

TRAITÉ, D'ÉCONOMÉTRIE FINANCIÈRE

MODÉLISATION FINANCIÈRE

François-Éric Racicot
Raymond Théoret



Presses de l'Université du Québec

TRAITÉ, D'ÉCONOMÉTRIE FINANCIÈRE

MODÉLISATION FINANCIÈRE

© 2001 – Presses de l'Université du Québec

Édifice Le Delta I, 2875, boul. Laurier, bureau 450, Québec, Québec G1V 2M2 • Tél. : (418) 657-4399 – www.puq.ca
Tiré : *Traité d'économétrie financière*, François-Éric Racicot et Raymond Théoret, ISBN 2-7605-1123-5 • D1123N

Tous droits de reproduction, de traduction ou d'adaptation réservés

PRESSES DE L'UNIVERSITÉ DU QUÉBEC

Le Delta I, 2875, boulevard Laurier, bureau 450

Sainte-Foy (Québec) G1V 2M2

Téléphone : (418) 657-4399 • Télécopieur : (418) 657-2096

Courriel : puq@puq.quebec.ca • Internet : www.puq.quebec.ca

Distribution :

CANADA et autres pays

DISTRIBUTION DE LIVRES UNIVERS S.E.N.C.

845, rue Marie-Victorin, Saint-Nicolas (Québec) G7A 3S8

Téléphone : (418) 831-7474 / 1-800-859-7474 • Télécopieur : (418) 831-4021

FRANCE

DIFFUSION DE L'ÉDITION QUÉBÉCOISE

30, rue Gay-Lussac, 75005 Paris, France

Téléphone : 33 1 43 54 49 02

Télécopieur : 33 1 43 54 39 15

SUISSE

GM DIFFUSION SA

Rue d'Etraz 2, CH-1027 Lonay, Suisse

Téléphone : 021 803 26 26

Télécopieur : 021 803 26 29



La *Loi sur le droit d'auteur* interdit la reproduction des œuvres sans autorisation des titulaires de droits. Or, la photocopie non autorisée – le « photocopillage » – s'est généralisée, provoquant une baisse des ventes de livres et compromettant la rédaction et la production de nouveaux ouvrages par des professionnels.

L'objet du logo apparaissant ci-contre est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit le développement massif du « photocopillage ».

© 2001 – Presses de l'Université du Québec

Édifice Le Delta I, 2875, boul. Laurier, bureau 450, Québec, Québec G1V 2M2 • Tél. : (418) 657-4399 – www.puq.ca
Tiré : *Traité d'économétrie financière*, François-Éric Racicot et Raymond Théoret, ISBN 2-7605-1123-5 • D1123N

Tous droits de reproduction, de traduction ou d'adaptation réservés

TRAITÉ, D'ÉCONOMÉTRIE FINANCIÈRE

MODÉLISATION FINANCIÈRE

**François-Éric Racicot
Raymond Théoret**

2001



Presses de l'Université du Québec

Le Delta I, 2875, boul. Laurier, bur. 450
Sainte-Foy (Québec) Canada G1V 2M2

© 2001 – Presses de l'Université du Québec

Édifice Le Delta I, 2875, boul. Laurier, bureau 450, Québec, Québec G1V 2M2 • Tél. : (418) 657-4399 – www.puq.ca
Tiré : *Traité d'économétrie financière*, François-Éric Racicot et Raymond Théoret, ISBN 2-7605-1123-5 • D1123N

Tous droits de reproduction, de traduction ou d'adaptation réservés

Données de catalogage avant publication (Canada)

Racicot, François-Éric

Traité d'économétrie financière : modélisation financière

Comprend des réf. bibliogr.

ISBN 2-7605-1123-5

1. Finances – Modèles économétriques. 2. Économétrie. 3. Mathématiques économiques. 4. Modèles linéaires (Statistique). 5. Hétéroscédasticité. 6. Statistique mathématique. I. Théoret, Raymond. II. Titre.

HG106.R32 2001

332'.01'5195

C2001-940210-4

Nous reconnaissons l'aide financière du gouvernement du Canada par l'entremise du Programme d'aide au développement de l'industrie de l'édition (PADIÉ) pour nos activités d'édition.

Mise en pages : INFO 1000 MOTS INC.

Couverture : RICHARD HODGSON

1 2 3 4 5 6 7 8 9 PUQ 2001 9 8 7 6 5 4 3 2 1

Tous droits de reproduction, de traduction et d'adaptation réservés

© 2001 Presses de l'Université du Québec

Dépôt légal – 2^e trimestre 2001

Bibliothèque nationale du Québec / Bibliothèque nationale du Canada

Imprimé au Canada

© 2001 – Presses de l'Université du Québec

Édifice Le Delta I, 2875, boul. Laurier, bureau 450, Québec, Québec G1V 2M2 • Tél. : (418) 657-4399 – www.puq.ca
Tiré : *Traité d'économétrie financière*, François-Éric Racicot et Raymond Théoret, ISBN 2-7605-1123-5 • D1123N

Tous droits de reproduction, de traduction ou d'adaptation réservés

TABLE DES MATIÈRES

Introduction	1
Chapitre 1 Rappels statistiques	5
1. Notion de variable aléatoire	5
2. Statistiques descriptives	8
2.1. Mesure de la tendance centrale d'une variable aléatoire	8
2.2. Mesures de dispersion d'une variable aléatoire	9
2.3. Mesure du degré d'asymétrie et d'aplatissement d'une distribution empirique	11
2.4. Test de Kolmogorov-Smirnov (K-S)	17
3. Modèles probabilistes	18
3.1. Lois de probabilité discrètes univariées	18
3.2. Lois de probabilité continues univariées et concepts bivariés	27
4. Notions d'indépendance, de densité jointe et de densité marginale	38
5. Probabilités conditionnelles et densités conditionnelles	40
6. Théorème central limite (cas univarié)	41
7. La loi normale multivariée	42
8. Estimation de la moyenne et de la variance dans un modèle simple	44

9. Théorème de Gauss-Markov	46
10. La méthode d'estimation du maximum de vraisemblance	47
11. Tests d'hypothèses et intervalles de confiance	51
12. Applications numériques	60
Annexe	67
Chapitre 2 Le modèle linéaire à deux variables	69
1. Spécification du modèle à deux variables et propriétés des erreurs résiduelles	69
2. Estimateur des MCO	71
3. Propriétés des MCO	73
4. Tests d'hypothèses et intervalles de confiance	78
5. Prévision	79
5.1. Prévision de $E(y_0)$	79
5.2. Prévision de y_0	81
6. Mesures du degré d'ajustement	82
7. Applications	89
Chapitre 3 Le modèle linéaire général	97
1. Formulation matricielle et hypothèses de base	97
2. Propriétés de l'estimateur des MCO	102
3. Hypothèses sur les erreurs et conséquences	104
4. Tests d'hypothèses et intervalles de confiance	105
5. Prévision dans le modèle linéaire général	114
6. Applications	115
Annexe : Rappels de calcul matriciel	117
1. Opérations matricielles	117
2. Matrices carrées importantes	119
3. Des matrices importantes : la matrice variance-covariance d'un portefeuille de titres et la covariance entre deux portefeuilles	126
4. Quelques applications du calcul matriciel en finance	129

Chapitre 4 Variations sur les modèles linéaire et non linéaire : Partie I	143
1. Erreurs de spécification	143
2. Tests reliés aux erreurs de spécification	146
2.1. Test sur la forme logarithmique	146
2.2. La transformation de Box-Cox	147
3. Méthodes des moindres carrés non linéaires et transformation Box-Cox	149
4. Critères de sélection des variables explicatives	156
5. Variables auxiliaires et modélisation des changements structurels	159
Chapitre 5 Variations sur les modèles linéaire et non linéaire : Partie II	169
1. Théorie asymptotique : convergence, tests asymptotiques et variables instrumentales	169
1.1 Convergence	169
1.2. Tests asymptotiques : LR, LM et Wald	173
2. Problèmes au chapitre des variables explicatives : multicollinéarité et endogénéité des variables explicatives	179
Chapitre 6 Les méthodes numériques en économétrie : une introduction	185
1. Simulation de Monte Carlo : le cas d'une option asiatique	185
2. La méthode du bootstrap	193
3. Régression non paramétrique : une simulation Monte Carlo	196
Chapitre 7 L'hétéroscédasticité	201
1. Propriétés de l'estimateur des MCO lorsque les erreurs sont hétéroscédastiques	201
2. L'estimateur des moindres carrés généralisés : MCG	205
3. Matrice de White pour l'hétéroscédasticité	207

4.	Tests d'hétéroscédasticité	208
4.1.	Test de Goldfeld et Quandt (1965)	208
4.2.	Test de Breusch-Pagan (1979)	209
4.3.	Test de White (1980)	211
5.	Applications	211
6.	Note sur l'inférence statistique en présence d'hétéroscédasticité	214

Chapitre 8 L'autocorrélation des erreurs résiduelles 215

1.	Propriétés de l'estimateur des MCO lorsque les résidus sont autocorrélés	215
2.	Correction du modèle original lorsque ρ n'est pas connu	222
3.	Tests d'autocorrélation	224
4.	Prévision dans le modèle linéaire avec erreur de la forme AR(1)	226
5.	Applications	227

Chapitre 9 Les séries temporelles 229

1.	Processus stochastiques	229
1.1.	Stationnarité	230
1.2.	Processus autorégressifs stationnaires : représentation et estimation	230
2.	Estimation du processus autorégressif AR(p)	232
3.	Fonction d'autocorrélation partielle (PACF)	234
3.	Processus de moyennes mobiles : MA(q)	236
4.	Estimation d'un MA(q)	239
5.	Modèles ARMA (p, q)	240
6.	Introduction aux processus stochastiques non stationnaires : modèles ARIMA (p, d, q)	242
7.	La méthode de Box et Jenkins	243
8.	Autres critères de sélection pour les modèles ARMA	246
9.	Prévisions à l'aide de modèles statistiques de séries chronologiques	247

10. Évaluation de la précision des prévisions	249
11. Applications	250
12. Processus stochastiques non stationnaires	258
13. Modèles de tendance	260
14. Racines unitaires et régressions fallacieuses	263
15. Tests de racine unitaire	265
16. Cointégration	268
Chapitre 10 L'hétéroscédasticité conditionnelle (ARCH) .	273
1. Notions d'espérances conditionnelle et non conditionnelle ; notions de variance conditionnelle et non conditionnelle	273
2. L'hétéroscédasticité conditionnelle et les faits	275
3. Le modèle ARCH	276
4. Estimation du modèle ARCH	278
5. Généralisation du modèle ARCH	280
5.1. Le modèle ARCH(q)	280
5.2. Le modèle ARCH-M	282
5.3. Le modèle EGARCH	283
5.4. Le modèle TARCH	284
5.5. Prévision à partir du modèle GARCH	284
5.6. Test ARCH	286
6. Une digression : la théorie de l'APT	308
6.1. Le principe de l'arbitrage	308
6.2. L'APT : aperçu général	314
6.3. Dérivation du modèle de l'APT	315
6.4. Tests de l'APT	322
Chapitre 11 La méthode des moments généralisés	325
1. Introduction à la méthode des moments	325
2. La méthode des moments et les MCO	328
3. La méthode des moments et l'estimateur des variables instrumentales	329

4.	GMM et conditions d'orthogonalité	332
5.	Maximum de vraisemblance et GMM	333
6.	Applications	335
	Annexe – Tables statistiques	353
A1.	Répartition de la loi normale centrée réduite	354
A2.	Répartition du t de Student	355
A3.	Répartition du χ^2	356
A4.	Répartition du F	358
A5.	Statistique de Durbin et Watson au seuil de 5 %	362
	Index	367

INTRODUCTION

Jusqu'au début des années 1980, l'économétrie s'est développée à un rythme relativement lent. Elle avait beaucoup de mal à se libérer du paradigme statistique classique. Mais avec la poussée fulgurante de l'informatique, l'économétrie a connu un essor fort appréciable ces vingt dernières années. Que l'on pense simplement à la multiplication effrénée des modèles économétriques non linéaires, des modèles de volatilité et des nouvelles techniques d'estimation comme le GMM ou la méthode des moments simulés, pour ne nommer que quelques nouveaux champs de l'économétrie contemporaine.

Mais ce qui est encore plus saisissant, c'est l'avancée au pas de charge de l'économétrie dans le domaine de la théorie financière. En effet, la théorie des produits dérivés, qui prend sa source au début des années 1970, fait de plus en plus appel aux modèles économétriques de volatilité, tels les modèles GARCH, et à la méthode du GMM pour estimer les paramètres des équations différentielles stochastiques qui servent à la détermination des prix des options, entre autres. L'économétrie a également permis au modèle du CAPM, bien connu en théorie financière, de s'affranchir de son cadre statique. On peut maintenant parler de bêtas variables dans le temps et la transposition de l'approche GARCH au CAPM a permis de le situer dans un cadre multivarié. La finance corporative emprunte également de plus en plus à l'économétrie. Ainsi, l'analyse des investissements des entreprises dans un contexte d'incertitude donne lieu à la formulation d'équations différentielles stochastiques dont l'estimation des paramètres exige le recours à l'économétrie, entre autres à la méthode économétrique du GMM.

L'incursion de l'économétrie dans le domaine de la finance a donné lieu à l'apparition d'une nouvelle discipline : l'économétrie financière. L'économètre financier, en plus de maîtriser l'économétrie moderne, doit disposer de bases solides en théorie financière de façon à pouvoir opérer une symbiose des deux disciplines que sont l'économétrie et la finance. La formation de l'économètre financier est donc très exigeante. Le présent *Traité d'économétrie financière* s'attaque à cette discipline complexe en visant à exposer au lecteur les fondements de l'économétrie financière. Les applications des méthodes économétriques présentées dans notre *Traité* seront donc tirées de la théorie financière moderne.

Il n'existe pas à notre avis de manuel rédigé en français qui se soit donné notre objectif. Du fait de l'importance de plus en plus grande de la finance empirique, notre *Traité* vient combler une grave lacune qui existe encore aujourd'hui au sein des outils pédagogiques à la disposition des étudiants de la finance et de l'économie financière. Il vise la clientèle des étudiants de troisième année du baccalauréat spécialisé en finance ou en économie financière et des étudiants des divers programmes de MBA, de maîtrise en finance appliquée ou de DESS en finance. Il s'adresse également au spécialiste de la finance – analyste financier, gestionnaire de portefeuille, ingénieur financier – qui souhaite effectuer un tour d'horizon complet et rigoureux de l'économétrie financière moderne.

Tout en se voulant une introduction à l'économétrie financière moderne, notre *Traité d'économétrie financière* vise également à approfondir certains domaines-clefs de cette discipline, parfois jugés complexes par l'étudiant, comme les modèles GARCH et le GMM. Dans son souci de rigueur, notre *Traité* fournit très souvent au lecteur les preuves des diverses formules qui y apparaissent. Dans son souci pédagogique, notre *Traité* renferme également des chapitres ou sections consacrés à des rappels de la statistique ou du calcul matriciel.

Voici un bref survol de notre *Traité d'économétrie financière*. Le chapitre 1 porte sur des rappels de notions statistiques de base qui sont utilisées par la suite dans notre manuel. On y expose, entre autres, une version étoffée de la méthode d'estimation du maximum de vraisemblance. Les chapitres 2 et 3 sont les chapitres classiques de tout manuel d'économétrie. Ils présentent le modèle linéaire à deux

variables et le modèle linéaire général. Les chapitres 4 et 5 ont traité à des variations sur les modèles linéaire et non linéaire. Y sont présentés, entre autres : le modèle des moindres carrés non linéaires et le modèle Box-Cox ; les tests J et RESET ; le test de Chow ; une introduction à la théorie asymptotique ; les tests LM, LR et de Wald ; une introduction à la théorie des variables instrumentales et au phénomène de la multicollinéarité.

Le chapitre 6 se penche sur les méthodes numériques utilisées en économétrie. On y aborde la simulation de Monte Carlo, la technique dite du *bootstrapping* et celle du kernel. On y montre incidemment comment évaluer le prix d'une option asiatique à partir d'une simulation de Monte Carlo. Les chapitres 7 et 8 s'attardent aux problèmes économétriques classiques de l'hétéroscédasticité et de l'autocorrélation des erreurs résiduelles. Le chapitre 9 concerne la théorie économétrique des séries temporelles. Y font figure les processus stochastiques, les modèles ARMA et ARIMA, les prévisions à l'aide de séries chronologiques, les tests de racines unitaires et le phénomène de la cointégration.

Le chapitre 10 dirige son collimateur sur un problème statistique important dans le domaine des séries financières : l'hétéroscédasticité conditionnelle. Une attention particulière est accordée aux modèles ARCH, ARCH-M, GARCH, EGARCH et TARARCH. La prévision des séries chronologiques dans un contexte d'hétéroscédasticité conditionnelle y est étudiée. Finalement, les applications que contient ce chapitre concernent le modèle financier du CAPM. On y montre entre autres comment estimer le modèle du CAPM dans le cadre d'un modèle GARCH multivarié.

Finalement, le chapitre 11 s'attaque à la méthode des moments généralisés, dont l'acronyme est : GMM. Nous y démontrons comment cette technique d'estimation intègre les modèles classiques d'estimation : modèle des moindres carrés linéaires, des doubles moindres carrés et du maximum de vraisemblance. Comme application de la méthode du GMM, nous estimons les paramètres du modèle stochastique de taux d'intérêt de Schaefer et Schwartz dans un contexte canadien.

L'économétrie financière est une discipline captivante. À en juger par l'évolution accélérée qu'elle connaît depuis vingt ans, elle est appelée à un brillant avenir. Nous espérons que le lecteur partagera, au fil de la lecture des chapitres de notre *Traité d'économétrie financière*, notre très vif intérêt pour cette nouvelle discipline.

RAPPELS STATISTIQUES¹

Ce chapitre vise à présenter les principaux outils probabilistes et statistiques qui sont essentiels à la compréhension de ce *Traité d'économétrie financière*. Nous présentons dans un premier temps les notions de variables aléatoires et de modèles probabilistes en temps discret et continu. Les modèles probabilistes regroupent les principales lois de probabilité en temps discret, soit les distributions binômiale et de Poisson, et les lois de probabilité en temps continu, soit les lois normales univariée et bivariée, le chi-carré, le t de Student, le F de Fisher, la loi uniforme. Dans un second temps, nous nous penchons sur les moments de certaines distributions et sur le théorème central limite. L'estimation de certains de ces moments, entre autres par la méthode des moindres carrés ordinaires et celle du maximum de vraisemblance, est abordée. Leurs intervalles de confiance sont calculés.

1. NOTION DE VARIABLE ALÉATOIRE

Il existe deux définitions pour une variable aléatoire, l'une heuristique, l'autre basée sur la théorie de la mesure. Selon la définition

-
1. Les références des chapitres 1 et 2 sont les suivantes: Amemiya, T. (1994), *Introduction to Statistics and Econometrics*, Harvard University Press, Cambridge, Massachusetts; Baillargeon, G. et J. Rainville (1979), *Statistique appliquée*, tomes 1 et 2, 5^e édition, Éditions SMG, Trois-Rivières; Judge, G.G. *et al.* (1988), *Introduction to the Theory and Practice of Econometrics*, 2^e édition, Wiley, New York; Kendall's *Advanced Theory of Statistics* (1999), Arnold, London; Rao, C.R. (1973), *Linear Statistical Inference and Its Applications*, 2^e édition, John Wiley and Sons, New York.

heuristique, une variable aléatoire est une variable qui prend des valeurs suivant une certaine fonction de distribution². Dans sa version formelle, une fonction à valeur réelle $X(\cdot)$ définie sur l'espace (Ω, β, P) est appelée variable aléatoire (ou mesurable dans le langage de la théorie de la mesure) si l'ensemble $\{\omega : X(\omega) < x\} \in \beta$ pour tout x dans \mathfrak{R} où le triplet (Ω, β, P) est appelé espace probabiliste, Ω étant défini comme l'espace échantillonnal et β étant l'ensemble de Borel. Un ensemble de Borel est une famille de sous-ensembles contenant tous les événements de la droite des réels pour lesquels on peut calculer une probabilité de réalisation. Formellement, on appelle aussi β une σ -algèbre. Pour sa part, P est une mesure de probabilité sur β . Donc, on peut voir la fonction $X(\cdot)$ comme une application de Ω à \mathfrak{R} : $\Omega \xrightarrow{x} \mathfrak{R}^3$.

On distingue les variables aléatoires continues et discrètes. Par exemple, les réalisations de l'indice S&P's 500 aux États-Unis et du TSE 300 au Canada font partie des variables aléatoires continues, car elles peuvent prendre n'importe quelle valeur dans l'ensemble des réels. Les variables dichotomiques font partie de l'ensemble des variables aléatoires discrètes. Par exemple, émettre ou ne pas émettre un dividende est un exemple de variable dichotomique. On donnerait la valeur 1 lorsqu'il y a émission de dividende et 0 autrement.

On distingue également les variables déterministes des variables aléatoires. Une variable $y^*_t = f(x_t)$, où x_t est connu et parfaitement contrôlé, est dite déterministe et donc parfaitement prévisible. Par exemple, si $f(x_t) = mx_t + b$, alors $y^*_t = mx_t + b$. On peut illustrer cette relation par la figure 1.1.

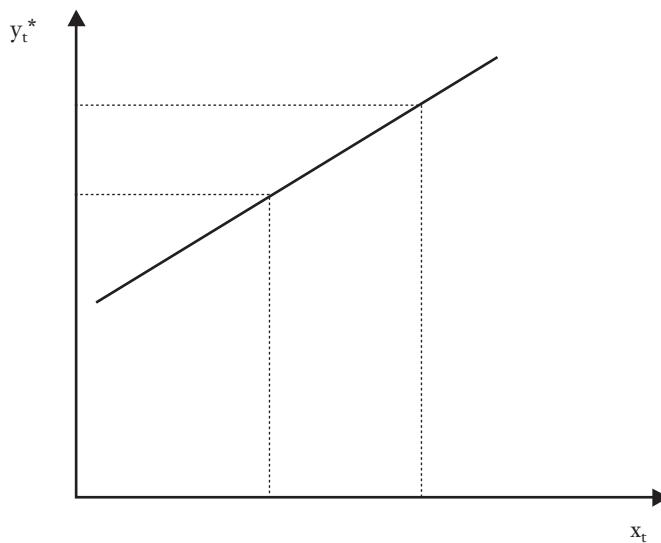
On observe sur cette figure que pour une valeur donnée de x , la valeur de y est automatiquement déterminée. Connaissant la valeur de x , on peut donc prévoir parfaitement la valeur de y . Il existe des formes beaucoup plus complexes de variables déterministes. Par exemple, dans la théorie du chaos déterministe, les formes fonctionnelles sont de nature hautement non linéaire, mais elles nous amènent

2. La notion de fonction de distribution sera définie ultérieurement.

3. Pour des détails additionnels, voir : Rao, C. (1973), *Linear Statistical Inference and Its Applications*, John Wiley and Sons, New York.

à la prévision exacte de certains phénomènes physiques⁴. Tel n'est pas le cas pour une variable aléatoire. Une variable définie comme : $y_t = f(x_t) + e_t$, où e_t est une variable aléatoire IID⁵(0, σ^2) y_t étant la somme d'une fonction non stochastique et d'une composante aléatoire (stochastique) est donc une variable aléatoire.

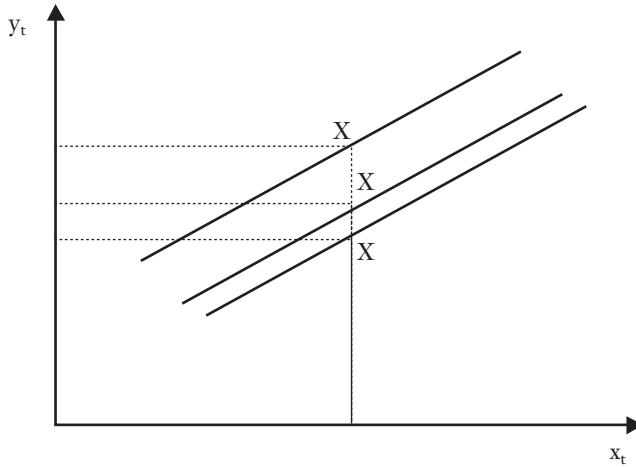
FIGURE 1.1



La figure 1.2 représente la fonction : $y_t = b + mx_t + e_t$. y_t peut prendre plusieurs valeurs pour une valeur donnée de x_t . Ceci est dû à la présence du terme aléatoire e_t dans la fonction de y_t . y_t n'étant plus prévisible parfaitement, mais seulement à l'intérieur d'un intervalle de confiance, il s'agit donc d'une variable aléatoire.

4. Les exemples classiques de fonctions de variables déterministes en physique sont : la *tent map*, la *logistic map* et le modèle du chaos déterministe de Makey et Glass (1977) qui a servi, entre autres, à modéliser la reproduction des cellules rouges du sang. À ce sujet, on consultera également : Racicot, F.E. (2000), Notes on Nonlinear Dynamics, document de travail, CRG, 16-2000, ESG, UQAM.
5. IID pour « identiquement et indépendamment distribué ».

FIGURE 1.2



2. STATISTIQUES DESCRIPTIVES

À une variable aléatoire donnée sont associées plusieurs statistiques descriptives. Dans ce qui suit, nous analysons les plus utilisées en finance empirique.

2.1. Mesure de la tendance centrale d'une variable aléatoire

Soit une variable aléatoire X et ses réalisations $x_i \in \{x_1, \dots, x_T\}$. Alors la moyenne des réalisations se définit comme suit :

$$\bar{x} = \frac{\sum_{i=1}^T x_i}{T}$$

Supposons maintenant que l'on ait plusieurs variables aléatoires : $X_i \in \{X_1, \dots, X_T\}$ où les $X_i \sim \text{IID}(m, \sigma^2)$, où m est la moyenne de la

population et s^2 , sa variance. Alors l'estimateur \bar{X} est dit sans biais si

$$E(\bar{X}) = \frac{\sum_{i=1}^T E(X_i)}{T} = \frac{T\mu}{T} = \mu. \text{ Le biais de } \bar{X} \text{ se définit comme suit :}$$

$$\text{biais}(\bar{X}) = E(\bar{X}) - \mu$$

D'autres mesures de la tendance centrale sont la médiane et le mode. La médiane se définit comme étant la valeur qui sépare l'échantillon en deux. Si le nombre d'observations est impair, la médiane est

égale à : $\frac{N+1}{2}$. Pour sa part, le mode est la valeur la plus fréquente

observée dans un échantillon. Soulignons que la médiane fait figure d'estimateur robuste de la tendance centrale en ce sens qu'elle ne dépend pas de la normalité d'une distribution, contrairement à la moyenne qui, elle, dépend de cette hypothèse. En effet, si la distribution échantillonnale diffère de la normale, la moyenne est alors un mauvais estimateur de la tendance centrale, ce qui n'est pas le cas de la médiane.

2.2. Mesures de dispersion d'une variable aléatoire

Nous voulons calculer la variance échantillonnale, mesure de dispersion de cet échantillon, désignée par s^2 . Pour les réalisations x_i de X , celle-ci est égale à :

$$s^2 = \frac{\sum_{i=1}^T (x_i - \bar{x})^2}{T-1}$$

Supposons, comme dans le cas précédent, que l'on ait plusieurs variables aléatoires : $X_i \in \{X_1, \dots, X_T\}$ où les $X_i \sim \text{IID}(\mu, \sigma^2)$, où μ est la moyenne de la population et s^2 , sa variance. Alors l'estimateur S^2 est

$$\text{dit sans biais si : } S^2 = \frac{\sum_{i=1}^T (X_i - \bar{X})^2}{T-1} \text{ et } E(S^2) = \sigma^2.^6$$

6. La preuve de cette formule est donnée à l'annexe de ce chapitre.

L'écart-type est désigné par $\sqrt{s^2}$. Intuitivement, s^2 peut être vu comme une moyenne d'écart par rapport à la moyenne au carré, le carré éliminant les signes négatifs de cette moyenne. Cette statistique nous donne l'étendue d'une distribution. Pour un petit écart-type, les observations seront concentrées autour de la moyenne alors que dans le cas d'un grand écart-type, elles seront plus dispersées. Les figures 1.3 et 1.4 illustrent les distributions empiriques associées à un petit écart-type et à un grand écart-type.

FIGURE 1.3

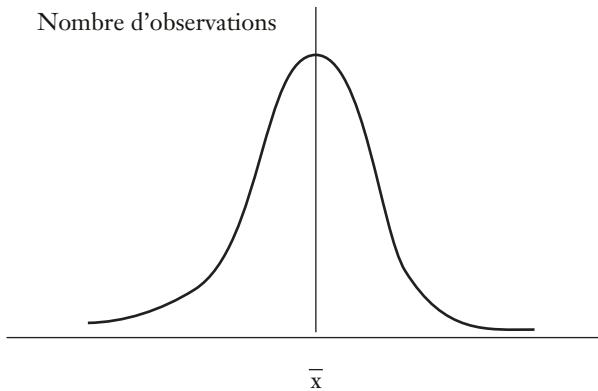
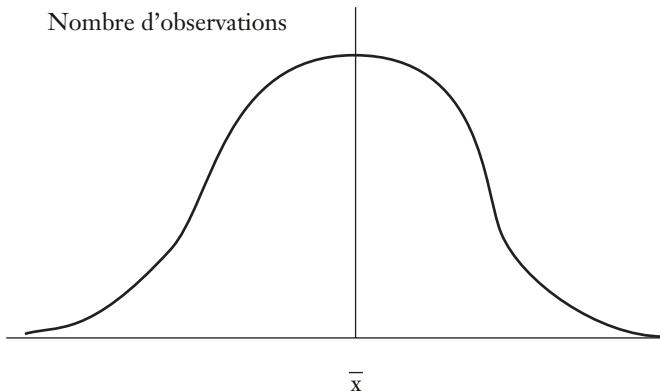


FIGURE 1.4



Une autre mesure de la variabilité des réalisations d'une variable aléatoire est le coefficient de variation, désigné par CV. Il se définit comme suit pour la variable X :

$$CV = \frac{s_X}{\bar{x}} \times 100$$

Si l'on veut comparer la variabilité des réalisations de deux variables aléatoires X et Y, on recourt au CV de chacune de ces variables pour pallier le problème de l'échelle des mesures. En effet, supposons que s_X soit égal à 3,6 et s_Y , à 631,4, on pourrait être porté à croire que Y est plus variable que X, la variabilité étant mesurée par l'écart-type. Tel n'est pas le cas puisqu'il y a ici un problème d'unité de mesure : la variable X est mesurée en pourcentage (p. ex., le taux d'intérêt) et la variable Y, en dollars (p. ex., le volume des transactions boursières). On peut remédier à ce problème en dégonflant s_X et s_Y par leur moyenne respective : \bar{x} est égal à 8,8 et \bar{y} à 2915,3. Les coefficients respectifs de variation pour X et Y sont de 0,409 et de 0,217. On réalise donc après coup que X est plus variable que Y sur la base du coefficient de variation même si, a priori, on concluait l'inverse. Comme autre exemple, on peut noter que les variables mesurées en millions de dollars sont plus volatiles en termes absolus que celles mesurées en unités de dollars, ce qui n'est pas le cas sur une base relative, qui s'obtient en divisant ces variables par leur moyenne respective.

2.3. Mesure du degré d'asymétrie et d'aplatissement d'une distribution empirique

2.3.1. Le coefficient d'asymétrie (*skewness*)

Ce coefficient mesure le degré d'asymétrie d'une distribution. Il se définit comme suit :

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

où E(.) est l'espérance et V(.), la variance. E, dans les petits échantillons, est estimé par la moyenne arithmétique des réalisations de la variable aléatoire :

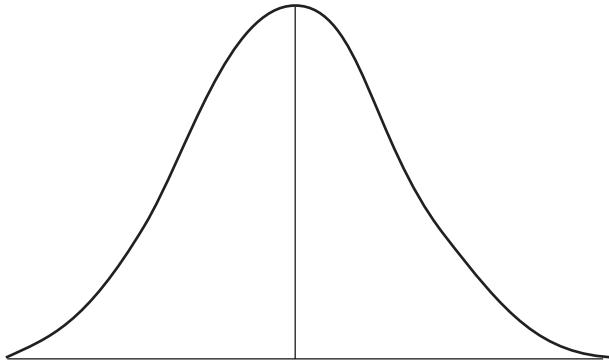
$$E\left[(X - \mu)^3\right] \cong \frac{\sum_{i=1}^T (x_i - \bar{x})^3}{T} = \hat{\mu}_3$$

où μ est estimé par \bar{x} .

Par ailleurs, on estime σ^2 par $\hat{\sigma}^2 = \frac{\sum_{i=1}^T (x_i - \bar{x})^2}{T}$.

Si $\frac{\mu_3}{\sigma^3} = 0$, alors la distribution est symétrique, à l'instar de la normale, qui apparaît à la figure 1.5.

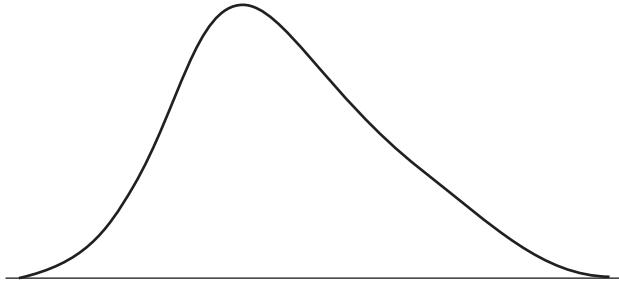
FIGURE 1.5



Notons que dans le cas d'une distribution normale, tous les moments impairs sont nuls.

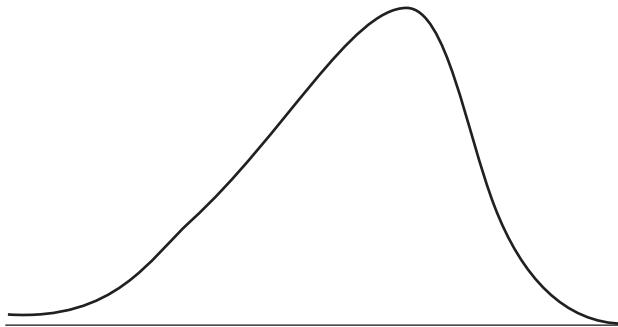
Si $\frac{\mu_3}{\sigma^3} > 0$, alors la densité de la distribution est concentrée vers la droite, comme le montre la figure 1.6.

FIGURE 1.6



Si $\frac{\mu_3}{\sigma^3} < 0$, alors la densité de la distribution est concentrée vers la gauche, comme le montre la figure 1.7.

FIGURE 1.7



L'intuition ici est que l'on compare les moments empiriques de nos données aux moments théoriques d'une distribution qui est la distribution normale.

2.3.2. Le coefficient d'aplatissement (Kurtosis)

Comme son nom l'indique, le coefficient d'aplatissement d'une distribution mesure son degré d'aplatissement. Il est associé à l'épaisseur des queues (*tails*) de la distribution. On le définit comme suit :

$$\frac{\mu_4}{\sigma^4} = \frac{E[(X - \mu)^4]}{(V(X))^2} = \frac{E[(X - \mu)^4]}{(\sigma^2)^2}$$

Dans la pratique on estime ce coefficient de la façon suivante. L'espérance E est estimée par la moyenne échantillonnale, c'est-à-dire :

$$E[(X - \mu)^4] \cong \frac{\sum_{i=1}^T (x_i - \bar{x})^4}{T} = \hat{\mu}_4$$

où μ et σ^2 sont estimés comme ci-devant.

Si $\frac{\mu_4}{\sigma^4} = 3$, il n'y a pas de biais leptocurtique. On dit alors que la distribution est mésocurtique comme c'est le cas pour la distribution normale qui sert de point de référence. Par ailleurs, si $\frac{\mu_4}{\sigma^4} > 3$, on est confronté au cas d'une distribution leptocurtique. Plus communément, on dit qu'une telle distribution présente des queues épaisses, toujours en rapport avec les extrémités d'une distribution normale, comme on peut le constater à la figure 1.8.

Finalement, si $\frac{\mu_4}{\sigma^4} < 3$, on parle alors de distribution platicurtique. Plus communément, on dit qu'une telle distribution présente des queues minces (*thin tails*), toujours en rapport avec les extrémités d'une distribution normale, comme on peut le constater à la figure 1.9.

FIGURE 1.8

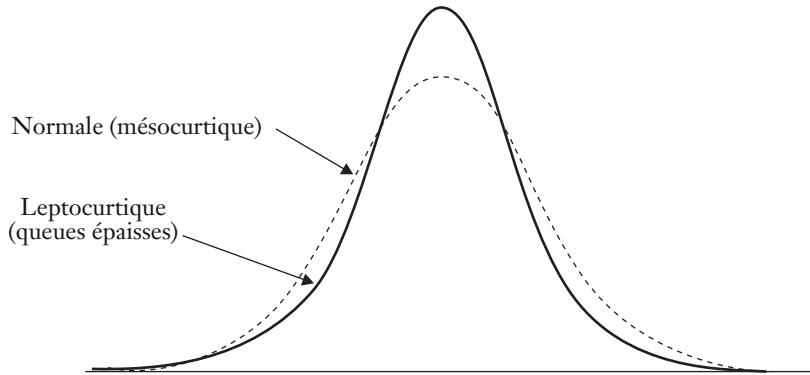
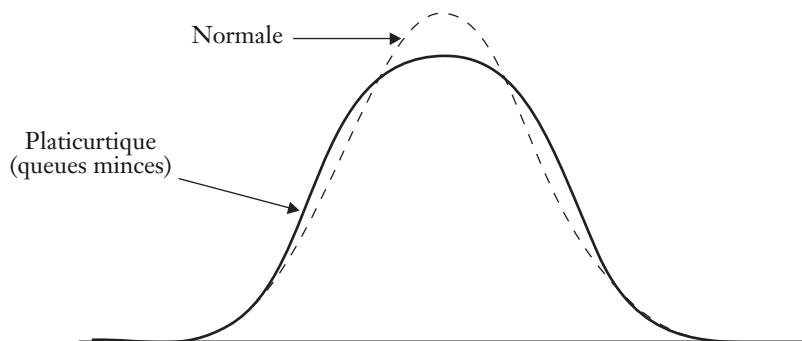


FIGURE 1.9



Si les coefficients estimés d'asymétrie et d'aplatissement sont respectivement près de 0 et de 3 pour une distribution donnée, on pourrait conclure qu'on est en présence d'une distribution gaussienne (normale). Certains logiciels très connus comme EViews, SAS et RATS sont déjà préprogrammés pour le calcul de ces coefficients.

Comme c'est toujours le cas en statistique, un seul coup d'œil graphique ne suffit pas à mesurer les déviations de ces coefficients par rapport à la normale. Comme à l'accoutumée, il faut développer un test pour juger du caractère significatif de ces déviations. Le test de Jarque et Bera (1984) est conçu à cette fin. Ce test est défini sur la somme des coefficients d'asymétrie et d'aplatissement élevés au carré. Plus précisément, le test de Jarque et Bera est basé sur la statistique suivante :

$$JB = \frac{T-K}{6} \left[AS^2 + \frac{1}{4} (KUR-3)^2 \right] \overset{a}{\sim} \chi^2(2)$$

où AS est le coefficient d'asymétrie et KUR, le coefficient de kurtosis.

Le test d'hypothèses est le suivant. L'hypothèse nulle H0 est que la distribution est normale alors que l'hypothèse alternative H1 est que la distribution n'est pas normale. La règle consiste à rejeter H0 si JB est plus grand que χ^2 avec deux degrés de liberté au seuil de signification habituel de 5 % ou si la p-value associée à la statistique JB est inférieure à 0,05. Une mise en garde vis-à-vis l'utilisation de ce test s'impose cependant. Ce test est asymptotique comme l'indique le symbole $\overset{a}{\sim}$ dans la formule de JB. Ce test n'est donc pas exact parce que l'on ne connaît pas la distribution de JB dans de petits échantillons. On ne connaît sa distribution que dans les grands échantillons⁷. On peut effectuer directement ce test de normalité dans le logiciel EViews.

7. Si l'on régressait les y sur X , on obtiendrait le vecteur $\hat{e} = y - X\hat{\beta}$. Pour tester la normalité des erreurs résiduelles, on calcule les inputs de la formule de JB

comme suit : $\hat{\mu}_3 = \frac{\sum \hat{e}_i^3}{T}$; $\hat{\mu}_4 = \frac{\sum \hat{e}_i^4}{T}$; $\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{T}$. Ce test fait partie de la catégorie des tests LM. Notons que l'on pourrait aussi le voir comme un test de Wald.

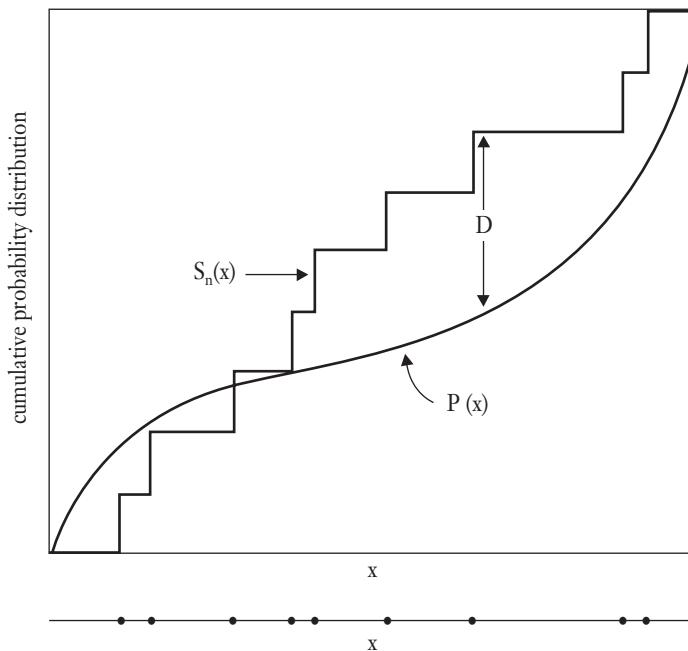
2.4. Test de Kolmogorov-Smirnov⁸ (K-S)

Ce test est utilisé pour comparer une distribution empirique à une distribution donnée. Ce test peut notamment servir de test de normalité. La statistique reliée au test de K-S, désignée par D , se calcule comme suit :

$$D = \max_{-\infty < x < \infty} |S_N(x) - P(x)|$$

où $S_N(x)$ est un échantillon de données et $P(x)$ est une cdf⁹ connue. La représentation graphique des variables du test apparaît à la figure 1.10.

FIGURE 1.10



8. Cette section s'inspire du livre suivant : Press, W.H. *et al.* (1989), *Numerical Recipes: The Art of Scientific Computing (FORTRAN Version)*, Cambridge University Press, Cambridge (reproduit avec la permission de l'éditeur).
9. cdf est l'abréviation de *cumulative distribution function*.

Sur cette figure, D est la statistique de K-S. La distribution empirique des valeurs de x , soit le graphique de $S_N(x)$, est comparée à une distribution théorique où la densité de probabilité cumulée est représentée par $P(x)$. $S_N(x)$ est une fonction par paliers qui augmente d'un montant identique pour chaque observation mesurée. La valeur de D représente la plus grande distance entre ces deux distributions cumulées, soit $S_N(x)$ et $P(x)$. Pour le cas où il y a deux fonctions de distribution cumulatives d'observations différentes, $S_{N1}(x)$ et $S_{N2}(x)$, la statistique K-S s'écrit :

$$D = \max_{-\infty < x < \infty} |S_{N1}(x) - S_{N2}(x)|$$

Les hypothèses à tester sont les suivantes. H_0 : les deux distributions sont identiques; H_1 : les deux distributions sont différentes. Pour ce faire, il faut calculer le caractère significatif de D (p -value), c'est-à-dire : $\text{Prob}(D > \text{valeur observée}) = Q_{K-S}(\sqrt{ND})$, où

$Q_{K-S}(\lambda) = 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2\lambda^2}$ et où $Q_{K-S}(0) = 1$ et $Q_{K-S}(\infty) = 0$. N représente le nombre d'observations.

Dans le cas où l'on compare deux distributions, le niveau de signification se calcule comme suit : $\text{Prob}(D > \text{valeur observée}) =$

$$Q_{K-S}\left(\sqrt{\frac{N_1 N_2}{N_1 + N_2}} D\right). \text{ À remarquer que ce test est asymptotique}$$

mais en pratique, on peut considérer que $N = 20$ est un seuil tolérable, d'autant plus si le degré de conservatisme est important (seuil de signification de 0,01 ou moins).

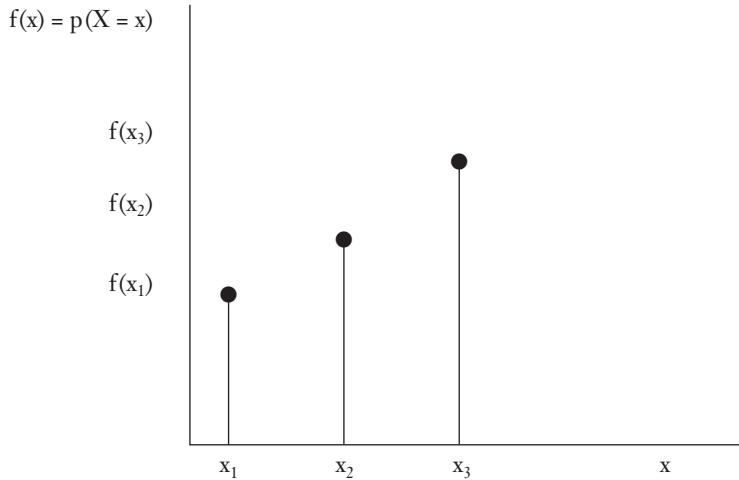
3. MODÈLES PROBABILISTES

3.1. Lois de probabilité discrètes univariées

3.1.1. La fonction de distribution et de répartition

Rappelons d'abord les statistiques descriptives associées aux lois de probabilité discrètes univariées. La distribution d'une variable aléatoire discrète qui rend compte de ses probabilités de réalisation peut être représentée par la figure en bâtonnets 1.11.

FIGURE 1.11 Distribution discrète de X



La probabilité que X soit plus petit ou égal à x est représentée par :
 $P(X \leq x) = F(x) = \sum_{X \leq x} f(x)$, où $F(\cdot)$ est la fonction de répartition (probabilité cumulative). La représentation de la probabilité cumulative apparaît à la figure 1.12.

Notons sur cette figure que les probabilités se cumulent à mesure que les x_i augmentent puisque la fonction $F(x_i)$ est la probabilité de réalisation jusqu'à x_i . Par exemple $F(x_3) = f(x_1) + f(x_2) + f(x_3) = 1$. Voici quelques propriétés de ces distributions.

$$\text{i) } F(\infty) = \sum_{i=1}^{\infty} f(x_i) = 1 ; F(-\infty) = 0$$

$$\text{ii) } \text{Si } x > y, \Rightarrow F(x) \geq F(y)$$

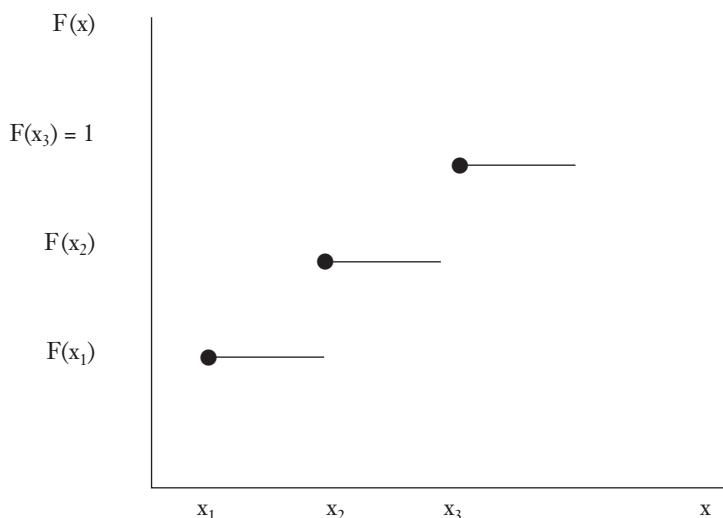
$$\text{iii) } P(x_i \leq X \leq x_j) = F(x_j) - F(x_{i-1}). \text{ Par exemple,}$$

$$P(x_2 \leq X \leq x_4) = F(x_4) - F(x_1)$$

iv) $P(x_i < X) = 1 - P(X \leq x_i)$

v) $P(x_i \leq X \leq x_i) = f(x_i)$

FIGURE 1.12 Fonction de répartition



3.1.2. Espérance et variance en discret

L'espérance

L'espérance mathématique pour une variable aléatoire discrète se définit comme suit :

$$E(X) = \sum_i f(x_i) x_i \text{ où } f(x_i) = P(X = x_i)$$

Plus généralement, $E[g(X)] = \sum_i g(x_i) f(x_i)$. Par exemple, dans le cas précédent : $g(x_i) = x_i$. L'espérance mathématique est donc la

moyenne pondérée des réalisations de la variable aléatoire X sur la population, les facteurs de pondération étant les probabilités respectives de ces réalisations.

Les propriétés de l'opérateur *espérance* sont les suivantes :

i) Pour n variables aléatoires X_1, X_2, \dots, X_n ,

$$E\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n c_i E(X_i) \text{ où } c_i \text{ est une constante.}$$

$$\text{Par exemple, } E(c_1 X_1 + c_2 X_2) = c_1 E(X_1) + c_2 E(X_2).$$

Par ailleurs, la moyenne d'une suite de variables aléatoires $X_1,$

X_2, \dots, X_n , où $X_i \sim \text{NID}^{10}(\mu, \sigma^2)$, est de : $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$,

$$E(\bar{X}) = \frac{\sum_{i=1}^n E(X_i)}{n} = \frac{n\mu}{n} = \mu.$$

La variance

La variance d'une variable aléatoire discrète se définit comme suit.

$$\begin{aligned} V(X) &= E\left[(X - E(X))^2\right] = \sum_i f(x_i)(x_i - E(X))^2 \\ &= \sum_i f(x_i) x_i^2 - [E(X)]^2 = E(X^2) - [E(X)]^2 \end{aligned}$$

Plus généralement, la variance s'écrit : $E[g(X)] = \sum_i g(x_i) f(x_i)$

où $g(x_i) = (x_i - E(X))^2$.

Les propriétés de l'opérateur variance sont les suivantes :

10. NID : variables normales indépendamment distribuées.

Pour n variables aléatoires X_1, X_2, \dots, X_n , on a :

$$i) \quad V\left(\sum_i c_i X_i\right) = \sum_i c_i^2 V(X_i) + 2 \sum_i \sum_j c_i c_j \operatorname{cov}(X_i, X_j) \quad \forall i \neq j$$

où les c_i sont des constantes.

$$ii) \quad \text{Si } c \text{ est une constante} \Rightarrow V(c) = 0$$

$$iii) \quad V(a \pm cX) = c^2 V(X) \text{ où } a \text{ et } c \text{ sont des constantes.}$$

Par exemple, supposons que $X_i \sim \text{NID}(\mu, \sigma^2)$. Alors,

$$V(\bar{X}) = V\left(\frac{\sum_i X_i}{n}\right) = \frac{\sum_i V(X_i)}{n^2} = \frac{\sum_i \sigma^2}{n^2} = \frac{n}{n^2} \sigma^2 = \frac{\sigma^2}{n}$$

3.1.3. Distribution binômiale

La loi binômiale se définit comme suit :

$$b(i; n, p) = P(X = i) = C_i^n p^i (1-p)^{n-i}, \quad i = 0, \dots, n, \text{ où } 0 < p < 1$$

et où $C_i^n = \binom{n}{i} = \frac{n!}{i!(n-i)!}$. Cette fonction permet de calculer la

probabilité associée à i succès parmi n expériences. La distribution binômiale sert entre autres à modéliser des expériences du type succès-échec, chaque expérience étant indépendante l'une de l'autre. L'exemple classique de ce type d'expérience est le lancer répété d'une pièce de monnaie. C'est le jeu communément appelé pile ou face. Si on la lance 3 fois, alors la probabilité d'avoir pile deux fois est de :

$$P(X = 2) = \frac{3!}{2!(3-2)!} \left(\frac{1}{2}\right)^2 \left(1 - \frac{1}{2}\right)^{3-2} = \frac{3}{8} \text{ où } p, \text{ soit la probabilité}$$

d'avoir pile, est égale à $\frac{1}{2}$. Il est à noter que l'on pourrait également utiliser cette formule pour calculer la probabilité d'avoir au moins i succès parmi n expériences.

La distribution binômiale présente les propriétés suivantes. Son espérance $E(X) = np$. Sa variance est de : $V(X) = np(1-p)$.

Voici maintenant un exemple d'utilisation de la loi binômiale en finance, soit celui de l'évaluation d'une option européenne¹¹. Cet exemple est celui du modèle binomial de Cox, Ross et Rubinstein (1979). Considérons d'abord l'évaluation d'une option d'achat. En vertu de la loi binômiale, la prime de cette option s'évalue par la formule suivante :

$$c = e^{-rt} E^* [\text{Max}(S_t - X, 0)] = e^{-rt} \sum_{i=0}^{i=n} \binom{n}{i} p^i (1-p)^{n-i} \text{Max}[S_0 u^i d^{n-i} - X, 0]$$

où r est le taux sans risque ; t , la durée en années de l'option ; n , le nombre de périodes ; S_0 , la valeur du prix de l'action au temps 0 ; u , le multiple de hausse du prix de l'action ; d , le multiple de baisse ; et X , le prix d'exercice. Les u et d se calculent comme suit : $u = e^{\sigma\sqrt{t/n}}$; $d = e^{-\sigma\sqrt{t/n}}$. Dans ces relations, s représente la volatilité annuelle du rendement de l'action sous-jacente à l'option. Les p sont ici calculés dans un univers neutre au risque et s'obtiennent comme suit :

$$p = \frac{e^{\frac{r}{n}} - d}{u - d}$$

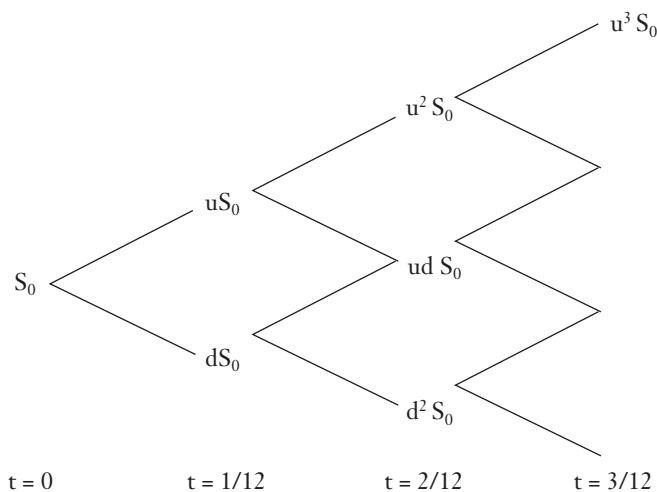
Expliquons plus précisément la formule du prix de l'option d'achat. Selon celle-ci, le prix de cette option est la valeur actualisée de l'espérance des cash-flows à l'échéance de cette option. Cette espérance, notée E^* pour la distinguer de l'espérance classique, est dérivée dans un univers sans risque. Cette formule montre encore une fois que le prix de tout titre est la valeur actualisée de ses cash-flows.

11. L'option européenne, par opposition à l'option américaine, ne peut être exercée avant son échéance.

La version binômiale du prix de l'option précise le calcul de l'espérance. On note que les facteurs de pondération des cash-flows sont

ceux de la binômiale, soit : $\binom{n}{i} p^i (1-p)^{n-i}$.

Illustrons la procédure numérique du calcul du prix de l'option lorsque l'on recourt à la loi binômiale. Il faut alors construire un arbre binômial de l'évolution du prix de l'action puis du prix de l'option. Nous nous en tenons ici à l'arbre des prix de l'action puisque celui des prix de l'option obéit au même principe. Voici comment se présente l'arbre binômial du prix de l'action si l'on suppose trois périodes d'un mois.



On laisse au lecteur le soin de remplir les cases manquantes. Pour plus de détails, on consultera : Racicot, F.-É. et R. Théoret (2000)¹².

12. Racicot, F.-É. et R. Théoret (2000), *op. cit.*

3.1.4. La distribution de Poisson

La distribution de Poisson n'est rien d'autre que la limite de la distribution binômiale sous certaines conditions. Elle se formule comme suit :

$$\lim_{n \rightarrow \infty} b(i; n, p) = \frac{(n \times p)^i e^{-n \times p}}{i!} = P^*(i; n \times p) = P(X = i)$$

Si on pose $\lambda = n \times p$, on a :

$$P(X = i) = P^*(i; \lambda) = \frac{\lambda^i e^{-\lambda}}{i!}$$

L'espérance de cette distribution est de : $E(X) = \lambda$ et sa variance est de : $V(X) = \lambda$. L'espérance de la distribution de Poisson est donc égale à la variance.

Il est à noter que la loi de Poisson peut servir d'approximation à la loi binômiale. La loi de Poisson peut donc servir à modéliser des expériences du type succès-échec si les conditions suivantes sont réalisées : $p \leq 10\%$ et $np < 5$. Par exemple, on veut calculer la probabilité (p) d'obtenir exactement deux pièces défectueuses dans un procédé de production qui maintient en moyenne 4 % de pièces défectueuses lorsque la taille de l'échantillon est $n=100$. Notons que les conditions d'approximation sont ici respectées : $\lambda = 4$, donc plus petit que la borne supérieure de 5, et p est égal à 4 %, donc plus petit que la borne supérieure de 10 %. On obtient le résultat suivant à l'aide de la binômiale :

$$P(X = 2) = b(2; 100, 0,04) = \binom{100}{2} (0,04)^2 (1 - 0,04)^{100-2} = 0,1449$$

L'approximation de cette probabilité par la loi de Poisson est de :

$$P(X = 2) = P^*(2; 4) = \frac{4^2 e^{-4}}{2} = 0,1465$$

On voit que la loi de Poisson se rapproche beaucoup de la loi binômiale.

La distribution de Poisson a donné lieu à de multiples applications. Donnons quelques exemples. Hausman *et al.* (1984)¹³ utilisent la loi de Poisson pour modéliser la distribution des brevets accordés aux entreprises pour une année. Leur distribution de Poisson pour la variable y_i s'écrit comme suit :

$$P(y_i = j) = f(j | R_i) = \frac{(\alpha + \beta R_i)^j e^{-(\alpha + \beta R_i)}}{j!}$$

où $\lambda_i = \alpha + \beta R_i = E(y_i | R_i)$. Dans cette équation, R_i représente les dépenses en recherche et développement d'une entreprise représentative et j , le nombre de brevets accordés à cette entreprise dans une année. Selon cette équation, plus une entreprise investit en recherche, c'est-à-dire plus R est important, plus l'espérance des brevets qui lui sont accordés est élevée. Une autre application, cette fois-ci financière, a été formulée par Greene (1995)¹⁴ et a trait aux cartes de crédit. La variable qui dans son modèle obéit à une loi de Poisson est le nombre de défauts enregistrés dans l'histoire du crédit d'un échantillon de clients pour une catégorie de cartes de crédit. Cette variable est en fait la plus importante pour déterminer si une demande d'emprunt ou de carte de crédit sera acceptée. La variable dépendante est une variable discrète de score qui mesure le nombre de défauts observés durant l'histoire du crédit d'un client. Par exemple, si un client n'a jamais fait défaut, son score est nul. S'il a enregistré deux défauts de paiement, son score est de 2. Et ainsi de suite. La probabilité que y_i prenne la valeur j est la suivante :

$$P(y_i = j) = \frac{\lambda_i^j e^{-\lambda_i}}{j!}$$

Encore une fois, l'espérance de y_i est λ_i qui peut être représentée par une forme fonctionnelle linéaire, non linéaire ou tout autre forme pertinente.

-
13. Hausman, J., B. Hall et Z. Griliches (1984), « Economic Models for Count Data with an Application to the Patents R&D Relationship », *Econometrica*, 52, p. 909-938.
14. Greene, W.A. (1995), Sample Selection in the Poisson Regression Model, document de travail , #EC-95-6, Stern School of Business.

3.2. Loïs de probabilité continues univariées et concepts bivariés

La loi normale ou gaussienne est traditionnellement la plus utilisée en économétrie. Sa fonction de densité (pdf)¹⁵ s'écrit comme suit :

$$f(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2}$$

La variable X suit une loi normale centrée réduite désignée par : $X \sim N(\mu, \sigma^2)$. Les propriétés de la pdf normale sont les suivantes :

- i) La fonction $f(x) \geq 0, \forall x \in (-\infty, \infty)$.¹⁶
- ii) $\int_{-\infty}^{\infty} f(x) dx = F(\infty) = 1$. F représente ici la cdf¹⁷ d'une variable X quelconque.
- iii) $F(-\infty) = 0$
- iv) $f(x) = \frac{dF(x)}{dx}$
- v) $P(x < X < x + dx) = f(x)dx$
- vi) $P(X = x) = \int_x^x f(x)dx = F(x) - F(x) = 0$.
- vii) $P(x_i \leq X \leq x_j) = \int_{x_i}^{x_j} f(x)dx = F(x_j) - F(x_i)$.

Cette relation est aussi égale à :

$$\begin{aligned} P(x_i \leq X \leq x_j) &= P(x_i < X < x_j) \\ &= P(x_i \leq X < x_j) = P(x_i < X \leq x_j). \end{aligned}$$

15. pdf est l'abréviation de *probability density function*.

16. Pour simplifier la notation, nous omettons l'indice i à la variable x .

17. cdf est l'abréviation de *cumulative density function* et représente la probabilité suivante : $P(X \leq x)$.

À la suite de ces relations, plusieurs remarques s'imposent.

- a) Lorsqu'une variable est continue, les probabilités ne se calculent pas par référence à un point mais bien par rapport à une densité. Ainsi, la probabilité que la variable aléatoire X prenne une valeur ponctuelle est égale à 0 alors que ce n'était pas le cas pour les variables discrètes.
- b) Si la variable X est définie par : $X \sim N(\mu, \sigma^2)$, alors, $E(X) = \mu$ et $V(X) = \sigma^2$. La statistique, définie comme : $Z = \frac{X - \mu}{\sigma}$, suit une loi normale centrée réduite notée $N(0,1)$. L'espérance de Z , notée $E(Z)$, est de :

$$E(Z) = E\left(\frac{X - \mu}{\sigma}\right) = \frac{E(X) - \mu}{\sigma} = \frac{\mu - \mu}{\sigma} = 0.$$

Par ailleurs, la variance de Z , notée $V(Z)$, est de :

$$V(Z) = V\left(\frac{X}{\sigma}\right) + V\left(\frac{\mu}{\sigma}\right) = \frac{1}{\sigma^2} V(X) + 0 = 1.$$

- c) Le kernel¹⁸ gaussien, qui s'écrit : $K(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$, est une

fonction telle que $\int_{-\infty}^{\infty} K(x)dx = 1$. Notons qu'il existe d'autres kernels statistiques importants : le bipondéré, l'epanechnikov, le rectangulaire et le triangulaire. Ces kernels sont utilisés pour estimer des distributions empiriques à l'intérieur d'une classe très générale d'estimateurs non paramétriques du type Rosenblatt-Parzen.

Nous abordons maintenant les définitions de l'espérance et de la variance en temps continu.

18. La définition du kernel est la suivante. Soit une fonction composée $f(g(x))$. La fonction $g(x)$ est le kernel et f , la fonction extérieure.

L'espérance

Soit une variable aléatoire X continue. L'espérance de X se définit comme suit :

$$E(X) = \int_{-\infty}^{\infty} xf(x) dx$$

Cette intégrale est une intégrale de Riemann habituelle. Par ailleurs, si l'on formulait l'intégrale de la façon suivante : $\int_{-\infty}^{\infty} x dF$, ce serait alors une intégrale Lebesgue-Stieljes.

La variance

La variance de X se définit comme suit :

$$V(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx$$

Plus généralement :

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x) dx$$

Dans le cas de l'espérance simple, $g(x) = x$; dans le cas de la variance : $g(x) = (x - E(X))^2$.

À noter qu'en temps continu, les propriétés des opérateurs de l'espérance et de la variance sont les mêmes qu'en temps discret.

Supposons maintenant deux variables aléatoires X et Y . Calculons la covariance et la corrélation de ces deux variables en temps discret et continu, ce qui nous introduit dans le domaine des concepts bivariés.

La covariance

La covariance entre X et Y se définit comme suit :

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$$

En temps discret, cette expression est égale à :

$$\text{Cov}(X, Y) = \sum_x \sum_y (x - E(X))(y - E(Y))f(x, y)$$

Comme dans le cas de l'espérance, il existe une formulation plus générale pour la covariance qui la ramène dans le domaine de l'espérance :

$$\text{Cov}(X, Y) = \sum_x \sum_y g(x, y)f(x, y)$$

où $g(x, y) = (x - E(X))(y - E(Y))$.

Par ailleurs, en temps continu, la covariance entre X et Y est donnée par :

$$\text{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f(x, y)$$

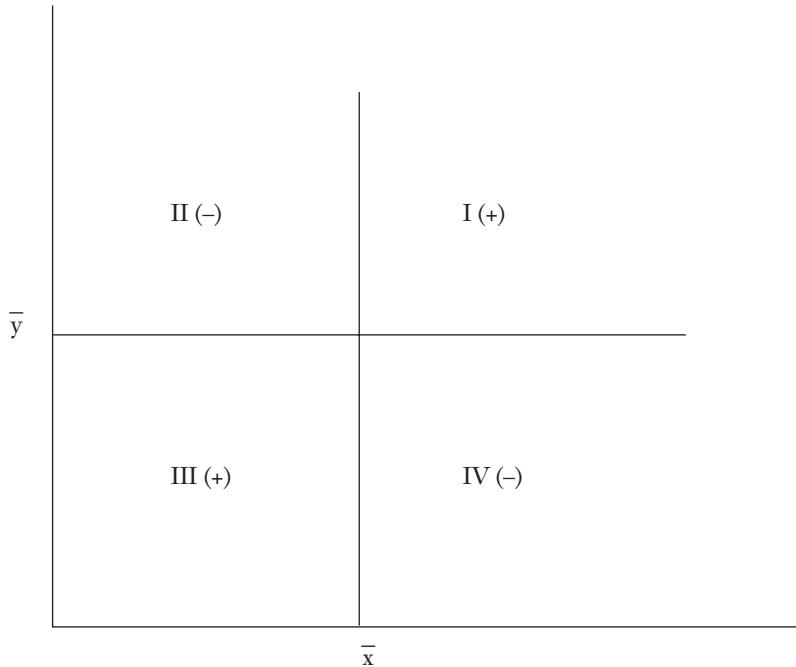
Au plan échantillonnal, la covariance s_{XY} se calcule comme suit :

$$s_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

La figure 1.13 est une représentation graphique de la covariance :

Si les réalisations de X et de Y se distribuent majoritairement dans les quadrants I et III, alors la covariance entre ces deux variables est positive. Si par contre les réalisations de X et de Y se distribuent majoritairement dans les quadrants II et IV, la covariance entre ces deux variables est négative. Finalement, si les réalisations de ces deux variables sont distribuées également entre ces quadrants, la covariance est nulle.

FIGURE 1.13



Corrélation de Pearson

La notion de covariance donne lieu à celle de la corrélation de Pearson $\hat{\rho}$, soit :

$$\hat{\rho} = \frac{s_{XY}}{s_X s_Y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

La corrélation nous indique s'il existe un lien linéaire entre deux variables. Elle corrige la covariance de l'influence des unités de mesure des variables X et Y. Plus la valeur absolue de rho, soit $|\hat{\rho}|$, se

rapproche de 1, plus le lien linéaire entre les deux variables est important. Certes, si ρ est égal à 0, il n'y a pas de lien linéaire. Un exemple théorique montre que si le lien linéaire est parfait, le coefficient de corrélation est égal à 1. Soit $Y = a + bX$. On peut alors écrire :

$$\begin{aligned}\text{Cov}(X, Y) &= E\left[(X - E(X))(Y - E(Y))\right] \\ &= E\left[(X - E(X))(a + bX - E(a + bX))\right] \\ &= bE\left[X - E(X)\right]^2 = bV(X)\end{aligned}$$

Par ailleurs on a :

$$V(Y) = b^2V(X)$$

Cela implique que :

$$\rho = \frac{bV(X)}{\sqrt{V(X)}\sqrt{V(Y)}} = \frac{bV(X)}{b\sqrt{V(X)}\sqrt{V(X)}} = 1$$

Le même exercice peut être refait pour obtenir $\rho = -1$ en supposant que $Y = a - bX$. Il est important de souligner que la corrélation ne mesure que le lien linéaire entre deux variables et que si la corrélation se rapproche de 0, cela n'implique pas qu'il n'existe pas d'autres formes de liens entre ces deux variables. À cet effet, il existe un coefficient de corrélation non linéaire pour capter le lien non linéaire entre deux variables. Ce coefficient sera discuté plus en profondeur à l'intérieur du chapitre ayant trait aux méthodes de régression non linéaires.

Précisons une règle de calcul de la covariance pour les cas continu et discret. Supposons des constantes a, b, c et d , et deux variables aléatoires X et Y , alors on peut écrire :

$$\text{Cov}(aX + bY, cX + dY) = acV(X) + bdV(Y) + (ad + bc)\text{Cov}(X, Y)$$

La distribution uniforme (continue)

La variable aléatoire U est uniformément distribuée dans l'intervalle $[a, b]$ désigné par $U \sim U(a, b)$ si elle obéit à la pdf suivante :

$$f(u) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq u \leq b \\ 0 & \text{autrement} \end{cases}$$

où u est une réalisation de U .

Les propriétés de cette distribution sont les suivantes :

i) L'espérance de U , soit $E(U)$, est égale à :

$$E(U) = \frac{a+b}{2}$$

La démonstration est la suivante.

$$\begin{aligned} E(U) &= \int_a^b u f(u) du = \int_a^b u \frac{1}{b-a} du = \frac{u^2}{2} \times \frac{1}{b-a} + c \Big|_a^b \\ &= \frac{b^2 - a^2}{b-a} \frac{1}{2} = \frac{(a+b)(b-a)}{b-a} \frac{1}{2} = \frac{a+b}{2} \end{aligned}$$

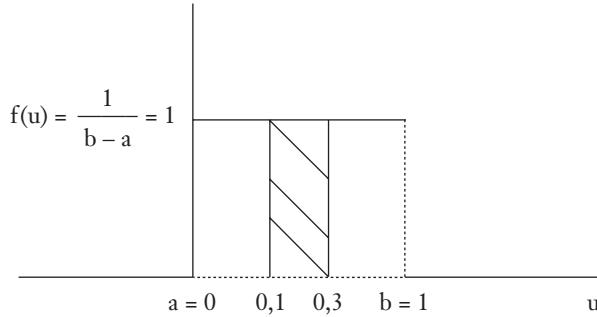
ii) La variance de U , soit $V(U)$, est égale à :

$$V(U) = E\left[(U - E(U))^2\right] = \frac{(b-a)^2}{12}$$

$$\text{où } E(U^2) = \int_a^b u^2 f(u) du = \int_a^b \frac{u^2}{b-a} du.$$

Illustrons la distribution uniforme à partir de l'exemple suivant. On suppose que $a = 0$ et $b = 1$. On obtient alors la distribution uniforme très connue $U(0,1)$. Cette distribution est représentée à la figure 1.14.

FIGURE 1.14



Sur la figure 1.14, la région hachurée nous indique la probabilité que la variable aléatoire U soit inférieure ou égale à $0,3$ ou supérieure ou égale à $0,1$.

La cdf de la distribution uniforme se calcule comme suit :

$$F(u) = P(U \leq u) = \int_{-\infty}^u f(t)dt$$

où :

$$F(u) = \begin{cases} 0 & \text{si } u < a \\ \frac{u - a}{b - a} & \text{si } a \leq u < b \\ 1 & \text{si } u \geq b \end{cases}$$

La fonction $F(u)$ apparaît à la figure 1.15.

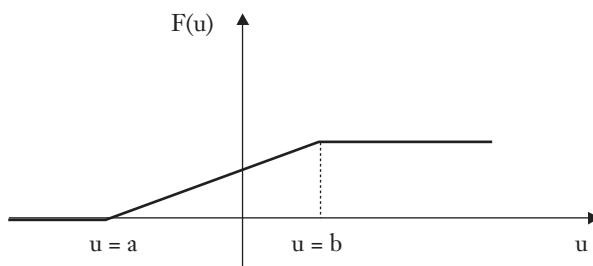
Il importe de fournir quelques exemples d'utilisation de la distribution uniforme :

a) Génération de nombres aléatoires

Les programmes informatiques qui servent à calculer des variables qui obéissent à la loi normale et à ses dérivées sont basés sur un généra-

teur de variables pseudo-aléatoires de loi uniforme. On applique alors une transformation Box-Müller pour passer de la loi uniforme à la loi normale.

FIGURE 1.15



b) Calcul du coefficient de corrélation de Spearman

Le coefficient de corrélation de Pearson n'est pas robuste, en ce sens que lorsque les données dévient de la loi normale, ce coefficient doit être utilisé avec circonspection. Un coefficient de corrélation (non paramétrique) robuste aux déviations de la normalité est celui de Spearman. Il est souhaitable de toujours comparer le coefficient de Pearson à celui de Spearman ; si le premier dévie significativement du second, il serait préférable d'établir des conclusions à partir du coefficient de Spearman.

Pour calculer le coefficient de corrélation non paramétrique de Spearman, on doit dans un premier temps transformer les données en rang ; pour ce faire, on utilise la distribution uniforme $U(a,b)$. Plus précisément, le calcul du coefficient de Spearman s'établit comme suit pour deux variables X et Y :

$$\rho_{\text{Spear}} = \frac{\text{Cov}(R,S)}{\sqrt{V(R)V(S)}}$$

L'estimateur de ce coefficient, soit $\hat{\rho}_{\text{Spear}}$, est de :

$$\hat{\rho}_{\text{Spear}} = \frac{\sum_{i=1}^n ((R_i - \bar{R})(S_i - \bar{S}))}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2} \sqrt{\sum_{i=1}^n (S_i - \bar{S})^2}}$$

où R est le rang de X et S, le rang de Y.

On peut illustrer cette formule par un exemple. Les réalisations de X sont les suivantes :

$$x = \begin{pmatrix} 33 \\ 44 \\ 32,3 \\ 24 \\ 55 \end{pmatrix}$$

À partir de ce vecteur, illustrons la transformation de X en rang R. Pour obtenir R, on transforme X comme suit et on numérote les réalisations. On obtient alors R^* :

$$x^* = \begin{pmatrix} 24 \\ 32,3 \\ 33 \\ 44 \\ 55 \end{pmatrix} \rightarrow r^* = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}$$

Par conséquent, x implique le vecteur r suivant :

$$x = \begin{pmatrix} 33 \\ 44 \\ 32,3 \\ 24 \\ 55 \end{pmatrix} \rightarrow r = \begin{pmatrix} 3 \\ 4 \\ 2 \\ 1 \\ 5 \end{pmatrix}$$

où $\bar{r} = 3 = \frac{n+1}{2} = \frac{\sum r_i}{n}$ où $n = 5$. De façon analogue, le vecteur s est égal à :

$$s = \begin{pmatrix} 3 \\ 1 \\ 4 \\ 2 \\ 5 \end{pmatrix}$$

Donc, par cette transformation des données, on obtient une distribution uniforme.

Comme le coefficient de Spearman, le coefficient de Kendall τ est également utilisé dans la pratique comme alternative au coefficient de Pearson et se définit comme suit :

$$\tau = \frac{S}{S^*} = \frac{\# \text{ de paires concordantes} - \# \text{ de paires discordantes}}{\binom{n}{2}}$$

L'avantage de τ sur le coefficient de Spearman est qu'il n'est point besoin d'utiliser le rang pour calculer cette statistique. On compare simplement les grandeurs relatives des combinaisons possibles :

$$C_2^n = \frac{n!}{2!(n-2)!} = \frac{n(n-1)(n-2)!}{2!(n-2)!} = \frac{1}{2}n(n-1) \text{ de paires : } (x_i, y_i).$$

Il faut noter que l'on n'utilise ici aucune distribution pour effectuer les comparaisons. En ce sens, le coefficient de Kendall est encore moins restrictif que celui de Spearman.

4. NOTIONS D'INDÉPENDANCE, DE DENSITÉ JOINTE ET DE DENSITÉ MARGINALE

Pour n variables aléatoires X_1, X_2, \dots, X_n , discrètes ou continues, on dit que l'on a indépendance entre ces variables si l'on peut écrire la densité jointe (continue) des réalisations de ces variables aléatoires comme le produit des distributions marginales de ses réalisations. La densité jointe de X_1, X_2, \dots, X_n s'écrit donc :

$$f(x_1, x_2, \dots, x_n) = f_1(x_1)f_2(x_2)\dots f_n(x_n)$$

où, par exemple,

$f_1(x_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, x_2, x_3, x_4, \dots, x_n) dx_2 dx_3 dx_4 \dots dx_n$ est la densité marginale de la variable aléatoire continue X_1 . Dans le cas discret, la fonction de distribution marginale s'écrit

$$f_1(x_1) = \sum_{x_2} \sum_{x_3} \sum_{x_4} \dots \sum_{x_n} f(x_1, x_2, x_3, x_4, \dots, x_n).$$

Pour deux variables aléatoires X_1 et X_2 , on a comme fonctions de distribution marginales :

$$f(x_1, x_2) = f_1(x_1)f_2(x_2)$$

où $f_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_1$ si X_1 et X_2 sont continues et

$$f_2(x_2) = \sum_{x_1} f(x_1, x_2) \text{ si } X_1 \text{ et } X_2 \text{ sont discrètes.}$$

La conséquence de l'indépendance au chapitre de la covariance est la suivante. Dans le cas de deux variables aléatoires X_1 et X_2 , l'indépendance implique une covariance nulle :

$$\begin{aligned} \text{Cov}(X_1, X_2) &= E\left[(X_1 - E(X_1))(X_2 - E(X_2))\right] \\ &= E(X_1 X_2) - E(X_1)E(X_2) = 0 \end{aligned}$$

Ceci implique :

$$E(X_1 X_2) = E(X_1)E(X_2)$$

Prouvons maintenant ce résultat, à savoir que la covariance est nulle si les deux variables aléatoires sont indépendantes, en utilisant les résultats précédents ayant trait à la forme des distributions jointes et marginales dans le cas où il y a indépendance entre deux variables aléatoires. La preuve est la suivante :

$$E[g(X_1, X_2)] = \sum_{x_1} \sum_{x_2} f(x_1, x_2) g(x_1, x_2)$$

où $g(X_1, X_2) = (X_1 - E(X_1))(X_2 - E(X_2))$ pour la covariance. Dans ce cas,

$$\begin{aligned} E[g(X_1, X_2)] &= \text{Cov}(X_1, X_2) = E\left[(X_1 - E(X_1))(X_2 - E(X_2))\right] \\ &= \sum_{x_1} \sum_{x_2} (x_1 - E(X_1))(x_2 - E(X_2)) f(x_1, x_2) \end{aligned}$$

En vertu de l'indépendance, on a : $f(x_1, x_2) = f_1(x_1)f_2(x_2)$. Alors,

$$\begin{aligned}
\text{Cov}(X_1, X_2) &= \sum_{x_1} \sum_{x_2} (x_1 - E(X_1))(x_2 - E(X_2))f_1(x_1)f_2(x_2) \\
&= \sum_{x_1} (x_1 - E(X_1))f_1(x_1) \sum_{x_2} (x_2 - E(X_2))f_2(x_2) \\
&= \left(\sum_{x_1} f_1(x_1)x_1 - \sum_{x_1} E(X_1)f_1(x_1) \right) \\
&\quad \left(\sum_{x_2} f_2(x_2)x_2 - \sum_{x_2} E(X_2)f_2(x_2) \right) \\
&= (E(X_1) - E(X_1))(E(X_2) - E(X_2)) = 0
\end{aligned}$$

Expliquons ces dernières équations. Par définition,

$$\sum_{x_1} f_1(x_1)x_1 = E(X_1). \text{ Par ailleurs,}$$

$$\sum_{x_1} E(X_1) f_1(x_1) = E(X_1) \sum_{x_1} f_1(x_1) = E(X_1) \times 1 = E(X_1),$$

ce qui démontre la formule. On se rend donc compte que l'indépendance implique une covariance nulle. Mais l'inverse n'est généralement pas vrai sauf pour des variables aléatoires normalement distribuées. Par exemple, supposons une combinaison non linéaire des variables aléatoires X et Y , soit $X^2 + Y^2 = 1$. Leur covariance peut être très rapprochée de 0, mais il existe pourtant un lien évident entre ces deux variables.

5. PROBABILITÉS CONDITIONNELLES ET DENSITÉS CONDITIONNELLES

Soit deux variables aléatoires X et Y . La probabilité de X condition-

$$\text{nelle à } Y \text{ s'écrit : } f(x|y) = \frac{f(x,y)}{f_y(y)} = P(x|y) = \frac{P((X=x) \cap (Y=y))}{P(Y=y)}$$

où $f_y(y)$ est la distribution marginale de y . Ces égalités sont valables en totalité dans le cas discret mais seules les deux premières le sont dans

le continu. Pour n variables aléatoires, le principe est le même. Considérons par exemple trois variables aléatoires X , Y et Z . La densité conditionnelle de Z s'écrit comme suit :

$$f(z|x,y) = \frac{f(x,y,z)}{f(x,y)}$$

De cette équation, il découle que la densité jointe peut s'écrire comme le produit de densités conditionnelles :

$$f(x,y,z) = f(z|x,y)f(x,y) = f(z|x,y)f(y|x)f(x)$$

Si l'on suppose que les variables X , Y et Z sont indépendantes, cette dernière équation s'écrit alors comme suit :

$$f(x,y,z) = f(x)f(y)f(z)$$

ce qui représente le produit des densités marginales.

6. THÉORÈME CENTRAL LIMITE (CAS UNIVARIÉ)

Nous allons maintenant introduire un théorème très important en théorie statistique, soit le théorème central limite. Ce théorème justifie l'utilisation de la loi normale en économétrie. Ce théorème est très important parce que beaucoup de concepts économétriques sont basés sur lui. Par exemple, en théorie asymptotique, on se sert de ce théorème pour formuler certains tests statistiques comme le test de Jarque et Béra qui, à la limite, suit une distribution χ^2 qui, elle, est une normale au carré. Le théorème central limite s'énonce comme suit.

Si X_1, X_2, \dots, X_n sont des variables aléatoires IID (μ, σ^2) et que $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$, alors la statistique $Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n} \frac{\bar{X} - \mu}{\sigma}$

dispose d'une pdf qui s'approche de la loi normale centrée et réduite $N(0,1)$ à mesure que n tend vers l'infini.

Selon ce théorème, la moyenne de n variables indépendantes qui obéissent à une distribution quelconque (en autant que la distribution ait une moyenne et une variance) s'approche donc d'une $N(0,1)$ après l'avoir centrée et réduite sur un échantillon suffisamment grand. Plus formellement, ces résultats se formulent comme suit.

$$\lim_{n \rightarrow \infty} P \left[\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq y \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{1}{2}Z^2} dZ$$

Ceci implique que :

$$\sqrt{n}\bar{X}_n \xrightarrow{d} N(\sqrt{n}\mu, \sigma^2)$$

ou, en termes équivalents,

$$\bar{X}_n \underset{a}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

Il est à souligner qu'il s'agit ici de la convergence en distribution. Il existe en effet d'autres modes de convergence, soit la convergence en probabilité, en moyenne quadratique et la convergence presque sûre. Ces différents types de convergence seront discutés ultérieurement.

7. LA LOI NORMALE MULTIVARIÉE

Nous avons présenté précédemment la loi normale univariée qui s'applique à une série d'observations sur une seule variable aléatoire. Mais il convient de généraliser cette distribution à plusieurs variables aléatoires. On parle alors de distribution normale multivariée. La pdf multivariée s'écrit comme suit :

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{k}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu)\right)$$

où \mathbf{x} est un vecteur de k variables aléatoires supposées dépendantes, μ est le vecteur de leurs moyennes et Σ , la matrice variance-covariance. La distribution de \mathbf{x} s'établit comme suit : $\mathbf{x} \sim N(\mu, \Sigma)$. Si les variables

aléatoires sont indépendantes, alors $\Sigma = \sigma^2 \mathbf{I}$, \mathbf{I} étant la matrice identité.

Sachant que dans ce cas particulier $|\Sigma| = \sigma^{2k}$, la pdf multivariée s'écrit :

$$\begin{aligned} & \frac{1}{(2\pi\sigma^2)^{\frac{k}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{I}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \\ &= \prod_{i=1}^k \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(x_i - \mu_i)^2}{\sigma^2}\right) = f(x_1)f(x_2)\dots f(x_k) \end{aligned}$$

Par conséquent, la densité jointe est le produit des densités marginales dans le cas de variables indépendantes. On retrouve le résultat précédent, à l'effet qu'une covariance (corrélation) nulle entre des variables normalement distribuées donne lieu à l'indépendance statistique.

Dans le cas bivarié, c'est-à-dire que lorsqu'il n'y a que deux variables aléatoires, la pdf s'écrit :

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right]$$

La distribution bivariée est utilisée en finance pour calculer par exemple le prix d'options écrites sur deux actifs ou encore dans le modèle de Whaley¹⁹, qui consiste à calculer de façon analytique le prix d'une option d'achat américaine écrite sur une action versant un dividende.

19. Pour plus de détails, voir : Racicot, F.-É. et R. Théoret (2000), *Traité de gestion de portefeuille : titres à revenus fixes et produits dérivés*, Presses de l'Université du Québec, chapitre 6.

8. ESTIMATION DE LA MOYENNE ET DE LA VARIANCE DANS UN MODÈLE SIMPLE

Supposons que la variable aléatoire Y se formule comme suit :

$$Y_i = \beta + e_i, \quad i = 1, \dots, n$$

où $e_i \sim \text{IID}(0, \sigma^2)$. β est une constante et Y_i est alors une variable aléatoire. e_i est une variable aléatoire appelée terme d'erreur et supposée provenir d'une population de moyenne 0 et de variance σ^2 . Par hypothèse, les e_i sont supposés indépendants, c'est-à-dire $\text{Cov}(e_i, e_j) = 0 \forall i \neq j$. En termes matriciels, on peut réécrire ce modèle comme suit :

$$\mathbf{Y} = \beta \mathbf{1} + \mathbf{e}$$

où

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix}_{n \times 1}, \quad \mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}_{n \times 1} \quad \text{et} \quad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_n \end{bmatrix}_{n \times 1}$$

Notre objectif est d'estimer la moyenne de \mathbf{Y} à partir de ce modèle. Comme $E(Y_i) = E(\beta + e_i) = \beta$, il suffit de trouver un estimateur de β pour obtenir un estimateur de la moyenne de Y . Pour ce faire, nous recourons à la méthode des moindres carrés ordinaires (MCO). Selon cette méthode, il faut minimiser la somme des e_i au carré pour trouver l'estimateur recherché.

Nous disposons de n réalisations de Y_i notées par y_i , $i = 1, \dots, n$. Le principe des MCO consiste à minimiser la somme suivante :

$$S = \sum_{i=1}^n d_i^2$$

où $d_i = \sqrt{(y_i - \beta)^2}$ est une distance euclidienne. Plus précisément, le problème de minimisation est le suivant :

$$\text{Min}_{\beta} S = \text{Min}_{\beta} \sum_{i=1}^n d_i^2 = \text{Min}_{\beta} \sum_{i=1}^n (y_i - \beta)^2$$

On peut développer la sommation comme suit :

$$\sum_{i=1}^n (y_i - \beta)^2 = \sum_{i=1}^n (y_i^2 - 2\beta y_i + \beta^2) = \sum_{i=1}^n y_i^2 - 2\beta \sum_{i=1}^n y_i + n\beta^2$$

où la somme des β^2 est égale à $n\beta^2$. Pour obtenir le minimum de S , il suffit d'égaliser sa dérivée première par rapport à β à 0 :

$$\frac{dS}{d\beta} = 0 - 2 \sum_{i=1}^n y_i + 2n\hat{\beta} = 0$$

Il résulte que :

$$\hat{\beta} = \frac{\sum_{i=1}^n y_i}{n}$$

Donc, par ce problème de minimisation, on a trouvé que l'estimateur de β , soit $\hat{\beta}$, est la moyenne arithmétique de l'échantillon. Cet esti-

mateur est valable pour tout échantillon, c'est-à-dire que : $\hat{\beta} = \frac{\sum_{i=1}^n Y_i}{n}$.

L'espérance de $\hat{\beta}$ est de :

$$E(\hat{\beta}) = E\left(\frac{\sum_{i=1}^n Y_i}{n}\right) = \sum_{i=1}^n E\left(\frac{Y_i}{n}\right) = E(Y_i) = \beta$$

Le biais d'un estimateur est égal à : $\text{biais}(\hat{\beta}) = E(\hat{\beta}) - \beta$. Dans le cas qui nous intéresse, le biais est nul. Par conséquent, si on simule des échantillons en recourant à la méthode de Monte Carlo et que l'on calcule $\hat{\beta}$ pour chaque échantillon, alors en moyenne on trouve que $\hat{\beta}$ est égal à la moyenne de la population, c'est-à-dire β .

La variance de $\hat{\beta}$, ici égal à la moyenne, se calcule comme suit :

$$V(\hat{\beta}) = V\left(\frac{\sum_{i=1}^n Y_i}{n}\right) = \frac{\sum_{i=1}^n V(Y_i)}{n^2} = \frac{\sum_{i=1}^n V(\beta + e_i)}{n^2} = \frac{\sigma^2}{n}.$$

Pour estimer σ^2 , on utilise $\hat{\sigma}^2$ qui est égal à :

$$\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{n-1}$$

où $\hat{e}_i = Y_i - \hat{\beta}$. L'estimateur $\hat{\sigma}^2$ de σ^2 est sans biais parce que $E(\hat{\sigma}^2) = \sigma^2$. La démonstration de ce résultat est présentée à l'appendice de ce chapitre.

9. THÉORÈME DE GAUSS-MARKOV

L'estimateur des MCO de β est l'estimateur BLUE²⁰ de ce paramètre. Un estimateur est dit BLUE si, parmi tous les estimateurs linéaires non biaisés, il est le plus efficace. Cette proposition repose sous l'hypothèse peu restrictive que les $e_i \sim \text{IID}(0, \sigma^2)$. Un estimateur BLUE est par définition : i) sans biais ; ii) efficace, c'est-à-dire à variance minimale dans sa classe. Incidemment, on peut trouver des estimateurs non linéaires biaisés mais à variance inférieure à celle des MCO. L'écart quadratique de tels estimateurs peut même être plus faible que celui des MCO. Notons que l'écart quadratique moyen (EQM) de $\hat{\beta}$ se définit comme suit : $\text{EQM}(\hat{\beta}) = \text{Biais}^2(\hat{\beta}) + V(\hat{\beta})$.

20. BLUE est l'acronyme anglais de l'expression : *Best Linear Unbiased Estimator*.

10. LA MÉTHODE D'ESTIMATION DU MAXIMUM DE VRAISEMBLANCE

On veut estimer le paramètre β de l'équation de y présentée à la section précédente par la méthode du maximum de vraisemblance. On rappelle que y^{21} est égal à :

$$y_i = \beta + e_i$$

e_i suit une $NID(0, \sigma^2)$. La méthode du maximum de vraisemblance requiert que la distribution des résidus soit spécifiée, ce qui n'est pas le cas des MCO comme nous l'indique le théorème de Gauss-Markov. La densité jointe des résidus s'exprime donc comme suit :

$$f(e_1, e_2, \dots, e_n) = f(e_1)f(e_2)\dots f(e_n)$$

en vertu de l'indépendance des résidus. La méthode du maximum de vraisemblance exige que l'on passe des résidus à y . Pour ce faire, on recourt à la transformation jacobienne. Pour circonscrire ce concept, supposons une fonction monotone reliant y à e , représentée à la figure 1.16.

Démontrons informellement la provenance de la transformée jacobienne. À partir de la figure 1.15, on observe que la probabilité de se situer dans l'intervalle Δy est égale à la probabilité de se situer dans l'intervalle Δe , ceci étant dû au fait que y est lié à e par une fonction monotone. Mathématiquement, on peut représenter cette relation comme suit :

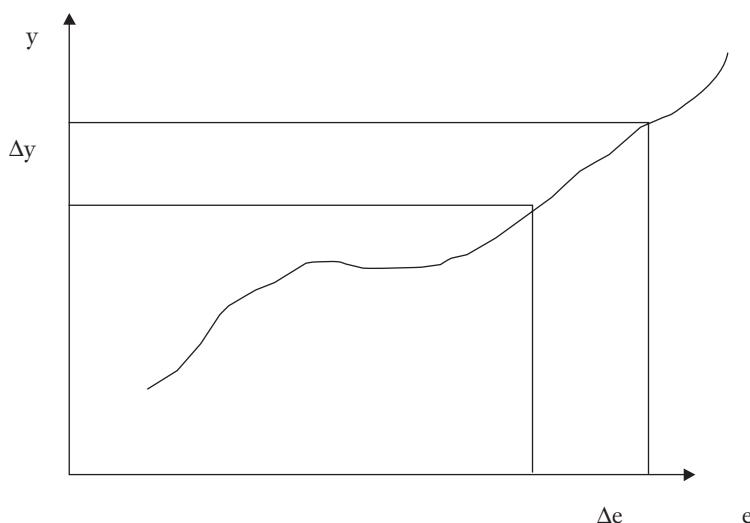
$$\Delta y \times f(y) = \Delta e \times f(e)$$

À partir de cette expression, on obtient la transformation jacobienne en recourant aux transformations suivantes :

$$f(y) = f(e) \frac{\Delta e}{\Delta y}$$

21. Par la suite, dans le but d'alléger l'exposé, on ne fait plus de distinction entre variable aléatoire et réalisation.

FIGURE 1.16



Si on prend la limite quand $\Delta e \rightarrow 0$, on obtient le résultat suivant :

$$f(y) = f(e) \left| \frac{de}{dy} \right|$$

où $|\cdot|$ désigne ici valeur absolue. Si l'on prend la valeur absolue de $\frac{de}{dy}$, c'est pour s'assurer de la positivité de $f(y)$ puisque l'on a supposé une fonction monotone croissante. Généralisons ce résultat au cas de plusieurs variables aléatoires $e_1, e_2, e_3, \dots, e_n$. On a alors, en termes vectoriels :

$$f(\mathbf{y}) = f(\mathbf{e}) \left| \frac{\partial \mathbf{e}}{\partial \mathbf{y}} \right|$$

où $\left| \frac{\partial \mathbf{e}}{\partial \mathbf{y}} \right|$ représente la valeur absolue du déterminant de la matrice de dérivées partielles suivante :

$$\begin{bmatrix} \frac{\partial e_1}{\partial y_1} & \frac{\partial e_1}{\partial y_2} & \cdot & \cdot & \cdot & \frac{\partial e_1}{\partial y_n} \\ \frac{\partial e_2}{\partial y_1} & \frac{\partial e_2}{\partial y_2} & \cdot & \cdot & \cdot & \frac{\partial e_2}{\partial y_n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{\partial e_n}{\partial y_1} & \frac{\partial e_n}{\partial y_2} & \cdot & \cdot & \cdot & \frac{\partial e_n}{\partial y_n} \end{bmatrix}$$

La valeur absolue du déterminant de cette matrice est connue sous le nom de jacobien de transformation de \mathbf{e} à \mathbf{y} . On recourt à ce jacobien pour passer de la fonction de densité de \mathbf{e} à la fonction de densité de \mathbf{y} . Pour simplifier, considérons le cas univarié. On a :

$$e_i = y_i - \beta$$

Alors, $\frac{de_i}{dy_i} = 1$. Ce qui implique la transformation jacobienne suivante : $f(y_i) = 1 \times f(e_i)$. Dans le cas multivarié, la matrice des dérivées partielles est égale à la matrice identité.

Les développements qui précèdent servent à écrire la fonction de densité jointe du vecteur \mathbf{y} à partir de celle du vecteur \mathbf{e} . La densité jointe de \mathbf{y} est de :

$$f(y_1, y_2, \dots, y_n) = f(y_1) f(y_2) \dots f(y_n)$$

où $f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \times \left(\frac{y_i - \beta}{\sigma}\right)^2\right)$. On a ici utilisé la règle de

transformation de la fonction de densité des y_i à celle des e_i qui vient d'être établie. Puisque la densité jointe est ici le produit des densités

marginales, les e_i étant indépendants par hypothèse, et par conséquent les y_i , on peut écrire :

$$f(y_1, \dots, y_n | \beta, \sigma^2) = (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \sum_{i=1}^n \left(\frac{y_i - \beta}{\sigma}\right)^2\right)$$

Notons que $f(\cdot)$ est déterminé en fonction de valeurs données pour β et σ^2 , d'où le caractère conditionnel de la densité jointe. La fonction de vraisemblance est pour sa part une fonction des paramètres, conditionnelle aux valeurs observées de y , c'est-à-dire :

$$L(\beta, \sigma^2 | y_1, y_2, \dots, y_n) = (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \sum_{i=1}^n \left(\frac{y_i - \beta}{\sigma}\right)^2\right)$$

La méthode du maximum de vraisemblance consiste à trouver les paramètres β et σ^2 qui maximisent la probabilité (densité de probabilité) de générer l'échantillon observé. Pour alléger, exprimons cette fonction en termes de logarithme :

$$\ell = \ln L = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2} \sum_{i=1}^n \left(\frac{y_i - \beta}{\sigma}\right)^2$$

Pour estimer β et σ^2 , il suffit de maximiser ℓ par rapport à ces paramètres, c'est-à-dire $\text{Max}_{\beta, \sigma^2} \ell$. Pour ce faire, on égalise les dérivées de ℓ par rapport à ces paramètres à 0 :

$$\frac{\partial \ell}{\partial \beta} = \frac{1}{2} \sum \left(2(y_i - \hat{\beta}) \frac{1}{\sigma^2} \right) = 0$$

On obtient la valeur de β par les manipulations suivantes :

$$\sum (y_i - \hat{\beta}) = 0 \Rightarrow \sum y_i - n\hat{\beta} = 0 \Rightarrow \hat{\beta} = \frac{\sum y_i}{n}$$

Pour obtenir un estimateur de σ^2 , on suit la même procédure que pour β . Pour simplifier le calcul, posons $\sigma^2 = u$. On dérive donc la fonction de vraisemblance par rapport à u et on l'égalise à zéro :

$$\frac{\partial \ell}{\partial \mathbf{u}} = -\frac{n}{2} \frac{1}{\mathbf{u}} - \frac{1}{2} \sum -\mathbf{u}^{-2} (y_i - \hat{\beta})^2 = 0$$

Si on remplace dans cette équation \mathbf{u} par σ^2 , on a :

$$\frac{\partial \ell}{\partial \mathbf{u}} = -\frac{n}{2} \frac{1}{\tilde{\sigma}^2} - \frac{1}{2} \sum -(\tilde{\sigma}^2)^{-2} (y_i - \hat{\beta})^2 = 0$$

En simplifiant, on obtient :

$$\tilde{\sigma}^2 = \frac{\sum (y_i - \hat{\beta})^2}{n}$$

On constate donc que l'estimateur de β donné par le maximum de vraisemblance est identique à celui qui résulte des MCO. Par conséquent, nous nous sommes permis de substituer $\hat{\beta}$ à $\tilde{\beta}$. Cependant, l'estimateur du maximum de vraisemblance de σ^2 est biaisé alors que celui associé des MCO, noté $\hat{\sigma}^2$, ne l'est pas. Donc, en supposant la normalité des résidus, utiliser la méthode des MCO revient à utiliser celle du maximum de vraisemblance en ce qui concerne l'estimateur de β , c'est-à-dire : $\tilde{\beta} = \hat{\beta}$. Mais par contre, l'estimateur du maximum de vraisemblance de la variance est biaisé dans les petits échantillons mais convergent, c'est-à-dire que le biais disparaît dans les grands échantillons.

11. TESTS D'HYPOTHÈSES ET INTERVALLES DE CONFIANCE

Supposons le modèle suivant :

$$y_i = \beta + e_i, \quad i = 1, \dots, n$$

et $e_i \sim \text{NID}(0, \sigma^2)$, où σ^2 est connu. Si on applique les MCO à ce modèle, on obtient alors $\hat{\beta}$. Selon les hypothèses qui concernent les résidus, on peut écrire :

$$\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{n}\right) \Rightarrow (\hat{\beta} - \beta) \sim N\left(0, \frac{\sigma^2}{n}\right) \Rightarrow z = \frac{\hat{\beta} - \beta}{\sigma / \sqrt{n}} \sim N(0,1).$$

Les tests d'hypothèse sont basés sur cette statistique z . Pour tester si β est significativement différent d'une valeur donnée (test bilatéral), c'est-à-dire tester les hypothèses suivantes :

$$H_0: \beta = \beta_0$$

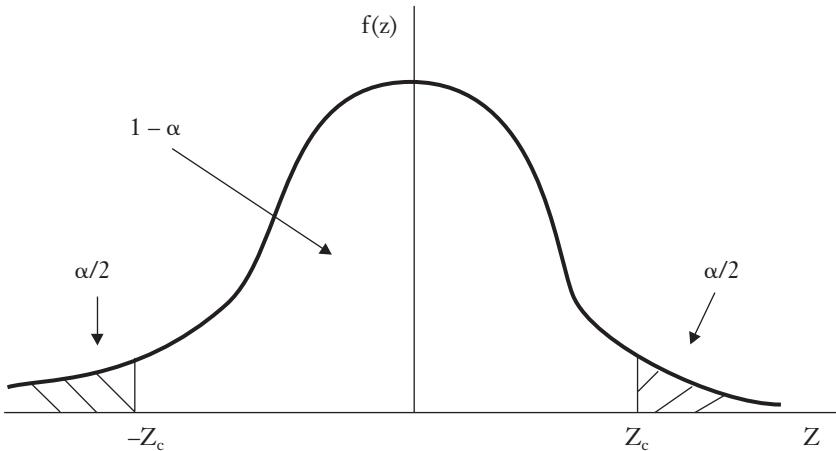
$$H_1: \beta \neq \beta_0,$$

on construit z en remplaçant β par β_0 et le test sera de rejeter H_0 au niveau de signification α si et seulement si :

$$|z| = \left| \frac{\hat{\beta} - \beta_0}{\sigma / \sqrt{n}} \right| > z_c$$

Le test bilatéral est représenté à la figure 1.17.

FIGURE 1.17



$$\frac{\alpha}{2} = P(z > z_c) = P(z < -z_c) = \int_{z_c}^{\infty} f(z) dz = \int_{-\infty}^{-z_c} f(z) dz$$

où $f(z)$ est le kernel de la normale. On peut aussi écrire, en recourant à la cdf :

$$\frac{\alpha}{2} = F(-z_c) = 1 - F(z_c)$$

Par conséquent,

$$F(-z_c) + [1 - F(z_c)] = 1 - P(-z_c < z < z_c) = \alpha$$

Notons que si $F(z_c)$ est monotone, $F(-z_c) = \frac{\alpha}{2} \Rightarrow -z_c = F^{-1}\left(\frac{\alpha}{2}\right)$.

En vertu de cette dernière équation, l'intervalle de confiance de β est : $1 - (1 - P(-z_c \leq z \leq z_c))$. On a donc, en substituant z par sa valeur :

$$P\left(-z_c < \frac{\hat{\beta} - \beta}{\sigma / \sqrt{n}} < z_c\right) = 1 - \alpha$$

En effectuant quelques manipulations, on obtient l'intervalle de confiance pour β :

$$P\left(\hat{\beta} - z_c \frac{\sigma}{\sqrt{n}} < \beta < \hat{\beta} + z_c \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Antérieurement, on a supposé que σ^2 était connu, ce qui n'est plus le cas ici.

Tests et intervalles de confiance lorsque σ^2 n'est pas connu

Test t

En substituant $\hat{\sigma}$ à σ dans l'équation de z , soit $z = \frac{\hat{\beta} - \beta}{\hat{\sigma} / \sqrt{n}}$, on obtient la statistique t de Student :

$$t = \frac{\hat{\beta} - \beta}{\hat{\sigma} / \sqrt{n}} \sim t(n-1)$$

où $\hat{\sigma} = \sqrt{\frac{\sum \hat{e}_i^2}{(n-1)}}$, $\hat{e}_i = y_i - \hat{\beta}$ et $(n-1)$ étant le nombre de degrés de

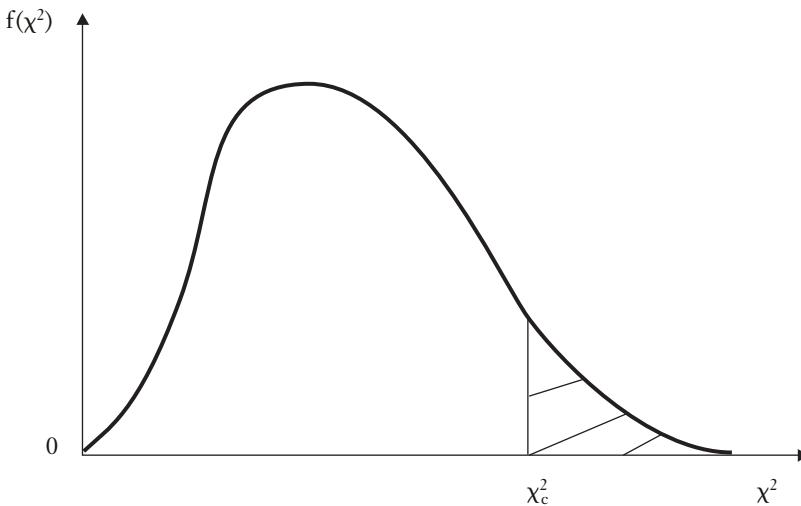
liberté. La statistique t est en fait le ratio d'une normale $N(0,1)$ et d'une

$\sqrt{\frac{\chi^2}{n-1}}$. Ouvrons ici une parenthèse sur la distribution χ^2 . Une $\chi^2(1)$

est une $N(0,1)$ au carré. Plus généralement, la somme de n variables normales au carré suit une distribution $\chi^2(n)$. Plus formellement,

$\sum_{i=1}^n z^2 \sim \chi^2(n)$. L'allure de cette distribution apparaît à la figure 1.18.

FIGURE 1.18



Pour fixer les idées, supposons que : $e_i \sim N(0, \sigma^2)$. Ceci implique que : $\frac{e_i}{\sigma} \sim N(0,1)$.

En vertu des explications antérieures, $\sum_{i=1}^n \left(\frac{e_i}{\sigma}\right)^2 \sim \chi^2(n)$.

Dans le cas où l'on substitue \hat{e}_i à e_i , $\sum_{i=1}^n \left(\frac{\hat{e}_i}{\sigma}\right)^2 = \frac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-1)$,

où $\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{n-1}$. Il reste $(n-1)$ degrés de liberté en raison de l'estimation de la variance qui consomme un degré de liberté parce qu'elle requiert l'estimation de la moyenne. Par conséquent,

$$t = \frac{N(0,1)}{\sqrt{\frac{\chi^2(n-1)}{n-1}}} = \frac{\frac{\hat{\beta} - \beta}{\sigma / \sqrt{n}}}{\sqrt{\frac{(n-1)\hat{\sigma}^2}{(n-1)\sigma^2}}} = \frac{\hat{\beta} - \beta}{\hat{\sigma} / \sqrt{n}}. \text{ CQFD.}$$

Nous voulons maintenant tester si $\hat{\beta}$ est significatif au niveau de confiance α . Ce test, à caractère bilatéral, est le suivant :

$$H_0: \beta = \beta_0$$

$$H_1: \beta \neq \beta_0$$

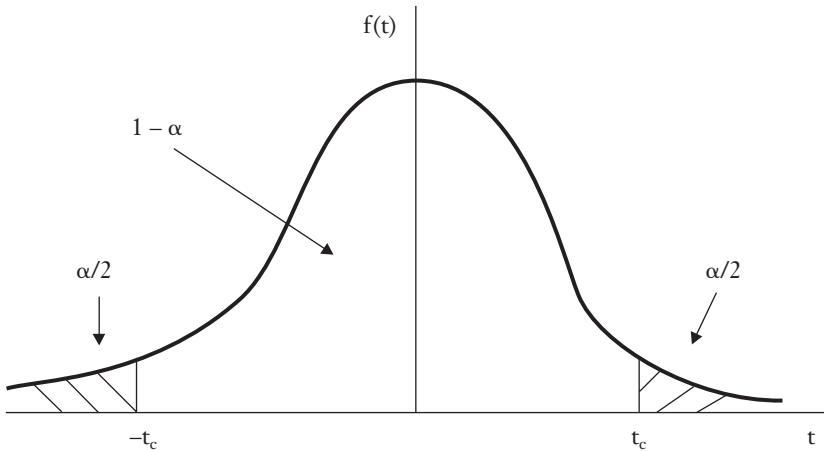
Nous voulons ici effectuer ce test pour le cas du modèle précédent. La règle du test est ici de rejeter H_0 au niveau α , qui est égal à 5 % sauf indication contraire, si :

$$|t| = \left| \frac{\hat{\beta} - \beta_0}{\hat{\sigma} / \sqrt{n}} \right| > t_c(n-1)$$

où t_c est la valeur critique de la distribution t associée à α . La figure 1.19 représente la distribution t . Cette distribution a grosso modo la même forme que la normale à partir d'un nombre de degrés de liberté égal à 120 et s'en rapproche de plus en plus par la suite pour devenir identique à la normale lorsque le nombre de degrés de liberté tend

vers l'infini. En deçà de 120 degrés de liberté, la forme de cette distribution dépend du nombre d'observations et a incidemment des queues plus épaisses que la normale.

FIGURE 1.19



L'intervalle de confiance de β se construit exactement comme dans le cas où σ est connu, sauf qu'il faut remplacer σ par $\hat{\sigma}$. Cet intervalle se construit comme suit :

$$P\left[-t_c < \frac{\hat{\beta} - \beta}{\hat{\sigma} / \sqrt{n}} < t_c\right] = 1 - \alpha$$

Ce qui implique :

$$P\left[\hat{\beta} - t_c \hat{\sigma} / \sqrt{n} < \beta < \hat{\beta} + t_c \hat{\sigma} / \sqrt{n}\right] = 1 - \alpha$$

Cette équation signifie que la probabilité que β soit contenu dans l'intervalle aléatoire est égale à $(1 - \alpha)$. C'est-à-dire que dans 100 échantillons générés, l'intervalle contiendra 95 fois β si α est fixé à 5 %. On écrit aussi de manière plus succincte cet intervalle comme suit :

$$\hat{\beta} \pm t_c \hat{\sigma} / \sqrt{n} .$$

Passons maintenant au test unilatéral sur β . Ce test est le suivant.

$$H_0: \beta \leq \beta_0$$

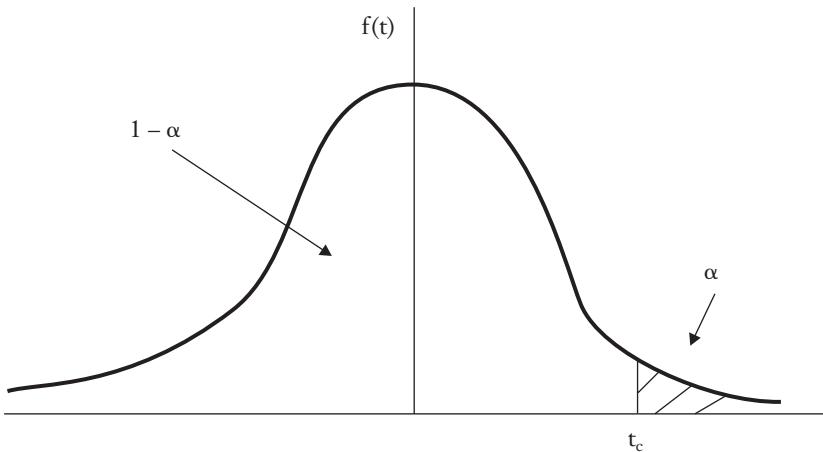
$$H_1: \beta > \beta_0$$

Ce test se formule comme suit :

$$t = \frac{\hat{\beta} - \beta_0}{\hat{\sigma} / \sqrt{n}} \sim t_c(n-1)$$

La règle de décision est la suivante. On rejette H_0 si $t > t_c(n-1)$. La figure 1.20 illustre le test unilatéral sur β .

FIGURE 1.20



Soit l'exemple suivant pour illustrer les deux tests précédents. Dans le cas du test bilatéral, si le nombre de degrés de liberté est supérieur à 120 et que α est supérieur à 5 %, t_c est égal à 1,96, puisque la probabilité d'excéder ce nombre, qui est mesurée par la surface sous la distribution t comprise entre t_c et l'infini, est égale à 0,025. Dans le cas du test unilatéral, tel que l'illustre la figure 1.20, le t_c sera de 1,645, toujours pour un nombre de degrés de liberté supérieur à 120. On trouve ce nombre dans la table de la distribution t qui apparaît à l'annexe X.

Erreur de type I et II et puissance d'un test

À ce stade-ci, il convient d'introduire certains concepts utiles dans la théorie des tests, en l'occurrence l'erreur de type I et la puissance d'un test. L'erreur de type I se définit comme la probabilité de rejeter H_0 alors que cette hypothèse est vraie. Par conséquent, la probabilité de l'erreur de type I est égale à α . Plus précisément,

$$P[|t| > t_c] = P[t > t_c] + P[t < -t_c] = \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha$$

Pour sa part, l'erreur de type II est la probabilité de ne pas rejeter H_0 alors que H_0 est fautive. La probabilité de l'erreur de type I est inversement reliée à la probabilité de l'erreur de type II.

La puissance est définie en termes de l'erreur de type I :

Puissance = nombre de fois qu'un test a commis une erreur de type I

On peut évaluer la puissance d'un test en recourant à la simulation de Monte Carlo. Dans ce contexte, une simulation de Monte Carlo consiste à générer plusieurs échantillons et à calculer le nombre total de fois que le test a commis une erreur de type I sur l'ensemble des échantillons générés.

Un autre concept fréquemment utilisé dans la pratique est la *p-value*. La *p-value* associée à une statistique donnée est la probabilité d'obtenir (par hasard), pour une expérience aléatoire quelconque, une valeur au moins aussi grande, en valeur absolue, que la valeur obtenue pour ladite statistique.

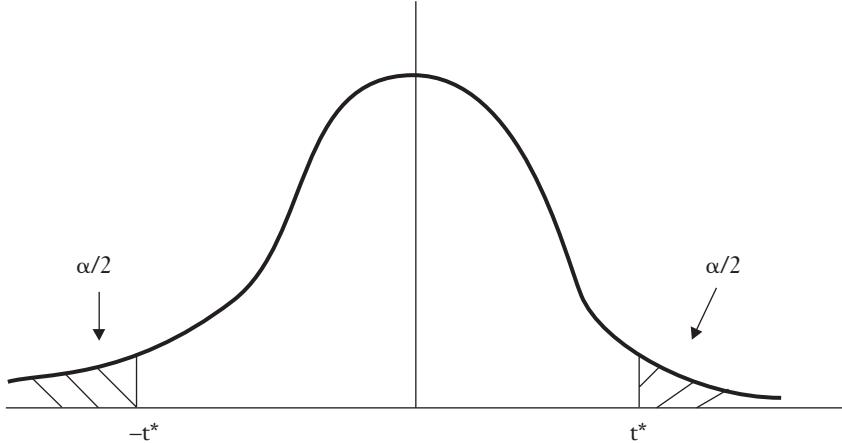
$$\begin{aligned} p\text{-value}(t^*) &= 2(1 - F(|t|, n - 1)) = 2 \times P(|t| > t^*) \\ &= 2 \times P(1 - P(t < t^*)) \end{aligned}$$

où t^* est la valeur estimée de la statistique t .

La figure 1.21 illustre le calcul de la *p-value*. La *p-value* représente la probabilité que la variable aléatoire t soit dans l'une ou l'autre des deux zones hachurées. Au seuil de confiance de 5 %, la *p-value* doit être inférieure à 0,05 pour que la statistique associée à la *p-value* soit significative. Par exemple, dans le cas de la statistique t , une *p-value* inférieure à 0,05 est associée à un t supérieur à 1,96 pour un

nombre de degrés de liberté supérieur à 120. Incidemment, la p-value peut être calculée pour toute autre statistique, telles la F et la chi-carré.

FIGURE 1.21



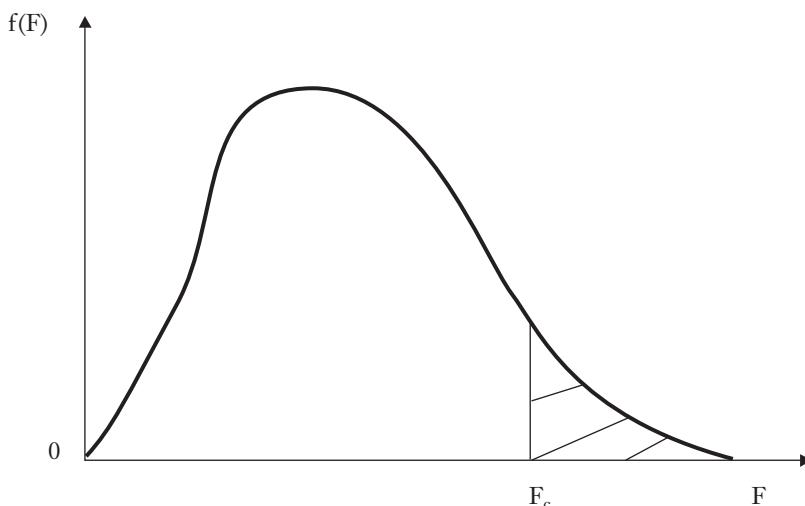
Test F

À partir de la statistique t de Student, soit : $t = \frac{z}{\sqrt{w/(n-1)}}$ où $w = \frac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-1)$ et $z \sim N(0,1)$, en mettant le t au carré, on obtient :

$$t^2 = \frac{z^2}{w/(n-1)} = \frac{\chi^2(1)/1}{\chi^2(n-1)/(n-1)} = F(1, n-1)$$

La distribution F est donc le ratio de deux variables aléatoires de distributions chi-carré indépendantes divisées par leurs degrés de liberté respectifs, soit 1 et $(n-1)$. La distribution F est représentée à la figure 1.22.

FIGURE 1.22



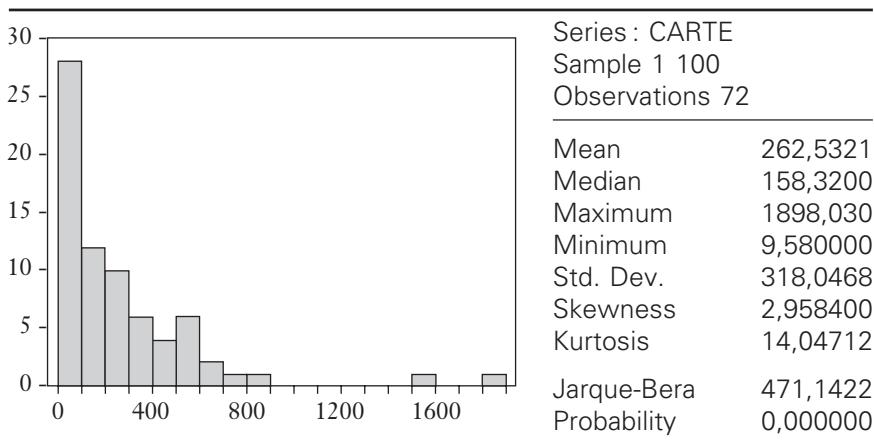
On pourrait remettre en question l'utilité de la statistique t étant donné la relation directe entre t^2 et F . La réponse est simple. Alors que la statistique t permet d'effectuer des tests unilatéraux, ce n'est pas le cas pour la statistique F , puisqu'elle est le carré du t . Nous exposons plus en détails le test F ultérieurement.

12. APPLICATIONS NUMÉRIQUES

Les données utilisées pour illustrer les concepts exposés antérieurement concernent les dépenses mensuelles effectuées à partir de cartes de crédit de 72 individus. Ces données ont été colligées par Greene (1992)²². Au tableau 1.1, on retrouve un histogramme des observations ainsi que les différents moments de la distribution et le test JB de normalité ainsi que sa p -value.

22. Greene, W. (1992), A Statistical Model for Credit Scoring, document de recherche n° EC-92-29, Université de New York, département d'économie, Stern School of Business.

TABLEAU 1.1



Sur ce tableau, on constate que la moyenne des dépenses mensuelles effectuées à partir de cartes de crédit est égale à 262,53 \$. La médiane, à hauteur de 158,32 \$, s'avère ici très différente de la moyenne, ce qui est un signe de non-normalité. L'écart-type de la variable CARTE indiqué par Std.Dev., est de 318,05 \$. Cela représente une variation importante puisque le minimum de cette variable est de 9,58 \$ et le maximum de 1898,03 \$. En effet, les coefficients d'asymétrie et d'aplatissement sont respectivement de 2,96 et de 14,05, ce qui diffère substantiellement des valeurs correspondantes pour la normale, soit 0 et 3. La valeur de la statistique Jarque-Bera est de 471,14 (p-value : 0,00), ce qui indique que la distribution empirique des données n'est pas normale. Ceci était anticipé puisque cette statistique est basée sur l'asymétrie et l'aplatissement.

Au tableau 1.2, l'estimation par les moindres carrés ordinaires du paramètre β représenté par la variable C est de 262,53, ce qui correspond à la moyenne de la variable CARTE qui apparaît au tableau 1.1. L'écart-type de ce coefficient est de 37,48. Supposons que l'on veut tester l'hypothèse $H_0 : \beta \geq 158,32$ (médiane) contre l'hypothèse $H_1 : \beta < 158,32$. Ce test est donc unilatéral. La statistique t associée à la

moyenne est de $\frac{262,53 - 158,32}{37,48} = 2,78$. Par ailleurs, la valeur critique

de t avec $(n - 1) = 71$ degrés de liberté est approximativement de 1,66. Ce test invalide donc l'hypothèse nulle $H_0: \beta = 158,32$ au niveau de $\alpha = 5\%$.

TABLEAU 1.2

Dependent Variable: CARTE				
Method: Least Squares				
Date: 04/13/00 Time: 19:21				
Sample: 1 100 IF X3 > 0				
Included observations: 72				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	262.5321	37.48218	7.004184	0.0000
R-squared	0.000000	Mean dependent var		262.5321
Adjusted R-squared	0.000000	S.D. dependent var		318.0468
S.E. of regression	318.0468	Akaike info criterion		14.37607
Sum squared resid	7181919.	Schwarz criterion		14.40769
Log likelihood	-516.5384	Durbin-Watson stat		1.785861

Dependent Variable: CARTE				
Method: Least Squares				
Date: 04/13/00 Time: 18:15				
Sample: 1 100 IF X3 > 0				
Included observations: 72				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-304.1486	160.7096	-1.892535	0.0626
REVENU	225.6555	74.60287	3.024757	0.0035
REVENU^2	-14.24923	7.185293	-1.983111	0.0513
R-squared	0.239702	Mean dependent var		262.5321
Ajusted R-squared	0.217665	S.D. dependent var		318.0468
S.E. of regression	281.3115	Akaike info criterion		14.15758
Sum squared resid	5460396.	Schwarz criterion		14.25244
Log likelihood	-506.6727	F-statistic		10.87696
Durbin-Watson stat	1.679384	Prob(F-statistic)		0.000078

Dependent Variable: CARTE
Method: Least Squares
Date: 04/13/00 Time: 18:14
Sample: 1 100 IF X3 > 0
Included observations: 72
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-304.1486	167.3797	-1.817118	0.0735
REVENU	225.6555	86.57141	2.606582	0.0112
REVENU^2	-14.24923	7.084070	-2.011447	0.0482
R-squared	0.239702	Mean dependent var	262.5321	
Ajusted R-squared	0.217665	S.D. dependent var	318.0468	
S.E. of regression	281.3115	Akaike info criterion	14.15758	
Sum squared resid	5460396.	Schwarz criterion	14.25244	
Log likelihood	-506.6727	F-statistic	10.87696	
Durbin-Watson stat	1.679384	Prob(F-statistic)	0.000078	

Nous présentons maintenant un exemple qui illustre le calcul de la p-value. On veut tester l'hypothèse que la moyenne des dépenses sur cartes de crédit est égale à 0, soit $H_0: \beta = 0$ (test bilatéral). La statistique t, toujours calculée à partir de l'échantillon précédent, est

égale à : $t^* = \frac{262,53 - 0}{37,48} = 7,00$. La p-value associée à cette statistique

est :

$$p\text{-value}(7,00) = 2P[1 - P(t < 7,00)] = 2P[1 - F(7,00; 71)] = 0,0000,$$

où F est la cdf. Au seuil $\alpha = 5\%$, on rejette fortement H_0 car la p-value est très nettement en deçà de ce seuil. Par ailleurs, l'intervalle de confiance de β est de :

$$\hat{\beta} \pm t_c \frac{\hat{\sigma}}{\sqrt{n}} = 262,53 \pm 1,99 \times \frac{318,04}{\sqrt{72}} = 262,53 \pm 74,58.$$

Le coefficient de corrélation de Pearson pour le cas où l'on veut évaluer la relation linéaire entre les dépenses sur cartes de crédit et le niveau du revenu des individus est de :

$$\hat{\rho} = \frac{\text{Cov}(\text{CARTE}, \text{REVENU})}{\sqrt{V(\text{CARTE}) \times V(\text{REVENU})}} = 0,44.$$

Par ailleurs, le coefficient de corrélation de Spearman, calculé par

$$\hat{\rho}_{\text{Spear}} = \frac{\text{Cov}(\hat{R}, \hat{S})}{\sqrt{\hat{V}(\hat{R})\hat{V}(\hat{S})}} = 0,53. \text{ Finalement le coefficient de corrélation}$$

de Kendall, calculé par $\hat{\tau} = \frac{S}{S^*} = 0,37$. Ces calculs apparaissent au tableau 1.3.

TABLEAU 1.3 **Correlations**

		CARTE	REVENU
CARTE	Pearson Correlation	1.000	.443 **
	Sig. (2-tailed)	.	.000
	N	72	72
REVENU	Pearson Correlation	.443 **	1.000
	Sig. (2-tailed)	.000	.
	N	72	72

** . Correlation is significant at the 0.01 level (2-tailed).

Nonparametric Correlations

	REVENU		REVENU	CARTE
Kendall's tau_b	CARTE	Correlation Coefficient	1.000	.374 **
		Sig. (2-tailed)	.	.000
		N	72	72
	REVENU	Correlation Coefficient	.374 **	1.000
		Sig. (2-tailed)	.000	.
		N	72	72
Spearman's rho	CARTE	Correlation Coefficient	1.000	.534 **
		Sig. (2-tailed)	.	.000
		N	72	72
	REVENU	Correlation Coefficient	.534 **	1.000
		Sig. (2-tailed)	.000	.
		N	72	72

** . Correlation is significant at the .01 level (2-tailed).

© 2001 – Presses de l'Université du Québec

Édifice Le Delta I, 2875, boul. Laurier, bureau 450, Québec, Québec G1V 2M2 • Tél. : (418) 657-4399 – www.puq.ca
Tiré : *Traité d'économétrie financière*, François-Éric Racicot et Raymond Théoret, ISBN 2-7605-1123-5 • D1123N

Tous droits de reproduction, de traduction ou d'adaptation réservés

ANNEXE

Dans cette annexe, nous voulons démontrer que l'estimateur $\hat{\sigma}^2$ de σ^2 est sans biais, c'est-à-dire : $E(\hat{\sigma}^2) = \sigma^2$. Supposons une suite de variables aléatoires IID(μ, σ^2): X_1, X_2, \dots, X_n . On a : $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$. On

élabore le numérateur comme suit :

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_i [(X_i - \mu) - (\bar{X} - \mu)]^2. \text{ On a donc :}$$

$$\begin{aligned} E(\hat{\sigma}^2) &= E \left[\frac{\sum_i [(X_i - \mu) - (\bar{X} - \mu)]^2}{n-1} \right] \\ &= \frac{1}{n-1} E \left[\sum_i [(X_i - \mu) - (\bar{X} - \mu)]^2 \right] \\ &= \frac{1}{n-1} E \left[\sum_i \left[(X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2 \right] \right] \\ &= \frac{1}{n-1} E \left[\sum_i (X_i - \mu)^2 - 2n(\bar{X} - \mu)^2 + n(\bar{X} - \mu)^2 \right] \\ &= \frac{1}{n-1} \left[\sum_i E(X_i - \mu)^2 - nE(\bar{X} - \mu)^2 \right] \end{aligned}$$

Par définition, $E(X_i - \mu)^2 = \sigma^2$ et $E(\bar{X} - \mu)^2 = V(\bar{X}) = \frac{\sigma^2}{n}$. Alors

on a :

$$E(\hat{\sigma}^2) = \frac{1}{n-1} \sum_i \sigma^2 - \frac{n}{n-1} \frac{\sigma^2}{n} = \frac{n}{n-1} \sigma^2 - \frac{\sigma^2}{n-1} = \frac{n-1}{n-1} \sigma^2 = \sigma^2$$

LE MODÈLE LINÉAIRE À DEUX VARIABLES

Ce chapitre vise à présenter le modèle de régression linéaire à une variable explicative. Dans ce contexte, nous exposerons les moindres carrés ordinaires (MCO)¹ pour estimer les paramètres de cette équation. Nous nous attarderons aux diverses formes fonctionnelles que peut prendre l'équation à estimer ainsi qu'aux tests statistiques et intervalles de confiance pertinents. Par la suite, nous dirigerons notre collimateur vers les indices de performance d'une régression et les diverses techniques de prévision en coupe instantanée. Finalement, quelques exemples financiers serviront à illustrer l'estimation du modèle linéaire à deux variables.

1. SPÉCIFICATION DU MODÈLE À DEUX VARIABLES ET PROPRIÉTÉS DES ERREURS RÉSIDUELLES

Le terme *régression* tire son origine de Galton (1889), qui a utilisé ce terme pour désigner le phénomène de régression vers la médiocrité qu'il notait dans une étude sur les tailles relatives des individus. Une régression linéaire se représente comme suit :

$$y_t = \beta_1 + \beta_2 x_t + e_t$$

Dans cette équation, y_t est la variable dépendante ou expliquée. x_t est le régresseur, dite encore variable indépendante ou explicative, où

1. Soit les *Ordinary Least Squares (OLS)* en anglais.

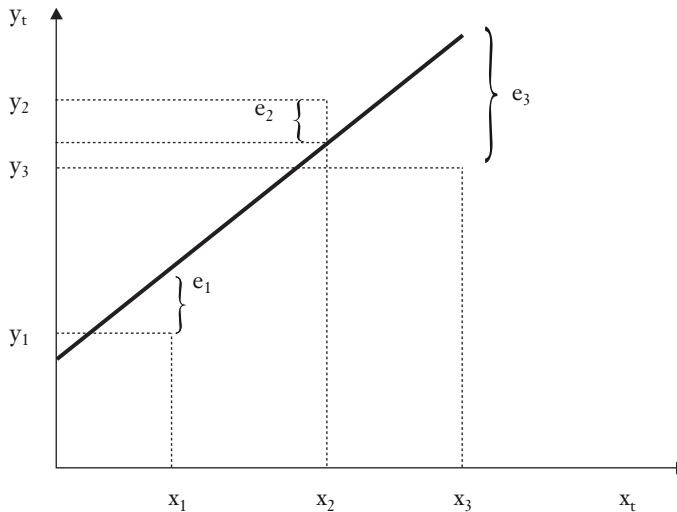
$t = 1, \dots, T$, soit l'indice des observations contenues dans l'échantillon. Dans cette équation linéaire, deux paramètres sont à estimer : β_1 et β_2 .

Les hypothèses du modèle de régression linéaire simple sont les suivantes :

- i) e_t est une variable aléatoire appelée erreur résiduelle IID ($0, \sigma^2$). Cette hypothèse implique que $E(e_t) = 0, \forall t$; que $Cov(e_t, e_s) = E(e_t, e_s) = 0, \forall t \neq s$; que $V(e_t) = E(e_t^2) = \sigma^2, \forall t$. Cette dernière hypothèse est celle de l'homoscédasticité. Cette propriété désigne la constance de la variance des erreurs résiduelles, une fonction scédastique désignant une fonction de la variance, le préfixe *homo* se rapportant au caractère constant de cette fonction.
- ii) x_t est supposé non stochastique, ce qui signifie que dans des échantillons répétés, x_t est fixé. Cette hypothèse implique que : $Cov(x_t, e_t) = 0 \Rightarrow E(x_t e_t) = 0$

La figure 2.1 donne une représentation du modèle de régression linéaire à une variable explicative.

FIGURE 2.1



2. On dit alors que e_t est orthogonal à e_s (ou que e_t est indépendant de e_s).

Comme on peut le constater sur cette figure, il n'existe pas de relation exacte entre y_t et x_t . Par exemple, si $x_t = x_1$, y_1 ne tombe pas exactement sur la droite de régression. Il existe un écart de e_1 par rapport à la droite qui constitue le terme d'erreur de la régression pour cette valeur de x . Et ainsi de suite pour tous les autres x . La droite apparaissant à la figure 2.1 est l'espérance mathématique suivante :

$$E(y_t) = \beta_1 + \beta_2 x_t$$

en vertu des hypothèses précédentes. La variance de y_t est donnée par l'expression suivante :

$$V(y_t) = E\left[(y_t - E(y_t))^2\right] = E\left[(y_t - \beta_1 - \beta_2 x_t)^2\right] = E(e_t^2) = \sigma^2$$

Par conséquent, la variance de la variable aléatoire y_t se confond avec celle des erreurs résiduelles, c'est-à-dire : $y_t \sim \text{IID}(\beta_1 + \beta_2 x_t, \sigma^2)$

2. ESTIMATEUR DES MCO

On vise à estimer les paramètres β_1 et β_2 du modèle linéaire de régression à deux variables. La règle de la minimisation de la somme des erreurs résiduelles au carré introduite au chapitre 1 est de nouveau utilisée ici. Elle se formule comme suit :

$$\text{Min}_{\beta_1, \beta_2} S(\beta_1, \beta_2)$$

où $S(\beta_1, \beta_2) = \sum_{t=1}^T e_t^2 = \sum_{t=1}^T (y_t - \beta_1 - \beta_2 x_t)^2$. En élaborant le carré de

S , on obtient :

$$S(\beta_1, \beta_2) = \sum_t y_t^2 + T\beta_1^2 + \beta_2^2 \sum_t x_t^2 - 2\beta_1 \sum_t y_t - 2\beta_2 \sum_t x_t y_t + 2\beta_1 \beta_2 \sum_t x_t.$$

Pour minimiser $S(\cdot)$, il suffit de calculer les dérivées partielles de cette fonction par rapport aux deux paramètres et de les égaliser à zéro, c'est-à-dire :

$$\frac{\partial S}{\partial \beta_1} = 2T\beta_1 - 2 \sum_t y_t + 2\beta_2 \sum_t x_t; \quad \frac{\partial S}{\partial \beta_2} = 2\beta_2 \sum_t x_t^2 - 2 \sum_t x_t y_t + 2\beta_1 \sum_t x_t$$

$$\frac{\partial S}{\partial \hat{\beta}_1} = 0 \Rightarrow -2\left(-T\hat{\beta}_1 + \sum y_t - \hat{\beta}_2 \sum x_t\right) = 0$$

$$\frac{\partial S}{\partial \hat{\beta}_2} = 0 \Rightarrow -2\left(-\hat{\beta}_2 \sum x_t^2 + \sum y_t x_t - \hat{\beta}_1 \sum x_t\right) = 0$$

Ces deux équations peuvent être réécrites comme suit :

$$\sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t) = \sum \hat{e}_t = 0$$

$$\sum (x_t \hat{e}_t) = 0$$

Ces deux équations sont appelées : équations normales. Exprimée sous forme vectorielle, la dernière équation s'écrit : $\mathbf{x}^T \hat{\mathbf{e}} = 0$. Un vecteur tel que $\hat{\mathbf{e}}$, orthogonal à tout vecteur de l'hyperplan engendré par \mathbf{x} , est dit normal à l'hyperplan. D'où le qualificatif « normal ».

Pour déterminer $\hat{\beta}_1$ et $\hat{\beta}_2$, on solutionne les deux équations normales pour ces deux inconnues. Selon la première équation normale, on a :

$$\sum \hat{e}_t = 0 \Rightarrow \sum y_t = T\hat{\beta}_1 + \hat{\beta}_2 \sum x_t$$

En multipliant cette équation par $\sum x_t$, on a :

$\sum x_t \sum y_t = \sum x_t T\hat{\beta}_1 + \sum x_t \sum x_t \hat{\beta}_2$. Par ailleurs, selon la deuxième équation normale :

$$\sum (x_t \hat{e}_t) = 0 \Rightarrow \sum x_t y_t = \sum x_t \hat{\beta}_1 + \sum x_t^2 \hat{\beta}_2$$

En multipliant cette équation par T et en soustrayant le dernier résultat de cette équation, on a :

$$T \sum x_t y_t - \sum x_t \sum y_t = \hat{\beta}_2 \left(T \sum x_t^2 - \left(\sum x_t \right)^2 \right)$$

Il en résulte que :

$$\hat{\beta}_2 = \frac{T \left(\sum x_t y_t \right) - \left(\sum x_t \sum y_t \right)}{T \sum x_t^2 - \left(\sum x_t \right)^2}$$

On peut simplifier cette expression de la façon suivante :

$$\begin{aligned} \hat{\beta}_2 &= \frac{T(\sum(x_t - \bar{x})(y_t - \bar{y}))}{T(\sum(x_t - \bar{x})^2)} \\ &= \frac{(\sum(x_t - \bar{x})(y_t - \bar{y})) / (T-1)}{(\sum(x_t - \bar{x})^2) / (T-1)} \\ &= \frac{\text{Côv}(x, y)}{\hat{V}(x)} = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x^2} \end{aligned}$$

Connaissant $\hat{\beta}_2$, on déduit $\hat{\beta}_1$ à partir de la première équation normale en la divisant préalablement par T :

$$\frac{\sum y_t}{T} = \hat{\beta}_1 + \hat{\beta}_2 \frac{\sum x_t}{T} \Rightarrow \bar{y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{x} \Rightarrow \hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$

3. PROPRIÉTÉS DES MCO

On peut représenter sous forme matricielle l'équation de la régression linéaire :

$$y = X\beta + e$$

où :

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_T \end{bmatrix}_{T \times 1} \quad X = \begin{bmatrix} x_1 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_T \end{bmatrix}_{T \times 2} \quad e = \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_T \end{bmatrix}_{T \times 1} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}_{2 \times 1}$$

De façon équivalente, on peut représenter ce système comme suit :

$$\mathbf{y} = \mathbf{x}_1\beta_1 + \mathbf{x}_2\beta_2 + \mathbf{e} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \mathbf{e} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

où $\mathbf{e} \sim \text{IID}(\mathbf{0}, \sigma^2\mathbf{I})$: $E(\mathbf{e}) = \mathbf{0}$ et $V(\mathbf{e}) = E(\mathbf{e}\mathbf{e}^T) = \sigma^2\mathbf{I}$. Comme auparavant, \mathbf{X} est une matrice de variables non stochastiques.

Sous forme matricielle, l'estimateur des MCO s'écrit comme suit :

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Les propriétés des MCO sont les suivantes :

- i) les MCO sont sans biais sous les hypothèses énoncées ci-dessus. Cette propriété se démontre comme suit³ :

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}\boldsymbol{\beta} + \mathbf{e}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{e} \\ &= \boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{e} \end{aligned}$$

L'espérance de $\hat{\boldsymbol{\beta}}$ est donc de :

$$E(\hat{\boldsymbol{\beta}}) = E(\boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{e}) = \boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^TE(\mathbf{e}) = \boldsymbol{\beta}$$

cette dernière opération se justifiant par le caractère non stochastique de \mathbf{X} . Il s'ensuit que le biais de $\hat{\boldsymbol{\beta}}$, donné par : $\text{biais}(\hat{\boldsymbol{\beta}}) = E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta}$, est égal à zéro.

- ii) L'estimateur $\hat{\boldsymbol{\beta}}$ est un estimateur linéaire. En effet :

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{A}\mathbf{y} = f(\mathbf{y})$$

la fonction $f(\cdot)$ est une fonction linéaire de \mathbf{y} .

3. Les principales règles du calcul matriciel concernant ce chapitre apparaissent dans l'appendice de celui-ci.

- iii) À l'intérieur de la classe des estimateurs linéaires sans biais, les MCO sont efficaces, c'est-à-dire que pour tout autre vecteur $\hat{\beta}^*$ associé à un autre estimateur linéaire, on a :

$$V(\mathbf{c}^T \hat{\beta}) \leq V(\mathbf{c}^T \hat{\beta}^*)$$

où $\mathbf{c}^T \hat{\beta}$ est une combinaison linéaire des coefficients estimés par la méthode des MCO. Dans le cas qui nous concerne, $\mathbf{c}^T \hat{\beta}$ est égal à :

$$\mathbf{c}^T \hat{\beta} = \begin{bmatrix} c_1 & c_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$$

Si la variance de cet estimateur est égale à celle des MCO et que cet estimateur est sans biais, cet estimateur ne peut être que celui des MCO.

Il convient maintenant de calculer la matrice variance-covariance de $\hat{\beta}$. On a : $V(\hat{\beta}) = E\left[\left(\hat{\beta} - E(\hat{\beta})\right)\left(\hat{\beta} - E(\hat{\beta})\right)^T\right]$. Sachant que

$\hat{\beta} - \beta = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{e}$, on a :

$$\begin{aligned} V(\hat{\beta}) &= E\left[\left((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{e}\right)\left((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{e}\right)^T\right] \\ &= E\left[\left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{e} \mathbf{e}^T \mathbf{X} \left(\mathbf{X}^T \mathbf{X}\right)^{-1}\right] \\ &= \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T E(\mathbf{e} \mathbf{e}^T) \mathbf{X} \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \\ &= \sigma^2 \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \end{aligned}$$

Les principales règles du calcul matriciel apparaissent dans l'appendice.

Telles sont les principales propriétés de l'estimateur des moindres carrés ordinaires. Il importe de spécifier la composition de la matrice variance-covariance de $\hat{\beta}$.

$$\begin{aligned}
\mathbf{V}(\hat{\boldsymbol{\beta}}) &= \mathbf{E} \left[(\hat{\boldsymbol{\beta}} - \mathbf{E}(\boldsymbol{\beta}))(\hat{\boldsymbol{\beta}} - \mathbf{E}(\boldsymbol{\beta}))^T \right] \\
&= \mathbf{E} \left[\begin{bmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 - \beta_1 & \hat{\beta}_2 - \beta_2 \end{bmatrix} \right]_{2 \times 2} \\
&= \mathbf{E} \begin{bmatrix} (\hat{\beta}_1 - \beta_1)^2 & (\hat{\beta}_1 - \beta_1)(\hat{\beta}_2 - \beta_2) \\ (\hat{\beta}_2 - \beta_2)(\hat{\beta}_1 - \beta_1) & (\hat{\beta}_2 - \beta_2)^2 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{E}(\hat{\beta}_1 - \beta_1)^2 & \mathbf{E}((\hat{\beta}_1 - \beta_1)(\hat{\beta}_2 - \beta_2)) \\ \mathbf{E}((\hat{\beta}_2 - \beta_2)(\hat{\beta}_1 - \beta_1)) & \mathbf{E}(\hat{\beta}_2 - \beta_2)^2 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{V}(\hat{\beta}_1) & \mathbf{Cov}(\hat{\beta}_1, \hat{\beta}_2) \\ \mathbf{Cov}(\hat{\beta}_1, \hat{\beta}_2) & \mathbf{V}(\hat{\beta}_2) \end{bmatrix}
\end{aligned}$$

Étant donné que l'on se situe encore dans le cas de la régression à deux variables, on peut représenter, comme on l'a vu antérieurement, la matrice variance-covariance⁴ de façon compacte comme suit :

$$\mathbf{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$$

où $\mathbf{X} = [\mathbf{i} \quad \mathbf{x}]$. Plus explicitement :

4. À noter que l'on utilise indifféremment $\mathbf{Cov}(\cdot)$ ou $\mathbf{V}(\cdot)$ (en gras) pour désigner la matrice variance-covariance.

$$\begin{aligned} \mathbf{Cov}(\hat{\beta}) &= \sigma^2 \left[\begin{matrix} \mathbf{i} & \mathbf{x} \end{matrix} \right]^T \left[\begin{matrix} \mathbf{i} & \mathbf{x} \end{matrix} \right]^{-1} \\ \mathbf{Cov}(\hat{\beta}) &= \sigma^2 \left(\begin{matrix} \mathbf{i}^T & \\ & \mathbf{i} & \mathbf{x} \end{matrix} \right)^{-1} \\ \mathbf{Cov}(\hat{\beta}) &= \sigma^2 \left(\begin{matrix} \mathbf{i}^T \mathbf{i} & \mathbf{i}^T \mathbf{x} \\ \mathbf{x}^T \mathbf{i} & \mathbf{x}^T \mathbf{x} \end{matrix} \right)^{-1} \end{aligned}$$

où $\mathbf{i}^T \mathbf{i} = T$. On a finalement :

$$\begin{aligned} \mathbf{Cov}(\hat{\beta}) &= \sigma^2 \left(\begin{matrix} T & \sum x_t \\ \sum x_t & \sum x_t^2 \end{matrix} \right)^{-1} \\ \mathbf{Cov}(\hat{\beta}) &= \sigma^2 \left(\begin{matrix} \frac{\sum x_t^2}{T \sum (x_t - \bar{x})^2} & \frac{-\bar{x}}{\sum (x_t - \bar{x})^2} \\ \frac{-\bar{x}}{\sum (x_t - \bar{x})^2} & \frac{1}{\sum (x_t - \bar{x})^2} \end{matrix} \right) \end{aligned}$$

Pour calculer la covariance de $\hat{\beta}$, il suffit maintenant d'estimer σ^2 . On a le modèle suivant :

$$\mathbf{y} = \beta_1 \mathbf{i} + \beta_2 \mathbf{x} + \mathbf{e}$$

En appliquant les MCO sur ce modèle, on obtient le vecteur des erreurs résiduelles :

$$\begin{aligned} \hat{\mathbf{e}} &= \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\beta} \\ \hat{\mathbf{e}} &= \mathbf{y} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = \mathbf{M}\mathbf{y} \end{aligned}$$

où $\mathbf{M} = (\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T)$. Cette matrice a la propriété d'être symétrique et idempotente. On a également :

$$\begin{aligned} \mathbf{M}\mathbf{X} &= \mathbf{0} \\ \hat{\mathbf{e}} &= \mathbf{M}\mathbf{y} = \mathbf{M}\mathbf{e} \end{aligned}$$

D'où :

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{e}}^T \hat{\mathbf{e}}}{T-2} = \frac{\sum \hat{e}_i^2}{T-2}$$

$\hat{\sigma}^2$ étant un estimateur non biaisé de σ^2 , on a :

$$E(\hat{\sigma}^2) = \sigma^2$$

Démontrons ce résultat.

$$E(\hat{\sigma}^2) = \frac{E(\hat{\mathbf{e}}^T \hat{\mathbf{e}})}{T-2}$$

Élaborons le numérateur de cette équation.

$$\begin{aligned} E(\hat{\mathbf{e}}^T \hat{\mathbf{e}}) &= E(\mathbf{e}^T \mathbf{M}^T \mathbf{M} \mathbf{e}) = E(\mathbf{e}^T \mathbf{M} \mathbf{e}) = E(\text{tr}(\mathbf{e}^T \mathbf{M} \mathbf{e})) = E(\text{tr}(\mathbf{M} \mathbf{e} \mathbf{e}^T)) \\ &= \text{tr}(\mathbf{M} E(\mathbf{e} \mathbf{e}^T)) = \text{tr}(\mathbf{M} \sigma^2 \mathbf{I}) = \text{tr}(\mathbf{M} \sigma^2) = \sigma^2 \text{tr}(\mathbf{M}) = \sigma^2 (T-2) \end{aligned}$$

où $\text{tr}(\mathbf{M}) = \text{tr}(\mathbf{I}) - \text{tr}(\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T) = \text{tr}(\mathbf{I}) - \text{tr}(\mathbf{X}^T \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1}) = T - 2$.

En substituant ce résultat dans l'espérance de $\hat{\sigma}^2$, on a :

$$E(\hat{\sigma}^2) = \frac{\sigma^2 (T-2)}{(T-2)} = \sigma^2$$

Il va de soi que la démonstration est la même dans le cas multivarié (plusieurs variables explicatives).

4. TESTS D'HYPOTHÈSES ET INTERVALLES DE CONFIANCE

De manière à construire les tests d'hypothèses et les intervalles de confiance sur le modèle de régression suivant :

$$y_t = \beta_1 + \beta_2 x_t + e_t$$

il faut identifier la distribution des résidus. En vertu du théorème central-limite, on suppose que $e_t \sim \text{NID}(0, \sigma^2)$. Nous rappelons que cette spécification peut être vérifiée empiriquement par le test Bera-Jarque. Si $e_t \sim \text{NID}(0, \sigma^2)$, en vertu des MCO, on peut écrire :

$$\hat{\beta}_1 \sim N(\beta_1, V(\hat{\beta}_1)); \hat{\beta}_2 \sim N(\beta_2, V(\hat{\beta}_2)) \text{ où: } V(\hat{\beta}_1) = \frac{\sigma^2 \sum x_t}{T \sum (x_t - \bar{x})^2};$$

$$V(\hat{\beta}_2) = \frac{\sigma^2}{\sum (x_t - \bar{x})^2}; \text{ Cov}(\hat{\beta}_1, \hat{\beta}_2) = \frac{-\sigma^2 \bar{x}}{\sum (x_t - \bar{x})^2}.$$

On dispose maintenant des outils pour construire les tests. On applique la même procédure qu’au chapitre précédent. Étant donné la

distribution de $\hat{\beta}_2$, on a: $\frac{\hat{\beta}_2 - \beta_2}{\sqrt{V(\hat{\beta}_2)}} \sim N(0,1)$. Mais comme σ^2 est

généralement inconnu, on a: $\frac{\hat{\beta}_2 - \beta_2}{\sqrt{\hat{V}(\hat{\beta}_2)}} \sim t(T - 2)$ où on a remplacé

$V(\cdot)$ par $\hat{V}(\cdot)$ et où σ^2 est estimé par $\hat{\sigma}^2 = \frac{\hat{e}'\hat{e}}{T - 2}$. Par conséquent, l’intervalle de confiance pour β_2 est de :

$$\beta_2 \in \left[\hat{\beta}_2 - t_c \sqrt{\hat{V}(\hat{\beta}_2)}, \hat{\beta}_2 + t_c \sqrt{\hat{V}(\hat{\beta}_2)} \right]. \text{ De façon plus concise, on}$$

peut écrire: $\hat{\beta}_2 \pm t_c \sqrt{\hat{V}(\hat{\beta}_2)}$.

5. PRÉVISION

5.1. Prévision de $E(y_0)$

On veut prévoir les dépenses moyennes à partir des cartes de crédit désignées par $E(y_0)$ en fonction du revenu moyen représenté par x_0 . Dans le cadre du modèle de régression linéaire, on a :

$$y_0 = \beta_1 + \beta_2 x_0 + e_0$$

$$\text{où } e_0 \sim N(0, \sigma^2) \Rightarrow y_0 \sim N(\beta_1 + \beta_2 x_0, \sigma^2)$$

Par conséquent,

$$E(y_0) = \beta_1 + \beta_2 x_0 \Rightarrow y_0 = E(y_0) + e_0$$

y_0 est constitué de deux parties : i) $E(y_0)$, soit les dépenses moyennes à partir de cartes de crédit pour un revenu y_0 ; ii) e_0 , une erreur hors-échantillon. La valeur exacte de y_0 est donc inconnue car β_1 , β_2 et e_0 sont inconnus. Par contre, on obtient $\hat{\beta}_1$ et $\hat{\beta}_2$ par les moindres carrés ordinaires et on sait que $E(e_0) = 0$ et que cette erreur n'est corrélée avec aucune autre, i.e. $\text{Cov}(e_0, e_t) = 0 \quad \forall t \neq 0$. L'estimateur de $E(y_0)$ est obtenu en remplaçant β_1 et β_2 par $\hat{\beta}_1$ et $\hat{\beta}_2$. On obtient donc la prévision suivante : $\hat{y}_0 = \hat{E}(y_0) = \hat{\beta}_1 + \hat{\beta}_2 y_0$. L'estimateur \hat{y}_0 est un estimateur sans biais de y_0 puisque :

$E(\hat{y}_0) = E(\hat{\beta}_1) + x_0 E(\hat{\beta}_2) = \beta_1 + \beta_2 x_0$. En fait, l'estimateur $E(\hat{y}_0)$ est le meilleur estimateur linéaire de $E(y_0)$.

La variance de cet estimateur se calcule comme suit :

$$V(\hat{y}_0) = E[\hat{y}_0 - E(y_0)]^2 = E\left(\left(\mathbf{x}_0^T (\hat{\beta} - \beta)\right)\left(\mathbf{x}_0^T (\hat{\beta} - \beta)\right)^T\right)$$

$$\text{où } \mathbf{x}_0 = \begin{bmatrix} 1 \\ x_0 \end{bmatrix} \text{ et } \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}.$$

$$\begin{aligned} V(\hat{y}_0) &= E\left(\mathbf{x}_0^T (\hat{\beta} - \beta)(\hat{\beta} - \beta)^T \mathbf{x}_0\right) \\ &= \sigma^2 \mathbf{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_0 = \sigma^2 \left[\frac{1}{T} + \frac{(x_0 - \bar{x})^2}{\sum (x_0 - \bar{x})^2} \right] \end{aligned}$$

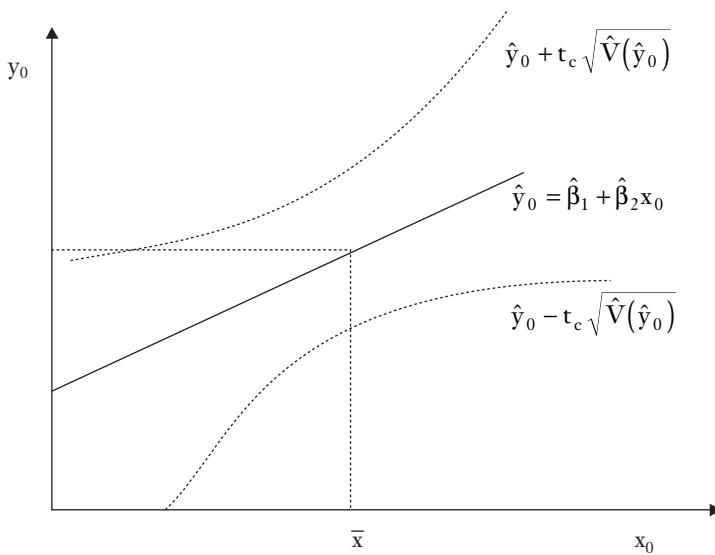
$$\text{L'estimateur de } V(\hat{y}_0) \text{ est } \hat{V}(\hat{y}_0) = \hat{\sigma}^2 \left[\frac{1}{T} + \frac{(x_0 - \bar{x})^2}{\sum (x_t - \bar{x})^2} \right]$$

Fort de ces résultats, nous sommes en mesure de construire l'intervalle de confiance de $E(y_0)$:

$$\hat{y}_0 \pm t_c \sqrt{\hat{V}(\hat{y}_0)}$$

Cet intervalle est représenté à la figure 2.2.

FIGURE 2.2



5.2. Prédiction de y_0

Nous voulons prédire les dépenses effectuées à partir de cartes de crédit et non les dépenses moyennes comme dans la section précédente pour un revenu x_0 donné. On a le modèle suivant:

$$y_0 = \beta_1 + \beta_2 x_0 + e_0$$

Pour obtenir la prédiction de y_0 , on procède comme dans la section précédente. On obtient:

$$\hat{y}_0 = \hat{\beta}_1 + \hat{\beta}_2 x_0$$

On calcule ici la variance de l'erreur de prévision.

$$V(\hat{y}_0 - y_0) = \sigma^2 \left[1 + \left(x_0^T (\mathbf{X}^T \mathbf{X})^{-1} x_0 \right) \right] = \sigma^2 \left[1 + \frac{1}{T} + \frac{(x_0 - \bar{x})^2}{\sum (x_t - \bar{x})^2} \right]$$

6. MESURES DU DEGRÉ D'AJUSTEMENT

La mesure la plus utilisée pour évaluer le degré d'ajustement d'une régression est le coefficient de détermination, désigné par R^2 . Il se définit comme suit :

$$R^2 = 1 - \frac{SCR}{SCT} = 1 - \frac{\sum \hat{e}_t^2}{\sum (y_t - \bar{y})^2} = 1 - \frac{\hat{e}^T \hat{e}}{y' y - T \bar{y}^2} = 1 - \frac{\tilde{\sigma}_e^2}{\tilde{\sigma}_y^2}$$

On peut démontrer cette relation comme suit. Soit le modèle de régression suivant : $y_t = \beta_1 + \beta_2 z_t + e_t$. On exprime ce modèle en déviation de la moyenne et on obtient : $y_t - \bar{y} = \beta_1 + \beta_2 x_t + e_t - \bar{y}$, où $\bar{y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{x}$. On a aussi : $y_t - \bar{y} = \hat{y}_t - \bar{y} + \hat{e}_t$.

D'où : $\hat{y}_t - \bar{y} = \hat{\beta}_2 (x_t - \bar{x}) + \hat{e}_t$. En élevant au carré cette dernière équation et en sommant, on obtient :

$$\sum (y_t - \bar{y})^2 = \sum (\hat{y}_t - \bar{y})^2 + 2 \sum (\hat{e}_t (\hat{y}_t - \bar{y})) + \sum (\hat{e}_t)^2,$$

où $(\hat{y}_t - \bar{y})^2 = \hat{\beta}_2^2 (x_t - \bar{x})^2$. Puisque $2 \sum (\hat{e}_t (\hat{y}_t - \bar{y})) = 0$, on obtient alors :

$$\sum (y_t - \bar{y})^2 = \sum (\hat{y}_t - \bar{y})^2 + \sum (\hat{e}_t)^2$$

$$SCT = SCE + SCR$$

où SCT désigne la somme des carrés totale, SCE, la somme des carrés expliqués et SCR, la somme des carrés résiduels. En divisant les deux côtés par SCT, on a :

$1 = \frac{SCE}{SCT} + \frac{SCR}{SCT} = R^2 + \frac{SCR}{SCT} \Rightarrow R^2 = 1 - \frac{SCR}{SCT}$.⁵ Pour sa part, le R^2 ajusté de Theil se calcule comme suit :

$$\bar{R}^2 = 1 - \frac{\frac{\hat{\mathbf{e}}^T \hat{\mathbf{e}}}{T-2}}{\frac{(y'y - T\bar{y}^2)}{T-1}} = 1 - \frac{\hat{\sigma}_e^2}{\hat{\sigma}_y^2}$$

On notera que l'on a ici corrigé pour les degrés relatifs de liberté. Le R^2 traditionnel augmentera tant que l'on ajoutera des variables même si celles-ci sont non significatives, alors que le R^2 ajusté pénalise l'ajout de variables explicatives à un modèle, qu'elles soient significatives ou pas.

Le R^2 s'interprète comme suit. Dans le cas de deux variables aléatoires, ce coefficient est égal au coefficient de corrélation de Pearson au carré. Il mesure donc la force du lien entre les deux variables.

$$R^2 = \frac{(\sum (x_t - \bar{x})(y_t - \bar{y}))^2}{\sum (y_t - \bar{y})^2 \sum (x_t - \bar{x})^2} = \frac{\sum (\hat{\beta}_2 (x_t - \bar{x}))^2}{\sum (y_t - \bar{y})^2} = \hat{\rho}^2. \text{ Dans le cas}$$

où il existe plusieurs variables explicatives, il mesure l'effet qu'ont les variations des variables explicatives sur la variable dépendante. Par exemple, si R^2 est égal à 30 %, cela signifie que 30 % de la variance de la variable dépendante est attribuable à la variance de l'ensemble des variables explicatives.

Autres formes fonctionnelles pour la relation dépenses sur carte de crédit et revenus qui se ramènent au modèle linéaire standard

Le modèle de régression linéaire qui vient d'être exposé ne s'applique pas seulement à l'équation décrivant une relation linéaire entre deux ou plusieurs variables. En effet, certaines formes fonctionnelles peuvent

5. Ce R^2 est dit centré car il est calculé en déviation de la moyenne.

être linéarisées très facilement. Une fois cette transformation opérée, on peut alors estimer ces équations par les MCO. Voici les formes fonctionnelles les plus courantes qui sont facilement linéarisables.

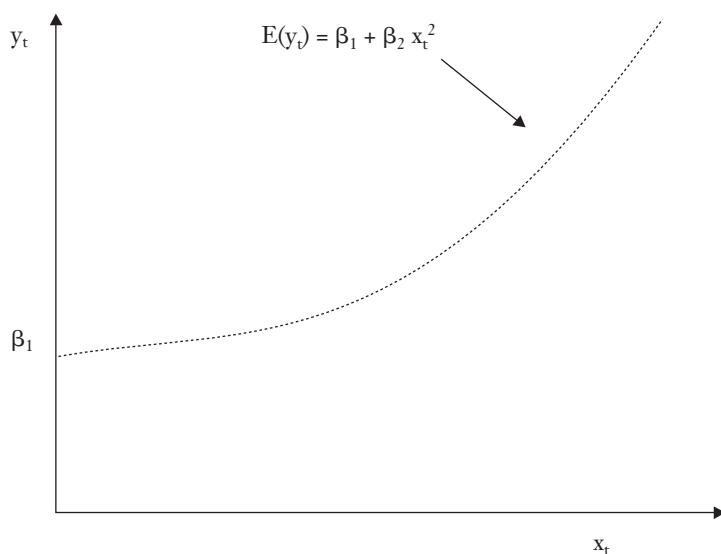
1) Forme quadratique

La forme quadratique s'écrit comme suit :

$$y_t = \beta_1 + \beta_2 x_t^2 + e_t$$

Le graphique de cette relation apparaît à la figure 2.3.

FIGURE 2.3



Mettre les données au carré pourrait viser à capter les non-linéarités que pourraient incorporer les données. La contribution marginale de x dans l'explication de y est évaluée par la dérivée première, soit :

$\frac{dy_t}{dx_t} = 2\beta_2 x_t$. On voit ici que la dérivée n'est pas cons-

tante mais proportionnelle à x . La dérivée seconde : $\frac{d^2 y_t}{(dx_t)^2} = 2\beta_2$.

Dans cet exemple, β_2 est positif. On peut ramener cette forme à la formulation standard antérieure du modèle de régression par la simple transformation suivante : $z = x^2$. D'où on obtient : $y_t = \beta_1 + \beta_2 z_t + e_t$. Du point de vue de la régression, on retrouve le modèle linéaire bien connu.

2) La forme log-linéaire (exponentielle)

Cette relation s'exprime comme suit :

$$y_t = \exp(\beta_1 + \beta_2 x_t + e_t)$$

Par la transformation logarithmique, on obtient le modèle linéaire suivant :

$$\ln y_t = \beta_1 + \beta_2 x_t + e_t$$

La forme log-linéaire contribue à corriger les problèmes d'asymétrie au chapitre de la distribution empirique des observations ayant trait à la variable dépendante. Elle permet également de stabiliser la variance de la variable dépendante.

3) La forme semi-log

Cette forme s'écrit comme suit :

$$\exp(y_t) = \exp(\beta_1 + e_t) x_t^{\beta_2}$$

Après transformation logarithmique, on obtient :

$$y_t = \beta_1 + \beta_2 \ln(x_t) + e_t$$

Cette forme peut servir au traitement du problème de l'hétéroscédasticité dont il sera question dans un autre chapitre.

4) La forme log-log

Cette forme s'écrit comme suit :

$$\exp(y_t) = \alpha_1 x_t^{\beta_2} \exp(e_t)$$

En appliquant la transformation logarithmique, on obtient :

$$\ln y_t = \beta_1 + \beta_2 \ln x_t + e_t$$

ce qui se ramène à un modèle linéaire standard. Cette forme est souvent utilisée pour représenter les fonctions de demande et de production (type Cobb-Douglas). En valeur absolue, l'élasticité de la

fonction est ici mesurée par : $\left| \frac{d \ln y_t}{d \ln x_t} \right| = \left| \frac{\frac{dy_t}{y_t}}{\frac{dx_t}{x_t}} \right| = |\beta_2|$. Quand $|\beta_2| > 1$, la

fonction est dite élastique. Dans le cas inverse, elle est dite inélastique.

L'élasticité est un concept que l'on retrouve fréquemment en finance. À titre d'exemple, dans la théorie des options, on peut calculer l'élasticité du prix d'une option d'achat ou de vente. Pour l'option d'achat, l'élasticité de son prix à celui de l'action sous-jacente se calcule comme suit :

$$\frac{\frac{\partial c}{c}}{\frac{\partial S}{S}} = \frac{S}{c} N(d_1)$$

où c représente le prix de l'option d'achat ; S , le prix de l'action sous-jacente et $N(d_1)$, la probabilité cumulative sous la normale de moins l'infini à d_1 ⁶. Briys *et al.*⁷ donnent l'exemple suivant. On suppose que

6. Pour plus de détails sur ce sujet, voir : Racicot, F.-É. et R. Théoret (2000), *Traité de gestion de portefeuille : titres à revenus fixes et produits dérivés*, Presses de l'Université du Québec, Ste-Foy, chap. 6. On consultera également : Briys, É., M. Bellalah, H.M. Mai et F. de Varenne (1998), *Options, Futures and Exotic Derivatives: Theory, Application and Practice*, Wiley, New York.

7. Briys *et al.*, *op. cit.*

$S = 18$, $X = 15$, $r_f = 10\%$, $T = 0,25$, $\sigma = 15\%$. Alors, $N(2,8017) = 0,997$ et $c = 3,3659$. Alors, l'élasticité du prix de l'option d'achat est de :

$$\text{élasticité} = \frac{18}{3,3659} \times 0,997 = 5,3317. \text{ L'élasticité indique le change-}$$

ment procentuel du prix de l'option quand le prix de l'action varie de 1 %. Ici, un accroissement de 1 % du prix de l'action, c'est-à-dire de 0,18, induit un accroissement de 5,33 % au chapitre du prix de l'option d'achat. Il s'ensuit que lorsque le prix de l'action passe de 18 à 18,18, le prix de l'option d'achat se voit modifié de 3,3659 à 3,5453 ($3,3659 \times (1 + 0,0533)$). Pour sa part, l'élasticité du prix de l'option de vente se calcule comme suit :

$$\text{élasticité} = \frac{S}{p} \frac{\partial p}{\partial S} = -\frac{S}{p} [N(d_1) - 1]$$

En utilisant les données précédentes, on obtient le résultat suivant pour l'élasticité du prix de l'option de vente :

$$\frac{18}{0,0045} [0,997 - 1] = -12. \text{ Cela signifie que le prix de l'option de}$$

vente diminue de 12 % quand le prix de l'action augmente de 1 %.

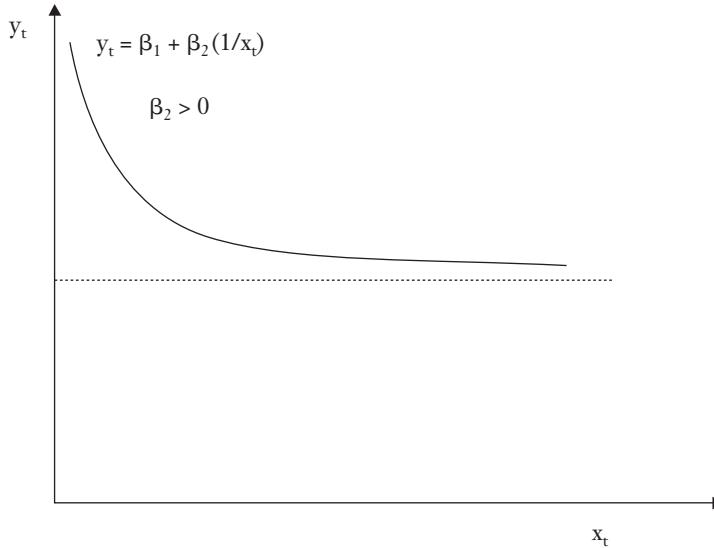
5) La forme réciproque

Cette forme s'écrit comme suit :

$$y_t = \beta_1 + \beta_2 \left(\frac{1}{x_t} \right) + e_t$$

Cette spécification est notamment utilisée pour estimer la courbe de Phillips, qui est la relation entre l'inflation et le taux de chômage. La représentation graphique de cette fonction pour le cas où β_2 est positif apparaît à la figure 2.4.

FIGURE 2.4



6) La forme log-inverse

Cette forme s'exprime comme suit :

$$y_t = \exp\left(\beta_1 + \beta_2 \frac{1}{x_t} + e_t\right)$$

En appliquant la transformation logarithmique sur cette forme, on obtient :

$$\ln(y_t) = \beta_1 + \beta_2 \frac{1}{x_t} + e_t$$

Cette forme s'apparente beaucoup à la forme réciproque sauf que la variable dépendante est exprimée sous forme logarithmique. Quand x augmente, $\ln(y_t)$ diminue.

7. APPLICATIONS

Fort des enseignements précédents, nous revenons dans ce chapitre à l'exemple des dépenses sur carte de crédit. On doit effectuer une régression du type suivant :

$$y_t = \beta_1 + \beta_2 x_t + e_t$$

où y_t représente les dépenses moyennes mensuelles sur carte de crédit⁸ et x_t , le revenu des détenteurs de carte. Sur le fichier original, celui-ci est exprimé sur une base annuelle et est divisé par 10000. On régresse cette équation selon les MCO, qui comprennent 72 individus. Le résultat est présenté au tableau 2.1.

TABLEAU 2.1

Dependent Variable : CARTE				
Method : Least Squares				
Date : 04/13/00 Time : 18:10				
Sample : 1 100 IF X3 >0				
Included observations : 72				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-22.50933	76.78067	-0.293164	0.7703
REVENU	82.93119	20.05219	4.135768	0.0001
R-squared	0.196368	Mean dependent var		262.5321
Ajusted R-squared	0.184888	S.D. dependent var		318.0468
S.E. of regression	287.1440	Akaike info criterion		14.18523
Sum squared resid	5771618.	Schwarz criterion		14.24847
Log likelihood	-508.6683	F-statistic		17.10458
Durbin-Watson stat	1.668577	Prob(F-statistic)		0.000097

À la lecture du tableau 2.1, on constate que le R^2 (R-squared) est de 0,20. Pour sa part, le R^2 ajusté se situe à 0,18. Comme il s'agit ici de données microéconomiques, on s'attend à des R^2 plutôt modérés, de l'ordre de 0,25. En dépit du caractère très modéré du R^2 , on observe que le coefficient associé au revenu, à hauteur de 82,9, a une

8. Soit les dépenses annuelles divisées par 12.

statistique t significative égale à $\frac{82,93 - 0}{20,05} = 4,14$ avec 70 degrés de liberté et une p-value de 0,00, inférieure au seuil critique de 0,05. Pour illustrer, considérons deux exemples.

Exemple 1

On veut maintenant prévoir les dépenses sur carte de crédit associées à un niveau de revenu de 40 000 \$. Les dépenses prévues pour ce niveau de revenu seront de :

$$\hat{y}_0 = -22,51 + 82,93 \left(\frac{40\,000}{10\,000} \right) = 309,21 \$$$

Nous voulons calculer l'intervalle de confiance de cette prévision. Nous devons donc évaluer la variance de celle-ci. Elle est égale à :

$$\hat{V}(\hat{y}_0) = \left[\frac{1}{72} + \frac{(4,00 - 3,44)^2}{205,06} \right] \times (287,14)^2 = 1269,76$$

L'intervalle de confiance de la prévision de $E(y_0)$ est donc de : $309,21 \pm 1,99 \sqrt{1269,76}$ où 1,99 est le t critique correspondant à 70 degrés de liberté.

Exemple 2

L'exercice suivant vise à calculer le ratio de couverture optimal à partir de contrats à terme. Les contrats choisis sont les BAX qui sont transigés à la Bourse de Montréal. Ces contrats sont écrits sur des acceptations bancaires à trois mois. Nous disposons de données journalières sur le prix de ce contrat pour l'année 1999, ce qui équivaut à 250 observations. De façon à calculer le ratio de couverture optimale, on recourt à la méthode de l'appariement des positions au comptant et à terme. En vertu de cette méthode, le nombre optimal de contrats à

terme (NF) que l'on doit détenir pour couvrir de façon optimale le nombre de contrats détenus au comptant (NS) est de :

$$|\text{NF}| = \frac{\Delta S}{\Delta F} \times |\text{NS}|. \text{ Or, } \frac{\Delta S}{\Delta F} = \rho_{\text{SF}} \frac{\sigma_S}{\sigma_F} = \gamma. \text{ On peut calculer ce coeffi-}$$

cient en régressant S sur F, c'est-à-dire : $S_t = \lambda + \gamma F_t + \varepsilon_t$. Pour la période 1999, on obtient l'estimation suivante :

$$S_t = 76,00 + 0,24 F_t + \hat{\varepsilon}_t. \text{ Entre parenthèses, on retrouve les statis-}$$

tiques t. La Durbin-Watson est de 1,77 et le R^2 est 0,99, ce qui indique que les séries semblent cointégrées⁹. Le ratio de couverture, à hauteur de 0,24, est donc plutôt faible pour cette période mais historiquement, on remarque qu'il peut beaucoup fluctuer, du moins à en juger par certaines études de la Banque du Canada¹⁰.

Exemple 3

L'exemple suivant est un test de la relation du CAPM¹¹. On suppose qu'il existe un actif sans risque dont le taux d'intérêt est de r_f . On a la relation suivante qui relie le rendement espéré du titre i à celui du portefeuille du marché¹² :

$$E(r_i) = r_f + \beta_i [E(r_m) - r_f]$$

où $E(r_i)$ est le rendement espéré du titre i, $(E(r_m) - r_f)$, la prime de risque du portefeuille du marché, $E(r_m)$ étant le rendement espéré du portefeuille du marché et β_i , le bêta du titre i. Cette équation est appelée *SML (Security Market Line)*. Cette relation exprime la relation

9. Pour l'explication de ce concept, voir le chapitre 9.

10. Watt, D.G. (1997), *Canadian Short-Term Interest Rates and the BAX Futures Market: An Analysis of the Impact of Volatility on Hedging Activity and the Correlation of Returns between Markets*, document de travail, Banque du Canada.

11. On retrouvera ce test dans Benninga, S. (1997), *Financial Modeling*, MIT Press. Pour une introduction au modèle du CAPM, on consultera : Gagnon, J.M. et N. Khoury (1988), *Traité de gestion financière*, 3^e édition révisée, Gaëtan Morin, Boucherville, chap. 8.

12. À noter qu'il est très courant de déduire le taux sans risque du rendement espéré du titre i pour formuler la SML. Benninga (1997) n'a toutefois pas retenu cette approche.

d'équilibre entre le rendement espéré du titre i et son risque. Selon le CAPM, cette relation est linéaire. β_i mesure la sensibilité du rendement du titre i à celui du portefeuille du marché. Plus β_i est important, plus le rendement espéré du titre i réagit à une variation donnée du rendement espéré du portefeuille du marché. On estime le bêta à partir de l'équation suivante :

$$r_{it} = \alpha_i + \beta_i r_{mt} + e_{it}$$

On suppose les hypothèses habituelles pour le terme d'erreur, entre autres $e_t \sim \text{NID}(0, \sigma^2)$, ce qui est une hypothèse de base du CAPM¹³. Pour estimer cette équation, on dispose de la matrice de données qui apparaît au tableau 2.2.

TABLEAU 2.2

	AMR	BS	GE	HR	MO	UK	S & P's
1974	-0,3505	-0,1154	-0,4246	-0,2107	-0,0758	0,2331	-0,2647
1975	0,7083	0,2472	0,3719	0,2227	0,0213	0,3569	0,3720
1976	0,7329	0,3665	0,2550	0,5815	0,1276	0,0781	0,2384
1977	-0,2034	-0,4271	-0,0490	-0,0938	0,0712	-0,2721	-0,0718
1978	0,1663	-0,0452	-0,0573	0,2751	0,1372	-0,1346	0,0656
1979	-0,2659	0,0158	0,0898	0,0793	0,0215	0,2254	0,1844
1980	0,0124	0,4751	0,3350	-0,1894	0,2002	0,3657	0,3242
1981	-0,0264	-0,2042	-0,0275	-0,7427	0,0913	0,0479	-0,0491
1982	1,0642	-0,1493	0,6968	-0,2615	0,2243	0,0456	0,2141
1983	0,1942	0,3680	0,3110	1,8682	0,2066	0,2640	0,2251

13. Si l'on exclut cette hypothèse, il faut supposer que la fonction d'utilité des individus est quadratique, ce qui semble cependant une hypothèse encore plus restrictive que celle de la normalité du terme d'erreur.

où AMR : American Airlines ; BS : Bethlehem Steel ; GE : General Electric ; HR : International Harvester ; MO : Philip Morris ; UK : Union Carbide ; S&P's : l'indice Standard & Poor's 500, considéré ici comme l'approximation du rendement du portefeuille du marché.

Les rendements sont calculés par la formule suivante : $r_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$.

Précisons davantage la méthode d'estimation des bêtas. Nous sommes ici confrontés à 6 compagnies, donc 6 actions. Dans l'équation précédente, $i = 1, \dots, 6$. Nous disposons de dix années. Considérons le calcul du bêta de la première compagnie ($i = 1$), ici AMR. Sous forme matricielle, la régression s'écrit comme suit pour cette compagnie :

$$r_1 = \alpha_1 + \beta_1 r_{S\&P's} + e_1$$

$$\text{où : } r_1 = r_{AMR} = \begin{pmatrix} r_{11} \\ r_{12} \\ \cdot \\ \cdot \\ \cdot \\ r_{110} \end{pmatrix} = \begin{pmatrix} -0,3505 \\ 0,7083 \\ \cdot \\ \cdot \\ \cdot \\ 0,1942 \end{pmatrix}, r_m = r_{S\&P's} = \begin{pmatrix} r_{m1} \\ r_{m2} \\ \cdot \\ \cdot \\ \cdot \\ r_{m10} \end{pmatrix} = \begin{pmatrix} -0,2647 \\ 0,3720 \\ \cdot \\ \cdot \\ \cdot \\ 0,2251 \end{pmatrix}$$

Les régressions des rendements des compagnies sont effectuées par les MCO. On obtient alors le vecteur suivant des bêtas estimés pour ces six compagnies :

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\beta}_5 \\ \hat{\beta}_6 \end{pmatrix} = \begin{pmatrix} 1,48 \\ 1,08 \\ 1,31 \\ 1,30 \\ 0,26 \\ 0,49 \end{pmatrix}$$

Par ailleurs, pour estimer $E(r_i)$ dans la relation du CAPM, on calcule la moyenne des rendements sur l'ensemble de la période pour la

compagnie i : $\bar{r}_i = \frac{\sum_{t=1}^{10} r_{it}}{10} = \hat{E}(r_i)$. Pour l'ensemble des six compagnies,

on obtient : $\bar{\mathbf{R}} = \begin{pmatrix} \bar{r}_1 \\ \bar{r}_2 \\ \bar{r}_3 \\ \bar{r}_4 \\ \bar{r}_5 \\ \bar{r}_6 \end{pmatrix} = \begin{pmatrix} 0,2032 \\ 0,0531 \\ 0,1501 \\ 0,1529 \\ 0,1025 \\ 0,1210 \end{pmatrix}$

Une fois les vecteurs $\bar{\mathbf{R}}$ et $\hat{\beta}$ estimés, on peut estimer la prime de risque du marché en régressant $\bar{\mathbf{R}}$ sur $\hat{\beta}$ et sur une constante, qui est supposée le taux sans risque. On obtient : $\bar{\mathbf{R}} = r_f + [E(r_m) - r_f] \hat{\beta} + e$, soit : $\hat{E}(r_i) = 0,0766 + 0,0545 \hat{\beta}_i$. Si la constante est le taux sans risque et comme la prime de risque estimée est égale à 0,0545, on peut déduire l'espérance du rendement du marché comme suit : $\hat{E}(r_m) = 0,0766 + 0,0545 = 0,1311$.

La statistique t de la constante est de 1.61 et celle de la prime de risque, 1,24. Le R^2 est de 0,28 et le R^2 ajusté, de 0,10. On en conclut que ce modèle n'est pas vérifié si l'on en juge par la faiblesse des statistiques t et du R^2 . Notons cependant que le nombre de degrés de liberté est très réduit mais il n'en reste pas moins qu'avec un échantillon plus grand, les conclusions pourraient être similaires. Nous avons également recouru ici à une méthode d'estimation très simple. Nous aurions pu sophistiquer davantage en recourant aux techniques du panel et à leurs variantes. Pour estimer le bêta de chaque titre, on aurait pu se servir de la technique ARCH-M sur chaque série en ajoutant la racine carrée de la variance conditionnelle dans l'équation qui sert à estimer le vecteur β . Cette procédure sera analysée au chapitre 10.

LE MODÈLE LINÉAIRE GÉNÉRAL

Jusqu'ici, nous ne nous sommes intéressés qu'à une seule variable explicative. Le chapitre que voici se veut plus général en introduisant plusieurs variables explicatives. Encore une fois, ce chapitre se penche sur les problèmes d'estimation, de spécification, d'inférence et de prévision. Lorsque l'on passe au niveau de plusieurs variables explicatives, force est d'utiliser le calcul matriciel. Les principes du calcul matriciel sont présentés à l'annexe de ce chapitre.

1. FORMULATION MATRICIELLE ET HYPOTHÈSES DE BASE

Soit le modèle linéaire suivant, qui incorpore plusieurs variables explicatives :

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + e_t$$

où $x_{t1} = 1, \forall t$. Si l'on dispose de T observations sur y_t et les x_{tk} , on peut écrire :

$$\begin{aligned} y_1 &= \beta_1 + \beta_2 x_{12} + \beta_3 x_{13} + \dots + \beta_k x_{1k} + e_1 \\ y_2 &= \beta_1 + \beta_2 x_{22} + \beta_3 x_{23} + \dots + \beta_k x_{2k} + e_2 \\ &\dots \\ y_T &= \beta_1 + \beta_2 x_{T2} + \beta_3 x_{T3} + \dots + \beta_k x_{Tk} + e_T \end{aligned}$$

En exprimant ce système d'équations sous forme matricielle, on a :

$$\begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_T \end{pmatrix} = \begin{pmatrix} 1 & x_{12} & x_{13} & \cdot & \cdot & \cdot & x_{1k} \\ 1 & x_{22} & x_{23} & \cdot & \cdot & \cdot & x_{2k} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{T2} & x_{T3} & \cdot & \cdot & \cdot & x_{Tk} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_T \end{pmatrix}$$

Sous forme compacte, ce système s'écrit :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

Les hypothèses de base du modèle classique linéaire général de la régression multiple sont les suivantes :

- i) Les innovations e sont IID $\sim (0, \sigma^2 \mathbf{I}_T)$. Cette hypothèse implique que $E(\mathbf{e}) = \mathbf{0}$; $V(\mathbf{e}) = E(\mathbf{e}\mathbf{e}^T) = \sigma^2 \mathbf{I}_T$. Cette hypothèse implique qu'il n'y a pas de corrélation entre les résidus : $Cov(e_t, e_s) = 0, \forall t \neq s$. De plus, $\sigma_t^2 = \sigma_s^2, \forall t, s$. C'est là l'hypothèse de l'homoscédasticité des résidus.
- ii) Les variables explicatives x sont supposées non stochastiques, c'est-à-dire qu'elles sont fixes dans des échantillons répétés. Cette hypothèse implique que $E(\mathbf{x}^T \mathbf{e}) = \mathbf{0}$. Cette hypothèse est nécessaire pour assurer que l'estimateur des MCO sera sans biais. C'est aussi une condition de moments utilisée dans la méthode des moments généralisés (GMM) qui sera traitée au chapitre 11.

On veut estimer le paramètre $\boldsymbol{\beta}$ du modèle de régression : $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$. À l'instar du chapitre 2, la méthode des MCO consiste à solutionner le problème de minimisation suivant :

$$\text{MIN}_{\boldsymbol{\beta}} S(\boldsymbol{\beta}) = \text{MIN}_{\boldsymbol{\beta}} \sum_{t=1}^T e_t^2, \text{ où } \sum_{t=1}^T e_t^2 = \mathbf{e}^T \mathbf{e}. \text{ On a :}$$

$$\begin{aligned}
 \mathbf{e}^T \mathbf{e} &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\
 &= (\mathbf{y}^T - \boldsymbol{\beta}^T \mathbf{X}^T) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\
 &= (\mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y} - \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta}) \\
 &= \mathbf{y}^T \mathbf{y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y} - \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta}
 \end{aligned}$$

Pour calculer $\hat{\boldsymbol{\beta}}$, on calcule la dérivée suivante :

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^T \mathbf{y} - 2\mathbf{X}^T \mathbf{X}\boldsymbol{\beta}. \text{ On égalise ce résultat à zéro et on obtient :}$$

$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$. Précisons ce calcul en introduisant les formes quadratiques et les techniques de dérivation matricielle.

D’abord la forme quadratique. Sa forme générale est la suivante :

$\mathbf{x}^T \mathbf{A}\mathbf{x} = \sum_i \sum_j x_i x_j a_{ij}$, où \mathbf{A} est une matrice symétrique. Soit l’exemple suivant où $i = 1, 3$ et $j = 1, 3$. Soit $\mathbf{A} = \mathbf{X}^T \mathbf{X}$, cette matrice étant tirée de l’estimateur des MCO et $\mathbf{x} = \boldsymbol{\beta}$ dans cette forme quadratique. Par conséquent :

$$\begin{aligned}
 \boldsymbol{\beta}^T \mathbf{A}\boldsymbol{\beta} &= \sum_{j=1}^3 \sum_{i=1}^3 \beta_i \beta_j a_{ij} = \sum_{j=1}^3 (\beta_1 \beta_j a_{1j} + \beta_2 \beta_j a_{2j} + \beta_3 \beta_j a_{3j}) \\
 &= (\beta_1^2 a_{11} + \beta_2 \beta_1 a_{21} + \beta_3 \beta_1 a_{31}) + \\
 &(\beta_2 \beta_1 a_{12} + \beta_2^2 a_{22} + \beta_3 \beta_2 a_{32}) + (\beta_3 \beta_1 a_{13} + \beta_3 \beta_2 a_{23} + \beta_3^2 a_{33})
 \end{aligned}$$

Parce que \mathbf{A} est une matrice symétrique, on a :

$$\begin{aligned}
 \boldsymbol{\beta}^T \mathbf{A}\boldsymbol{\beta} &= a_{11}\beta_1^2 + 2a_{12}\beta_1\beta_2 + 2a_{13}\beta_1\beta_3 \\
 &+ a_{22}\beta_2^2 + 2a_{23}\beta_2\beta_3 \\
 &+ a_{33}\beta_3^2
 \end{aligned}$$

La dérivée de la forme quadratique s'effectue comme à l'accoutumée, c'est-à-dire :

$$\frac{\partial(\beta^T \mathbf{A}\beta)}{\partial\beta} = \begin{pmatrix} \frac{\partial(\beta^T \mathbf{A}\beta)}{\partial\beta_1} \\ \frac{\partial(\beta^T \mathbf{A}\beta)}{\partial\beta_2} \\ \frac{\partial(\beta^T \mathbf{A}\beta)}{\partial\beta_3} \end{pmatrix} = 2 \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = 2\mathbf{A}\beta = 2\mathbf{X}^T \mathbf{X}\beta.$$

C'est là la dérivée recherchée qui constitue le second terme de $\frac{\partial S(\beta)}{\partial\beta}$, en la faisant certes précéder d'un terme négatif. Pour sa part, son premier terme se calcule comme suit. Posons :

$$\frac{\partial(\beta^T \mathbf{X}^T \mathbf{y})}{\partial\beta} = \frac{\partial(\beta^T \mathbf{a})}{\partial\beta}, \text{ cette substitution :}$$

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \mathbf{X}^T \mathbf{y}$$

étant effectuée pour simplifier les calculs. On a donc : $\beta^T \mathbf{a} = \beta_1 a_1 + \beta_2 a_2 + \beta_3 a_3$. La dérivée partielle recherchée se réduit donc à un simple calcul habituel, c'est-à-dire :

$$\frac{\partial(\beta^T \mathbf{a})}{\partial\beta} = \begin{pmatrix} \frac{\partial(\beta^T \mathbf{a})}{\partial\beta_1} \\ \frac{\partial(\beta^T \mathbf{a})}{\partial\beta_2} \\ \frac{\partial(\beta^T \mathbf{a})}{\partial\beta_3} \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \mathbf{a} = \mathbf{X}^T \mathbf{y}.$$

Nous avons donc démontré comment calculer la dérivée recherchée :

$$\frac{\partial S(\beta)}{\partial \beta} = -2\mathbf{X}^T \mathbf{y} - 2\mathbf{X}^T \mathbf{X} \beta$$

Comme cela a été évoqué plus haut, en égalant cette dérivée à 0, on obtient sous forme matricielle les équations normales énoncées au chapitre 2, c'est-à-dire :

$$\begin{aligned} \mathbf{X}^T \mathbf{y} &= \mathbf{X}^T \mathbf{X} \hat{\beta} \\ \Rightarrow \hat{\beta} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \end{aligned}$$

ce qui est l'équation des MCO. En développant les matrices, on a :

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\beta}_k \end{pmatrix} = \begin{pmatrix} \sum x_{t1}^2 & \sum x_{t1}x_{t2} & \cdot & \cdot & \cdot & \sum x_{t1}x_{tk} \\ \sum x_{t1}x_{t2} & \sum x_{t2}^2 & \cdot & \cdot & \cdot & \cdot \sum x_{t2}x_{tk} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sum x_{t1}x_{tk} & \sum x_{t2}x_{tk} & \cdot & \cdot & \cdot & \sum x_{tk}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum x_{t1}Y_t \\ \sum x_{t2}Y_t \\ \cdot \\ \cdot \\ \cdot \\ \sum x_{tk}Y_t \end{pmatrix}$$

Pour être certain d'obtenir un minimum, on requiert la condition second ordre suivante : $\frac{\partial^2 S(\beta)}{\partial \beta \partial \beta^T} = 2\mathbf{X}^T \mathbf{X}$, soit une matrice définie positive, c'est-à-dire que les déterminants des sous-matrices de $\mathbf{X}^T \mathbf{X}$ soient tous positifs.

2. PROPRIÉTÉS DE L'ESTIMATEUR DES MCO

Dans le cas de la régression multiple, l'estimateur des MCO possède les mêmes propriétés que celles évoquées au chapitre 2 : i) l'estimateur des MCO du vecteur β est sans biais et efficient dans la classe des estimateurs linéaires. Cela se résume par le théorème de Gauss-Markov : sous les hypothèses du modèle classique linéaire général, l'estimateur des MCO $\hat{\beta}$ est l'estimateur BLUE¹ de bêta. Par conséquent, en supposant que les résidus $e_i \sim \text{IID}(0, \sigma^2)$ et que \mathbf{X} est non stochastique, on obtient le meilleur estimateur linéaire du modèle de régression.

Il reste à démontrer ces propriétés. D'abord, celle ayant trait à l'absence de biais. On sait que $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$, où $\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$. En remplaçant \mathbf{y} par sa valeur dans l'expression de $\hat{\beta}$ et en insérant l'espérance à l'intérieur de l'expression, on a : $E(\hat{\beta}) = \beta + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E(\mathbf{e})$ puisque $\mathbf{X} \perp \mathbf{e}$. Par conséquent, $\hat{\beta}$ ne présente pas de biais.

Passons maintenant à la propriété de l'efficacité. Rappelons le calcul de la variance de $\hat{\beta}$. On a : $V(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T] = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E(\mathbf{e}\mathbf{e}^T) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$. À partir de ce résultat, on peut déduire que l'estimateur des MCO est l'estimateur linéaire dont la variance est la plus faible dans la classe des estimateurs linéaires. En effet, si l'on suppose un autre estimateur linéaire $\hat{\beta}^*$, cela revient à dire que $V(\hat{\beta}) \leq V(\hat{\beta}^*)$ et que pour toute autre combinaison linéaire : $V(\mathbf{c}^T \hat{\beta}) \leq V(\mathbf{c}^T \hat{\beta}^*)$.

Envisageons maintenant l'estimateur de la variance de $\hat{\beta}$,

$$\hat{V}(\hat{\beta}) = \hat{\sigma}^2 (\mathbf{X}^T \mathbf{X})^{-1}, \text{ où } \hat{\sigma}^2 = \frac{\hat{\mathbf{e}}^T \hat{\mathbf{e}}}{T - k}, \text{ et où } \hat{\sigma}^2 \text{ est l'estimateur sans}$$

biais de σ^2 puisque le dénominateur de $\hat{\sigma}^2$ représente le nombre de degrés de liberté.

1. Soit le *best linear unbiased estimator*.

R^2 et \bar{R}^2

Le coefficient de détermination R^2 se calcule comme suit à partir de l'équation de SCT, soit la somme des carrés totale: $SCT = SCE + SCR$, où SCE désigne la somme des carrés expliquée et SCR, la somme des carrés résiduelle. En divisant les deux membres de l'équation par SCT, on obtient :

$$R^2 = 1 - \frac{SCR}{SCT} = 1 - \frac{\sum \hat{e}_i^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{\hat{\mathbf{e}}^T \hat{\mathbf{e}}}{\mathbf{y}^T \mathbf{A} \mathbf{y}}$$

où $\mathbf{A} = \mathbf{I} - \frac{\mathbf{u} \mathbf{u}^T}{T}$ est la matrice de transformation des données en déviation de la moyenne. Par exemple, si $R^2 = 87\%$, cela signifie que 87 % de la variation de la variable expliquée est attribuable aux variations des variables explicatives. Il peut être montré que :

$$R^2 = \left(\frac{\text{Cov}(\hat{y}_t, y_t)}{\sqrt{V(\hat{y}_t)} \sqrt{V(y_t)}} \right)^2, \text{ qui est le coefficient de corrélation au}$$

carré entre \hat{y}_t et y_t . Comme \hat{y}_t est une prévision de la valeur de y_t , R^2 est un indicateur du caractère explicatif de l'équation de régression à bien modéliser y_t . La valeur de R^2 est comprise entre 0 et 1. Plus R^2 est rapproché de 1, plus le caractère explicatif du modèle est important. Comme cela a été expliqué au chapitre 2, cette mesure a la déficience d'augmenter au fur et à mesure que l'on ajoute des variables explicatives même si celles-ci ne sont pas significatives. À noter que l'interprétation du R^2 comme coefficient de corrélation de Pearson entre une variable explicative et la variable dépendante n'est plus valable, puisque l'on est évidemment ici en présence de plusieurs variables explicatives. Une mesure alternative qui corrige ce problème associé aux degrés de liberté est le R^2 ajusté de Theil. Elle se définit comme suit :

$$\bar{R}^2 = 1 - \frac{\sum \hat{e}_i^2 / (T - k)}{\sum (y_i - \bar{y})^2 / (T - 1)} = 1 - \frac{\hat{\sigma}_e^2}{\hat{\sigma}_y^2}$$

Cette mesure utilise les estimateurs sans biais de la variance des erreurs résiduelles et de la variance de y . Il faut également souligner que les deux mesures du degré d'ajustement du modèle de régression que nous avons présentées sont centrées, c'est-à-dire qu'elles sont définies en déviation de la moyenne. On pourrait également exprimer ces mesures en données brutes.

3. HYPOTHÈSES SUR LES ERREURS ET CONSÉQUENCES

Supposons le modèle de régression suivant :

$$y_t = \beta_1 + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + e_t, \quad t = 1, \dots, T$$

où $e_t \sim N(0, \sigma^2)$. L'application des moindres carrés ordinaires (MCO) à ce modèle implique que $\hat{\beta} \sim N\left(\beta, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}\right)$. Précisons les implications de ces hypothèses :

- i) Sous l'hypothèse de normalité des erreurs, non seulement l'estimateur des MCO est BLUE par le théorème Gauss-Markov, mais il devient le meilleur estimateur sans biais de β . La variance des MCO atteint la borne Cramer-Rao, borne inférieure pour tous les estimateurs. Cela signifie que sans l'hypothèse de normalité, il peut exister un estimateur non linéaire biaisé de β , mais qui comporte une variance échantillonnale inférieure à celle des MCO.
- ii) L'estimateur des MCO de β se confond avec celui du maximum de vraisemblance. Pour ce qui concerne la variance, l'estimateur de la variance de σ^2 du maximum de vraisemblance est toutefois biaisé, ce qui n'est pas le cas pour les MCO.
- iii) Sous l'hypothèse de normalité, on obtient des tests exacts.

Sachant que $\hat{\beta} \sim N\left(\beta, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}\right)$, cela revient à dire que l'on connaît les distributions exactes des tests. On peut donc construire les tests t , de χ^2 et de Fisher dans les petits échantillons. Advenant le cas où l'on ne connaît pas la distribution des erreurs, on recourt aux distributions asymptotiques de nos estimateurs pour ainsi effectuer les tests LM, LR et de

Wald. Il existe d'autres méthodes de tests pour les petits échantillons, tels les tests non paramétriques.

4. TESTS D'HYPOTHÈSES ET INTERVALLES DE CONFIANCE

Tel qu'on vient de le mentionner, l'hypothèse de la normalité des résidus nous permet d'effectuer des tests d'inférence et de calculer des intervalles de confiance.

Tests t

Considérons le cas où σ^2 est inconnu. Soit $H_0 : \beta_k = 0$ et $H_1 : \beta_k \neq 0$, c'est-à-dire que l'on veut tester si β_k est significativement différent de 0. Pour construire ce test, on procède de la même façon qu'au chapitre 2, c'est-à-dire, sous H_0 :

$$t = \frac{\hat{\beta}_k - 0}{\sqrt{\hat{V}(\hat{\beta}_k)}} = \frac{\hat{\beta}_k - 0}{\sqrt{c_{kk} \times \hat{\sigma}^2}} \sim t_c (T - k)$$

où c_{kk} provient de la diagonale de la matrice $(\mathbf{X}^T \mathbf{X})^{-1}$. La représentation matricielle de la variance de $\hat{\beta}$ est la suivante :

$$\hat{V}(\hat{\beta}) = \hat{\sigma}^2 \begin{pmatrix} T & \sum x_{t2} & \cdot & \cdot & \cdot & \sum x_{tk} \\ \sum x_{t2} & \sum x_{t2}^2 & \cdot & \cdot & \cdot & \cdot \sum x_{t2} x_{tk} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sum x_{tk} & \sum x_{t2} x_{tk} & \cdot & \cdot & \cdot & \sum x_{tk}^2 \end{pmatrix}^{-1}$$

Par exemple, pour $k = 2$, c_{22} est égal à : $c_{22} = \sum x_{t2}^2$.

Intervalles de confiance

Pour σ^2 inconnu, la technique pour construire un intervalle de confiance est exactement la même que celle présentée aux chapitres 1 et 2, c'est-à-dire :

$$P[-t_c < t < t_c] = 1 - \alpha$$

$$\Rightarrow P\left[\hat{\beta}_k - t_c \sqrt{\hat{V}(\hat{\beta}_k)} < \beta_k < \hat{\beta}_k + t_c \sqrt{\hat{V}(\hat{\beta}_k)}\right] = 1 - \alpha$$

L'intervalle de confiance pour $\hat{\beta}_k$ s'écrit : $\hat{\beta}_k \pm se(\hat{\beta}_k) \times t_c$, où se est l'écart-type de $\hat{\beta}_k$. Le test qui vient d'être présenté est bilatéral. Pour ce qui concerne les tests unilatéraux, la technique est identique à celle exposée aux chapitres 1 et 2.

Test F

Dans le chapitre précédent, on a présenté le test F dans le cas univarié comme étant un t^2 . Mais dans sa forme la plus classique, le test F est utilisé pour effectuer un test conjoint sur l'ensemble des paramètres. En effet, pour tester l'hypothèse $H_0 : \beta_2 = \beta_3 = \dots = \beta_k = 0 : \beta_s = 0$ contre $H_1 : \text{au moins l'un des } \beta_k \text{ est différent de } 0 : \beta_s \neq 0$. Le test F correspondant est le suivant. Tout d'abord, nous savons que : $\hat{\beta}_s \sim N(\beta_s, \sigma^2 M)$. Posons :

$$w_1 = (\hat{\beta}_s - \beta_s)^T [V(\hat{\beta}_s)]^{-1} (\hat{\beta}_s - \beta_s)$$

$$= (\hat{\beta}_s - \beta_s)^T \frac{[M]^{-1}}{\sigma^2} (\hat{\beta}_s - \beta_s) \sim \chi_c^2(k-1)$$

où $M = (X_2^T A X_2)^{-1} = (X_*^T X_*)^{-1}$. Posons :

$$w_2 = \frac{\hat{\sigma}^2(T-k)}{\sigma^2} \sim \chi_c^2(T-k)$$

Sachant que le test F est construit à partir du ratio de deux χ^2 divisées par leurs degrés de liberté respectifs, on a alors :

$$\begin{aligned} F &= \frac{\frac{w_1}{(k-1)}}{\frac{w_2}{(T-k)}} = \frac{\frac{(\hat{\beta}_s - \beta_s)^T \mathbf{M}^{-1} (\hat{\beta}_s - \beta_s)}{\sigma^2}}{\frac{\hat{\sigma}^2(T-k)}{\sigma^2}} \\ &= \frac{(\hat{\beta}_s - \beta_s)^T (\mathbf{M}\hat{\sigma}^2)^{-1} (\hat{\beta}_s - \beta_s)}{k-1} \\ &= \frac{(\hat{\beta}_s - \beta_s)^T (\hat{\mathbf{V}}(\hat{\beta}_s))^{-1} (\hat{\beta}_s - \beta_s)}{k-1} \sim F_c(k-1, T-k) \end{aligned}$$

Étant donné que l'on teste $H_0: \beta_s = 0$, on peut écrire :

$$F = \frac{\hat{\beta}_s^T (\hat{\mathbf{V}}(\hat{\beta}_s))^{-1} \hat{\beta}_s}{k-1} \sim F_c(k-1, T-k)$$

Comme $F = \frac{\frac{w_1}{(k-1)}}{\frac{w_2}{(T-K)}}$, que $w_1 = \frac{\hat{\beta}_s^T \mathbf{X}_*^T \mathbf{X}_* \hat{\beta}_s}{\sigma^2} = \frac{\text{SCE}}{\sigma^2}$ et que

$$w_2 = \frac{\hat{\sigma}^2(T-k)}{\sigma^2} = \frac{\text{SCR}}{\sigma^2}, \text{ on a :}$$

$$F = \frac{SCE / k - 1}{SCR / T - k} = \frac{R^2 / k - 1}{(1 - R^2) / T - k}$$

puisque $R^2 \times SCT = SCE$ et $(1 - R^2) \times SCT = SCR$.

Estimation en présence de contraintes linéaires sur les paramètres

Dans la section précédente, nous avons présenté la procédure générale pour tester l'ensemble des paramètres. Dans cette section, nous nous attaquons à la procédure pour tester un sous-ensemble de paramètres faisant partie de cet ensemble. À cet effet, supposons que l'on

veut tester $H_0: \beta_s = \begin{pmatrix} \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ qui est le sous-ensemble de l'ensemble

des paramètres $(\beta_2, \beta_3, \dots, \beta_k)$, contre l'alternative H_1 :

$\beta_s = \begin{pmatrix} \beta_2 \\ \beta_3 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. En utilisant la notation habituelle, on peut écrire les

contraintes linéaires sur les β comme suit :

$$R\beta = r$$

Le test H_0 est ici le suivant si l'on suppose un ensemble de cinq paramètres :

$$H_0: R\beta = r$$

$$H_1: R\beta \neq r$$

où :

$$\mathbf{R}\boldsymbol{\beta} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} = \begin{bmatrix} \beta_2 \\ \beta_3 \end{bmatrix}$$

Pour construire le test F en utilisant cette notation, il faut estimer la covariance de $\hat{\boldsymbol{\beta}}_s$, c'est-à-dire :

$$\hat{V}(\hat{\boldsymbol{\beta}}_s) = \mathbf{R}\hat{V}(\hat{\boldsymbol{\beta}})\mathbf{R}^T = \mathbf{R}\hat{\sigma}^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{R}^T$$

$$= \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \hat{V}(\hat{\beta}_1) & \text{Cov}(\hat{\beta}_1, \hat{\beta}_2) & \cdot & \cdot & \cdot & \text{Cov}(\hat{\beta}_1, \hat{\beta}_5) \\ \text{Cov}(\hat{\beta}_2, \hat{\beta}_1) & \hat{V}(\hat{\beta}_2) & \cdot & \cdot & \cdot & \text{Cov}(\hat{\beta}_2, \hat{\beta}_5) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \text{Cov}(\hat{\beta}_5, \hat{\beta}_1) & \text{Cov}(\hat{\beta}_5, \hat{\beta}_2) & \cdot & \cdot & \cdot & \hat{V}(\hat{\beta}_5) \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$= \begin{bmatrix} \hat{V}(\hat{\beta}_2) & \text{Cov}(\hat{\beta}_2, \hat{\beta}_3) \\ \text{Cov}(\hat{\beta}_2, \hat{\beta}_3) & \hat{V}(\hat{\beta}_3) \end{bmatrix}$$

Après avoir calculé cette variance et sachant que $\mathbf{R}\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_s$ et que $\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r} = (\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s)$, on peut écrire le test F basé sur la notation générale des contraintes linéaires sur les paramètres comme suit :

$$\begin{aligned}
 F &= \frac{(\mathbf{R}\hat{\beta} - \mathbf{r})^T [\mathbf{RC}\hat{\sigma}_v(\hat{\beta})\mathbf{R}^T]^{-1} (\mathbf{R}\hat{\beta} - \mathbf{r})}{q} \\
 &= \frac{(\mathbf{R}\hat{\beta} - \mathbf{r})^T [\mathbf{R}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{R}^T]^{-1} (\mathbf{R}\hat{\beta} - \mathbf{r})}{q\hat{\sigma}^2} \\
 &= \frac{\text{SCR}_R - \text{SCR}_U}{\frac{\text{SCR}_U}{T - k}} \sim F_c(q, T - k)
 \end{aligned}$$

Envisageons maintenant l'estimation en présence de contraintes. Nous traitons toujours le cas du modèle de régression linéaire à cinq paramètres où $H_0 : \beta_2 = \beta_3 = 0$. Si l'on incorpore ces contraintes, dans ce modèle, on a :

$$y_t = \beta_1 + \beta_4 x_{4t} + \beta_5 x_{5t} + e_{Rt}$$

En appliquant les MCO sur cette équation, on obtient les résidus suivants :

$$\hat{\mathbf{e}}_R = \mathbf{y} - \mathbf{X}_R \hat{\beta}_R$$

où

$$\hat{\beta}_R = \begin{bmatrix} \hat{\beta}_{1R} \\ \hat{\beta}_{4R} \\ \hat{\beta}_{5R} \end{bmatrix} \quad \mathbf{X}_R = \begin{bmatrix} 1 & x_{41} & x_{51} \\ 1 & x_{42} & x_{52} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ 1 & x_{4T} & x_{5T} \end{bmatrix}$$

La somme des carrés résiduels contraints est donc égale à :

$$SCR_R = (\mathbf{y} - \mathbf{X}\hat{\beta}_R)^T (\mathbf{y} - \mathbf{X}\hat{\beta}_R) = \hat{\mathbf{e}}_R^T \hat{\mathbf{e}}_R$$

Par ailleurs la somme des carrés résiduels non contraints, soit SCR_U , est égale à : $SCR_U = (\mathbf{y} - \mathbf{X}\hat{\beta}_U)^T (\mathbf{y} - \mathbf{X}\hat{\beta}_U) = \hat{\mathbf{e}}_U^T \hat{\mathbf{e}}_U$. Par consé-

quent, le test d'hypothèse $H_0 : \beta_s = \begin{bmatrix} \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ contre $H_1 : \beta_s \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

s'écrit comme suit : $F = \frac{SCR_R - SCR_U}{T - k} \sim F_c(2, T - k)$. Si H_0 est

vraie, on devrait s'attendre à ce que SCR_R soit très rapproché de SCR_U . En effet, les variables 2 et 3 n'exercent pas alors d'incidence sur la régression. La statistique F est alors rapprochée de 0, très en deçà du seuil critique au-delà duquel H_0 serait rejetée. Par ailleurs, si $F > F_c$, SCR_U est sensiblement plus faible que SCR_R et les contraintes ne sont pas valables. H_0 est alors rejeté².

Supposons maintenant que l'on veut estimer et tester le modèle contraint suivant. Dans le modèle précédent, nous imposons les contraintes suivantes. Sous

$$H_0 : \beta_2 + \beta_3 + \beta_4 + \beta_5 = 0$$

2. La variance de l'estimateur contraint sera toujours inférieure ou égale à celle de l'estimateur non contraint puisque, à l'instar de la statistique bayésienne, on fait appel à de l'information supplémentaire à l'extérieur de l'échantillon.

C'est-à-dire, en notation matricielle :

$$\mathbf{R}\boldsymbol{\beta} = \mathbf{r}, \text{ où}$$

$$\mathbf{R}\boldsymbol{\beta} = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} = 0 + \beta_2 + \beta_3 + \beta_4 + \beta_5$$

et $\mathbf{r} = \mathbf{0}$. Par ailleurs, sous $H_1 : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}$. L'interprétation de H_0 est la suivante. Imaginons que le modèle précédent concerne la demande d'un bien, par exemple la demande de contrats au comptant sur les carcasses de porc, sur lesquelles peuvent être aussi écrits des contrats à terme. Les variables explicatives x_2 , x_3 et x_4 sont les prix relatifs habituels et x_5 est le revenu. Supposons que l'on double le prix des carcasses (x_2), de même que les prix des autres viandes (x_3) et les prix des biens et services restants (x_4). On double également le niveau du revenu (x_5). Dans ce contexte, l'hypothèse H_0 signifie que la quantité demandée de carcasses de porc demeurerait la même en dépit de ces variations de prix et de revenu.

Pour tester H_0 , on estime l'équation restreinte suivante. Remplaçons β_4 par $-\beta_2 - \beta_3 - \beta_5$.³ En substituant cette valeur dans l'équation présentée précédemment, nous obtenons le modèle contraint suivant :

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + (-\beta_2 - \beta_3 - \beta_5) x_{4t} + \beta_5 x_{5t} + e_{Rt}$$

En réarrangeant cette équation, on obtient :

$$y_t = \beta_1 + \beta_2 (x_{2t} - x_{4t}) + \beta_3 (x_{3t} - x_{4t}) + \beta_5 (x_{5t} - x_{4t}) + e_{Rt}$$

3. Ce qui est bien sûr une façon d'exprimer l'hypothèse $H_0 : \beta_2 + \beta_3 + \beta_4 + \beta_5 = 0$.

Dans une première étape, on estime ce modèle par les MCO pour obtenir les résidus contraints $\hat{\mathbf{e}}_R$ et l'on peut ainsi calculer la somme des carrés résiduels contraints : $SCR_R = \hat{\mathbf{e}}_R^T \hat{\mathbf{e}}_R$. Dans une deuxième étape, on estime le modèle non contraint d'où l'on déduit la somme des carrés résiduels non contraints : $SCR_U = \hat{\mathbf{e}}_U^T \hat{\mathbf{e}}_U$. À l'aide de ces informations, on formule le test F comme suit :

$$F = \frac{(SCR_R - SCR_U) / 1}{SCR_U / (T - 5)} \sim F_c(1, (T - 5))$$

où $q = 1$ et $k = 5$. La règle est de rejeter H_0 si $F > F_c(1, (T - 5))$ ou si la $p\text{-value}(F) < 0,05$. Notons que pour $J = 1$, un test de Student aurait pu convenir. En effet, l'hypothèse H_0 se formule alors comme suit :

$$H_0 : \beta_2 + \beta_3 + \beta_4 + \beta_5 = 0$$

contre l'alternative $H_1 : \beta_2 + \beta_3 + \beta_4 + \beta_5 \neq 0$. La procédure du test s'effectue en deux temps.

- 1) On estime la régression : $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \beta_5 x_{5t}$ par les MCO, d'où on obtient $\hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$ et $\hat{\beta}_5$.
- 2) On fait la somme de ces coefficients estimés et on construit le test t de Student de la façon habituelle.

$$t = \frac{(\hat{\beta}_2 + \hat{\beta}_3 + \hat{\beta}_4 + \hat{\beta}_5) - 0}{\sqrt{\hat{V}(\hat{\beta}_2 + \hat{\beta}_3 + \hat{\beta}_4 + \hat{\beta}_5)}} \sim t_c(T - 5)$$

où \hat{V} se calcule comme suit :

$$\hat{V}(\hat{\beta}_2 + \hat{\beta}_3 + \hat{\beta}_4 + \hat{\beta}_5) = \sum_{i=2}^5 \hat{V}(\hat{\beta}_i) + \sum_{i=2}^4 \sum_{j=3}^5 \text{C}\hat{\text{ov}}(\hat{\beta}_i, \hat{\beta}_j)$$

et où $\text{C}\hat{\text{ov}}(\hat{\beta}_i, \hat{\beta}_j)$ provient de la matrice variance-covariance de $\hat{\beta}$:

$\hat{\beta} : \hat{V}(\hat{\beta}) = \hat{\sigma}^2 (\mathbf{X}^T \mathbf{X})^{-1}$. La règle est de rejeter H_0 au seuil $\alpha = 5 \%$ si

$|t| > t_c(T - k)$, ou lorsque la $p\text{-value}(t) < 5 \%$.

Il est à signaler que le test F aurait pu être formulé en termes de R^2 . Le test F pour tester un sous-ensemble de paramètres s'exprime comme suit :

$$F = \frac{(\text{SCR}_R - \text{SCR}_U) / q}{\text{SCR}_U / (T - k)} = \frac{(R_U^2 - R_R^2) / q}{(1 - R_U^2) / (T - k)}$$

Cette dernière égalité repose sur les relations suivantes :

$$\text{SCR}_R = (1 - R_R^2) \text{SCT}_R; \quad \text{SCR}_U = (1 - R_U^2) \text{SCT}_U.$$

5. PRÉVISION DANS LE MODÈLE LINÉAIRE GÉNÉRAL

Nous transposons ici les techniques de prévision introduites au chapitre 2 au modèle linéaire général. Soit le modèle suivant à quatre variables explicatives :

$$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \beta_5 x_{t5} + e_t$$

où $e_t \sim N(0, \sigma^2)$. Supposons que l'on veuille prédire y à partir des valeurs suivantes des variables explicatives : $\mathbf{x}_0 = [1 \ x_{02} \ x_{03} \ x_{04} \ x_{05}]^T$. On veut donc prédire $E(y_0)$. En substituant le vecteur \mathbf{x}_0 dans l'équation de y_t , on obtient :

$$y_0 = \mathbf{x}_0^T \boldsymbol{\beta} + e_0$$

où $\boldsymbol{\beta} = [\beta_2 \ \beta_3 \ \beta_4 \ \beta_5]^T$. Le meilleur estimateur linéaire sans biais de $E(y_0)$ est :

$$\hat{y}_0 = \mathbf{x}_0^T \hat{\boldsymbol{\beta}}$$

où $\hat{\boldsymbol{\beta}}$ est l'estimateur de la régression de y_t sur le terme constant et ses quatre variables explicatives. L'intervalle de confiance de cet estimateur est :

$$y_0 \pm t_{\alpha/2} \hat{\sigma} \sqrt{\mathbf{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_0}$$

6. APPLICATIONS

Pour illustrer ce chapitre, nous renouons avec l'exemple des dépenses sur les cartes de crédit que nous avons traité en partie dans les chapitres précédents. Nous voulons effectuer la régression suivante, qui renferme plusieurs variables explicatives :

$$\mathbf{y} = \mathbf{x}_1\beta_1 + \mathbf{x}_2\beta_2 + \mathbf{x}_3\beta_3 + \mathbf{x}_4\beta_4 + \mathbf{x}_5\beta_5 + \mathbf{e}$$

où \mathbf{x}_1 est un vecteur unitaire ; \mathbf{x}_2 est le vecteur des données sur les revenus ; \mathbf{x}_3 , le vecteur des revenus au carré qui sert à capter la non-linéarité de la fonction des dépenses ; \mathbf{x}_4 , un vecteur des âges en années ; \mathbf{x}_5 , un vecteur de variables dichotomiques qui prend la valeur 1 si l'individu possède une maison et 0 s'il est à loyer. En appliquant les MCO sur ces données, on obtient les résultats qui apparaissent au tableau 3.1.

TABLEAU 3.1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-237.1465	199.3517	-1.189589	0.2384
REVENU	234.3470	80.36595	2.915999	0.0048
REVENU ^2	-14.99684	7.469337	-2.007788	0.0487
AGE	-3.081814	5.514717	-0.558835	0.5781
LOYERMAISON	27.94091	82.92232	0.336953	0.7372
R-squared	0.243578	Mean dependent var		262.5321
Ajusted R-squared	0.198418	S.D. dependent var		318.0468
S.E. of regression	284.7508	Akaike info criterion		14.20802
Sum squared resid	5432562.	Schwarz criterion		14.36612
Log likelihood	-506.4888	F-statistic		5.393722
Durbin-Watson stat	1.682310	Prob(F-statistic)		0.000795
<i>smp1 if x3 = 0</i>		<i>smp1 if x3 > 0</i>		
<i>series carte = na</i>		<i>series carte = x3</i>		
<i>series loyermaison = na</i>		<i>series loyermaison = x4</i>		
<i>series age = na</i>		<i>series age = x1</i>		
<i>series revenu = na</i>		<i>series revenu = x2</i>		

On observe entre autres que seules les variables revenus et revenus au carré sont significatives au seuil de 5 % puisque leur p-value est respectivement de 0,005 et de 0,049, soit inférieures au seuil critique de 0,05. Les R^2 et R^2 ajusté sont respectivement de 24,3 % et de 19,8 %, soit du même ordre que dans l'exemple du chapitre 2. En regard de cet exemple, l'augmentation du R^2 ajusté est faible puisque dans cet exemple, celui-ci se situait à 18,5 %. Par ailleurs, le R^2 ordinaire a beaucoup plus augmenté puisque dans l'exemple du chapitre 2 il se situait à 19,6 %. Cette situation était anticipée puisque le R^2 ordinaire n'est pas corrigé pour les degrés de liberté.

Finalement, le test F sur l'ensemble des variables explicatives dégage une statistique F de 5,39 et sa p-value est de 0,000. Les variables explicatives sont donc significatives et tel que mentionné plus haut, les tests t nous permettent de discriminer entre les variables explicatives.

RAPPELS DE CALCUL MATRICIEL

1. OPÉRATIONS MATRICIELLES

1.1. Addition et soustraction

Soit deux matrices carrées **A** et **B** d'ordre 2×2 . La somme de ces deux matrices s'effectue comme suit :

$$\begin{aligned} \mathbf{A} + \mathbf{B} &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \\ &= \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{bmatrix} \end{aligned}$$

Pour la soustraction des deux matrices **A** et **B**, on remplace les + par des – au niveau des éléments de la matrice **B**.

Soit trois matrices **A**, **B** et **C** conformables. La loi de l'associativité vaut également pour la somme des matrices, c'est-à-dire :

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$$

1.2. Multiplication et produit Kronecker

Nous voulons maintenant multiplier les matrices \mathbf{A} et \mathbf{B} . Ces deux matrices sont conformables pour la multiplication puisque le nombre de lignes de \mathbf{A} est égal au nombre de colonnes de \mathbf{B} . Nous procédons comme suit pour multiplier \mathbf{A} et \mathbf{B} :

$$\mathbf{AB} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}_{2 \times 2}$$

Par ailleurs, $\mathbf{A} \otimes \mathbf{B}$, ou le produit Kronecker de \mathbf{A} et \mathbf{B} , désigne le produit de chaque élément de \mathbf{A} par \mathbf{B} , c'est-à-dire :

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} \mathbf{a}_{11}\mathbf{B} & \mathbf{a}_{12}\mathbf{B} \\ \mathbf{a}_{21}\mathbf{B} & \mathbf{a}_{22}\mathbf{B} \end{bmatrix}_{4 \times 4}$$

Soit trois matrices conformables \mathbf{A} , \mathbf{B} et \mathbf{C} . La loi de l'associativité vaut pour le produit de ces trois matrices :

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$$

1.3. Transposée d'une matrice

Soit la matrice précédente \mathbf{A} . Sa transposée est l'inversion de ses lignes et colonnes, c'est-à-dire :

$$\mathbf{A}^T = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}$$

Soit trois matrices conformables \mathbf{A} , \mathbf{B} et \mathbf{C} . On a :

$$(\mathbf{A} + \mathbf{B} + \mathbf{C})^T = \mathbf{A}^T + \mathbf{B}^T + \mathbf{C}^T$$

Par ailleurs, la règle concernant la transposée d'une multiplication est la suivante :

$$(\mathbf{ABC})^T = \mathbf{C}^T \mathbf{B}^T \mathbf{A}^T$$

2. MATRICES CARRÉES IMPORTANTES

2.1. La matrice I

La matrice identité de dimension $(n \times n)$ s'écrit comme suit :

$$\mathbf{I}_n = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & 1 \end{bmatrix}$$

Cette matrice est diagonale car les triangles supérieurs et inférieurs contiennent des éléments qui sont nuls. Seuls les éléments de la diagonale principale ne sont pas nuls. Supposons maintenant un vecteur \mathbf{y} de dimension $(n \times 1)$. On a : $\mathbf{I}_n \mathbf{y} = \mathbf{y}$. La transposée de cette opération est la suivante : $(\mathbf{I}_n \mathbf{y})^T = \mathbf{y}^T \mathbf{I}_n^T = \mathbf{y}^T$.

La matrice \mathbf{I}_n est un cas de matrice symétrique. En effet, une matrice symétrique est une matrice carrée, c'est-à-dire de dimension $(n \times n)$, qui est égale à sa transposée. Soit une matrice carrée \mathbf{A} . \mathbf{A} est symétrique si : $\mathbf{A}^T = \mathbf{A}$. Pour fixer les idées, dans le cas de la méthode

des MCO, la matrice $(\mathbf{X}^T\mathbf{X})$ est une matrice symétrique puisque la transposée de cette matrice est égale au résultat initial.

2.2. Matrice idempotente

La matrice \mathbf{A} est idempotente si $\mathbf{A} = \mathbf{A}^2 = \dots = \mathbf{A}^n$. C'est-à-dire que lorsque l'on multiplie la matrice \mathbf{A} par elle-même autant de fois que l'on veut, on revient à la matrice originale. Une matrice idempotente très connue est la matrice dite *residuals maker matrix*⁴ que l'on peut traduire en français par *matrice génératrice des résidus* :

$$\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$$

Cette propriété signifie, par exemple, que $\mathbf{M}^2 = \mathbf{M}$. Cette matrice a également la propriété d'être symétrique, c'est-à-dire que : $\mathbf{M}^T = \mathbf{M}$.

À ce stade-ci, il convient d'introduire la trace d'une matrice carrée. Elle se définit comme la somme des éléments de sa diagonale principale, c'est-à-dire : $\text{tr}(\mathbf{A}) = \sum_i a_{ii}$, où \mathbf{A} est une matrice carrée.

Soit maintenant deux matrices \mathbf{A} et \mathbf{B} , dont les dimensions respectives sont de $(m \times n)$ et de $(n \times m)$. Par conséquent, \mathbf{AB} et \mathbf{BA} sont deux matrices carrées et :

$$\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$$

Pour trois matrices \mathbf{A} , \mathbf{B} et \mathbf{C} , si le produit donne des matrices carrées, on a :

$$\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{CAB}) = \text{tr}(\mathbf{BCA})$$

Cette propriété nous a servi à démontrer que $\hat{\sigma}^2 = \frac{\hat{\mathbf{e}}^T\hat{\mathbf{e}}}{T-k}$ est un estimateur non biaisé de σ^2 . En effet, pour démontrer cette propriété, il faut calculer l'espérance de $\hat{\sigma}^2$, et pour effectuer cette opération, on doit calculer $E(\hat{\mathbf{e}}^T\hat{\mathbf{e}})$, c'est-à-dire :

4. Cette matrice est ainsi appelée car lorsqu'elle prémultiplie \mathbf{y} , on obtient les résidus estimés. En effet, $\mathbf{My} = [\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T] \mathbf{y} = \mathbf{y} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \hat{\mathbf{e}}$.

$$\begin{aligned}
 E(\hat{\mathbf{e}}^T \hat{\mathbf{e}}) &= E(\mathbf{e}^T \mathbf{M}^T \mathbf{M} \mathbf{e}) \\
 &= E(\mathbf{e}^T \mathbf{M} \mathbf{e}) \\
 &= E(\text{tr}(\mathbf{e}^T \mathbf{M} \mathbf{e}))
 \end{aligned}$$

car la trace d'un scalaire est égale à ce scalaire⁵. Il s'ensuit que :

$$\begin{aligned}
 &= E(\text{tr}(\mathbf{M} \mathbf{e} \mathbf{e}^T)) \\
 &= \text{tr}(\mathbf{M} E(\mathbf{e} \mathbf{e}^T)) \\
 &= \text{tr}(\mathbf{M} \sigma^2 \mathbf{I}) \\
 &= \sigma^2 \text{tr}(\mathbf{M} \mathbf{I}) \\
 &= \sigma^2 \text{tr}(\mathbf{M})
 \end{aligned}$$

Pour terminer ce calcul, il nous suffit de calculer la trace de la matrice \mathbf{M} . Ce calcul s'effectue comme suit :

$$\begin{aligned}
 \text{tr}(\mathbf{M}) &= \text{tr}(\mathbf{I}) - \text{tr}\left(\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T\right) \\
 &= n - \text{tr}\left(\left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{X}\right) = n - \text{tr}(\mathbf{I}_k) = n - k
 \end{aligned}$$

2.3. Le déterminant d'une matrice

Soit une matrice \mathbf{B} de dimension 3×3 :

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

5. En effet, $\mathbf{e}^T \mathbf{M} \mathbf{e}$ est un scalaire.

Le déterminant de \mathbf{B} se calcule comme suit :

$$\begin{aligned} |\mathbf{B}| &= \sum_{i,j,k} \pm b_{1i} b_{2j} b_{3k} \\ &= b_{11} b_{22} b_{33} + b_{12} b_{23} b_{31} + b_{13} b_{21} b_{32} \\ &\quad - b_{11} b_{23} b_{32} - b_{12} b_{21} b_{33} - b_{13} b_{22} b_{31} \end{aligned}$$

L'opérateur $\sum \pm$ commande toutes les permutations des indices 1, 2 et 3, le signe moins s'imposant lorsque l'ordre naturel des indices est inversé. De façon plus générale, le déterminant d'une matrice \mathbf{A} est :

$$|\mathbf{A}| = \sum_{i,j,\dots,s} \pm a_{1i} a_{2j} \dots a_{ns}$$

Voici quelques propriétés des déterminants :

- i) Le déterminant d'une matrice triangulaire \mathbf{B}

$$\mathbf{B} = \begin{bmatrix} b_{11} & 0 & 0 \\ b_{21} & b_{22} & 0 \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

est égal au produit des éléments de sa diagonale principale, soit sa trace :

$$|\mathbf{B}| = b_{11} b_{22} b_{33}$$

- ii) Le déterminant du produit de deux matrices carrées \mathbf{A} et \mathbf{B} de même dimension est égal à :

$$|\mathbf{AB}| = |\mathbf{A}| |\mathbf{B}|$$

Un corollaire à cette règle est que :

$$|\mathbf{A}^{-1}| = \frac{1}{|\mathbf{A}|}$$

- iii) La multiplication de toute ligne ou colonne d'une matrice par une constante a le même effet sur le déterminant de cette matrice. Pour illustrer, multiplions par c la première ligne de la matrice \mathbf{B} précédente :

$$\begin{bmatrix} cb_{11} & cb_{12} & cb_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \Rightarrow c|\mathbf{B}|$$

Par ailleurs, multiplier tous les éléments d'une matrice ($n \times n$) par la constante c revient à multiplier son déterminant par c^n .

- iv) Si une ou plusieurs lignes ou colonnes d'une matrice sont linéairement dépendantes, alors le déterminant de cette matrice est nul. On dit qu'une telle matrice est singulière par rapport à une matrice régulière dont le déterminant est différent de 0.

2.4. L'inverse d'une matrice

L'inverse d'une matrice se calcule comme suit en utilisant la méthode de l'adjointe (adj). Soit une matrice \mathbf{B} de dimension (3×3). Son inverse désigné par \mathbf{B}^{-1} se calcule comme suit :

$$\mathbf{B}^{-1} = \frac{1}{|\mathbf{B}|} \times (\text{adj}(\mathbf{B})) = \frac{1}{|\mathbf{B}|} \times \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix}^T$$

où

$$C_{ij} = (-1)^{i+j} M_{ij}$$

C_{ij} est appelé le cofacteur et M_{ij} est le mineur, obtenu en calculant les déterminants des sous-matrices obtenues en supprimant la i^{e} ligne et la j^{e} colonne. À titre d'exemple, pour la matrice \mathbf{B} (3×3) précédente, C_{11} est égal à :

$$C_{11} = (-1)^{1+1} \begin{vmatrix} b_{22} & b_{23} \\ b_{32} & b_{33} \end{vmatrix}$$

Voici les propriétés de l'inverse.

- i) L'inverse de l'inverse d'une matrice carrée \mathbf{B} donne la matrice \mathbf{B} , c'est-à-dire : $(\mathbf{B}^{-1})^{-1} = \mathbf{B}$.
- ii) Le produit d'une matrice par son inverse donne la matrice unitaire : $\mathbf{B}\mathbf{B}^{-1} = \mathbf{I}$.
- iii) L'inverse de la transposée d'une matrice est égale à la transposée de l'inverse : $(\mathbf{B}^T)^{-1} = (\mathbf{B}^{-1})^T$. On a donc : $(\mathbf{B}^{-1})^T \mathbf{B}^T = \mathbf{I}$.
- iv) L'inverse d'une matrice triangulaire supérieure ou inférieure est également une matrice triangulaire supérieure ou inférieure.
- v) L'inverse d'une matrice partitionnée est également une matrice partitionnée. Soit la partition suivante de \mathbf{A} :

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \text{ où } \mathbf{A}_{11} \text{ et } \mathbf{A}_{22} \text{ sont des matrices carrées régulières. L'inverse de cette matrice est alors :}$$

$$\begin{aligned} \mathbf{A}^{-1} &= \begin{bmatrix} \mathbf{B}_{11} & -\mathbf{B}_{11}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{B}_{11} & \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{B}_{11}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22} \\ -\mathbf{B}_{22}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{B}_{22} \end{bmatrix} \end{aligned}$$

où : $\mathbf{B}_{11} = (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}$ et $\mathbf{B}_{22} = (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}$.

Pour illustrer, prenons l'exemple suivant. Dans le cas du maximum de vraisemblance d'une régression linéaire à plusieurs variables explicatives, la matrice d'information⁶, qui sert à calculer la matrice variance-covariance des estimateurs

de β et de σ^2 , est la suivante : $\mathbf{I} \begin{pmatrix} \beta \\ \sigma^2 \end{pmatrix} = \begin{bmatrix} \frac{1}{\sigma^2}(\mathbf{X}^T \mathbf{X}) & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix}$.

Cette matrice partitionnée est de la forme : $\begin{bmatrix} \mathbf{A}_{11} & 0 \\ 0 & \mathbf{A}_{22} \end{bmatrix}$.

En vertu de la formule de l'inverse de la matrice partitionnée qui vient d'être exposée, l'inverse d'une telle matrice est la

suivante : $\begin{bmatrix} \mathbf{A}_{11}^{-1} & 0 \\ 0 & \mathbf{A}_{22}^{-1} \end{bmatrix}$. $\mathbf{I}^{-1}(\cdot)$ est donc égale à :

$\mathbf{I}^{-1} \begin{pmatrix} \beta \\ \sigma^2 \end{pmatrix} = \begin{pmatrix} \sigma^2(\mathbf{X}^T \mathbf{X})^{-1} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}$. Cette matrice est parti-

culièrement utile pour le calcul des tests asymptotiques de Wald et LM où elle apparaît explicitement.

6. Cette matrice est ainsi appelée car elle informe sur la courbure de la fonction de vraisemblance.

3. DES MATRICES IMPORTANTES : LA MATRICE VARIANCE-COVARIANCE D'UN PORTEFEUILLE DE TITRES ET LA COVARIANCE ENTRE DEUX PORTEFEUILLES

Imaginons qu'un portefeuille soit composé de deux titres : le titre 1 et le titre 2. La pondération du titre 1 dans ce portefeuille est w_1 et celle du titre 2, w_2 . De plus : $w_1 + w_2 = 1$. L'espérance du rendement de ce portefeuille est de :

$$E(R_p) = w_1 E(R_1) + w_2 E(R_2)$$

Par ailleurs, la variance du rendement du portefeuille est pour sa part :

$$\text{Var}(R_p) = E\left[R_p - E(R_p)\right]^2$$

En remplaçant R_p et $E(R_p)$ par leur valeur respective, on a :

$$\text{Var}(R_p) = E\left[w_1 R_1 + w_2 R_2 - w_1 E(R_1) - w_2 E(R_2)\right]^2$$

En regroupant les termes, on obtient :

$$\begin{aligned} \text{Var}(R_p) &= E\left[w_1 (R_1 - E(R_1)) + w_2 (R_2 - E(R_2))\right]^2 \\ &= w_1^2 \left[R_1 - E(R_1)\right]^2 + 2w_1 w_2 E\left[(R_1 - E(R_1))(R_2 - E(R_2))\right] \\ &\quad + w_2^2 \left[R_2 - E(R_2)\right]^2 \\ &= w_1^2 \sigma_1^2 + 2w_1 w_2 \sigma_{12} + w_2^2 \sigma_2^2 \end{aligned}$$

où σ_1^2 désigne la variance du rendement du titre 1 ; σ_2^2 , la variance du rendement du titre 2 et σ_{12} , la covariance entre les rendements des titres 1 et 2. Plus généralement, dans le cas d'un portefeuille de N titres, la variance du rendement de ce portefeuille s'écrit :

$$\text{Var}(R_p) = \sum_{i=1}^N \sum_{j=1}^N w_i w_j \sigma_{ij}$$

L'équation de la variance du rendement d'un portefeuille est une forme quadratique. On peut recourir à la forme matricielle quadratique pour l'exprimer de façon plus compacte et combien plus facile à manipuler.

Dans le cas de deux titres, l'espérance du rendement du portefeuille antérieur s'écrit :

$$E(R_p) = [w_1 \quad w_2] \begin{bmatrix} E(R_1) \\ E(R_2) \end{bmatrix} = \mathbf{w}^T \mathbf{E}$$

où \mathbf{E} est le vecteur des espérances de rendement. La variance de R_p est de :

$$\text{Var}(R_p) = [w_1 \quad w_2] \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \mathbf{w}^T \mathbf{\Omega} \mathbf{w}$$

où $\mathbf{\Omega}$ désigne la matrice variance-covariance des rendements des deux titres. La variance du rendement d'un portefeuille est donc une forme quadratique matricielle. Certes, cette matrice est symétrique puisque $\sigma_{12} = \sigma_{21}$. Le lecteur généralisera très facilement les résultats antérieurs au cas de N titres.

Nous envisageons maintenant deux portefeuilles qui renferment les titres 1 et 2, mais avec des pondérations différentes. Dans le premier, que nous désignons par p , les pondérations des deux titres sont de w_1 et w_2 . Dans le second, que nous désignons par s , les pondérations sont de z_1 et z_2 . Certes, pour chacun de ces deux portefeuilles, la somme des pondérations est égale à 1.

Nous voulons calculer la covariance entre les rendements des portefeuilles p et s . Cette covariance en vertu même de la définition de la covariance, est égale à :

$$\text{Cov}(R_p, R_s) = E\left[\left(R_p - E(R_p)\right)\left(R_s - E(R_s)\right)\right]$$

En remplaçant R_p et R_s par leur équivalent en termes des rendements des titres 1 et 2, on obtient :

$$\text{Cov}(R_p, R_s) = E \left[\begin{array}{l} (w_1 R_1 + w_2 R_2 - w_1 E(R_1) - w_2 E(R_2)) \\ (z_1 R_1 + z_2 R_2 - z_1 E(R_1) - z_2 E(R_2)) \end{array} \right]$$

En regroupant les termes, on a :

$$\begin{aligned} \text{Cov}(R_p, R_s) &= E \left\{ \begin{array}{l} [w_1 (R_1 - E(R_1)) + w_2 (R_2 - E(R_2))] \\ [z_1 (R_1 - E(R_1)) + z_2 (R_2 - E(R_2))] \end{array} \right\} \\ &= w_1 z_1 E [R_1 - E(R_1)]^2 \\ &\quad + w_2 z_1 [(R_2 - E(R_2))(R_1 - E(R_1))] \\ &\quad + w_1 z_2 [(R_1 - E(R_1))(R_2 - E(R_2))] \\ &\quad + w_2 z_2 [R_2 - E(R_2)]^2 \end{aligned}$$

On obtient finalement :

$$\text{Cov}(R_p, R_s) = w_1 z_1 \sigma_1^2 + w_2 z_1 \sigma_{21} + w_1 z_2 \sigma_{12} + w_2 z_2 \sigma_2^2$$

Cette expression finale de la covariance entre les rendements de deux portefeuilles est déjà lourde et pourtant les portefeuilles ne comprennent que deux titres. Le recours au calcul matriciel simplifie de beaucoup le calcul de cette covariance. Pour le cas précédent, on a :

$$\text{Cov}(R_p, R_s) = \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \mathbf{w}^T \mathbf{\Omega} \mathbf{z}$$

où w désigne le vecteur de pondérations du portefeuille p et z , celui du portefeuille s . Cette expression se transpose immédiatement au cas de N titres.

4. QUELQUES APPLICATIONS DU CALCUL MATRICIEL EN FINANCE

Dans le but d'illustrer l'utilisation du calcul matriciel en finance, nous présentons dans cette section trois applications : i) la couverture optimale d'un bilan bancaire par des contrats à terme ; 2) la construction de la frontière efficiente ; 3) le modèle de l'erreur de suivi (*tracking error*) de Roll.

4.1. Couverture optimale d'un bilan bancaire⁴

Le modèle que nous présentons dans cette section est une simple transposition de la théorie du portefeuille de Markowitz au bilan bancaire. En effet, un bilan peut être considéré comme un portefeuille de titres. Les actifs sont assimilables à des titres que détient un investisseur, ici une banque. Pour leur part, les passifs peuvent être considérés comme des actifs négatifs, soit des titres vendus à découvert dans le cadre de notre analogie. L'objectif de la banque est de maximiser le rendement de l'avoir des actionnaires. Le bilan de la banque se présente comme suit :

Bilan de la banque XYZ

Actifs	Passifs
V_1	V_3
V_2	V_4
	S

où V_1 désigne les actifs à court terme ; V_2 , les actifs à long terme ; V_3 , les passifs à court terme ; V_4 , les passifs à long terme et S, l'équité. L'avoir des actionnaires peut pour sa part, être considéré comme un investissement dans un portefeuille qui est *en compte* (*long*) dans les actifs et à *découvert* (*short*) dans les passifs.

4. Pour rédiger cette sous-section, nous nous inspirons de : Copeland, T.E. et J.F. Weston (1988), *Financial Theory and Corporate Policy*, Addison Wesley, New York, chap. 6.

- Chacune des rubriques du bilan, actif et passif, comporte un taux de rendement R_i et un coefficient de pondération w_i . Ces coefficients sont définis par rapport à l'avoir des actionnaires :

$$w_i = \frac{V_i}{S}$$

où V_i désigne la valeur de l'actif (ou du passif) et S , l'équité. La somme de ces coefficients de pondération est bien sûr égale à l'unité en vertu de l'identité comptable du bilan.

Le rendement de l'avoir des actionnaires s'écrit comme suit :

$$R_s = w_1 R_1 + w_2 R_2 + w_3 R_3 + w_4 R_4$$

où w_3 et w_4 , associés à des passifs, sont négatifs.

La banque veut minimiser la variance de l'avoir de ses actionnaires. Pour ce faire, elle recourt à des contrats à terme. Le prix d'un contrat à terme est de P_5 et son rendement, de R_5 . On veut déterminer le nombre de contrats à terme, désigné par N , qui minimise la variance du rendement des actionnaires.

Les contrats à terme, selon qu'ils soient achetés ou vendus, deviennent un nouvel actif ou passif dans le portefeuille que constitue le bilan bancaire. À la suite de l'introduction des contrats à terme, le rendement de l'avoir des actionnaires s'écrit :

$$R_s = w_1 R_1 + w_2 R_2 + w_3 R_3 + w_4 R_4 + \frac{NP_5}{S} R_5$$

Pour trouver le nombre de contrats qui minimise la variance de R_s , on égale la dérivée de la variance de R_s par rapport à N à 0. La variance de R_s est égale à :

$$\text{Var}(R_s) = \mathbf{w}^T \mathbf{\Omega} \mathbf{w}$$

où :

$$\mathbf{w}^T = \left[w_1 \quad w_2 \quad w_3 \quad w_4 \quad \frac{NP_5}{S} \right]$$

soit la transposée des coefficients de pondérations et $\mathbf{\Omega}$ est la matrice variance-covariance des rendements de tous les actifs et passifs présents dans le bilan de la banque, y compris les contrats à terme, c'est-à-dire :

$$\mathbf{\Omega} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdot & \cdot & \sigma_{15} \\ \sigma_{21} & \sigma_{22} & \cdot & \cdot & \sigma_{25} \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ \sigma_{51} & \sigma_{52} & \cdot & \cdot & \sigma_{55} \end{bmatrix}$$

Nous tirons parti des règles de la dérivation matricielle pour calculer \mathbf{N} . Rappelons rapidement l'une d'elles que nous utilisons à l'intérieur de cet exercice. La variance de R_s est une forme matricielle quadratique du type :

$$y = \mathbf{x}^T \mathbf{A} \mathbf{x}$$

où \mathbf{A} est une matrice ($N \times N$) et \mathbf{x} , un vecteur ($N \times 1$). La dérivée de y par rapport à \mathbf{x} est égale à :

$$\frac{dy}{d\mathbf{x}} = 2\mathbf{A}\mathbf{x}$$

– Dans le cas où \mathbf{A} est une matrice (2×2), cette dérivée est égale à :

$$\frac{dy}{d\mathbf{x}} = \begin{bmatrix} \frac{dy}{dx_1} \\ \frac{dy}{dx_2} \end{bmatrix} = \begin{bmatrix} 2a_{11}x_1 + 2a_{12}x_2 \\ 2a_{21}x_1 + 2a_{22}x_2 \end{bmatrix}$$

Dans le cas qui nous intéresse, si l'on dérive la variance de R_s par rapport à \mathbf{w} , on obtient :

$$\frac{d\text{Var}(R_s)}{d\mathbf{w}} = 2\mathbf{\Omega}\mathbf{w}$$

Pour le cinquième actif, soit le contrat à terme, cette dérivée est égale à :

$$\frac{d\text{Var}(R_s)}{dw_5} = 2\Omega_5 \mathbf{w} = 2\sigma_{51}w_1 + 2\sigma_{52}w_2 + \dots + 2\sigma_{55}w_5$$

où $w_5 = \frac{NP_5}{S}$ et Ω_5 désigne la 5^e ligne de la matrice V .

- Revenons au problème qui nous intéresse. On veut déterminer le N qui minimise la variance de R_s . Pour ce faire, on recourt à la règle de la chaîne :

$$\frac{d\text{Var}(R_s)}{dN} = \frac{d\text{Var}(R_s)}{dw_5} \frac{dw_5}{dN} = 0$$

soit,

$$2\Omega_5 \mathbf{w} \times \frac{P_5}{S} = 0$$

- Puisque (P_5/S) ne peut être nul, cette égalité implique que :

$$2\Omega_5 \mathbf{w} = 0$$

Et comme :

$$w_i = \frac{V_i}{S}$$

cette condition d'optimisation s'écrit, en termes de V :

$$2\Omega_5 V = 0$$

En isolant la valeur du cinquième actif, soit celle du contrat à terme qui est égale à NP_5 , on obtient :

$$\sigma_5^2 NP_5 = - \sum_i^4 V_i \sigma_{i5}$$

La valeur de N qui minimise la variance de l'avoir des actionnaires est donc égale à :

$$N = - \sum \frac{V_i}{P_5} \frac{\sigma_{i5}}{\sigma_5^2}$$

On est donc à même de constater que ce sont les covariances entre le rendement du contrat à terme et celles des autres actifs et passifs du

bilan qui importent du point de vue de la couverture. À l'instar de la théorie de Markowitz, les covariances des rendements tiennent le haut du pavé comme mesure du risque dans ce modèle de couverture optimale d'un bilan.

Il est plus simple de manipuler cette dernière formule lorsqu'elle est exprimée en termes de corrélation des rendements. La relation entre la covariance σ_{i5} et la corrélation r_{i5} s'écrit bien sûr comme suit :

$$\sigma_{i5} = r_{i5} \sigma_i \sigma_5$$

En termes de corrélation, le nombre optimal de contrats à terme requis pour la couverture du bilan est égal à :

$$N = - \sum \frac{V_i}{P_5} \times \frac{\sigma_i}{\sigma_5} \times r_{i5}$$

Cette formule nous indique que la couverture optimale d'un bilan par des contrats à terme dépend de trois facteurs :

- i) le ratio de la valeur de la rubrique du bilan à couvrir au prix d'un contrat à terme : $\frac{V_i}{P_5}$. En fait ce facteur est associé à une couverture naïve. Celui qui est étranger à la finance divise la valeur à couvrir par le prix d'un contrat pour déterminer le nombre de contrats optimal. Mais la formule signale que ce dilettante oublie deux éléments dont tout bon spécialiste de la finance doit tenir compte :
- ii) le rapport entre la volatilité de l'actif à couvrir et celle du contrat à terme : $\frac{\sigma_i}{\sigma_5}$. En effet, plus la volatilité de l'actif (ou du passif) à couvrir est élevée par rapport à celle du contrat à terme, plus il faut acheter de contrats. On exerce alors un effet de levier sur la volatilité du contrat à terme, ce qui la rapproche de celle de l'actif (ou du passif) à couvrir ;
- iii) la corrélation entre le contrat et l'actif à couvrir : r_{i5} . S'il n'existe aucune corrélation entre le rendement du contrat à terme et celui d'un actif (ou d'un passif), inutile de couvrir cet actif (ou ce passif). Par ailleurs, si la corrélation est de -1 ,

cela correspond à la situation idéale au plan de la couverture. S'il appert que cette corrélation est de -1 pour chaque rubrique du bilan, la variance de R_s s'avérera nulle après couverture, c'est-à-dire que la variance minimale de l'avoir des actionnaires est alors nulle.

La formule de la couverture optimale nous révèle d'autres informations. Considérons d'abord le cas d'un actif. Le V correspondant est alors positif dans la formule. Si la corrélation entre le rendement du contrat à terme et celui de cet actif s'avère positive, la formule indique qu'il faut alors vendre des contrats à terme pour les fins de la couverture (N négatif) puisque la formule est précédée d'un signe négatif. Dans pareil cas, les prix de l'actif et du contrat à terme ont tendance à évoluer dans la même direction. De façon à couvrir l'actif par le contrat à terme de manière à diminuer sa volatilité, il faut donc inverser cette corrélation entre les prix de l'actif et ceux du contrat à terme. C'est en vendant des contrats à terme qu'on pourra y parvenir. Les pertes que l'on essuiera sur l'actif seront alors compensées par les gains réalisés sur la vente de contrats à terme, ce qui est le principe même de la couverture (*hedging*).

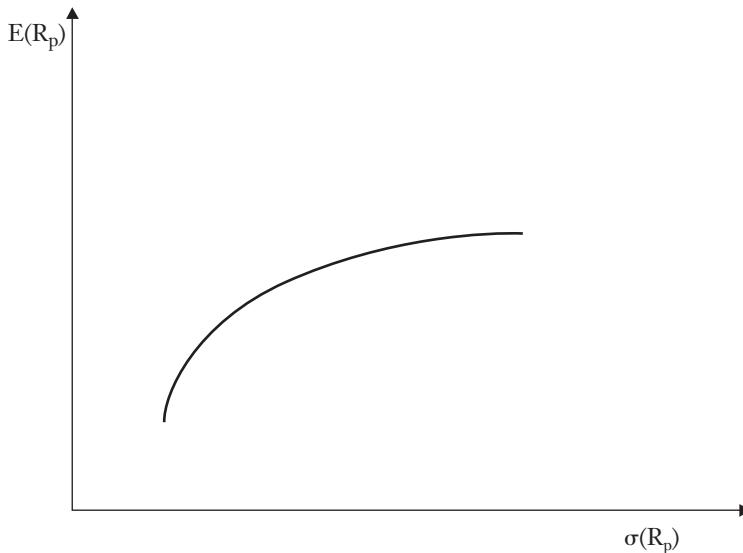
Supposons maintenant un cas de corrélation positive entre le rendement d'un passif et celui du contrat à terme. V entre alors négativement dans la formule de la couverture. Celle-ci indique alors qu'il faut acheter des contrats à terme pour les fins de la couverture (N positif). En vertu de la corrélation positive, les prix du passif et ceux du contrat à terme tendent à évoluer à l'unisson. Mais le passif étant une dette, lorsque la valeur du passif augmente, on espère récupérer cette perte par un gain sur le contrat à terme. Comme les prix du passif et du contrat à terme évoluent dans le même sens, c'est en détenant des contrats à terme que l'on pourra alors compenser cette perte par un gain sur le contrat à terme. Donc, lorsqu'il existe une corrélation positive entre le rendement d'un passif et celui d'un contrat à terme, il faut acheter des contrats pour se couvrir. On laisse au lecteur le soin de développer les cas de couverture qui se rapportent à une corrélation négative entre le rendement d'un actif (ou d'un passif) et celui du contrat à terme⁵.

5. Nous avons estimé le ratio de *hedging* pour le cas des BAX à la fin du chapitre 3. Cette section permet évidemment de mieux l'interpréter.

4.2. Une méthode générale pour calculer une frontière efficiente

Nous nous attaquons maintenant au cas de la construction d'une frontière efficiente. Celle-ci est le lieu des combinaisons rendement-risque optimales de portefeuilles. Pour chaque espérance de rendement donnée, la frontière efficiente donne le portefeuille qui comporte l'écart-type minimal, soit le risque minimal, pour cette espérance de rendement. La frontière efficiente classique apparaît à la figure A3-1.

FIGURE A3-1 La frontière efficiente classique



où $E(R_p)$ désigne l'espérance du rendement d'un portefeuille et $\sigma(R_p)$, son écart-type. Le problème de la construction d'une frontière efficiente consiste donc à minimiser la variance du rendement du portefeuille :

$$\text{Var}(R_p) = \sum_{i,j} w_i w_j \sigma_{ij}$$

w_i étant la pondération du titre i dans le portefeuille p , sous deux contraintes :

- i) la contrainte d'un niveau donné de rendement espéré E^* :

$$\sum w_i E(R_i) = E^*$$

$E(R_i)$ désignant l'espérance de rendement du titre i ;

- ii) la contrainte que la somme des pondérations des titres dans le portefeuille soit égale à l'unité :

$$\sum w_i = 1$$

Pour solutionner ce problème en termes des w_i , nous devons écrire la fonction de Lagrange qui lui correspond :

$$z = \sum_{i,j} w_i w_j \sigma_{ij} + \lambda_1 \left[\sum w_i E(r_i) - E^* \right] + \lambda_2 \left[\sum w_i - 1 \right]$$

Pour trouver le minimum, nous devons évaluer les dérivées premières de cette fonction par rapport aux w_i et aux λ_i à 0. La solution donne alors les w_i optimaux associés à E^* . L'on insère ces w_i dans la formule de l'écart-type, ce qui nous donne l'écart-type minimal associé à E^* . On obtient un point de la frontière efficiente : (σ^*, E^*) . Et l'on refait cet exercice pour d'autres niveaux de E^* de façon à générer toute la frontière efficiente.

Notre propos est ici de formuler ce problème d'optimisation sous forme matricielle. En termes matriciels, la fonction de Lagrange précédente s'écrit :

$$z = \mathbf{w}^T \mathbf{\Omega} \mathbf{w} + \lambda_1 \left[\mathbf{w}^T \mathbf{E} - E^* \right] + \lambda_2 \left[\mathbf{w}^T \mathbf{1} - 1 \right]$$

Les dérivées pour trouver le vecteur \mathbf{w}^* optimal associé à E^* sont les suivantes :

$$\begin{aligned} \frac{\partial z}{\partial \mathbf{w}} &= 2\mathbf{\Omega} \mathbf{w} + \lambda_1 \mathbf{E} + \lambda_2 \mathbf{1} = 0 \\ \frac{\partial z}{\partial \lambda_1} &= \mathbf{w}^T \mathbf{E} - E^* = 0 \\ \frac{\partial z}{\partial \lambda_2} &= \mathbf{w}^T \mathbf{1} - 1 = 0 \end{aligned}$$

Pour mieux visualiser comment solutionner ce système d'équations en termes des w_i , concentrons-nous sur le cas d'un portefeuille qui ne comprend que trois titres. En termes matriciels, ce système s'écrit alors :

$$\begin{bmatrix} 2\sigma_1^2 & 2\sigma_{12} & 2\sigma_{13} & E_1 & 1 \\ 2\sigma_{21} & 2\sigma_2^2 & 2\sigma_{23} & E_2 & 1 \\ 2\sigma_{31} & 2\sigma_{32} & 2\sigma_{33} & E_3 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ E_1 & E_2 & E_3 & 0 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ E^* \end{bmatrix}$$

En termes plus compacts, ce système s'écrit comme suit :

$$\mathbf{C}\mathbf{y} = \mathbf{k}$$

où \mathbf{y} est le vecteur des inconnues, soit les w_i et les λ_i . La solution s'obtient en inversant la matrice \mathbf{C} :

$$\mathbf{y} = \mathbf{C}^{-1}\mathbf{k}$$

On obtient alors le vecteur \mathbf{w}^* associé à E^* . On peut dès lors calculer le σ^* correspondant, ce qui nous fournit un point de la frontière efficiente : (σ^*, E^*) . On refait par la suite le même exercice pour d'autres niveaux de E^* de manière à pouvoir tracer toute la frontière efficiente⁶.

6. À remarquer qu'il est très facile de construire une frontière efficiente sur Excel lorsque l'on dispose de la matrice variance-covariance des rendements des titres analysés en recourant aux fonctions matricielles PRODUITMAT (produit matriciel) et INVERSEMAT (inversion matricielle). En effet, le calcul de la frontière efficiente ne met en jeu que ces deux opérations matricielles.

4.3. Le modèle de l'erreur de suivi (*tracking error*) de Roll⁷

Nous envisageons ici le cas du gestionnaire qui contrôle un fonds investi dans une catégorie d'actifs donnée. Il apparaît logique de mesurer sa performance par rapport à un indice de référence (*benchmark*) qui se compare bien à la composition dudit fonds. À cet effet, plusieurs maisons de courtage publient des indices de portefeuilles spécialisés. On peut alors comparer la performance d'un gestionnaire de portefeuille à l'indice publié qui correspond le mieux à la composition de son portefeuille.

Roll a proposé une telle comparaison dans un article publié en 1992 dans le *Journal of Portfolio Management* : « A Mean/Variance Analysis of Tracking Error ». Son but : est d'évaluer la performance d'un gestionnaire en le comparant au rendement d'un indice de référence (*benchmark*) apparenté au comportement du fonds géré par ledit gestionnaire.

Mais avant de poursuivre, définissons un concept utilisé dans cet article : l'erreur de suivi ou *tracking error*. Cet indicateur traduit une déviation de performance. C'est la différence entre le taux de rendement périodique du fonds et le taux de rendement de l'indice de référence. Un critère pertinent pour évaluer un gestionnaire est la minimisation du *tracking error*, mais cette minimisation doit certes s'effectuer sous contrainte. Fixons au gestionnaire l'objectif suivant : minimiser la variance du *tracking error* sous la contrainte d'un certain écart positif de rendement relativement à l'indice de référence.

La valeur espérée de l'erreur de performance (G) se définit comme suit :

$$G = (\mathbf{q}_p - \mathbf{q}_b)^T \mathbf{R} = \mathbf{x}^T \mathbf{R}$$

où \mathbf{q}_p est le vecteur de pondération du portefeuille du gestionnaire vis-à-vis des N titres qui constituent le *benchmark* et \mathbf{q}_b , le facteur de pondération de ces titres à l'intérieur du *benchmark*. Dans le portefeuille du gestionnaire comme dans le *benchmark*, la somme des pon-

7. Roll, R. (1992), « A Mean/Variance Analysis of Tracking Error », *Journal of Portfolio Management*, Summer 1992.

dérations des titres est égale à l'unité. Le vecteur \mathbf{x} : de dimension $(N \times 1)$ représente des altérations de portefeuille qui s'autofinancent, c'est-à-dire :

$$\mathbf{x}^T \mathbf{1} = 0$$

où $\mathbf{1}$ est un vecteur $(N \times 1)$ d'unités et \mathbf{x} , l'écart entre les vecteurs \mathbf{q}_p et \mathbf{q}_b . Finalement, \mathbf{R} : $(N \times 1)$ est le vecteur des rendements anticipés des N titres.

Fixons G à un niveau cible, soit à l'excédent positif de rendement imposé au gestionnaire en regard du *benchmark*. Soit $(R_p - R_b)^*$ cet excédent-cible. On a donc :

$$G = (R_p - R_b)^*$$

Soit à désigner par \mathbf{V} la matrice variance-covariance des N titres en cause. La variance du *tracking-error* est donc :

$$(\mathbf{q}_p - \mathbf{q}_b)^T \mathbf{V} (\mathbf{q}_p - \mathbf{q}_b) = \mathbf{x}^T \mathbf{V} \mathbf{x}$$

Le problème d'optimisation auquel est soumis le gestionnaire est le suivant. Il doit minimiser la variance du *tracking error* :

$$\mathbf{x}^T \mathbf{V} \mathbf{x}$$

sous les deux contraintes suivantes :

- i) les réaménagements de portefeuille qu'effectue le gestionnaire par rapport au *benchmark* doivent comporter une espérance de gain G positive, ce qui correspond à la cible recherchée :

$$\mathbf{x}^T \mathbf{R} = G$$

- ii) les réaménagements par rapport au portefeuille de référence doivent s'autofinancer :

$$\mathbf{x}^T \mathbf{1} = 0$$

La fonction de Lagrange qui correspond à ce problème d'optimisation est la suivante :

$$\mathbf{x}^T \mathbf{V} \mathbf{x} + \lambda_1 [G - \mathbf{x}^T \mathbf{R}] + \lambda_2 [0 - \mathbf{x}^T \mathbf{1}]$$

Pour trouver le vecteur \mathbf{x} optimal, on égale les dérivées premières de la fonction de Lagrange par rapport à \mathbf{x} et aux λ_i à 0. Si nous dérivons cette fonction par rapport au vecteur \mathbf{x} et égalons cette dérivée à 0, nous obtenons :

$$\mathbf{V}\mathbf{x} - \lambda_1 \mathbf{R} - \lambda_2 \mathbf{1} = \mathbf{0}$$

Mettons le vecteur \mathbf{x} en évidence :

$$\mathbf{V}\mathbf{x} = [\mathbf{R} \quad \mathbf{1}] \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$$

En multipliant les deux côtés par \mathbf{V}^{-1} , on obtient :

$$\mathbf{x} = \mathbf{V}^{-1} [\mathbf{R} \quad \mathbf{1}] \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$$

Pour trouver la solution, nous devons éliminer les deux multiplicateurs de Lagrange. Pour y arriver, on multiplie les deux côtés de la dernière équation par le vecteur $[\mathbf{R} \quad \mathbf{1}]^T$:

$$[\mathbf{R} \quad \mathbf{1}]^T \mathbf{x} = [\mathbf{R} \quad \mathbf{1}]^T \mathbf{V}^{-1} [\mathbf{R} \quad \mathbf{1}] \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$$

Nous remplaçons le terme de gauche par les dérivées de la fonction de Lagrange par rapport aux λ_i , qui sont égales à 0 dans la solution optimale. Ces dérivées correspondent évidemment aux contraintes du problème d'optimisation. En effectuant cette substitution, on obtient :

$$\begin{bmatrix} \mathbf{G} \\ 0 \end{bmatrix} = \mathbf{A} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$$

En multipliant les deux côtés de la dernière équation par \mathbf{A}^{-1} , on trouve finalement les valeurs des deux λ recherchées :

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \mathbf{A}^{-1} \begin{bmatrix} \mathbf{G} \\ 0 \end{bmatrix}$$

Et en se référant aux calculs précédents, on trouve l'expression de la matrice \mathbf{A} :

$$\mathbf{A} = \begin{bmatrix} \mathbf{R}^T \mathbf{V}^{-1} \mathbf{R} & \mathbf{R}^T \mathbf{V}^{-1} \mathbf{1} \\ \mathbf{R}^T \mathbf{V}^{-1} \mathbf{1} & \mathbf{1}^T \mathbf{V}^{-1} \mathbf{1} \end{bmatrix} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

Le vecteur optimal \mathbf{x} des réaménagements est le suivant :

$$\mathbf{x} = \mathbf{V}^{-1} \begin{bmatrix} \mathbf{R} & \mathbf{1} \end{bmatrix} \mathbf{A}^{-1} \begin{bmatrix} \mathbf{G} \\ 0 \end{bmatrix}$$

On a donc trouvé les réaménagements du portefeuille de référence (*benchmark*) qui minimisent la variance du *tracking error* et qui assurent un rendement espéré positif G .

VARIATIONS SUR LES MODÈLES LINÉAIRE ET NON LINÉAIRE

PARTIE I

Dans ce chapitre, nous développons plusieurs thèmes rattachés au modèle linéaire général. Dans un premier temps, nous abordons les erreurs de spécification de la forme fonctionnelle et les tests qui s'y rattachent. Certains de ces tests font appel aux techniques d'estimation non linéaires : ces techniques seront donc introduites dans ce chapitre. Puis, nous nous penchons sur les critères de sélection des variables explicatives dans le modèle de régression et sur d'autres critères associés : critères d'Akaike, de Schwartz, test J, test RESET. Ensuite, nous considérons les variables dichotomiques (*dummy*) et les tests ayant trait aux changements structurels.

1. ERREURS DE SPÉCIFICATION

Nous abordons ici les erreurs de spécification de la forme fonctionnelle reliées au cas des variables explicatives omises. Supposons que le vrai modèle à estimer soit le suivant :

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e}_2$$

alors que, par ignorance, on estime le modèle suivant :

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{e}_1$$

où $\mathbf{e}_1 = \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e}_2 \sim \text{IID}(0, \sigma^2\mathbf{I}_T)$. Quelles sont les conséquences de cette erreur de spécification sur l'estimateur des MCO ? En estimant par les MCO la dernière équation, on a :

$$\begin{aligned}\hat{\beta}_1 &= (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{y} \\ &= (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T (\mathbf{X}_1 \beta_1 + \mathbf{X}_2 \beta_2 + \mathbf{e}_2) \\ &= \beta_1 + (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{X}_2 \beta_2 + (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{e}_2\end{aligned}$$

Calculons maintenant l'espérance de $\hat{\beta}_1$:

$$\begin{aligned}E(\hat{\beta}_1) &= \beta_1 + E\left[(\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{X}_2 \beta_2\right] + \mathbf{0} \\ &= \beta_1 + \left[(\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{X}_2 \beta_2\right] \neq \beta_1\end{aligned}$$

De ces développements, il résulte que si l'on omet une variable explicative et si \mathbf{X}_1 n'est pas orthogonal à \mathbf{X}_2 et $\beta_2 \neq 0$, alors β_1 est biaisé.

Examinons maintenant le cas où l'on ajoute trop de variables explicatives. Le vrai modèle est le suivant :

$$\mathbf{y} = \mathbf{X}_1 \beta_1 + \mathbf{e}_1$$

Or, on estime plutôt le modèle suivant :

$$\mathbf{y} = \mathbf{X}_1 \beta_1 + \mathbf{X}_2 \beta_2 + \mathbf{e}_2$$

où $\mathbf{e}_1 = \mathbf{e}_2$ puisque $\beta_2 = 0$. Appliquons les MCO sur le dernier modèle et dégageons-en les conséquences.

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}^T \mathbf{X})^{-1} \begin{bmatrix} \mathbf{X}_1^T \\ \mathbf{X}_2^T \end{bmatrix} (\mathbf{X}_1 \beta_1 + \mathbf{e}_1) \\ &= \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}_1^T \mathbf{X}_1 \beta_1 + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}_1^T \mathbf{X}_1 \mathbf{e}_1 \\ (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}_2^T \mathbf{X}_1 \beta_1 + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}_2^T \mathbf{X}_1 \mathbf{e}_1 \end{bmatrix}\end{aligned}$$

Calculons l'espérance de $\hat{\beta}$:

$$E(\hat{\beta}) = \begin{bmatrix} E(\hat{\beta}_1) \\ E(\hat{\beta}_2) \end{bmatrix} = \begin{bmatrix} \beta_1 \\ 0 \end{bmatrix}$$

puisque \mathbf{X}_2 est supposé orthogonal à \mathbf{X}_1 , \mathbf{X}_2 n'apparaissant pas dans le vrai modèle. L'estimateur de β est donc sans biais. Cependant, il est inefficace parce que l'on sait que c'est le premier modèle qui est correctement spécifié. Ce problème est donc moindre que celui de l'omission de variables explicatives.

Passons maintenant aux problèmes suscités par les erreurs de spécification sur la variable dépendante. Supposons que l'on ait estimé le modèle suivant :

$$y_t = \beta_1 x_t + e_{1t}$$

alors que le vrai modèle est :

$$\ln(y_t) = \beta_2 x_t + e_{2t} \Rightarrow y_t = \exp(\beta_2 x_t + e_{2t}) = \exp(\beta_2 x_t) \exp(e_{2t})$$

En appliquant les MCO sur l'équation correspondant à la mauvaise spécification, on a :

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \exp(\beta_2 \mathbf{X}) \exp(e_2)$$

En calculant l'espérance de $\hat{\beta}$, on a :

$$\begin{aligned} E(\hat{\beta}) &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \exp(\beta_2 \mathbf{X}) E(\exp(e_2)) \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \exp(\beta_2 \mathbf{X}) \exp\left(\frac{\sigma^2}{2}\right) \neq \beta \end{aligned}$$

Nous recourons à certains résultats de la distribution lognormale pour calculer l'espérance de e_2 . Soit $x = \ln(y)$, où $x \sim N(0, \sigma^2)$. Alors

$$E(y) = E(e^x) = e^{\left(E(x) + \frac{1}{2}V(x)\right)} = e^{\frac{1}{2}\sigma^2}. \text{ La variance de } y \text{ est par ailleurs}$$

de: $V(y) = e^{(2E(x)+V(x))} \left(e^{(V(x))} - 1 \right) = e^{\sigma^2} \left(e^{\sigma^2} - 1 \right)$. Une erreur de spécification sur la variable dépendante entraîne également un biais au chapitre de l'estimateur des MCO.

2. TESTS RELIÉS AUX ERREURS DE SPÉCIFICATION

2.1. Test sur la forme logarithmique

On veut choisir l'un des deux modèles suivants :

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + e_{Lt}$$

$$\ln(y_t) = \gamma_1 + \gamma_2 x_{2t} + \gamma_3 x_{3t} + e_{LLt}$$

Box et Cox ont formulé un test pour discriminer entre ces deux modèles. Il faut d'abord exprimer le premier modèle en unités comparables au second. Pour ce faire, on calcule la moyenne géométrique de

y_t : $\bar{y}_G = (y_1 \times y_2 \times \dots \times y_T)^{\frac{1}{T}}$. On peut écrire une formule équivalente de la moyenne géométrique en exprimant cette dernière équation sous forme logarithmique :

$\ln(\bar{y}_G) = \frac{1}{T} \sum_{t=1}^T \ln y_t \Rightarrow \bar{y}_G = e^{\frac{1}{T} \sum_{t=1}^T \ln y_t}$. Ce résultat est utilisé pour calculer la somme des carrés résiduels transposés en unités comparables à celles du deuxième modèle. On obtient, après application des MCO sur le premier modèle :

$$SCR_L = \frac{\hat{e}_L^T \hat{e}_L}{\bar{y}_G^2}$$

Par ailleurs, en appliquant les MCO sur la deuxième équation, on a :

$$SCR_{LL} = \hat{e}_{LL}^T \hat{e}_{LL}$$

On est maintenant en mesure de construire le test suggéré par Box-Cox en 1964 :

$$l = \frac{T}{2} \left| \ln \frac{\text{SCR}_L / \bar{y}_G^2}{\text{SCR}_{LL}} \right|^a \sim \chi^2(1)$$

La formulation du test est la suivante :

H0 : les deux modèles sont équivalents

H1 : les deux modèles ne sont pas équivalents

La première étape du test consiste à calculer la statistique l . On rejette H0 si $l > l_c$ au niveau α . Dans une deuxième étape, on discrimine entre

les deux modèles. On choisit le modèle linéaire si : $\frac{\text{SCR}_L}{\bar{y}_G^2} < \text{SCR}_{LL}$.
On choisit le modèle logarithmique dans le cas inverse.

2.2. La transformation de Box-Cox

Le test précédent est plutôt restrictif car il nous limite au couple modèle linéaire–modèle logarithmique. La transformation Box-Cox s'avère beaucoup plus flexible, car elle nous laisse le choix entre plusieurs modèles, entre autres les modèles linéaire simple, logarithmique et beaucoup d'autres comme le modèle non linéaire. Pour effectuer une telle transformation, il faut reformuler les variables dépendantes et indépendantes comme suit. Supposons à cet effet le modèle suivant :

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + e_t$$

La transformation Box-Cox des variables de ce modèle est la suivante :

$$y_t^{(\lambda)} = \frac{y_t^\lambda - 1}{\lambda}; \quad x_{2t}^{(\lambda)} = \frac{x_{2t}^\lambda - 1}{\lambda}; \quad x_{3t}^{(\lambda)} = \frac{x_{3t}^\lambda - 1}{\lambda}$$

Cette transformation implique les relations suivantes pour y :

$$y_t^{(\lambda)} = \begin{cases} \frac{y_t^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \ln(y_t) & \text{si } \lambda = 0 \end{cases}$$

Et ces relations valent également pour les x_{it} .

Démontrons ces relations. Pour ce faire on doit calculer la limite de l'expression suivante :

$$\lim_{\lambda \rightarrow 0} \frac{y_t^\lambda - 1}{\lambda} = \frac{d(y_t^\lambda - 1)}{d\lambda}$$

Pour pouvoir effectuer ce calcul, on recourt à deux règles : la règle de l'Hôpital et une règle de dérivation. La règle de l'Hôpital s'écrit comme suit :

$$\lim_{x \rightarrow 0} \frac{m(x)}{n(x)} = \lim_{x \rightarrow 0} \frac{m'(x)}{n'(x)}$$

où $m(x)$ et $n(x)$ sont des fonctions continûment différentiables.

Par ailleurs, la règle de dérivation requise est la suivante. Soit la fonction :

$$f(\cdot) = x^b$$

La dérivée de cette fonction est :

$$\frac{df(\cdot)}{db} = \ln(x) x^b$$

On se sert de cette règle pour déterminer le numérateur de la règle de l'Hôpital :

$$\frac{d(y_t^\lambda - 1)}{d\lambda} = y_t^\lambda \ln(y_t)$$

La dérivée du dénominateur de la règle de l'Hôpital est triviale. Ces calculs impliquent que :

$$\lim_{\lambda \rightarrow 0} y_t^\lambda \ln(y_t) = \ln(y_t)$$

Précisons les diverses formes d'équations que peut accommoder la transformation Box-Cox. Si $\lambda = 1$, on revient au modèle linéaire. Si $\lambda = 0$, on a la forme logarithmique. Si $\lambda = 2$, on a une forme quadratique. Et ainsi de suite. Ce modèle non linéaire s'estime par la méthode du maximum de vraisemblance. Il convient de souligner que pour les modèles non linéaires, les tests ne valent qu'asymptotiquement.

On veut maintenant tester les deux hypothèses suivantes :

$$H_0 : \lambda = 0$$

$$H_1 : \lambda \neq 0$$

On construit le ratio t asymptotiquement normal :

$$t = \frac{\hat{\lambda} - 0}{\sqrt{\hat{V}(\hat{\lambda})}} \xrightarrow{d} N(0,1). \text{ On rejette } H_0 \text{ si : } t > 1,96 \text{ pour } \alpha = 5 \%.$$

3. MÉTHODES DES MOINDRES CARRÉS NON LINÉAIRES ET TRANSFORMATION BOX-COX

La méthode des moindres carrés non linéaires se présente comme suit. Soit le modèle de régression non linéaire suivant :

$$y_t = f(x_t, \beta) + e_t$$

Pour estimer le vecteur de paramètres b , on minimise :

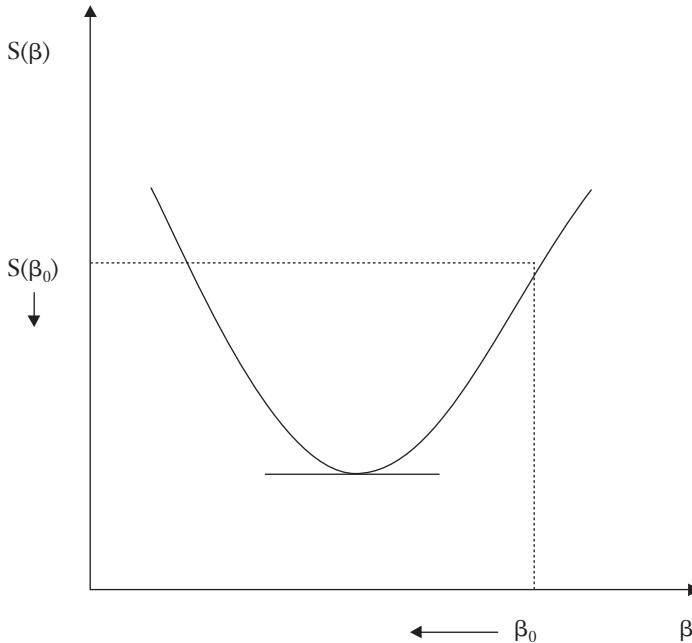
$$\text{Min}_{\beta} S(\beta) = \text{Min}_{\beta} \sum_{t=1}^T (y_t - f(x_t, \beta))^2$$

Les méthodes non linéaires recourent à des algorithmes d'optimisation. Nous présentons d'abord le plus connu, soit l'algorithme de Newton-Raphson, qui fait appel au processus itératif suivant :

$$\beta_{n+1} = \beta_n - \frac{S(\beta_n)}{S'(\beta_n)}$$

Représentons à la figure 4.1 la procédure décrite par cette équation.

FIGURE 4.1



On veut trouver le minimum de $S(\beta)$ qui correspond au point où $S'(\beta) = 0$. La procédure Newton-Raphson consiste à initier l'algorithme au point β_0 et à évaluer à ce point $S(\beta)$ et $S'(\beta)$. Si $S'(\beta_0) > 0$,

cela implique que $\frac{S(\beta_0)}{S'(\beta_0)} > 0$. Mais comme ce ratio est affecté d'un

signe négatif, alors sur la figure 4.1 la direction sera vers la gauche. On obtient alors β_1 , que l'on remet dans l'équation précédente pour obtenir β_2 et ainsi de suite jusqu'à ce que l'écart entre β_n et β_{n+1} soit très petit.

La méthode de Newton-Raphson qui vient d'être présentée ne tient pas compte de la courbure de $S(\beta)$. Une méthode qui en tient compte est celle de Newton. Cet algorithme s'écrit comme suit :

$$\beta_{n+1} = \beta_n - \frac{S'(\beta_n)}{S''(\beta_n)}$$

Cette équation est dérivée de l'approximation de Taylor. Dans cette expression, la courbure de $S(\beta)$ est évaluée par sa dérivée seconde.

Envisageons ici le cas des moindres carrés linéaires¹. Situons-nous à la première itération. La formule de Newton s'écrit alors :

$$\beta_2 = \beta_1 - \frac{S'(\beta_1)}{S''(\beta_1)}$$

Il suffit alors de minimiser la somme des carrés résiduels : $(\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta)$. Calculons les dérivées premières qui apparaissent dans la

formule itérative. $S'(\beta_1) = \frac{\partial(\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta)}{\partial\beta} = 2\mathbf{X}^T (\mathbf{y} - \mathbf{X}\beta)$, soit

les équations normales. $S''(\beta) = \frac{\partial^2(\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta)}{(\partial\beta)^2} = 2\mathbf{X}^T \mathbf{X}$. La for-

mule de Newton s'écrit donc dans ce cas :

$$\beta_2 = \beta_1 + (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{y} - \mathbf{X}^T \mathbf{X}\beta_1) = \beta_1 + \hat{\beta} - \beta_1 = \hat{\beta}$$

Pour les MCO, la méthode de Newton converge en une seule itération. La dérivée première représente la direction du déplacement et la dérivée seconde, l'accélération du mouvement².

Le problème relié aux calculs numériques est qu'ils s'avèrent intensifs en calculs. La méthode de Gauss-Newton vient pallier en partie à ce problème en évaluant la dérivée seconde par un produit de dérivées premières. Cette méthode s'écrit comme suit :

$$\beta_{n+1} = \beta_n - p_n d_n$$

-
1. Hendry, D.F. (1995), *Dynamic Econometrics*, Oxford University Press, Oxford.
 2. Ou encore, elle tient compte de la courbure de la fonction à minimiser.

$$\text{où } d_n = \left. \frac{dS}{d\beta} \right|_{\beta_n} \text{ et } p_n = \frac{1}{2} \left[\sum_{t=1}^T \left(\frac{df_t}{d\beta} \right)^2 \right]^{-1}.$$

Pour illustrer la méthode Gauss-Newton et, par le fait même, les autres méthodes, supposons la régression non linéaire suivante, qui nous sert à estimer le coefficient β_1^3 .

$$y_t = \beta x_{1t} + \beta^2 x_{2t} + e_t$$

Ce modèle est de la forme: $y_t = f(\beta) + e_t$. Cette équation fait apparaître un produit de paramètres, c'est-à-dire que lorsque l'on minimise la somme des carrés résiduels, on n'obtient pas une forme explicite pour l'estimateur de β . Supposons en effet qu'on minimise $S(\beta)$, qui est égal à :

$$S(\beta) = \sum_{t=1}^T (y_t - \beta x_{1t} - \beta^2 x_{2t})^2$$

En égalant la dérivée de cette fonction à 0, on obtient :

$$\frac{dS}{d\beta} = 0 \Rightarrow \sum_{t=1}^T (y_t - \beta x_{1t} - \beta^2 x_{2t})(-x_{1t} - 2\beta x_{2t}) = 0$$

On remarque qu'il n'y a pas de solution analytique pour β . On doit donc recourir aux méthodes numériques pour trouver une solution pour β . Dans ce cas-ci, f_t , qui apparaît dans l'algorithme de Gauss-

Newton, est égal à : $f_t = \beta x_{1t} + \beta^2 x_{2t}$. D'où : $\frac{df_t}{d\beta} = x_{1t} + 2\beta x_{2t}$, ce qui

est requis pour le calcul de p_n . Finalement, d_n est égal à la dérivée de S par rapport à β évalué à β_n . Par conséquent, le processus itératif est le suivant :

$$\beta_{n+1} = \beta_n - \left[\frac{1}{2} \left(\sum_{t=1}^T (x_{1t} + 2\beta x_{2t})^2 \right) \right]^{-1} \left(\left. \frac{dS}{d\beta} \right|_{\beta_n} \right)$$

3. Cet exemple est emprunté à : Griffiths, W.E., C. Hill et G.G. Judge (1993), *Learning and Practicing Econometrics*, John Wiley, New York.

Une autre application des moindres carrés non linéaires est d'estimer la transformation Box-Cox du modèle de régression suivant de la demande de monnaie aux États-Unis⁴:

$$\ln(M_t) = \beta_1 + \beta_2 r_t^{(\lambda)} + \beta_3 y_t^{(\lambda)} + e_t$$

où M_t est la masse monétaire américaine au sens de M2, r_t , le taux d'escompte de la Réserve fédérale américaine et y_t , le produit national brut en dollars constants. Pour estimer les paramètres de cette équation, on applique les moindres carrés non linéaires. Les hypothèses à tester sont les suivantes: $H_0: \lambda = 0$; $H_1: \lambda \neq 0$. Pour ce faire, il faut construire la statistique t qui obéit asymptotiquement à la distribution normale. On rejettera H_0 si le t calculé est supérieur au seuil critique donné par la loi normale.

Jusqu'ici, à l'intérieur de la présentation des méthodes numériques, nous n'avons envisagé que le cas univarié. Le cas de la régression de Box-Cox en est cependant un où il existe plusieurs paramètres à estimer. La formulation générale du modèle de régression non linéaire multivarié est la suivante: $y_t = f(x_t, \beta) + e_t$. Pour le cas de la régression Box-Cox: $\beta^T = [\beta_1 \quad \beta_2 \quad \beta_3 \quad \lambda]$. Pour estimer ces paramètres, l'algorithme d'optimisation requiert l'évaluation

$$\frac{\partial S}{\partial \beta} = \left[\frac{\partial S}{\partial \beta_1} \quad \frac{\partial S}{\partial \beta_2} \quad \frac{\partial S}{\partial \beta_3} \quad \frac{\partial S}{\partial \lambda} \right]^T. \text{ Ce vecteur est appelé : gradient,}$$

qui est un vecteur de dérivées premières. Dans le cas multivarié, pour utiliser les algorithmes de Newton-Raphson, de Gauss-Newton ou de Newton, il suffit donc simplement de remplacer les dérivées simples par les dérivées partielles.

À titre d'exemple d'une régression non linéaire, considérons un modèle de détermination de la structure à terme des taux d'intérêt. Ce modèle, associé à Haugen (1993)⁵, relie le rendement des obligations américaines à leur échéance, de façon à déterminer la courbe de leur structure à terme. Cette équation est la suivante :

$$y_i = (a_1 + a_2 t_i) e^{-a_3 t_i} + a_4$$

4. Voir Greene, W.H. (2000), *Econometric Analysis*, Prentice Hall, New York.

5. Haugen, R.A. (1993), *Modern Investment Theory*, troisième édition, Prentice Hall, New York.

où y_i est le rendement à l'échéance de l'obligation de catégorie i , t_i , l'échéance correspondante de cette obligation et a_i , les divers coefficients qui servent à ajuster la courbe. Cette régression est à l'évidence non linéaire. Les paramètres a_1 à a_4 sont estimés par le logiciel Eviews par la méthode des moindres carrés non linéaires. Il faut bien sûr fixer des valeurs de départ (*seed values*) pour les coefficients a_1 à a_4 . Les commandes pour effectuer cette régression sur EViews apparaissant au tableau 4.1.

TABLEAU 4.1

 simpl 1 50

 ' Estimation de la courbe des rendements à l'échéance à partir de yld⁶ et tm

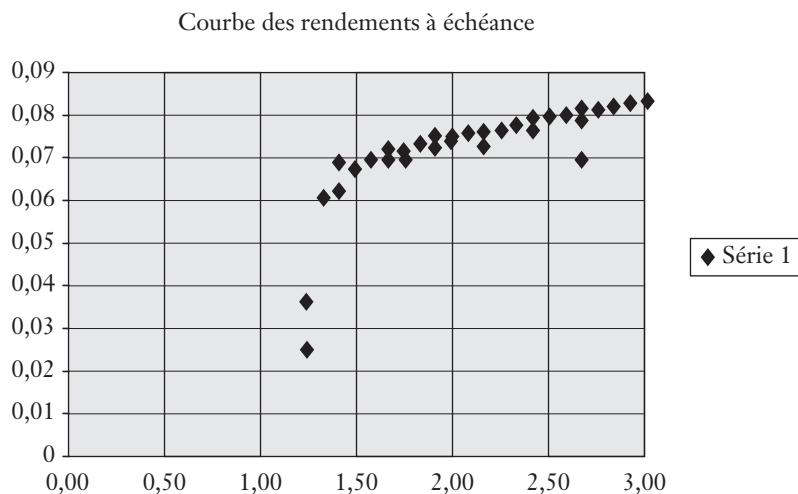
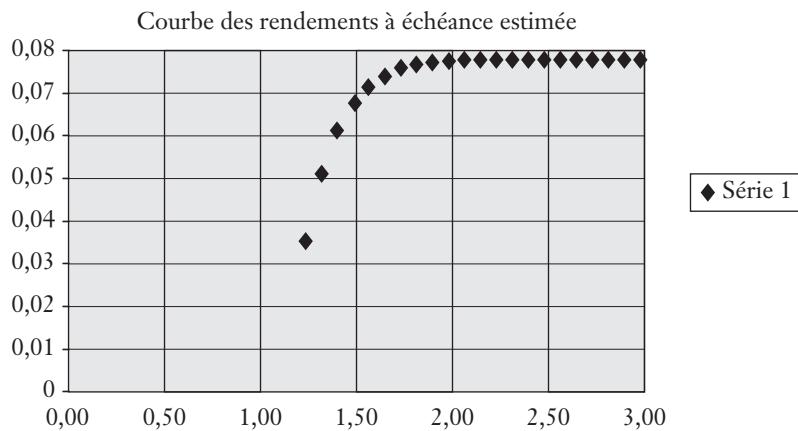
' Déclaration des paramètres de départ :

PARAM c(1) -.49 c(2) .1 c(3) .1 c(4) .1

 NLS yld=((c(1)+c(2)*tm)/exp(c(3)*tm))+c(4)

Les coefficients estimés par NLS sont les suivants : $a_1 = -45,96$; $a_2 = -0,6285$; $a_3 = 5,5842$; $a_4 = 0,0783$. La représentation graphique de la structure à terme observée un 13 mars pour les obligations fédérales américaines de diverses échéances et de la structure à terme estimée apparaît aux figures 4.2 et 4.3.

6. yld désigne le rendement à l'échéance et tm, l'échéance.

FIGURE 4.2 **Structure à terme observée**FIGURE 4.3 **Structure à terme estimée**

4. CRITÈRES DE SÉLECTION DES VARIABLES EXPLICATIVES

Dans cette section, nous abordons les différents tests statistiques qui servent à vérifier les modèles théoriques. Rappelons d'abord rapidement les différents tests examinés jusqu'ici pour la même fin. Nous avons étudié le test t , servant à tester un seul paramètre dans sa présentation classique, le test F , servant à tester un sous-ensemble ou l'ensemble des paramètres d'une régression, et le R^2 ajusté, qui sert également à évaluer l'ensemble des paramètres d'une régression.

Il convient maintenant de présenter d'autres critères classiques pour sélectionner les variables explicatives d'un modèle. Dans la même veine que le critère du R^2 ajusté, les critères d'information d'Akaike et de Schwartz visent également à évaluer le degré d'ajustement d'un modèle tout en tenant compte de l'ajout de nouvelles variables explicatives. Le critère d'Akaike pour le modèle i , désigné par AIC_i ⁷, se définit comme suit :

$$AIC_i = \ln\left(\frac{SCR_i}{T}\right) + 2\frac{k_i}{T}$$

où SCR_i est la somme des carrés résiduels du modèle i ; T , la taille de l'échantillon; k_i , le nombre de variables explicatives. On est confronté à plusieurs modèles et celui qui minimise le critère AIC est celui qui sera retenu. À l'instar du R^2 ajusté, ce critère prend en compte le nombre de degrés de liberté. En raison de la présence du logarithme, ce critère peut donner lieu à des nombres négatifs. On choisit alors le modèle qui comporte l'AIC le plus faible. Un autre critère, celui de Schwarz, désigné par SC ou SBC⁸, se définit comme suit :

$$SC_i = \ln\left(\frac{SCR_i}{T}\right) + \frac{k_i}{T} \ln(T)$$

Le critère SC s'interprète de la même façon que le critère AIC, c'est-à-dire que l'on choisit le modèle qui minimise le critère SC. Le critère SC comporte des propriétés asymptotiques supérieures au critère AIC. Le critère AIC est biaisé en faveur des modèles surparamétrisés.

7. Abréviation de *Akaike information criterion*.

8. Abréviation de *Schwarz criterion* (SC) et de *Schwarz Bayesian Criterion* (SCB).

Le test J

Le test J proposé par Davidson et McKinnon⁹ tire son appellation de l'expression anglaise *jointly estimating*. Si les critères de décision précédents favorisent un modèle, il n'en reste pas moins que la procédure suivie est plus ou moins rigoureuse. Une procédure plus satisfaisante consiste à tester conjointement (simultanément) deux modèles concurrents. Le test J est alors tout désigné.

La procédure du test J est la suivante. On veut tester les deux hypothèses. $H_0 : \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}_1$ contre $H_1 : \mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e}_2$. Théoriquement, on teste ces hypothèses en recourant à la stratégie du *nesting*, que l'on peut traduire par « stratégie de l'imbrication », pour écrire la relation suivante :

$$\mathbf{y} = (1 - \alpha)\mathbf{X}\boldsymbol{\beta} + \alpha\mathbf{Z}\boldsymbol{\gamma} + \mathbf{e}$$

Tester H_0 revient à vérifier si $\alpha = 0$. En pratique, le test, qui fait appel à la technique des régressions artificielles, s'effectue en trois étapes : i) régresser y sur Z . On obtient : $\hat{\mathbf{y}}^* = \mathbf{Z}\hat{\boldsymbol{\gamma}}$; ii) régresser \mathbf{y} sur \mathbf{X} et $\hat{\mathbf{y}}^*$: $\mathbf{y} = \mathbf{X}\boldsymbol{\Omega} + \alpha\hat{\mathbf{y}}^* + \boldsymbol{\varepsilon}$. On obtient $\hat{\boldsymbol{\alpha}}$; iii) on construit le test t asymptotique :

$t = \frac{\hat{\boldsymbol{\alpha}} - 0}{\sqrt{\hat{\mathbf{V}}(\hat{\boldsymbol{\alpha}})}} \xrightarrow{d} N(0,1)$. On rejette H_0 si la valeur absolue de t

excède la valeur critique.

Le test RESET

Par ailleurs, le test RESET¹⁰, qui est l'acronyme de l'expression anglaise *REgression Specification Error Test*, est, comme son nom l'indique, un test général d'erreurs de spécification, par exemple les erreurs omises ou les erreurs de forme fonctionnelle. Pour effectuer ce test,

9. Davidson, R. et J. McKinnon (1981), « Several Tests for Model Specification in the Presence of Alternatives Hypotheses », *Econometrica*, 49, p. 781-793.

10. Ce test est redevable à Ramsey, J.B. (1969), « Tests for Specification Error in Classical Linear Least Squares Analysis », *Journal of the Royal Statistical Society, Series B*, 31, p. 350-371. Voir aussi : Ramsey, J.B. et P. Schmidt (1976), « Some Further Results on the Use of OLS and BLUS in Residuals in Specification Error Tests », *Journal of the American Statistical Association*, 71, p. 389-390.

on suit la procédure suivante. On veut tester si le modèle $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, où $\mathbf{e} \sim N(0, \sigma^2\mathbf{I})$, est spécifié correctement. Ce test comporte trois étapes: i) on applique d'abord les MCO sur cette équation et on obtient $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$; ii) on écrit la régression suivante:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \gamma_1\hat{\mathbf{y}}^2 + \gamma_2\hat{\mathbf{y}}^3 + \gamma_3\hat{\mathbf{y}}^4 + \mathbf{u}$$

où $\mathbf{u} \sim N(0, \sigma^2\mathbf{I})$. On applique les MCO sur cette équation pour obtenir les $\hat{\boldsymbol{\theta}}$ et les $\hat{\gamma}$; iii) on construit la statistique F pour tester les hypothèses suivantes en supposant que le vecteur \mathbf{q} comporte trois paramètres: $H_0: \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ contre $H_1: \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}$, soit:

$$\mathbf{R}\boldsymbol{\beta} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Le test F s'écrit:

$$F = \frac{SCR_R - SCR_U / q}{SCR_U / T - k} \sim F(q, T - k)$$

où q est égal au nombre de restrictions, 3 dans le cas qui nous intéresse, et k , le nombre de paramètres à estimer, est égal à 6. SCR_R est la somme des carrés résiduels du modèle contraint ($\gamma = 0$) et SCR_U est la somme des carrés résiduels du modèle non contraint ($\gamma \neq 0$). On rejette H_0 si F est supérieur au F critique. Si ce test rejette l'hypothèse H_0 , cela indique la présence probable d'une erreur de spécification au chapitre du modèle. Mais le test reste muet sur l'identité de l'erreur de spécification.

5. VARIABLES AUXILIAIRES ET MODÉLISATION DES CHANGEMENTS STRUCTURELS

La théorie des options s'est énormément développée depuis que Black et Scholes ont présenté leur fameux modèle sur la détermination du prix d'une option d'achat européenne en 1973. Depuis cette époque, les catégories d'options se sont démultipliées. Elles ont même débordé le secteur financier pour envahir le secteur réel. À titre d'exemple, il existe plusieurs catégories d'options dans le secteur de l'énergie tels les caps sur le prix de la gazoline. Il est donc important dans ce cas de se donner un modèle de fixation du prix de la gazoline qui représente ici le sous-jacent du cap. Examinons comment l'économétrie peut nous aider à définir un tel modèle. Cet exemple nous sert également de prétexte pour introduire les variables auxiliaires¹¹.

Supposons à cet effet un modèle très simple de dépenses en gazoline définies en fonction du revenu et du prix, les deux premières variables étant définies sur une base réelle et per capita. Le modèle se présente comme suit :

$$\ln \text{dep}_t = \beta_1 + \beta_2 \ln y_t + \beta_3 \ln p_t + e_t \quad t = 1930, \dots, 1970$$

où dep_t représente les dépenses en gazoline, y_t , le revenu et p_t , le prix de la gazoline. Nous savons que de 1939 à 1945, il y eut une guerre mondiale qui s'est traduite par une diminution marquée des dépenses de consommation, entre autres de la gazoline. Durant cette période, les paramètres de la fonction des dépenses en gazoline se sont modifiés. Une façon de capter ce changement est d'introduire dans le modèle une variable auxiliaire. Les variables de cette catégorie sont du type dichotomique ou binaire qui, dans notre cas, prend la forme suivante :

$$D = \begin{cases} 1 & \text{si la caractéristique est présente} \\ 0 & \text{sinon} \end{cases}$$

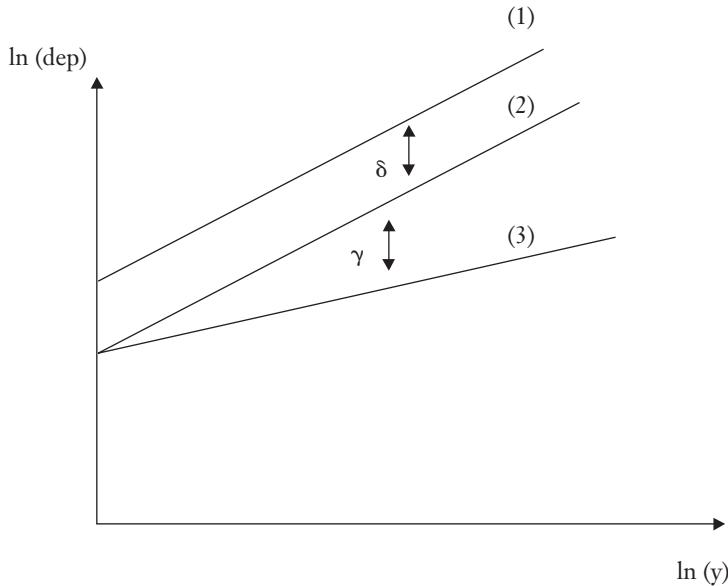
Dans le cadre de notre modèle, $D = 1$ si $t = 1939, \dots, 1945$ et $D = 0$ ailleurs. Cette variable s'introduit comme suit dans notre modèle de régression :

$$\ln \text{dep}_t = \beta_{1t} + \beta_{2t} \ln y_t + \beta_{3t} \ln p_t + e_t \quad t = 1920, \dots, 1970$$

11. Qui sont désignées par *dummies* en anglais.

Ce modèle est dit à coefficients variables où $\beta_{1t} = \beta_1 + \delta D_t$ et $\beta_{2t} = \beta_2 + \gamma D_t$. Pour simplifier, nous supposons que : $\beta_{3t} = \beta_3$. Ce modèle est représenté à la figure 4.4 en fixant p au niveau constant \bar{p} .

FIGURE 4.4



- (1): $E(\ln dep) = b_1 + b_2 \ln y + b_3 \ln \bar{p}$
 (2): $E(\ln dep) = b_1 + d + b_2 \ln y + b_3 \ln \bar{p}$
 (3): $E(\ln dep) = b_1 + d + (b_2 + g) \ln y + b_3 \ln \bar{p}$

On peut tester les hypothèses suivantes : i) $H_0 : \gamma = 0$ contre $H_1 : \gamma \neq 0$; ii) $H_0 : \delta = 0$ contre $H_1 : \delta \neq 0$;

iii) $H_0: \begin{bmatrix} \delta \\ \gamma \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ contre $H_1: \begin{bmatrix} \delta \\ \gamma \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Le premier test concerne le

changement de la pente de la droite de régression. Ce test peut être effectué à l'aide d'un test t standard. Il peut être formulé comme suit :

$t = \frac{\hat{\gamma} - 0}{\sqrt{\hat{V}(\hat{\gamma})}} \sim t(T - 5)$, en supposant certes que les résidus sont normalement distribués.

Le second test concerne le déplacement parallèle de la droite de régression. C'est là le test classique d'un changement structurel. Si l'on rejette H_0 , il y a alors ici diminution des dépenses en période de guerre. À l'instar du premier, ce test peut aussi s'effectuer par le test t . Finalement, le troisième test est le F de Fisher. Cette statistique se construit comme à l'accoutumée, en faisant appel aux résidus contraints et non contraints. Ce test sert à vérifier les hypothèses conjointes d'un déplacement parallèle et d'un changement de pente.

Le cas suivant sert également à utiliser l'emploi des variables instrumentales. Il consiste à estimer l'impact de certains jours de la semaine sur le cours des actions. Comme on le sait, les cours des actions a tendance à baisser systématiquement le lundi. Pour tester cet effet, on recourt à l'équation suivante associée à Connolly (1989)¹² :

$$r_t = \beta_0 + \beta_1 m_t + \beta_2 T_t + \beta_3 th_t + \beta_4 f_t + e_t$$

où r_t désigne le rendement journalier de l'indice Standard & Poor's 500 pour les années 1972 et 1973 ; m_t , T_t , th_t , et f_t sont des variables auxiliaires qui identifient respectivement les jours suivants : lundi, mardi, jeudi et vendredi ; et e_t est le terme d'erreur. Pour générer les variables auxiliaires, nous avons effectué en recourant au logiciel EViews le programme apparaissant au tableau 4.2.

12. Connolly, R.A. (1989), «An Examination of the Robustness of the Weekend Effect», *Journal of Financial and Quantitative Analysis*, juin, vol. 24, n° 2.

TABLEAU 4.2

```

smp1 1 503
genr m72=0
genr t72=0
genr w72=0
genr th72=0
genr f72=0

' Debut de la boucle for
for !i=0 to 502
if day72(!i+1)=1 then
genr m72(!i)=1
else
genr m72(!i)=0
endif
if day72(!i+1)=1 then
genr t72(!i)=1
else
genr t72(!i)=0
endif
if day72(!i+1)=1 then
genr w72(!i)=1
else
genr w72(!i)=0
endif
if day72(!i+1)=1 then
genr th72(!i)=1
else
genr th72(!i)=0
endif
if day72(!i+1)=1 then
genr f72(!i)=1
else
genr f72(!i)=0
endif
next

```

En appliquant les MCO à l'équation précédente, on obtient les résultats apparaissant au tableau 4.3.

TABLEAU 4.3

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.043659	0.080335	0.543455	0.5871
M72	-0.319821	0.117947	-2.711565	0.0069
T72	0.028880	0.114167	0.252962	0.8004
TH72	0.015556	0.113887	0.136589	0.8914
F72	-0.017636	0.113887	-0.154857	0.8770
R-squared	0.023408	Mean dependent var	-0.008239	
Ajusted R-squared	0.015548	S.D. dependent var	0.825708	
S.E. of regression	0.819264	Akaike info criterion	-0.388788	
Sum squared resid	333.5829	Schwarz criterion	-0.346770	
Log likelihood	-609.7212	F-statistic	2.978156	
Durbin-Watson stat	1.562695	Prob(F-statistic)	0.018937	

On observe au tableau 4.3 que seule la variable m_t , dont le coefficient estimé est de $-0,3198$, est significative avec une p-value égale à $0,0069$. On vérifie donc que, parmi les jours de la semaine, seul le lundi présente une anomalie. Les rendements boursiers n'auraient donc tendance à diminuer que le lundi. Cependant, il semble que cette anomalie ait eu tendance à se résorber par la suite.

Le test de Chow

Ce test, issu de Chow¹³, vise à vérifier la stabilité des paramètres d'une régression. Supposons le modèle suivant :

$$\mathbf{y}_t = \beta^T \mathbf{x}_t + \mathbf{e}_t \quad t = 1, \dots, T$$

où $\mathbf{x}_t = [x_{1t} \quad x_{2t} \quad \dots \quad x_{kt}]$. Le test de Chow consiste, dans le cadre de ce modèle temporel, à séparer l'échantillon en deux :

13. Chow, G.C. (1960), «Test of Equality between Sets of Coefficients in Two Linear Regressions», *Econometrica*, 52, p. 211-222.

1) la première partition s'étire du début de l'échantillon jusqu'à la date du changement structurel ; 2) la deuxième partition a trait à l'échantillon observé après la date du changement structurel. On effectue une régression sur chacune de ces deux sous-périodes et sur l'ensemble de la période, soit la somme des deux sous-périodes. Le test F se formule alors comme suit :

$$F = \frac{(SCR_T - (SCR_1 + SCR_2)) / k}{(SCR_1 + SCR_2) / T - 2k}$$

où SCR_T est la somme des erreurs résiduelles sur l'ensemble de l'échantillon ; SCR_1 , la somme des erreurs résiduelles sur la première sous-période et SCR_2 , la somme des erreurs résiduelles sur la deuxième sous-période.

On retrouve un test de Chow appliqué à un exemple financier dans Mills (1999)¹⁴. Il vise à vérifier la stabilité des paramètres de l'équation du CAPM en faisant appel à des statistiques corporatives. Un extrait de son étude apparaît au tableau 4.4, où NONLIN signifie RESET ; NORM : Jarque-Bera ; et HET : test d'hétéroscédasticité. Prenons le cas de Citicorp. La régression suivante de l'équation de la SML est estimée pour cette compagnie sur des données mensuelles de janvier 1978 à décembre 1987.

$$r_{it} - r_{ft} = \alpha + \beta[r_{mt} - r_{ft}] + e_t$$

où r_{it} désigne le rendement de la compagnie i , ici Citicorp ; r_{ft} est le taux sans risque ; r_{mt} , le taux de rendement du marché et e_t est le terme d'erreur. On veut tester la stabilité conjointe des coefficients α et β . Selon le CAPM, α doit être égal à 0 et β doit demeurer constant dans le temps. Pour effectuer le test de Chow, on fixe le point de rupture à décembre 1984. On divise donc l'échantillon en deux parties, soit de janvier 1978 à décembre 1984 et de janvier 1985 à décembre 1987. On effectue des régressions sur l'ensemble de l'échantillon et sur ses deux partitions et on calcule la somme des carrés résiduels pour ces trois régressions. On obtient :

14. Mills, T.C. (1999), *The Econometric Modelling of Financial Time Series*, deuxième édition, Cambridge University Press, Cambridge.

$$F = \frac{(\text{SCR}_T - (\text{SCR}_1 + \text{SCR}_2)) / k}{(\text{SCR}_1 + \text{SCR}_2) / T - 2k} = 6,44$$

Ce test est significatif au niveau de 5 %. On rejette l'hypothèse nulle de la stabilité du β d'un sous-échantillon à l'autre. L'auteur effectue également d'autres tests sur cette équation, dont le test RESET que nous avons déjà présenté. Ce test vise ici à tester la forme fonctionnelle de l'équation de la SML. Sous H_0 : le modèle est bien spécifié. Et sous H_1 : le modèle est mal spécifié. La valeur du test RESET¹⁵ est ici de 0,40, ce qui donne à penser que le modèle est bien spécifié. On retrouve ce test au tableau 4.4.

Le test de Chow présente le désavantage suivant. Dans certains sous-échantillons, il est possible qu'il y ait une pénurie d'observations puisque le point de rupture peut séparer l'échantillon total en deux partitions très inégales. Le test de Chow (*predictive*) corrigé par Fisher en 1970 vient pallier ce désavantage. Ce test s'écrit comme suit :

$$F = \frac{\hat{\mathbf{e}}_*^T \hat{\mathbf{e}}_* - \hat{\mathbf{e}}^T \hat{\mathbf{e}} / n_1}{\hat{\mathbf{e}}^T \hat{\mathbf{e}} / n_2 - k} \sim F(n_1, n_2 - k)$$

où $\hat{\mathbf{e}}_*^T \hat{\mathbf{e}}_*$ est la somme des carrés résiduels (contraints)¹⁶ sur l'ensemble de l'échantillon (SCR_T) et $\hat{\mathbf{e}}^T \hat{\mathbf{e}}$, la somme des carrés résiduels (non contraints) définie sur le sous-échantillon qui comporte le plus grand nombre d'observations.

15. Pour effectuer le test RESET, on a régressé les résidus de l'équation de la SML sur $(r_m - r_f)$ et $(r_i - r_f)^2$ estimés.

16. La contrainte étant que les paramètres sont identiques à travers l'ensemble de l'échantillon.

TABLEAU 4.4

Table 6.1. Estimates of the CAPM regression (6.13)

Company	$\hat{\alpha}$	$\hat{\beta}$	R ²	dw	NONLIN	NORM	HET	ARCH	CHOW
BOISE	0.0031 (0.0068) [0.0053]	0.94 (0.10) [0.13]	0.43	2.17	2.95	4.72	9.57*	8.69*	2.69
CITCRP	0.0025 (0.0062) [0.0058]	0.67 (0.09) [0.14]	0.32	1.84	0.40	1.20	10.33*	2.50	6.44*
CONED	0.0110 (0.0046) [0.0036]	0.09 (0.07) [0.06]	0.02	2.15	0.74	1.12	0.20	5.04	0.05
CONTIL	-0.0132 (0.0131) [0.0128]	0.73 (0.19) [0.24]	0.11	2.07	0.45	2245*	0.37	0.06	2.56
DATGEN	-0.0067 (0.0098) [0.0094]	1.03 (0.14) [0.19]	0.31	2.08	7.20*	5.03	3.21	0.15	1.89
DEC	0.0068 (0.0074) [0.0066]	0.85 (0.11) [0.13]	0.34	2.14	0.72	9.23*	0.66	16.03*	5.67
DELTA	0.0014 (0.0083) [0.0082]	0.49 (0.12) [0.15]	0.12	1.99	0.01	2.55	0.43	2.46	3.67
GENMIL	0.0078 (0.0058) [0.0050]	0.27 (0.08) [0.10]	0.08	2.08	0.14	2.64	0.94	2.16	14.90*
GERBER	0.0051 (0.0071) [0.0065]	0.63 (0.10) [0.11]	0.24	2.25	8.38*	7.14*	0.27	1.72	6.62*
IBM	-0.0005 (0.0046) [0.0054]	0.46 (0.07) [0.07]	0.28	1.88	0.06	1.14	0.17	3.06	6.68*

MOBIL	0.0042 (0.0059) [0.0051]	0.72 (0.09) [0.09]	0.37	2.09	0.55	34.6*	6.93*	1.68	0.29
MOTOR	0.0069 (0.0083) [0.0077]	0.10 (0.12) [0.10]	0.01	1.86	0.90	2.23	0.97	0.73	2.05
PANAM	-0.0086 (0.0112) [0.0103]	0.74 (0.16) [0.15]	0.15	2.21	0.51	10.9*	2.52	4.46	0.14
PSNH	-0.0126 (0.0100) [0.0105]	0.21 (0.15) [0.10]	0.02	1.88	0.25	92.5*	1.64	10.46*	0.03
TANDY	0.0107 (0.0097) [0.0100]	1.05 (0.14) [0.14]	0.32	1.89	3.27	6.13*	2.76	0.66	0.13
TEXACO	0.0007 (0.0062) [0.0049]	0.61 (0.09) [0.10]	0.28	2.02	0.00	127.5*	2.59	3.10	0.14
WEYER	-0.0031 (0.0059) [0.0046]	0.82 (0.09) [0.10]	0.43	2.29*	1.76	1.44	15.07*	9.97*	9.88*
Asymptotic distribution					χ^2_1	χ^2_2	χ^2_2	χ^2_3	χ^2_2
Critical 0.05 value					3.84	5.99	5.99	7.81	5.99

Notes:

* : Significant at 0.05 level.

(. . .) : Conventional standard error ; [...] : Newey-West (1987) standard error from (6.3).

dw : Durbin-Watson statistic.

NONLIN : Ramsey's (1996) RESET test for functional form, calculated from the regression of \hat{u}_t on x_t and \hat{y}_t^2 .

NORM : Jarque-Bera (1980) test for normality.

HET : Test for heteroskedasticity, calculated from the regression of \hat{u}_t^2 on a constant. \hat{y}_t and \hat{y}_t^2 .

CHOW : Chow's (1960) test for coefficient stability: break point taken to be December 1984.

Source : Mills, T.C. (1999), *The Econometric Modelling of Financial Time Series*, 2^e edition, Cambridge University Press, Cambridge.

(Reproduit avec la permission de l'éditeur)

VARIATIONS SUR LES MODÈLES LINÉAIRE ET NON LINÉAIRE

PARTIE II

Dans ce chapitre, nous introduisons d'abord la théorie asymptotique et les tests s'y rapportant : LM, LR et Wald. Puis nous examinons les problèmes causés par la multicollinéarité au chapitre des variables explicatives. Finalement, nous nous penchons sur le phénomène de l'endogénéité des variables explicatives.

1. THÉORIE ASYMPTOTIQUE : CONVERGENCE, TESTS ASYMPTOTIQUES ET VARIABLES INSTRUMENTALES

1.1 Convergence

La théorie asymptotique est maintenant à ce point avancée que la plupart des propriétés linéaires et non linéaires des estimateurs et, partant, des tests, font appel à cette théorie. Les propriétés des estimateurs reliés aux petits échantillons sont beaucoup moins connues même si, ces dernières décennies, il s'est produit une grande amélioration au chapitre des techniques de travail empirique en matière d'évaluation de ces propriétés¹. Il existe quatre modes de convergence.

-
1. On pense ici par exemple à l'utilisation beaucoup plus poussée de certaines techniques de simulation de type Monte Carlo et *bootstrapping* et des méthodes de réduction de la variance.

- i) *La convergence en probabilité.* C'est le mode de convergence le plus connu. Elle se définit comme suit. Supposons un estimateur $\hat{\theta}_T$ de θ . On dit que l'estimateur $\hat{\theta}_T$ converge en probabilité vers θ , c'est-à-dire : $\hat{\theta}_T \xrightarrow{p} \theta$, si :

$$\lim_{T \rightarrow \infty} \Pr \left[\left| \hat{\theta}_T - \theta \right| < \varepsilon \right] = 1$$

au fur et à mesure que la taille de l'échantillon devient importante. Une notation plus compacte de ce mode de convergence fréquemment utilisée dans la littérature est la suivante :

$$p \lim \hat{\theta}_T = \theta$$

Les conditions suffisantes pour vérifier la convergence asymptotique en probabilité sont les suivantes, en termes de l'espérance et la variance de $\hat{\theta}_T$:

$$\lim_{T \rightarrow \infty} E \left(\hat{\theta}_T \right) = \theta \quad \text{et} \quad \lim_{T \rightarrow \infty} V \left(\hat{\theta}_T \right) = 0.$$

Sous certaines conditions très générales, on peut vérifier la convergence asymptotique en probabilité seulement en calculant l'espérance asymptotique. Pour un traitement rigoureux de ce dernier sujet, on consultera White (1984)². Et pour une application de ce concept, on se reportera à : Racicot (2000)³.

- ii) *La convergence en moyenne quadratique.* On dit qu'un estimateur $\hat{\theta}_T$ converge en moyenne quadratique vers θ , i.e. $\hat{\theta}_T \xrightarrow{m.s.} \theta$, si :

$$\lim_{T \rightarrow \infty} E \left[\left| \hat{\theta}_T - \theta \right|^r \right] = 0$$

Ce mode de convergence est plus fort que celui de la convergence en probabilité. En effet, si les conditions antérieures ayant trait à l'espérance et à la variance asymptotique sont

-
2. White, H. (1984), *Asymptotic Theory for Econometricians*, Academic Press, New York.
 3. Racicot, F.-É. (2000), Estimation et tests en présence d'erreurs de mesure sur les variables explicatives : vérification empirique par la méthode de simulation Monte Carlo, Centre de recherche en gestion, série Finance empirique et quantitative, document de travail CRG-09-2000.

respectées, alors $\hat{\theta}_T \xrightarrow{\text{m.s.}} \theta$. Il s'ensuit que : $p \lim \hat{\theta}_T = \theta$, soit en notation courante : $\hat{\theta}_T \xrightarrow{p} \theta$;

iii) *la convergence quasi sûre (almost sure, abrégé pas a.s.)*. On a ici la forme de convergence la plus forte. On dit qu'un estimateur $\hat{\theta}$ converge de façon quasi sûre vers θ , i.e. $\hat{\theta}_T \xrightarrow{\text{a.s.}} \theta$, si :

$$\Pr \left[\lim_{T \rightarrow \infty} (\hat{\theta}_T = \theta) \right] = 1$$

iv) *la convergence en distribution*

Supposons deux séries : $\hat{\theta}_T$ et θ . Une série $\hat{\theta}_T$ converge en distribution vers une série θ si la fonction de distribution f_T de $\hat{\theta}_T$ converge vers la distribution f de θ pour chaque point de f . Par exemple, supposons un estimateur $\hat{\theta}_T$ de θ . La notation utilisée pour montrer que $(\hat{\theta}_T - \theta)$ converge par exemple en distribution vers une normale d'espérance nulle et de variance σ^2 est la suivante : $(\hat{\theta}_T - \theta) \xrightarrow{d} N(0, \sigma^2)$.

En résumé, établissons la structure hiérarchique des modes de convergence. Elle s'établit comme suit⁴ :

$$\begin{aligned} \text{m.s.} &\Rightarrow p \Rightarrow d \\ \text{et a.s.} &\Rightarrow p \end{aligned}$$

Exceptionnellement, on peut observer :

$$d \Rightarrow p$$

Concluons cette section en précisant qu'un estimateur sans biais n'est pas nécessairement convergent. En effet, l'une des conditions pour la convergence en probabilité est que la variance asymptotique soit nulle. Supposons que l'on ait le modèle de régression suivant :

$$y = c + e$$

4. Voir à cet effet : Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press, Cambridge.

où e est une variable aléatoire de moyenne nulle et de variance constante et la corrélation entre ses valeurs est constante et égale à ρ , ce qui peut se représenter par :

$$\Omega = \begin{bmatrix} 1 & \rho & \rho & \cdot & \cdot & \cdot & \rho \\ \rho & 1 & \rho & \cdot & \cdot & \cdot & \rho \\ \rho & \rho & 1 & \cdot & \cdot & \cdot & \rho \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho & \rho & \rho & \cdot & \cdot & \cdot & 1 \end{bmatrix}$$

La matrice X est ici égale à un vecteur colonne dont toutes les composantes sont égales à 1, ce qui implique que l'estimateur des MCO de c est \bar{y} . En insérant Ω dans la formule de la matrice variance-covariance de $\hat{\beta}$, ici c , estimé à partir du modèle de régression généralisé, on a :

$$V(\hat{\beta}) = \frac{\sigma^2}{T} \left(\frac{1}{T} (\mathbf{X}^T \mathbf{X}) \right)^{-1} \left(\frac{1}{T} (\mathbf{X}^T \Omega \mathbf{X}) \right) \left(\frac{1}{T} (\mathbf{X}^T \mathbf{X}) \right)^{-1}$$

$$V(\bar{y}) = \frac{\sigma^2}{T} (1 - \rho + T\rho)$$

Si l'on prend la limite de cette expression, on obtient : $\rho\sigma^2$. Dans cet exemple, l'application des MCO en présence d'autocorrélation fournit un estimateur sans biais mais non convergent puisque la variance asymptotique n'est pas nulle. L'autocorrélation est en effet fort prononcée dans cet exemple. Dans les modèles de séries temporelles, on

exige que la corrélation entre les observations diminue au fur et à mesure que la distance entre celles-ci augmente. Ici, certe, cette condition est violée.

1.2. Tests asymptotiques : LR, LM et Wald

1.2.1. Test du ratio de vraisemblance : LR

Le terme LR est l'abréviation anglaise de *likelihood ratio*. On veut tester l'hypothèse $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ contre $H_1 : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}$. \mathbf{R} est une matrice de dimension $(q \times k)$, où $q < k$, et \mathbf{r} est un vecteur de dimension $(q \times 1)$. De manière à effectuer ce test, on doit obtenir la valeur estimée de la fonction de vraisemblance non contrainte : $L(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\sigma}}^2)$ et sa valeur contrainte : $L(\hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\sigma}}_R^2)$. Le ratio de vraisemblance se définit comme suit :

$$\lambda = \frac{L(\hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\sigma}}_R^2)}{L(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\sigma}}^2)}$$

Intuitivement, on s'attend à rejeter l'hypothèse nulle si λ est faible. Le test asymptotique général se formule comme suit :

$$LR = -2 \ln \lambda = 2 \left[\ln L(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\sigma}}^2) - \ln L(\hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\sigma}}_R^2) \right]^a \sim \chi^2(q)$$

où $L(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\sigma}}^2) = (2\pi e)^{-\frac{T}{2}} (\hat{\boldsymbol{\sigma}}^2)^{-\frac{T}{2}} = \left(\frac{(2\pi e)}{T} \right)^{-\frac{T}{2}} (\hat{\mathbf{e}}^T \hat{\mathbf{e}})^{-\frac{T}{2}} = \text{cste} \times (\hat{\mathbf{e}}^T \hat{\mathbf{e}})^{-\frac{T}{2}}$.

Le maximum de vraisemblance contraint est obtenu en maximisant la fonction suivante :

$$l^* = \ell - \mathbf{u}^T (\mathbf{R}\boldsymbol{\beta} - \mathbf{r})$$

où \mathbf{u} est un vecteur de dimension $(q \times 1)$ de multiplicateurs de Lagrange et ℓ , le logarithme de la fonction de vraisemblance, c'est-à-dire :

$$\begin{aligned}\ell &= \ln f(y_1, y_2, \dots, y_n | \mathbf{X}) = \ln f(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) \\ &= -\frac{T}{2} \ln 2\pi - \frac{T}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\end{aligned}$$

On peut montrer que $\hat{\boldsymbol{\beta}}_R$ est simplement le vecteur contraint obtenu des MCO satisfaisant l'équation : $\mathbf{R}\hat{\boldsymbol{\beta}}_R = \mathbf{r}$. Les résidus qui en résultent sont les suivants : $\hat{\mathbf{e}}_R = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_R$. Le MV contraint de σ^2 est : $\hat{\boldsymbol{\sigma}}_R^2 = \hat{\mathbf{e}}_R^T \hat{\mathbf{e}}_R / T$. Par conséquent, la fonction de vraisemblance est :

$$L(\hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\sigma}}_R^2) = \text{cste} \times \left(\hat{\mathbf{e}}_R^T \hat{\mathbf{e}}_R \right)^{-\frac{T}{2}}$$

En substituant $L(\hat{\boldsymbol{\beta}}_R, \hat{\boldsymbol{\sigma}}_R^2)$ et $L(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\sigma}}^2)$ dans LR, on obtient :

$$\begin{aligned}\text{LR} &= T \left(\ln \hat{\mathbf{e}}_R^T \hat{\mathbf{e}}_R - \ln \hat{\mathbf{e}}^T \hat{\mathbf{e}} \right) = T \ln \left(1 + \frac{\hat{\mathbf{e}}_R^T \hat{\mathbf{e}}_R - \hat{\mathbf{e}}^T \hat{\mathbf{e}}}{\hat{\mathbf{e}}^T \hat{\mathbf{e}}} \right) \\ &= T \ln \left(\frac{1}{1 - \frac{\hat{\mathbf{e}}_R^T \hat{\mathbf{e}}_R - \hat{\mathbf{e}}^T \hat{\mathbf{e}}}{\hat{\mathbf{e}}_R^T \hat{\mathbf{e}}_R}} \right)\end{aligned}$$

Cette équation exige donc d'estimer les modèles contraint et non contraint. On peut également considérer le test LR comme une transformation du test F :

$$\text{LR} = T \ln \left[1 + \frac{1}{T-k} \times qF \right] \cong \frac{T}{T-k} qF$$

1.2.2. Test de Wald

Dans le test de Wald, seul le modèle non contraint est estimé alors que pour effectuer le test LR, il fallait estimer les modèles contraint et non contraint. Le test de Wald s'écrit comme suit :

$$W = \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})^T \left[\mathbf{R}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \right]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})}{\hat{\boldsymbol{\sigma}}^2} = \frac{\mathbf{T}(\hat{\mathbf{e}}_R^T \hat{\mathbf{e}}_R - \hat{\mathbf{e}}^T \hat{\mathbf{e}})_a}{\hat{\mathbf{e}}^T \hat{\mathbf{e}}} \sim \chi^2(q)$$

Dans la première égalité de cette expression, le vecteur $(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})$ nous indique si l'estimateur du maximum de vraisemblance non contraint est rapproché ou éloigné de \mathbf{r} , associé à l'hypothèse nulle. Si $(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})$ est voisin de $\mathbf{0}$, cela implique que l'hypothèse nulle ne sera pas rejetée. À l'opposé, si $(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})$ est élevé, cela milite en faveur du rejet de l'hypothèse nulle.

Par ailleurs, on remarque dans la deuxième égalité que les erreurs contraintes interviennent dans le test de Wald. Cela est attribuable à la relation suivante : $\hat{\mathbf{e}}_R^T \hat{\mathbf{e}}_R - \hat{\mathbf{e}}^T \hat{\mathbf{e}} = (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})^T \left[\mathbf{R}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R} \right]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})$. Par conséquent, le test sur les contraintes peut être effectué en faisant la régression non contrainte et en substituant la valeur de $\hat{\boldsymbol{\beta}}$ dans la première égalité. Alternativement, une régression contrainte peut être estimée et le test portera sur la différence $\hat{\mathbf{e}}_R^T \hat{\mathbf{e}}_R - \hat{\mathbf{e}}^T \hat{\mathbf{e}}$. $\hat{\boldsymbol{\beta}}$ est asymptotiquement distribué comme une $N(\boldsymbol{\beta}, \mathbf{I}^{-1}(\boldsymbol{\beta}))$, où \mathbf{I} est la matrice d'information. Sous H_0 , $\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}$ est asymptotiquement distribué comme une loi normale multivariée de moyenne $\mathbf{0}$ et une matrice variance-covariance $\mathbf{R}\mathbf{I}^{-1}(\boldsymbol{\beta})\mathbf{R}^T$, où $\mathbf{I}^{-1}(\boldsymbol{\beta}) = \boldsymbol{\sigma}^2(\mathbf{X}^T \mathbf{X})^{-1}$. Soit, en notation standard : $(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})^a \sim N(0, \mathbf{R}\mathbf{I}^{-1}(\boldsymbol{\beta})\mathbf{R}^T)$. Élaborons davantage sur la matrice dite d'information. Cette matrice peut se calculer comme suit :

$$\mathbf{I}(\theta) = \mathbf{I} \begin{pmatrix} \beta \\ \sigma^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2} (\mathbf{X}^T \mathbf{X}) & 0 \\ 0 & \frac{T}{2\sigma^4} \end{pmatrix}$$

Puisque cette matrice est diagonale, l'inverse de cette matrice est donc de :

$$\mathbf{I}^{-1}(\theta) = \mathbf{I}^{-1} \begin{pmatrix} \beta \\ \sigma^2 \end{pmatrix} = \begin{pmatrix} \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} & 0 \\ 0 & \frac{2\sigma^4}{T} \end{pmatrix}$$

Étant donné que cette matrice est diagonale, on peut se concentrer sur la partition reliée à β et écrire que :

$(\mathbf{R}\hat{\beta} - \mathbf{r})^T [\mathbf{R}\mathbf{I}^{-1}(\beta)\mathbf{R}^T]^{-1} (\mathbf{R}\hat{\beta} - \mathbf{r}) \sim \chi^2(q)$. La distribution asymptotique vaut encore quand σ^2 est remplacé par sa valeur estimée : $\hat{\sigma}^2 = \frac{\hat{\mathbf{e}}^T \hat{\mathbf{e}}}{T}$. Il en résulte la statistique W présentée antérieurement :

$$W = \frac{(\mathbf{R}\hat{\beta} - \mathbf{r})^T [\mathbf{R}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T]^{-1} (\mathbf{R}\hat{\beta} - \mathbf{r})}{\hat{\sigma}^2}$$

1.2.3. Test LM⁵

Le test LM, aussi connu sous le nom de *test score*, se formule de la façon suivante :

$$LM = \mathbf{S}^T (\hat{\theta}_R^2) \mathbf{I}^{-1} (\hat{\theta}_R^2) \mathbf{S} (\hat{\theta}_R^2) \sim \chi^2(q)$$

où $\hat{\theta}_R^2$ représente l'estimateur contraint et $\mathbf{S}(\theta)$, le vecteur score. À l'inverse du test de Wald, on n'a ici qu'à calculer l'estimateur contraint. Le vecteur score :

5. LM est l'abréviation anglaise de l'expression *Lagrange Multiplier Test*.

$$\mathbf{S}(\theta) = \begin{bmatrix} \frac{\partial \ell}{\partial \beta} \\ \frac{\partial \ell}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{e} \\ -\frac{T}{\sigma^2} + \frac{\mathbf{e}^T \mathbf{e}}{2\sigma^4} \end{bmatrix}$$

évalué à $\hat{\theta}_R^2$, est obtenu en remplaçant \mathbf{e} par $\hat{\mathbf{e}}_R = \mathbf{y} - \mathbf{X}\hat{\beta}_R$ et σ^2 par

$$\hat{\sigma}_R^2 = \frac{\hat{\mathbf{e}}_R^T \hat{\mathbf{e}}_R}{T}.$$

$$\mathbf{S}(\hat{\theta}_R) = \begin{bmatrix} \frac{1}{\hat{\sigma}_R^2} \mathbf{X}^T \hat{\mathbf{e}}_R \\ 0 \end{bmatrix}$$

Expliquons le calcul du deuxième élément de ce vecteur. En effet :

$$-\frac{T}{2\hat{\sigma}_R^2} + \frac{\hat{\mathbf{e}}_R^T \hat{\mathbf{e}}_R}{2\left(\frac{\hat{\mathbf{e}}_R^T \hat{\mathbf{e}}_R}{T}\right)^2} = -\frac{T}{2\hat{\sigma}_R^2} + \frac{T}{2\hat{\sigma}_R^2} = 0$$

En substituant ces calculs dans l'expression de LM, on obtient :

$$\begin{aligned} \text{LM} &= \begin{bmatrix} \frac{\hat{\mathbf{e}}_R^T \mathbf{X}}{\hat{\sigma}_R^2} & 0 \end{bmatrix} \begin{bmatrix} \hat{\sigma}_R^2 (\mathbf{X}^T \mathbf{X})^{-1} & 0 \\ 0 & \frac{2\hat{\sigma}_R^4}{T} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{X}^T \hat{\mathbf{e}}_R \\ 0 \end{bmatrix} = \frac{\hat{\mathbf{e}}_R^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{e}}_R}{\hat{\sigma}_R^2} \\ &= \frac{T \mathbf{e}_R^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{e}}_R}{\hat{\mathbf{e}}_R^T \hat{\mathbf{e}}_R} = TR^2 \end{aligned}$$

6. Fournissons une interprétation intuitive de ce résultat qui a déjà été démontré. La matrice $\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ effectue une projection dans l'espace des résidus expliqués. Le numérateur est donc la somme des carrés expliquée. En divisant cette somme par la somme des carrés résiduelle, on obtient R^2 .

En termes de résidus, la représentation du test LM est de :

$$LM = \frac{T(\hat{\mathbf{e}}_R^T \hat{\mathbf{e}}_R - \hat{\mathbf{e}}^T \hat{\mathbf{e}})}{\hat{\mathbf{e}}_R^T \hat{\mathbf{e}}_R}$$

Enfin, on peut montrer que les tests asymptotiques de Wald, du ratio de vraisemblance et du multiplicateur de Lagrange sont reliés comme suit : $W \geq LR \geq LM$. Ces tests sont asymptotiquement équivalents, mais en général, ils diffèrent dans les petits échantillons.

1.2.4. Variables instrumentales et estimation

Dans cette section, on s'intéresse au cas où : $\text{plim} \frac{1}{T} \mathbf{X}^T \mathbf{e} \neq 0$. On pense ici à une situation où les \mathbf{X} ne sont pas orthogonaux aux \mathbf{e} , c'est-à-dire dans le cas où, par exemple, les \mathbf{X} sont stochastiques et sont corrélés avec les \mathbf{e} . Pour traiter ce problème, on recourt aux variables dites instrumentales. Supposons le modèle de régression linéaire suivant :

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$$

où l'on suppose que $E(\mathbf{X}^T \mathbf{e}) \neq 0$. Une matrice de variables instrumentales \mathbf{Z} se définit par les deux critères suivants : i) les variables incluses dans \mathbf{Z} sont à la limite non corrélées avec les résidus \mathbf{e} , c'est-à-dire :

$\text{plim} \frac{1}{T} \mathbf{Z}^T \mathbf{e} = 0$; ii) les variables incluses dans \mathbf{Z} sont corrélées avec

\mathbf{X} et, à la limite : $\text{plim} \frac{1}{T} \mathbf{Z}^T \mathbf{X} = \Sigma_{ZX}$, cette dernière matrice étant finie et de plein rang⁷. L'estimateur des variables instrumentales IV s'obtient par la procédure suivante : i) on régresse d'abord \mathbf{X} sur \mathbf{Z} , d'où on obtient : $\hat{\mathbf{X}} = \mathbf{Z}\hat{\beta} = \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{X}$; ii) on régresse par la suite \mathbf{Y} sur $\hat{\mathbf{X}}$. On obtient alors : $\hat{\beta} = (\hat{\mathbf{X}}^T \hat{\mathbf{X}})^{-1} \hat{\mathbf{X}}^T \mathbf{Y}$. On peut également utiliser la notation suivante : $\hat{\beta} = \hat{\beta}_{IV} = \hat{\beta}_{2SLS}$. En effet, l'esti-

7. C'est-à-dire qu'aucune ligne ou colonne n'est linéairement dépendante des autres.

mateur des variables instrumentales est ici égal à l'estimateur des doubles moindres carrés (2SLS : *two-stage least squares*), l'estimateur des doubles moindres carrés s'obtenant en effet en appliquant les deux étapes précédentes. On élaborera davantage sur ces sujets dans le chapitre 11 consacré à la méthode des moments généralisés⁸.

2. PROBLÈMES AU CHAPITRE DES VARIABLES EXPLICATIVES : MULTICOLLINÉARITÉ ET ENDOGÉNÉITÉ DES VARIABLES EXPLICATIVES

Multicollinéarité

Dans le modèle de régression linéaire classique : $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, la multicollinéarité parfaite signifie qu'il existe une relation linéaire entre les variables explicatives. Dans la pratique, la multicollinéarité n'est pas parfaite. Elle présente cependant les conséquences suivantes : i) la présence de collinéarité au chapitre des variables de la matrice \mathbf{X} , que l'on désigne par multicollinéarité, se traduit par une matrice $(\mathbf{X}^T\mathbf{X})$, singulière lorsque cette multicollinéarité est parfaite, c'est-à-dire $|\mathbf{X}^T\mathbf{X}| = 0$. Il en résulte que $(\mathbf{X}^T\mathbf{X})^{-1}$ n'est pas définie. Dans le cas où la multicollinéarité est très élevée sans être parfaite, $(\mathbf{X}^T\mathbf{X})^{-1}$ sera important, ce qui se traduira par une matrice de variances des MCO très élevées ; ii) la deuxième conséquence découle de la première. En effet, si les écarts-types des estimateurs $\boldsymbol{\beta}$ sont élevés, les statistiques t seront faibles même si les R^2 ou les F indiquent que l'ensemble des variables explicatives sont significatives ; iii) les estimateurs pourront être très sensibles à l'ajout ou au retrait de variables explicatives apparemment sans signification ; iv) on peut effectuer des prévisions suffisamment précises en présence de multicollinéarité si la collinéarité entre les variables demeure stable.

Pour identifier la multicollinéarité, il existe plusieurs méthodes. Nous en retenons deux. La plus simple consiste à calculer la matrice de corrélation de Pearson. Cette matrice donne une mesure approximative de la liaison linéaire entre les paires de variables explicatives.

8. On trouvera un exemple d'application des variables instrumentales dans : Racicot, F.-É. (2000), *op. cit.*, CRG 09-2000.

Le seuil à partir duquel la collinéarité peut être jugée importante s'établit au voisinage de 0,8. La deuxième méthode pour juger de l'importance de la multicollinéarité fait appel aux régressions auxiliaires. L'avantage de cette méthode sur la première est qu'elle permet d'exprimer une variable en fonction de plusieurs autres :

$x_k = a_1x_1 + a_2x_2 + \dots + a_{k-1}x_{k-1} + v$. Cette régression est dite auxiliaire car elle exprime une variable explicative à l'aide d'autres variables explicatives. Le terme v est une variable aléatoire appelé différence. Une collinéarité importante sera associée à une somme des carrés résiduelle faible ou à un R^2 élevé.

Quelles sont les solutions possibles au problème de multicollinéarité? Nous en retenons quatre: i) on peut d'abord ajouter des variables explicatives de bonne qualité, mais celles-ci ne sont pas toujours disponibles; ii) on peut exclure les variables qui font problème si tant est que cela soit possible. Cette procédure équivaut à imposer des contraintes sur les paramètres, mais nous avons vu antérieurement que si nous procédons de la sorte, les MCO sont biaisés dans le cas où les contraintes ne sont pas valables. Les MCO seront cependant sans biais dans le cas où les contraintes sont valables; iii) si $(\mathbf{X}^T\mathbf{X})$ est singulière et que l'on a besoin d'inverser cette matrice, on peut utiliser l'inverse généralisé de Moore-Penrose, aussi appelé pseudo-inverse. Posons: $\mathbf{Z} = (\mathbf{X}^T\mathbf{X})$, une matrice carrée, alors \mathbf{Z}^+ est l'inverse généralisé de Moore-Penrose. Cette matrice a la propriété d'être unique et se définit comme suit: $\mathbf{Z}^+ = \mathbf{C}_1\mathbf{\Lambda}_1^{-1}\mathbf{C}_1^T$, où $\mathbf{\Lambda}_1$ est la matrice diagonale qui renferme les valeurs propres (*eigenvalues*) de $(\mathbf{X}^T\mathbf{X})$ non nulles, c'est-à-dire :

$$\mathbf{\Lambda}_1 = \begin{bmatrix} \lambda_1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \lambda_2 & 0 & \cdot & \cdot & 0 \\ \cdot & & \cdot & & & \cdot \\ \cdot & & & \cdot & & \cdot \\ \cdot & & & & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & \lambda_k \end{bmatrix}$$

et \mathbf{C}_1 est la matrice regroupant les vecteurs caractéristiques de $(\mathbf{X}^T\mathbf{X})$.

Pour fixer les idées, considérons un cas du calcul des valeurs propres et des vecteurs propres. Soit \mathbf{A} et \mathbf{c} , un vecteur dont les éléments sont non nuls. On cherche à solutionner le système d'équations suivant : $(\mathbf{A} - \lambda\mathbf{I})\mathbf{c} = \mathbf{0}$. Pour obtenir une solution non triviale à ce système, il faut que : $|\mathbf{A} - \lambda\mathbf{I}| = 0$. On trouve les valeurs de λ qui satisfont à la valeur de ce déterminant. Une fois ces λ calculés, on trouve le vecteur \mathbf{c} à partir de la relation précédente : $(\mathbf{A} - \lambda\mathbf{I})\mathbf{c} = \mathbf{0}$. Pour illustrer ces calculs, considérons le cas d'une matrice \mathbf{A} de dimension (2×2) . L'équation : $(\mathbf{A} - \lambda\mathbf{I})\mathbf{c} = \mathbf{0}$ peut alors se représenter comme suit :

$$\begin{bmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Le déterminant $|\mathbf{A} - \lambda\mathbf{I}| = 0$ est ici de :

$(a_{11} - \lambda)(a_{22} - \lambda) - (a_{12}a_{21}) = 0$. On appelle cette équation : l'équation caractéristique. Le λ est obtenu en solutionnant ce polynôme du second degré, qui est de la forme : $ax^2 + bx + c = 0$. La solution est :

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

iv) la quatrième solution pour traiter le problème de multicollinéarité consiste à utiliser l'estimateur de la *ridge regression*. Cet estimateur, désigné par $\hat{\beta}_r$, est égal à : $\hat{\beta}_r = (\mathbf{X}^T\mathbf{X} + r\mathbf{D})^{-1}\mathbf{X}^T\mathbf{y}$ où r est approximativement égal à 0,01 et est obtenu par simulation et où \mathbf{D} est une matrice diagonale qui contient les éléments de la diagonale de $(\mathbf{X}^T\mathbf{X})$. L'espérance de cet estimateur est égal à :

$E(\hat{\beta}_r) = (\mathbf{X}^T\mathbf{X} + r\mathbf{D})^{-1}(\mathbf{X}^T\mathbf{X})\beta$, ce qui montre que l'estimateur de la *ridge regression* est biaisé. Par contre, sa variance est inférieure à celle des MCO, c'est-à-dire :

$$V(\hat{\beta}_r) = \sigma^2(\mathbf{X}^T\mathbf{X} + r\mathbf{D})^{-1}(\mathbf{X}^T\mathbf{X})(\mathbf{X}^T\mathbf{X} + r\mathbf{D})^{-1} < V(\hat{\beta}).$$

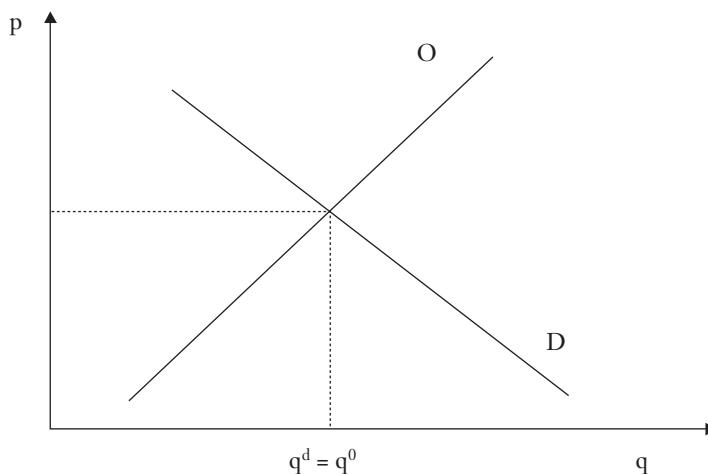
Ici, l'intuition est la suivante. Lorsque l'on a multicollinéarité quasi parfaite, le déterminant de $(\mathbf{X}^T\mathbf{X})$ est très rapproché de 0. Le stratagème est ici d'ajouter à la matrice $(\mathbf{X}^T\mathbf{X})$ des éléments de nature à rendre ce déterminant différent de 0, en l'occurrence rajouter $r\mathbf{D}$ à $(\mathbf{X}^T\mathbf{X})$. On obtiendra ce faisant un estimateur biaisé mais ayant une variance plus faible que les MCO, c'est-à-dire comportant un écart quadratique moyen plus faible que les MCO, i.e. :

$$\text{EQM}(\hat{\beta}_r) < \text{EQM}(\hat{\beta}).$$

Endogénéité des variables explicatives

Pour présenter ce problème, nous recourons au système d'équations de l'offre et de la demande. On veut ici estimer la demande. Or, l'estimateur des MCO des paramètres de la fonction de demande sera généralement biaisé et non convergent. Les équations d'offre et de demande sont représentées à la figure 5.1.

FIGURE 5.1



Le système d'équations d'offre et de demande s'exprime comme suit :

$$q_t^d = \beta_{11} + \beta_{12}p_t + e_{1t} \quad (\text{D})$$

$$q_t^o = \beta_{21} + \beta_{22}p_t + e_{2t} \quad (\text{O})$$

On suppose que les innovations sont des bruits blancs. Sur la figure 5.1, l'équilibre se situe au point de jonction de l'offre et de la demande. À ce point, on a :

$$q_t^d = q_t^o = q_t$$

En exprimant la demande sous sa forme inverse (appelée *demande inverse*) et après avoir remplacé q_t^d par sa valeur d'équilibre q_t , on obtient :

$$p_t = -\frac{\beta_{11}}{\beta_{12}} + \frac{1}{\beta_{12}}q_t - \frac{1}{\beta_{12}}e_{1t}$$

Par conséquent, une variation dans l'innovation e_{1t} se répercute sur le prix (p_t) et par ricochet, par l'intermédiaire de la fonction d'offre, sur q_t . Cela démontre que :

$$E(p_t, e_{1t}) \neq 0$$

Par conséquent, l'estimateur des MCO appliqué aux paramètres de l'équation de demande est biaisé et non convergent. La non-prise en compte de l'offre se traduit ici par ce qu'on appelle un biais de simultanéité puisque la détermination du prix et de la quantité correspondant à l'équilibre provient d'un système d'équations. Pour traiter ce problème, on recourt généralement aux doubles moindres carrés (2SLS) qui se traduisent par un estimateur convergent. En finance corporative, l'estimation de la demande et de l'offre d'obligations, qui donne lieu à la détermination du taux d'intérêt d'équilibre du marché des obligations, est confrontée au même problème. Il faut recourir à des variables instrumentales pour identifier ces deux fonctions. Ces fonctions découlent surtout des travaux de Miller (1977)⁹.

9. Miller, M. (1977), Debt and Taxes, *Journal of Finance*, mai, p. 261-276.

LES MÉTHODES NUMÉRIQUES EN ÉCONOMÉTRIE UNE INTRODUCTION

Dans ce chapitre, nous envisageons certaines techniques de simulation qui sont utilisées en économétrie financière pour générer des distributions, telles la simulation de Monte Carlo et la technique dite du *bootstrapping*. Ce chapitre s'attarde également sur une brève introduction au calcul stochastique dans ses rapports avec l'économétrie financière.

1. SIMULATION DE MONTE CARLO : LE CAS D'UNE OPTION ASIATIQUE¹

Une simulation de Monte Carlo vise généralement à générer la distribution d'une variable économique ou financière, du moins dans ses rapports avec la science économique et la théorie financière. En économétrie, l'input d'une simulation de Monte Carlo est généralement une distribution tandis qu'en finance, l'input est généralement une équation différentielle stochastique discrétisée. Le but est alors par exemple de calculer le prix théorique d'un instrument financier, telle une option, ou d'évaluer le risque théorique d'un portefeuille, par exemple lors d'une simulation de Monte Carlo qui met en cause la VaR².

-
1. Ce chapitre est basé sur le cahier de recherche suivant : Racicot, F.E. et R. Théoret (2001), Les méthodes numériques en économétrie financière. Analyse de quelques cas : La simulation de Monte Carlo, le bootstrap et le kernel, Document de travail du CRG, École des sciences de la gestion, UQAM.
 2. VaR est l'acronyme anglais de *Value at Risk*.

Pour mieux comprendre la technique de la simulation de Monte Carlo, nous recourons à un cas, celui de la détermination du prix d'une option de vente asiatique. Le prix d'exercice d'une telle option est la moyenne des prix de l'action sous-jacente du début de la durée de l'option jusqu'à son échéance. L'expression générale du prix d'une option, comme d'ailleurs du prix de tout instrument financier, en termes de ses cash-flows est la suivante :

$$\text{Prix} = e^{-r(T-t)} E^* [\text{Cash-flows}(S)]$$

où r désigne le taux sans risque ; $(T - t)$, le temps qu'il reste jusqu'à l'échéance de l'option ; E^* , l'espérance neutre au risque et Cash-flows, les cash-flows finaux de l'option selon les divers états de la nature donnés ici par les prix probables de l'action sous-jacente.

Avant de progresser davantage dans cette présentation de la simulation de Monte Carlo appliquée au calcul du prix d'une option asiatique, effectuons une courte digression³ sur l'équation différentielle stochastique du prix d'une action dont nous nous servons ici pour construire notre simulation. Dans sa forme générale, cette équation, appelée *mouvement brownien géométrique*, est la suivante :

$$dS_t = rS_t dt + \sigma S_t dW_t$$

où S désigne le prix de l'action ; dt désigne la période de temps ; s est l'écart-type du rendement de l'action ; dW est un processus de Wiener d'espérance nulle et de variance dt . Pour en déduire l'équation du prix de l'action, soit S , divisons d'abord cette dernière équation par S et intégrons cette équation de 0 à t :

$$\int_0^t \frac{dS_u}{S_u} = \int_0^t r du + \int_0^t \sigma dW_u$$

La première intégrale est une intégrale de Riemann standard égale à rt et la seconde intégrale contient un terme aléatoire dW , mais son coefficient est constant dans le temps. Cette intégrale peut donc se

3. Pour plus de détails, voir : Neftci, S.N. (1996), *An Introduction to the Mathematics of Financial Derivatives*, Academic Press, New York ; Wilmot, P. (1998), *Derivatives: The Theory and Practice of Financial Engineering*, John Wiley and Sons, New York ; Briys, E. *et al.*, (1998), *Options, Futures and Exotic Derivatives*, John Wiley and Sons, New York.

calculer de la façon habituelle : $\sigma(W_t - W_0) = \sigma W_t$ puisque $W_0 = 0$. Par conséquent,

$$\int_0^t \frac{dS_u}{S_u} = rt + \sigma W_t$$

Toute solution de cette intégrale stochastique doit satisfaire cette intégrale. En particulier, l'une des solutions est la solution suivante :

$$S_t = S_0 e^{\left(\left(r - \frac{1}{2} \sigma^2 \right) t + \sigma W_t \right)}$$

Cette solution est fonction des paramètres σ , r et W_t . On peut vérifier l'exactitude de cette solution en différenciant cette dernière équation par le biais du lemme d'Itô. Ouvrons ici une parenthèse pour introduire ce lemme très important en calcul stochastique. Supposons que la variable aléatoire x suive le processus d'Itô suivant :

$$dx = a(x, t)dt + b(x, t)dW$$

où W désigne un processus de Wiener et a et b des fonctions de x et de t . La tendance (*drift*) de x est de a et sa variance est de b^2 . Soit une autre fonction de x et t désignée par G . En vertu du lemme d'Itô, la fonction G suit le processus suivant :

$$dG = \left(\frac{\partial G}{\partial x} a + \frac{\partial G}{\partial t} + \frac{1}{2} \frac{\partial^2 G}{\partial x^2} b^2 \right) dt + \frac{\partial G}{\partial x} b dW$$

où dW désigne le même processus de Wiener que celui auquel obéit x . G suit également un processus d'Itô. Son terme tendanciel est le

contenu de la parenthèse de dt et sa variance est $\left(\frac{\partial G}{\partial x} \right)^2 b^2$. Par exemple, supposons que S suive le processus suivant : $dS = adt + b dW = uSdt + \sigma SdW$. Supposons que $G = \ln S$. En appliquant le lemme d'Itô, on a :

$$dG = \left(u - \frac{\sigma^2}{2} \right) dt + \sigma dW$$

puisque $\frac{\partial G}{\partial S} = \frac{1}{S}$ et $\frac{\partial^2 G}{\partial S^2} = -\frac{1}{S^2}$ et $\frac{\partial G}{\partial t} = 0$. On peut montrer que le lemme d'Itô est au calcul stochastique ce qu'est l'expansion de Taylor au calcul différentiel classique.

On obtient, en appliquant le lemme d'Itô à S qui, on le rappelle, est

$$\text{égal à : } S_t = S_0 e^{\left(\left(r - \frac{1}{2}\sigma^2 \right) t + \sigma W_t \right)} ;$$

$$dS_t = \left[S_0 e^{\left(\left(r - \frac{1}{2}\sigma^2 \right) t + \sigma W_t \right)} \right] \left[\left(r - \frac{1}{2}\sigma^2 \right) dt + \sigma dW_t + \frac{1}{2}\sigma^2 dt \right]$$

On retrouve alors l'équation différentielle initiale : $dS_t = rS_t dt + \sigma S_t dW_t$. Cette digression avait pour but de justifier la forme de l'équation du prix de l'action qui sert de base à la simulation de Monte

Carlo : $S_t = S_0 e^{\left(\left(r - \frac{1}{2}\sigma^2 \right) t + \sigma W_t \right)}$. Comme S_t suit une loi lognormale,

$\ln S_t$ suit une loi normale. Ce dernier est obtenu en prenant directement le logarithme de la dernière équation, écrite sous forme différentielle :

$$d \ln S_t = \left(r - \frac{1}{2}\sigma^2 \right) dt + \sigma dW_t$$

Cette équation participe de la nature des mouvements browniens généralisés⁴. Cependant, avant d'enclencher la simulation de Monte Carlo, nous devons discrétiser⁵ cette équation. On peut à cet effet utiliser la méthode d'Euler pour obtenir une discrétisation du premier degré⁶. Pour y parvenir, discrétisons dans un premier temps l'équation différentielle qui sert de base à l'équation du prix, soit :

4. Un mouvement brownien généralisé est de la forme : $dX = udt + \sigma dW$ alors qu'un mouvement brownien géométrique est de la forme : $dX = uXd + \sigma XdW$.
5. Discrétiser signifie convertir en temps discret une équation en temps continu.
6. À remarquer qu'il existe des méthodes d'approximation plus exactes comme celle de Milstein, qui est du second degré. Pour des éclaircissements sur ce sujet, voir : Jegadeesh, N. et B. Tuckman (2000), *Advanced Fixed-Income Valuation Tools*, Wiley, New York.

$$\Delta S_t = rS_t \Delta t + \sigma S_t \sqrt{\Delta t} \phi$$

où Δ représente une variation finie et $\phi \sim N(0, 1)$. On discrétise également sa solution, soit :

$$S_{t+\Delta t} = S_t e^{\left(\left(r - \frac{1}{2}\sigma^2 \right) \Delta t + \sigma \sqrt{\Delta t} \phi \right)}$$

Les développements antérieurs avaient pour but de donner quelques fondements théoriques au cas de simulation de Monte Carlo qui nous intéresse, soit la détermination du prix d'une option de vente asiatique. Nous effectuons cette simulation en recourant au logiciel Excel. La méthode de Monte Carlo consiste à générer des séries de prix d'actions en utilisant la formule du prix d'une action qui vient d'être dérivée. Pour déterminer la valeur du put asiatique, on calcule ensuite la moyenne des prix de l'action que l'on retranche au prix d'exercice. Une série est générée sur une période de durée 1 par tranche de 0,01 période (Δt). Chaque série comporte donc 100 périodes. Au bout de chacune de ces périodes, on calcule la moyenne de ces prix, que l'on retranche au prix d'exercice. Pour les résultats, on consultera les colonnes moyenne et profit du chiffrier Excel qui apparaît au tableau 6.1.

Attardons-nous maintenant sur la technique du calcul d'une série. Pour calculer le prix d'un put asiatique⁷, nous avons besoin des données suivantes. Le prix de départ de l'action sous-jacente au put asiatique est ici fixé à 80 et le prix d'exercice du put est de 85. Le terme tendanciel⁸, soit r ou taux sans risque puisque l'on effectue une évaluation neutre au risque, est fixé à 5 %. La volatilité du rendement de l'action sous-jacente au put asiatique s'établit à 0,20 (20 %). Δt est de 0,01. Calculons, pour la première simulation⁹, qui apparaît au tableau 6.1 à la ligne SIM1 ($t = 0,01$), la formule qui nous sert à calculer l'évolution du prix de l'action par incrément d'une période à partir de son prix

7. Voici quelques références en français dans le domaine de la théorie des options : Khoury, N. et P. Laroche avec la collaboration de E. Briys et M. Crouhy (1990), *Options et contrats à terme*, Nathan ; Khoury, N. et P. Laroche, *Options et contrats à terme*, 2^e édition, Presses de l'Université Laval, Ste-Foy.

8. *Drift* en anglais.

9. 5 000 simulations similaires seront effectuées. Nous donnons ici un cas-type de simulation.

initial de 80. La formule mathématique utilisée a été donnée précédemment, soit : $S_{t+\Delta t} = S_t e^{\left(\left(r - \frac{1}{2}\sigma^2 \right) \Delta t + \sigma \sqrt{\Delta t} \phi \right)}$. En terme du langage d'Excel, elle s'écrit :

$$= E8*EXP((\$B\$10 - 0.5*(\$B\$8^2))*\$B\$9 + \$B\$8*(\$B\$9^(0.5))*(LOLNORMALE.STANDARD.INVERSE(ALEA()))))$$

TABLEAU 6.1

Simulation de Monte Carlo effectuée pour le calcul d'un put asiatique
 Programmé par : François-E. Racicot

P. l'action	80	Simulation	Temps		t=	t=
			0		0.01	0.02
Drift (u=r)	0.05	Sim				
Volatilité	0.2		1	80	79.2891205	79.27126577
Incrément	0.01	delta t	2	80	78.66499431	81.95555147
Interet r	0.05		3	80	81.17816896	81.77546605
			4	80	79.25448317	81.57007378
PV PUT	5.19 \$		5	80	80.33680191	78.61477857
			6	80	78.31094261	78.56076837
P.exercice	85		7	80	79.8205457	79.23789559
			8	80	80.82829744	79.19169645
			9	80	78.6009229	77.80161597
			10	80	79.85180274	79.96214356
			11	80	80.5444967	82.36397261
			12	80	82.41977267	82.92348779
			13	80	78.7899294	79.51720458
			14	80	82.94541457	83.96704888
			15	80	80.88513922	81.31031068
			16	80	81.86096848	83.49307978
			17	80	81.06315519	80.34279928
			18	80	78.69561097	77.63004399
			19	80	79.56165587	78.62794456
			20	80	79.62973861	81.73153194
			21	80	80.85553489	79.22527776
			22	80	76.24670934	78.53087555
			23	80	81.01945084	82.32192909
			24	80	80.0000912	78.49258051
			25	80	79.27836197	79.81361149
			26	80	81.47731027	80.59582096
			27	80	79.65430822	80.22714671
			28	80	80.67799503	79.62513833
			29	80	80.72992357	80.77637017
			30	80	80.38681647	79.93164664

où le contenu des cellules est le suivant: E8 est le prix initial de l'action, soit 80; B10, le taux d'intérêt sans risque (*drift*); B8, l'écart-type du rendement de l'action (σ); B9, l'incrément périodique (Δt); ALEA, le générateur de variables aléatoires d'Excel. La fonction *LOI.NORMALE.STANDARD.INVERSE()* exprime ALEA en une valeur obéissant à une loi normale $N(0, 1)$. Pour ce qui concerne la première simulation, on copie la cellule dans les colonnes de 0,01 jusqu'à 1. Rendu à 1, on ouvre une autre cellule (DC8 dans le cas de notre chiffrier) dans laquelle on calcule la moyenne des prix obtenus dans l'intervalle $[0, 1]$. La formule de DC8 est la suivante :

$$=MOYENNE(E8:DA8)$$

t = 0.03	t = 1	Moyenne	cash-flow	Espérance
80.85290647	121.1143472	96.52243117	0	5.4577
80.48461538	92.71114888	76.75135139	8.248648605	
80.49076078	87.95487337	83.61684531	1.383154694	
79.62603839	94.88866876	78.89180296	6.108197036	
74.29885889	61.70866471	73.06788345	11.93211655	
81.2310313	87.44423868	77.69186473	7.308135266	
77.63445005	64.58821663	76.93485993	8.065140071	
78.26866563	59.96705756	64.53758041	20.46241959	
79.38815434	90.62379039	81.72143152	3.278568483	
79.37309059	82.5227778	75.06513338	9.93486622	
84.87106345	102.7100865	90.46294143	0	
85.01618452	73.09447992	83.18424521	1.815754793	
78.54671554	130.097093	108.1235819	0	
82.75009397	61.4196164	72.65938023	12.34061977	
80.11606546	96.81254171	93.34845228	0	
84.7212354	120.8615132	97.96073286	0	
80.06404572	78.51878392	77.69767106	7.302328944	
79.22190591	70.70213568	75.25370363	9.746296368	
78.5616171	62.461891	71.48802002	13.51197998	
81.82941669	64.18020385	69.87013862	15.12986138	
78.17742773	61.4891354	74.121112	10.878888	
80.27823954	62.80009254	69.24875683	15.75124317	
80.93960323	109.0163873	94.25784774	0	
76.39817994	92.98605881	84.8702347	0.129765301	
81.01382557	47.250664197	65.34312409	19.65687591	
82.14333057	70.41452489	76.38069633	8.619303674	
81.46243778	97.62815574	100.1983943	0	
81.40768541	106.275939	88.3070333	0	
81.26770213	61.93170253	70.60894538	14.39105462	
81.09882507	82.88289369	83.16753639	1.832463614	

On peut alors calculer le cash-flow final du put associé à la première simulation dans la cellule DD8, toujours dans notre chiffrier Excel, dont la formule est la suivante :

$$=MAX(\$B\$14-DC8,0)$$

où la cellule B14 contient le prix d'exercice. On reproduit cette simulation 5000 fois dans notre exemple.

Finalement, on calcule l'espérance risque-neutre (E^*), qui, on le rappelle, est égale à : $E^* [Cash-flows(S)]$, s'exprime, dans le langage d'Excel :

$$=SOMME(DD8:DD5007)/NB(DD8:DD5007).$$

où la colonne s'étirant de DD8 jusqu'à DD5007 renferme les cash-flows du put. Le résultat de cette opération apparaît dans la cellule DG6 en ce qui concerne notre chiffrier. Pour obtenir le prix du put asiatique, il reste à actualiser l'espérance E^* , c'est-à-dire :

$$=SDG\$6*EXP(-\$B\$10*\$DA\$6)$$

où DA6 est la cellule qui renferme la période 1, soit le temps terminal de chaque simulation. Ce résultat est reporté dans la cellule B12 qui nous fournit le prix du put asiatique, soit 5,24 \$.

Dans l'exemple précédent, nous avons effectué 5000 simulations. Dans certains cas, ce nombre peut s'avérer insuffisant. Par exemple, dans le cas de la discrétisation d'Euler du modèle de Cox, Ingersoll et Ross (1985)¹⁰ servant à la modélisation de la structure à terme des taux d'intérêt, il peut être requis d'effectuer jusqu'à 10 millions de simulations. En effet, l'approximation cause deux types d'erreurs : l'erreur systématique et l'erreur statistique. Plus précisément : $\hat{e} = e_{SYS} + \hat{e}_{STAT}$, où \hat{e} désigne l'erreur totale, e_{SYS} , l'erreur systématique et \hat{e}_{STAT} , l'erreur statistique. Dans le modèle de Cox, Ingersoll et Ross, l'erreur statistique ne se résorbe qu'au bout de 10 millions de simulations. Cela donne à penser que l'approximation d'Euler fait problème et qu'il faudrait peut-être recourir à des approximations de degrés plus élevés que 1, telle l'approximation de Milstein qui est une discrétisation du second degré¹¹.

10. Cox, J.C., J.E. Ingersoll et S.A. Ross (1985), « A Theory of Term Structure of Interest Rates », *Econometrica*, vol. 53.

11. Pour plus de détails, voir : Jegadeesh, N. et B. Tuckman (2000), *Advanced Fixed-Income Valuation Tools*, John Wiley and Sons, New York, chap. 13.

Mentionnons finalement que certains générateurs de variables aléatoires, telle la fonction ALEA d'Excel, finissent par reproduire les mêmes variables à partir d'un certain nombre de simulations. À titre d'exemple, si l'on veut générer 1 million de variables aléatoires à partir de la version d'Excel 1995, il pourrait y avoir jusqu'à 30 copies du même ensemble de nombres aléatoires. Si l'on veut par exemple générer 10 millions de nombres aléatoires, le logiciel Mathematica serait requis dans ce cas. Il peut en effet générer 10^{445} nombres aléatoires avant de se répéter¹². Incidemment, pour le cas de l'évaluation du prix d'un titre, l'erreur de mesure décroît en proportion inverse de la racine carrée du nombre (N) de simulations. Prenons par exemple le cas de la simulation du prix d'un produit dérivé désigné par f par la méthode de Monte Carlo. On peut alors construire l'intervalle de confiance de f qui nous donnera le nombre de simulations requises pour en arriver à une précision suffisamment élevée. Cet intervalle, au seuil de 95 %, est le suivant :

$$\mu - \frac{1,96\sigma}{\sqrt{N}} < f < \mu + \frac{1,96\sigma}{\sqrt{N}}$$

où μ est la valeur estimée de f lors de la simulation et σ , l'écart-type¹³.

2. LA MÉTHODE DU BOOTSTRAP

La méthode du *bootstrap* est due à Efron (1979)¹⁴. Cette méthode est très apparentée à la simulation de Monte Carlo, mais elle a l'avantage de ne pas requérir de distribution a priori dans le mécanisme de génération des variables aléatoires. Nous avons déjà mentionné que la médiane est un estimateur robuste de la tendance centrale alors que la moyenne ne l'est pas dans le cas d'une distribution non normale. Supposons que l'on ait un échantillon de départ X dont la distribution théorique est inconnue. La méthode du *bootstrap* peut être utilisée

-
12. Bhansani, V. (1998), *Pricing and Managing Exotic and Hybrid Options*, McGraw-Hill, New York, chap. 5.
 13. Pour des informations additionnelles, voir : Hull J.C., (2000), *Options, Futures and Other Derivatives*, quatrième édition, Prentice Hall, New Jersey, chap. 16.
 14. Efron, B. (1979), « Bootstrap Methods: Another Look at the Jackknife », *Annals of Statistics*, n° 7, p. 1-26.

pour générer certaines informations reliées à cette distribution. Par exemple¹⁵, la méthode du *bootstrap* peut servir à calculer l'écart-type de la médiane de cet échantillon. Pour ce faire, on recourt à la formule suivante :

$$\sigma_{\text{BOOT}} = \sqrt{\frac{1}{B-1} \sum_{i=1}^B [\hat{M}_i - \hat{M}_{(\cdot)}]^2}$$

où σ_{BOOT} désigne l'écart-type de la médiane ; B, le nombre de simulations ; \hat{M}_i , la médiane du nouvel échantillon ; $\hat{M}_{(\cdot)} = \frac{1}{B} \sum_{i=1}^B \hat{M}_i$, soit la moyenne des \hat{M}_i découlant des simulations. Attardons-nous à expliquer la procédure du *bootstrap* pour le calcul de l'écart-type de la médiane.

- i) La première étape consiste à générer un échantillon aléatoire X_i de taille n avec remise à partir de l'échantillon initial, également de taille n . Toute observation de cet échantillon de nombre aléatoires comporte la même probabilité d'occurrence, soit $1/n$.
- ii) La seconde étape revient à calculer la médiane pour l'échantillon qui vient d'être généré.
- iii) Et l'on répète ces deux étapes B fois. On peut alors calculer l'écart-type de la médiane à partir de la formule pertinente.

On peut généraliser cette procédure à toute statistique $\theta(\mathbf{y})$ calculée à partir d'un échantillon y_t ($t = 1, \dots, n$) dont on ne connaît pas la distribution dans un petit échantillon¹⁶. L'opération *bootstrap* vise ici à approximer la distribution de cette statistique à partir de l'échantillon des y observés. Pour ce faire, on doit tirer un nombre B d'échantillons de taille n . Cet échantillonnage est effectué avec remise. Certains de ces échantillons pourront contenir les mêmes observations plus d'une fois et les observations qui y apparaîtront seront dans un ordre com-

15. Nous nous inspirons ici de : Johnston, J. et J. Dinardo (1997), *Econometric Methods*, 4^e édition, McGraw-Hill, New York, chap. 11.

16. Nous nous inspirons ici de : Davidson, R. et J.G. McKinnon (1993), *Estimation and Inference in Econometrics*, Oxford University Press, New York, chap. 21.

plètement différent de l'échantillon initial. On calcule $\theta(\mathbf{y}^*(i))$ pour chacun de ces échantillons, où $i = 1, \dots, B$. En bout de piste, on obtient ainsi B statistiques $\theta(\mathbf{y}^*(i))$. On peut alors calculer toute statistique reliée à la distribution de $\theta(\mathbf{y})$, comme l'écart-type ou autres moments.

On peut également illustrer la méthode du *bootstrap* dans le cas d'un modèle de régression linéaire¹⁷. Le but est ici de calculer l'espérance de $\hat{\beta}$. Soit le modèle de régression linéaire suivant :

$$y_t = x_t \beta + e_t \quad t = 1, \dots, n$$

Nous disposons d'un échantillon de départ pour les y et les x . On peut calculer le premier $\hat{\beta}$, désigné par $\hat{\beta}(1)$ à partir de ces observations. On effectue des *bootstraps* sur les y et x et à chaque fois on recalcule $\hat{\beta}(i)$, $i = 1, \dots, B$. On peut alors calculer à partir de ces derniers l'espérance de $\hat{\beta}$, c'est-à-dire :

$$E(\hat{\beta}) \cong \frac{1}{B} \sum_{i=1}^B \hat{\beta}(i).$$

Wilmott (1998)¹⁸ donne un exemple de *bootstrap* appliqué à une série de rendements d'actions. Il dispose de 1000 rendements journaliers sur chaque action. Il reporte ces rendements dans un chiffrier. Les cellules sont numérotées de 1 à 1000, une pour chaque journée de l'échantillon et à chacune de ces journées sont associés les rendements des actions retenues pour cette journée. Il effectue des tirages avec remise 1000 fois en utilisant une distribution uniforme. Cela constitue un premier scénario pour les rendements de son échantillon d'actions sur l'horizon d'investissement envisagé. Et il refait ce processus autant de fois qu'il le faut pour générer une distribution précise des rendements futurs de son portefeuille d'actions. Il peut alors calculer par exemple la VaR de son portefeuille, qui correspond à la perte maximale que peut subir celui-ci à un seuil de confiance généralement fixé à 95 % ou à 99 %.

17. On peut également utiliser la méthode du *bootstrap* dans le cadre du modèle de régression non linéaire, mais celle-ci est alors très intensive en calculs.

18. Wilmott, P. (1998), *Derivatives: The Theory and Practice of Financial Engineering*, Wiley, New York, chap. 42.

3. RÉGRESSION NON PARAMÉTRIQUE : UNE SIMULATION MONTE CARLO

Dans un chapitre précédent, nous avons présenté les corrélations non paramétriques de Spearman et de Kendall. Pour poursuivre la discussion sur les méthodes non paramétriques, on peut montrer comment ces méthodes peuvent s'appliquer au domaine de la régression. À cette fin, nous envisageons un cas basé sur des observations générées par la simulation de Monte Carlo et mettant en cause la régression recourant aux kernels.

Soit le modèle de régression suivant :

$$y_t = m(\mathbf{x}_t) + e_t$$

où $m(\cdot)$ est une fonction arbitraire non linéaire et e_t , un terme d'erreur. Il s'agit ici d'estimer m par la méthode de régression non paramétrique, le kernel. Dans les régressions du type kernel, la fonction de poids $w_{t,T}(x)$ construite à partir de fonctions de densité de probabilité (pdf) $k(x)$ et justement appelée kernel possède les propriétés suivantes :

- i) $k(x) \geq 0$;
- ii) $\int_{-\infty}^{\infty} k(u) du = 1$;

Et même si le kernel ne joue aucun rôle sur le plan probabiliste dans l'analyse qui suit, on recourt à une fonction de densité connue pour déterminer les pondérations comme le kernel gaussien qui est retenu dans le cas de notre exemple. Ce dernier a la forme suivante :

$$k_h(x) = \frac{1}{h} k\left(\frac{x}{h}\right) = \frac{1}{h\sqrt{2\pi}} e^{-\frac{x^2}{2h^2}}$$

Il est à remarquer que nous avons ici effectué un changement d'échelle de la variable x en la reportant dans un intervalle h compris entre $0,1 \times \hat{\sigma}_x$ et $0,5 \times \hat{\sigma}_x$ où $\hat{\sigma}_x$ est l'écart-type de $\{\mathbf{x}_t\}$. Notons que h est choisi de façon à minimiser un critère : $h = \arg \min_h Q(h)$. Ce critère nous incite à choisir la valeur $0,1 \times \hat{\sigma}_x$. La propriété ii) du kernel s'exprime, dans le cas de la transmutation d'échelle : $\int k_h(u) du = 1$

où $k_h(u) = \frac{1}{h} k\left(\frac{u}{h}\right)$. Définissons la fonction de poids $w_{t,T}(x)$ qui entre dans notre moyenne pondérée comme suit :

$$w_{t,T}(x) = \frac{k_h(x - x_t)}{g_h(x)}$$

où $g_h(x) = \frac{1}{T} \sum_{t=1}^T k_h(x - x_t)$.

La technique de lissage des données de Nadaraya-Watson, l'estimateur kernel de m , se définit comme suit :

$$\hat{m}_h(x) = \frac{1}{T} \sum_{t=1}^T w_{t,T}(x) y_t = \frac{\sum_{t=1}^T k_h(x - x_t) y_t}{\sum_{t=1}^T k_h(x - x_t)}$$

On remarquera dans cette formule que l'on effectue une moyenne pondérée des y_t . Si h est très petit, la moyenne sera calculée dans un voisinage très concentré autour de x . Par ailleurs, quand l'intervalle est important, la moyenne sera calculée dans un voisinage plus étendu autour de x . h est donc le paramètre de lissage.

Passons maintenant à un exemple programmé dans le logiciel Excel. Dans un premier temps, nous générons les variables y à partir d'une simulation de Monte Carlo. Nous générons à cette fin 200 variables y à partir de la relation suivante :

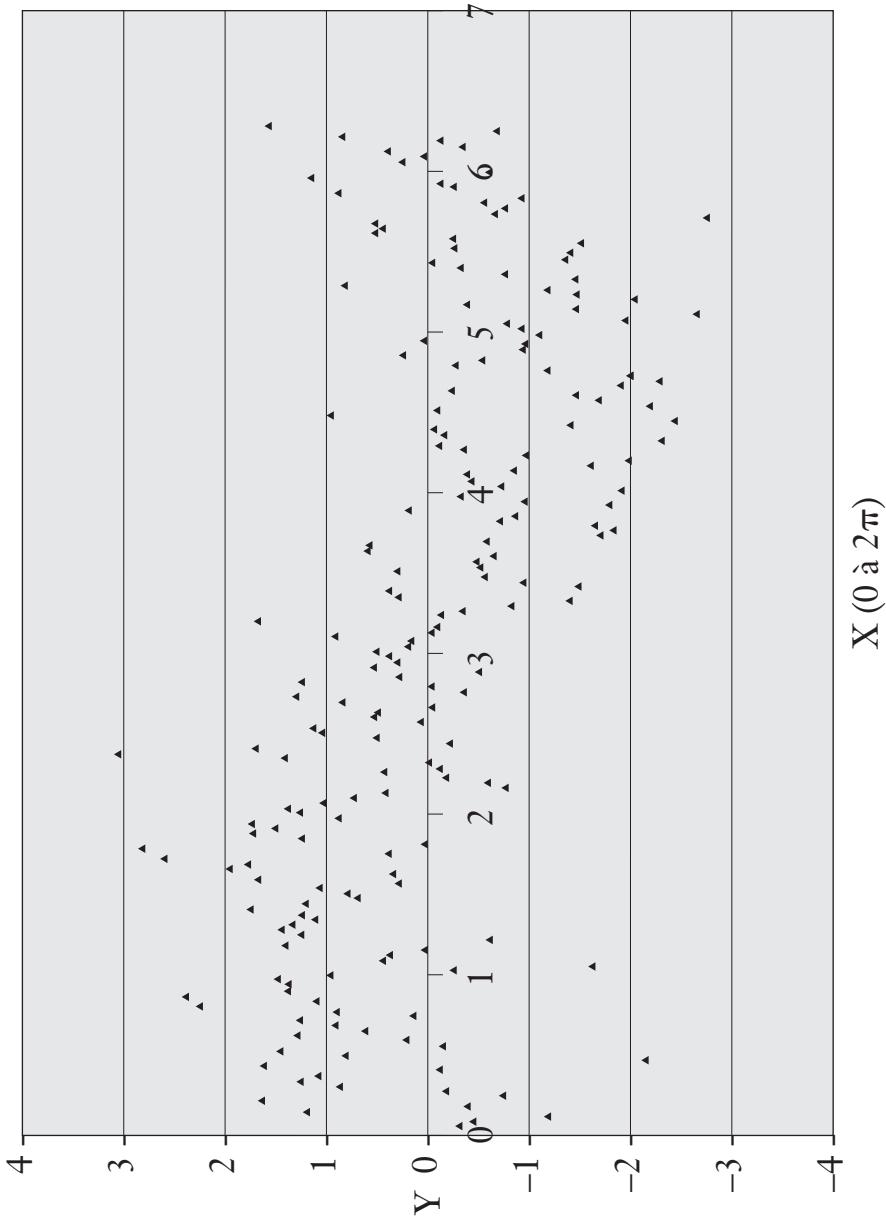
$$y_t = \sin(x_t) + e_t$$

où $e_t \sim N(0, 1)$. Dans Excel, en faisant varier x_t de 0 à 2π , on écrit :

$$= \text{SIN}(B4) + \text{LOI.NORMALE.STANDARD.INVERSE}(\text{ALEA}())$$

et on copie cette cellule autant de fois qu'on le désire, 200 dans notre cas. Ce faisant, on obtient la variable y de notre simulation de Monte Carlo. La figure 6.1 reproduit les y obtenus par cette simulation :

FIGURE 6.1



On se sert ensuite de ces données pour calculer l'estimateur kernel obtenu, rappelons-le, à partir de la formule suivante :

$$\hat{m}_h(x) = \frac{1}{T} \sum_{t=1}^T w_{t,T}(x) y_t = \frac{\sum_{t=1}^T k_h(x - x_t) y_t}{\sum_{t=1}^T k_h(x - x_t)}. \text{ Incidemment, } \hat{m}_h(x)$$

est l'estimateur de: $E(y_t | x_t) = \sin(x_t)$. Le résultat apparaît à la figure 6.2.

Pour générer $k_h(x - x_t)$ dans le logiciel Excel, on recourt à la formule suivante :

$$=(1/(\$E\$4*RACINE(2*PI())))/(EXP(((\$F\$2-B4)^2)/(2*\$E\$4^2)))$$

La cellule E4 renferme la valeur de h. Elle se calcule comme suit :

$$=ECARTYPE(B4:B204)*0.1$$

Les valeurs de x sont situées dans les cellules B4 à B204. Pour calculer $\hat{m}_h(x)$, on recourt à la formule suivante :

$$=SOMMEPROD(K4:K204; \$C \$4: \$C \$204)/NB(B4:B204)$$

où les cellules K4 à K204 renferment les valeurs de $k_h(\cdot)$, et les cellules C4 à C204, les valeurs générées de y.

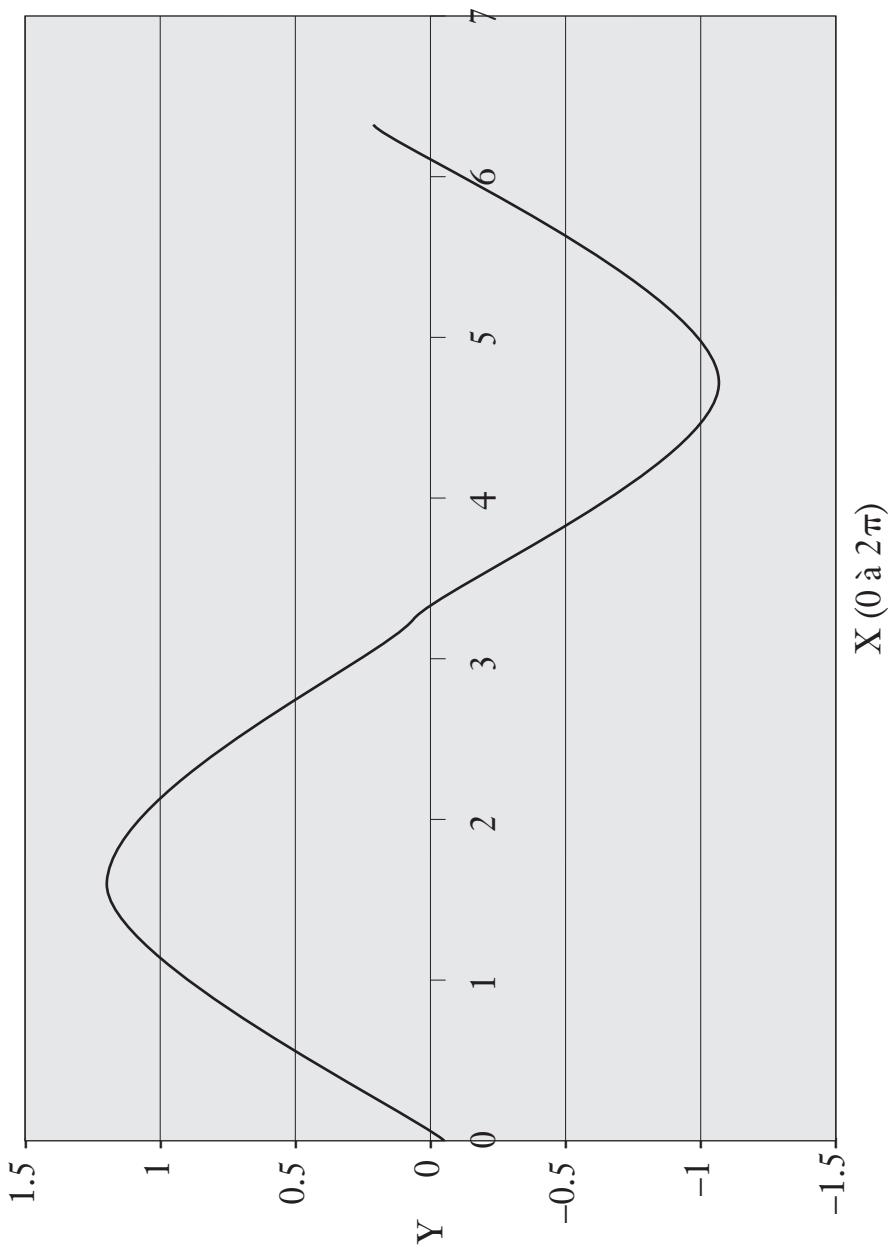
Finalement, il faut donner des valeurs à x pour le calcul de $\hat{m}_h(x)$. Nous avons choisi les cinq valeurs suivantes programmées dans Excel : 0, =PI()/2, =PI(), =3PI()/2, =2PI().

En finance, la régression kernel peut servir à estimer les bêtas des titres. Par exemple, si nous disposons des rendements mensuels, disons de IBM et du TSE300, on peut générer l'estimateur kernel du bêta d'IBM. Les méthodes non paramétriques peuvent servir aussi à estimer des distributions empiriques et à effectuer des prévisions¹⁹. À cet effet, Diebold et Nason appliquent les méthodes non paramétriques pour prévoir le taux de change²⁰.

19. Pour cette section, nous nous sommes référés aux documents suivants : Campbell, J.Y. *et al.* (1997), *The Econometrics of Financial Markets*, Princeton University Press ; Oxley, L. *et al.* (1995), *Surveys in Econometrics*, Blackwell ; Johnston et Dinardo (1997), *op. cit.*, chap. 11.

20. Diebold, F.X. et J.A. Nason (1990), « Nonparametric Exchange Rate Prediction ? », *Journal of International Economics*, 28, p. 315-332.

FIGURE 6.2



L'HÉTÉROSCÉDASTICITÉ

Dans ce chapitre, nous examinons d'abord comment la présence de l'hétéroscédasticité vient modifier les propriétés classiques des MCO. Puis nous envisageons les correctifs au problème de l'hétéroscédasticité, soit les estimateurs des moindres carrés pondérés et des moindres carrés généralisés. Dans le cas pour lequel la matrice variance-covariance est inconnue, nous montrons comment on peut construire l'estimateur des moindres carrés quasi généralisés à partir de différentes formes d'hétéroscédasticité en guise de correction de ce problème. À cet effet, nous présentons la matrice de White pour traiter ledit problème. Suivront les tests classiques pour circonscrire l'hétéroscédasticité, soit les tests de Goldfeld et Quandt, de Breusch-Pagan, de White et un test général asymptotique. Finalement, nous présentons quelques applications.

1. PROPRIÉTÉS DE L'ESTIMATEUR DES MCO LORSQUE LES ERREURS SONT HÉTÉROSCÉDASTIQUES

Le terme *hétéroscédasticité* comporte les racines suivantes. D'abord « scédastique » est associé à « fonction scédastique », qui signifie « variance conditionnelle ». La racine « hétéro » fait référence à « plusieurs ». Hétéroscédasticité signifie donc *différentes variances*. Bien que l'hétéroscédasticité se retrouve à la fois dans les séries en coupe instantanée et dans les séries temporelles, nous n'étudierons dans ce chapitre que la première forme. La seconde sera abordée ultérieurement.

Précisons maintenant ce que nous entendons par hétéroscédasticité. Soit le modèle de régression linéaire suivant ne comportant qu'une seule variable explicative :

$$y_t = \beta_1 + \beta_2 x_t + e_t$$

où $E(e_t) = 0$ et $V(e_t) = \sigma_t^2 = \sigma^2 x_t$. Analysons maintenant ces nouvelles hypothèses. Nous supposons donc que les résidus sont d'espérance nulle mais que la variance n'est pas fixe comme dans le cas homoscédastique mais bien variable. Elle varie ici en proportion directe de x_t . En termes matriciels, la matrice variance-covariance s'écrit :

$$E(\mathbf{e}\mathbf{e}^T) = \mathbf{V} = \sigma^2 \begin{bmatrix} x_1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & x_2 & & & & \cdot \\ \cdot & & \cdot & & & \cdot \\ \cdot & & & \cdot & & \cdot \\ \cdot & & & & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & x_T \end{bmatrix} = \sigma^2 \mathbf{\Omega}$$

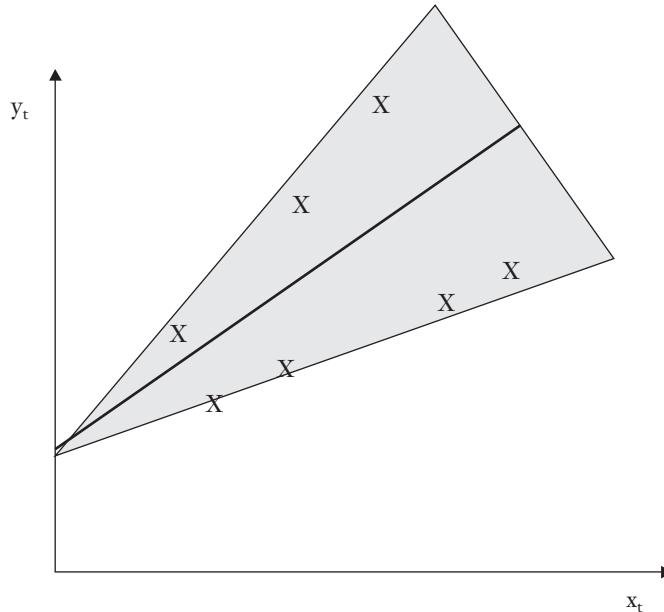
Si on applique les MCO sur l'équation écrite plus haute, on obtient encore un estimateur sans biais, c'est-à-dire : $E(\hat{\beta}) = \beta$. Mais la variance de $\hat{\beta}$ est de :

$$\begin{aligned} & E\left[(\hat{\beta} - E(\hat{\beta})) (\hat{\beta} - E(\hat{\beta}))^T \right] \\ &= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{\Omega} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \neq \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \end{aligned}$$

Cela implique que les MCO sont inefficients en présence d'hétéroscédasticité puisque la variance des MCO est biaisée. Représentons graphiquement le phénomène de l'hétéroscédasticité en reprenant l'exemple

de la relation entre les dépenses sur cartes de crédit et les revenus. Cette relation apparaît à la figure 7.1.

FIGURE 7.1



où y_t désigne les dépenses sur cartes de crédit et x_t , les revenus. La figure 7.1 indique que la variance des dépenses sur cartes de crédit, circonscrite par le cône, augmente en fonction du revenu individuel. En effet, plus le revenu est élevé, plus les dépenses font montre d'une dispersion importante en raison de la marge de manœuvre plus élevée des gens riches.

Comme l'hétéroscédasticité est ici de la forme : $\sigma_t^2 = \sigma^2 x_t$ et que l'on voudrait revenir à l'hypothèse d'homoscédasticité, c'est-à-dire : $E(e_t^2) = \sigma^2$, ce qu'il faut alors faire pour corriger ce problème est de

diviser e_t par $\sqrt{x_t}$: $V\left(\frac{e_t}{\sqrt{x_t}}\right) = \frac{1}{x_t} V(e_t) = \frac{1}{x_t} \sigma^2 x_t = \sigma^2$. On en déduit

que la solution au problème de l'hétéroscédasticité des erreurs est de corriger la régression par $(1/\sqrt{x_t})$:

$$\frac{y_t}{\sqrt{x_t}} = \beta_1 \frac{1}{\sqrt{x_t}} + \beta_2 \frac{x_t}{\sqrt{x_t}} + \frac{e_t}{\sqrt{x_t}}$$

où $\frac{e_t}{\sqrt{x_t}} \sim \text{IID}(0, \sigma^2)$. On peut réécrire le modèle corrigé comme suit :

$$y^*_t = \beta_1 x^*_{1t} + \beta_2 x^*_{2t} + e^*_t$$

Et sous forme matricielle, ce modèle s'écrit comme suit :

$$\mathbf{y}^* = \mathbf{X}^* \boldsymbol{\beta} + \mathbf{e}^*$$

où $\mathbf{e}^* \sim \text{IID}(0, \sigma^2 \mathbf{I})$. Si on applique les MCO sur cette régression, on obtient les moindres carrés pondérés :

$$\hat{\boldsymbol{\beta}}_{\text{WLS}} = (\mathbf{X}^{*\text{T}} \mathbf{X}^*)^{-1} \mathbf{X}^* \mathbf{y}^*$$

Les propriétés de cet estimateur sont les suivantes :

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}_{\text{WLS}}) &= \boldsymbol{\beta} \\ V(\hat{\boldsymbol{\beta}}_{\text{WLS}}) &= \sigma^2 (\mathbf{X}^{*\text{T}} \mathbf{X}^*)^{-1} \end{aligned}$$

où l'estimateur de σ^2 est de : $\frac{1}{T-K} \hat{\mathbf{e}}^{*\text{T}} \hat{\mathbf{e}}^*$, où k est ici de 2.

2. L'ESTIMATEUR DES MOINDRES CARRÉS GÉNÉRALISÉS : MCG

Soit \mathbf{P} la matrice diagonale de transformation des variables de la régression :

$$\mathbf{y}^* = \mathbf{P}\mathbf{y}; \mathbf{X}^* = \mathbf{P}\mathbf{X} \text{ et } \mathbf{P}^T\mathbf{P} = \begin{bmatrix} x_1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & x_2 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & x_T \end{bmatrix}^{-1} = \mathbf{\Omega}^{-1}.$$

Puisque $\mathbf{\Omega}^{-1}$ est diagonale, \mathbf{P} est égal à l'inverse de la racine carrée de la diagonale de $\mathbf{\Omega}$. Si l'on applique les MCO sur la régression transformée par \mathbf{P} , on obtient l'estimateur des moindres carrés généralisés (MCG) :

$$\hat{\beta}_{\text{GLS}} = (\mathbf{X}^T \mathbf{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{\Omega}^{-1} \mathbf{y} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}$$

$\hat{\beta}_{\text{GLS}}$ a la propriété d'être l'estimateur BLUE de β . On a en effet les relations suivantes : $E(\hat{\beta}_{\text{GLS}}) = \beta$; $V(\hat{\beta}_{\text{GLS}}) = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1}$. La variance de $E(\hat{\beta}_{\text{GLS}}) = \beta$ est ici la variance minimale dans la classe des estimateurs linéaires.

On a ici supposé que \mathbf{V} est connu. Si \mathbf{V} est inconnu, on peut l'estimer par $\hat{\mathbf{V}}$. On obtient ainsi les moindres carrés quasi généralisés (MCQG). Cet estimateur s'écrit comme suit :

$$\hat{\beta} = (\mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{y} = \left(\sum_{t=1}^T \hat{\sigma}_t^{-2} \mathbf{x}_t \mathbf{x}_t^T \right)^{-1} \left(\sum_{t=1}^T \hat{\sigma}_t^{-2} \mathbf{x}_t y_t \right)$$

où \mathbf{x}_t est un vecteur de variables explicatives.

Pour cet estimateur, σ_t^2 peut prendre un grand nombre de spécifications. Par exemple, il peut être égal à $\mathbf{z}_t^T \boldsymbol{\alpha}$ ou encore à $(\mathbf{z}_t^T \boldsymbol{\alpha})^2$. Considérons par exemple le second cas. On a :

$\sigma_t^2 = (\alpha_1 + \alpha_2 x_{2t} + \alpha_3 x_{3t} + \dots + \alpha_k x_{kt})^2 = (\mathbf{z}_t^T \boldsymbol{\alpha})^2$. Cette spécification représente une forme linéaire d'hétéroscédasticité qui dépend des variables explicatives. Pour rendre opérationnelle cette spécification, nous devons calculer les résidus estimés de la régression : $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ et effectuer la régression suivante :

$$|\hat{e}_t| = \mathbf{z}_t^T \boldsymbol{\alpha} + v_t$$

On estime cette régression par les MCO et l'on obtient les $\hat{\boldsymbol{\alpha}}$ du modèle d'hétéroscédasticité. On peut alors écrire le modèle d'hétéroscédasticité comme suit :

$$\hat{\sigma}_t^2 = (\mathbf{z}_t^T \hat{\boldsymbol{\alpha}})^2$$

Ayant postulé cette forme d'hétéroscédasticité, nous pouvons substituer cette valeur dans l'estimateur des MCQG et obtenir :

$$\hat{\boldsymbol{\beta}} = \left(\sum_{t=1}^T (\mathbf{z}_t^T \hat{\boldsymbol{\alpha}})^{-2} \mathbf{x}_t \mathbf{x}_t^T \right)^{-1} \left(\sum_{t=1}^T (\mathbf{z}_t^T \hat{\boldsymbol{\alpha}})^{-2} \mathbf{x}_t y_t \right)$$

Cette technique est utilisée dans le cas où plusieurs variables explicatives sont susceptibles de causer l'hétéroscédasticité. Donnons un exemple simplifié de cette technique de correction de l'hétéroscédasticité. Soit le modèle déjà considéré à deux variables explicatives qui relie les dépenses sur cartes de crédit aux revenus et aux revenus au carré.

$$\frac{y_t}{\hat{\sigma}_t} = \frac{1}{\hat{\sigma}_t} \beta_1 + \frac{x_{2t}}{\hat{\sigma}_t} \beta_2 + \frac{x_{3t}}{\hat{\sigma}_t} \beta_3 + \frac{e_t}{\hat{\sigma}_t}$$

Il reste à appliquer à cette équation les MCO pour obtenir les MCQG.

Notons ici deux propriétés des MCQG. D'abord, ils sont sans biais :

$E\left(\hat{\beta} \mid \mathbf{X}^*\right) = \beta$. On ne peut déterminer si cet estimateur est à variance minimale dans les petits échantillons parce qu'il n'est guère possible de calculer la matrice variance-covariance. Par contre, cet estimateur est convergent : $p\lim\left(\hat{\beta}\right) = \beta$. Ensuite, les propriétés asymptotiques de cet estimateur sont les mêmes que celles des MCG.

3. MATRICE DE WHITE POUR L'HÉTÉROSCÉDASTICITÉ

Au vu des développements précédents, la correction pour l'hétéroscédasticité peut s'avérer un processus ardu, ne serait-ce qu'en raison de la spécification de la forme de l'hétéroscédasticité. White (1980)¹ a proposé un estimateur pour la matrice écrite antérieurement : $V(\hat{\beta}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{\Omega} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$. Cet estimateur a la forme suivante :

$$\hat{V}(\hat{\beta}) = T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{S}_0 (\mathbf{X}^T \mathbf{X})^{-1}$$

où $\mathbf{S}_0 = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t^2 \mathbf{x}_t \mathbf{x}_t^T$ où \mathbf{x}_t est un vecteur de variables explicatives.

Cet estimateur est pré-programmé dans le logiciel EViews. En pratique, il suffira simplement d'estimer la régression linéaire : $\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$ par la méthode des MCO et de calculer la matrice de White de façon à obtenir des écarts-types corrigés pour l'hétéroscédasticité qui serviront à la construction des tests t corrigés. Un exemple de cet estimateur sera fourni dans la section des applications².

1. White, H. (1980), « A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity », *Econometrica*, vol. 48, p. 817-838.
2. Notons que la transformation Box-Cox peut être également utilisée pour traiter le problème de l'hétéroscédasticité. À cet effet, on consultera par exemple : Gaudry, M.J.I. et M.G. Dagenais (1979), « Heteroscedasticity and the Use of Box-Cox Transformations », *Economics Letters*, vol. 2, p. 225-229.

4. TESTS D'HÉTÉROSCÉDASTICITÉ

Dans cette section seront présentés les tests de Goldfeld et Quandt, de Breusch-Pagan, un test général et finalement celui de White.

4.1. Test de Goldfeld et Quandt (1965)

Lorsque l'on peut identifier la variable à la source de l'hétéroscédasticité, le test de Goldfeld et Quandt (GQ) est approprié. Soit le modèle de régression linéaire suivant: $y_t = \beta_1 + \beta_2 x_t + e_t$, où $e_t \sim N(0, \sigma_t^2)$, $\sigma_t^2 = x_t \sigma^2$. x_t étant la cause de l'hétéroscédasticité, on ordonne les observations par ordre décroissant des x_t . On sépare par la suite l'échantillon en trois partitions égales et l'on effectue la régression linéaire sur les partitions extrêmes, numérotées respectivement 1 et 2. On obtient alors la variance des termes d'erreur de chacune de ces régressions, c'est-à-dire :

$$\hat{\sigma}_1^2 = \frac{\hat{\mathbf{e}}_1^T \hat{\mathbf{e}}_1}{T_1 - 2} \text{ et } \hat{\sigma}_2^2 = \frac{\hat{\mathbf{e}}_2^T \hat{\mathbf{e}}_2}{T_2 - 2}$$

où $k=2$ puisque, dans notre régression, nous avons deux paramètres à estimer: le paramètre du terme constant et celui de la variable explicative. Le test GQ s'écrit comme suit :

$$\text{GQ} = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \sim F(T_1 - k, T_2 - k)$$

L'hypothèse testée par GQ est la suivante: $H_0: \sigma_1^2 = \sigma_2^2$ versus $H_1: \sigma_1^2 > \sigma_2^2$. Ce test est unilatéral et il consiste à rejeter H_0 si $F > F_c$ pour un seul de confiance α de 5 %. La version bilatérale de ce test existe également et consiste aussi à ordonner les observations par ordre décroissant et à mettre au numérateur de F la variance estimée la plus élevée. Notons que la puissance du test dépend du nombre d'observations exclues³.

3. Il est à signaler que l'on peut également effectuer le test GQ en ordonnant les observations par ordre croissant de x_t . La variance de la partition 2, qui est éventuellement la plus élevée, apparaît alors au numérateur de F .

4.2. Test de Breusch-Pagan (1979)

Dans la version présentée ci-devant, le test de Goldfeld-Quandt a pour défaut que la variable qui cause l'hétéroscédasticité doit être connue. Dans le cas où la variance de l'innovation dépend de plus d'une variable explicative, le test de Breusch-Pagan (BP) s'avère plus approprié. Soit le modèle de régression suivant: $y_t = \mathbf{x}_t^T \boldsymbol{\beta} + e_t$ où $\mathbf{x}_t^T = [1 \ x_{2t} \ x_{3t} \ \dots \ x_{kt}]$ et où $e_t \sim N(0, \sigma_t^2)$ et $\sigma_t^2 = h(\mathbf{z}_t^T \boldsymbol{\alpha})$, $\boldsymbol{\alpha}^T$, est un vecteur: $\boldsymbol{\alpha}^T = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_k]$. La fonction $h(\cdot)$ peut inclure toutes les variables explicatives. Cette formulation prend également en compte les différentes spécifications suivantes. Une première spécification est la forme multiplicative de l'hétéroscédasticité due à Harvey (1976)⁴, qui est incidemment une forme très flexible :

$$\sigma_t^2 = \sigma^2 e^{(\mathbf{x}_t^T \boldsymbol{\alpha})}$$

Deux autres spécifications de l'hétéroscédasticité sont les suivantes :

$$\sigma_t^2 = (\mathbf{x}_t^T \boldsymbol{\alpha})^2$$

$$\sigma_t^2 = (\mathbf{x}_t^T \boldsymbol{\alpha})$$

Le test de BP vise à tester l'hypothèse suivante: H_0 :

$\boldsymbol{\alpha}^T = [0 \ \dots \ 0]$ contre $H_1: \boldsymbol{\alpha}^T \neq [0 \ \dots \ 0]$. Si H_0 n'est pas rejetée, alors $\sigma_t^2 = \alpha_1$, ce qui implique l'hypothèse d'homoscédasticité. Pour effectuer ce test, on n'a qu'à suivre la procédure suivante :

- i) Obtenir les résidus estimés du modèle: $\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$;
- ii) Substituer \hat{e}_t^2 à σ_t^2 pour obtenir :

$$\hat{e}_t^2 = \mathbf{z}_t^T \boldsymbol{\alpha} + v_t$$

Ce type de régression est qualifié d'*auxiliaire* ou d'*artificielle*. Cette régression respecte le test BP puisque BP ne dépend

4. Harvey, A. (1976), « Estimating Regression Models with Multiplicative Heteroscedasticity », *Econometrica*, vol. 44, p. 461-465.

pas de $h(\cdot)$. En appliquant les MCO sur cette régression, on obtient ainsi la somme des carrés expliqués (SCE) :

$$SCE = \frac{SCR}{1 - R^2} - SCR = SCT - SCR. \text{ On calcule par la suite :}$$

$$\tilde{\sigma}^2 = \frac{\sum \hat{e}_t^2}{T}, \text{ qui est incidemment l'estimateur maximum de vraisemblance de } \sigma^2.$$

iii) Calculer la statistique BP sous $H_0 : \alpha = 0$. On a donc :

$$BP = \frac{SCE}{2\tilde{\sigma}^4} \overset{a}{\sim} \chi^2(s)$$

où s est le nombre de variables incluses dans la régression artificielle, incluant le terme constant. On rejette H_0 si $BP > \chi^2_c$.

Pour fixer les idées, dirigeons notre collimateur vers l'exemple suivant. On suppose que l'on a deux variables explicatives pour expliquer y , y étant les dépenses sur carte de crédit de l'individu i et les deux variables explicatives : le revenu et le revenu au carré de l'individu i . En appliquant la procédure décrite précédemment, on suit les étapes suivantes pour implanter le test BP :

i) On effectue la régression suivante :

$$\hat{e}_t^2 = \alpha_1 + \alpha_2 x_{2t} + \alpha_3 x_{3t} + v_t$$

ii) On obtient la SCE de cette régression et on calcule également $\tilde{\sigma}^2$, tel qu'expliqué précédemment.

iii) Sous l'hypothèse que les résidus sont normalement distribués et sous $H_0 : \alpha_2 = \alpha_3 = 0$, on construit la statistique BP :

$$BP = \frac{SCE}{2\tilde{\sigma}^4} \overset{a}{\sim} \chi^2(s). \text{ On rejette } H_0, \text{ soit l'hypothèse d'homoscédasticité, si } BP > \chi^2_c.$$

Si les résidus ne sont pas distribués normalement, on recourt à la régression artificielle présentée précédemment et l'on obtient le R^2 . On construit alors le test général asymptotique suivant :

$$TR^2 \overset{a}{\sim} \chi^2(s)$$

Ce test fait partie de la classe des tests LM. Le test BP et le test général LM qui se confondent asymptotiquement sous l'hypothèse de la non-normalité des résidus sont des tests assez généraux, on le rappelle, pour intégrer les formes d'hétéroscédasticité non linéaire.

4.3. Test de White (1980)

Le test de White est un test général de l'hétéroscédasticité. Ce test recourt à la matrice de White exposée précédemment :

$\hat{V}(\hat{\beta}) = T(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{S}_0 (\mathbf{X}^T \mathbf{X})^{-1}$. Il consiste à comparer cet estimateur à celui de la variance des moindres carrés ordinaires. Sous l'hypothèse d'homoscédasticité, l'estimateur de la variance des MCO est convergent tandis que sous l'hypothèse d'hétéroscédasticité, c'est l'estimateur de White qui l'est. Le test de White se fonde sur ce résultat. Ce test consiste comporte donc les étapes suivantes :

- i) Effectuer la régression de \hat{e}_t^2 sur une constante et chaque variable de $\mathbf{X} \otimes \mathbf{X}$.
- ii) Calculer TR^2 de cette régression et formuler le test :
 $\text{TR}^2 \overset{a}{\sim} \chi^2(k-1)$, où k est le nombre de régresseurs incluant la constante.
- iii) On rejette l'hypothèse $H_0 : \sigma_t^2 = \sigma^2 \forall t$ contre $H_1 : \sigma_t^2 \neq \sigma^2$, si $\text{TR}^2 > \chi_c^2$.

5. APPLICATIONS

Pour illustrer le problème de l'hétéroscédasticité dans les coupes instantanées, nous allons revenir au modèle des dépenses sur cartes de crédit en fonction du revenu de l'individu i , de son revenu au carré, de son âge et de la variable loyer-maison. Les résultats apparaissent au tableau 7.1.

TABLEAU 7.1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Dependent Variable: CARTE				
Method: Least Squares				
Date: 04/13/00 Time: 18:09				
Sample: 1 100 IF X3 > 0				
Included observations: 72				
C	-237.1465	199.3517	-1.189589	0.2384
REVENU	234.3470	80.36595	2.915999	0.0048
REVENU^2	-14.99684	7.469337	-2.007788	0.0487
AGE	-3.081814	5.514717	-0.558835	0.5781
LOYERMAISON	27.94091	82.92232	0.336953	0.7372
R-squared	0.243578	Mean dependent var	262.5321	
Ajusted R-squared	0.198418	S.D. dependent var	318.0468	
S.E. of regression	284.7508	Akaike info criterion	14.20802	
Sum squared resid	5432562.	Schwarz criterion	14.36612	
Log likelihood	-506.4888	F-statistic	5.393722	
Durbin-Watson stat	1.682310	Prob(F-statistic)	0.000795	

Le test de White

La matrice $\mathbf{X} \otimes \mathbf{X}$ qui est à la base de ce test comprend 13 variables, soit les variables initiales, les variables initiales élevées au carré et les produits croisés des variables initiales. La régression suivante devra donc être effectuée :

$$\hat{\varepsilon}_t^2 = \beta_1 + \beta_2(\hat{\text{age}}) + \beta_3(\hat{\text{age}})^2 + \beta_4(\hat{\text{age}} \times \text{revenu}) + \dots + v_t$$

À la suite de cette régression, on calcule le R^2 que l'on multiplie par le nombre d'observations et que l'on compare ensuite au χ_c^2 , c'est-à-dire dans notre cas :

$$TR^2 = 72 \times 0,199 = 14,33$$

Par ailleurs, le $\chi^2(12)$ est égal à 21,03 au seuil $\alpha = 5\%$. Selon H_0 : $\sigma_t^2 = \sigma^2$, soit l'hypothèse d'homoscédasticité. Contrairement à nos attentes, l'on ne peut ici rejeter l'hypothèse H_0 , soit l'absence

d'hétéroscédasticité. Signalons que les tests de Breusch-Pagan et de Goldfeld et Quandt rejettent pour leur part H_0 , ce qui est conforme au graphique antérieur reliant les dépenses sur cartes de crédit au revenu individuel sur lequel il y avait apparence d'hétéroscédasticité.

Finalement, le tableau 7.2 présente les statistiques t corrigées construites à partir des écarts-types de la matrice de White.

TABLEAU 7.2

Dependent Variable : CARTE				
Method : Least Squares				
Date : 04/13/00 Time : 18:12				
Sample : 1 100 IF X3 > 0				
Included observations : 72				
White Heteroskedasticity-Consistent Standard Errors & Covariance				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-237.1465	220.7950	-1.074058	0.2866
REVENU	234.3470	92.12260	2.543860	0.0133
REVENU^2	-14.99684	7.199027	-2.083177	0.0411
LOYERMAISON	27.94091	95.56573	0.292374	0.7709
AGE	-3.081814	3.422641	-0.900420	0.3711
R-squared	0.243578	Mean dependent var	262.5321	
Ajusted R-squared	0.198418	S.D. dependent var	318.0468	
S.E. of regression	284.7508	Akaike info criterion	14.20802	
Sum squared resid	5432562.	Schwarz criterion	14.36612	
Log likelihood	-506.4888	F-statistic	5.393722	
Durbin-Watson stat	1.682310	Prob(F-statistic)	0.000795	

Pour obtenir ces résultats, il suffit d'utiliser l'option du logiciel Eviews permettant de calculer la matrice de White tout en effectuant une régression ordinaire pour obtenir les coefficients de régression. On ne remarque ici que peu de différence entre les statistiques t corrigées et non corrigées : après correction, les variables non significatives demeurent non significatives et les variables significatives le demeurent.

6. NOTE SUR L'INFÉRENCE STATISTIQUE EN PRÉSENCE D'HÉTÉROSCÉDASTICITÉ

Les tests t et F de la régression corrigée ne sont pas valables, car on ignore la forme exacte de la matrice variance-covariance dans les petits échantillons. Il faut se méfier du R^2 associé à cette régression. Ce R^2 tendra en effet à être plus élevé après régression, mais cette augmentation peut être due au calcul de \mathbf{y}^* qui est utilisé dans le calcul de ce R^2 . Qui plus est, ce R^2 n'est pas nécessairement situé entre 0 et 1, mais on peut remédier à cette situation en utilisant le R^2 corrigé suivant :

$$R_{\text{GLS}}^2 = 1 - \frac{\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{GLS}}\right)^T \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{GLS}}\right)}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

où \mathbf{X} et \mathbf{y} sont les variables originales.

L'AUTOCORRÉLATION DES ERREURS RÉSIDUELLES

L'autocorrélation des erreurs est un phénomène qui se retrouve dans le domaine des séries temporelles. Un autre type plus rare de corrélation des résidus retrouvé dans les coupes instantanées est appelé *corrélation spatiale*¹. Nous allons cependant nous consacrer ici au premier type de corrélation des résidus. Nous envisagerons d'abord les propriétés de l'estimateur des MCO lorsque les résidus sont autocorrélés. Puis nous verrons comment on peut traiter ce problème à partir des moindres carrés généralisés.

1. PROPRIÉTÉS DE L'ESTIMATEUR DES MCO LORSQUE LES RÉSIDUS SONT AUTOCORRÉLÉS

Il y a essentiellement deux causes classiques au phénomène de l'autocorrélation. D'abord, l'omission d'une variable pertinente au modèle envisagé ou l'addition de variables inadéquates peut causer l'autocorrélation. Ensuite, le traitement des données effectué par les agences statistiques pourrait se traduire par l'autocorrélation des erreurs résiduelles lorsque l'on estime un modèle.

Considérons maintenant un modèle de régression linéaire appliqué à des séries temporelles. Ce modèle s'écrit :

$$y_t = \mathbf{x}_t^T \boldsymbol{\beta} + e_t$$

-
1. Voir à cet effet : Racicot, F.-É. (2000), *Notes on Nonlinear Dynamics*, CRG, *op. cit.* On consultera aussi : Jonhston, J. (1988), *Méthodes économétriques*, tome 2, Economica, Paris.

où t représente le temps. Nous introduisons ici une modification au processus suivi par l'innovation par rapport au modèle classique. En effet, on postule que e_t suit un processus autorégressif d'ordre p , désigné par $AR(p)$. Nous nous limiterons dans cette section au processus autorégressif d'ordre 1, c'est-à-dire : $p = 1$. Ce processus stochastique se définit ainsi :

$$e_t = \rho e_{t-1} + \varepsilon_t$$

où ε_t est un bruit blanc (*white noise*) et où $|\rho| < 1$. Considérons d'abord certains aspects du processus $AR(1)$ avant de nous déplacer vers les propriétés de l'estimateur MCO comme tel en présence d'auto-corrélation.

Dans le processus $AR(1)$, nous avons postulé que ε_t est un bruit blanc. Cela signifie que :

$$\begin{aligned} E(\varepsilon_t) &= 0 \\ V(\varepsilon_t) &= \sigma_\varepsilon^2 \\ \text{Cov}(\varepsilon_t, \varepsilon_s) &= 0 \quad \forall t, s \end{aligned}$$

Avoir supposé que $|\rho| < 1$ implique que le processus stochastique de e_t est stationnaire. La stationnarité d'une série signifie entre autres que sa moyenne et sa variance sont constantes et que l'autocovariance ne dépend que de la distance entre les réalisations. Nous reconsidérerons plus en détails ces allégations dans le prochain chapitre.

Pour dégager les propriétés de l'estimateur des MCO en présence d'autocorrélation, écrivons un modèle simple qui ne renferme qu'une seule variable explicative, soit :

$$y_t = \beta_1 + \beta_2 x_{2t} + e_t \quad t = 1, \dots, T$$

où ε_t suit un processus $AR(1)$. La matrice variance-covariance des résidus se présente comme suit :

$$E(\mathbf{e}\mathbf{e}^T) = \mathbf{V} = \sigma_\varepsilon^2 \boldsymbol{\Omega} = \sigma_\varepsilon^2 \frac{1}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \cdot & \cdot & \rho^{T-1} \\ \rho & 1 & \rho & \cdot & \cdot & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdot & \cdot & \rho^{T-3} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho^{T-1} & \rho^{T-2} & \cdot & \cdot & \cdot & 1 \end{bmatrix}$$

Voyons comment se calculent l'un des éléments de la diagonale principale et un autre hors-diagonale. Les éléments sur la diagonale sont les variances de e_t . On a :

$$V(e_t) = V(\rho e_{t-1} + \varepsilon_t) = \rho^2 V(e_{t-1}) + V(\varepsilon_t)$$

puisque $\text{Cov}(e_{t-1}, \varepsilon_t) = 0$. Comme nous avons supposé que le processus est stationnaire, on a :

$$\begin{aligned} V(e_t) &= \rho^2 V(e_t) + \sigma_\varepsilon^2 \\ \Rightarrow V(e_t) &= \frac{\sigma_\varepsilon^2}{1-\rho^2} = \sigma_e^2 \end{aligned}$$

Pour obtenir les éléments hors-diagonale, nous avons calculé l'autocovariance qui est requise pour le calcul de l'autocorrélation.

$$\begin{aligned} \text{Cov}(e_t, e_{t-1}) &= E[(e_t - E(e_t))(e_{t-1} - E(e_{t-1}))] = E(e_t e_{t-1}) \\ &= E((\rho e_{t-1} + \varepsilon_t)e_{t-1}) = \rho E(e_{t-1})^2 + E(\varepsilon_t e_{t-1}) = \rho \sigma_e^2 = \rho \frac{\sigma_\varepsilon^2}{1-\rho^2} \end{aligned}$$

Il en résulte la formule suivante pour l'autocorrélation :

$$\text{Corr}(e_t, e_{t-1}) = \frac{\text{Cov}(e_t, e_{t-1})}{V(e_t)} = \frac{\rho\sigma_e^2}{\sigma_e^2} = \rho$$

Par conséquent, ρ représente la corrélation entre l'erreur d'une période à l'autre et est donc désigné : coefficient d'autocorrélation. Par ailleurs, l'autocovariance entre e_t et e_{t-k} est égale à : $\rho^k\sigma_e^2$ et est obtenue par substitutions répétées. On peut donc écrire :

$$\text{Corr}(e_t, e_{t-k}) = \rho^k$$

pour $k = 1, 2, \dots$

Par rapport au modèle classique des MCO, la matrice variance-covariance des résidus n'est plus égale à $\sigma^2\mathbf{I}$ mais à $\sigma_e^2\mathbf{\Omega}$. Comme dans le cas de l'hétéroscédasticité, en présence d'autocorrélation, l'estimateur des MCO de β est sans biais, c'est-à-dire : $E(\hat{\beta}) = \beta$. Mais cet estimateur est inefficace puisque la variance de $\hat{\beta}$ qui découle de l'estimateur des MCO est égale à : $\sigma_e^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Omega}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}$, ce qui est certes différent de $\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$. La variance donnée par les MCO comporte donc un biais, d'où l'inefficacité de cet estimateur.

Revenons au modèle simple à une variable où le terme d'erreur suit un processus AR(1). Nous voulons démontrer comment traiter le problème d'autocorrélation dans ce cas. Substituons la valeur de e_t dans ce modèle. Nous avons :

$$y_t = \beta_1 + \beta_2 x_{2t} + \rho e_{t-1} + \varepsilon_t$$

où $e_{t-1} = y_{t-1} - \beta_1 - \beta_2 x_{2,t-1}$. En remplaçant e_{t-1} par sa valeur, on a :

$$y_t - \rho y_{t-1} = \beta_1(1 - \rho) + \beta_2(x_{2t} - \rho x_{2,t-1}) + \varepsilon_t$$

Une telle transformation est appelée : *transformation en quasi-différences*. Une telle équation peut être estimée en recourant aux moindres carrés non linéaires. En effectuant les regroupements suivants, cette dernière s'écrit :

$$y_t^* = \beta_1 x_{1t}^* + \beta_2 x_{2t}^* + \varepsilon_t \quad t = 2, 3, \dots, T$$

où ε_t est un bruit blanc. Si on applique les MCO sur ce dernier modèle, on obtient un estimateur BLUE pour β quand ρ est connu. Par contre, deux problèmes subsistent. En premier lieu, la transformation en quasi-différences se traduit par une perte d'information, en l'occurrence y_1^* et x_{21}^* . En deuxième, la valeur de ρ n'est généralement pas connue. Pour pallier à ce dernier problème, Cochrane et Orcutt (1949)² ont suggéré la procédure itérative suivante qui fait litière de la première observation. Soit les deux équations :

$$y_t - \rho y_{t-1} = \beta_1(1 - \rho) + \beta_2(x_{2t} - \rho x_{2,t-1}) + \varepsilon_t$$

$$y_t - \beta_1 - \beta_2 x_{2t} = \rho(y_{t-1} - \beta_1 - \beta_2 x_{2,t-1}) + \varepsilon_t$$

La première étape de la procédure itérative consiste à postuler une valeur initiale hypothétique pour $\hat{\rho}$, désignée par $\hat{\rho}_{(1)}$. On obtient les valeurs : $y_t - \hat{\rho}_{(1)}y_{t-1}$ et $x_{2t} - \hat{\rho}_{(1)}x_{2,t-1}$ que l'on substitue dans la première équation. On applique les MCO sur cette équation et l'on obtient : $\hat{\beta}_{1(1)}$ et $\hat{\beta}_{2(1)}$. Ceci constitue la première étape de l'itération. Dans une deuxième itération, on substitue ces deux coefficients dans la deuxième équation et l'on obtient, en appliquant les MCO, une seconde valeur de ρ , désignée par $\hat{\rho}_{(2)}$. Le processus itératif se termine quand on note la convergence, c'est-à-dire quand les coefficients estimés ne varient plus sensiblement d'une régression à l'autre. Comme cette méthode néglige la première observation, elle peut donner lieu à des estimateurs relativement inefficients dans les échantillons restreints. Ce problème tend à s'estomper à mesure que la taille de l'échantillon augmente.

Prais et Winsten (1954) ont proposé de prendre en compte la première observation en utilisant la procédure suivante. On calcule comme suit la première observation :

$$\sqrt{1 - \rho^2} y_1 = \beta_1 \sqrt{1 - \rho^2} + \beta_2 \sqrt{1 - \rho^2} x_{21} + \sqrt{1 - \rho^2} e_1$$

2. En présence d'erreurs de mesure sur les variables, la procédure de Cochrane et Orcutt et celle de Prais et Winsten pourraient s'avérer perverses, en ce sens qu'elles amplifieraient les problèmes causés par les erreurs de mesure. Pour plus de détails, voir : Dagenais, M.G. (1994), « Parameter Estimation in Regression Models with Errors in the Variables and Autocorrelated Disturbances », *Journal of Econometrics*, vol. 64, p. 145-163.

En transposant le tout sous forme matricielle, on obtient :

$$\mathbf{y}^* = \mathbf{X}^* \boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

où :

$$\mathbf{y}^* = \begin{bmatrix} \sqrt{1-\rho^2} y_1 \\ y_2 - \rho y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_T - \rho y_{T-1} \end{bmatrix} = \mathbf{P} \mathbf{y} \quad \mathbf{X}^* = \begin{bmatrix} \sqrt{1-\rho^2} & \sqrt{1-\rho^2} x_{21} \\ 1-\rho & x_{22} - \rho x_{21} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1-\rho & x_{2T} - \rho x_{2,T-1} \end{bmatrix} = \mathbf{P} \mathbf{X}$$

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \sqrt{1-\rho^2} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_T \end{bmatrix} = \mathbf{P} \mathbf{e}$$

où $\mathbf{P} = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & \cdot & \cdot & \cdot & 0 \\ -\rho & 1 & 0 & \cdot & \cdot & 0 \\ 0 & -\rho & 1 & 0 & \cdot & 0 \\ \cdot & & & \cdot & & \cdot \\ \cdot & & & & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & -\rho & 1 \end{bmatrix}_{T \times T}$

En appliquant les MCO sur les variables transformées par \mathbf{P} , on obtient l'estimateur BLUE de β :

$$\begin{aligned} \hat{\beta} &= (\mathbf{X}^{*T} \mathbf{X}^*)^{-1} \mathbf{X}^{*T} \mathbf{y}^* = (\mathbf{X}^T \mathbf{P}^T \mathbf{P} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{P}^T \mathbf{P} \mathbf{y} \\ &= (\mathbf{X}^T \mathbf{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{\Omega}^{-1} \mathbf{y} = \hat{\beta}_{GLS} \end{aligned}$$

où $\mathbf{\Omega}^{-1} = \mathbf{P}^T \mathbf{P}$. On est ici en présence de l'estimateur des moindres carrés généralisés pour un r donné. La variance de cet estimateur est de :

$$V(\hat{\beta}_{GLS}) = \sigma_\varepsilon^2 (\mathbf{X}^T \mathbf{\Omega}^{-1} \mathbf{X})^{-1}$$

Dans cette équation, on estime σ_ε^2 par :

$$\hat{\sigma}_\varepsilon^2 = \frac{(\mathbf{y}^* - \mathbf{X}^* \hat{\beta}_{GLS})^T (\mathbf{y}^* - \mathbf{X}^* \hat{\beta}_{GLS})}{T - k} \text{ où } k = 2 \text{ pour le cas qui nous intéresse.}$$

En résumé, si ρ est connu, en remplaçant \mathbf{y} , \mathbf{X} et \mathbf{e} par \mathbf{y}^* , \mathbf{X}^* et \mathbf{e}^* dans le modèle ci-haut présenté et en appliquant les MCO sur le modèle ainsi transformé, on obtient l'estimateur des moindres carrés généralisés, qui est l'estimateur BLUE de β . Quant à eux, les tests d'hypothèses et les intervalles de confiance valent toujours tant et

aussi longtemps que les estimateurs MCG de β et de σ_ε^2 sont utilisés. Par contre, si ρ est inconnu et estimé, alors les tests ne sont valables qu'asymptotiquement. Le R^2 souffre du même problème que dans le cas de l'hétéroscédasticité. Après correction pour l'autocorrélation, le R^2 tendra à diminuer. La procédure proposée pour corriger ce problème est la même que celle établie dans le cas du problème de l'hétéroscédasticité, qui, on le rappelle, est de calculer le R^2 comme suit :

$$R_{\text{GLS}}^2 = 1 - \frac{(\mathbf{y} - \mathbf{X}\hat{\beta}_{\text{GLS}})^T (\mathbf{y} - \mathbf{X}\hat{\beta}_{\text{GLS}})}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

où \mathbf{y} et \mathbf{X} sont les variables non transformées.

2. CORRECTION DU MODÈLE ORIGINAL LORSQUE ρ N'EST PAS CONNU

Tel que mentionné, la méthode Cochrane-Orcut est un algorithme itératif pour déterminer la valeur des paramètres du modèle de régression avec autocorrélation. La méthode de Prais et Winsten, quant à elle, tient compte de la première observation. Une façon d'enclencher l'algorithme est de prendre comme ρ initial celui qui résulte de la régression suivante :

$$\hat{e}_t = \rho \hat{e}_{t-1} + \xi_t$$

L'estimateur de ρ , biaisé mais convergent, est donné par :

$$\hat{\rho} = \frac{\sum_{t=2}^T \hat{e}_t \hat{e}_{t-1}}{\sum_{t=2}^T \hat{e}_{t-1}^2}$$

qui est le coefficient d'autocorrélation estimé entre e_t et e_{t-1} . Bien qu'il soit préférable de se servir de cet estimé comme point de départ à l'algorithme itératif Cochrane-Orcut, de façon à simplifier l'exposé,

nous substituons directement cette valeur dans la formule des MCG de manière à obtenir les moindres carrés quasi généralisés (MCQG)³, ce qui est une procédure qui est également acceptable. On obtient :

$$\hat{\beta}_{\text{MCQG}} = (\mathbf{X}^T \hat{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\Omega}^{-1} \mathbf{y}$$

où $V(\hat{\beta}) = \sigma_\varepsilon^2 (\mathbf{X}^T \hat{\Omega}^{-1} \mathbf{X})^{-1}$ et où l'estimateur de σ_ε^2 est :

$$\hat{\sigma}_\varepsilon^2 = \frac{\left(\mathbf{y}^* - \mathbf{X}^* \hat{\beta}_{\text{MCQG}} \right)^T \left(\mathbf{y}^* - \mathbf{X}^* \hat{\beta}_{\text{MCQG}} \right)}{T - k}$$

Procédures récentes pour la correction de l'autocorrélation

Comme nous l'avons mentionné antérieurement, le modèle de régression classique peut s'écrire comme suit en présence d'autocorrélation :

$$y_t = (1 - \rho)\beta_1 + \rho y_{t-1} + \beta_2 x_{2t} - \rho \beta_2 x_{2,t-1} + \varepsilon_t \quad t = 2, 3, \dots, T$$

Cette équation contient des produits de paramètres, ce qui implique qu'elle est non linéaire. On peut donc recourir aux moindres carrés non linéaires (NLS) pour estimer cette équation :

$$\text{Min}_{\beta_1, \beta_2, \rho} S^*(\beta_1, \beta_2, \rho) = \text{Min}_{\beta_1, \beta_2, \rho} \sum_{t=2}^T \left(y_t - (1 - \rho)\beta_1 - \rho y_{t-1} - \beta_2 x_{2t} + \rho \beta_2 x_{2,t-1} \right)^2$$

Cet estimateur, on le rappelle, ne comporte pas de solution analytique. Pour estimer les paramètres, on aura recours à l'algorithme d'optimisation dont il fut question dans la section des moindres carrés non linéaires. Si les résidus sont normalement distribués, l'estimateur des moindres carrés non linéaires se confondra avec celui du maximum de vraisemblance. Sa performance équivaudra alors à celle des FGLS.

3. Soit les FGLS en anglais, ce qui est l'abréviation de : *Feasible Generalized Least Squares*.

La méthode d'intégration de la première observation développée par Prais et Winsten peut être incorporée dans le modèle de régression non linéaire. On obtient alors :

$$S(\beta_1, \beta_2, \rho) = S^*(\beta_1, \beta_2, \rho) + (e^*_{1})^2$$

où $(e^*_{1})^2 = (y^*_{1} - \beta_1 x^*_{11} - \beta_2 x^*_{21})^2$. Pour déterminer la valeur des paramètres, on minimise la fonction S en ayant recours à l'algorithme d'optimisation. Dans les cas où le terme d'erreur suit un processus AR(p), il existe une procédure en deux étapes pour corriger l'autocorrélation.

3. TESTS D'AUTOCORRÉLATION

Pour détecter la présence d'autocorrélation d'ordre 1, le test Durbin-Watson (1951) demeure le plus populaire. La statistique associée à ce test est :

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

Pour comprendre pourquoi d est une statistique pertinente pour tester l'autocorrélation, on réécrit d comme suit :

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

En éclatant cette somme en ses composantes, on a :

$$d = \frac{\sum_{t=2}^T \hat{e}_t^2}{\sum_{t=1}^T \hat{e}_t^2} + \frac{\sum_{t=2}^T \hat{e}_{t-1}^2}{\sum_{t=1}^T \hat{e}_t^2} - 2 \frac{\sum_{t=2}^T \hat{e}_t \hat{e}_{t-1}}{\sum_{t=1}^T \hat{e}_t^2} \cong 1 + 1 - 2\hat{\rho} = 2(1 - \hat{\rho})$$

Si $\hat{\rho} = 0$, il en résulte donc que d est environ égal à 2 et si $\hat{\rho} = 1$, d est approximativement égal à 0. Le calcul de la statistique d nous indique donc si les résidus sont autocorrélés ou non. Il reste que cette démarche s'avère plus ou moins rigoureuse.

La procédure classique pour tester la présence d'autocorrélation est la suivante.

- i) Soit d_u la borne supérieure du test et d_L , la borne inférieure, le test comportant en effet deux bornes. Soit $H_0: \rho = 0$ et $H_1: \rho \neq 0$. On a les cas suivants :
 - a) $d_u < d < 4 - d_u \rightarrow$ ne pas rejeter H_0 ;
 - b) $0 < d < d_L \rightarrow$ rejeter $H_0 \Rightarrow \rho > 0$;
 - c) $4 - d_L < d < 4 \rightarrow$ rejeter $H_0 \Rightarrow \rho < 0$;
 - d) $d_L < d < d_u$ ou $4 - d_u < d < 4 - d_L \rightarrow$ zone d'indétermination

Il faut cependant remarquer que dans le cas où la régression comporte, parmi les variables explicatives, la variable dépendante retardée (y_{t-1}) et que les résidus sont autocorrélés d'ordre 1, la statistique d est alors biaisée vers 2. Elle ne peut donc être utilisée directement pour tester l'autocorrélation. La statistique h développée par Durbin (1970) doit alors être utilisée. Celle-ci se définit alors comme suit :

$$h = \hat{\rho} \sqrt{\frac{T}{1 - T\hat{V}(\hat{\beta}_2)}}$$

où $\hat{\rho}$ est le coefficient d'autocorrélation du processus AR(1) calculé à l'aide des résidus des MCO et $\hat{V}(\hat{\beta}_2)$ est l'estimé de la variance du coefficient de y_{t-1} . Sous H_0 , la distribution de h est de : $h \underset{d}{\sim} N(0, 1)$. Si l'on détecte de l'autocorrélation en présence d'une variable dépendante retardée, l'estimateur des MCO est alors biaisé et non convergent : le terme d'erreur n'est pas dans ce cas orthogonal à y_{t-1} . Il faut alors recourir à d'autres méthodes pour estimer une telle régression⁴.

4. Pour plus de détails, on consultera : Judge, G.G. *et al.*, (1985), *The Theory and Practice of Econometrics*, John Wiley and Sons, New York.

4. PRÉVISION DANS LE MODÈLE LINÉAIRE AVEC ERREUR DE LA FORME AR(1)

Dans les chapitres précédents, nous avons étudié comment prévoir à l'intérieur du modèle linéaire. Nous suivons la même procédure ici en venant toutefois nous situer dans le domaine temporel. Quand les résidus sont autocorrélés, l'estimateur des MCO de y_0 , désigné par \hat{y}_0 , souffre d'inefficience. On recourra alors aux moindres carrés généralisés (MCQG). Pour fixer les idées, reprenons le modèle linéaire de base :

$$y_t = \mathbf{x}_t^T \boldsymbol{\beta} + e_t, \quad t = 1, \dots, T$$

où $e_t = \rho e_{t-1} + \varepsilon_t$, où $\varepsilon_t \sim \text{WN}(0, \sigma_\varepsilon^2)$ et où $|\rho| < 1$, soit l'hypothèse de stationnarité. Si l'on veut prédire y_{T+1} , on peut écrire :

$$y_{T+1} = \mathbf{x}_{T+1}^T \boldsymbol{\beta} + \rho e_T + \varepsilon_{T+1}$$

Pour obtenir la prévision, on substitue $\hat{\boldsymbol{\beta}}_{\text{MCQG}}$ à $\boldsymbol{\beta}$, $\hat{\rho}$ à ρ et \tilde{e}_T à e_T . On trouve \tilde{e}_T comme suit :

$$\tilde{e}_T = y_T - \mathbf{x}_T^T \hat{\boldsymbol{\beta}}_{\text{MCQG}}$$

On dégage la prévision suivante de y_{T+1} :

$$\hat{y}_{T+1} = \mathbf{x}_{T+1}^T \hat{\boldsymbol{\beta}}_{\text{MCQG}} + \hat{\rho} \tilde{e}_T$$

$\hat{\rho} \tilde{e}_T$ tenant lieu de prévision de e_{T+1} , qui incidemment n'est valable qu'asymptotiquement.

Si l'on désire effectuer une prévision plusieurs périodes à l'avance, par exemple h périodes, on procède de la façon suivante. Il faut pour ce faire calculer l'espérance conditionnelle du terme d'erreur du modèle, qui est égale à :

$$\hat{e}_{T+h} = E_T(e_{T+h}) = \rho^h e_T + \rho^{h-1} E_T(\varepsilon_{T+1}) + \dots + E_T(\varepsilon_{T+h})$$

Cette équation est obtenue par substitutions répétées. Comme l'espérance des ε_{T+j} est égale à 0, on obtient la prévision de y_{T+h} qui est de :

$$\hat{y}_{T+h} = E_T(y_{T+h}) = \mathbf{x}_{T+h}^T \hat{\boldsymbol{\beta}}_{\text{MCQG}} + \hat{\rho}^h \tilde{e}_T$$

La variance de cette prévision (σ_{AF}^2), qui vaut asymptotiquement, se calcule comme suit :

$$\hat{\sigma}_{AF}^2 = \hat{\sigma}_F^2 + \left(\hat{\sigma}_\varepsilon^2 h^2 \hat{\rho}^{2(h-1)} / T \right)$$

où :

$$\hat{\sigma}_F^2 = \hat{\sigma}_\varepsilon^2 \left[(1 - \hat{\rho}^{2h}) / (1 - \hat{\rho}^2) \right] + \hat{\sigma}_\varepsilon^2 (\mathbf{x}_{T+h} - \hat{\rho}^h \mathbf{x}_T)^T (\mathbf{X}^T \hat{\Omega}^{-1} \mathbf{X})^{-1} (\mathbf{x}_{T+h} - \hat{\rho}^h \mathbf{x}_T)$$

5. APPLICATIONS

Nous recourons de nouveau au modèle de l'impact des jours de la semaine sur les indices boursiers que nous avons introduit dans la section sur les variables auxiliaires. Nos résultats sont compilés au tableau 8.1.

TABLEAU 8.1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.043659	0.080335	0.543455	0.5871
M72	-0.319821	0.117947	-2.711565	0.0069
T72	0.028880	0.114167	0.252962	0.8004
TH72	0.015556	0.113887	0.136589	0.8914
F72	-0.017636	0.113887	-0.154857	0.8770
R-squared	0.023408	Mean dependent var	-0.008239	
Ajusted R-squared	0.015548	S.D. dependent var	0.825708	
S.E. of regression	0.819264	Akaike info criterion	-0.388788	
Sum squared resid	333.5829	Schwarz criterion	-0.346770	
Log likelihood	-609.7212	F-statistic	2.978156	
Durbin-Watson stat	1.562695	Prob(F-statistic)	0.018937	

La statistique Durbin-Watson (d) est de 1,56, ce qui, en vertu des tables qui apparaissent à la fin de ce volume, donne lieu au rejet de l'hypothèse H_0 (absence d'autocorrélation). d est en effet inférieur à d_L qui est de 1,73 pour un nombre de degrés de liberté $k^5 = 4$, c'est-à-dire 4 variables explicatives excluant la constante, au niveau de confiance de 5 % et pour un nombre d'observations (T) de 200. Cette conclusion vaut toujours si nous excluons les variables qui ne sont pas significatives dans le tableau 8.1. Dans le logiciel EViews, il est suggéré une procédure simple de correction du problème d'autocorrélation, à savoir rajouter à la régression la variable dépendante retardée et réestimer le modèle. On vérifie alors, par le biais de la statistique h , s'il y a encore autocorrélation, tout en tenant compte du problème qui se crée s'il y a autocorrélation. S'il n'y a pas autocorrélation, on conserve alors la variable dépendante retardée comme variable explicative du modèle.

5. Nous mettons un ' à k car le test exclut la constante du modèle.

LES SÉRIES TEMPORELLES

La modélisation des séries chronologiques vise la prévision à court terme sans s'intéresser au caractère structurel des modèles économiques et financiers, c'est-à-dire les modèles qui expliquent le comportement des variables économiques et financières. La modélisation vise simplement le *data-fitting*, c'est-à-dire l'ajustement d'une série à ses propres observations de façon à maximiser la vraisemblance de cet échantillon. La modélisation en séries temporelles remonte à l'étude publiée par Box et Jenkins¹ en 1976. Bien que les économètres aient alors accueilli cet ouvrage avec une certaine froideur, les techniques reliées aux séries temporelles ont par la suite envahi le champ de l'économétrie financière. L'ouvrage de Hamilton² (1994) prend acte de la complexité qu'a atteint cette branche de l'économétrie.

1. PROCESSUS STOCHASTIQUES

Un processus stochastique est une suite de variables aléatoires définie sur un même espace W , appelé *espace fondamental* ou *espace des états de la nature*. Un processus stochastique se formule comme suit :

$$y = (y_t, t \in \mathbb{N})$$

1. Box, G.E.P. et G.M. Jenkins (1976), *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.
2. Hamilton, J.D. (1994), *Time Series Analysis*, Princeton University Press, New Jersey.

où \mathbb{N} désigne l'ensemble des nombres naturels. Nous recourons à la notation plus compacte suivante :

$$y = \{y_t\}$$

Par exemple : $\{y_1, y_2, \dots, y_T\}$ est une réalisation du processus stochastique $\{y_t\}$. Une telle réalisation est souvent appelée *trajectoire* du processus.

1.1. Stationnarité

Il sera ici question de *stationnarité faible* ou *stationnarité de second ordre*. Un processus $\{y_t\}$ est dit stationnaire si et seulement si :

- i) $E(y_t) = E(y_{t-s}) = \mu$, c'est-à-dire que la moyenne du processus est de μ , donc constante ;
- ii) $E(y_t - \mu)^2 = E(y_{t-s} - \mu)^2 = \sigma_y^2$, c'est-à-dire que la variance est invariable dans le temps ;
- iii) $E[(y_t - \mu)(y_{t-s} - \mu)] = E[(y_{t-j} - \mu)(y_{t-s-j} - \mu)] = \gamma_s$. L'auto-covariance ne dépend donc que de la distance entre deux points dans le temps et non d'une date particulière.

1.2. Processus autorégressifs stationnaires : représentation et estimation

Un processus autorégressif d'ordre p , noté $AR(p)$, se définit comme suit :

$$y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + e_t \quad t = 1, \dots, T$$

où t dénote le temps $e_t \sim WN(0, \sigma_e^2)$, et WN est un bruit blanc homoscédastique, c'est-à-dire que $\text{Cov}(e_t, e_s) = 0$ pour tout t et s , $E(e_t) = E(e_s) = 0$ pour tout t et s et $V(e_t) = V(e_s) = \sigma_e^2$ pour tout t et s . Une autre représentation de $AR(p)$ couramment utilisée est celle qui recourt à l'opérateur de retard L , qui retarde une variable d'une période :

$$\theta(L)y_t = \delta + e_t$$

où $\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_p L^p$. Voici quelques propriétés de l'opérateur L :

- i) $Lc = c$, où c est une constante ;
- ii) $L \times L \times \dots \times L = L^n$;
- iii) $L^n y_t = y_{t-n}$
- iv) $(1 - L)^{-1} = 1 + L + L^2 + \dots$, ce qui est la formule de la progression géométrique de raison L .

On peut exprimer AR(p) sous la forme d'une moyenne mobile infinie :

$$y_t = \frac{\delta}{\theta(L)} + \frac{e_t}{\theta(L)}$$

Par exemple, dans le cas particulier d'un AR(1), on a :

$$y_t = \frac{\delta}{1 - \theta_1} + \sum_{i=0}^{\infty} e_t (L\theta_1)^i$$

Nous abordons maintenant les propriétés statistiques du processus AR(1). L'espérance du processus AR(1) est la suivante :

$$E(y_t) = \delta + \theta_1 E(y_{t-1}) + E(e_t)$$

Et si le processus est stationnaire, alors $E(y_t) = E(y_{t-1})$, ce qui implique :

$$E(y_t) = \delta + \theta_1 E(y_t) + 0 = \frac{\delta}{1 - \theta_1} = \mu$$

Pour que $E(y_t)$ soit finie, il faut que $\theta_1 \neq 1$. En fait, pour respecter les conditions de stationnarité, il faut que : $0 < \theta < 1$.

Par ailleurs, la variance de y_t est égale à :

$$V(y_t) = \sigma_y^2 = \theta_1^2 V(y_{t-1}) + V(e_t) = \theta_1^2 V(y_t) + \sigma_e^2$$

D'où,

$$V(y_t) = \frac{\sigma_e^2}{1 - \theta_1^2} = \sigma_y^2$$

L'autocovariance entre y_t et y_{t-1} se calcule comme suit :

$$\text{Cov}(y_t, y_{t-1}) = E[(y_t - \mu)(y_{t-1} - \mu)]$$

Pour simplifier les calculs, supposons que μ est égal à 0. En remplaçant y_t par sa valeur, on obtient :

$$\begin{aligned} \text{Cov}(y_t, y_{t-1}) &= E(y_t y_{t-1}) = E[(\theta_1 y_{t-1} + e_t) y_{t-1}] \\ &= \theta_1 E(y_{t-1})^2 + E(e_t y_{t-1}) = \theta_1 E(y_{t-1})^2 + 0 = \theta_1 \sigma_y^2 = \theta_1 \left(\frac{\sigma_e^2}{1 - \theta_1^2} \right) \end{aligned}$$

La covariance entre y_t et y_{t-k} est égale à $\theta_1^k \sigma_y^2 = \gamma_k$.

À partir de la fonction d'autocovariance, on peut définir la fonction d'autocorrélation (ACF)³ comme suit :

$$\begin{aligned} \rho_k &= \text{corr}(y_t, y_{t-k}) = \frac{\text{Cov}(y_t, y_{t-k})}{\sqrt{V(y_t)V(y_{t-k})}} \\ &= \frac{\text{Cov}(y_t, y_{t-k})}{V(y_t)} = \frac{\gamma_k}{\gamma_0} = \frac{\theta_1^k \sigma_y^2}{\sigma_y^2} = \theta_1^k \end{aligned}$$

$k = 0, \pm 1, \pm 2, \dots$ On constate que la corrélation s'atténue avec k , soit la mesure de la distance. La fonction d'autocorrélation est symétrique, c'est-à-dire que : $\rho_{-k} = \rho_k$. Comme $\rho_0 = 1$ et que la fonction d'autocorrélation est symétrique, on utilisera $k \geq 1$ lorsqu'on analyse une fonction d'autocorrélation.

2. ESTIMATION DU PROCESSUS AUTORÉGRESSIF AR(p)

On peut réécrire le processus AR(p) sous la forme suivante :

$$\theta(L)y_t - \delta = e_t$$

3. ACF est l'abréviation de *Autocorrelation function* en anglais.

Pour estimer les paramètres du polynôme $\theta(L)$ et du paramètre δ , il suffit de minimiser la fonction suivante :

$$\text{Min}_{\theta, \delta} S(\theta, \delta) = \text{Min}_{\theta, \delta} \sum_{t=1}^T e_t^2$$

La minimisation de cette somme nous ramène à la méthode des MCO. Cependant, dans le contexte de ce processus autorégressif, l'estimateur des MCO est biaisé mais convergent. De façon à illustrer ce problème de biais et pour expliciter davantage comment on estime ce type de modèle, on réécrit le processus AR(p) pour chaque observation de la façon suivante :

$$\begin{aligned} y_{p+1} &= \delta + \theta_1 y_p + \theta_2 y_{p-1} + \dots + \theta_p y_{p-(p-1)} + e_{p+1} \\ y_{p+2} &= \delta + \theta_1 y_{p+1} + \theta_2 y_p + \dots + \theta_p y_2 + e_{p+2} \\ &\cdot \qquad \qquad \qquad \cdot \\ &\cdot \qquad \qquad \qquad \cdot \\ &\cdot \qquad \qquad \qquad \cdot \\ y_T &= \delta + \theta_1 y_{T-1} + \theta_2 y_{T-2} + \dots + \theta_p y_{T-p} + e_T \end{aligned}$$

Sous forme matricielle, ce système s'écrit :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

Considérons le cas d'un processus AR(1). Sous forme matricielle, on a alors :

$$\mathbf{y} = \begin{bmatrix} y_2 \\ y_3 \\ \cdot \\ \cdot \\ \cdot \\ y_T \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & y_1 \\ 1 & y_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & y_{T-1} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \delta \\ \theta_1 \end{bmatrix} \quad \mathbf{e} = \begin{bmatrix} e_2 \\ e_3 \\ \cdot \\ \cdot \\ \cdot \\ e_T \end{bmatrix}$$

On remarque que la variable explicative y_2 est reliée à e_2 , que la variable explicative y_3 est reliée à e_3 et ainsi de suite. En appliquant les MCO sur ce processus, on obtient évidemment que :

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

Or,

$$\text{plim } \hat{\beta} = \beta$$

Mais $E(\mathbf{X}^T \mathbf{e}) \neq 0 \Rightarrow E(\hat{\beta}) \neq \beta$. Les MCO sont par conséquent biaisés. Ce cas se généralise facilement à celui du processus AR(p).

Dans le modèle AR(p), la variance de $\hat{\beta}$ est donnée par :

$$\hat{V}(\hat{\beta}) = \hat{\sigma}_e^2 (\mathbf{X}^T \mathbf{X})^{-1}$$

où $\hat{\sigma}_e^2$ est égal à :

$$\hat{\sigma}_e^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\beta})^T (\mathbf{y} - \mathbf{X}\hat{\beta})}{T - 2p - 1}$$

3. FONCTION D'AUTOCORRÉLATION PARTIELLE (PACF)⁴

Pour déterminer l'ordre p d'un processus autorégressif, on recourt à la fonction d'autocorrélation partielle. Dans un processus AR(2), θ_2 évalue l'autocorrélation partielle notée par θ_{22} qui peut être estimée par les MCO. Si l'on estime un processus AR(3) alors que le vrai processus est un AR(2), on devrait observer que θ_3 n'est pas significatif. Pour un modèle AR(p), l'autocorrélation partielle d'ordre p , notée par θ_{pp} , est estimée par $\hat{\theta}_p$.

Plus précisément, l'autocorrélation partielle θ_{pp} mesure l'association linéaire entre y_t et y_{t-p} , compensée pour l'effet des autres variables, soit y_{t-1} , y_{t-2} jusqu'à y_{t-p-1} . La fonction d'autocorrélation

4. PACF est l'abréviation anglaise de *partial autocorrelation function*.

partielle est la suite $\{\theta_{11}, \theta_{22}, \dots, \theta_{pp}\}$. La figure 9.1 représente le graphique d'une PACF pour un AR(1) d'une série y_1 simulée dans le logiciel EViews.

FIGURE 9.1

Nombre d'observations : 100

Autocorrélation	Corrélation partielle		AC	PAC	Q-stat	Prob.
		1	0.882	0.882	80.153	0.000
		2	0.781	0.015	143.67	0.000
		3	0.706	0.062	196.04	0.000
		4	0.593	-0.196	233.45	0.000
		5	0.535	0.177	264.21	0.000
		6	0.474	-0.070	288.58	0.000
		7	0.415	0.047	307.49	0.000
		8	0.361	-0.100	321.93	0.000
		9	0.296	-0.019	331.73	0.000
		10	0.256	0.043	339.18	0.000
		11	0.185	-0.166	343.10	0.000
		12	0.092	-0.145	344.08	0.000
		13	0.021	-0.048	344.13	0.000
		14	-0.055	-0.022	344.49	0.000
		15	-0.146	-0.186	347.05	0.000
		16	-0.185	0.130	351.21	0.000
		17	-0.194	0.089	355.81	0.000
		18	-0.204	0.043	360.97	0.000

Les droites en pointillés verticales représentent une marge de deux fois l'erreur.

AC : coefficient d'autocorrélation

PAC : coefficient d'autocorrélation partielle

Q-stat : statistique de Box-Pierce-Ljung [16.581]

Prob. : probabilité de l'hypothèse que tous les coefficients d'autocorrélation soient nuls en ce point.

Source : Johnston, J. et J. Dinardo (1997), *op. cit.*, p. 217.

L'ordre p d'un modèle AR(p) sera choisi de telle sorte que $\theta_{kk} \neq 0$ pour $k = p$ et $\theta_{kk} = 0$ pour $k > p$. Donc, si θ_{kk} est significativement différent de 0 pour $p = k$ et égal à 0 pour $k > p$, on peut alors déterminer l'ordre par le test suivant :

$$H_0 : \theta_{kk} = 0 \text{ pour } k > p$$

$$H_1 : \theta_{kk} \neq 0 \text{ pour } k = p$$

Sous l'hypothèse nulle, on sait que : $\hat{\theta}_{kk}$ est asymptotiquement normalement distribué avec moyenne 0 et variance égale à : $V(\hat{\theta}_{kk}) = \frac{1}{T}$. On peut alors construire la statistique t asymptotique suivante :

$$t = \frac{\hat{\theta}_{kk}}{\sqrt{\frac{1}{T}}} = \sqrt{T}\hat{\theta}_{kk} \xrightarrow{d} N(0,1)$$

Le test bilatéral se formule comme à l'accoutumée. On rejette H_0 pour $\alpha = 5\%$ si $t > 2$. D'où l'intervalle de confiance : $\hat{\theta}_{kk} \pm \frac{2}{\sqrt{T}}$ ⁵.

3. PROCESSUS DE MOYENNES MOBILES : MA(q)⁶

Un processus de moyenne mobile, noté MA(q), stationnaire par définition, s'écrit comme suit⁷ :

$$y_t = \mu + e_t + \alpha_1 e_{t-1} + \alpha_2 e_{t-2} + \dots + \alpha_q e_{t-q}$$

On peut exprimer cette équation sous forme plus compacte en utilisant l'opérateur de retard :

$$y_t = \mu + \alpha(L)e_t$$

où

$$\alpha(L) = 1 + \alpha_1 L + \alpha_2 L^2 + \dots + \alpha_q L^q$$

et $e_t \sim WN(0, \sigma_e^2)$.

-
5. Nous avons retenu comme seuil critique pour $\alpha = 5\%$ la valeur 2. En fait, quand T se dirige vers l'infini, le seuil critique est de 1,96. À 60 observations, ce seuil est de 2. D'où l'approximation de 2 que nous avons retenue.
 6. MA est l'abréviation de l'expression anglaise : *moving average*. q représente l'ordre de la moyenne mobile.
 7. Nous avons ici incorporé une constante dans le modèle général d'un MA(q). Notons toutefois qu'il est d'usage de l'omettre.

Le théorème de Wold

Nous ouvrons ici une parenthèse sur le théorème de la décomposition de Wold (1938), considéré comme fondamental dans le domaine des séries temporelles. En vertu de ce théorème, tout processus stochastique $\{y_t - \mu\}$ stationnaire au sens faible et purement non déterministe peut être écrit comme une combinaison linéaire (dite encore *filtre linéaire*) d'une séquence non corrélée de variables aléatoires. Ce filtre linéaire est justement le modèle de moyenne mobile que nous avons représenté précédemment mais dont l'ordre q est infini :

$$y_t - \mu = e_t + \alpha_1 e_{t-1} + \alpha_2 e_{t-2} + \dots = \sum_{i=0}^{\infty} \alpha_i e_{t-i}$$

où $\alpha_0 = 1$. En termes plus concrets, on peut inverser tout processus stationnaire sous forme d'un MA infini qui peut être approximé par un ARMA(p, q) d'ordre faible. Ce théorème prend acte de l'importance du phénomène de stationnarité dans l'analyse des séries temporelles.

Propriétés du processus MA(1)

Pour démontrer qu'un processus MA est stationnaire, on calcule d'abord l'espérance d'un tel processus, ici d'ordre 1⁸.

$$y_t = \mu + e_t + \alpha_1 e_{t-1}$$

et $e_t \sim \text{WN}(0, \sigma_e^2)$. Son espérance est de μ puisque e_t est un processus stationnaire. La variance de y_t est de :

$$V(y_t) = V(\mu + e_t + \alpha_1 e_{t-1}) = 0 + \sigma_e^2 + \alpha_1^2 \sigma_e^2 = (1 + \alpha_1^2) \sigma_e^2 = \gamma_0$$

8. Les MA(q) de q faible sont dits de *courte mémoire* (*short memory*).

Par ailleurs,

$$\begin{aligned}
 \gamma_1 &= \text{Cov}(y_t, y_{t-1}) = E\left[(y_t - E(y_t))(y_{t-1} - E(y_{t-1}))\right] \\
 &= E\left[(\mu + e_t + \alpha_1 e_{t-1} - \mu)(\mu + e_{t-1} + \alpha_1 e_{t-2} - \mu)\right] \\
 &= E\left[(e_t + \alpha_1 e_{t-1})(e_{t-1} + \alpha_1 e_{t-2})\right] \\
 &= E\left[e_t e_{t-1} + \alpha_1 e_t e_{t-2} + \alpha_1 e_{t-1}^2 + \alpha_1^2 e_{t-1} e_{t-2}\right] \\
 &= E\left[\alpha_1 e_{t-1}^2\right] \\
 &= \alpha_1 \sigma_e^2
 \end{aligned}$$

Pour sa part, $\gamma_2 = \text{Cov}(y_t, y_{t-2}) = 0$. On peut démontrer que pour tout k plus grand que 1, toutes les autocovariances sont nulles pour le cas du MA(1). On en déduit la propriété suivante pour la fonction d'autocorrélation d'un MA(1) :

$$\rho_k = \begin{cases} \frac{\gamma_k}{\gamma_0} = \frac{\alpha_1}{1 + \alpha_1^2} & \text{pour } k = 1 \\ 0 & \text{pour } k > 1 \end{cases}$$

Généralisons maintenant pour le cas d'un MA(q). Sa fonction d'autocorrélation se présente comme suit :

$$\rho_k = \begin{cases} \frac{\gamma_k}{\gamma_0} = \frac{\sum_{i=0}^{q-k} \alpha_i \alpha_{i+k}}{\sum_{i=0}^q \alpha_i^2} & \text{pour } k = 0, 1, 2, \dots, q \\ 0 & \text{pour } k > q \end{cases}$$

Parce que la fonction d'autocorrélation est égale à 0 pour les retards supérieurs à q , on peut donc identifier l'ordre d'un MA(q) à partir de la fonction d'autocorrélation. En pratique, on estimera la fonction d'autocorrélation par :

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

Pour identifier l'ordre d'un MA(q), on utilise le test t comme on l'a vu précédemment pour tester $H_0: r_k = 0$ versus $H_1: \rho_k \neq 0$. Le test se formule comme suit :

$$t = \frac{\hat{\rho}_k - \rho_k}{\sqrt{V(\hat{\rho}_k)}} = \sqrt{T} \hat{\rho}_k \xrightarrow{d} N(0,1)$$

On rejette H_0 au seuil de $\alpha = 5\%$ si la valeur absolue de t excède 2. On peut également construire l'intervalle de confiance de ρ_k comme suit :

$$\hat{\rho}_k \pm \frac{2}{\sqrt{T}}$$

4. ESTIMATION D'UN MA(q)

À l'instar du modèle autorégressif, pour estimer les paramètres d'un MA(q) on écrit l'équation des e_t et on minimise par la suite la somme des erreurs au carré. Plus précisément :

$$\text{MIN}_{\alpha_1, \alpha_2, \dots, \alpha_q} S(\alpha_1, \alpha_2, \dots, \alpha_q) = \text{MIN}_{\alpha_1, \alpha_2, \dots, \alpha_q} \sum_{t=1}^T e_t^2$$

où $e_t = y_t - \alpha(L)e_t$. $S(\cdot)$ étant une fonction non linéaire de α , on aura alors recours aux méthodes d'estimation non linéaires, telles que les moindres carrés non linéaires ou encore le maximum de vraisemblance.

Pour constater que le modèle d'un MA(q) requiert une méthode d'estimation non linéaire, on peut se servir d'un modèle MA(1) qui se généralisera par la suite au MA(q). Celui-ci s'écrit :

$$y_t = e_t + \alpha_1 e_{t-1}$$

où on a supposé que μ est égal à 0. Substituons la valeur de e_{t-1} de façon répétée, c'est-à-dire :

$$\begin{aligned} y_t &= e_t + \alpha_1 (y_{t-1} - \alpha_1 (y_{t-2} - \alpha_1 e_{t-3})) \\ &= e_t + \alpha_1 y_{t-1} - \alpha_1^2 y_{t-2} + \alpha_1^3 e_{t-3} \end{aligned}$$

Nous pourrions continuer ces substitutions en remplaçant e_{t-3} par sa valeur. On obtient alors que :

$$e_t = y_t - \alpha_1 y_{t-1} + \alpha_1^2 y_{t-2} - \alpha_1^3 y_{t-3} + \dots - \alpha_1^{n-1} y_{t-n+1} + \alpha_1^n e_{t-n}$$

On peut réécrire une MA(1) de façon compacte en recourant à l'opérateur de retard :

$$y_t = \alpha_1(L)e_t = (1 + \alpha_1 L)e_t$$

Le processus MA(1) peut être inversé pour donner un processus AR d'ordre infini à condition bien sûr que $|\alpha_1| < 1$. On a :

$$(1 + \alpha_1 L)^{-1} y_t = e_t$$

où

$$(1 + \alpha_1 L)^{-1} = [1 - (-\alpha_1)L]^{-1} = 1 + (-\alpha_1)L + (-\alpha_1)^2 L^2 + (-\alpha_1)^3 L^3 + \dots$$

On voit donc ici que e_t est effectivement une fonction non linéaire de ses paramètres. C'est pourquoi l'on recourt à une méthode d'estimation non linéaire.

5. MODÈLES ARMA (p, q)

Le modèle ARMA(p, q) est mixte en ce sens qu'il combine modèle autorégressif (AR) et modèle de moyenne mobile (MA), d'où son appellation ARMA. Le modèle ARMA(p, q) s'écrit comme suit :

$$y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + e_t \\ + \alpha_1 e_{t-1} + \alpha_2 e_{t-2} + \dots + \alpha_q e_{t-q}$$

Sous forme compacte, ce processus s'écrit :

$$\theta(L)y_t = \delta + \alpha(L)e_t$$

où e_t est un bruit blanc homoscédastique de moyenne nulle.

Propriétés du modèle ARMA(1, 1)

Avec le modèle AR(1), le modèle ARMA(1, 1) est celui que l'on rencontre le plus souvent dans la littérature financière. Il s'écrit comme suit :

$$y_t = \delta + \theta_1 y_{t-1} + e_t + \alpha_1 e_{t-1}$$

où e_t est un bruit blanc de moyenne nulle et de variance σ^2 . L'espérance non conditionnelle de y_t nous donne l'expression suivante :

$$E(y_t) = \frac{\delta}{1 - \theta_1}$$

Pour le calcul de la variance de y_t , nous faisons l'hypothèse que δ est nul.

$$V(y_t) = \gamma_0 = E\left[(y_t - E(y_t))^2\right] = \sigma_e^2 \left(\frac{1 + \alpha_1^2 + 2\theta_1 \alpha_1}{1 - \theta_1^2} \right)$$

où l'on suppose que $E(y_t)$ est nulle. Alors l'autocovariance se calcule comme suit :

$$\begin{aligned} \text{Cov}(y_t, y_{t-1}) &= \gamma_1 \\ &= E\left[(y_t - E(y_t))(y_{t-1} - E(y_{t-1}))\right] = \theta_1 \gamma_0 + \alpha_1 \sigma_e^2 \end{aligned}$$

et où

$$\gamma_k = \theta_1 \gamma_{k-1} \quad \text{pour } k \geq 2$$

La fonction d'autocorrélation pour le modèle ARMA(1, 1) est de :

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(1 + \theta_1 \alpha_1)(\theta_1 + \alpha_1)}{1 + \alpha_1^2 + 2\theta_1 \alpha_1}$$

et

$$\rho_k = \theta_1 \rho_{k-1} \quad \text{pour } k \geq 2$$

Généralisons maintenant au cas du ARMA(p, q). La fonction d'autocorrélation de ce modèle est donnée par :

$$\rho_k = \theta_1 \rho_{k-1} + \theta_2 \rho_{k-2} + \dots + \theta_p \rho_{k-p}$$

pour $k > q$.

6. INTRODUCTION AUX PROCESSUS STOCHASTIQUES NON STATIONNAIRES : MODÈLES ARIMA (p, d, q)

Jusqu'ici, nous avons considéré des processus stationnaires. Cependant, en finance, les séries financières sont bien souvent non stationnaires. Le modèle ARIMA⁹ (p, d, q) est tout désigné pour cette catégorie de séries. L'ajout du I dans l'acronyme ARMA désigne l'ordre d'intégration requis pour atteindre la stationnarité. En effet, on peut rendre stationnaire une série non stationnaire simplement en la différenciant. Par exemple, un processus intégré d'ordre 1 ($d = 1$) doit être différencié une fois pour atteindre la stationnarité. Le qualificatif *intégré* provient du calcul différentiel et intégral. En effet, comme :

$$x_t = \frac{dy_t}{dt}$$

alors, y_t est égal à :

$$y_t = \int x_t dt$$

Transposons ces équations au cas discret. On a :

$$x_t = \Delta y_t$$

où, par analogie avec le cas continu, y_t est une somme infinie de x_t , c'est-à-dire :

$$x_t = \Delta y_t = (1-L)y_t \Rightarrow y_t = (1-L)^{-1} x_t = (1+L+L^2+\dots)x_t$$

ce qui est en fait une somme infinie de x_t . On voit donc qu'une série intégrée d'ordre 1 se ramène à une somme infinie de retards, d'où le

9. ARIMA est l'abréviation de : *autoregressive integrated moving average*.

concept d'intégration. Ce cas se généralise à une série intégrée d'ordre d , qui revient à différencier la série d fois, donc à effectuer d intégrales.

Un modèle ARIMA (1,1,0) a la forme suivante :

$$y_t = y_{t-1} + e_t$$

et est en fait un modèle AR(1) où $\theta_1 = 1$, aussi appelé *marche aléatoire*. Ce modèle doit être différencié une fois pour atteindre la stationnarité, c'est-à-dire :

$$\Delta^1 y_t = y_t - y_{t-1} = e_t$$

où e_t , nous l'avons mentionné, est un bruit blanc, par conséquent stationnaire. Par conséquent, un processus intégré d'ordre d , c'est-à-dire : $\Delta^d y_t = (1-L)^d y_t$, doit être différencié d fois pour atteindre la stationnarité. À titre d'exemple, supposons un processus stochastique x_t qui est intégré d'ordre 2, soit :

$$\Delta^2 x_t = (1-L)^2 x_t = (1-2L+L^2)x_t = x_t - 2x_{t-1} + x_{t-2}$$

Pour les processus intégrés (non différenciés), les ACF sont très significatives jusqu'à des k (retards) contrairement aux ACF de processus stationnaires qui se dirigent vers 0 pour des k élevés.

7. LA MÉTHODE DE BOX ET JENKINS¹⁰

La méthode de Box et Jenkins consiste à déceler la forme du modèle ARIMA qui reproduit le mieux la série financière analysée. Cette méthode comporte trois étapes : i) l'identification ; ii) l'estimation ; iii) les tests et diagnostics. Dans ce qui suit, nous allons développer ces trois étapes.

10. Box, G.E.P. et G.M. Jenkins (1976), *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco. Un autre livre qui est aussi à l'origine de l'analyse des séries temporelles et qui adapte la méthode de Box et Jenkins aux sciences de la gestion est : Nelson, C.R. (1973), *Applied Time Series Analysis for Managerial Forecasting*, Holden-Day, San Francisco.

i) L'identification

L'identification se base principalement sur l'analyse des ACF et PACF des séries économiques et financières considérées. On peut distinguer les cas d'espèce suivants :

- 1) Si l'ACF décroît lentement vers 0, c'est-à-dire que pour des retards éloignés, les coefficients d'autocorrélation demeurent significatifs, on considère alors que cette série est non stationnaire. On devra alors différencier cette série une fois et parfois même deux fois pour la rendre stationnaire. Un modèle ARMA d'ordre p et q peu élevé pourra être ensuite estimé sur la série stationnarisée.
- 2) Pour un processus MA(q), l'ACF : $\rho(k)$ est égale à 0 pour $k > q$ et la PACF décroît géométriquement vers 0. Pour déterminer l'ordre q de ce processus, il faut comparer l'ACF : $\rho(k)$ avec $\pm \frac{2}{\sqrt{T}}$.
- 3) Pour un processus AR(p), la PACF : θ_{kk} est égale à 0 pour $k > p$ et l'ACF décroît géométriquement vers 0. Pour déterminer l'ordre p de ce processus, on compare la PACF : θ_{kk} avec $\pm \frac{2}{\sqrt{T}}$.
- 4) Si l'on ne trouve pas de point de rupture précis, un modèle ARMA pourrait être pertinent. Par exemple, étant donné qu'un ARMA(1, 1) est une combinaison d'un modèle AR et d'un modèle MA, on s'attend à ce que l'ACF ait les caractéristiques des modèles AR et MA combinés. La portion MA disposant d'une mémoire d'une période seulement, le point de rupture devrait se produire après une période. Par ailleurs, la composante AR dispose d'une ACF qui décroît géométriquement et on anticipe ce profil pour des retards supérieurs à une période.

ii) L'estimation

Si un modèle AR a été identifié, alors l'estimation se fera par les MCO. Par ailleurs, si on a identifié un modèle MA, l'estimation se

fera par les moindres carrés non linéaires (NLS) ou par la méthode ML. Si un modèle ARMA a été identifié, l'estimation s'effectuera par la méthode des NLS ou la méthode ML, ou encore par la méthode des MCO en deux étapes¹¹.

iii) Tests et diagnostics

Dans cette sous-section, nous traitons de la surparamétrisation (*overfitting*) et des tests sur les résidus.

La surparamétrisation

Si un modèle ARIMA(p, d, q) a été identifié, la surparamétrisation consiste à estimer un modèle ARIMA(p + 1, d, q) ou un modèle ARIMA(p, d, q + 1) ou les deux et faire un test sur le ou les paramètres additionnels. Si le vrai modèle est un ARIMA(p, d, q), les tests sur les paramètres additionnels ne devraient pas être significativement différents de 0.

Analyse des résidus

Une fois l'estimation des modèles ARIMA complétée, on calcule les résidus de façon à dégager les ACF et PACF de ces résidus. On devrait trouver que la fonction d'autocorrélation est non significative pour

tous les retards en comparant $\hat{\rho}_k$ à $\frac{2}{\sqrt{T}}$. Une procédure plus scienti-

fique est de calculer le test Ljung-Box (1978). Ce test, que l'on désigne par la statistique Q, se formule comme suit :

$$Q = T(T + 2) \sum_{k=1}^K \frac{\hat{\rho}_k^2}{T - k}$$

11. Au sujet de la méthode des moindres carrés en deux étapes, on consultera : Gouriéroux, C. et A. Monfort (1990), *Séries temporelles et modèles dynamiques*, Economica, Paris.

L'hypothèse H_0 est la suivante : $\rho_1 = \rho_2 = \dots = \rho_k = 0$. Ce test est asymptotiquement distribué comme une $\chi^2(k-p-q)$.

8. AUTRES CRITÈRES DE SÉLECTION POUR LES MODÈLES ARMA

On peut distinguer deux critères de choix de modèles : celui d'Akaike (1973) et celui de Schwarz (1978). Le critère d'Akaike, désigné par AIC¹², se définit comme suit :

$$\text{AIC} = \ln\left(\frac{\hat{\mathbf{e}}^T \hat{\mathbf{e}}}{T}\right) + \frac{2(p+q+1)}{T}$$

Par ailleurs, le critère de sélection de Schwarz, désigné par SC¹³, se calcule comme suit :

$$\text{SC} = \ln\left(\frac{\hat{\mathbf{e}}^T \hat{\mathbf{e}}}{T}\right) + \frac{p+q+1}{T} \ln(T)$$

Ces deux critères remplacent le R^2 dans l'analyse des séries temporelles. On cherche le modèle qui donne la valeur minimale à ces deux statistiques. Par exemple, si un ARMA(1, 1) se traduit par des statistiques AIC et SC plus faibles qu'un ARMA(2, 1), on choisira le modèle ARMA(1, 1).

On remarquera que les deux critères pénalisent l'ajout de degrés de liberté, ce qui milite en faveur du principe de la parcimonie dans l'établissement d'un modèle, ce qui se compare davantage au R^2 ajusté qu'au R^2 non ajusté.

12. AIC est l'abréviation anglaise de : *Akaike Information Criterion*.

13. SC est l'abréviation anglaise de : *Schwarz Criterion*.

9. PRÉVISIONS À L'AIDE DE MODÈLES STATISTIQUES DE SÉRIES CHRONOLOGIQUES

Une fois un modèle spécifié, estimé et testé, on peut l'utiliser pour effectuer des prévisions. Envisageons d'abord le cas d'un modèle ARMA(1, 1).

Prévision à l'aide d'un modèle ARMA(1, 1)

Le modèle ARMA(1, 1) se formule comme suit :

$$y_t = \delta + \theta_1 y_{t-1} + e_t + \alpha_1 e_{t-1}$$

Soit h le nombre de périodes à l'avance au chapitre de la prévision. Considérons d'abord le cas où h est égal à 1. Nous sommes à la fin de l'échantillon T . Nous voulons prévoir l'observation en $(T+1)$ qui se trouve dans le futur immédiat. On a, en vertu du modèle ARMA(1, 1) :

$$y_{T+1} = \delta + \theta_1 y_T + e_{T+1} + \alpha_1 e_T$$

L'espérance conditionnelle (E_T) de cette équation correspond à la prévision MMSE (*minimum mean square error*) de y_{T+1} :

$$\hat{y}_{T+1} = E_T(y_{T+1}) = \delta + \theta_1 y_T + 0 + \alpha_1 e_T$$

où l'espérance conditionnelle¹⁴ de e_{T+1} est nulle et $E_T(\alpha_1 e_T) = \alpha_1 e_T$. Incidemment, on se sert de l'espérance conditionnelle pour formuler des prévisions à court terme et de l'espérance inconditionnelle pour formuler des prévisions à long terme. La variance de l'erreur de prévision se calcule comme suit :

$$V(\hat{e}_{T+1}) = V(y_{T+1} - \hat{y}_{T+1}) = V(e_{T+1}) = \sigma_e^2$$

Soit ensuite le cas où $h=2$, cas où l'on calcule la prévision de y deux périodes à l'avance. Toujours en vertu du modèle ARMA(1, 1), on a :

$$\begin{aligned} \hat{y}_{T+2} &= E_T(y_{T+2}) = E_T(\delta + \theta_1 y_{T+1} + e_{T+2} + \alpha_1 e_{T+1}) \\ &= \delta + \theta_1 E_T(y_{T+1}) = \delta + \theta_1 \hat{y}_{T+1} \end{aligned}$$

14. Conditionnelle à l'information disponible jusqu'à la fin de l'échantillon T .

L'erreur de prévision est bien sûr :

$$\begin{aligned}\hat{e}_{T+2} &= y_{T+2} - \hat{y}_{T+2} = \delta + \theta_1 y_{T+1} + e_{T+2} + \alpha_1 e_{T+1} - \delta - \theta_1 \hat{y}_{T+1} \\ &= \theta_1 (y_{T+1} - \hat{y}_{T+1}) + e_{T+2} + \alpha_1 e_{T+1} \\ &= (\theta_1 + \alpha_1) e_{T+1} + e_{T+2}\end{aligned}$$

Fort de ces calculs, on calcule la variance de l'erreur de prévision :

$$V(\hat{e}_{T+2}) = (\theta_1 + \alpha_1)^2 \sigma_e^2 + \sigma_e^2 = \sigma_e^2 \left((\theta_1 + \alpha_1)^2 + 1 \right)$$

En procédant de la sorte, pour $h = 3$, on obtient :

$$\hat{y}_{T+3} = \delta + \theta_1 \hat{y}_{T+2}$$

et la variance de l'erreur de prévision est de :

$$V(\hat{e}_{T+3}) = \sigma_e^2 \left[1 + (\theta_1 + \alpha_1)^2 + (\theta_1^2 + \theta_1 \alpha_1)^2 \right]$$

Comme on l'aura constaté, les prévisions sont obtenues par substitutions répétées dans le modèle initial ARMA(1, 1). Déplaçons-nous vers le modèle général ARMA(p, q) qui se formule comme suit :

$$y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + e_t + \alpha_1 e_{t-1} + \dots + \alpha_q e_{t-q}$$

On peut effectuer une prévision pour h périodes à l'avance ; on calcule l'espérance conditionnelle et on obtient :

$$\begin{aligned}\hat{y}_{T+h} &= \delta + \theta_1 \hat{y}_{T+h-1} + \theta_2 \hat{y}_{T+h-2} + \dots + \theta_{h-1} \hat{y}_{T+1} \\ &+ \theta_h \hat{y}_T + \dots + \theta_p \hat{y}_{T-p+h} + \alpha_h e_T + \dots + \alpha_q e_{T-q+h}\end{aligned}$$

L'intervalle de confiance de cette prévision est de :

$$\hat{y}_{T+h} \pm z_c V(e_{T+h})^{\frac{1}{2}}$$

où z_c est la valeur critique d'une $N(0, 1)$ au niveau α retenu. Mentionnons que cet intervalle de confiance est asymptotique, c'est-à-dire que l'échantillon doit être suffisamment grand si l'on utilise des estimés des paramètres à la place des paramètres comme tels pour effectuer les calculs précédents.

10. ÉVALUATION DE LA PRÉCISION DES PRÉVISIONS

On distingue quatre critères pour évaluer la précision des prévisions : 1) le RMSE (*root mean square error*); 2) le MAE (*mean absolute error*); les deux statistiques U de Theil. Ces critères donnent lieu au choix d'un modèle qui minimise lesdits critères.

Le RMSE est égal à :

$$\text{RMSE} = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{n_0}}$$

où n_0 représente le nombre de périodes prédites.

Pour sa part, MAE se définit comme suit :

$$\text{MAE} = \frac{\sum_i |y_i - \hat{y}_i|}{n_0}$$

Remarquons que ces deux critères sont sujets à des problèmes d'unités de mesure. Les deux statistiques U de Theil corrigent pour ce problème. La première, dénotée par U, est égale à :

$$U = \sqrt{\frac{(1/n_0) \sum_i (y_i - \hat{y}_i)^2}{(1/n_0) \sum_i y_i^2}}$$

Remarquons que cette mesure est reliée au R^2 mais n'est pas limitée à l'intervalle (0, 1). Une mesure apparentée consiste à calculer les variations de y :

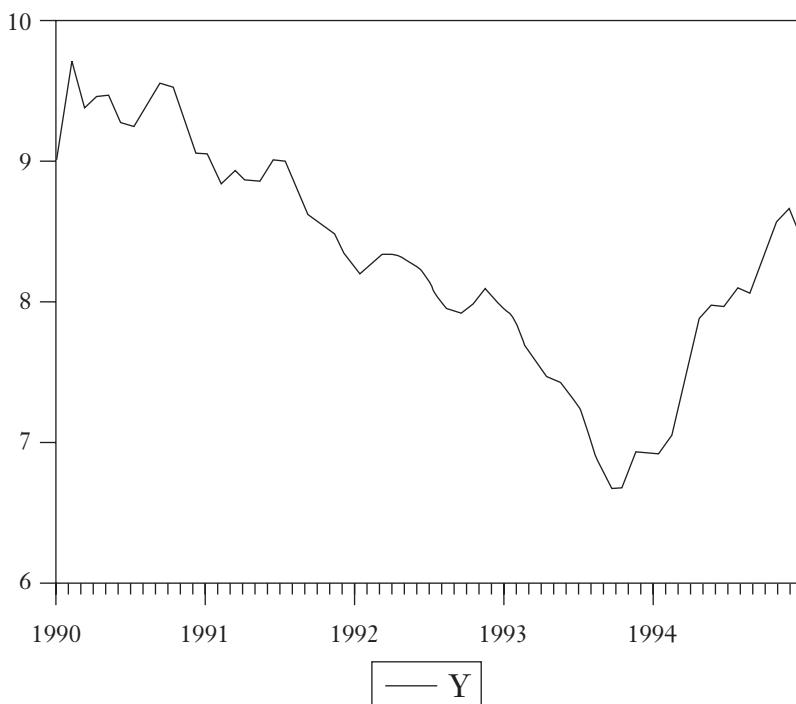
$$U_{\Delta} = \sqrt{\frac{(1/n_0) \sum_i (\Delta y_i - \Delta \hat{y}_i)^2}{(1/n_0) \sum_i \Delta y_i^2}}$$

Toutes ces mesures mesurent la capacité des modèles à capter les différents points de retournement des séries temporelles.

11. APPLICATIONS

Pour illustrer ce chapitre ayant trait aux séries chronologiques, nous nous basons sur une série chronologique ayant trait au marché obligataire. Le rendement retenu est celui de l'obligation corporative américaine cotée Aaa selon le système de notation Moody's. Les données sont mensuelles et s'étirent de 1990 à 1994¹⁵. La figure 9.2 retrace l'évolution de ladite série sur la période considérée, dénommée Y. Ce graphique a été généré par le logiciel EViews.

FIGURE 9.2



15. Cette série est tirée de Greene (2000), *op. cit.*

Un examen rapide de la figure 9.2 semble révéler que cette série n'est pas stationnaire. Pour vérifier l'hypothèse de stationnarité, nous présentons, à la figure 9.3, les calculs et graphiques quant aux autocorrélations totales et partielles.

FIGURE 9.3 Corrélogramme de Y

Date : 08/30/00 Time : 11:48
 Sample : 1990:01 1994:12
 Included observations : 60

Autocorrélation	Corrélation partielle		AC	PAC	Q-stat	Prob.
		1	0.967	0.967	58.904	0.000
		2	0.909	-0.384	111.90	0.000
		3	0.853	0.143	159.37	0.000
		4	0.795	-0.150	201.33	0.000
		5	0.736	0.020	237.93	0.000
		6	0.674	-0.101	269.26	0.000
		7	0.606	-0.123	295.02	0.000
		8	0.530	-0.090	315.15	0.000
		9	0.451	-0.082	330.00	0.000
		10	0.379	0.110	340.70	0.000
		11	0.318	0.015	348.36	0.000
		12	0.260	-0.043	353.59	0.000
		13	0.209	0.106	357.04	0.000
		14	0.165	-0.025	359.25	0.000
		15	0.127	0.053	360.57	0.000
		16	0.095	0.000	361.33	0.000
		17	0.059	-0.207	361.63	0.000
		18	0.014	-0.131	361.65	0.000
		19	-0.030	-0.020	361.74	0.000
		20	-0.069	0.035	362.17	0.000
		21	-0.103	-0.072	363.19	0.000
		22	-0.136	0.003	365.00	0.000
		23	-0.165	0.026	367.75	0.000
		24	-0.190	0.092	371.47	0.000
		25	-0.218	-0.116	376.50	0.000
		26	-0.251	-0.070	383.42	0.000
		27	-0.284	-0.094	392.54	0.000
		28	-0.311	0.028	403.81	0.000

En examinant la structure des autocorrélations totales, on peut observer des autocorrélations persistantes pour des retards éloignés, ce qui pourrait indiquer que l'hypothèse de non-stationnarité pourrait être violée. Mais, pour l'instant, nous mettons en veilleuse ce problème

pour nous concentrer sur les ACF de la série qui semblent indiquer qu'un modèle AR(2) est approprié. L'estimation de ce modèle apparaît au tableau 9.1.

TABLEAU 9.1

Dependent Variable: Y				
Method: Least Squares				
Date: 08/30/00 Time: 11:51				
Sample: (adjusted): 1990:03 1994:12				
Included observations: 58 after adjusting endpoints				
Convergence achieved after 3 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.876779	0.441819	17.82807	0.0000
AR(1)	1.156648	0.110660	10.45226	0.0000
AR(2)	-0.208293	0.110161	-1.890803	0.0639
R-squared	0.962990	Mean dependent var	8.254138	
Adjusted R-squared	0.961644	S.D. dependent var	0.794039	
S.E. of regression	0.155511	Akaike info criterion	-0.833866	
Sum squared resid	1.330096	Schwarz criterion	-0.727292	
Log likelihood	27.18212	F-statistic	715.5335	
Durbin-Watson stat	1.418410	Prob(F-statistic)	0.000000	
Inverted AR Roots	.93	.22		

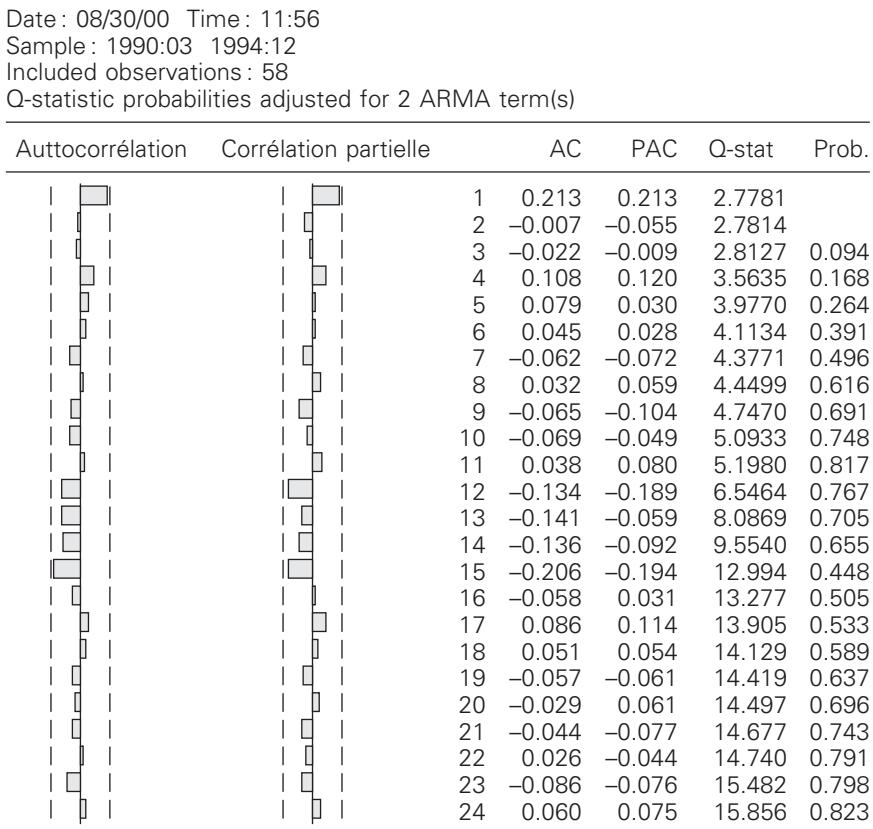
Examinons la stationnarité de cette série à partir des coefficients estimés des retards sur la variable dépendante. Mais auparavant, rappelons les conditions de stationnarité pour un AR(2). Dans ce cas, la série converge si les coefficients du AR(2) respectent les conditions suivantes: 1) $|\theta_2| < 1$; 2) $\theta_1 + \theta_2 < 1$; 3) $\theta_2 - \theta_1 < 1$. Nous pouvons vérifier au tableau 9.1 que ces trois conditions sont respectées pour la série analysée. De façon plus générale un processus autorégressif d'ordre p est stationnaire si les racines de l'équation caractéristique suivante :

$$C(z) = 1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_p z^p$$

ont un module supérieur à 1 ou, de façon équivalente, se situent à l'extérieur du cercle unitaire.

La figure 9.4 présente l'analyse des résidus du modèle estimé. On observe que les ACF et les PACF pour tous les retards ne sont pas significatifs. Cela confirme que les résidus obéissent à un processus de bruit blanc. Et l'on en conclut que le modèle choisi est le bon puisqu'il capte toute l'autocorrélation présente dans la série.

FIGURE 9.4 Corrélogramme des résidus



Séries temporelles non linéaires

Selon Franses (1998)¹⁶, une série temporelle est non linéaire quand des chocs importants ont un impact différent de chocs de moindre

16. Franses, P.H. (1998), *Time Series Models for Business and Economic Forecasting*, Cambridge University Press, Cambridge.

envergure dans le sens que l'impact d'un choc n'est pas proportionnel à son importance. La non-linéarité peut aussi signifier que l'impact d'un choc dépend de son signe. Pour modéliser le caractère non linéaire des séries temporelles, nous envisageons ici deux méthodes d'estimation : la méthode STAR, soit un processus autorégressif de transition lisse ou sans cassures (*smooth*), et le modèle économétrique basé sur les réseaux de neurones, que nous désignerons par ANN.

Le modèle STAR

Supposons que la série temporelle y_t suive le processus suivant :

$$y_t = \delta + \phi y_{t-1} + F(y_{t-d})(\psi + \lambda y_{t-1}) + \varepsilon_t$$

où $F(y_{t-d})$ est une fonction quelconque de la variable y_{t-d} , dite variable de transition. Ce modèle, désigné par STAR¹⁷, est la combinaison d'un modèle linéaire AR(1) et d'un modèle non linéaire quand $F(y_{t-d})$ n'est pas nul. On peut interpréter $\psi F(y_{t-d})$ et $\lambda F(y_{t-d}) y_{t-1}$ comme des composantes qui rendent variables dans le temps l'intercept et le paramètre autorégressif de premier ordre. La fonction de transition $F(\cdot)$ est habituellement choisie de telle sorte que ses réalisations se situent dans l'intervalle $[0, 1]$. Par conséquent, quand $F(\cdot) = 0$, y_t peut être décrit comme un processus autorégressif AR(1) pur : $\delta + \phi y_{t-1}$. Par ailleurs, quand $F(\cdot) = 1$, un processus AR(1) avec les paramètres $(\delta + \psi)$ et $(\phi + \lambda)$ s'avère pertinent. Finalement, quand $0 < F(\cdot) < 1$, y_t peut être décrit comme une somme pondérée de deux processus AR linéaires. Cela suggère une forme alternative pour y_t :

$$y_t = (1 - F(y_{t-d}))(\delta_1 + \alpha_1 y_{t-1}) + F(y_{t-d})(\delta_2 + \alpha_2 y_{t-1}) + \varepsilon_t$$

On remarquera que lorsque $\alpha_1 = 1$ (ou même supérieur à 1), les effets des innovations ε_t sur la série y_t peuvent encore n'être que transitoires. En d'autres mots, la stationnarité de y_t dépend de α_1 , α_2 et de la forme spécifique de la fonction $F(\cdot)$. Franses (1998) note que tester la stationnarité de y_t est une entreprise difficile.

17. Soit l'abréviation de l'expression anglaise : *smooth transition autoregressive*.

La fonction de transition $F(\cdot)$ peut prendre plusieurs formes : i) la forme exponentielle suivante, qui donne lieu au modèle ESTAR :

$$F(y_{t-d}) = 1 - e^{(-\gamma(y_{t-d}-c)^2)}$$

où $\gamma > 0$; ii) la fonction logistique suivante, qui donne lieu au modèle LSTAR :

$$F(y_{t-d}) = \left[1 + e^{(-\gamma(y_{t-d}-c))} \right]^{-1}$$

où $\gamma > 0$; la fonction suivante, qui donne lieu au modèle AR à paliers ou à seuils, encore désigné par modèle TAR¹⁸ :

$$F(y_{t-d}) = 0 \quad \text{pour } y_{t-d} \leq c$$

$$F(y_{t-d}) = 1 \quad \text{pour } y_{t-d} > c$$

Le paramètre c est appelé : paramètre de palier (*threshold parameter*). Le modèle AR est un cas spécial du modèle LSTAR quand $\gamma \rightarrow \infty$. Dans le modèle ESTAR, y_t réagit de façon symétrique aux valeurs positives et négatives de $(y_{t-d} - c)$ tandis que dans le modèle TAR, le changement de régime est abrupt. Le modèle LSTAR comporte pour sa part des changements lisses ou sans cassures (*smooth*) et des réactions asymétriques aux chocs, ce qui rehausse son attrait par rapport aux deux autres modèles.

Le modèle économétrique basé sur les réseaux de neurones (ANN)

Le modèle ANN (*artificial neural network*) peut être combiné à une série temporelle qui obéit à une dynamique autorégressive de premier ordre. En supposant qu'il n'y ait qu'une seule couche cachée dans le réseau de neurones, ce modèle économétrique s'écrit comme suit :

$$y_t = \delta + \phi y_{t-1} + \sum_{j=1}^q \beta_j G(\psi_j + \lambda_j y_{t-1}) + \varepsilon_t$$

18. TAR est l'abréviation anglaise de : *threshold autoregressive*.

où q désigne le nombre de cellules cachées dans la couche. Contrairement au modèle STAR, le modèle ANN se traduit par un intercept variable dans le temps. La fonction $G(\cdot)$ prend habituellement la forme de la fonction logistique suivante :

$$G(\psi_j + \lambda_j y_{t-1}) = \left[1 + e^{-(\psi_j + \lambda_j y_{t-1})} \right]^{-1}$$

La fonction $G(\cdot)$ est appelée *fonction d'activation logistique* puisque la composante non linéaire du modèle ANN ne devient active que si l'argument $(\psi_j + \lambda_j y_{t-1})$ est assez important. À l'instar du modèle LSTAR, le modèle ANN peut décrire des changements de régimes dans les séries temporelles, en autant que ces changements se confinent aux intercepts.

Quand q devient relativement important, le modèle ANN peut approximer de façon très précise toute fonction :

$$f(y_t) = f(y_{t-1}) + \varepsilon_t$$

Par conséquent, le R^2 d'un modèle ANN peut se rapprocher facilement de 1 lorsque q est élevé. Puisque le modèle ANN fait la somme d'un ensemble de fonctions logistiques, il est toutefois difficile d'interpréter les valeurs des paramètres des fonctions $G(\cdot)$.

Pour estimer les paramètres des modèles STAR et ANN, on recourt à un algorithme d'estimation non linéaire. Franses (1998) suggère d'adopter la stratégie suivante pour estimer ces paramètres. On fixe d'abord les paramètres γ et c à des niveaux donnés et on estime les autres paramètres par la méthode des MCO. On estime ensuite les paramètres γ et c par étapes par la méthode des moindres carrés non linéaires. Certes, la valeur de d dans y_{t-d} est incertaine et il faut effectuer plusieurs essais sur cet indice de façon à déterminer la régression qui comporte le R^2 le plus élevé. Franses (1998)¹⁹ a appliqué le modèle LSTAR à la première différence du taux de chômage trimestriel allemand sur une période s'étirant de 1962 à 1990. Il a trouvé que le d optimal était de 1. Le résultat de son estimation est le suivant :

19. Franses (1998), *op. cit.*, p. 183.

$$\Delta y_t = (1 - \hat{F}(\Delta y_{t-1}))(-0,040 + 0,454\Delta y_{t-1}) \\ + \hat{F}(\Delta y_{t-1})(1,901 - 1,840\Delta y_{t-1}) + \hat{\varepsilon}_t$$

où $\hat{F}(\Delta y_{t-1}) = \left[1 + e^{(-9,231(\Delta y_{t-1} - 0,500))} \right]^{-1}$. On peut tirer plusieurs informations d'une telle estimation qui alterne entre le régime de reprise économique et celui de récession. D'abord, en vertu de l'estimation, si $\Delta y_{t-1} > 0,500$, la fonction \hat{F} prend une valeur rapprochée de 1, ce qui se traduit dans ce modèle par une récession du point de vue de l'emploi. En traçant le graphique de \hat{F} en fonction du temps, on peut ainsi identifier les périodes de récession et de reprise économique. Ensuite, on peut relier graphiquement \hat{F} et Δy_{t-1} , \hat{F} étant sur l'ordonnée et Δy_{t-1} sur l'abscisse. On obtient alors la fonction de transition des récessions aux reprises économiques. Finalement, si l'on identifie un trimestre de récession à une valeur de \hat{F} excédant 0,5, on peut identifier les sommets et les creux du cycle économique classique.

L'ajout de composantes non linéaires aux modèles autorégressifs élargit donc de beaucoup le champ de l'analyse économétrique des séries chronologiques et augmente d'autant l'attrait de ce type d'analyse. En recourant aux réseaux de neurones, on peut même en arriver à reproduire quasi exactement toute série temporelle y_t ! Par ailleurs, les réseaux de neurones ont également été utilisés dans un contexte GARCH par Donaldson et Kamstra (1997)²⁰ pour estimer la volatilité du S&P500, entre autres. Pour plus de détails sur ce domaine captivant de l'économétrie, le lecteur consultera avec intérêt : Granger et Teräsvirta (1993); Teräsvirta, Tjostheim et Granger (1994); De Gooijer et Kumar (1992); Tong (1990); Bishop (1995); Ripley (1994); Kuan et Liu (1995); Kuan et White (1994); Swanson et White (1995)²¹.

20. Donaldson, R.G. et M. Kamstra (1997), « An Artificial Neural Network-GARCH Model for International Stock Return Volatility », *Journal of Empirical Finance*, 4, p. 17-46.

21. Granger, C.W.J. et T. Teräsvirta (1993), *Modelling Nonlinear Economic Relationships*, Oxford University Press, Oxford; Teräsvirta, T., D. Tjostheim et C.W.J. Granger (1994), « Aspects of Modelling Nonlinear Time Series », dans : Engle, R.F. et D.L. McFadden (dir.), *Handbook of Econometrics*, volume IV, North Holland, Amsterdam; De Gooijer, J.G. et K. Kumar (1992), « Some Recent

12. PROCESSUS STOCHASTIQUES NON STATIONNAIRES

Dans cette section, nous nous penchons sur les sujets suivants : processus de marche aléatoire, processus de marche aléatoire avec tendance, tests de stationnarité et cointégration.

Marche aléatoire (*random walk*)

Une marche aléatoire est un processus stochastique non stationnaire qui se définit comme suit dans sa version simplifiée :

$$y_t = y_{t-1} + e_t$$

où $e_t \sim WN(0, \sigma^2)$. Ce processus est un AR(1) sans constante où $\theta_1 = 1$. Si l'on différencie une fois ce modèle, i.e. $\Delta y_t = y_t - y_{t-1} = e_t$, on obtient alors un processus stationnaire. Si l'on veut tester l'efficacité des marchés financiers dans ce contexte, l'hypothèse nulle est alors : $H_0 : \theta_1 = 1$, où y_t est le rendement d'un titre. Si cette hypothèse est vérifiée, les marchés financiers sont alors efficaces. Un autre façon de vérifier l'efficacité est d'examiner le profil de l'ACF de Δy_t . S'il correspond à celui d'un bruit blanc, on a alors efficacité.

Le processus qui vient d'être décrit correspond à celui d'une marche aléatoire avec tendance stochastique (tendance non temporelle). Précisons ce point. Par substitutions successives dans l'équation de base de la marche aléatoire, on obtient :

$$y_t = y_0 + \sum_{i=1}^t e_i$$

Developments in Non-linear Time Series Modelling, Testing and Forecasting », *International Journal of Forecasting*, 8, p. 135-156 ; Tong. H. (1990), *Non-linear Time Series. A Dynamical System Approach*, Oxford University Press, Oxford ; Bishop, C.M. (1995), *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford ; Ripley, B.D. (1994), « Neural Networks and Related Methods for Classification », *Journal of the Royal Statistical Society*, 56, p. 409-456 ; Kuan, C.M. et T. Liu (1995), *Forecasting Exchange Rates using Feedforward and Recurrent « Neural Networks »*, *Journal of Applied Econometrics*, 10, p. 347-364 ; Kuan, C.M. et H. White (1994), « Artificial Neural Networks : An Econometric Perspective », *Econometric Reviews*, 13, p. 1-91 ; Swanson, N.R. et H. White (1995), « A Model Selection Approach to Assessing the Information in the Term Structure Using Linear Models and Artificial Neural Networks », *Journal of Business and Economic Statistics*, 13, p. 265-275.

Selon cette équation, ce processus de marche aléatoire ne comporte pas de tendance temporelle particulière. Si on calcule l'espérance non conditionnelle de cette valeur, on a :

$$E(y_t) = y_0$$

On en déduit donc la propriété que la moyenne d'une marche aléatoire avec tendance stochastique est constante. L'espérance conditionnelle, quant à elle, du processus de marche aléatoire se calcule :

$$E_t(y_{t+1}) = y_t$$

Par conséquent, du processus de marche aléatoire avec tendance stochastique, on déduit le résultat que l'espérance conditionnelle, qui est une prévision à court terme, est non constante alors que l'espérance non conditionnelle est pour sa part constante.

De façon similaire :

$$y_{t+s} = y_t + \sum_{i=1}^s e_{t+i}$$

Ce processus comporte un trend stochastique parce que le trend stochastique est la somme des termes d'erreur. Il en résulte que l'espérance conditionnelle n'est pas affectée par ce trend de termes d'erreur inconnus, c'est-à-dire :

$$E_t(y_{t+s}) = y_t$$

Pour sa part, la variance de y_t est de :

$$V(y_t) = 0 + \sum_{i=1}^t V(e_i) = t\sigma^2$$

D'après ce résultat, la marche aléatoire n'est donc un processus stationnaire puisque sa variance augmente avec le temps.

En finance, une martingale correspond en temps discret à une marche aléatoire. On dit en effet qu'un processus stochastique est une martingale relativement à l'ensemble d'information I_t si :

$$E_t(y_{t+1}) = y_t$$

Autrement dit, la meilleure prévision de y_{t+1} est y_t , soit l'observation antérieure. On retrouve bien ici le processus de marche aléatoire. En finance, des exemples de séries suivant une martingale sont les cours des actions et les taux d'intérêt.

Marche aléatoire avec tendance

Nous présentons ici le modèle avec drift. Ce modèle est celui d'une marche aléatoire auquel s'ajoute une constante. Ce modèle est le suivant :

$$y_t = a_0 + y_{t-1} + e_t$$

où a_0 est une constante. Il convient de montrer la présence d'une tendance dans un tel processus. Par substitutions répétées, on a :

$$y_t = y_0 + a_0 t + \sum_{i=1}^t e_i$$

Ce processus présente deux formes de trends : i) un trend déterministe représenté par $a_0 t$; ii) un trend stochastique, représenté par $\sum_{i=1}^t e_i$. On voit que l'espérance d'un tel processus se modifie dans le temps. La prévision $E(y_{t+h})$ est :

$$E_t(y_{t+h}) = y_t + a_0 h$$

13. MODÈLES DE TENDANCE

Une équation de différence peut être éclatée en trois composantes :

$$y_t = \text{trend} + \text{composante saisonnière} + \text{composante irrégulière}$$

Nous avons vu auparavant comment modéliser la composante irrégulière à partir d'un modèle ARMA(p, q). Nous faisons ici litière de la composante saisonnière, abordée sommairement dans une section antérieure ayant trait à l'effet du lundi sur les cours boursiers. Nous nous concentrons plutôt ici sur la composante tendancielle ou trend d'une série chronologique. Nous faisons ici référence à l'indice boursier américain S&P500 qui fait montre d'une tendance à la hausse.

Comment peut-on modéliser cette tendance de façon rigoureuse ? Deux techniques sont couramment utilisées à cet effet : i) le trend temporel linéaire ; ii) le trend temporel polynomial.

Le trend linéaire se présente comme suit :

$$y_t = a_0 + a_1 t + e_t$$

Pour sa part, le trend polynomial se formule comme suit :

$$y_t = a_0 + a_1 t + a_2 t^2 + \dots + a_n t^n + e_t$$

Ces deux équations peuvent être estimées par les MCO. Ces deux équations peuvent également comporter des retards sur y_t et e_t . On obtient alors un modèle ARMA(p, q) auquel s'ajoute un trend polynomial.

Pour enlever la tendance, il existe deux techniques : i) la différenciation ; ii) la technique du retrait de la tendance (*detrending*).

D'abord la différenciation. Considérons la solution du modèle de marche aléatoire avec tendance non temporelle : $y_t = y_0 + a_0 t + \sum_{i=1}^t e_i$. En calculant la différence première, on obtient :

$$\Delta y_t = 0 + a_0 + e_t$$

L'espérance de Δy_t est de :

$$E(\Delta y_t) = a_0$$

La variance de Δy_t est de :

$$V(\Delta y_t) = \sigma^2$$

et :

$$\text{Cov}(\Delta y_t, \Delta y_{t-s}) = 0$$

Toutes les propriétés d'un modèle stationnaire sont donc au rendez-vous.

L'autre technique utilisée pour enlever la tendance est le *detrending*. Nous venons de voir qu'un modèle non stationnaire peut être transformé en un modèle stationnaire seulement en le différenciant.

Cela ne signifie pas que tous les modèles non stationnaires peuvent être transformés en modèles stationnaires. Pour illustrer, considérons le modèle suivant.

$$y_t = a_0 + a_1 t + e_t$$

Pour extraire le trend, on estime le modèle par les MCO. On obtient :

$$y_t - \hat{y}_t = \hat{e}_t$$

où

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

On pourrait par la suite appliquer un modèle ARMA sur les résidus et déterminer les techniques étudiées auparavant pour déterminer l'ordre de ce modèle. Au lieu d'effectuer cette régression, on aurait pu exprimer y_t en première différence, soit :

$$\Delta y_t = a_0 + e_t - e_{t-1}$$

On constate que la partie en moyenne mobile (MA) comporte une racine unitaire. Ce modèle ne peut donc pas être inversé sous une forme autorégressive.

De façon générale, pour éliminer la tendance, on estime le modèle à trend polynomial par les MCO et, pour déterminer le degré de ce polynôme, on recourt aux tests t, aux tests F ou aux critères Akaike et Schwartz. Le modèle à trend polynomial s'écrit :

$$y_t = a_0 + a_1 t + a_2 t^2 + \dots + a_n t^n + e_t$$

En appliquant les MCO sur cette équation, on obtient \hat{y}_t . En soustrayant \hat{y}_t de y_t , on obtient alors la série stationnaire : $\{\hat{e}_t\}$.

Séries en différence stationnaire (DS) ou en tendance stationnaire (TS)

Jusqu'ici, nous avons examiné deux catégories de séries non stationnaires :

- i) la série en différence stationnaire (DS)
- ii) la série en tendance stationnaire (TS)

Les séries DS sont transformées en différence stationnaire en les différenciant. Quant à elles, les séries en tendance stationnaire sont rendues stationnaires en éliminant le trend par la modélisation de ce dernier. Mais un sérieux problème se présente lorsque, dans ce dernier cas, l'on enlève le trend en différenciant plutôt qu'en le modélisant. Pour le constater, considérons le modèle ARMA(p, q) suivant :

$$\Theta(L)y_t = \alpha_0 + \alpha_1 t + \Phi(L)e_t$$

où les racines du polynôme $\Theta(L)$ et $\Phi(L)$ se situent en dehors du cercle unitaire. Ce modèle est TS puisqu'il ne comporte aucune racine unitaire. Supposons quand même que l'on élimine le trend en le différenciant. On obtient :

$$\Theta(L)y_t^* = \alpha_1 + (1-L)\Phi(L)e_t$$

La différenciation a fait apparaître une racine unitaire dans la partie MA du processus. De façon parallèle, éliminer par modélisation du trend le trend d'un modèle DS ne se traduit par aucun retrait du trend stochastique.

14. RACINES UNITAIRES ET RÉGRESSIONS FALLACIEUSES

Comme nous l'avons vu, une racine unitaire dans un modèle AR(1) correspond à un coefficient unitaire dans l'autorégression. Nous obtenons alors une marche aléatoire reliée à l'efficiencia des marchés financiers. Attardons-nous maintenant au problème de racine unitaire dans une régression linéaire. Soit le modèle de régression suivant :

$$y_t = a_0 + a_1 z_t + e_t$$

où l'on suppose que $E(z_t e_t) = 0$. Les modèles où les variables explicatives sont stochastiques requièrent qu'elles soient stationnaires. Mais si y_t et z_t sont non stationnaires, on pourrait être confronté à une régression fallacieuse (*spurious regression*). Une telle régression n'a aucune valeur. Elle présente les symptômes suivants : 1) une faible Durbin-Watson ; 2) un R^2 élevé ; 3) des statistiques t significatives. Les résultats pourraient paraître acceptables à prime abord mais l'estimateur des MCO est alors non convergent et les tests statistiques usuels sont non valables. Par exemple, Granger et Newbold (1974)

relatent des régressions qu'ils ont effectuées sur des variables suivant chacune une marche aléatoire et ont pourtant obtenu un R^2 important même si ces deux processus n'ont apparemment aucun lien entre eux.

Considérons maintenant tour à tour quatre cas d'espèce ayant trait à deux séries chronologiques y_t et z_t : 1) les séries y_t et z_t sont stationnaires; 2) les deux séries sont intégrées de différents ordres; 3) les deux séries sont non stationnaires mais intégrées du même ordre; les résidus emmagasinent un trend stochastique; 4) les deux séries sont intégrées du même ordre et les résidus sont stationnaires.

Examinons d'abord le premier cas. Quand les deux séries sont stationnaires, alors les MCO peuvent être utilisées. Pour ce qui concerne le second cas, soit celui où les séries ont des ordres d'intégration différents, la régression par les MCO n'a aucune valeur. Pour fixer les idées, analysons la régression suivante :

$$y_t = a_1 z_t + e_t$$

où z_t obéit au processus AR(1) suivant :

$$z_t = \rho z_{t-1} + \varepsilon_t$$

et y_t suit la marche aléatoire suivante :

$$y_t = y_{t-1} + v_t$$

Les résidus e_t de cette régression peuvent être écrits sous la forme suivante :

$$e_t = \sum_{i=0}^t v_{t-i} - a_1 \sum_{i=0}^t \rho^i \varepsilon_{t-i}$$

Dans cette équation, la première somme ne converge pas alors que la deuxième somme converge en supposant que $|\rho| < 1$. On en conclut que e_t n'est pas stationnaire : cette régression n'a donc aucune valeur.

Passons maintenant au troisième cas, soit celui de deux séries non stationnaires mais intégrées du même ordre et où les résidus incorporent un trend stochastique. Dans un pareil cas, l'application des MCO à l'équation reliant y_t à z_t se traduira par une régression fallacieuse. Il est alors recommandé de différencier les deux séries :

$$\Delta y_t = a_1 \Delta z_t + \Delta e_t$$

où

$$\begin{aligned}y_t &= y_{t-1} + v_t \\z_t &= z_{t-1} + \zeta_t \\e_t &= e_{t-1} + \zeta_t\end{aligned}$$

Finalement, dans le quatrième cas, les séries sont intégrées du même ordre et les résidus sont censés stationnaires. Dans ce cas, on dit que les deux séries sont cointégrées. L'estimateur des MCO pour la régression de y sur z est alors convergent²². Les séries cointégrées sont reliées entre elles par un équilibre de long terme. Par exemple, on a constaté que les taux d'intérêt sont des séries cointégrées en vertu des théories de la structure à terme des taux d'intérêt. Incidemment Wilmott²³ voit la cointégration comme une mesure de corrélation.

15. TESTS DE RACINE UNITAIRE

Dans cette section, nous portons notre attention sur deux tests principaux : le test de Dickey-Fuller et le test de Dickey-Fuller *augmenté*. Il existe également le test de Perron, dont nous ferons une brève mention.

Pour tester la présence d'une racine unitaire, on utilise le test dû à Dickey-Fuller et publié en 1979. Ces auteurs se basent sur les régressions suivantes, qui sont des transformations du modèle autorégressif AR(1) :

$$\begin{aligned}\Delta y_t &= \gamma y_{t-1} + e_t \\ \Delta y_t &= a_0 + \gamma y_{t-1} + e_t \\ \Delta y_t &= a_0 + a_1 t + \gamma y_{t-1} + e_t\end{aligned}$$

Les différences entre ces trois équations sont évidentes. Par rapport à la première, un *drift* a été ajouté dans la seconde et la troisième comporte en plus un trend. Le test de racine unitaire est le

22. Notons que l'estimateur est même super-convergent.

23. Wilmott (1998), *op. cit.*

suivant. Sous H_0 , $\gamma = 0$, ce qui correspond à une racine unitaire puisque $\gamma = \rho - 1$. On effectue le test de racine unitaire en effectuant le test t standard :

$$t = \frac{\hat{\gamma}}{\sqrt{V(\hat{\gamma})}} \sim \tau$$

Selon cette dernière équation, sous H_0 , la statistique t n'a pas sa distribution habituelle. Elle obéit plutôt à une distribution calculée par Dickey-Fuller, nommée τ . Les valeurs critiques respectives de t pour les trois équations pour des seuils de 1 % à 10 % sont reportées au tableau 9.2.

TABLEAU 9.2 Valeurs critiques asymptotiques pour les tests de racine unitaire

Statistique du test	1 %	2,5 %	5 %	10 %
τ_{sc}	-2,56	-2,23	-1,94	-1,62
τ_c	-3,43	-3,12	-2,86	-2,57
τ_{ct}	-3,96	-3,66	-3,41	-3,13

Source : Davidson, R. et J.G. McKinnon (1993), *Estimation and Inference in Econometrics*, Oxford University Press, New York.

À titre d'exemple, si l'on applique le test Dickey-Fuller à la troisième régression, on rejettera l'hypothèse nulle pour un seuil $\alpha = 5\%$ si la statistique t de γ est supérieure au τ_{ct} qui est égal en valeur absolue à 3,41. Cela indiquerait l'absence d'une racine unitaire au niveau $\alpha = 5\%$.

Envisageons maintenant le test Dickey-Fuller élargi publié en 1981, abrégé par ADF²⁴. Pour construire ce test, on effectue la régression suivante :

$$\Delta y_t = a_0 + a_1 t + a_2 t^2 + \dots + a_n t^n + \gamma y_{t-1} + \sum_{i=2}^p \beta_i \Delta y_{t-i+1} + e_t$$

La version élargie du test de Dickey-Fuller comporte donc des retards additionnels sur la variable Δy . Cet ajout vise à prendre en compte la

24. ADF est l'abréviation de *Augmented Dickey-Fuller*.

présence éventuelle d'autocorrélation dans les résidus. Dans la pratique, on se limite au cas pour lequel $n = 1$ et $p = 2$. Il en résulte :

$$\Delta y_t = a_0 + a_1 t + \gamma y_{t-1} + \beta_2 \Delta y_{t-1} + e_t$$

Les hypothèses testées sont :

$$H_0 : \gamma = 0$$

contre :

$$H_1 : \gamma \neq 0$$

où $\gamma = \rho - 1$. On rejette H_0 si la valeur absolue de t est supérieure à τ_{ct} . Sous H_0 , la distribution asymptotique du test t associée à cette régression est la même que celle du test DF présentée précédemment. On peut aussi recourir à cette dernière régression pour tester si la série financière est TS ou DS de la façon suivante. On calcule les statistiques t associées aux paramètres γ et a_1 . Si ces coefficients sont statistiquement différents de zéro, on est présence d'une série TS. Par ailleurs, si le coefficient γ n'est pas significativement différent de 0, alors la série est DS, ce qui implique que ρ est alors égal à 1. Ce test est cependant plus ou moins rigoureux.

Phillips et Perron (1988) ont proposé une généralisation des tests Dickey-Fuller qui impose moins de restrictions sur la distribution des termes d'erreur. Plus précisément, Dickey et Fuller supposaient que les résidus de leurs régressions étaient indépendants et de distribution homogène. Le test Phillips-Perron suppose seulement que les résidus sont faiblement dépendants et qu'ils peuvent être distribués de façon hétérogène. La régression qu'ils proposent est la suivante :

$$y_t = \mu + \beta \left(\frac{t-T}{2} \right) + \alpha y_{t-1} + e_t$$

où $t = 1, \dots, T$. Sous H_0 , les données sont générées par :

$$y_t = y_{t-1} + e_t$$

soit une marche aléatoire où $E(e_t) = 0$. Perron et Phillips ont modifié les statistiques t de Dickey-Fuller pour prendre en compte l'hétérogénéité des termes d'erreur. Il faut calculer les statistiques t_μ , t_α et t_β qui sont les tests t usuels pour l'hypothèse nulle : $\mu = 0$, $\alpha = 1$ et $\beta = 0$. Par la suite, il suffit simplement de calculer les statistiques de Perron-

Phillips associées à ces trois tests. Les valeurs critiques de ces statistiques de Perron-Phillips sont identiques à celles qui apparaissent dans la table de Dickey-Fuller. Pour plus de détails à ce sujet, on consultera Enders (1995)²⁵. Par la suite, Perron (1989) a élargi le test de racine unitaire au cas des changements structurels. Il considère l'hypothèse d'un seul sursaut dans le niveau du processus de racine unitaire contre l'alternative d'un changement au niveau du terme constant d'un processus TS.

16. COINTÉGRATION

La cointégration se rattache à la corrélation de deux séries dans le temps. On dit que deux séries y_t et x_t sont cointégrées si les résidus de la régression de y_t sur x_t sont stationnaires :

$$y_t = \alpha + \beta x_t + e_t$$

Par exemple, si y_t et x_t sont intégrées d'ordre 1 et si e_t est intégrée d'ordre 0, on a cointégration des deux séries. Dans ce cas, même si les deux séries sont non stationnaires, la régression par les MCO ne sera pas fallacieuse en autant que les dites séries soient cointégrés. L'une des implications est la suivante. Si les séries ne sont pas stationnaires, il n'est pas nécessairement obligatoire de les stationnariser pour autant qu'elles soient cointégrés. Par ailleurs, lorsque les séries sont cointégrées, l'estimateur des MCO est super-convergent : $\hat{\beta}$ est alors un excellent estimateur de β puisque lorsque les deux séries sont cointégrées, l'estimateur des MCO converge plus rapidement qu'autrement. La régression par les MCO décrit alors un équilibre à long terme, soit un équilibre stationnaire entre y_t et x_t .

Pour tester si y_t et x_t sont cointégrées, il suffit de recourir à un test de racine unitaire sur les résidus, tel le test Dickey-Fuller. Les étapes du test tel que développé par Engle et Granger (1987)²⁶ sont les suivantes : 1) on régresse y_t sur x_t :

$$y_t = \alpha + \beta x_t + e_t$$

25. Enders, W. (1995), *Applied Econometric Time-Series*, John Wiley and Sons, New York.

26. Engle, R.F. et C.W.J. Granger (1987), « Cointegration and Error Correction : Representation, Estimation and Testing », *Econometrica*, 55, p. 251-276.

on en déduit $\hat{\varepsilon}_t$; 2) on calcule :

$$\Delta \hat{\varepsilon}_t = \hat{\varepsilon}_t - \hat{\varepsilon}_{t-1}$$

et on régresse ensuite Δx_t sur x_{t-1} :

$$\Delta \hat{\varepsilon}_t = (\rho - 1)\hat{\varepsilon}_{t-1} + v_t$$

$$\Delta \hat{\varepsilon}_t = \rho^* \hat{\varepsilon}_{t-1} + v_t$$

3) on effectue le test bilatéral suivant :

$$H_0 : \rho = 1 \text{ ou } \rho^* = 0$$

$$H_1 : \rho \neq 1 \text{ ou } \rho^* \neq 0$$

Ce test est le test t habituel :

$$t = \frac{\rho^* - 0}{\sqrt{V(\rho^*)}}$$

4) on rejette H_0 si : $|t| > |\tau_c|$. Il est à noter que les valeurs critiques doivent être recalculées puisque $\hat{\varepsilon}_t$ est le résultat d'une régression dans laquelle le vecteur de cointégration est lui-même estimé. Pour plus de détails, voir Mills (1999)²⁷.

Application

Pour illustrer les tests de racines unitaires et de cointégrations, considérons l'exemple suivant. Cet exemple se fonde sur la relation entre le prix à terme et le prix au comptant d'un instrument financier. Nous abordons le cas de l'acceptation bancaire²⁸ canadienne à trois mois et du contrat à terme écrit sur cet instrument. Le contrat à terme sur l'acceptation bancaire se transige à la Bourse de Montréal et est désigné par l'acronyme BAX. Théoriquement, il existe une relation stricte entre le prix au comptant d'un instrument financier et son prix à terme. Mais dans la pratique, les prix peuvent dévier de cette relation. Il reste que ces deux catégories de prix devraient être des séries

27. Mills, T.C. (1999), *The Econometric Modelling of Financial Time Series*, 2^e édition, Cambridge University Press, p. 269.

28. Une acceptation bancaire est un titre de financement à court terme émis par une entreprise et garanti par une banque.

cointégrées. Cette cointégration est d'ailleurs à la base de la couverture d'un portefeuille de titres à court terme par les BAX. N'était la cointégration, la couverture serait impensable. Pour tester la cointégration entre ces deux séries, nous appliquons une version du test de Engle et Granger qui, en même temps, illustre bien le test de racine unitaire, ici la version Dickey-Fuller élargie.

Dans un premier temps, nous régressons le prix au comptant (S) sur le prix à terme (F). Les données sont journalières et s'étirent du début à la fin de 1997. Les résultats de cette régression apparaissent au tableau 9.3.

TABLEAU 9.3

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	76.30243	0.116647	654.1325	0.0000
F	0.236687	0.001209	195.7090	0.0000
R-squared	0.993567	Mean dependent var	99.13103	
Ajusted R-squared	0.993541	S.D. dependent var	0.100913	
S.E. of regression	0.008110	Akaike info criterion	-6.783403	
Sum squared resid	0.016313	Schwarz criterion	-6.755231	
Log likelihood	849.9254	F-statistic	38302.01	
Durbin-Watson stat	1.828987	Prob(F-statistic)	0.000000	

Nous retenons les résidus de cette régression et nous effectuons la régression auxiliaire suivante qui est en fait le test ADF :

$$\Delta \hat{e}_t = c + \rho^* \hat{e}_{t-1} + \rho_2 \Delta \hat{e}_{t-1} + \beta^* t + u_t$$

Les résultats apparaissent au tableau 9.4.

TABLEAU 9.4 Augmented Dickey-Fuller Unit Root Test on RES

ADF Test Statistic	-8.668564	1 % Critical Value*	-3.9983
		5 % Critical Value	-3.4292
		10 % Critical Value	-3.1378

* MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation

Dependent Variable : D(RES)

Method : Least Squares

Date : 10/03/00 Time : 13:17

Sample (ajusted) : 3 250

Included observations : 248 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RES(-1)	-0.740128	0.085381	-8.668564	0.0000
D(RES(-1))	-0.191772	0.063091	-3.039631	0.0026
C	0.000404	0.001021	0.395437	0.6929
@TREND(1)	-2.98E-06	7.07E-06	-0.422159	0.6733
R-squared	0.479433	Mean dependent var	1.45E-05	
Ajusted R-squared	0.473033	S.D. dependent var	0.010974	
S.E. of regression	0.007966	Akaike info criterion	-6.811198	
Sum squared resid	0.015485	Schwarz criterion	-6.754530	
Log likelihood	848.5886	F-statistic	74.90664	
Durbin-Watson stat	2.000846	Prob(F-statistic)	0.000000	

La statistique t associée à $\hat{\epsilon}_{t-1}$ est égale à -8.66. Cette statistique doit être comparée aux valeurs critiques asymptotiques pour les tests de cointégration. En effet, ces statistiques diffèrent des valeurs critiques habituelles des tests de racines unitaires. Un aperçu de ces valeurs critiques est présenté au tableau 9.5.

TABLEAU 9.5

Statistique du test	1 %	2,5 %	5 %	10 %	97,5 %
m = 2					
τ_c	-3,90	-3,59	-3,34	-3,04	-0,30
τ_{ct}	-4,32	-4,03	-3,78	-3,50	-1,03

Source : Davidson, R. et J.G. MacKinnon (1993), *op. cit.*

Comme on peut le constater, la statistique t associée à $\hat{\epsilon}_{t-1}$ dépasse largement la valeur de τ_{ct} au seuil de 5 %, ce qui nous conforte dans notre attente d'une forte cointégration entre le prix au comptant et le prix à terme des acceptations bancaires.

Revenons à la régression de S sur F . On constate que le coefficient de régression associé à F est approximativement égal à 0,24. Or, cette information est très importante pour une opération de couverture, car elle représente la valeur absolue du ratio de couverture (h) :

$$h = -\frac{\Delta S}{\Delta F} = -0,24$$

Ce ratio nous indique que pour chaque contrat détenu au comptant, il faut vendre 0,24 contrat à terme pour couvrir sa position. En effet, la relation qui relie le nombre de contrats à terme (NF) au nombre de contrats au comptant (NS) est la suivante :

$$NF = -NS \times \frac{\Delta S}{\Delta F}^{29}$$

29. Pour plus de détails, voir : Racicot, F.-É. et R. Théoret (2000), *op. cit.*

L'HÉTÉROSCÉDASTICITÉ CONDITIONNELLE (ARCH)

Ce chapitre se penche sur une propriété de certaines séries économiques ou financières, à savoir que leur volatilité ou leur variance se modifie dans le temps. Pour traiter ce problème en économétrie financière furent introduits les modèles d'hétéroscédasticité conditionnelle. Ces modèles sont relativement récents. Le principal instigateur est Engle (1982). Dans un article intitulé «Autoregressive Conditional Heteroscedasticity with Estimate of the Variance of United Kingdom Inflation», paru dans *Econometrica* en 1982, Engle a proposé le modèle ARCH, soit l'acronyme de : *AutoRegressive Conditional Heteroscedasticity*. Bollerslev, dans un article paru en 1986 dans le *Journal of Econometrics* et intitulé «Generalized Autoregressive Conditional Heteroscedasticity», a présenté une version généralisée du modèle d'Engle : le GARCH. Toute une panoplie de versions plus sophistiquées de ce modèle devait s'ensuivre. L'ensemble de ces modèles fait l'objet de ce chapitre.

1. NOTIONS D'ESPÉRANCES CONDITIONNELLE ET NON CONDITIONNELLE ; NOTIONS DE VARIANCES CONDITIONNELLE ET NON CONDITIONNELLE

Pour introduire ces notions, nous considérons un modèle du taux d'intérêt à court terme. Pour ce faire, on recourt à un processus autorégressif AR(1) stationnaire :

$$y_t = a_0 + a_1 y_{t-1} + e_t$$

où $e_t \sim \text{WN}(0, \sigma^2)$. Dans cette équation, y_t est la valeur du taux d'intérêt au temps t . Nous voulons prévoir y_{t+1} en utilisant l'espérance non conditionnelle. On procède comme suit :

$$y_{t+1} = a_0 + a_1 y_t + e_{t+1}$$

$$E(y_{t+1}) = a_0 + a_1 E(y_t) + E(e_{t+1})$$

L'espérance non conditionnelle sert à calculer des prévisions à long terme d'une variable, soit sa moyenne à long terme. Sachant par ailleurs que la moyenne d'une série stationnaire est constante, on peut écrire :

$$E(y_{t+1}) = a_0 + a_1 E(y_t)$$

puisque :

$$E(e_t) = E(e_{t+1}) = 0 \quad \forall t$$

Il en résulte que :

$$E(y_{t+1}) = \frac{a_0}{1 - a_1}$$

C'est là l'espérance non conditionnelle de y_{t+1} .

Nous nous intéressons maintenant à la prévision à court terme de y_{t+1} . L'espérance conditionnelle nous donne alors une prévision supérieure à l'espérance non conditionnelle. L'espérance conditionnelle prend en effet compte de toute l'information disponible jusqu'au temps t . Elle suppose que toutes les variables sont connues et fixées jusqu'à cette période. On peut donc écrire

$$E_t(y_{t+1}) = a_0 + a_1 E_t(y_t) + E_t(e_{t+1}) = a_0 + a_1 y_t + 0$$

y_t est en effet connu au temps t et n'est donc pas aléatoire. Par ailleurs e_{t+1} est inconnu à la période t . Son espérance conditionnelle est donc nulle.

Pour introduire les notions de variances conditionnelle et non conditionnelle, nous au recourons au processus autorégressif suivant :

$$y_t = \lambda y_{t-1} + \varepsilon_t$$

où $\varepsilon_t \sim N(0, \sigma^2)$. Nous supposons également que y est une série stationnaire, c'est-à-dire que λ est inférieur à 1. En vertu des calculs antérieurs nous pouvons écrire la variance conditionnelle comme suit :

$$V_t(y_t) = E_t[y_t - E_t(y_t)]^2 = \sigma^2$$

Pour calculer la variance non conditionnelle, nous recourons à l'équation qui relie y_t au décalage de l'innovation. Pour établir cette relation, nous exprimons l'équation de y_t comme suit :

$$y_t(1 - \lambda L) = \varepsilon_t$$

où L désigne l'opérateur de retard. En multipliant les deux côtés de cette expression par $(1 - \lambda L)^{-1}$, on obtient :

$$y_t = (1 - \lambda L)^{-1} \varepsilon_t$$

et en utilisant une propriété bien connue de l'opérateur de retard, on a :

$$y_t = \varepsilon_t + \lambda \varepsilon_{t-1} + \lambda^2 \varepsilon_{t-2} + \dots + \lambda^n \varepsilon_{t-n} \quad n \rightarrow \infty$$

La variance non conditionnelle est donc égale à :

$$V(y_t) = \sigma^2(1 + \lambda^2 + \lambda^4 + \dots)$$

L'expression entre parenthèses est une progression géométrique de raison λ^2 . La variance non conditionnelle de y_t se simplifie donc comme suit :

$$v(y_t) = \frac{\sigma^2}{1 - \lambda^2}$$

La variance non conditionnelle est donc différente de la variance conditionnelle.

2. L'HÉTÉROSCÉDASTICITÉ CONDITIONNELLE ET LES FAITS

Traditionnellement, l'hétéroscédasticité était associée aux données en coupe instantanée (*cross sectional data*), les séries temporelles étant étudiées dans un contexte d'homoscédasticité. En analysant les données macroéconomiques, Engle (1982) et Cragg (1982) ont trouvé que la variance, dans les séries temporelles, était moins stable qu'il n'était

généralement supposé. Pour pallier à ce problème dans son modèle d'inflation, Engle a réalisé que les grandes et petites erreurs de prévision étaient groupées (*clustered*), ce qui suggérait une forme d'hétéroscédasticité dans laquelle la variance de l'erreur de prévision dépend de l'importance de l'erreur précédente. En examinant certaines séries financières comme les taux d'intérêt et les cours boursiers, on remarque le même phénomène. On peut donc dire que la variance de ces séries est hétéroscédastique dans le temps.

Une autre caractéristique des séries financières est leur leptokurticité, ce qui les éloigne d'une distribution normale. Gouriéroux (1992)¹ rapporte les coefficients de kurtosis de certains titres cotés en bourse ainsi que ceux de certains métaux. Tous comportent un coefficient plus élevé que celui qui est associé à la loi normale. Les titres aurifères ont même un coefficient de 11,4, donc très éloigné de celui qui est associé à la loi normale, qui se situe à trois. Selon Gouriéroux, les modèles ARCH sont de nature à modéliser la leptokurticité présente dans les données financières puisque si l'on calcule le coefficient de kurtosis associé au modèle ARCH, on trouvera que celui-ci génère des coefficients supérieurs à 3.

3. LE MODÈLE ARCH

On a le modèle de régression suivant :

$$y_t = \beta^T \mathbf{x}_t + e_t$$

où \mathbf{x}_t n'est pas aléatoire et où e_t suit le processus suivant :

$$e_t = u_t \left[\alpha_0 + \alpha_1 e_{t-1}^2 \right]^{1/2}$$

où $\alpha_0 > 0$ et $0 < \alpha_1 < 1$, ces hypothèses assurant une variance conditionnelle finie et non négative. On postule ici que $u_t \sim N(0, 1)$ et $E(u_t e_{t-1}) = 0$.

Calculons l'espérance conditionnelle de l'innovation :

$$E(e_t | e_{t-1}, e_{t-2}, \dots) = E \left[u_t (\alpha_0 + \alpha_1 e_{t-1}^2)^{1/2} | e_{t-1}, e_{t-2}, \dots \right]$$

1. Gouriéroux, C. (1992), *Modèles ARCH et applications financières*, Economica, Paris.

Désignons par E_{t-1} l'espérance conditionnelle à l'information disponible jusqu'en $t-1$: on suppose alors que cette information est une donnée pour les fins de l'analyse. On peut donc écrire :

$$\begin{aligned} E_{t-1}(e_t) &= E_{t-1}\left[u_t(\alpha_0 + \alpha_1 e_{t-1}^2)^{1/2}\right] \\ &= E_{t-1}(u_t)E_{t-1}\left((\alpha_0 + \alpha_1 e_{t-1}^2)^{1/2}\right) \\ &= 0 \end{aligned}$$

Passons maintenant au calcul l'espérance non conditionnelle de l'innovation.

$$E(e_t) = E\left[u_t(\alpha_0 + \alpha_1 e_{t-1}^2)^{1/2}\right]$$

et puisque u_t et e_{t-1} sont indépendants :

$$E(e_t) = E(u_t)E\left((\alpha_0 + \alpha_1 e_{t-1}^2)^{1/2}\right) = 0$$

Notons encore une fois que $E(\cdot)$, soit l'espérance non conditionnelle, sert à calculer une prévision à long terme associée à la moyenne. Par contre, $E_{t-1}(\cdot)$, soit l'espérance conditionnelle, sert à calculer une prévision à court terme en utilisant l'information disponible jusqu'au temps $t-1$.

La variance conditionnelle de l'innovation, identifiée par h_t est égale à :

$$\begin{aligned} h_t = \sigma_t^2 &= V(e_t | e_{t-1}, e_{t-2}, \dots) = E(e_t^2 | e_{t-1}, e_{t-2}, \dots) \\ &= E_{t-1}\left(u_t^2(\alpha_0 + \alpha_1 e_{t-1}^2)\right) = E_{t-1}(u_t^2)E_{t-1}(\alpha_0 + \alpha_1 e_{t-1}^2) = 1 \times (\alpha_0 + \alpha_1 e_{t-1}^2) \\ &= (\alpha_0 + \alpha_1 e_{t-1}^2) \end{aligned}$$

Il est certes évident que :

$$\begin{aligned} V(y_t | y_{t-1}, y_{t-2}, \dots) &= V(e_t | e_{t-1}, e_{t-2}, \dots) \\ &= \alpha_0 + \alpha_1 e_{t-1}^2 = \sigma_t^2 \end{aligned}$$

Tel est le modèle ARCH(1) proposé par Engle en 1982. Il est à noter que l'on suppose que $\alpha_0 > 0$ et $0 < \alpha_1 < 1$, ceci pour obtenir une variance conditionnelle finie, c'est-à-dire s'assurer de la stationnarité du pro-

cessus et de la non-négativité de la variance conditionnelle. On pourrait par ailleurs démontrer la supériorité des prévisions effectuées à partir de l'espérance conditionnelle, cela parce que l'erreur de prévision est alors inférieure à celle qui est associée à l'espérance non conditionnelle. En effet, comme nous l'avons démontré dans la section précédente, l'erreur de prévision associée à l'espérance conditionnelle est de σ^2 et celle reliée à l'espérance non conditionnelle est

de $\frac{\sigma^2}{1-\lambda^2}$. Comme $0 < \lambda < 1$, la variance conditionnelle est donc inférieure à la variance non conditionnelle.

4. ESTIMATION DU MODÈLE ARCH

Tel qu'il vient d'être présenté, le modèle ARCH respecte les hypothèses du modèle classique des MCO. Ses paramètres peuvent donc être estimés par la méthode des MCO puisque la variance non conditionnelle n'est pas hétéroscédastique : l'estimateur des MCO est alors efficace. On estimerait alors les paramètres de la variance conditionnelle par une régression auxiliaire ou artificielle. Par contre, il existe un estimateur non linéaire qui prend en compte l'hétéroscédasticité conditionnelle et qui est plus efficace asymptotiquement que les MCO. Cette méthode, telle que préconisée par Engle (1982), est le maximum de vraisemblance. Écrivons cette fonction de vraisemblance en posant que $e_t \sim N(0, \sigma_t^2) = N(0, \alpha_0 + \alpha_1 e_{t-1}^2)$. La densité conjointe des innovations est :

$$f(e_1 e_2 \dots e_T | \beta, \alpha_0, \alpha_1) = \prod_{t=1}^T f(e_t | \beta, \alpha_0, \alpha_1)$$

où

$$f(e_t) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp \left[-\frac{1}{2} \left(\frac{e_t - E(e_t)}{\sigma_t} \right)^2 \right]$$

Pour passer à la densité conjointe des observations sur y , on utilise la transformation jacobienne suivante :

$$f(y_t) = f(e_t) \left| \frac{de_t}{dy_t} \right|$$

On pourra vérifier facilement que $\left| \frac{de_t}{dy_t} \right| = 1$. Il s'ensuit :

$$f(y_t | \beta, \alpha_0, \alpha_1) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp \left[-\frac{1}{2} \left(\frac{y_t - \beta^T \mathbf{x}_t}{\sigma_t} \right)^2 \right]$$

Ceci implique :

$$L(\beta, \alpha_0, \alpha_1 | y_1, y_2, \dots, y_T) = \prod_{t=1}^T f(y_t | \beta, \alpha_0, \alpha_1)$$

On recherche donc les valeurs des paramètres qui maximisent la probabilité de générer l'échantillon observé. Élaborons la fonction de vraisemblance :

$$\begin{aligned} L(\beta, \alpha_0, \alpha_1 | y_1, y_2, \dots, y_T) &= \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp \left[-\frac{1}{2} \left(\frac{y_t - \beta^T \mathbf{x}_t}{\sigma_t} \right)^2 \right] \\ &= (2\pi)^{-T/2} \sum_{t=1}^T (\sigma_t^2)^{-1/2} \exp \left(-\frac{1}{2} \sum_{t=1}^T \left(\frac{y_t - \beta^T \mathbf{x}_t}{\sigma_t} \right)^2 \right) \end{aligned}$$

Exprimons cette expression sous forme logarithmique :

$$\ln L = -\frac{T}{2} \ln 2\pi - \frac{1}{2} \ln \left(\sum_{t=1}^T h_t \right) - \frac{1}{2} \sum_{t=1}^T \frac{(y_t - \beta^T \mathbf{x}_t)^2}{h_t}$$

où $h_t = \sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2$. L'algorithme d'optimisation utilisé pour solutionner cette équation à déjà été présenté au chapitre 4. Enders (1995) montre comment on peut trouver les valeurs des paramètres qui maximisent la fonction de vraisemblance en utilisant le logiciel RATS. Ce programme apparaît au tableau 10.1.

TABLEAU 10.1

```

NONLIN β α0 α1
FRML ε = y - βx
FRML h = α0 + α1*εt-12
FRML LIKELIHOOD = - 0.5*[log(ht) + (εt2/ht)]
COMPUTE β = initial guess, α0 = initial guess, α1 = initial guess
MAXIMISE(RECURSIVE) LIKELIHOOD 2 end

```

Source : Enders (1995).

Dans le langage EViews, la fonction de vraisemblance d'un modèle de régression linéaire simple se formule comme au tableau 10.2.

TABLEAU 10.2

```

smpl 1 n
logl fctvrai
fctvrai.append @logl logl1
fctvrai.append res=y-c(1)-c(2)*x
fctvrai.append logl1=log(@dnorm(res/@sqrt(c(3)))) -log(c(3))/2
fctvrai.ml(b, d)

```

Pour élargir ce programme au modèle ARCH, le lecteur n'aura qu'à ajouter une ligne qui définit le modèle ARCH des résidus de la régression et à introduire cette variable dans la fonction de vraisemblance *logl1*.

5. GÉNÉRALISATION DU MODÈLE ARCH

5.1. Le modèle ARCH(q)

Le modèle ARCH(1) ne comporte qu'un seul retard sur l'innovation élevée au carré, c'est-à-dire :

$$\sigma_t^2 = h_t = \alpha_0 + \alpha_1 e_{t-1}^2$$

Par contre, dans le modèle ARCH(q), q retards sont utilisés :

$$E_{t-1}(e_t^2) = \sigma_t^2 = h_t = \alpha_0 + \alpha_1 e_{t-1}^2 + \dots + \alpha_q e_{t-q}^2 = \alpha_0 + \alpha(L)e_t^2$$

De cette dernière équation, il résulte que l'on peut écrire l'innovation comme suit :

$$e_t = u_t h_t^{1/2} = u_t \left(\sqrt{\alpha_0 + \sum_{i=1}^q \alpha_i e_{t-i}^2} \right)$$

Pour avoir une variance conditionnelle finie et positive, on impose que $\alpha_0 > 0$ et que les racines caractéristiques du $\alpha(L)$ se situent à l'extérieur du cercle unitaire.

Bollerslev (1986)² généralise le modèle de Engle (1982) en permettant à la variance conditionnelle de suivre un processus ARMA(p, q). Le modèle général GARCH(p, q) s'écrit comme suit :

$$h_t = \alpha_0 + \sum_{i=1}^q \alpha_i e_{t-i}^2 + \sum_{j=1}^p \gamma_j h_{t-j} = \alpha_0 + \alpha(L)e_t^2 + \gamma(L)h_t$$

Il est à noter que le modèle GARCH(1, 1) approxime de très près un modèle ARCH(q) pour un indice q assez élevé. Ce modèle s'écrit comme suit :

$$h_t = \alpha_0 + \alpha_1 e_{t-1}^2 + \gamma_1 h_{t-1}$$

La variance conditionnelle de l'innovation se calcule comme suit :

$$V(e_t | e_{t-1}, e_{t-2}, \dots) = h_t = \alpha_0 + \alpha_1 e_{t-1}^2 + \gamma_1 h_{t-1}$$

Pour s'assurer de la stationnarité du processus et du caractère fini de la variance conditionnelle, on impose aux trois paramètres d'être positifs ainsi que la relation suivante : $\alpha_1 + \gamma_1 < 1$. Si $\alpha_1 + \gamma_1 = 1$, on a le modèle IGARCH(1, 1), soit le *integrated* GARCH.

Soulignons finalement qu'un modèle GARCH peut représenter de façon parcimonieuse un processus ARCH(q), par exemple, un

2. Bollerslev, T. (1986), « Generalized Autoregressive Conditional Heteroscedasticity », *Journal of Econometrics*, vol. 31, p. 307-327.

processus GARCH(1, 1) a pu se révéler une approximation parcimonieuse d'un processus ARCH(8). Il en résulte moins de restrictions sur les paramètres estimés.

5.2. Le modèle ARCH-M

Ce modèle, dû à Engle, Lilien et Robins (1987)³, s'écrit comme suit :

$$y_t = \mathbf{X}_t \boldsymbol{\beta} + \delta h_t + e_t$$

Dans cette équation, e_t obéit au modèle GARCH suivant :

$$e_t = \left(\alpha_0 + \sum_{i=1}^q \alpha_i e_{t-i}^2 + \sum_{j=1}^p \gamma_j h_{t-j} \right)^{1/2} u_t = h_t^{1/2} u_t$$

On voit que, dans ce modèle, la moyenne de y_t dépend aussi de la variance conditionnelle. Cette classe de modèles se prête bien à l'analyse des actifs financiers. Contrairement au CAPM classique, la prime de risque n'est pas ici considérée comme fixe mais devient plutôt une fonction croissante de la variance conditionnelle des rendements de l'actif financier. Engle, Lilien et Robins se sont servis de ce modèle pour modéliser l'écart de rendement entre un actif à long terme et un actif à court terme. Leur équation est la suivante :

$$y_t = u_t + e_t = \boldsymbol{\beta} + \delta h_t + e_t$$

où

y_t = écart de rendement entre un actif à long terme et celui d'un bon du Trésor ;

u_t = prime de risque incorporée dans l'actif à long terme ;

e_t = terme d'erreur.

3. Engle, R., D. Lilien et R. Robins (1987), « Estimating Time Varying Risk Premia in the Term Structure : The ARCH-M Model », *Econometrica*, 55, p. 391-407.

5.3. Le modèle EGARCH

Le modèle EGARCH (*exponential GARCH*), dû à Nelson (1991)⁴, a été développé pour modéliser non seulement l'excès de leptokurticité mais aussi les effets asymétriques qu'ont les rendements sur la volatilité. Dans la littérature, cet effet est appelé : *leverage effect*. Il s'agit d'une corrélation négative entre les rendements présents et la volatilité future. En effet, une baisse du rendement d'une action est associée à une diminution de la valeur marchande de l'avoir des actionnaires de l'entreprise qui l'a émise. Il en résulte une hausse du levier de cette entreprise, soit le ratio de sa dette à l'équité. L'entreprise est alors perçue plus risquée sur le marché, ce qui augmente la volatilité future du rendement de cette action. La baisse du rendement de cette action exerce donc un effet de levier sur sa volatilité future⁵. Le modèle EGARCH(p, q) se formule comme suit :

$$\ln(h_t) = \ln(\sigma_t^2) = \alpha_0 + \sum_{i=1}^q \alpha_i \left[\phi u_{t-i} + \theta \left(|u_{t-i} - E|u_{t-i}| \right) \right] + \sum_{j=1}^p \gamma_j \ln(\sigma_{t-j}^2)$$

où $u_t \sim \text{NID}(0, 1)$ et $e_t = u_t h_t^{1/2} = u_t \sigma_t$. On suppose ici que les conditions concernant la stationnarité du processus sont respectées. Expliquons sommairement le deuxième terme du logarithme de la variance conditionnelle. Rappelons auparavant que dans le modèle GARCH classique, la volatilité dépend de l'importance des chocs sur l'innovation e_t et non de leur signe. Cela néglige le *leverage effect* présent dans les séries financières. Le modèle EGARCH le prend en compte par le biais du deuxième terme de son équation. Contrairement au modèle GARCH, le modèle EGARCH n'impose aucune contrainte de non-négativité sur les paramètres α_i et γ_j de façon à s'assurer que la variance conditionnelle soit positive. Toutefois, comme cela a été spécifié auparavant, le processus est supposé stationnaire.

4. Nelson, D.B. (1991), « Conditional Heteroskedasticity in Asset Returns: A New Approach », *Econometrica*, 59, p. 347-370.

5. Pour plus de détails sur ce sujet, on consultera : Bollerslev, T., R.Y. Chou et K.F. Kroner (1992), « ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence », *Journal of Econometrics*, 52, p. 55-59.

Voici comment le modèle EGARCH intègre le *leverage effect*. Si $\alpha_i \phi < 0$, alors la variance conditionnelle (σ_t^2) tend à augmenter (diminuer) si e_{t-1} est négatif (positif), ceci en conformité avec les faits empiriques qui font apparaître le *leverage effect* au niveau du comportement des rendements des actifs financiers.

5.4. Le modèle TARCH

Le modèle TARCH (*threshold ARCH*), dû à Zakoian (1994)⁶ s'écrit comme suit :

$$y_t = \beta^T \mathbf{x}_t + e_t$$

La variance conditionnelle de l'innovation s'écrit comme suit à l'intérieur de modèle :

$$\sigma_t^2 = \omega + \alpha e_{t-1}^2 + \gamma e_{t-1}^2 d_{t-1} + \beta \sigma_{t-1}^2$$

où $d_t = 1$ si $e_t < 0$ (mauvaise nouvelle) et 0 autrement. Ce modèle veut intégrer l'observation suivante sur les séries temporelles des rendements financiers. En effet, on a remarqué que les mauvaises nouvelles affectent davantage la volatilité que les bonnes. Si γ est significativement positif, il y aura alors évidence de la présence d'un tel phénomène.

Les modèles GARCH ont trouvé de multiples applications en finance moderne. En plus de celle reliée au CAPM que nous relate-rons dans la section « Applications », d'autres chercheurs les ont transposés à la théorie des options, là où la volatilité joue un rôle primordial⁷.

5.5. Prévision à partir du modèle GARCH

On pourrait montrer que le modèle GARCH(1, 1) revient à prévoir y_t à partir de la moyenne pondérée, à poids géométriquement décrois-

6. Zakoian, J.M. (1994), « Threshold Heteroskedastic Models », *Journal of Economic Dynamics and Control*, 18, p. 931-955.

7. Voir à cet effet : Engle, R. et C. Moustafa (1992), « Implied ARCH models from Options Prices », *Journal of Econometrics*, 52, p. 289-311 ; Duan, J.C. (1995), « The GARCH Option Pricing Model », *Mathematical Finance*, 5, p. 13-32. Pour une synthèse des articles dans le domaine des modèles GARCH, voir : Racicot, F.É. (2000), Notes on Nonlinear Dynamics, *op. cit.*

sants, des retards sur l'innovation au carré. Cela revient presque alors à un modèle MA où la variable indépendante est l'innovation au carré. Pour le montrer, supposons le modèle GARCH(1, 1) suivant :

$$\sigma_t^2 = \omega + \alpha e_{t-1}^2 + \beta \sigma_{t-1}^2$$

Par substitution répétée de σ_{t-i}^2 , on obtient :

$$\sigma_t^2 = \omega + \beta\omega + \beta^2\omega + \dots + \alpha e_{t-1}^2 + \alpha\beta e_{t-2}^2 + \alpha\beta^2 e_{t-3}^2 + \dots$$

On voit que les poids appliqués par à e_{t-i}^2 sont $\alpha\beta^{i-1}$. Les poids diminuent donc au taux exponentiel β .

Cette méthode est plus satisfaisante puisque notre objectif est ici d'effectuer le monitoring de la volatilité actuelle. Il est donc approprié de donner davantage de poids aux observations plus récentes pour calculer la volatilité. Tel n'est pas le cas pour la variance classique qui donne un poids de $1/(T - 1)$ à toutes les observations.

Nous voulons maintenant prévoir la volatilité conditionnelle d'une série à partir d'un modèle GARCH(1, 1). Soit :

$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta \sigma_{t-1}^2$$

Nous voulons prévoir dans un premier temps σ_{t+1}^2 . On a :

$$E_t(\sigma_{t+1}^2) = \alpha_0 + \alpha_1 E_t(e_t^2) + \beta E_t(\sigma_t^2) = \alpha_0 + \sigma_t^2(\alpha_1 + \beta)$$

La prévision de la variance conditionnelle dans deux périodes peut s'écrire comme suit :

$$E_t(\sigma_{t+2}^2) = \alpha_0 + \alpha_1 E_t(\sigma_{t+1}^2) + \beta E_t(\sigma_{t+1}^2) = \alpha_0 + (\alpha_1 + \beta) E_t(\sigma_{t+1}^2)$$

En substituant dans cette équation la valeur déjà trouvée pour σ_{t+1}^2 , on a :

$$E_t(\sigma_{t+2}^2) = \alpha_0(1 + \alpha_1 + \beta) + \sigma_t^2(\alpha_1 + \beta)^2$$

On peut réécrire cette expression comme suit :

$$\begin{aligned} E_t(\sigma_{t+2}^2) &= \frac{(1 + \alpha_1 + \beta)(1 - \alpha_1 - \beta)}{(1 - \alpha_1 - \beta)} \alpha_0 + \sigma_t^2(\alpha_1 + \beta)^2 \\ &= \frac{1 - (\alpha_1 + \beta)^2}{1 - \alpha_1 - \beta} \alpha_0 + \sigma_t^2(\alpha_1 + \beta)^2 \end{aligned}$$

Finalement, la prévision dans n périodes de la variance conditionnelle se calcule en remplaçant simplement l'exposant 2 dans l'expression précédente par l'exposant n :

$$\begin{aligned} &= \frac{1 - (\alpha_1 + \beta)^n}{1 - \alpha_1 - \beta} \alpha_0 + \sigma_t^2 (\alpha_1 + \beta)^n \\ &= \frac{\alpha_0}{1 - \alpha_1 - \beta} - \frac{\alpha_0 (\alpha_1 + \beta)^n}{1 - \alpha_1 - \beta} + \sigma_t^2 (\alpha_1 + \beta)^n \\ &= \frac{\alpha_0}{1 - \alpha_1 - \beta} + (\alpha_1 + \beta)^n \left(\sigma_t^2 - \frac{\alpha_0}{1 - \alpha_1 - \beta} \right) \end{aligned}$$

Cette dernière équation peut être utilisée pour calculer des prévisions pour n'importe lequel horizon en fixant n à l'horizon désiré. Quand $\alpha_1 + \beta = 1$, l'espérance conditionnelle de la volatilité se simplifie pour devenir :

$$E_t(\sigma_{t+n}^2) = \sigma_t^2 + n\alpha_0$$

Le modèle GARCH(1, 1) pour lequel $\alpha_1 + \beta = 1$ comporte une racine unitaire de telle sorte que la volatilité présente affecte les prévisions de volatilité dans un futur indéfini. On est alors en présence du GARCH dit intégré, encore désigné par IGARCH(1, 1). Pour des GARCH d'ordres supérieurs : GARCH(p , q), des prévisions sur de multiples périodes peuvent être effectuées de façon similaire.

5.6. Test ARCH

Pour tester la présence d'erreurs de type ARCH dans le modèle suivant :

$$y_t = x_t \beta + e_t$$

on peut effectuer le test LM proposé par Engle (1982). Les étapes du test sont les suivantes :

- 1) On estime d'abord l'équation précédente par les MCO.

- 2) Les résidus estimés de la régression précédente sont alors élevés au carré :

$$\hat{e}_t^2 = (y_t - x_t \hat{\beta})^2$$

- 3) On effectue la régression artificielle suivante :

$$\hat{e}_t^2 = \alpha_0 + \alpha_1 \hat{e}_{t-1}^2 + \dots + \alpha_q \hat{e}_{t-q}^2 + u_t$$

On se sert du R^2 de cette régression pour effectuer le test. En effet, la statistique $(T \times R^2) \xrightarrow{d} \chi^2(q)$. S'il n'y a pas d'effet ARCH, le R^2 de la régression artificielle sera faible. La statistique $(T \times R^2)$ se situera alors sous la valeur critique de la distribution χ^2 . L'hypothèse H_0 , à l'effet de l'absence d'effet ARCH, se formule comme suit :

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_q = 0$$

On rejettera H_0 si $(T \times R^2) > \chi^2$. Par exemple, on rejettera l'hypothèse de l'absence de l'effet ARCH(1) si $T \times R^2 > \chi^2(1) = 3,84$.

Application

Dans cette section, nous coulons le CAPM dans le creuset de la méthode d'estimation dite GARCH-M (GARCH-in-Mean)⁸. Mais auparavant, forts des concepts que nous avons introduits depuis que nous avons examiné le CAPM, nous reconsidérons l'estimation de ce modèle par la méthode des MCO.

Les hypothèses du CAPM sont au nombre de cinq : 1) les investisseurs éprouvent de l'aversion pour le risque ; 2) leurs attentes sont homogènes au plan des rendements, ce qui signifie qu'ils prévoient la même matrice variance-covariance des rendements qui sont censés obéir à une distribution normale ; 3) il existe un actif sans risque ; 4) l'information est gratuite ; 5) il n'existe aucune imperfection de marché. Entre autres, il n'existe pas d'impôts ou de restrictions sur les ventes à découvert.

8. Par association au modèle ARCH-M présenté auparavant dans ce chapitre.

Le modèle du CAPM comprend grosso modo deux équations. La première est la CML (*Capital Market Line*) et la seconde, la SML (*Security Market Line*). Considérons ces deux relations tour à tour.

La CML

La CML est une équation qui relie, lorsque l'équilibre sur les marchés financiers est atteint, les rendements des portefeuilles dits efficients⁹. L'équation de la CML s'écrit comme suit :

$$E(R_p) = R_f + \left[\frac{E(R_m) - R_f}{\sigma_m} \right] \sigma_p$$

où $E(R_p)$ représente la valeur espérée du portefeuille p ; R_f , le taux sans risque; $E(R_m)$, la valeur espérée du rendement du portefeuille du marché; σ_m , l'écart-type du rendement du portefeuille du marché et σ_p , l'écart-type du rendement du portefeuille p . La représentation graphique de la CML apparaît à la figure 10.1.

La pente de la CML est donc : $\left[\frac{E(R_m) - R_f}{\sigma_m} \right]$. Elle représente

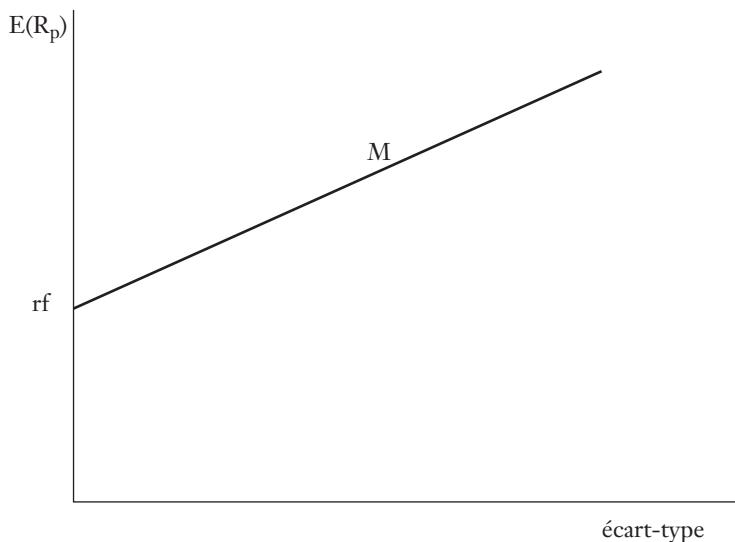
le prix du risque. C'est le rendement excédentaire¹⁰ du portefeuille du marché par unité de risque, ici mesuré par l'écart-type du rendement du portefeuille du marché. Ce prix est commun à tous les portefeuilles. Ce qui diffère d'un portefeuille à l'autre, c'est la quantité de risque qui y est emmagasinée, mesurée par σ_p .

On l'aura constaté à la lecture de la figure 10.1, les portefeuilles efficients sont de simples combinaisons linéaires de l'actif sans risque et du portefeuille du marché. Entre les points R_f et M , les investisseurs détiennent pour partie l'actif sans risque et pour partie le porte-

-
9. Un portefeuille est dit effICIENT s'il ne comporte pas de risque non systématique, dit encore risque résiduel ou non relié au risque présenté par le portefeuille du marché.
 10. C'est-à-dire le rendement au-delà du taux sans risque que les investisseurs exigent pour supporter le risque incorporé dans le portefeuille du marché.

feuille du marché. Au-delà de M , les investisseurs ont vendu à découvert l'actif sans risque, c'est-à-dire qu'ils ont emprunté de manière à investir davantage dans le portefeuille du marché.

FIGURE 10-1 La CML

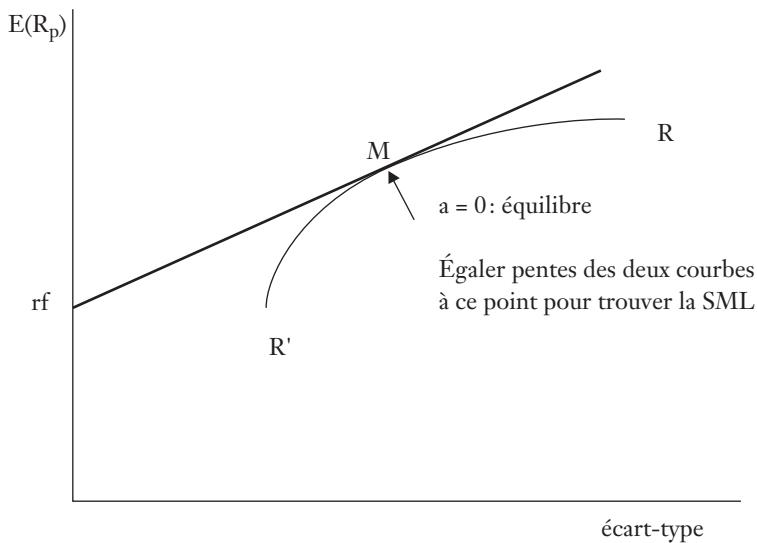


La CML aboutit donc au phénomène dit de la *séparation* de portefeuilles. En effet, en vertu de la CML, les investisseurs n'exercent leurs choix que sur deux catégories d'actifs ou fonds : le fonds constitué par l'actif sans risque et le fonds constitué par le portefeuille du marché. Tout investisseur détient donc les actifs risqués dans les mêmes proportions, qui sont égales aux pondérations que revêtent ces actifs dans le portefeuille du marché, et cela quel que soit le niveau de son degré d'aversion au risque. Ce qui diffère d'un individu à l'autre, ce sont les pondérations respectives qu'il accorde au fonds sans risque (actif sans risque) et au fonds risqué (portefeuille de marché). Ces pondérations sont déterminées par son degré d'aversion au risque. Plus un individu craint le risque, plus il donnera la part belle au fonds sans risque dans son portefeuille.

La SML

La SML est une relation d'équilibre entre le rendement d'un titre (ou d'un portefeuille) et son bêta respectif. Pour la dériver, nous nous reportons à la figure 10.2.

FIGURE 10.2 Dérivation de la SML



Supposons qu'un portefeuille soit composé de a % d'un actif risqué i et de $(1 - a)$ % du portefeuille du marché M . La frontière efficiente représentée par la courbe $R'MR$ sur la figure 10.2 fournit ces combinaisons rendement-risque de l'actif i et du portefeuille du marché M . Or, l'équilibre des marchés financiers ne peut se situer qu'au point M . En effet, l'actif i a déjà sa pondération d'équilibre au sein du portefeuille du marché. La proportion a n'est donc qu'une demande excédentaire qui prend une valeur nulle lorsque l'équilibre sur les marchés financiers est atteint.

Au point M , il y a égalité entre les pentes de la CML et de la frontière $R'MR$. Pour dériver la SML, soit la relation d'équilibre entre l'espérance du rendement d'un titre et son risque, nous devons donc évaluer les pentes de la CML et de $R'MR$ au point $a = 0$. Nous

savons que la pente de la CML est de : $\frac{E(R_m) - R_f}{\sigma_m}$. Par ailleurs, la

pente de R'MR est de : $\frac{\partial E(R_p) / \partial a}{\partial \sigma(R_p) / \partial a}$. La SML doit donc satisfaire à l'égalité suivante des pentes :

$$\frac{E(R_m) - R_f}{\sigma_m} = \frac{\partial E(R_p) / \partial a}{\partial \sigma(R_p) / \partial a} \Bigg|_{a=0}$$

Nous devons effectuer les deux dérivées qui apparaissent à droite de cette équation. Effectuons d'abord la dérivée qui apparaît au numérateur pour ensuite passer à la dérivée présente au dénominateur. Nous savons que :

$$E(R_p) = aE(R_i) + (1-a)E(R_m)$$

La dérivée du numérateur est donc de :

$$\frac{\partial E(R_p)}{\partial a} = E(R_i) - E(R_m)$$

Passons maintenant à la dérivée présente au dénominateur du terme de droite de l'égalité des pentes. Nous savons que :

$$\sigma(R_p) = \left[a^2 \sigma_i^2 + (1-a)^2 \sigma_m^2 + 2a(1-a)\sigma_{im} \right]^{1/2}$$

Nous devons dériver cette expression par rapport à a . Pour ce faire, nous recourons à la règle de la chaîne. Représentons par y le terme entre crochets dans cette équation. On a donc :

$$\sigma(R_p) = y^{1/2}$$

La règle de la chaîne est donc, dans le cas qui nous intéresse :

$$\frac{\partial \sigma_p}{\partial a} = \frac{\partial \sigma_p}{\partial y} \frac{\partial y}{\partial a}$$

On peut écrire :

$$\frac{\partial \sigma_p}{\partial a} = \frac{1}{2} y^{1/2-1} \frac{\partial y}{\partial a}$$

Et, en remplaçant y par sa valeur¹¹ :

$$\frac{\partial \sigma_p}{\partial a} = \frac{1}{2} \left[a^2 \sigma_i^2 + (1-a)^2 \sigma_m^2 + 2a(1-a) \sigma_{im} \right]^{-1/2} \times \left[2a\sigma_i^2 - 2\sigma_m^2 + 2a\sigma_m^2 + 2\sigma_{im} - 4a\sigma_{im} \right]$$

Cette dérivée doit être évaluée au point $a = 0$, soit à l'équilibre des marchés financiers :

$$\left. \frac{\partial \sigma_p}{\partial a} \right|_{a=0} = \frac{1}{2} (\sigma_m^2)^{-1/2} \left[-2\sigma_m^2 + 2\sigma_{im} \right] = \frac{\sigma_{im} - \sigma_m^2}{\sigma_m}$$

En substituant ces dérivées dans l'expression de l'égalité des pentes de la CML et de R'MR, on obtient finalement :

$$\begin{aligned} \frac{E(R_m) - R_f}{\sigma_m} &= \left. \frac{\partial E(R_p) / \partial a}{\partial \sigma(R_p) / \partial a} \right|_{a=0} \\ \frac{E(R_m) - R_f}{\sigma_m} &= \frac{E(R_i) - E(R_m)}{\frac{\sigma_{im} - \sigma_m^2}{\sigma_m}} \\ \frac{E(R_m) - R_f}{\sigma_m^2} (\sigma_{im} - \sigma_m^2) &= E(R_i) - E(R_m) \\ \left[E(R_m) - R_f \right] \frac{\sigma_{im}}{\sigma_m^2} - E(R_m) + R_f &= E(R_i) - E(R_m) \end{aligned}$$

11. À remarquer que l'on recourt également à la règle de la chaîne pour dériver $(1-a)^2 \sigma_m^2$ par rapport à a. On pose $(1-a) = y$. La règle de la chaîne devient dans ce cas : $\frac{\partial \sigma_m^2 y^2}{\partial y} \frac{\partial y}{\partial a} = 2y\sigma_m^2 \times -1$. En remplaçant y par sa valeur, on obtient finalement : $-2(1-a)\sigma_m^2 = -2\sigma_m^2 + 2a\sigma_m^2$.

À l'équilibre, l'espérance du rendement du titre i est égale à :

$$E(R_i) = R_f + [E(R_m) - R_f] \frac{\sigma_{im}}{\sigma_m^2}$$

Définissons comme suit le bêta du titre i , soit son risque systématique :

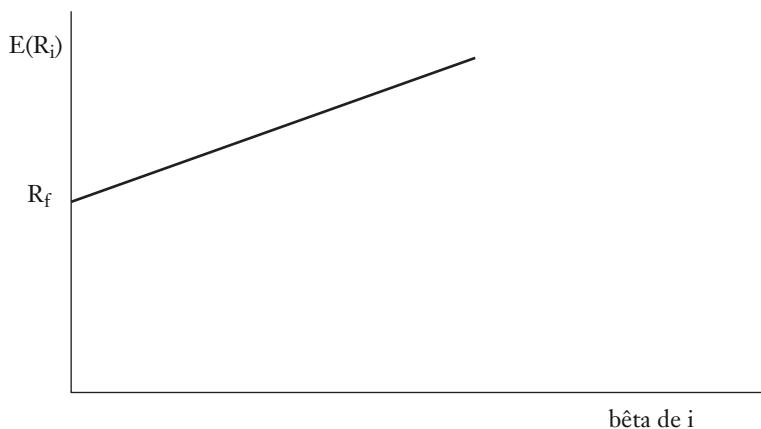
$$\beta_i = \frac{\sigma_{im}}{\sigma_m^2}$$

On obtient l'expression finale de la SML :

$$E(R_i) = R_f + [E(R_m) - R_f] \beta_i$$

En situation d'équilibre sur les marchés financiers, il existe par conséquent une relation linéaire entre le rendement espéré d'un titre et son risque systématique, représenté par son bêta¹². La représentation graphique de la SML apparaît à la figure 10.3.

FIGURE 10.3 **La SML**



12. Pour un titre particulier, l'écart-type de son rendement est une mesure inadéquate de son risque puisque cette mesure comporte une part de risque non systématique. Or, les marchés financiers ne rémunèrent que le risque systématique. Le bêta d'un titre, qui est une mesure du risque systématique d'un titre, fait donc figure de mesure plus appropriée de son risque.

On constate que le modèle du CAPM a beaucoup emprunté à la théorie de Markowitz ayant trait à la diversification de portefeuille. La mesure pertinente du risque d'un titre est la covariance entre son rendement et celui du portefeuille du marché, tout comme chez Markowitz la mesure pertinente du risque est la covariance entre son rendement et ceux des autres titres qui constituent un portefeuille. En formulant le CAPM, Sharpe a simplifié la matrice variance-covariance des rendements des titres en reliant le rendement d'un titre à celui du portefeuille du marché et non à celui de chacun des titres du portefeuille. Mais le message est le même. Le marché ne rémunérera que le risque non diversifiable, qui est mesuré par la covariance entre les rendements des titres. Le risque diversifiable est dilué dans un portefeuille bien diversifié et ne sera pas rémunéré sur des marchés financiers efficients.

Relation entre le bêta et la variance du rendement d'un titre

Pour établir la relation entre le bêta d'un titre et la variance de son rendement, nous recourons au *modèle* du marché, qui relie, au plan empirique, le rendement d'un titre à celui du portefeuille du marché :

$$R_{it} = \alpha_i + \beta_i R_{mt} + \varepsilon_{it}$$

où R_{it} est le rendement du titre i au temps t ; R_{mt} , celui du portefeuille du marché et ε_{it} , le terme d'erreur de la régression. Il est postulé que : $E(R_{mt}, \varepsilon_{it}) = 0$, c'est-à-dire que le rendement du portefeuille du marché est indépendant de l'innovation¹³. L'estimateur des MCO de β_i correspond bien à la définition théorique du bêta d'un titre puisque la méthode des MCO donne au bêta la valeur suivante :

$$\hat{\beta}_i = \frac{\text{Cov}(R_i, R_m)}{\text{Var}(R_m)}$$

13. Nous omettons l'indice t par la suite pour alléger la présentation.

C'est pourquoi on recourt généralement au modèle du marché pour calculer le bêta d'un titre¹⁴. Calculons, à partir de cette équation, la variance du rendement du titre i . L'input à ce calcul est l'espérance de R_i , qui est égale à :

$$E(R_i) = \alpha_i + \beta_i E(R_m)$$

La variance de R_i est donc égale à :

$$\begin{aligned} E[R_i - E(R_i)]^2 &= E[\alpha_i + \beta_i R_m + \varepsilon_i - \alpha_i - \beta_i E(R_m)]^2 \\ &= E[\beta_i (R_m - E(R_m)) + \varepsilon_i]^2 \\ &= E\left\{ \beta_i^2 E[R_m - E(R_m)]^2 + 2\beta_i [R_m - E(R_m)]\varepsilon_i + \varepsilon_i^2 \right\} \\ &= \beta_i^2 E[R_m - E(R_m)]^2 + E(\varepsilon_i^2) \end{aligned}$$

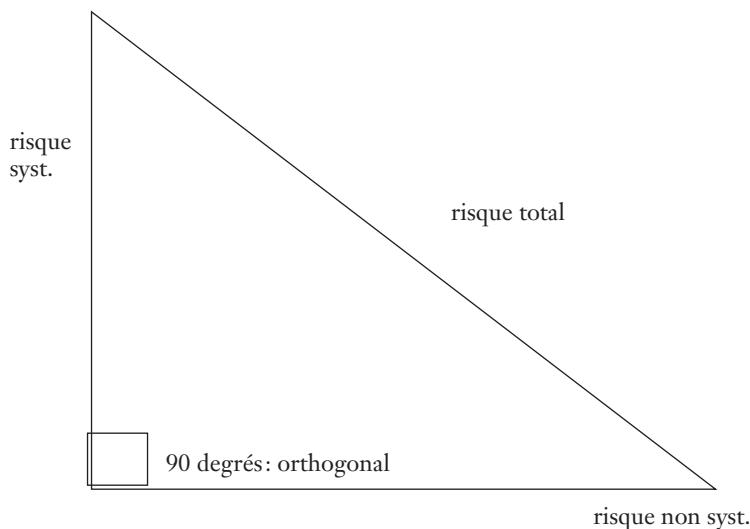
puisque R_m et ε_i sont supposés indépendants. Il suit que :

$$\sigma_i^2 = \beta_i^2 \sigma_m^2 + \sigma_{\varepsilon_i}^2$$

Cette équation indique que le risque total d'un titre comprend deux composantes : le risque systématique, qui est relié à la variance du portefeuille du marché et au bêta du titre, et le risque non systématique, qui est idiosyncratique au titre en question. Le risque systématique est orthogonal au risque non systématique. La représentation graphique de cette orthogonalité apparaît à la figure 10.4 qui n'est rien d'autre qu'une application du célèbre théorème de Pythagore.

14. Attention. Le modèle du marché ne repose sur aucune théorie particulière. Ce n'est pas à proprement parler la version empirique du CAPM, même si on recourt la plupart du temps au modèle du marché pour calculer le bêta. Il y a en effet deux autres modèles qui sont également utilisés dans les tests du CAPM. Le premier est à proprement parler la version empirique du CAPM. Nous le présentons dans la section qui suit. L'équation empirique de ce modèle est la suivante : $R_{it} = R_{ft} + \beta_i [R_{mt} - R_{ft}] + \varepsilon_{it}$. Le second est la ligne empirique du marché (*empirical market line*). Son équation est la suivante : $R_{it} = \hat{\gamma}_{0t} + \hat{\gamma}_{1t} \beta_{it} + \varepsilon_{it}$. Pour plus de détails à ce sujet, on consultera : Copeland, T.E. et J.F. Weston (1988), *Financial Theory and Corporate Policy*, Addison-Wesley, New York, chap. 6, chap. 7 et chap. 10.

FIGURE 10.4



Un portefeuille efficace ne comporte pas de risque résiduel. En vertu de la formule qui vient d'être établie, sa variance s'écrit comme suit :

$$\sigma_p^2 = \beta_p^2 \sigma_m^2$$

Le bêta d'un portefeuille efficace est donc de :

$$\beta_p = \frac{\sigma_p}{\sigma_m}$$

soit le simple ratio des écart-types respectifs des rendements du portefeuille p et du portefeuille du marché. Une autre façon d'arriver à ce résultat est d'écrire le bêta en termes du coefficient de corrélation entre le rendement du portefeuille p et celui du portefeuille du marché, soit ρ_{pm} :

$$\beta_p = \rho_{pm} \frac{\sigma_p}{\sigma_m}$$

Comme la corrélation entre le rendement d'un portefeuille efficient et celui du portefeuille du marché est de 1, on retrouve le résultat précédent :

$$\beta_p = 1 \times \frac{\sigma_p}{\sigma_m} = \frac{\sigma_p}{\sigma_m}$$

Le modèle du marché donne lieu à une expression relativement simple pour la covariance entre les rendements des titres i et j . Par définition, cette covariance est égale à :

$$\sigma_{ij} = E\left\{ \left[R_i - E(R_i) \right] \left[R_j - E(R_j) \right] \right\}$$

En remplaçant R_i , R_j , $E(R_i)$ et $E(R_j)$ par leurs valeurs respectives dans le modèle du marché, on obtient :

$$\begin{aligned} \sigma_{ij} &= E\left\{ \left[\beta_i (R_m - E(R_m)) + \varepsilon_i \right] \left[\beta_j (R_m - E(R_m)) + \varepsilon_j \right] \right\} \\ &= E\left\{ \begin{array}{l} \beta_i \beta_j [R_m - E(R_m)]^2 + \beta_j [R_m - E(R_m)] \varepsilon_i \\ + \beta_i [R_m - E(R_m)] \varepsilon_j + \varepsilon_i \varepsilon_j \end{array} \right\} \end{aligned}$$

Comme $E(R_m, \varepsilon_i) = 0$, $E(R_m, \varepsilon_j) = 0$ et $E(\varepsilon_i, \varepsilon_j) = 0$, on peut écrire :

$$\sigma_{ij} = \beta_i \beta_j E\left[R_m - E(R_m) \right]^2 = \beta_i \beta_j \sigma_m^2$$

Par conséquent, pour calculer la covariance entre les rendements de deux titres, il suffit de connaître les bêtas respectifs des deux titres et la variance du rendement du portefeuille du marché. Si un portefeuille comprend N titres, on n'a besoin que de $(N + 1)$ données pour calculer la matrice variance-covariance, en supposant qu'il n'existe pas de risque résiduel¹⁵ : les N bêtas des N titres et la variance du rendement du portefeuille du marché. Dans le cadre de la théorie de Markowitz, comme on ne relie pas le rendement d'un titre à celui du portefeuille

15. En effet, comme on l'a vu précédemment, les variances des rendements des titres sont affectés par le risque résiduel. Mais on peut supposer que celui-ci est dilué dans un portefeuille bien diversifié.

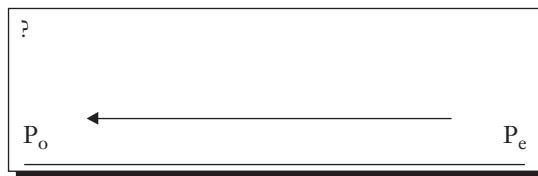
du marché, il faut calculer N variances et $\frac{N^2 - N}{2}$ covariances¹⁶ pour en arriver à construire la matrice variance-covariance. Dans le cadre du modèle du marché, la matrice variance-covariance des rendements des titres s'écrit donc, pour un portefeuille composé de N titres, toujours en supposant la dilution du risque résiduel :

$$\Omega = \sigma_m^2 \begin{bmatrix} \beta_1^2 & \beta_1\beta_2 & \dots & \beta_1\beta_N \\ \beta_2\beta_1 & \beta_2^2 & \dots & \beta_2\beta_N \\ \cdot & \cdot & \cdot & \cdot \\ \beta_N\beta_1 & \beta_N\beta_2 & \dots & \beta_N^2 \end{bmatrix}$$

Cette matrice, à l'évidence symétrique, est donc d'un calcul fort aisé.

Transposition de la théorie du CAPM au calcul des prix des actions¹⁷

La théorie du CAPM est formulée en termes de rendements espérés. Nous voulons maintenant la transposer à la détermination des prix d'équilibre des titres. Nous nous situons dans le contexte d'une période, disons un an. Nous voulons déterminer le prix d'un titre qui paiera dans un an un cash-flow-risqué, désigné par P_e . Ce dernier est donc une variable aléatoire.



16. La matrice variance-covariance est en effet symétrique.

17. Pour rédiger cette section, nous nous inspirons de : Copeland et Weston (1988), *op. cit.*, chap. 7, p. 202-203.

Pour ce faire, nous écrivons la SML dans une forme où apparaît le prix du risque, désigné par λ .

$$E(R_j) = R_f + [E(R_m) - R_f] \frac{\text{Cov}(R_j, R_m)}{\text{Var}(R_m)} = R_f + \lambda \text{Cov}(R_j, R_m)$$

λ est le rendement excédentaire du marché par unité de risque. On peut donc l'assimiler au prix du risque. Par ailleurs, le rendement du titre j est égal à :

$$R_j = \frac{P_e - P_0}{P_0}$$

L'espérance du rendement du titre j est donc de :

$$E(R_j) = \frac{E(P_e) - P_0}{P_0}$$

En substituant cette valeur dans le CAPM, on obtient :

$$\frac{E(P_e) - P_0}{P_0} = R_f + \lambda \text{Cov}(R_j, R_m)$$

En mettant P_0 en évidence, on décrypte l'expression du prix d'équilibre du titre j dans le contexte du modèle du CAPM :

$$P_0 = \frac{E(P_e)}{1 + R_f + \lambda \text{Cov}(R_j, R_m)}$$

Cette expression obéit à la forme générale de l'équation du prix d'un titre dans le contexte d'une période, forme égale à :

$$P_s = \frac{E(\text{CF})}{1 + E(R_s)}$$

où P_s est le prix d'un titre ; $E(\text{CF})$, l'espérance du cash-flow qu'il paiera à la fin de la période et $E(R_s)$, le taux de rendement requis par les investisseurs pour détenir cet actif. Dans sa forme générale, ce taux est égal à :

$$E(R_s) = R_f + \text{prime de risque}$$

Chaque modèle donne une forme particulière à la prime de risque. Dans le cadre du CAPM, l'équation antérieure indique que cette prime est égale à : $\lambda \text{Cov}(R_j, R_m)$, soit le produit du prix du risque par la quantité de risque du titre j , quantité mesurée par la covariance du rendement du titre j avec celui du portefeuille du marché.

Il existe une autre façon de calculer le prix d'équilibre d'un actif risqué dans le cadre du modèle du CAPM qui fait appel cette fois-ci à la technique dite de l'équivalent-certain. Pour y arriver, il suffit d'exprimer la covariance entre les rendements du titre j et du portefeuille du marché en termes du prix du titre j . On a donc :

$$\text{Cov}(R_j, R_m) = \text{Cov}\left(\frac{P_e - P_0}{P_0}, R_m\right)$$

Par définition, cette covariance est égale à :

$$= E\left\{ \left[\frac{P_e - P_0}{P_0} - \frac{E(P_e) - P_0}{P_0} \right] [R_m - E(R_m)] \right\}$$

Cette expression se simplifie comme suit :

$$\begin{aligned} \text{Cov}(R_j, R_m) &= \frac{1}{P_0} E\left\{ [P_e - E(P_e)] [R_m - E(R_m)] \right\} \\ &= \frac{1}{P_0} \text{Cov}(P_e, R_m) \end{aligned}$$

Il suffit maintenant de remplacer, dans la formule antérieure de P_0 , $\text{Cov}(\cdot)$ par cette valeur :

$$P_0 = \frac{E(P_e)}{1 + R_f + \lambda(1/P_0) \text{Cov}(P_e, R_m)}$$

En mettant P_0 en évidence, on trouve finalement :

$$P_0 = \frac{E(P_e) - \lambda \text{Cov}(P_e, R_m)}{1 + R_f}$$

C'est là l'expression finale du prix d'un titre en termes d'équivalent-certain. C'est-à-dire qu'au numérateur de cette formule, on calcule l'équivalent-certain de l'espérance du cash-flow promis par le titre en lui soustrayant une prime de risque. Cette prime est égale au produit du prix du risque et de la quantité de risque du titre, quantité mesurée par la covariance entre le cash-flow aléatoire du titre et le rendement du portefeuille du marché. Après prise en compte du risque du cash-flow au numérateur, on peut alors actualiser l'espérance du cash-flow ainsi corrigée au taux sans risque. Au numérateur de P_0 , on a donc retranché une pénalité pour prendre en compte le risque que présente le cash-flow du titre analysé¹⁸.

18. Dans la pratique, il est courant de relier le prix d'un titre à son bêta respectif à l'aide de l'équation économétrique suivante :

$$\left(\frac{P}{E}\right)_i = \alpha_0 + \alpha_1 (\text{PAYOUT})_i + \alpha_2 (\text{bêta})_i + \alpha_3 (\text{EGR})_i + \varepsilon_i$$

où le ratio $(P/E)_i$ représente le rapport cours-bénéfices de l'action, P étant son prix et E , le bénéfice par action ; PAYOUT , le taux de paiement du dividende à la fin de l'année de l'émetteur de l'action i ; EGR , le taux de croissance des profits au cours des cinq dernières années de l'émetteur de l'action et bêta , bien entendu le bêta de l'action en cause. En fait, cette équation est une transposition du modèle de Gordon au plan empirique. Cette régression est effectuée en coupe instantanée en recourant à diverses actions cotées en bourse. Lorsque l'on effectue cette régression, on s'attend à ce que le bêta ait un signe négatif. En effet, plus le bêta d'un titre est élevé, plus son risque est important, ce qui devrait se traduire par une baisse de son rapport cours-bénéfices ou par une hausse de l'inverse de ce rapport qui est une mesure du rendement du titre. Plus le risque d'un titre est important, plus son prix doit s'ajuster à la baisse pour inciter un investisseur qui éprouve de l'aversion pour le risque à le détenir. Par ailleurs, dans cette équation EGR mesure la croissance anticipée des profits d'une compagnie qui est évaluée par la croissance observée de ceux-ci au cours des cinq dernières années. Certes, un tel indicateur des profits prévus peut se révéler inadéquat. Quoi qu'il en soit, une hausse des profits anticipés devrait évidemment se traduire par une hausse du rapport cours-bénéfices. Une compagnie donnée, qui œuvre dans le secteur de la biotechnologie, peut avoir un bêta important mais quand même jouir d'un (P/E) élevé si on anticipe une croissance marquée de ses profits au cours des prochaines années, ce qui s'observe souvent dans le secteur de la biotechnologie. Finalement, le modèle de Gordon indique que la variable *taux de paiement du dividende* devrait exercer un impact positif sur le rapport cours-bénéfices. Mais le recours à cette variable dans la pratique ne va pas sans failles, puisqu'elle peut être porteuse d'effets de signalisation. Notons rapidement les limites d'une telle régression. Celle-ci postule d'abord une relation linéaire entre le rapport cours-bénéfices et ses facteurs explicatifs, ce qui n'est peut-être pas le cas. Ensuite, l'analyse des résidus peut signaler de

Comme on a pu le constater dans le chapitre 6, consacré aux méthodes numériques utilisées en économétrie, l'actualisation au taux sans risque jouit d'une très grande popularité dans le domaine de l'évaluation des produits dérivés. Pour déterminer le prix d'un produit dérivé, on calcule l'espérance de ses cash-flows dans un univers neutre au risque. On peut alors actualiser lesdits cash-flows au taux sans risque pour en arriver à la détermination du prix d'équilibre du produit dérivé. Nous envisageons maintenant l'estimation de la SML par la méthode des MCO, en insistant sur les conditions qui doivent être satisfaites pour que l'on puisse recourir à une telle méthode d'estimation.

L'estimation de la SML par la méthode des MCO

La SML est formulée en termes de rendements espérés. Or, ceux-ci ne sont pas directement observables. Pour arriver à exprimer la SML en termes de données observables, on suppose que le rendement d'un titre obéit à un jeu non biaisé (*fair game*)¹⁹ représenté par le processus stochastique suivant :

$$R_{it} = E(R_{it}) + \beta_i [R_{mt} - E(R_{mt})] + \varepsilon_{it}$$

On remplace ensuite $E(R_{it})$ par son expression en termes de la SML et après certaines manipulations évidentes, on obtient :

$$R_{it} - R_{ft} = \beta_i [R_{mt} - R_{ft}] + \varepsilon_{it}$$

signaler de l'hétéroscédasticité ou d'autres problèmes statistiques. Cette analyse peut entre autres indiquer certaines transformations souhaitables des variables indépendantes (variables élevées au carré, transformation logarithmique et plus généralement transformation Box-Cox) qui sont de nature à mieux expliquer le rapport cours-bénéfices. La présence de multicollinéarité, qui est fréquente dans ce type de régression, diminue la fiabilité des coefficients estimés, pouvant même se traduire par des signes contraires à ceux qui sont anticipés pour les différentes variables et par une forte instabilité dans l'estimation des coefficients d'une période à l'autre. Pour plus de détails sur ce type d'analyse empirique, on consultera : Damodaran, A., (1996), *Investment Valuation*, John Wiley & Sons, New York, chapitre 14.

19. Pour plus de détails à ce sujet, voir : Copeland, T.E. et J.E. Weston (1988), *Financial Theory and Corporate Policy*, Addison Wesley, New York, p. 212 et suivantes.

Après introduction d'une constante²⁰, la forme empirique de la SML est donc de :

$$R_{it} - R_{ft} = \alpha_i + \beta_i [R_{mt} - R_{ft}] + \varepsilon_{it}$$

Soit à définir : $r_{it} = R_{it} - R_{ft}$ et $r_{mt} = R_{mt} - R_{ft}$. On a donc finalement :

$$r_{it} = \alpha_i + \beta_i [r_{mt}] + \varepsilon_{it}$$

Cette régression est donc formulée en termes de rendements excédentaires, ou de primes de risque. Forts des enseignements délivrés par les chapitres antérieurs, nous pouvons maintenant énoncer les conditions qui sont requises pour que l'on puisse recourir à la méthode des MCO pour estimer les paramètres de cette équation :

- i) Les primes de risque doivent être stationnaires. En effet, si elles ne l'étaient pas, on tomberait dans le piège des régressions fallacieuses²¹.
- ii) Les résidus de cette régression doivent être homoscedastiques et non autocorrélés. S'ils ne l'étaient pas, l'estimateur des MCO ne serait pas efficient. De plus, ces résidus doivent obéir à une distribution normale. Dans le cas contraire, les tests ne seraient valables qu'asymptotiquement, entre autres.

De plus, en vertu du CAPM classique, la constante de la régression ne doit pas être significativement différente de 0, la relation entre les primes de risque doit être linéaire et l'estimé du bêta ne doit pas varier dans le temps.

Nous avons auparavant esposé le test du CAPM tel qu'effectué par Berndt (1991). Selon Mills (1993)²², l'absence d'hétéroscédasticité dans les estimations de Berndt, notamment sous sa forme conditionnelle, est plutôt surprenante du fait des travaux antérieurs à ce sujet.

20. Certes, la constante ne devra pas être significativement différente de 0 pour que le CAPM soit valable. Cette constante est appelée *alpha de Jensen*. Elle mesure le rendement exceptionnel sur un titre, c'est-à-dire un rendement en sus de celui que justifie l'ampleur du bêta de ce titre. Un tel rendement exceptionnel n'a pas sa place sur des marchés efficients.

21. Cette allégation doit cependant être nuancée si les primes de risque sont cointégrées, comme cela devrait être. À ce sujet, voir le chapitre 9.

22. Mills, T.C. (1993), *The Econometric Modelling of Financial Time Series*, Cambridge University Press, Cambridge.

De plus, d'autres études ont révélé que le bêta d'un titre est variable dans le temps. L'analyse du CAPM dans un contexte GARCH-M multivarié apparaît donc appropriée. C'est cette formulation que nous considérons maintenant.

Analyse du CAPM dans le contexte GARCH-M multivarié²³

Envisageons d'abord ce que l'on entend par un modèle de régression multivarié. Pour y arriver, on remplace la variable dépendante y_t du modèle suivant :

$$y_t = \alpha_0 + \sum_{i=1}^m \alpha_i y_{t-i} + \sum_{i=0}^m \beta_i x_{t-i} + u_t$$

par un vecteur :

$$\mathbf{y}_t = [y_{1t} \quad y_{2t} \quad \dots \quad y_{nt}]^T$$

On obtient alors le modèle de régression multivarié²⁴ (dynamique) suivant :

$$\mathbf{y}_t = \mathbf{C} + \sum_{i=1}^m \mathbf{A}_i^T \mathbf{y}_{t-i} + \sum_{i=0}^m \mathbf{B}_i^T \mathbf{x}_{t-i} + \mathbf{u}_t$$

Dans cette expression, \mathbf{C} est un vecteur ($n \times 1$) de constantes ; $\mathbf{A}_1, \dots, \mathbf{A}_m$ sont des matrices ($n \times n$) de coefficients retardés, $\mathbf{B}_0, \mathbf{B}_1, \dots, \mathbf{B}_m$ sont des matrices ($k \times n$) de coefficients et \mathbf{u}_t est le vecteur ($n \times n$) des résidus. Sans entrer dans les détails, disons que l'estimation d'un modèle multivarié est une simple extension du cas univarié.

Qu'entend maintenant par un GARCH-M multivarié ? Dans sa forme simple, un modèle GARCH-M peut s'exprimer de la façon suivante :

$$\mathbf{y}_t = \mathbf{X}_t \boldsymbol{\beta} + \delta h_t + \boldsymbol{\varepsilon}_t$$

23. Cette section s'inspire de Mills (1993), *op. cit.*

24. Un tel modèle est donc multivarié (plusieurs variables) du point de vue des variables dépendantes. Se pose alors le problème de la modélisation des covariances entre les résidus de ces variables. Le modèle GARCH-M se présente comme une tentative de modéliser l'hétéroscédasticité conditionnelle de ces résidus.

où

$$h_t = \gamma_0 + \sum_{i=1}^p \gamma_i \varepsilon_{t-i}^2 + \sum_{i=1}^q \phi_i h_{t-i}$$

où $\varepsilon_t | \Phi_{t-1} \sim N(0, h_t)$, F_{t-1} étant l'ensemble d'informations disponibles jusqu'à $t-1$. Ce modèle est une simple adaptation du modèle ARCH-M. Dans le cadre d'un modèle multivarié, modèle qui a été présenté au début de cette section, chaque élément de la matrice variance-covariance conditionnelle peut s'exprimer comme suit, en supposant un processus GARCH(1, 1) :

$$h_{ijt} = c_{ij} + a_{ij} \varepsilon_{i,t-1} \varepsilon_{j,t-1} + b_{ij} h_{ij,t-1}$$

ce qui complète la présentation du modèle GARCH-M multivarié. Pour estimer ce modèle, on suit la procédure habituelle en maximisant la fonction de vraisemblance en recourant à l'algorithme BHHH.

Cette digression étant faite, nous pouvons maintenant nous pencher sur la mise en forme du CAPM selon la méthode du GARCH-M multivarié qui permet, entre autres, de rendre le bêta variable dans le temps. Cette formulation du CAPM est surtout due à Bollerslev, Engle et Woolridge (1988)²⁵. Hall, Miles et Taylor (1989) ont également présenté une version de ce modèle²⁶.

De manière à transposer le CAPM dans le cadre d'un modèle GARCH-M, il faut d'abord l'exprimer sous une forme conditionnelle. L'espérance des rendements excédentaires devient donc conditionnelle à l'ensemble des informations disponibles au temps $t-1$, désigné par F_{t-1} . Sous forme conditionnelle, l'équation de la SML devient :

$$E(R_{pt} | \Phi_{t-1}) - R_{f,t-1} = \beta_{pt} [E(R_{mt} | \Phi_{t-1}) - R_{f,t-1}]$$

et le bêta :

$$\beta_{pt} = \frac{\text{Cov}(R_{pt}, R_{mt} | \Phi_{t-1})}{V(R_{mt} | \Phi_{t-1})}$$

25. Bollerslev, T., R.F. Engle et J.M. Woolridge (1988), « A Capital Asset Pricing Model with Time-Varying Covariances », *Journal of Political Economy*, 96, p. 116-131.

26. Hall, S.G., D.K. Miles et M.P. Taylor (1989), « Modelling Asset Prices with Time-Varying Betas », *Manchester School*, 57, p. 340-356.

Dans ce qui suit, de manière à simplifier la notation, nous omettrons l'ensemble d'informations Φ_{t-1} dans la notation des moments conditionnels : l'indice t rattaché à ces moments en tiendra lieu. La matrice variance-covariance conditionnelle des rendements est de :

$$\mathbf{H}_t = \begin{bmatrix} V_t(R_{pt}) & \text{Cov}_t(R_{pt}, R_{mt}) \\ \text{Cov}_t(R_{pt}, R_{mt}) & V_t(R_{mt}) \end{bmatrix}$$

Comme l'on suppose que cette matrice peut maintenant varier dans le temps, il en résulte que les rendements espérés et les bêtas seront eux-mêmes variables dans le temps.

Cette formulation du CAPM n'est pas encore opérationnelle puisqu'il n'existe pas de séries sur les rendements espérés. Pour pallier à ce problème, Bollerslev, Engle et Woolridge ont supposé que le prix du risque est constant. Par rapport à la formulation présentée antérieurement, ce prix comporte la variance du rendement du portefeuille du marché à son dénominateur et non son écart-type. Le prix du risque λ est donc égal à :

$$\lambda = \frac{[E_t(R_{mt}) - R_{f,t-1}]}{V_t(R_{mt})}$$

La prime de risque conditionnelle du portefeuille du marché peut alors être exprimée en termes de sa variance conditionnelle :

$$E_t(R_{mt}) - R_{f,t-1} = \lambda V_t(R_{mt})$$

Suite à cette hypothèse, l'équation de la SML conditionnelle devient :

$$\begin{aligned} E_t(R_{pt}) &= R_{f,t-1} + \text{Cov}_t(R_{pt}, R_{mt}) \frac{E_t(R_{mt}) - R_{f,t-1}}{V_t(R_{mt})} \\ &= R_{f,t-1} + \text{Cov}_t(R_{pt}, R_{mt}) \lambda \end{aligned}$$

En définissant u_{pt} comme :

$$u_{pt} = R_{pt} - E_t(R_{pt})$$

on obtient finalement :

$$R_{pt} = R_{f,t-1} + \lambda \text{Cov}_t(R_{pt}, R_{mt}) + u_{pt}$$

De la même façon on peut écrire :

$$R_{mt} = R_{f,t-1} + \lambda V_t(R_{mt}) + u_{mt}$$

Ces deux dernières équations servent à estimer la version GARCH-M multivariée du CAPM. On constate que :

$$V_t(R_{mt}) = E_t(u_{mt}^2) = h_{mmt}$$

et que :

$$\text{Cov}_t(R_{pt}, R_{mt}) = E_t(u_{pt}, u_{mt}) = h_{pmt}$$

Ce modèle se formule comme suit :

$$\begin{bmatrix} R_{pt} - R_{f,t-1} \\ R_{mt} - R_{f,t-1} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \lambda \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} h_{ppt} \\ h_{mmt} \\ h_{pmt} \end{bmatrix} + \begin{bmatrix} u_{pt} \\ u_{mt} \end{bmatrix}$$

ou sous forme matricielle compacte :

$$y_t = \alpha + \mathbf{dvech}(\mathbf{H}_t) + u_t$$

où :

$$\begin{aligned} y_t &= [R_{pt} - R_{f,t-1}, R_{mt} - R_{f,t-1}]^T \quad 27 \\ \text{vech}(\mathbf{H}_t) &= [h_{ppt}, h_{mmt}, h_{pmt}]^T \\ u_t &= (u_{pt}, u_{mt})^T \\ \mathbf{d} &= \lambda \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \end{aligned}$$

27. L'opérateur matriciel *vech* vectorise le triangle inférieur de la matrice variance-covariance conditionnelle.

Bollerslev, Engle et Woolridge (1988) ont estimé ce modèle sur un échantillon américain comprenant des bons du Trésor à 6 mois, des obligations du Trésor à 20 ans et des rendements d'actions cotées à la NYSE. Leur échantillon consistait en des données trimestrielles s'étirant de 1959 à 1984, le taux sans risque choisi étant celui des bons du Trésor à 3 mois. L'estimation du prix du risque (λ) fut d'environ 0,5. Les estimés des paramètres de la variance conditionnelle furent très significatifs pour les bons et les obligations. Pour les actions, ils ne l'étaient pas individuellement mais ils étaient hautement significatifs lorsque regroupés. Ces estimés sont utilisés pour calculer les primes de risque : $\alpha + d \text{vech}(\mathbf{H}_t)$ et les bêtas variables dans le temps implicites. Les bons et les obligations ont des primes de risque croissantes après octobre 1979, le bêta des actions est rapproché de 1, le bêta des obligations légèrement supérieur à 1 et celui des bons près de 0. Soulignons finalement que Hall, Miles et Taylor ont donné crédit à la formulation du modèle intertemporel du CAPM de Merton, qui donne la part belle à la fonction de consommation des agents économiques, cela dans le cadre du GARCH-M multivarié²⁸.

6. UNE DIGRESSION : LA THÉORIE DE L'APT²⁹

Nous ne saurions quitter ce chapitre sans faire état d'un modèle important en finance moderne : l'APT (*Arbitrage Pricing Theory*). En effet, celui-ci se veut plus général que le CAPM qui peut donc être vu comme un cas particulier de l'APT. La section qui suit nous permettra par conséquent de mieux maîtriser le CAPM.

6.1. Le principe de l'arbitrage

Le modèle de l'APT repose sur une notion bien connue en finance moderne : l'arbitrage. Lorsque l'équilibre des marchés financiers est

28. Voir à ce sujet : Merton, R.C. (1973), « An Intertemporal Capital Asset Pricing Model », *Econometrica*, 41, p. 867-887.

29. Pour cette section, les références sont les suivantes : Copeland et Weston (1988), *op. cit.*; Ingersoll, J.E. (1987), *Theory of Financial Decision Making*, Rowman et Littlefield, Maryland, 4^e édition, chap.7.; Elton, E.J. et M.J. Gruber (1995), *Modern Portfolio Theory and Investment Analysis*, New York, Wiley; Haugen, R.A. (1997), *Modern Investment Theory*, Prentice Hall, Englewoods Cliffs.

atteint, il ne doit subsister aucune possibilité d'arbitrage. On dira qu'un arbitrage est possible quand un investisseur qui réaménage son portefeuille sans injecter d'argent frais et qui n'assume aucun risque additionnel peut obtenir un rendement positif additionnel. Autrement dit, il y a alors possibilité d'un repas gratuit (*free lunch*). L'équilibre sur les marchés financiers est caractérisé par l'absence de telles opportunités.

Pour mieux comprendre le principe de l'arbitrage, nous envisageons le cas de trois placements, dont les cash-flows respectifs apparaissent au tableau 10.3.

TABLEAU 10.3

Cash-flows au temps	t_0	t_1	t_2
Placement 1	-1	0,049	1,05
Placement 2	-1	1,045	0
Placement 3	-1	0	1,1025

Trois placements sans risque s'offrent donc. Tous les trois coûtent 1 \$. Le premier placement génère un cash-flow de 0,049 \$ au temps 1 et de 1,05 \$ au temps 2. Le placement 2 donne pour sa part un cash-flow de 1,045 \$ au temps 1 et aucun au temps 2. Finalement, le placement 3 fournit un cash-flow nul au temps 1 et un cash-flow de 1,1025 \$ au temps 2.

Nous nous demandons s'il y a ici une possibilité d'arbitrage. Autrement dit, en réaménageant son portefeuille composé de ces trois placements, un investisseur peut-il dégager un profit net (*free lunch*) ? À cet effet, imaginons que notre investisseur effectue le réaménagement de portefeuille suivant. Il diminue le montant qu'il a investi dans le placement 1 de 10 000 \$. Selon le tableau 10.3, il se prive alors d'un cash-flow de 490 \$ au temps t_0 et de 10 500 \$ au temps t_1 . Pour reconstituer ces cash-flows, il investit d'abord un montant de :

$$\frac{490}{1,045} = 468,90\$$$

dans le placement 1, ce qui lui permet de reconstituer le cash-flow de 490 \$ au temps t_1 qui a été abandonné en sabrant dans le placement 1. Et pour reconstituer le cash-flow abandonné au temps t_2 , il investit le montant suivant dans le placement 2 :

$$\frac{10500}{1,1025} = 9523,81$$

La facture de ce réaménagement de portefeuille se monte à :

$$468,90 + 9523,81 = 9992,71 \$$$

ce qui est inférieur au cash-flow libéré lors du prélèvement dans le placement 1, soit 10 000 \$. En réaménageant son portefeuille, l'individu a encaissé un profit net de 7,29 \$. Il a ainsi réalisé un arbitrage !

À hauteur de 1 \$, le coût du placement 1 est donc trop élevé. Pour calculer le prix juste de ce placement, nous devons actualiser ses cash-flows aux taux au comptant (taux *spot*) appropriés. Chaque cash-flow doit être actualisé au taux au comptant dont la durée est égale à l'échéance de ce cash-flow. Or, les placements 2 et 3 sont assimilables à des obligations à coupon 0^{30} dont les durées sont respectivement de un et de deux ans. Selon le tableau 10-3, le taux au comptant d'un an est de 4,5 % et le taux au comptant de deux ans, de :

$$\sqrt{1,1025} - 1 = 0,05 \text{ ou } 5\%$$

Le prix d'équilibre du placement 1 est donc de :

$$\frac{0,049}{1,045} + \frac{1,05}{1,1025} = 0,9993\$$$

et non 1 \$ comme l'indique le tableau 10.3.

Une autre façon de constater que le placement 1 est trop dispendieux est de calculer sa valeur future. Pour ce faire, il faut connaître le taux de réinvestissement du premier cash-flow du placement 1 qui se produit au temps t_1 . Selon la théorie des taux à terme, le taux de réinvestissement, ou taux à terme d'un an dans un an, est ici égal à :

30. Rappelons ici que les taux *spots* se calculent à partir des obligations à coupon 0. Pour plus de détails à ce sujet, on consultera : Racicot, F.-É. et R. Théoret (2000), *Traité de gestion de portefeuille*, *op. cit.*

$$f_1 = \frac{(1 + R_2)^2}{(1 + R_1)} - 1 = \frac{1,1025}{1,045} - 1 = 5,50\%$$

où f_1 désigne le taux à terme (taux *forward*) d'un an dans un an ; R_1 , le taux *spot* de un an et R_2 , le taux *spot* de deux ans. La valeur future du placement 1 est de :

$$0,049(1,055) + 1,05 = 1,1017\$$$

Or, le placement 3, qui est pourtant de même coût que le placement 1, comporte une valeur future de 1,1025 \$. Le placement 1 est donc surévalué en regard du placement 3. L'opération d'arbitrage consiste alors à vendre à découvert (*short*) le placement 1 et à acheter le placement 3, jusqu'à ce que l'équilibre sur les marchés financiers soit restauré. Celui-ci percera à jour quand le prix du placement 1 se sera abaissé à 0,9993 \$ à la suite des ventes à découvert dont il est victime, cela en vertu des calculs précédents.

Abordons maintenant le phénomène de l'arbitrage sur les marchés financiers dans une perspective beaucoup plus générale. Posons-nous la question suivante : comment peut-on calculer les prix d'équilibre des titres en recourant au principe de l'arbitrage ? Quelle est la structure de prix qui doit prévaloir pour qu'il n'y ait plus de repas gratuit ? Or, un théorème bien connu en finance avance qu'il n'existera plus d'occasions d'arbitrage s'il existe un vecteur de prix \mathbf{p} tel que :

$$\mathbf{B}\mathbf{p} = -\mathbf{a}$$

où \mathbf{B} désigne la matrice des cash-flows des titres et \mathbf{a} , le vecteur de leur coût (néгатif).

Forts de ce théorème, recalculons le coût du placement 1 de sorte qu'il n'y ait plus trace d'arbitrage. Le système d'équations $\mathbf{B}\mathbf{p} = -\mathbf{a}$ s'écrit en faisant appel au tableau 10.3 :

$$\begin{bmatrix} 0,049 & 1,05 \\ 1,045 & 0 \\ 0 & 1,1025 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} -a_1 \\ 1 \\ 1 \end{bmatrix}$$

où a_1 désigne le coût (négalif) du placement 1. Les équations correspondant à ce système matriciel sont :

$$\begin{aligned} 0,049p_1 + 1,05p_2 &= -a_1 \\ 1,045p_1 &= 1 \\ 1,1025p_2 &= 1 \end{aligned}$$

Selon les deux dernières équations, $p_1 = 0,9569$ et $p_2 = 0,9070$. En substituant ces deux valeurs dans l'équation du placement 1, on trouve que $a_1 = -0,9993$ \$, ce qui correspond bien au résultat trouvé auparavant.

Le système matriciel $\mathbf{Bp} = -\mathbf{a}$ qui doit prévaloir lorsque l'équilibre s'est instauré sur les marchés financiers s'interprète donc très facilement. Il signifie tout simplement que dans un univers sans risque, le prix d'un titre est la valeur actualisée de ses cash-flows, les facteurs d'actualisation étant calculés au moyen des taux *spots* correspondants. Le vecteur \mathbf{p} est en effet un vecteur de facteurs d'actualisation, ou de prix des périodes. Dans notre exemple :

$$p_1 = \frac{1}{1,045} = 0,9569$$

alors que p_2 est de :

$$p_2 = \frac{1}{(1,05)^2} = 0,9070$$

Les facteurs d'actualisation des périodes p_1 et p_2 sont donc calculés à partir des taux *spots* correspondants. Le théorème précédent indique qu'à l'équilibre, la relation entre les cash-flows d'un titre et son prix est linéaire.

Nous pouvons poser le problème d'une autre façon. Supposons que le coût du placement 1 soit de 1 \$. Quelle valeur doit alors prendre le cash-flow du placement 2 pour qu'il y ait absence d'arbitrage ? L'équation $\mathbf{Bp} = -\mathbf{a}$ s'écrit, dans ce cas :

$$\begin{bmatrix} 0,049 & 1,05 \\ C_{21} & 0 \\ 0 & 1,1025 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

où C_{21} désigne le cash-flow recherché du placement 2. En solutionnant, on trouve que $C_{21} = 1,0284 \$$, ce qui implique que le taux spot d'un an doit être de 2,84 %. Le coût du placement 1 est alors de :

$$\frac{0,049}{1,0284} + \frac{1,05}{(1,05)^2} = 1 \$$$

Pour que le prix d'équilibre du placement 1 soit de 1 \$, il faut donc que le taux *spot* d'un an soit de 2,84 %³¹ si le taux *spot* de deux ans s'établit à 5 %.

Nous sommes maintenant en mesure de formuler le théorème de l'arbitrage de façon plus rigoureuse en nous situant dans un univers sans risque. Soit le vecteur \mathbf{x} suivant qui renferme les coefficients de réaménagement des n titres que comprend l'univers :

$$\mathbf{x} = (x_1, x_2, \dots, x_n)$$

Un tel réaménagement constitue un arbitrage si deux conditions sont réalisées :

- i) ce réaménagement doit générer des cash-flows au moins aussi élevés que le portefeuille initial :

$$\mathbf{B}^T \mathbf{x} \geq \mathbf{0}$$

où \mathbf{B} est la matrice des cash-flows des titres.

- ii) Le réaménagement de portefeuille doit libérer des fonds³² :

$$\mathbf{a}^T \mathbf{x} > 0$$

31. Autrement dit, le premier cash-flow du placement 1 doit être moins actualisé de façon à relever la valeur de ce placement. On peut également aborder ce problème en termes de valeur future. Étant donné que le taux *spot* de deux ans est alors beaucoup plus élevé que celui d'un an, on pourra réinvestir le premier cash-flow du placement 1 à un taux plus important que dans le tableau 10.3. Ce taux de réinvestissement est de 7,21 %, ce qui donne au placement 1 une valeur future identique à celle du placement 3. Il n'y a donc plus possibilité d'arbitrage.

32. N'oublions que le vecteur \mathbf{a} est constitué des coûts des titres. Ses entrées sont donc négatives.

Et en se basant sur des théorèmes d'algèbre linéaire établis par H. Minkowski et J. Farkas, on peut énoncer les propositions suivantes :

- i) ou bien il existe un vecteur \mathbf{x} tel que :

$$\mathbf{a}^T \mathbf{x} > 0$$

et que :

$$\mathbf{B}^T \mathbf{x} \geq \mathbf{0}$$

c'est-à-dire qu'il existe possibilité d'arbitrage ;

- ii) ou bien, il existe un vecteur \mathbf{p} de prix d'équilibre qui obéit au système linéaire suivant :

$$\mathbf{B}\mathbf{p} = -\mathbf{a}$$

et il n'y a plus possibilité d'arbitrage.

Comme nous serons à même de le constater ultérieurement, cette condition s'avère très importante dans la formulation de l'équation de base de l'APT : *Arbitrage Pricing Theory*.

6.2. L'APT : aperçu général

L'APT peut être présentée comme une théorie concurrente au CAPM, qui vise comme elle à déterminer les rendements d'équilibre des titres. Mais contrairement au CAPM, l'APT ne se situe pas dans le cadre de l'analyse moyenne-variance mis de l'avant par Markowitz au début des années 1950. Elle ne suppose donc pas que la distribution des rendements est normale. Elle ne donne pas non plus un rôle particulier au portefeuille du marché comme c'est le cas pour le CAPM où le portefeuille du marché, pièce maîtresse de la CML, doit être efficient. L'APT se veut beaucoup plus générale que le CAPM. Elle repose sur le principe de l'arbitrage tel qu'il vient d'être énoncé. Lorsque l'équilibre des marchés financiers est atteint, il doit en effet y avoir absence d'arbitrage. Le modèle de l'APT essaie de dégager l'incidence de cet impératif sur les rendements d'équilibre des titres. Cependant, comme nous le verrons, en se voulant plus général que CAPM, il en arrive également à des conclusions moins spécifiques. C'est pourquoi ce modèle n'a pas réussi à détrôner le CAPM, qui jouit encore d'une grande popularité après plus de quarante ans d'existence.

6.3. Dérivation du modèle de l'APT

La notion de matrice orthogonale

La dérivation de l'APT fait appel à la notion de matrice orthogonale. Rappelons-la. La matrice carrée \mathbf{A} est orthogonale si

$$\mathbf{A}^T \mathbf{A} = \mathbf{A} \mathbf{A}^T = \mathbf{I}$$

Si la matrice \mathbf{A} est orthogonale, on peut donc écrire :

$$\mathbf{A}^{-1} = \mathbf{A}^T$$

Cette égalité matricielle implique que les vecteurs qui forment les lignes (ou les colonnes) d'une matrice orthogonale constituent un ensemble orthonormal. Désignons la i^{e} ligne (ou colonne) de \mathbf{A} par \mathbf{a}_i . Un ensemble orthonormal se définit comme suit :

$$\begin{aligned} \mathbf{a}_i^T \mathbf{a}_j &= 1 \quad \text{si } i = j \\ \mathbf{a}_i^T \mathbf{a}_j &= 0 \quad \text{si } i \neq j \end{aligned}$$

Dans une matrice orthogonale, chaque vecteur est donc normalisé pour avoir une longueur unitaire et est orthogonal à tout autre vecteur.

Les équations de l'APT

Le modèle de l'APT fut proposé par Ross³³ en 1976. On se souvient que dans le modèle classique du CAPM, l'espérance du taux de rendement d'un titre ne dépend que d'une seule variable : la prime de risque du portefeuille du marché, soit $[E(R_m) - R_f]$. Le modèle de l'APT se veut beaucoup plus général. À l'intérieur de celui-ci, le rendement du titre i est une fonction linéaire de k facteurs. L'équation du rendement du titre i (R_i) est la suivante :

$$R_i = E(R_i) + b_{i1}F_1 + \dots + b_{ik}F_k + \varepsilon_i$$

33. Ross, S.A. (1976), « The Arbitrage Theory of Capital Asset Pricing », *Journal of Economic Theory*, décembre, p. 343-362.

où F_j sont les divers facteurs et b_{ij} , les bêtas respectifs de ces facteurs. ε_i , l'innovation, représente le risque non systématique du titre i .

Les prévisions du marché sont emmagasinées dans $E(R_i)$. Les facteurs représentent des variables non anticipées pertinentes pour prédire R_i . Par exemple, l'inflation non anticipée peut être l'un de ces facteurs ou encore la croissance non anticipée du Pib.

Reste à déterminer l'équation de $E(R_i)$. Pour y arriver, il faut se rappeler que l'APT repose sur le principe de l'arbitrage. Ce principe s'énonce comme suit. À l'équilibre, les réaménagements de portefeuilles qui ne requièrent aucune injection d'argent frais et qui, de plus, ne comportent aucun risque doivent dégager un rendement nul. Nous nous attaquons maintenant à la construction de portefeuilles d'arbitrage³⁴ en mettant en application ce principe.

- i) Un portefeuille d'arbitrage ne doit comporter aucune injection d'argent frais. Soit \mathbf{x} le vecteur des coefficients de réaménagement d'un portefeuille composé de n titres. Cette condition s'écrit comme suit :

$$\sum_{i=1}^n x_i = 0$$

Transposons-la en termes matriciels :

$$\begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix} = 0$$

soit :

$$\mathbf{x}^T \mathbf{1} = 0$$

34. Attention ! Un portefeuille d'arbitrage désigne ici un portefeuille pour lequel il y a absence d'arbitrage.

ce qui signifie que le vecteur \mathbf{x} doit être orthogonal au vecteur unitaire $\mathbf{1}$ à l'équilibre.

ii) Un portefeuille d'arbitrage doit être sans risque.

Les x_i doivent être choisis de façon telle que pour tout facteur (ici k), la somme des b_{ik} soit nulle.

$$\sum_{i=1}^n x_i b_{ik} = 0$$

Les b_{ik} sont les bêtas du facteur k pour chaque titre i . Cette équation signifie que la somme de ces bêtas pour le facteur k , pondérés par les facteurs de réaménagement de portefeuille, doit être nulle à l'équilibre. Autrement dit, le réaménagement n'accroît pas le risque relié au facteur k à travers l'ensemble du portefeuille.

Le vecteur des bêtas des n titres associé au facteur k s'écrit comme suit :

$$\begin{bmatrix} b_{1k} \\ b_{2k} \\ \dots \\ b_{nk} \end{bmatrix}$$

La deuxième condition de l'absence d'arbitrage est donc la suivante :

$$\begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} b_{1k} \\ b_{2k} \\ \dots \\ b_{nk} \end{bmatrix} = 0$$

soit, pour l'ensemble des facteurs :

$$\mathbf{x}^T \mathbf{B} = \mathbf{0}$$

où en supposant k facteurs, \mathbf{B} est égal à :

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & \cdot & \cdot & \cdot & b_{1k} \\ b_{21} & b_{22} & \cdot & \cdot & \cdot & b_{2k} \\ \cdot & & \cdot & & & \cdot \\ \cdot & & & \cdot & & \cdot \\ \cdot & & & & \cdot & \cdot \\ b_{n1} & b_{n2} & \cdot & \cdot & \cdot & b_{nk} \end{bmatrix}$$

À l'équilibre, le vecteur \mathbf{x} doit donc être également orthogonal à la matrice \mathbf{B} qui renferme les bêtas des n titres pour les k facteurs.

- iii) Le taux de rendement d'un portefeuille d'arbitrage doit être nul.

Cette condition s'écrit :

$$R_p = \sum_i x_i E(R_i) = 0$$

soit, en termes matriciels :

$$\mathbf{R} = \mathbf{x}^T \mathbf{E}(\mathbf{R}) = 0$$

À l'équilibre, le vecteur \mathbf{x} doit finalement être orthogonal au vecteur $\mathbf{E}(\mathbf{R})$.

Les trois équations que nous venons d'écrire caractérisent les portefeuilles d'arbitrage. Elles sont :

$$\begin{aligned} \mathbf{x}^T \mathbf{1} &= 0 \\ \mathbf{x}^T \mathbf{B} &= \mathbf{0} \\ \mathbf{x}^T \mathbf{E}(\mathbf{R}) &= 0 \end{aligned}$$

Le vecteur \mathbf{x} est donc orthogonal à trois vecteurs ou matrice à l'équilibre : le vecteur unitaire, la matrice \mathbf{B} et le vecteur $\mathbf{E}(\mathbf{R})$. La conséquence algébrique est que le rendement espéré du titre i est une combinaison linéaire du vecteur vecteur unitaire et du vecteur de ses bêtas factoriels, c'est-à-dire :

$$E(R_i) = \lambda_0 + \lambda_1 b_{i1} + \dots + \lambda_k b_{ik}$$

C'est là le résultat principal de la théorie de l'APT.

Pour mieux comprendre ce résultat algébrique, situons-nous dans un monde de trois titres. Supposons également qu'un seul facteur explique le rendement de ces trois titres. Les trois conditions de l'absence d'arbitrage ou de l'équilibre s'écrivent alors, en termes matriciels :

$$\begin{bmatrix} 1 & 1 & 1 \\ b_1 & b_2 & b_3 \\ E(R_1) & E(R_2) & E(R_3) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Nous sommes ici en présence d'un système d'équations linéaires homogènes³⁵ du type $\mathbf{Ax} = \mathbf{0}$. Pour que ce système admette une solution non triviale, il faut que la matrice \mathbf{A} soit singulière, ce qui revient à dire dans notre exemple que son rang doit être inférieur à 3. La conséquence algébrique est que la troisième ligne de cette matrice est une

35. La solution d'un tel système d'équations est reliée au rang de la matrice \mathbf{A} . Une matrice est d'ordre r si et seulement si l'un de ses déterminants d'ordre r n'est pas nul et que tout déterminant d'ordre supérieur à r est nul. Supposons que \mathbf{A} soit une matrice carrée de dimension $(n \times n)$. Si le rang de cette matrice est de n , c'est-à-dire que cette matrice admet un inverse, alors le système d'équations homogènes admet une seule solution dite triviale : le vecteur \mathbf{x} est alors égal au vecteur nul. Si le rang de la matrice \mathbf{A} s'avère inférieur à n , le nombre d'inconnues est alors supérieur au nombre d'équations et le système admet alors plusieurs solutions non triviales. Dans le cas spécial où le rang de \mathbf{A} est égal à $(n-1)$, la solution est unique à un vecteur de proportionnalité près. L'espace vectoriel qui renferme l'ensemble des solutions d'un système d'équations homogènes est appelé : le kernel de \mathbf{A} . Pour plus de détails sur la solution d'un système d'équations linéaires homogènes, on consultera par exemple : Judge, G.G. *et al.* (1988), *Introduction to the Theory and Practice of Econometrics*, 2^e édition, John Wiley and Sons, New York.

combinaison linéaire des deux premières, celles-ci ne pouvant en effet être colinéaires. Il en résulte que :

$$\begin{bmatrix} E(R_1) \\ E(R_2) \\ E(R_3) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \lambda_0 + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \lambda_1$$

À l'équilibre, l'espérance du rendement d'un titre est donc une fonction linéaire de son bêta respectif. Si l'on retient la prime de risque du marché pour expliquer l'espérance du rendement d'un titre, on renoue alors avec le CAPM : λ_0 est alors le taux sans risque et λ_1 , la prime de risque du marché. C'est dans ce sens que le CAPM est un cas particulier de l'APT lorsqu'il n'existe qu'un seul facteur explicatif du rendement d'un titre : celui du portefeuille du marché.

S'il existe k facteurs explicatifs, la même méthodologie s'applique pour trouver l'espérance du rendement du titre i .

$$E(R_i) = \lambda_0 + \lambda_1 b_{i1} + \dots + \lambda_k b_{ik}$$

Le coefficient λ_0 est assimilable au taux sans risque. En effet, si le titre i ne comporte aucun risque, tous ses b_{ij} sont nuls. L'espérance du rendement du titre i se réduit alors à :

$$E(R_i) = \lambda_0$$

ce qui ne peut être que le taux sans risque.

Les autres λ_j de l'équation peuvent être interprétés en termes de rendement excédentaire, c'est-à-dire en termes de rendement requis au-delà du taux sans risque pour supporter une unité du risque associé au facteur j . Par exemple, λ_1 est le rendement excédentaire requis pour supporter une unité supplémentaire du risque associé au facteur 1. Il peut donc être assimilé au prix du risque³⁶ du facteur 1. On peut désigner ce rendement excédentaire par l'expression suivante :

$$\lambda_1 = \delta_1 - R_f$$

36. On insiste ici sur la correspondance entre rendement excédentaire par unité de risque et prix du risque, cela pour chaque catégorie de risque.

où δ_j est le rendement d'un portefeuille qui comprend une unité du risque associé au facteur 1 et une *quantité* nulle de tout autre facteur. Et ainsi de suite pour les autres λ_j .

L'équation de l'espérance du rendement du titre i peut donc être réécrite comme suit :

$$E(R_i) - R_f = (\delta_1 - R_f)b_{i1} + (\delta_2 - R_f)b_{i2} + \dots + (\delta_k - R_f)b_{ik}$$

En résumé, le modèle de l'APT renferme deux catégories d'équations³⁷ :

i) L'équation du rendement du titre i :

$$R_i = E(R_i) + b_{i1}F_1 + \dots + b_{ik}F_k + \varepsilon_i$$

ii) L'équation de l'espérance du rendement du titre i :

$$E(R_i) - R_f = (\delta_1 - R_f)b_{i1} + (\delta_2 - R_f)b_{i2} + \dots + (\delta_k - R_f)b_{ik}$$

Il est à noter que cette dernière équation peut être interprétée comme une régression linéaire si l'on formule les hypothèses suivantes : 1) la distribution des rendements est normale ; 2) les facteurs ont été transformés linéairement de telle sorte que les vecteurs transformés soient orthonormaux. Les b_{ik} sont alors égaux à :

$$b_{ik} = \frac{\text{COV}(R_i, \delta_k)}{\text{VAR}(\delta_k)}$$

où δ_k est la transformation linéaire du k^{e} facteur.

Remarquons finalement que le modèle de l'APT est plus robuste que le CAPM : 1) il n'émet pas d'hypothèses quant à la distribution des rendements des titres. Le CAPM suppose pour sa part que cette distribution est normale ; 2) l'APT ne formule pas d'hypothèses fortes quant aux fonctions d'utilité des individus³⁸ ; 3) les taux de rendement

37. Les équations correspondantes pour le CAPM sont les suivantes. L'équation du rendement du titre i s'écrit, dans ce modèle : $R_i - R_f = \beta_i(R_m - R_f) + \varepsilon_i$. Alors que l'équation de l'espérance du rendement du titre i est de :

$$E(R_i) = R_f + \beta_i[E(R_m) - R_f].$$

38. En effet, pour que le CAPM soit valable en l'absence de l'hypothèse de la distribution normale des rendements, il faut que la fonction d'utilité d'un individu soit quadratique de façon à récupérer l'analyse moyenne-variance.

d'équilibre peuvent dépendre de plusieurs facteurs dans l'APT, et non d'un seul comme c'est le cas dans le cadre du CAPM classique ; 4) l'APT n'accorde aucun rôle particulier au portefeuille du marché alors qu'il tient le haut du pavé dans le modèle du CAPM. De plus, le portefeuille du marché doit être efficient dans le modèle du CAPM, sinon ce dernier modèle n'est pas valable.

En se voulant plus général, le modèle de l'APT aboutit toutefois à des conclusions moins spécifiques que le CAPM. L'APT reste en effet muet sur l'identité des facteurs. C'est là son talon d'Achille. Le CAPM identifie pour sa part un unique facteur pour expliquer le rendement d'un titre : la prime de risque du marché. C'est pourquoi le CAPM jouit encore d'une grande popularité même si l'APT voulait au départ dominer le CAPM.

6.4. Tests de l'APT

La procédure habituelle pour tester le modèle de l'APT est la suivante : 1) colliger des séries statistiques sur les rendements d'un groupe d'actions ; 2) recourir à l'analyse factorielle pour déterminer simultanément le nombre de facteurs et leurs bêtas pour chaque titre, c'est-à-dire estimer les paramètres et déterminer les facteurs de la première équation de l'APT :

$$R_i = E(R_i) + b_{i1}F_1 + \dots + b_{ik}F_k + \varepsilon_i$$

3) utiliser les bêtas ainsi calculés pour déterminer les primes de risque ou les λ de chacun des facteurs. On recourt à la régression pour estimer ces primes de risque (rendements excédentaires) ou prix des diverses catégories de risque. C'est-à-dire que l'on estime les λ_j dans la seconde équation de l'APT :

$$E(R_i) = \lambda_0 + \lambda_1 b_{i1} + \dots + \lambda_k b_{ik}$$

C'est là une régression en coupe instantanée où les $E(R_i)$ sont approximés par une moyenne de rendements calculée sur la période de temps choisie et où les b_{ij} proviennent de la première étape du test. On régresse donc les rendements moyens sur les bêtas pour obtenir les lambdas.

Comme cela fut mentionné auparavant, le modèle de l'APT ne spécifie pas la nature des facteurs : cela relève des tests empiriques. Les premières vérifications de l'APT ont démontré que le marché ne rémunérerait qu'un nombre limité de facteurs, trois ou quatre, sans plus. L'APT ne spécifie pas la nature des facteurs. Le test empirique de l'APT qui a retenu le plus d'attention est celui effectué par Chen, Roll et Ross³⁹ en 1983. Leur étude permet d'identifier quatre variables macroéconomiques comme candidates au titre de facteurs : 1) la production industrielle ; 2) le changement dans une prime de risque de défaut, mesurée par la différence⁴⁰ entre les rendements des obligations de cote AAA et celles de cote Baa ; 3) la différence⁴¹ entre les rendements à l'échéance des obligations gouvernementales à long et à court termes ; 4) l'inflation non anticipée. À première vue, le choix de telles variables surprend, c'est le moins qu'on puisse dire. Mais il se justifie par le concept même du prix d'un titre, qui peut être défini comme la valeur escomptée de ses cash-flows anticipés. La production industrielle est reliée aux cash-flows. Les autres variables retenues par Chen, Roll et Ross ont trait au taux d'escompte utilisé pour escompter les cash-flows dudit titre. Une telle analyse se révèle utile sur le plan empirique. Comme les facteurs sont orthogonaux⁴², du moins l'analyse factorielle les a réduits à cette condition, il est en principe possible de choisir un portefeuille qui est protégé contre l'inflation sans que son exposition aux autres facteurs ne soit modifiée. On peut donc couvrir (*hedge*) les facteurs !

39. Chen, N.F. *et al.* (1983), « Economic Forces and the Stock Market : Testing the APT and Alternative Asset Pricing Theories », document de travail, UCLA, décembre.

40. Cet écart représentant le risque de crédit d'une entreprise.

41. Cet écart tenant lieu de la structure à terme des taux d'intérêt.

42. C'est-à-dire qu'il y a absence de corrélation entre les facteurs.

LA MÉTHODE DES MOMENTS GÉNÉRALISÉS

Ce chapitre présente une classe d'estimateurs aux propriétés asymptotiques regroupés sous le vocable : méthodes des moments généralisés, que nous désignerons par l'acronyme anglais GMM (*Generalized Method of Moments*). Cette méthode présente plusieurs avantages : i) la méthode du GMM fait la synthèse de plusieurs estimateurs bien connus, comme l'estimateur des MCO, l'estimateur des variables instrumentales, l'estimateur des doubles moindres carrés et l'estimateur des doubles moindres carrés non linéaires. De plus, il nous fournit un cadre d'analyse général pour la comparaison de ces divers estimateurs et leur évaluation ; ii) l'estimateur du GMM nous procure une solution de rechange simple à d'autres estimateurs, spécialement dans les cas où il est ardu d'écrire la fonction de vraisemblance dans le but de calculer l'estimateur du maximum de vraisemblance.

Il faut toutefois souligner que l'estimateur du GMM vaut pour les grands échantillons. C'est-à-dire que les propriétés désirables de cet estimateur ne seront obtenues que dans les grands échantillons. Ainsi, l'estimateur du GMM est asymptotiquement efficace mais il n'est que rarement efficace.

1. INTRODUCTION À LA MÉTHODE DES MOMENTS

Selon Hamilton (1994)¹, une description générale de la méthode classique des moments peut s'énoncer comme suit : soit un vecteur ($a \times 1$)

1. Hamilton, J.D. (1994), *Time Series Analysis*, Princeton University Press, Princeton.

de paramètres inconnus θ qui caractérise la fonction de densité d'une variable observée y_t . Supposons que l'on puisse exprimer les moments de la population comme des fonction de θ , telles que :

$$E(y_t^i) = \mu_i(\theta) \text{ pour } i = i_1, i_2, \dots, i_a.$$

La méthode classique des moments estime θ comme la valeur $\hat{\theta}_T$ pour laquelle ces moments de la population sont égaux aux moments échantillonnaires observés. C'est-à-dire que $\hat{\theta}_T$ est la valeur pour laquelle les équations suivantes sont réalisées :

$$\mu_i(\hat{\theta}_T) = \left(\frac{1}{T}\right) \sum_{t=1}^T y_t^i \text{ pour } i = i_1, i_2, \dots, i_a$$

Illustrons la méthode des moments à partir d'un exemple simple, tiré de Judge *et al.* (1988).² Soit Y_1, \dots, Y_T , un échantillon aléatoire d'une population $N(\theta, \sigma^2)$. On sait que :

$$E[Y] = \theta - \mu_1^2$$

et que :

$$\sigma^2 = E[Y^2] - (E[Y])^2 = \mu_2 - \mu_1^2$$

où μ_i désigne le i^{e} moment par rapport à l'origine. En égalant les moments de la population aux moments échantillonnaires, ce qui est le principe même de la méthode des moments, on obtient :

$$\hat{\mu}_1 = \frac{\sum_{i=1}^T Y_i}{T} = \bar{Y} = \hat{\theta}$$

et

$$\hat{\mu}_2 = \frac{\sum_{i=1}^T Y_i^2}{T} = \hat{\sigma}^2 + \hat{\theta}^2$$

2. Judge, G.G., R.C. Hill, W.E. Griffiths, H. Lütkepohl et T.-C. Lee (1988), *Introduction to the Theory and Practice of Econometrics*, 2^e édition, John Wiley and Sons, New York, p. 61-62.

Par conséquent,

$$\hat{\sigma}^2 = \left(\frac{\sum_{i=1}^T Y_i^2}{T} \right) - \bar{Y}^2 = \frac{1}{T} \sum_{i=1}^T (Y_i - \bar{Y})^2$$

La méthode des moments nous a donc permis d'estimer les deux premiers moments de notre population³.

Le cas précédent est le plus simple, désigné en anglais par *location model*. Mais il se peut que, pour estimer un paramètre, il faille recourir à plusieurs moments. Il faut alors pondérer ces moments d'une façon ou d'une autre pour estimer lesdits paramètres. Soit \mathbf{g} le vecteur des moments. Pour estimer les paramètres d'un modèle, il faudra alors minimiser une fonction critère représentée par la forme quadratique suivante (\mathbf{Q}):

$$\mathbf{Q} = \mathbf{g}^T \mathbf{W} \mathbf{g}$$

où \mathbf{W} est la matrice de pondération de ces moments. Tout repose alors sur la sélection de la matrice \mathbf{W} . Hamilton (1994) relate l'histoire de l'évolution de cette matrice qui devait aboutir à la méthode du GMM. Un estimateur basé sur la minimisation d'une expression de type \mathbf{Q} fut appelé *minimum chi-square* par Cramer en 1946. Ferguson en 1958 et Rothenberg en 1973 ont retenu la même appellation alors que Malinvaud en 1970 le nommera : *estimateur de distance minimale*. Mais il faudra attendre Hansen en 1982 pour avoir la forme la plus sophistiquée et la plus rigoureuse de cette approche, soit la méthode des moments généralisés, dont l'acronyme est le GMM⁴. Selon

3. À remarquer cependant que l'estimateur de la variance est biaisé. L'estimateur

$$\text{non biaisé est de : } \hat{\sigma}^2 = \frac{\sum_{i=1}^T (Y_i - \bar{Y})^2}{T - 1}.$$

4. L'article de base en ce qui concerne l'approche du GMM est : Hansen, L. (1982), « Large Sample Properties of Generalized Method of Moments Estimator », *Econometrica*, 50, p. 1029-1054. Pour une présentation simplifiée, voir : Hall, A. (1992), « Some Aspects of Generalized Method of Moments Estimation », document de travail, North Carolina State University ; Ogaki, M. (1992), « Generalized Method of Moments : Econometric Applications », document de travail, University of Rochester.

Davidson et McKinnon (1993)⁵, l'idée essentielle qui sous-tend la méthode du GMM est que des conditions de moments peuvent être utilisées non seulement pour spécifier un modèle mais également pour définir les paramètres de celui-ci. La meilleure façon de comprendre cette méthode est de la positionner par rapport aux méthodes d'estimation classiques, sujets qui font l'objet des prochaines sections.

2. LA MÉTHODE DES MOMENTS ET LES MCO

La méthode du GMM tire son intérêt du fait que plusieurs problèmes d'estimation sont une simple fonction des moments. Pour fixer les idées, considérons le modèle de régression linéaire suivant :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

où \mathbf{y} est un vecteur de dimension $(n \times 1)$, \mathbf{X} est une matrice de dimension $(n \times k)$, $\boldsymbol{\beta}$ est le vecteur des paramètres à estimer de dimension $(k \times 1)$ et \mathbf{e} , le vecteur des résidus, de dimension $(n \times 1)$. $\mathbf{e} \sim \mathbf{Q}(0, \sigma^2)$, \mathbf{Q} étant une distribution quelconque. Supposons également que \mathbf{X} soit orthogonale à \mathbf{e} , c'est-à-dire :

$$\mathbf{E}(\mathbf{X}^T \mathbf{e}) = \mathbf{0}$$

Nous voulons ici estimer le vecteur $\boldsymbol{\beta}$. Nous savons qu'au niveau de la population :

$$\mathbf{E}[\mathbf{X}^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})] = \mathbf{0}$$

puisque $\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$. La méthode transpose cette dernière équation au niveau échantillonnal. Pour ce faire, on remplace l'opérateur \mathbf{E} par son équivalent échantillonnal, soit la moyenne. On a donc :

$$\frac{1}{n} \mathbf{X}^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}$$

Comme nous savons que \mathbf{b} réalise l'orthogonalité entre la matrice d'observations \mathbf{X} et le vecteur des résidus \mathbf{e} , il est donc loisible de

5. Davidson, R. et J.G. Mackinnon (1993), *Estimation and Inference in Econometrics*, Oxford University Press, New York.

transposer ce résultat au niveau échantillonnal. $\hat{\beta}$, soit l'estimateur de β , devra donc réaliser l'égalité suivante :

$$\frac{1}{n} \mathbf{X}^T (\mathbf{y} - \mathbf{X}\hat{\beta}) = \mathbf{0}$$

On peut donc solutionner cette équation de moments pour $\hat{\beta}$:

$$\hat{\beta}_{\text{mom}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

soit l'estimateur des MCO ! On retrouve donc l'estimateur des MCO non pas en mimimisant la somme des résidus mais bien en recourant à l'équation implicite des moments.

3. LA MÉTHODE DES MOMENTS ET L'ESTIMATEUR DES VARIABLES INSTRUMENTALES

Supposons que :

$$E(\mathbf{X}^T \mathbf{e}) \neq 0$$

Nous devons alors recourir à des variables instrumentales pour renouer avec les conditions d'orthogonalité entre ces variables et les résidus :

$$E(\mathbf{Z}^T \mathbf{e}) = 0$$

On transpose facilement la section précédente au problème qui nous intéresse, soit déterminer un estimateur convergent de β . L'orthogonalité entre les variables instrumentales et les résidus au niveau échantillonnal s'écrit donc comme suit :

$$\frac{1}{n} \mathbf{Z}^T (\mathbf{y} - \mathbf{X}\hat{\beta}) = \mathbf{0}$$

La solution :

$$\hat{\beta} = (\mathbf{Z}^T \mathbf{X})^{-1} \mathbf{Z}^T \mathbf{y}$$

n'est envisageable que lorsqu'il est possible d'inverser la matrice $\mathbf{Z}^T\mathbf{X}$. Cette condition est valable si le système d'équations est exactement identifié, c'est-à-dire que les variables instrumentales permettent d'identifier exactement le vecteur β des paramètres. L'estimateur qui résulte de la condition des moments est appelé *estimateur des moindres carrés indirects*.

Dans la pratique, le nombre d'instruments excède habituellement le nombre de paramètres à estimer. On est alors confronté à un système suridentifié et la matrice $\mathbf{Z}^T\mathbf{X}$, qui n'est plus alors une matrice carrée, ne peut donc plus être inversée. On pourrait à la limite se délester d'instruments de telle sorte que le système d'équations soit exactement identifié, mais on ferait alors abstraction d'informations pertinentes pour l'estimation.

Par simple analogie avec les MCO, on pourrait estimer le vecteur β en minimisant la forme quadratique suivante, qui représente, grosso modo, la somme des résidus au carré :

$$\frac{1}{n} \left[\mathbf{Z}^T (\mathbf{y} - \mathbf{X}\hat{\beta}) \right]^T \mathbf{I} \left[\mathbf{Z}^T (\mathbf{y} - \mathbf{X}\hat{\beta}) \right]$$

On remarquera que l'on pondère ici les moments d'égale façon. L'estimateur qui en résulte s'avère convergent, mais comme le système de pondération des moments n'est pas optimal, l'estimateur obtenu ne fait pas figure de meilleur de sa classe. Hansen (1982)⁶ a dérivé l'estimateur optimal dans sa classe pour le problème précédent. Il suffit de minimiser le système suivant. L'estimateur qui en résulte est appelé GMM :

$$\text{Min}_{\hat{\beta}} \left\{ \frac{1}{n} \left[\mathbf{Z}^T (\mathbf{y} - \mathbf{X}\hat{\beta}) \right]^T \hat{\mathbf{V}}^{-1} \frac{1}{n} \left[\mathbf{Z}^T (\mathbf{y} - \mathbf{X}\hat{\beta}) \right] \right\}$$

6. Hansen, L. (1982), « Large Sample Properties of Generalized Method of Moments Estimator », *Econometrica*, 50, p. 1029-1054. Pour une présentation simplifiée, voir : Hall, A. (1992), « Some Aspects of Generalized Method of Moments Estimation », document de travail, North Carolina State University ; Ogaki, M. (1992), « Generalized Method of Moments : Econometric Applications », document de travail, University of Rochester. Ce chapitre est basé sur le cahier de recherche suivant : Racicot, F.-E. et R. Théoret (2001), « La méthode économétrique du GMM et l'estimation des paramètres de modèles financiers stochastiques. Analyse d'un cas : le modèle de taux d'intérêt de Schaefer et Schwartz », document de travail du CRG, École des sciences de la gestion.

où $\hat{\mathbf{V}}^{-1}$ est un estimateur convergent de $\left\{ \mathbf{V} \left[\frac{1}{n} (\mathbf{Z}^T \mathbf{e}) \right] \right\}^{-1}$. Selon cette formulation, une restriction estimée de façon moins précise⁷ se verra attribuer une pondération plus faible, par l'intermédiaire de la matrice $\hat{\mathbf{V}}^{-1}$, qu'une restriction qui est estimée avec plus de précision⁸.

Dans le cas où les erreurs sont homoscédastiques et indépendantes, l'estimateur qui découle de la minimisation du critère associé au GMM se confond avec l'estimateur des doubles moindres carrés (2SLS) :

$$\frac{\partial}{\partial \hat{\boldsymbol{\beta}}} \left\{ \frac{1}{n} \left[\mathbf{Z}^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \right]^T \hat{\mathbf{V}}^{-1} \frac{1}{n} \left[\mathbf{Z}^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \right] \right\} = 0$$

$$(\mathbf{X}^T \mathbf{Z}) \hat{\mathbf{V}}^{-1} (\mathbf{Z}^T \mathbf{y} - \mathbf{Z}^T \mathbf{X}\hat{\boldsymbol{\beta}}) = 0$$

De cette équation, il résulte que $\hat{\boldsymbol{\beta}}$ est égal à :

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{Z} \hat{\mathbf{V}}^{-1} \mathbf{Z}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{Z}) \hat{\mathbf{V}}^{-1} \mathbf{Z}^T \mathbf{y}$$

Il reste à spécifier la valeur de $\hat{\mathbf{V}}^{-1}$. Sous les hypothèses actuelles d'homoscédasticité et du caractère IID de l'innovation, la méthode des moments suggère que : $\hat{\mathbf{V}}^{-1} = \left((\hat{\sigma}^2 / n^2) \mathbf{Z}^T \mathbf{Z} \right)^{-1}$. La forme finale de $\hat{\boldsymbol{\beta}}$ est de :

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{X} \right)^{-1} \left(\mathbf{X}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{y} \right)$$

Ce qui est bien l'estimateur des doubles moindres carrés.

Ouvrons ici une parenthèse les doubles moindres carrés non linéaires. Comme cela fut mentionné auparavant, \mathbf{y} a alors la forme :

$$\mathbf{y} = f(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}$$

7. C'est-à-dire qui comporte une variance plus élevée.

8. C'est-à-dire qui comporte une variance plus faible.

Pour obtenir les doubles moindres carrés non linéaires selon la méthode des GMM, il suffit de remplacer dans la fonction précédente $\underset{\hat{\beta}}{\text{Min}}(\cdot)$ l'expression $\left[\mathbf{Z}^T (\mathbf{y} - \mathbf{X}\hat{\beta}) \right]$ par $\left[\mathbf{Z}^T (\mathbf{y} - f(\mathbf{X}, \hat{\beta})) \right]$.

4. GMM ET CONDITIONS D'ORTHOGONALITÉ

De façon générale, la méthode du GMM est basée sur des conditions d'orthogonalité de la population :

$$E[g(\mathbf{y}, \mathbf{X}, \theta_0)] = \mathbf{0}$$

où $g(\cdot)$ est une fonction continue de (\mathbf{y}, \mathbf{X}) et de paramètres qui sont uniques et qui font que l'espérance égale zéro.

Précisons les propriétés de l'estimateur des GMM en les reliant aux conditions d'orthogonalité. L'estimateur des GMM résulte du problème de minimisation suivant :

$$\underset{\hat{\theta}}{\text{Min}} \left[\mathbf{m}(\mathbf{y}, \mathbf{X}, \hat{\theta})^T \mathbf{W}_n \mathbf{m}(\mathbf{y}, \mathbf{X}, \hat{\theta}) \right]$$

où $\mathbf{m}(\mathbf{y}, \mathbf{X}, \theta) = 1/n \sum_{i=1}^n g(\mathbf{y}_i, \mathbf{X}_i, \theta)$ et où \mathbf{W}_n peut être estimée par la matrice de White (1980), dite encore HCCM, ceci pour avoir le meilleur estimateur convergent de cette matrice. On trouve $\hat{\theta}$ en solutionnant la condition suivante de premier ordre :

$$\frac{\partial \mathbf{m}(\cdot)}{\partial \hat{\theta}^T} \mathbf{W}_n \mathbf{m}(\cdot) = \mathbf{0}$$

où l'on symbolise la matrice des dérivées premières par $\mathbf{G} = \frac{\partial \mathbf{m}}{\partial \theta}$.

La distribution de $\hat{\theta}$ résulte d'une approximation de Taylor de $g(\hat{\theta})$ autour de θ_0 . Dans son article, Hansen (1982) a démontré que :

$$\hat{\theta} \overset{a}{\sim} N \left(\theta_0, (\mathbf{G}^T \mathbf{W} \mathbf{G})^{-1} \mathbf{G}^T \mathbf{W} \Omega \mathbf{W} \mathbf{G} (\mathbf{G}^T \mathbf{W} \mathbf{G})^{-1} \right)$$

où $\Omega = E\left[g(\mathbf{y}, \mathbf{X}, \theta_0)g(\mathbf{y}, \mathbf{X}, \theta_0)^T\right]$ qui est tout simplement la variance de la condition des moments. Hansen (1982) a démontré qu'un choix optimal de W est simplement un estimé hétéroscédastique et convergent de :

$$E\left[g(\mathbf{y}, \mathbf{X}, \theta_0)g(\mathbf{y}, \mathbf{X}, \theta_0)^T\right]^{-1} = \Omega^{-1}$$

Si l'on dispose d'un estimé convergent de $\hat{\theta}$, alors un estimé de Ω^{-1} peut être obtenu. La distribution asymptotique de $\hat{\theta}$ se réduit alors à :

$$\hat{\theta} \overset{a}{\sim} N\left(\theta_0, \left(\mathbf{G}^T \Omega^{-1} \mathbf{G}\right)^{-1}\right)$$

Quelle que soit la matrice de pondération \mathbf{W} , l'estimateur du GMM est toujours convergent et non biaisé asymptotiquement et lorsque l'on recourt au W optimal, l'estimateur du GMM est alors asymptotiquement efficient dans la classe des estimateurs définis par les conditions d'orthogonalité. Finalement, la matrice de White (1980), dite encore HCCM, semble être le meilleur candidat pour cette matrice de pondération.

5. MAXIMUM DE VRAISEMBLANCE ET GMM

La méthode du maximum de vraisemblance s'intègre facilement dans l'univers du GMM. Pour estimer les paramètres selon cette méthode, il suffit de réaliser la condition suivante :

$$\frac{\partial \ln L(\theta, \mathbf{X})}{\partial \theta} = 0$$

c'est-à-dire qu'au maximum de vraisemblance, le score est nul. Dans le langage du GMM, cette condition s'écrit comme suit :

$$m(\mathbf{y}, \mathbf{X}, \theta) \equiv \frac{\partial \ln L}{\partial \theta} = \mathbf{0}$$

ce qui est simplement une condition de moments.

Précisons d'avantage. Dans le cas le plus simple, la méthode du GMM est :

$$\underset{\hat{\theta}}{\text{Min}} \left(m(\mathbf{y}, \mathbf{X}, \theta)^T \mathbf{H}^{-1} m(\mathbf{y}, \mathbf{X}, \theta) \right)$$

où \mathbf{H} est la matrice des poids, soit la variance des conditions des moments. En termes de la méthode du maximum de vraisemblance, cette matrice s'écrit :

$$\mathbf{H} = -\text{E} \left(\frac{\partial^2 \ln L}{\partial \theta \partial \theta^T} \right)$$

soit la matrice d'information. L'opération Min s'écrit dans ce cas :

$$\frac{\partial^2 \ln L}{\partial \theta \partial \theta^T} \mathbf{H}^{-1} \frac{\partial \ln L}{\partial \theta} = 0$$

soit l'équation qui définit l'estimateur du maximum de vraisemblance. Par conséquent, l'estimateur du maximum de vraisemblance peut être vu comme un estimateur GMM.

Certains chercheurs préfèrent recourir au GMM plutôt qu'à la méthode du maximum de vraisemblance. Voici leurs motifs : 1) l'estimateur du maximum de vraisemblance est parfois difficile à calculer. Il existe alors un GMM qui est asymptotiquement moins efficace que l'estimateur du maximum de vraisemblance, mais qui est toutefois convergent et plus facile à calculer ; 2) quelquefois, l'on ne dispose pas assez d'information sur la distribution empirique des données pour spécifier complètement la fonction de vraisemblance. Par contre, on dispose habituellement d'une quantité suffisante d'information pour spécifier les conditions des moments et recourir ainsi à l'estimateur GMM.

L'estimateur du GMM s'intègre donc dans la classe des estimateurs semi-paramétriques.

6. APPLICATIONS

Dans cette section, nous visons à montrer comment les paramètres d'une équation différentielle stochastique peuvent être estimés en recourant à la méthode du GMM. En effet, l'évaluation de ces paramètres s'avère de toute première importance pour estimer les prix des options. Les paramètres que nous voulons estimer dans cette section sont ceux du modèle stochastique bifactoriel de taux d'intérêt de Schaefer et Schwartz (1984)⁹. Mais avant de présenter ce modèle, nous rappelons les formes générales des mouvements browniens en temps continu, puis en temps discret. Nous verrons alors comment la discrétisation d'une équation différentielle stochastique peut aboutir à un processus autorégressif. Nous serons alors à même de nous attaquer à l'estimation des paramètres du modèle de Schaefer et Schwartz par la méthode du GMM.

Le recours à la méthode du GMM marque une avancée importante pour l'estimation des paramètres d'une équation différentielle stochastique dans le domaine de la finance empirique, notamment dans le champ de la théorie des options et de la finance corporative. En effet, bien souvent par le passé, on estimait ces paramètres en les égalisant directement à leur moment empirique correspondant. Ainsi, on estimait le *drift* d'un processus stochastique par la moyenne de la variable analysée et sa volatilité par son écart-type historique. Cette procédure n'apparaît pas des plus satisfaisantes, il s'en faut de beaucoup, quand on sait que la méthode du GMM permet d'estimer ces paramètres de façon robuste.

La théorie des options est basée sur la résolution analytique ou numérique d'équations différentielles stochastiques¹⁰. L'une des formes les plus simples de cette catégorie d'équations est mouvement brownien avec *drift*. L'équation différentielle de ce mouvement est la suivante :

$$dx = a(x, t)dt + b(x, t)dz$$

9. Schaefer, S.M. et Schwartz, E.S. (1984), «A Two Factor Model of the Term Structure : An Approximate Analytical Solution», *Journal of Financial and Quantitative Analysis*, 19, p. 413-424.

10. Pour plus détails à ce sujet, on consultera : Dixit, A.K. et R.S. Pindyck (1994), *Investment under Uncertainty*, Princeton University Press, Princeton, chap. 3 et 4.

où $a(\cdot)$ désigne le drift et où dz suit un processus de Wiener, c'est-à-dire : $dz = \varepsilon \sqrt{dt}$, $\varepsilon \sim N(0, 1)$ ¹¹. Le processus stochastique $x(t)$, défini par l'équation précédente, est appelé *processus d'Ito*.

Considérons l'espérance et la variance de dx . L'espérance de dx est égale à :

$$E(dx) = a(x, t)dt$$

puisque $E(dz) = 0$. La variance de dx est égale à :

$$V(dx) = E(dx)^2 - (E(dx))^2$$

Cette variance contient des termes en dt , en $(dt)^2$ et en $(dt)(dz)$, ce produit étant de l'ordre $(dt)^{3/2}$. Comme dt est infinitésimalement petit, les termes en $(dt)^2$ et en $(dt)^{3/2}$ peuvent être ignorés. La variance de dx est donc de :

$$V(dx) = b^2(x, t)dt$$

On appelle $a(x, t)$ le *drift* instantané du processus d'Ito et $b^2(x, t)$, la variance instantanée.

Un cas spécial du processus d'Ito est le mouvement brownien géométrique qui s'écrit comme suit :

$$dx = \alpha x dt + \sigma x dz$$

Nous nous intéressons maintenant à la relation entre x et son logarithme. Soit : $F(x) = \log(x)$. En appliquant le lemme d'Ito étudié auparavant, l'équation différentielle de F est la suivante :

$$dF = \left(\alpha - \frac{1}{2} \sigma^2 \right) dt + \sigma dz$$

11. L'espérance mathématique d'un processus de Wiener est donc de : $E(dz) = 0$, et sa variance : $V(dz) = E[(dz)^2] = dt$. Par conséquent, l'écart-type de dz est de \sqrt{dt} , ce qui revient à dire que l'incertitude augmente avec la racine carrée du temps.

En appliquant les règles antérieures, on calcule que le changement dans le logarithme de x est distribué normalement avec une espérance

$\left(\alpha - \frac{1}{2}\sigma^2\right)t$ et une variance σ^2t . On peut montrer que l'espérance de $x(t)$ est égale à :

$$E[x(t)] = x_0 e^{\alpha t}$$

et sa variance :

$$V[x(t)] = x_0^2 e^{2\alpha t} \left(e^{\sigma^2 t} - 1 \right)$$

Ces résultats sont d'une grande utilité pour calculer l'espérance d'un mouvement géométrique brownien :

$$E\left[\int_0^\infty x(t) e^{-rt} dt\right] = \int_0^\infty x_0 e^{-(r-\alpha)t} dt = x_0 (r - \alpha)$$

Les mouvements browniens géométriques ont une grande utilité dans la modélisation des variables financières comme les prix des titres et les taux d'intérêt.

Nous sommes maintenant en mesure d'aborder les processus browniens de retour vers la moyenne, dits encore processus Ornstein-Uhlenbeck. Ces processus sont très plausibles pour modéliser certaines variables financières qui ont tendance à retourner à long terme vers un niveau dit *normal*, tels les taux d'intérêt. L'équation différentielle générale de tels processus est la suivante :

$$dx = \theta(x - \bar{x})dt + \sigma dz$$

où

θ : vitesse de retour à la moyenne

\bar{x} : niveau normal de x (niveau auquel il retourne à long terme)

Calculons l'espérance mathématique et la variance d'un processus de retour vers la moyenne. L'espérance de x_t est égale à :

$$E(x_t) = \bar{x} + (x_0 - \bar{x})e^{-\theta t}$$

Et sa variance :

$$V(x_t - \bar{x}) = \frac{\sigma^2}{2\theta} (1 - e^{-2\theta t})$$

On se rend compte que la valeur espérée de x_t tend vers \bar{x} quand $t \rightarrow \infty$ et que sa variance converge vers $\frac{\sigma^2}{2\theta}$. Si $\theta \rightarrow \infty$, alors $V(x_t) \rightarrow 0$, c'est-à-dire que x_t ne dévie jamais de \bar{x} , même temporairement. Par ailleurs, si $\theta \rightarrow 0$, alors $V(x_t) \rightarrow \sigma^2 t$. x_t suit dans ce cas un mouvement brownien simple à la limite.

Broze *et al.* (1993)¹² ont écrit une forme générale pour les processus browniens de taux d'intérêt qui synthétise l'ensemble des modèles bien connus. Cette forme est la suivante :

$$dr_t = (\alpha + \beta r_t) dt + \sigma_0 (r_t^\gamma + \sigma_1) dW_t$$

Le tableau 11.1, tiré de Broze *et al.* (1993), montre comment cette forme générale intègre les modèles stochastiques de taux d'intérêt bien connus.

Pour estimer les paramètres d'un processus brownien de retour vers la moyenne, nous devons discrétiser un tel processus. En fait, le processus continu est la valeur limite quand $\Delta t \rightarrow \infty$ du processus autorégressif suivant en temps discret :

$$x_t - x_{t-1} = \bar{x}(1 - e^{-\theta}) + (e^{-\theta} - 1)x_{t-1} + \varepsilon_t$$

où

$$\varepsilon_t \sim N\left(0, \frac{\sigma^2}{2\theta} (1 - e^{-2\theta})\right)$$

12. Broze, L., O. Scaillet et J.-M. Zakoian (1993), « Testing for Continuous-Time Models of the Short-Term Interest Rate », Centre d'économie mathématique et d'économétrie, Université de Bruxelles.

TABLEAU 11.1 Modèles stochastiques de taux d'intérêt

Modèle	α	β	σ_0	σ_1	γ
Merton (1973) $dr = \alpha dt + \sigma_0 dW_t$		0		0	0
Vasicek (1977) $dr_t = (\alpha + \beta r_t) dt + \sigma_0 dW_t$				0	0
Cox, Ingersoll et Ross (1985) $dr_t = (\alpha + \beta r_t) dt + \sigma_0 \sqrt{r_t} dW_t$				0	$\frac{1}{2}$
Dothan (1978) $dr_t = \sigma_0 r_t dW_t$	0	0		0	1
Processus brownien géométrique $dr_t = \beta r_t dt + \sigma_0 r_t dW_t$	0			0	1
Brennan et Schwartz (1980) $dr_t = (\alpha + \beta r_t) dt + \sigma_0 r_t dW_t$				0	1
Cox, Ingersoll et Ross (1980) $dr_t = \sigma_0 r_t^{\frac{3}{2}} dW_t$	0	0		0	$\frac{3}{2}$
Élasticité constante de la variance $dr_t = \beta r_t dt + \sigma_0 r_t^\gamma dW_t$	0			0	
Chan, Karolyi, Longstaff et Sanders (1992) $dr_t = (\alpha + \beta r_t) dt + \sigma_0 r_t^\gamma dW_t$				0	

On peut donc estimer les paramètres d'un processus brownien de retour vers la moyenne en utilisant des données discrètes et en recourant à l'équation suivante :

$$x_t - x_{t-1} = a + bx_{t-1} + \varepsilon_t$$

On peut récupérer facilement la structure initiale du mouvement brownien comme suit, en sachant que $\hat{a} = \bar{x}(1 - e^{-\hat{\theta}})$ et $\hat{b} = e^{-\hat{\theta}} - 1$. \bar{x} est égal à l'expression suivante :

$$\bar{x} = -\frac{\hat{a}}{\hat{b}} = -\frac{\bar{x}(1 - e^{-\hat{\theta}})}{e^{-\hat{\theta}} - 1}$$

En procédant de la sorte, on trouve :

$$\hat{\theta} = -\log(1 + \hat{b})$$

et

$$\hat{\sigma} = \hat{\sigma}_\varepsilon \sqrt{\frac{2\log(1 + \hat{b})}{(1 + \hat{b}^2) - 1}}$$

Il est facile de généraliser le processus brownien de retour vers la moyenne qui vient d'être exposé. Par exemple, on peut supposer que x_t retourne vers \bar{x} à long terme, mais que la variance du processus augmente avec x . Le processus brownien s'écrit dans ce cas :

$$dx = \theta(\bar{x} - x)dt + \sigma x dz$$

Avant de poursuivre, nous ouvrons une parenthèse sur les procédures de discrétisation puisque l'on doit utiliser de telles procédures pour passer de la forme théorique d'une équation différentielle stochastique, généralement exprimée en temps continu et qui ne se prête pas comme telle à l'estimation empirique, à sa forme discrète, qui peut alors faire l'objet d'une estimation¹³.

Soit une équation de la forme suivante :

$$dx(t) = adt + bdW_t$$

où W_t désigne un processus de Wiener. Cette équation est exprimée en temps continu et on désire la discrétiser. Nous envisageons dans un premier temps la discrétisation de premier ordre, encore appelée discrétisation d'Euler. Ce type de discrétisation est qualifié de *crude* par Gouriéroux (1996), car elle n'est pas une forme de discrétisation exacte, forme que nous avons exposée antérieurement pour le processus Ornstein-Uhlenbeck¹⁴. La discrétisation d'Euler et la discrétisation

13. Pour cette section, nous nous inspirons de : Jegadeesh, N. et B. Tuckman (2000), *Advanced Fixed-Income Valuation Tools*, John Wiley and Sons, New York, chap. 13.

14. Voir à cet effet : Gouriéroux, C. et A. Montfort (1996), *Simulation-Based Econometric Methods*, Oxford University Press, Oxford, chap. 6. Ce chapitre est excellent en ce qui concerne l'estimation d'équations différentielles stochastiques à partir de données discrètes ou de conditions de moments.

exacte sont toutefois réconciliables, comme le lecteur sera à même de le constater¹⁵.

Pour ce qui concerne l'approximation d'Euler, l'idée de base est ici d'approximer une équation différentielle stochastique par une série de Taylor du premier degré. Pour ce faire, on divise l'intervalle $[0, T]$ en k sous-intervalles identiques de longueur h . L'approximation d'Euler de la fonction différentielle stochastique dx_t est de :

$$\hat{x}_{k+1} - \hat{x}_k = ah + bz_{k+1} \sqrt{h}$$

où $z \sim N(0, 1)$. Cette équation fournit une approximation pour la distribution de \hat{x}_{k+1} . L'approximation est plus précise quand $h \rightarrow 0$.

Comme nous l'avons mentionné auparavant, Broze *et al.* (1993) ont écrit une forme générale pour les mouvements browniens de taux d'intérêt, c'est-à-dire :

$$dr_t = (\alpha + \beta r_t)dt + \sigma_0 (r_t^\gamma + \sigma_1) dW_t$$

La discrétisation d'Euler de cette équation est la suivante :

$$r_{t+1}^{(h)} = \alpha_h + (\beta_h + 1)r_t^{(h)} + \sigma_{0,h} \left(|r_t^{(h)}|^\gamma + \sigma_{1,h} \right) Z_{t+1}$$

où (Z_t) est une séquence de variables gaussiennes indépendantes et où :

$$\alpha_h = \alpha h; \beta_h = \beta h; \sigma_{0,h} = \sigma_0 \sqrt{h}; \sigma_{1,h} = \sigma_1$$

Comme cela vient d'être mentionné, l'approximation d'Euler est une approximation de Taylor du premier degré aux variations observées de $x(t)$. Mais on peut obtenir une discrétisation plus précise et qui converge plus rapidement en recourant à une série de Taylor du second degré pour convertir en temps discret la variation de $x(t)$. Cette procédure est appelée *schème de Milstein*. Elle s'applique lorsque les paramètres de l'équation différentielle dépendent de x et de t . Mais du fait des différences entre le calcul différentiel classique et le calcul

15. En fait, la discrétisation exacte permet d'identifier les coefficients d'un processus brownien dont la version continue a été discrétisée par une approximation d'Euler pour des fins d'estimation.

différentiel stochastique, le schème de Milstein est beaucoup plus complexe que les schèmes de second degré généralement utilisés dans le cas des équations différentielles ordinaires ou non stochastiques. Les schèmes de Milstein débordent de l'objet de l'actuel *Traité d'économétrie financière*. Pour plus de détails à ce sujet, le lecteur se référera à Jegadeesh et Tuckman (2000).

Forts de ces développements, nous pouvons maintenant envisager l'estimation du modèle de Schaefer et Schwartz (SS), un modèle bifactoriel de taux d'intérêt. Les deux facteurs sont le taux à long terme et l'écart de rendement qui est la différence entre le taux à long terme et le taux à court terme. Ils utilisent l'écart de rendement plutôt que le taux à court terme, car il a été démontré empiriquement que l'écart de rendement est orthogonal au taux à long terme, ce qui facilite l'estimation. Le modèle SS s'écrit comme suit :

$$\begin{aligned} ds &= m(\mu - s)dt + \gamma dz_1 \\ dl &= \beta_2 dt + \sigma \sqrt{l} dz_2 \\ dz_1 dz_2 &= 0 \end{aligned}$$

où s est l'écart de rendement, l est le taux à long terme, m est le coefficient de retour vers la moyenne pour l'écart de rendement, μ est la moyenne à long terme de l'écart de rendement, γ est le coefficient de variance pour l'écart de rendement, s est le coefficient de variance pour le taux à long terme et β_2 représente la tendance (*drift*) du taux à long terme.

Envisageons d'abord l'estimation des deux équations du modèle de SS par la méthode des MCO. Considérons d'abord l'équation de l'écart de rendement :

$$ds_t = m(u - s_t)dt + \gamma dw_t$$

La discrétisation d'Euler de cette équation est la suivante :

$$\begin{aligned} s_t - s_{t-1} &= m^*(u^* - y_{t-1}) + \gamma^* \varepsilon_t^* = m^* u^* - m^* y_{t-1} \\ &\quad + \gamma^* \varepsilon_t^* = c + m^{**} y_{t-1} + \varepsilon_t^{**} \end{aligned}$$

or, la discrétisation exacte de ds_t est la suivante :

$$s_t - s_{t-1} = u(1 - e^{-m}) + (e^{-m} - 1)s_{t-1} + \varepsilon_t$$

$$\text{où } \varepsilon_t \sim N\left(0, \frac{\sigma^2}{2m} (1 - e^{-2m})\right)$$

On peut estimer cette dernière équation par les MCO de la façon suivante :

$$s_t - s_{t-1} = a + bs_{t-1} + \varepsilon_t$$

ce qui est un processus autorégressif d'ordre 1. En comparant la discrétisation d'Euler avec la discrétisation exacte, on identifie les paramètres de l'équation initiale comme suit :

$$u = -\frac{\hat{a}}{\hat{b}}$$

$$m = -\log(1 + \hat{b})$$

$$\hat{\sigma} = \hat{\sigma}_\varepsilon \sqrt{\frac{2\log(1 + \hat{b})}{(1 + \hat{b})^2 - 1}}$$

Pour estimer cette équation, nous avons utilisé des données mensuelles canadiennes s'étirant de 1949 à 1999. Le taux à court terme retenu est celui des bons du Trésor à trois mois et celui à long terme, le taux des obligations fédérales à dix ans. Nous avons d'abord estimé la première équation du modèle de SS selon la méthode des MCO. Les résultats apparaissent au tableau 11.2.

Comme on peut le constater, les deux coefficients estimés \hat{a} et \hat{b} sont hautement significatifs, à en juger par le niveau des statistiques t et des p -values. L'écart de rendement à long terme est égal, en vertu de la discrétisation exacte à :

$$u = -\frac{\hat{a}}{\hat{b}} = 1,2486$$

TABLEAU 11.2

System: STRUC				
Estimation Method: Least Squares				
Date: 04/26/00 Time: 19:00				
Sample: 2 627				
Included observations: 626				
Total system (balanced) observations 1252				
	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.043168	0.020411	2.114887	0.0346
C(2)	-0.034634	0.010405	-3.328516	0.0009
C(3)	0.009544	0.007851	1.215625	0.2244
Determinant residual covariance		0.000911		
Equation: $D(\text{SPR}) = C(1) + C(2) * \text{SPR}(-1)$				
Observations: 626				
R-squared	0.017445	Mean dependent var	-0.002602	
Ajusted R-squared	0.015870	S.D. dependent var	0.380444	
S.E. of regression	0.377413	Sum squared resid	88.88281	
Durbin-Watson stat	1.274598			
Equation: $D(L)/(L-1)^{(1/2)} = C(3)/(L-1)^{(1/2)}$				
Observations: 626				
R-squared	0.000951	Mean dependent var	0.003014	
Ajusted R-squared	0.000951	S.D. dependent var	0.080303	
S.E. of regression	0.080265	Sum squared resid	4.026560	
Durbin-Watson stat	1.324424			

Cette valeur pour l'écart de rendement sur longue période entre les obligations à dix ans et les bons du Trésor à 3 mois nous conforte dans nos attentes, quand on sait que le Canada devait maintenir traditionnellement un écart de rendement substantiel pour encourager l'offre de fonds à long terme, nécessaire au financement des très nombreux projets à long terme, le Canada n'ayant pas encore atteint sa période de maturité industrielle comme les États-Unis. Ce raisonnement est corroboré par l'estimation du même coefficient par Jagadeesh et Tuckman (2000) qui ont obtenu, pour ce même coefficient, un résultat de -0,0171 pour les États-Unis, ce qui démontre la maturité industrielle avancée de ce pays.

Passons maintenant à l'estimation de la vitesse d'ajustement de l'écart de rendement vers sa valeur de longue période. Toujours selon la discrétisation exacte, ce coefficient est égal à :

$$m = -\log(1 + \hat{b}) = -\log(1 - 0,0346) = 0,0352$$

Nous prenons donc acte que la vitesse d'ajustement de l'écart de rendement est plutôt lente. Ce résultat était attendu, car on sait que cet écart peut dévier de sa moyenne pendant de longues périodes de temps au Canada. Ce résultat est appuyé par l'estimation de l'écart-type du mouvement brownien de l'écart de rendement canadien dont nous donnons maintenant le résultat.

L'écart-type estimé de l'équation de l'écart de rendement est égal à :

$$\hat{\gamma} = \hat{\sigma}_\varepsilon \sqrt{\frac{2\log(1 + \hat{b})}{(1 + \hat{b})^2 - 1}} = 0,3774 \sqrt{\frac{2\log(0,9654)}{(0,9654)^2 - 1}} = 0,3840$$

L'écart-type estimé de l'équation stochastique de l'écart de rendement, soit $\hat{\gamma}$, est donc relativement important, ce qui implique que l'écart de rendement peut dévier de sa valeur à long terme pendant une période de temps appréciable.

Envisageons maintenant l'estimation de l'équation stochastique de la différentielle du taux à long terme du modèle SS, soit :

$$dl = udt + \sigma\sqrt{l}dW$$

Pour être en mesure d'estimer cette équation différentielle stochastique, nous recourons à la discrétisation d'Euler. À la suite de cette discrétisation, cette dernière équation s'écrit :

$$l_t - l_{t-1} = u + \sigma\sqrt{l_{t-1}}\varepsilon_t$$

Cette équation incorpore à l'évidence une forme d'hétéroscédasticité conditionnelle. Pour y pallier, nous pondérons chacune des observa-

tions des variables de l'équation par $\sqrt{\frac{1}{l_{t-1}}}$. À la suite de cette transformation, l'équation différentielle de la variation du taux à long terme devient :

$$\frac{l_t - l_{t-1}}{\sqrt{l_{t-1}}} = \frac{u}{\sqrt{l_{t-1}}} + \sigma \varepsilon_t = \frac{u}{\sqrt{l_{t-1}}} + \varepsilon_t^*$$

où $\varepsilon_t^* = \sigma \varepsilon_t$. Nous estimons cette équation en recourant dans un premier temps aux moindres carrés pondérés.

Comme cela apparaît au tableau 11.2, la valeur estimée de u , soit la valeur à long terme ou *drift* de la différentielle du taux à long terme, est égale à 0,0095 et n'est pas significative au seuil $\alpha = 5\%$. Pour estimer le σ de l'équation différentielle stochastique du taux à long terme, nous recourons au résultat suivant :

$$V(\varepsilon_t^*) = V(\sigma \varepsilon_t) = \sigma^2 V(\varepsilon_t) = \sigma^2$$

puisque $\varepsilon_t \sim N(0, 1)$. L'écart-type σ qui apparaît dans l'équation différentielle stochastique est donc égal à l'écart-type du terme d'erreur de la régression effectuée, qui est égal à : 0,0802¹⁶.

Comme le modèle de SS comporte deux équations, nous avons également estimé ses paramètres en recourant à la méthode des doubles moindres carrés. Les variables instrumentales retenues sont des décalages sur l'écart de rendement entre les obligations fédérales à long terme américaines et le taux de rendement des bons du Trésor américain. Cet écart semblait en effet bien jouer son rôle de variable instrumentale, à savoir une corrélation élevée avec les variables endogènes du modèle et une orthogonalité avec les termes d'erreur des équations du modèle.

Les résultats de cette estimation apparaissent au tableau 11.3.

En refaisant les calculs précédents pour identifier les paramètres des équations du modèle de SS, le lecteur sera à même de constater que les résultats de cette estimation ne diffèrent pas sensiblement de ceux obtenus par la méthode des moindres carrés pondérés.

16. À remarquer que, pour ce qui concerne l'équation différentielle stochastique du taux à long terme, les coefficients dont les valeurs sont données dans le texte correspondent à la discrétisation d'Euler et non à la discrétisation exacte.

TABLEAU 11.3

System : STRUC				
Estimation Method : Two-Stage Least Squares				
Date : 04/26/00 Time : 19:00				
Sample : 4 627				
Included observations : 624				
Total system (balanced) observations 1248				
	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.077456	0.020874	3.710549	0.0002
C(2)	-0.061026	0.010851	-5.623929	0.0000
C(3)	0.008725	0.007933	1.099815	0.2716
Determinant residual covariance		0.000924		
Equation : D(STR)= C(1)+C(2)*STR(-1)				
Observations : 624				
R-squared	0.007758	Mean dependent var	-0.002979	
Ajusted R-squared	0.006163	S.D. dependent var	0.380993	
S.E. of regression	0.379817	Sum squared resid	89.73033	
Durbin-Watson stat	1.230532			
Equation : D(L)/(L-1)^(1/2) = C(3)/(L-1)^(1/2)				
Observations : 624				
R-squared	0.000788	Mean dependent var	0.002804	
Ajusted R-squared	0.000788	S.D. dependent var	0.080341	
S.E. of regression	0.080309	Sum squared resid	4.018078	
Durbin-Watson stat	1.326180			

Nous en arrivons maintenant à l'objet principal de cette section, soit l'estimation des paramètres du modèle de SS par la méthode du GMM¹⁷. Comme nous l'avons vu auparavant, cette méthode exige l'estimation d'une matrice de pondération des moments, appelons-la

17. Les conditions d'orthogonalité ou moments du modèle de SS sont les suivantes :

$$E[\varepsilon_s Z_{t-1}] = 0; E[(\varepsilon_s^2 - \gamma^2) Z_{t-1}] = 0; E[\varepsilon_1 Z_{t-1}] = 0; E[(\varepsilon_1^2 - \sigma^2) Z_{t-1}] = 0;$$

$$E[(\varepsilon_s \varepsilon_1 - \rho \sigma \gamma \sqrt{1}) Z_{t-1}] = 0, \text{ où } Z_{t-1} \text{ désigne la matrice des variables instrumentales décalées d'une période; } \varepsilon_s, \text{ le terme d'erreur de la régression de l'écart de rendement; } \varepsilon_1, \text{ le terme d'erreur de la régression du taux à long terme et } \rho, \text{ la corrélation entre } \varepsilon_s \text{ et } \varepsilon_1.$$

Ω^{-1} , où Ω est la matrice variance-covariance des moments. Le logiciel EViews 3.1 offre plusieurs options pour estimer Ω .

- i) La matrice de White, encore appelée HCCM (*Heteroskedasticity Consistent Covariance Matrix*). Nous avons déjà frayé avec cette matrice. Rappelons ici le principe de sa construction dans le contexte de la méthode du GMM.

Soit les moments suivants, qui sont en fait des conditions d'orthogonalité entre le vecteur des variables instrumentales \mathbf{Z} retenu pour l'estimation et les résidus de la régression :

$$\mathbf{m}(\boldsymbol{\theta}, \mathbf{y}, \mathbf{X}, \mathbf{z}) = \mathbf{Z}^T \mathbf{u}(\boldsymbol{\theta}, \mathbf{y}, \mathbf{X}) = \mathbf{0}$$

où $\boldsymbol{\theta}$ est le vecteur des paramètres à estimer. Par exemple, pour l'estimateur des MCO, la condition d'orthogonalité est :

$$\mathbf{X}^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}$$

Pour des raisons d'identification des paramètres, il faut que le nombre de variables instrumentales soit au moins égal au nombre de paramètres à estimer.

Si la matrice de White est retenue pour calculer les moments, EViews calcule cette matrice comme suit :

$$\hat{\boldsymbol{\Omega}} = \frac{1}{T - k} \sum_{t=1}^T \mathbf{Z}_t^T \mathbf{u}_t \mathbf{u}_t^T \mathbf{Z}_t$$

où \mathbf{u}_t est le vecteur des résidus estimés¹⁸.

- ii) La matrice dite HAC (*heteroskedastic and autocorrelation consistent covariance matrix*). Pour les aspects techniques de la construction de cette matrice, nous reportons le lecteur au manuel de référence du logiciel EViews. Cette construction repose entre autres sur le recours à des kernels de façon à ce que la matrice HAC soit semi-définie positive.

L'estimation des paramètres du modèle de SS par la méthode du GMM apparaît au tableau 11.4.

18. Pour un exposé plus élaboré de la matrice de White dans le contexte de la méthode du maximum de vraisemblance, on consultera : Hendry, D.F. (1995), *Dynamic Econometrics: Advanced Texts in Econometrics*, Oxford University Press, Oxford, chap. 10, section 10.10.

TABLEAU 11.4

System : STRUC
 Estimation Method : Generalized Method of Moments
 Date : 04/26/00 Time : 19:00
 Sample : 4 627
 Included observations : 624
 Total system (balanced) observations 1248
 No prewhitening
 Bandwidth : Fixed (6)
 Kernel : Bartlett
 Convergence achieved after : 4 weight matrices. 5 total coef iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.076434	0.028534	2.678673	0.0075
C(2)	-0.054968	0.015209	-3.614102	0.0003
C(3)	0.015096	0.007348	2.054310	0.0402

Determinant residual covariance 0.000921
 J-statistic 0.013064

Equation : $D(\text{SPR}) = C(1) + C(2) * \text{SPR}(-1)$
 Observations : 624

R-squared	0.011489	Mean dependent var	-0.002979
Ajusted R-squared	0.009899	S.D. dependent var	0.380993
S.E. of regression	0.379102	Sum squared resid	89.39297
Durbin-Watson stat	1.242401		

Equation : $D(L)/(L(-1)^{(1/2)}) = C(3)/(L(-1)^{(1/2)})$
 Observations : 624

R-squared	-0.000229	Mean dependent var	0.002804
Ajusted R-squared	-0.000229	S.D. dependent var	0.080341
S.E. of regression	0.080350	Sum squared resid	4.022167
Durbin-Watson stat	1.324358		

On remarquera que les trois coefficients estimés sont maintenant significatifs, ce qui n'était pas le cas pour les deux autres méthodes d'estimation. Les résultats de cette estimation se rapprochent de ceux de la méthode des doubles moindres carrés.

L'estimation des paramètres des équations différentielles stochastiques trouve un très grand nombre d'applications dans le domaine de la théorie des produits dérivés. Par exemple, les scénarios de taux d'intérêt, qui entrent dans la détermination des prix des produits dérivés, font appel à de tels paramètres. Il faut également recourir à de tels paramètres pour évaluer les prix des obligations qui entrent dans l'évaluation des options sur taux d'intérêt. Dans le modèle de SS, le prix d'une obligation se calcule comme suit, une fois les paramètres du modèle stochastique de taux d'intérêt estimés :

$$\begin{aligned}
 P(s, l, \tau) &= X(s, \tau)Y(l, \tau) \\
 X(s, \tau) &= e^{[C(\tau) - sD(\tau)]} \\
 C(\tau) &= \frac{(D - \tau) \left(m \frac{(m\mu - \lambda\gamma) - \gamma^2}{2} \right)}{m^2} - \frac{\gamma^2 D^2}{4m} \\
 D(\tau) &= \frac{1 - e^{m\tau}}{m} \\
 Y(l, \tau) &= e^{[A(\tau) - lB(\tau)]} \\
 A(\tau) &= \left[\frac{2\alpha e^{(\hat{s} + \alpha)\tau/2}}{(\hat{s} + \alpha)(e^{\alpha\tau} - 1) + 2\alpha} \right]^2 \\
 B(\tau) &= \frac{2(e^{\alpha\tau} - 1)}{(\hat{s} + \alpha)(e^{\alpha\tau} - 1) + 2\alpha} \\
 \alpha &= \sqrt{\hat{s}^2 + 2\sigma^2}
 \end{aligned}$$

où τ désigne l'échéance de l'obligation et \hat{s} est défini implicitement par l'équation suivante :

$$\frac{l_0 \hat{\mu} - \sigma^2}{\hat{\mu}^2 \tau} (1 - e^{\hat{\mu}\tau}) + \frac{\sigma^2}{\hat{\mu}} = \frac{l_0 \hat{s} - \sigma^2}{\hat{s}^2 \tau} (1 - e^{-\hat{s}\tau}) + \frac{\sigma^2}{\hat{s}}$$

Par simulation, Jegadeesh et Tuckman (2000) ont obtenu une valeur de 0,7 pour λ , soit le prix du risque.

Test sur les restrictions de suridentification

Un modèle GMM est suridentifié quand le nombre des conditions d’orthogonalité est supérieur au nombre de paramètres à estimer. Hansen (1982) a suggéré un test pour vérifier si les moments échantillonnaux ne sont pas significativement différents de 0. Si les conditions d’orthogonalité de la population sont toutes vraies, alors :

$$\left[\sqrt{T} \mathbf{g}(\theta_0; y_T) \right]^T \mathbf{W} \left[\sqrt{T} \mathbf{g}(\theta_0; y_T) \right]^a \sim \chi^2(r)$$

où $g(\cdot)$ désigne le vecteur des r conditions d’orthogonalité. Dans cette équation, la fonction $g(\cdot)$ est évaluée à la vraie valeur θ_0 . Par ailleurs, l’estimateur GMM, désigné par $\hat{\theta}_T$, est typiquement une solution au système suivant :

$$\left\{ \begin{array}{c} \frac{\partial \mathbf{g}(\theta; y_T)}{\partial \theta^T} \Big|_{\theta = \hat{\theta}_T} \\ (a \times r) \end{array} \right\}^T \times \hat{\mathbf{W}}_T \times \left[\begin{array}{c} \mathbf{g}(\hat{\theta}_T; y_T) \\ (r \times 1) \end{array} \right] = \mathbf{0}_{(a \times 1)}$$

où a désigne le nombre de paramètres à estimer. Comme le système est suridentifié, $r > a$, on pourrait penser que l’expression antérieure ayant trait à la distribution asymptotique des moments pondérés vaut également lorsque le tout est évalué à $\hat{\theta}_T$. Mais cela est faux puisque le vecteur $\mathbf{g}(\hat{\theta}_T; y_T)$ contient $(r - a)$ variables aléatoires non dégénérées¹⁹. Évaluée à $\hat{\theta}_T$, la distribution asymptotique des moments est plutôt de :

$$\left[\sqrt{T} \mathbf{g}(\hat{\theta}_T; y_T) \right]^T \mathbf{W} \left[\sqrt{T} \mathbf{g}(\hat{\theta}_T; y_T) \right]^a \sim \chi^2(r - a)$$

19. Pour la preuve de ce résultat, voir : Hamilton, J.D. (1994), *op. cit.*

Le calcul de la statistique du test est trivial puisqu'elle est égale au produit de T et de la valeur de la fonction critère GMM évaluée à $\hat{\theta}_T$. Ce test sert à vérifier si le modèle est bien spécifié ou non. Mais selon Hamilton (1994), il faut compléter ce test car il peut facilement échouer.

Les hypothèses de ce test sont les suivantes :

H_0 : les conditions d'orthogonalité ne sont pas significativement différentes de 0.

H_1 : les conditions d'orthogonalité sont significativement différentes de 0, auquel cas le modèle est mal spécifié.

Si la statistique χ^2 calculée excède la χ_c^2 au seuil retenu, alors on rejette H_0 en faveur de H_1 .

ANNEXE

Tables statistiques

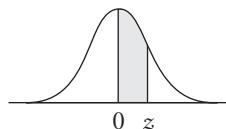
© 2001 – Presses de l'Université du Québec

Édifice Le Delta I, 2875, boul. Laurier, bureau 450, Québec, Québec G1V 2M2 • Tél. : (418) 657-4399 – www.puq.ca
Tiré : *Traité d'économétrie financière*, François-Éric Racicot et Raymond Théoret, ISBN 2-7605-1123-5 • D1123N

Tous droits de reproduction, de traduction ou d'adaptation réservés

TABLEAU A.1 Répartition de la loi normale centrée réduite

Les chiffres figurant dans le tableau donnent le rapport entre la surface en dessous de la courbe, comprise entre 0 et z et la surface totale se trouvant en dessous de cette courbe. Si z est négatif, on procède par symétrie.

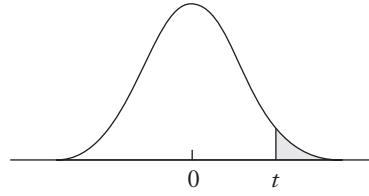


z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	,0000	,0040	,0080	,0120	,0160	,0199	,0239	,0279	,0319	,0359
0,1	,0398	,0438	,0478	,0517	,0557	,0596	,0636	,0675	,0714	,0753
0,2	,0793	,0832	,0871	,0910	,0948	,0987	,1026	,1064	,1103	,1141
0,3	,1179	,1217	,1255	,1293	,1331	,1368	,1406	,1443	,1480	,1517
0,4	,1554	,1591	,1628	,1664	,1700	,1736	,1772	,1808	,1844	,1879
0,5	,1915	,1950	,1985	,2019	,2054	,2088	,2123	,2157	,2190	,2224
0,6	,2257	,2291	,2324	,2357	,2389	,2422	,2454	,2486	,2517	,2549
0,7	,2580	,2611	,2642	,2673	,2703	,2734	,2764	,2794	,2823	,2852
0,8	,2881	,2910	,2939	,2967	,2995	,3023	,3051	,3078	,3106	,3133
0,9	,3159	,3186	,3212	,3238	,3264	,3289	,3315	,3340	,3365	,3389
1,0	,3413	,3438	,3461	,3485	,3508	,3531	,3554	,3577	,3599	,3621
1,1	,3643	,3665	,3686	,3708	,3729	,3749	,3770	,3790	,3810	,3830
1,2	,3849	,3869	,3888	,3907	,3925	,3944	,3962	,3980	,3997	,4015
1,3	,4032	,4049	,4066	,4082	,4099	,4115	,4131	,4147	,4162	,4177
1,4	,4192	,4207	,4222	,4236	,4251	,4265	,4279	,4292	,4306	,4319
1,5	,4332	,4345	,4357	,4370	,4382	,4394	,4406	,4418	,4429	,4441
1,6	,4452	,4463	,4474	,4484	,4495	,4505	,4515	,4525	,4535	,4545
1,7	,4554	,4564	,4573	,4582	,4591	,4599	,4608	,4616	,4625	,4633
1,8	,4641	,4649	,4656	,4664	,4671	,4678	,4686	,4693	,4699	,4706
1,9	,4713	,4719	,4726	,4732	,4738	,4744	,4750	,4756	,4761	,4767
2,0	,4772	,4778	,4783	,4788	,4793	,4798	,4803	,4808	,4812	,4817
2,1	,4821	,4826	,4830	,4834	,4838	,4842	,4846	,4850	,4854	,4857
2,2	,4861	,4864	,4868	,4871	,4875	,4878	,4881	,4884	,4887	,4890
2,3	,4893	,3896	,4898	,4901	,4904	,4906	,4909	,4911	,4913	,4916
2,4	,4918	,4920	,4922	,4925	,4927	,4929	,4931	,4932	,4934	,4936
2,5	,4938	,4940	,4941	,4943	,4945	,4946	,4948	,4949	,4951	,4952
2,6	,4953	,4955	,4956	,4957	,4959	,4960	,4961	,4962	,4963	,4964
2,7	,4965	,4966	,4967	,4968	,4969	,4970	,4971	,4972	,4973	,4974
2,8	,4974	,4975	,4976	,4977	,4977	,4978	,4979	,4979	,4980	,4981
2,9	,4981	,4982	,4982	,4983	,4984	,4984	,4985	,4985	,4986	,4986
3,0	,4987	,4987	,4987	,4988	,4988	,4989	,4989	,4989	,4990	,4990

Tiré de P.G. Hoel, *Introduction to Mathematical Statistics*, 4th edition, Wiley, 1971, avec la permission de l'éditeur.

TABLEAU A.2 Répartition du t de Student

La première colonne donne le nombre des degrés de liberté (v). Les autres colonnes correspondent à la probabilité P que t dépasse les valeurs données. Pour des t négatifs, on procède par symétrie.

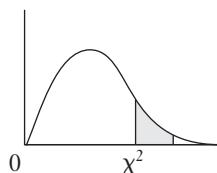


v	P	0,10	0,05	0,025	0,1	0,005
1		3,078	6,314	12,706	31,821	63,657
2		1,886	2,920	4,303	6,965	9,925
3		1,638	2,353	3,182	4,541	5,841
4		1,533	2,132	2,776	3,747	4,604
5		1,476	2,015	2,571	3,365	4,032
6		1,440	1,943	2,447	3,143	3,707
7		1,415	1,895	2,365	2,998	3,499
8		1,397	1,860	2,306	2,896	3,355
9		1,383	1,833	2,262	2,821	3,250
10		1,372	1,812	2,228	2,764	3,169
11		1,363	1,796	2,201	2,718	3,106
12		1,356	1,782	2,179	2,681	3,055
13		1,350	1,771	2,160	2,650	3,012
14		1,345	1,761	2,145	2,624	2,977
15		1,341	1,753	2,131	2,602	2,947
16		1,337	1,746	2,120	2,583	2,921
17		1,333	1,740	2,110	2,567	2,898
18		1,330	1,734	2,101	2,552	2,878
19		1,328	1,729	2,093	2,539	2,861
20		1,325	1,725	2,086	2,528	2,845
21		1,323	1,721	2,080	2,518	2,831
22		1,321	1,717	2,074	2,508	2,819
23		1,319	1,714	2,069	2,500	2,807
24		1,318	1,711	2,064	2,492	2,797
25		1,316	1,708	2,060	2,485	2,787
26		1,315	1,706	2,056	2,479	2,779
27		1,314	1,703	2,052	2,473	2,771
28		1,313	1,701	2,048	2,467	2,763
29		1,311	1,699	2,045	2,462	2,756
30		1,310	1,697	2,042	2,457	2,750
40		1,303	1,684	2,021	2,423	2,704
60		1,296	1,671	2,000	2,390	2,660
120		1,289	1,658	1,980	2,358	2,617
∞		1,282	1,645	1,960	2,326	2,576

Tiré de P.G. Hoel, *Introduction to Mathematical Statistics*, 4th edition, Wiley, 1971, avec la permission de l'éditeur.

TABLEAU A.3 Répartition du χ^2

Pour un nombre de degrés de liberté supérieur à 30, l'expression suit approximativement une loi normale centrée réduite (n étant le nombre de degrés de liberté).



Degrés de liberté	0,99	0,98	0,95	0,90	0,80	0,70
1	0,000157	0,000628	0,00393	0,0158	0,0642	0,148
2	0,0201	0,0404	0,103	0,211	0,446	0,713
3	0,115	0,185	0,352	0,584	1,005	1,424
4	0,297	0,429	0,711	1,064	1,649	2,195
5	0,554	0,752	1,145	1,610	2,343	3,000
6	0,872	1,134	1,635	2,204	3,070	3,828
7	1,239	1,564	2,167	2,833	3,822	4,671
8	1,646	2,032	2,733	3,490	4,594	5,527
9	2,088	2,532	3,325	4,168	5,380	6,393
10	2,558	3,059	3,940	4,865	6,179	7,267
11	3,053	3,609	4,575	5,578	6,989	8,148
12	3,571	4,178	5,226	6,304	7,807	9,034
13	4,107	4,765	5,892	7,042	8,634	9,926
14	4,660	5,368	6,571	7,790	9,467	10,821
15	5,229	5,985	7,261	8,547	10,307	11,721
16	5,812	6,614	7,962	9,312	11,152	12,624
17	6,408	7,255	8,672	10,085	12,002	13,531
18	7,015	7,906	9,390	10,865	12,857	14,440
19	7,633	8,567	10,117	11,651	13,716	15,352
20	8,260	9,237	10,851	12,443	14,578	16,266
21	8,897	9,915	11,591	13,240	15,445	17,182
22	9,542	10,600	12,338	14,041	16,314	18,101
23	10,196	11,293	13,091	14,848	17,187	19,021
24	10,856	11,992	13,848	15,659	18,062	19,943
25	11,524	12,697	14,611	16,473	18,940	20,867
26	12,198	13,409	15,379	17,292	19,820	21,792
27	12,879	14,125	16,151	18,114	20,703	22,719
28	13,565	14,847	16,928	18,939	21,588	23,647
29	14,256	15,574	17,708	19,768	22,475	24,577
30	14,953	16,306	18,493	20,599	23,364	25,508

Tiré de R.A. Fisher, *Statistical Methods for Research Workers*, 14th ed., New York, Macmillan Publishing Co., Inc.

TABLEAU A.3 (suite)

Degrés de liberté	0,50	0,30	0,20	0,10	0,05	0,02	0,01
1	0,455	1,074	1,642	2,706	3,841	5,412	6,635
2	1,386	2,408	3,219	4,605	5,991	7,824	9,210
3	2,366	3,665	4,642	6,251	7,815	9,837	11,341
4	3,357	4,878	5,989	7,779	9,488	11,668	13,277
5	4,351	6,064	7,289	9,236	11,070	13,388	15,086
6	5,348	7,231	8,558	10,645	12,592	15,033	16,812
7	6,346	8,383	9,803	12,017	14,067	16,622	18,475
8	7,344	9,524	11,030	13,362	15,507	18,168	20,090
9	8,343	10,656	12,242	14,684	16,919	19,679	21,666
10	9,342	11,781	13,442	15,987	18,307	21,161	23,209
11	10,341	12,899	14,631	17,275	19,675	22,618	24,725
12	11,340	14,011	15,812	18,549	21,026	24,054	26,217
13	12,340	15,119	16,985	19,812	22,362	25,472	27,688
14	13,339	16,222	18,151	21,064	23,685	26,873	29,141
15	14,339	17,322	19,311	22,307	24,996	28,259	30,578
16	15,338	18,418	20,465	23,542	26,296	29,633	32,000
17	16,338	19,511	21,615	24,769	27,587	30,995	33,409
18	17,338	20,601	22,760	25,989	28,869	32,346	34,805
19	18,338	21,689	23,900	27,204	30,144	33,687	36,191
20	19,337	22,775	25,038	28,412	31,410	35,020	37,566
21	20,337	23,858	26,171	29,615	32,671	36,343	38,932
22	21,337	24,939	27,301	30,813	33,924	37,659	40,289
23	22,337	26,018	28,429	32,007	35,172	38,968	41,638
24	23,337	27,096	29,553	33,196	36,415	40,270	42,980
25	24,337	28,172	30,675	34,382	37,652	41,566	44,314
26	25,336	29,246	31,795	35,563	38,885	42,856	45,642
27	26,336	30,319	32,912	36,741	40,113	44,140	46,963
28	27,336	31,391	34,027	37,916	41,337	45,419	48,278
29	28,336	32,461	35,139	39,087	42,557	46,693	49,588
30	29,336	33,530	36,250	40,256	43,773	47,962	50,892

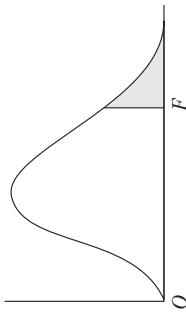


TABLEAU A.4 Répartition du F

Seuil de 5 % (en caractères romans) et de 1 % (italiques).

Degrés de liberté du dénominateur	Degré de liberté du numérateur																								
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
1	161	200	216	225	230	234	237	239	241	243	244	245	246	248	249	250	251	252	253	254	254	254	254	254	254
2	4 052	4 999	5 403	5 625	5 764	5 859	5 928	5 981	6 022	6 056	6 082	6 106	6 142	6 169	6 208	6 234	6 258	6 286	6 302	6 323	6 334	6 352	6 361	6 366	
3	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41	19,42	19,43	19,44	19,45	19,46	19,47	19,47	19,48	19,49	19,49	19,50	19,50	
4	98,49	99,01	99,17	99,25	99,30	99,33	99,34	99,36	99,38	99,40	99,41	99,42	99,43	99,44	99,45	99,46	99,47	99,48	99,48	99,49	99,49	99,50	99,50	99,50	
5	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74	8,71	8,69	8,66	8,64	8,62	8,60	8,58	8,57	8,56	8,54	8,54	8,53	
6	34,12	30,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05	26,92	26,83	26,69	26,60	26,50	26,41	26,30	26,27	26,23	26,18	26,14	26,12	
7	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91	5,87	5,84	5,80	5,77	5,74	5,71	5,70	5,68	5,66	5,65	5,64	5,63	
8	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,45	14,37	14,24	14,15	14,02	13,93	13,83	13,74	13,69	13,61	13,57	13,52	13,48	13,46	
9	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68	4,64	4,60	4,56	4,53	4,50	4,46	4,44	4,42	4,40	4,38	4,37	4,36	
10	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89	9,77	9,68	9,55	9,47	9,38	9,29	9,24	9,17	9,13	9,07	9,04	9,02	
11	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,96	3,92	3,87	3,84	3,81	3,77	3,75	3,72	3,71	3,69	3,68	3,67	
12	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72	7,60	7,52	7,39	7,31	7,23	7,14	7,09	7,02	6,99	6,94	6,90	6,88	
13	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57	3,52	3,49	3,44	3,41	3,38	3,34	3,32	3,29	3,28	3,25	3,24	3,23	
14	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47	6,35	6,27	6,15	6,07	5,98	5,90	5,85	5,78	5,75	5,70	5,67	5,65	
15	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28	3,23	3,20	3,15	3,12	3,08	3,05	3,03	3,00	2,98	2,96	2,94	2,93	
16	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67	5,56	5,48	5,36	5,28	5,20	5,11	5,06	5,00	4,96	4,91	4,88	4,86	
17	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07	3,02	2,98	2,93	2,90	2,86	2,82	2,80	2,77	2,76	2,73	2,72	2,71	
18	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11	5,00	4,92	4,80	4,73	4,64	4,56	4,51	4,45	4,41	4,36	4,33	4,31	
19	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91	2,86	2,82	2,77	2,74	2,70	2,67	2,64	2,61	2,59	2,56	2,55	2,54	
20	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71	4,60	4,52	4,41	4,33	4,25	4,17	4,12	4,05	4,01	3,96	3,93	3,91	

11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79	2,74	2,70	2,65	2,61	2,57	2,53	2,50	2,47	2,45	2,42	2,41	2,40
	9,65	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40	4,29	4,21	4,10	4,02	3,94	3,86	3,80	3,74	3,70	3,66	3,62	3,60
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69	2,64	2,60	2,54	2,50	2,46	2,42	2,40	2,36	2,35	2,32	2,31	2,30
	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16	4,05	3,98	3,86	3,78	3,70	3,61	3,56	3,49	3,46	3,41	3,38	3,36
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60	2,55	2,51	2,46	2,42	2,38	2,34	2,32	2,28	2,26	2,24	2,22	2,21
	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96	3,85	3,78	3,67	3,59	3,51	3,42	3,37	3,30	3,27	3,21	3,18	3,16
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,53	2,48	2,44	2,39	2,35	2,31	2,27	2,24	2,21	2,19	2,16	2,15	2,14	2,14	2,13
	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80	3,70	3,62	3,51	3,43	3,34	3,26	3,21	3,14	3,11	3,06	3,02	3,00
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48	2,43	2,39	2,33	2,29	2,25	2,21	2,18	2,15	2,12	2,12	2,10	2,07
	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67	3,56	3,48	3,36	3,29	3,20	3,12	3,07	3,00	2,97	2,92	2,89	2,87
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42	2,37	2,33	2,28	2,24	2,20	2,16	2,13	2,09	2,07	2,04	2,02	2,01
	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55	3,45	3,37	3,25	3,18	3,10	3,01	2,96	2,89	2,86	2,80	2,77	2,75
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38	2,33	2,29	2,23	2,19	2,15	2,11	2,08	2,04	2,02	1,99	1,97	1,96
	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,45	3,35	3,27	3,16	3,08	3,00	2,92	2,86	2,79	2,76	2,70	2,67	2,65
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,29	2,25	2,19	2,15	2,11	2,07	2,04	2,00	1,98	1,95	1,93	1,92
	8,28	6,01	5,09	4,58	4,25	4,01	3,85	3,71	3,60	3,51	3,44	3,37	3,27	3,19	3,07	3,00	2,91	2,83	2,78	2,71	2,68	2,62	2,59	2,57
19	4,38	3,52	3,13	2,90	2,74	2,63	2,55	2,48	2,43	2,38	2,34	2,31	2,26	2,21	2,15	2,11	2,07	2,02	2,00	1,96	1,94	1,91	1,90	1,88
	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36	3,30	3,19	3,12	3,00	2,92	2,84	2,76	2,70	2,63	2,60	2,54	2,51	2,49
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,35	2,31	2,28	2,23	2,18	2,12	2,08	2,04	1,99	1,96	1,92	1,90	1,87	1,85	1,84
	8,10	5,85	4,94	4,43	4,10	3,87	3,71	3,56	3,45	3,37	3,30	3,23	3,13	3,05	2,94	2,86	2,77	2,69	2,63	2,56	2,53	2,47	2,44	2,42
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25	2,20	2,15	2,09	2,05	2,00	1,96	1,93	1,89	1,87	1,84	1,82	1,81
	8,02	5,78	4,87	4,04	3,81	3,65	3,51	3,40	3,31	3,24	3,17	3,07	2,99	2,88	2,80	2,72	2,63	2,58	2,51	2,47	2,42	2,38	2,36	2,36
22	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40	2,35	2,30	2,26	2,23	2,18	2,13	2,07	2,03	1,98	1,93	1,91	1,87	1,84	1,81	1,80	1,78
	7,94	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,18	3,12	3,02	2,94	2,83	2,75	2,67	2,58	2,53	2,46	2,42	2,37	2,33	2,31
23	4,28	3,42	3,03	2,80	2,64	2,53	2,45	2,38	2,32	2,28	2,24	2,20	2,14	2,10	2,04	2,00	1,96	1,91	1,88	1,84	1,82	1,79	1,77	1,76
	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,14	3,07	2,97	2,89	2,78	2,70	2,62	2,53	2,48	2,41	2,37	2,32	2,28	2,26
24	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	2,30	2,26	2,22	2,18	2,13	2,09	2,02	1,98	1,94	1,89	1,86	1,82	1,80	1,76	1,74	1,73
	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,25	3,17	3,09	3,03	2,93	2,85	2,74	2,66	2,58	2,49	2,44	2,36	2,33	2,27	2,23	2,21
25	4,24	3,38	2,99	2,76	2,60	2,49	2,41	2,34	2,28	2,24	2,20	2,16	2,11	2,06	2,00	1,96	1,92	1,87	1,84	1,80	1,77	1,74	1,72	1,71
	7,77	5,57	4,68	4,18	3,86	3,63	3,46	3,32	3,21	3,13	3,05	2,99	2,89	2,81	2,70	2,62	2,54	2,45	2,40	2,32	2,29	2,23	2,19	2,17

TABLEAU A.4 (suite)

Degrés de liberté du dénominateur	Degré de liberté du numérateur																							
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
26	4,22	3,37	2,89	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	2,10	2,05	1,99	1,95	1,90	1,85	1,82	1,78	1,76	1,72	1,70	1,69
27	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,17	3,09	3,02	2,96	2,86	2,77	2,66	2,58	2,50	2,41	2,36	2,28	2,25	2,19	2,15	2,13
28	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,30	2,25	2,20	2,16	2,13	2,08	2,03	1,97	1,93	1,88	1,84	1,80	1,76	1,74	1,71	1,68	1,67
29	7,68	5,49	4,60	4,11	3,79	3,56	3,39	3,26	3,14	3,06	2,98	2,93	2,83	2,74	2,63	2,55	2,47	2,38	2,33	2,25	2,21	2,16	2,12	2,10
30	4,20	3,34	2,95	2,71	2,56	2,44	2,36	2,29	2,24	2,19	2,15	2,12	2,06	2,02	1,96	1,91	1,87	1,81	1,78	1,75	1,72	1,69	1,67	1,65
31	7,64	5,45	4,57	4,07	3,76	3,53	3,36	3,23	3,11	3,03	2,95	2,90	2,80	2,71	2,60	2,52	2,44	2,35	2,30	2,22	2,18	2,13	2,09	2,06
32	4,18	3,33	2,93	2,70	2,54	2,43	2,35	2,28	2,22	2,18	2,14	2,10	2,05	2,00	1,94	1,90	1,85	1,80	1,77	1,73	1,71	1,68	1,66	1,64
33	7,60	5,52	4,54	4,04	3,73	3,50	3,33	3,20	3,08	3,00	2,92	2,87	2,77	2,68	2,57	2,49	2,41	2,32	2,27	2,19	2,15	2,10	2,06	2,03
34	4,17	3,32	2,92	2,69	2,53	2,42	2,34	2,27	2,21	2,16	2,12	2,09	2,04	1,99	1,93	1,89	1,84	1,79	1,76	1,72	1,69	1,66	1,64	1,62
35	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,06	2,98	2,90	2,84	2,74	2,66	2,55	2,47	2,38	2,29	2,24	2,16	2,13	2,07	2,03	2,01
36	4,15	3,30	2,90	2,67	2,51	2,40	2,32	2,25	2,19	2,14	2,10	2,07	2,02	1,97	1,91	1,86	1,82	1,76	1,74	1,69	1,67	1,64	1,61	1,59
37	7,50	5,34	4,46	3,97	3,66	3,42	3,25	3,12	3,01	2,94	2,86	2,80	2,70	2,62	2,51	2,42	2,34	2,25	2,20	2,12	2,08	2,02	1,98	1,96
38	4,13	3,28	2,88	2,65	2,49	2,38	2,30	2,23	2,17	2,12	2,08	2,05	2,00	1,95	1,89	1,84	1,80	1,74	1,71	1,67	1,64	1,61	1,59	1,57
39	7,44	5,29	4,42	3,93	3,61	3,38	3,21	3,08	2,97	2,89	2,82	2,76	2,66	2,58	2,47	2,38	2,30	2,21	2,15	2,08	2,04	1,98	1,94	1,91
40	4,11	3,26	2,86	2,63	2,48	2,36	2,28	2,21	2,15	2,10	2,06	2,03	1,98	1,93	1,87	1,82	1,78	1,72	1,69	1,65	1,62	1,59	1,56	1,55
41	7,39	5,25	4,38	3,89	3,58	3,35	3,18	3,04	2,94	2,86	2,78	2,72	2,62	2,54	2,43	2,35	2,26	2,17	2,12	2,04	2,00	1,94	1,90	1,87
42	4,10	3,25	2,85	2,62	2,46	2,35	2,26	2,19	2,14	2,09	2,05	2,02	1,96	1,92	1,85	1,80	1,76	1,71	1,67	1,63	1,60	1,57	1,54	1,53
43	7,35	5,21	4,34	3,86	3,54	3,32	3,15	3,02	2,91	2,82	2,75	2,69	2,59	2,51	2,40	2,32	2,22	2,14	2,08	2,00	1,97	1,90	1,86	1,84
44	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,07	2,04	2,00	1,95	1,90	1,84	1,79	1,74	1,69	1,66	1,61	1,59	1,55	1,53	1,51
45	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,88	2,80	2,73	2,66	2,56	2,49	2,37	2,29	2,20	2,11	2,05	1,97	1,94	1,88	1,84	1,81
46	4,07	3,22	2,83	2,59	2,44	2,32	2,24	2,17	2,11	2,06	2,02	1,99	1,94	1,89	1,82	1,78	1,73	1,68	1,64	1,60	1,57	1,54	1,51	1,49
47	7,27	5,15	4,29	3,80	3,49	3,26	3,10	2,96	2,86	2,77	2,70	2,64	2,54	2,46	2,35	2,26	2,17	2,08	2,02	1,94	1,91	1,85	1,80	1,78
48	4,06	3,21	2,82	2,58	2,43	2,31	2,23	2,16	2,10	2,05	2,01	1,98	1,92	1,88	1,81	1,76	1,72	1,66	1,63	1,58	1,56	1,52	1,50	1,48
49	7,24	5,12	4,26	3,78	3,46	3,24	3,07	2,94	2,84	2,75	2,68	2,62	2,52	2,44	2,32	2,24	2,15	2,06	2,00	1,92	1,88	1,82	1,78	1,75

46	4,05 7,21	3,20 5,10	2,81 4,24	2,57 3,76	2,42 3,44	2,30 3,22	2,22 3,05	2,14 2,92	2,09 2,82	2,04 2,73	2,00 2,66	1,97 2,60	1,91 2,50	1,87 2,42	1,80 2,40	1,75 2,22	1,71 2,13	1,65 2,04	1,62 1,98	1,57 1,90	1,54 1,86	1,51 1,80	1,48 1,76	1,46 1,72	
48	4,04 7,19	3,19 5,08	2,80 4,22	2,56 3,74	2,41 3,42	2,30 3,20	2,21 3,04	2,14 2,90	2,08 2,80	2,03 2,71	1,99 2,64	1,96 2,58	1,90 2,48	1,86 2,40	1,79 2,28	1,74 2,20	1,70 2,11	1,64 2,02	1,61 1,96	1,56 1,88	1,53 1,84	1,50 1,78	1,47 1,73	1,45 1,70	
50	4,03 7,17	3,18 5,06	2,79 4,20	2,56 3,72	2,40 3,41	2,29 3,18	2,20 3,02	2,13 2,88	2,07 2,78	2,02 2,70	1,98 2,62	1,95 2,56	1,90 2,46	1,85 2,39	1,78 2,26	1,74 2,18	1,69 2,10	1,63 2,00	1,60 1,94	1,55 1,82	1,52 1,80	1,48 1,76	1,46 1,71	1,44 1,68	
55	4,02 7,12	3,17 5,01	2,78 4,16	2,54 3,68	2,38 3,37	2,27 3,15	2,18 2,98	2,11 2,85	2,05 2,75	2,00 2,66	1,97 2,59	1,93 2,53	1,88 2,43	1,83 2,35	1,76 2,23	1,72 2,15	1,67 2,06	1,61 1,96	1,58 1,90	1,52 1,82	1,50 1,78	1,46 1,71	1,43 1,66	1,41 1,64	
60	4,00 7,08	3,15 4,98	2,76 4,13	2,52 3,65	2,37 3,34	2,25 3,12	2,17 2,95	2,10 2,82	2,04 2,72	1,99 2,63	1,95 2,56	1,92 2,50	1,86 2,40	1,81 2,32	1,75 2,20	1,70 2,12	1,65 2,03	1,59 1,93	1,56 1,87	1,50 1,79	1,48 1,74	1,44 1,68	1,41 1,63	1,39 1,60	
65	3,99 7,04	3,14 4,95	2,75 4,10	2,51 3,62	2,36 3,31	2,24 3,09	2,15 2,93	2,08 2,79	2,02 2,70	1,98 2,61	1,94 2,54	1,90 2,47	1,85 2,30	1,80 2,27	1,73 2,18	1,68 2,09	1,63 2,00	1,57 1,90	1,54 1,84	1,49 1,76	1,46 1,71	1,42 1,64	1,39 1,60	1,37 1,56	
70	3,98 7,01	3,13 4,92	2,74 4,08	2,50 3,60	2,35 3,29	2,22 3,07	2,14 2,91	2,07 2,77	2,01 2,67	1,97 2,59	1,93 2,51	1,89 2,45	1,84 2,35	1,79 2,28	1,72 2,15	1,67 2,07	1,62 2,00	1,56 1,88	1,53 1,82	1,47 1,74	1,45 1,69	1,40 1,63	1,37 1,56	1,35 1,53	
80	3,96 6,96	3,11 4,88	2,72 4,04	2,48 3,56	2,33 3,25	2,21 3,04	2,12 2,87	2,05 2,74	1,99 2,64	1,95 2,55	1,91 2,48	1,88 2,41	1,82 2,32	1,77 2,24	1,70 2,11	1,65 2,03	1,60 1,94	1,54 1,84	1,51 1,78	1,45 1,70	1,42 1,65	1,38 1,57	1,35 1,52	1,32 1,49	
100	3,94 6,90	3,09 4,82	2,70 3,98	2,46 3,51	2,30 3,20	2,19 2,99	2,10 2,82	2,03 2,69	1,97 2,59	1,92 2,51	1,88 2,43	1,85 2,36	1,79 2,26	1,75 2,19	1,68 2,06	1,63 1,98	1,57 1,73	1,51 1,79	1,48 1,73	1,42 1,64	1,39 1,59	1,34 1,51	1,30 1,46	1,28 1,43	
125	3,92 6,84	3,07 4,78	2,68 3,94	2,44 3,47	2,29 3,17	2,17 2,95	2,08 2,79	2,01 2,65	1,95 2,56	1,90 2,47	1,86 2,40	1,83 2,33	1,77 2,23	1,72 2,15	1,65 2,03	1,60 1,94	1,55 1,85	1,49 1,75	1,45 1,68	1,39 1,59	1,36 1,54	1,31 1,46	1,27 1,40	1,25 1,37	
150	3,91 6,81	3,06 4,75	2,67 3,91	2,43 3,44	2,27 3,13	2,16 2,92	2,07 2,76	2,00 2,62	1,94 2,53	1,89 2,44	1,85 2,37	1,82 2,30	1,76 2,20	1,71 2,12	1,64 2,00	1,59 1,91	1,54 1,83	1,47 1,72	1,44 1,66	1,37 1,56	1,34 1,51	1,29 1,43	1,25 1,37	1,22 1,33	
200	3,89 6,76	3,04 4,71	2,65 3,88	2,41 3,41	2,26 3,11	2,14 2,90	2,05 2,73	1,98 2,60	1,92 2,50	1,87 2,41	1,83 2,34	1,80 2,28	1,74 2,17	1,69 2,09	1,62 1,97	1,57 1,88	1,45 1,79	1,45 1,69	1,42 1,62	1,35 1,53	1,32 1,48	1,26 1,39	1,22 1,33	1,19 1,28	
400	3,86 6,70	3,02 4,66	2,62 3,83	2,39 3,36	2,23 3,06	2,12 2,85	2,03 2,69	1,96 2,55	1,90 2,46	1,85 2,37	1,81 2,29	1,78 2,23	1,72 2,12	1,67 2,04	1,60 1,92	1,54 1,84	1,49 1,74	1,42 1,64	1,38 1,57	1,32 1,47	1,28 1,42	1,22 1,32	1,16 1,24	1,13 1,19	
1000	3,85 6,66	3,00 4,62	2,61 3,80	2,38 3,34	2,22 3,04	2,10 2,82	2,02 2,66	1,95 2,53	1,89 2,43	1,84 2,34	1,80 2,26	1,76 2,20	1,70 2,09	1,65 2,01	1,58 1,89	1,53 1,81	1,47 1,71	1,41 1,61	1,36 1,54	1,30 1,44	1,26 1,38	1,19 1,28	1,13 1,19	1,08 1,11	
∞	3,84 6,64	2,99 4,60	2,60 3,78	2,37 3,32	2,21 3,02	2,09 2,80	2,01 2,64	1,94 2,51	1,88 2,41	1,83 2,32	1,79 2,24	1,75 2,18	1,69 2,07	1,64 1,99	1,57 1,87	1,52 1,79	1,46 1,69	1,40 1,59	1,35 1,52	1,28 1,41	1,24 1,36	1,17 1,25	1,11 1,15	1,00 1,00	

TABLEAU A.5 Statistique de Durbin et Watson au seuil de 5 %

<i>n</i>	K' = 1		K' = 2		K' = 3		K' = 4		K' = 5	
	d _L	d _V								
6	0,610	1,400	–	–	–	–	–	–	–	–
7	0,700	1,356	0,467	1,896	–	–	–	–	–	–
8	0,763	1,332	0,559	1,777	0,368	2,287	–	–	–	–
9	0,824	1,320	0,629	1,699	0,455	2,128	0,296	2,588	–	–
10	0,879	1,320	0,697	1,641	0,525	2,016	0,376	2,414	0,243	2,822
11	0,927	1,324	0,758	1,604	0,595	1,928	0,444	2,283	0,316	2,645
12	0,971	1,331	0,812	1,579	0,658	1,864	0,512	2,177	0,379	2,506
13	1,010	1,340	0,861	1,562	0,715	1,816	0,574	2,094	0,445	2,390
14	1,045	1,350	0,905	1,551	0,767	1,779	0,632	2,030	0,505	2,296
15	1,077	1,361	0,946	1,543	0,814	1,750	0,685	1,977	0,562	2,220
16	1,106	1,371	0,982	1,539	0,857	1,728	0,734	1,935	0,615	2,157
17	1,133	1,381	1,015	1,536	0,897	1,710	0,779	1,900	0,664	2,104
18	1,158	1,391	1,046	1,535	0,933	1,696	0,820	1,872	0,710	2,060
19	1,180	1,401	1,074	1,536	0,967	1,685	0,859	1,848	0,752	2,023
20	1,201	1,411	1,100	1,537	0,998	1,676	0,894	1,828	0,792	1,991
21	1,221	1,420	1,125	1,538	1,026	1,669	0,927	1,812	0,829	1,964
22	1,239	1,429	1,147	1,541	1,053	1,664	0,958	1,797	0,863	1,940
23	1,257	1,437	1,168	1,543	1,078	1,660	0,986	1,785	0,895	1,920
24	1,273	1,446	1,188	1,546	1,101	1,656	1,013	1,775	0,925	1,902
25	1,288	1,454	1,206	1,550	1,123	1,654	1,038	1,767	0,953	1,886
26	1,302	1,461	1,224	1,553	1,143	1,652	1,062	1,759	0,979	1,873
27	1,316	1,469	1,240	1,556	1,162	1,651	1,084	1,753	1,004	1,861
28	1,328	1,476	1,255	1,560	1,181	1,650	1,104	1,747	1,028	1,850
29	1,341	1,483	1,270	1,563	1,198	1,650	1,124	1,743	1,050	1,841
30	1,352	1,489	1,284	1,567	1,214	1,650	1,143	1,739	1,071	1,833
31	1,363	1,496	1,297	1,570	1,229	1,650	1,160	1,735	1,090	1,825
32	1,373	1,502	1,309	1,574	1,244	1,650	1,177	1,732	1,109	1,819
33	1,383	1,508	1,321	1,577	1,258	1,651	1,193	1,730	1,127	1,813
34	1,393	1,514	1,333	1,580	1,271	1,652	1,208	1,728	1,144	1,808
35	1,402	1,519	1,343	1,584	1,283	1,653	1,222	1,726	1,160	1,803
36	1,411	1,525	1,354	1,587	1,295	1,654	1,236	1,724	1,175	1,799
37	1,419	1,530	1,364	1,590	1,307	1,655	1,249	1,723	1,190	1,795
38	1,427	1,535	1,373	1,594	1,318	1,656	1,261	1,722	1,204	1,792
39	1,435	1,540	1,382	1,597	1,328	1,658	1,273	1,722	1,218	1,789
40	1,442	1,544	1,391	1,600	1,338	1,659	1,285	1,721	1,230	1,786
45	1,475	1,566	1,430	1,615	1,383	1,666	1,336	1,720	1,287	1,776
50	1,503	1,585	1,462	1,628	1,421	1,674	1,378	1,721	1,335	1,771
55	1,528	1,601	1,490	1,641	1,452	1,681	1,414	1,724	1,374	1,768
60	1,549	1,616	1,514	1,652	1,480	1,689	1,444	1,727	1,408	1,767
65	1,567	1,629	1,536	1,662	1,503	1,696	1,471	1,731	1,438	1,767
70	1,583	1,641	1,554	1,672	1,525	1,703	1,494	1,735	1,464	1,768
75	1,598	1,652	1,571	1,680	1,543	1,709	1,515	1,739	1,487	1,770
80	1,611	1,662	1,586	1,688	1,560	1,715	1,534	1,743	1,507	1,772
85	1,624	1,671	1,600	1,696	1,575	1,721	1,550	1,747	1,525	1,774
90	1,635	1,679	1,612	1,703	1,589	1,726	1,566	1,751	1,542	1,776
95	1,645	1,687	1,623	1,709	1,602	1,732	1,579	1,755	1,557	1,778
100	1,654	1,694	1,634	1,715	1,613	1,736	1,592	1,758	1,571	1,780
150	1,720	1,746	1,706	1,760	1,693	1,774	1,679	1,788	1,665	1,802
200	1,758	1,778	1,748	1,789	1,738	1,799	1,728	1,810	1,718	1,820

= nombre de régresseurs (hors constante).

Tiré de *Econometrica*, vol. 45, n° 1977, 1992-1995.

TABLEAU A.5 (suite)

<i>n</i>	K' = 6		K' = 7		K' = 8		K' = 9		K' = 10	
	<i>d</i> _L	<i>d</i> _V								
6	-	-	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-	-	-
8	-	-	-	-	-	-	-	-	-	-
9	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-
11	0,203	3,005	-	-	-	-	-	-	-	-
12	0,268	2,832	0,171	3,149	-	-	-	-	-	-
13	0,328	2,692	0,230	2,985	0,147	3,266	-	-	-	-
14	0,389	2,572	0,286	2,848	0,200	3,111	0,127	3,360	-	-
15	0,447	2,472	0,343	2,727	0,251	2,979	0,175	3,216	0,111	3,438
16	0,502	2,388	0,398	2,624	0,304	2,860	0,222	3,090	0,155	3,304
17	0,554	2,318	0,451	2,537	0,356	2,757	0,272	2,975	0,198	3,184
18	0,603	2,257	0,502	2,461	0,407	2,667	0,321	2,873	0,244	3,073
19	0,649	2,206	0,459	2,396	0,456	2,589	0,369	2,783	0,290	2,974
20	0,692	2,162	0,595	2,339	0,502	2,521	0,416	2,704	0,336	2,885
21	0,732	2,124	0,637	2,290	0,547	2,460	0,461	2,633	0,380	2,806
22	0,769	2,090	0,677	2,246	0,588	2,407	0,504	2,571	0,424	2,734
23	0,804	2,061	0,715	2,208	0,628	2,360	0,545	2,514	0,465	2,670
24	0,837	2,035	0,751	2,174	0,666	2,318	0,584	2,464	0,506	2,613
25	0,868	2,012	0,784	2,144	0,702	2,280	0,621	2,419	0,544	2,560
26	0,897	1,992	0,816	2,117	0,735	2,246	0,657	2,379	0,581	2,513
27	0,925	1,974	0,845	2,093	0,767	2,216	0,691	2,342	0,616	2,470
28	0,951	1,958	0,874	2,071	0,798	2,188	0,723	2,309	0,650	2,431
29	0,975	1,944	0,900	2,052	0,826	2,164	0,753	2,278	0,682	2,396
30	0,998	1,931	0,926	2,034	0,854	2,141	0,782	2,251	0,712	2,363
31	1,020	1,920	0,950	2,018	0,879	2,120	0,810	2,226	0,741	2,333
32	1,041	1,909	0,972	2,004	0,904	2,102	0,836	2,203	0,769	2,306
33	1,061	1,900	0,994	1,991	0,927	2,085	0,861	2,181	0,795	2,281
34	1,080	1,891	1,015	1,979	0,950	2,069	0,885	2,162	0,821	2,257
35	1,097	1,884	1,034	1,967	0,971	2,054	0,908	2,144	0,845	2,236
36	1,114	1,877	1,053	1,957	0,991	2,041	0,930	2,127	0,868	2,216
37	1,131	1,870	1,071	1,948	1,011	2,029	0,951	2,112	0,891	2,198
38	1,146	1,864	1,088	1,939	1,029	2,017	0,970	2,098	0,912	2,180
39	1,161	1,859	1,104	1,932	1,047	2,007	0,990	2,085	0,932	2,164
40	1,175	1,854	1,120	1,924	1,064	1,997	1,008	2,072	0,945	2,149
45	1,238	1,835	1,189	1,895	1,139	1,958	1,089	2,002	1,038	2,088
50	1,291	1,822	1,246	1,875	1,201	1,930	1,156	1,986	1,110	2,044
55	1,334	1,814	1,294	1,861	1,253	1,909	1,212	1,959	1,170	2,010
60	1,372	1,808	1,335	1,850	1,298	1,894	1,260	1,939	1,222	1,984
65	1,404	1,805	1,370	1,843	1,336	1,882	1,301	1,923	1,266	1,964
70	1,433	1,802	1,401	1,837	1,369	1,873	1,337	1,910	1,305	1,948
75	1,458	1,801	1,428	1,834	1,399	1,867	1,369	1,901	1,339	1,935
80	1,480	1,801	1,453	1,831	1,425	1,861	1,397	1,893	1,369	1,925
85	1,500	1,801	1,474	1,829	1,448	1,857	1,422	1,886	1,396	1,916
90	1,518	1,801	1,494	1,827	1,469	1,854	1,445	1,881	1,420	1,909
95	1,535	1,802	1,512	1,827	1,489	1,852	1,465	1,877	1,442	1,903
100	1,550	1,803	1,528	1,826	1,506	1,850	1,484	1,874	1,462	1,898
150	1,651	1,817	1,637	1,832	1,622	1,847	1,608	1,862	1,594	1,877
200	1,707	1,831	1,697	1,841	1,686	1,852	1,675	1,863	1,665	1,874

TABLEAU A.5 (suite)

<i>n</i>	<i>K'</i> = 11		<i>K'</i> = 12		<i>K'</i> = 13		<i>K'</i> = 14		<i>K'</i> = 15	
	<i>d_L</i>	<i>d_V</i>								
16	0,098	3,503	–	–	–	–	–	–	–	–
17	0,138	3,378	0,087	3,557	–	–	–	–	–	–
18	0,177	3,265	0,123	3,441	0,078	3,603	–	–	–	–
19	0,220	3,159	0,160	3,335	0,111	3,496	0,070	3,642	–	–
20	0,263	3,063	0,200	3,234	0,145	3,395	0,100	3,542	0,063	3,676
21	0,307	2,976	0,240	3,141	0,182	3,300	0,132	3,448	0,091	3,583
22	0,349	2,897	0,281	3,057	0,220	3,211	0,166	3,358	0,120	3,495
23	0,391	2,826	0,322	2,979	0,259	3,128	0,202	3,272	0,153	3,409
24	0,431	2,761	0,362	2,908	0,297	3,053	0,239	3,193	0,186	3,327
25	0,470	2,702	0,400	2,844	0,335	2,983	0,275	3,119	0,221	3,251
26	0,508	2,649	0,438	2,784	0,373	2,919	0,312	3,051	0,256	3,179
27	0,544	2,600	0,475	2,730	0,409	2,859	0,348	2,987	0,291	3,112
28	0,578	2,555	0,510	2,680	0,445	2,805	0,383	2,928	0,325	3,050
29	0,612	2,515	0,544	2,634	0,479	2,755	0,418	2,874	0,359	2,992
30	0,643	2,477	0,577	2,592	0,512	2,708	0,451	2,823	0,392	2,937
31	0,674	2,443	0,608	2,553	0,545	2,665	0,484	2,776	0,425	2,887
32	0,703	2,411	0,638	2,517	0,576	2,625	0,515	2,733	0,457	2,840
33	0,731	2,382	0,668	2,484	0,606	2,588	0,546	2,692	0,488	2,796
34	0,758	2,355	0,695	2,454	0,634	2,554	0,575	2,654	0,518	2,754
35	0,783	2,330	0,722	2,425	0,662	2,521	0,604	2,619	0,547	2,716
36	0,808	2,306	0,748	2,398	0,689	2,492	0,631	2,586	0,575	2,680
37	0,831	2,285	0,772	2,374	0,714	2,464	0,657	2,555	0,602	2,646
38	0,854	2,265	0,796	2,351	0,739	2,438	0,683	2,526	0,628	2,614
39	0,875	2,246	0,819	2,329	0,763	2,413	0,707	2,499	0,653	2,585
40	0,896	2,228	0,840	2,309	0,785	2,391	0,731	2,473	0,678	2,557
45	0,988	2,156	0,938	2,225	0,887	2,296	0,838	2,367	0,788	2,439
50	1,064	2,103	1,019	2,163	0,973	2,225	0,927	2,287	0,882	2,350
55	1,129	2,062	1,087	2,116	1,045	2,170	1,003	2,225	0,961	2,281
60	1,184	2,031	1,145	2,079	1,106	2,127	1,068	2,177	1,029	2,227
65	1,231	2,006	1,195	2,049	1,160	2,093	1,124	2,138	1,088	2,183
70	1,272	1,986	1,239	2,026	1,206	2,066	1,172	2,106	1,139	2,148
75	1,308	1,970	1,277	2,006	1,247	2,043	1,215	2,080	1,184	2,118
80	1,340	1,957	1,311	1,991	1,283	2,024	1,253	2,059	1,224	2,093
85	1,369	1,946	1,342	1,977	1,315	2,009	1,287	2,040	1,260	2,073
90	1,395	1,937	1,369	1,966	1,344	1,995	1,318	2,025	1,292	2,055
95	1,418	1,929	1,394	1,956	1,370	1,984	1,345	2,012	1,321	2,040
100	1,434	1,923	1,416	1,948	1,393	1,974	1,371	2,000	1,347	2,026
150	1,579	1,892	1,564	1,908	1,550	1,924	1,535	1,940	1,519	1,956
200	1,654	1,885	1,643	1,896	1,632	1,908	1,621	1,919	1,610	1,931

TABLEAU A.5 (suite)

<i>n</i>	<i>K'</i> = 16		<i>K'</i> = 17		<i>K'</i> = 18		<i>K'</i> = 19		<i>K'</i> = 20	
	<i>d_L</i>	<i>d_V</i>								
16	-	-	-	-	-	-	-	-	-	-
17	-	-	-	-	-	-	-	-	-	-
18	-	-	-	-	-	-	-	-	-	-
19	-	-	-	-	-	-	-	-	-	-
20	-	-	-	-	-	-	-	-	-	-
21	0,058	3,705	-	-	-	-	-	-	-	-
22	0,083	3,619	0,052	3,731	-	-	-	-	-	-
23	0,110	3,535	0,076	3,650	0,048	3,753	-	-	-	-
24	0,141	3,454	0,101	3,572	0,070	3,678	0,044	3,773	-	-
25	0,172	3,376	0,130	3,494	0,094	3,604	0,065	3,702	0,041	3,790
26	0,205	3,303	0,160	3,420	0,120	3,531	0,087	3,632	0,060	3,724
27	0,238	3,233	0,191	3,349	0,149	3,460	0,112	3,563	0,081	3,658
28	0,271	3,168	0,222	3,283	0,178	3,392	0,138	3,495	0,104	3,592
29	0,305	3,107	0,254	3,219	0,208	3,327	0,166	3,431	0,129	3,528
30	0,337	3,050	0,286	3,160	0,238	3,266	0,195	3,368	0,156	3,465
31	0,370	2,996	0,317	3,103	0,269	3,208	0,224	3,309	0,183	3,406
32	0,401	2,946	0,349	3,050	0,299	3,153	0,253	3,252	0,211	3,348
33	0,432	2,899	0,379	3,000	0,329	3,100	0,283	3,198	0,239	3,293
34	0,462	2,854	0,409	2,954	0,359	3,051	0,312	3,147	0,267	3,240
35	0,492	2,813	0,439	2,910	0,388	3,005	0,340	3,099	0,295	3,190
36	0,520	2,774	0,467	2,868	0,417	2,961	0,369	3,053	0,323	3,142
37	0,548	2,738	0,495	2,829	0,445	2,920	0,397	3,009	0,351	3,097
38	0,575	2,703	0,522	2,792	0,472	2,880	0,424	2,968	0,378	3,054
39	0,600	2,671	0,549	2,757	0,499	2,843	0,451	2,929	0,404	3,013
40	0,626	2,641	0,575	2,724	0,525	2,808	0,477	2,892	0,430	2,974
45	0,740	2,512	0,692	2,586	0,644	2,659	0,598	2,733	0,553	2,807
50	0,836	2,414	0,792	2,479	0,747	2,544	0,703	2,610	0,660	2,675
55	0,919	2,338	0,877	2,396	0,836	2,454	0,795	2,512	0,754	2,571
60	0,990	2,278	0,951	2,330	0,913	2,382	0,874	2,434	0,836	2,487
65	1,052	2,229	1,016	2,276	0,980	2,323	0,944	2,371	0,908	2,419
70	1,105	2,189	1,072	2,232	1,038	2,275	1,005	2,318	0,971	2,362
75	1,153	2,156	1,121	2,195	1,090	2,235	1,058	2,275	1,027	2,315
80	1,195	2,129	1,165	2,165	1,136	2,201	1,106	2,238	1,076	2,275
85	1,232	2,105	1,205	2,139	1,177	2,172	1,149	2,206	1,121	2,241
90	1,266	2,085	1,240	2,116	1,213	2,148	1,187	2,179	1,160	2,211
95	1,296	2,068	1,271	2,097	1,247	2,126	1,222	2,156	1,197	2,186
100	1,324	2,053	1,301	2,080	1,277	2,108	1,253	2,135	1,229	2,164
150	1,504	1,972	1,489	1,989	1,474	2,006	1,458	2,023	1,443	2,040
200	1,599	1,943	1,588	1,955	1,576	1,967	1,565	1,979	1,554	1,991

INDEX

A

acceptation bancaire, 90, 269
algorithme BHHH, 305
algorithme
 d'estimation non linéaire, 256
 d'optimisation, 149
 de Newton, 150
 de Newton-Raphson, 149
 itératif Cochrane-Orcut, 222
analyse factorielle, 242
approximation
 de Taylor, 151, 332
 parcimonieuse, 282
APT (Arbitrage Pricing Theory), 228
arbitrage, 228
arbre binomial, 24
ARCH-M, 95
autocorrélation, 215
 en présence d'une variable
 dépendante retardée, 225
 partielle, 234
autocovariance, 216, 230
autorégression, 263

B

BAX, 90, 269
benchmark, 138
bêta, 92, 293
biais, 9
 de simultanéité, 183

bootstrap, 193
borne Cramer-Rao, 104
bruit blanc, 183, 216

C

calcul stochastique, 187
CAPM, 91, 164, 234, 287
caps, 159
carrés résiduels contraints, 113
classe des estimateurs linéaires, 102
CML, 288
coefficient
 d'autocorrélation, 218
 de corrélation de Pearson, 83
 de détermination, 82
 de Kendall, 37
 de kurtosis, 276
 de retour vers la moyenne, 342
 de variance, 342
 R^2 ajusté, 83
 R^2 ajusté de Theil, 103
cofacteur, 124
cointégration, 265, 268
combinaisons rendement-risque
 optimales, 135
conditions
 de moments, 328
 d'orthogonalité, 329, 332, 348
contraintes linéaires sur les
 paramètres, 108

- contrat
à terme, 90, 272
au comptant, 272
- convergence, 169
en distribution, 171
en moyenne quadratique, 170
en probabilité, 170
quasi sûre, 171
- corrélation
de Pearson, 31
de Spearman, 35
non paramétrique, 196
spatiale, 215
- couche cachée, 255
- coupe instantanée, 242, 275
- courbe de Phillips, 87
- couverture d'un portefeuille de titres, 270
- couverture
naïve, 133
optimale d'un bilan bancaire, 129
- covariance, 29
entre les rendements de deux portefeuilles, 126
- critère
d'Akaike, 246
d'information d'Akaike, 156
de sélection de Schwarz, 246
d'information de Schwartz, 156
- D**
- data-fitting*, 229
- days of the week effect (DOW), 161
- degrés de liberté, 62
- demande de contrats au comptant, 112
- demande inverse*, 183
- densité
jointe, 38
marginale, 38
- densités conditionnelles, 40
- dérivation matricielle, 99, 131
- déterminant, 121
- detrending*, 261
- Dickey-Fuller *augmenté*, 265
- différenciation, 261
- discrétisation
d'Eule, 340
d'Euler, 188, 342, 345
d'une équation différentielle stochastique, 335
de Milstein, 192
de premier ordre, 340
d'un processus stochastique, 338
exacte, 340
- distance euclidienne, 45
- distribution
 χ^2 , 62
de Poisson, 25
F, 67
lognormale, 145
uniforme, 33, 195
- diversification de portefeuille, 294
- doubles moindres carrés non linéaires, 331
- doubles moindres carrés, 179
- drift*, 191, 335
- écart quadratique moyen, 46
- écart-type, 10
- effets asymétriques, 283
- efficience des marchés financiers, 263
- efficients, 75
- E**
- élasticité-prix d'une option, 86
- endogénéité, 179, 182
- ensemble de Borel, 6
- ensemble orthonormal, 235
- équation
caractéristique, 181
de différence, 260
différentielle stochastique discrétisée, 185
normales, 72, 101
- équilibre des marchés financiers, 292
- erreur
de prévision., 82
de suivi, 129, 138
de type I, 66
de type II, 66

- erreurs
 de prévision groupées
 (*clustered*), 276
 de spécification, 143
espace des états de la nature, 229
espace fondamental, 229
 espérance, 29
 conditionnelle et prévision à court
 terme, 277
 conditionnelle, 259, 276
 du rendement du titre i , 241
 mathématique, 20
 non conditionnelle et prévision à
 long terme, 277
 non conditionnelle, 259
 non conditionnelle, 274, 277
 risque-neutre, 192
- estimateur
 BLUE, 46, 102, 219
 convergent, 183
 de distance minimale, 327
 des doubles moindres carrés, 331
 des MCO, 71
 des moindres carrés indirects, 330
 kernel du bêta, 199
 super-convergent, 268
 semi-paramétriques, 334
- EViews, 15
 excès de leptokurticité, 283
 expansion de Taylor, 188
- F**
- facteurs, 236
 FGLS, 224
 filtre linéaire, 237
 finance corporative, 335
- fonction
 d'activation logistique, 256
 d'autocorrélation partielle, 234
 d'autocorrélation, 232
 d'autocovariance, 232
 de lagrange, 136, 139
 de répartition, 19
 de transition, 254, 257
 de vraisemblance, 50
 logistique, 255, 256
 scédastique, 70
- forme
 linéaire d'hétéroscédasticité, 206
 quadratique, 99, 127
- formes fonctionnelles, 83
 frontière efficiente, 129, 135
- G-H**
- GARCH(1,1), 282
 GARCH, 257
 GARCH-M multivarié, 304
 générateur de variables aléatoires, 193
 GMM, 327
 gradient, 153
 hétéroscédasticité, 201
 homoscédasticité, 70, 98, 203, 275
- I**
- impact des jours de la semaine sur les
 indices boursiers, 227
 indépendance, 38
 indice de référence, 138
 instruments, 330
 intégrale
 de Riemann, 29, 186
 stochastique, 187
 intégration, 243
 intégré, 242
 intervalle de confiance, 61
 Intervalles de confiance, 106
 integrated GARCH, 281
 inverse
 d'une matrice, 123
 généralisé de Moore-Penrose, 180
- J-K**
- jacobien, 49
 Jarque et Bera, 16
 jeu non biaisé (*fair game*), 302
 kernel, 196, 348
 gaussien, 28
 Kurtosis, 14

L

lemme d'Ito, 187, 336
 leptokurticité, 276
leverage effect, 283
location model, 327
 logiciel
 Eviews, 154, 207, 228, 235, 250, 348
 Excell, 189
 Mathematica, 193
 RATS, 279
 loi
 binômiale, 22
 lognormale, 188
 normale centrée réduite, 28
 normale multivariée, 42
 normale ou gaussienne, 27

M

MAE (*mean absolute error*), 249
 marche
 aléatoire avec tendance
 stochastique, 258
 aléatoire avec tendance, 260
 aléatoire, 243, 258, 263
 aléatoire, 263
 martingale, 259
 matrice
 HAC (*heteroskedastic and autocorrelation consistent covariance matrix*), 348
 adjointe, 123
 conformable, 117
 d'information, 125
 de corrélation de Pearson, 179
 de pondération des moments, 347
 de transformation, 103
 de White, 207, 332, 348
 définie positive, 101
 diagonale de transformation, 205
 diagonale, 119
 génératrice des résidus, 120
 HCCM, 332
 idempotente, 77, 120
 orthogonale, 235

partitionnée, 124
 semi-définie positive, 348
 singulière, 239
 symétrique, 77, 119
 transposée, 118
 triangulaire, 122
 variance-covariance d'un portefeuille, 126
 variance-covariance, 42, 75
 maximum de vraisemblance, 47, 104
 médiane, 9
 méthode
 classique des moments, 325
 de Box et Jenkins, 243
 de Gauss-Newton, 151
 de Prais et Winsten, 222
 des doubles moindres carrés, 346, 349
 des MCO en deux étapes, 245
 des moindres carrés ordinaires, 44, 346
 des moments généralisés, 98, 179, 327
 d'estimation GARCH-M, 287
 STAR, 254
 mineur, 124
minimum chi-square, 327
 mode, 9
 modèle
 ANN (*artificial neural network*), 255
 ANN, 256
 ARCH(1), 277
 ARCH(q), 280
 ARCH, 276
 ARCH-M, 282
 ARIMA (p,d,q), 242
 ARMA(1,1), 241
 ARMA(p,q), 242
 bifactoriel de taux d'intérêt, 342
 binomial de Cox, Ross et Rubinstein, 23
 classique linéaire général, 98
 contraint, 112
 de Cox, Ingersoll et Ross, 192
 de régression non linéaire multivarié, 153
 de Schaefer et Schwartz, 342

- EGARCH(p,q), 283
 EGARCH, 283
 ESTAR, 255
 GARCH(1,1), 281, 284
 général GARCH(p,q), 281
 IGARCH(1,1), 281, 286
 LSTAR, 255
 STAR, 254
 stochastique bifactoriel de taux
 d'intérêt de Schaefer et
 Schwartz, 335
 TAR, 255
 TARCH, 284
 contraint et non contraint, 174
 moindres
 carrés généralisés, 205, 211
 carrés non linéaires, 149, 218, 223,
 245
 carrés pondérés, 204
 carrés quasi généralisés, 205
 moments
 d'une population, 326
 échantillonaux, 326
 mouvement
 brownien avec drift, 335
 brownien généralisé, 188
 brownien géométrique, 186
 brownien géométrique, 336
 browniens de taux d'intérêt, 341
 moyenne
 géométrique, 146
 mobile, 231
 multicollinéarité, 179
- N-O**
- nombre optimal de contrats à
 terme, 133
 notation Moody's, 250
 opérateur de retard L, 230, 275
 options, 335
 d'achat, 86
 de vente asiatique, 186
 de vente, 87
 sur taux d'intérêt, 350
 ordre d'intégration, 242
 oubles moindres carrés, 183
- P**
- paramètre
 de lissage, 197
 de palier (*threshold parameter*), 255
 portefeuille
 d'arbitrage, 236
 efficient, 296
position
 long, 129
 short, 129
 positions au comptant et à terme, 90
 prévision, 79, 114
 à court terme d'une variable, 274
 à long terme d'une variable, 274
 à partir du modèle GARCH, 284
 de court terme, 259
 en présence d'autocorrélation, 226
 MMSE, 247
 à l'aide de modèles statistiques de
 processus de marche
 aléatoire, 258
 prime de risque, 95
 du marché, 240
 d'un facteur, 242
 principe de l'arbitrage, 236
 prix
 à terme, 269
 au comptant, 269
 d'exercice, 186
 d'équilibre du titre j dans le contexte
 du modèle du CAPM, 299
 du risque, 240, 288, 306
 Probabilités conditionnelles, 40
 problème d'optimisation, 139
 procédure
 itérative Cochrane-Orcutt, 219
 de discrétisation, 340
 processus
 ARCH(8), 282
 ARMA(p, q), 281
 autorégressif d'ordre 1, 343
 autorégressif d'ordre p, 216
 browniens de retour vers la
 moyenne, 337
 browniens de taux d'intérêt, 338
 d'Ito, 187
 d'Ito, 336

- processus
 de marche aléatoire avec
 tendance, 258
 de moyenne mobile MA(q), 236
 de Wiener, 186, 187, 336, 340
 intégré d'ordre 1, 242
 Ornstein-Uhlenbeck, 337, 340
 stochastique non
 stationnaire, 242, 258
 stochastique, 229
 produit Kronecker, 118
 progression géométrique, 231
 pseudo-inverse, 180
 puissance, 66
 put asiatique, 189
p-value, 66
- Q-R**
- quantité de risque, 288
 racine unitaire, 262, 263
 ratio
 de couverture optimal, 90
 de couverture, 272
 de la dette à l'équité, 283
 de vraisemblance, 173
 règle
 de l'Hôpital, 148
 de la chaîne, 291
 régresseur, 69
 régression, 69
 artificielle, 210, 287
auxiliaire ou *artificielle*, 209
 auxiliaire, 180, 270
 fallacieuse (*spurious regression*), 263
 multiple, 98
 non linéaire, 149
 non paramétrique, 196
 rendement, 93
 excédentaire, 240, 288
 réseaux de neurones, 254
 résidus stationnaires, 268
 retrait de la tendance (*detrending*), 261
ridge regression, 181
 risque
 diversifiable, 294
 non diversifiable, 294
 non systématique, 236, 295
 systématique, 293, 295
 RMSE (*root mean square error*), 249
- S**
- schème de Milstein*, 341
 série
 de Taylor du premier degré, 341
 de Taylor du second degré, 341
 DS, 267
 en coupe instantanée, 201
 en différence stationnaire, 262
 en tendance stationnaire, 262
 stationnarisée, 244
temporelle non linéaire, 253
 temporelle, 201
 TS, 267
 cointégrées, 265
 simulation de Monte Carlo, 185
skewness, 11
 SML, 91, 165
 solution non triviale, 239
 stationnarité, 216, 230
de second ordre, 230
 faible, 230
 statistique
 h de Durbin, 225
 Q, 245
 TR², 287
 U de Theil, 249
 structure à terme des taux
 d'intérêt, 153
 surparamétrisation (*overfitting*), 245
 système
 d'équations exactement
 identifié, 330
 d'équations linéaires
 homogènes, 239
 suridentifié, 330, 351
- T**
- taux à terme, 231
 taux au comptant, 230
 taux sans risque, 95
 technique de lissage des données de
 Nadaraya-Watson, 197

- technique dite de l'équivalent-certain, 300
- test
- Bera-Jarque, 79
 - Box-Cox, 146
 - d'un changement structurel, 161
 - de Breusch-Pagan, 209
 - de Chen, Roll et Ross, 243
 - de Chow (predictive), 165
 - de Chow, 163
 - de cointégration d' Engle et Granger, 268
 - de Dickey-Fuller, 265
 - de Goldfeld et Quandt, 208
 - de Kolmogorov-Smirnov, 17
 - de Perron, 265
 - de racine unitaire, 265
 - de Wald, 175
 - de White, 211
 - Durbin-Watson, 224
 - F, 67, 106
 - J, 157
 - LM d'Engle, 286
 - LM, 176
 - LR, 173
 - Phillips-Perron, 267
 - score, 176
 - sur les restrictions de suridentification, 351
 - t, 61
- tests
- bilatéraux, 106
 - exacts, 104
 - LM, LR et Wald, 105
 - non paramétriques, 105
 - unilatéraux, 106
- théorème
- central limite, 41, 78
 - de Gauss-Markov, 46, 102
 - de l'arbitrage, 233
 - de Pythagore, 295
 - de Wold, 237
- théorie
- asymptotique, 169
 - de Markowitz, 294
 - des produits dérivés, 350
 - tirage avec remise, 195
 - TR^2 , 211
 - trace
 - d'un scalaire, 121
 - d'une matrice, 120
 - tracking error, 129, 138
 - transformation
 - de Box-Cox, 147
 - en quasi-différences, 218
 - jacobienne, 47, 278
 - trend
 - d'une série chronologique, 260
 - stochastique, 259, 264
 - temporel linéaire, 261
 - temporel polynomial, 261
- V**
- valeur
- critique, 63
 - propre, 180
- VaR, 185, 195
- variable
- aléatoire, 6
 - auxiliaire, 159, 227
 - de transition, 254
 - dichotomique, 115
 - gaussienne, 341
 - instrumentale, 178, 346
 - déterministes, 6
 - instrumentales, 329
 - variance, 21, 29
 - conditionnelle, 95, 201, 275
 - du tracking-error, 139
 - échantillonnale, 9
 - non conditionnelle, 275
 - caractéristique, 181
 - GARCH-M multivariée du CAPM, 307
 - vecteur orthonormal, 241