} = -0,1752 \times y_{1t-1} + 0,2992 \times y_{2t-1} - 12,862$$

Les deux résidus d'estimation sont alors calculés : $e_{1,t}$ et $e_{2,t}$ ainsi la matrice des variances covariances des résidus est donnée par :

$$\begin{aligned} \Sigma_e &= \frac{1}{n} \begin{bmatrix} \sum_1^n e_1^2 & \sum_1^n e_1 e_2 \\ \sum_1^n e_1 e_2 & \sum_1^n e_2^2 \end{bmatrix} \\ &= \frac{1}{71} \begin{bmatrix} 106500,16 & -47878,12 \\ -47878,12 & 85205,25 \end{bmatrix} = \begin{bmatrix} 1500,002 & -674,3398 \\ -674,339 & 1200,074 \end{bmatrix} \end{aligned}$$

Attention, certains logiciels (par exemple Eviews à partir de la version 4) estiment cette matrice des variances covariances en divisant la somme des carrés ou du produit par le degré de liberté de chaque équation estimée soit :

$$\sum_e = \frac{1}{71 - 2 - 1} \begin{bmatrix} 106500,16 & -47878,12 \\ -47878,12 & 85205,25 \end{bmatrix} = \begin{bmatrix} 1566,18 & -704,09 \\ -704,09 & 1253,02 \end{bmatrix}$$

Dans toute la suite des calculs nous retenons la méthode 1, soit :

$$\text{Var}(e_{1t}) = 1\,500,002 ; \text{Var}(e_{2t}) = 1\,200,074 ; \text{Cov}(e_{1t}, e_{2t}) = -674,339$$

$$\text{Soit } \det(\sum_e) = \begin{vmatrix} 1\,500,002 & -674,339 \\ -674,339 & 1\,200,074 \end{vmatrix} = 1\,345\,379.$$

Pour $p = 1$:

$$\text{AIC}(p) = \text{Ln} [\det(\sum_e)] + \frac{2k^2 p}{n} = \text{Ln}(1\,345\,379) + (2 \times 2^2 \times 1)/71 = 14,28$$

et

$$\begin{aligned} \text{SC}(p) &= \text{Ln} [\det(\sum_e)] + \frac{k^2 p \text{Ln}(n)}{n} = \text{Ln}(1\,345\,379) + (2^2 \times 1 \times \text{Ln}(71))/71 \\ &= 14,47 \end{aligned}$$

Nous pouvons calculer de la même manière :

$$\text{AIC}(2) = 14,31 ; \text{AIC}(3) = 14,39 ; \text{AIC}(4) = 14,45$$

$$\text{SC}(2) = 14,56 ; \text{SC}(3) = 14,78 ; \text{SC}(4) = 14,97$$

Le nombre de décalages retenu correspond à la valeur la plus faible des critères soit : $p = 1$.

2) Estimation des paramètres

Le modèle VAR(1) estimé s'écrit :

$$y_{1t} = 0,00676 \times y_{1t-1} - 0,6125 \times y_{2t-1} + 17,1292 + e_{1t}$$

(0,06) (4,47) (2,87)

$$R^2 = 0,28 ; n = 71 ; (\cdot) = t \text{ de Student}$$

$$y_{2t} = -0,1752 \times y_{1t-1} + 0,2992 \times y_{2t-1} - 12,862 + e_{2t}$$

(1,84) (2,44) (2,41)

$$R^2 = 0,21 ; n = 71 ; (\cdot) = t \text{ de Student}$$

3) Prévision

La prévision est calculée en faisant « tourner » le modèle :

$$\hat{y}_{19:1} = 0,0067 \times y_{18:4} - 0,6125 \times y_{18:4} + 17,129$$

$$\hat{y}_{19:1} = -0,1752 \times y_{18:4} + 0,2992 \times y_{18:4} - 12,862$$

$$\hat{y}_{19:1} = 0,0067 \times 50,4 - 0,6125 \times 3,6 + 17,129 = 15,3$$

$$\hat{y}_{19:1} = -0,1752 \times 50,4 + 0,2992 \times 3,6 - 12,862 = -20,6$$

$$\hat{y}_{19:2} = 0,0067 \times 15,3 - 0,6125 \times -20,6 + 17,129 = 29,9$$

$$\hat{y}_{19:2} = -0,1752 \times 15,3 + 0,2992 \times -20,6 - 12,862 = -21,7$$

$$\hat{y}_{19;3} = 0,0067 \times 29,9 - 0,6125 \times -21,7 + 17,129 = 30,6$$

$$\hat{y}_{19;3} = -0,1752 \times 29,9 + 0,2992 \times -21,7 - 12,862 = -24,6$$

$$\hat{y}_{19;4} = 0,0067 \times 30,6 - 0,6125 \times -24,6 + 17,129 = 32,4$$

$$\hat{y}_{19;4} = -0,1752 \times 30,6 + 0,2992 \times -24,6 - 12,862 = -25,6$$

La variance de l'erreur de prévision est donnée par la formule :

$\sum_e(h) = \sum_e + M_1 \sum_e M'_1 + \dots + M_{h-1} \sum_e M'_{h-1}$ où M_i est calculé par la formule de récurrence suivante : $M_i = \sum_{j=1}^{\min(p,i)} \hat{A}_j M_{i-j}$ $i = 1, 2, \dots$ et $M_0 = I$. Les variances se lisent sur la première diagonale de la matrice $\sum_e(h)$.

Ainsi, il vient : $M_1 = \hat{A}_1$; $M_2 = \hat{A}_1 M_1 = \hat{A}_1^2$; $M_3 = \hat{A}_1 M_2 = \hat{A}_1^3$; etc., car le modèle est un VAR(1) et donc seul \hat{A}_1 est estimé.

Soit à l'horizon $h = 1$, nous avons $\sum_e(1) = \begin{bmatrix} 1\,500,02 & -674,34 \\ -674,34 & 1\,200,1 \end{bmatrix}$. La variance de l'erreur de prévision pour $\hat{y}_{19;1}$ est égale à 1 500,02 et la variance de l'erreur de prévision pour $\hat{y}_{19;1}$ est égale à 1 200,1. L'intervalle de confiance pour $y_{19;1}$ est donné par : $15,3 \pm 1,96 \times 38,73 = [-60,64 ; 91,17]$ et pour $y_{19;1}$ est donné par :

$$-20,6 \pm 1,96 \times 34,64 = [-88,52 ; 47,27].$$

Soit à l'horizon $h = 2$, nous avons :

$$\sum_e(2) = \sum_e + \hat{A}_1 \sum_e \hat{A}'_1, \text{ soit :}$$

$$\sum_e(2) = \begin{bmatrix} 1\,500,02 & -674,34 \\ -674,34 & 1\,200,1 \end{bmatrix} + \begin{bmatrix} 0,00676 & -0,6126 \\ -0,1753 & 0,2992 \end{bmatrix} \begin{bmatrix} 1\,500,02 & -674,34 \\ -674,34 & 1\,200,1 \end{bmatrix}$$

$$\begin{bmatrix} 0,00676 & -0,1753 \\ -0,6126 & 0,2992 \end{bmatrix} = \begin{bmatrix} 1\,955,99 & -969,88 \\ -969,88 & 1\,424,37 \end{bmatrix}$$

$$\text{Pour } h = 3 : \sum_e(3) = \sum_e + \hat{A}_1 \sum_e \hat{A}'_1 + \hat{A}_1^2 \sum_e \hat{A}'_1$$

$$\sum_e(3) = \begin{bmatrix} 1\,500,02 & -674,34 \\ -674,34 & 1\,200,1 \end{bmatrix} + \begin{bmatrix} 0,00676 & -0,6126 \\ -0,1753 & 0,2992 \end{bmatrix} \begin{bmatrix} 1\,500,02 & -674,34 \\ -674,34 & 1\,200,1 \end{bmatrix}$$

$$\begin{bmatrix} 0,00676 & -0,1753 \\ -0,6126 & 0,2992 \end{bmatrix} + \begin{bmatrix} 0,00676 & -0,6126 \\ -0,1753 & 0,2992 \end{bmatrix}^2 \begin{bmatrix} 1\,500,02 & -674,34 \\ -674,34 & 1\,200,1 \end{bmatrix}$$

$$\begin{bmatrix} 0,00676 & -0,1753 \\ -0,6126 & 0,2992 \end{bmatrix}^2 = \begin{bmatrix} 2\,042,63 & -1\,043,88 \\ -1\,043,88 & 1\,489,48 \end{bmatrix}$$

$$\text{Pour } h = 4 : \sum_e(4) = \sum_e + \hat{A}_1 \sum_e \hat{A}'_1 + \hat{A}_1^2 \sum_e \hat{A}'_1 + \hat{A}_1^3 \sum_e \hat{A}'_1$$

$$\sum_e(4) = \begin{bmatrix} 1\,500,02 & -674,34 \\ -674,34 & 1\,200,1 \end{bmatrix} + \begin{bmatrix} 0,00676 & -0,6126 \\ -0,1753 & 0,2992 \end{bmatrix} \begin{bmatrix} 1\,500,02 & -674,34 \\ -674,34 & 1\,200,1 \end{bmatrix}$$

$$\begin{bmatrix} 0,00676 & -0,1753 \\ -0,6126 & 0,2992 \end{bmatrix} + \begin{bmatrix} 0,00676 & -0,6126 \\ -0,1753 & 0,2992 \end{bmatrix}^2 \begin{bmatrix} 1\,500,02 & -674,34 \\ -674,34 & 1\,200,1 \end{bmatrix}$$

$$\begin{bmatrix} 0,00676 & -0,1753 \\ -0,6126 & 0,2992 \end{bmatrix}^2 + \begin{bmatrix} 0,00676 & -0,6126 \\ -0,1753 & 0,2992 \end{bmatrix}^3 \begin{bmatrix} 1\,500,02 & -674,34 \\ -674,34 & 1\,200,1 \end{bmatrix}$$

$$\begin{bmatrix} 0,00676 & -0,1753 \\ -0,6126 & 0,2992 \end{bmatrix}^3 = \begin{bmatrix} 2\,067,67 & -1\,064,00 \\ -1\,064,00 & 1\,505,73 \end{bmatrix}$$

$$IC_{y19;2} = 29,9 \pm 1,96 \times 44,22 = [-56,82; 116,54]$$

$$Ic_{y19;2} = -21,7 \pm 1,96 \times 37,74 = [-95,68; 52,26]$$

$$IC_{y19;3} = 30,6 \pm 1,96 \times 45,19 = [-57,95; 119,21]$$

$$Ic_{y19;3} = -24,6 \pm 1,96 \times 38,59 = [-100,23; 51,05]$$

$$IC_{y19;4} = 32,4 \pm 1,96 \times 45,47 = [-56,72; 121,52]$$

$$Ic_{y19;4} = -25,6 \pm 1,96 \times 38,80 = [-101,64; 50,46]$$

III. Dynamique d'un modèle VAR

Les modèles VAR permettent d'analyser les effets de la politique économique, cela au travers de simulations de chocs aléatoires¹ et de la décomposition de la variance de l'erreur. Cependant, cette analyse s'effectue en postulant la constance de l'environnement économique « toutes choses étant égales par ailleurs ».

A. Représentation VMA d'un processus VAR

Nous avons vu qu'un modèle AR(1) a une représentation MA(∞). Par analogie, nous pouvons démontrer qu'un VAR(1) a une représentation VMA(∞). Le modèle sous cette forme va permettre de mesurer l'impact sur les valeurs présentes d'une variation des innovations (ou des chocs) v_{1t} et de v_{2t} .

Soit la représentation VAR(p) stationnaire :

$$Y_t = A_0 + A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + v_t$$

Sa représentation VMA(∞) est donnée par :

$$Y_t = \mu + v_t + M_1 v_{t-1} + M_2 v_{t-2} + \dots = \mu + \sum_{i=0}^{\infty} M_i v_{t-i}$$

avec : $\mu = (I - A_1 - A_2 - \dots - A_p)^{-1} \times A_0$

et $M_i = \sum_{j=1}^{\min(p,i)} A_j M_{i-j}$ $i = 1, 2, \dots$ et $M_0 = I$

1. Le terme d'innovation est aussi souvent employé.

Sous cette forme, la matrice M apparaît comme un « multiplicateur d'impact », c'est-à-dire que c'est au travers de cette matrice qu'un choc se répercute tout le long du processus. Une variation à un instant donné t de v_t affecte toutes les valeurs suivantes de Y_t , l'effet d'un choc (ou d'une innovation) est donc permanent et va en s'amortissant.

B. Analyse et orthogonalisation des « chocs »

L'analyse d'un choc consiste à mesurer l'impact de la variation d'une innovation sur les variables. Prenons l'exemple du modèle estimé de l'exercice 2 :

$$y_{1t} = 0,00676 * y_{1t-1} - 0,6125 * y_{2t-1} + 17,129 + e_{1t}$$

$$y_{2t} = -0,1752 * y_{1t-1} + 0,2992 * y_{2t-1} - 12,863 + e_{2t}$$

Une variation à un instant donné de e_{1t} a une conséquence immédiate sur y_{1t} puis sur y_{2t+1} et y_{2t+2} , par exemple s'il se produit en t un choc sur e_{1t} égal à 1, nous avons l'impact suivant :

$$\text{En } t : \begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{À la période } t + 1 : \begin{bmatrix} \Delta y_{1t+1} \\ \Delta y_{2t+1} \end{bmatrix} = \begin{bmatrix} 0,00676 & -0,6126 \\ -0,1753 & 0,2992 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0,0067 \\ -0,175 \end{bmatrix}$$

$$\text{À la période } t + 2 : \begin{bmatrix} \Delta y_{1t+2} \\ \Delta y_{2t+2} \end{bmatrix} = \begin{bmatrix} 0,00676 & -0,6126 \\ -0,1753 & 0,2992 \end{bmatrix} \begin{bmatrix} 0,0067 \\ -0,175 \end{bmatrix} = \begin{bmatrix} 0,107 \\ -0,054 \end{bmatrix}$$

etc.

Les différentes valeurs ainsi calculées constituent la « fonction de réponse impulsionnelle ». Dans ce schéma, nous faisons l'hypothèse que les résidus e_{1t} et e_{2t} sont indépendants entre eux. Or, cette hypothèse est rarement vérifiée ; en effet, nous avons montré en section I., qu'il peut exister une corrélation entre les erreurs v_{1t} et de v_{2t} , cette corrélation pouvant être mesurée à l'aide des résidus d'estimation : $\rho_{e_1 e_2} = \frac{\text{Cov}(e_1, e_2)}{\sigma_{e_1} \times \sigma_{e_2}}$.

Cependant si ce coefficient nous renseigne bien sur la liaison contemporaine entre les deux résidus, il n'indique pas le sens de la causalité. C'est pourquoi, nous devons faire une hypothèse supplémentaire concernant la relation entre les erreurs : par exemple, dans l'exercice 2, la réflexion économique laisse présager qu'une variation du prix (y_{2t}) a un impact sur la demande (y_{1t}). Un choc sur y_{1t} n'a donc pas d'impact contemporain sur y_{2t} ; en revanche un choc sur y_{2t} a un impact contemporain (mesuré par le coefficient de corrélation entre les résidus) sur y_{1t} . Dans cet exemple, le coefficient de corrélation est égal à $\rho_{e_1 e_2} = \frac{\text{Cov}(e_1, e_2)}{\sigma_{e_1} \times \sigma_{e_2}} = -0,503$; en faisant l'hypothèse que $\sigma_{e_1} \approx \sigma_{e_2}$, un choc aléatoire de 1 sur y_{2t} se traduit par une baisse de $-0,503$ sur y_{1t} . Le tableau 2 illustre cette dissymétrie des innovations sur les variables.

Tableau 2 – Analyse des chocs sur les variables y_{1t} et y_{2t}
avec $v_{1t} = -0,503 \times v_{2t}$

t	y_{1t}	y_{2t}	t	y_{1t}	y_{2t}
1	1,000	0,000	1	-0,530	1,000
2	0,007	-0,175	2	-0,616	0,387
3	0,107	-0,054	3	-0,241	0,224
4	0,034	-0,035	4	-0,139	0,109
5	0,022	-0,016	5	-0,068	0,057
6	0,010	-0,009	6	-0,035	0,029
7	0,005	-0,004	7	-0,018	0,015
8	0,003	-0,002	8	-0,009	0,008
9	0,001	-0,001	9	-0,005	0,004
10	0,001	-0,001	10	-0,002	0,002
11	0,000	0,000	11	-0,001	0,001
12	0,000	0,000	12	-0,001	0,001

Le choix du sens de l'impact est donc très important et conditionne les valeurs obtenues. Nous pouvons observer que l'effet d'une innovation s'estompe au cours du temps. Cela caractérise un processus VAR stationnaire.

Le graphique 1 présente les deux fonctions de réponses impulsionnelles.

Le problème de la corrélation contemporaine des erreurs et donc de l'impact d'un choc sur une variable est traité, d'une manière générale, par la recherche d'une représentation à erreurs orthogonales. Reprenons l'exemple d'un modèle VAR à deux variables :

$$\begin{aligned}y_{1,t} &= a_1 y_{1,t-1} + b_1 y_{2,t-1} + v_{1,t} \\ y_{2,t} &= a_2 y_{1,t-1} + b_2 y_{2,t-1} + v_{2,t}\end{aligned}$$

avec $\text{Var}(v_{1,t}) = \sigma_{v_1}^2$, $\text{Var}(v_{2,t}) = \sigma_{v_2}^2$ et $\text{Cov}(v_{1,t}, v_{2,t}) = k \neq 0$.

En calculant $y_{2,t} - (k/\sigma_{v_1}^2)y_{1,t}$ on obtient :

$$y_{2,t} = k/\sigma_{v_1}^2 y_{1,t} + (a_2 - a_1 \times k/\sigma_{v_1}^2)y_{1,t-1} + (b_2 - b_1 \times k/\sigma_{v_1}^2)y_{2,t-1} + v_{2,t} - k/\sigma_{v_1}^2 v_{1,t}$$

On pose : $v_t = v_{2,t} - k/\sigma_{v_1}^2 v_{1,t}$

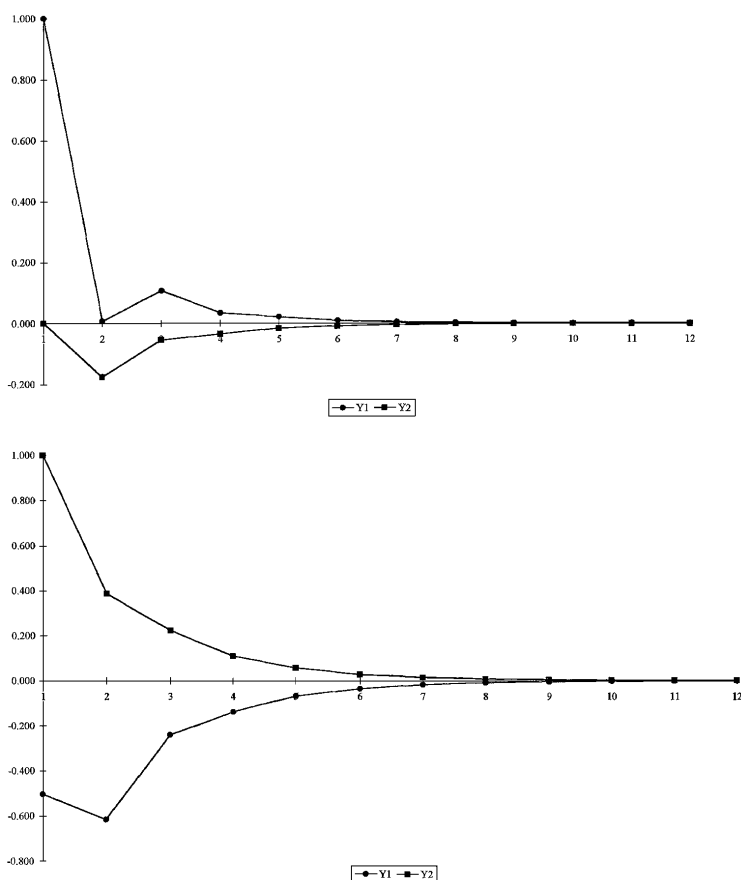
$$\text{Cov}(v_{1,t}, v_t) = \text{E}(v_{1,t} v_t) = \text{Cov}(v_{1,t}, v_{2,t}) - k/\sigma_{v_1}^2 \text{E}(v_{1,t}^2) = k - k = 0.$$

Les erreurs (ou innovations) ne sont plus corrélées. L'analyse des chocs peut donc s'effectuer sur les deux équations suivantes dont les erreurs sont orthogonales :

$$\begin{aligned}y_{1,t} &= a_1 y_{1,t-1} + b_1 y_{2,t-1} + v_{1,t} \\ y_{2,t} &= k/\sigma_{v_1}^2 y_{1,t} + (a_2 - a_1 \times k/\sigma_{v_1}^2)y_{1,t-1} + (b_2 - b_1 \times k/\sigma_{v_1}^2)y_{2,t-1} + v_{2,t} - k/\sigma_{v_1}^2 v_{1,t}\end{aligned}$$

Nous remarquons que, sous cette forme, nous avons fait apparaître dans la deuxième équation la causalité instantanée entre $y_{2,t}$ et $y_{1,t}$ telle qu'elle existait dans le VAR structurel (cf. I. A.). Cependant, la causalité instantanée entre $y_{1,t}$ et $y_{2,t}$ n'apparaît pas dans la première équation. Dans cette procédure d'orthogonalisation, les résidus sont indépendants, nous pouvons faire apparaître soit une causalité instantanée entre $y_{2,t}$ et $y_{1,t}$ soit une causalité instantanée entre $y_{1,t}$ et $y_{2,t}$ mais pas les deux à la fois : il s'agit de l'ordre de décomposition.

La généralisation à un modèle VAR à k variables nécessite le recours à des procédures d'orthogonalisation¹ de matrice et s'avère donc complexe. Il convient de noter que les résultats sont influencés par le choix de l'équation servant de base à la transformation. Les résultats sont différents si la transformation affecte y_{1t} à la place de y_{2t} , c'est pourquoi le choix² de l'ordre des variables modifie les résultats obtenus.



Graphique 1 – Fonctions de réponses impulsionnelles

1. Cf. Hamilton J. D. page 318, 1994.

2. Les logiciels d'économétrie offrent la possibilité de choisir l'ordre des variables et permettent ainsi de simuler les différents cas de figure.

C. Décomposition de la variance

La décomposition de la variance de l'erreur de prévision a pour objectif de calculer pour chacune des innovations sa contribution à la variance de l'erreur. Par une technique mathématique¹, on peut écrire la variance de l'erreur de prévision à un horizon h en fonction de la variance de l'erreur attribuée à chacune des variables ; il suffit ensuite de rapporter chacune de ces variances à la variance totale pour obtenir son poids relatif en pourcentage.

Reprenons notre modèle VAR(1) à deux variables y_{1t} et y_{2t} , la variance de l'erreur de prévision pour y_{1t+h} peut s'écrire :

$$\sigma_{y_1}^2(h) = \sigma_{\varepsilon_1}^2 [m_{11}^2(0) + m_{11}^2(1) + \dots + m_{11}^2(h-1)] \\ + \sigma_{\varepsilon_2}^2 [m_{22}^2(0) + m_{22}^2(1) + \dots + m_{22}^2(h-1)]$$

où les m_{ii} sont les termes de la matrice M de la section III.A.

À l'horizon h , la décomposition de la variance, en pourcentage, des propres innovations de y_{1t} sur y_{1t} , est donnée par :

$$\frac{\sigma_{\varepsilon_1}^2 [m_{11}^2(0) + m_{11}^2(1) + \dots + m_{11}^2(h-1)]}{\sigma_{y_1}^2(h)}.$$

Et la décomposition de la variance, en pourcentage, des innovations de y_{1t} sur y_{2t} , est donnée par : $\frac{\sigma_{\varepsilon_2}^2 [m_{22}^2(0) + m_{22}^2(1) + \dots + m_{22}^2(h-1)]}{\sigma_{y_1}^2(h)}$

L'interprétation des résultats est importante :

- si un choc sur ε_{1t} n'affecte pas la variance de l'erreur de y_{2t} quel que soit l'horizon de prévision, alors y_{2t} peut être considéré comme exogène car y_{2t} évolue indépendamment de ε_{1t} ;
- *a contrario*, si un choc sur ε_{1t} affecte fortement – voire totalement – la variance de l'erreur de y_{2t} , alors y_{2t} est considéré comme endogène.

Dans la pratique, les résultats ne sont pas aussi marqués mais indiquent la contribution de chacune des variables à la variance de l'erreur.

D. Choix de l'ordre de décomposition

Il est à noter que le problème de la corrélation contemporaine des erreurs, et donc de l'impact d'un choc sur une variable, implique un choix de décomposi-

1. La matrice Σ étant toujours définie positive, la décomposition de Cholesky permet de diagonaliser la matrice des variances covariances de l'erreur de prévision, et ainsi d'identifier les variances propres à chacune des variables. Nous ne pouvons pas aborder dans le cadre de ce manuel cette technique, le lecteur peut se référer à Hamilton J. D. (1994) pour tous les compléments mathématiques.

tion qui fournit des résultats dissymétriques en fonction de l'ordre des variables. Le problème est plus complexe si le nombre de variables est important.

Prenons l'exemple d'un VAR à 4 variables Y1, Y2, Y3 et Y4.

Si nous choisissons l'ordre de décomposition de Cholesky suivant : Y3 Y2 Y1 Y4, cela entraîne :

- un choc en période 1 sur Y3 (la variable la plus exogène) a un impact sur l'ensemble des quatre variables ;
- un choc en période 1 sur Y2 a un impact sur Y2, Y3 et Y4, absence de corrélation contemporaine avec Y3 ;
- un choc en période 1 sur Y1 a un impact sur Y1 et Y4, absence de corrélation contemporaine avec Y3 et Y2 ;
- et, enfin, un choc en période 1 sur Y4 (la moins exogène) n'a un impact que sur elle-même, absence de corrélation contemporaine avec Y3, Y2 et Y1.

Le tableau 3 illustre cet exemple de décomposition et d'analyse des chocs. Nous ne présentons que les deux premières périodes.

Tableau 3 – Exemple de décomposition Y3, Y2, Y1, Y4

Cholesky Ordering : Y3 Y2 Y1 Y4				
Response of Y1 :				
Period	Y1	Y2	Y3	Y4
1	1.603662	1.022459	1.402923	0.000000
2	– 0.395786	– 0.151091	0.488095	– 0.275476
Response of Y2 :				
Period	Y1	Y2	Y3	Y4
1	0.000000	5.434671	4.125310	0.000000
2	0.875767	– 1.251545	0.115017	– 0.105811
Response of Y3 :				
Period	Y1	Y2	Y3	Y4
1	0.000000	0.000000	4.544690	0.000000
2	– 0.241257	– 0.954324	– 1.584595	– 0.476602
Response of Y4 :				
Period	Y1	Y2	Y3	Y4
1	0.768803	1.686351	2.315711	2.091513
2	– 0.129818	– 1.104756	– 0.927459	– 0.186928

L'ordre de décomposition s'effectue donc de la variable supposée être la plus exogène vers la variable la moins exogène. Dans le doute, il convient d'effectuer plusieurs combinaisons différentes d'ordre de décomposition et d'analyser la robustesse des résultats.

Exercice n° 3

↓ fichier C10EX2

Analyse d'une fonction de réponse impulsionnelle et décomposition de la variance

À partir de la représentation VAR(1) estimée à l'exercice 2, on demande de calculer et d'interpréter les fonctions de réponses impulsionnelles et la décomposition de la variance des erreurs.

Solution

Les résultats sont issus de Eviews 3 (RATS possède une procédure « *IMPULSE.PRG* » permettant elle aussi de calculer tous ces éléments).

Nous considérons qu'une innovation sur $y_{1,t}$ n'influence pas de manière instantanée $y_{2,t}$, en revanche une innovation sur $y_{2,t}$ a une influence contemporaine sur $y_{1,t}$ (un choc sur la variable prix influence la demande mais pas réciproquement).

La matrice des variances covariances des résidus est égale (cf. exercice 2) à $\Sigma_e = \begin{pmatrix} 1\,500,002 & -674,3398 \\ -674,3398 & 1\,200,074 \end{pmatrix}$.

Calcul des chocs initiaux

L'ordre de décomposition des variables est le suivant Y2 Y1. L'analyse des chocs doit donc s'effectuer sur les deux équations suivantes dont les résidus sont orthogonalisés (cf. III. B.) :

$$y_{1,t} \text{ dépend de } v_{1,t} = (e_{1,t} - k/\sigma_{e_2}^2 e_{2,t})$$

$$y_{2,t} \text{ dépend de } v_{2,t} = e_{2,t}$$

$$\text{Var}(v_{1,t}) = \text{Var}(e_{1,t} - \text{Cov}(e_{1,t}, e_{2,t})/\sigma_{e_2}^2 e_{2,t})$$

$$\text{Var}(v_{1,t}) = \text{Var}(e_{1t}) - \frac{2\text{Cov}(e_{1t}, e_{2t})}{\text{Var}(e_{2t})} \text{Cov}(e_{1t}, e_{2t}) + \frac{[\text{Cov}(e_{1t}, e_{2t})]^2}{[\text{Var}(e_{2t})]^2} \times \text{Var}(e_{2t})$$

$$\text{Var}(v_{1,t}) = \text{Var}(e_{1t}) - \frac{2[\text{Cov}(e_{1t}, e_{2t})]^2}{\text{Var}(e_{2t})} + \frac{[\text{Cov}(e_{1t}, e_{2t})]^2}{[\text{Var}(e_{2t})]}$$

$$\text{Var}(v_{1,t}) = \text{Var}(e_{1t}) - \text{Cov}(e_{1t}, e_{2t})^2/\text{Var}(e_{2t})$$

L'amplitude du choc sur $y_{1,t}$ est égale à une fois l'écart type orthogonalisé de $v_{1,t}$ soit $\sqrt{\text{Var}(e_{1t}) - \text{Cov}(e_{1t}, e_{2t})^2/\text{Var}(e_{2t})} = \sqrt{1\,500,002 - 674,3398^2/1\,200,074} = 33,48$.
Ce choc n'a pas d'effet instantané sur $y_{2,t}$.

L'amplitude du choc sur $y_{2,t}$ est égale à une fois l'écart type orthogonalisé :

Or, $\text{Var}(v_{2,t}) = \text{Var}(e_{2,t}) = 1\,200,074$, le choc sur $y_{2,t}$ est donc égal à 34,62.

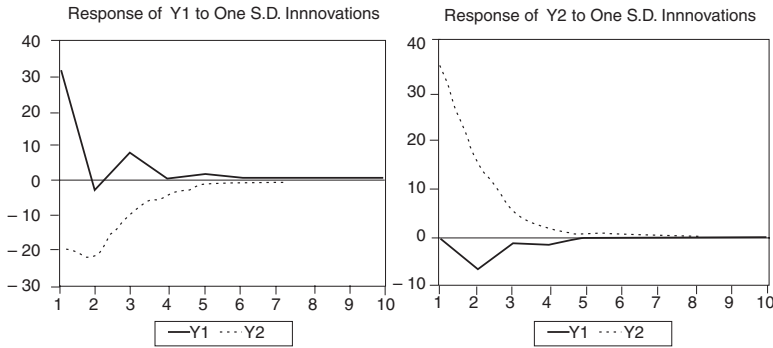
Ce choc se répercute instantanément sur $y_{1,t}$, il est donc égal, après orthogonalisation, à $34,62 \times \text{Cov}(e_{1t}, e_{2t})/\text{Var}(e_{2,t})$ soit $34,62 \times -674,3398/1\,200,074 = -19,4659$.

Une fois les chocs initiaux calculés, nous calculons les fonctions de réponse impulsionnelle comme précédemment, les chocs se répercutent ensuite sur les deux processus en s'amortissant, signe de la stationnarité du processus VAR.

Fonctions de réponse impulsionnelle à un écart type orthogonalisé (Eviews 3)

Réponse de Y1 :			Réponse de Y2 :		
Période	à un choc sur Y1	à un choc sur Y2	Période	à un choc sur Y1	à un choc sur Y2
1	33,48254	-19,46591	1	0,000000	34,64208
2	0,226363	-21,35271	2	-5,868694	13,77894
3	3,596584	-8,585080	3	-1,795947	7,866119
4	1,124479	-4,876681	4	-1,167853	3,858783
5	0,723007	-2,396789	5	-0,546587	2,009549
6	0,339717	-1,247217	6	-0,290298	1,021480
7	0,180128	-0,634172	7	-0,146419	0,524297
8	0,090912	-0,325462	8	-0,075390	0,268057
9	0,046797	-0,166407	9	-0,038496	0,137265
10	0,023898	-0,085211	10	-0,019723	0,070245
Ordre : Y2 Y1 (Y2 est la cause de Y1)					

Le graphique 2 présente ces fonctions de réponse impulsionnelles.



Graphique 2 – Fonctions de réponses impulsionnelles

Décomposition de la variance (Eviews 3)

Décomposition de la variance de Y1 :				Décomposition de la variance de Y2 :			
Période	SE	Y1	Y2	Période	SE	Y1	Y2
1	38,72986	74,73859	25,26141	1	34,64208	0,000000	100,0000
2	44,22659	57,31782	42,68218	2	37,74089	2,418013	97,58199
3	45,19547	55,51993	44,48007	3	38,59373	2,528876	97,47112
4	45,47172	54,90856	45,09144	4	38,80374	2,592156	97,40784
5	45,54058	54,76783	45,23217	5	38,85958	2,604496	97,39550
6	45,55892	54,72930	45,27070	6	38,87409	2,608129	97,39187
7	45,56369	54,71940	45,28060	7	38,87790	2,609036	97,39096
8	45,56494	54,71679	45,28321	8	38,87889	2,609278	97,39072
9	45,56527	54,71611	45,28389	9	38,87916	2,609341	97,39066
10	45,56536	54,71593	45,28407	10	38,87922	2,609358	97,39064
Ordre : Y2 Y1 (Y2 est la cause de Y1)							

La décomposition de la variance indique que la variance de l'erreur de prévision de $y_{1,t}$ est due à 55 % à ses propres innovations et à 45 % à celles de $y_{2,t}$. La variance de l'erreur de prévision de $y_{2,t}$ est due à 2,6 % à $y_{1,t}$ et à 97,4 % à $y_{2,t}$. Cette dissymétrie d'effet va être étudiée par l'analyse de la causalité.

Enfin, nous remarquons que ce tableau nous indique les écarts types de l'erreur de prévision pour $y_{1,t}$ et $y_{2,t}$ que nous avons calculés de manière fastidieuse lors de l'exercice 2.

IV. La causalité

Au niveau théorique, la mise en évidence de relations causales entre les variables économiques fournit des éléments de réflexion propices à une meilleure compréhension des phénomènes économiques. De manière pratique, « *the causal knowledge* » est nécessaire à une formulation correcte de la politique économique. En effet, connaître le sens de la causalité est aussi important que de mettre en évidence une liaison entre des variable économiques.

A. Causalité au sens de Granger

Granger (1969) a proposé les concepts de causalité et d'exogénéité : la variable y_{2t} est la cause de y_{1t} , si la prédictibilité¹ de y_{1t} est améliorée lorsque l'information relative à y_{2t} est incorporée dans l'analyse. Soit le modèle VAR(p) pour lequel les variables y_{1t} et y_{2t} sont stationnaires :

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} + \begin{bmatrix} a_1^1 & b_1^1 \\ a_1^2 & b_1^2 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} a_2^1 & b_2^1 \\ a_2^2 & b_2^2 \end{bmatrix} \begin{bmatrix} y_{1t-2} \\ y_{2t-2} \end{bmatrix} + \dots \\ + \begin{bmatrix} a_p^1 & b_p^1 \\ a_p^2 & b_p^2 \end{bmatrix} \begin{bmatrix} y_{1t-p} \\ y_{2t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

Le bloc de variables ($y_{2t-1}, y_{2t-2}, \dots, y_{2t-p}$) est considéré comme exogène par rapport au bloc de variables ($y_{1t-1}, y_{1t-2}, \dots, y_{1t-p}$) si le fait de rajouter le bloc y_{2t} n'améliore pas significativement la détermination des variables y_{1t} . Ceci consiste à effectuer un test de restrictions sur les coefficients des variables y_{2t} de la représentation VAR (noté RVAR = Restricted VAR). La détermination du retard p est effectuée par les critères AIC ou SC (cf. II.B). Soit :

- y_{2t} ne cause pas y_{1t} si l'hypothèse suivante est acceptée H_0 :
 $b_1^1 = b_2^1 = \dots = b_p^1 = 0$.
- y_{1t} ne cause pas y_{2t} si l'hypothèse suivante est acceptée H_0 :
 $a_1^2 = a_2^2 = \dots = a_p^2 = 0$.

Si nous sommes amenés à accepter les deux hypothèses que y_{1t} cause y_{2t} et que y_{2t} cause y_{1t} , on parle de boucle rétroactive « *feedback effect* ».

1. Le terme de prédictibilité semble préférable au terme de causalité ; en effet, dire que y_t cause x_t , signifie seulement qu'il est préférable de prédire x_t en connaissant y_t que sans le connaître.

Ces tests peuvent être conduits à l'aide d'un test de Fisher classique (*cf.* chapitre 3, test de Wald) de nullité des coefficients, équation par équation ou bien directement par comparaison entre un modèle VAR non contraint (UVAR) et le modèle VAR contraint (RVAR). On calcule le ratio de vraisemblance suivant :

$L^* = (n - c) \times (\text{Ln}|\sum_{\text{RVAR}}| - \text{Ln}|\sum_{\text{UVAR}}|)$ qui suit un χ^2 à $2 \times p$ degrés de liberté avec :

\sum_{RVAR} = matrice des variances covariances des résidus du modèle contraint,

\sum_{UVAR} = matrice des variances covariances des résidus du modèle non contraint,

n = nombre d'observations,

c = nombre de paramètres estimés dans chaque équation du modèle non contraint.

Si $L^* > \chi^2$ lu dans la table, alors on rejette l'hypothèse de validité de la contrainte.

B. Causalité au sens de Sims

Sims (1980) présente une spécification de test légèrement différente, en considérant que si les valeurs futures de y_{1t} permettent d'expliquer les valeurs présentes de y_{2t} , alors y_{2t} est la cause de y_{1t} .

Ceci se traduit par la représentation suivante :

$$y_{1t} = a_1^0 + \sum_{i=1}^p a_{1i}^1 y_{1t-i} + \sum_{i=1}^p a_{1i}^2 y_{2t-i} + \sum_{i=1}^p b_i^2 y_{2t+i} + \varepsilon_{1t}$$

$$y_{2t} = a_2^0 + \sum_{i=1}^p a_{2i}^1 y_{1t-i} + \sum_{i=1}^p a_{2i}^2 y_{2t-i} + \sum_{i=1}^p b_i^1 y_{1t+i} + \varepsilon_{2t}$$

- y_{1t} ne cause pas y_{2t} si l'hypothèse suivante est acceptée H_0 :
 $b_1^2 = b_2^2 = \dots = b_p^2 = 0$.
- y_{2t} ne cause pas y_{1t} si l'hypothèse suivante est acceptée H_0 :
 $b_1^1 = b_2^1 = \dots = b_p^1 = 0$.

Il s'agit là encore d'un test de Fisher classique de nullité de coefficients.

Exercice n° 4

 **fichier C10EX2**

Tests de causalité de Granger et de Sims

On demande, à partir de la représentation VAR(1) estimée lors de l'exercice 2, de procéder aux tests de Granger et de Sims.

Solution

Nous procédons au test de Fisher équation par équation.

– *Test de Granger*

• $H_0 : y_{2t}$ ne cause pas y_{1t}

Nous estimons les modèles suivants :

$$y_{1t} = 0,006760 * y_{1t-1} - 0,6125 * y_{2t-1} + 17,129632 + e_{1t}$$

$$R^2 = 0,28 ; n = 71 ; SCRU = 106500,2 \text{ (non contraint)}$$

$$y_{1t} = 0,2359 * y_{1t-1} + 26,9448 + e_{1t}$$

$$R^2 = 0,067 ; n = 71 ; SCRR = 13792,6 \text{ (contraint)}$$

$$F^* = \frac{(SCRR - SCRU)/c}{SCRU/(n - k - 1)} = \frac{13792,6 - 106500,2)/1}{106500,2/(71 - 2 - 1)} = 20,064$$

c = nombre de restrictions (nombre de coefficients dont on teste la nullité).

$F^* > F_{1;68}^{0,05} \approx 4$, on rejette l'hypothèse H_0 , y_{2t} explique significativement la variable y_{1t} , il y a donc causalité au sens de Granger de y_{2t} vers y_{1t} .

Ou bien :

$$L^* = (n - c)(\ln|\sum_{RVAR}| - \ln|\sum_{UVAR}|) = (71 - 1)(14,44 - 14,11) = 23,29$$

$> \chi^2(2) = 5,99$, on rejette l'hypothèse de validité de la contrainte.

• $H_0 : y_{1t}$ ne cause pas y_{2t}

$$y_{2t} = -0,17527 * y_{1t-1} + 0,29926096 * y_{2t-1} - 12,862837 + e_{2t}$$

$$R^2 = 0,21 ; n = 71 ; SCRU = 85205,25 \text{ (non contraint)}$$

$$y_{2t} = 0,4075 * y_{2t-1} - 15,391 + e_{2t}$$

$$R^2 = 0,168 ; n = 71 ; SCRR = 89452,77 \text{ (contraint)}$$

$$F^* = \frac{(SCRR - SCRU)/c}{SCRU/(n - k - 1)} = \frac{(89452,77 - 85205,25)/1}{85205,22/(71 - 2 - 1)} = 3,389$$

$F^* < F_{1;68}^{0,05} \approx 4$, on accepte l'hypothèse H_0 , y_{1t} n'explique pas significativement la variable y_{2t} , il n'y a donc pas causalité au sens de Granger de y_{1t} vers y_{2t} .

Ou bien :

$$L^* = (n - c)(\ln|\sum_{RVAR}| - \ln|\sum_{UVAR}|) = (71 - 1)(14,17 - 14,11) = 4,52$$

$< \chi^2(2) = 5,99$, on accepte l'hypothèse de validité de la contrainte.

Ce résultat ne doit pas nous surprendre puisque y_{1t} est la variable représentative de la demande et y_{2t} du prix.

– *Test de Sims*

• $H_0 : y_{1t}$ ne cause pas y_{2t}

Nous estimons les modèles suivants (sur la même période de 78:2 à 95:3) :

$$y_{1t} = -0,01435 * y_{1t-1} - 0,579 * y_{2t-1} - 0,357 * y_{2t+1} + 8,713 + e_{1t}$$

$$R^2 = 0,37 ; n = 70 ; SCRU = 92676,66 \text{ (non contraint)}$$

$$y_{1t} = 0,007654 * y_{1t-1} - 0,6145 * y_{2t-1} + 16,714 + e_{1t}$$

$$R^2 = 0,28 ; n = 70 ; SCRR = 105952,6 \text{ (contraint)}$$

$$F^* = \frac{(105952,6 - 92676,66)/1}{92676,66/(70 - 3 - 1)} = 9,45$$

$F^* > F_{1;66}^{0,05} \approx 4$, on rejette l'hypothèse H_0 , y_{2t+1} explique significativement la variable y_{1t} , il y a donc causalité au sens de Sims de y_{1t} vers y_{2t} .

- H_0 : y_{2t} ne cause pas y_{1t}

$$y_{2t} = -0,0967 * y_{1t-1} + 0,2259 * y_{2t-1} - 0,390 * y_{1t+1} - 3,328 + e_{2t}$$

$$R^2 = 0,40 ; n = 70 ; SCRU = 63742,31 \text{ (non contraint)}$$

$$y_{2t} = -0,174 * y_{1t-1} + 0,297 * y_{2t-1} - 13,30 + e_{2t}$$

$$R^2 = 0,21 ; n = 71 ; SCRR = 84586,73 \text{ (contraint)}$$

$$F^* = \frac{(84586,73 - 63742,31)/1}{84586,73/(70 - 3 - 1)} = 21,58$$

$F^* > F_{1;66}^{0,05} \approx 4$, on refuse l'hypothèse H_0 , y_{1t+1} explique significativement la variable y_{2t} , il y a donc causalité au sens de Sims de y_{2t} vers y_{1t} .

Nous observons une divergence dans ces tests, en ce qui concerne la causalité de y_{1t} sur y_{2t} ; le test de Granger nous fait rejeter l'hypothèse de causalité et celui de Sims laisse présager d'un lien de causalité de type *feedback* : les prix de la période suivante influencent la demande de la période actuelle, s'agit-il d'une anticipation des agents ?

11. La cointégration et le modèle à correction d'erreur

L'analyse de la cointégration, présentée par Granger (1983) et Engle et Granger (1987), est considérée par beaucoup d'économistes comme un des concepts nouveaux les plus importants dans le domaine de l'économétrie et de l'analyse de séries temporelles. Dans ce chapitre, nous abordons en section I. des exemples permettant une introduction au problème de la cointégration. La section II. est consacrée à la présentation de la cointégration en termes de définition et de spécification. Les tests de cointégration ainsi que les méthodologies d'estimation sont ensuite développés dans le cas à deux variables (section III.) et enfin (section IV) dans les cas à k variables.

I. Exemples introductifs

– Premier exemple

Soit les variables y_t et x_t définies de la manière suivante :

$$\begin{aligned}y_1 &= 1 \text{ avec comme variable explicative : } x_1 = 1 \\y_2 &= 2 \text{ avec comme variable explicative : } x_2 = 2^2 \\&\dots \\y_n &= n \text{ avec comme variable explicative : } x_n = n^2\end{aligned}$$

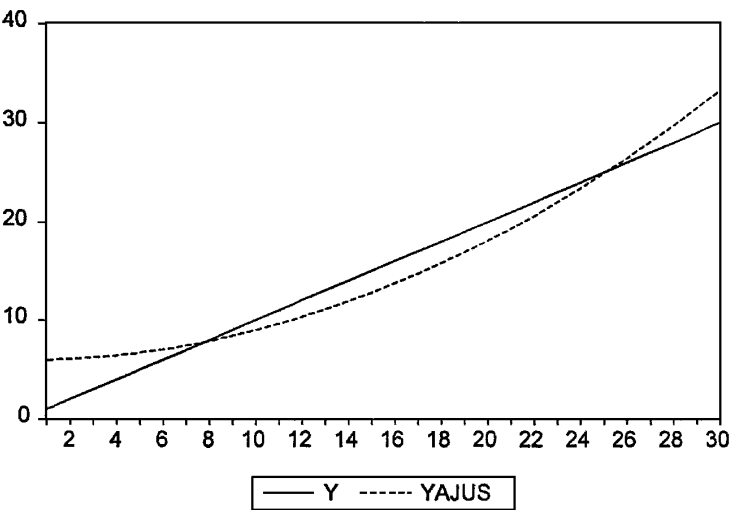
La tendance de y_t est de type linéaire et celle de x_t est quadratique.

Le résultat de la régression de y_t sur x_t ($n = 30$) conduit à l'estimation suivante :

$$\begin{aligned}y_t &= 5,92 + 0,03x_t \\(8,5) \quad (19,8) \\R^2 &= 0,94 - DW = 0,057\end{aligned}$$

Les coefficients de régression sont significatifs, la valeur du R^2 est élevée, cependant il apparaît évident que ce modèle a un pouvoir prédictif très mauvais. En effet, sur le plan statistique, la statistique de Durbin et Watson (proche de 0) présage d'une autocorrélation forte des erreurs¹.

Ce premier exemple illustre le danger d'interpréter et d'utiliser une régression entre deux variables affectées de tendances déterministes de degré différent. Le graphique 1 présente la série brute (tendance linéaire) et la série ajustée par une tendance quadratique.



Graphique 1 – Régression d'une tendance linéaire sur une tendance quadratique

– Deuxième exemple

On génère deux processus aléatoires :

$$y_t = y_{t-1} + \varepsilon_{1t} \quad \text{avec} \quad \varepsilon_t \rightarrow N(0; \sigma_{\varepsilon_1}^2)$$

$$x_t = x_{t-1} + \varepsilon_{2t} \quad \text{avec} \quad \varepsilon_t \rightarrow N(0; \sigma_{\varepsilon_2}^2)$$

Sur 1 000 régressions, nous obtenons les résultats suivants : 670 sont significatives d'après la statistique du t de Student, cependant la statistique de DW est toujours faible (la moyenne sur les 1 000 régressions est : $\overline{DW} = 0,33$). Donc apparemment, ces régressions donnent de bons résultats (hormis la statistique DW). Mais ces résultats sont purement fortuits : ils découlent de la corrélation entre les séries qui sont affectées d'une tendance stochastique. En effet, si on passe en différences premières ($\Delta y_t = f(\Delta x_t)$), plus aucune régression n'est significative.

$$\Delta y_t = y_t - y_{t-1} = \varepsilon_{1t}$$

$$\Delta x_t = x_t - x_{t-1} = \varepsilon_{2t}$$

1. Cela souligne l'importance à ne pas oublier d'interpréter la statistique de Durbin et Watson.

En régressant une série non stationnaire (de type DS) sur une autre série du même type, on obtient des coefficients significatifs mais avec une statistique DW proche de 0. Ce deuxième exemple illustre le risque de régresser entre elles deux séries affectées d'une tendance stochastique. Il faut donc toujours, au préalable, stationnariser des séries non stationnaires ; dans le cas contraire, il existe un risque de « régression fallacieuse » (« *spurious regression* »).

II. Le concept de cointégration

L'analyse de la cointégration permet d'identifier clairement la relation véritable entre deux variables en recherchant l'existence d'un vecteur de cointégration et en éliminant son effet, le cas échéant.

A. Propriétés de l'ordre d'intégration d'une série

Une série est intégrée d'ordre d (notée $x_t \rightarrow I(d)$), s'il convient de la différencier d fois afin de la stationnariser¹.

Soit une série x_{1t} stationnaire et une série x_{2t} intégrée d'ordre 1 :

$$\begin{aligned} x_{1t} &\rightarrow I(0) \\ x_{2t} &\rightarrow I(1) \end{aligned} \Rightarrow x_{1t} + x_{2t} \rightarrow I(1)$$

La série $y_t = x_{1t} + x_{2t}$ est non stationnaire puisque l'on somme une série affectée d'une tendance et une série stationnaire.

Soit deux séries x_{1t} et x_{2t} intégrées d'ordre d :

$$\begin{aligned} x_{1t} &\rightarrow I(d) \\ x_{2t} &\rightarrow I(d) \end{aligned} \Rightarrow x_{1t} + x_{2t} \rightarrow I(?)$$

La combinaison linéaire $\alpha x_{1t} + \beta x_{2t} \rightarrow I(?)$.

En effet, le résultat dépend des signes des coefficients α , β et de l'existence d'une dynamique non stationnaire commune.

Examinons un autre cas de figure :

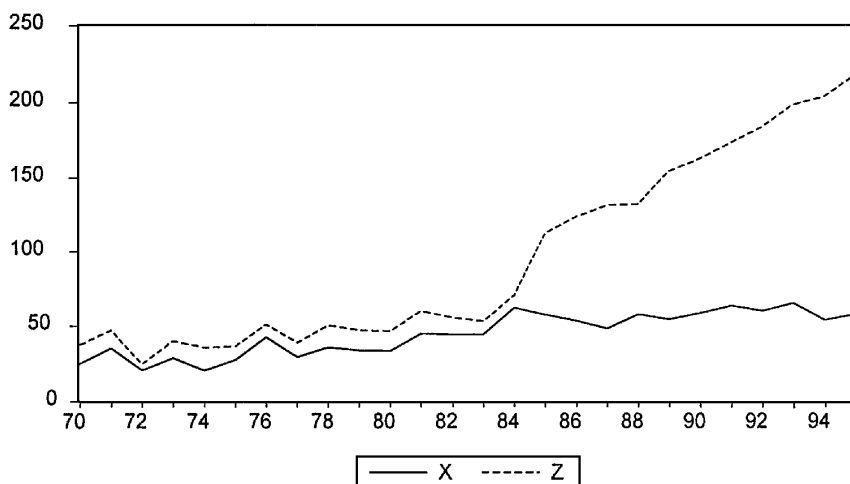
$$\begin{aligned} x_{1t} &\rightarrow I(d) \\ x_{2t} &\rightarrow I(d') \quad d' \neq d \end{aligned} \Rightarrow x_{1t} + x_{2t} \rightarrow I(?)$$

Il s'avère impossible de conclure car on somme deux séries d'ordre d'intégration différent.

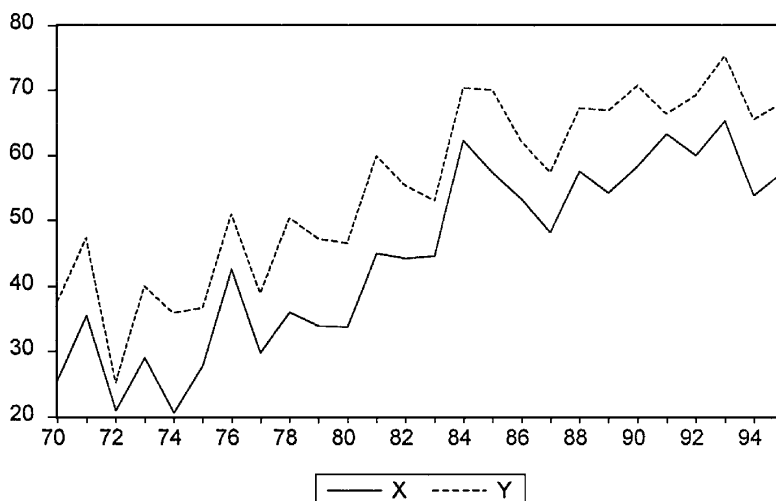
1. Cf. le test de Dickey-Fuller du chapitre 9 afin de déterminer l'existence d'une tendance stochastique ou déterministe et de rechercher les ordres d'intégration d'une série.

Soit deux séries affectées d'une tendance :

- dans un premier cas, les deux séries ont une tendance d'évolution constante sur une première période et puis une tendance d'évolution divergente en une deuxième période, les séries ne sont pas cointégrées (cf. graphique 2) ;
- dans un deuxième cas, les deux séries ont une évolution constante sur toute la période, les séries sont cointégrées (cf. graphique 3), il existe une évolution de long terme constante entre les chroniques.



Graphique 2 – Les variables x_t et z_t ne sont pas cointégrées



Graphique 3 – Les variables x_t et y_t sont cointégrées

B. Conditions de cointégration

Deux séries x_t et y_t sont dites cointégrées si les deux conditions sont vérifiées :

- elles sont affectées d'une tendance stochastique de même ordre d'intégration d ,
- une combinaison linéaire de ces séries permet de se ramener à une série d'ordre d'intégration inférieur.

Soit :

$$\begin{aligned}x_t &\rightarrow I(d) \\ y_t &\rightarrow I(d)\end{aligned}$$

tel que $\alpha_1 x_t + \alpha_2 y_t \rightarrow I(d - b)$,
avec $d \geq b > 0$.

On note : $x_t, y_t \rightarrow CI(d, b)$

où $[\alpha_1 \alpha_2]$ est le vecteur de cointégration.

Dans le cas général à k variables, on a :

$$\begin{aligned}x_{1,t} &\rightarrow I(d) \\ x_{2,t} &\rightarrow I(d) \\ &\dots \\ x_{k,t} &\rightarrow I(d)\end{aligned} \quad \text{on note } X_t = [x_{1,t} \ x_{2,t} \ \dots \ x_{k,t}]$$

s'il existe un vecteur de cointégration $\alpha = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_k]$ de dimension $(k, 1)$ tel que $\alpha X_t \rightarrow I(d - b)$, alors les k variables sont cointégrées et le vecteur de cointégration est α . On note que $X_t \rightarrow CI(d, b)$ avec $b > 0$.

C. Le modèle à correction d'erreur (ECM)

Examinons le cas particulier suivant : $x_t, y_t \rightarrow CI(1, 1)$ et $[\beta, -1]$ (on normalise en posant $\beta = -\frac{\alpha_1}{\alpha_2}$ soit $\beta x_t - y_t \rightarrow I(0)$) est le vecteur de cointégration.

Dans ce type de spécification, le fait que les séries soient cointégrées et non stationnaires soulève un problème d'estimation. La bonne qualité statistique du modèle (R^2 élevé et coefficients significatifs) est due au fait que les séries sont non stationnaires (puisque cointégrées). En régression directe de y_t sur x_t – lorsque $y_t, x_t \rightarrow CI(1, 1)$ – l'utilisation de ce modèle à des fins prévisionnelles s'avère désastreuse ; en effet, la relation mise en évidence par cette régression n'est pas réelle, elle découle simplement d'une relation entre deux tendances¹.

Le problème est donc, d'une part de retirer la relation commune de cointégration (la tendance commune), d'autre part, de rechercher la liaison réelle entre

1. Cf. les exemples introductifs sur les problèmes de régressions fallacieuses.

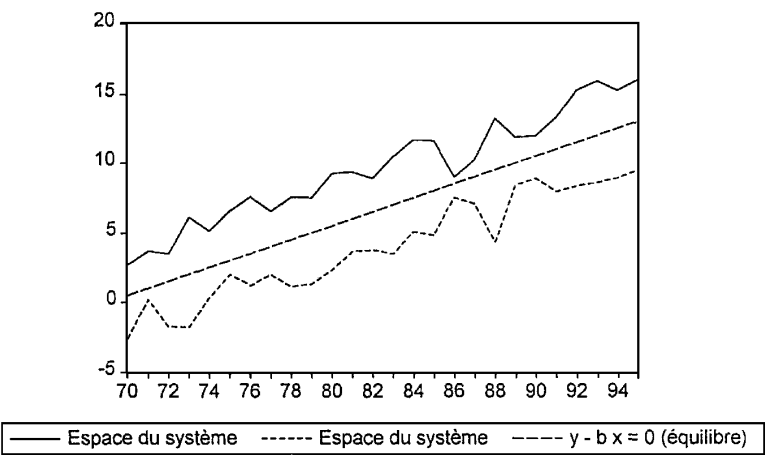
les variables : c'est l'objectif du Modèle à Correction d'Erreur (ECM¹). Cette représentation est à la fois un modèle statique ($\beta_1 \Delta x_t$) et un modèle dynamique ($\beta_2(y_{t-1} - \beta x_{t-1})$).

Nous pouvons spécifier la relation :

$$\Delta y_t = \beta_1 \Delta x_t + \beta_2 (y_{t-1} - \beta x_{t-1})$$

Autour de la relation de long terme, le modèle à correction d'erreur permet d'intégrer les fluctuations de court terme. Le coefficient β_2 – qui doit être négatif – rend compte d'une force de rappel vers l'équilibre de long terme.

Le graphique 4 illustre la relation de long terme entre x_t et y_t ($y_t - \beta x_t = 0$), cette ligne représente l'équilibre de long terme du système ; la zone d'évolution du système en dehors de l'équilibre (dynamique de court terme) est définie par l'erreur constatée entre y_{t-1} et x_{t-1} .



Graphique 4 – Représentation du modèle à correction d'erreur

III. Cointégration entre deux variables

Le traitement de séries chronologiques longues impose de tester une éventuelle cointégration entre les variables. En effet, le risque d'estimer des relations « fallacieuses » et d'interpréter les résultats de manière erronée est très élevé. Nous allons tout d'abord examiner le cas à deux variables : test de cointégration et estimation du modèle à correction d'erreur.

1. Nous utilisons l'abréviation anglo-saxonne « *Error Correction Model* ».

A. Test de cointégration entre deux variables

Nous reprenons ici l'algorithme en deux étapes de Engle et Granger.

– Étape 1 : tester l'ordre d'intégration des variables

Une condition nécessaire de cointégration est que les séries doivent être intégrées de même ordre. Si les séries ne sont pas intégrées de même ordre, elles ne peuvent pas être cointégrées.

Il convient donc (*cf.* tests de Dickey-Fuller et Dickey-Fuller Augmenté) de déterminer très soigneusement le type de tendance déterministe ou stochastique de chacune des variables, puis l'ordre d'intégration d des chroniques étudiées. Si les séries statistiques étudiées ne sont pas intégrées de même ordre, la procédure est arrêtée, il n'y a pas de risque de cointégration.

Soit :

$$x_t \rightarrow I(d) \text{ et } y_t \rightarrow I(d)$$

– Étape 2 : estimation de la relation de long terme

Si la condition nécessaire est vérifiée, on estime par les MCO la relation de long terme entre les variables : $y_t = a_1 x_t + a_0 + \varepsilon_t$.

Pour que la relation de cointégration soit acceptée, le résidu e_t issu de cette régression doit être stationnaire : $e_t = y_t - \hat{a}_1 x_t - \hat{a}_0$.

La stationnarité du résidu est testée à l'aide des tests DF ou DFA (*cf.* chapitre 9).

Dans ce cas, nous ne pouvons plus utiliser les tables de Dickey et Fuller. En effet, le test porte sur les résidus estimés à partir de la relation statique et non pas sur les « vrais » résidus de la relation de cointégration. MacKinnon (1991) a donc simulé des tables qui dépendent du nombre d'observations et du nombre de variables explicatives figurant dans la relation statistique.

Si le résidu est stationnaire nous pouvons alors estimer le modèle à correction d'erreur.

B. Estimation du modèle à correction d'erreur

Lorsque des séries sont non stationnaires et cointégrées, il convient d'estimer leurs relations au travers d'un modèle à correction d'erreur (ECM, « *Error Correction Model* »). Engle et Granger (1987) ont démontré que toutes les séries cointégrées peuvent être représentées par un ECM (théorème de la représentation de Granger).

Nous allons présenter la méthode la plus répandue d'estimation d'un ECM. Nous nous bornons, ici, au cas d'une seule variable explicative. La généralisation à k variables est abordée à la section suivante.

Soit les séries y_t et $x_t \rightarrow I(1)$, l'estimation par les MCO de la relation de long terme indique une stationnarité du résidu. Les séries y_t et x_t sont donc notées : $CI(1, 1)$.

Nous pouvons, dès lors, estimer le modèle à correction d'erreur.

Estimation du modèle à correction d'erreur (ECM) en deux étapes

- **Étape 1** : estimation par les MCO de la relation de long terme :

$$y_t = \hat{\alpha} + \hat{\beta} x_t + e_t \quad (\text{ECM}) \quad [1]$$

- **Étape 2** : estimation par les MCO de la relation du modèle dynamique (court terme) :

$$\Delta y_t = \alpha_1 \Delta x_t + \alpha_2 e_{t-1} + u_t \quad \alpha_2 < 0 \quad [2]$$

Le coefficient α_2 (force de rappel vers l'équilibre) doit être significativement négatif ; dans le cas contraire, il convient de rejeter une spécification de type ECM. En effet, le mécanisme de correction d'erreur (rattrapage qui permet de tendre vers la relation de long terme) irait alors en sens contraire et s'éloignerait de la cible de long terme. La procédure en deux étapes conduit à une estimation convergente des coefficients du modèle et les écarts types des coefficients peuvent s'interpréter de manière classique¹.

L'estimation d'un modèle à correction d'erreur ne pose donc pas de problème particulier et nécessite simplement le recours à une technique de moindres carrées ordinaires.

En cas de variables explicatives supplémentaires dont l'hypothèse de cointégration ne peut pas être retenue (comme des variables indicatrices par exemple), la procédure d'estimation est la même sauf pour l'estimation du modèle [1] (équation de long terme) dans lequel figurent ces variables explicatives.

Exercice n° 1

↓ fichier C11EX1

Test de cointégration et estimation d'un modèle à correction d'erreur

Soit deux séries statistiques y_t et x_t dont les observations sont présentées au tableau 1. On demande d'estimer la relation entre ces deux variables ($y_t = \hat{\alpha}_0 + \hat{\alpha}_1 x_t + e_t$) en testant une éventuelle cointégration (dans ce cas estimer le modèle à correction d'erreur).

1. Engle et Granger, 1987.

Tableau 1 – Valeurs observées de x_t et y_t (extrait des données)

Obs	x_t	y_t
1	0,000	10,890
2	2,851	12,188
...
29	6,744	12,347
30	7,270	15,194

Solution

La première étape consiste à étudier les propriétés des deux séries en termes de stationnarité. Le tableau 2 présente les résultats des tests :

- de Dickey-Fuller (aucun retard n'est significatif dans l'éventualité d'un test de Dickey-Fuller Augmenté),
- de Phillips-Perron (troncature = 2),
- et de KPSS (troncature = 2).

Ces tests sont présentés au chapitre 9.

Tableau 2 – Résultats des tests de Dickey-Fuller, Phillips-Perron et KPSS

Test	Type de Modèle	x_t en niveau – t statistique	y_t en niveau – t statistique
Test DF	[1]	0,81	0,17
Test DF	[2]	–0,05	–2,67
Test DF	[3]	–0,82	–3,05
Test PP	[1]	0,88	0,62
Test PP	[2]	–0,11	–2,60
Test PP	[3]	–0,80	–2,96
		x_t en niveau – LM statistique	y_t en niveau – LM statistique
Test KPSS	[2]	0,53	0,39
Test KPSS	[3]	0,22	0,20

La comparaison des $t_{\hat{\phi}_1}$ calculés aux t lus (Tables de MacKinnon) ou des LM statistiques aux valeurs critiques indique que les deux séries x_t et y_t sont non stationnaires en niveau¹. Des tests similaires sur les différentes premières de x_t et y_t indiquent qu'elles sont stationnaires, les deux séries étudiées sont $I(1)$, il existe donc un risque de cointégration.

Le test de cointégration est effectué à partir du résidu d'estimation du modèle :

$$y_t = a_1 x_t + a_0 + \varepsilon_t$$

Soit : $y_t = 0,55x_t + 10,38 + e_t$

(6,3) (41,46)

$n = 30$; $R^2 = 0,58$; $(.) = t$ de Student.

1. Nous rejetons l'hypothèse d'une tendance déterministe.

Nous pouvons vérifier que le résidu est bien stationnaire, il existe donc un risque de cointégration entre les deux variables.

	Test DF		Test PP ($l = 2$)	
e_t	Modèle [1]	Modèle [2]	Modèle [1]	Modèle [2]
$t_{\hat{\phi}_1}$	-5,39	-5,30	-5,42	-5,31

Nous procédons donc maintenant à l'estimation du modèle à correction d'erreur.

Nous calculons, d'abord, le résidu (provenant du modèle précédent) décalé d'une période, soit : $e_{t-1} = y_{t-1} - 0,55x_{t-1} - 10,38$

Puis nous estimons (par les MCO) le modèle :

$$\Delta y_t = \alpha_1 \Delta x_t + \alpha_2 e_{t-1} + u_t$$

$$\Delta y_t = 0,610 \Delta x_t - 1,02 e_{t-1} + \hat{u}_t$$

$$(3,09) \qquad (5,22)$$

$$n = 29 ; \quad R^2 = 0,60 ; \quad (.) = t \text{ de Student.}$$

Le coefficient (terme de rappel) de e_{t-1} est bien significativement négatif, la représentation à correction d'erreur est validée.

IV. Généralisation à k variables

La littérature économique récente concernant la théorie de l'équilibre démontre qu'une série macro-économique stationnaire peut être le résultat d'une combinaison de variables non stationnaires, d'où l'importance actuelle de l'analyse de la cointégration à k variables. La généralisation de deux à k variables s'avère assez complexe du fait du nombre de possibilités de vecteurs de cointégration possibles.

A. La cointégration entre k variables

Dans un modèle économétrique à k variables explicatives :

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_k x_{kt} + \varepsilon_t$$

si les variables (y_t et x_{kt}) sont non stationnaires, $I(1)$ par exemple, il existe alors un risque de cointégration. En effet, l'existence d'une éventuelle cointégration implique que les variables doivent être non stationnaires¹. Comme dans le cas à

1. Généralement, en économie, les variables sont $I(1)$; cependant certaines séries financières peuvent être de type $I(2)$.

deux variables, si une combinaison linéaire de ces variables est stationnaire, elles sont alors cointégrées. L'estimation par les MCO du modèle permet de calculer le résidu :

$$e_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{1t} - \dots - \hat{\beta}_k x_{kt}$$

Si ce résidu est stationnaire, nous acceptons l'hypothèse d'une cointégration entre les variables. Les tests de Dickey-Fuller de stationnarité sur le résidu doivent s'effectuer à partir des valeurs critiques tabulées par MacKinnon (1991) en fonction du nombre de variables total du modèle. Le vecteur de cointégration est donné par : $[1, -\hat{\beta}_0, -\hat{\beta}_1, \dots, -\hat{\beta}_k]$.

Cependant, le cas multivarié est plus complexe que le cas à deux variables du fait des possibilités combinatoires de cointégration. En effet si, par exemple, les variables y_t, x_{1t}, x_{2t} et x_{3t} sont cointégrées par paires $y_t, x_{1t} \rightarrow CI(1, 1)$ et $x_{2t}, x_{3t} \rightarrow CI(1, 1)$, leur combinaison linéaire est $I(0)$.

Nous avons : $e_t^1 = y_t - \hat{\alpha}_0 - \hat{\alpha}_1 x_{1t}$ et $e_t^2 = x_{2t} - \hat{\gamma}_0 - \hat{\gamma}_1 x_{3t}$ et par conséquent :

$$e_t = e_t^1 + e_t^2 = y_t - \hat{\alpha}_0 - \hat{\alpha}_1 x_{1t} + x_{2t} - \hat{\gamma}_0 - \hat{\gamma}_1 x_{3t} \rightarrow I(0)$$

Nous obtenons, dans ce cas de figure, un autre vecteur de cointégration possible : $[1, -\hat{\alpha}_0 - \hat{\gamma}_0, -\hat{\alpha}_1, 1, -\hat{\gamma}_1]$. D'une manière générale, dans un modèle à une variable à expliquer et k variables explicatives (soit $k + 1$ variables au total), il peut exister k vecteurs de cointégration linéairement indépendants. Le nombre de vecteurs de cointégration linéairement indépendants est appelé le *rang de la cointégration*.

Si les variables sont de même ordre d'intégration – $I(1)$ par exemple – l'existence d'un seul vecteur de cointégration est possible ; en revanche, si les séries ne sont pas toutes intégrées du même ordre, nous pouvons être certains que le vecteur de cointégration n'est pas unique.

De manière pratique, pour tester une éventuelle cointégration entre plusieurs variables, il convient tout d'abord de la tester sur l'ensemble des $k + 1$ variables, puis – en cas de cointégration – de la tester par combinatoire entre les variables.

B. Estimation du modèle à correction d'erreur

La cointégration ayant été révélée par les tests précédents, deux cas de figure sont à envisager :

- il existe un vecteur unique de cointégration,
- plusieurs vecteurs de cointégration existent.

Si le vecteur de cointégration est unique, nous pouvons employer les méthodes d'estimation envisagées à la section précédente, par exemple, celle en deux étapes de Engle et Granger.

- **Étape 1** : estimation par les MCO de la relation de long terme et calcul du résidu :

$$e_t = y_t - \widehat{\beta}_0 - \widehat{\beta}_1 x_{1t} - \dots - \widehat{\beta}_k x_{kt}$$

- **Étape 2** : estimation par les MCO de la relation du modèle dynamique (court terme) :

$$\Delta y_t = \alpha_1 \Delta x_{1t} + \alpha_2 \Delta x_{2t} + \dots + \alpha_k \Delta x_{kt} + \gamma_1 e_{t-1} + u_t$$

Le coefficient γ_1 (force de rappel vers l'équilibre) doit être, comme dans le cas à une variable explicative, significativement négatif.

Cependant, le plus souvent, le vecteur de cointégration n'est pas unique et la méthode de Engle-Granger n'est plus valide. En effet, les estimateurs des MCO ne sont plus consistants quels que soient les vecteurs de cointégration. Nous devons, dans ce cas, faire appel à la représentation vectorielle à correction d'erreur (VECM, « *Vector Error Correction Model* »).

C. Le modèle à correction d'erreur vectoriel

Examinons au préalable le cas d'un processus VAR(2) à k variables sous forme matricielle :

$$Y_t = A_0 + A_1 Y_{t-1} + A_2 Y_{t-2} + \varepsilon$$

avec :

Y_t : vecteur de dimension $(k \times 1)$ constitué des k variables $(y_{1t}, y_{2t}, \dots, y_{kt})$,

A_0 : vecteur de dimension $(k \times 1)$,

A_i : matrice de dimension $(k \times k)$.

Ce modèle peut s'écrire en différence première :

$$\begin{aligned} Y_t - Y_{t-1} &= A_0 + A_1 Y_{t-1} - Y_{t-1} + A_2 Y_{t-2} + \varepsilon \\ \Delta Y_t &= A_0 + (A_1 - I) Y_{t-1} + A_2 Y_{t-2} + \varepsilon \end{aligned} \quad [3]$$

Afin de faire apparaître des différences premières à droite de l'équation, nous ajoutons et soustrayons $A_1 Y_{t-2} - Y_{t-2}$ de la manière suivante

$$\begin{aligned} \Delta Y_t &= A_0 + (A_1 - I) Y_{t-1} + A_1 Y_{t-2} - Y_{t-2} - A_1 Y_{t-2} + Y_{t-2} + A_2 Y_{t-2} + \varepsilon \\ \Delta Y_t &= A_0 + A_1 Y_{t-1} - Y_{t-1} + A_1 Y_{t-2} - Y_{t-2} - A_1 Y_{t-2} + Y_{t-2} + A_2 Y_{t-2} + \varepsilon \end{aligned} \quad [4]$$

Or $\Delta Y_{t-1} = Y_{t-1} - Y_{t-2}$.

En regroupant les termes en ΔY_{t-1} et après simplification, on obtient :

$$\Delta Y_t = A_0 + (A_1 - I) \Delta Y_{t-1} + (A_1 + A_2 - I) Y_{t-2} + \varepsilon \quad [5]$$

Cependant, afin de faire figurer les relations de cointégration – qui relient entre elles les variables en niveau décalées d’une période – nous cherchons à écrire l’équation [4] en fonction de Y_{t-1} , soit dans l’équation [3] nous pouvons ajouter et soustraire $A_2 Y_{t-1}$ et après simplification nous obtenons :

$$\Delta Y_t = A_0 - A_2 \Delta Y_{t-1} + (A_1 + A_2 - I) Y_{t-1} + \varepsilon \quad [6]$$

Ou encore :

$$\Delta Y_t = A_0 + B_1 \Delta Y_{t-1} + \pi Y_{t-1} + \varepsilon \quad [7]$$

avec : $A_2 = -B_1$ et $\pi = (A_1 + A_2 - I)$.

Ce résultat peut être généralisé à une représentation VAR(p) à k variables sous forme matricielle :

$$Y_t = A_0 + A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + \varepsilon$$

avec :

Y_t : vecteur de dimension $(k \times 1)$ constitué des k variables $(y_{1t}, y_{2t}, \dots, y_{kt})$,

A_0 : vecteur de dimension $(k \times 1)$,

A_i : matrice de dimension $(k \times k)$.

Ce modèle peut s’écrire en différences premières de deux manières :

$$\Delta Y_t = A_0 + (A_1 - I) \Delta Y_{t-1} + (A_2 + A_1 - I) \Delta Y_{t-2} + \dots + (A_{p-1} + \dots + A_2 + A_1 - I) \Delta Y_{t-p+1} + \pi Y_{t-p} + \varepsilon$$

ou encore en fonction de Y_{t-1} :

$$\Delta Y_t = A_0 + B_1 \Delta Y_{t-1} + B_2 \Delta Y_{t-2} + \dots + B_{p-1} \Delta Y_{t-p+1} + \pi Y_{t-1} + \varepsilon$$

Les matrices B_i étant des fonctions des matrices A_i et $\pi = \left(\sum_{i=1}^p A_i - I \right)$.

La matrice π peut s’écrire sous la forme $\pi = \alpha \beta'$ où le vecteur α est la force de rappel vers l’équilibre et β le vecteur dont les éléments sont les coefficients des relations de long terme des variables. Chaque combinaison linéaire représente donc une relation de cointégration.

Si tous les éléments de π sont nuls (le rang de la matrice π est égal à 0 et donc $A_{p-1} + \dots + A_2 + A_1 = I$), alors nous ne pouvons pas retenir une spécification à correction d’erreur, nous estimons un VAR classique en différences premières afin d’éliminer les tendances. Si le rang de π est égal à k , cela implique alors que les variables sont toutes $I(0)$ et le problème de la cointégration ne se pose donc pas, il convient d’estimer un modèle VAR en niveau.

Si le rang de la matrice π (noté r) est compris entre 1 et $k-1$ ($1 \leq r \leq k-1$), alors il existe r relations de cointégration et la représentation ECM est valide soit :

$$\Delta Y_t = A_0 + B_1 \Delta Y_{t-1} + B_2 \Delta Y_{t-2} + \dots + B_{p-1} \Delta Y_{t-p+1} + \alpha e_{t-1} + \varepsilon$$

avec $e_t = \beta' Y_t$.

Exemple de spécification : soit un processus VAR(2) à trois variables $y_{1,t}$, $y_{2,t}$ et $y_{3,t}$ et une relation de cointégration, le VECM s'écrit alors :

$$\begin{aligned}\Delta y_{1,t} &= a_0^1 + b_1^1 \Delta y_{1,t-1} + b_2^1 \Delta y_{2,t-1} + b_3^1 \Delta y_{3,t-1} + \alpha^1 (y_{1,t-1} - \beta_2 y_{2,t-1} - \beta_3 y_{3,t-1} - \beta_0) + \varepsilon_t^1 \\ \Delta y_{2,t} &= a_0^2 + b_1^2 \Delta y_{1,t-1} + b_2^2 \Delta y_{2,t-1} + b_3^2 \Delta y_{3,t-1} + \alpha^2 (y_{1,t-1} - \beta_2 y_{2,t-1} - \beta_3 y_{3,t-1} - \beta_0) + \varepsilon_t^2 \\ \Delta y_{3,t} &= a_0^3 + b_1^3 \Delta y_{1,t-1} + b_2^3 \Delta y_{2,t-1} + b_3^3 \Delta y_{3,t-1} + \alpha^3 (y_{1,t-1} - \beta_2 y_{2,t-1} - \beta_3 y_{3,t-1} - \beta_0) + \varepsilon_t^3\end{aligned}$$

Nous avons normalisé l'équation de cointégration par rapport au coefficient de $y_{1,t}$.

Selon les caractéristiques des données traitées, ce processus VAR peut connaître les variantes suivantes :

- existence d'une constante soit dans la relation de cointégration (β_0) soit dans le VAR (a_0^i),
- existence d'une tendance ($t = 1, 2, \dots, n$) dans le VAR et/ou dans la relation de cointégration,
- ou encore intégrer des variables exogènes de type indicatrice pour corriger d'un mouvement saisonnier.

Sur ce type de modèle, nous ne pouvons pas appliquer la méthode des MCO car nous avons des problèmes d'identification similaires à ceux évoqués lors du chapitre 8 concernant les modèles à équations simultanées. Il convient d'employer une méthode du maximum de vraisemblance¹.

D. Tests de relation de cointégration

Pour déterminer le nombre de relations de cointégration, Johansen (1988) propose deux tests fondés sur les valeurs propres d'une matrice issue d'un calcul en deux étapes :

Etape 1 : calcul de deux résidus u_t et v_t

Nous effectuons deux régressions :

Première régression :

$$\Delta Y_t = \hat{A}_0 + \hat{A}_1 \Delta Y_{t-1} + \hat{A}_2 \Delta Y_{t-2} + \dots + \hat{A}_p \Delta Y_{t-p} + u_t$$

Deuxième régression :

$$Y_{t-1} = \hat{A}'_0 + \hat{A}'_1 \Delta Y_{t-1} + \hat{A}'_2 \Delta Y_{t-2} + \dots + \hat{A}'_p \Delta Y_{t-p} + v_t$$

$$\text{Avec } Y_t = \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \dots \\ y_{k,t} \end{bmatrix}$$

1. Johansen, 1988.

Nous avons les mêmes variables explicatives, seule la spécification du bloc de la variable à expliquer est modifiée.

u_t et v_t sont donc les matrices des résidus de dimension (k, n) avec k = nombre de variables, n = nombre d'observations.

Etape 2 : calcul de la matrice permettant le calcul des valeurs propres

Nous calculons quatre matrices des variances-covariances de dimension (k, k) à partir des résidus u_t et v_t .

$$\widehat{\Sigma}_{uu} = (1/n) \sum_{t=1}^n u_t u_t'$$

$$\widehat{\Sigma}_{vv} = (1/n) \sum_{t=1}^n v_t v_t'$$

$$\widehat{\Sigma}_{uv} = (1/n) \sum_{t=1}^n u_t v_t'$$

$$\widehat{\Sigma}_{vu} = (1/n) \sum_{t=1}^n v_t u_t'$$

Puis nous extrayons les k valeurs propres de la matrice M de dimension k, k calculée de la manière suivante :

$$M = \widehat{\Sigma}_{vv}^{-1} \widehat{\Sigma}_{vu} \widehat{\Sigma}_{uu}^{-1} \widehat{\Sigma}_{uv}$$

1) Test de la trace

A partir de ces valeurs propres, on calcule une statistique :

$$\lambda_{\text{trace}} = -n \sum_{i=r+1}^k \text{Ln}(1 - \lambda_i) \text{ avec } n = \text{nombre d'observations, } \lambda_i = i^{\text{ème}} \text{ valeur}$$

propre de la matrice M , k = nombre de variables, r = rang de la matrice.

Cette statistique suit une loi de probabilité (similaire à un χ^2) tabulée à l'aide de simulations par Johansen et Juselius (1990). Ce test de Johansen fonctionne par exclusion d'hypothèses alternatives :

- rang de la matrice π égal 0 ($r = 0$), soit $H_0 : r = 0$ contre $H_1 : r > 0$;
si H_0 est refusé, on passe au test suivant¹ (si $\lambda_{\text{trace}} >$ à la valeur critique lue dans la table, on rejette H_0) ;
- rang de la matrice π égal 1 ($r = 1$), soit $H_0 : r = 1$ contre $H_1 : r > 1$;
si H_0 est refusé, on passe au test suivant ;
- rang de la matrice π égal 2 ($r = 2$), soit $H_0 : r = 2$ contre $H_1 : r > 2$;
si H_0 est refusé, on passe au test suivant, etc.

Si, après avoir refusé les différentes hypothèses H_0 à la fin de la procédure, on teste $H_0 : r = k - 1$ contre $H_1 : r = k$ et que l'on soit amené à refuser H_0 , alors le rang de la matrice est $r = k$ et il n'existe pas de relation de cointégration car les variables sont toutes $I(0)$.

Les logiciels d'économétrie fournissent directement l'ensemble de ces informations : la statistique λ_{trace} calculée entre 1 et k ainsi que les valeurs critiques associées.

1. Dans le cas contraire, la procédure est arrêtée, et le rang de la matrice est $r = 0$.

Pour mener ce test, Johansen propose cinq spécifications concernant soit les vecteurs cointégrants soit les séries (le VAR proprement dit) :

Absence de tendance linéaire dans les données (les séries sont toutes DS sans dérive) :

- a) Absence d'une tendance linéaire dans les séries et d'une constante dans les relations de cointégration (la constante dans la relation de long terme est non significative).

$$\text{Exemple : } \Delta y_{1,t-1} = b_1^1 \Delta y_{1,t-1} + b_2^1 \Delta y_{2,t-1} + b_3^1 \Delta y_{3,t-1} + \alpha^1 (y_{1,t-1} - \beta_2 y_{2,t-1} - \beta_3 y_{3,t-1}) + \varepsilon_t^1$$

- b) Absence d'une tendance linéaire dans les séries mais présence d'une constante dans les relations de cointégration (la constante dans la relation de long terme est significative).

$$\text{Exemple : } \Delta y_{1,t} = b_1^1 \Delta y_{1,t-1} + b_2^1 \Delta y_{2,t-1} + b_3^1 \Delta y_{3,t-1} + \alpha^1 (y_{1,t-1} - \beta_2 y_{2,t-1} - \beta_3 y_{3,t-1} - \beta_0) + \varepsilon_t^1$$

Présence d'une tendance linéaire dans les données (au moins une série est un DS avec dérive) :

- c) Présence d'une tendance linéaire dans les séries et d'une constante dans les relations de cointégration.

$$\text{Ex. : } \Delta y_{1,t} = a_0^1 + b_1^1 \Delta y_{1,t-1} + b_2^1 \Delta y_{2,t-1} + b_3^1 \Delta y_{3,t-1} + \alpha^1 (y_{1,t-1} - \beta_2 y_{2,t-1} - \beta_3 y_{3,t-1} - \beta_0) + \varepsilon_t^1$$

- d) Présence d'une tendance linéaire dans les séries et dans les relations de cointégration (au moins une série est un TS).

$$\text{Ex : } \Delta y_{1,t} = a_0^1 + b_1^1 \Delta y_{1,t-1} + b_2^1 \Delta y_{2,t-1} + b_3^1 \Delta y_{3,t-1} + \alpha^1 (y_{1,t-1} - \beta_2 y_{2,t-1} - \beta_3 y_{3,t-1} - \beta_0 + ct) + \varepsilon_t^1$$

Présence d'une tendance quadratique dans les données :

- e) Présence d'une tendance quadratique dans les séries et d'une tendance linéaire dans les relations de cointégration.

$$\text{Ex : } \Delta y_{1,t} = a_0^1 + bt + b_1^1 \Delta y_{1,t-1} + b_2^1 \Delta y_{2,t-1} + b_3^1 \Delta y_{3,t-1} + \alpha^1 (y_{1,t-1} - \beta_2 y_{2,t-1} - \beta_3 y_{3,t-1} - \beta_0 + ct) + \varepsilon_t^1$$

Le choix d'une de ces spécifications s'effectue en fonction des données et de la forme supposée de la tendance (une analyse des propriétés stochastiques des séries ou un simple examen visuel des graphiques permettent de se déterminer).

Le tableau 3 synthétise le choix de la spécification du VECM en fonction de la typologie des processus.

Tableau 3 – Choix de la spécification en fonction du type de processus

Type de processus	Spécification				
	1	2	3	4	5
Tous les processus sont des DS sans dérive	X	X			
Au moins un des processus est un DS avec dérive			X		
Au moins un des processus est un TS				X	
Au moins un processus a une tendance quadratique					X

2) Test de la valeur propre maximale

Le deuxième test proposé par Johansen est donné par la statistique :

$$\lambda_{\max} = -n \text{Log}(1 - \lambda_{r+1}) \quad r = 0, 1, 2, \dots$$

Le test s'effectue comme précédemment de manière séquentielle par exclusion d'hypothèses alternatives.

En cas de divergence des deux tests (valeur propre maximum et trace), nous privilégions le test de la trace dont la puissance est la plus élevée.

E. Test d'exogénéité faible

Le test d'exogénéité faible consiste à vérifier si les variables sont bien endogènes. Ce test porte sur le coefficient α de la force de rappel (s'il existe une seule relation de cointégration) ou les coefficients α dans le cas de plusieurs relations de cointégration. Effectuer un test sur α revient à vérifier si la relation de cointégration figure dans toutes les équations du modèle. C'est un test d'exogénéité faible des différentes variables du système, pour les paramètres d'intérêt donnés par la relation de long terme de la matrice $\pi = \alpha' \beta$, c'est-à-dire β les relations de cointégration et α les poids de ces relations dans chacune des équations du système.

Prenons par exemple un VECM à deux relations de cointégration :

$$\begin{pmatrix} \Delta y_{1t} \\ \Delta y_{2t} \\ \Delta y_{3t} \end{pmatrix} = A \begin{pmatrix} \Delta y_{1t-1} \\ \Delta y_{2t-1} \\ \Delta y_{3t-1} \end{pmatrix} + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{pmatrix} \begin{pmatrix} \beta_{11} & \beta_{21} & \beta_{31} \\ \beta_{12} & \beta_{22} & \beta_{32} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{pmatrix}$$

Pour tester l'exogénéité de la variable y_{2t} , il convient d'effectuer le test d'hypothèse $H_0 : \alpha_{21} = \alpha_{22} = 0$, si l'hypothèse H_0 est acceptée cela signifie que la force de rappel α n'intervient dans aucune relation de cointégration et donc que la variable y_{2t} est faiblement exogène. Dans ce cas nous estimons un VECM partiel en imposant la contrainte $\alpha_{21} = \alpha_{22} = 0$.

Ce test est réalisé par le calcul de la statistique du ratio de vraisemblance du modèle contraint sous H_0 et non contraint.

F. Synthèse de la procédure d'estimation

Nous essayons ici de synthétiser les grandes étapes relatives à l'estimation d'un modèle VECM.

Étape 1 : Détermination du nombre de retards p du modèle (en niveau ou en Log) selon les critères AIC ou SC (cf. chapitre 10).

Étape 2 : Estimation de la matrice Π et test de Johansen permettant de connaître le nombre de relations de cointégration (les logiciels proposent un certain nombre de spécifications alternatives, telles que l'existence d'un terme constant dans la relation de cointégration, contraindre $A_0 = 0$, l'existence d'une tendance déterministe, etc.).

Étape 3 : Identification des relations de cointégration, c'est-à-dire des relations de long terme entre les variables.

Étape 4 : Estimation par la méthode du maximum de vraisemblance du modèle vectoriel à correction d'erreur et validation à l'aide des tests usuels : significativité des coefficients¹ et vérification que les résidus sont des bruits blancs (test de Ljung-Box), tests d'exogénéité faible.

Enfin, nous pouvons vérifier que l'estimation par les MCO de la relation de long terme fournit des résultats à peu près similaires (en termes de significativité et de valeurs estimées des coefficients) à ceux obtenus par la méthode du maximum de vraisemblance².

Exercice n° 2

↓ fichier C11EX2

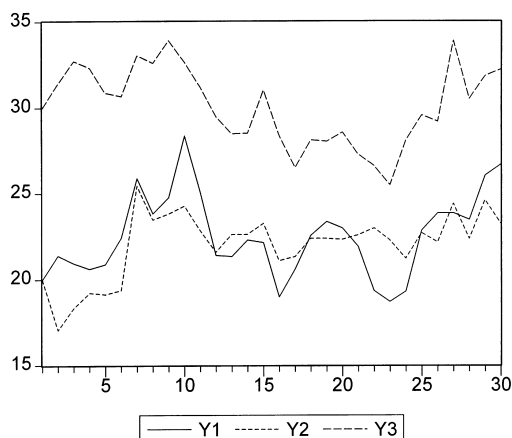
Tests de cointégration et estimation d'un modèle vectoriel à correction d'erreur

Soit trois variables y_{1t} , y_{2t} et y_{3t} observées sur 30 périodes dont les données sont consignées au tableau 3. On demande de tester une éventuelle cointégration et d'estimer un modèle VAR ou un modèle vectoriel à correction d'erreur s'il y a lieu.

Solution

– *Première étape* : détermination du nombre de retards de la représentation VAR en niveau (la transformation logarithmique ne s'impose pas au vue des données (cf. graphique 5))

1. Évidemment on tolère que quelques coefficients ne soient pas significativement de 0.
2. De manière asymptotique, les deux estimateurs sont équivalents



Graphique 5 – Représentation de y_{1t} , y_{2t} et y_{3t}

Tableau 4 – Données observées de y_{1t} , y_{2t} et y_{3t} (extrait des données)

OBS	$y_{1,t}$	$y_{2,t}$	$y_{3,t}$
1	20,00	20,00	30,00
2	21,39	17,05	31,40
...
29	26,04	24,61	31,84
30	26,70	23,25	32,22

Le calcul des critères d'information AIC et SC pour des retards allant de 1 à 3 – nous n'allons pas plus loin compte tenu du faible nombre d'observations – ne pose pas de difficulté.

$$\text{AIC}(1) = 2,45 ; \quad \text{SC}(1) = 2,88$$

$$\text{AIC}(2) = 1,93 ; \quad \text{SC}(2) = 2,78$$

$$\text{AIC}(3) = 2,04 ; \quad \text{SC}(3) = 3,34$$

Le retard retenu est donc de 2 (minimum des critères AIC et SC), nous allons donc procéder au test de Johansen sur un VECM(1).

– Deuxième étape : Test de Johansen

Nous procédons au test de Johansen sous deux hypothèses :

a) Existence d'une constante dans la relation de long terme et non dans les données (pas de tendance déterministe), spécification b.

Détaillons ce premier cas.

Premier test : Rang de la matrice π égal 0 ($r = 0$), soit $H_0 : r = 0$ contre $H_1 : r > 0$.

Les trois valeurs propres de la matrice π , estimée par le maximum de vraisemblance, sont égales à $\lambda_1 = 0,605$; $\lambda_2 = 0,257$; $\lambda_3 = 0,139$.

Calculons la statistique de Johansen : $\lambda_{\text{trace}} = -n \sum_{i=r+1}^k \text{Ln}(1 - \lambda_i)$ pour $r = 0$:

$$\lambda_{\text{trace}} = -n \times \{(\text{Ln}(1 - \lambda_1) + \text{Ln}(1 - \lambda_2) + \text{Ln}(1 - \lambda_3))\}$$

$$\lambda_{\text{trace}} = -28 \times \{0,928 + 0,297 + 0,150\} = 38,50$$

Les valeurs critiques sont égales à 34,91 pour un seuil de 5 % et 41,07 pour un seuil de 1 % ; on rejette donc l'hypothèse H_0 , le rang de la matrice n'est pas 0 (les séries ne sont donc pas stationnaires).

Rang de la matrice π égal 1 ($r = 1$), soit $H_0 : r = 1$ contre $H_1 : r > 1$.

$$\lambda_{\text{trace}} = -28 \times \{0,297 + 0,150\} = 12,51$$

Les valeurs critiques sont égales à 19,96 pour un seuil de 5 % et 24,60 pour un seuil de 1 % ; on ne peut pas rejeter l'hypothèse H_0 ni à 5 % ni à 1 %, on considère donc que le rang de la matrice π est égal à 1.

Nous acceptons donc l'hypothèse d'une relation de cointégration.

b) Existence d'une constante dans la relation de long terme et dans les données, spécification c.

Nous vérifions la robustesse de ce résultat sur cette autre spécification (compte tenu des données, les autres possibilités sont exclues).

Les résultats sont les suivants :

Valeur propre	λ_{trace}	Seuil critique à 5%	Seuil critique à 5%
0,603749	37,38888	29,68	35,65
0,229311	11,46904	15,41	20,04
0,138550	4,175870	3,76	6,65

Là encore, on constate que le rang de la matrice n'est pas 0 (ligne 1), mais qu'en revanche, on ne peut pas rejeter l'hypothèse H_0 ni à 5 % ni à 1 % (ligne 2) dans l'hypothèse d'un rang de la matrice π égal à 1.

Le test de la valeur propre maximale corrobore les résultats précédents :

Hypothesized		Max-Eigen	0.05	
No. of CE(s)	Eigenvalue	Statistic	Critical Value	Prob.**
None *	0.603749	25.91984	21.13162	0.0098
At most 1	0.229311	7.293174	14.26460	0.4550
At most 2 *	0.138550	4.175870	3.841466	0.0410

L'hypothèse d'une seule relation de cointégration est validée.

- *Troisième étape* : estimation du modèle vectoriel à correction d'erreur

Les deux spécifications précédentes sont estimées avec ou sans constante dans les données et donc, dans les deux cas, avec une seule relation de cointégration entre $y_{1,t}$, $y_{2,t}$ et $y_{3,t}$.

La première spécification est rejetée du fait que les trois constantes des trois équations ne sont pas significativement différentes de 0.

L'estimation finale du VECM sur 28 observations est donc la suivante :

$$\begin{aligned}
 \Delta y_{1,t} &= -0,85 \times (y_{1,t-1} - 0,87 \times y_{2,t-1} - 0,54 \times y_{3,t-1} + 12,92) \\
 &\quad (2,64) \quad (8,9) \quad (7,1) \quad (4,19) \\
 &\quad + 0,77 \times \Delta y_{1,t-1} - 0,71 \times \Delta y_{2,t-1} - 0,26 \times \Delta y_{3,t-1} \\
 &\quad (2,82) \quad (2,61) \quad (1,07) \\
 \Delta y_{2,t} &= 0,32 \times (y_{1,t-1} - 0,87 \times y_{2,t-1} - 0,54 \times y_{3,t-1} + 12,92) \\
 &\quad (1,18) \quad (8,9) \quad (7,1) \quad (4,19) \\
 &\quad + 0,10 \times \Delta y_{1,t-1} - 0,34 \times \Delta y_{2,t-1} - 0,08 \times \Delta y_{3,t-1} \\
 &\quad (0,43) \quad (1,44) \quad (0,39) \\
 \Delta y_{3,t} &= 0,0089 \times (y_{1,t-1} - 0,87 \times y_{2,t-1} - 0,54 \times y_{3,t-1} + 12,92) \\
 &\quad (0,03) \quad (8,9) \quad (7,1) \quad (4,19) \\
 &\quad + 0,53 \times \Delta y_{1,t-1} - 0,46 \times \Delta y_{2,t-1} - 0,25 \times \Delta y_{3,t-1} \\
 &\quad (2,16) \quad (1,86) \quad (1,13)
 \end{aligned}$$

Les chiffres entre parenthèses sont les t de Student.

– *Quatrième étape* : validation de la représentation vectorielle à correction d'erreur

Les coefficients des termes de rappel ont bien le signe attendu (cependant seul le terme de rappel de l'équation est significatif) et les coefficients de la relation de long terme sont significatifs.

Le fait que seul le coefficient α_1 soit significativement différent de 0, indique que les variables y_{2t} et y_{3t} sont faiblement exogènes ce qui est confirmé par le test de contrainte portant sur les coefficients de la force de rappel α_1 , α_2 et α_3 :

Cointegration Restrictions :		
A(1,1) = 0	A(2,1) = 0	A(3,1) = 0
Not all cointegrating vectors are identified		
LR test for binding restrictions (rank = 1) :		
Chi-square(1) : 5,557	Chi-square(1) : 1,275770	Chi-square(1) : 0,001005
Probability : 0,0184	Probability : 0,258687	Probability : 0,974714

Nous acceptons l'hypothèse H0 de nullité pour les coefficients α_2 et α_3 .

Le VECM partiel est estimé en imposant la contrainte $\alpha_2 = \alpha_3 = 0$.

Vector Error Correction Estimates			
Sample (adjusted) : 3 30			
Included observations : 28 after adjustments			
Standard errors in () & t-statistics in []			
Cointegration Restrictions :			
A(3,1)=0, A(2,1) = 0			
Not all cointegrating vectors are identified			
LR test for binding restrictions (rank = 1) :			
Chi-square(2)	1.558967		
Probability	0.458643		
Cointegrating Eq :	CointEq1		
Y1(- 1)	0.997283		
Y2(- 1)	- 0.797187		
Y3(- 1)	- 0.555922		
C	11.94414		
Error Correction :	D(Y1)	D(Y2)	D(Y3)
CointEq1	- 1.187476	0.000000	0.000000
	(0.20103)	(0.00000)	(0.00000)
	[- 5.90704]	[NA]	[NA]
D(Y1(- 1))	0.813973	0.139232	0.527616
	(0.26543)	(0.23701)	(0.24591)
	[3.06663]	[0.58745]	[2.14553]
D(Y2(- 1))	- 0.715308	- 0.372227	- 0.456821
	(0.26279)	(0.23465)	(0.24346)
	[- 2.72203]	[- 1.58631]	[- 1.87634]
D(Y3(- 1))	- 0.299747	- 0.097043	- 0.244642
	(0.23644)	(0.21113)	(0.21906)
	[- 1.26773]	[- 0.45964]	[- 1.11679]

Quelques coefficients du modèle ne sont pas significatifs, cela était probable et ne remet pas en cause la validité globale du modèle.

Enfin, les trois résidus issus de chaque équation sont des bruits blancs d'après la Q-statistique de Ljung-Box :

- première équation : $Q(12) = 5,87$ ($\alpha = 0,92$) \implies Acceptation de H_0 ;
- deuxième équation : $Q(12) = 4,60$ ($\alpha = 0,97$) \implies Acceptation de H_0 ;
- troisième équation : $Q(12) = 5,99$ ($\alpha = 0,92$) \implies Acceptation de H_0 ;

La spécification VECM est donc validée.

12. Introduction à l'économétrie des variables qualitatives

Nous avons, lors des chapitres précédents (chapitre 3, partie 5), utilisé des variables explicatives particulières appelées variables dichotomiques¹, dont la caractéristique est de prendre deux valeurs 0 ou 1. Leurs utilisations ne posent aucun problème en tant que variable explicative, en revanche lorsqu'elles sont utilisées en tant que variable à expliquer la méthode des moindres carrés ordinaires est défaillante.

Dans ce chapitre, qui constitue une simple introduction² à l'économétrie des variables qualitatives, nous allons donc aborder, dans une première partie, la problématique particulière de l'économétrie des variables qualitatives et les problèmes rencontrés dans ce type de modélisation. Puis nous présentons les modèles spécifiques lorsque la variable à expliquer est binaire tels que : décision d'acheter ou de ne pas acheter un produit, risque de défaillance de paiement, obtenir un diplôme, Il s'agit des modèles à choix binaires (Probit et Logit).

Puis, dans une troisième partie, nous abordons les modèles à choix multiples (pour quel candidat voter, quel moyen de transport adopter, ..). Enfin, la quatrième partie est consacrée aux variables tronquées et censurées.

1. Les termes de variables indicatrices, binaires, muettes, ou « *dummy* » sont aussi utilisés.

2. Pour des développements, nous recommandons au lecteur de se référer au livre de Thomas A. « *Économétrie des variables qualitatives* », Dunod, 2002.

I. Les problèmes et les conséquences de la spécification binaire

Le recours à des modèles et des méthodes d'estimation spécifiques lorsque la variable endogène est qualitative (dichotomique ou à choix multiples) est lié au fait que le modèle linéaire général ne peut pas s'appliquer dans ce contexte.

Prenons l'exemple d'un modèle de régression simple permettant de relier le fait d'être, pour un individu i , propriétaire d'un logement à son revenu :

$$y_i = a_0 + a_1 x_i + \varepsilon_i \quad i = 1, \dots, n \quad [1]$$

avec :

y_i = une variable endogène (variable à expliquer) qui prend les valeurs 1 si l'individu i est propriétaire de son logement et 0 dans le cas contraire,

x_i = une variable exogène (variable explicative) qui représente le revenu en euros de l'individu i ,

ε_i = l'erreur de spécification du modèle,

a_0 et a_1 = les paramètres à estimer.

Ce modèle, appelé aussi modèle à probabilité linéaire, présente les propriétés suivantes.

a) – En faisant l'hypothèse classique de l'espérance de l'erreur nulle :

$$E(\varepsilon_i) = 0, \text{ alors } E(y_i) = a_0 + a_1 x_i.$$

b) – La valeur prévue de la variable à expliquer y_i peut s'interpréter de la manière suivante :

Soit $P_i = \text{Prob}(y_i = 1)$, d'où $\text{Prob}(y_i = 0) = 1 - P_i$

$$E(y_i) = 1 \times \text{Prob}(y_i = 1) + 0 \times \text{Prob}(y_i = 0) = P_i$$

Soit : $P_i = a_0 + a_1 x_i \quad \forall i$, d'où le nom du modèle.

c) – La variable y_i ne pouvant prendre que deux valeurs (0 et 1), par voie de conséquence, l'erreur ne peut donc prendre que deux valeurs :

$$\varepsilon_i = 1 - (a_0 + a_1 x_i) \text{ avec la probabilité } P_i$$

$$\varepsilon_i = -(a_0 + a_1 x_i) \text{ avec la probabilité } 1 - P_i$$

La variance de l'erreur est donnée par :

$$\text{Var}(\varepsilon_i) = E(\varepsilon_i^2) = P_i \times (\text{valeur de } \varepsilon_i \text{ si } y_i = 1)^2 + (1 - P_i) \times (\text{valeur de } \varepsilon_i \text{ si } y_i = 0)^2.$$

$$\text{D'où : } \text{Var}(\varepsilon_i) = E(\varepsilon_i^2) = P_i \times (1 - a_0 - a_1 x_i)^2 + (1 - P_i) \times (-a_0 - a_1 x_i)^2$$

$$\text{Or } P_i = a_0 + a_1 x_i$$

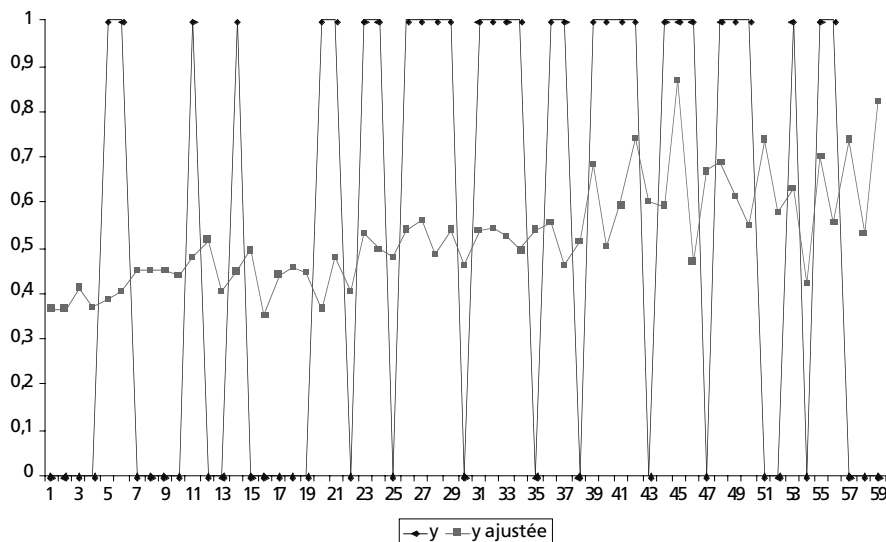
$$\text{Var}(\varepsilon_i) = E(\varepsilon_i^2) = P_i \times (1 - P_i)^2 + (1 - P_i) \times (P_i)^2 \quad [2]$$

$$\text{Soit : } \text{Var}(\varepsilon_i) = E(\varepsilon_i^2) = P_i \times (1 - P_i) \quad [3]$$

Examinons les problèmes soulevés par l'application d'une méthode des moindres carrés ordinaires afin d'estimer ce modèle.

a) – Dans l'équation [1], le codage de la variable endogène (0 ou 1) est totalement arbitraire, par exemple si nous prenions un codage sous la forme (0 ou 10) les valeurs estimées des coefficients a_k seraient évidemment différentes ($10 \times a_k$).

Un examen graphique (cf. graphique 1) de la série à expliquer (y_i) et de la série ajustée (\hat{y}_i) illustre parfaitement la non adéquation de l'estimation linéaire des paramètres par la méthode des moindres carrés.



Graphique 1 – Série brute (y_i) et de la série ajustée (\hat{y}_i)

b) – Puisque l’erreur ne peut prendre que deux valeurs, elle suit donc une loi discrète, l’hypothèse de normalité des erreurs n’est donc pas vérifiée.

c) – D’après la relation [3] : $E(\varepsilon_i^2) = P_i \times (1 - P_i)$, il existe de fait une hétéroscédasticité. Cependant nous ne pouvons pas appliquer la méthode des moindres carrés généralisés car P_i dépend des paramètres a_0 et a_1 du modèle.

d) – Enfin, nous devons imposer une contrainte au modèle : $0 \leq P_i = a_0 + a_1 x_i \leq 1$ qui peut se révéler non compatible avec les données.

Tous ces éléments indiquent clairement que nous sommes dans l’impossibilité d’utiliser la méthode des moindres carrés ordinaires.

II. Les modèles de choix binaires

Les modèles de choix binaires (ou encore appelés modèles dichotomiques) sont utilisés dans des secteurs très divers dès que la variable à expliquer ne peut prendre que deux modalités, par exemple :

- médecine : guérison ou non guérison,
- marketing : achat ou non achat,
- finance : remboursement du prêt ou non remboursement, défaillance d’une entreprise ou non défaillance,
- science politique : répondre oui ou non à un référendum, voter pour X ou Y à une élection,
- sociologique : réussir ou rater un examen,
- etc.

Dans un modèle de choix binaire, nous cherchons à modéliser une alternative ($y_i = 0$ ou 1) et donc à estimer la probabilité P_i associée à l’événement ($y_i = 1$).

A. Le modèle linéaire sur variable latente

Les variables latentes sont une première réponse aux problèmes liés à l’utilisation des moindres carrés ordinaires dans un modèle dont la variable à expliquer est binaire.

Qu’est-ce qu’une variable latente ? Il s’agit d’une variable continue non observable et représentative du phénomène étudié. Par exemple la richesse d’un individu peut être expliquée par le fait qu’il soit propriétaire ou non de son appartement.

On introduit donc la variable latente y_i^* (la richesse) et on suppose que :

- 1) l'individu i est propriétaire si sa richesse est suffisante, soit $y_i = 1$ si $y_i^* > 0$
- 2) y_i^* est une fonction linéaire des x_i , $y_i^* = a_0 + a_1x_i + \varepsilon_i$.

La variable à expliquer binaire y_i est alors définie par le modèle de décision suivant :

$$\begin{cases} y_i = 1 & \text{si } y_i^* > 0 \\ y_i = 0 & \text{si } y_i^* \leq 0 \end{cases}$$

Intuitivement, cette règle de décision consiste simplement à supposer que la proportion des ($y_i = 1$) est élevée pour $a_0 + a_1x_i + \varepsilon_i > 0$.

Soit P_i la probabilité que $y_i^* > 0$.

$$\begin{aligned} P_i &= \text{Prob}(y_i = 1) = \text{Prob}(y_i^* > 0) = \text{Prob}(a_0 + a_1x_i + \varepsilon_i > 0) \\ &= \text{Prob}(\varepsilon_i > -(a_0 + a_1x_i)) \end{aligned}$$

Si la distribution de ε_i est centrée par rapport à la moyenne, nous avons l'équivalence :

$$\text{Prob}(\varepsilon_i > -(a_0 + a_1x_i)) = \text{Prob}(\varepsilon_i < a_0 + a_1x_i)$$

$$\text{Soit : } P_i = \text{Prob}(y_i = 1) = \text{Prob}(\varepsilon_i < a_0 + a_1x_i)$$

L'ensemble de ces résultats peut être généralisé dans le cas d'un modèle à plusieurs variables.

La probabilité P_i dépend ainsi de la distribution du terme de l'erreur ε_i du modèle de décision, nous pouvons alors distinguer deux cas :

- le modèle Probit si la fonction de répartition de l'erreur suit une loi normale,
- le modèle Logit si la fonction de répartition de l'erreur suit une loi de type logistique.

B. Les modèles Probit et Logit

Dans un modèle Probit la fonction de répartition de l'erreur ε_i est donnée par :

$$P_i = \int_{-\infty}^{a_0 + a_1x_i} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

Il s'agit d'une loi normale centrée et réduite $N(0,1)$.

La fonction logistique « *Logit model* » ou encore appelée courbe en « S » est donnée par l'expression suivante :

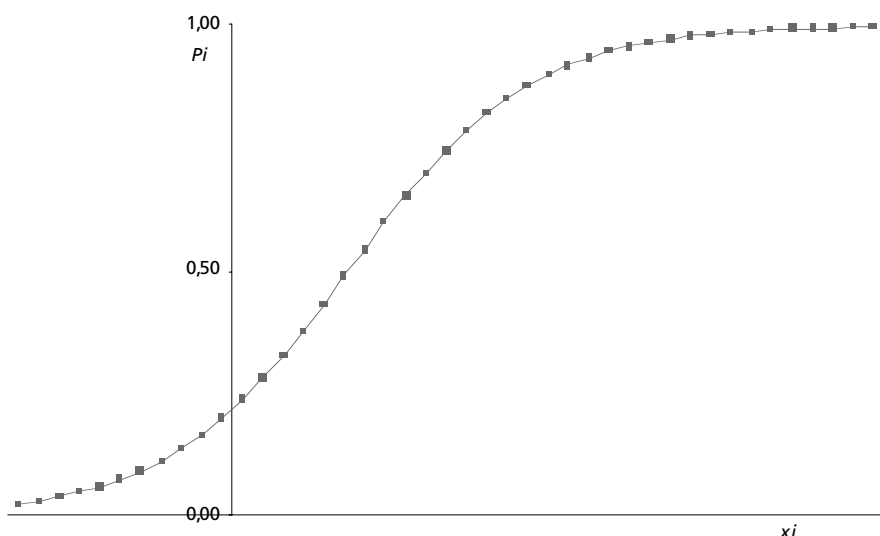
$$\begin{aligned}
 P_i &= \text{Prob}(y_i = 1) = \text{Prob}(y_i^* > 0) = \text{Prob}(a_0 + a_1x_i + \varepsilon_i > 0) \\
 &= \text{Prob}(\varepsilon_i > -(a_0 + a_1x_i)) = \Phi(a_0 + a_1x_i)
 \end{aligned}$$

avec $\Phi(\cdot)$ la fonction de répartition de la loi logistique.

$$P_i = \Phi(a_0 + a_1x_i) = \frac{\exp(a_0 + a_1x_i)}{1 + \exp(a_0 + a_1x_i)} = \frac{1}{1 + \exp(-(a_0 + a_1x_i))}$$

Les propriétés de cette équation sont les suivantes, si $\alpha_1 > 0$:

$\lim_{x_i \rightarrow \infty} P_i \rightarrow 1$ et $\lim_{x_i \rightarrow -\infty} P_i \rightarrow 0$, ce qui assure bien que les valeurs de P_i sont comprises entre 0 et 1, comme l'illustre le graphique 2.



Graphique 2 – La fonction logistique

Nous pouvons observer, après une transformation analogue à celle de l'exercice 2 du chapitre 6, que la fonction Logit s'écrit :

$\text{Ln} \left(\frac{P_i}{1 - P_i} \right) = y_i^* = a_0 + a_1x_i + \varepsilon_i$ avec $\left(\frac{P_i}{1 - P_i} \right)$ la probabilité relative du choix $y_i = 1$.

Ces spécifications peuvent être généralisées dans le cas de plusieurs variables explicatives.

L'estimation des paramètres des modèles Probit ou Logit est effectuée à l'aide de des algorithmes de maximisation d'une fonction log-vraisemblance¹. Quels que soient les modèles retenus (Probit ou Logit) les résultats d'estimation sont relativement proches, cependant les coefficients estimés ne sont pas directement comparables.

C. Interprétation des résultats et tests statistiques

Contrairement aux modèles linéaires estimés par la méthode des moindres carrés ordinaires pour lesquels les coefficients ont des interprétations économiques immédiates en termes de propension marginale, les valeurs des coefficients des modèles ne sont pas directement interprétables. Seuls les signes des coefficients indiquent si la variable agit positivement ou négativement sur la probabilité P_i .

Cependant, il est possible de calculer les effets marginaux² afin de connaître la sensibilité de la variation d'une variable explicative sur la probabilité P_i .

La significativité des coefficients est appréciée à l'aide des ratios appelés « *z-Statistique* » car la distribution des rapports du coefficient sur son écart type ne suit pas une loi de Student, comme dans le modèle linéaire général, mais une loi normale. Cette *z-Statistique* s'interprète de manière classique à partir des probabilités critiques et permet la tenue de tous les tests de significativité concernant les coefficients.

Afin de tester l'hypothèse : $H_0 : a_1 = a_2 = a_3 = \dots = a_k = 0$, nous utilisons le ratio du Log vraisemblance. Soit la statistique suivante :

$LR = -2(\ln(L_R) - \ln(L_U))$ avec L_R = valeur de la fonction du Log-vraisemblance contrainte sous H_0 et L_U = valeur de la fonction du Log-vraisemblance non contrainte.

LR suit, sous l'hypothèse nulle H_0 , une distribution d'un χ^2 à k degrés de liberté. Si la statistique LR est supérieure au χ^2 lu dans la table pour un seuil déterminé, généralement de 5 %, alors nous refusons l'hypothèse H_0 , le modèle estimé comporte au moins une variable explicative de significative.

Compte tenu de la caractéristique de la variable à expliquer codée en 0 ou 1, le coefficient de détermination R^2 n'est pas interprétable en termes d'ajustement du modèle, c'est pourquoi on utilise une statistique appelée le pseudo- R^2 donnée par :

$$R^2 = 1 - \frac{\text{Log}(L_U)}{\text{Log}(L_R)}.$$

1. Cf. Thomas A. « *Économétrie des variables qualitatives* », Dunod, 2002, pages 56-57.

2. Cf. Thomas A. (2002), pages 60-61.

Exercice n° 1

↓ fichier C12EX1

Estimation des modèles Probit et Logit binaires explicatifs des facteurs de la réussite en Licence

Nous avons relevé sur un échantillon de 60 étudiants inscrits en dernière année de Licence d'Économie, les variables suivantes susceptibles d'expliquer la réussite ou l'échec à l'examen de Licence (variable REUSSITE = 0 si échec, 1 sinon) :

NENFANTS = variable discrète représentant le nombre de frères et soeurs de l'étudiant,

NECONO = la note d'économétrie sur 20 obtenue en Licence,

NMICRO = la note de micro-économie sur 20 obtenue en Licence,

GENRE = variable muette, (1 = masculin, 0 = féminin).

Un extrait des données est présenté dans le tableau 1.

Tableau 1 – Extrait de données

OBS	REUSSITE	NENFANTS	NECONO	NMICRO	GENRE
1	0	2	3,6	0	1
2	0	5	3,8	0	1
...
59	1	0	16,2	12	0
60	1	2	17	4	0

On demande :

1) d'estimer un modèle de type Logit permettant de prévoir la probabilité de réussite d'un étudiant en Licence,

2) de comparer les résultats avec un modèle de type Probit et le modèle linéaire général,

3) de donner la probabilité de réussite, à l'aide du modèle Logit estimé, pour un étudiant dont les caractéristiques sont les suivantes : NENFANTS = 1 ; NECONO = 12 ; NMICRO = 13,5 ; GENRE = masculin.

Solution

1) Une première estimation d'un modèle Logit, conduit aux résultats suivants :

Dependent Variable : REUSSITE
Method: ML – Binary Logit
Included observations: 60

Variable	Coefficient	Std. Error	z-Statistic	Prob.
NENFANTS	– 0.682523	0.378870	– 1.801470	0.0716
NECONO	0.632062	0.239564	2.638382	0.0083
GENRE	– 3.761718	1.437068	– 2.617633	0.0089
NMICRO	0.155322	0.188916	0.822173	0.4110
C	– 3.265634	2.020060	–1.616602	0.1060

À la lecture des résultats, nous constatons que :

- la variable NMICRO à une probabilité critique de 0,41, elle n’est donc pas significative,
- la variable NENFANTS à une probabilité critique de 0,07, elle est donc faiblement significative.

Nous procédons à une nouvelle estimation en retirant la variable NMICRO dont le coefficient n’est pas significativement différent de 0.

Les résultats complets fournis par Eviews sont les suivants :

Dependent Variable : REUSSITE
Method: ML – Binary Logit
Included observations: 60

Variable	Coefficient	Std. Error	z-Statistic	Prob.
NENFANTS	– 0.746742	0.378942	– 1.970596	0.0488
NECONO	0.695857	0.231789	3.002112	0.0027
GENRE	– 3.634605	1.410945	– 2.576008	0.0100
C	– 2.859277	1.910377	– 1.496708	0.1345
Mean dependent var	0.516667	S.D. dependent var		0.503939
S.E. of regression	0.287086	Akaike info criterion		0.645890
Sum squared resid	4.615432	Schwarz criterion		0.785512
Log likelihood	– 15.37669	Hannan-Quinn criter.		0.700504
Restr. log likelihood	– 41.55549	Avg. log likelihood		– 0.25627
LR statistic (3 df)	52.35761	McFadden R-squared		0.629972
Probability(LR stat)	2.51E – 11			
Obs with Dep=0	29	Total obs		60
Obs with Dep=1	31			

Avec : L_U = Log likelihood ; L_R = Restr. log likelihood ; LR = LR statistic ; L_U/n = Avg. log likelihood. Le critère d’information de Hannan-Quinn permet des comparaisons entre les modèles (comme les critères de Akaike ou Schwarz) en termes d’arbitrage : apport d’information lié à l’ajout de variables explicatives et perte de degrés de liberté. En cas de modèle concurrent, celui ayant le plus faible critère d’information sera retenu.

a) Interprétation statistique

Les coefficients sont tous significativement différents de 0, hormis le terme constant.

La statistique de la Log vraisemblance est égale à $LR = 52,35$ que l'on compare à un χ^2 lu dans la table à un seuil de 0,95 % et à 3 degrés de liberté, $\chi_3^{20,95} = 9,28 < 52,35 \rightarrow$ rejet de H_0 .

Le pseudo- R^2 est donné par :

$$R^2 = 1 - \frac{\text{Log}(Lu)}{\text{Log}(L_R)} = 1 - \frac{-15,38}{-41,56} = 1 - 0,37 = 0,63$$

Le modèle est validé sur le plan statistique.

Le logiciel Eviews propose une table permettant d'appréhender les qualités prévisionnelles du modèle sur l'échantillon en comparant la probabilité estimée pour un individu i d'être $y_i = 1 (P(Dep = 1))$ au seuil arbitraire de 50 % à la valeur observée des $y_i = 0$ ou 1.

Dependent Variable : REUSSITE

Method: ML – Binary Logit

Included observations: 60

Prediction Evaluation (success cutoff C = 0.5)

	Estimated Equation			Constant Probability		
	Dep = 0	Dep = 1	Total	Dep = 0	Dep = 1	Total
P(Dep = 1) <= C	26	4	30	0	0	0
P(Dep = 1) > C	3	27	30	29	31	60
Total	29	31	60	29	31	60
Correct	26	27	53	0	31	31
% Correct	89.66	87.10	88.33	0.00	100.00	51.67
% Incorrect	10.34	12.90	11.67	100.00	0.00	48.33
Total Gain*	89.66	- 12.90	36.67			
Percent Gain**	89.66	NA	75.86			

Ici, pour les individus (29) pour lesquels $y_i = 0$, le modèle indique que 26 individus ont une probabilité estimée de réussite inférieure à 50 %. Dans 89,66 % des cas, les échecs sont donc correctement prévus.

Pour les individus (31) pour lesquels $y_i = 1$, le modèle indique que 27 individus ont une probabilité estimée supérieure à 50 %. Dans 87,10 % des cas, les réussites sont correctement prévues.

Le taux d'erreur est donc faible.

b) Interprétation économique

Le modèle s'écrit :

$$\text{Ln} \left(\frac{P_i}{1 - P_i} \right) = \underset{(1,97)}{-0,75} \text{NENFANTS} + \underset{(3,00)}{0,70} \text{NECONO} - \underset{(2,57)}{3,63} \text{GENRE} - 2,86 + e_i$$

(.) = z-Statistique

e_i = Résidu d'estimation

– Le nombre de frères et sœurs du foyer agit négativement, les étudiants issus de familles nombreuses ont un taux de réussite plus faible.

– La note d'économétrie est un facteur positif de réussite.

– Enfin, les étudiants de genre masculin réussissent en général moins bien (signe négatif) que les étudiants de genre féminin.

2) Estimation d'un modèle Probit et du modèle linéaire général

a) L'estimation d'un modèle Probit conduit aux résultats suivants :

Dependent Variable : REUSSITE

Method: ML – Binary Logit

Included observations: 60

Variable	Coefficient	Std. Error	z-Statistic	Prob.
NENFANTS	– 0.428197	0.219223	– 1.953247	0.0508
NECONO	0.363148	0.110230	3.294454	0.0010
GENRE	– 1.824203	0.650426	– 2.804629	0.0050
C	– 1.491466	1.108767	– 1.345157	0.1786

Les valeurs des coefficients sont de même signe mais différentes par rapport au modèle Logit car la spécification n'est pas la même. Cependant, nous pouvons retrouver, approximativement, les valeurs estimées du modèle Logit en multipliant chacun des coefficients des variables explicatives par la constante¹ $\pi \sqrt{3} \approx 1,81$.

b) Sans tenir compte des spécificités du modèle à variable à expliquer binaire et en utilisant une méthode (non appropriée) des moindres carrés avec correction pour hétéroscédasticité, nous obtenons les résultats suivants :

Dependent Variable : REUSSITE

Method: ML – Binary Logit

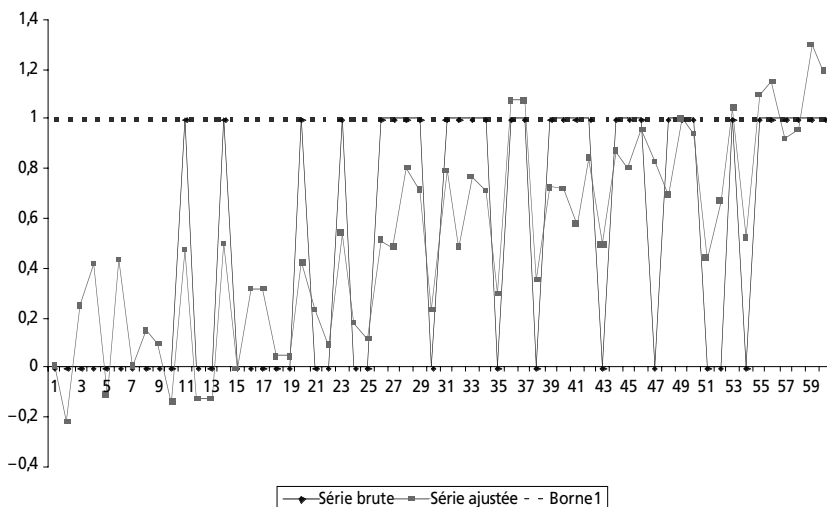
Included observations: 60

White Heteroskedasticity–Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
NENFANTS	– 0.078499	0.037854	– 2.073727	0.0427
NECONO	0.066672	0.016580	4.021299	0.0002
GENRE	– 0.295568	0.092934	– 3.180407	0.0024
C	0.217243	0.235111	0.924003	0.3594
R-squared	0.611335	F-statistic		29.36094
Durbin-Watson stat	1.741795	Prob(F-statistic)		0.000000

Les coefficients sont affectés du même signe que dans les modèles Probit et Logit, cependant le graphique des valeurs ajustées (cf. graphique 3) indique que les valeurs sont parfois négatives ou supérieures à 1 ce qui est évidemment incompatible avec la valeur d'une probabilité.

1. Une littérature abondante concerne la comparaison entre les modèles Probit et Logit, cf. Amemiya T., 1981.



Graphique 3 – Séries brute et ajustée de la variable binaire « Réussite »

3) Soit les caractéristiques de l'étudiant : NENFANTS = 1 ; NECONO = 12 ; NMICRO = 13,5 ; GENRE = masculin.

Le modèle Logit estimé (la note de micro-économie ne figurant pas dans le modèle final, elle n'est pas intégrée dans le calcul, cf. question 1) est le suivant :

$$\text{Ln} \left(\frac{P_i}{1 - P_i} \right) = -0,75 \underset{(1,97)}{\text{NENFANTS}} + 0,70 \underset{(3,00)}{\text{NECONO}} - 3,63 \underset{(2,57)}{\text{GENRE}} - 2,86 + e_i$$

$$\text{Ln} \left(\frac{\hat{P}_i}{1 - \hat{P}_i} \right) = -0,75 \times 1 + 0,70 \times 12 - 3,63 \times 1 - 2,86 = 1,109$$

$$\left(\frac{\hat{P}_i}{1 - \hat{P}_i} \right) = e^{1,109} = 3,033 \longrightarrow \hat{P}_i = 3,033 / (1 + 3,033) = 0,75$$

La probabilité de réussite de cet étudiant de licence est donc de 75 %.

III. Les modèles à choix multiples

Les modèles à choix multiples sont une généralisation des modèles binaires. Dans ces modèles la variable à expliquer, qualitative, n'est donc plus binaire (0 ou 1), mais multinomiale (ou polytomique) comme par exemple :

- le choix lors d'un vote entre quatre candidats,
- la possibilité pour un bachelier de choisir de s'inscrire dans une préparation aux concours des écoles, à l'université, dans un IUT, dans un BTS,
- appartenir à une classe de revenu,
- etc.

Nous différencions, en fonction du type de la variable à expliquer, les modèles ordonnés et les modèles non ordonnés.

A. Les modèles Probit et Logit ordonnés

Dans un modèle ordonné, les modalités de la variable à expliquer sont hiérarchisées. Elles indiquent l'appartenance de l'individu à une classe ou à une catégorie, par exemple l'appartenance à une tranche de revenu.

Nous pouvons distinguer deux classes de modèles à choix multiples ordonnés en fonction de la variable à expliquer qui est issue, soit d'une « discrétisation » d'une variable continue telle que l'appartenance à une tranche de salaire, soit directement d'une appartenance à une catégorie (faire du sport : une fois par semaine, une fois par mois, une fois par an, ...).

Comme pour le modèle de choix binaire nous modélisons¹ une variable latente continue :

$$y_i^* = a_0 + a_1 x_i + \varepsilon_i.$$

Les valeurs prises par la variable y_i correspondent à des intervalles dans lesquels se trouve y_i^* définissant ainsi le modèle de décision suivant à $M + 1$ modalités :

$$\begin{cases} y_i = 0 \text{ si } y_i^* \leq c_1 \\ y_i = 1 \text{ si } c_1 < y_i^* \leq c_2 \\ y_i = 2 \text{ si } c_2 < y_i^* \leq c_3 \\ \vdots \\ y_i = M \text{ si } c_M < y_i^* \end{cases}$$

Soit P_i la probabilité d'apparition de chaque événement pour l'individu i :

$$P_{i0} = \text{Prob}(y_i = 0) = \Phi(c_1 - (a_0 + a_1 x_i))$$

$$P_{i1} = \text{Prob}(y_i = 1) = \Phi(c_2 - (a_0 + a_1 x_i)) - \Phi(c_1 - (a_0 + a_1 x_i))$$

$$P_{i2} = \text{Prob}(y_i = 2) = \Phi(c_3 - (a_0 + a_1 x_i)) - \Phi(c_2 - (a_0 + a_1 x_i))$$

...

$$P_{iM} = \text{Prob}(y_i = M) = 1 - \Phi(c_M - (a_0 + a_1 x_i))$$

1. Afin d'alléger l'écriture nous ne faisons figurer qu'une seule variable explicative, la généralisation à k variables ne pose aucun problème comme l'illustre l'exercice n°2.

Avec Φ la fonction de répartition de la loi de probabilité normale ou logistique $\left(\Phi(t) = \frac{e^t}{1 + e^t} \right)$ et $\sum_{i=0}^M P_i = 1$

Comme pour le modèle de choix binaire, le recours à une fonction de répartition normale, permet de définir un modèle de type Probit et une fonction de répartition de type logistique permet de définir un modèle Logit. L'estimation de tous les paramètres, les coefficients de régression (a_i) et les valeurs des seuils (c_i) des modèles ordonnés (Probit ou Logit) est effectuée à l'aide des algorithmes de maximisation d'une fonction de Log-vraisemblance définie par les P_{ij} .

Les valeurs des coefficients des modèles ne sont pas directement interprétables en termes de propension marginale, seuls les signes des coefficients indiquent si la variable agit positivement ou négativement sur la variable latente.

Les résultats d'estimation s'apprécient de la même manière que pour les modèles de choix binaires :

- la significativité des coefficients à l'aide des ratios *z-Statistique*,
- la significativité globale de l'ajustement (l'hypothèse : $H_0 : a_1 = a_2 = a_3 = \dots = a_k = 0$) par la statistique $LR = -2(Ln(L_R) - Ln(L_U))$ qui suit, sous l'hypothèse nulle H_0 , une distribution d'un χ^2 à k degrés de liberté.

Le pseudo- R^2 est donné par : $R^2 = 1 - \frac{Log(L_U)}{Log(L_R)}$.

Exercice n° 2

↓ fichier C12EX2

Estimation d'un modèle à choix multiples de prévision des ventes

La société Télé-Ventes (ventes par téléphone lors d'une émission à la télévision) désire estimer le niveau des ventes par article pour chaque émission afin de dimensionner la charge de l'entrepôt et prévoir ainsi le nombre d'équipes.

L'émission est diffusée tous les jours sauf le dimanche. Les ventes sont réparties en trois classes : faible, moyenne, forte.

L'objectif est d'estimer un modèle permettant de prévoir à quelle classe de vente (faible, moyenne, forte) appartient un article présenté lors d'une émission. Pour ce faire, on dispose des informations suivantes sur 82 émissions passées :

VENTES : classe de l'article (faible = 0, moyenne = 1, forte = 2),

WE : variable indicatrice du type de jour de diffusion de l'émission (1 les jours de semaine, 0 le samedi),

EXPO : temps d'exposition du produit en minutes,

REDUC : % de réduction proposé sur le prix,

DIRECT = variable indicatrice d'émission enregistrée (0 pas direct, 1 direct).

Un extrait des données est présenté dans le tableau 2.

Tableau 2 – Extrait de données

Obs	VENTES	WE	EXPO	REDUC	DIRECT
1	0	1	3,5	0,20	0
2	1	0	3,5	0,20	0
...
81	0	1	7	0	0
82	0	1	3,5	0	0

- 1) On demande d'estimer un modèle Logit multinomial permettant de prévoir la classe de vente d'un article à partir des facteurs explicatifs proposés.
- 2) D'effectuer une prévision pour un article présenté lors d'une émission en différé diffusée en semaine dont le temps d'exposition est de 7 minutes et sans réduction.

Solution

- 1) Une première estimation à l'aide d'un modèle de type Logit conduit au résultat suivant :

Dependent Variable : VENTES

Method: ML – Ordered Logit (Quadratic hill climbing)

Included observations: 82 after adjusting endpoints

	Coefficient	Std. Error	z-Statistic	Prob.
DIRECT	0.226809	1.325269	0.171142	0.8641
EXPO	0.644804	0.150121	4.295228	0.0000
REDUC	8.922215	2.759073	3.233772	0.0012
WE	- 1.395159	0.512093	- 2.724426	0.0064

Nous constatons que la variable DIRECT dont le coefficient est affecté d'une probabilité critique de 0,86 n'est pas significative (les téléspectateurs ne sont pas sensibles aux émissions diffusées en direct), elle est donc retirée du modèle. La nouvelle estimation est alors la suivante.

Dependent Variable : VENTES

Method: ML – Ordered Logit (Quadratic hill climbing)

Included observations: 82 after adjusting endpoints

	Coefficient	Std. Error	z-Statistic	Prob.
EXPO	0.641023	0.148189	4.325700	0.0000
REDUC	8.982926	2.735164	3.284237	0.0010
WE	- 1.377785	0.500824	- 2.751038	0.0059

Limit Points

LIMIT_1:C(4)	3.679581	0.994565	3.699689	0.0002
LIMIT_2:C(5)	6.170926	1.174856	5.252497	0.0000
Akaike info criterion	1.605306	Schwarz criterion		1.752057
Log likelihood	- 60.81754	Hannan-Quinn criter.		1.664224
Restr. log likelihood	- 79.48219	Avg. log likelihood		- 0.74167
LR statistic (3 df)	37.32930	LR index (Pseudo-R2)		0.234828
Probability(LR stat)	3.92E-08			

a) Interprétation statistique

Les coefficients sont maintenant tous significativement différents de 0 (probabilités critiques inférieures à 0,05).

La statistique de la Log vraisemblance est égale à $LR = 37,33$ que l'on compare à un χ^2 lu dans la table à un seuil de 0,95 % et à 3 degrés de liberté, $\chi^2_{3,0.95} < 37,33 \rightarrow$ rejet de H_0 .

Le modèle est donc validé sur le plan statistique.

b) Interprétation économique

La durée d'exposition et le pourcentage de réduction agissent positivement sur les ventes.

Une émission diffusée un jour de semaine engendre moins de ventes qu'une émission diffusée le samedi.

Les seuils c_1 et c_2 sont respectivement de 3,679 et 6,170.

Les signes des coefficients sont conformes à l'intuition économique.

Le logiciel Eviews propose une table permettant d'appréhender les qualités prévisionnelles du modèle sur l'échantillon :

Dependent Variable : VENTES

Method: ML – Ordered Logit (Quadratic hill climbing)

Included observations: 82 after adjusting endpoints

Prediction table for ordered dependent variable

Value	Count	Count of obs with Max Prob	Error	Sum of all Probabilities	Error
0	44	47	- 3	44.381	- 0.381
1	27	31	- 4	26.834	0.166
2	11	4	7	10.785	0.215

À la lecture des résultats, nous constatons que les qualités prévisionnelles de ce modèle sont satisfaisantes car le taux d'erreur est assez faible pour les ventes de niveau 0 et 1. En revanche, pour les ventes de niveau 2 (forte) nous constatons que sur 11 ventes réalisées, le modèle en a prévues correctement seulement 4.

Enfin pour chaque individu Eviews indique :

VENTES_0_0 = la probabilité associée à la modalité 0,

VENTES_1_0 = la probabilité associée à la modalité 1,

VENTES_2_0 = la probabilité associée à la modalité 2,

I_VENTES_0 = la réalisation de l'estimation de la variable latente.

Un exemple est donné¹ pour les cinq premières ventes :

1. Un exemple de calcul est détaillé lors de la réponse à la question 2.

obs	VENTES	VENTES_0_0	VENTES_1_0	VENTES_2_0	I_VENTES_0
1	0	0.7342647	0.2366418	0.0290935	2.663212
2	1	0.4106169	0.4831615	0.1062216	4.040996
3	0	0.4106169	0.4831615	0.1062216	4.040996
4	2	0.4106169	0.4831615	0.1062216	4.040996
5	1	0.7342647	0.2366418	0.0290935	2.663212

2) Nous calculons l'estimation de la variable latente correspondante aux caractéristiques de l'émission avec EXPO = 7, REDUC = 0 et WE = 1 :

$$\hat{y}^* = 0,641 \times \text{EXPO} + 8,982 \times \text{REDUC} - 1,377 \times \text{WE}$$

$$\hat{y}^* = 0,641 \times 7 + 8,982 \times 0 - 1,377 \times 1 = 3,109$$

Puis nous calculons les probabilités :

$$P_1 = \text{Prob}(y_i = 0) = \Phi(c_1 - x_i \hat{a}) = \Phi(3,679 - 3,109)$$

$$= \Phi(0,57) = \frac{e^{0,57}}{1 + e^{0,57}} = 0,639$$

$$P_2 = \text{Prob}(y_i = 1) = \Phi(c_2 - x_i \hat{a}) - \Phi(c_1 - x_i \hat{a}) = \Phi(6,17 - 3,109) - 0,639$$

$$P_2 = \Phi(3,06) - 0,639 = \frac{e^{3,06}}{1 + e^{3,06}} - 0,639 = 0,955 - 0,639 = 0,31$$

$$P_3 = \text{Prob}(y_i = 2) = 1 - \Phi(6,17 - 3,109) = 1 - 0,955 = 0,045$$

Soit les résultats suivants :

	VENTES = 0	VENTES = 1	VENTES = 2
P_i	0,639	0,316	0,045

Les ventes de l'émission prévue ont 64% de probabilité d'appartenir à la classe 0 de faible vente.

B. Le modèle de choix multiples non ordonné : le Logit multinomial

Dans les modèles non ordonnés la variable à expliquer représente les possibilités de choix d'un individu parmi $M + 1$ possibilités et donc ses préférences, c'est pourquoi on les dénomme aussi modèle à utilités aléatoires.

Soit l'utilité¹ de l'individu i pour la possibilité j ($j = 0, 1, \dots, M$) :

$$U_{ij} = a_0 + a_1 x_i + \varepsilon_i$$

1. Afin d'alléger l'écriture nous ne faisons figurer qu'une seule variable explicative, la généralisation à k variables ne pose aucun problème.

La décision $y_i = j$ est retenue si $U_{ij} = \text{Max}(U_{i0}, U_{i1}, \dots, U_{iM})$.

La loi des utilités peut être normale (Probit multinomial) ou log-normale (Logit multinomial).

Dans le cas le plus fréquent d'un modèle Logit multinomial à $M + 1$ modalités, la probabilité du choix j pour l'individu i est donnée¹ par :

$$\text{Prob}(y_i = j) = \frac{\exp(x'_i a_j)}{\sum_{k=0}^M \exp(x'_i a_k)} = \frac{1}{1 + \sum_{\substack{k=0 \\ k \neq j}}^M \exp(x'_i (a_k - a_j))}$$

Avec x_i le vecteur des caractéristiques de l'individu i et a_j le vecteur caractéristique de la modalité j .

Dans un modèle Logit multinomial, la probabilité d'apparition de la $j^{\text{ème}}$ modalité dépend de la différence $a_k - a_j$, nous pouvons donc poser sans restriction $a_0 = 0$.

Par exemple, prenons le cas d'un individu i qui a trois choix ($M = 2$), les probabilités s'écrivent :

$$\text{Prob}(y_i = 0) = \frac{1}{1 + \exp(x'_i (a_1 - a_0)) + \exp(x'_i (a_2 - a_0))}$$

$$\text{Prob}(y_i = 1) = \frac{1}{1 + \exp(x'_i (a_0 - a_1)) + \exp(x'_i (a_2 - a_1))}$$

$$\text{Prob}(y_i = 2) = \frac{1}{1 + \exp(x'_i (a_0 - a_2)) + \exp(x'_i (a_1 - a_2))}$$

Or, par construction $\sum_{i=0}^2 P_i = 1$ et comme nous avons posé la contrainte $a_0 = 0$, nous pouvons estimer les deux paramètres a_1 et a_2 qui s'interprètent comme étant des écarts par rapport au paramètre a_0 .

Une autre caractéristique fondamentale du Logit multinomial réside dans la propriété d'indépendance – par rapport aux autres événements – du rapport des deux probabilités associées à deux possibilités j et k :

$$\frac{\text{Prob}(y_i = j)}{\text{Prob}(y_i = k)} = \frac{\exp(x_i a_j)}{\exp(x_i a_k)} = \exp(x_i (a_j - a_k))$$

Ce rapport est indépendant des possibilités autres que j et k , cette hypothèse s'appelle l'indépendance des alternatives non pertinentes « *Independence of*

1. Cf. Thomas A. (2002), pages 92-93.

Irrelevant Alternative ». Cette propriété est vérifiée¹ si le fait de changer l'ensemble des choix (par exemple ajouter une possibilité) ne modifie pas les ratios de chances.

L'estimation des paramètres du modèle est effectuée à l'aide des algorithmes de maximisation de la fonction log-vraisemblance. Les valeurs des coefficients ne sont pas directement interprétables en termes de propension marginale, seuls les signes des coefficients indiquent si la variable agit positivement ou négativement sur la probabilité relative de choisir j plutôt que 0.

Les résultats d'estimation s'apprécient de la même manière que pour les autres modèles :

- la significativité des coefficients à l'aide des ratios *z-Statistique*,
- la significativité globale de l'ajustement (l'hypothèse : $H_0 : a_1 = a_2 = a_3 = \dots = a_k = 0$) par la statistique $LR = -2(Ln(L_R) - Ln(L_U))$ qui suit, sous l'hypothèse nulle H_0 , une distribution d'un χ^2 à k degrés de liberté.

Le pseudo- R^2 est donné par : $R^2 = 1 - \frac{Log(L_U)}{Log(L_R)}$.

IV. Les modèles à variable dépendante limitée : le modèle Tobit

Dans les modèles à variable dépendante limitée (« *limited dependent variable models* ») les valeurs de la variable à expliquer continue ne sont observables que sur un intervalle donné. Par exemple :

- La demande d'un produit n'est connue (les ventes observées) que si le magasin possède ce produit en stock, au-delà la demande non satisfaite (les ventes perdues) est inconnue.
- Le salaire d'un jeune diplômé n'est observable que si il travaille (salaire > 0).
- Lors d'une enquête sur les dépenses relatives à l'achat de biens d'équipement, seules les dépenses des ménages ayant effectivement acheté un bien d'équipement sont prises en compte.
- etc.

Le recours à l'économétrie des variables qualitatives se justifie par l'estimation de la probabilité que la variable à expliquer se trouve à l'intérieur de l'intervalle pour lequel elle est observable. Ces modèles, appelés modèle Tobit par référence à J. Tobin², ont connu beaucoup d'extensions, nous pouvons citer :

1. Cf. Thomas A. (2002), pages 101-103.

2. Tobin J. (1958).

- le modèle censuré et le modèle tronqué (modèle Tobit simple) qui font seuls l'objet d'une présentation dans le cadre de ce manuel,
- le modèle de sélection (modèle Tobit généralisé),
- le modèle avec censure et sélection.

A. Le modèle Tobit simple : modèle de régression tronqué ou censuré

Définitions préliminaires :

- Une distribution est tronquée si les valeurs de la variable à expliquer et des variables explicatives ne sont pas observables lorsque les valeurs de la variable à expliquer dépassent un intervalle.
- Une distribution est censurée si seules les valeurs de la variable à expliquer ne sont pas connues lorsqu'elles sortent d'un intervalle.

Soit une variable latente y_i^* – donc non systématiquement observable – et des variables explicatives x_i , le modèle s'écrit de manière classique comme dans le cas des modèles à choix discrets : $y_i^* = x_i a + \varepsilon_i$ avec :

x_i la matrice des valeurs des variables explicatives,

a les coefficients du modèle,

ε_i le terme d'erreur avec $\varepsilon_i \rightarrow N(0, \sigma_\varepsilon^2)$ et donc $y_i^* \rightarrow N(x_i a, \sigma_\varepsilon^2)$.

Les valeurs observées de la variable explicative y_i sont données par :

$$\begin{cases} y_i = y_i^* \text{ si } y_i^* > 0 \\ y_i = 0 \text{ si } y_i^* \leq 0 \end{cases}$$

Lorsque les valeurs de y_i^* sont nulles ou négatives, y_i est égale à 0 mais l'on connaît néanmoins les valeurs des variables explicatives. Les données dans ce type de modèle sont dites censurées à gauche, y_i suit alors une loi normale censurée.

Comme pour les moindres carrés ordinaires :

$$E(y_i^*) = x_i a \text{ mais } E(y_i / y_i > 0) \neq x_i a \text{ et } E(y_i) \neq x_i a.$$

En effet, calculons l'espérance de y_i dans le cas d'un modèle censuré à gauche, non pas à 0, mais à une valeur seuil égale à c_1 , le modèle s'écrit :

$$\begin{cases} y_i = y_i^* \text{ si } y_i^* > c_1 \\ y_i = c_1 \text{ si } y_i^* \leq c_1 \end{cases}$$

$$y_i^* = x_i a + \varepsilon_i \text{ avec } \varepsilon_i \rightarrow N(0, \sigma_\varepsilon^2).$$

L'espérance a pour expression :

$$E(y_i) = \text{Prob}(y_i = c_1) \times E(y_i/y_i = c_1) + \text{Prob}(y_i > c_1) \times E(y_i/y_i > c_1)$$

Soit Φ la fonction de répartition de la loi de probabilité des erreurs (normale ou logistique) et ϕ la fonction de densité associée. Nous avons :

$$\text{Prob}(y_i = c_1) = \text{Prob}(y_i^* \leq c_1) = \text{Prob}\left(\frac{y_i^* - x_i a}{\sigma} \leq \frac{c_1 - x_i a}{\sigma}\right) = \Phi\left(\frac{c_1 - x_i a}{\sigma}\right)$$

$$\text{Prob}(y_i > c_1) = \text{Prob}(y_i = y_i^*) = \frac{1}{\sigma} \phi\left(\frac{c_1 - x_i a}{\sigma}\right) = 1 - \Phi\left(\frac{c_1 - x_i a}{\sigma}\right)$$

$$E(y_i/y_i = c_1) = c_1$$

$$E(y_i/y_i > c_1) = E(y_i^*/y_i^* > c_1) = \left(x_i a + \sigma \frac{\phi[(c_1 - x_i a)/\sigma]}{1 - \Phi[(c_1 - x_i a)/\sigma]}\right) \neq x_i a$$

Soit :

$$E(y_i) = \Phi\left(\frac{c_1 - x_i a}{\sigma}\right) \times c_1 + \left[1 - \Phi\left(\frac{c_1 - x_i a}{\sigma}\right)\right] \times \left(x_i a + \sigma \frac{\phi[(c_1 - x_i a)/\sigma]}{1 - \Phi[(c_1 - x_i a)/\sigma]}\right) \neq x_i a \quad [4]$$

Le recours à la méthode des moindres carrés n'est donc pas possible.

Le modèle de Tobit peut alors s'écrire de la manière suivante.

En posant $\hat{y}_i^* = x\hat{a}$ le modèle estimé de la variable latente et en effectuant les termes du deuxième membre de [4] et après simplification, nous obtenons l'écriture suivante :

$$y_i = c_1 \times \Phi((c_1 - \hat{y}_i^*)/\sigma) + (1 - \Phi((c_1 - \hat{y}_i^*)/\sigma) > 0) \times (\hat{y}_i^* \times (1 - \Phi((c_1 - \hat{y}_i^*)/\sigma)) + \sigma \times (\phi((c_1 - \hat{y}_i^*)/\sigma))) \quad [5]$$

Plus généralement, et par extension, il est possible de formaliser un modèle de Tobit avec censure à gauche et censure à droite tel que :

$$\begin{cases} y_i = c_1 & \text{si } y_i^* \leq c_1 \\ y_i = y_i^* & \text{si } c_1 < y_i^* \leq c_2 \\ y_i = c_2 & \text{si } c_2 < y_i^* \end{cases}$$

Avec c_1 et c_2 deux valeurs numériques qui représentent les seuils de censure, l'une de ces valeurs peut être égale à ∞ , si il n'existe pas de seuil à droite ou à gauche.

Dans le cas particulier où $c_1 = 0$ et $c_2 = \infty$, nous retrouvons le modèle canonique de Tobit.

Le modèle de Tobit à deux seuils s'écrit à l'aide d'une démonstration analogue à la précédente :

$$\begin{aligned}\hat{y}_i^* &= x\hat{a} \\ y_i &= c_1 \times \Phi((c_1 - \hat{y}_i^*)/\sigma) + c_2 \times (1 - \Phi((c_2 - \hat{y}_i^*)/\sigma)) \\ &\quad + (\Phi((c_2 - \hat{y}_i^*)/\sigma) - \Phi((c_1 - \hat{y}_i^*)/\sigma) > 0) \\ &\quad \times (\hat{y}_i^* * (\Phi((c_2 - \hat{y}_i^*)/\sigma) - \Phi((c_1 - \hat{y}_i^*)/\sigma)) + \sigma \\ &\quad \times (-\phi((c_2 - \hat{y}_i^*)/\sigma) + \phi((c_1 - \hat{y}_i^*)/\sigma)))\end{aligned}\quad [6]$$

B. Estimation et interprétation des résultats

La méthode des moindres carrés ordinaires n'est pas adaptée du fait de l'apparition fréquente, dans les observations de la variable à expliquer, des valeurs seuils. L'estimation des paramètres (a et σ) du modèle est effectuée à l'aide d'un algorithme de maximisation de la fonction de la Log-vraisemblance.

Les résultats d'estimation s'apprécient directement par la significativité des coefficients à l'aide des ratios appelés *z-Statistique*.

Les valeurs des coefficients du modèle estimé sont directement interprétables en termes de propension marginale sur y_i^* , du fait de la continuité dans l'intervalle des seuils de la variable à expliquer.

Exercice n° 3

↓ fichier C12EX3

Prévision de la demande d'électricité pour un fournisseur à capacité limitée

Dans le cadre de l'ouverture du marché de l'électricité à destination des industriels, un fournisseur d'électricité, qui n'est pas l'opérateur historique, propose de l'énergie électrique à bas prix dans la limite de ses capacités fixées à 3 000 mégawatts : la demande supérieure à ce seuil ne peut donc pas être servie.

Au-delà de ses capacités les clients sont délestés et sont donc dans l'obligation de basculer vers une autre source. La demande (y_t) exprimée en mégawatts à la période t est fonction de trois facteurs explicatifs :

x_{1t} = indicateur d'écart de prix par rapport à la concurrence en t , la valeur indique le % de réduction, pour le jour considéré, accordé par l'opérateur historique (si 0 pas de réduction de prix),

x_{2t} = nombre de clients industriels alimentés en t ,

x_{3t} = variable indicatrice signalant les jours particuliers à forte consommation tels que le lendemain de jour férié, ...

Soit les données quotidiennes sur 60 jours (cf. un extrait dans le tableau 3) dont cet opérateur dispose.

Tableau 3 – Extrait de données

Jour	y	x_1	x_2	x_3
1	2717	0	61	0
2	2126	0	32	0
...
59	2683	0,1	61	0
60	3000	0	79	0
61		1	98	1
62		0	60	0

On demande :

- 1) d'estimer un modèle Logit permettant de prévoir la demande quotidienne à partir des facteurs explicatifs proposés et de commenter les résultats ;
- 2) de comparer les résultats avec l'estimation par la méthode, non adaptée, des moindres carrés ordinaires ;
- 3) d'effectuer une prévision pour les jours 61 et 62 sachant que :

$$x_{1,61} = 1 ; x_{2,61} = 98 ; x_{3,61} = 1 \text{ et } x_{1,62} = 0 ; x_{2,62} = 60 ; x_{3,62} = 0.$$

Solution

- 1) La consommation est censurée car les valeurs de la variable à expliquer (la demande) ne sont pas connues lorsqu'elles sortent de l'intervalle $[0 ; 3\,000]$ puisque au-delà de 3 000, la demande ne peut être satisfaite.

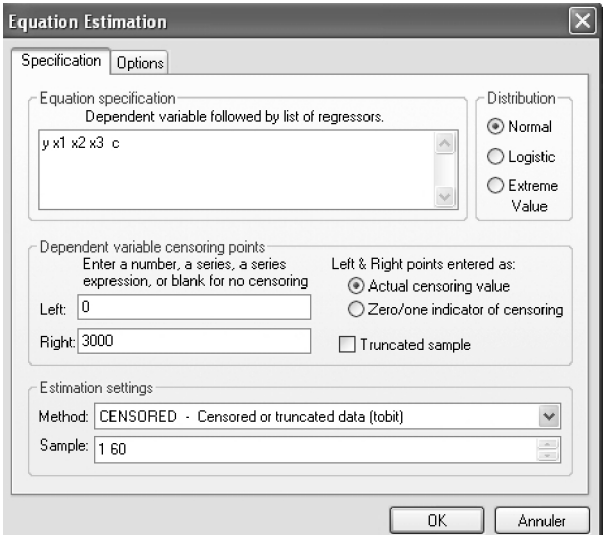


Figure 1 – Fenêtre de paramétrage pour un modèle Tobit (logiciel Eviews)

Pour estimer le modèle Tobit, le logiciel Eviews (cf. figure 1) permet de paramétrer les seuils de censure à gauche ($c_1 = 0$) et à droite ($c_2 = 3\,000$) et de choisir la distribution de l'erreur (ici normale). Nous ne sélectionnons donc pas l'option distribution tronquée puisque la distribution est censurée.

Les résultats d'estimation sont les suivants :

Dependent Variable : Y
 Method: ML – Censored Normal (TOBIT) (Quadratic hill climbing)
 Included observations: 60 after adjustments
 Left censoring (value) series: 0
 Right censoring (value) series: 3000

	Coefficient	Std. Error	z-Statistic	Prob.
X1	– 27.69316	7.582649	– 3.652175	0.0003
X2	20.50361	0.480937	42.63261	0.0000
X3	186.2357	34.88046	5.339256	0.0000
C	1473.642	23.60984	62.41642	0.0000
Error Distribution				
SCALE: SIG	47.41983	4.617038	10.27062	0.0000
R-squared	0.981952	Mean dependant var		2542.667
S.E. of regression	46.09561	Akaike info criterion		9.355239
Sum squared resid	116864.3	Schwarz criterion		9.529768
Log likelihood	– 275.6572	Hannan-Quinn criter		9.423507
Avg. log likelihood	– 4.594286			
Left censored obs	0	Right censored obs		8
Uncensored obs	52	Total obs		60

a) Interprétation statistique

Les coefficients sont tous significativement différents de 0 (les probabilités critiques des coefficients sont toutes inférieures à 0,05), le modèle est validé sur le plan statistique.

Eviews indique, sur l'avant dernière ligne, le nombre de données censurées : 0 à gauche et 8 à droite.

b) Interprétation économique

- L'indicateur d'écart de prix agit négativement sur la demande : en cas de réduction tarifaire de la concurrence, la demande diminue.
- Le nombre de clients connectés au réseau et la variable muette « type de jour » ont un effet positif sur la demande.
- La variable d'échelle (estimateur de σ) est égale à 47,41.

Les coefficients ont bien le signe attendu, le modèle est validé sur le plan économique.

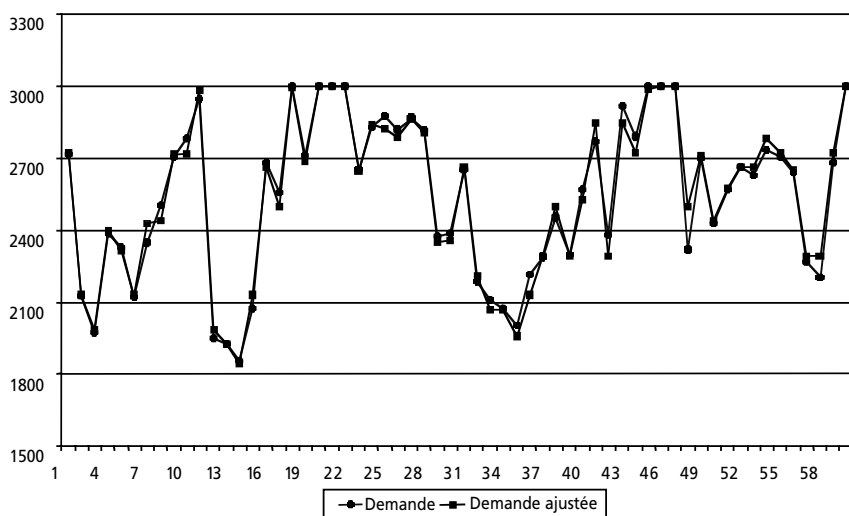
Le modèle Tobit s'écrit d'après [6] :

$$\hat{y}_t^* = -27,69 \times x_{1t} + 20,50 \times x_{2t} + 186,23 \times x_{3t} + 1473,64$$

$$\begin{aligned}
y_t = & 0 \times \Phi((0 - \hat{y}_t^*)/47,41) + 3000 \times (1 - \Phi((3000 - \hat{y}_t^*)/47,41)) \\
& + (\Phi((3000 - \hat{y}_t^*)/47,41) - \Phi((0 - \hat{y}_t^*)/47,41) > 0) \\
& \times (\hat{y}_t^* \times (\Phi((3000 - \hat{y}_t^*)/47,41) - \Phi((0 - \hat{y}_t^*)/47,41)) \\
& + 47,41 \times (-\phi((3000 - \hat{y}_t^*)/47,41) + \phi((0 - \hat{y}_t^*)/47,41)))
\end{aligned}$$

avec Φ la fonction de répartition de la loi normale et ϕ la fonction de densité associée.

Les graphiques (cf. graphique 4) de la série brute et de la série ajustée présentent la qualité de l'ajustement : les valeurs sont limitées par le seuil maximum, soit 3 000 mégawatts.



Graphique 4 – Séries brute et ajustée de la variable (y_t) demande (Modèle Tobit)

- 2) L'application de la méthode des moindres carrés ordinaires conduit aux résultats suivants :

Dependent Variable : Y

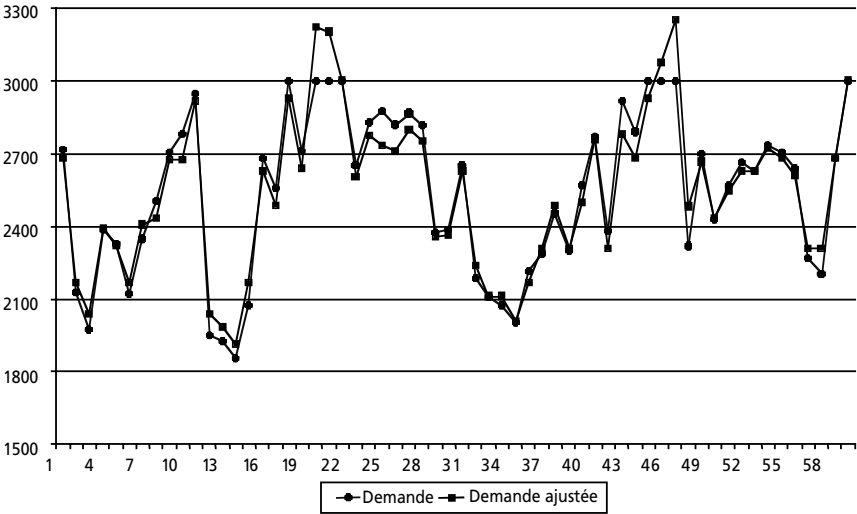
Method: Least Squares

Included observations: 60 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X1	- 31.95356	11.44578	- 2.791732	0.0072
X2	17.83772	0.659615	27.04260	0.0000
X3	156.0019	60.45055	2.580653	0.0125
C	1594.516	34.80522	45.81254	0.0000
R-squared	0.941213	F-statistic		298.8649
Sum squared resid	380662.1	Prob(F-statistic)		0.000000

Les coefficients du modèle, tous significativement différents de 0, sont biaisés. Nous constatons que les valeurs estimées des coefficients sont légèrement différentes, bien que de même signe, de celles des valeurs estimées à l'aide du modèle Tobit.

Les graphiques (cf. graphique 5) de la série brute et de la série ajustée indiquent que l'utilisation de la méthode des MCO peut fournir des valeurs estimées parfois supérieures à 3 000 mégawatts ce qui est techniquement impossible.



Graphique 5 – Séries brute et ajustée de la variable (y_t) demande (Méthode MCO)

3) Le calcul de la prévision pour les jours 61 et 62 est directement effectué par application du modèle Tobit estimé.

Sachant que :

$$x_{1,61} = 1 \ ; \ x_{2,61} = 98 \ ; \ x_{3,61} = 1 \ \text{et} \ x_{1,62} = 0 \ ; \ x_{2,62} = 60 \ ; \ x_{3,62} = 0.$$

La prévision pour le jour 61 est donnée par :

$$y_{61}^* = -27,69 \times 1 + 20,50 \times 98 + 186,23 \times 1 + 1473,64 = 3641,53$$

$$y_{61}^* > c_2 \longrightarrow y_{61} = c_2 = 3\,000$$

La prévision pour le jour 62 est donnée par :

$$y_{62}^* = -27,69 \times 0 + 20,50 \times 60 + 186,23 \times 0 + 1473,64 = 2703,85$$

$$c_1 < y_{62}^* < c_2 \longrightarrow y_{62} = y_{62}^* = 2703,85$$

Le calcul de la prévision à l'aide du modèle estimé par la méthode des MCO conduit à des résultats légèrement différents comme l'illustre le tableau 4.

Tableau 4 – Comparaison des prévisions par le modèle Tobit et par la méthode des MCO

Jour	Modèle Tobit	MCO
61	3000	3544,63
62	2703,86	2687,26

13. Introduction à l'économétrie des données de panel

Dans le chapitre 1 nous avons présenté les données de panel comme étant des modèles dans lesquels les variables représentent les valeurs prises par un échantillon d'individus à intervalle de temps réguliers. En économétrie nous pouvons nous intéresser à des données temporelles ou à des données en coupe instantanée, les données de panel mélangent les deux dimensions : séries temporelles et données en coupe. Les exemples sont nombreux : estimation économétrique entre les exportations et l'ouverture commerciale sur 20 ans (le temps) et 15 pays (les individus), le revenu d'un échantillon de ménages sur plusieurs années, l'évolution du chiffre d'affaires de 30 grandes surfaces sur 36 mois... L'intérêt de cette spécification réside dans la prise en compte d'un effet temporel et d'un effet individuel.

Les individus sont l'unité statistique observée (consommateur, entreprise, région, pays, etc.) à intervalle de temps régulier.

Dans ce chapitre, qui constitue une simple introduction à l'économétrie des données de panel¹, nous présentons à la section I les modèles à données de panel et la méthode d'estimation SUR ; puis en section II, les tests d'homogénéité permettant de déterminer la structure du panel. Enfin la section III traite des panels à effets fixes et aléatoires, des méthodes d'estimation et du test d'Hausman de spécification des effets individuels.

1. Le lecteur désirant approfondir ce chapitre peut se référer à Sevestre P. 2002.

I. Présentation des modèles à données de panel

A. Spécificités des données de panel

1) Définitions préliminaires

Les données de panel (ou données longitudinales) sont représentatives d'une double dimension : individuelle et temporelle. Un panel équilibré (*balanced panel*) a le même nombre d'observations pour tous les individus, un panel déséquilibré (*unbalanced panel*) est un panel où il manque des observations pour certains individus.

2) Exemple introductif

Prenons un exemple introductif très simple, un panel composé de 2 individus ($N = 2$) et connu sur 3 périodes ($T = 3$), soit un total de $N \times T = 6$ observations à $k = 2$ variables explicatives.

Le modèle général s'écrit :

$$y_{it} = a_{0i} + a_{1i}x_{1it} + a_{2i}x_{2it} + \varepsilon_{it},$$

y_{it} = variable endogène observée pour l'individu i à la période t ,

x_{1it}, x_{2it} = variables explicatives observées pour l'individu i à l'instant t ,

a_{0i} = terme constant pour l'individu i ,

a_{1i}, a_{2i} = coefficients des 2 variables exogènes pour l'individu i ,

ε_{it} = terme d'erreur pour l'individu i à la période t .

Soit en l'écrivant pour l'ensemble des 6 observations¹ :

$$\text{Individu n}^\circ 1 \quad \begin{cases} y_{11} = a_{01} + a_{11}x_{111} + a_{21}x_{211} + \varepsilon_{11} \\ y_{12} = a_{01} + a_{11}x_{112} + a_{21}x_{212} + \varepsilon_{12} \\ y_{13} = a_{01} + a_{11}x_{113} + a_{21}x_{213} + \varepsilon_{13} \end{cases}$$

$$\text{Individu n}^\circ 2 \quad \begin{cases} y_{21} = a_{02} + a_{12}x_{121} + a_{22}x_{221} + \varepsilon_{21} \\ y_{22} = a_{02} + a_{12}x_{122} + a_{22}x_{222} + \varepsilon_{22} \\ y_{23} = a_{02} + a_{12}x_{123} + a_{22}x_{223} + \varepsilon_{23} \end{cases}$$

1. Nous constatons que nous avons 6 observations et 6 paramètres à estimer, l'objectif de cet exemple est simplement d'illustrer l'écriture d'un modèle en panel...

Sous forme matricielle le modèle s'écrit :

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix} = \begin{bmatrix} 1 & 0 & x_{111} & 0 & x_{211} & 0 \\ 1 & 0 & x_{112} & 0 & x_{212} & 0 \\ 1 & 0 & x_{113} & 0 & x_{213} & 0 \\ 0 & 1 & 0 & x_{121} & 0 & x_{221} \\ 0 & 1 & 0 & x_{122} & 0 & x_{222} \\ 0 & 1 & 0 & x_{123} & 0 & x_{223} \end{bmatrix} \begin{bmatrix} a_{01} \\ a_{02} \\ a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \end{bmatrix}$$

Si les hypothèses classiques sur les erreurs sont respectées (homoscédasticité et indépendance temporelle) et quelles sont indépendantes d'un individu à l'autre ($\text{Cov}(\varepsilon_{it}, \varepsilon_{jt}) = 0$ pour $i \neq j$), nous pouvons appliquer la méthode des moindres carrés pour chacune des N équations relatives aux individus. Dans le cas où l'hypothèse d'indépendance entre les individus ($\text{Cov}(\varepsilon_{it}, \varepsilon_{jt}) = \sigma_{ij}^2 \neq 0$ pour $i \neq j$) n'est plus vérifiée, l'estimateur des MCO n'est plus BLUE, nous devons appliquer la méthode SUR (*Seemingly Unrelated Regressions*) de Zellner.

B. La méthode SUR

La méthode SUR, *Seemingly Unrelated Regressions*, de Zellner (1962) est utilisée lorsque les erreurs des équations individuelles sont corrélées : la covariance individuelle $\text{Cov}(\varepsilon_{it}, \varepsilon_{jt}) = \sigma_{ij}^2 \neq 0$ pour $i \neq j$. Les individus sont alors interdépendants. Cette méthode consiste à appliquer les MCG : $\widehat{a}_{SUR} = (X' \Omega^{-1} X)^{-1} (X' \Omega^{-1} Y)$

$$\text{Avec } \Omega_{(N \times T, N \times T)} = \begin{pmatrix} \sigma_1^2 I & \sigma_{12} I & \dots & \sigma_{1N} I \\ \sigma_{21} I & \sigma_2^2 I & \dots & \sigma_{2N} I \\ \dots & \dots & \dots & \dots \\ \sigma_{N1} I & \sigma_{N2} I & \dots & \sigma_N^2 I \end{pmatrix}, \text{ chaque élément } \sigma_{ij} I \text{ étant de}$$

dimension (T, T) .

Pratiquement, la procédure est la suivante :

- Estimation par les MCO des N équations individuelles.
- Calcul des N résidus e_{it} .

$$\text{– Calcul des } \widehat{\sigma}_i^2 = \frac{\sum_{t=1}^T e_{it}^2}{T} \text{ et } \widehat{\sigma}_{ij} = \frac{\sum_{t=1}^T e_{it} e_{jt}}{T}.$$

- Application des MCG : $\widehat{a}_{SUR} = (X' \widehat{\Omega}^{-1} X)^{-1} (X' \widehat{\Omega}^{-1} Y)$.

Si les erreurs sont, de plus, autocorrélées dans le temps, la correction peut s'effectuer par une des méthodes présentées au chapitre 5.

La méthode d'estimation SUR présente l'avantage de prendre en compte l'hétérogénéité individuelle par l'intégration de la dépendance entre les individus mais présente l'inconvénient – surtout si le nombre d'observations temporelles T est faible – d'un grand nombre de paramètres à estimer : $N \times (k + 1)$ coefficients et $N \times (N + 1)/2$ éléments de la matrice Ω .

Les modèles à effets individuels fixes ou aléatoires permettent de trouver un compromis entre le nombre de paramètres à estimer et l'intégration de l'hétérogénéité individuelle.

C. Le modèle linéaire simple

Le modèle en données de panel peut s'écrire pour N individus ($i = 1, \dots, N$) et T observations temporelles ($t = 1, \dots, T$), soit $n = N \times T$ observations totales, de la manière suivante :

$$y_{it} = a_{0i} + a'_i x_{it} + \varepsilon_{it}.$$

y_{it} = variable endogène observée pour l'individu i à la période t ,

x_{it} = vecteur des k variables exogènes $x'_{it} = (x_{1it}, x_{2it}, \dots, x_{kit})$; x_{kit} est donc la valeur observée pour la k^e variable exogène pour l'individu i à l'instant t ,

a_{0i} = terme constant pour l'individu i ,

a'_i = vecteur des k coefficients des k variables exogènes $a'_i = (a_{1i}, a_{2i}, \dots, a_{ki})$,

ε_{it} = terme d'erreur.

À partir de cette spécification générale nous pouvons envisager quatre possibilités.

Cas n° 1 : homogénéité totale.

Les constantes a_{0i} et les coefficients a'_i sont tous identiques pour tous les individus, nous avons $a_{0i} = a_0$ et $a'_i = a'$ pour toutes les valeurs de i . Le modèle ne comporte qu'une seule équation estimée sur $n = N \times T$ observations empilées par les MCO (ou les MCG selon la structure de la matrice des variances et covariances des erreurs).

Cas n° 2 : hétérogénéité totale.

Les constantes a_{0i} et les coefficients a'_i sont tous différents pour toutes les valeurs de i , la structure en panel est rejetée. Le modèle doit être estimé équation par équation pour les N équations (une équation par individu) par les MCO (ou les MCG selon la structure de la matrice des variances et covariances des erreurs).

Cas n° 3 : hétérogénéité des coefficients des variables explicatives et homogénéité des termes constants.

Les constantes a_{0i} sont toutes identiques ($a_{0i} = a_0$) pour les individus, mais les coefficients a'_i des variables explicatives sont différents pour chaque individu.

Comme au cas n° 2, le modèle doit être estimé sur les N équations (une équation par individu) par les MCO (ou les MCG selon la structure de la matrice des variances et covariances des erreurs).

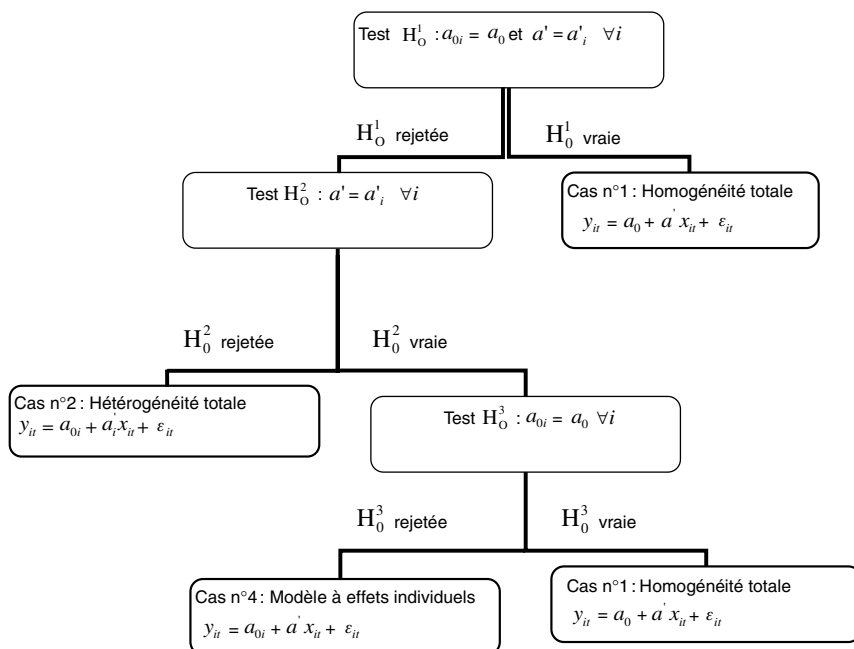
Cas n° 4 : hétérogénéité des termes constants et homogénéité des coefficients des variables explicatives – le modèle à effets individuels.

Les constantes a_{0i} sont différentes pour les individus, mais les coefficients a'_i des variables explicatives sont constants pour les individus ($a'_i = a'$). Ce modèle est appelé « modèle à effets individuels » et nous le présenterons à la section III.

II. Les tests d'homogénéité

A. Procédure séquentielle de tests

Le choix de la spécification (homogénéité, hétérogénéité) est donc très important. Afin de déterminer la structure du panel, Hsiao (1986) propose une procédure séquentielle de tests (cf. graphique 1) permettant de définir le cas dans lequel nous nous situons.



Graphique 1 – Procédure séquentielle des tests

B. Construction des tests

Les tests d'hypothèse sont construits à partir des statistiques de Fisher (test de Wald de restrictions sur les coefficients).

1) Test $H_0^1 : a_{0i} = a_0$ et $a' = a'_i \quad \forall i$

Ce test d'hypothèses jointes se ramène à un test de Fisher dont la statistique est donnée par :

$$F_1 = \frac{(SCR_{c1} - SCR)/(N - 1)(k + 1)}{SCR/(N \times T - N(k + 1))}$$

SCR_{c1} = somme des carrés des résidus du modèle contraint sous l'hypothèse H_0^1 , soit à estimer par les MCO le modèle en empilant toutes les observations. Le degré de liberté est égal à : ($N \times T$ = nombre total d'observations) $-(k + 1)$ = nombre de coefficients à estimer).

SCR = somme des carrés des résidus du modèle non contraint, elle est égale à la somme des N sommes des carrés des résidus des modèles estimés sur les T observations de chaque équation individuelle, soit $SCR = \sum_{i=1}^N SCR_i$. Le degré de liberté est donc la somme des N degrés de liberté de chaque équation estimée, soit $ddl = \sum_{i=1}^N (T - (k + 1)) = N \times T - N(k + 1)$.

Le degré de liberté du numérateur est donc égal à la différence des degrés de liberté de SCR_{c1} et SCR :

$$ddl_n = [(N \times T) - (k + 1)] - [(N \times T) - N(k + 1)] = (N - 1)(k + 1)$$

La statistique F_1 est à comparer à la valeur lue dans la table de Fisher aux degrés de liberté du numérateur et du dénominateur. Si $F_1 > F_{ddl_n; ddl_d}^\alpha$, nous rejetons l'hypothèse H_0^1 au seuil α .

2) Test $H_0^2 : a' = a'_i \quad \forall i$

Ce test d'hypothèses jointes se ramène au test de Fisher dont la statistique est donnée par :

$$F_2 = \frac{(SCR_{c2} - SCR)/((N - 1) \times k)}{SCR/(N \times T - N(k + 1))}$$

SCR_{c2} = somme des carrés des résidus du modèle contraint sous l'hypothèse H_0^2 , soit à estimer le modèle à effets fixes individuels. Le degré de liberté est égal

à $(N \times T = \text{nombre d'observations}) - (k + N = \text{nombre de coefficients à estimer})$, nous estimons k coefficients et N termes constants.

SCR = somme des carrés des résidus du modèle non contraint. Le degré de liberté du numérateur est donc égal à :

$$ddl_n = [(N \times T) - (k + N)] - [(N \times T) - N(k + 1)] = (N - 1) \times k$$

La statistique F_2 est à comparer à la valeur lue dans la table de Fisher aux degrés de liberté du numérateur et du dénominateur. Si $F_2 > F_{ddl_n; ddld}^\alpha$, nous rejetons l'hypothèse H_0^2 au seuil α .

3) Test $H_0^3 : a_{0i} = a_0 \quad \forall i$

Ce test d'hypothèses jointes se ramène au test de Fisher dont la statistique est donnée par :

$$F_3 = \frac{(SCR_{c1} - SCR_{c2}) / (N - 1)}{SCR_{c2} / (N \times (T - 1) - k)}$$

SCR_{c2} = somme des carrés des résidus du modèle contraint sous l'hypothèse H_0^2 .

SCR_{c1} = somme des carrés des résidus du modèle contraint sous l'hypothèse H_0^1 . Le degré de liberté du numérateur est donc égal à :

$$ddl_n = [(N \times T) - (k + 1)] - [(N \times T) - (k + N)] = N - 1$$

La statistique F_3 est à comparer à la valeur lue dans la table de Fisher aux degrés de liberté du numérateur et du dénominateur. Si $F_3 > F_{ddl_n; ddld}^\alpha$, nous rejetons l'hypothèse H_0^3 au seuil α .

Exercice n° 1

↓ fichier C13EX1 = données en panel
C13EX1-E = données empilées

Procédure séquentielle des tests d'homogénéité

Nous nous intéressons à l'évolution des dépenses d'éducation sur 25 ans (variable à expliquer notée y) pour neuf pays. Les facteurs explicatifs proposés sont :

- x_{1it} = évolution des dépenses militaires en milliers de dollars pour le pays i à l'année t ,
- x_{2it} = évolution du PIB (Produit intérieur brut) en milliers de dollars pour le pays i à l'année t .

Les variables sont toutes connues sur 25 ans.

Le modèle à étudier est donc :

$$y_{it} = a_{0i} + a_{1i}x_{1it} + a_{2i}x_{2it} + \varepsilon_{it} \text{ pour } i = 1, \dots, N \text{ et } t = 1, \dots, T$$

avec $N = 9$ pays et $T = 25$ années, le nombre total d'observations est donc $n = N \times T = 225$.

— y_{it} = évolution des dépenses d'éducation en milliers de dollars pour le pays i à l'année t ,

— ε_{it} = erreur de spécification pour le pays i à l'année t , ces erreurs répondent aux hypothèses classiques (homoscédasticité et indépendance temporelle) et sont indépendantes d'un individu à l'autre ($\text{Cov}(\varepsilon_{it}, \varepsilon_{jt}) = 0$ pour $i \neq j$).

Un extrait des données est présenté au Tableau 1.

Tableau 1 – Extrait des données de l'évolution en dépenses d'éducation, militaires et du PIB pour 9 pays sur 25 ans

Pays	Année	y_{it}	x_{1it}	x_{2it}
1	1	512	3,15	25,92
1	2	687	5,32	22,51
1
1	24	2 166	6,29	19,76
1	25	1 214	11,62	19,76
2	1	74	8,47	35,99
2	2	67	11,37	17,8
2
2	24	5	2,66	12,43
2	25	3	4,11	11,13
3	1	419	5,81	13,61
3	2	10238	7,02	3,27
3
3	24	1 939	12,83	14,27
3	25	2 254	16,7	14,14
...
9	23	5	6,05	8,77
9	24	28	4,84	12,96
9	25	170	3,87	12,83

On demande d'appliquer la stratégie des tests d'homogénéité afin de déterminer la structure du panel.

Solution

Nous commençons par tester l'hypothèse $H_0^1 : a_{0i} = a_0 \text{ et } a' = a'_i \quad \forall i$.

Soit à calculer : SCR_{c1} = somme des carrés des résidus du modèle contraint sous l'hypothèse H_0^1 , nous estimons par les MCO le modèle en empilant toutes les observations (fichier de données C13EX1-E).

Les résultats sont les suivants :

Dependent Variable: Y
Method: Least Squares
Included observations: 225

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	146.2167	138.0735	1.058977	0.2908
X1	29.32698	13.82629	2.121103	0.0350
X2	21.51258	11.47806	1.874234	0.0622

$SCR_{c1} = 233\,782\,555,085$; degré de libertés = $225 - 3 = 222$.

Puis à calculer SCR = somme des carrés des résidus du modèle non contraint, elle est égale à la somme des 9 sommes des carrés des résidus des modèles estimés sur les équations individuelles, soit $SCR = \sum_{i=1}^N SCR_i$.

Les résultats sont les suivants, le programme EvIEWS est en téléchargement (C13EX1.PRG) :

Pays	Somme des carrés des résidus
1	2 757 341
2	211 063,8
3	84 306 563
4	82 876 299
5	150 121,7
6	7 419 022
7	2 771 928
8	2 793 154
9	297 885,1
Total SCR	183 583 378

À titre d'exemple, l'estimation pour le deuxième pays (individu n° 2) est la suivante :

Dependent Variable: Y				
Sample: 26 50				
Included observations: 25				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	58.36178	47.94892	1.217166	0.2364
X1	- 0.180573	6.411318	- 0.028165	0.9778
X2	0.506402	2.722052	0.186037	0.8541
Sum squared resid	211 063,8	Degré de liberté = 25 - 3 = 22		

$SCR = 183\,583\,378,162$; degré de libertés = $9 \times (25-3) = 198$.

$$F_1 = \frac{(SCR_{c1} - SCR)/(N-1)(k+1)}{SCR/(N \times T - N(k+1))} = \frac{(233\,782\,555,08 - 183\,583\,378,162)/24}{183\,583\,378,162/198}$$

$$= \frac{2\,091\,632,372}{927\,188,778\,6} = 2,25, F_1 > F_{24;198}^{0,05} = 1,52, \text{ nous rejetons l'hypothèse } H_0^1.$$

Nous nous dirigeons vers la branche de gauche du graphique 1, soit à tester : H_0^2 :
 $a' = a'_i \quad \forall i$.

Nous calculons SCR_{c2} = somme des carrés des résidus du modèle contraint sous l'hypothèse H_0^2 , soit à estimer le modèle à effets fixes individuels (la méthode d'estimation sera étudiée à la section III, fichier de données C13EX1).

Dependent Variable: Y?				
Method: Pooled Least Squares				
Sample: 1 25				
Included observations: 25				
Cross-sections included: 9				
Total pool (balanced) observations: 225				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	654.9109	181.7605	3.603153	0.0004
X1?	- 49.50377	23.64297	- 2.093805	0.0375
X2?	22.17142	11.02105	2.011733	0.0455
Fixed Effects (Cross)				
_1-C	49.20967			
_2-C	- 438.3058			
_3-C	1 157.176			
_4-C	102.6206			
_5-C	- 409.4617			
_6-C	794.9591			
_7-C	- 273.2759			
_8-C	- 501.3735			
_9-C	- 481.5489			
$SCR_{c2} = 195\,511\,374,365$				

Le degré de liberté de ce modèle est de $225 - 2 - 9 = 214$, car la constante $a_0 = C = 654,91$ est la moyenne des 9 coefficients des effets fixes.

$$F_2 = \frac{(SCR_{c2} - SCR)/(N - 1) \times k}{SCR/(N \times T - N(k + 1))} = \frac{(233\,782\,555,08 - 195\,511\,374,365)/16}{183\,583\,378,162/198}$$

$$= \frac{74\,599,762\,7}{927\,188,778\,6} = 0,804, F_2 > F_{7;198}^{0,05} = 1,90, \text{ nous acceptons l'hypothèse } H_0^2. \text{ Nous}$$

nous dirigeons donc vers la branche de droite du graphique 1.

$$\text{Test } H_0^3 : a_{0i} = a_0 \quad \forall i$$

Soit à calculer la statistique : $F_3 = \frac{(SCR_{c1} - SCR_{c2})/(N - 1)}{SCR_{c2}/(N \times (T - 1) - k)}$ dont nous connaissons tous les éléments.

$$F_3 = \frac{(233\,782\,555,08 - 195\,511\,374,365)/8}{195\,511\,374,365/213} = \frac{4\,783\,897,589}{917\,893,776\,4} = 5,24,$$

$F_3 > F_{8;213}^{0,05} \approx 1,90$ nous rejetons l'hypothèse H_0^3 .

Le panel à une structure à effets individuels.

III. Spécifications et estimations des modèles à effets individuels

Les modèles à effets individuels supposent que les modèles estimés ne diffèrent par individu que par la valeur de la constante $a_{0i} = a_0 + a_i$. Nous allons distinguer deux cas : les modèles à effets fixes (l'effet individuel est constant au cours du temps) et les modèles à effets aléatoires (le terme constant est une variable aléatoire).

Les modèles à effets individuels peuvent être transposés au modèle à effets temporels, la constante s'écrit alors $a_{0t} = a_0 + a_t$ (cf. Sevestre P., 2002).

A. Le modèle à effets fixes individuels

Le modèle en données de panel à effets fixes individuels peut s'écrire de la manière suivante :

$$y_{it} = a_{0i} + a'x_{it} + \varepsilon_{it}$$

y_{it} = variable endogène observée pour l'individu i à la période t ,

x_{it} = vecteur des k variables exogènes $x'_{it} = (x_{1it}, x_{2it}, \dots, x_{kit})$; x_{kit} est donc la valeur observée pour la k - i^e variable exogène pour l'individu i à l'instant t ,

a_{0i} = terme constant pour l'individu i ,

a' = vecteur des k coefficients des k variables exogènes, $a' = (a_1, a_2, \dots, a_k)$,

ε_{it} = terme d'erreur.

La méthode d'estimation des paramètres va dépendre de la structure des termes d'erreurs :

- Si les erreurs sont homoscédastiques, non autocorrélées dans la dimension temporelle $\text{Cov}(\varepsilon_{it}, \varepsilon_{it'}) = 0$ pour $t \neq t'$ et dans la dimension individuelle $\text{Cov}(\varepsilon_{it}, \varepsilon_{jt}) = 0$ pour $i \neq j$: méthode des Moindres Carrés Ordinaires (MCO) sur les variables indicatrices (en anglais LSDV = *Least Square Dummy Variable*) ou sur les estimateurs *Within*.
- Si les erreurs sont hétéroscéastiques et/ou autocorrélées dans la dimension temporelle mais indépendantes dans la dimension individuelle : méthode des Moindres Carrés Généralisés (MCG) sur les variables indicatrices (LSDV) ou sur les estimateurs *Within*.

1) Estimateur LSDV

L'estimateur LSDV consiste à appliquer la méthode des MCO sur le modèle avec variables indicatrices spécifiques pour chacun des N individus. Nous construisons donc N variables indicatrices tel que : $D_i = 1$ pour l'individu i et 0 pour les autres.

Le modèle s'écrit : $y_{it} + a_0 + a_{01}D_1 + a_{02}D_2 + \dots + a_{0N}D_N + a'x_{it} + \varepsilon_{it}$

Pratiquement, nous estimons le modèle sans la constante¹ a_0 : $y_{it} = a_1D_1 + a_2D_2 + \dots + a_ND_N + a'x_{it} + \varepsilon_{it}$ par les MCO ou les MCG si les erreurs sont hétéroscéastiques ou/et autocorrélées.

Nous pouvons ensuite calculer les coefficients $a_{0i} = a_0 + a_i$ du modèle initial, la valeur de la constante a_0 étant égale à la moyenne des coefficients a_i estimés.

2) Estimateur *Within*

L'estimateur *Within* (estimateur intra-individuel) consiste à centrer préalablement toutes les variables à expliquer et explicatives sur leurs moyennes individuelles et à appliquer la méthode des MCO (ou MCG si les erreurs sont hétéroscéastiques ou autocorrélées) sur le modèle ainsi transformé :

$$(y_{it} - \bar{y}_i) = a'(x_{it} - \bar{x}_i) + \varepsilon_{it} \text{ pour } i = 1, \dots, N \text{ et } t = 1, \dots, T.$$

Après l'estimation des coefficients a' les coefficients fixes individuels a_{0i} s'obtiennent par les relations : $\hat{a}_{01} = \bar{y}_1 - \hat{a}'\bar{x}_1$; $\hat{a}_{02} = \bar{y}_2 - \hat{a}'\bar{x}_2$; ... ; $\hat{a}_{0N} = \bar{y}_N - \hat{a}'\bar{x}_N$

Nous calculons ensuite les coefficients $a_{0i} = a_0 + a_i$ du modèle initial comme précédemment.

Cette méthode d'estimation conduit exactement aux mêmes résultats que la méthode LSDV.

1. Sinon, il y aurait alors colinéarité entre le vecteur du terme constant (a_0) et la somme des N variables indicatrices, la matrice $(X'X)$ serait alors singulière (non inversible).

B. Le modèle à effets aléatoires

1) Présentation du modèle à effets aléatoires

Le modèle à effets aléatoires suppose que la relation entre la variable à expliquer et les variables explicatives ne soit plus fixe mais aléatoire, l'effet individuel n'est plus un paramètre fixe a_{0i} mais une variable aléatoire.

Le terme d'erreur se décompose de la manière suivante : $\varepsilon_{it} = a_{0i} + \lambda_t + v_{it}$ d'où le nom donné aussi à ce type de modèle : modèle à erreur composée (*error components model*).

Les a_{0i} représentent les effets individuels aléatoires, la variable aléatoire λ_t représente les effets temporels identiques pour tous les individus et enfin v_{it} est un terme d'erreur qui est orthogonal aux effets individuels et temporels.

En outre, nous faisons les hypothèses suivantes concernant les variables aléatoires a_{0i} , λ_t et v_{it} : espérances nulles, homoscedasticités, indépendance temporelle et orthogonalité entre les trois composantes.

Dans le cas simple où l'effet temporel n'existe pas ($\lambda_t = 0$), le modèle à effets aléatoires individuels s'écrit alors : $y_{it} = a_0 + a'x_{it} + \varepsilon_{it}$ avec $\varepsilon_{it} = a_{0i} + v_{it}$.

Ou encore $y_{it} = a_0 + a_{0i} + a'x_{it} + v_{it}$. Le coefficient a_0 est un coefficient fixe identique pour tous les individus.

2) Estimation du modèle à effets aléatoires

La méthode d'estimation adéquate est celle des MCG car la composante aléatoire a_{0i} est présente dans ε_{it} et $\varepsilon_{it'}$, d'où $\text{Cov}(\varepsilon_{it}, \varepsilon_{it'}) \neq 0$. Nous avons donc une autocorrélation des erreurs (cf. Sevestre 2002).

On démontre que l'estimateur des MCG se ramène à une moyenne pondérée des estimateurs *Within* (présenté précédemment) et *Between*. L'estimateur *Between* (estimateur inter-individuel noté \widehat{a}_{Bet}) est l'estimateur des MCO appliqué sur le modèle dont les données sont les N moyennes individuelles des variables à expliquer et explicatives :

$$\bar{y}_i = a_0 + a_{0i} + a'\bar{x}_i + \bar{v}_i \text{ pour } i = 1, \dots, N.$$

L'estimateur des MCG¹ est donné par : $\widehat{a}_{MCG} = \Delta \widehat{a}_{Bet} + (1 - \Delta) \widehat{a}_{LSDV}$

La valeur des poids Δ , une matrice de dimension (k, k) , est inversement proportionnelle à la matrice des covariances de \widehat{a}_{Bet} .

L'introduction d'effets individuels aléatoires permet donc de combiner une spécification intermédiaire entre le modèle sans effet individuel et le modèle avec effets fixes. La structure du panel n'est ni totalement homogène, ni totalement hétérogène.

1. On démontre aussi que l'estimateur des MCO sur les données empilées se ramène aussi à une moyenne pondérée des estimateurs *Within* et *Between*.

C. Effets fixes ou effets aléatoires ? Le test d'Hausman

Le test d'Hausman (cf. chapitre 5) est un test de spécification qui permet de déterminer si les coefficients des deux estimations (fixes et aléatoires) sont statistiquement différents.

Sous l'hypothèse nulle H_0 d'orthogonalité entre les variables explicatives et le terme d'erreur du modèle à effets aléatoires, les deux estimateurs – *Within* et MCG – sont des estimateurs non biaisés et dans ce cas il ne doit pas y avoir de différence significative entre les estimations *Within* et MCG des différents coefficients. La méthode des MCG est alors retenue : le modèle est à effets aléatoires.

Soit le test d'hypothèses :

$H_0 : \widehat{a}_{LSDV} - \widehat{a}_{MCG} = 0 \rightarrow$ le modèle est à effets aléatoires


$H_1 : \widehat{a}_{LSDV} - \widehat{a}_{MCG} \neq 0 \rightarrow$ le modèle est à effets fixes

Nous calculons la statistique :

$$H = (\widehat{a}_{LSDV} - \widehat{a}_{MCG})' [\text{Var}(\widehat{a}_{LSDV}) - \text{Var}(\widehat{a}_{MCG})]^{-1} (\widehat{a}_{LSDV} - \widehat{a}_{MCG})$$

La statistique H est distribuée selon un chi-deux à k degré de liberté. Si $H > \chi^2(k)$ pour un seuil α % fixé, nous rejetons l'hypothèse H_0 , l'estimateur LSDV (*Within*) est non biaisé, nous rejetons alors la spécification à effets aléatoires et nous choisissons un modèle à effets fixes individuels.

Exercice n° 2

 fichiers C13EX1, C13EX1-E

Estimation du modèle à effets fixes individuels et à effets aléatoires individuels, estimateur *Between*, test d'Hausman

En reprenant le modèle et les données de l'exercice n° 1, on demande :

- 1) D'estimer le modèle à effets fixes individuels.
- 2) D'estimer le modèle à effets aléatoires individuels.
- 3) De calculer l'estimateur *Between*.
- 4) De procéder au test d'Hausman.

Solution

1) Nous créons les neuf variables indicatrices relatives à chacun des pays et nous procédons à l'estimation par les MCO du modèle (hors terme constant) sur les données empiriques (fichier C13EX1-E) :

$$y_t = a_1x_{1t} + a_2x_{2t} + b_1D_{1t} + b_2D_{2t} + \dots + b_9D_{9t} + \varepsilon_t$$

Soit le résultat suivant :

Dependent Variable: Y				
Sample: 1 225				
Included observations: 225				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
X1	- 49.50377	23.64297	- 2.093805	0.0375
X2	22.17142	11.02105	2.011733	0.0455
DUM1	704.1205	240.8149	2.923907	0.0038
DUM2	216.6050	250.8295	0.863555	0.3888
DUM3	1 812.087	340.8004	5.317151	0.0000
DUM4	757.5315	227.2174	3.333951	0.0010
DUM5	245.4491	232.6545	1.054994	0.2926
DUM6	1 449.870	429.0331	3.379390	0.0009
DUM7	381.6349	224.4928	1.699987	0.0906
DUM8	153.5374	215.6064	0.712119	0.4772
DUM9	173.3620	234.4385	0.739477	0.4604

Nous calculons maintenant les coefficients $a_i = a_{0i} - a_0$:

Pays	a_{0i}	a_i
1	704,12	49,21
2	216,61	- 438,31
3	1 812,09	1 157,18
4	757,53	102,62
5	245,45	- 409,46
6	1 449,87	794,96
7	381,63	- 273,28
8	153,54	- 501,37
9	173,36	- 481,55
Moyenne = a_0	654,91	

Nous retrouvons bien exactement les mêmes résultats que lors du calcul de la SCR_{c2} = somme des carrés des résidus du modèle contraint sous l'hypothèse H_0^2 de l'exercice n° 1 (fichier de données C13EX1, instruction Eviews : panel.ls(cx =f) y? x1? x2?).

2) L'estimation du modèle à effets aléatoires individuels est effectuée par l'instruction Eviews : panel.ls(cx =R) y? x1? x2?

Les résultats sont les suivants :

Dependent Variable: Y?				
Method: Pooled EGLS (Cross-section random effects)				
Sample: 1 25				
Included observations: 25				
Cross-sections included: 9				
Total pool (balanced) observations: 225				
Cross-section SUR (PCSE) standard errors & covariance (d.f. corrected)				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	400.3119	208.0501	1.924112	0.0556
X1?	− 9.907703	22.74786	− 0.435544	0.6636
X2?	21.70457	11.17599	1.942071	0.0534
Random Effect (Cross)				
_1-C	76.20577			
_2-C	− 326.3516			
_3-C	729.8104			
_4-C	171.7304			
_5-C	− 279.7010			
_6-C	327.8098			
_7-C	− 109.6655			
_8-C	− 269.7585			
_9-C	− 320.0798			

Nous observons que le modèle à effets aléatoires conduit à des estimations des coefficients très différents du modèle à effets fixes, cela peut provenir d'une corrélation entre les effets individuels et les variables explicatives qui introduit un biais d'estimation lors de l'application de la méthode des MCG.

3) L'estimateur *Between* consiste à estimer par les MCO le modèle suivant :

$$\bar{y}_i = a_0 + a_{0i} + a'\bar{x}_i + \bar{v}_i$$

Les moyennes calculées sur les 25 années pour les 9 individus sont présentées au tableau 2.

Tableau 2 – Moyennes sur les 25 ans par pays

Pays	\bar{y}_i	\bar{x}_{1i}	\bar{x}_{2i}
1	602,00	5,26	7,13
2	60,64	6,15	6,69
3	1 330,56	11,76	4,54
4	769,68	3,49	8,33
5	70,20	5,31	3,96
6	842,72	15,85	7,99
7	403,88	3,22	8,20
8	175,00	2,70	7,00
9	81,44	4,85	6,68

Les résultats de l'estimation *Between* (application des MCO sur les données du tableau 2) sont les suivants avec pour variable à expliquer \bar{y}_i :

	Coefficients	Écart-type	t-Statistique	Probabilité
Constante	- 125,65	694,52	- 0,183	0,86
\bar{x}_{1i}	62,93	32,29	1,94	0,09
\bar{x}_{2i}	29,41	92,22	0,318	0,76

4) Nous calculons la statistique d'Hausman :

$$H = (\widehat{a}_{LSDV} - \widehat{a}_{MCG})' [\text{Var}(\widehat{a}_{LSDV}) - \text{Var}(\widehat{a}_{MCG})]^{-1} (\widehat{a}_{LSDV} - \widehat{a}_{MCG})$$

Avec :

$$\widehat{a}_{LSDV} = \begin{bmatrix} -49,503 \\ 22,171 \end{bmatrix}; \widehat{a}_{MCG} = \begin{bmatrix} -9,907 \\ 21,704 \end{bmatrix}; \text{Var}(\widehat{a}_{LSDV}) = \begin{bmatrix} 558,99 & -2,30 \\ -2,30 & 121,46 \end{bmatrix};$$

$$\text{Var}(\widehat{a}_{MCG}) = \begin{bmatrix} 361,49 & 0,566 \\ 0,566 & 119,71 \end{bmatrix}$$

$$(\widehat{a}_{LSDV} - \widehat{a}_{MCG}) = \begin{bmatrix} -39,59 \\ 0,466 \end{bmatrix}; [\text{Var}(\widehat{a}_{LSDV}) - \text{Var}(\widehat{a}_{MCG})] = \begin{bmatrix} 197,50 & -2,869 \\ -2,869 & 1,751 \end{bmatrix}$$

$$H = \begin{bmatrix} -39,59 & 0,466 \end{bmatrix} \begin{bmatrix} 0,00518 & 0,8499 \\ 0,8499 & 0,585 \end{bmatrix} \begin{bmatrix} -39,59 \\ 0,466 \end{bmatrix} = 7,94 > \chi^2(2) \text{ pour un seuil de } 5\% = 5,99.$$

Nous rejetons l'hypothèse H_0 , le modèle est donc à effets individuels fixes.

Le programme Eviews en téléchargement (C13EX2) permet de calculer directement la statistique d'Hausman.

Liste des exercices

Chapitre 1

<i>Exercice 1</i> — Calcul d'un coefficient de corrélation	10
--	----

Chapitre 2

<i>Exercice 1</i> — Génération d'une consommation aléatoire	15
<i>Exercice 2</i> — Estimations des coefficients de régression	20
<i>Exercice 3</i> — Test de coefficient et intervalle de confiance	29
<i>Exercice 4</i> — Équivalence des tests et analyse de la variance	36
<i>Exercice 5</i> — Préviation dans un modèle de régression simple	41
<i>Exercice 6</i> — Comparaison de coefficients de régression	43
<i>Exercice 7</i> — Apprendre à manipuler les formules	44
<i>Exercice 8</i> — Apprendre à utiliser les formules	45

Chapitre 3

<i>Exercice 1</i> — Mise sous forme matricielle d'un modèle et calcul des coefficients de régression	55
<i>Exercice 2</i> — Tests statistiques sur les coefficients et la variance de l'erreur : calcul des leviers et des résidus studentisés	64
<i>Exercice 3</i> — Tests à partir de l'analyse de la variance	69
<i>Exercice 4</i> — Détection et correction de valeurs anormales par variable indicatrice	76
<i>Exercice 5</i> — Intégration d'une variable qualitative	78
<i>Exercice 6</i> — Étude de saisonnalité par variables indicatrices	79
<i>Exercice 7</i> — Préviation à partir du modèle linéaire	83
<i>Exercice 8</i> — Tests de stabilité structurelle, test de spécification de Ramsey	87
<i>Exercice 9</i> — Estimation d'un modèle et tests de validité à partir de la connaissance de quelques éléments statistiques	90
<i>Exercice 10</i> — Test d'une combinaison linéaire de coefficients de régression (test de contrainte linéaire)	93
<i>Exercice 11</i> — Reconstitution d'informations manquantes	95

<i>Exercice 12</i> — Modèle à deux régimes : estimation par variable muette ou par analyse de la variance	96
<i>Exercice 13</i> — Estimation d'une fonction de production : impact de la privatisation	97
<i>Exercice 14</i> — Fonction économétrique d'investissement	101

Chapitre 4

<i>Exercice 1</i> — Calcul des coefficients de corrélation partielle du premier ordre et du deuxième ordre	109
<i>Exercice 2</i> — Relation entre coefficients de corrélation simple, partielle et multiple	113
<i>Exercice 3</i> — Conséquences de la multicollinéarité	114
<i>Exercice 4</i> — Tests de multicollinéarité	117
<i>Exercice 5</i> — Méthodes de sélection de variables explicatives	121

Chapitre 5

<i>Exercice 1</i> — Tests d'indépendance des erreurs	131
<i>Exercice 2</i> — Procédures d'estimation en cas d'autocorrélation des erreurs d'ordre 1	139
<i>Exercice 3</i> — Correction de l'hétéroscédasticité lorsque les données représentent des moyennes	144
<i>Exercice 4</i> — Tests de détection d'une hétéroscédasticité et correction	147
<i>Exercice 5</i> — Test d'exogénéité d'Hausman, modèle à erreurs sur les variables : la technique des variables instrumentales	158
<i>Exercice 6</i> — Préviation de part de marché	160
<i>Exercice 7</i> — Un peu de réflexion économétrique...	162

Chapitre 6

<i>Exercice 1</i> — Estimation d'une fonction de production de type Cobb-Douglas	165
<i>Exercice 2</i> — Estimation des paramètres d'un modèle de diffusion Logistique et de Gompertz	172
<i>Exercice 3</i> — Estimation des paramètres d'une fonction de production de type CES	174

Chapitre 7

<i>Exercice 1</i> — Estimation des paramètres d'un modèle autorégressif à erreurs liées	181
<i>Exercice 2</i> — Détermination du nombre de retards dans un modèle à retards échelonnés	186
<i>Exercice 3</i> — Estimation des coefficients à l'aide d'un polynôme (méthode d'Almon)	191
<i>Exercice 4</i> — Estimation des coefficients d'un modèle selon une spécification des retards de Koyck et une distribution de Pascal	195
<i>Exercice 5</i> — Estimation d'un modèle d'ajustement partiel et d'anticipations adaptatives	200
<i>Exercice 6</i> — Modélisation du prix d'une matière première	201
<i>Exercice 7</i> — Un modèle d'évaluation de la politique marketing	203
<i>Exercice 8</i> — Une comparaison des courbes de Phillips française et américaine	208

Chapitre 8

<i>Exercice 1</i> — Comparaison des résultats par moindres carrés ordinaires, moindres carrés indirects et doubles moindres carrés	225
<i>Exercice 2</i> — Le problème de l'identification lorsque la structure d'un modèle change	229
<i>Exercice 3</i> — Le modèle de Klein	230

Chapitre 9

<i>Exercice 1</i> — Exemple d'application des tests de racine unitaire au CAC40	252
<i>Exercice 2</i> — Génération de processus ARMA et analyse des corrélogrammes	260
<i>Exercice 3</i> — Analyse par la méthode de Box et Jenkins et prévision de trois processus aléatoires	264

Chapitre 10

<i>Exercice 1</i> — Recherche des conditions de stationnarité d'un modèle VAR	278
<i>Exercice 2</i> — Spécification, estimation et prévision d'un modèle VAR	281
<i>Exercice 3</i> — Analyse d'une fonction de réponse impulsionnelle et décomposition de la variance	290
<i>Exercice 4</i> — Tests de causalité de Granger et de Sims	293

Chapitre 11

<i>Exercice 1</i> — Test de cointégration et estimation d'un modèle à correction d'erreur	304
<i>Exercice 2</i> — Tests de cointégration et estimation d'un modèle vectoriel à correction d'erreur	314

Chapitre 12

<i>Exercice 1</i> — Estimation des modèles Probit et Logit binaires explicatifs des facteurs de la réussite en Licence	326
<i>Exercice 2</i> — Estimation d'un modèle à choix multiples de prévision des ventes	332
<i>Exercice 3</i> — Prévision de la demande d'électricité pour un fournisseur à capacité limitée	340

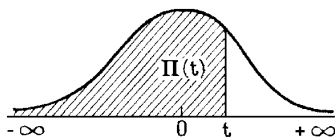
Chapitre 13

<i>Exercice 1</i> — Procédure séquentielle des tests d'homogénéité	351
<i>Exercice 2</i> — Estimation du modèle à effets fixes individuels et à effets aléatoires individuels, estimateur <i>Between</i> , test d'Hausman	358

T**ables statistiques**

1. Table de la loi de Laplace-Gauss
2. Table de la loi de Student
3. Table de la loi du Chi-Deux
4. Table de la loi de Fisher-Snedecor
5. Table de la loi de Fisher-Snedecor (suite)
6. Table de Durbin-Watson
7. Tables de Dickey-Fuller

1. TABLE DE LA LOI DE LAPLACE-GAUSS
(Probabilité de trouver une valeur inférieure à x)



$$\pi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-t^2/2} dt .$$

t	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7290	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9779	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

TABLE POUR LES GRANDES VALEURS DE x

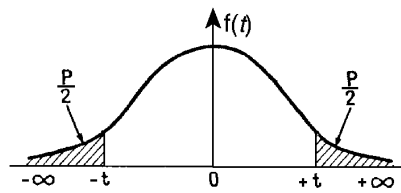
x	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,8	4,0	4,5
$F(x)$	0,998 65	0,999 04	0,999 31	0,999 52	0,999 66	0,999 76	0,999 841	0,999 928	0,999 968	0,999 997

Nota. — La table donne les valeurs de $F(x)$ pour x positif. Lorsque x est négatif, il faut prendre le complément à l'unité de la valeur lue dans la table.

Exemples : pour $x = 1,37$ $F(x) = 0,9147$
pour $x = -1,37$ $F(x) = 0,0853$

2. TABLE DE LA LOI DE STUDENT

Valeurs de T ayant la probabilité P d’être dépassées en valeur absolue

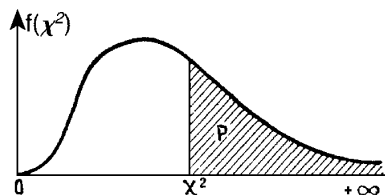


ν	$P = 0,90$	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750
∞	0,12566	0,25335	0,38532	0,52440	0,67449	0,84162	1,03643	1,28155	1,64485	1,95996	2,32634	2,57582

Nota. — ν est le nombre de degrés de liberté.

3. TABLE DE LA LOI DU CHI-DEUX

Valeurs de χ^2 ayant la probabilité P d'être dépassées



ν	$P = 0,90$	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01
1	0,0158	0,0642	0,148	0,455	1,074	1,642	2,706	3,841	5,412	6,635
2	0,211	0,446	0,713	1,386	2,408	3,219	4,605	5,991	7,824	9,210
3	0,584	1,005	1,424	2,366	3,665	4,642	6,251	7,815	9,837	11,345
4	1,064	1,649	2,195	3,357	4,878	5,989	7,779	9,488	11,668	13,277
5	1,610	2,343	3,000	4,351	6,064	7,289	9,236	11,070	13,388	15,086
6	2,204	3,070	3,828	5,348	7,231	8,558	10,645	12,592	15,033	16,812
7	2,833	3,822	4,671	6,346	8,383	9,803	12,017	14,067	16,662	18,475
8	3,490	4,594	5,527	7,344	9,524	11,030	13,362	15,507	18,168	20,090
9	4,168	5,380	6,393	8,343	10,656	12,242	14,684	16,919	19,679	21,666
10	4,865	6,179	7,267	9,342	11,781	13,442	15,987	18,307	21,161	23,209
11	5,578	6,989	8,148	10,341	12,899	14,631	17,275	19,675	22,618	24,725
12	6,304	7,807	9,034	11,340	14,011	15,812	18,549	21,026	24,054	26,217
13	7,042	8,634	9,926	12,340	15,119	16,985	19,812	22,362	25,472	27,688
14	7,790	9,467	10,821	13,339	16,222	18,151	21,064	23,685	26,873	29,141
15	8,547	10,307	11,721	14,339	17,322	19,311	22,307	24,996	28,259	30,578
16	9,312	11,152	12,624	15,338	18,418	20,465	23,542	26,296	29,633	32,000
17	10,085	12,002	13,531	16,338	19,511	21,615	24,769	27,587	30,995	33,409
18	10,865	12,857	14,440	17,338	20,601	22,760	25,989	28,869	32,346	34,805
19	11,651	13,716	15,352	18,338	21,689	23,900	27,204	30,144	33,687	36,191
20	12,443	14,578	16,266	19,337	22,775	25,038	28,412	31,410	35,020	37,566
21	13,240	15,445	17,182	20,337	23,858	26,171	29,615	32,671	36,343	38,932
22	14,041	16,314	18,101	21,337	24,939	27,301	30,813	33,924	37,659	40,289
23	14,848	17,187	19,021	22,337	26,018	28,429	32,007	35,172	38,968	41,638
24	15,659	18,062	19,943	23,337	27,096	29,553	33,196	36,415	40,270	42,980
25	16,473	18,940	20,867	24,337	28,172	30,675	34,382	37,652	41,566	44,314
26	17,292	19,820	21,792	25,336	29,246	31,795	35,563	38,885	42,856	45,642
27	18,114	20,703	22,719	26,336	30,319	32,912	36,741	40,113	44,140	46,963
28	18,939	21,588	23,647	27,336	31,391	34,027	37,916	41,337	45,419	48,278
29	19,768	22,475	24,577	28,336	32,461	35,139	39,087	42,557	46,693	49,588
30	20,599	23,364	25,508	29,336	33,530	36,250	40,256	43,773	47,962	50,892

Lorsque $\nu > 30$, on peut admettre que la quantité $\sqrt{2} \chi^2 - \sqrt{2\nu - 1}$ suit la loi normale réduite.

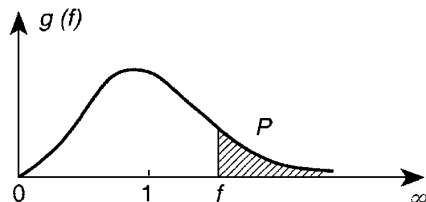
Exemple :

Calculez la valeur de χ^2 correspondant à une probabilité $P = 0,10$ de dépassement lorsque $\nu = 41$. À l'aide de la table 1, on calcule, pour $P = 0,10$, $x = 1,2816$.

$$\text{D'où : } \chi^2 = \frac{[x + \sqrt{2\nu - 1}]^2}{2} = \frac{1}{2}[1,2816 + \sqrt{82 - 1}]^2 = \frac{1}{2}(10,2816)^2 = 52,85.$$

4. TABLE DE LA LOI DE FISHER-SNEDECOR

Valeurs de F ayant la probabilité P d'être dépassées ($F = s_1^2/s_2^2$)

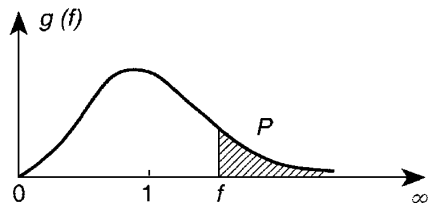


ν_2	$\nu_1 = 1$		$\nu_1 = 2$		$\nu_1 = 3$		$\nu_1 = 4$		$\nu_1 = 5$	
	$P = 0,05$	$P = 0,01$	$P = 0,05$	$P = 0,01$	$P = 0,05$	$P = 0,01$	$P = 0,05$	$P = 0,01$	$P = 0,05$	$P = 0,01$
1	161,4	4 052	199,5	4 999	215,7	5 403	224,6	5 625	230,2	5 764
2	18,51	98,49	19,00	99,00	19,16	99,17	19,25	99,25	19,30	99,30
3	10,13	34,12	9,55	30,81	9,28	29,46	9,12	28,71	9,01	28,24
4	7,71	21,20	6,94	18,00	6,59	16,69	6,39	15,98	6,26	15,52
5	6,61	16,26	5,79	13,27	5,41	12,06	5,19	11,39	5,05	10,97
6	5,99	13,74	5,14	10,91	4,76	9,78	4,53	9,15	4,39	8,75
7	5,59	12,25	4,74	9,55	4,35	8,45	4,12	7,85	3,97	7,45
8	5,32	11,26	4,46	8,65	4,07	7,59	3,84	7,01	3,69	6,63
9	5,12	10,56	4,26	8,02	3,86	6,99	3,63	6,42	3,48	6,06
10	4,96	10,04	4,10	7,56	3,71	6,55	3,48	5,99	3,33	5,64
11	4,84	9,65	3,98	7,20	3,59	6,22	3,36	5,67	3,20	5,32
12	4,75	9,33	3,88	6,93	3,49	5,95	3,26	5,41	3,11	5,06
13	4,67	9,07	3,80	6,70	3,41	5,74	3,18	5,20	3,02	4,86
14	4,60	8,86	3,74	6,51	3,34	5,56	3,11	5,03	2,96	4,69
15	4,54	8,68	3,68	6,36	3,29	5,42	3,06	4,89	2,90	4,56
16	4,49	8,53	3,63	6,23	3,24	5,29	3,01	4,77	2,85	4,44
17	4,45	8,40	3,59	6,11	3,20	5,18	2,96	4,67	2,81	4,34
18	4,41	8,28	3,55	6,01	3,16	5,09	2,93	4,58	2,77	4,25
19	4,38	8,18	3,52	5,93	3,13	5,01	2,90	4,50	2,74	4,17
20	4,35	8,10	3,49	5,85	3,10	4,94	2,87	4,43	2,71	4,10
21	4,32	8,02	3,47	5,78	3,07	4,87	2,84	4,37	2,68	4,04
22	4,30	7,94	3,44	5,72	3,05	4,82	2,82	4,31	2,66	3,99
23	4,28	7,88	3,42	5,66	3,03	4,76	2,80	4,26	2,64	3,94
24	4,26	7,82	3,40	5,61	3,01	4,72	2,78	4,22	2,62	3,90
25	4,24	7,77	3,38	5,57	2,99	4,68	2,76	4,18	2,60	3,86
26	4,22	7,72	3,37	5,53	2,98	4,64	2,74	4,14	2,59	3,82
27	4,21	7,68	3,35	5,49	2,96	4,60	2,73	4,11	2,57	3,78
28	4,20	7,64	3,34	5,45	2,95	4,57	2,71	4,07	2,56	3,75
29	4,18	7,60	3,33	5,42	2,93	4,54	2,70	4,04	2,54	3,73
30	4,17	7,56	3,32	5,39	2,92	4,51	2,69	4,02	2,53	3,70
40	4,08	7,31	3,23	5,18	2,84	4,31	2,61	3,83	2,45	3,51
60	4,00	7,08	3,15	4,98	2,76	4,13	2,52	3,65	2,37	3,34
120	3,92	6,85	3,07	4,79	2,68	3,95	2,45	3,48	2,29	3,17
∞	3,84	6,64	2,99	4,60	2,60	3,78	2,37	3,32	2,21	3,02

Nota. — s_1^2 est la plus grande des deux variances estimées, avec ν_1 degrés de liberté.

5. TABLE DE LA LOI DE FISHER-SNEDECOR (suite)

Valeurs de F ayant la probabilité P d'être dépassées ($F = s_1^2/s_2^2$)



ν_2	$\nu_1 = 6$		$\nu_1 = 8$		$\nu_1 = 12$		$\nu_1 = 24$		$\nu_1 = \infty$	
	$P = 0,05$	$P = 0,01$	$P = 0,05$	$P = 0,01$	$P = 0,05$	$P = 0,01$	$P = 0,05$	$P = 0,01$	$P = 0,05$	$P = 0,01$
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	27,49	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	5,63	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	2,53	3,80	2,35	3,43	2,13	3,00
15	2,79	4,32	2,64	4,00	2,48	3,67	2,29	3,29	2,07	2,87
16	2,74	4,20	2,59	3,89	2,42	3,55	2,24	3,18	2,01	2,75
17	2,70	4,10	2,55	3,79	2,38	3,45	2,19	3,08	1,96	2,65
18	2,66	4,01	2,51	3,71	2,34	3,37	2,15	3,00	1,92	2,57
19	2,63	3,94	2,48	3,63	2,31	3,30	2,11	2,92	1,88	2,49
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	1,90	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
120	2,17	2,96	2,01	2,66	1,83	2,34	1,61	1,95	1,25	1,38
∞	2,09	2,80	1,94	2,51	1,75	2,18	1,52	1,79	1,00	1,00

Nota. — s_1^2 est la plus grande des deux variances estimées, avec ν_1 degrés de liberté.

6. TABLE DE DURBIN-WATSON
Risque $\alpha = 5 \%$

n	$k = 1$		$k = 2$		$k = 3$		$k = 4$		$k = 5$	
	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
15	1,08	1,36	0,95	1,54	0,82	1,75	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	1,74	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	1,76	1,57	1,78

k est le nombre de variables exogènes (constante exclue).
 n est la taille de l'échantillon.

7. TABLES DE DICKEY-FULLER¹

Modèle [1] sans tendance et sans terme constant

Modèle [2] sans tendance et avec terme constant

Modèle [3] avec tendance et avec terme constant

Tables de la distribution du t_{ϕ_1}

Nombre observations	Probabilités								
n	0,01	0,025	0,05	0,10	0,90	0,95	0,975	0,99	
25	-2,66	-2,26	-1,95	-1,60	0,92	1,33	1,70	2,16	Modèle [1]
50	-2,62	-2,25	-1,95	-1,61	0,91	1,31	1,66	2,08	
100	-2,60	-2,4	-1,95	-1,61	0,91	1,29	1,64	2,03	
250	-2,58	-2,23	-1,95	-1,62	0,89	1,29	1,63	2,01	
500	-2,58	-2,23	-1,95	-1,62	0,89	1,28	1,62	2,00	
∞	-2,58	-2,23	-1,95	-1,62	0,89	1,28	1,62	2,00	
25	-3,75	-3,33	-3,00	-2,63	-0,37	0,00	0,34	0,72	Modèle [2]
50	-3,58	-3,22	-2,93	-2,60	-0,40	-0,03	0,29	0,66	
100	-3,51	-3,17	-2,89	-2,58	-0,42	-0,05	0,26	0,63	
250	-3,46	-3,14	-2,88	-2,57	-0,42	-0,06	0,24	0,62	
500	-3,44	-3,13	-2,87	-2,57	-0,43	-0,07	0,24	0,61	
∞	-3,43	-3,12	-2,86	-2,57	-0,44	-0,07	0,23	0,60	
25	-4,38	-3,95	-3,60	-3,24	-1,14	-0,80	-0,50	-0,15	Modèle [3]
50	-4,15	-3,80	-3,50	-3,18	-1,19	-0,87	-0,58	-0,24	
100	-4,04	-3,73	-3,45	-3,15	-1,22	-0,90	-0,62	-0,28	
250	-3,99	-3,69	-3,43	-3,13	-1,23	-0,92	-0,64	-0,31	
500	-3,98	-3,68	-3,42	-3,13	-1,24	-0,93	-0,65	-0,32	
∞	-3,96	-3,66	-3,41	-3,12	-1,25	-0,94	-0,66	-0,33	

Tables de la distribution des t_c et t_b (test bilatéral)

	Modèle [2]			Modèle [3]					
	Constante c			Constante c			Tendance b		
n	2 %	5 %	10 %	2 %	5 %	10 %	2 %	5 %	10 %
25	3,41	2,97	2,61	4,05	3,59	3,20	3,74	3,25	2,85
50	3,28	2,89	2,56	3,87	3,47	3,14	3,60	3,18	2,81
100	3,22	2,86	2,54	3,78	3,42	3,11	3,53	3,14	2,79
250	3,19	2,84	2,53	3,74	3,39	3,09	3,49	3,12	2,79
500	3,18	2,83	2,52	3,72	3,38	3,08	3,48	3,11	2,78
∞	3,18	2,83	2,52	3,71	3,38	3,08	3,46	3,11	2,78

1. Source : Fuller W.A., *Introduction to Statistical Times Series*, John Wiley, 1976.

Bibliographie

- AKAIKE H., « Information theory and an extension of the maximum likelihood principle », in B.N. Petrov and F. Csáki, eds, *2nd International Symposium on Information Theory*, Akadémia Kiadó, Budapest, 1973.
- AKAIKE H., « A new look at the statistical model identification », *IEEE Transactions on automatic Control*, Vol. 19, 1974.
- AKAIKE H., « A bayesian extension of the minimum AIC procedure », *Biometrika*, Vol. 66, 1979.
- ALMON S., « The distributed lag between capital appropriation and expenditures », *Econometrica*, Vol. 33, n°1, janvier 1965.
- AMEMIYA T., « Qualitative response models: a survey », *Journal of Economic Literature*, n°19, 1981.
- ARROW K.J., CHENERY H.B., MINHAS B.S., SOLOW R.M., « Capital-labor substitution and economic efficiency », *Review of Economics and Statistics*, Vol. 43, 1961.
- BEVERIDGE S., NELSON C. R., « A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle », *Journal of the Monetary Economics*, 7, 2, 1981.
- BOURBONNAIS R., TERRAZA M., *Analyse des séries temporelles*, Dunod, Paris, 3^e éd., 2010.
- BOURBONNAIS R., USUNIER J.-C., *Prévision des ventes*, 5^e éd., Economica, Paris, 2013.
- BOX G.E.P., JENKINS G.M., *Time series analysis : forecasting and control*, Holdenday, 1976.
- BOX G.E.P., PIERCE D.A., « Distribution of residual autocorrelations in autoregressive-integrated moving average time series models », *Journal of the American Statistical Association*, Vol. 65, 1970.
- BREUSCH T., « Testing for autocorrelation in dynamic linear models », *Australian Economic Papers*, Vol. 17, 1978.
- DICKEY D., FULLER W., « Distribution of the estimators for autoregressive time series with unit root », *Journal of the American Statistical Association*, Vol. 74, n°366, 1979.
- DICKEY D., FULLER W., « Likelihood ratio statistics for autoregressive time series with unit root », *Econometrica*, Vol. 49, n°4, 1981.
- DURBIN J., « Testing for serial correlation in least-squares regression when some of the regressors are lagged dependant variables », *Biometrika*, Vol. 38, 1970.
- ENGLE R.E., GRANGER C.W.J., « Cointegration and Error-correction : representation, estimation and testing », *Econometrica*, Vol. 55, mars 1987.
- ENGLE R.F., « Autoregressive conditional heteroscedasticity with estimate of the variance of U.K. inflation », *Econometrica*, Vol. 50, 1982.

- FARRAR D.E., GLAUBER R.R., « Multicollinearity in regression analysis », *Review of Economics and Statistics*, Vol. 49, 1967.
- FULLER W.A., *Introduction to statistical time series*, John Wiley, 1976.
- GLEISJER H., « A new test for heteroscedasticity », *Journal of American Statistical Association*, Vol. 64, 1969.
- GODFREY L.G., « Testing for higher order serial correlation in regression equation when the regressors contain lagged dependant variables », *Econometrica*, Vol. 46, 1978.
- GOLDFELD S.M., QUANDT R.E., « Some tests of homoscedasticity », *Journal of the American Statistical Association*, Vol. 60, 1965.
- GRANGER C.W.J., « Co-integrated variables and Error-correcting models », *Document de travail*, université de San Diego, 1983.
- GRANGER C.W.J., « Investigating causal relations by econometrics models and cross spectral methods », *Econometrica*, Vol. 37, 1969.
- GREENE W.H., *Econometric Analysis*, Prentice Hall, 4^e éd., 2000.
- GRIFFITHS W.E., CARTER HILL R., JUDGE G.G., *Learning and practicing econometrics*, John Wiley, New York, 1993.
- HAMILTON J.D., *Time series analysis*, Princeton University Press, 1994.
- HARVEY A.C., *Forecasting Structural Time Series Models and the Kalman Filter*, Cambridge University Press, 1988.
- HAUSMAN J.A., « Specification Tests in Econometrics », *Econometrica*, Vol. 46, 1978.
- HSIAO C., « Analysis of Panel Data », *Econometric Society Monographs*, N°11, Cambridge University Press, 1986.
- JARQUE C.M., BERA A.K., « Efficient test for normality homoscedasticity and serial independance of regression residuals », *Applied Statistics*, Vol. 31, n°2, 1980.
- JARQUE C.M., BERA A.K., « Testing the normality assumption in limited dependant variable models », *International Economic Review*, Vol. 25, n°3, 1984.
- JOHANSEN S., « Statistical analysis of cointegration vectors », *Journal of economic dynamics and control*, Vol. 12, 1988.
- JOHANSEN S., JUSELIUS, « Maximum likelihood estimation and to the demand for money inference on cointegration with application », *Oxford Bulletin of Economics and Statistics*, Vol. 52, 1990.
- JOHANSEN S., « Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models », *Econometrica*, Vol. 59, 1991.
- JOHNSTON J. et DINARDO J., *Méthodes économétriques*, (traduit par B. Guerrien) Economica, 4^e éd., Paris, 1999.
- JUDGE G.G., CARTER HILL R., GRIFFITHS W.E., LÜTKEPOHL H., LEE T.-C., *Introduction to the theory and practice of econometrics*, John Wiley, New York, 1988.
- KENDALL M.G., STUART A., *The advanced theory of statistics*, Vol. 2, Griffin, Londres, 1961.
- KLEIN L.R., *An introduction to econometrics*, Prentice Hall, 1962.
- KLEIN L.R., *Economic fluctuations in the United States 1921-1941*, John Wiley, New York, 1950.
- KOYCK L.M., *Distributed lags and investment analysis*, North-Holland, Amsterdam, 1954.

- KWIATKOWSKI D., PHILLIPS P., SCHMIDT P., et SHIN Y., « Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root ? » *Journal of econometric*, Vol. 54, 1992.
- LJUNG G.M., BOX G.E.P., « On a mesure of lack of fit in time series models », *Biometrika*, Vol. 65, 1978.
- MACKINNON, J.G., *Critical Values for Cointegration Tests*, Chapitre 13 in *Long-run Economic Relationships : Readings in Cointegration*, R.F. Engle et C.W.J. Granger éditeurs, Oxford University Press, 1991.
- MADDALA G.S., RAO A.S., « Maximum likelihood estimation of Solow's and Jorgenson's distributed lag models », *The Review of Economics and Statistics*, février 1971.
- NELSON C.R., PLOSSER C., « Trends and Random Walks in Macroeconomics Time Series : some Evidence and Applications », *Journal of Monetary Economics*, 10, 1982.
- PHILLIPS P., PERRON P., « Testing for unit root in time series regression », *Biometrika*, Vol. 75, 1988.
- QUENOUILLE M.H., « The joint distribution of serial correlation coefficients », *Annals of mathematical statistics*, Vol. 20, 1949.
- SCHWARZ G., « Estimating the dimension of a model », *The annals of Statistics*, Vol. 6, 1978.
- SEVESTRE P., *Économétrie des données de Panel*, Dunod, 2002.
- SIMS C.A., « Macroeconomics and reality », *Econometrica*, Vol. 48, 1980.
- SOLOW R.M., « On a family of lag distributions », *Econometrica*, avril 1960.
- THOMAS A., *Économétrie des variables qualitatives*, Dunod, 2000.
- TOBIN J., « Estimation of relationships for limited dependent variables », *Econometrica*, n°26, 1958.
- WHITE H., « A heteroskedasticity-consistent covariance estimator and a direct test for heteroskedasticity », *Econometrica*, Vol. 48, 1980.
- WOLD H., *A study in the analysis of stationary time series*, Almqvist-Wiksell, 1954.
- WONNACOTT T.H., WONNACOTT R.J., *Statistique*, Economica, 4^e éd., Paris, 1998.
- ZELLNER A., « An efficient Method of Estimating Seemingly Unrelated Regression Equations and Tests for Aggregation Bias », *Journal of the American Statistical Association*, 57, pp. 348-368, 1962.

ndex

A

Algorithme de Gauss-Newton 171
Analyse de la variance 33, 34, 38, 54,
67, 68, 69, 70, 73
Autocorrélation des erreurs 127, 179

B, C

Backward Elimination 120
Bruit Blanc 242
Causalité 292, 293
Coefficient
de corrélation 8, 10, 11, 37, 108
de corrélation multiple 54, 92
de corrélation partielle 109, 113
de détermination 41, 54, 92, 112
Conditions d'identification 222, 225
Corrélogramme 243, 256, 258
Critère 119
d'information de Hannan-Quinn 327
de Akaike 119, 185, 279
de Schwarz 119, 185

D

Décomposition
de Cholesky 288
de la variance 290
Degré
de liberté 27, 34, 54, 70
Données centrées 54
Doubles Moindres Carrés 223

E

Effet de levier 62, 63
Équations
de Yule-Walker 261
réduites 219
structurelles 219

Estimateur

Between 357, 358, 360
de Aitken 127
LSDV 356, 358
Within 356

F

Fonction

d'autocorrélation 240
d'autocorrélation partielle 241
de réponse impulsionnelle 285, 290
de vraisemblance 74

Forward Regression 120

H

Hétéroscédasticité 18, 142, 144
Homoscédasticité 18

I

Intervalle

de confiance 29, 39, 62, 66
de prédiction 40
de prévision 82

L

Leptokurtique 244

Loi

de Fisher 35, 60, 61
de Student 9, 27, 30, 36, 59, 60

M

Matrice « *HAT* » 62
Maximum de vraisemblance 224
Méthode
de Box et Jenkins 260
de Cochran-Orcutt 138, 140
des Moments Généralisés 156, 157

- des triples moindres carrés 234
- des variables instrumentales 155, 160
- du « balayage » 138, 141
- du maximum de vraisemblance 235
- SUR 347
- Modèle
 - à choix multiples 330
 - à correction d'erreur 301, 303, 307
 - à correction d'erreur vectoriel 308
 - à effets aléatoires 355, 357
 - à effets aléatoires individuels 358
 - à effets fixes 355
 - à effets fixes individuels 355, 358
 - à erreur composée 357
 - à erreurs sur les variables 154
 - à retards échelonnés 183
- AR 256
- ARIMA 256, 259
- ARMA 257
- autorégressif 178, 180
- d'ajustement partiel 198
- d'anticipations adaptatives 199
- de choix binaire 322
- de Gompertz 169
- de Klein 230
- de Koyck 193
- de marche au hasard 246
- de Solow 194
- linéaire sur variable latente 322
- logistique 169
- Logit 323, 326, 341
- Logit multinomial 336
- MA 257
- Probit 323, 329
- SARIMA 259
- Tobit 337, 338, 341
- VAR 276
- Moindres Carrés Généralisés 126
- Moindres carrés indirects 223
- Multicolinéarité 114, 118

O, P, R

- Orthogonalisation 287
- Point
 - d'influence 63
 - de levier 63

- Processus
 - DS 246, 248
 - TS 245, 248
- Pseudo- R^2 325
- Racine unitaire 245
- Régression
 - augmentée 157
 - fallacieuse 299
 - pondérée 144
 - réursive 84
- Résidus
 - standardisés 63
 - studentisés 63
- Retards d'Almon 188
- Risque de première espèce 30

S

- Saisonnalité 79, 261
- Schwarz 279
- Stagewise Regression 121
- Stationnarité 239, 277
- Stepwise Regression 120

T

- Test 244
 - ARCH 153
 - bilatéral 32, 60, 64
 - Box-Pierce 131, 243
 - bruit blanc 241
 - d'autocorrélation 178
 - d'égalité des variances 148
 - d'exogénéité d'Hausman 156, 158
 - d'exogénéité faible 313
 - d'Hausman 158, 159, 358
 - de Breusch-Godfrey 130, 132
 - de CHOW 68, 69, 71, 85, 97, 100, 144
 - de cointégration 303
 - de Dickey-Fuller 249, 299
 - de Dickey et Fuller Augmentés 249
 - de différence 156
 - de Durbin et Watson 129, 130, 133, 178
 - de Farrar et Glauber 116, 118
 - de Fisher 36, 44, 67, 74, 93, 184, 350
 - de « h » de Durbin 178, 182
 - de Gleisjer 148, 150
 - de Goldfeld-Quandt 148, 149

- de Granger 294
- de Jarque et Bera 244
- de Johansen 315
- de Klein 115, 117
- de la trace 311
- de la valeur propre maximale 313
- de Ljung-Box 131, 243, 314
- de normalité 244
- de Phillips et Perron 250
- de racines unitaires 249
- de Ramsey 86, 89
- de relation de cointégration 310
- de RESET 86
- de Sims 294
- de Wald 94, 350
- de White 148, 151
- du « CUSUM » 85
- du Kurtosis 244
- du multiplicateur de Lagrange 74, 153
- du Skewness 244

- KPSS 251
- de stationnarité 241
- unilatéral 32
- Théorème
 - de Frisch, Waugh et Lovell 53
 - de Gauss-Markov 53
- Triples moindres carrés 224

V

- Valeur anormale 62
- Variable
 - auxiliaire 75
 - centrée 20
 - endogène 4, 14, 47, 177, 219, 320
 - exogène 4, 14, 47, 177, 219, 320
 - indicatrice 75, 76, 80
 - latente 322, 338
 - muette 75, 96
 - qualitative 77

Dans la même collection

- B. Bernier et Y. Simon, *Initiation à la macroéconomie*, 9^e éd., 2009.
- B. Bernier et H.-L. Védie, *Initiation à la microéconomie*, 3^e éd., 2009.
- R. Bourbonnais et M. Terraza, *Analyse des séries temporelles*, 3^e éd., 2010.
- R. Bourbonnais, *Économétrie*, 9^e éd., 2015.
- M. Delaplace, *Monnaie et financement de l'économie*, 4^e édition 2013.
- G. Deleplace, *Histoire de la pensée économique*, 2^e éd., 2009.
- B. Goldfarb et C. Pardoux, *Introduction à la méthode statistique*, 7^e éd., 2013.
- B. Guillochon, A. Kawecki et B. Venet, *Économie internationale*, 7^e éd., 2012.
- N. Hayek et J.-P. Leca, *Mathématiques pour l'économie*, 4^e éd., 2011.
- F. Huart, *Économie des finances publiques*, 2012.
- J.-P. Lecoutre, *Statistique et probabilités*, 5^e édition, 2012.
- P. de Lima, *Économie bancaire et croissance économique*, 2012.
- T. de Montbrial et E. Fauchart-Foray, *Introduction à l'économie*, 4^e éd., 2009.
- P. Petauton et M. Fromenteau, *Théorie et pratique de l'assurance vie*, 4^e éd., 2012.
- F. Poulon, *Économie générale*, 7^e éd., 2011.
- M. Raffinot, *Économie du développement*, 2015.
- M. Yildizoglu, *Introduction à la théorie des jeux*, 2^e éd., 2011.

TD Éco Sup

- S. Brana et M.-C. Bergouignan, *TD Comptabilité nationale*, 3^e éd., 2011.
- S. Brana, M. Cazals et P. Kauffmann, *TD Économie monétaire et financière*, 4^e éd., 2012.
- S. Brana et M.-C. Bergouignan, *TD Macroéconomie*, 5^e éd., 2015.
- J. Héricourt et J. Reynaud, *TD Économétrie*, 2007.
- K. Jouaber Snoussi et M.-J. Rigobert, *TD Finance d'entreprise*, 2^e éd., 2010.
- J.-P. Lecoutre et Ph. Pilibossian, *TD Algèbre*, 4^e éd., 2014.
- J.-P. Lecoutre et Ph. Pilibossian, *TD Analyse*, 5^e éd., 2013.
- J.-P. Lecoutre, *TD Statistique et probabilités*, 6^e éd., 2015.
- P. Médan, *TD Microéconomie*, 5^e éd., 2015.
- J.-L. Monino, J.-M. Kosianski et F. Le Cornu, *TD Statistique descriptive*, 4^e éd., 2010.
- G. Neuberg, *TD Mathématiques financières et actuarielles*, 2012.
- F. Poulon, *TD Économie générale*, 2011.