

Université de Montréal

Quelques contributions en théorie de l'échantillonnage et dans
l'analyse des séries chronologiques multidimensionnelles

par

Pierre Duchesne

Département de mathématiques et de statistique
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en Statistique

février 2000

© Pierre Duchesne, 2000



QA
3
U54
2000
v.013

Université de Montréal

Quelques contributions au thème de l'échantillonnage et dans
l'analyse des séries chronologiques multivariées

Pierre Duchesne

École des Études Supérieures
à Montréal

7 rue Jean-Jacques, 1205
Montréal, Québec H3T 1J2
Canada

1999



1999

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée

**Quelques contributions en théorie de l'échantillonnage et dans
l'analyse des séries chronologiques multidimensionnelles**

présentée par

Pierre Duchesne

a été évaluée par un jury composé des personnes suivantes :

Christian Léger

(président-rapporteur)

Roch Roy

(directeur de recherche)

Jean-Marie Dufour

(membre du jury)

Marc Hallin

(examineur externe)

Christian Léger

(représentant du doyen)

Thèse acceptée le :

9 juin 2000

SOMMAIRE

Cette thèse comporte quatre essais en statistique appliquée. Les deux premiers portent sur la théorie de l'échantillonnage alors que les deux derniers portent sur l'analyse des séries chronologiques dans un contexte multidimensionnel.

Le premier essai porte sur les estimateurs de calage d'usage fortement répandu. Les résultats de cet essai ont fait l'objet de l'article Duchesne (1999) paru dans le numéro de juin 1999 de *Survey Methodology*, revue publiée par Statistique Canada. Les estimateurs de calage du total consistent à considérer comme estimateur du total de la variable y la quantité $\sum_s w_k y_k$, où les poids w_k sont obtenus d'une mesure de distance, sujet aux contraintes d'étalonnage et sur les poids. Ils sont faciles à interpréter car les poids w_k ont une signification en pratique. Cependant, les estimateurs de calage usuels ne sont pas robustes. On montre comment l'on peut construire des estimateurs de calage avec des propriétés raisonnables de robustesse en choisissant adéquatement des poids implicites de façon robuste qui sont transformés en poids calés. Les estimateurs proposés sont étudiés de manière empirique.

Dans le second essai, on établit à l'aide de la technique des groupes aléatoires des formules explicites de l'estimateur de variance jackknife de l'estimateur par régression. Les résultats de cet essai sont acceptés pour publication dans *Journal of Official Statistics*, revue publiée par *Statistics Sweden*. Les formules obtenues donnent de nouveaux exemples de la règle bien connue affirmant que les estimateurs de variance jackknife sur-estiment la vraie variance. Un estimateur modifié afin de corriger la sur-estimation est présenté. Dans l'illustration numérique, des résultats raisonnables sont obtenus avec ce nouvel estimateur.

Dans le troisième essai, on étudie les modèles autorégressifs multidimensionnels avec variables explicatives (VARX), qui se retrouvent fréquemment dans les applications statistiques. Plusieurs propriétés des statistiques de base pour cette classe de modèles reposent sur l'hypothèse d'indépendance des erreurs. En s'inspirant de Hong (1996a), une statistique de test pour vérifier cette dernière hypothèse est proposée. Elle est obtenue en comparant la densité spectrale des erreurs sous l'hypothèse nulle d'indépendance à un estimateur de la densité spectrale calculé à l'aide de la méthode du noyau. La loi asymptotique de la statistique de

test est obtenue sous l'hypothèse nulle. Le test résultant généralise la statistique portmanteau de Hosking (1980). Cependant, certains noyaux permettent d'obtenir une meilleure puissance. Le niveau et la puissance des tests résultants sont étudiés par simulation et une application est présentée.

Dans le quatrième essai, on s'intéresse au problème d'indépendance de deux séries. Une approche classique est celle de Haugh (1976), mais sa statistique repose sur un nombre fixé de délais. Hong (1996*b*) considère une statistique pondérée qui généralise la statistique de Haugh. Dans bien des situations la puissance est supérieure. Cependant, le problème des valeurs aberrantes survenant en séries chronologiques est important et les statistiques de Haugh et Hong sont susceptibles d'être sensibles dans ce contexte. Une statistique de test robuste pour vérifier l'indépendance de deux séries et prenant en compte un nombre fixé de délais a été considérée dans Li et Hui (1994). La statistique que nous proposons généralise celle de Hong et celle de Li et Hui. Elle repose sur l'ensemble des délais et offre de meilleures propriétés de puissance. Le niveau et la puissance des tests sont étudiés empiriquement. Une application est présentée pour illustrer la méthodologie.

REMERCIEMENTS

Cette thèse n'aurait pu voir le jour sans des circonstances favorables, où les difficultés furent surmontées grâce à une confiance témoignée dans un climat rendant propice l'accomplissement. Après des travaux en échantillonnage, je tiens en premier lieu à remercier mon directeur de recherche Roch Roy pour avoir accepté de superviser mes travaux subséquents en séries chronologiques. Les échanges et discussions furent toujours réalisés dans un climat de grande confiance, rendant possible le développement de mes idées en séries chronologiques. Je le remercie particulièrement pour son intérêt dans tous mes travaux. Ses commentaires judicieux et ses suggestions pertinentes témoignent d'une grande expérience en recherche, et ce fut un privilège de travailler sous sa direction.

Je remercie en second lieu Christian Léger, avec qui les échanges très nombreux m'ont permis d'avoir une meilleure compréhension et vision de la recherche ainsi que de la statistique en général. Je le remercie pour m'avoir fait part à de très nombreuses occasions de ses conseils, toujours avec une grande intelligence et une grande justesse. Plusieurs conversations furent déterminantes pour moi, à des moments divers tout au long de mon cheminement. Je le remercie sincèrement pour sa confiance, son modèle d'honnêteté et de rigueur, et son soutien.

Je tiens également à remercier Carl-Erik Särndal et Martin Bilodeau, pour les échanges profitables qui ont marqués à plusieurs points de vue mes travaux de recherche. Je remercie aussi François Perron et Yannis Yatracos pour les nombreuses conversations.

Je remercie le CRSNG et le Fonds FCAR pour les bourses de doctorat accordées. Je remercie également l'ISM et la Faculté des Études Supérieures pour l'appui financier sous forme de bourses. Je remercie l'Équipe FCAR, l'Équipe MITACS ainsi que le CRSNG pour le soutien financier accordé à travers les fonds de mon directeur de recherche.

Je remercie ma mère, pour avoir toujours eu une grande confiance en mes moyens. Merci à la famille et aux amis, pour leurs encouragements et le soutien moral. Finalement, je remercie mon épouse Ampha pour avoir été présente à mes côtés durant tout ce trajet. Attentive et dévouée, elle fut toujours présente dans les moments difficiles comme dans les moments heureux.

Table des matières

Sommaire	iii
Remerciements	v
Table des figures	x
Liste des tableaux	xi
Introduction	1
Théorie de l'échantillonnage	1
Estimateurs de calage robustes	2
Une note sur l'estimation de variance jackknife pour le GREG	6
Analyse des séries chronologiques	6
Tests de corrélation sérielle dans les modèles <i>VARX</i>	6
Tests d'indépendance robustes entre deux séries chronologiques	8
Chapitre 1. Estimateurs de calage robustes	12
1.1. Introduction	12
1.2. Estimateurs de la classe RQR	16
1.3. Construction d'estimateurs robustes et calés	19
1.3.1. Méthodes reposant sur la réduction de poids et la modification des valeurs .	19
1.3.2. Estimateurs du total basés sur des statistiques robustes	22
1.3.2.1. Choix de q_k basé sur un <i>GM</i> -estimateur	23
1.3.2.2. Choix de q_k basé sur un estimateur avec haut point de rupture	24
1.3.2.3. Choix de r_k	25
1.3.3. Estimateur modéliste de Chambers	27
1.4. Étude empirique	28
1.4.1. Populations à l'étude	29

1.4.2. Description des estimateurs	30
1.4.3. Mesures fréquentistes	31
1.4.4. Discussion	31
1.5. Conclusion	36
Remerciements	37
A. Preuve de la proposition 1	37
B. Liste des abréviations	38
C. Liste des principales constantes	38
Bibliographie	39
Annexe A. Complément à “Estimateurs de calage robustes”	42
A.1. Estimation de l’erreur quadratique moyenne	42
A.2. Nouvelle métrique	44
A.3. Simulations dans un plan poisson séquentiel	45
Bibliographie	48
Chapitre 2. A note on jackknife variance estimation for the general regression estimator	49
2.1. Introduction	49
2.2. Jackknife variance estimation	50
2.3. Corrected estimator for the model bias	52
2.4. Illustration	53
2.5. Discussion	54
2.6. Acknowledgements	55
Bibliographie	56
Chapitre 3. Tests de corrélation sérielle dans les modèles VARX	57
3.1. Introduction	57

3.2. Préliminaires.....	59
3.2.1. Théorème central limite pour les différences de martingales.....	61
3.3. Estimation par les moindres carrés calculables.....	62
3.4. Tests dans le modèle linéaire dynamique.....	64
3.5. Puissance du test de Hong.....	68
3.5.1. Puissance asymptotique locale.....	68
3.5.2. Puissance asymptotique globale.....	70
3.6. Simulations.....	72
3.6.1. Choix des modèles.....	73
3.6.2. Discussion de l'étude de niveau.....	74
3.6.3. Discussion de l'étude de puissance.....	76
3.7. Application.....	80
3.8. Preuve des théorèmes.....	83
3.8.1. Preuve du théorème 2.....	83
3.8.2. Preuve du théorème 3.....	105
3.8.3. Preuve du théorème 4.....	106
Bibliographie.....	110
Chapitre 4. Tests d'indépendance robustes entre deux séries chronologiques	112
4.1. Introduction.....	112
4.2. Préliminaires.....	115
4.2.1. Hypothèses sur les processus.....	115
4.2.2. Hypothèses sur les noyaux.....	115
4.2.3. Valeurs aberrantes en séries chronologiques.....	116
a) Aberrants dans le processus innovation.....	117
b) Aberrants additifs.....	117
4.2.4. Estimateurs robustes et fonction d'autocovariance robuste.....	118
4.3. Tests basés sur la distribution asymptotique des corrélations croisées robustes .	120
4.3.1. Distribution asymptotiques des corrélations croisées robustes.....	120
4.3.2. Test basé sur des corrélations croisées robustes à un délai.....	121

4.3.3. Liens avec la causalité	122
4.4. Procédure basée sur une statistique de test pondérée	124
4.4.1. Distribution asymptotique sous l'hypothèse nulle.....	125
4.5. Étude de simulation.....	126
4.5.1. Choix des modèles.....	127
4.5.2. Emplacement des valeurs aberrantes	128
4.5.3. Statistiques considérées.....	128
4.5.4. Discussion des résultats sur l'étude de niveau	129
4.5.5. Tableaux pour l'étude de niveau	133
4.5.6. Discussion des résultats sur l'étude de puissance	142
4.5.7. Tableaux pour l'étude de puissance	144
4.6. Application.....	153
4.6.1. Analyse préliminaire des données	154
4.6.2. Tests de non-corrélation entre les deux séries.....	155
4.6.3. Tests de non-corrélation entre les deux séries avec une valeur aberrante	156
4.7. Preuve du théorème.....	159
Bibliographie	174
Conclusion	176
Chapitre 1: Estimateurs de calage robustes.....	176
Chapitre 2: A note on jackknife variance estimation for the general regression estimator	177
Chapitre 3: Tests de corrélation sérielle dans les modèles <i>VARX</i>	177
Chapitre 4: Tests d'indépendance robustes entre deux séries chronologiques.....	178
Bibliographie.....	179

Table des figures

1.1	Les quatre populations à l'étude.....	29
3.1	Les quatre variables à l'étude.....	80
4.1	Types de valeurs aberrantes. Le graphique de gauche correspond aux valeurs aberrantes dans les innovations alors que le graphique de droite correspond aux valeurs aberrantes additives.	118
4.2	Séries chronologiques portant sur le nombre de peaux d'animaux commercées annuellement pour la période 1848-1911, pour le vison et le rat musqué.	153
4.3	Tests à un délai basés sur les carrés des corrélations croisées entre les visons et les rats musqués.	156
4.4	Les deux variables à l'étude en présence d'une valeur aberrante.	157
4.5	Tests à un délai basés sur les carrés des corrélations croisées usuelles et robustes entre les visons et les rats musqués, avec une valeur aberrante.	158

Liste des tableaux

1.1	Estimateur (1.12) reposant sur la réduction de poids et la modification des valeurs.	21
1.2	Totaux des diverses populations et totaux connus de l'information auxiliaire.	28
1.3	Résultats de Monte-Carlo lors de l'échantillonnage dans la population <i>POPUSA</i> .	32
1.4	Résultats de Monte-Carlo lors de l'échantillonnage dans la population <i>AREA</i>	33
1.5	Résultats de Monte-Carlo lors de l'échantillonnage dans la population <i>MU284</i>	34
1.6	Résultats de Monte-Carlo lors de l'échantillonnage dans la population <i>MU281</i>	35
A.1	Résultats de Monte-Carlo lors de l'échantillonnage dans la population <i>POPUSA</i> , <i>AREA</i> , <i>MU284</i> et <i>MU281</i> pour l'estimateur robuste avec métrique logit.	45
A.2	Résultats de Monte-Carlo lors de l'échantillonnage dans la population <i>MU284</i> sous un plan poisson séquentiel	47
A.3	Résultats de Monte-Carlo lors de l'échantillonnage dans la population <i>MU281</i> sous un plan poisson séquentiel	47
2.1	Results of the Simulation	54
3.1	Zones de non-rejet (en pourcentage) de ' H_0 : niveau est α (en pourcentage)', basées sur 10000 réalisations.....	72
3.2	Niveaux en pourcentage de la statistique de Hong multivariée avec le modèle donné dans la formule 3.20	75
3.3	Puissance en pourcentage basée sur les quantiles asymptotiques (empiriques) lorsque $\delta = 0.75$	77
3.4	Puissance en pourcentage basée sur les quantiles asymptotiques (empiriques) lorsque $\delta = 1.0$	78
3.5	Puissance en pourcentage basée sur les quantiles asymptotiques (empiriques) lorsque $\delta = 1.25$	79
3.6	Estimation du modèle de Gilbert (1993) portant sur divers indicateurs canadiens.	82

3.7	Valeurs de la statistique portmanteau	82
4.1	Zones de non-rejet de ' H_0 : niveau est α (en pourcentage)', basées sur 10000 réalisations.....	127
4.2	Nombres de rejets (sous l'hypothèse nulle d'indépendance) basés sur 10000 réalisations en utilisant la fonction ψ donnant les fonctions de corrélation usuelles avec les estimateurs des moindres carrés et les statistiques de Hong Q_n données par (4.2).	133
4.3	Nombres de rejets (sous l'hypothèse nulle d'indépendance) basés sur 10000 réalisations en utilisant la fonction ψ donnant les fonctions de corrélation usuelles avec les estimateurs des moindres carrés et les statistiques de Hong Q_n^* données par (4.11).	134
4.4	Nombres de rejets (sous l'hypothèse nulle d'indépendance) basés sur 10000 réalisations en utilisant la fonction ψ donnant les fonctions de corrélation usuelles avec les estimateurs des moindres carrés et les statistiques de Haugh S_M et S_M^* données par (4.1) et (4.10).	135
4.5	Nombres de rejets (sous l'hypothèse nulle d'indépendance) basés sur 10000 réalisations en utilisant la fonction ψ de type Huber avec les estimateurs RA et les statistiques robustes Q_{Rn} données par (4.8).	136
4.6	Nombres de rejets (sous l'hypothèse nulle d'indépendance) basés sur 10000 réalisations en utilisant la fonction ψ de type Huber avec les estimateurs RA et les statistiques robustes Q_{Rn}^* données par (4.9).	137
4.7	Nombres de rejets (sous l'hypothèse nulle d'indépendance) basés sur 10000 réalisations en utilisant la fonction ψ de type Huber avec les estimateurs RA et les statistiques de Li et Hui S_{RM} et S_{RM}^* données par (4.6) et (4.7).	138
4.8	Nombres de rejets (sous l'hypothèse nulle d'indépendance) basés sur 10000 réalisations en utilisant la fonction ψ de type bicarré avec les estimateurs RA et les statistiques robustes Q_{Rn} données par (4.8).	139
4.9	Nombres de rejets (sous l'hypothèse nulle d'indépendance) basés sur 10000 réalisations en utilisant la fonction ψ de type bicarré avec les estimateurs RA et les statistiques robustes Q_{Rn}^* données par (4.9).	140
4.10	Nombres de rejets (sous l'hypothèse nulle d'indépendance) basés sur 10000 réalisations en utilisant la fonction ψ de type bicarré avec les estimateurs RA et les statistiques de Li et Hui S_{RM} et S_{RM}^* données par (4.6) et (4.7).	141

4.11	Nombres de rejets (sous l'hypothèse alternative) basés sur 10000 réalisations en utilisant la fonction ψ donnant les fonctions de corrélation usuelles avec les estimateurs des moindres carrés et les statistiques de Hong Q_n données par (4.2).	144
4.12	Nombres de rejets (sous l'hypothèse alternative) basés sur 10000 réalisations en utilisant la fonction ψ donnant les fonctions de corrélation usuelles avec les estimateurs des moindres carrés et les statistiques de Hong Q_n^* données par (4.11).	145
4.13	Nombres de rejets (sous l'hypothèse alternative) basés sur 10000 réalisations en utilisant la fonction ψ donnant les fonctions de corrélation usuelles avec les estimateurs des moindres carrés et les statistiques de Haugh S_M et S_M^* données par (4.1) et (4.10).	146
4.14	Nombres de rejets (sous l'hypothèse alternative) basés sur 10000 réalisations en utilisant la fonction ψ de type Huber avec les estimateurs RA et les statistiques robustes Q_{Rn} données par (4.8).	147
4.15	Nombres de rejets (sous l'hypothèse alternative) basés sur 10000 réalisations en utilisant la fonction ψ de type Huber avec les estimateurs RA et les statistiques robustes Q_{Rn}^* données par (4.9).	148
4.16	Nombres de rejets (sous l'hypothèse alternative) basés sur 10000 réalisations en utilisant la fonction ψ de type Huber avec les estimateurs RA et les statistiques de Li et Hui S_{RM} et S_{RM}^* données par (4.6) et (4.7).	149
4.17	Nombres de rejets (sous l'hypothèse alternative) basés sur 10000 réalisations en utilisant la fonction ψ de type bicarré avec les estimateurs RA et les statistiques robustes Q_{Rn} données par (4.8).	150
4.18	Nombres de rejets (sous l'hypothèse alternative) basés sur 10000 réalisations en utilisant la fonction ψ de type bicarré avec les estimateurs RA et les statistiques robustes Q_{Rn}^* données par (4.9).	151
4.19	Nombres de rejets (sous l'hypothèse alternative) basés sur 10000 réalisations en utilisant la fonction ψ de type bicarré avec les estimateurs RA et les statistiques de Li et Hui S_{RM} et S_{RM}^* données par (4.6) et (4.7).	152

INTRODUCTION

Deux domaines importants de la statistique appliquée sont sans aucun doute la théorie de l'échantillonnage et l'analyse des séries chronologiques. Nous allons dans les deux sections qui suivent introduire les concepts de base utiles dans notre contexte, et donner un aperçu du contenu des quatre essais.

THÉORIE DE L'ÉCHANTILLONNAGE

La théorie de l'échantillonnage cherche souvent à estimer des quantités inconnues dans le contexte d'une population finie, que l'on note $U = \{1, \dots, N\}$. Étant donné une variable d'intérêt y , un but fréquent est l'estimation du total $T_y = \sum_U y_k$ inconnu (Si A est un ensemble d'unités, $A \subseteq U$, alors \sum_A est une notation signifiant $\sum_{k \in A}$). Un échantillon s est observé selon un plan d'échantillonnage. Les probabilités d'inclusion d'ordre un et deux sont notées π_k et π_{kl} . Le plan d'échantillonnage pourrait être un tirage aléatoire simple (où tous les π_k sont égaux et valent n/N), ou encore un plan πps .

Un estimateur bien connu dans ce contexte est l'estimateur Horvitz-Thompson (HT), défini par $\hat{T}_{yHT} = \sum_s d_k y_k$, où $d_k = \pi_k^{-1}$ est le poids d'échantillonnage. Cependant, nous avons souvent en plus de l'information auxiliaire x de dimension p , qui peut être utilisée afin d'obtenir de meilleures estimations. Par exemple, un estimateur utilisant l'information auxiliaire est l'estimateur par la régression généralisée (GREG) de T_y donné par

$$\hat{T}_{GREG} = \sum_s d_k g_{ks} y_k,$$

où

$$g_{ks} = 1 + (T_x - \hat{T}_{x\pi})' \left(\sum_s d_k x_k x_k' / c_k \right)^{-1} x_k / c_k,$$

est le poids-g, $\hat{T}_{x\pi} = \sum_s x_k / \pi_k$ est l'estimateur HT pour T_x , et c_k est choisi par l'utilisateur. Les constantes c_k sont des facteurs de pondération qui peuvent prendre en compte (par exemple) des problèmes d'hétéroscédasticité. Särndal (1996) discute le choix des c_k . La

variance asymptotique AV pour le GREG est donnée par

$$AV(\hat{T}_{GREG}) = \sum \sum_U \Delta_{kl} \check{E}_k \check{E}_l,$$

où $E_k = y_k - x'_k B$, $B = (\sum_U x_k x'_k / c_k)^{-1} \sum_U x_k y_k / c_k$, $\Delta_{kl} = \pi_{kl} - \pi_k \pi_l$ et $\check{E}_k = E_k / \pi_k$.

Nous allons considérer dans le premier essai des estimateurs du total T_y s'écrivant comme $\sum_s w_k y_k$, appelés estimateurs de calage. Ils sont développés par exemple dans Deville et Särndal (1992). On cherche les poids w_k aussi près que possible des poids d'échantillonnage $d_k = \pi_k^{-1}$ mais qui respectent les contraintes d'étalonnage, notée CE (dites aussi contraintes de calage),

$$\sum_s w_k x_k = T_x,$$

où x_k est l'information auxiliaire disponible de total $T_x = \sum_U x_k$ connu. Ces estimateurs sont populaires car ils s'interprètent aisément, puisque les méthodologistes sont habitués d'apposer des poids w_k aux unités y_k . Plusieurs métriques sont étudiées afin de mesurer la proximité entre d_k et w_k . L'estimateur GREG est un exemple important avec $w_k = d_k g_{ks}$. Il est obtenu en minimisant la métrique quadratique $\sum_s c_k (w_k - d_k)^2 / d_k$. Cependant, puisque les poids-g $g_k = w_k / d_k$ du GREG ne sont pas bornés en général, d'autres métriques sont proposées afin de les borner pour qu'ils satisfassent certaines restrictions applicables à la fourchette des valeurs (RAFV). Ceci permet en particulier d'éviter la présence indésirable de poids w_k négatifs. Voir Deville et Särndal (1992), Singh et Mohl (1996) et Stukel, Hidiroglou et Särndal (1996).

La variance asymptotique est habituellement une quantité inconnue. Ainsi, Särndal, Swensson et Wretman (1989) ont suggéré un estimateur pondéré par les poids-g, pour estimer la variance du GREG:

$$\hat{V}_g = \hat{V}_g(\hat{T}_{GREG}) = \sum \sum_s \check{\Delta}_{kl} (g_{ks} \check{e}_{ks})(g_{ls} \check{e}_{ls}),$$

où $e_{ks} = y_k - x'_k \hat{B}_s$, $\hat{B}_s = (\sum_s d_k x_k x'_k / c_k)^{-1} \sum_s d_k x_k y_k / c_k$, $\check{\Delta}_{kl} = \Delta_{kl} / \pi_{kl}$ et $\check{e}_{ks} = e_{ks} / \pi_k$. Avec \hat{V}_g , on peut construire un intervalle de confiance T_y donné par $\hat{T}_{GREG} \pm z_{1-\alpha/2} [\hat{V}_g]^{1/2}$, valide approximativement au niveau de confiance $1 - \alpha$. Une autre méthode pour estimer la variance est la méthode du jackknife, que nous étudions pour le GREG dans l'essai 2.

Estimateurs de calage robustes

L'estimateur GREG repose essentiellement sur l'estimateur des moindres carrés généralisés, qui est bien connu être sensible à la présence de valeurs aberrantes. Le premier essai

s'intéresse à la recherche d'estimateurs alternatifs au GREG, satisfaisant des propriétés raisonnables de robustesse.

Le problème des valeurs aberrantes est important dans toutes les branches de la statistique. En théorie des sondages, le contexte est différent de celui de la statistique paramétrique puisque l'on s'intéresse souvent à l'estimation du total T_y . Or une valeur aberrante peut avoir son plein poids dans le total de la population. De plus, le méthodologiste qui procède à la phase d'estimation peut supposer que les valeurs des unités sont enregistrées sans erreur, puisque souvent les unités une fois ramassées sont traitées dans un système d'édition (Särndal, Swensson et Wretman 1992, Section 1.7). Cette étape fait partie de la réalisation d'un sondage dans les grandes agences statistiques, comme par exemple à Statistique Canada. Lee (1995) fait un survol des développements de la robustesse en théorie des sondages.

Cependant, puisqu'il arrive fréquemment que les populations dans les enquêtes économiques sont asymétriques, quelques unités risquent d'être *extrêmes* par rapport aux autres, sujet abordé dans Kish (1965). Retirer complètement ces unités entraîne des estimations biaisées, alors que si on les conserve avec leur plein poids un estimateur comme le GREG risque d'être très variable. Ceci suggère un compromis entre le biais et la variance. En présence de valeurs aberrantes, le défi est de proposer des estimateurs *robustes* du total peu affectés par certaines unités très à l'écart des autres. Ces estimateurs devront avoir un petit biais et une petite erreur quadratique moyenne. Traditionnellement, la théorie de l'échantillonnage a consacré beaucoup d'efforts à développer des estimateurs sans biais ou asymptotiquement sans biais (ADU). Voir par exemple Särndal et al. (1992, Section 7.12). Cependant, la propriété ADU n'est peut-être pas souhaitable dans un contexte de valeurs aberrantes. Voir la discussion de Chambers et Kokic (1993) qui montrent qu'un conflit est présent entre la propriété ADU et la robustesse d'un estimateur.

Supposons la variable d'intérêt y positive et la population asymétrique. L'estimateur HT étant une moyenne pondérée par les d_k , il est vulnérable aux grandes valeurs de y . Une unité avec un poids d_k élevé peut aussi avoir un grand impact sur l'estimation en entraînant des estimations variables. Lee (1995) définit ces unités comme *influentes*. Une unité extrême n'est pas forcément influente si son poids d_k est suffisamment petit. Traditionnellement, les méthodologistes cherchent à limiter l'impact des unités influentes lorsqu'elles sont connues avant le sondage, en donnant par exemple des poids d'échantillonnage proche de un aux unités extrêmes. Gambino (1987) et Lee (1995) discutent cependant de situations où nous ne pouvons pas agir ainsi. Dans un article important, Hidiroglou et Srinath (1981) considèrent

l'altération des poids d'échantillonnage en présence de valeurs aberrantes. Leur approche a grandement légitimé la modification des poids dans le monde des sondages.

Parmi les premières alternatives robustes du total, plusieurs reposent sur les M -estimateurs et les GM -estimateurs. Cependant, beaucoup d'attention a été consacrée récemment à des estimateurs possédant en plus de bonnes propriétés de robustesse globale, telle que mesurée par le point de rupture d'un estimateur. Ces notions sont discutées par exemple dans Donoho et Huber (1983), Hampel, Ronchetti, Rousseeuw et Stahel (1986) et Rousseeuw et Leroy (1987). Le point de rupture mesure le pourcentage de valeurs aberrantes dans l'échantillon que l'estimateur peut tolérer tout en donnant tout de même une bonne estimation d'une certaine caractéristique de la population. Lee, Ghangurbe, Mach et Yung (1992) réclament des estimateurs du total basés sur des estimateurs robustes avec un haut point de rupture.

Les estimateurs de calage ont un lien avec les méthodes robustes comme noté dans Fuller, Loughin et Baker (1994, p. 81). Cependant, il est inexact de prétendre que les estimateurs de calage possèdent nécessairement de bonnes propriétés de robustesse, puisqu'il suffit de se rappeler que tous les estimateurs de calage considérés par Deville et Särndal (1992) sont asymptotiquement équivalents au GREG qui étant ADU n'est pas robuste. De plus, on voit qu'un estimateur traditionnel de calage n'est pas robuste puisqu'il dépend linéairement de w_k , et w_k ne prend pas en compte y_k .

Le but du premier essai est de construire des estimateurs de la forme $\sum_s w_k y_k$ où les poids w_k assurent en même temps la robustesse et le respect des contraintes sur les variables de calage et sur les poids w_k . Le point de départ de notre approche est la classe des estimateurs QR de Wright (1983). Supposons que nous disposions des constantes $\{(q_k, r_k), q_k > 0, r_k \geq 0, \forall k \in U\}$, telles que $\sum_U \pi_k q_k x_k x_k' > 0$ et $\sum_s q_k x_k x_k' > 0, \forall s$. (Si A est une matrice symétrique, $A > 0$ signifie que A est définie positive). Les estimateurs QR sont définis à partir de q_k et r_k par la relation

$$\hat{T}_{yQR} = T_x' \hat{B}_q + \sum_s r_k e_k,$$

où \hat{B}_q admet une forme pondérée par les q_k

$$\hat{B}_q = \left(\sum_s q_k x_k x_k' \right)^{-1} \sum_s q_k x_k y_k,$$

où $e_k = y_k - x_k' \hat{B}_q$. Nous montrons que les estimateurs QR sont des estimateurs de calage et introduisons une nouvelle classe d'estimateurs, notée RQR, reposant également sur le choix des constantes q_k et r_k . Elle généralise d'une certaine façon les estimateurs QR ainsi que la classe des estimateurs de Deville et Särndal (1992). L'intérêt de la classe RQR est qu'elle permet

d'obtenir des poids w_k restreints à un intervalle donné, disons $[L, U]$. Quelques propriétés des classes QR et RQR sont discutées.

Un aspect intéressant de la classe RQR est qu'elle fournit un cadre pour la construction d'estimateurs de calage robustes. L'idée principale consiste à modifier des poids implicites robustes pour qu'ils respectent les contraintes de calage. Nous discutons du choix des constantes q_k et r_k en utilisant un raisonnement propre aux estimateurs de calage. Cette approche s'avère nouvelle et unificatrice, et nous guide dans le choix des q_k et r_k en présence d'information auxiliaire. Un élément important est l'utilisation d'un estimateur robuste admettant une forme pondérée fournissant les q_k . Nous remarquons que c'est le cas pour les *GM*-estimateurs. Les estimateurs avec un haut point de rupture n'ont habituellement pas une forme pondérée. Une repondération de ces estimateurs est considérée, permettant de conserver le point de rupture sous contrôle et d'avoir des estimateurs s'écrivant sous la forme pondérée. Voir Rousseeuw et Leroy (1987) et Simpson et Chang (1997). Nous discutons alors des choix de q_k et r_k permettant de calculer un estimateur RQR afin d'obtenir un estimateur de calage robuste avec des poids bornés. Dans une étude empirique, nous comparons divers estimateurs robustes, dont l'estimateur de Lee (1991) et celui de Chambers (1986), aux estimateurs RQR ainsi qu'au GREG et à un estimateur de calage considéré dans Deville et Särndal (1992) dont les poids sont restreints. L'estimateur de Lee (1991) peut être considéré comme un cas particulier de notre approche. Il nous permet aussi de considérer un nouvel estimateur avec des poids bornés. Quatre populations déjà étudiées dans la littérature sont considérées. On notera que les estimateurs sans contrainte sur les poids sont sujets à des problèmes de pondération négative. Les estimateurs de la classe RQR permettent d'obtenir des estimateurs robustes possédant des poids positifs, se comparant bien aux estimateurs sans contrainte sur les poids. Nous concluons l'essai en donnant quelques recommandations ainsi qu'en discutant quelques avenues de recherche possible.

Dans l'annexe A se trouve un complément à l'essai 1. L'estimation de l'erreur quadratique moyenne des estimateurs RQR est considérée et une nouvelle métrique est introduite. Des simulations d'estimateurs robustes avec cette métrique sont considérées. Finalement, l'effet du plan d'échantillonnage est abordé, en considérant des simulations sous un plan **SPS**, ce dernier plan étant une modification du plan **PO**.

Une note sur l'estimation de variance jackknife pour le GREG

Dans le second essai, nous nous intéressons à l'estimation de la variance de l'estimateur GREG avec la méthode du jackknife. Le jackknife est une autre technique populaire pour obtenir un estimateur de variance. Cette méthode est décrite dans Wolter (1985) et Särndal et al. (1992, chap. 11). Nous obtenons des formules explicites pour les estimateurs de variance jackknife du GREG. Une version corrigée est proposée qui retire une grande partie du biais positif calculé selon le modèle. Dans une brève étude empirique, nous obtenons des résultats raisonnables avec la version corrigée.

ANALYSE DES SÉRIES CHRONOLOGIQUES

L'étude des séries chronologiques cherche à expliquer des phénomènes observés dans le temps. Plusieurs phénomènes physiques ou économiques peuvent être décrits à l'aide de modèles de séries chronologiques. On s'intéresse à deux types de problèmes dans un contexte multidimensionnel, supposant que nous avons plusieurs variables à l'étude. Dans le premier problème nous nous intéressons à des tests de corrélation sérielle dans le terme d'erreur dans une certaine classe de modèles, alors que dans le second nous nous intéressons à des tests d'indépendance robustes entre deux séries chronologiques.

Tests de corrélation sérielle dans les modèles *VARX*

Dans le premier problème on s'intéresse à la classe des processus autorégressifs multivariés avec variables explicatives, notée $VARX(r,s)$. Ces processus s'écrivent sous la forme

$$\mathbf{\Lambda}(B)\mathbf{y}_t = \mathbf{c} + \mathbf{V}(B)\mathbf{x}_t + \mathbf{u}_t, \quad t \in \mathbb{Z},$$

où $\mathbf{\Lambda}(B) = \mathbf{\Lambda}_0 - \sum_{j=1}^r \mathbf{\Lambda}_j B^j$, $\mathbf{V}(B) = \sum_{j=0}^s \mathbf{V}_j B^j$, B désigne l'opérateur retard et $\mathbf{u} = \{\mathbf{u}_t : t \in \mathbb{Z}\}$ est un bruit blanc fort de dimension d , c'est-à-dire que les \mathbf{u}_t sont des vecteurs aléatoires indépendants et identiquement distribués. La classe des modèles *VARX* est très utilisée dans les applications touchant entre autres l'économie. Ils sont aussi appelés modèles à équations simultanées dynamiques dans la littérature économétrique. Ces modèles généralisent les modèles de régression linéaire multiple où l'on inclut parmi les variables explicatives des variables endogènes et exogènes retardées. Ainsi, dictés par des considérations théoriques ou empiriques, ces modèles permettent de formuler des situations où l'on reconnaît qu'il peut exister des relations causales entre des variables économiques présumées stochastiques. Dans de tels modèles, on reconnaît également que des données économiques sont susceptibles

d'être influencées par les états actuels et passés des variables impliquées. Dans cette classe de modèles, nous pouvons être amené à vérifier si toutes les hypothèses de base entourant ce modèle ne sont pas violées de façon flagrante, comme par exemple l'absence de corrélation du terme d'erreur.

Dans les modèles ARX univarié, Hong (1996a) a proposé plusieurs classes de tests convergents pour vérifier l'hypothèse que le terme d'erreur est un bruit blanc contre une alternative de dépendance de forme inconnue. Rappelons qu'un test convergent admet une puissance asymptotique tendant vers un lorsque le nombre d'observations n tend vers l'infini, pour un niveau α fixé. Ces tests sont motivés par le fait que toute forme de corrélation sérielle dans le terme d'erreur risque d'entraîner des problèmes dans l'identification des paramètres ainsi que la non convergence de l'estimateur des moindres carrés. Ainsi, Hong (1996a) a proposé trois classes de statistiques de type portmanteau pour l'hypothèse de bruit blanc pour le terme d'erreur d'un modèle ARX. Son approche repose sur une comparaison d'un estimateur de la densité spectrale calculé par la méthode du noyau à la densité spectrale sous l'hypothèse nulle. Plusieurs mesures de distance sont considérées par Hong (1996a). Lorsque la fonction de distance choisie est la norme quadratique, la statistique de Hong (1996a) s'écrit comme

$$M_{1n} = \frac{n \sum_{j=1}^{n-1} k^2(j/p_n) \hat{\rho}^2(j) - M_n(k)}{\sqrt{2V_n(k)}},$$

où $\hat{\rho}(j) = \hat{R}(j)/\hat{R}(0)$ et $\hat{R}(j) = n^{-1} \sum_{t=|j|+1}^n \hat{u}_t \hat{u}_{t-|j|}$, k est un noyau et

$$M_n(k) = \sum_{j=1}^{n-1} (1 - j/n) k^2(j/p_n),$$

$$V_n(k) = \sum_{j=1}^{n-2} (1 - j/n)(1 - (j+1)/n) k^4(j/p_n).$$

Hong (1996a) considère aussi d'autres mesures de distance, dont une basée sur la métrique de Hellinger et une autre basée sur le critère d'information de Kullback-Leibler.

En présence de plusieurs équations, il incombe également de vérifier l'hypothèse que le terme d'erreur est un bruit blanc. Ainsi, nous étendons l'approche de Hong (1996a) aux modèles à équations simultanées dynamiques écrits sous leur forme réduite. Tout comme dans le cas univarié, nous développons un test convergent pour une alternative de corrélation sérielle de forme inconnue. L'approche consiste à comparer un estimateur de la densité spectrale multidimensionnel obtenue avec la méthode du noyau et la densité spectrale sous l'hypothèse

nulle, en utilisant une certaine norme quadratique. Nous obtenons la distribution asymptotique sous l'hypothèse nulle qui est $N(0, 1)$ sans avoir à spécifier un modèle pour les erreurs sous l'hypothèse alternative. Le test reste invariant lorsque les régresseurs contiennent des variables endogènes retardées. Ce test permet aussi d'obtenir une nouvelle interprétation de la statistique portmanteau multivariée de Hosking (1980), ce dernier test étant la version multivariée de la statistique de Box et Pierce (1970).

La puissance est étudiée théoriquement, suivant l'approche décrite dans Hong (1996a), et on démontre que certains noyaux permettent d'obtenir une meilleure puissance que celle du test de Hosking (1980).

Une approche différente abordant les tests d'ajustement d'une série univariée est décrite dans Paparoditis (1999). L'approche consiste à comparer avec une mesure de distance un estimateur par le noyau du ratio de la véritable densité spectrale et de la densité spectrale hypothétique, à la valeur espérée de l'estimateur, sous l'hypothèse nulle. L'étude de la puissance des tests d'ajustement est étudiée dans Paparoditis (2000).

Le niveau et la puissance des nouveaux tests sont également étudiés dans une étude de simulation. L'étude empirique suggère également que des gains peuvent être obtenus en choisissant un noyau autre que le noyau uniforme tronqué. Une application est présentée, illustrant l'utilisation des nouveaux tests.

Tests d'indépendance robustes entre deux séries chronologiques

Dans le second problème on s'intéresse aux tests robustes d'indépendance entre deux séries chronologiques unidimensionnelles. On suppose que les deux processus $\{X_t\}$ et $\{Y_t\}$ sont dans la classe des processus ARMA:

$$\begin{aligned}\phi_1(B)X_t &= \theta_1(B)u_t, \\ \phi_2(B)Y_t &= \theta_2(B)v_t,\end{aligned}$$

où $\phi_1(B) = \sum_{i=0}^{p_1} \phi_{1i}B^i$, $\theta_1(B) = \sum_{i=0}^{q_1} \theta_{1i}B^i$, où B est l'opérateur retard. Les définitions pour $\phi_2(B)$ et $\theta_2(B)$ sont similaires. On suppose que $\theta_{10} = \theta_{20} = \phi_{10} = \phi_{20} = 1$. Il est présumé que $\{X_t\}$ et $\{Y_t\}$ sont stationnaires et inversibles. Lorsque nous avons à notre disposition deux séries chronologiques, une question naturelle porte sur l'existence de relations entre elles. En économie, l'élucidation des relations de causalité entre les séries chronologiques est souvent très importante dans un contexte de prévision. Une autre question importante est l'indépendance entre deux séries chronologiques. Comme l'hypothèse d'indépendance entre deux séries

implique la non-causalité, alors si des procédures raisonnablement puissantes pour vérifier l'indépendance sont disponibles, alors une analyse de causalité peut devenir redondante. Les approches connues pour tester l'indépendance entre deux séries reposent habituellement sur l'utilisation de la fonction de corrélation croisée. Par exemple, Haugh (1976) a développé une procédure dans laquelle des modèles ARMA sont ajustés aux séries. La statistique de test est basée sur les corrélations croisées résiduelles, notées $r_{\hat{u}\hat{v}}(k)$, qui sont définies par

$$r_{\hat{u}\hat{v}}(k) = \frac{\sum_{t=k+1}^n \hat{u}_t \hat{v}_{t-k}}{(\sum_{t=1}^n \hat{u}_t^2 \sum_{t=1}^n \hat{v}_t^2)^{1/2}},$$

pour $|k| \leq n - 1$. La distribution asymptotique d'un nombre fixé de corrélations croisées résiduelles est établie par ce dernier sous l'hypothèse nulle d'indépendance. Il considère alors une statistique de type portmanteau, basée sur la somme d'un nombre fixé de corrélations croisées prises au carré calculées à l'aide des séries résiduelles. Plus précisément, il considère la statistique

$$S_M = n \sum_{k=-M}^M r_{\hat{u}\hat{v}}^2(k),$$

où $M \leq n - 1$ est un entier fixé. La distribution asymptotique est de type khi-carré, et l'on rejette pour de grandes valeurs de la statistique de test. Voir aussi El Himdi et Roy (1997), qui généralisent l'approche de Haugh à un contexte multidimensionnel.

Hong (1996b) a généralisé l'approche de Haugh dans plusieurs directions. La statistique de test est une somme pondérée des carrés des corrélations croisées, c'est-à-dire que la statistique est de la forme

$$Q_n(k) = \frac{n \sum_{j=1-n}^{n-1} k^2(j/m) r_{\hat{u}\hat{v}}^2(j) - M_n(k)}{(2V_n(k))^{1/2}},$$

où $M_n(k) = \sum_{j=1-n}^{n-1} (1 - |j|/n) k^2(j/m)$ et $V_n(k) = \sum_{j=2-n}^{n-2} (1 - |j|/n)(1 - (|j|+1)/n) k^4(j/m)$. La pondération s'effectue à l'aide d'un noyau k , et dans ce sens le test de Haugh peut être considéré comme un cas particulier lorsque l'on choisit le noyau uniforme tronqué. Cependant beaucoup de noyaux offrent une puissance supérieure au noyau uniforme tronqué. Aussi, Hong (1996b) évite l'utilisation de modèles ARMA en ajustant des modèles autorégressifs d'un ordre suffisamment élevé. La statistique de test est sous l'hypothèse nulle asymptotiquement $N(0, 1)$ et le test est unilatéral, rejetant pour les grandes valeurs.

En pratique, la présence de valeurs aberrantes dans les séries chronologiques est un phénomène non-négligeable. Il peut être occasionné par exemple par des erreurs dans l'écriture des données, ou des défaillances temporaires dans les équipements. Les procédures de test

d'indépendance comme celles de Haugh ou de Hong reposent sur des méthodes d'estimation sensibles à la présence de valeurs aberrantes. Le choix de la fonction de corrélation croisée usuelle n'est pas souhaitable en présence de valeurs aberrantes. Des procédures d'estimation robuste dans les modèles ARMA ont été développées par Bustos et Yohai (1986), et une fonction d'autocovariance robustifiée y est proposée. En présence de deux séries chronologiques, Li et Hui (1994) proposent une statistique de test généralisant Haugh (1976), ainsi que les tests considérés par McLeod (1979). Li et Hui (1994) proposent une fonction de corrélation croisée robustifiée et établissent la distribution asymptotique d'un nombre fixé de corrélations croisées résiduelles robustifiées. Des statistiques de test dont la distribution asymptotique est khi-carré sont alors obtenues.

Le but de l'essai 4 est de décrire deux statistiques de test afin de tester la non-corrélation de manière robuste. La première est basée sur la corrélation croisée résiduelle robuste à un délai particulier. Cette approche utilise les résultats généraux obtenus par Li et Hui (1994). La seconde est basée sur une statistique de type portmanteau qui généralise d'une certaine manière au cas robuste la statistique de Hong (1996*b*). La statistique de test est une somme pondérée des carrés des corrélations croisées résiduelles robustes. Nous présumons que les deux séries chronologiques sont ARMA et estimons les paramètres suivant la procédure robuste de Bustos et Yohai (1986). Nous discutons aussi comment adapter les procédures pour tester la causalité au sens de Granger (1969) entre les deux séries, suivant une approche similaire à El Himdi et Roy (1997), en utilisant des résultats de Li et Hui (1994). Les nouveaux tests sont étudiés empiriquement relativement au niveau et à la puissance, avec plusieurs choix des noyaux et de fonctions intervenant dans la définition des corrélations croisées robustifiées. Les résultats suggèrent que des gains peuvent être obtenus en choisissant un noyau autre que le noyau uniforme tronqué. Une application est présentée afin d'illustrer la méthodologie.

Je dédie le premier essai au peuple québécois, qui vivait une crise du verglas alors que je m'étais réfugié au laboratoire pour compléter ce travail qui est devenu mon article sur les estimateurs de calage.

Chapitre 1

ESTIMATEURS DE CALAGE ROBUSTES

Résumé

Nous considérons l'utilisation d'estimateurs de calage en présence de valeurs aberrantes. Une extension de la classe des estimateurs de calage de Deville et Särndal (1992) reposant sur les estimateurs QR de Wright (1983) est obtenue. Elle s'obtient aussi en minimisant une métrique générale sous des contraintes sur les variables de calage et sur les poids. Comme application, cette classe d'estimateurs nous permet de considérer des estimateurs de calage robustes en choisissant judicieusement ses paramètres. Il est ainsi possible, pour des besoins éventuellement cosmétiques, de restreindre les poids robustes à un intervalle spécifié à l'avance. L'utilisation d'estimateurs robustes avec haut point de rupture est considéré. Dans le cas particulier de la métrique quadratique, l'estimateur que nous suggérons est une généralisation d'une proposition de Lee (1991). Une brève étude de simulation illustre la nouvelle méthodologie.

Mots Clés: Estimateur de calage, Estimateur par la régression, Pondération restreinte, Robustesse.

1.1. INTRODUCTION

Le problème des valeurs aberrantes est important dans toutes les branches de la statistique. En théorie des sondages, le contexte est différent de celui de la statistique paramétrique puisque l'on s'intéresse souvent à l'estimation du total d'une variable d'intérêt y . Or une valeur aberrante peut avoir son plein poids dans le total de la population. De plus, le méthodologiste qui procède à la phase d'estimation peut supposer que les valeurs des unités sont enregistrées sans erreur, puisque souvent les unités une fois ramassées sont traitées dans un système d'édition (Särndal et al. 1992, Section 1.7). Cette étape fait partie de la réalisation d'un sondage dans les grandes agences statistiques, comme par exemple à Statistique Canada. Lee (1995) fait un survol des développements de la robustesse en théorie des sondages.

Cependant, puisqu'il arrive fréquemment que les populations dans les enquêtes économiques sont asymétriques, quelques unités risquent d'être *extrêmes* par rapport aux autres,

sujet abordé dans Kish (1965). Retirer complètement ces unités entraîne des estimations biaisées, alors que si on les conserve avec leur plein poids un estimateur comme celui par la régression généralisée (GREG) risque d'être très variable. Ceci suggère un compromis entre le biais et la variance. En présence de valeurs aberrantes, le défi est de proposer des estimateurs *robustes* du total peu affectés par certaines unités très à l'écart des autres. Ces estimateurs devront avoir un petit biais et une petite erreur quadratique moyenne. Traditionnellement, la théorie de l'échantillonnage a consacré beaucoup d'efforts à développer des estimateurs sans biais ou asymptotiquement sans biais (ADU). Voir par exemple Särndal et al. (1992, Section 7.12). Cependant, la propriété ADU n'est peut-être pas souhaitable dans un contexte de valeurs aberrantes. Voir la discussion de Chambers et Kokic (1993) qui montrent qu'un conflit est présent entre la propriété ADU et la robustesse d'un estimateur.

Considérons l'estimateur Horvitz-Thompson (*HT*), défini par $\hat{T}_{yHT} = \sum_s d_k y_k$, où $d_k = \pi_k^{-1}$, π_k étant la probabilité d'inclusion (Si A est un ensemble d'unités, $A \subseteq U$, alors \sum_A est une notation signifiant $\sum_{k \in A}$). Supposons la variable d'intérêt y positive et la population asymétrique. L'estimateur *HT* étant une moyenne pondérée par les d_k , il est vulnérable aux grandes valeurs de y . Une unité avec un poids d_k élevé peut aussi avoir un grand impact sur l'estimation en entraînant des estimations variables. Lee (1995) définit ces unités comme *influentes*. Une unité extrême n'est pas forcément influente si son poids d_k est suffisamment petit. Traditionnellement, les méthodologistes cherchent à limiter l'impact des unités influentes lorsqu'elles sont connues avant le sondage, en donnant par exemple des poids d'échantillonnage proche de un aux unités extrêmes. Gambino (1987) et Lee (1995) discutent cependant de situations où nous ne pouvons pas agir ainsi. Dans un article important, Hidiroglou et Srinath (1981) considèrent l'altération des poids d'échantillonnage en présence de valeurs aberrantes. Leur approche a grandement légitimé la modification des poids dans le monde des sondages.

Parmi les premières alternatives robustes du total, plusieurs reposent sur les *M*-estimateurs et les *GM*-estimateurs. Cependant, beaucoup d'attention a été consacrée récemment à des estimateurs possédant en plus de bonnes propriétés de robustesse globale, telle que mesurée par le point de rupture d'un estimateur. Ces notions sont discutées par exemple dans Donoho et Huber (1983), Hampel et al. (1986) et Rousseeuw et Leroy (1987). Le point de rupture mesure le pourcentage de valeurs aberrantes dans l'échantillon que l'estimateur peut tolérer tout en donnant tout de même une bonne estimation d'une certaine caractéristique de la population. Lee et al. (1992) réclament des estimateurs du total basés sur des estimateurs robustes avec un haut point de rupture.

Nous allons considérer des estimateurs de calage du total T_y s'écrivant comme $\sum_s w_k y_k$. Ces estimateurs sont développés par exemple dans Deville et Särndal (1992). On cherche les poids w_k aussi près que possible des poids d'échantillonnage $d_k = \pi_k^{-1}$ mais en respectant les contraintes d'étalonnage, notée CE (dites aussi contraintes de calage),

$$\sum_s w_k x_k = T_x, \quad (1.1)$$

où x_k est un vecteur de dimension m correspondant à l'information auxiliaire disponible de total $T_x = \sum_U x_k$ connu. Ces estimateurs sont populaires car ils s'interprètent aisément, puisque les méthodologistes sont habitués d'apposer des poids w_k aux unités y_k . Plusieurs métriques sont étudiées afin de mesurer la proximité entre d_k et w_k . L'estimateur GREG est un exemple important avec $w_k = d_k(1 + (T_x - \hat{T}_{xHT})' M_s^{-1} x_k / c_k)$, où $M_s = \sum_s d_k x_k x_k' / c_k$. Il est obtenu en minimisant la métrique quadratique $\sum_s c_k (w_k - d_k)^2 / d_k$. Les constantes c_k sont des facteurs de pondération qui peuvent prendre en compte (par exemple) des problèmes d'hétéroscédasticité. Särndal (1996) discute du choix de ces constantes. Cependant, puisque les poids-g $g_k = w_k / d_k$ du GREG ne sont pas bornés en général, d'autres métriques sont proposées afin de les borner pour qu'ils satisfassent certaines restrictions applicables à la fourchette des valeurs (RAFV). Ceci permet en particulier d'éviter la présence indésirable de poids w_k négatifs. Voir Deville et Särndal (1992), Singh et Mohl (1996) et Stukel et al. (1996).

Les estimateurs de calage ont un lien avec les méthodes robustes comme noté dans Fuller et al. (1994, p. 81). Cependant, il est inexact de prétendre que les estimateurs de calage possèdent nécessairement de bonnes propriétés de robustesse, puisqu'il suffit de se rappeler que tous les estimateurs de calage considérés par Deville et Särndal (1992) sont asymptotiquement équivalents au GREG qui étant ADU n'est pas robuste. De plus, on voit qu'un estimateur traditionnel de calage n'est pas robuste puisqu'il dépend linéairement de w_k , et w_k ne prend pas en compte y_k .

Le but de cet article est de construire des estimateurs de la forme $\sum_s w_k y_k$ où les poids w_k assurent en même temps la robustesse et le respect des contraintes sur les variables de calage et sur les poids w_k . Le point de départ de notre approche est la classe des estimateurs QR de Wright (1983). Supposons que nous disposions des constantes $\{(q_k, r_k), q_k > 0, r_k \geq 0, \forall k \in U\}$, telles que $\sum_U \pi_k q_k x_k x_k' > 0$ et $\sum_s q_k x_k x_k' > 0, \forall s$. (Si A est une matrice symétrique, $A > 0$ signifie que A est définie positive). Les estimateurs QR sont définis à partir de q_k et r_k par la relation

$$\hat{T}_{yQR} = T_x' \hat{B}_q + \sum_s r_k e_k, \quad (1.2)$$

où \hat{B}_q admet une forme pondérée par les q_k

$$\hat{B}_q = \left(\sum_s q_k x_k x_k' \right)^{-1} \sum_s q_k x_k y_k, \quad (1.3)$$

et

$$e_k = y_k - x_k' \hat{B}_q. \quad (1.4)$$

Nous montrons à la section 1.2 que les estimateurs QR sont des estimateurs de calage et introduisons une nouvelle classe d'estimateurs, notée RQR, reposant également sur le choix des constantes q_k et r_k . Elle généralise d'une certaine façon les estimateurs QR ainsi que la classe des estimateurs de Deville et Särndal (1992). L'intérêt de la classe RQR est qu'elle permet d'obtenir des poids w_k restreints à un intervalle donné, disons $[L, U]$. Quelques propriétés des classes QR et RQR sont données à la section 1.2.

La section 1.3 contient des applications de la classe RQR pour la construction d'estimateurs de calage robustes. L'idée principale consiste à modifier des poids implicites robustes pour qu'ils respectent les contraintes de calage. Nous discutons à la section 1.3.1 du choix des constantes q_k et r_k en utilisant un raisonnement propre aux estimateurs de calage. Cette approche s'avère nouvelle et unificatrice, et nous guide à la section 1.3.2 dans le choix des q_k et r_k en présence d'information auxiliaire. Un élément important est l'utilisation d'un estimateur robuste admettant la forme pondérée (1.3), fournissant les q_k . Nous remarquons que c'est le cas pour les *GM*-estimateurs. Les estimateurs avec un haut point de rupture n'ont habituellement pas une forme pondérée. Une repondération de ces estimateurs est considérée, permettant de conserver le point de rupture sous contrôle et d'avoir des estimateurs s'écrivant sous la forme (1.3). Voir Rousseeuw et Leroy (1987) et Simpson et Chang (1997). Nous discutons alors des choix de q_k et r_k permettant de calculer un estimateur RQR afin d'obtenir un estimateur de calage robuste avec des poids bornés. Nous comparons dans la section 1.4 divers estimateurs robustes, dont l'estimateur de Lee (1991) et celui de Chambers (1986), aux estimateurs RQR ainsi qu'au *GREG* et à un estimateur de calage considéré dans Deville et Särndal (1992) dont les poids sont restreints. L'estimateur de Lee (1991) peut être considéré comme un cas particulier de notre approche. Il nous permet aussi de considérer un nouvel estimateur avec des poids bornés. Quatre populations déjà étudiées dans la littérature sont considérées. On notera que les estimateurs sans contrainte sur les poids sont sujets à des problèmes de pondération négative. Les estimateurs de la classe RQR permettent d'obtenir des estimateurs robustes possédant des poids positifs, se comparant bien aux estimateurs sans

contrainte sur les poids. Finalement, nous concluons dans la section 1.5. Notons que l'appendice B contient une liste des abréviations et que l'appendice C contient une liste des diverses constantes contenues dans l'article avec leurs définitions.

1.2. ESTIMATEURS DE LA CLASSE RQR

Soit une population finie $U = \{1, 2, \dots, N\}$ de taille N dont nous désirons estimer le total $T_y = \sum_U y_k$ pour une variable d'intérêt y positive. On sélectionne un échantillon s de taille n_s selon un plan d'échantillonnage $p(s)$. La probabilité d'inclusion d'une unité k est notée π_k et les probabilités d'inclusion d'ordre deux sont notées π_{kl} . Nous supposons que l'information auxiliaire x_k est *unitaire*, c'est-à-dire que x_k est connu de source sûre, $\forall k \in U$.

Wright (1983) introduit la classe des estimateurs QR s'écrivant sous la forme (1.2) avec pour objectif principal d'unifier un grand nombre d'estimateurs familiaux. On retrouve le meilleur estimateur linéaire sans biais (*BLUP*) de Royall (1970) issu de la théorie basée sur les modèles, obtenu en posant $(q_k, r_k) = (1/c_k, 1)$ et le GREG de Cassel, Särndal et Wretman (1976) en considérant le choix $(q_k, r_k) = (d_k/c_k, d_k)$. Alternativement, (1.2) peut s'écrire comme

$$\hat{T}_{yQR} = \sum_s d_k g_k y_k,$$

où $d_k g_k$ satisfait

$$d_k g_k = r_k + (T_x - \hat{T}_{xr})' (\sum_s q_k x_k x_k')^{-1} q_k x_k, \quad (1.5)$$

avec $\hat{T}_{xr} = \sum_s r_k x_k$. Si on pose $r_k = d_k$, g_k correspond au poids-g du GREG.

Les estimateurs QR sont des estimateurs de calage, obtenus de la minimisation de la métrique quadratique sujet aux CE

$$\min \frac{1}{2} \sum_s (w_k - r_k)^2 / q_k, \text{ tel que } \sum_s w_k x_k = T_x. \quad (1.6)$$

Les poids w_k sont choisis aussi près que possible des r_k et les q_k sont des facteurs de pondération. Autrement formulé, les poids de départ r_k sont transformés en poids calés w_k . La solution au problème (1.6) est $w_k = d_k g_k$, où $d_k g_k$ est donné dans la formule (1.5).

Cependant, rien ne garantit que les poids w_k de l'estimateur QR soient positifs, ce qui peut être indésirable pour le praticien. Voir Brewer (1994), qui formalise l'interprétation des poids. Afin de restreindre les poids w_k dans $[L, U]$, nous considérons de résoudre

$$\min \sum_s G(w_k; q_k, r_k), \text{ tel que } \sum_s w_k x_k = T_x \text{ et } w_k \in [L, U]. \quad (1.7)$$

L'estimateur de calage du total est

$$\hat{T}_{yRQR} = \sum_s w_k y_k, \quad (1.8)$$

où les w_k sont obtenus de la résolution du problème (1.7). La fonction $G(w; q, r)$ est supposée strictement convexe et dérivable en w pour r et q fixés. Nous notons $g(u; q, r) = G'(u; q, r)$ et $h(u; q, r) = g^{-1}(u; q, r)$. On suppose de plus que $h(0; q, r) = r$ et $h'(0; q, r) = q$. Nous appelons les estimateurs obtenus les estimateurs de calage restreints QR (RQR).

Fuller et al. (1994) privilégient des estimateurs de régression possédant des propriétés raisonnables d'invariance. On peut montrer que les estimateurs RQR sont équivariants de régression et d'échelle lorsque les constantes q_k et r_k sont invariantes aux transformations. Voir pour les définitions utiles Bolfarine et Zacks (1992)¹.

L'existence d'une solution au problème (1.7) n'est pas garantie. Voir l'étude de simulation de Stukel et al. (1996). Il peut arriver par exemple des réalisations de l'échantillon pour lequel on ne peut même pas satisfaire les CE (1.1). Ainsi, l'échantillon est tellement débalancé qu'il est impossible que la somme pondérée des composantes de chaque dimension puisse donner le total de la population correspondant. Le seul recours du praticien est alors de relaxer les contraintes en réduisant la dimension du nombre de variables auxiliaires. Voir aussi la discussion de Fuller et al. (1994). Dans le cas des estimateurs de calage considérés dans Deville et Särndal (1992), les auteurs montrent dans leur résultat 1 qu'une solution existe avec probabilité tendant vers un. Sous certaines conditions, ce résultat peut être adapté aux estimateurs de la classe RQR.

1. Les estimateurs RQR sont équivariants de régression et d'échelle lorsque les constantes q_k et r_k sont invariantes aux transformations. Considérons un estimateur du total \hat{t} . Un estimateur \hat{t} possède l'équivariance de régression si

$$\hat{t}\{(x_k, y_k + x'_k \gamma), \forall k \in s\} = \hat{t}\{(x_k, y_k), \forall k \in s\} + t'_x \gamma,$$

alors que \hat{t} est équivariant d'échelle si

$$\hat{t}\{(x_k, c y_k), \forall k \in s\} = c \hat{t}\{(x_k, y_k), \forall k \in s\}.$$

Lorsque $x_k \equiv 1$, l'équivariance de régression se réduit à l'équivariance de position. Notons que l'estimateur HT n'est pas équivariant de position. Cependant, la forme de Hájek, c'est-à-dire $C_s(d) \sum_s d_k y_k$, où $C_s(d) = N / \sum_s d_k$ possède cette propriété. On montre qu'un estimateur RQR est équivariant de régression en notant les relations suivantes

$$\begin{aligned} \hat{t}_{RQR}\{(x_k, y_k + x'_k \gamma), \forall k \in s\} &= \sum_s w_k (y_k + x'_k \gamma) \\ &= \sum_s w_k y_k + \sum_s w_k x'_k \gamma \\ &= \hat{t}_{RQR}\{(x_k, y_k), \forall k \in s\} + t'_x \gamma, \end{aligned}$$

et l'équivariance d'échelle se montre de la même manière.

La métrique qui retiendra notre attention afin que les poids satisfassent les RAFV est une légère modification du cas 7 de Deville et Särndal (1992). Nous l'appelons la métrique quadratique restreinte. La fonction G correspondant au choix de cette métrique est

$$G(w_k; q_k, r_k) = \begin{cases} \frac{1}{2}(w_k - r_k)^2/q_k & \text{si } w_k \in [L, U], \\ \infty & \text{sinon,} \end{cases}$$

alors que la fonction h est

$$h(x'_k \lambda; q_k, r_k) = \begin{cases} L & r_k + q_k x'_k \lambda < L, \\ r_k + q_k x'_k \lambda & r_k + q_k x'_k \lambda \in [L, U], \\ U & r_k + q_k x'_k \lambda > U. \end{cases}$$

Avec cette modification, c'est le poids w_k qui est contraint et non pas seulement w_k/d_k comme dans le cas 7 de Deville et Särndal (1992). Dans notre situation, w_k peut "corriger" un poids initial qui est aberrant. Notons que telle que formulée la métrique de Deville et Särndal (1992) insère subtilement les contraintes sur les w_k dans la fonction G . Pour calculer l'estimateur (1.8) selon cette métrique, il suffit de respecter la même démarche que Deville et Särndal (1992), ce qui nous amène à résoudre avec la méthode de Newton l'équation suivante en λ

$$\sum_s h(x'_k \lambda; q_k, r_k) x_k = T_x. \quad (1.9)$$

L'estimateur final est $\hat{T}_{yRQR} = \sum_s h(x'_k \lambda_\infty; q_k, r_k) y_k$ où λ_∞ est la solution de l'équation (1.9)².

Il est d'intérêt de savoir si la contrainte des poids change les propriétés de l'estimateur comparativement à l'estimateur QR sans contrainte de poids. Le résultat suivant (dont la preuve est en appendice) montre sous certaines conditions que les deux estimateurs sont asymptotiquement équivalents. En pratique, avec la métrique quadratique restreinte, nous n'avons pas observé d'écarts importants.

Proposition 1. *Sous les hypothèses C_1 et C_2 données en appendice,*

$$N^{-1} |\hat{T}_{yQR} - \hat{T}_{yRQR}| = o_p(n^{-1/2}). \quad (1.10)$$

2. Notons que résoudre l'équation 1.9 avec la méthode de Newton revient à calculer

$$\lambda_{\nu+1} = \lambda_\nu - \left(\sum_s h'(x'_k \lambda_\nu; q_k, r_k) x_k x'_k \right)^{-1} \left(\sum_s h(x'_k \lambda_\nu; q_k, r_k) x_k - T_x \right).$$

Il est bien connu que si la valeur initiale λ_0 est suffisamment près de la solution, alors la méthode de Newton converge à un taux quadratique. Notons que l'estimateur QR correspond à la première itération de la méthode de Newton en choisissant $\lambda_0 = 0$.

Ce résultat s'obtient possiblement de la démarche pour obtenir le résultat 5 dans Deville et Särndal (1992) portant sur l'équivalence asymptotique entre le GREG et les estimateurs de calage considérés par ces auteurs. Cependant, la proposition 1 peut être utile pour comprendre le type de conditions nécessaires pour avoir le résultat annoncé dans notre contexte.

Puisque (1.2) a le même comportement asymptotique que la quantité $T'_x B_q + \sum_s w_k E_k$, où $E_k = y_k - x'_k B_q$ et $B_q = (\sum_U \pi_k q_k x_k x'_k)^{-1} \sum_U \pi_k q_k x_k y_k$, ceci suggère pour estimateur de variance

$$v_L = \sum \sum_s \check{\Delta}_{kl} (w_k e_k)(w_l e_l), \quad (1.11)$$

où $\Delta_{kl} = \pi_{kl} - \pi_k \pi_l$, $\check{\Delta}_{kl} = \Delta_{kl} / \pi_{kl}$ et e_k est donné dans (1.4). Voir Särndal et al. (1989) et Särndal et al. (1992, p. 234). On peut montrer que le biais asymptotique d'un estimateur QR est donné sous des conditions générales par

$$E_p(\hat{T}_{yQR} - T_y) = \sum_U (\pi_k r_k - 1) E_k.$$

Un estimateur possible du biais est alors $b = \sum_s d_k (\pi_k r_k - 1) e_k$, que l'on peut utiliser conjointement avec la formule (1.11) pour construire un estimateur de l'erreur quadratique moyenne pour un estimateur QR et RQR, en utilisant la proposition 1.

Les estimateurs RQR permettent d'obtenir des estimateurs de calage avec des poids contraints. Avec q_k et r_k déterminées, il suffit de procéder à la résolution du problème (1.7). Les sections qui suivent sont des applications de la classe RQR dans un contexte de robustesse. Nous verrons comment devrait s'orienter le choix des constantes q_k et r_k , choisies en pratique avec l'aide de l'échantillon s .

1.3. CONSTRUCTION D'ESTIMATEURS ROBUSTES ET CALÉS

1.3.1. Méthodes reposant sur la réduction de poids et la modification des valeurs

Lee (1995) discute de diverses propositions reposant sur la méthode de réduction de poids dans un tirage aléatoire simple. Une fois les observations aberrantes détectées, ces méthodes consistent à réduire les poids des observations extrêmes. Ces méthodes sont préférables à celles qui retirent complètement les observations douteuses puisque toutes les observations dans l'échantillon sont légitimes, comme discuté dans Lee et al. (1992).

Traduit dans le langage des estimateurs de calage, commençons par considérer la situation où nous ne disposons pas d'information auxiliaire et où la seule contrainte est $\sum_s w_k = N$. Ce cas sert à motiver notre démarche. Considérons l'estimateur QR avec $q_k = r_k$. Pour les fins de

notre discussion, on considère les constantes r_k connues et fixées. Les poids minimisant (1.6) sujet à $\sum_s w_k = N$ sont $w_k = C_s(r)r_k$, où $C_s(r) = N/\sum_s r_k$, de sorte que \hat{T}_{yQR} devient

$$\hat{T}_{yQR} = C_s(r) \sum_s r_k y_k. \quad (1.12)$$

Lorsqu'une observation est extrême, elle représente potentiellement peu d'unités comme elle dans la population et on devrait alors peut-être réduire son poids. Afin de satisfaire les CE, ceci suggère de trouver des poids w_k les plus près possible des poids d'échantillonnage d_k pour les unités qui ne sont pas aberrantes mais aussi près que possible d'un facteur de réduction r pour les unités aberrantes, où r est choisi par le statisticien. Plus précisément, notons $s = s_1 \cup s_2$, où s_1 de cardinalité n_1 représente les unités qui ne sont pas déclarées aberrantes, alors que $s_2 = s - s_1$ de cardinalité $n_2 = n - n_1$ représente les unités aberrantes de s . Le facteur de réduction r satisfera typiquement $r \leq d_k, \forall k \in s_2$. Par exemple, considérons l'estimateur (1.12) avec $q_k = r_k = B_k = d_k I_{k1} + r(1 - I_{k1})$, où I_{k1} est la variable indicatrice de l'appartenance à s_1 . Ainsi les constantes q_k et r_k sont réduites pour les unités de s_2 afin de refléter que les unités de s_2 sont extrêmes. L'estimateur (1.12) devient

$$\hat{T}_{yQR} = C_s(B) \left(\sum_{s_1} d_k y_k + r \sum_{s_2} y_k \right).$$

Dans le cas du tirage aléatoire simple, $d_k = N/n$ et on obtient

$$\hat{T}_{yQR} = C_s(B) \hat{T}_{yB},$$

où $\hat{T}_{yB} = N/n \sum_{s_1} y_k + r \sum_{s_2} y_k$ est l'estimateur de Bershada (1960) discuté dans Lee (1995). D'autres méthodes reposant sur la méthode de réduction de poids sont discutées dans Lee (1995), qui discute également du choix de r .

Un désavantage des méthodes reposant sur la réduction de poids est que l'analyste doit lui-même identifier les unités aberrantes. Les méthodes reposant sur la modification des valeurs permettent d'éviter cette difficulté en réduisant graduellement le poids des unités plus extrêmes. Considérons un tirage aléatoire simple. Posons

$$m(y_k; t, a, b) = b + (a - b) \min(1, t/y_k). \quad (1.13)$$

Ainsi, cette fonction attribue un poids de départ valant a pour les $y_k < t$, et graduellement réduit jusqu'à un poids final b , à mesure que y_k devient extrême. La valeur t est appelée le seuil. Les constantes a , b et t sont choisies par le statisticien. Plusieurs valeurs pour a et b ont été considérées dans la littérature. Ainsi, plutôt que d'attribuer un facteur de réduction fixe

aux unités de s_2 , choisissons $q_k = r_k = W_k = m(y_k; t, N/n, fN/n)$, où f est une constante comprise entre 0 et 1. L'estimateur (1.12) devient

$$\begin{aligned}\hat{T}_{yQR} &= C_s(W) \sum_s W_k y_k \\ &= C_s(W) \hat{T}_{yW}.\end{aligned}$$

L'estimateur \hat{T}_{yW} est discuté dans Gross, Bode, Taylor et Lloyd-Smith (1986) ainsi que dans Chambers et Kokic (1993), qui l'appellent l'estimateur winsorisé. Il est un cas particulier de l'approche de Chambers (1982, 1986). Lorsque $f = 0$, l'estimateur (1.12) devient $\hat{T}_{yQR} = C_s(W_I) \hat{T}_{yW_I}$ avec $q_k = r_k = W_{Ik} = m(y_k; t, N/n, 0)$, où $\hat{T}_{yW_I} = N/n(\sum_{s_1} y_k + n_2 t)$, s_1 dénotant la partie de s contenant les unités satisfaisant $y_k < t$. L'estimateur \hat{T}_{yW_I} est discuté dans Lee (1995), ainsi que dans Gross et al. (1986) qui l'appellent l'estimateur winsorisé de type I. Lorsque $f = n/N$, Gross et al. (1986) l'appellent l'estimateur winsorisé de type II. Il est également discuté dans Bruce (1991).

Dans un plan πps , Dalén (1987) a inséré le plan en posant $D_k = m(y_k; \pi_k t, d_k, 1)$. Ainsi, si k et l sont deux observations extrêmes telles que $y_k = y_l$, alors celle dont le poids d'échantillonnage est le plus grand aura un poids D_k plus élevé. Le choix $r_k = q_k = D_k$ permet d'obtenir essentiellement l'estimateur de Dalén, $\hat{T}_{yQR} = C_s(D) \sum_s D_k y_k$. L'estimateur $\hat{T}_{yD} = \sum_s D_k y_k$ a été étudié notamment dans Tambay (1988).

TABLE 1.1. *Estimateur (1.12) reposant sur la réduction de poids et la modification des valeurs.*

Estimateur	Valeurs de $q_k = r_k$
Bershad	$B_k = d_k I_{k1} + r(1 - I_{k1})$
Winsorisé	$W_k = m(y_k; t, N/n, fN/n)$
Winsorisé, type I	$W_{Ik} = m(y_k; t, N/n, 0)$
Winsorisé, type II	$W_{IIk} = m(y_k; t, N/n, 1)$
Dalén	$D_k = m(y_k; \pi_k t, d_k, 1)$

Note: $m(y_k; t, a, b) = b + (a - b) \min(1, t/y_k)$.

La démarche de cette section suggère que nous pouvons à l'occasion vouloir des estimateurs dont les poids sont près des r_k plutôt que des poids d'échantillonnage d_k . Les constantes r_k seront choisies elle-mêmes près des d_k pour les bonnes unités, mais réduites lorsqu'une unité est déclarée extrême. Les estimateurs QR unifient les méthodes de réduction de poids et de modification des valeurs. Les méthodes reposant sur la modification des valeurs nous amènent à choisir les poids de manière adaptative à l'échantillon particulier s choisi. Comme

noté dans Chambers et Kokic (1993) cela n'est pas surprenant puisque le problème des valeurs aberrantes survient après la sélection de l'échantillon s . On doit utiliser l'échantillon à notre disposition pour tenter de le régler. Dans la prochaine section nous généralisons ces méthodes en présence d'information auxiliaire.

1.3.2. Estimateurs du total basés sur des statistiques robustes

Parmi les premières tentatives vers l'obtention d'alternatives robustes pour le total de la population en présence d'information auxiliaire on retrouve Chambers (1982, 1986) qui propose un estimateur par le ratio robuste basé sur une décomposition de l'estimateur *BLUP*. Une extension récente des travaux de Chambers se retrouve dans Welsh et Ronchetti (1998). Gwet et Rivest (1992) proposent également une version robuste de l'estimateur par le ratio en basant leur approche sur le plan d'échantillonnage dans un tirage aléatoire simple. Rivest et Rouillard (1991) font une étude comparative de plusieurs estimateurs robustes et étudient plusieurs estimateurs de l'erreur quadratique moyenne. Dans les plans avec probabilités inégales, Hulliger (1995) considère une robustification de l'estimateur *HT* lorsque les probabilités d'inclusion sont obtenues avec l'information auxiliaire. Gwet et Rivest (1992) et Hulliger (1995) considèrent une version de la fonction d'influence pour populations finies, insistant sur le besoin de procédures possédant de bonnes propriétés de robustesse locales et l'utilisation d'estimateurs ayant des fonctions d'influence bornées. Les fonctions d'influence sont discutées dans un contexte général dans Hampel et al. (1986).

Nous allons voir dans les sections suivantes comment construire des estimateurs robustes avec des poids contraints. La construction de ces estimateurs repose sur les étapes suivantes:

- Identification des constantes q_k et r_k . Ceci nous donne un estimateur QR.
- Résolution du problème (1.7) afin d'obtenir un estimateur RQR.

D'un point de vue de la robustesse, les coefficients q_k sont choisis de telle sorte que \hat{B}_q est un estimateur robuste. De cette manière, la première partie de l'estimateur QR, $T'_x \hat{B}_q$, donne une bonne valeur prédite pour toute la population. La deuxième partie de l'estimateur QR, $\sum_s r_k e_k$, corrige la première partie pour les y_k observés dans l'échantillon. Les constantes r_k évitent qu'avec cette correction les valeurs aberrantes dans l'échantillon reviennent avec plein poids.

1.3.2.1. Choix de q_k basé sur un GM -estimateur

Considérons l'estimateur (1.2) dans lequel on remplace \hat{B}_g par un estimateur robuste d'un coefficient de régression. Ces estimateurs sont discutés par exemple dans Huber (1981) ou Hampel et al. (1986). On obtient ainsi

$$\hat{T}_g = T'_x \hat{B}_g + \sum_s r_k(y_k - x'_k \hat{B}_g). \quad (1.14)$$

L'estimateur (1.14) n'est pas sous la forme des estimateurs QR à moins que \hat{B}_g admette une forme pondérée. C'est le cas si \hat{B}_g est un GM -estimateur défini par l'équation

$$\sum_s d_k h_k x_k \psi((y_k - x'_k B)/(\sigma h_k^\alpha \sqrt{c_k}))/\sqrt{c_k} = 0, \quad (1.15)$$

puisque la solution de (1.15) peut s'exprimer comme

$$\hat{B}_g = \left(\sum_s d_k h_k^{1-\alpha} u_k x_k x'_k / c_k \right)^{-1} \sum_s d_k h_k^{1-\alpha} u_k x_k y_k / c_k,$$

où

$$u_k = \frac{\psi((y_k - x'_k \hat{B}_g)/(\sigma h_k^\alpha \sqrt{c_k}))}{(y_k - x'_k \hat{B}_g)/(\sigma h_k^\alpha \sqrt{c_k})}.$$

Les propriétés des GM -estimateurs sont discutées dans Simpson et Chang (1997). On suppose que σ est connu afin de simplifier notre discussion, et le rôle de c_k est le même que dans le cas du GREG. La fonction ψ est déterminée par l'analyste. Un exemple d'un usage courant est la fonction de Huber

$$\psi_{Hub}(x; c) = \begin{cases} c & \text{si } x > c, \\ x & \text{si } |x| \leq c, \\ -c & \text{si } x < -c. \end{cases} \quad (1.16)$$

Une valeur de c autour de 2 est souvent utilisée dans le calcul des GM -estimateurs. Voir par exemple Hampel et al. (1986), Gwet et Rivest (1992) et Hulliger (1995).

Le choix de h_k permet de borner l'influence d'une information auxiliaire trop extrême. La constante $\alpha = 0$ donne lieu au choix de Mallows alors que $\alpha = 1$ permet d'obtenir la version de Schweppe. La version de Schweppe est parfois privilégiée. Voir Coakley et Hettmansperger (1993) et Hampel et al. (1986, p. 322). En présence d'information auxiliaire minimale, c'est-à-dire lorsque nous disposons que d'une variable réelle $x_k, \forall k \in U$, un choix possible de fonction h_k est

$$h_k = \min\left(1, \frac{t}{x_k / \text{med}(x_k)}\right). \quad (1.17)$$

Dans un plan πps , une modification de h_k inspirée de Dalén (1987) pour tenir compte des différents poids d'échantillonnage serait peut-être souhaitable. La constante t doit être spécifiée par le statisticien. Une valeur de t autour de 1.5 est rencontrée dans les applications. Voir par exemple Rivest et Rouillard (1991) qui donnent aussi d'autres choix de fonctions h_k .

L'écriture de \hat{B}_g comme un estimateur pondéré permet d'écrire l'estimateur (1.14) comme un estimateur QR avec

$$(q_k, r_k) = (d_k h_k^{1-\alpha} u_k / c_k, r_k).$$

Nous discuterons du choix des constantes r_k dans la section 1.3.2.3.

1.3.2.2. Choix de q_k basé sur un estimateur avec haut point de rupture

Le choix d'un *GM*-estimateur n'est qu'un premier pas afin d'obtenir un estimateur très robuste du total. En effet, bien que la fonction d'influence des *GM*-estimateurs soit bornée, il n'en demeure pas moins que ces estimateurs ne possèdent pas un haut point de rupture, qui décroît habituellement avec la dimension de l'information auxiliaire (Rousseeuw et Leroy 1987, p. 13). Nous verrons dans cette section comment construire des estimateurs de calage robustes basés sur des estimateurs avec un haut point de rupture. Comme ces estimateurs ne possèdent habituellement pas la forme pondérée, nous considérons une repondération de ces estimateurs. Ceci nous permettra de dégager, comme dans la section précédente, les constantes q_k nécessaires au calcul de la métrique de l'estimateur RQR. Plus précisément, on considère les poids \hat{u}_k suivants

$$\hat{u}_k = \frac{\psi((y_k - x_k' \hat{B}_0) / (\sigma h_k^\alpha \sqrt{c_k}))}{(y_k - x_k' \hat{B}_0) / (\sigma h_k^\alpha \sqrt{c_k})}, \quad (1.18)$$

où \hat{B}_0 est un estimateur équivariant avec un haut point de rupture satisfaisant certaines conditions de régularité. L'estimateur repondéré est

$$\hat{B}_r = \left(\sum_s d_k h_k^{1-\alpha} \hat{u}_k x_k x_k' / c_k \right)^{-1} \sum_s d_k h_k^{1-\alpha} \hat{u}_k x_k y_k / c_k. \quad (1.19)$$

Les propriétés asymptotiques de ce type d'estimateurs sont étudiées dans Simpson et Chang (1997).

L'estimateur \hat{B}_0 que l'on considère est le *GM*-estimateur en une étape de Coakley et Hettmansperger (1993). Cet estimateur possède un haut point de rupture. Il s'obtient comme la première itération de la formule de Newton dans l'équation (1.15), où l'on adopte la version

de Schweppe en posant $\alpha = 1^3$. D'autres choix d'estimateurs robustes auraient pu être utilisés. Cependant, l'efficacité jumelée avec les propriétés de robustesse de l'estimateur de Coakley et Hettmansperger (1993) en font un bon choix. Ainsi, la constante q_k que l'on propose est

$$q_k = d_k h_k^{1-\alpha} \hat{u}_k / c_k,$$

avec $\hat{B}_0 = \hat{B}_{CH}$, \hat{B}_{CH} désignant l'estimateur de Coakley et Hettmansperger (1993).

1.3.2.3. Choix de r_k

Ayant déterminé les constantes q_k , nous devons maintenant choisir les constantes r_k . Si $d_k = r_k$ alors l'estimateur QR est sous des conditions générales un estimateur ADU. Cependant ce choix de r_k donne un estimateur sensible aux valeurs aberrantes. Alternativement le choix $r_k = 0$ procure un estimateur robuste qui risque d'être très biaisé comme souligné dans Gwet et Rivest (1992, p. 1180). Lee (1991) suggère le choix $r_k = \theta d_k$, où $\theta \in [0, 1]$. Le biais asymptotique devient sous des conditions générales $(\theta - 1) \sum_U E_k$, où E_k représente les résidus obtenus en ajustant un estimateur robuste sur toute la population. Le choix de θ permet de contrôler le biais de l'estimateur. La discussion de la section 1.3 nous amène à suggérer des constantes r_k proches des d_k pour les bonnes unités et graduellement réduites pour les observations suspectes. Nous suggérons le choix

$$r_k = d_k u_k^*, \quad (1.20)$$

où

$$u_k^* = \frac{\psi^*((y_k - x'_k \hat{B}_r) / (\sigma h_k^\alpha \sqrt{c_k}))}{(y_k - x'_k \hat{B}_r) / (\sigma h_k^\alpha \sqrt{c_k})}.$$

La fonction ψ^* que nous allons considérer est une modification de la fonction de Huber

$$\psi^*(x) = \begin{cases} x & \text{si } |x| \leq a, \\ a \operatorname{sign}(x) & \text{si } |x| > a \text{ et } |x| < a/b, \\ bx & \text{si } |x| > a/b. \end{cases} \quad (1.21)$$

3. Il s'obtient de la formule

$$\hat{B}_{CH} = \hat{\beta}_0 + \left(\sum_s d_k x_k x'_k \psi' \left(\frac{y_k - x'_k \hat{\beta}_0}{\sigma h_k \sqrt{c_k}} \right) / c_k \right)^{-1} \sum_s d_k u_{0k} x_k (y_k - x'_k \hat{\beta}_0) / c_k,$$

où ψ' est la dérivée première de ψ et

$$u_{0k} = \frac{\psi((y_k - x'_k \hat{\beta}_0) / (\sigma h_k \sqrt{c_k}))}{(y_k - x'_k \hat{\beta}_0) / (\sigma h_k \sqrt{c_k})},$$

et $\hat{\beta}_0$ est l'estimateur par les moindres résidus tronqués (LTS) qui est disponible dans le logiciel S-PLUS (Statistical Sciences 1991).

Nous choisissons $a = 9$, $b = 1/4$. La raison de cette modification est que nous ne voulons pas que les valeurs aberrantes possédant de gros résidus ou une information auxiliaire extrême aient des poids trop réduits. Ainsi le poids d'échantillonnage est pleinement conservé lorsque l'argument de u_k^* est entre -9 et 9, et graduellement réduit jusqu'au quart. Si on réduit trop le poids des grands résidus alors le biais devient trop grand, ce qui motive le choix de ψ^* . Le choix des constantes r_k s'est effectué de manière empirique et semble bien fonctionner en pratique.

Ainsi, on se propose de considérer le choix des constantes q_k et r_k suivant

$$(q_k, r_k) = (d_k h_k^{1-\alpha} \hat{u}_k / c_k, d_k u_k^*). \quad (1.22)$$

Notre suggestion est une généralisation de la proposition de Lee (1991) puisque plutôt que de considérer $r_k = d_k \theta$ où θ est fixe, $r_k = d_k u_k^*$ s'adaptera de manière automatique (ou adaptative) avec l'échantillon. Ayant ce choix des constantes q_k et r_k à notre disposition et avec la métrique quadratique usuelle, nous obtenons un estimateur QR. Il est cependant sujet à des problèmes de pondération négative. Cependant avec les constantes (1.22) nous pouvons considérer la métrique quadratique restreinte, résoudre le problème (1.7) et obtenir un estimateur robuste satisfaisant les CE et les RAFV.

Le comportement asymptotique de l'estimateur RQR résultant comparativement à l'estimateur QR sans contrainte de poids trouve des éléments de réponse dans la proposition 1. Cependant, puisque les constantes (1.22) dépendent de s de façon compliquée, ceci ne nous permet pas de conclure automatiquement à l'équivalence asymptotique. Cependant, l'étude de simulation de la section 1.4 semble suggérer un comportement très comparable entre l'estimateur avec et sans contrainte sur les poids, en ce qui a trait à l'erreur quadratique moyenne de Monte-Carlo. Ainsi, l'évidence empirique montre que si les q_k et r_k sont choisies de telle sorte que l'estimateur sans contrainte sur les poids est robuste, alors la version avec des contraintes de poids sera aussi robuste.

Finalement, résumons les étapes de la méthode proposée afin d'obtenir un estimateur RQR robuste.

- i. Choix des constantes q_k et r_k . Nous suggérons les constantes que l'on retrouve à l'équation (1.22). Cette étape nécessite le calcul de \hat{B}_{CH} .
- ii. Choix de la métrique. Si besoin est, choix des constantes L et U . On choisit ces constantes telles que $L \leq r_k \leq U$, $\forall k \in s$.
- iii. Résolution avec la méthode de Newton de l'équation (1.9).

- iv. Poser $w_k = h(x_k' \lambda_\infty; q_k, r_k)$ pour λ_∞ solution de l'étape 3.
- v. Poser $\hat{T}_{yT} = \sum_s w_k y_k$, qui est l'estimateur RQR que l'on propose.

La procédure nécessite un certain nombre de constantes. On retrouve les constantes α , t et c dans le calcul de q_k et r_k . Le choix de ces valeurs est cependant justifié avec l'aide de la théorie de la robustesse, ce qui permet de guider le praticien. Par exemple, la valeur du c de la fonction de Huber peut être obtenue en tenant compte de préoccupations d'efficacité sous erreurs normales. Voir Hampel et al. (1986, p. 333) et Gwet et Rivest (1992). On retrouve aussi les constantes a et b , qui sont plus directement reliées aux estimateurs proposés. La constante b représente la réduction maximale de poids que l'on est prêt à accorder dans la spécification des poids implicites r_k , et a dans ce sens un lien avec la suggestion de Lee (1991). La constante la plus importante à spécifier est possiblement la valeur de a . Nous suggérons ici $a = 9$. Cependant, dans nos simulations, une valeur de a entre 6 et 12 donne des résultats relativement comparables. Le choix des bornes L et U repose sur des considérations cosmétiques, afin d'avoir des poids restreints à un intervalle. Cette dernière considération est peut-être secondaire pour le praticien. Il semble donc que l'aspect le plus important est de choisir une valeur de r_k proche de d_k pour les bonnes valeurs, et réduite à mesure qu'une observation est extrême, et c'est ce but qui a motivé le choix de r_k de cette section. Cependant il serait utile d'avoir un choix de r_k satisfaisant un certain critère d'optimalité.

1.3.3. Estimateur modéliste de Chambers

Une autre approche est basée sur une décomposition proposée par Chambers (1982, 1986) que nous appliquons maintenant aux estimateurs QR. Remarquons que nous pouvons toujours écrire un estimateur QR sous la forme

$$\hat{T}_{yQR} = \sum_s r_k y_k + (T_x - \hat{T}_{xr})' B + \sum_s z_k \sqrt{q_k} (y_k - x_k' B),$$

où $z_k = (T_x - \hat{T}_{xr})' (\sum_s q_k x_k x_k')^{-1} x_k \sqrt{q_k}$, $\hat{T}_{xr} = \sum_s r_k x_k$ et B est arbitraire. Chambers (1986) avait considéré le cas particulier $(q_k, r_k) = (1/\sigma_k^2, 1)$ pour l'estimateur par le ratio. Afin de limiter l'influence des unités aberrantes, Chambers propose

$$\hat{T}_{yCHAM} = \sum_s r_k y_k + (T_x - \hat{T}_{xr})' B + \sum_s z_k \psi_p(\sqrt{q_k} (y_k - x_k' B)). \quad (1.23)$$

La fonction ψ_p permet de limiter l'influence des gros résidus. Le choix de B est un estimateur robuste, comme par exemple \hat{B}_g . Une fonction ψ_p considérée dans Chambers (1986) est

$$\psi_p(t) = t \exp(-0.25(|t| - 6)^2). \quad (1.24)$$

Il est intéressant de noter que (1.23) peut s'écrire comme

$$\hat{T}_{yCHAM} = T'_x B + \sum_s (r_k + (d_k g_k - r_k) \lambda_k) e_k(B),$$

où $e_k(B) = y_k - x'_k B$, g_k est défini dans la formule (1.5) calculé à l'aide de q_k et r_k , et

$$\lambda_k = \frac{\psi_p(\sqrt{q_k}(y_k - x'_k B))}{\sqrt{q_k}(y_k - x'_k B)}. \quad (1.25)$$

Ainsi, les résidus $e_k(B)$ sont pondérés par une relation rappelant la formule (1.13). Si $\lambda_k \equiv 1$, alors $d_k g_k$ est appliqué aux résidus $e_k(B)$ et on vérifie facilement que nous retrouvons l'estimateur \hat{T}_{yQR} . Alternativement, si $\lambda_k \equiv 0$, nous obtenons (1.14) si nous posons $B = \hat{B}_g$. Si dans (1.23) on pose $(q_k, r_k) = (1/c_k, 1)$ et $B = \hat{B}_g$, alors l'estimateur de Chambers représente un compromis entre le *BLUP* et un estimateur robuste basé sur un *GM*-estimateur. Remarquons que (1.23) est formellement un estimateur QR avec

$$(q_k^*, r_k^*) = (d_k h_k^{1-\alpha} u_k / c_k, r_k + (d_k g_k - r_k) \lambda_k).$$

Cependant, puisque r_k^* n'est pas forcément positif, il n'est pas toujours possible de procéder à un changement de métrique dans ce cas.

1.4. ÉTUDE EMPIRIQUE

Afin d'étudier la performance des estimateurs de calage robustes, nous avons entrepris une étude de simulation de Monte Carlo. Nous avons considéré quatre populations composées de données provenant d'ouvrages facilement accessibles en théorie de l'échantillonnage. Pour chacune des populations, $K = 2000$ échantillons ont été tirés par tirage aléatoire simple pour différentes tailles échantillonnales. Notre objectif principal est de savoir s'il est possible d'obtenir des estimateurs possédant de bonnes propriétés empiriques (biais, erreur quadratique moyenne) mais satisfaisant aussi les CE et les RAFV. Notons que tous les programmes ont été écrit en S-PLUS (Statistical Sciences 1991) et sont disponibles en communiquant avec l'auteur.

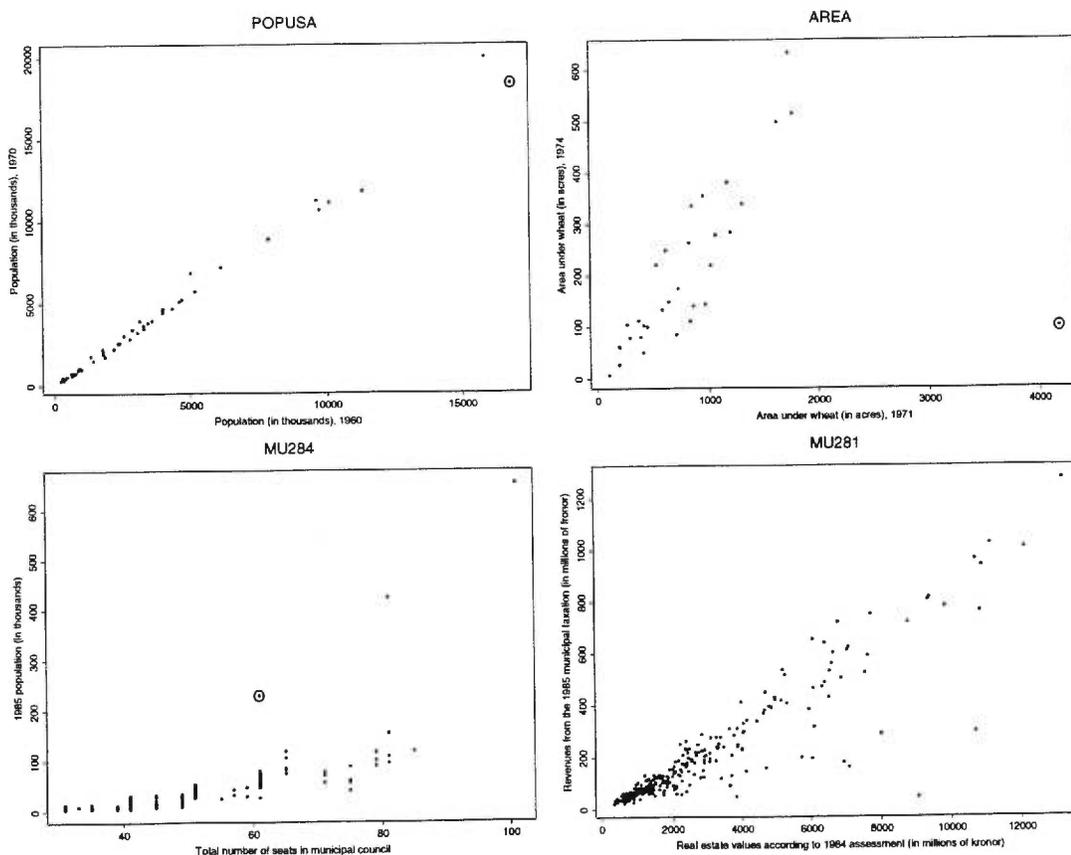
TABLE 1.2. *Totaux des diverses populations et totaux connus de l'information auxiliaire.*

Population	T_x	T_y	N
POPUSA	179972	203923	51
AREA	29118	6781	34
MU284	13500	8339	284
MU281	757246	53124	281

1.4.1. Populations à l'étude

Les graphiques des populations se retrouvent dans la figure (1.1). La première population possède 51 unités et se retrouve dans Mosteller et Tukey (1977, p. 560). Elle consiste en la population américaine en 1960 et en 1970 pour chacun des 50 états et du district fédéral de Columbia. Nous la notons *POPUSA*. Du nuage de points de la population de 1970 en fonction de celle de 1960, nous remarquons que toutes les unités semblent être sur la même droite, avec la présence de quelques bons points de levier. Un exemple d'un bon point de levier est le point entouré dans cette population. La seconde population possède 34 unités et se retrouve dans Singh et Chaudhary (1986, p. 177). Il est question de superficie de champs ensemencés en 1971 et en 1974. Nous appellons cette population *AREA*. Un mauvais point de levier (voir le point entouré) est présent dans cette population puisque le point (4170, 99) ne respecte pas la tendance linéaire de la majorité des unités. Des échantillons de taille 10 et

FIG. 1.1. Les quatre populations à l'étude.



15 sont tirés dans *POPUSA* et *AREA*. La troisième population est la population *MU284* de Särndal et al. (1992) qui porte sur les 284 municipalités de Suède. Nous considérons les variables $x = S82$ concernant le nombre total de sièges dans le conseil municipal et $y = P85$ la population suédoise en 1985. On note la présence d'aberrants verticaux (voir par exemple le point entouré) et d'un mauvais point de levier. Finalement, nous considérons la population *MU281* consistant de *MU284* de laquelle les trois plus grosses municipalités ont été exclues. Les variables considérées sont $x = REV84$ qui représente les valeurs des biens fonciers suivant l'évaluation de 1984 et $y = RMT85$ qui représente les revenus des taxes municipales en 1985. L'unité de mesure est le million de kronors pour ces deux variables. Plusieurs mauvais points de levier semblent présents dans cette population. Des échantillons de taille $n = 30$ et $n = 60$ sont tirés dans *MU284* et *MU281*. Le tableau (1.2) contient les totaux des diverses populations.

1.4.2. Description des estimateurs

Les deux estimateurs de base sont le GREG ainsi que l'estimateur obtenu en considérant le cas 7 de Deville et Särndal (1992) qui est un GREG avec des poids bornés. Nous notons ces estimateurs *GREG/U* et *GREG/R* respectivement. Nous avons choisi $c_k \equiv 1$ pour les populations *POPUSA* et *AREA*, alors que nous avons optés pour $c_k = x_k$ pour les populations *MU284* et *MU281*. Notre choix des c_k est motivé par la relation existant entre ces constantes et l'hétéroscédasticité du modèle de superpopulation. Parmi les estimateurs robustes, nous avons étudié l'estimateur de Chambers (1986) en considérant

$$(q_k^*, r_k^*) = (d_k \hat{u}_k(\hat{B}_{CH}) / c_k, 1 + (d_k g_k - 1) \lambda_k(\hat{B}_\tau)),$$

où dans la formule (1.5) $(q_k, r_k) = (1/c_k, 1)$, noté *CHAM*, basé sur \hat{B}_τ . Les constantes $\hat{u}_k(\hat{B}_{CH})$ sont obtenus de la formule (1.18). Le choix $\alpha = 1$ a été adopté tout au long de la simulation. La fonction ψ de Huber a été utilisé avec la constante $c = 1.345$ pour \hat{B}_τ . Les fonctions h_k sont celles données par la formule (1.17), où nous avons choisis $t = 1.46$. La fonction λ_k est définie par l'équation (1.25). La fonction ψ_p considérée est celle donnée dans l'équation (1.24). L'échelle a été estimée comme dans Coakley et Hettmansperger (1993). Nous avons également considéré l'estimateur modéliste *BLUP* dans lequel nous avons remplacé l'estimateur par les moindres carrés généralisés par l'estimateur \hat{B}_τ , que nous appelons *MODEL*. De plus, l'estimateur de Lee (1991) a été considéré basé sur \hat{B}_τ où $r_k = 0.25d_k$, en utilisant la métrique quadratique. Nous étudions également une extension de l'estimateur de Lee (1991) en considérant la métrique quadratique restreinte. Nous notons ces estimateurs

LEE25/U et *LEE25/R* respectivement. Finalement, nous avons considéré la nouvelle méthode de la section 1.3.2.3 avec le choix (q_k, r_k) donné dans l'équation (1.22) selon la métrique quadratique et la métrique quadratique restreinte. Nous les notons *QRROB/U* et *QRROB/R* respectivement. Le choix de la fonction ψ^* est donné dans la formule (1.21).

1.4.3. Mesures fréquentistes

Pour chaque échantillon, les huit estimateurs de la section (1.4.2) sont calculés. Les résultats sont dans les tableaux 1.3, 1.4, 1.5 et 1.6. Puisqu'un attrait des nouvelles méthodes concerne les RAFV, des statistiques sont calculées sur ces poids. Les colonnes *MIN* et *MAX* des tableaux de résultats contiennent les valeurs minimales et maximales des poids calculées lors de la simulation et ce pour chaque estimateur. Nous rapportons aussi le pourcentage d'échantillons pour lesquels les poids sont à l'intérieur des RAFV dans la colonne RAFV des tableaux de résultats. Nous avons également considéré le pourcentage d'échantillons pour lesquels les estimateurs restreints ont convergé dans la colonne *CONV*. Les intervalles adoptés $[L, U]$ pour les intervalles restreints sont précisés dans les divers tableaux. Dans tous les cas, les diverses statistiques ont été calculées sur les échantillons pour lesquels tous les estimateurs ont convergé.

Un autre aspect important est relié au biais et à l'efficacité des méthodes proposées. Soit \hat{T} un estimateur du total T_y . Posons \hat{T}_i l'estimateur du total calculé à l'aide de l'échantillon i , $i = 1, \dots, K$. Le biais relatif de Monte Carlo BR_M , la valeur moyenne E_M et la variance V_M sont donnés par les formules usuelles, à savoir $BR_M = (E_M(\hat{T}) - T_y)/T_y \times 100$, $E_M(\hat{T}) = \frac{1}{K} \sum_{i=1}^K \hat{T}_i$ et $V_M = \frac{1}{K} \sum_{i=1}^K (\hat{T}_i - E_M(\hat{T}))^2$. Notre critère principal d'efficacité sera l'erreur quadratique moyenne de Monte Carlo, définie par $MSE_M = \frac{1}{K} \sum_{i=1}^K (\hat{T}_i - T_y)^2$. Les coefficients de variation CV_M sont calculés selon $\sqrt{MSE_M}/T_y$. Les variances et erreurs quadratiques moyennes sont exprimées en millions. Les coefficients de variation, les biais relatifs, les RAFV et la convergence des versions restreintes sont exprimés en pourcentage.

1.4.4. Discussion

La population *POPUSA* ne comportait pas de valeurs aberrantes ne respectant pas le modèle linéaire. Lors de l'échantillonnage, les coefficients de variation des estimateurs étaient petits, ce qui était prévisible compte tenu de l'allure de la population. Les colonnes *MSE* et *VAR* sont très semblables, confirmant que le biais n'est pas un problème dans cette population. Tous les biais relatifs sont inférieurs à 1%. L'estimateur *QRROB/U* offre une réduction de

TAB. 1.3. Résultats de Monte-Carlo lors de l'échantillonnage dans la population POPUSA.

ESTIMATEURS	VAR_M	MSE_M	CV_M	BR_M	MIN	MAX	$RAFV^1$	$CONV$
$n = 10$								
<i>GREG/U</i>	34.90	34.92	2.90	-0.07	-6.24	26.75	86.7	
<i>GREG/R</i>	35.29	35.30	2.91	-0.04	0.20	32.00	100.0	98.4
<i>CHAM</i>	32.43	33.75	2.85	-0.56	-19.61	40.96	84.0	
<i>MODEL</i>	27.66	30.69	2.72	-0.85	-19.71	40.86	82.8	
<i>LEE25/U</i>	27.48	30.07	2.69	-0.79	-19.38	39.64	83.2	
<i>LEE25/R</i>	28.67	30.90	2.73	-0.73	0.20	32.00	100.0	98.4
<i>QRROB/U</i>	27.40	28.40	2.61	-0.49	-15.68	40.10	83.2	
<i>QRROB/R</i>	28.33	29.18	2.65	-0.45	0.20	32.00	100.0	98.4
$n = 15$								
<i>GREG/U</i>	21.90	21.95	2.30	-0.10	-3.13	15.32	94.7	
<i>GREG/R</i>	22.12	22.15	2.31	-0.09	0.20	16.00	100.0	99.5
<i>CHAM</i>	18.11	20.14	2.20	-0.70	-5.79	16.44	92.4	
<i>MODEL</i>	15.43	19.03	2.14	-0.93	-6.09	16.92	91.0	
<i>LEE25/U</i>	15.44	19.54	2.17	-0.99	-6.19	17.06	90.8	
<i>LEE25/R</i>	15.72	19.68	2.18	-0.98	0.20	16.00	100.0	99.5
<i>QRROB/U</i>	14.68	16.44	1.99	-0.65	-4.48	16.41	90.9	
<i>QRROB/R</i>	14.85	16.56	2.00	-0.64	0.20	16.00	100.0	99.5

¹ Les bornes pour les RAFV sont [0.20, 32] pour $n = 10$ et [0.20, 16] pour $n = 15$.

variance par rapport à *GREG/U* de plus de 21% lorsque $n = 10$, et de plus de 30% lorsque $n = 15$.

La population *AREA* est de petite taille. Elle comportait un mauvais point de levier qui a occasionné des biais relatifs empiriques très élevés pour tous les estimateurs. L'estimateur *GREG/U* affiche un biais relatif de plus de 7% malgré un échantillonnage de 44% dans cette population. Les estimateurs robustes affichent les biais les plus importants, mais relativement comparables au biais de l'estimateur *GREG/U*. La réduction de variance la plus importante a été enregistrée par l'estimateur *QRROB/U*, mais au prix d'un biais relatif autour de 10%.

La population *MU284* contient un aberrant vertical et des mauvais points de levier. Les estimateurs robustes réduisent de manière radicale la variance, puisqu'ils ne sont pas affectés

TAB. 1.4. Résultats de Monte-Carlo lors de l'échantillonnage dans la population AREA.

ESTIMATEURS	VAR_M	MSE_M	CV_M	BR_M	MIN	MAX	$RAFV^1$	$CONV$
$n = 10$								
<i>GREG/U</i>	1.334	1.700	19.23	8.92	-3.35	14.94	86.6	
<i>GREG/R</i>	1.295	1.629	18.82	8.53	0.20	14.00	100.0	99.0
<i>CHAM</i>	1.187	1.541	18.30	8.77	-4.09	14.90	87.2	
<i>MODEL</i>	1.291	1.580	18.54	7.93	-5.23	16.75	86.8	
<i>LEE25/U</i>	1.279	1.593	18.61	8.26	-5.28	16.89	86.6	
<i>LEE25/R</i>	1.284	1.596	18.63	8.24	0.20	14.00	100.0	99.0
<i>QRROB/U</i>	1.026	1.440	17.70	9.50	-4.74	15.38	87.6	
<i>QRROB/R</i>	1.028	1.437	17.68	9.43	0.20	14.00	100.0	99.0
$n = 15$								
<i>GREG/U</i>	0.940	1.178	16.00	7.18	-1.40	7.03	93.0	
<i>GREG/R</i>	0.928	1.154	15.85	7.01	0.20	6.00	100.0	99.8
<i>CHAM</i>	0.708	0.989	14.67	7.82	-1.52	7.92	93.7	
<i>MODEL</i>	0.757	0.997	14.73	7.22	-1.66	8.39	93.1	
<i>LEE25/U</i>	0.672	1.059	15.18	9.18	-1.68	9.40	92.0	
<i>LEE25/R</i>	0.671	1.056	15.15	9.15	0.20	6.00	100.0	99.8
<i>QRROB/U</i>	0.485	0.990	14.68	10.48	-1.59	8.90	93.9	
<i>QRROB/R</i>	0.485	0.986	14.64	10.44	0.20	6.00	100.0	99.8

¹ Les bornes pour les RAFV sont [0.20, 14] pour $n = 10$ et [0.20, 6] pour $n = 15$.

par les trois unités extrêmes en y qui s'éloignent de façon marquée de la tendance linéaire. Les estimateurs *CHAM*, *QRROB/R* et *QRROB/U* sont plus que quatre fois moins variables que l'estimateur *GREG/U*. Cependant cela occasionne un biais négatif qui est beaucoup plus élevé. Tous les estimateurs robustes sont sévèrement biaisés. L'estimateur *MODEL* affiche un biais négatif de plus de 13% alors que *QRROB/U* possède un biais négatif de l'ordre de 11%. Dans le cas de *QRROB*, un meilleur choix des constantes dans la fonction ψ^* pourrait réduire une plus grande part du biais au prix d'une moins grande réduction de variance. L'augmentation de la taille échantillonnale à $n = 60$ a permis de réduire le biais sous les 10% pour les estimateurs *CHAM* et *QRROB*, mais les autres estimateurs robustes restent plus biaisés.

TAB. 1.5. Résultats de Monte-Carlo lors de l'échantillonnage dans la population MU284.

ESTIMATEURS	VAR_M	MSE_M	CV_M	BR_M	MIN	MAX	RAFV ¹	CONV
<i>n</i> = 30								
<i>GREG/U</i>	2.833	2.925	20.51	-3.64	-6.83	23.30	89.8	
<i>GREG/R</i>	2.813	2.910	20.46	-3.73	0.20	16.00	100.0	99.2
<i>CHAM</i>	0.645	1.639	15.35	-11.95	-11.80	31.26	77.0	
<i>MODEL</i>	0.709	2.037	17.11	-13.82	-12.06	31.91	68.4	
<i>LEE25/U</i>	0.887	1.877	16.43	-11.93	-11.06	30.93	73.5	
<i>LEE25/R</i>	0.871	1.847	16.30	-11.85	0.20	26.00	100.0	99.2
<i>QRROB/U</i>	0.719	1.532	14.84	-10.81	-9.46	25.84	86.5	
<i>QRROB/R</i>	0.720	1.525	14.81	-10.76	0.20	16.00	100.0	99.2
<i>n</i> = 60								
<i>GREG/U</i>	1.473	1.489	14.63	-1.49	-1.19	10.03	90.1	
<i>GREG/R</i>	1.467	1.484	14.61	-1.57	0.20	7.00	100.0	99.7
<i>CHAM</i>	0.357	0.990	11.93	-9.54	-2.53	15.59	69.8	
<i>MODEL</i>	0.380	1.255	13.43	-11.22	-4.93	14.52	58.1	
<i>LEE25/U</i>	0.403	1.201	13.14	-10.72	-4.80	14.20	60.3	
<i>LEE25/R</i>	0.396	1.203	13.16	-10.78	0.20	7.00	100.0	99.7
<i>QRROB/U</i>	0.308	0.976	11.85	-9.80	-2.36	10.99	86.1	
<i>QRROB/R</i>	0.308	0.979	11.87	-9.82	0.20	7.00	100.0	99.7

¹ Les bornes pour les RAFV sont [0.20, 16] pour *n* = 30 et [0.20, 7] pour *n* = 60.

La population MU281 contient un nombre assez grand de mauvais points de levier. La variance domine la part du *MSE* dans cette population. L'estimateur *LEE25* est le moins variable, avec une réduction de plus de 35% comparativement à *GREG/U*. Cependant, bien que $\theta = 0.25$ fonctionne bien dans cette population, notre étude illustre bien que ce n'est pas toujours le meilleur choix.

On remarque que tous les estimateurs robustes sont plus efficaces que le GREG ou sa version restreinte. Comme confirmé par les résultats de Deville et Särndal (1992), la version restreinte du GREG se comporte essentiellement de la même manière que le GREG, tant au niveau du biais que de la variance Monte Carlo dans chacune des populations. Parmi tous les estimateurs considérés, *GREG/U* et *GREG/R* sont les estimateurs les moins biaisés.

TAB. 1.6. Résultats de Monte-Carlo lors de l'échantillonnage dans la population MU281.

ESTIMATEURS	VAR_M	MSE_M	CV_M	BR_M	MIN	MAX	$RAFV^1$	$CONV$
$n = 30$								
<i>GREG/U</i>	17.33	17.35	7.84	-0.26	-38.97	34.56	86.0	
<i>GREG/R</i>	17.40	17.41	7.86	-0.24	0.20	25.00	100.0	99.8
<i>CHAM</i>	13.23	13.26	6.86	-0.33	-47.09	39.08	56.9	
<i>MODEL</i>	11.30	11.91	6.50	1.47	-66.22	41.43	47.9	
<i>LEE25/U</i>	11.21	11.60	6.41	1.17	-59.75	37.03	53.3	
<i>LEE25/R</i>	11.26	11.73	6.45	1.29	0.20	25.00	100.0	99.8
<i>QRROB/U</i>	12.92	13.29	6.86	1.15	-54.14	39.73	70.8	
<i>QRROB/R</i>	12.94	13.34	6.88	1.20	0.20	25.00	100.0	99.8
$n = 60$								
<i>GREG/U</i>	7.57	7.57	5.18	-0.10	-12.77	15.34	86.4	
<i>GREG/R</i>	7.58	7.58	5.18	-0.09	0.20	9.00	100.0	99.9
<i>CHAM</i>	5.85	5.90	4.57	-0.43	-22.97	11.49	51.4	
<i>MODEL</i>	4.53	5.23	4.30	1.57	-24.02	14.58	38.7	
<i>LEE25/U</i>	4.55	5.18	4.28	1.49	-23.74	14.41	41.2	
<i>LEE25/R</i>	4.50	5.21	4.30	1.58	0.20	9.00	100.0	99.9
<i>QRROB/U</i>	5.40	6.16	4.67	1.64	-21.08	21.07	68.6	
<i>QRROB/R</i>	5.39	6.17	4.67	1.66	0.20	9.00	100.0	99.9

¹ Les bornes pour les RAFV sont [0.20, 25] pour $n = 30$ et [0.20, 9] pour $n = 60$.

Les versions robustes présentent toutes un biais plus important. Cependant la réduction de variance compense plus que largement de telle sorte que l'efficacité des estimateurs robustes est toujours plus grande que *GREG/U* ou *GREG/R*.

Concernant les restrictions sur les poids, on remarque que les estimateurs *GREG/U*, *CHAM*, *MODEL*, *LEE25* et *QRROB/U* sont tous sujets aux problèmes de pondération négative comme en fait foi la colonne *MIN*. Les estimateurs restreints évitent ce problème. La colonne *RAFV* montre que les restrictions ne furent pas satisfaites relativement fréquemment, dépendant de la population et de la taille de l'échantillon, allant de 5% à 60%. Les deux estimateurs robustes restreints se comportent en général de façon comparable à leur

version non-restreinte. De plus, $QRROB/R$, en plus de satisfaire les RAFV offre des propriétés d'efficacité intéressantes, lorsque comparé aux autres estimateurs robustes. Les versions restreintes ont été moins sujettes aux problèmes de convergence lorsque les tailles échantillonnelles étaient plus grandes. Notons que nous avons dû prendre des bandes plus larges dans le cas de *POPUSA* afin d'obtenir des taux de convergence satisfaisants.

1.5. CONCLUSION

Le but de cet article est d'introduire des estimateurs de calage possédant de bonnes propriétés de robustesse. Les estimateurs traditionnels de calage sont faciles d'emploi, puisqu'il suffit d'un ensemble de poids de départ, habituellement les poids d'échantillonnage d_k , qui sont transformés en poids calés. La démarche de cet article suit ces mêmes étapes, où les poids implicites robustes r_k sont transformés en poids calés, et où les constantes q_k sont choisies de sorte que \hat{B}_q soit un estimateur robuste. Le choix de r_k proposé est donné par la formule (1.20), avec $a = 9$, $b = 1/4$. Une théorie pour un choix de r_k optimal reste à faire. Il est suggéré dans les applications de varier la constante a , entre par exemple 6 et 12, pour voir l'influence de la constante sur l'estimation. Si désiré, les bornes L et U permettent de restreindre les poids, afin qu'ils soient tous positifs par exemple. Nous suggérons d'utiliser de manière générale $L = 0.2$, $U = kN/n$, où k est autour de 3.

Notons que le but des estimateurs de calage robustes n'est pas de remplacer le GREG, mais bien d'être utilisé conjointement avec ce dernier. Ainsi, si l'estimateur robuste et l'estimateur GREG sont très différents, alors une analyse plus poussée pourrait permettre d'expliquer le phénomène. Les estimateurs proposés sont potentiellement utiles comme outils de diagnostic.

Il serait intéressant de pousser les études empiriques de la section 1.4, en étudiant par exemple l'effet du plan d'échantillonnage sur les procédures proposées. L'estimation de la variance est un autre sujet important qui nécessite des développements. Une autre avenue intéressante concerne les sondages "tout usage" (multipurpose). En effet, dans les applications, nous disposons rarement que d'une seule variable d'intérêt. Les méthodologistes désirent alors disposer d'un unique ensemble de poids utile pour toutes les variables d'intérêt. Dans un contexte de robustesse, une solution a été proposée dans la conclusion de l'article de Gwet et Rivest (1992), où des poids robustes sont calculés pour chaque variable d'intérêt $y^{(i)}$, $i = 1, \dots, I$. Pour une unité, le poids final correspond au poids minimal parmi les poids obtenus. Alternativement, afin d'obtenir des estimateurs robustes et calés, nous pourrions

calculer les poids implicites robustes pour chaque variable d'intérêt, fournissant un ensemble de $r_k(y^{(i)})$, et poser $r_k = \min r_k(y^{(i)})$, où le minimum est sur $i = 1, \dots, I$. On pourrait par la suite transformer ces poids en poids calés. Cette procédure nécessite d'être évaluée plus en détail.

REMERCIEMENTS

Je tiens à remercier Carl-Erik Särndal pour m'avoir initié à la théorie de l'échantillonnage et pour m'avoir suggéré de considérer le problème des valeurs aberrantes en théorie des sondages. Je remercie Roch Roy et Christian Léger pour m'avoir aidé à diverses étapes dans la réalisation de cet article. Je remercie sincèrement le rédacteur associé, le rédacteur adjoint et deux arbitres anonymes pour leurs commentaires constructifs qui ont permis d'améliorer substantiellement le contenu et la présentation de cet article.

A. PREUVE DE LA PROPOSITION 1

Soit $\Delta_h(u; q, r) = r + qu - h(u; q, r)$ et z_k une variable d'intérêt. On suppose les conditions suivantes

$$C_1. N^{-1} \sum_s q_k z_k = O_p(1).$$

$$C_2. N^{-1} \sum_s \Delta_h(x'_k \lambda_\infty; q_k, r_k) z_k = o_p(n^{-1/2}), \text{ où } \lambda_\infty \text{ est solution de l'équation (1.9).}$$

On remarque que $\sum_s (r_k + q_k x'_k \lambda_1) x_k = T_x$, où $\lambda_1 = -(\sum_s q_k x_k x'_k)^{-1} (\hat{T}_{xr} - T_x)$, et aussi que $\sum_s h(x'_k \lambda_\infty; q_k, r_k) x_k = T_x$. Ainsi, on trouve en utilisant C_2 que

$$N^{-1} \sum_s q_k x_k x'_k (\lambda_1 - \lambda_\infty) = o_p(n^{-1/2}),$$

et donc $\lambda_1 - \lambda_\infty = o_p(n^{-1/2})$, en utilisant C_1 avec $z_k = x_k x'_k$. On montre aussi aisément que

$$\begin{aligned} N^{-1} (\hat{T}_{QR} - \hat{T}_{RQR}) &= N^{-1} \sum_s (r_k + q_k x'_k \lambda_1) y_k - N^{-1} \sum_s h(x'_k \lambda_\infty; q_k, r_k) y_k \\ &= N^{-1} \sum_s q_k x'_k y_k (\lambda_1 - \lambda_\infty) + N^{-1} \sum_s \Delta_h(x'_k \lambda_\infty; q_k, r_k) y_k \\ &= o_p(n^{-1/2}). \end{aligned}$$

B. LISTE DES ABRÉVIATIONS

- ADU: Absence de biais asymptotique (Asymptotically Design Unbiased).
- BLUP: Meilleur estimateur linéaire sans biais de Royall (1970).
- CE: Contraintes d'étalonnage, $\sum_s w_k x_k = T_x$, où $T_x = \sum_U x_k$.
- CH: Estimateur robuste proposé par Coakley et Hettmansperger (1993) qui un GM-estimateur en une étape. Il est robuste et efficace.
- CHAM: Estimateur robuste de Chambers (1982, 1986).
- GM: Estimateurs M généralisés, issus de la théorie de la robustesse. Voir par exemple Hampel et al. (1986).
- GREG: Estimateur par la régression généralisée proposé par Cassel et al. (1976).
- HT: Estimateur Horvitz-Thompson $\sum_s d_k y_k$, où $d_k = \pi_k^{-1}$.
- QR: Estimateurs de Wright (1983), de la forme $T'_x \hat{B}_q + \sum_s r_k e_k$.
- RAFV: Restrictions applicables à la fourchette des valeurs pour les poids w_k , en imposant par exemple que tous les $w_k \in [L, U]$.
- RQR: Généralisation des estimateurs de Wright (1983), qui s'obtiennent d'une métrique générale en plus de restrictions sur les poids.

C. LISTE DES PRINCIPALES CONSTANTES

- c_k : Facteur pouvant tenir compte de problèmes d'hétéroscédasticité.
- d_k : Poids d'échantillonnage.
- g_k : Poids-g défini par w_k/d_k .
- h_k : Quantité permettant de réduire l'influence d'une information auxiliaire aberrante dans \hat{B}_q .
- π_k, π_{kl} : Probabilités d'inclusion d'ordre un et deux, respectivement.
- q_k, r_k : Quantités définissant un estimateur QR. Les q_k servent à construire les coefficients de régression intervenant dans la première partie, $T'_x \hat{B}_q$. Les r_k servent pour la seconde partie, $\sum_s r_k e_k$.
- u_k, \hat{u}_k : Poids servant à construire \hat{B}_q de manière robuste.
- u_k^* : Poids servant à considérer un facteur de correction $\sum_s r_k e_k$ robuste.
- w_k : Poids calé à être apposé à y_k pour former $\sum_s w_k y_k$.

BIBLIOGRAPHIE

- Bershad, M. A. (1960), Some observations on outliers, Unpublished Memorandum, Statistical Research Division, U.S. Bureau of Census.
- Bolfarine, H. et Zacks, S. (1992), *Prediction Theory for Finite Population*, New-York: Springer-Verlag.
- Brewer, K. R. W. (1994), 'Survey sampling inference: Some past perspectives and present prospects', *Pakistan Journal of Statistics* **10**, 213–233.
- Bruce, A. G. (1991), 'Robust estimation and diagnostics for repeated sample surveys'. Mathematical Statistics Working Paper 1991/1, Wellington, Statistics New Zealand.
- Cassel, C. M., Särndal, C. E. et Wretman, J. H. (1976), 'Some results on generalized difference estimation and generalized regression estimation for finite population', *Biometrika* **63**, 615–620.
- Chambers, R. L. (1982), Robust Finite Population Estimation, PhD thesis, Johns Hopkins University, Dept of Biostatistics.
- Chambers, R. L. (1986), 'Outlier robust finite population estimation', *Journal of the American Statistical Association* **81**, 1063–1069.
- Chambers, R. L. et Kokic, P. N. (1993), An integrated approach for the treatment of outliers in sub-annual surveys, in 'Proceedings on the 49th Session, International Statistical Institute'.
- Coakley, C. W. et Hettmansperger, T. P. (1993), 'A bounded influence, high breakdown, efficient regression estimator', *Journal of the American Statistical Association* **88**, 872–880.
- Dalén, J. (1987), 'Practical estimators of a population total which reduce the impact of large observations'. R & D Report, Stockholm: Statistics Sweden.
- Deville, J.-C. et Särndal, C. E. (1992), 'Calibration estimators in survey sampling', *Journal of the American Statistical Association* **87**, 376–382.
- Donoho, D. L. et Huber, P. J. (1983), The notion of breakdown point, in P. J. Bickel, K. A. Doksum et J. L. Hodges, eds, 'A Festschrift for Erich Lehmann', Belmont, CA: Wadsworth.
- Fuller, W. A., Loughin, M. M. et Baker, H. D. (1994), 'Regression weighting in the presence of nonresponse with application to the 1987-1988 nationwide food consumption survey', *Survey Methodology* **20**, 75–85.
- Gambino, J. (1987), Dealing with outliers: A look at some methods used at statistics canada, Technical report, Business Survey Methods Division, Statistics Canada. Technical Report prepared for the Fifth Meeting of the Advisory Committee on Statistical Methods.
- Gross, W. F., Bode, G., Taylor, J. M. et Lloyd-Smith, C. W. (1986), Some finite population estimators which reduce the contribution of outliers, in 'Proceedings of the Pacific Statistical Congress'. Auckland, New Zealand, 20-24 May 1985.

- Gwet, J. P. et Rivest, L.-P. (1992), 'Outlier resistant alternatives to the ratio estimator', *Journal of the American Statistical Association* **87**, 1174-1182.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. et Stahel, W. A. (1986), *Robust Statistics: The Approach Based on Influence Functions*, New-York: Wiley.
- Hidiroglou, M. A. et Srinath, K. P. (1981), 'Some estimators of the population total from simple random samples containing large units', *Journal of the American Statistical Association* **76**, 690-695.
- Huber, P. J. (1981), *Robust Statistics*, New-York: Wiley.
- Hulliger, B. (1995), 'Outlier robust Horvitz-Thompson estimators', *Survey Methodology* **21**, 79-87.
- Kish, L. (1965), *Survey Sampling*, New-York: Wiley.
- Lee, H. (1991), Model-based estimators that are robust to outliers, in 'Proceedings of the 1991 Annual Research Conference'. U. S. Bureau of Census.
- Lee, H. (1995), Outliers in business surveys, in Cox, Binder, Chinnappa, Christiansono, Colledge et Kott, eds, 'Business Surveys Methods', New-York: Wiley.
- Lee, H., Ghangurbe, P. D., Mach, L. et Yung, W. (1992), 'Outliers in sample surveys'. Methodology Branch Working Paper BSMD-92-008E, Ottawa: Statistics Canada.
- Mosteller, F. et Tukey, J. W. (1977), *Data Analysis and Regression, A Second Course in Statistics*, Addison-Wesley Publishing Company.
- Rivest, L. P. et Rouillard, E. (1991), M-estimators and outlier resistant alternatives to the ratio estimator, in 'Proceedings of the 1990 Symposium of Statistics Canada', pp. 271-285.
- Rousseeuw, P. J. et Leroy, A. M. (1987), *Robust Regression and Outlier Detection*, New-York: Wiley.
- Royall, R. M. (1970), 'On finite population sampling under certain linear regression models', *Biometrika* **57**, 377-387.
- Särndal, C. E. (1996), 'Efficient estimators with simple variance in unequal probability sampling', *Journal of the American Statistical Association* **91**, 1289-1300.
- Särndal, C. E., Swensson, B. et Wretman, J. H. (1989), 'The weighted residual technique for estimating the variance of the general regression estimator of the finite population total', *Biometrika* **76**, 527-537.
- Särndal, C. E., Swensson, B. et Wretman, J. H. (1992), *Model Assisted Survey Sampling*, New-York: Springer-Verlag.
- Simpson, D. G. et Chang, Y.-C. I. (1997), 'Reweighted approximate GM-estimators: asymptotics and residual-based graphics', *Journal of Statistical Planning and Inference* **57**, 273-293.
- Singh, A. C. et Mohl, C. A. (1996), 'Understanding calibration estimators in survey sampling', *Survey Methodology* **22**, 107-115.
- Singh, D. et Chaudhary, F. S. (1986), *Theory and Analysis of Sample Survey Designs*, India: Wiley.
- Statistical Sciences, Inc. (1991), *S-PLUS Reference Manual*. Seattle: Statistical Science, Inc.
- Stukel, D. M., Hidiroglou, M. A. et Särndal, C. E. (1996), 'Variance estimation for calibration estimators: A comparison of jackknifing versus Taylor linearization', *Survey Methodology* **22**, 117-125.
- Tambay, J.-L. (1988), An integrated approach for the treatment of outliers in sub-annual surveys, in 'Proceedings on the Survey Research Methods Section, American Statistical Association', pp. 229-234.

- Welsh, A. H. et Ronchetti, E. (1998), 'Bias-calibrated estimation from sample surveys containing outliers', *Journal of the Royal Statistical Society, Series B* **60**, 413–428.
- Wright, R. L. (1983), 'Finite population sampling with multivariate auxiliary information', *Journal of the American Statistical Association* **78**, 879–884.

Annexe A

COMPLÉMENT À “ESTIMATEURS DE CALAGE ROBUSTES”

A.1. ESTIMATION DE L'ERREUR QUADRATIQUE MOYENNE

L'estimation de la variance peut s'établir en remarquant qu'un estimateur RQR peut toujours s'écrire comme

$$\hat{T}_{yRQR} = T'_x \beta + \sum_s w_k E_k,$$

où $E_k = y_k - x'_k \beta$, et β arbitraire. Ceci suggère qu'un *prototype* pour l'estimateur de la variance est

$$v_L = \sum \sum_s \check{\Delta}_{kl} (w_k E_k) (w_l E_l), \quad (\text{A.1})$$

où $\Delta_{kl} = \pi_{kl} - \pi_k \pi_l$, $\check{\Delta}_{kl} = \Delta_{kl} / \pi_{kl}$. Ainsi, la formule A.1 serait pleinement justifiée si les poids w_k étaient indépendants de s , ce qui n'est pas le cas. Cependant, si nous oublions un instant que les poids dépendent de l'échantillon et appliquons les résultats usuels sur l'estimation de la variance des estimateurs Horvitz-Thompson, nous obtenons la formule A.1. Cette manière de construire un estimateur de variance est empruntée à Särndal et al. (1992, p. 237). En pratique, nous pouvons considérer de remplacer les E_k par les résidus de l'estimateur QR .

Nous allons maintenant nous attarder à la formule A.1 dans certains contextes particuliers. Nous allons considérer les cas où $r_k = d_k$, $r_k = 0$ et $r_k = 1$.

Lorsque nous considérons l'estimateur QR avec $(q_k, r_k) = (d_k/c_k, d_k)$, on obtient les poids

$$w_k = d_k g_k,$$

où

$$g_k = 1 + (T_x - \hat{T}_{HT})' \left(\sum_s d_k x_k x'_k / c_k \right)^{-1} x_k / c_k,$$

est le poids-g usuel de l'estimateur GREG. Ainsi, la formule A.1 se réduit alors à la proposition de Särndal et al. (1989) pour estimer la variance de l'estimateur GREG. Notons également

que dans le cas de l'estimateur RQR avec ces mêmes constantes (q_k, r_k) et pour les métriques et les estimateurs de calage considérés dans Deville et Särndal (1992), l'estimateur de variance donné par A.1 se réduit à une proposition de Deville et Särndal (1992) pour estimer la variance de ces estimateurs.

Lorsque nous considérons un estimateur QR avec $(q_k, r_k) = (d_k/c_k, 0)$, les poids w_k deviennent

$$w_k = T'_x \left(\sum_s d_k x_k x'_k / c_k \right)^{-1} d_k x_k / c_k,$$

de telle manière que la formule A.1 devient

$$v_L = T'_x M_s^{-1} \sum \sum_s \check{\Delta}_{kl} (d_k x_k E_k / c_k) (d_l x'_l E_l / c_l) M_s^{-1} T_x$$

ce qui correspond à estimer la variance de $\hat{T}_{yQR} = T'_x \hat{\beta}_q$ par $\hat{V}(T'_x \hat{\beta}_q) = T'_x \hat{V}(\hat{\beta}_q) T_x$ avec l'aide d'un estimateur de variance d'un vecteur correspondant à des coefficients de régression, comme décrit dans Särndal et al. (1992, section 5.10.2, pp.192-197).

Discutons maintenant le cas $r_k = 1$ pour l'estimateur QR lorsque x_k est une quantité scalaire. Nous allons considérer les cas où $q_k = 1/x_k$ et $q_k = (1 - \pi_k)/(\pi_k x_k)$. Lorsque $(q_k, r_k) = (1/x_k, 1)$, l'estimateur QR se réduit à $\hat{T}_{yQR} = T_x \sum_s y_k / \sum_s x_k$ qui n'est rien d'autre que l'estimateur par le ratio utilisé dans un tirage aléatoire simple. Voir Särndal et al. (1992, section 7.3.1). Dans cette situation $w_k = T_x / \sum_s x_k$ et la formule A.1 devient dans un tirage aléatoire simple

$$v_L = N^2 \left(\frac{1}{n} - \frac{1}{N} \right) \left(\frac{\bar{x}_U}{\bar{x}_s} \right)^2 \frac{\sum_s e_k^2}{n-1},$$

où $e_k = y_k - \hat{\beta} x_k$, où $\hat{\beta} = \sum_s y_k / \sum_s x_k$. Cet estimateur est considéré dans Särndal et al. (1992, p. 249).

Discutons finalement du cas $(q_k, r_k) = ((1 - \pi_k)/(\pi_k x_k), 1)$, proposé par Brewer (1979). Cet estimateur est étudié notamment dans Wright (1983) et Särndal et al. (1992). Cet estimateur est ADU, puisque la condition ADU

$$1 - \pi_k r_k = \pi_k q_k x_k \lambda$$

est satisfaite pour $\lambda = 1$. Voir Wright (1983). L'estimateur QR devient dans ces conditions

$$\hat{T}_{yQR} = \sum_s y_k + (T_x - \sum_s x_k) \hat{\beta}$$

où $\hat{\beta} = \sum_s (d_k - 1)y_k / \sum_s (d_k - 1)x_k$. On peut également écrire l'estimateur de Brewer (1979) de la manière suivante

$$\hat{T}_{yQR} = T_x \hat{\beta} + \sum_s d_k (y_k - x_k \hat{\beta}),$$

puisque l'on vérifie facilement que

$$\sum_s (y_k - x_k \hat{\beta}) = \sum_s d_k (y_k - x_k \hat{\beta})$$

qui n'est rien d'autre qu'un estimateur GREG avec $c_k = x_k / (1 - \pi_k)$. On peut ainsi considérer comme estimateur de variance l'estimateur de Särndal et al. (1989) pour l'estimateur GREG avec les poids-g

$$g_k = 1 + (T_x - \hat{T}_{xHT}) \frac{(1 - \pi_k)}{\sum_s (d_k - 1)x_k} \quad (\text{A.2})$$

Dans cette situation, si on utilise la formule A.1 avec $(q_k, r_k) = ((1 - \pi_k) / (\pi_k x_k), 1)$, on vérifie que

$$w_k = g_k / \pi_k,$$

où les poids-g g_k sont ceux donnés dans A.2. L'estimateur de variance A.1 se spécialise ainsi dans cette situation à l'estimateur de variance de Särndal et al. (1989).

A.2. NOUVELLE MÉTRIQUE

Nous généralisons le cas 6 de Deville et Särndal (1992). Cette métrique est parfois appelé la métrique logit. La fonction G correspondant au choix de cette métrique est

$$G(w_k; q_k, r_k) = \frac{w_k - L}{q_k A_k} \log\left(\frac{w_k - L}{r_k - L}\right) + \frac{U - w_k}{q_k A_k} \log\left(\frac{U - w_k}{U - r_k}\right),$$

où $A_k = (U - L) / ((r_k - L)(U - r_k))$.

La fonction h correspondant au choix de cette métrique est

$$h(x'_k \lambda; q_k, r_k) = \frac{L(U - r_k) + U(r_k - L) \exp(A_k q_k x'_k \lambda)}{(U - r_k) + (r_k - L) \exp(A_k q_k x'_k \lambda)}$$

Des simulations ont été entreprises avec cette métrique. Nous avons effectué des simulations complémentaires en utilisant le même germe, mais avec la métrique logit. Les résultats montrent que les estimateurs QR et RQR selon cette métrique se comportent essentiellement de la même manière. Cependant, les résultats sont un peu moins bon comparativement au cas de la métrique quadratique restreinte.

TAB. A.1. Résultats de Monte-Carlo lors de l'échantillonnage dans la population POPUSA, AREA, MU284 et MU281 pour l'estimateur robuste avec métrique logit.

POPUSA	VAR_M	MSE_M	CV_M	BR_M	MIN	MAX	RAFV ¹	CONV
$n = 10$								
LOGIT	29.08	29.86	2.68	-0.43	0.20	31.86	100.0	98.4
$n = 15$								
LOGIT	15.29	16.93	2.02	-0.63	0.20	15.89	100.0	99.5
AREA	VAR_M	MSE_M	CV_M	BR_M	MIN	MAX	RAFV ²	CONV
$n = 10$								
LOGIT	1.046	1.452	17.77	9.40	0.20	14.00	100.0	99.0
$n = 15$								
LOGIT	0.505	0.995	14.71	10.32	0.20	6.00	100.0	99.8
MU284	VAR_M	MSE_M	CV_M	BR_M	MIN	MAX	RAFV ³	CONV
$n = 30$								
LOGIT	0.719	1.530	14.84	-10.80	0.20	16.00	100.0	99.2
$n = 60$								
LOGIT	0.311	0.994	11.96	-9.91	0.20	7.00	100.0	99.7
MU281	VAR_M	MSE_M	CV_M	BR_M	MIN	MAX	RAFV ⁴	CONV
$n = 30$								
LOGIT	12.98	13.40	6.89	1.23	0.20	25.00	100.0	99.8
$n = 60$								
LOGIT	5.40	6.17	4.68	1.65	0.20	9.00	100.0	99.9

¹ Les bornes pour les RAFV sont [0.20, 32] pour $n = 10$ et [0.20, 16] pour $n = 15$.

² Les bornes pour les RAFV sont [0.20, 14] pour $n = 10$ et [0.20, 6] pour $n = 15$.

³ Les bornes pour les RAFV sont [0.20, 16] pour $n = 30$ et [0.20, 7] pour $n = 60$.

⁴ Les bornes pour les RAFV sont [0.20, 25] pour $n = 30$ et [0.20, 9] pour $n = 60$.

A.3. SIMULATIONS DANS UN PLAN POISSON SÉQUENTIEL

Dans la conclusion du premier essai, il est suggéré de pousser les études empiriques, en étudiant par exemple l'effet du plan d'échantillonnage. Puisque dans les simulations le plan d'échantillonnage était le tirage aléatoire simple, nous étions dans une situation où tous les π_k étaient égaux. Afin de considérer un plan avec des π_k inégaux, nous avons repris les simulations dans la population MU284 et MU281 avec $n = 30$ et $n = 60$, mais avec un plan de Poisson séquentiel, comme introduit et décrit dans Ohlsson (1998).

Le plan Poisson séquentiel (noté **SPS**) est un plan inspiré du plan de Poisson (**PO**). Rappelons que pour exécuter un plan **PO**, avec les probabilités π_1, \dots, π_N données, il suffit de considérer N réalisations u_1, \dots, u_N d'une loi uniforme sur $[0, 1]$, et de choisir l'unité k si $u_k < \pi_k$, $k = 1, \dots, N$. Ainsi, la taille de l'échantillon s est aléatoire dans un plan **PO**. Cette propriété n'est pas trop grave selon certains. Voir par exemple Särndal (1996). Cependant, elle n'est pas désirable pour d'autres. À cette fin, Ohlsson (1998) propose une modification du plan **PO**. Posons $p_k = nx_k/T_x$, où x_k est l'information auxiliaire (scalaire), par exemple une mesure de taille. Posons $\xi_k = T_x u_k / x_k$. La règle d'un plan **PO** reformulée avec les variables ξ_k consiste à prendre l'unité k si $\xi_k < n$. La modification de Ohlsson (1998) fournissant le plan **SPS** consiste à choisir les n plus petites valeurs de ξ_k .

Le plan **SPS** n'est pas strictement πps , mais il l'est approximativement. De plus, l'étude empirique de Ohlsson (1998) suggère fortement que les probabilités d'inclusion des unités k sont proches de p_k . Pour plus de détails, voir Ohlsson (1998).

Les estimateurs considérés sont le *GREG* usuel et la version avec les poids restreints. On considère également les estimateurs robustes *QRROB/U* et *QRROB/R* décrits dans l'étude empirique de l'essai 1. Cependant, le plan d'échantillonnage considéré est le plan **SPS**. Nous avons choisi comme mesure de taille dans *MU284* et *MU281* $\pi_k \propto \sqrt{x_k}$, afin d'éviter d'obtenir des probabilités d'inclusion trop grandes. Nous avons choisit aussi

$$h_k = \min[1, 1.46N \text{med}_U(x_k) / (nd_k x_k)], \quad (\text{A.3})$$

afin d'insérer le plan d'échantillonnage dans les h_k . Ce choix s'est effectué de manière empirique. Nous avons considéré aussi $h_k \equiv 1$ et aussi les mêmes h_k que dans l'essai 1. Cependant A.3 faisait raisonnablement bien et ce sont les résultats basés sur cette formule que nous reproduisons ici.

Les résultats de simulation obtenus dans le plan **SPS** sont intéressants, puisque ce plan règle en partie le problème des grandes unités. Voir l'introduction de l'essai 1. Cependant, malgré cela, les estimateurs robustes permettent des gains en erreur quadratique moyenne dans la population *MU284*, puisque les valeurs aberrantes semblent surtout occasionnées par de grandes valeurs de la variable y . Dans la population *MU281*, les estimateurs robustes et le *GREG* se comportent de façon similaire, avec un léger gain pour les estimateurs robustes.

TABLE A.2. Résultats de Monte-Carlo lors de l'échantillonnage dans la population MU284 sous un plan poisson séquentiel

ESTIMATEURS	VAR_M	MSE_M	CV_M	BR_M	MIN	MAX	RAFV ¹	CONV
<i>n</i> = 30								
GREG/U	2.482	2.511	19.00	-2.04	-5.17	28.10	84.1	
GREG/R	2.475	2.509	18.99	-2.21	0.20	16.00	100.0	99.8
QRROB/U	0.669	1.356	13.96	-9.94	-12.46	27.05	85.1	
QRROB/R	0.665	1.354	13.95	-9.96	0.20	16.00	100.0	99.8
<i>n</i> = 60								
GREG/U	1.094	1.101	12.59	-1.03	-0.13	11.45	91.3	
GREG/R	1.096	1.104	12.60	-1.07	0.20	8.00	100.0	100.0
QRROB/U	0.258	0.876	11.22	-9.43	-2.92	10.27	94.5	
QRROB/R	0.258	0.877	11.23	-9.43	0.20	8.00	100.0	100.0

¹ Les bornes pour les RAFV sont [0.20, 16] pour *n* = 30 et [0.20, 8] pour *n* = 60.

TABLE A.3. Résultats de Monte-Carlo lors de l'échantillonnage dans la population MU281 sous un plan poisson séquentiel

ESTIMATEURS	VAR_M	MSE_M	CV_M	BR_M	MIN	MAX	RAFV ¹	CONV
<i>n</i> = 30								
GREG/U	9.603	9.617	5.84	-0.23	-50.1	77.5	95.9	
GREG/R	9.613	9.627	5.84	-0.22	0.20	75.0	100.0	100.0
QRROB/U	9.355	9.371	5.76	0.24	-57.3	83.6	93.4	
QRROB/R	9.359	9.377	5.76	0.26	0.20	75.0	100.0	100.0
<i>n</i> = 60								
GREG/U	3.417	3.420	3.48	-0.09	-11.25	29.40	87.1	
GREG/R	3.414	3.417	3.48	-0.10	0.20	15.00	100.0	100.0
QRROB/U	3.282	3.371	3.46	0.56	-19.96	30.80	83.9	
QRROB/R	3.284	3.370	3.46	0.55	0.20	15.00	100.0	100.0

¹ Les bornes pour les RAFV sont [0.20, 75] pour *n* = 30 et [0.20, 15] pour *n* = 60.

BIBLIOGRAPHIE

- Brewer, K. R. W. (1979), 'A class of robust sampling designs for large-scale surveys', *Journal of the American Statistical Association* **74**, 911–915.
- Deville, J.-C. et Särndal, C. E. (1992), 'Calibration estimators in survey sampling', *Journal of the American Statistical Association* **87**, 376–382.
- Ohlsson, O. (1998), 'Sequential Poisson Sampling', *Journal of Official Statistics* **14**, 149–162.
- Särndal, C. E. (1996), 'Efficient estimators with simple variance in unequal probability sampling', *Journal of the American Statistical Association* **91**, 1289–1300.
- Särndal, C. E., Swensson, B. et Wretman, J. H. (1989), 'The weighted residual technique for estimating the variance of the general regression estimator of the finite population total', *Biometrika* **76**, 527–537.
- Särndal, C. E., Swensson, B. et Wretman, J. H. (1992), *Model Assisted Survey Sampling*, New-York: Springer-Verlag.
- Stukel, D. M., Hidioglou, M. A. et Särndal, C. E. (1996), 'Variance estimation for calibration estimators: A comparison of jackknifing versus taylor linearization', *Survey Methodology* **22**, 117–125.
- Wright, R. L. (1983), 'Finite population sampling with multivariate auxiliary information', *Journal of the American Statistical Association* **78**, 879–884.

Chapitre 2

A NOTE ON JACKKNIFE VARIANCE ESTIMATION FOR THE GENERAL REGRESSION ESTIMATOR

Résumé

We derive in this note explicit jackknife variance estimators of the general regression estimator (*GREG*) using the random group technique. A corrected version is proposed that removes a large part of the positive model bias. A small simulation is presented.

Some key words: Confidence interval; Jackknife; Regression estimator; Survey sampling; Variance estimation.

2.1. INTRODUCTION

Let $U = \{1, \dots, N\}$ be a finite population. Suppose that we know the total T_x of an auxiliary variable x of dimension p . A sample s is observed from a πps sampling plan. Let π_k and π_{kl} be the first and second inclusion probabilities, respectively. Our goal is to estimate the total $T_y = \sum_U y_k$ of a positive variable y with $\{(x_k, y_k), k \in s\}$ and T_x .

The general regression estimator (*GREG*) of T_y is given by

$$\hat{T}_{GREG} = \sum_s d_k g_{ks} y_k,$$

where

$$g_{ks} = 1 + (T_x - \hat{T}_{x\pi})' \left(\sum_s d_k x_k x_k' / c_k \right)^{-1} x_k / c_k,$$

is the 'g-weight', $d_k = \pi_k^{-1}$ is the sampling weight, $\hat{T}_{x\pi} = \sum_s x_k / \pi_k$ is the Horvitz-Thompson estimator of T_x , and c_k is chosen by the user. Särndal (1996) discusses the choice of c_k . The asymptotic variance *AV* for the *GREG* is given by

$$AV(\hat{T}_{GREG}) = \sum \sum_U \Delta_{kl} \check{E}_k \check{E}_l,$$

where $E_k = y_k - x_k' B$, $B = \left(\sum_U x_k x_k' / c_k \right)^{-1} \sum_U x_k y_k / c_k$, $\Delta_{kl} = \pi_{kl} - \pi_k \pi_l$ and $\check{E}_k = E_k / \pi_k$.

Since the asymptotic variance is an ordinarily unknown quantity, Särndal et al. (1989) suggested the following g -weighted variance estimator given by

$$\hat{V}_g = \hat{V}_g(\hat{T}_{GREG}) = \sum \sum_s \check{\Delta}_{kl}(g_{ks}\check{e}_{ks})(g_{ls}\check{e}_{ls}), \quad (2.1)$$

where $e_{ks} = y_k - x'_k \hat{B}_s$, $\hat{B}_s = (\sum_s d_k x_k x'_k / c_k)^{-1} \sum_s d_k x_k y_k / c_k$, $\check{\Delta}_{kl} = \Delta_{kl} / \pi_{kl}$ and $\check{e}_{ks} = e_{ks} / \pi_k$. With \hat{V}_g , we can construct a confidence interval for T_y given by $\hat{T}_{GREG} \pm z_{1-\alpha/2} [\hat{V}_g]^{1/2}$, that is expected to be valid approximately to the $1 - \alpha$ confidence level.

The jackknife technique is another popular method to obtain a variance estimator. That method is described in Wolter (1985) and Särndal et al. (1992, chap. 11). We derive in the next section explicit jackknife variance estimators of the *GREG*. A corrected version is proposed that removes a large part of the positive model bias in section 3. A small simulation is given in section 4 to illustrate the proposed estimator. We conclude with a discussion in section 5.

2.2. JACKKNIFE VARIANCE ESTIMATION

In this section, we obtain explicit formulas for the jackknife variance estimators of the *GREG*. Let the sample be divided into A groups of size m partitioning the sample, where $Am = n$, where n is the sample size. The two jackknife variance estimators advocated by Särndal et al. (1992) are given by

$$\begin{aligned} \hat{V}_{JK1} &= \frac{A-1}{A} \sum_{a=1}^A (\hat{T}_{GREG}(a) - \hat{T}_{GREG,JK})^2, \\ \hat{V}_{JK2} &= \frac{A-1}{A} \sum_{a=1}^A (\hat{T}_{GREG}(a) - \hat{T}_{GREG})^2, \end{aligned}$$

where $\hat{T}_{GREG}(a)$ is the *GREG* calculated without the group a and

$$\hat{T}_{GREG,JK} = \frac{1}{A} \sum_{a=1}^A \hat{T}_{GREG}(a).$$

Since the two formulas \hat{V}_{JK1} and \hat{V}_{JK2} are related by the following relation

$$\hat{V}_{JK1} = \hat{V}_{JK2} - (A-1)(\hat{T}_{GREG} - \hat{T}_{GREG,JK})^2, \quad (2.2)$$

we can conclude that $\hat{V}_{JK1} \leq \hat{V}_{JK2}$ and it is easy to pass from one form to the other. In practice, the two formulas give very similar results.

We consider in the following the maximal number of groups, that is the case $A = n$, $m = 1$. See the remark 11.5.3 of Särndal et al. (1992, pp. 441-442). With that hypothesis, we now use the random group technique to obtain explicit formulas for \hat{V}_{JK2} and for \hat{V}_{JK1}

using expression (2.2). Under that technique, we suppose that conditional on s , each group $\{i\}$ is obtained by simple random sampling. In that case, there are n random subsamples $s_{(i)} = s - \{i\}$. The inclusion probability that the unit k will be in the final subsample, denoted $\pi_k(i)$, is $\pi_k(i) = \frac{n-1}{n}\pi_k$. Using that technique, we obtain the following result:

Proposition 1. *The jackknife variance estimator \hat{V}_{JK2} of the GREG estimator is given by*

$$\hat{V}_{JK2} = \frac{n}{n-1} \sum_s (\tilde{g}_{is}\tilde{e}_{is} - n^{-1}\hat{T}_{e\pi})^2, \quad (2.3)$$

where $\tilde{g}_{is} = (g_{is} - n^{-1}T'_x M_s^{-1}x_i/c_i)/(1 - h_i)$, $\hat{T}_{e\pi} = \sum_s e_{ks}/\pi_k$, $h_i = d_i x'_i M_s^{-1}x_i/c_i$, $e_{is} = y_i - x'_i \hat{B}_s$, $M_s = \sum_s d_k x_k x'_k / c_k$.

Proof. Let $s_{(i)}$ the sample s without unit i . Since

$$\begin{aligned} \hat{T}_{GREG} &= \sum_s d_k g_{ks} y_k, \\ &= \hat{T}_{y\pi} + (T_x - \hat{T}_{x\pi})' \hat{B}_s, \end{aligned}$$

the GREG without unit i can be written as

$$\hat{T}_{GREG}(i) = \hat{T}_{y\pi}(i) + (T_x - \hat{T}_{x\pi}(i))' \hat{B}_s(i),$$

where $\hat{T}_{y\pi}(i) = \sum_{s_{(i)}} y_k / \pi_k(i)$ and similarly for $\hat{T}_{x\pi}(i)$, and

$$\hat{B}_s(i) = \left\{ \sum_{s_{(i)}} x_k x'_k / (c_k \pi_k(i)) \right\}^{-1} \sum_{s_{(i)}} x_k y_k / (c_k \pi_k(i)).$$

With some algebra¹, we can show that

$$\hat{T}_{GREG}(i) = \hat{T}_{GREG} + \frac{1}{n-1} \sum_s e_{ks}/\pi_k - \frac{n}{n-1} (g_{is} - n^{-1}T'_x M_s^{-1}x_i/c_i)\tilde{e}_{is}/(1 - h_i).$$

Finally, using the relation (2.2), we obtain the following corollary

Corollary 1.

$$\hat{V}_{JK1} = \hat{V}_{JK2} - \frac{1}{n-1} (\hat{T}_{e\pi} - \sum_s \tilde{g}_{is}\tilde{e}_{is})^2.$$

It is interesting to note that with the exception of the factor \tilde{g}_{is} , formula (2.3) looks like the simplified variance estimator \hat{V}_0 of Särndal et al. (1992, ex. (11.2.6), p. 422), where e_{is} now replaces y_i .

1. Note that using standard linear regression results, we obtain $\hat{B}_s(i) = \hat{B}_s - M_s^{-1}x_i\tilde{e}_{is}/((1 - h_i)c_i)$. Note also that $\hat{T}_{y\pi}(i) = \frac{n}{n-1}(\hat{T}_{y\pi} - d_i y_i)$ and similarly for $\hat{T}_{x\pi}(i)$.

2.3. CORRECTED ESTIMATOR FOR THE MODEL BIAS

Särndal et al. (1992) note that, with the exception of the Horvitz-Thompson estimator, there are no exact results concerning the properties of the jackknife variance estimator. We study in this section the model bias of the formula (2.3). Let the ξ regression model for y_1, \dots, y_N given by

$$y_k = x_k' \beta + \epsilon_k,$$

where the ϵ_k are independent under the model, and such that $E_\xi(\epsilon_k) = 0$, $V_\xi(\epsilon_k) = \sigma^2 c_k$, where E_ξ and V_ξ indicate the mean and variance under the model. We assume like Särndal et al. (1989) that for some λ independent of k

$$c_k = \lambda' x_k, \tag{2.4}$$

and we define like them the *prototype* \hat{V}^* for \hat{V}_g as

$$\hat{V}^* = \sum \sum_s \check{\Delta}_{kl}(g_{ks} \check{\epsilon}_k)(g_{ls} \check{\epsilon}_l).$$

Särndal et al. (1992, p. 232) give several examples of variance structures satisfying condition (2.4). Note that under that condition we have $\sum_s e_{ks}/\pi_k = 0$. We now recall a result that will be useful in the sequel.

Lemma 1. *Under the model ξ , for any given realized sample s , the model mean, the model mean squared error and the relative model bias of the prototype \hat{V}^* are given by*

$$\begin{aligned} (i) \quad E_\xi(\hat{V}^*) &= \sigma^2 \left(\sum_s g_{ks}^2 c_k / \pi_k^2 - \sum_U g_{ks} c_k \right) \\ (ii) \quad MSE_\xi(\hat{T}_{GREG}) &= E_\xi(\hat{T}_{GREG} - T_y)^2 = \sigma^2 \left(\sum_s g_{ks}^2 c_k / \pi_k^2 - \sum_U c_k \right) \\ (iii) \quad RMB_\xi(\hat{V}^*) &= \frac{E_\xi(\hat{V}^*) - MSE_\xi(\hat{T}_{GREG})}{MSE_\xi(\hat{T}_{GREG})} = \frac{-\sum_U (g_{ks} - 1) c_k}{\sum_s g_{ks}^2 c_k / \pi_k^2 - \sum_U c_k}. \end{aligned}$$

Proof. See Särndal et al. (1989).

We study properties of the jackknife variance estimator *prototype*. Under condition (2.4), it is given by

$$\hat{V}_{JK2}^* = \frac{n}{n-1} \sum_s \tilde{g}_{ks}^2 \tilde{\epsilon}_k^2.$$

Under the model, note that $E_\xi(\hat{V}_{JK2}^*) = \frac{n}{n-1} \sigma^2 \sum_s \tilde{g}_{ks}^2 c_k / \pi_k^2 = A_s$. Suppose that all h_i are negligible (their sample mean is $n^{-1} \sum_s h_i = p/n$). Then $g_{ks} \approx \tilde{g}_{ks}$ and A_s will be of the same

order as the first term in the right member of (i) in the lemma. We have approximately

$$E_{\xi}(\hat{V}_{JK2}^*) - E_{\xi}(\hat{V}^*) \approx \sigma^2 \sum_U g_{ks} c_k,$$

suggesting that the jackknife variance estimator over-estimates the true variance, as it is well-known. The relative model bias of the jackknife variance estimator can also be calculated using (iii) in the lemma:

$$RMB_{\xi}(\hat{V}_{JK2}^*) \approx \frac{\sum_U c_k}{\sum_s g_{ks}^2 c_k / \pi_k^2 - \sum_U c_k}. \quad (2.5)$$

Looking at the numerator of (2.5), the relative model bias $RMB_{\xi}(\hat{V}_{JK2}^*)$ is expected to be more important than for \hat{V}_g . It may however be small if the first term in the denominator dominates the second term in formula (2.5). It can be seen that under simple random sampling, if the sampling fraction $f = n/N$ is small, then $RMB_{\xi}(\hat{V}_{JK2}^*)$ can be negligible.

However, since the positive bias may be more important in practice, we consider the following modification:

$$\hat{V}_{JK3}^* = \frac{n}{n-1} \sum_s (1 - \pi_k) \tilde{g}_{ks}^2 \check{\epsilon}_k^2. \quad (2.6)$$

We can justify that modification with the following argument. Under the model, we now obtain

$$E_{\xi}(\hat{V}_{JK3}^*) = A_s - \frac{n}{n-1} \sigma^2 \sum_s \tilde{g}_{ks}^2 c_k / \pi_k. \quad (2.7)$$

If $g_{ks} \approx \tilde{g}_{ks}$, then the second term of that right member of expression (2.7) will be of the same order as the second member of (i) in the lemma since $\sum_s \tilde{g}_{ks}^2 c_k / \pi_k = \sum_U g_{ks} c_k$ (see Särndal et al. (1989, expression 5.6)). Let us remark that in the case of the simple random sampling, \hat{V}_{JK3}^* is simply \hat{V}_{JK2}^* affected by the finite population correction. However, our analysis is in the more general setting of a πps sampling plan. Wolter (1985) discusses some methods to remove the bias of the jackknife variance estimator in the Horvitz-Thompson case. See also Särndal et al. (1992, pp. 439-440). These ideas are applied here to the *GREG*.

2.4. ILLUSTRATION

We consider a small Monte Carlo simulation for the variables $y = RMT85 \times 10^{-4}$, $x_1 = CS82$ and $x_2 = SS82$ for the *MU281* population (of size $N = 281$) in Särndal et al. (1992). This study is a complement to the one in Särndal et al. (1992, pp. 278-280). Like them, we carried out 5000 repeated simple random samples, each of sample size $n = 100$. The main

objective of the simulation study is to evaluate coverage properties of confidence intervals at the 95% level

$$\hat{T} \pm 1.96[\hat{V}(\hat{T})]^{1/2},$$

where \hat{T} is the *GREG* estimator, and $\hat{V}(\hat{T})$ is a variance estimator. We consider the *GREG* estimator with only x_1 , the *GREG* estimator with only x_2 , and the *GREG* estimator with x_1 and x_2 . We always included an intercept and let $c_k \equiv 1$ in all the study. We consider the variance estimator given in formula (2.1), the jackknife variance estimators given in formula (2.3) and the corrected version (2.6). Formula (2.1) becomes under simple random sampling

$$\hat{V}_g = N^2 \left(\frac{1}{n} - \frac{1}{N} \right) \frac{\sum_s g_{ks}^2 e_k^2}{n-1}.$$

Results are presented in table 2.1, where $\bar{\hat{T}}$ and $S_{\hat{T}}^2$ are the sample mean and sample variance of the 5000 estimates \hat{T} ; $\bar{\hat{V}}_g$, $\bar{\hat{V}}_{JK2}$ and $\bar{\hat{V}}_{JK3}$ are the sample mean of the 5000 variance estimates, \hat{V}_g , \hat{V}_{JK2} and \hat{V}_{JK3} , respectively; and ECR_g , ECR_{JK2} and ECR_{JK3} are the respective coverage rates for the *GREG* based on \hat{V}_g , \hat{V}_{JK2} , \hat{V}_{JK3} , respectively. The final column gives the approximate variance for \hat{T} given by

$$AV(\hat{T}) = N^2 \left(\frac{1}{n} - \frac{1}{N} \right) \frac{\sum_U E_k^2}{N-1}.$$

The results in table 2.1 show that in our experience, \hat{V}_{JK3} gives good coverage properties and in the three cases the variance of the 5000 estimates is close to $\bar{\hat{V}}_{JK3}$. In that limited simulation, \hat{V}_{JK3} seems to compare reasonably well with \hat{V}_g .

TABLE 2.1. *Results of the Simulation*

Estimator	$\bar{\hat{T}}$	$S_{\hat{T}}^2$	$\bar{\hat{V}}_g$	ECR_g	$\bar{\hat{V}}_{JK2}$	ECR_{JK2}	$\bar{\hat{V}}_{JK3}$	ECR_{JK3}	AV
$\hat{T}_{GREG}(x_1)$	5.31	0.122	0.115	0.937	0.189	0.977	0.121	0.942	0.116
$\hat{T}_{GREG}(x_2)$	5.30	0.124	0.118	0.934	0.191	0.978	0.123	0.941	0.117
$\hat{T}_{GREG}(x_1, x_2)$	5.31	0.056	0.052	0.929	0.088	0.978	0.057	0.939	0.052

Note: The total T_y is 5.315.

2.5. DISCUSSION

In this note, explicit jackknife *GREG* variance estimators are exhibited. These formulas give new examples of the well-known rule of thumb that jackknifing leads to overestimation of variance. An idea for possible overestimation 'correction' is presented, leading to a modified estimator. In the numerical illustration, we obtain reasonable properties with the corrected

version. We do not claim that the proposed estimator is superior to other estimators. In fact, in a vast majority of situations occurring in practice, \hat{V}_g may be preferable. Jackknife estimators are maybe for exceptional situations (shortage of time, one-time use, etc.). It seems to appear that their chief merit is that they require less programming efforts. For example, there is no need to evaluate all the π_{kl} as in \hat{V}_g . See Särndal (1996) for a discussion of that problem. However, if jackknife variance estimators for the *GREG* are needed, it is hoped that proposition 2.1 will be useful.

2.6. ACKNOWLEDGEMENTS

I would like to thank Prof. Carl-Erik Särndal, Prof. Bengt Rosén, the Chief Editor and two anonymous referees for helpful comments.

BIBLIOGRAPHIE

- Särndal, C. E. (1996). 'Efficient estimators with simple variance in unequal probability sampling', *Journal of the American Statistical Association* **91**, 1289–1300.
- Särndal, C. E., Swensson, B. et Wretman, J. H. (1989), 'The weighted residual technique for estimating the variance of the general regression estimator of the finite population total', *Biometrika* **76**, 527–537.
- Särndal, C. E., Swensson, B. et Wretman, J. H. (1992), *Model Assisted Survey Sampling*, New-York: Springer-Verlag.
- Wolter, K. M. (1985), *Introduction to Variance Estimation*, New-York: Springer-Verlag.

Chapitre 3

TESTS DE CORRÉLATION SÉRIELLE DANS LES MODÈLES *VARX*

Résumé

Nous considérons des tests d'absence de corrélation dans le terme d'erreur d'un modèle *VARX*. Il est possible dans cette classe de généraliser au cas multivarié l'approche de Hong (1996a). Le test est obtenu en comparant un estimateur de la densité spectrale multivariée calculé par la méthode du noyau avec la densité spectrale sous l'hypothèse nulle d'absence de corrélation dans le terme d'erreur, en utilisant une certaine norme quadratique. Le test généralise la statistique portmanteau de Hosking (1980), qui peut être vu comme un test basé sur la norme quadratique utilisant le périodogramme tronqué. Cependant, tout comme dans le cas univarié, plusieurs noyaux offrent une meilleure puissance que le test basé sur le noyau uniforme tronqué.

3.1. INTRODUCTION

La classe des modèles *VARX* est très utilisée dans les applications touchant entre autres l'économie. Ils sont aussi appelés modèles à équations simultanées dynamiques dans la littérature économétrique. Dans la terminologie économétrique, les variables dépendantes sont souvent dites endogènes alors que les variables explicatives sont appelées les variables exogènes. Ces modèles généralisent les modèles de régression linéaire multiple où l'on inclut parmi les variables explicatives des variables endogènes et exogènes retardées. Ainsi, dictés par des considérations théoriques ou empiriques, ces modèles permettent de formuler des situations où l'on reconnaît qu'il peut exister des relations causales entre des variables économiques présumées stochastiques. Dans de tels modèles, on reconnaît que des données économiques sont susceptibles d'être influencées par les états actuels et passés des variables impliquées. Ces modèles sont discutés dans de nombreux ouvrages comme par exemple Judge et al. (1985, 1988), Lütkepohl (1993) et Hendry (1995).

Dans les modèles *ARX*, Hong (1996a) a proposé plusieurs classes de tests convergents pour vérifier l'hypothèse que le terme d'erreur est un bruit blanc contre une alternative de

dépendance de forme inconnue. Rappelons qu'un test convergent admet une puissance asymptotique tendant vers un lorsque le nombre d'observations n tend vers l'infini, pour un niveau α fixé. Ces tests sont motivés par le fait que toute forme de corrélation sérielle dans le terme d'erreur risque d'entraîner des problèmes d'identification des paramètres du modèle. De plus, elle risque d'occasionner la non convergence de l'estimateur des moindres carrés. Ainsi, Hong (1996a) a proposé trois classes de statistiques de type portmanteau pour l'hypothèse de bruit blanc pour le terme d'erreur d'un modèle ARX . L'approche repose sur une comparaison d'un estimateur de la densité spectrale calculé par la méthode du noyau à la densité spectrale sous l'hypothèse nulle. Plusieurs mesures de distance sont possibles. Lorsque la fonction de distance choisie est la norme quadratique, la statistique proposée s'écrit comme

$$M_{1n} = \frac{n \sum_{j=1}^{n-1} k^2(j/p_n) \hat{\rho}^2(j) - M_n(k)}{\sqrt{2V_n(k)}},$$

où $\hat{\rho}(j) = \hat{R}(j)/\hat{R}(0)$ est l'autocorrélation échantillonnale, $\hat{R}(j) = n^{-1} \sum_{t=|j|+1}^n \hat{u}_t \hat{u}_{t-|j|}$ est l'autocovariance échantillonnale, où k est un noyau et où

$$M_n(k) = \sum_{j=1}^{n-1} (1 - j/n) k^2(j/p_n), \quad (3.1)$$

$$V_n(k) = \sum_{j=1}^{n-2} (1 - j/n)(1 - (j+1)/n) k^4(j/p_n). \quad (3.2)$$

Parmi les autres mesures de distance possibles, on peut entre autres considérer la métrique de Hellinger ou encore le critère d'information de Kullback-Leibler.

En présence de plusieurs équations, il incombe également de vérifier l'hypothèse que le terme d'erreur est un bruit blanc. Dans la section 3.4, l'approche de Hong (1996a) est étendue aux modèles à équations simultanées dynamiques écrits sous leur forme réduite. Tout comme dans le cas univarié, nous développons un test convergent contre une alternative de corrélation sérielle de forme inconnue. L'approche consiste à comparer un estimateur de la densité spectrale multidimensionnel calculé avec la méthode du noyau et la densité spectrale sous l'hypothèse nulle, avec une certaine norme quadratique. Nous obtenons la distribution asymptotique sous l'hypothèse nulle qui est $N(0, 1)$ sans avoir à spécifier un modèle pour les erreurs sous l'hypothèse alternative. Le test est unilatéral et rejette pour des grandes valeurs. Le test reste invariant lorsque les régresseurs contiennent des variables endogènes retardées. Ce test permet aussi d'obtenir une nouvelle interprétation de la statistique portmanteau multivariée de Hosking (1980), ce dernier test étant la version multivariée de la statistique de

Box et Pierce (1970). La puissance est étudiée de manière théorique, suivant une approche similaire à celle de Hong (1996a). Il est possible dans une certaine classe de trouver le noyau optimal.

Une approche différente abordant les tests d'ajustement d'une série univariée est décrite dans Paparoditis (1999). L'approche consiste à comparer avec une mesure de distance un estimateur par le noyau du ratio de la véritable densité spectrale et de la densité spectrale hypothétique, à la valeur espérée de l'estimateur, sous l'hypothèse nulle. L'étude de la puissance des tests d'ajustement est étudiée dans Paparoditis (2000).

Nous étudions aussi les niveaux et les puissances des tests de manière empirique. Il ressort de l'étude empirique que l'on peut obtenir des résultats de puissance supérieure en choisissant un noyau autre que le noyau uniforme tronqué.

Une application est présentée, en utilisant des données réelles provenant de Gilbert (1993).

3.2. PRÉLIMINAIRES

Soit $\mathbf{y} = \{\mathbf{y}_t : t \in \mathbb{Z}\}$ et $\mathbf{x} = \{\mathbf{x}_t : t \in \mathbb{Z}\}$ des processus multivariés de dimension d et m respectivement, que l'on suppose stationnaires au second ordre.

Définition 1. *Le processus \mathbf{y} est dit processus autorégressif multivarié avec variables explicatives, noté VARX(r, s), s'il existe des matrices $\mathbf{\Lambda}_j$ de dimension $d \times d$, $j = 0, \dots, r$ et des matrices \mathbf{V}_j , de dimension $d \times m$, $j = 0, \dots, s$ telles que $\mathbf{\Lambda}_r \neq \mathbf{0}$ et $\mathbf{V}_s \neq \mathbf{0}$, et*

$$\mathbf{\Lambda}(B)\mathbf{y}_t = \mathbf{c} + \mathbf{V}(B)\mathbf{x}_t + \mathbf{u}_t, \quad t \in \mathbb{Z} \quad (3.3)$$

où $\mathbf{\Lambda}(B) = \mathbf{\Lambda}_0 - \sum_{j=1}^r \mathbf{\Lambda}_j B^j$, $\mathbf{V}(B) = \sum_{j=0}^s \mathbf{V}_j B^j$, B désigne l'opérateur retard et $\mathbf{u} = \{\mathbf{u}_t : t \in \mathbb{Z}\}$ est un bruit blanc fort de dimension d , c'est-à-dire que les \mathbf{u}_t sont des vecteurs aléatoires centrés en zéro, indépendants et de matrice de covariance $\mathbf{\Sigma}_u$ définie positive.

La fonction d'autocovariance de $\{\mathbf{u}_t\}$ est notée $\mathbf{\Gamma}_u(j) = E(\mathbf{u}_t \mathbf{u}'_{t-j})$. Dans la suite, nous supposons de plus que toutes les racines de $\mathbf{\Lambda}(z)$ sont à l'extérieur du disque unité, où z désigne une variable complexe. Lorsque la matrice $\mathbf{\Lambda}_0$ reflète les effets instantanés du système, on dit que le modèle (3.3) est formulé dans sa forme structurelle. Lorsque $\mathbf{\Lambda}_0$ est non-singulière et que l'on pré-multiplie (3.3) par $\mathbf{\Lambda}_0^{-1}$, on obtient alors la forme réduite du système. La forme structurelle est peut-être plus satisfaisante du point de vue de l'interprétation des divers paramètres. Cependant, l'utilisation de la méthode des moindres carrés est problématique avec ce type de représentation. La forme réduite évite les problèmes liés à l'identification des paramètres du modèle et permet également une utilisation immédiate de la méthode des

moindres carrés. Lütkepohl (1993, chap. 10) discute de l'estimation et des avantages de la forme réduite pour des fins de prévision, d'analyse des multiplicateurs et de contrôle. On présume dans la suite que le modèle (3.3) est formulé dans la forme réduite, en supposant que $\Lambda_0 = \mathbf{I}_d$.

Nous allons supposer dans la suite que les moments d'ordre quatre des composantes de $\mathbf{u}_t = (u_t(1), \dots, u_t(d))'$ existent, c'est-à-dire que

$$E(|u_t(i)u_t(j)u_t(k)u_t(l)|) < \infty, \quad i, j, k, l = 1, \dots, d.$$

Nous notons aussi

$$E(u_t(i)u_t(j)u_t(k)u_t(l)) = \mu_4(i, j, k, l)$$

dans la suite.

L'hypothèse d'existence des moments d'ordre quatre est équivalente à l'existence des cumulants d'ordre quatre, que nous notons

$$\kappa_{pqrs}(i, j, k, l) = \text{cum}(u_i(p), u_j(q), u_k(r), u_l(s)).$$

Si le processus \mathbf{u} est gaussien, il est bien connu que les cumulants d'ordre quatre sont nuls.

Étant donné une réalisation de taille n du processus \mathbf{u} , nous pouvons alors calculer la fonction d'autocovariance pour les délais $0 \leq |j| \leq n-1$, selon

$$\mathbf{C}_u(j) = \begin{cases} n^{-1} \sum_{t=j+1}^n \mathbf{u}_t \mathbf{u}'_{t-j}, & j = 0, 1, \dots, n-1, \\ \mathbf{C}'_u(-j), & j = -1, \dots, -n+1. \end{cases}$$

Des définitions similaires peuvent être obtenues une fois le modèle (3.3) ajusté. Ainsi, lorsque les résidus $\hat{\mathbf{u}}_t$ résultant de l'ajustement sont déterminés nous pouvons calculer la fonction d'autocovariance échantillonnale $\mathbf{C}_{\hat{\mathbf{u}}}(j)$.

La densité spectrale du processus \mathbf{u} est définie par

$$\mathbf{f}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \Gamma_u(h) e^{-i\omega h}. \quad (3.4)$$

L'estimateur de la densité spectrale pour \mathbf{f} que nous considérons est calculé selon la méthode du noyau. Il est obtenu selon

$$\hat{\mathbf{f}}_n(\omega) = \frac{1}{2\pi} \sum_{j=-n+1}^{n-1} k(j/p_n) \mathbf{C}_{\hat{\mathbf{u}}}(j) e^{-i\omega j}, \quad (3.5)$$

où $k(\cdot)$ est le noyau. Le paramètre p_n correspond au point de troncature lorsque le noyau est à support compact ou au paramètre de lissage lorsque le support du noyau n'est pas borné.

On suppose que $p_n \rightarrow \infty$ et $p_n/n \rightarrow 0$. Des exemples de p_n satisfaisant ces critères sont $p_n \propto n^\theta$, avec $\theta \in (0, 1)$, ainsi que $p_n \propto \log(n)$. Les hypothèses sur le noyau sont regroupées dans l'hypothèse A suivante.

Nous définissons le périodogramme tronqué comme étant

$$\hat{f}_{Tn}(\omega) = \frac{1}{2\pi} \sum_{|j| \leq p_n} C_{\hat{a}}(j) e^{-i\omega j}, \quad (3.6)$$

correspondant à l'estimateur de la densité spectrale (3.5) avec le choix du noyau uniforme tronqué.

Hypothèse A: On suppose que le noyau $k : \mathbb{R} \rightarrow [-1, 1]$ est une fonction symétrique, continue à 0 et ayant au plus un nombre fini de discontinuités, telle que $k(0) = 1$ et $\int_{-\infty}^{\infty} k^2(z) dz < \infty$.

Nous concluons cette section en donnant les définitions des noyaux que nous étudierons par la suite. Pour plus de détails sur les différents noyaux, on peut consulter Priestley (1981a, 1981b).

$$\begin{aligned} \text{Noyau tronqué:} \quad k(z) &= \begin{cases} 1 & \text{si } |z| \leq 1, \\ 0 & \text{sinon,} \end{cases} \\ \text{Bartlett:} \quad k(z) &= \begin{cases} 1 - |z| & \text{si } |z| \leq 1, \\ 0 & \text{sinon,} \end{cases} \\ \text{Daniell:} \quad k(z) &= \frac{\sin(\pi z)}{\pi z}, \\ \text{Parzen:} \quad k(z) &= \begin{cases} 1 - 6(\pi z/6)^2 + 6|\pi z/6|^3 & \text{si } |z| \leq 3/\pi, \\ 2(1 - |z|)^3 & \text{si } 3/\pi \leq |z| \leq 6/\pi, \\ 0 & \text{sinon,} \end{cases} \\ \text{Bartlett-Priestley:} \quad k(z) &= \frac{9}{5\pi^2 z^2} \left\{ \frac{\sin(\pi\sqrt{5/3}z)}{\pi\sqrt{5/3}z} - \cos(\pi\sqrt{5/3}z) \right\}. \end{aligned}$$

3.2.1. Théorème central limite pour les différences de martingales

Afin de démontrer le résultat de normalité asymptotique de la section 3.4, nous ferons appel à un théorème central limite pour les différences de martingale. On retrouve les définitions utiles et ce genre de résultats dans Brown (1971), White (1984) et Hamilton (1994).

Considérons le processus $Y = \{Y_t : t \in \mathbb{Z}\}$ tel que $E(Y_t) = 0, \forall t$. Soit $\{\mathcal{F}_t, t \in \mathbb{N}\}$ une suite croissante de σ -algèbres, $\mathcal{F}_t \subset \mathcal{F}_{t+1}$ où Y_t est mesurable par rapport à \mathcal{F}_t . Alors si

$$E(Y_t | \mathcal{F}_{t-1}) = 0, \quad t = 2, 3, \dots,$$

Y est dite une différence de martingales.

Plusieurs théorèmes limites sont disponibles lorsque nous sommes en présence de différences de martingales. Nous allons utiliser le théorème de Brown (1971).

Théorème 1 (Brown, 1971). *Soit Y une différence de martingales. Posons*

$$\begin{aligned} V_n^2 &= \sum_{j=1}^n \sigma_j^2, \\ s_n^2 &= E(V_n^2), \end{aligned}$$

où $\sigma_j^2 = E(Y_j^2 | \mathcal{F}_{j-1})$. Si $V_n^2 s_n^{-2} \rightarrow_P 1$ et si la condition de Lindeberg suivante est satisfaite

$$s_n^{-2} \sum_{j=1}^n E[Y_j^2 I(|Y_j| \geq \epsilon s_n)] \rightarrow 0, \quad \forall \epsilon > 0,$$

où $I(A)$ désigne la fonction indicatrice de l'ensemble A . Alors, on a

$$s_n^{-1} \sum_{j=1}^n Y_j \rightarrow_L N(0, 1).$$

En pratique, la condition de Lyapounov suivante est plus facile à vérifier et implique la condition de Lindeberg

$$s_n^{-2-\delta} \sum_{j=1}^n E(|Y_j|^{2+\delta}) \rightarrow 0.$$

Pour plus de détails sur les conditions de Lindeberg et de Lyapounov, on peut consulter Shiryayev (1995, p. 332). Pour des théorèmes limites sur les différences de martingales avec des conditions légèrement différentes, voir par exemple White (1984) et Hamilton (1994).

3.3. ESTIMATION PAR LES MOINDRES CARRÉS CALCULABLES

Nous allons dans cette section décrire la méthode d'estimation des moindres carrés calculables pour estimer les paramètres d'un modèle $VARX(r, s)$ écrit sous sa forme réduite. Le traitement de cette section est décrit dans Lütkepohl (1993, Chap. 10) et est inclus ici dans un souci de clarté pour la suite.

Posons $\mathbf{\Lambda} = (\mathbf{\Lambda}_1, \mathbf{\Lambda}_2, \dots, \mathbf{\Lambda}_r)$, $\mathbf{V} = (\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_s)$, $\mathbf{Y}_t = (\mathbf{y}_t' \dots \mathbf{y}_{t-r+1}')'$, $\mathbf{X}_t = (\mathbf{x}_t' \dots \mathbf{x}_{t-s+1}')'$, \mathbf{Y}_t et \mathbf{X}_t sont de dimension $dr \times 1$ et $ms \times 1$, respectivement. On peut alors écrire le modèle (3.3) sous la forme

$$\mathbf{y}_t = \mathbf{\Lambda} \mathbf{Y}_{t-1} + \mathbf{V} \mathbf{X}_{t-1} + \mathbf{V}_0 \mathbf{x}_t + \mathbf{u}_t.$$

Si on pose $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)$ et $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n)$ des matrices de dimension $d \times n$, et si nous posons

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Y}_0 & \mathbf{Y}_1 & \cdots & \mathbf{Y}_{n-1} \\ \mathbf{X}_0 & \mathbf{X}_1 & \cdots & \mathbf{X}_{n-1} \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{pmatrix}$$

une matrice de dimension $(dr + ms + m) \times n$, on peut alors écrire le modèle sous la forme

$$\mathbf{Y} = (\mathbf{\Lambda}, \mathbf{V}, \mathbf{V}_0)\mathbf{Z} + \mathbf{U}. \quad (3.7)$$

Si on suppose que les contraintes sur les paramètres satisfont $\boldsymbol{\beta} = \text{vec}(\mathbf{\Lambda}, \mathbf{V}, \mathbf{V}_0) = \mathbf{R}\boldsymbol{\gamma}$, alors on peut écrire en vectorisant (3.7)

$$\mathcal{Y} = (\mathbf{Z}' \otimes \mathbf{I}_d)\mathbf{R}\boldsymbol{\gamma} + \mathcal{U}. \quad (3.8)$$

où $\mathcal{Y} = \text{vec}(\mathbf{Y})$ et $\mathcal{U} = \text{vec}(\mathbf{U})$ et en utilisant la relation bien connue $\text{vec}(\mathbf{AXB}) = (\mathbf{B}' \otimes \mathbf{A})\text{vec}(\mathbf{X})$. S'il n'y a aucune contrainte, alors \mathbf{R} est la matrice identité. Le contexte avec des contraintes linéaires sur les paramètres est légèrement plus général et permet une utilisation simple de l'estimation des moindres carrés. De ce point de vue il est d'intérêt en soi. Cependant, d'un point de vue économique, il peut également y avoir des situations où des contraintes linéaires a priori sont connues sur les paramètres. Par exemple Lütkepohl (1993, Chap. 5) fait un traitement détaillé de l'estimation des modèles VAR avec des contraintes linéaires. Finalement, l'estimation avec des contraintes sur les paramètres peut être utile lorsque certains d'entre eux sont fixés à zéro afin d'obtenir par exemple des modèles plus parcimonieux. Ceci est illustré dans l'application de la section 3.7. Si $\boldsymbol{\Sigma}_u$ était connu, alors nous pourrions calculer un estimateur par les moindres carrés généralisés du modèle (3.8) fournissant ainsi

$$\hat{\boldsymbol{\gamma}} = (\mathbf{R}'(\mathbf{Z}'\mathbf{Z} \otimes \boldsymbol{\Sigma}_u^{-1})\mathbf{R})^{-1}\mathbf{R}'(\mathbf{Z} \otimes \boldsymbol{\Sigma}_u^{-1})\mathcal{Y}.$$

Puisque l'on connaît rarement $\boldsymbol{\Sigma}_u$ dans les applications, on le remplace par un estimateur convergent afin d'obtenir un estimateur par les moindres carrés calculables. On considère alors l'estimateur $\tilde{\boldsymbol{\Sigma}}_u$ obtenu suite à l'estimation par moindres carrés ordinaires du modèle (3.8):

$$\tilde{\boldsymbol{\Sigma}}_u = n^{-1}\tilde{\mathbf{U}}\tilde{\mathbf{U}}',$$

où $\tilde{\mathbf{u}}$ est tel que $\tilde{\mathbf{u}} = \text{vec}\tilde{\mathbf{U}}$ et

$$\tilde{\mathbf{u}} = \mathcal{Y} - (\mathbf{Z}' \otimes \mathbf{I}_d)\mathbf{R}\tilde{\boldsymbol{\gamma}},$$

et

$$\tilde{\gamma} = (\mathbf{R}'(\mathbf{Z}'\mathbf{Z} \otimes \mathbf{I}_d)\mathbf{R})^{-1}\mathbf{R}'(\mathbf{Z} \otimes \mathbf{I}_d)\mathcal{Y}.$$

Ainsi l'estimateur que l'on considère est obtenu en remplaçant Σ_u par $\tilde{\Sigma}_u$

$$\hat{\gamma} = (\mathbf{R}'(\mathbf{Z}'\mathbf{Z} \otimes \tilde{\Sigma}_u^{-1})\mathbf{R})^{-1}\mathbf{R}'(\mathbf{Z} \otimes \tilde{\Sigma}_u^{-1})\mathcal{Y}.$$

Les résidus du modèle (3.3) sont alors

$$\hat{u}_t = \mathbf{y}_t - \hat{\Lambda}\mathbf{Y}_{t-1} - \hat{\mathbf{V}}\mathbf{X}_{t-1} - \hat{\mathbf{V}}_0\mathbf{x}_t,$$

fournissant comme estimateur de Σ_u

$$\hat{\Sigma}_u = n^{-1}\hat{\mathbf{U}}\hat{\mathbf{U}}',$$

où $\hat{\mathbf{u}} = \text{vec}\hat{\mathbf{U}}$. L'estimateur $\hat{\gamma}$ est sous certaines conditions convergent et asymptotiquement normal. Voir Lütkepohl (1993, pp. 331-332). Pour l'estimateur du vecteur β , on pose alors simplement $\hat{\beta} = \mathbf{R}\hat{\gamma}$ qui est aussi convergent et asymptotiquement normalement distribué. En particulier, l'estimateur $\hat{\beta}$ satisfait la condition suivante

$$\hat{\beta} - \beta = O_p(1/\sqrt{n}). \quad (3.9)$$

Nous avons discuté le cas le plus simple de l'estimation des paramètres d'un modèle VARX écrit sous sa forme réduite lorsque des contraintes linéaires sont présentes. Lütkepohl (1993, pp. 332-333) discute d'autres aspects, tel la présence de contraintes non-linéaires dans les paramètres, l'estimation de la forme structurelle (plutôt que la forme réduite) et le contexte plus général des modèles VARMAX. La méthode d'estimation décrite dans cette section sera celle utilisée dans la simulation de la section 3.6. Cependant, toute méthode d'estimation fournissant des estimateurs satisfaisant la relation (3.9) sera suffisante pour la validité du théorème 2. Ainsi, on formule l'hypothèse suivante:

Hypothèse B: Les estimateurs du modèle (3.8) sont tels que la relation (3.9) est satisfaite.

3.4. TESTS DANS LE MODÈLE LINÉAIRE DYNAMIQUE

Pour un modèle *ARX*, Hong (1996a) a proposé trois classes de statistiques de type port-manteau pour l'hypothèse de bruit blanc. Son approche repose sur une comparaison à l'aide d'une mesure de distance entre la densité spectrale sous l'hypothèse nulle et un estimateur de la densité spectrale calculé par la méthode du noyau.

Le but de cette section est de développer une statistique de test pour l'hypothèse du bruit blanc du processus \mathbf{u} dans le modèle (3.3). Ainsi, les hypothèses d'intérêt peuvent se formuler comme

$$\begin{aligned} H_0: \Gamma_u(j) &= 0, \forall j \neq 0, \text{ versus} \\ H_1: \Gamma_u(j) &\neq 0, \text{ pour un certain } j \neq 0. \end{aligned}$$

Reformulée à l'aide de la densité spectrale, l'hypothèse H_0 est équivalente à $\mathbf{f}(\omega) = \mathbf{f}_0(\omega)$, $\forall \omega \in [-\pi, \pi]$, où $\mathbf{f}_0(\omega) = \Gamma_u(0)/(2\pi)$, pour $\omega \in [-\pi, \pi]$.

Considérons une mesure de divergence pour deux densités spectrales multivariées $\mathbf{f}_1, \mathbf{f}_2$, telle que $D(\mathbf{f}_1; \mathbf{f}_2) \geq 0$ et $D(\mathbf{f}_1; \mathbf{f}_2) = 0$ si et seulement si $\mathbf{f}_1 = \mathbf{f}_2$. On construira alors notre test de H_0 en s'inspirant de la quantité $D(\hat{\mathbf{f}}_n; \mathbf{f}_0)$, où $\hat{\mathbf{f}}_n$ est donné par (3.5). La norme que l'on va considérer est la norme quadratique suivante

$$\begin{aligned} Q^2(\mathbf{f}_1; \mathbf{f}_2) &= 2\pi \int_{-\pi}^{\pi} \text{vec}[\bar{\mathbf{f}}_1 - \bar{\mathbf{f}}_2]' \Gamma_u(0)^{-1} \otimes \Gamma_u(0)^{-1} \text{vec}[\mathbf{f}_1 - \mathbf{f}_2] d\omega, \\ &= 2\pi \int_{-\pi}^{\pi} \text{tr}[(\bar{\mathbf{f}}_1 - \bar{\mathbf{f}}_2)' \Gamma_u^{-1}(0) (\mathbf{f}_1 - \mathbf{f}_2) \Gamma_u^{-1}(0)] d\omega, \\ &= 2\pi \int_{-\pi}^{\pi} \text{tr}[\Gamma_u^{-1}(0) (\bar{\mathbf{f}}_1 - \bar{\mathbf{f}}_2)' \Gamma_u^{-1}(0) (\mathbf{f}_1 - \mathbf{f}_2)] d\omega. \end{aligned} \quad (3.10)$$

où $\mathbf{f}^* = \bar{\mathbf{f}}'$, $\bar{\mathbf{f}}$ étant la conjuguée de \mathbf{f} . La seconde identité a été obtenue par le théorème 16.2.2 dans Harville (1997).

On peut motiver cette mesure de distance de la manière suivante. Pour ω fixé,

$$Q_{\omega}^2(\mathbf{f}_1; \mathbf{f}_2) = \text{vec}(\mathbf{f}_1(\omega) - \mathbf{f}_2(\omega))^* (\Gamma_u(0)^{-1} \otimes \Gamma_u(0)^{-1}) \text{vec}(\mathbf{f}_1(\omega) - \mathbf{f}_2(\omega))$$

est une mesure standardisée entre les densités spectrales \mathbf{f}_1 et \mathbf{f}_2 évaluées à la fréquence ω . Nous avons alors que (3.10) est simplement une mesure intégrée sur l'ensemble des fréquence $[-\pi, \pi]$.

Il est intéressant de noter que $Q^2(\mathbf{f}; \mathbf{f}_0)$ peut s'écrire alternativement en terme de la fonction d'autocovariance, comme montré dans le résultat suivant:

Proposition 1. *Considérons la mesure de distance $Q^2(\mathbf{f}; \mathbf{f}_0)$ donnée par (3.10), où \mathbf{f} est donné dans l'expression (3.4) et $\mathbf{f}_0 = \Gamma_u(0)/(2\pi)$. Nous avons alors que*

$$Q^2(\mathbf{f}; \mathbf{f}_0) = 2 \sum_{h=1}^{\infty} \text{tr}[\Gamma_u(h) \Gamma_u(0)^{-1} \Gamma_u(h)' \Gamma_u(0)^{-1}]. \quad (3.11)$$

DÉMONSTRATION. En reprenant les détails menant à la formule (3.10), nous pouvons écrire $Q_\omega^2(\mathbf{f}; \mathbf{f}_0)$ avec l'aide de la trace:

$$Q_\omega^2(\mathbf{f}; \mathbf{f}_0) = \text{tr}[\mathbf{\Gamma}_u^{-1}(0)(\bar{\mathbf{f}}(\omega) - \bar{\mathbf{f}}_0(\omega))'\mathbf{\Gamma}_u^{-1}(0)(\mathbf{f}(\omega) - \mathbf{f}_0(\omega))].$$

Puisque $\mathbf{\Gamma}_u(0)$ est définie positive, nous pouvons trouver une matrice \mathbf{L} telle que $\mathbf{\Gamma}_u^{-1}(0) = \mathbf{L}\mathbf{L}'$ et alors nous pouvons écrire

$$\begin{aligned} Q_\omega^2(\mathbf{f}; \mathbf{f}_0) &= \text{tr}[(\mathbf{L}'(\mathbf{f}(\omega) - \mathbf{f}_0(\omega))\mathbf{L})(\mathbf{L}'(\mathbf{f}(\omega) - \mathbf{f}_0(\omega))\mathbf{L})^*], \\ &= \text{tr}[(\mathbf{f}_L - \mathbf{I}_d/(2\pi))(\mathbf{f}_L - \mathbf{I}_d/(2\pi))^*], \end{aligned}$$

où A^* dénote la conjuguée transposée de la matrice A , et $\mathbf{f}_L = \mathbf{L}'\mathbf{f}\mathbf{L}$. En intégrant $Q_\omega^2(\mathbf{f}; \mathbf{f}_0)$ entre $-\pi$ et π , on trouve que

$$\begin{aligned} Q^2(\mathbf{f}; \mathbf{f}_0) &= 2\pi \int_{-\pi}^{\pi} Q_\omega^2(\mathbf{f}; \mathbf{f}_0) d\omega, \\ &= \sum_{h=-\infty}^{\infty} \text{tr}[\mathbf{\Gamma}_u(h)\mathbf{\Gamma}_u(0)^{-1}\mathbf{\Gamma}_u(h)'\mathbf{\Gamma}_u(0)^{-1}] - d, \\ &= 2 \sum_{h=1}^{\infty} \text{tr}[\mathbf{\Gamma}_u(h)\mathbf{\Gamma}_u(0)^{-1}\mathbf{\Gamma}_u(h)'\mathbf{\Gamma}_u(0)^{-1}]. \end{aligned}$$

□

Ainsi, Q^2 est une mesure globale tenant compte de l'ensemble des délais. En refaisant le même genre de calcul, on montre que lorsque l'on remplace \mathbf{f} par $\hat{\mathbf{f}}_n$, on obtient

$$\begin{aligned} Q^2(\hat{\mathbf{f}}_n; \mathbf{f}_0) &= \sum_{j \leq |n-1|} k^2(j/p_n) \text{tr}[\mathbf{C}'_{\hat{u}}(j)\mathbf{\Gamma}_u^{-1}(0)\mathbf{C}_{\hat{u}}(j)\mathbf{\Gamma}_u^{-1}(0)] - 2\text{tr}[\mathbf{\Gamma}_u^{-1}(0)\mathbf{C}'_{\hat{u}}(0)] + d, \\ &= 2 \sum_{j=1}^{n-1} k^2(j/p_n) \text{tr}[\mathbf{C}'_{\hat{u}}(j)\mathbf{\Gamma}_u^{-1}(0)\mathbf{C}_{\hat{u}}(j)\mathbf{\Gamma}_u^{-1}(0)] + \\ &\quad \text{tr}[\mathbf{C}'_{\hat{u}}(0)\mathbf{\Gamma}_u^{-1}(0)\mathbf{C}_{\hat{u}}(0)\mathbf{\Gamma}_u^{-1}(0)] - 2\text{tr}[\mathbf{\Gamma}_u^{-1}(0)\mathbf{C}'_{\hat{u}}(0)] + d, \\ &= 2 \sum_{j=1}^{n-1} k^2(j/p_n) \text{tr}[\mathbf{C}'_{\hat{u}}(j)\mathbf{C}_{\hat{u}}^{-1}(0)\mathbf{C}_{\hat{u}}(j)\mathbf{C}_{\hat{u}}^{-1}(0)] + o_p(\sqrt{p_n}/n), \end{aligned}$$

ce résultat se déduisant facilement des relations (3.34) et (3.36) dans la section 3.8. Le test proposé est essentiellement une version standardisée de $Q^2(\hat{\mathbf{f}}_n; \mathbf{f}_0)$, donné par

$$T_n = \frac{n \sum_{j=1}^{n-1} k^2(j/p_n) \text{tr}[\mathbf{C}'_{\hat{u}}(j)\mathbf{C}_{\hat{u}}^{-1}(0)\mathbf{C}_{\hat{u}}(j)\mathbf{C}_{\hat{u}}^{-1}(0)] - d^2 M_n(k)}{\sqrt{2d^2 V_n(k)}} \quad (3.12)$$

où $M_n(k)$ et $V_n(k)$ sont donnés par (3.1) et (3.2). En effet, on constate que l'on a la relation suivante:

$$T_n = \frac{\frac{n}{2}Q^2(\hat{\mathbf{f}}_n; \mathbf{f}_0) - d^2 M_n(k)}{[2d^2 V_n(k)]^{1/2}} + o_p(1). \quad (3.13)$$

Notons que Hong (1996a) discute de statistiques où l'on remplace dans la statistique (3.12) les valeurs $M_n(k)$ et $V_n(k)$ par $p_n M(k)$ et $p_n V(k)$, respectivement, où $M(k) = \int_0^\infty k^2(z) dz$ et $V(k) = \int_0^\infty k^4(z) dz$. En effet, avec $p_n \rightarrow \infty$, $p_n/n \rightarrow 0$, il est possible de montrer sous certaines conditions sur le noyau et sur p_n , que $p_n^{-1} V_n(k) \rightarrow V(k)$ et $p_n^{-1} M_n(k) = M(k) + o(p_n^{-1/2})$. Voir Hong (1996a, p. 840). En particulier, $M_n(k) = O(p_n)$ et $V_n(k) = O(p_n)$. Ainsi, ces changements n'affectent pas la distribution asymptotique. Ces remplacements peuvent offrir de légers avantages de calcul de la statistique de test, et potentiellement une meilleure approximation en échantillons finis.

La statistique portmanteau multivariée de Hosking (1980) peut être interprétée comme un test basé sur T_n , à des facteurs de normalisation près, utilisant le périodogramme tronqué. Elle est obtenue selon la formule suivante

$$P = n \sum_{j=1}^M \text{tr}[\mathbf{C}'_{\hat{u}}(j) \mathbf{C}_{\hat{u}}^{-1}(0) \mathbf{C}_{\hat{u}}(j) \mathbf{C}_{\hat{u}}^{-1}(0)],$$

où M est une constante fixée correspondant au nombre de délais tenu en compte dans la statistique. Cette statistique a été étudiée notamment dans Chitturi (1974) dans les modèles VAR, et étendue aux modèles VARMA par Hosking (1980). Cette statistique est discutée également dans Lütkepohl (1993).

Ceci nous amène au théorème 2:

Théorème 2. *Sous l'hypothèse que le processus \mathbf{y} est VAR(r, s) tel que défini dans la définition 1 et que les hypothèses A et B sont satisfaites, $p_n \rightarrow \infty$ et $p_n/n \rightarrow 0$, alors la statistique T_n définie par (3.12) converge en loi vers une normale centrée réduite, c'est-à-dire que*

$$T_n \rightarrow_L N(0, 1).$$

Nous allons donner ici seulement les deux grandes étapes de la preuve. La preuve détaillée se retrouve dans l'appendice. La démonstration respecte les deux mêmes grandes étapes que Hong (1996a). On commence par montrer la normalité asymptotique d'une pseudo-statistique ayant la même forme que T_n , mais construite avec l'aide du bruit blanc indépendant \mathbf{u} inobservable.

Étape 1.

$$\frac{n \sum_{j=1}^{n-1} k^2(j/p_n) \text{tr}[\mathbf{C}_u^{-1}(0) \mathbf{C}_u(j) \mathbf{C}_u^{-1}(0) \mathbf{C}'_u(j)] - d^2 M_n(k)}{\sqrt{2d^2 V_n(k)}} \rightarrow_L N(0, 1). \quad (3.14)$$

Nous montrons l'étape 1 en utilisant le théorème 1. Notons que ce résultat est établi sans l'intervention d'aucun modèle.

On fait intervenir les données observées et le modèle dans la seconde étape.

Étape 2.

$$\sum_{j=1}^{n-1} k^2(j/p_n) \{ \text{tr}[\mathbf{C}_u^{-1}(0) \mathbf{C}_u(j) \mathbf{C}_u^{-1}(0) \mathbf{C}'_u(j)] - \text{tr}[\mathbf{C}_{\hat{u}}^{-1}(0) \mathbf{C}_{\hat{u}}(j) \mathbf{C}_{\hat{u}}^{-1}(0) \mathbf{C}'_{\hat{u}}(j)] \} = o_p(\sqrt{p_n/n}). \quad (3.15)$$

Ainsi, on montre essentiellement que la différence entre T_n et la version reposant sur le processus inobservable \mathbf{u} est $o_p(1)$.

3.5. PUISSANCE DU TEST DE HONG

Dans le contexte des modèles ARX, Hong (1996a) a étudié la puissance asymptotique des tests qu'il propose. L'analyse lui a permis de conclure (entre autres) que ses tests sont convergents. Cependant, lorsqu'un test est convergent, sa puissance asymptotique tend vers un lorsque le nombre d'observations tend vers l'infini. Un objectif étant précisément de proposer des tests offrant de bonnes propriétés de puissance avec un noyau adéquatement choisi, on doit trouver une façon de comparer les puissances asymptotiques en fonction des noyaux. Afin de réaliser cet objectif, il faut obtenir une puissance asymptotique strictement inférieure à un. Deux types d'analyse permettent d'obtenir des puissances inférieures à un. La première est dite analyse locale alors que la seconde est dite analyse globale. Nous allons rappeler les résultats obtenus par Hong (1996a), et allons les étendre dans le contexte multidimensionnel. Le but principal de cette section est d'établir la convergence des tests généralisés et aussi de déterminer le noyau optimal.

3.5.1. Puissance asymptotique locale

L'analyse locale est aussi connue sous le nom d'analyse de Pitman (1979). Elle consiste essentiellement à garder le niveau fixé et à considérer une suite d'hypothèses alternatives convergeant vers l'hypothèse nulle lorsque $n \rightarrow \infty$.

Considérons les hypothèses alternatives suivantes:

$$H_{1n} : f_{0n}(\omega) = f_0 + a_n g(\omega),$$

où f_0 est la densité spectrale sous l'hypothèse nulle, et où g est une fonction symétrique, périodique (de période 2π), continue, bornée et telle que $\int_{-\pi}^{\pi} g(\omega) d\omega = 0$. En considérant $a_n = p_n^{1/4}/n^{1/2}$, il est possible de montrer que

$$M_{1n}^a = \frac{\frac{n}{2} Q^2(\hat{f}_n; f_{0n}) - M_n(k)}{\sqrt{2V_n(k)}} \rightarrow_L N(\mu(k), 1),$$

où $\mu(k) = 2\pi \int_{-\pi}^{\pi} g^2(\omega) d\omega / (2V(k))^{1/2}$.

La puissance asymptotique est alors

$$\lim_{n \rightarrow \infty} Pr(M_{1n}^a > z_{1-\alpha}) = 1 - \Phi(z_{1-\alpha} - \mu(k)), \quad (3.16)$$

où Φ est la fonction de répartition d'une loi $N(0, 1)$, et $z_{1-\alpha}$ est son quantile d'ordre $1 - \alpha$. Puisque (3.16) dépend de k , on peut alors s'attendre à ce que les puissances asymptotiques varient selon le noyau utilisé. En reprenant un raisonnement de Pitman (1979), si on suppose que $p_n = cn^\nu$, pour $0 < \nu < 1$, il est possible de calculer l'efficacité relative asymptotique du noyau k_2 par rapport au noyau k_1 , et on peut montrer qu'elle devient

$$ARE_P(k_2; k_1) = [V(k_1)/V(k_2)]^{1/(2-\nu)}.$$

Ainsi, $ARE_P(k_B; k_T) > 2.23$, où k_B est le noyau de Bartlett, et k_T est le noyau tronqué. Ceci permet d'affirmer que la puissance peut être supérieure en adoptant un noyau autre que le noyau tronqué.

Il est aussi possible de maximiser la puissance (3.16) comme une fonction de k , dans la classe

$$\kappa(\tau) = \{k(\cdot) \text{ satisfait l'hypothèse A avec } k^{(2)} = \tau^2/2 > 0\}, \quad (3.17)$$

où $k^{(r)} = \lim_{z \rightarrow 0} (1 - k(z))/|z|^r$ est l'indice caractéristique de k . La classe $\kappa(\tau)$ contient les noyaux de Daniell, Parzen et spectral quadratique. Elle exclut cependant les noyaux tronqués, de Bartlett et le noyau général de Tukey. En optimisant la puissance asymptotique, il est alors possible de montrer que le noyau de Daniell est le noyau optimal dans la classe (3.17).

Dans un contexte multidimensionnel, il est naturel de considérer la suite d'hypothèses alternatives suivante:

$$H_{1n} : \mathbf{f}_{0n}(\omega) = \mathbf{f}_0 + a_n \mathbf{g}(\omega), \quad (3.18)$$

où $\mathbf{f}_0 = \Gamma_u(0)/(2\pi)$, et \mathbf{g} est une matrice hermitienne de dimension $d \times d$, telle que $\int_{-\pi}^{\pi} \mathbf{g}(\omega) d\omega = \mathbf{0}$. On considère le taux $a_n = p_n^{1/4}/n^{1/2}$. On peut alors montrer le théorème suivant, qui est une généralisation d'une partie d'un théorème de Hong (1996a).

Théorème 3. *Soit $\hat{\mathbf{f}}$ donné par (3.5) et considérons la suite d'hypothèses alternatives H_{1n} données par (3.18). Sous les hypothèses A et B, $p_n \rightarrow \infty$ et $p_n/n \rightarrow 0$, on a que*

$$\frac{\frac{n}{2} Q^2(\hat{\mathbf{f}}_n; \mathbf{f}_{0n}) - d^2 M_n(k)}{\sqrt{2d^2 V_n(k)}} \rightarrow_L N(\mu(k), 1),$$

où $\mu(k) = 2\pi \int_{-\pi}^{\pi} \text{tr}[\Gamma_u(0)^{-1} \mathbf{g}^*(\omega) \Gamma_u(0)^{-1} \mathbf{g}(\omega)] d\omega / [2d^2 V(k)]^{1/2}$.

Ainsi, la forme de la puissance asymptotique est la même que dans le cas ARX univarié et est donné par (3.16). Il n'y a que le numérateur de $\mu(k)$ qui est différent. Cependant, l'argument de Hong (1996a, p. 861) pour établir son théorème 5 s'applique intégralement, ce qui nous permet d'affirmer que le noyau optimal est le noyau de Daniell, obtenu par la minimisation de $V(k)$ dans la classe (3.17).

3.5.2. Puissance asymptotique globale

Afin de comparer ses diverses classes de tests, Hong (1996a), considère une analyse de puissance globale. Cette analyse lui permet de comparer les différents tests obtenus en choisissant la métrique quadratique, la métrique de Hellinger et le test basé sur l'information de Kullback-Leibler. Aussi, cela lui permet d'étudier la convergence des tests et le taux de convergence vers l'infini des statistiques de test. Il considère également le calcul des pentes asymptotiques de Bahadur (1960), permettant entre autres la comparaison des tests paramétriques et non paramétriques.

Pour étudier la convergence des tests sous l'hypothèse alternative, considérons le modèle de régression statique

$$y_t = c + \sum_{j=1}^q \nu^{(j)}(B) x_{jt} + u_t,$$

où les $\nu^{(j)}(B)$ sont des polynômes en B . Ainsi, c'est un modèle ARX mais avec une absence de retards de la variable y_t parmi les variables explicatives. On aura alors dans les développements que l'estimateur des moindres carrés des paramètres du modèle de régression statique est convergent sous l'hypothèse alternative, même s'il est inefficace. Supposons que le processus

$\{u_t, t \in \mathbb{Z}\}$ est tel que $\sum \rho^2(j) < \infty$, avec la structure de dépendance suivante

$$\sum_i \sum_j \sum_l |\kappa(t, t+i, t+j, t+l)| < \infty.$$

Sous l'hypothèse alternative, il est possible de montrer que

$$\frac{p_n^{1/2}}{n} M_{1n} \rightarrow_P \frac{1}{2} Q^2(\mathbf{f}; \mathbf{f}_0) / [2V(k)]^{1/2}.$$

Nous allons nous attarder à généraliser ce théorème pour les modèles VARX. Dans cette perspective, nous considérons le modèle VARX statique suivant:

$$\mathbf{y}_t = \mathbf{c} + \mathbf{V}(B)\mathbf{x}_t + \mathbf{u}_t, t \in \mathbb{Z}. \quad (3.19)$$

Les hypothèses sur le processus sont regroupées dans l'hypothèse C.

Hypothèse C: Supposons alors que la structure de dépendance du processus \mathbf{u} soit telle que $\sum_j \|\Gamma_u(j)\|^2 < \infty$ et supposons que les cumulants satisfont la condition

$$\sum_i \sum_j \sum_l |\kappa_{pqrs}(t, t+i, t+j, t+l)| < \infty,$$

où $p, q, r, s \in \{1, \dots, d\}$.

Nous pouvons alors montrer le résultat suivant:

Théorème 4. *Considérons le modèle (3.19). Soit T_n donnée dans l'équation (3.12). Supposons aussi que les hypothèses A et B sont satisfaites, que $p_n \rightarrow \infty$ et $p_n/n \rightarrow 0$. Sous l'hypothèse alternative $H_1 : \Gamma_u(j) \neq 0$, pour un certain $j \neq 0$, supposant que la structure de dépendance est telle que décrite dans l'hypothèse C, nous avons alors*

$$\frac{p_n^{1/2}}{n} T_n \rightarrow_P \frac{1}{2} Q^2(\mathbf{f}; \mathbf{f}_0) / [2d^2 V(k)]^{1/2},$$

où \mathbf{f} est la densité spectrale donnée par (3.4).

Ainsi, à la lumière du théorème 4, nous pouvons conclure que le test basé sur (3.12) est convergent. Le taux auquel la statistique (3.12) diverge à l'infini est donné par $n/\sqrt{p_n}$. Ainsi, si p_n tend plus lentement vers l'infini, nous aurons que (3.12) tendra plus rapidement vers l'infini, et cela résultera en un test plus puissant. Voir aussi Hong (1996a, p. 847).

3.6. SIMULATIONS

Nous avons proposé dans la section 3.4 une statistique pour tester l'hypothèse d'absence de corrélation dans le terme d'erreur d'un modèle $VARX$. Cette statistique généralise au cas multivarié la démarche de Hong (1996a). De plus, la statistique proposée généralise la statistique portmanteau de Hosking (1980) pour cette classe de modèles.

La statistique de test possède une distribution asymptotique normale. On se propose dans cette section d'étudier par simulation le comportement des distributions exactes de la statistique pour différents noyaux et de les comparer à la distribution asymptotique normale, lorsque les observations proviennent d'un modèle $VARX$. À cette fin, nous avons calculé les niveaux empiriques sous l'hypothèse nulle $H_0 : \Gamma(j) = \mathbf{0}, \forall j \neq 0$, lorsque nous utilisons les quantiles issus de la théorie asymptotique. Nous avons utilisé les niveaux nominaux 1%, 5%, 10%, qui sont largement utilisés en pratique. Afin d'évaluer l'influence sur les résultats de la longueur de la série chronologique, nous avons généré pour l'étude de niveau 10000 réalisations de séries de longueur $n = 50, 100, 200$. Les écarts type (en pourcentage) correspondant à 10000 réalisations pour les trois niveaux 1%, 5%, 10% sont respectivement 0.099, 0.218 et 0.300. Les zones de non-rejet de l'hypothèse nulle ' H_0 : niveau est α (en pourcentage)' au niveau 5% (notée NR5) et au niveau 1% (notée NR1) sont données dans le tableau 3.1.

TABLEAU 3.1. Zones de non-rejet (en pourcentage) de ' H_0 : niveau est α (en pourcentage)', basées sur 10000 réalisations.

α	NR5	NR1
1	[0.8, 1.2]	[0.7, 1.3]
5	[4.6, 5.4]	[4.4, 5.6]
10	[9.4, 10.6]	[9.2, 10.8]

Nous avons également évalué par simulation la puissance des nouveaux tests. Nous avons ainsi généré des observations selon un modèle $VARX$, mais avec un terme d'erreur qui n'est pas un bruit blanc. Pour l'étude de puissance, nous avons généré 2000 réalisations des processus. Dans un premier temps, nous avons calculé les taux de rejets de l'hypothèse nulle en utilisant les valeurs critiques asymptotiques. Cependant, puisque les niveaux exacts sont différents des niveaux nominaux, risquant de fausser les comparaisons, nous avons également utilisé les 10000 réalisations de l'étude de niveau afin de calculer les valeurs critiques exactes des tests. Ceci a permis de calculer les puissances utilisant les points critiques empiriques.

3.6.1. Choix des modèles

Afin d'évaluer la performance de la statistique avec différents noyaux, nous avons généré des observations selon le modèle suivant:

$$\mathbf{y}_t = \mathbf{c} + \mathbf{\Lambda}_1 \mathbf{y}_{t-1} + \mathbf{V}_0 \mathbf{x}_t + \mathbf{V}_1 \mathbf{x}_{t-1} + \mathbf{a}_t, \quad (3.20)$$

où $\mathbf{x}_t = \mathbf{\Phi}_x \mathbf{x}_{t-1} + \mathbf{b}_t$ et \mathbf{b}_t est iid $N_2(\mathbf{0}, \mathbf{\Sigma}_b)$. Nous avons considéré deux situations pour le terme d'erreur \mathbf{a}_t , à savoir

- a) $\mathbf{a}_t = \mathbf{e}_t$,
- b) $\mathbf{a}_t = \mathbf{e}_t - \mathbf{\Theta}_\delta \mathbf{e}_{t-1}$,

où \mathbf{e}_t est un bruit blanc indépendant $N_2(\mathbf{0}, \mathbf{\Sigma}_a)$. Les paramètres choisis sont

$$\mathbf{\Lambda}_1 = \begin{pmatrix} -0.5 & 0.5 \\ -1.4 & -0.2 \end{pmatrix}, \quad \mathbf{V}_0 = \begin{pmatrix} 0.0 & 0.3 \\ 0.1 & 0.6 \end{pmatrix}, \quad \mathbf{V}_1 = \begin{pmatrix} 0.7 & 0.0 \\ 0.0 & 0.0 \end{pmatrix}, \quad \mathbf{\Phi}_x = \begin{pmatrix} -1.5 & 1.2 \\ -0.9 & 0.5 \end{pmatrix},$$

$$\mathbf{\Theta}_\delta = \begin{pmatrix} 0.18\delta & 0.04\delta \\ 0.0 & 0.02\delta \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} 3.0 \\ 2.0 \end{pmatrix}, \quad \mathbf{\Sigma}_a = \begin{pmatrix} 1.0 & 0.75 \\ 0.75 & 1.0 \end{pmatrix}, \quad \mathbf{\Sigma}_b = \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix},$$

La situation a) permet d'évaluer le niveau des tests, alors que le cas b) permet de considérer des situations où \mathbf{a}_t n'est pas un bruit blanc, et ainsi nous permet d'évaluer la puissance. Le paramètre δ permet de faire varier la dépendance dans \mathbf{a}_t . Nous avons considéré $\delta = 0.75, 1.00, 1.25$. En effet, la fonction d'autocovariance $\mathbf{\Gamma}_a$ est donnée par

$$\mathbf{\Gamma}_a(j) = \begin{cases} \mathbf{\Sigma}_a + \mathbf{\Theta}_\delta \mathbf{\Sigma}_a \mathbf{\Theta}'_\delta & j = 0, \\ -\mathbf{\Theta}_\delta \mathbf{\Sigma}_a & j = 1, \\ \mathbf{0} & j > 1. \end{cases}$$

Afin d'illustrer la procédure d'estimation de la section 3.3 avec des contraintes linéaires sur les paramètres, nous avons supposé que les paramètres nuls dans les matrices \mathbf{V}_0 et \mathbf{V}_1 étaient connus. Ainsi, ces valeurs ont été fixées à 0 durant la simulation. Afin de calculer les estimateurs des moindres carrés dans ce modèle, il suffisait de choisir judicieusement la matrice \mathbf{R} de la section 3.3.

Pour le modèle (3.20), 10000 réalisations de taille n sont générées. Une réalisation de taille n est obtenue en supposant que le processus conjoint $\mathbf{z}_t = (\mathbf{y}'_t, \mathbf{x}'_t)'$ est VARMA gaussien. Afin de s'assurer que la structure de corrélation pour les valeurs initiales de la série \mathbf{z}_t et le reste des observations est la même, \mathbf{z}_1 est générée suivant une distribution gaussienne centrée de covariance $\mathbf{\Gamma}_z(0)$ obtenue à l'aide de l'algorithme de Ansley (1980).

Les statistiques T_n reposent sur le choix d'un noyau. Nous considérons le noyau quadratique spectral (QS), le noyau de Bartlett (BAR), le noyau de Daniell (DAN), le noyau de

Parzen (PAR) et le noyau tronqué (TR). Un paramètre à déterminer est la valeur de p_n . Nous avons considéré pour $n = 50$ les choix $p_n = 4, 8, 10$; pour $n = 100$ les valeurs $p_n = 5, 9, 12$ et finalement $n = 200$ les valeurs $p_n = 6, 10, 15$. Ces taux correspondent à $p_n = \lceil \log(n) \rceil$, $p_n = \lceil 3.5n^{0.2} \rceil$ et $p_n = \lceil 3n^{0.3} \rceil$. Les choix des taux sont discutés dans Hong (1996a).

Une procédure de sélection automatique de p_n est souhaitable. Dans un contexte unidimensionnel, Hong (1996a) a utilisé la procédure de Beltrao et Bloomfield (1987) et de Robinson (1991). Cette dernière est une procédure de validation croisée. Elle est basée sur la minimisation d'un critère de pseudo-log-vraisemblance. Les résultats de Beltrao et Bloomfield (1987) se situent dans le cas gaussien, alors que Robinson (1991) a étendu considérablement l'approche au cas non-gaussien. Il est alors montré que le choix de p_n est convergent pour un choix de largeur de fenêtre optimale sous une certaine erreur quadratique moyenne intégrée, cette dernière mesurant l'ajustement entre l'estimateur de la densité spectrale et la densité spectrale théorique. Voir Robinson (1991, pp. 1343-1345). Dans un contexte d'estimation de la densité spectrale multidimensionnelle, les résultats disponibles sont à notre connaissance beaucoup plus rares. Cependant, Robinson (1991, p. 1346) discute l'extension de ses résultats dans ce contexte, et offre plusieurs solutions pratiques. Pour notre étude de simulation, nous avons minimisé l'analogue multivarié discuté dans Robinson (1991), et qui en donne une justification heuristique. Plus précisément, nous avons minimisé par rapport à m la pseudo-log-vraisemblance $\sum_{j=1}^n [\log \det \hat{\mathbf{f}}_{(j)}^m(\lambda_j) + \text{tr}\{\mathbf{I}(\lambda_j) \hat{\mathbf{f}}_{(j)}^m(\lambda_j)^{-1}\}]$, où $\lambda_j = 2\pi j/n$, $\hat{\mathbf{f}}_{(j)}^m(\cdot)$ étant un estimateur de la densité spectrale basé sur le périodogramme lissé de type *leave-two-out* et $\mathbf{I}(\cdot)$ étant le périodogramme. Pour l'optimisation, nous avons considéré la palette de valeurs $m = 2, 3, \dots, 20$.

3.6.2. Discussion de l'étude de niveau

Les résultats de l'étude de niveau se retrouvent dans le tableau 3.2. On note que pour des séries de taille $n = 50$, nous obtenons des résultats très moyens pour les niveaux 1% et 5%, mais que dans l'ensemble les résultats sont très bons pour le niveau 10%, lorsque le noyau est autre que le noyau tronqué. Ainsi, on constate que les niveaux observés sont nettement à l'extérieur des zones de non-rejet NR1, avec un sur-rejet marqué. Cependant, pour le niveau 10%, les niveaux empiriques observés se situent très près des zones de non-rejet NR1, et pour $p_n = 4$ se situent dans les zones de non-rejet NR5, lorsque le noyau est autre que le noyau tronqué. La procédure de validation croisée ici semble fournir des sur-rejets pour les trois niveaux nominaux.

TABLEAU 3.2. Niveaux en pourcentage de la statistique de Hong multivariée avec le modèle donné dans la formule 3.20

	$\alpha = 1\%$					$\alpha = 5\%$					$\alpha = 10\%$				
$n = 50$															
p_n	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
4	2.4	2.5	2.4	2.4	2.4	6.5	6.5	6.6	6.6	6.5	10.2	10.1	10.2	10.3	11.0
8	2.3	2.5	2.4	2.4	2.8	6.8	6.7	6.8	6.9	7.4	10.9	10.9	11.1	11.1	12.3
10	2.5	2.4	2.6	2.6	3.0	7.0	6.9	6.8	7.2	8.0	11.3	11.1	11.4	11.6	12.4
CV	2.9	3.2	2.9	2.8	ND	7.3	7.7	7.2	7.2	ND	11.5	12.2	11.6	11.5	ND
$n = 100$															
p_n	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
5	2.2	2.2	2.2	2.1	2.1	5.7	5.7	5.7	5.7	5.7	9.6	9.4	9.6	9.6	9.8
9	2.2	2.2	2.2	2.3	2.2	5.6	5.6	5.5	5.7	6.1	9.7	9.7	9.6	9.5	10.1
12	2.2	2.2	2.3	2.2	2.2	5.8	5.7	5.9	5.9	6.3	9.9	9.7	9.9	10.1	10.5
CV	2.7	3.0	2.7	2.7	ND	6.7	7.5	6.6	7.1	ND	10.8	11.9	10.7	11.6	ND
$n = 200$															
p_n	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
6	2.1	2.3	2.1	2.1	1.8	5.3	5.4	5.2	5.2	5.1	8.6	8.7	8.5	8.5	9.0
10	1.8	1.8	1.9	1.9	1.7	5.1	5.4	5.2	5.1	5.3	8.8	8.7	8.8	8.7	9.2
15	1.8	1.8	1.8	1.7	1.9	5.2	5.0	5.3	5.3	5.8	8.8	8.6	8.8	8.9	9.8
CV	2.9	3.4	2.8	3.0	ND	6.7	7.7	6.5	7.4	ND	10.6	11.6	10.3	11.8	ND

L'effet de la taille de la série semble jouer un rôle non-négligeable. Ainsi, pour $n = 100$, nous observons de meilleurs résultats pour les niveaux 5% et 10%. Un léger sur-rejet est observé pour le niveau 5%. Les niveaux observés sont soit à l'intérieur des zones de non-rejet NR1, ou très près de ces intervalles. Le niveau 10% est quant à lui très bien contrôlé. On observe de meilleurs niveaux avec une grande valeur de p_n , et tous les niveaux sont dans les zones de non-rejet NR5. La procédure de validation croisée fait raisonnablement bien, particulièrement pour les noyaux QS et de Daniell.

Pour $n = 200$, nous obtenons de très bons résultats pour les niveaux 5%. Ainsi, pour les trois valeurs de p_n , les niveaux observés se situent à l'intérieur des zones NR5. Nous obtenons cependant une légère sous-estimation pour les niveaux 10%. La procédure de validation croisée indique un sur-rejet pour le niveau 5%, mais fonctionne raisonnablement bien pour le niveau 10%.

3.6.3. Discussion de l'étude de puissance

Lorsque le terme d'erreur satisfait $\mathbf{a}_t = \mathbf{e}_t - \Theta_\delta \mathbf{e}_{t-1}$, pour $\delta = 0.75, 1.00, 1.25$ nous sommes alors dans des situations où le terme d'erreur n'est pas un bruit blanc. Plus δ est élevé, plus la dépendance est forte et plus le test semble avoir de la facilité à la détecter, comme en font foi les tableaux 3.3, 3.4 et 3.5. Nous avons calculé la puissance en utilisant les points critiques asymptotiques ainsi que les points critiques exacts obtenus de l'étude de niveau. Les puissances basées sur les points critiques empiriques sont entre parenthèses.

Les résultats des tableaux 3.3, 3.4 et 3.5 montrent que les puissances se comportent essentiellement de la même manière, pour un noyau et une valeur de p_n donnés, lorsque le noyau est autre que le noyau tronqué. Les résultats pour les noyaux DAN, PAR et QS sont plutôt similaires. Lorsque p_n est fixé, et surtout pour $n = 100$ et $n = 200$, il semble que le noyau de DAN soit légèrement supérieur. Plus particulièrement, DAN et QS se comportent essentiellement de la même façon, et semblent légèrement plus puissants que PAR. Ce phénomène semble perceptible tant en utilisant les quantiles asymptotiques qu'empiriques. Les conclusions ne sont pas les mêmes lorsque la procédure de validation croisée est utilisée, où le noyau QS semble être plus puissant que DAN et PAR. Le noyau BAR semble plus puissant que les autres, ce qui n'est pas en désaccord avec l'affirmation entourant l'optimalité du noyau DAN, puisque le noyau BAR ne remplit pas les conditions du théorème. Ceci est en accord avec les analyses de Hong (1996a).

Les résultats basés sur les quantiles empiriques et asymptotiques diffèrent de manière notable pour les niveaux 1% pour $n = 50$. Les résultats sont plus similaires pour les niveaux 5% et 10%, et les résultats s'améliorent lorsque $n = 100$ et sont satisfaisants lorsque $n = 200$.

Puisque la dépendance du terme d'erreur est d'ordre un, il semble intuitivement raisonnable que les tests accordant beaucoup d'importance aux petits délais seront plus puissants que ceux prenant en compte inefficacement un grand nombre de délais. C'est confirmé dans notre étude, puisque nous remarquons qu'une valeur de p_n plus faible semble fournir des puissances plus élevées. La procédure de validation croisée semble très bien fonctionner ici, puisque les puissances obtenues sont plus élevées. Il ressort que le noyau tronqué est celui pour lequel on obtient la puissance la plus faible. Ainsi, un choix autre que le noyau tronqué est approprié. La validation croisée de Robinson (1991) semble fournir ici de très bons résultats, fournissant une sorte de compromis entre les erreurs de type I et de type II.

TABLEAU 3.3. Puissance en pourcentage basée sur les quantiles asymptotiques (empiriques) lorsque $\delta = 0.75$

		$\alpha = 1\%$					$\alpha = 5\%$					$\alpha = 10\%$				
$n = 50$	p_n	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
	4	7.1 (3.6)	7.4 (3.6)	7.0 (3.4)	6.3 (3.6)	4.3 (2.1)	14.3 (11.6)	14.8 (12.1)	14.4 (11.5)	13.7 (11.4)	9.7 (7.6)	20.8 (20.6)	21.1 (21.0)	21.0 (20.6)	20.0 (19.8)	16.3 (15.1)
	8	5.3 (2.8)	5.5 (3.1)	5.3 (2.7)	5.2 (2.6)	4.4 (1.7)	11.5 (9.0)	12.0 (9.2)	11.6 (9.0)	11.0 (8.6)	10.7 (7.4)	17.3 (15.9)	18.4 (16.7)	17.0 (15.5)	17.8 (15.6)	16.7 (13.5)
	10	5.0 (2.6)	5.4 (2.9)	5.1 (2.6)	5.0 (2.0)	4.4 (1.7)	10.9 (8.4)	11.4 (8.7)	10.8 (8.4)	10.7 (8.5)	10.6 (6.8)	17.5 (15.2)	18.4 (16.6)	17.6 (15.2)	17.6 (15.0)	16.2 (13.6)
	CV	8.4 (3.8)	8.8 (3.4)	8.2 (3.6)	6.4 (3.0)	ND	16.8 (12.6)	17.6 (12.9)	16.4 (12.6)	16.1 (12.1)	ND	23.1 (20.6)	24.3 (20.5)	23.1 (21.2)	22.9 (20.2)	ND
$n = 100$	p_n	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
	5	13.9 (8.2)	14.7 (8.8)	13.8 (8.3)	12.9 (7.8)	7.1 (3.5)	25.5 (23.8)	27.1 (25.2)	25.5 (24.2)	24.4 (23.2)	15.4 (14.1)	33.8 (34.7)	35.5 (36.3)	34.1 (34.8)	32.5 (33.1)	21.8 (22.2)
	9	10.0 (5.9)	11.3 (6.7)	10.2 (5.9)	9.6 (5.3)	6.1 (2.7)	19.3 (17.7)	22.0 (20.2)	19.3 (17.9)	18.6 (17.5)	12.9 (11.0)	27.0 (27.9)	29.5 (30.4)	27.2 (27.9)	26.4 (27.2)	20.4 (20.0)
	12	8.7 (4.2)	10.5 (5.6)	8.7 (4.1)	8.6 (3.7)	5.3 (2.9)	17.5 (16.3)	19.3 (18.1)	17.8 (16.2)	16.8 (15.5)	12.5 (10.3)	23.8 (24.0)	27.8 (28.2)	24.1 (24.2)	23.0 (22.9)	18.7 (17.9)
	CV	20.0 (10.4)	20.8 (10.0)	18.9 (9.7)	16.9 (8.4)	ND	33.2 (28.9)	34.6 (27.6)	31.4 (27.4)	30.8 (25.1)	ND	42.1 (40.5)	43.5 (40.1)	40.4 (38.9)	40.8 (37.6)	ND
$n = 200$	p_n	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
	6	32.7 (22.3)	36.3 (24.5)	33.1 (22.1)	31.2 (20.6)	13.4 (9.7)	49.2 (48.5)	52.8 (51.5)	49.9 (48.8)	47.4 (46.0)	26.8 (26.5)	58.8 (61.7)	62.0 (64.8)	59.0 (61.7)	57.0 (60.1)	35.9 (37.8)
	10	21.9 (15.9)	27.0 (19.5)	22.2 (15.4)	20.9 (15.4)	9.3 (6.2)	38.8 (38.1)	44.0 (42.7)	38.9 (38.2)	36.5 (36.3)	20.8 (20.3)	48.3 (51.3)	53.0 (56.3)	48.6 (51.1)	46.2 (49.0)	30.0 (31.3)
	15	16.6 (11.6)	21.4 (14.9)	16.9 (11.5)	15.6 (10.7)	7.6 (5.0)	30.5 (30.1)	36.5 (36.5)	30.8 (30.0)	29.4 (28.4)	18.3 (16.4)	40.3 (42.6)	46.3 (48.9)	40.6 (42.9)	38.8 (41.4)	26.0 (26.4)
	CV	46.5 (29.8)	50.0 (29.9)	43.0 (27.4)	41.9 (24.9)	ND	62.1 (56.7)	64.9 (57.5)	60.0 (54.2)	60.0 (51.9)	ND	69.2 (68.5)	72.0 (69.7)	67.9 (67.4)	70.6 (66.7)	ND

TABLEAU 3.4. Puissance en pourcentage basée sur les quantiles asymptotiques (empiriques) lorsque $\delta = 1.0$

	$\alpha = 1\%$					$\alpha = 5\%$					$\alpha = 10\%$				
$n = 50$															
p_n	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
4	11.7 (6.3)	12.0 (6.3)	11.9 (6.2)	10.9 (5.9)	5.6 (3.2)	22.3 (18.7)	22.5 (19.1)	22.2 (18.9)	21.4 (17.7)	14.1 (10.9)	29.4 (29.0)	30.0 (30.0)	30.0 (29.5)	28.8 (28.4)	21.5 (19.6)
8	7.7 (4.3)	8.7 (5.3)	7.9 (4.0)	7.7 (3.8)	6.0 (2.4)	16.0 (13.5)	17.9 (14.4)	16.1 (13.5)	15.9 (12.3)	13.3 (9.5)	25.0 (23.7)	27.0 (25.1)	25.0 (23.5)	24.0 (22.1)	20.0 (17.0)
10	7.4 (3.7)	8.1 (4.6)	7.4 (3.6)	7.0 (3.4)	6.0 (2.5)	15.5 (12.1)	16.8 (13.0)	15.8 (12.2)	15.1 (11.6)	13.0 (9.0)	23.6 (20.6)	25.2 (23.6)	23.4 (20.7)	22.8 (20.3)	19.0 (16.5)
CV	14.0 (6.7)	15.0 (6.3)	13.5 (6.7)	12.4 (5.4)	ND	25.4 (19.4)	27.2 (19.4)	24.7 (19.4)	24.6 (19.1)	ND	34.1 (31.6)	35.1 (31.2)	33.6 (31.0)	33.5 (30.3)	ND
$n = 100$															
p_n	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
5	27.7 (18.7)	30.3 (20.4)	28.2 (18.6)	26.0 (17.5)	12.5 (7.1)	43.3 (41.2)	45.0 (43.4)	43.6 (41.5)	41.6 (40.0)	24.2 (22.2)	52.6 (53.6)	54.7 (55.6)	52.6 (53.8)	50.5 (51.4)	32.8 (33.1)
9	18.9 (11.9)	22.2 (14.5)	18.9 (11.9)	17.8 (10.8)	9.5 (4.8)	32.9 (30.2)	37.6 (34.9)	33.2 (30.7)	31.2 (29.1)	20.1 (16.9)	43.1 (43.8)	47.3 (47.7)	43.1 (44.2)	41.6 (42.4)	27.7 (27.5)
12	15.8 (9.0)	19.2 (11.7)	16.1 (8.9)	14.7 (8.0)	8.3 (4.5)	28.5 (25.9)	32.7 (31.1)	28.7 (25.7)	27.1 (24.3)	17.9 (15.2)	37.8 (37.9)	43.5 (43.8)	37.6 (37.7)	36.1 (36.1)	25.2 (24.7)
CV	38.7 (23.6)	39.7 (22.5)	36.2 (22.4)	31.9 (18.7)	ND	52.2 (48.5)	53.9 (47.5)	51.0 (47.3)	49.0 (43.4)	ND	61.1 (59.6)	62.3 (59.7)	60.1 (58.7)	60.5 (57.0)	ND
$n = 200$															
p_n	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
6	63.5 (53.4)	67.6 (55.8)	63.8 (52.7)	61.3 (49.7)	30.0 (21.9)	77.2 (76.4)	80.1 (79.1)	77.4 (76.4)	75.4 (74.5)	47.7 (47.4)	83.9 (86.1)	86.4 (88.0)	84.1 (86.5)	82.4 (84.5)	58.8 (61.3)
10	48.2 (38.5)	56.3 (46.1)	48.8 (37.6)	45.2 (36.6)	20.7 (15.3)	66.2 (65.7)	71.7 (70.3)	66.5 (65.6)	63.4 (63.4)	37.7 (36.6)	74.3 (76.8)	79.6 (81.7)	74.5 (76.8)	72.2 (74.5)	46.8 (49.0)
15	36.2 (28.9)	45.3 (37.0)	36.8 (29.0)	33.8 (27.1)	16.7 (11.9)	54.5 (53.9)	63.7 (63.6)	55.0 (53.4)	52.2 (51.4)	31.1 (28.2)	65.1 (68.1)	72.7 (74.9)	65.6 (67.9)	63.1 (65.6)	42.1 (42.6)
CV	76.9 (62.7)	79.1 (62.6)	74.9 (60.5)	72.0 (55.5)	ND	87.5 (83.8)	89.4 (84.1)	84.8 (81.4)	85.5 (79.6)	ND	90.8 (90.7)	92.7 (91.4)	89.8 (89.6)	91.4 (89.7)	ND

TABLEAU 3.5. Puissance en pourcentage basée sur les quantiles asymptotiques (empiriques) lorsque $\delta = 1.25$

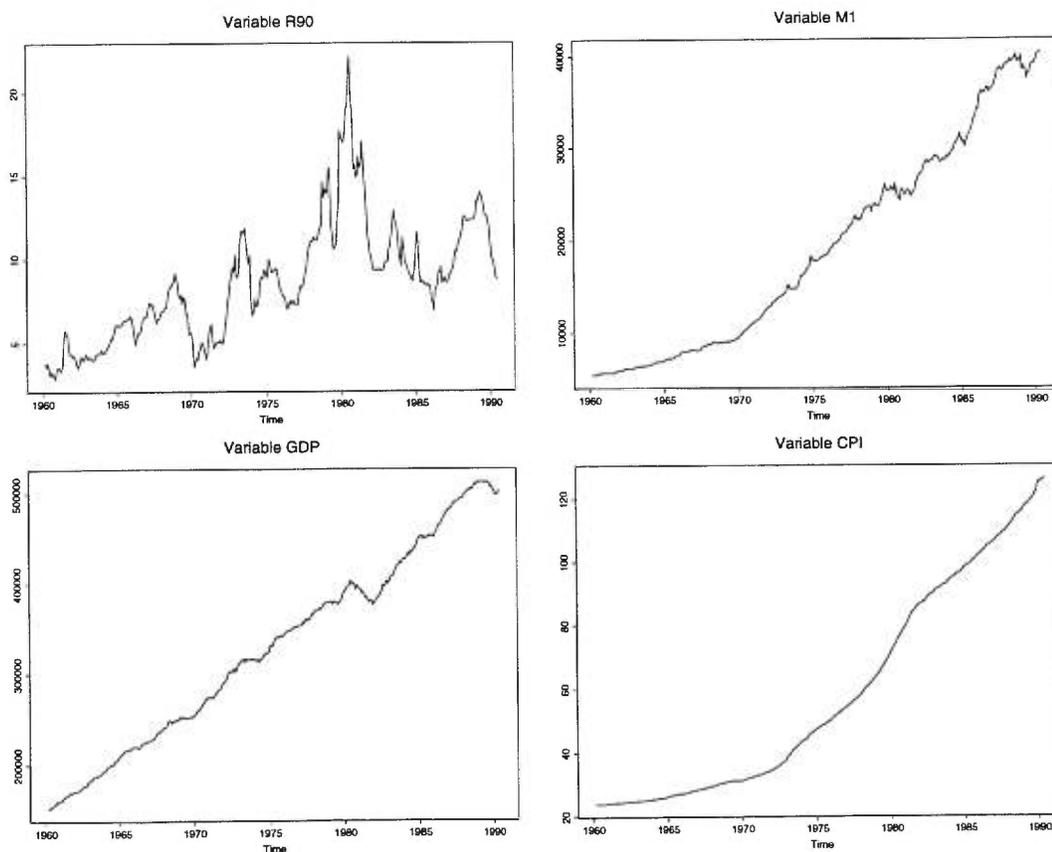
	$\alpha = 1\%$					$\alpha = 5\%$					$\alpha = 10\%$				
$n = 50$															
p_n	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
4	18.6	18.9	18.8	17.3	8.2	31.7	32.2	31.7	30.5	19.5	41.4	42.0	41.7	39.8	28.3
	(10.8)	(11.2)	(10.6)	(10.2)	(4.4)	(27.3)	(28.5)	(28.0)	(26.0)	(16.4)	(41.1)	(41.9)	(41.2)	(39.1)	(25.8)
8	11.8	13.5	12.1	10.9	7.9	24.2	27.2	24.0	23.1	16.6	34.1	36.3	34.1	33.2	23.2
	(6.6)	(7.9)	(6.1)	(5.7)	(3.4)	(19.6)	(22.1)	(19.6)	(18.3)	(12.0)	(32.3)	(34.2)	(32.2)	(30.3)	(20.5)
10	10.7	11.9	10.8	10.5	8.1	22.1	25.0	21.8	20.8	16.5	31.7	34.6	32.0	30.9	22.9
	(5.5)	(6.8)	(5.6)	(5.2)	(3.6)	(17.3)	(19.8)	(17.4)	(16.2)	(11.2)	(28.7)	(32.3)	(28.6)	(27.3)	(19.1)
CV	22.2	23.5	20.9	19.3	ND	37.3	38.2	36.2	34.8	ND	45.9	47.8	45.7	44.7	ND
	(12.4)	(11.6)	(11.6)	(9.5)	ND	(30.1)	(30.2)	(29.6)	(28.4)	ND	(43.2)	(43.0)	(42.7)	(40.3)	ND
$n = 100$															
p_n	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
5	46.7	49.3	47.2	43.7	21.2	64.2	66.2	64.1	61.7	36.4	72.4	73.5	72.3	70.6	47.9
	(34.7)	(37.7)	(34.6)	(32.0)	(13.0)	(61.5)	(64.0)	(61.9)	(59.7)	(33.1)	(72.9)	(74.6)	(72.9)	(71.4)	(48.5)
9	32.1	38.8	32.6	30.4	14.9	49.7	55.8	50.0	47.8	28.2	60.7	66.1	61.1	58.7	38.0
	(22.2)	(26.4)	(22.0)	(20.6)	(8.5)	(46.9)	(53.1)	(47.9)	(45.2)	(25.3)	(61.6)	(67.0)	(62.0)	(59.5)	(37.8)
12	26.3	33.0	26.9	24.8	13.3	43.3	50.2	43.8	41.8	25.0	53.8	61.2	54.7	52.4	36.6
	(16.7)	(21.9)	(16.9)	(14.6)	(7.0)	(40.7)	(47.4)	(40.2)	(38.5)	(22.0)	(54.1)	(62.0)	(54.7)	(52.2)	(34.8)
CV	58.3	59.7	56.8	50.7	ND	72.7	74.2	71.6	69.4	ND	80.2	81.7	79.0	78.7	ND
	(43.1)	(42.2)	(40.3)	(34.8)	ND	(68.4)	(66.8)	(66.9)	(63.0)	ND	(78.7)	(78.8)	(78.0)	(76.1)	ND
$n = 200$															
p_n	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
6	87.7	90.5	88.1	86.3	52.7	94.4	95.6	94.5	93.9	72.5	96.4	97.3	96.6	96.0	80.8
	(80.3)	(82.7)	(79.9)	(77.4)	(42.6)	(94.2)	(95.2)	(94.4)	(93.5)	(72.3)	(97.1)	(97.9)	(97.3)	(96.5)	(82.4)
10	75.6	82.7	75.8	73.3	38.3	87.8	91.7	87.9	85.8	57.3	92.1	95.0	92.3	90.8	68.6
	(66.3)	(74.0)	(65.7)	(63.5)	(31.3)	(87.5)	(91.1)	(87.6)	(85.8)	(56.3)	(93.1)	(95.5)	(93.3)	(92.0)	(70.0)
15	62.7	73.4	62.9	59.4	29.5	78.5	85.8	78.7	76.7	49.1	85.9	91.2	85.7	83.8	61.0
	(54.0)	(64.3)	(53.5)	(50.8)	(22.5)	(78.1)	(85.8)	(78.1)	(75.9)	(45.5)	(87.1)	(92.2)	(86.9)	(85.7)	(61.5)
CV	93.7	94.9	92.0	92.3	ND	97.4	97.9	97.1	96.5	ND	98.7	98.8	98.4	98.4	ND
	(87.6)	(87.4)	(85.5)	(82.8)	ND	(96.4)	(96.8)	(96.0)	(94.9)	ND	(98.6)	(98.4)	(98.3)	(97.8)	ND

3.7. APPLICATION

Afin d'illustrer l'application du test de corrélation sérielle, nous considérons une application provenant de la littérature en série chronologique. Cet exemple a été considéré dans Gilbert (1993) et concerne la modélisation de variables économiques canadiennes avec un modèle *VARX*.

Gilbert (1993) considère un modèle mensuel pour modéliser l'économie canadienne. La variable servant de variable explicative est le taux d'intérêt aux 90 jours (R90). La masse monétaire (M1), le produit intérieur brut (GDP) et l'indice des prix à la consommation (CPI) servent de variables dépendantes. Dans cette étude, la période d'observation est de 20 ans, allant de mars 1961 à mars 1981. Les graphiques des variables à l'étude se retrouvent dans la figure 3.1.

FIG. 3.1. Les quatre variables à l'étude.



Une analyse préliminaire des quatre séries semble indiquer qu'une transformation logarithmique et une différenciation des variables dépendantes M1, GDP et CPI sont appropriées. Une différenciation de la variable explicative R90 est aussi considérée afin que la série résultante soit stationnaire. Le modèle postulé par Gilbert (1993) est

$$(I - \Lambda_1 B - \Lambda_2 B^2 - \Lambda_3 B^3) \mathbf{y}_t = (V_0 + V_1 B + V_2 B^2) x_t + \mathbf{u}_t, \quad (3.21)$$

où

$$\mathbf{y}_t = \begin{bmatrix} 100(1 - B) \log(M1_t) \\ 100(1 - B) \log(GDP_{t-2}) \\ 100(1 - B) \log(CPI_t) \end{bmatrix}$$

et

$$x_t = (1 - B)R90_t.$$

Les facteurs multiplicatifs pour les trois séries composant les variables dépendantes sont choisis de telle sorte que les variances des séries soient d'un ordre de grandeur satisfaisant.

Gilbert (1993) explique les délais sur la variable *GDP* compte tenu de la disponibilité de l'information de cette variable. Cependant, l'interprétation économique que la variable *R90* influence *GDP* deux périodes en avance n'a peut-être pas un sens réel, à moins que les agents économiques puissent anticiper les changements dans les taux. Cependant, selon Gilbert (1993), l'impact estimé des taux d'intérêt sur les valeurs passées de la variable *GDP* semble très négligeable. Notons que notre modèle diffère de celui de Gilbert (1993) concernant la différenciation de la variable *R90*. Notre analyse du corrélogramme de cette variable semblait montrer qu'une différenciation était nécessaire afin d'avoir la stationnarité.

Nous avons procédé à l'estimation par la méthode des moindres carrés suivant la méthode de la section 3.3. Par la suite, les paramètres estimés plus petits qu'une fois leur écart type en valeur absolu ont été fixés à zéro. Les contraintes résultantes étant linéaires, il était possible de les formuler en choisissant adéquatement la matrice *R* de la section 3.3. Dans ce contexte, l'estimation avec des contraintes linéaires permettait surtout d'obtenir un modèle plus parcimonieux. Le modèle final retenu est donné dans le tableau 3.6. Les écarts type des estimateurs sont donnés entre parenthèses.

3.8. PREUVE DES THÉORÈMES

3.8.1. Preuve du théorème 2

Nous allons adopter les notations suivantes au long de la preuve de ces deux étapes. Un produit scalaire est représenté par $\langle \mathbf{x}_t, \mathbf{x}_s \rangle = \mathbf{x}_t' \mathbf{x}_s$ et la norme euclidienne d'un vecteur est représentée par $\|\mathbf{x}_t\| = \sqrt{\langle \mathbf{x}_t, \mathbf{x}_t \rangle}$. La norme matricielle est la norme euclidienne notée $\|\mathbf{A}\|_E^2 = \text{tr}(\mathbf{A}\mathbf{A}') = \sum_{i=1}^n \sum_{j=1}^n a_{ij}^2$. Les notations O_p et o_p sont telles que définies dans Fuller (1996). Nous posons aussi $k_{nj} = k(j/p_n)$. Nous posons également $\mathbf{v}_t = \boldsymbol{\Sigma}_u^{-1/2} \mathbf{u}_t$. On note aussi $\boldsymbol{\Sigma}_u = \boldsymbol{\Gamma}_u(0)$. Le processus $\mathbf{v} = \{\mathbf{v}_t : t \in Z\}$ est centré en $\mathbf{0}$ et de variance \mathbf{I}_d .

Nous allons utiliser de manière très intensive les inégalités impliquant la fonction tr , qui ne sont en fait que des versions de l'inégalité de Cauchy-Schartz dans notre contexte. Les versions les plus utiles sont données ici. Pour plus de détails, voir Harville (1997, chap. 5 et chap. 6). Soient \mathbf{A} et \mathbf{B} des matrices quelconques, \mathbf{C} et \mathbf{D} des matrices symétriques définies semi-positives. On a alors

$$|\text{tr}(\mathbf{A}\mathbf{B}')| \leq \sqrt{\text{tr}(\mathbf{A}\mathbf{A}')} \sqrt{\text{tr}(\mathbf{B}\mathbf{B}')}, \quad (3.22)$$

$$\text{tr}(\mathbf{C}^2) \leq (\text{tr}(\mathbf{C}))^2, \quad (3.23)$$

$$\text{tr}(\mathbf{C}\mathbf{D}) \leq \text{tr}(\mathbf{C})\text{tr}(\mathbf{D}), \quad (3.24)$$

$$\text{tr}(\mathbf{A}\mathbf{C}) \leq \sqrt{\text{tr}(\mathbf{A}\mathbf{A}')} \text{tr}(\mathbf{C}), \quad (3.25)$$

$$|\text{tr}(\mathbf{A}'\mathbf{B}\mathbf{A}\mathbf{B})| \leq \text{tr}(\mathbf{A}'\mathbf{A})\text{tr}(\mathbf{B}'\mathbf{B}), \quad (3.26)$$

$$\text{tr}[(\mathbf{A} + \mathbf{B} + \mathbf{C})(\mathbf{A} + \mathbf{B} + \mathbf{C})'] \leq 4[\text{tr}(\mathbf{A}\mathbf{A}') + \text{tr}(\mathbf{B}\mathbf{B}') + \text{tr}(\mathbf{C}\mathbf{C}')]. \quad (3.27)$$

Nous prouvons maintenant l'étape 1. Notons en premier lieu que $\mathbf{C}_u(0) - \boldsymbol{\Sigma}_u = O_p(n^{-1/2})$ puisque $E(C_{u,ij}(0) - \sigma_{ij}) = 0$ et $\text{var}(C_{u,ij}(0)) = n^{-1}(\mu(i, i, j, j) - \sigma_{ij}^2)$. De cela découle aussi que $\mathbf{C}_u^{-1}(0) - \boldsymbol{\Sigma}_u^{-1} = O_p(n^{-1/2})$.

Nous allons montrer qu'asymptotiquement, nous pouvons remplacer $\mathbf{C}_u(0)$ par $\boldsymbol{\Sigma}_u$, et ainsi travailler par la suite sur la version faisant intervenir $\boldsymbol{\Sigma}_u$.

Résultat 1.

$$\sum_{j=1}^{n-1} k^2(j/p_n) \{ \text{tr}[\mathbf{C}_u^{-1}(0)\mathbf{C}_u(j)\mathbf{C}_u^{-1}(0)\mathbf{C}_u'(j)] - \text{tr}[\boldsymbol{\Sigma}_u^{-1}\mathbf{C}_u(j)\boldsymbol{\Sigma}_u^{-1}\mathbf{C}_u'(j)] \} = o_p(\sqrt{p_n}/n).$$

La preuve du résultat 1 repose sur le lemme suivant:

Montrons le lemme suivant:

Lemme 1. $\sum_{j=1}^{n-1} k^2(j/p_n) \mathbf{C}_v(j) \mathbf{C}'_v(j) = O_p(p_n/n)$.

DÉMONSTRATION. En effet,

$$\begin{aligned} \mathbf{C}_v(j) \mathbf{C}'_v(j) &= n^{-2} \sum_{t=j+1}^n \|\mathbf{v}_{t-j}\|^2 \mathbf{v}_t \mathbf{v}'_t + n^{-2} \sum_{t=j+2}^n \sum_{s=j+1}^{t-1} \mathbf{v}_t \mathbf{v}'_{t-j} \mathbf{v}_{s-j} \mathbf{v}'_s + \\ & n^{-2} \sum_{t=j+1}^{n-1} \sum_{s=t+1}^n \mathbf{v}_t \mathbf{v}'_{t-j} \mathbf{v}_{s-j} \mathbf{v}'_s. \end{aligned}$$

Ainsi en prenant l'espérance de chaque côté que

$$\begin{aligned} E[\mathbf{C}_v(j) \mathbf{C}'_v(j)] &= n^{-2} \sum_{t=j+1}^n E(\|\mathbf{v}_{t-j}\|^2) E(\mathbf{v}_t \mathbf{v}'_t), \\ &= \frac{(n-j)}{n^2} d \mathbf{I}_d. \end{aligned}$$

Ainsi

$$\begin{aligned} E\left[\sum_{j=1}^{n-1} k^2(j/p_n) \mathbf{C}_v(j) \mathbf{C}'_v(j)\right] &= n^{-1} d \sum_{j=1}^{n-1} (1-j/n) k^2(j/p_n), \\ &= O(p_n/n). \end{aligned}$$

□

Afin de montrer la résultat 1, il suffit de remarquer que

$$\begin{aligned} \mathbf{C}'_u(j) \mathbf{C}_u^{-1}(0) \mathbf{C}_u(j) \mathbf{C}_u^{-1}(0) &= \mathbf{C}'_u(j) \boldsymbol{\Sigma}_u^{-1} \mathbf{C}_u(j) \boldsymbol{\Sigma}_u^{-1} + \\ & \mathbf{C}'_u(j) \Delta_{un} \mathbf{C}_u(j) \boldsymbol{\Sigma}_u^{-1} + \\ & \mathbf{C}'_u(j) \boldsymbol{\Sigma}_u^{-1} \mathbf{C}_u(j) \Delta_{un} + \\ & \mathbf{C}'_u(j) \Delta_{un} \mathbf{C}_u(j) \Delta_{un}, \end{aligned}$$

où $\mathbf{C}_u^{-1}(0) - \boldsymbol{\Sigma}_u^{-1} = \Delta_{un}$. Il suffit ensuite de multiplier par $k^2(j/p_n)$ et de sommer sur j , de prendre la trace, d'utiliser (3.25) et (3.26), $p_n/n \rightarrow 0$ et le lemme 1.

Le résultat 1 nous permet de travailler avec le terme faisant intervenir $\boldsymbol{\Sigma}_u$ plutôt que $\mathbf{C}_u(0)$. On décompose maintenant $\sum_{j=1}^{n-1} k^2(j/p_n) \text{tr}[\boldsymbol{\Sigma}_u^{-1} \mathbf{C}_u(j) \boldsymbol{\Sigma}_u^{-1} \mathbf{C}'_u(j)]$ en deux morceaux

A_{1n} et A_{2n} de la manière suivante:

$$\begin{aligned}
\sum_{j=1}^{n-1} k_{nj}^2 \text{tr}[\mathbf{C}'_u(j) \boldsymbol{\Sigma}_u^{-1} \mathbf{C}_u(j) \boldsymbol{\Sigma}_u^{-1}] &= n^{-2} \sum_{j=1}^{n-1} k_{nj}^2 \left\{ \sum_{t=j+1}^n \|\mathbf{v}_{t-j}\|^2 \|\mathbf{v}_t\|^2 \right\} + \\
& n^{-2} \sum_{j=1}^{n-2} k_{nj}^2 \left\{ \sum_{t=j+2}^n \sum_{s=j+1}^{t-1} 2 \langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j} \rangle \langle \mathbf{v}_t, \mathbf{v}_s \rangle \right\}, \\
&= n^{-2} \sum_{j=1}^{n-1} k_{nj}^2 \left\{ \sum_{t=j+1}^n Z_{jt}^2 \right\} + \\
& n^{-2} \sum_{j=1}^{n-2} k_{nj}^2 \left\{ \sum_{t=j+2}^n \sum_{s=j+1}^{t-1} w_{jts} \right\}, \\
&= n^{-1} (A_{1n} + A_{2n}),
\end{aligned}$$

où $Z_{jt} = \|\mathbf{v}_{t-j}\| \times \|\mathbf{v}_t\|$ et $w_{jts} = 2 \langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j} \rangle \langle \mathbf{v}_t, \mathbf{v}_s \rangle$.

Résultat 2. $p_n^{-1/2} (A_{1n} - d^2 M_n(k)) \rightarrow_P 0$.

Le preuve repose sur l'application des deux lemmes suivants:

Lemme 2. $E(A_{1n}) = d^2 M_n(k)$.

DÉMONSTRATION.

$$\begin{aligned}
E(A_{1n}) &= n^{-1} \sum_{j=1}^{n-1} k_{nj}^2 \left\{ \sum_{t=j+1}^n E(\|\mathbf{v}_{t-j}\|^2) E(\|\mathbf{v}_t\|^2) \right\}, \\
&= n^{-1} \sum_{j=1}^{n-1} k_{nj}^2 (n-j) d^2, \\
&= d^2 M_n(k).
\end{aligned}$$

□

Lemme 3. $\text{var}(A_{1n}) = O(p_n^2/n)$.

DÉMONSTRATION.

$$\begin{aligned} \text{var}(A_{1n}) &= E\left\{n^{-1} \sum_{j=1}^{n-1} k_{nj}^2 \left[\sum_{t=j+1}^n (Z_{jt}^2 - d^2) \right]\right\}^2, \\ &\leq n^{-2} \left\{ \sum_{j=1}^{n-1} k_{nj}^2 \left[E \left(\sum_{t=j+1}^n (Z_{jt}^2 - d^2) \right)^2 \right]^{1/2} \right\}^2, \\ &= O(p_n^2/n). \end{aligned}$$

d'après le lemme qui suit. □

Lemme 4. $E[\sum_{t=j+1}^n (Z_{jt}^2 - d^2)]^2 = O(n)$.

DÉMONSTRATION. On note en premier lieu que

$$\begin{aligned} \left(\sum_{t=j+1}^n (Z_{jt}^2 - d^2) \right)^2 &= \sum_{t=j+1}^n (Z_{jt}^2 - d^2)^2 + \\ &\quad 2 \sum_{t=j+2}^n \sum_{s=j+1}^{t-1} (Z_{jt}^2 - d^2)(Z_{js}^2 - d^2). \end{aligned}$$

On a alors le résultat puisque

$$E[(Z_{jt}^2 - d^2)^2] = E(\|\mathbf{v}_1\|^4) - d^4,$$

et

$$E(Z_{jt}^2 - d^2)(Z_{js}^2 - d^2) = \begin{cases} (E(\|\mathbf{v}_1\|^4) - d^2)d^2 & \text{si } s = t - j, \\ 0 & \text{sinon.} \end{cases}$$

□

Le résultat 2 s'obtient alors des lemmes 2 et 3 et d'après la proposition 6.2.4 dans Brockwell et Davis (1991). Ainsi, on montre avec le résultat 2 que la contribution de $p_n^{-1/2}(A_{1n} - d^2 M_n(k))$ est $o_p(1)$. Pour établir l'étape 1, on doit travailler le terme A_{2n} . Plus précisément, on doit montrer que

$$\frac{A_{2n}}{\sqrt{2d^2 V_n(k)}} \rightarrow_L N(0, 1).$$

Afin de démontrer ce résultat, considérons l_n telle que $l_n/p_n \rightarrow \infty$ et $l_n/n \rightarrow 0$. On décompose alors A_{2n} en plusieurs morceaux. On cherche ici à se débarrasser des termes qui

convergent en probabilité vers zéro.

$$A_{2n} = B_n + \sum_{i=1}^4 C_{in},$$

où

$$\begin{aligned} B_n &= n^{-1} \sum_{j=1}^{l_n} k_{nj}^2 \left\{ \sum_{t=2l_n+3}^n \sum_{s=l_n+2}^{t-l_n-1} w_{jts} \right\}, \\ C_{1n} &= n^{-1} \sum_{j=l_n+1}^{n-2} k_{nj}^2 \left\{ \sum_{t=j+2}^n \sum_{s=j+1}^{t-1} w_{jts} \right\}, \\ C_{2n} &= n^{-1} \sum_{j=1}^{l_n} k_{nj}^2 \left\{ \sum_{t=2l_n+3}^n \sum_{s=t-l_n}^{t-1} w_{jts} \right\}, \\ C_{3n} &= n^{-1} \sum_{j=1}^{l_n} k_{nj}^2 \left\{ \sum_{t=l_n+3}^{2l_n+2} \sum_{s=l_n+2}^{t-1} w_{jts} \right\}, \\ C_{4n} &= n^{-1} \sum_{j=1}^{l_n} k_{nj}^2 \left\{ \sum_{t=j+1}^{l_n+1} \sum_{s=t+1}^n w_{jts} \right\}. \end{aligned}$$

Le lemme suivant est utile pour la suite. Notons que nous généralisons un résultat dans la preuve de Hong (1996a) dans notre contexte, et que nous apportons une légère correction au résultat de Hong (1996a), qui a oublié de distinguer les cas $j_1 \neq j_2$ et $j_1 = j_2$.

Lemme 5. *Posons*

$$\begin{aligned} w_{j_1 t_1 s_1}^{l_1 l_2 l_3 l_4} &= 2v_{t_1}(l_1)v_{s_1}(l_2)v_{t_1-j_1}(l_3)v_{s_1-j_1}(l_4), \\ w_{j_1 t_1 s_1}^{l_1 l_2} &= 2\langle \mathbf{v}_{t_1}, \mathbf{v}_{s_1} \rangle v_{t_1-j_1}(l_1)v_{s_1-j_1}(l_2), \\ w_{j_1 t_1 s_1} &= 2\langle \mathbf{v}_{t_1}, \mathbf{v}_{s_1} \rangle \langle \mathbf{v}_{t_1-j_1}, \mathbf{v}_{s_1-j_1} \rangle. \end{aligned}$$

Nous avons alors:

$$\begin{aligned} E(w_{j_1 t_1 s_1}^{l_1 l_2 l_3 l_4} w_{j_2 t_2 s_2}^{m_1 m_2 m_3 m_4}) &= \begin{cases} E(w_{j_1 t_1 s_1}^{l_1 l_2 l_3 l_4} w_{j_2 t_2 s_2}^{m_1 m_2 m_3 m_4}) \delta_{t_1, t_2} \delta_{s_1, t_1-j_2} \delta_{s_2, t_1-j_1}, & j_1 \neq j_2 \\ E(w_{j_1 t_1 s_1}^{l_1 l_2 l_3 l_4} w_{j_2 t_2 s_2}^{m_1 m_2 m_3 m_4}) \delta_{t_1, t_2} \delta_{s_1, s_2}, & j_1 = j_2 \end{cases} \\ E(w_{j_1 t_1 s_1}^{l_1 l_2} w_{j_2 t_2 s_2}^{m_1 m_2}) &= \begin{cases} E(w_{j_1 t_1 s_1}^{l_1 l_2} w_{j_2 t_2 s_2}^{m_1 m_2}) \delta_{t_1, t_2} \delta_{s_1, t_1-j_2} \delta_{s_2, t_1-j_1}, & j_1 \neq j_2 \\ E(w_{j_1 t_1 s_1}^{l_1 l_2} w_{j_2 t_2 s_2}^{m_1 m_2}) \delta_{t_1, t_2} \delta_{s_1, s_2}, & j_1 = j_2 \end{cases} \\ E(w_{j_1 t_1 s_1} w_{j_2 t_2 s_2}) &= \begin{cases} E(w_{j_1 t_1 s_1} w_{j_2 t_2 s_2}) \delta_{t_1, t_2} \delta_{s_1, t_1-j_2} \delta_{s_2, t_1-j_1}, & j_1 \neq j_2 \\ E(w_{j_1 t_1 s_1} w_{j_2 t_2 s_2}) \delta_{t_1, t_2} \delta_{s_1, s_2}, & j_1 = j_2 \end{cases} \end{aligned}$$

DÉMONSTRATION. La preuve est directe et se fait par cas. Nous l'omettons ici. \square

On montre alors le résultat suivant:

Résultat 3. $p_n^{-1/2}C_{in} = o_p(1)$, $i = 1, 2, 3, 4$.

DÉMONSTRATION. Commençons par C_{1n} . Afin de démontrer le résultat annoncé, il suffit de montrer que $E(C_{1n}^2) = o(p_n)$. En élevant C_{1n} au carré, en séparant la somme selon $j_1 = j_2$ et $j_1 \neq j_2$, en prenant l'espérance et en utilisant le lemme 5, on parvient à montrer que

$$\begin{aligned} E(C_{1n}^2) &\leq 4d^2\mu_4(\|\mathbf{v}\|)\left(\sum_{j=l_n+1}^{n-2} k_{nj}^4\right) + \frac{8d}{n}\left(\sum_{l_n+2}^{n-2} k_{nj}^2\right)^2, \\ &= o(p_n), \end{aligned}$$

puisque $p_n^{-1} \sum_{j=l_n+1}^{n-2} k_{nj}^4 \rightarrow 0$ et $p_n/n \rightarrow 0$. □

Montrons le résultat pour C_{2n} . En procédant comme pour C_{1n} , on montre que

$$\begin{aligned} E(C_{2n}^2) &\leq 4d^2\mu_4(\|\mathbf{v}\|)\frac{l_n}{n}\left(\sum_{j=1}^{l_n} k_{nj}^4\right) + \frac{8d}{n}\left(\sum_{j=1}^{l_n} k_{nj}^2\right)^2, \\ &= O\left(\frac{l_n p_n}{n} + \frac{p_n^2}{n}\right). \end{aligned}$$

On montre de même pour C_{3n} que

$$\begin{aligned} E(C_{3n}^2) &\leq 4d^2\mu_4(\|\mathbf{v}\|)\frac{l_n^2}{n^2}\left(\sum_{j=1}^{l_n} k_{nj}^4\right) + 8d\frac{l_n}{n^2}\left(\sum_{j=1}^{l_n} k_{nj}^2\right)^2, \\ &= O\left(\frac{l_n^2 p_n}{n^2} + \frac{p_n^2 l_n}{n^2}\right). \end{aligned}$$

Finalement on obtient de la même manière le résultat pour C_{4n} en montrant que

$$\begin{aligned} E(C_{4n}^2) &\leq 4d^2\mu_4(\|\mathbf{v}\|)\frac{l_n}{n}\left(\sum_{j=1}^{l_n} k_{nj}^4\right) + \frac{8d}{n}\left(\sum_{j=1}^{l_n} k_{nj}^2\right)^2, \\ &= O\left(\frac{l_n p_n}{n} + \frac{p_n^2}{n}\right). \end{aligned}$$

Ainsi à la lumière du résultat 3, le seul terme important dans la distribution asymptotique est B_n . Ainsi, la preuve sera complétée si on montre que $\sigma^{-2}(n)B_n \rightarrow_L N(0, 1)$, où $\sigma^2(n) = E(B_n^2)$. De plus, on montrera un peu plus loin que $E(B_n^2) = 2d^2 p_n V(k)[1 + o(1)]$. On commence par tenter d'écrire B_n comme une moyenne. On remarque que l'on peut écrire B_n de la manière

suivante:

$$\begin{aligned} B_n &= n^{-1} \sum_{j=1}^{l_n} k_{nj}^2 \left\{ \sum_{t=2l_n+3}^n \sum_{s=l_n+2}^{t-l_n-1} w_{jts} \right\}, \\ &= n^{-1} \sum_{t=2l_n+3}^n B_{nt}, \end{aligned}$$

où $B_{nt} = 2\mathbf{v}'_t \left\{ \sum_{j=1}^{l_n} k_{nj}^2 H_{j,t-l_n-1} \mathbf{v}_{t-j} \right\}$, où l'on pose $H_{j,t-l_n-1} = \sum_{s=l_n+2}^{t-l_n-1} \mathbf{v}_s \mathbf{v}'_{s-j}$. On montre alors que $\{B_{nt}, \mathcal{F}_{t-1}\}$ est une différence de martingales dans le lemme suivant:

Lemme 6. $\{B_{nt}, \mathcal{F}_{t-1}\}$ est une différence de martingales, où $\mathcal{F}_t = \sigma(\mathbf{v}_s, s \leq t)$.

DÉMONSTRATION. Il suffit de remarquer que $E(B_{nt}) = 0$ et $E(B_{nt} | \mathcal{F}_{t-1}) = 0$. \square

Lemme 7. \mathbf{v}_{t-i} est indépendant de $H_{j,t-l_n-1}$, $1 \leq i, j \leq l_n$.

DÉMONSTRATION. On remarque que $H_{j,t-l_n-1}$ est fonction de $\{\dots, \mathbf{v}_{t-l_n-2}, \mathbf{v}_{t-l_n-1}\}$ indépendamment de j , et que \mathbf{v}_{t-i} , pour $1 \leq i \leq l_n$, appartient à l'ensemble $\{\mathbf{v}_{t-l_n}, \mathbf{v}_{t-l_n+1}, \dots, \mathbf{v}_{t-1}\}$. \square

Lemme 8. $E(B_{nt}^2) = 4d^2(t - 2l_n - 2) \sum_{i=1}^{l_n} k_{ni}^4$.

DÉMONSTRATION.

$$\begin{aligned} E(B_{nt}^2) &= 4E\left(\sum_{i,j=1}^{l_n} k_{ni}^2 k_{nj}^2 \text{tr}[\mathbf{v}_{t-j} \mathbf{v}'_{t-i} H'_{i,t-l_n-1} H_{j,t-l_n-1}]\right), \\ &= 4\text{tr}\left[\sum_{i,j=1}^{l_n} k_{ni}^2 k_{nj}^2 E(\mathbf{v}_{t-j} \mathbf{v}'_{t-i}) E(H'_{i,t-l_n-1} H_{j,t-l_n-1})\right], \end{aligned}$$

en utilisant le lemme 7. Ainsi

$$\begin{aligned} E(B_{nt}^2) &= 4 \sum_{i=1}^{l_n} k_{ni}^4 E(\text{tr}[H'_{i,t-l_n-1} H_{i,t-l_n-1}]), \\ &= \sum_{i=1}^{l_n} k_{ni}^4 \sum_{s_1=l_n+2}^{t-l_n-1} \sum_{s_2=l_n+1}^{t-l_n-1} E(w_{is_1 s_2}), \\ &= \sum_{i=1}^{l_n} k_{ni}^4 \sum_{s_2=l_n+1}^{t-l_n-1} E(w_{is_1 s_2}), \\ &= 4d^2(t - 2l_n - 2) \sum_{i=1}^{l_n} k_{ni}^4. \end{aligned}$$

□

Ainsi, en utilisant les résultats du lemme précédent, on peut conclure que

$$\begin{aligned} E(B_n^2) &= n^{-2} \sum_{t=2l_n+3}^n E(B_{nt}^2), \\ &= 2d^2 \frac{(n-2l_n-2)(n-2l_n-1)}{n^2} \sum_{i=1}^{l_n} k_{ni}^4, \\ &= 2d^2 p_n V(k) [1 + o(1)]. \end{aligned}$$

On rappelle que l'on pose pour la suite $\sigma^2(n) = 2d^2 p_n V(k) [1 + o(1)]$.

Résultat 4. $\sigma^{-1}(n)B_n \rightarrow_L N(0, 1)$.

DÉMONSTRATION. On applique le théorème 1. Il faut donc montrer les deux points suivants:

- a) $\sigma^{-2}(n)n^{-2} \sum_{t=2l_n+3}^n E(B_{nt}^2 I[|B_{nt}| > \epsilon n \sigma(n)]) \rightarrow 0, \forall \epsilon > 0$.
 - b) $\sigma^{-2}(n)n^{-2} \sum_{t=2l_n+3}^n \ddot{B}_{nt}^2 \rightarrow_P 1$,
- où $\ddot{B}_{nt}^2 = E(B_{nt}^2 | \mathcal{F}_{t-1})$.

Montrons a). Pour ce faire, montrons la condition de Lyapounov avec $\delta = 2$. Remarquons que puisque $|\langle x, y \rangle| \leq \|x\| \times \|y\|$, nous avons alors

$$|B_{nt}| \leq 2\|\mathbf{v}_t\| \times \left\| \sum_{j=1}^{l_n} k_{nj}^2 H_{j,t-l_n-1} \mathbf{v}_{t-j} \right\|,$$

et nous obtenons ainsi

$$E(B_{nt}^4) \leq 16\mu_4(\|\mathbf{v}\|) E\left(\left\| \sum_{j=1}^{l_n} k_{nj}^2 H_{j,t-l_n-1} \mathbf{v}_{t-j} \right\|^4\right),$$

en utilisant le lemme 7. Remarquons maintenant que $E(\|x\|^4) \leq d \sum_{i=1}^d E[x^4(i)]$, où $x = (x(1), \dots, x(d))'$ est un vecteur de dimension d . La composante l de $\sum_{j=1}^{l_n} k_{nj}^2 H_{j,t-l_n-1} \mathbf{v}_{t-j}$ étant donné par la quantité $\sum_{j=1}^{l_n} k_{nj}^2 \sum_{s=l_n+2}^{t-l_n-1} v_s(l) \langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j} \rangle$, on montre alors le lemme suivant:

Lemme 9. $E\left(\left[\sum_{j=1}^{l_n} k_{nj}^2 \sum_{s=l_n+2}^{t-l_n-1} v_s(l) \langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j} \rangle\right]^4\right) = O(p_n^2 t^2)$, indépendamment de l et de j .

DÉMONSTRATION. On applique en premier lieu le lemme 10 aux variables

$$\left\{ k_{nj}^2 \sum_{s=l_n+2}^{t-l_n-1} v_s(l) \langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j} \rangle, j = 1, \dots, l_n \right\},$$

donnant ainsi

$$E\left[\sum_{j=1}^{l_n} (k_{nj}^2 \sum_{s=l_n+2}^{t-l_n-1} v_s(l) \langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j} \rangle)\right]^4 \leq 3 \left\{ \sum_{j=1}^{l_n} k_{nj}^4 [E(\sum_{s=l_n+2}^{t-l_n-1} v_s(l) \langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j} \rangle)^4]^{1/2} \right\}^2.$$

On applique de nouveau le lemme 10 aux variables

$$\{v_s(l) \langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j} \rangle, s = l_n + 2, \dots, t - l_n - 1\},$$

fournissant ainsi

$$E\left[\sum_{s=l_n+2}^{t-l_n-1} v_s(l) \langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j} \rangle\right]^4 \leq 3 \left\{ \sum_{s=l_n+2}^{t-l_n-1} [E(v_s(l) \langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j} \rangle)^4]^{1/2} \right\}^2.$$

Or $E(v_s(l) \langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j} \rangle)^4 \leq \mu_4^3(\|\mathbf{v}\|)$. Ainsi

$$E\left[\sum_{s=l_n+2}^{t-l_n-1} v_s(l) \langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j} \rangle\right]^4 = O(t^2),$$

et ce indépendamment de l et de j . Regroupant les résultats obtenus, nous obtenons

$$\begin{aligned} E(B_{nt}^4) &\leq 144d\mu_4(\|\mathbf{v}\|) \sum_{l=1}^d \left\{ \sum_{j=1}^{l_n} k_{nj}^4 \sum_{s=l_n+2}^{t-l_n-1} (E[v_s(l) \langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j} \rangle]^4)^{1/2} \right\}^2, \\ &= O(p_n^2 t^2). \end{aligned}$$

Ainsi,

$$\sigma^{-4}(n)n^{-4} \sum_{t=2l_n+3}^n E(B_{nt}^4) = O(n^{-1}),$$

puisque $\sigma^2(n) = O(p_n)$, ce qui montre a). □

Lemme 10. Soit X_1, \dots, X_n des variables aléatoires telles que $E(X_i) = 0, i = 1, \dots, n$, et si $E[X_i g(X_j, X_k, X_l)] = 0, i \neq j, k, l$ pour toute fonction g , alors

$$E\left[\left(\sum_{i=1}^n X_i\right)^4\right] \leq 3 \left\{ \sum_{i=1}^n [E(X_i^4)]^{1/2} \right\}^2.$$

□

Afin de montrer le point b), il est suffisant de montrer que

$$\sigma^{-4}(n)E([\tilde{B}_n^2 - \sigma^2(n)]^2) \rightarrow 0,$$

où $\ddot{B}_n^2 = E(B_n^2 | \mathcal{F}_{t-1}) = n^{-2} \sum_{t=2l_n+3}^n \ddot{B}_{nt}^2$. On commence par décomposer \ddot{B}_{nt}^2 en plusieurs morceaux de la manière suivante:

$$\ddot{B}_{nt}^2 = E(B_{nt}^2) + 4 \sum_{i=1}^4 D_{int},$$

où

$$\begin{aligned} D_{1nt} &= 2 \sum_{j=2}^{l_n} \sum_{i=1}^{j-1} k_{ni}^2 k_{nj}^2 \mathbf{v}'_{t-i} H'_{i,t-l_n-1} H_{j,t-l_n-1} \mathbf{v}_{t-j}, \\ D_{2nt} &= 2 \sum_{i=1}^{l_n} k_{ni}^4 \sum_{s_1=l_n+3}^{t-l_n-1} \sum_{s_2=l_n+2}^{s_1-1} \mathbf{v}'_{t-i} \mathbf{v}_{s_1-i} \mathbf{v}'_{s_1} \mathbf{v}_{s_2} \mathbf{v}'_{s_2-i} \mathbf{v}_{t-i}, \\ D_{3nt} &= \sum_{i=1}^{l_n} k_{ni}^4 \mathbf{v}'_{t-i} \left[\sum_{s=l_n+2}^{t-l_n-1} (\mathbf{v}_{s-i} \mathbf{v}'_s \mathbf{v}_s \mathbf{v}'_{s-i} - dI_d) \right] \mathbf{v}_{t-i}, \\ D_{4nt} &= d(t - 2l_n - 2) \sum_{i=1}^{l_n} k_{ni}^4 (\mathbf{v}'_{t-i} \mathbf{v}_{t-i} - d). \end{aligned}$$

On montre maintenant les deux lemmes suivants:

Lemme 11.

$$E(D_{1nt}^2) = O(t^2 p_n^2),$$

$$E(D_{2nt}^2) = O(t^2 p_n + t p_n^2),$$

$$E(D_{3nt}^2) = O(t p_n^2),$$

$$E(D_{4nt}^2) = O(t^2 p_n).$$

DÉMONSTRATION. Commençons par montrer le résultat pour D_{1nt} . Posons $\mathbf{a}_j = H_{j,t-l_n-1} \mathbf{v}_{t-j}$.

Nous avons alors:

$$\begin{aligned} E(D_{1nt}^2) &= 4E\left(\sum_{j=2}^{l_n} \sum_{i=1}^{j-1} k_{ni}^4 k_{nj}^4 |\langle \mathbf{a}_i, \mathbf{a}_j \rangle|^2\right), \\ &\leq 4 \sum_{j=2}^{l_n} \sum_{i=1}^{j-1} k_{ni}^4 k_{nj}^4 E(\|\mathbf{a}_i\|^2 \|\mathbf{a}_j\|^2), \\ &= 4 \sum_{l_1, l_2=1}^d \sum_{j=2}^{l_n} \sum_{i=1}^{j-1} k_{ni}^4 k_{nj}^4 E[a_i^2(l_1) a_j^2(l_2)], \\ &\leq 4 \sum_{l_1, l_2=1}^d \sum_{j=2}^{l_n} \sum_{i=1}^{j-1} k_{ni}^4 k_{nj}^4 \{E[a_i^4(l_1)] E[a_j^4(l_2)]\}^{1/2}. \end{aligned}$$

Or on remarque que $a_i(l) = \sum_{s=l_n+2}^{t-l_n-1} v_s(l) \langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j} \rangle$ et on a déjà montré dans la preuve du lemme 9 que $E[a_i^4(l)] = O(t^2)$. Ainsi, $E(D_{1nt}^2) = O(t^2 p_n^2)$.

Montrons maintenant le résultat pour D_{2nt} . Commençons par remarquer que nous pouvons écrire D_{2nt} de la manière suivante:

$$D_{2nt} = \sum_{l_1, l_2=1}^d \sum_{i=1}^{l_n} k_{ni}^4 b_{t-i}^{l_1 l_2} \sum_{s=l_n+3}^{t-l_n-1} \sum_{r=l_n+2}^{s-1} w_{isr}^{l_1 l_2}.$$

où $b_t^{l_1 l_2} = \mathbf{v}_t(l_1) \mathbf{v}_t(l_2)$. Ainsi,

$$\begin{aligned} D_{2nt}^2 &= \sum_{l_1, l_2, m_1, m_2=1}^d \sum_{i_1, i_2=1}^{l_n} k_{ni_1}^4 k_{ni_2}^4 b_{t-i_1}^{l_1 l_2} b_{t-i_2}^{m_1 m_2} \sum_{s_1, s_2=l_n+3}^{t-l_n-1} \sum_{r_1=l_n+2}^{s_1-1} \sum_{r_2=l_n+2}^{s_2-1} w_{i_1 s_1 r_1}^{l_1 l_2} w_{i_2 s_2 r_2}^{m_1 m_2}, \\ &= \sum_{l_1, l_2, m_1, m_2=1}^d \sum_{i=1}^{l_n} k_{ni}^8 b_{t-i}^{l_1 l_2} b_{t-i}^{m_1 m_2} \sum_{s_1, s_2=l_n+3}^{t-l_n-1} \sum_{r_1=l_n+2}^{s_1-1} \sum_{r_2=l_n+2}^{s_2-1} w_{i s_1 r_1}^{l_1 l_2} w_{i s_2 r_2}^{m_1 m_2} + \\ &\quad 2 \sum_{l_1, l_2, m_1, m_2=1}^d \sum_{i_1=2}^{l_n} \sum_{i_2=1}^{i_1-1} k_{ni_1}^4 k_{ni_2}^4 b_{t-i_1}^{l_1 l_2} b_{t-i_2}^{m_1 m_2} \sum_{s_1, s_2=l_n+3}^{t-l_n-1} \sum_{r_1=l_n+2}^{s_1-1} \sum_{r_2=l_n+2}^{s_2-1} w_{i_1 s_1 r_1}^{l_1 l_2} w_{i_2 s_2 r_2}^{m_1 m_2}, \\ &= D_{21nt} + D_{22nt}. \end{aligned}$$

Ainsi, en prenant l'espérance de D_{21nt} , utilisant un résultat similaire au lemme 7, utilisant le lemme 5, on montre que $E(D_{21nt}) = O(p_n t^2)$. En procédant de manière similaire, on montre aussi que $E(D_{22nt}) = O(t p_n^2)$.

Montrons maintenant le résultat pour D_{3nt} . Notons que

$$\begin{aligned} E(D_{3nt}^2) &\leq \left\{ \sum_{i=1}^{l_n} k_{ni}^4 [E(\mathbf{v}'_{t-i} (\sum_{s=l_n+2}^{t-l_n-1} \mathbf{v}_{s-i} \mathbf{v}'_s \mathbf{v}_s \mathbf{v}'_{s-i} - d\mathbf{I}_d) \mathbf{v}_{t-i})^2]^{1/2} \right\}^2, \\ &\leq \mu_4(\|\mathbf{v}\|) \left\{ \sum_{i=1}^{l_n} k_{ni}^4 [E(\|\sum_{s=l_n+2}^{t-l_n-1} \mathbf{v}_{s-i} \mathbf{v}'_s \mathbf{v}_s \mathbf{v}'_{s-i} - d\mathbf{I}_d\|_E^2)]^{1/2} \right\}^2, \\ &= O(t p_n^2). \end{aligned}$$

en utilisant le lemme 12 ci-après.

Lemme 12. $E(\|\sum_{s=l_n+2}^{t-l_n-1} (\mathbf{v}_{s-i} \mathbf{v}'_s \mathbf{v}_s \mathbf{v}'_{s-i} - d\mathbf{I}_d)\|_E^2) = O(t)$.

DÉMONSTRATION. Posons $c_{si}^{lm} = \mathbf{v}'_s \mathbf{v}_s \mathbf{v}_{s-i}(l) \mathbf{v}_{s-i}(m) - d\delta_{lm}$. Nous avons alors

$$\begin{aligned} E\left(\left\|\sum_{s=l_n+2}^{t-l_n-1} (\mathbf{v}_{s-i} \mathbf{v}'_s \mathbf{v}_s \mathbf{v}'_{s-i} - dI_d)\right\|_E^2\right) &= \sum_{l,m=1}^d \left\{ \sum_{s=l_n+2}^{t-l_n-1} E(c_{si}^{lm})^2 + \right. \\ &\quad \left. 2 \sum_{s=l_n+3}^{t-l_n-1} \sum_{r=l_n+2}^{s-1} E(c_{si}^{lm} c_{ri}^{lm}) \right\}, \\ &= O(t), \end{aligned}$$

puisque

$$\begin{aligned} E(c_{si}^{lm})^2 &= \begin{cases} \mu_4(\|\mathbf{v}\|) \mu_4(l, l, l, l) - d^2 & \text{si } l = m \\ \mu_4(\|\mathbf{v}\|) \mu_4(l, l, m, m) & \text{si } l \neq m \end{cases} \\ E(c_{si}^{lm} c_{ri}^{lm}) &= \begin{cases} d[E(\|\mathbf{v}_1\|^2 v_1(l)^2) - d] & \text{si } l = m, r = s - i \\ 0 & \text{sinon.} \end{cases} \end{aligned}$$

□

Nous montrons finalement le résultat pour D_{4nt} . On remarque alors que

$$\begin{aligned} E(D_{4nt}^2) &= d^2(t - 2l_n - 2)^2 \sum_{i=1}^{l_n} k_{ni}^8 E(\mathbf{v}'_{t-i} \mathbf{v}_{t-i} - d)^2, \\ &= O(t^2 p_n), \end{aligned}$$

puisque $E(\mathbf{v}'_{t-i} \mathbf{v}_{t-i} - d)^2 = \mu_4(\|\mathbf{v}\|) - d^2$.

□

Le lemme suivant sera aussi utile pour la suite:

Lemme 13. $E(D_{1nt} D_{1ns}) = 0, t - s > l_n$.

DÉMONSTRATION. Notons que dans D_{1nt} , $\mathbf{v}_{t-j} \in \{\mathbf{v}_{t-l_n}, \mathbf{v}_{t-l_n+1}, \dots, \mathbf{v}_{t-1}\}$. De même pour D_{1ns} , $\mathbf{v}_{s-j} \in \{\mathbf{v}_{s-l_n}, \mathbf{v}_{s-l_n+1}, \dots, \mathbf{v}_{s-1}\}$. Si $t-l_n > s$, alors nous avons la conclusion annoncée.

□

On remarque alors que nous avons:

$$\ddot{B}_n^2 = \sigma^2(n) + 4n^{-2} \sum_{j=1}^4 \sum_{t=2l_n+3}^n D_{jnt}.$$

La preuve du point b) sera complète une fois le lemme suivant démontré.

Lemme 14. $E[p_n^{-2} (n^{-2} \sum_{t=2l_n+3}^n D_{jnt})^2] \rightarrow 0, j = 1, 2, 3, 4$.

DÉMONSTRATION. On note que

$$\begin{aligned} E(n^{-2} \sum_{t=2l_n+3}^n D_{1nt})^2 &= n^{-4} \sum_{s-t \leq l_n} E(D_{1nt}D_{1ns}), \\ &\leq n^{-4} \sum_{s-t \leq l_n} [E(D_{1nt}^2)E(D_{1ns}^2)]^{1/2}, \\ &\leq O(p_n^2 l_n/n) = o(p_n^2). \end{aligned}$$

$$\begin{aligned} E(n^{-2} \sum_{t=2l_n+3}^n D_{2nt})^2 &\leq n^{-4} \left\{ \sum_{t=2l_n+3}^n [E(D_{2nt}^2)]^{1/2} \right\}^2, \\ &= O(p_n) = o(p_n^2). \end{aligned}$$

$$\begin{aligned} E(n^{-2} \sum_{t=2l_n+3}^n D_{3nt})^2 &\leq n^{-4} \left\{ \sum_{t=2l_n+3}^n [E(D_{3nt}^2)]^{1/2} \right\}^2, \\ &= O(p_n^2/n) = o(p_n^2). \end{aligned}$$

$$\begin{aligned} E(n^{-2} \sum_{t=2l_n+3}^n D_{4nt})^2 &\leq n^{-4} \left\{ \sum_{t=2l_n+3}^n [E(D_{4nt}^2)]^{1/2} \right\}^2, \\ &= O(p_n) = o(p_n^2). \end{aligned}$$

Pour un raisonnement similaire, voir aussi Hong (1996a, formules (A7)-(A10)). \square

Ainsi, par le théorème 1, nous avons la normalité asymptotique annoncée. Ceci complète la preuve de la première étape.

Nous allons attaquer maintenant la preuve de la seconde étape. Afin de simplifier la longueur de la preuve, nous allons nous restreindre à démontrer le résultat pour le modèle

$$\mathbf{y}_t = \mathbf{c} + \Lambda_1 \mathbf{y}_{t-1} + \mathbf{V}_0 \mathbf{x}_t + \mathbf{u}_t. \quad (3.28)$$

La preuve pour le modèle général (3.3) est en tout point analogue, avec des développements algébriques cependant un peu plus lourds.

Afin de démontrer l'étape 2, on commence par décomposer

$$\sum_{j=1}^{n-1} k_{nj}^2 (\text{tr}[\mathbf{C}_{\hat{v}}(j)\mathbf{C}'_{\hat{v}}(j)] - \text{tr}[\mathbf{C}_v(j)\mathbf{C}'_v(j)]). \quad (3.29)$$

Or, puisque $\text{tr}(AA') - \text{tr}(BB') = \text{tr}[(A-B)(A-B)'] + 2\text{tr}[B(A-B)']$, il suffit de démontrer les deux résultats suivants:

Résultat 5. $\sum_{j=1}^{n-1} k_{nj}^2 \text{tr}[(\mathbf{C}_{\hat{v}}(j) - \mathbf{C}_v(j))(\mathbf{C}_{\hat{v}}(j) - \mathbf{C}_v(j))'] = O_p(n^{-1})$.

Résultat 6. $\sum_{j=1}^{n-1} k_{nj}^2 \text{tr}[\mathbf{C}_v(j)(\mathbf{C}_{\hat{v}}(j) - \mathbf{C}_v(j))'] = o_p(\sqrt{p_n}/n)$.

Posons

$$\hat{\boldsymbol{\lambda}}_{nt} = (\hat{\mathbf{c}} - \mathbf{c}) + (\hat{\boldsymbol{\Lambda}}_1 - \boldsymbol{\Lambda}_1)\mathbf{y}_{t-1} + (\hat{\mathbf{V}}_0 - \mathbf{V}_0)\mathbf{x}_t, \quad (3.30)$$

$$\hat{\boldsymbol{\gamma}}_{nt} = \boldsymbol{\Sigma}_u^{-1/2} \hat{\boldsymbol{\lambda}}_{nt}. \quad (3.31)$$

Posons aussi $\hat{\mathbf{u}}_t = \mathbf{u}_t - \hat{\boldsymbol{\lambda}}_{nt}$, et $\hat{\mathbf{v}}_t = \mathbf{v}_t - \hat{\boldsymbol{\gamma}}_{nt}$.

Commençons par montrer le résultat 5. Notons que nous pouvons écrire

$$\begin{aligned} \mathbf{C}_{\hat{v}}(j) - \mathbf{C}_v(j) &= n^{-1} \sum_{t=j+1}^n \hat{\mathbf{v}}_t \hat{\mathbf{v}}'_{t-j} - n^{-1} \sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j}, \\ &= -n^{-1} \sum_{t=j+1}^n \hat{\boldsymbol{\gamma}}_{nt} \mathbf{v}'_{t-j} - n^{-1} \sum_{t=j+1}^n \mathbf{v}_t \hat{\boldsymbol{\gamma}}'_{n,t-j} + n^{-1} \sum_{t=j+1}^n \hat{\boldsymbol{\gamma}}_{nt} \hat{\boldsymbol{\gamma}}'_{n,t-j}. \end{aligned} \quad (3.32)$$

Utilisant la relation (3.27), nous avons alors que

$$\sum_{j=1}^{n-1} k_{nj}^2 \text{tr}[(\mathbf{C}_{\hat{v}}(j) - \mathbf{C}_v(j))(\mathbf{C}_{\hat{v}}(j) - \mathbf{C}_v(j))'] \leq 4E_{1n} + 4E_{2n} + 4E_{3n},$$

où

$$\begin{aligned} E_{1n} &= n^{-2} \sum_{j=1}^{n-1} k_{nj}^2 \text{tr}[(\sum_{t=j+1}^n \hat{\boldsymbol{\gamma}}_{nt} \mathbf{v}'_{t-j})(\sum_{t=j+1}^n \hat{\boldsymbol{\gamma}}_{nt} \mathbf{v}'_{t-j})'], \\ E_{2n} &= n^{-2} \sum_{j=1}^{n-1} k_{nj}^2 \text{tr}[(\sum_{t=j+1}^n \mathbf{v}_t \hat{\boldsymbol{\gamma}}'_{n,t-j})(\sum_{t=j+1}^n \mathbf{v}_t \hat{\boldsymbol{\gamma}}'_{n,t-j})'], \\ E_{3n} &= n^{-2} \sum_{j=1}^{n-1} k_{nj}^2 \text{tr}[(\sum_{t=j+1}^n \hat{\boldsymbol{\gamma}}_{nt} \hat{\boldsymbol{\gamma}}'_{n,t-j})(\sum_{t=j+1}^n \hat{\boldsymbol{\gamma}}_{nt} \hat{\boldsymbol{\gamma}}'_{n,t-j})']. \end{aligned}$$

On montre alors le résultat suivant:

Résultat 7. $E_{jn} = O_p(n^{-1}), j = 1, 2, 3$.

DÉMONSTRATION. Commençons par E_{1n} , que l'on borne de la manière suivante:

$$\begin{aligned} E_{1n} &\leq 4\text{tr}[(\hat{\mathbf{c}} - \mathbf{c})' \boldsymbol{\Sigma}_u^{-1} (\hat{\mathbf{c}} - \mathbf{c})] F_{1n} + 4\text{tr}[(\hat{\boldsymbol{\Lambda}}_1 - \boldsymbol{\Lambda}_1)' \boldsymbol{\Sigma}_u^{-1} (\hat{\boldsymbol{\Lambda}}_1 - \boldsymbol{\Lambda}_1)] F_{2n} + \\ &\quad 4\text{tr}[(\hat{\mathbf{V}}_0 - \mathbf{V}_0)' \boldsymbol{\Sigma}_u^{-1} (\hat{\mathbf{V}}_0 - \mathbf{V}_0)] F_{3n}, \end{aligned}$$

où

$$\begin{aligned}
F_{1n} &= \sum_{j=1}^{n-1} k_{nj}^2 \text{tr}[(n^{-1} \sum_{t=j+1}^n \mathbf{v}'_{t-j})(n^{-1} \sum_{t=j+1}^n \mathbf{v}'_{t-j})'], \\
F_{2n} &= \sum_{j=1}^{n-1} k_{nj}^2 \text{tr}[(n^{-1} \sum_{t=j+1}^n \mathbf{y}_{t-1} \mathbf{v}'_{t-j})(n^{-1} \sum_{t=j+1}^n \mathbf{y}_{t-1} \mathbf{v}'_{t-j})'], \\
F_{3n} &= \sum_{j=1}^{n-1} k_{nj}^2 \text{tr}[(n^{-1} \sum_{t=j+1}^n \mathbf{x}_t \mathbf{v}'_{t-j})(n^{-1} \sum_{t=j+1}^n \mathbf{x}_t \mathbf{v}'_{t-j})'],
\end{aligned}$$

en utilisant (3.24) et (3.27). Le résultat pour E_{1n} est basé sur le lemme suivant:

Lemme 15.

$$\begin{aligned}
F_{1n} &= O_p(p_n/n), \\
F_{2n} &= O_p(1), \\
F_{3n} &= O_p(p_n/n).
\end{aligned}$$

DÉMONSTRATION. Le résultat pour F_{1n} s'obtient directement en remarquant que

$$E(|F_{1n}|) = dn^{-2} \sum_{j=1}^{n-1} (n-j) k_{nj}^2 = O(p_n/n).$$

Pour montrer le résultat pour F_{2n} , on commence par ré-écrire le modèle (3.28) de la manière suivante

$$\mathbf{y}_t = \mathbf{c}_0 + \Psi(B) \mathbf{V}_0 \mathbf{x}_t + \Psi(B) \mathbf{u}_t,$$

où $\mathbf{c}_0 = (\mathbf{I}_d - \Lambda_1)^{-1} \mathbf{c}$, $\Psi(B) = (\mathbf{I}_d - \Lambda_1 B)^{-1} = \sum_{j \geq 0} \Lambda_1^j B^j$, avec $\|\Lambda_1\| < 1$. Ainsi on a

$$\begin{aligned}
n^{-1} \sum_{t=j+1}^n \mathbf{y}_{t-1} \mathbf{v}'_{t-j} &= n^{-1} \sum_{t=j+1}^n \mathbf{c}_0 \mathbf{v}'_{t-j} + n^{-1} \sum_{t=j+1}^n (\Psi(B) \mathbf{V}_0 \mathbf{x}_{t-1}) \mathbf{v}'_{t-j} + \\
&\quad n^{-1} \sum_{t=j+1}^n (\Psi(B) \mathbf{u}_{t-1}) \mathbf{v}'_{t-j}.
\end{aligned}$$

Ainsi, nous avons que

$$F_{2n} \leq 4(G_{1n} + G_{2n} + G_{3n}),$$

où

$$\begin{aligned}
G_{1n} &= \sum_{j=1}^{n-1} k_{nj}^2 \text{tr}[(n^{-1} \sum_{t=j+1}^n \mathbf{c}_0 \mathbf{v}'_{t-j})(n^{-1} \sum_{t=j+1}^n \mathbf{c}_0 \mathbf{v}'_{t-j})'], \\
G_{2n} &= \sum_{j=1}^{n-1} k_{nj}^2 \text{tr}[(n^{-1} \sum_{t=j+1}^n (\Psi(B)\mathbf{V}_0 \mathbf{x}_{t-1}) \mathbf{v}'_{t-j})(n^{-1} \sum_{t=j+1}^n (\Psi(B)\mathbf{V}_0 \mathbf{x}_{t-1}) \mathbf{v}'_{t-j})'], \\
G_{3n} &= \sum_{j=1}^{n-1} k_{nj}^2 \text{tr}[(n^{-1} \sum_{t=j+1}^n (\Psi(B)\mathbf{u}_{t-1}) \mathbf{v}'_{t-j})(n^{-1} \sum_{t=j+1}^n (\Psi(B)\mathbf{u}_{t-1}) \mathbf{v}'_{t-j})'].
\end{aligned}$$

On note que $G_{1n} = O_p(p_n/n)$, puisque $\mathbf{c}_0 \sum_{s,t=j+1}^n \mathbf{v}'_{t-j} \mathbf{v}_{s-j} \mathbf{c}'_0 = O_p(n)$.

Pour G_{2n} , remarquons que

$$\begin{aligned}
G_{2n} &= n^{-2} \sum_{j=1}^{n-1} k_{nj}^2 \sum_{t=j+1}^n \|\mathbf{v}_{s-j}\|^2 (\Psi(B)\mathbf{V}_0 \mathbf{x}_{t-1})' (\Psi(B)\mathbf{V}_0 \mathbf{x}_{t-1}) \\
&\quad + 2 \sum_{s=j+1}^n \sum_{j=2}^{t-1} \mathbf{v}'_{t-j} \mathbf{v}_{s-j} (\Psi(B)\mathbf{V}_0 \mathbf{x}_{t-1})' (\Psi(B)\mathbf{V}_0 \mathbf{x}_{s-1}),
\end{aligned}$$

impliquant que

$$\begin{aligned}
E(|G_{2n}|) &= dn^{-2} \sum_{j=1}^{n-1} k_{nj}^2 \sum_{t=j+1}^n E\{[(\Psi(B)\mathbf{V}_0 \mathbf{x}_{t-1})' (\Psi(B)\mathbf{V}_0 \mathbf{x}_{t-1})]\}, \\
&= O(p_n/n),
\end{aligned}$$

et donc $G_{2n} = O_p(p_n/n)$. La preuve pour G_{3n} repose sur le lemme suivant, qui est une généralisation du lemme A.1 de Hong (1996a).

Lemme 16.

$$E\{\text{tr}[(\sum_{t=j+1}^n (\Psi(B)\mathbf{u}_{t-1}) \mathbf{v}'_{t-j})(\sum_{t=j+1}^n (\Psi(B)\mathbf{u}_{t-1}) \mathbf{v}'_{t-j})']\} \leq \Delta_1 n + \Delta_2 n^2 \|\Lambda_1\|_E^{2(j-1)}.$$

DÉMONSTRATION. On note que

$$\begin{aligned}
&(\sum_{t=j+1}^n (\Psi(B)\mathbf{u}_{t-1}) \mathbf{v}'_{t-j})(\sum_{t=j+1}^n (\Psi(B)\mathbf{u}_{t-1}) \mathbf{v}'_{t-j})' = \\
&\quad \sum_{s=j+1}^n \|\mathbf{v}_{s-j}\|^2 \|\Psi(B)\mathbf{u}_{s-1}\|^2 + 2 \sum_{t=j+2}^n \sum_{s=j+1}^{t-1} \langle \mathbf{v}_{t-j}, \mathbf{v}_{s-j} \rangle \langle \Psi(B)\mathbf{u}_{t-1}, \Psi(B)\mathbf{u}_{s-1} \rangle.
\end{aligned}$$

On commence par étudier le premier terme. On commence par noter que

$$\begin{aligned} \|\Psi(B)\mathbf{u}_{t-1}\|^2 &= \sum_{j_1 \geq 0} \mathbf{u}'_{t-j_1-1} (\Lambda_1^{j_1})' (\Lambda_1^{j_1}) \mathbf{u}_{t-j_1-1} + \\ &\quad \sum_{j_1 \neq j_2} \mathbf{u}'_{t-j_1-1} (\Lambda_1^{j_1})' (\Lambda_1^{j_2}) \mathbf{u}_{t-j_2-1}. \end{aligned}$$

Ainsi,

$$\begin{aligned} E(\|\mathbf{v}_{t-j}\|^2 \|\Psi(B)\mathbf{u}_{t-1}\|^2) &= \sum_{j_1 \geq 0} E(\|\mathbf{v}_{t-j}\|^2 \mathbf{u}'_{t-j_1-1} (\Lambda_1^{j_1})' (\Lambda_1^{j_1}) \mathbf{u}_{t-j_1-1}), \\ &\leq \sum_{j_1 \geq 0} E(\|\mathbf{v}_{t-j}\|^2 \|\mathbf{u}_{t-j_1-1}\|^2) \|(\Lambda_1^{j_1})' (\Lambda_1^{j_1})\|_E, \\ &\leq \mu_4(\|\mathbf{v}\|) \|\Sigma_u\| \sum_{j_1 \geq 0} \|\Lambda_1^{j_1}\|_E \|\Lambda_1^{j_1}\|_E, \\ &\leq \mu_4(\|\mathbf{v}\|) \|\Sigma_u\| \sum_{j \geq 0} \|\Lambda_1\|_E^{2j}, \\ &\leq \Delta_1. \end{aligned}$$

On étudie maintenant le second terme. Ainsi,

$$\begin{aligned} \langle \Psi(B)\mathbf{u}_{t-1}, \Psi(B)\mathbf{u}_{s-1} \rangle &= \sum_{j_1 \geq 0} \mathbf{u}'_{t-j_1-1} (\Lambda_1^{j_1})' (\Lambda_1^{j_1}) \mathbf{u}_{s-j_1-1} + \\ &\quad \sum_{j_1 \neq j_2} \mathbf{u}'_{t-j_1-1} (\Lambda_1^{j_1})' (\Lambda_1^{j_2}) \mathbf{u}_{s-j_2-1}. \end{aligned}$$

Finalement, il reste à considérer ceci:

$$\begin{aligned} E(\mathbf{v}'_{t-j} \mathbf{v}_{s-j} \langle \Psi(B)\mathbf{u}_{t-1}, \Psi(B)\mathbf{u}_{s-1} \rangle) &= \\ &\quad \sum_{j_1 \geq 0} E[\mathbf{v}'_{t-j} \mathbf{v}_{s-j} \mathbf{u}'_{t-j_1-1} (\Lambda_1^{j_1})' (\Lambda_1^{j_1}) \mathbf{u}_{s-j_1-1}] + \\ &\quad \sum_{j_1 \neq j_2} E[\mathbf{v}'_{t-j} \mathbf{v}_{s-j} \mathbf{u}'_{t-j_1-1} (\Lambda_1^{j_1})' (\Lambda_1^{j_2}) \mathbf{u}_{s-j_2-1}], \\ &\leq 2 \|\Lambda_1\|_E^{2(j-1)} \text{tr}(\Sigma_u). \end{aligned}$$

Regroupant les résultats obtenus,

$$\begin{aligned}
E\{\text{tr}[(n^{-1} \sum_{t=j+1}^n (\Psi(B)\mathbf{u}_{t-1})\mathbf{v}'_{t-j})(n^{-1} \sum_{t=j+1}^n (\Psi(B)\mathbf{u}_{t-1})\mathbf{v}'_{t-j})']]\} &\leq \\
n^{-2} \sum_{t=j+1}^n |E(\|\Psi(B)\mathbf{u}_{t-1}\|^2 \|\mathbf{v}_{t-j}\|^2)| &+ \\
2n^{-2} \sum_{t=j+2}^n \sum_{s=j+1}^{t-1} |E(\langle \Psi(B)\mathbf{u}_{s-1}, \Psi(B)\mathbf{u}_{t-1} \rangle \langle \mathbf{v}_{t-j}, \mathbf{v}_{s-j} \rangle)| &, \\
\leq n^{-1} \Delta_1 + \Delta_2 \|\Lambda_1\|_E^{2(j-1)}. &
\end{aligned}$$

Ceci montre la généralisation du lemme A.1 de Hong (1996a). \square

Remarque 3.8.1. *En fait, il découle du lemme 16 le résultat suivant:*

$$E\{\text{tr}[(\sum_{t=j+1}^n \mathbf{y}_{t-1} \mathbf{v}'_{t-j})(\sum_{t=j+1}^n \mathbf{y}_{t-1} \mathbf{v}'_{t-j})']\} \leq \Delta_1 n + \Delta_2 n^2 \|\Lambda_1\|_E^{2(j-1)}. \quad (3.33)$$

Avec l'aide du lemme 16, on peut compléter le résultat pour G_{3n} qui montre que ce terme est borné en probabilité. Ainsi, combinant les résultats pour G_{1n} , G_{2n} et G_{3n} , on peut conclure que $F_{2n} = O_p(1)$.

Montrons maintenant le résultat pour F_{3n} . On vérifie facilement que

$$\begin{aligned}
E(|F_{3n}|) &= dn^{-1} E(\|x_1\|^2) \sum_{j=1}^{n-1} (1 - j/n) k_{nj}^2, \\
&= O(p_n/n),
\end{aligned}$$

par exogénéité stricte du processus \mathbf{x}_t . \square

Ceci montre ainsi le résultat pour E_{1n} . On montre E_{2n} de la même manière. Montrons maintenant le résultat pour E_{3n} . Pour ce faire, commençons par remarquer que

$$|E_{3n}| \leq \sum_{j=1}^{n-1} k_{nj}^2 (n^{-1} \sum_{t=1}^n \|\hat{\gamma}_{nt}\|^2)^2.$$

On montre maintenant le lemme suivant:

Lemme 17. $n^{-1} \sum_{t=1}^n \|\hat{\gamma}_{nt}\|^2 = O_p(n^{-1})$.

DÉMONSTRATION. On obtient le résultat en remarquant que $\|\hat{\gamma}_{nt}\|^2 = \text{tr}[\hat{\gamma}_{nt}\hat{\gamma}'_{nt}]$ et qu'en utilisant les relations (3.30) et (3.31), nous avons:

$$\begin{aligned} \text{tr}(\hat{\gamma}_{nt}\hat{\gamma}'_{nt}) &\leq 4(\hat{c} - \mathbf{c})'\Sigma_u^{-1}(\hat{c} - \mathbf{c}) + 4\text{tr}[(\hat{\Lambda}_1 - \Lambda_1)'\Sigma_u^{-1}(\hat{\Lambda}_1 - \Lambda_1)]\|\mathbf{y}_{t-1}\|^2 + \\ &\quad 4\text{tr}[(\hat{V}_0 - \mathbf{V}_0)'\Sigma_u^{-1}(\hat{V}_0 - \mathbf{V}_0)]\|\mathbf{x}_t\|^2. \end{aligned}$$

□

Ceci montre donc que $E_{3n} = O_p(p_n/n^2) = o_p(n^{-1})$.

□

Ceci complète la preuve du résultat 5.

Afin de montrer le résultat 6, on décompose $\sum_{j=1}^{n-1} k_{nj}^2 \text{tr}[\mathbf{C}_v(j)(\mathbf{C}_{\hat{v}}(j) - \mathbf{C}_v(j))']$ de la manière suivante en utilisant la relation (3.32):

$$\sum_{j=1}^{n-1} k_{nj}^2 \text{tr}[\mathbf{C}_v(j)(\mathbf{C}_{\hat{v}}(j) - \mathbf{C}_v(j))'] = -E_{4n} - E_{5n} + E_{6n},$$

où

$$\begin{aligned} E_{4n} &= \sum_{j=1}^{n-1} k_{nj}^2 \text{tr}[(n^{-1} \sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j})(n^{-1} \sum_{t=j+1}^n \hat{\gamma}_{nt} \mathbf{v}'_{t-j})'], \\ E_{5n} &= \sum_{j=1}^{n-1} k_{nj}^2 \text{tr}[(n^{-1} \sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j})(n^{-1} \sum_{t=j+1}^n \mathbf{v}_t \hat{\gamma}'_{n,t-j})'], \\ E_{6n} &= \sum_{j=1}^{n-1} k_{nj}^2 \text{tr}[(n^{-1} \sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j})(n^{-1} \sum_{t=j+1}^n \hat{\gamma}_{nt} \hat{\gamma}'_{n,t-j})']. \end{aligned}$$

On termine la preuve en montrant que

Résultat 8. $E_{jn} = o_p(\sqrt{p_n}/n), j = 4, 5, 6$.

DÉMONSTRATION. On commence par montrer le résultat pour E_{4n} . On le décompose de la façon suivante,

$$E_{4n} = F_{4n} + F_{5n} + F_{6n},$$

où

$$\begin{aligned}
F_{4n} &= \sum_{j=1}^{n-1} k_{nj}^2 \text{tr}[(n^{-1} \sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j})(\Sigma_u^{-1/2}(\hat{\mathbf{c}} - \mathbf{c})n^{-1} \sum_{t=j+1}^n \mathbf{v}'_{t-j})'] \\
F_{5n} &= \sum_{j=1}^{n-1} k_{nj}^2 \text{tr}[(n^{-1} \sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j})(\Sigma_u^{-1/2}(\hat{\Lambda}_1 - \Lambda_1)n^{-1} \sum_{t=j+1}^n \mathbf{y}_{t-1} \mathbf{v}'_{t-j})'] \\
F_{6n} &= \sum_{j=1}^{n-1} k_{nj}^2 \text{tr}[(n^{-1} \sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j})(\Sigma_u^{-1/2}(\hat{\mathbf{V}}_0 - \mathbf{V}_0)n^{-1} \sum_{t=j+1}^n \mathbf{x}_t \mathbf{v}'_{t-j})']
\end{aligned}$$

Notons que

$$\begin{aligned}
|F_{4n}| &\leq n^{-2} \{\text{tr}[(\hat{\mathbf{c}} - \mathbf{c})' \Sigma_u^{-1} (\hat{\mathbf{c}} - \mathbf{c})]\}^{1/2} \times \\
&\quad \sum_{j=1}^{n-1} k_{nj}^2 \{\text{tr}[(\sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j})(\sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j})']\}^{1/2} \{\text{tr}[(\sum_{t=j+1}^n \mathbf{v}'_{t-j})(\sum_{t=j+1}^n \mathbf{v}'_{t-j})']\}^{1/2}.
\end{aligned}$$

Ainsi, puisque $E[(|X||Y|)^{1/2}] \leq \{E|X|E|Y|\}^{1/2}$ par l'inégalité de Cauchy-Schartz, que

$$\begin{aligned}
E(\text{tr}[(\sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j})(\sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j})']) &= O(n), \\
E(\text{tr}[(\sum_{t=j+1}^n \mathbf{v}'_{t-j})(\sum_{t=j+1}^n \mathbf{v}'_{t-j})']) &= O(n),
\end{aligned}$$

on peut conclure que $F_{4n} = O_p(p_n/n^{3/2})$.

On note de même que

$$\begin{aligned}
|F_{5n}| &\leq \{\text{tr}[(\hat{\Lambda}_1 - \Lambda_1)' \Sigma_u^{-1} (\hat{\Lambda}_1 - \Lambda_1)]\}^{1/2} n^{-2} \sum_{j=1}^{n-1} k_{nj}^2 \{\text{tr}[(\sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j}) \times \\
&\quad (\sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j})']\}^{1/2} \{\text{tr}[(\sum_{t=j+1}^n \mathbf{y}_{t-1} \mathbf{v}'_{t-j})(\sum_{t=j+1}^n \mathbf{y}_{t-1} \mathbf{v}'_{t-j})']\}^{1/2}.
\end{aligned}$$

Puisque

$$\begin{aligned}
E[\sum_{j=1}^{n-1} k_{nj}^2 \{\text{tr}[(\sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j})(\sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j})']\}^{1/2} \{\text{tr}[(\sum_{t=j+1}^n \mathbf{y}_{t-1} \mathbf{v}'_{t-j}) \times \\
(\sum_{t=j+1}^n \mathbf{y}_{t-1} \mathbf{v}'_{t-j})']\}^{1/2}] &\leq \Delta_1^{1/2} n \sum_{j=1}^{n-1} k_{nj}^2 + n^{3/2} \Delta_2^{1/2} \sum_{j=1}^{n-1} k_{nj}^2 \|\Lambda_1\|_E^{j-1},
\end{aligned}$$

en utilisant la relation (3.33), nous avons alors que $F_{5n} = O_p(p_n/n^{3/2} + 1/n)$.

On note de même que

$$|F_{6n}| \leq \{\text{tr}[(\hat{V}_0 - V_0)' \Sigma_u^{-1} (\hat{V}_0 - V_0)]\}^{1/2} n^{-2} \sum_{j=1}^{n-1} k_{nj}^2 \left\{ \text{tr} \left[\left(\sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j} \right) \times \left(\sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j} \right)' \right] \right\}^{1/2} \left\{ \text{tr} \left[\left(\sum_{t=j+1}^n \mathbf{x}_t \mathbf{v}'_{t-j} \right) \left(\sum_{t=j+1}^n \mathbf{x}_t \mathbf{v}'_{t-j} \right)' \right] \right\}^{1/2}.$$

Puisque

$$E\left(\text{tr} \left[\left(\sum_{t=j+1}^n \mathbf{x}_t \mathbf{v}'_{t-j} \right) \left(\sum_{t=j+1}^n \mathbf{x}_t \mathbf{v}'_{t-j} \right)' \right] \right) = O(n),$$

par exogénéité stricte de \mathbf{x}_t , nous pouvons alors conclure que $F_{6n} = O_p(p_n/n^{3/2})$. Ainsi, ceci montre que $E_{4n} = o_p(p_n^{1/2}/n)$ puisque $p_n/n \rightarrow 0$. Pour E_{5n} , notons que

$$\begin{aligned} |E_{5n}| &\leq n^{-2} \sum_{t=j+1}^n k_{nj}^2 \left\{ \text{tr} \left[\left(\sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j} \right) \left(\sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j} \right)' \right] \right\}^{1/2} \times \\ &\quad \left\{ \text{tr} \left[\left(\sum_{t=j+1}^n \mathbf{v}_t \hat{\gamma}'_{n,t-j} \right) \left(\sum_{t=j+1}^n \mathbf{v}_t \hat{\gamma}'_{n,t-j} \right)' \right] \right\}^{1/2}, \\ &\leq n^{-2} \left\{ \sum_{j=1}^{n-1} k_{nj}^2 \text{tr} \left[\left(\sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j} \right) \left(\sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j} \right)' \right] \right\}^{1/2} \times \\ &\quad \left\{ \sum_{j=1}^{n-1} k_{nj}^2 \text{tr} \left[\left(\sum_{t=j+1}^n \mathbf{v}_t \hat{\gamma}'_{n,t-j} \right) \left(\sum_{t=j+1}^n \mathbf{v}_t \hat{\gamma}'_{n,t-j} \right)' \right] \right\}^{1/2}. \end{aligned}$$

Ainsi, ceci nous permet de conclure que $E_{5n} = o_p(p_n^{1/2}/n)$, puisque l'on montre facilement que

$$\sum_{j=1}^{n-1} k_{nj}^2 \text{tr} \left[\left(n^{-1} \sum_{t=j+1}^n \mathbf{v}_t \hat{\gamma}'_{n,t-j} \right) \left(n^{-1} \sum_{t=j+1}^n \mathbf{v}_t \hat{\gamma}'_{n,t-j} \right)' \right] = o_p(n^{-1}).$$

Pour E_{6n} , on procède de manière similaire en remarquant que

$$\begin{aligned}
|E_{6n}| &\leq n^{-2} \sum_{t=j+1}^n k_{nj}^2 \{ \text{tr} [(\sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j}) (\sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j})'] \}^{1/2} \times \\
&\quad \{ \text{tr} [(\sum_{t=j+1}^n \hat{\gamma}_{nt} \hat{\gamma}'_{n,t-j}) (\sum_{t=j+1}^n \hat{\gamma}_{nt} \hat{\gamma}'_{n,t-j})'] \}^{1/2}, \\
&\leq n^{-2} \{ \sum_{j=1}^{n-1} k_{nj}^2 \text{tr} [(\sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j}) (\sum_{t=j+1}^n \mathbf{v}_t \mathbf{v}'_{t-j})'] \}^{1/2} \times \\
&\quad \{ \sum_{j=1}^{n-1} k_{nj}^2 \text{tr} [(\sum_{t=j+1}^n \hat{\gamma}_{nt} \hat{\gamma}'_{n,t-j}) (\sum_{t=j+1}^n \hat{\gamma}_{nt} \hat{\gamma}'_{n,t-j})'] \}^{1/2}.
\end{aligned}$$

Ceci permet de conclure que $E_{6n} = o_p(p_n^{1/2}/n)$. \square

La preuve sera complète si on montre que

$$\sum_{j=1}^{n-1} k_{nj}^2 (\text{tr} [\mathbf{C}_{\hat{u}}^{-1}(0) \mathbf{C}'_{\hat{u}}(j) \mathbf{C}_{\hat{u}}^{-1}(0) \mathbf{C}_{\hat{u}}(j)] - \text{tr} [\boldsymbol{\Sigma}_u^{-1} \mathbf{C}'_u(j) \boldsymbol{\Sigma}_u^{-1} \mathbf{C}_u(j)]) = o_p(\sqrt{p_n}/n), \quad (3.34)$$

où dans la formule (3.15) on a inséré et retiré le terme $\text{tr} [\boldsymbol{\Sigma}_u^{-1} \mathbf{C}'_u(j) \boldsymbol{\Sigma}_u^{-1} \mathbf{C}_u(j)]$ et utilisé le résultat 1. Aussi, par la formule (3.29), en utilisant les résultats 5 et 6, il nous reste à montrer que

$$\sum_{j=1}^{n-1} k_{nj}^2 (\text{tr} [\mathbf{C}_{\hat{u}}^{-1}(0) \mathbf{C}'_{\hat{u}}(j) \mathbf{C}_{\hat{u}}^{-1}(0) \mathbf{C}_{\hat{u}}(j)] - \text{tr} [\boldsymbol{\Sigma}_u^{-1} \mathbf{C}'_{\hat{u}}(j) \boldsymbol{\Sigma}_u^{-1} \mathbf{C}_{\hat{u}}(j)]) = o_p(\sqrt{p_n}/n), \quad (3.35)$$

Or, $\mathbf{C}_{\hat{u}}(0) - \boldsymbol{\Sigma}_u = O_p(n^{-1/2})$, et aussi

$$\mathbf{C}_{\hat{u}}^{-1}(0) - \boldsymbol{\Sigma}_u^{-1} = O_p(n^{-1/2}). \quad (3.36)$$

Puisque $|\text{tr}(AB')| \leq \sqrt{\text{tr}(AA')} \text{tr}(B)$ lorsque B est symétrique et définie positive, ceci nous permet de conclure la preuve.

3.8.2. Preuve du théorème 3

On remarque que l'on a $\hat{\mathbf{f}}_n - \mathbf{f}_{0n} = (\hat{\mathbf{f}}_n - \mathbf{f}_0) - a_n \mathbf{g}$. Par un calcul direct, on obtient alors que

$$\begin{aligned} Q^2(\hat{\mathbf{f}}_n; \mathbf{f}_{0n}) &= Q^2(\hat{\mathbf{f}}_n; \mathbf{f}_0) + \\ & 2\pi \frac{p_n^{1/2}}{n} \int_{-\pi}^{\pi} \text{tr}[\Gamma_u(0)^{-1} \mathbf{g}(\omega) \Gamma_u(0)^{-1} \mathbf{g}(\omega)] d\omega - \\ & 4\pi \frac{p_n^{1/4}}{n^{1/2}} \int_{-\pi}^{\pi} \text{tr}[\Gamma_u(0)^{-1} \mathbf{g}(\omega) \Gamma_u(0)^{-1} (\hat{\mathbf{f}}_n - \mathbf{f}_0)] d\omega. \end{aligned}$$

On pose $\tilde{\mathbf{f}}_n(\omega) = \frac{1}{2\pi} \sum_{j=-n+1}^{n-1} k_{nj} \mathbf{C}_u(j) e^{-i\omega j}$. On écrit alors $\hat{\mathbf{f}}_n - \mathbf{f}_0 = (\hat{\mathbf{f}}_n - \tilde{\mathbf{f}}_n) + (\tilde{\mathbf{f}}_n - \mathbf{f}_0)$. Ceci nous permet d'écrire

$$\begin{aligned} \int_{-\pi}^{\pi} \text{tr}[\Gamma_u(0)^{-1} \mathbf{g}(\omega) \Gamma_u(0)^{-1} (\hat{\mathbf{f}}_n - \mathbf{f}_0)] d\omega &= \\ & \int_{-\pi}^{\pi} \text{tr}[\Gamma_u(0)^{-1} \mathbf{g}(\omega) \Gamma_u(0)^{-1} (\hat{\mathbf{f}}_n - \tilde{\mathbf{f}}_n)] d\omega + \\ & \int_{-\pi}^{\pi} \text{tr}[\Gamma_u(0)^{-1} \mathbf{g}(\omega) \Gamma_u(0)^{-1} (\tilde{\mathbf{f}}_n - \mathbf{f}_0)] d\omega. \end{aligned}$$

Lemme 18.

$$\int_{-\pi}^{\pi} \text{tr}[\Gamma_u(0)^{-1} \mathbf{g}(\omega) \Gamma_u(0)^{-1} (\hat{\mathbf{f}}_n - \tilde{\mathbf{f}}_n)] d\omega = O_p(n^{-1/2}).$$

DÉMONSTRATION. On note que

$$\hat{\mathbf{f}}_n - \tilde{\mathbf{f}}_n = \frac{1}{2\pi} \sum_{j=-n+1}^{n-1} k_{nj} [\mathbf{C}_{\hat{u}}(j) - \mathbf{C}_u(j)] e^{-i\omega j}.$$

Ainsi,

$$\begin{aligned} \int_{-\pi}^{\pi} \text{tr}[\Gamma_u(0)^{-1} \mathbf{g}(\omega) \Gamma_u(0)^{-1} (\hat{\mathbf{f}}_n - \tilde{\mathbf{f}}_n)] d\omega &= \\ & \frac{1}{2\pi} \sum_{j=-n+1}^{n-1} k_{nj} \text{tr}[\Gamma_u(0)^{-1} \mathbf{g}_j \Gamma_u(0)^{-1} (\mathbf{C}_{\hat{u}}(j) - \mathbf{C}_u(j))], \end{aligned}$$

où l'on pose $\mathbf{g}_j = \int_{-\pi}^{\pi} \mathbf{g}(\omega) e^{i\omega j} d\omega$.

$$\begin{aligned} & \left| \sum_{j=-n+1}^{n-1} k_{nj} \operatorname{tr}[\Gamma_u(0)^{-1} \mathbf{g}_j \Gamma_u(0)^{-1} (\mathbf{C}_{\hat{u}}(j) - \mathbf{C}_u(j))] \right| \leq \\ & \sum_{j=-n+1}^{n-1} k_{nj} \operatorname{tr}[\Gamma_u(0)^{-1} \mathbf{g}_j \Gamma_u(0)^{-1} \mathbf{g}'_j]^{1/2} \operatorname{tr}[(\mathbf{C}_{\hat{v}}(j) - \mathbf{C}_v(j))(\mathbf{C}_{\hat{v}}(j) - \mathbf{C}_v(j))']^{1/2} \leq \\ & \left\{ \sum_{j=-n+1}^{n-1} \operatorname{tr}[\Gamma_u(0)^{-1} \mathbf{g}_j \Gamma_u(0)^{-1} \mathbf{g}'_j] \right\}^{1/2} \left\{ \sum_{j=-n+1}^{n-1} k_{nj}^2 \operatorname{tr}[(\mathbf{C}_{\hat{v}}(j) - \mathbf{C}_v(j))(\mathbf{C}_{\hat{v}}(j) - \mathbf{C}_v(j))'] \right\}^{1/2}, \end{aligned}$$

puisque $\sum_{j=-\infty}^{\infty} \operatorname{tr}[\Gamma_u(0)^{-1} \mathbf{g}_j \Gamma_u(0)^{-1} \mathbf{g}'_j] < \infty$ et en utilisant le résultat 5 de l'étape 2. \square

Lemme 19.

$$\int_{-\pi}^{\pi} \operatorname{tr}[\Gamma_u(0)^{-1} \mathbf{g}(\omega) \Gamma_u(0)^{-1} (\tilde{\mathbf{f}}_n - \mathbf{f}_0)] d\omega = O_p(n^{-1/2}).$$

DÉMONSTRATION. On commence par remarquer que

$$\tilde{\mathbf{f}}_n - \mathbf{f}_0 = \frac{1}{2\pi} \sum_{j=-n+1}^{n-1} k_{nj} [\mathbf{C}_u(j) - \Gamma_u(j)] e^{-i\omega j}.$$

Ainsi,

$$\begin{aligned} & \int_{-\pi}^{\pi} \operatorname{tr}[\Gamma_u(0)^{-1} \mathbf{g}(\omega) \Gamma_u(0)^{-1} (\tilde{\mathbf{f}}_n - \mathbf{f}_0)] d\omega = \\ & \frac{1}{2\pi} \sum_{j=-n+1}^{n-1} k_{nj} \operatorname{tr}[\Gamma_u(0)^{-1} \mathbf{g}_j \Gamma_u(0)^{-1} (\mathbf{C}_u(j) - \Gamma_u(j))'], \end{aligned}$$

où \mathbf{g}_j est défini comme précédemment. Posons $\mathbf{g}_{jS} = \Gamma_u(0)^{-1} \mathbf{g}_j \Gamma_u(0)^{-1}$. On note que

$$\operatorname{tr}[\mathbf{g}_{jS} (\mathbf{C}_u(j) - \Gamma_u(j))'] = \sum_{l,m=1}^d g_{jS,lm} [C_{u,lm}(j) - \Gamma_{u,lm}(j)],$$

où $g_{jS,lm}$ et $C_{u,lm}(j) - \Gamma_{u,lm}(j)$ sont les composantes (l, m) de \mathbf{g}_{jS} et $\mathbf{C}_u(j) - \Gamma_u(j)$, respectivement. On note que $\sum g_{jS,lm}^2$ est une série convergente. Ainsi, puisque $E[(\sum_{j=1}^{n-1} g_{jS,lm} C_{u,lm}(j))^2] = O(n^{-1})$, nous obtenons la conclusion recherchée. \square

3.8.3. Preuve du théorème 4

On a que

$$Q^2(\hat{\mathbf{f}}_n; \mathbf{f}_0) = 2\pi \int_{-\pi}^{\pi} \operatorname{tr}[\Gamma_u(0)^{-1} (\hat{\mathbf{f}}_n - \mathbf{f}_0)^* \Gamma_u(0)^{-1} (\hat{\mathbf{f}}_n - \mathbf{f}_0)] d\omega.$$

On commence par écrire $\hat{\mathbf{f}}_n - \mathbf{f}_0 = (\hat{\mathbf{f}}_n - \mathbf{f}) + (\mathbf{f} - \mathbf{f}_0)$. On montre alors par un calcul direct que

$$\begin{aligned} Q^2(\hat{\mathbf{f}}_n; \mathbf{f}_0) &= Q^2(\mathbf{f}; \mathbf{f}_0) + \\ &4\pi \int_{-\pi}^{\pi} \text{tr}[\mathbf{\Gamma}_u(0)^{-1}(\mathbf{f} - \mathbf{f}_0)^* \mathbf{\Gamma}_u(0)^{-1}(\hat{\mathbf{f}}_n - \mathbf{f})] d\omega + \\ &2\pi \int_{-\pi}^{\pi} \text{tr}[\mathbf{\Gamma}_u(0)^{-1}(\hat{\mathbf{f}}_n - \mathbf{f})^* \mathbf{\Gamma}_u(0)^{-1}(\hat{\mathbf{f}}_n - \mathbf{f})] d\omega. \end{aligned}$$

On montre que $\int_{-\pi}^{\pi} \text{tr}[\mathbf{\Gamma}_u(0)^{-1}(\hat{\mathbf{f}}_n - \mathbf{f})^* \mathbf{\Gamma}_u(0)^{-1}(\hat{\mathbf{f}}_n - \mathbf{f})] d\omega = o_p(1)$. Par Cauchy-Schartz, ceci nous permet de montrer aussi que $4\pi \int_{-\pi}^{\pi} \text{tr}[\mathbf{\Gamma}_u(0)^{-1}(\mathbf{f} - \mathbf{f}_0)^* \mathbf{\Gamma}_u(0)^{-1}(\hat{\mathbf{f}}_n - \mathbf{f})] d\omega = o_p(1)$.

Résultat 9.

$$\int_{-\pi}^{\pi} \text{tr}[\mathbf{\Gamma}_u(0)^{-1}(\hat{\mathbf{f}}_n - \mathbf{f})^* \mathbf{\Gamma}_u(0)^{-1}(\hat{\mathbf{f}}_n - \mathbf{f})] d\omega = o_p(1).$$

DÉMONSTRATION. Puisque $\text{tr}[A(B + C)^* A(B + C)] \leq 2\text{tr}[AB^* AB] + 2\text{tr}[AC^* AC]$, où A est symétrique et inversible, par la décomposition $\hat{\mathbf{f}}_n - \mathbf{f} = (\hat{\mathbf{f}}_n - \tilde{\mathbf{f}}_n) + (\tilde{\mathbf{f}}_n - \mathbf{f})$, nous pouvons écrire

$$\begin{aligned} &\int_{-\pi}^{\pi} \text{tr}[\mathbf{\Gamma}_u(0)^{-1}(\hat{\mathbf{f}}_n - \mathbf{f})^* \mathbf{\Gamma}_u(0)^{-1}(\hat{\mathbf{f}}_n - \mathbf{f})] d\omega \leq \\ &2 \int_{-\pi}^{\pi} \text{tr}[\mathbf{\Gamma}_u(0)^{-1}(\hat{\mathbf{f}}_n - \tilde{\mathbf{f}}_n)^* \mathbf{\Gamma}_u(0)^{-1}(\hat{\mathbf{f}}_n - \tilde{\mathbf{f}}_n)] d\omega + \\ &2 \int_{-\pi}^{\pi} \text{tr}[\mathbf{\Gamma}_u(0)^{-1}(\tilde{\mathbf{f}}_n - \mathbf{f})^* \mathbf{\Gamma}_u(0)^{-1}(\tilde{\mathbf{f}}_n - \mathbf{f})] d\omega. \end{aligned}$$

Il est suffisant de montrer les deux lemmes suivants:

Lemme 20.

$$\int_{-\pi}^{\pi} \text{tr}[\mathbf{\Gamma}_u(0)^{-1}(\hat{\mathbf{f}}_n - \tilde{\mathbf{f}}_n)^* \mathbf{\Gamma}_u(0)^{-1}(\hat{\mathbf{f}}_n - \tilde{\mathbf{f}}_n)] d\omega = o_p(1).$$

DÉMONSTRATION. On remarque que nous avons

$$\hat{\mathbf{f}}_n - \tilde{\mathbf{f}}_n = \frac{1}{2\pi} \sum_{j=-n+1}^{n-1} k_{nj} [\mathbf{C}_{\hat{u}}(j) - \mathbf{C}_u(j)] e^{-i\omega j}.$$

En évaluant, nous avons

$$\begin{aligned} &\int_{-\pi}^{\pi} \text{tr}[\mathbf{\Gamma}_u(0)^{-1}(\hat{\mathbf{f}}_n - \tilde{\mathbf{f}}_n)^* \mathbf{\Gamma}_u(0)^{-1}(\hat{\mathbf{f}}_n - \tilde{\mathbf{f}}_n)] d\omega = \\ &\frac{1}{2\pi} \sum_{j=-n+1}^{n-1} k_{nj}^2 \text{tr}[(\mathbf{C}_{\hat{u}}(j) - \mathbf{C}_u(j))(\mathbf{C}_{\hat{u}}(j) - \mathbf{C}_u(j))']. \end{aligned}$$

Or on a déjà majoré

$$\sum_{j=1}^{n-1} k_{nj}^2 \text{tr}[(\mathbf{C}_{\hat{v}}(j) - \mathbf{C}_v(j))(\mathbf{C}_{\hat{v}}(j) - \mathbf{C}_v(j))'].$$

(notons que les cas $j > 0$ et $j < 0$ se font de la même manière et il n'y a pas de difficultés à traiter $j = 0$). En effet, on avait majoré par $4E_{1n} + 4E_{2n} + 4E_{3n}$, où E_{jn} sont donnés dans la preuve de l'étape 2. Cependant, étant sous l'hypothèse alternative, on ne peut les traiter de la même façon. On note cependant que par Cauchy-Schartz, nous pouvons affirmer que

$$\left| \sum_{k=1}^n w_k \text{tr} \left[\left(\sum_{t=k+1}^n a_t b'_{t-k} \right) \left(\sum_{t=k+1}^n a_t b'_{t-k} \right)' \right] \right| \leq \left(\sum_{k=1}^n w_k \right) \left(\sum_{t=1}^n \|b_t\|^2 \right) \left(\sum_{t=1}^n \|a_t\|^2 \right).$$

On a alors que

$$|E_{1n}| = E_{1n} \leq \left(\sum_{j=1}^n k_{nj}^2 \right) (n^{-1} \sum_{j=1}^n \|v_t\|^2) (n^{-1} \sum_{j=1}^n \|\hat{\gamma}_{nt}\|^2).$$

Or, $n^{-1} \sum_{j=1}^n \|\hat{\gamma}_{nt}\|^2 = O_p(n^{-1})$, puisque

$$\|\hat{\gamma}_{nt}\|^2 \leq 4(\hat{c} - c)' \Sigma_u^{-1} (\hat{c} - c) + 4 \text{tr}[(\hat{V}_0 - V_0)' \Sigma_u^{-1} (\hat{V}_0 - V_0)] \|\mathbf{x}_t\|^2.$$

Puisque l'estimateur des moindres carrés est convergent dans le modèle statique, ceci montre le résultat. On fait de même pour E_{2n} et E_{3n} . Ceci complète la preuve. \square

Lemme 21.

$$\int_{-\pi}^{\pi} \text{tr}[\Gamma_u(0)^{-1}(\tilde{\mathbf{f}}_n - \mathbf{f})^* \Gamma_u(0)^{-1}(\tilde{\mathbf{f}}_n - \mathbf{f})] d\omega = o_p(1).$$

DÉMONSTRATION. On commence par écrire

$$\tilde{\mathbf{f}}_n - \mathbf{f} = \frac{1}{2\pi} \sum_{|j| \leq n-1} [k_{nj} \mathbf{C}_u(j) - \Gamma_u(j)] e^{-ij\omega} - \frac{1}{2\pi} \sum_{|j| > n-1} \Gamma_u(j) e^{-ij\omega}.$$

Par la suite, en intégrant on trouve que

$$\begin{aligned} \int_{-\pi}^{\pi} \text{tr}[\Gamma_u(0)^{-1}(\tilde{\mathbf{f}}_n - \mathbf{f})^* \Gamma_u(0)^{-1}(\tilde{\mathbf{f}}_n - \mathbf{f})] d\omega = \\ \frac{1}{2\pi} \sum_{|j| \leq n-1} \text{tr}[\Gamma_u(0)^{-1} (k_{nj} \mathbf{C}_u(j) - \Gamma_u(j))' \Gamma_u(0)^{-1} (k_{nj} \mathbf{C}_u(j) - \Gamma_u(j))] \\ + \frac{1}{2\pi} \sum_{|j| \geq n} \text{tr}[\Gamma_u(0)^{-1} \Gamma_u(j)' \Gamma_u(0)^{-1} \Gamma_u(j)]. \end{aligned}$$

Par notre hypothèse sur la fonction d'autocovariance, on peut affirmer que

$$\sum_{|j| \geq n} \text{tr}[\mathbf{\Gamma}_u(0)^{-1} \mathbf{\Gamma}_u(j)' \mathbf{\Gamma}_u(0)^{-1} \mathbf{\Gamma}_u(j)] = o_p(1).$$

Il reste à vérifier que le premier terme du membre de droite est aussi $o_p(1)$. Cependant, en utilisant que $k_{nj} \mathbf{C}_u(j) - \mathbf{\Gamma}_u(j) = (k_{nj} - 1) \mathbf{\Gamma}_u(j) + k_{nj} (\mathbf{C}_u(j) - \mathbf{\Gamma}_u(j))$, on montre que

$$\begin{aligned} & \sum_{|j| \leq n-1} \text{tr}[\mathbf{\Gamma}_u(0)^{-1} (k_{nj} \mathbf{C}_u(j) - \mathbf{\Gamma}_u(j))' \mathbf{\Gamma}_u(0)^{-1} (k_{nj} \mathbf{C}_u(j) - \mathbf{\Gamma}_u(j))] \leq \\ & 2 \sum_{|j| \leq n-1} (k_{nj} - 1)^2 \text{tr}[\mathbf{\Gamma}_u(0)^{-1} \mathbf{\Gamma}_u(j)' \mathbf{\Gamma}_u(0)^{-1} \mathbf{\Gamma}_u(j)] + \\ & 2 \sum_{|j| \leq n-1} k_{nj}^2 \text{tr}[\mathbf{\Gamma}_u(0)^{-1} (\mathbf{C}_u(j) - \mathbf{\Gamma}_u(j))' \mathbf{\Gamma}_u(0)^{-1} (\mathbf{C}_u(j) - \mathbf{\Gamma}_u(j))]. \end{aligned}$$

Pour le premier terme de droite, le même argument que Hong (1996a, p. 861) s'applique, à savoir que ce terme est $o(1)$ par le théorème de convergence dominé de Lebesgue, par nos hypothèses sur le noyau k . Pour l'autre terme, on note que

$$\begin{aligned} & \text{tr}[\mathbf{\Gamma}_u(0)^{-1} (\mathbf{C}_u(j) - \mathbf{\Gamma}_u(j))' \mathbf{\Gamma}_u(0)^{-1} (\mathbf{C}_u(j) - \mathbf{\Gamma}_u(j))] = \\ & \text{tr}[(\mathbf{C}_v(j) - \mathbf{\Gamma}_v(j))' (\mathbf{C}_v(j) - \mathbf{\Gamma}_v(j))] = \\ & \sum_{t=1}^d \sum_{s=1}^d (C_{v,st}(j) - \Gamma_{v,st}(j))^2, \end{aligned}$$

où $C_{v,st}(j)$ et $\Gamma_{v,st}(j)$ sont les composantes (s, t) de $\mathbf{C}_v(j)$ et $\mathbf{\Gamma}_v(j)$, respectivement.

Nous avons alors besoin d'un résultat général sur la variance des autocovariances croisées. Des résultats sont disponibles dans Hannan (1970, pp. 208-211) et Hannan (1976). Dans notre contexte, la variance de l'autocovariance $C_{v,st}(j)$ est donnée par $\text{var}(C_{v,st}(j)) = n^{-1} \sum_{|i| \leq n-1} (1 - |i|/n) [\Gamma_{v,st}(i+j) \Gamma_{v,st}(i-j) + \kappa_{stst}(0, j, i, i+j)]$. Nous avons alors que $\sup_{j \geq 1} \text{var}[C_{v,st}(j)] = O(n^{-1})$ avec nos hypothèses, et ainsi

$$\sum_{|j| \leq n-1} k_{nj}^2 \sum_{t=1}^d \sum_{s=1}^d [C_{v,st}(j) - \Gamma_{v,st}(j)]^2 = O_p(p_n/n).$$

Ceci complète la preuve. □

Les lemmes 20 et 21 complète alors la preuve du résultat 9. □

BIBLIOGRAPHIE

- Ansley, C. F. (1980), 'Computation of the theoretical autocovariance function for a vector ARMA process', *Journal of Statistical Computation and Simulation* **12**, 15–24.
- Bahadur, R. R. (1960), 'Stochastic comparison of tests', *Annals of Mathematical Statistics* **31**, 276–295.
- Beltrao, K. et Bloomfield, P. (1987), 'Determining the bandwidth of a kernel spectrum estimate', *Journal of Time Series Analysis* **8**, 21–38.
- Box, G. E. P. et Pierce, D. A. (1970), 'Distribution of residual autocorrelations in autoregressive-integrated moving average time series models', *Journal of American Statistical Association* **65**, 1509–1526.
- Brockwell, P. J. et Davis, R. A. (1991), *Time Series. Theory and Methods*, second edn, Springer-Verlag:Berlin.
- Brown, B. M. (1971), 'Martingale central limit theorems', *Annals of Mathematical Statistics* **42**, 59–66.
- Chitturi, R. V. (1974), 'Distribution of residual autocorrelations in multiple autoregressive schemes', *Journal of American Statistical Association* **69**, 928–934.
- Fuller, W. A. (1996), *Introduction to Statistical Time Series*, second edn, Wiley:New-York.
- Gilbert, P. D. (1993), 'State space and ARMA models: An overview of the equivalence'. Working Paper 93-4, Bank of Canada.
- Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press:NJ.
- Hannan, E. (1970), *Multiple time series*, Wiley:New-York.
- Hannan, E. J. (1976), 'The asymptotic distribution of serial covariances', *Annals of Statistics* **4**, 396–399.
- Harville, D. A. (1997), *Matrix Algebra from a Statistician's perspective*, Springer-Verlag:Berlin.
- Hendry, D. F. (1995), *Dynamic Econometrics*, Oxford University Press.
- Hong, Y. (1996), 'Consistent testing for serial correlation of unknown form', *Econometrica* **64**, 837–864.
- Hosking, J. (1980), 'The multivariate portmanteau statistic', *Journal of American Statistical Association* **75**, 602–608.
- Judge, G. G., Hill, R. C., Grittiths, W. E., Lütkepohl, H. et Lee, T.-C. (1985), *The theory and practice of econometrics*, second edn, Wiley:New-York.
- Judge, G. G., Hill, R. C., Grittiths, W. E., Lütkepohl, H. et Lee, T.-C. (1988), *Introduction to the Theory and Practice of Econometrics*, second edn, Wiley:New-York.
- Lütkepohl, H. (1993), *Introduction to Multiple Time Series Analysis*, second edn, Springer-Verlag:Berlin.
- Paparoditis, E. (1999), 'Spectral density based goodness-of-fit tests for time series models'. À paraître dans *Scandinavian Journal of Statistics*.
- Paparoditis, E. (2000), 'On some power properties of goodness-of-fit tests in time series analysis'. À paraître dans a Festschrift by Chapman and Hall.
- Pitman, E. J. G. (1979), *Some Basic Theory for Statistical Inference*, Halsted:New-York.

- Priestley, M. B. (1981a), *Univariate Series*, Vol. 1 of *Spectral Analysis and Time Series*, Academic:New-York.
- Priestley, M. B. (1981b), *Multivariate Series, Prediction and Control*, Vol. 2 of *Spectral Analysis and Time Series*, Academic:New-York.
- Robinson, P. M. (1991), 'Automatic frequency domain inference on semiparametric and non-parametric models', *Econometrica* **59**, 1329–1363.
- Shiryaev, A. N. (1995), *Probability*, second edn, Springer-Verlag:Berlin.
- White, H. (1984), *Asymptotic Theory for Econometricians*, Academic:New-York.

Chapitre 4

TESTS D'INDÉPENDANCE ROBUSTES ENTRE DEUX SÉRIES CHRONOLOGIQUES

Résumé

Nous considérons des tests d'indépendance robustes entre deux séries chronologiques unidimensionnelles. Les approches classiques reposent sur l'utilisation des corrélations croisées résiduelles. Avec ces dernières, il est possible de considérer des tests à des délais individuels ou des statistiques de type portmanteau. Voir Haugh (1976) et El Himdi et Roy (1997). L'approche de Hong (1996*b*) généralise d'une certaine manière Haugh (1976) et permet d'obtenir des tests plus puissants. Les approches classiques sont sensibles à la présence de valeurs aberrantes. Les résultats de Li et Hui (1994) montrent bien que les niveaux et les puissances des tests peuvent être affectés quand des valeurs aberrantes surviennent dans les séries. Nous considérons ici deux approches. La première repose sur des statistiques de tests à des délais individuels, utilisant la fonction de corrélation croisée robustifiée et les résultats généraux de Li et Hui (1994). La seconde généralise Hong (1996*b*) au cas robuste, et généralise d'une certaine manière les résultats de Li et Hui (1994). Les tests proposés offrent de meilleures propriétés de puissance que les tests robustes de type portmanteau de Li et Hui (1994). L'étude de la causalité de façon robuste est discutée. Une étude empirique illustre les propriétés de niveaux et de puissance des tests robustes de type portmanteau. Une application avec des données réelles illustre la méthodologie.

4.1. INTRODUCTION

Plusieurs phénomènes physiques ou économiques peuvent être décrits avec des séries chronologiques. Voir par exemple pour des applications touchant la physique Akaike et Kitagawa (1999) ou pour des applications économétriques Judge, Hill, Grittiths, Lütkepohl et Lee (1985). Lorsque nous avons à notre disposition deux séries chronologiques, une question naturelle porte sur l'existence de relations entre elles. En économie, l'élucidation des relations de causalité entre les séries chronologiques est souvent très importante dans un contexte de prévision. Une autre question importante est l'indépendance entre deux séries chronologiques. Comme l'hypothèse d'indépendance entre deux séries implique la non-causalité, alors si des

procédures raisonnablement puissantes pour vérifier l'indépendance sont à notre disposition, alors une analyse de causalité peut devenir redondante.

La plupart des approches connues reposent habituellement sur l'utilisation de la fonction de corrélation croisée. Par exemple, Haugh (1976) a développé une procédure dans laquelle des modèles ARMA sont ajustés aux deux séries. La statistique de test est basée sur les corrélations croisées résiduelles, notées $r_{\hat{u}\hat{v}}(j)$, qui sont définies par

$$r_{\hat{u}\hat{v}}(j) = \frac{\sum_{t=j+1}^n \hat{u}_t \hat{v}_{t-j}}{(\sum_{t=1}^n \hat{u}_t^2 \sum_{t=1}^n \hat{v}_t^2)^{1/2}},$$

pour $|j| \leq n-1$, où $\hat{u}_t, \hat{v}_t, t = 1, \dots, n$, désignent les deux séries résiduelles et n est la longueur des deux séries. La distribution asymptotique d'un nombre fixé de corrélations croisées résiduelles est établie par Haugh (1976). Il considère alors une statistique de type portman-teau, basée sur la somme d'un nombre fixé de corrélations croisées résiduelles au carré. Plus précisément, il considère la statistique

$$S_M = n \sum_{j=-M}^M r_{\hat{u}\hat{v}}^2(j), \quad (4.1)$$

où $M \leq n-1$ est un entier fixé. La distribution asymptotique est de type khi-carré, et l'on rejette l'hypothèse d'indépendance pour les grandes valeurs de la statistique de test. El Himdi et Roy (1997) généralisent l'approche de Haugh à un contexte multidimensionnel.

Hong (1996b) a généralisé l'approche de Haugh dans plusieurs directions. La statistique de test qu'il propose est une somme pondérée des carrés des corrélations croisées de la forme

$$Q_n(k) = \frac{n \sum_{j=1-n}^{n-1} k^2(j/m) r_{\hat{u}\hat{v}}^2(j) - M_n(k)}{[2V_n(k)]^{1/2}}, \quad (4.2)$$

où $M_n(k) = \sum_{j=1-n}^{n-1} (1-|j|/n) k^2(j/m)$ et $V_n(k) = \sum_{j=2-n}^{n-2} (1-|j|/n)(1-(|j|+1)/n) k^4(j/m)$. La pondération s'effectue avec un noyau k , et dans ce sens le test de Haugh peut être considéré comme un cas particulier lorsque l'on choisit le noyau uniforme tronqué. Cependant beaucoup de noyaux offrent une puissance supérieure au noyau uniforme tronqué. Aussi, Hong (1996b) évite l'utilisation de modèles ARMA en ajustant des modèles autorégressifs d'un ordre suffisamment élevé. Il est important de noter que les techniques de preuves de Hong (1996b) sont différentes de celles de Haugh (1976). Ainsi, Hong (1996b) établit directement la distribution asymptotique de la statistique Q_n , sans recourir à la distribution asymptotique d'un vecteur de taille fixe de corrélations croisées résiduelles. La statistique de test est sous l'hypothèse

nulle asymptotiquement $N(0, 1)$ et le test est unilatéral, rejetant pour de grandes valeurs de Q_n .

En pratique, la présence de valeurs aberrantes dans les séries chronologiques est un phénomène non-négligeable. Il peut être occasionné par exemple par des erreurs dans l'écriture des données, ou des défaillances temporaires dans les équipements. Les procédures de test d'indépendance comme celle de Haugh ou de Hong reposent sur des méthodes d'estimation sensibles à la présence de valeurs aberrantes. Le choix de la fonction de corrélation croisée usuelle n'est pas souhaitable en présence de valeurs aberrantes. Des procédures d'estimation robuste dans les modèles ARMA ont été étudiées par Bustos et Yohai (1986), et une fonction d'autocovariance robustifiée est proposée. En présence de deux séries chronologiques, Li et Hui (1994) proposent une statistique de test généralisant Haugh (1976), ainsi que les tests considérés par McLeod (1979). Ils utilisent une fonction de corrélation croisée robustifiée. Leur approche consiste à établir d'abord la distribution asymptotique d'un nombre fixé de corrélations croisées résiduelles robustifiées. Des statistiques de test dont la distribution asymptotique est khi-carré sont alors obtenues.

Une autre approche de type nonparamétrique entre deux processus autorégressifs est élaborée dans Hallin, Jurečková, Picek et Zahaf (1999), et repose sur les scores de rangs d'autorégression.

Le but de cet article est de décrire deux statistiques de test afin de tester la non-corrélation de manière robuste. La première est basée sur la corrélation croisée résiduelle robuste à un délai particulier. Cette approche utilise les résultats généraux obtenus par Li et Hui (1994). La seconde est basée sur une statistique de type portmanteau qui généralise d'une certaine manière au cas robuste la statistique de Hong (1996*b*). La statistique de test est une somme pondérée des carrés des corrélations croisées résiduelles robustes. La distribution de la statistique sous l'hypothèse nulle d'indépendance est $N(0, 1)$. Nous présumons que les deux séries chronologiques sont ARMA. Nous discutons l'estimation des paramètres suivant la procédure robuste de Bustos et Yohai (1986), quoique toute méthode robuste et convergente à la vitesse $n^{-1/2}$ est suffisante pour que la distribution de la statistique sous l'hypothèse nulle soit asymptotiquement normale. Nous discutons aussi comment adapter les procédures pour tester la causalité au sens de Granger (1969) entre les deux séries, suivant une approche similaire à El Himdi et Roy (1997), avec les résultats de Li et Hui (1994). Les nouveaux tests sont étudiés empiriquement relativement au niveau et à la puissance, avec plusieurs choix des noyaux et de fonctions intervenant dans la définition des corrélations croisées robustifiées. Les résultats

suggèrent que des gains peuvent être obtenus en choisissant un noyau autre que le noyau uniforme tronqué. Une application est présentée afin d'illustrer la méthodologie.

4.2. PRÉLIMINAIRES

4.2.1. Hypothèses sur les processus

Soit $\{(X_t, Y_t), t \in \mathbb{Z}\}$ un processus bivarié conjointement stationnaire et linéaire. Sans perte de généralité, nous supposons que $\{(X_t, Y_t)\}$ est centré en zéro. Nous supposons également que $\{X_t\}$ et $\{Y_t\}$ admettent tous deux une représentation ARMA, c'est-à-dire qu'ils satisfont les équations

$$\begin{aligned}\phi_1(B)X_t &= \theta_1(B)u_t, \\ \phi_2(B)Y_t &= \theta_2(B)v_t,\end{aligned}$$

où $\phi_h(B) = \sum_{i=0}^{p_h} \phi_{hi}B^i$, $\theta_h(B) = \sum_{i=0}^{q_h} \theta_{hi}B^i$, $h = 1, 2$, et B est l'opérateur retard. On suppose que $\phi_{h0} = \theta_{h0} = 1$, $h = 1, 2$. Il est présumé que $\{X_t\}$ et $\{Y_t\}$ sont stationnaires et inversibles. Posons $\psi_i(z) = \{\phi_i(z)\}^{-1}\theta_i(z) = \sum_{r=0}^{\infty} \psi_{i,r}z^r$, et $\pi_i(z) = \theta_i(z)^{-1} = \sum_{r=0}^{\infty} \pi_{i,r}z^r$, $i = 1, 2$. Sous les hypothèses de stationnarité et d'inversibilité, c'est-à-dire que $\phi_h(z) \neq 0$, $|z| \leq 1$ et $\theta_h(z) \neq 0$, $|z| \leq 1$, les éléments $\pi_{i,r}$ et $\psi_{i,r}$ convergent vers zéro de manière exponentielle.

On suppose également que $\{u_t\}$ et $\{v_t\}$ sont des suites de variables aléatoires indépendantes et identiquement distribuées, centrées en zéro, de variance σ_u^2 et σ_v^2 , respectivement. On suppose que les moments d'ordre quatre de $\{u_t\}$ et $\{v_t\}$ existent. De plus, on suppose que u_t et v_t ont des lois symétriques autour de zéro. Cette dernière hypothèse est fréquente dans un contexte de robustesse. Voir Bustos et Yohai (1986, p. 167).

4.2.2. Hypothèses sur les noyaux

Les tests de Hong (1996b) peuvent être interprétés à l'aide de la fonction de cohérence, définie par

$$f(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \rho_{uv}(j)e^{-i\omega j},$$

où $\rho_{uv}(j) = \text{cor}(u_t, v_{t-j})$ est la corrélation croisée théorique, de délai j entre $\{u_t\}$ et $\{v_t\}$.

L'estimateur de la cohérence selon la méthode du noyau est donné par

$$\hat{f}_n(\omega) = \frac{1}{2\pi} \sum_{j=-n+1}^{n-1} k(j/m) \rho_{\hat{u}\hat{v}}(j) e^{-i\omega j},$$

où k est un noyau et m est un paramètre de troncature ou de lissage. Comme dans Hong (1996*b*), notre approche nécessite une fonction de noyau satisfaisant l'hypothèse A suivante:

Hypothèse A: On suppose que le noyau $k : \mathbb{R} \rightarrow [-1, 1]$ est une fonction symétrique, continue à 0 et ayant au plus un nombre fini de discontinuités, telle que $k(0) = 1$ et $\int_{-\infty}^{\infty} k^2(z) dz < \infty$.

Des exemples de noyaux sont:

$$\text{Noyau tronqué: } k(z) = \begin{cases} 1 & \text{si } |z| \leq 1, \\ 0 & \text{sinon,} \end{cases}$$

$$\text{Bartlett: } k(z) = \begin{cases} 1 - |z| & \text{si } |z| \leq 1, \\ 0 & \text{sinon,} \end{cases}$$

$$\text{Daniell: } k(z) = \frac{\sin(\pi z)}{\pi z},$$

$$\text{Parzen: } k(z) = \begin{cases} 1 - 6(\pi z/6)^2 + 6|\pi z/6|^3 & \text{si } |z| \leq 3/\pi, \\ 2(1 - |z|)^3 & \text{si } 3/\pi \leq |z| \leq 6/\pi, \\ 0 & \text{sinon,} \end{cases}$$

$$\text{Bartlett-Priestley: } k(z) = \frac{9}{5\pi^2 z^2} \left\{ \frac{\sin(\pi\sqrt{5/3}z)}{\pi\sqrt{5/3}z} - \cos(\pi\sqrt{5/3}z) \right\}.$$

Pour plus de détails sur les différents noyaux, on peut consulter Priestley (1981*a*, 1981*b*).

4.2.3. Valeurs aberrantes en séries chronologiques

En séries chronologiques, l'étude du comportement des estimateurs usuels en présence de valeurs aberrantes se fait habituellement en présumant un certain modèle pour l'apparition des valeurs aberrantes. Fox (1972) a introduit deux types de valeurs aberrantes, ces dernières étant également considérées dans Derby et Martin (1979) et Bustos et Yohai (1986). Ces deux types sont les valeurs aberrantes dans le processus innovation et les valeurs aberrantes additives. Rousseeuw et Leroy (1987, pp. 273-284) présentent une introduction aux méthodes robustes en séries chronologiques. Nous discutons les types de valeurs aberrantes en faisant intervenir des lois de probabilité. Ce n'est pas absolument nécessaire, et des discussions sans l'intervention de lois de probabilité se retrouvent dans Wei (1990, p. 195) et Abraham et Ledolter (1983, p. 356).

a) *Aberrants dans le processus innovation*

Supposons que $\{X_t\}$ est un processus ARMA(p,q), comme décrit dans la section 4.2.1. Cependant, le processus innovation $\{u_t\}$ possède une distribution F s'approchant de la distribution gaussienne mais avec des ailes plus lourdes, comme par exemple une normale contaminée:

$$F = (1 - \epsilon)N(0, \sigma^2) + \epsilon N(0, \tau^2),$$

où $\epsilon > 0$ est petit, et $\tau \geq \sigma$.

Ainsi, les innovations $\{u_t\}$ proviennent d'une $N(0, \sigma^2)$ avec probabilité $1 - \epsilon$, et d'une $N(0, \tau^2)$ dont la variance est grande avec probabilité ϵ . Dans cette situation, on peut considérer que les valeurs aberrantes sont les innovations provenant de la distribution $N(0, \tau^2)$.

L'aspect important avec les valeurs aberrantes dans les innovations est que le modèle ARMA(p,q) tient toujours, mais des valeurs aberrantes surviennent dans les u_t . Ainsi, si pour un certain t_0 une valeur aberrante survient, alors u_{t_0} affectera non seulement X_{t_0} , mais aussi les valeurs futures X_{t_0+1} , X_{t_0+2} , et ainsi de suite. Cependant, après un certain temps, cet effet s'amenuise. Voir le tableau de gauche dans la figure 4.1. En fait, Bustos et Yohai (1986, p. 156) énoncent plusieurs résultats qui montrent que les valeurs aberrantes dans les innovations ne sont pas trop problématiques, même pour l'estimateur des moindres carrés des paramètres autorégressifs et moyennes mobiles d'un ARMA.

b) *Aberrants additifs*

Supposons que $\{X_t\}$ est obtenu de la relation

$$X_t = \tilde{X}_t + V_t,$$

où $\{\tilde{X}_t\}$ est ARMA, c'est-à-dire $\phi_1(B)\tilde{X}_t = \theta_1(B)u_t$, avec $\{u_t\}$ des innovations gaussiennes de variance σ^2 , et $\{V_t\}$ est une suite de variable iid, indépendante de \tilde{X}_t . Les variables V_t ont la distribution

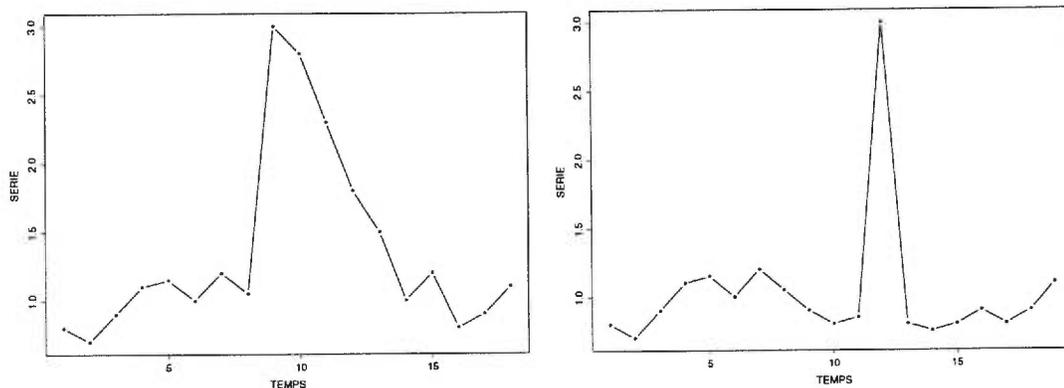
$$H = (1 - \epsilon)\delta_0 + \epsilon G,$$

où δ_0 est la distribution qui assigne zéro avec probabilité un, et G est arbitraire.

Ainsi, dans cette situation, le processus ARMA est observé avec probabilité $1 - \epsilon$ et avec probabilité ϵ le processus ARMA plus une erreur de distribution G est observée. Voir le tableau de droite dans la figure 4.1. L'estimateur des moindres carrés est sensible à la présence des

valeurs aberrantes additives, et Bustos et Yohai (1986) proposent des méthodes d'estimation qui ont un bon comportement en présence de ces dernières.

FIG. 4.1. *Types de valeurs aberrantes. Le graphique de gauche correspond aux valeurs aberrantes dans les innovations alors que le graphique de droite correspond aux valeurs aberrantes additives.*



4.2.4. Estimateurs robustes et fonction d'autocovariance robuste

Nous allons nous concentrer sur l'estimation des paramètres du processus $\{X_t\}$ de type ARMA(p, q) stationnaire et inversible. Soit $\phi(B)X_t = \theta(B)u_t$. Ainsi nous avons que $X_t = \phi^{-1}(B)\theta(B)u_t$ et également $u_t = \theta(B)^{-1}\phi(B)X_t$. Afin d'estimer les paramètres du processus $\{X_t\}$ de manière robuste, nous considérons les estimateurs robustes *RA* (pour *Residual Autocovariances*) définis dans Bustos et Yohai (1986). Ces estimateurs robustes des paramètres autorégressifs et moyennes mobiles sont obtenus du système (4.3) que satisfont les estimateurs des moindres carrés lorsque γ_u correspond à l'autocovariance usuelle. Bustos et Yohai (1986) remplacent γ_u par une version robustifiée. Posons $\phi^{-1}(z) = \sum_{j \geq 0} s_j z^j$ et $\theta^{-1}(z) = \sum_{j \geq 0} t_j z^j$. Ainsi, les estimateurs *RA* sont définis à travers la résolution du système suivant:

$$\begin{aligned}
 \sum_{h=0}^{n-j-p-1} s_j \gamma_u(h+j; \eta) &= 0, \quad j = 1, \dots, p; \\
 \sum_{h=0}^{n-j-p-1} t_j \gamma_u(h+j; \eta) &= 0, \quad j = 1, \dots, q; \\
 \sum_{t=p+1}^n \psi(u_t / \sigma_u) &= 0,
 \end{aligned} \tag{4.3}$$

où la fonction ψ est une fonction continue et impaire. Un exemple est la famille de Huber

$$\psi_H(u; c) = \text{signe}(u) \min(|u|, c).$$

Un autre choix discuté dans Bustos et Yohai (1986) est la famille bicarrée définie par

$$\psi_B(u; c) = \begin{cases} u(1 - u^2/c^2)^2 & \text{si } 0 \leq |u| \leq c, \\ 0 & \text{sinon.} \end{cases}$$

L'estimateur de $\hat{\sigma}_u$ est calculé simultanément en utilisant par exemple

$$\hat{\sigma}_u = \text{med}(|\hat{u}_{p+1}|, \dots, |\hat{u}_n|) / .6745.$$

La fonction $\gamma_u(i; \eta)$ est une robustification de la fonction d'autocorrélation, définie par

$$\gamma_u(i; \eta) = n^{-1} \sum_{t=p+1+i}^n \eta(u_t/\sigma_u, u_{t-i}/\sigma_u).$$

Nous définissons de façon similaire $\gamma_{\hat{u}}(i; \eta)$ calculée avec la série résiduelle, et où l'écart-type σ_u est estimé avec les observations.

Plusieurs choix de fonction η sont possibles, dont les choix de type Mallows et ceux de type Hampel:

$$\eta_M(u, v) = \psi(u)\psi(v), \quad \text{Mallows,}$$

$$\eta_H(u, v) = \psi(uv), \quad \text{Hampel,}$$

où la fonction ψ pourrait être dans la famille Huber ou bicarré. Notons que le choix $\eta(u, v) = uv$ permet de retrouver la définition usuelle de la fonction de corrélation. Aussi, il est utile de noter que l'estimateur des moindres carrés est obtenu en considérant la résolution du système (4.3) avec $\eta(u, v) = uv$ et $\psi(u) = u$.

Les estimateurs RA de Bustos et Yohai (1986) convergent à la vitesse $n^{-1/2}$ et sont asymptotiquement normaux. Cependant, pour que le résultat principal soit valide, il suffit que les estimateurs satisfassent l'hypothèse B suivante:

Hypothèse B: Soient $\{X_t\}$ et $\{Y_t\}$ ARMA comme décrits dans la section 4.2.1. On suppose que les estimateurs des paramètres satisfont les conditions suivantes:

$$\phi_{ij} - \hat{\phi}_{ij} = O_p(n^{-1/2}), j = 1, \dots, p; i = 1, 2;$$

$$\theta_{ij} - \hat{\theta}_{ij} = O_p(n^{-1/2}), j = 1, \dots, q; i = 1, 2;$$

$$\sigma_u - \hat{\sigma}_u = O_p(n^{-1/2}); \sigma_v - \hat{\sigma}_v = O_p(n^{-1/2}).$$

En présence de deux séries chronologiques, une mesure souvent adoptée pour mesurer la corrélation entre les deux processus est la fonction de corrélation croisée. Cependant, tout comme la fonction d'autocorrélation, elle est sensible à la présence de valeurs aberrantes dans les séries à l'étude. Dans un contexte de robustesse, Li et Hui (1994) ont proposé la fonction de corrélation croisée robustifiée suivante:

$$\gamma_{uv}(j; \eta) = \begin{cases} n^{-1} \sum_{t=j+1}^n \eta(u_t/\sigma_u, v_{t-j}/\sigma_v), & j \geq 0, \\ n^{-1} \sum_{t=-j+1}^n \eta(u_{t+j}/\sigma_u, v_t/\sigma_v), & j < 0, \end{cases}$$

où $\{u_t\}$ et $\{v_t\}$ correspondent aux deux séries d'innovation, de variance σ_u^2 et σ_v^2 respectivement. Lorsque $\eta(u, v) = uv$, nous obtenons la fonction de corrélation croisée échantillonnale usuelle, et nous la notons simplement $\gamma_{uv}(j; \eta) = r_{uv}(j)$. De façon similaire, les fonctions de corrélations croisées robustifiées résiduelles sont définies en remplaçant $\{u_t\}$ et $\{v_t\}$ par les séries résiduelles $\{\hat{u}_t\}$ et $\{\hat{v}_t\}$. Li et Hui (1994) ont étudié les propriétés asymptotiques d'un vecteur de taille fixé de corrélations croisées résiduelles robustes.

4.3. TESTS BASÉS SUR LA DISTRIBUTION ASYMPTOTIQUE DES CORRÉLATIONS CROISÉES ROBUSTES

Nous allons considérer dans cette section des procédures de test reposant sur la distribution asymptotique d'un vecteur de corrélations croisées robustifiées. Cette section est en grande partie une application des résultats généraux obtenus dans Li et Hui (1994). Nous rappelons les résultats asymptotiques utiles dans la section 4.3.1. Nous considérons une procédure de test basée sur les corrélations croisées robustes à des délais individuels dans la section 4.3.2. Nous effectuons dans la section 4.3.3 des liens avec la causalité, suivant une approche similaire à celle de El Himdi et Roy (1997).

4.3.1. Distribution asymptotiques des corrélations croisées robustes

La distribution asymptotique d'un vecteur fixé de corrélations croisées robustes a été établi dans un contexte général dans Li et Hui (1994). L'approche utilisée pour démontrer les résultats repose fortement sur l'approche de McLeod (1979). Considérons le vecteur $\gamma_{uv}^n = (\gamma_{uv}(-M; \eta), \dots, \gamma_{uv}(0; \eta), \dots, \gamma_{uv}(M; \eta))'$, de dimension $m = 2M + 1$. Le résultat suivant est un corollaire d'un théorème établi dans Li et Hui (1994, p. 106).

Théorème 1. *Soient $\{X_t\}$ et $\{Y_t\}$ deux processus ARMA stationnaires au second ordre et inversibles, tels que $\{u_t\}$ et $\{v_t\}$ sont des suites de variables aléatoires iid. Soient $\{\hat{u}_t\}$ et $\{\hat{v}_t\}$ les*

séries résiduelles obtenues suite à l'estimation des paramètres avec une méthode convergente à la vitesse $n^{-1/2}$. Si les deux processus sont non corrélés, alors les vecteurs $\sqrt{n}\gamma_{\hat{u}\hat{v}}^\eta$ et $\sqrt{n}\gamma_{\hat{v}\hat{u}}^\eta$ possèdent la même distribution asymptotique $N_m(\mathbf{0}, a\mathbf{I}_m)$, où $a = E[\eta^2(u_1/\sigma_u, v_1/\sigma_v)]$.

Un résultat similaire a été établi par Haugh (1976) et El Himdi et Roy (1997) pour le cas particulier $\mathbf{r}_{uv} = (r_{uv}(-M), \dots, r_{uv}(0), \dots, r_{uv}(M))'$ où les paramètres sont estimés par moindres carrés conditionnels. Cela explique pourquoi les statistiques portmanteau de Haugh convergent vers des lois khi-carré avec $2M + 1$ degrés de liberté, sans perte de degrés de liberté suite à l'estimation de modèles ARMA aux deux séries à l'étude. Les deux prochaines sections sont des applications du théorème 1.

4.3.2. Test basé sur des corrélations croisées robustes à un délai

Nous voulons tester de manière robuste que deux processus ARMA stationnaires et inversibles $\{X_t\}$ et $\{Y_t\}$ sont non-corrélés, c'est-à-dire:

$$H_0 : \rho_{uv}(j) = 0,$$

où $\rho_{uv}(j)$ est la corrélation croisée théorique, de délai j entre $\{u_t\}$ et $\{v_t\}$.

Pour tester cette hypothèse, El Himdi et Roy (1997) ont construit des statistiques utilisant la distribution asymptotique de $\mathbf{r}_{\hat{u}\hat{v}}$. Les statistiques de tests utilisées dépendent des alternatives considérées. Dans un contexte de robustesse, pour un ensemble fixé de délais $\{j_1, \dots, j_m\}$, la distribution asymptotique du vecteur $\gamma_{\hat{u}\hat{v}}^\eta$ donnée dans le théorème 1 permet de construire des tests qui seront des versions robustes de ceux de El Himdi et Roy (1997).

Sous H_0 , il découle du théorème 1 que $\sqrt{n}\gamma_{\hat{u}\hat{v}}(j_i; \eta)/\sqrt{a}$, $i = 1, \dots, m$ sont asymptotiquement indépendants et identiquement distribués $N(0, 1)$. Pour l'hypothèse alternative,

$$H_{1j} : \rho_{uv}(j) \neq 0,$$

il est naturel de considérer la statistique

$$S_R(j) = n\gamma_{\hat{u}\hat{v}}^2(j; \eta)/\hat{a},$$

et sous H_0 , $S_R(j)$ est distribué selon une loi χ_1^2 , où \hat{a} est un estimateur convergent de a . Par exemple, dans le cas où η est de type Mallows, un estimateur convergent pour a est $\hat{a} = [n^{-1} \sum_{t=1}^n \psi^2(\hat{u}_t/\hat{\sigma}_u)] \times [n^{-1} \sum_{t=1}^n \psi^2(\hat{v}_t/\hat{\sigma}_v)]$. Ainsi, pour un niveau de signification α , H_0 est rejeté si $S_R(j) > \chi_{1,1-\alpha}^2$, où $\chi_{m,p}^2$ dénote le quantile d'ordre p de la distribution χ_m^2 .

El Himdi et Roy (1997) ont construit des statistiques basées sur

$$S(j) = nr_{\hat{u}\hat{v}}^2(j),$$

et dans une étude de simulation, ils ont montré que le niveau se détériorait assez rapidement pour des délais élevés. Selon leurs résultats empiriques, la statistique modifiée

$$S^*(j) = \frac{n}{n - |j|} S(j) \quad (4.4)$$

est mieux approximée par la loi asymptotique. Ainsi, nous considérons également les statistiques modifiées

$$S_R^*(j) = \frac{n}{n - |j|} S_R(j). \quad (4.5)$$

Les distributions asymptotiques de $S_R(j)$ et $S_R^*(j)$ sont les mêmes.

En pratique, nous serons habituellement intéressé à considérer plusieurs délais, par exemple tous les délais tels que $|j| \leq M$, où $M \leq n - 1$. Ainsi, l'hypothèse alternative d'intérêt devient

$$H_1^{(M)} : \text{Il existe au moins un } j, |j| \leq M, \text{ pour lequel } \rho_{uv}(j) \neq 0.$$

Un test global pour H_0 basé sur les statistiques $S_R(j)$, $|j| \leq M$, consiste à rejeter H_0 si au moins pour un délai j , $S_R(j) > \chi_{1,1-\alpha}^2$. Pour obtenir un niveau global α , puisque les $S_R(j)$ sont asymptotiquement indépendants, le niveau α_0 de chaque test doit alors être $\alpha_0 = 1 - (1 - \alpha)^{1/(2M+1)}$.

Cette procédure de test a l'avantage de permettre un examen individuel de chaque délai de manière robuste. Les statistiques de type portmanteau n'offrent évidemment pas cette possibilité. Tout comme dans El Himdi et Roy (1997), nous pouvons représenter graphiquement la démarche de cette section, où on retrouve sur l'axe horizontal les délais considérés et sur l'axe vertical, les diverses valeurs des $S_R(j)$. Ce graphique permet d'apprécier la dépendance entre les deux séries résiduelles à plusieurs délais, et peut parfois être utile afin d'apprécier ce qui a permis de conclure à un rejet de l'hypothèse nulle. Il est souhaitable qu'il soit utilisé conjointement avec les statistiques globales de type portmanteau. Ce genre d'analyse peut également être utile dans les études de causalité.

4.3.3. Liens avec la causalité

Nous allons dans cette section établir des procédures de tests basées sur la distribution asymptotique de la section 4.3.1. Li et Hui (1994) discutent de procédures de causalité basées sur des résultats généraux généralisant dans un contexte de robustesse des résultats de

McLeod (1979). En effet, McLeod (1979) avait obtenu dans le cas non robuste la distribution asymptotique des corrélations croisées résiduelles dans le cas général où $\{X_t\}$ et $\{Y_t\}$ sont possiblement corrélés. L'approche décrite ici généralise au cas robuste l'approche décrite dans El Himdi et Roy (1997), où nous considérons cependant deux séries unidimensionnelles. L'approche a le double avantage d'être simple et de ne pas nécessiter l'estimation d'un modèle global ARMA bivarié pour les deux séries. L'estimation robuste des modèles ARMA multivariés est étudiée dans Ben, Martinez et Yohai (1999).

Supposons que $\{(X_t, Y_t), t \in \mathbb{Z}\}$ est stationnaire. Dans un cadre général, le processus $\{X_t\}$ ne cause pas $\{Y_t\}$ au sens de Granger (1969) si la projection affine de Y_t sur l'espace hilbertien engendré par $\{(X_s, Y_s), s \leq t-1\}$, est la même que la projection affine de Y_t sur l'espace hilbertien engendré par $\{Y_s, s \leq t-1\}$. On peut également la formuler en terme des erreurs quadratiques moyennes. Dans ce cas, la définition devient que l'erreur quadratique moyenne de $\{Y_t\}$ conditionnellement à $\{(X_s, Y_s), s \leq t-1\}$ est la même que l'erreur quadratique moyenne de $\{Y_t\}$ conditionnellement à $\{Y_s, s \leq t-1\}$. Ainsi, la connaissance de $\{X_s, s \leq t-1\}$ n'améliore pas l'erreur quadratique moyenne, comparativement au cas où nous connaissons seulement $\{Y_s, s \leq t-1\}$. Pour plus de détails, on peut consulter par exemple Granger (1969) et Boudjellaba et al. (1992, 1994).

Si on suppose que les processus $\{X_t\}$ et $\{Y_t\}$ sont deux processus ARMA stationnaires et inversibles, Pierce et Haugh (1977) ont obtenu une caractérisation de la non-causalité entre $\{X_t\}$ et $\{Y_t\}$ en fonction de la corrélation croisée entre les deux processus innovations correspondants, $\{u_t\}$ et $\{v_t\}$. Ainsi, il est montré que $\{X_t\}$ ne cause pas $\{Y_t\}$ si et seulement si $\text{cov}(u_t, v_{t+k}) = 0, \forall k > 0$. Par conséquent, la fonction de corrélation croisée basée sur les innovations indique la direction de la causalité entre $\{X_t\}$ et $\{Y_t\}$.

Supposons maintenant que nous sommes intéressé à étudier si $\{Y_t\}$ cause $\{X_t\}$. L'hypothèse nulle est l'absence de corrélation entre $\{Y_t\}$ et $\{X_t\}$.

Considérons dans un premier temps l'alternative

$$H_{1M}^+ : \rho_{uv}(j) \neq 0, \text{ pour au moins un } j \text{ tel que } 1 \leq j \leq M.$$

El Himdi et Roy (1997) ont construit des statistiques de test basés sur $r_{\hat{u}\hat{v}}$. Plus précisément, ils considèrent la statistique

$$S_M^+ = \sum_{j=1}^M S(j).$$

Si on rejette H_0 , on conclura que $\{Y_t\}$ cause $\{X_t\}$. Cependant, dans un contexte de robustesse, nous n'utiliserons pas la fonction de corrélation croisée usuelle, mais bien la version robustifiée. Une approche similaire se retrouve dans Li et Hui (1994). Il est alors naturel de contruire une statistique de test basée sur les corrélations croisées résiduelles robustes à des délais positifs. Ceci suggère comme statistique portmanteau la statistique

$$S_{RM}^+ = \sum_{j=1}^M S_R(j).$$

Cette statistique suit asymptotiquement une distribution χ_M^2 et on rejette H_0 pour de grandes valeurs de S_{RM}^+ .

Supposons maintenant que l'on s'intéresse à l'hypothèse alternative

$$H_{1M}^- : \rho_{uv}(j) \neq 0, \text{ pour au moins un } j \text{ tel que } -M \leq j \leq 1.$$

La statistique de test sera basée sur les corrélations croisées résiduelles robustes pour des délais négatifs et la statistique de test sera

$$S_{RM}^- = \sum_{j=-M}^{-1} S_R(j).$$

Dans le cas non-robuste, des approches similaires ont été considérées par Pierce (1977), Pierce et Haugh (1977), El Himdi et Roy (1997) pour identifier les directions de causalité entre deux séries chronologiques. Comme noté par El Himdi et Roy (1997), cette procédure ne constitue cependant pas un test statistique dans le sens usuel du terme, puisque l'hypothèse d'intérêt, l'absence de causalité, est exprimée sous l'hypothèse alternative, plutôt que sous l'hypothèse nulle. Dans un contexte de robustesse, les résultats de Li et Hui (1994) montrent bien que la distribution asymptotique fait intervenir une matrice de covariance asymptotique fort compliquée, difficile à utiliser en pratique. Li et Hui (1994) peuvent cependant l'exprimer plus simplement dans le cas où il y a seulement corrélation instantanée entre les deux séries.

4.4. PROCÉDURE BASÉE SUR UNE STATISTIQUE DE TEST PONDÉRÉE

Sous l'hypothèse nulle d'absence de corrélation, Li et Hui (1994) propose la statistique portmanteau suivante:

$$S_{RM} = \frac{n}{a} \sum_{j=-M}^M \gamma_{\hat{u}\hat{v}}^2(j; \eta). \quad (4.6)$$

Bien que non considéré par ces auteurs, afin que les niveaux soient mieux respectés, il semble préférable d'utiliser la statistique modifiée

$$S_{RM}^* = \frac{n}{a} \sum_{j=-M}^M \frac{n}{n-|j|} \gamma_{\hat{u}\hat{v}}^2(j; \eta). \quad (4.7)$$

L'étude de simulation de la section 4.5 illustre ce point. Cette procédure généralise au cas robuste la démarche de Haugh (1976). Hong (1996b) propose un test unilatéral dont la distribution asymptotique est $N(0, 1)$. Le but de cette section est de généraliser l'approche de Hong (1996b) à un contexte de robustesse. Plus précisément, nous énonçons dans la prochaine section les conditions pour lesquelles on établit la distribution asymptotique de la statistique suivante sous l'hypothèse nulle,

$$Q_{Rn} = \frac{na^{-1} \sum_{j=1-n}^{n-1} k^2(j/m) \gamma_{\hat{u}\hat{v}}^2(j; \eta) - M_n(k)}{[2V_n(k)]^{1/2}}, \quad (4.8)$$

où $M_n(k) = \sum_{j=1-n}^{n-1} (1-|j|/n) k^2(j/m)$ et $V_n(k) = \sum_{j=2-n}^{n-2} (1-|j|/n)(1-(|j|+1)/n) k^4(j/m)$. La constante m est le paramètre de lissage, et est telle que $m = m(n) \rightarrow \infty$. Cette statistique généralise Li et Hui (1994), qui peut être perçue comme le cas particulier avec le noyau uniforme tronqué. Elle généralise aussi la statistique de Hong (1996b) au cas robuste. On peut en pratique remplacer $M_n(k)$ et $V_n(k)$ par $mM(k)$ et $mV(k)$, où $M(k) = \int_{-\infty}^{\infty} k^2(z) dz < \infty$ et $V(k) = \int_{-\infty}^{\infty} k^4(z) dz < \infty$. On note la statistique obtenue

$$Q_{Rn}^* = \frac{na^{-1} \sum_{j=1-n}^{n-1} k^2(j/m) \gamma_{\hat{u}\hat{v}}^2(j; \eta) - mM(k)}{(2mV(k))^{1/2}}. \quad (4.9)$$

Elle possède sous certaines conditions la même distribution asymptotique que la statistique Q_{Rn} .

4.4.1. Distribution asymptotique sous l'hypothèse nulle

On suppose sous l'hypothèse nulle que $\{u_t\}$ et $\{v_t\}$ sont deux processus mutuellement indépendants. On obtient alors le théorème suivant:

Théorème 2. *Sous les hypothèses de la section 4.2.1 et sous les hypothèses A et B, $m \rightarrow \infty$, $m/n \rightarrow 0$, si $\{u_t\}$ est indépendant de $\{v_s\}$, alors $Q_{Rn} \rightarrow_L N(0, 1)$.*

La preuve détaillée du théorème 2 se trouve dans la section 4.7. Afin de démontrer ce résultat, nous adaptions la preuve de Hong (1996b) dans notre contexte. Tout comme Hong

(1996*b*), nous prouvons le résultat en deux étapes. Dans la première, nous établissons que $\tilde{Q}_{Rn} \rightarrow_L N(0, 1)$, où

$$\tilde{Q}_{Rn} = \frac{na^{-1} \sum_{j=1-n}^{n-1} k^2(j/m) \gamma_{uv}^2(j; \eta) - M_n(k)}{(2V_n(k))^{1/2}}$$

est définie comme Q_{Rn} , mais basée sur les processus innovations $\{u_t\}$ et $\{v_t\}$. Dans la seconde étape, nous établissons que $\tilde{Q}_{Rn} - Q_{Rn} = o_p(1)$. Ceci permet de conclure à la normalité asymptotique de la statistique sous l'hypothèse nulle.

4.5. ÉTUDE DE SIMULATION

Nous avons considéré dans la section 4.4 une statistique de test pondérée qui peut être considérée comme une version robustifiée de la statistique pondérée de Hong (1996*b*). Les nouveaux tests généralisent la statistique robuste considérée dans Li et Hui (1994), qui est en fait une version robuste du test de Haugh (1976). En effet, les statistiques robustes pondérées peuvent tenir compte de l'ensemble des délais, et le test de Li et Hui (1994) est essentiellement obtenu en choisissant le noyau uniforme tronqué.

L'étude empirique de cette section vise à vérifier l'impact de la présence de valeurs aberrantes sur le niveau et la puissance des nouveaux tests, que l'on compare au test de Li et Hui (1994). Notre façon de procéder consiste à générer des processus gaussiens et certaines observations sont systématiquement contaminées. Nous considérons aussi, pour fins de comparaison, la situation où des valeurs aberrantes ne sont pas ajoutées. Ceci permettra d'évaluer les performances de puissance quand on utilise une statistique robuste alors que le processus est gaussien. Les versions non robustes de Haugh (1976) et Hong (1996*b*) sont aussi incluses dans l'étude. En particulier, nous incluons les versions modifiées

$$S_M^* = n \sum_{j=-M}^M \frac{n}{(n-|j|)} r_{\hat{u}\hat{v}}^2(j), \quad (4.10)$$

ainsi que

$$Q_n^*(k) = \frac{n \sum_{j=1-n}^{n-1} k^2(j/m) r_{\hat{u}\hat{v}}^2(j) - mM(k)}{[2mV(k)]^{1/2}}, \quad (4.11)$$

4.5.1. Choix des modèles

Afin d'étudier la performance des diverses statistiques, nous avons simulé des réalisations suivant le processus gaussien bivarié VAR(1) suivant:

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 0.5 & 0.0 \\ 0.0 & 0.5 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} u_t \\ v_t \end{pmatrix}, t \in \mathbb{Z}.$$

La matrice des variances et covariances de $(u_t, v_t)'$ est

$$\Sigma = \begin{pmatrix} 1.0 & \rho \\ \rho & 1.0 \end{pmatrix}.$$

Nous avons considéré les deux valeurs $\rho = 0.0$ et $\rho = 0.2$. Le premier cas correspond au cas où deux processus AR(1) indépendants sont générés, et cette situation nous permet d'évaluer le niveau des tests. Le second cas nous situe dans un contexte où il y a dépendance entre les séries. Ainsi, nous pourrions évaluer la puissance des tests dans cette situation. Cette façon de créer de la dépendance entre deux séries se retrouve par exemple dans El Himdi et Roy (1997).

Pour l'étude de niveau, 10000 réalisations de longueur $n = 100$ et $n = 200$ ont été générées. Les écarts type (en pourcentage) correspondant à 10000 réalisations pour les trois niveaux nominaux 1%, 5%, 10% sont respectivement 0.099, 0.218 et 0.300. Les zones de non-rejet de l'hypothèse nulle ' H_0 : niveau est α (en pourcentage)'' sont données dans le tableau 4.1, pour les niveaux 5% (notée NR5) et 1% (notée NR1).

TABLEAU 4.1. Zones de non-rejet de ' H_0 : niveau est α (en pourcentage)'' basées sur 10000 réalisations.

α	NR5	NR1
1	[81, 119]	[75, 125]
5	[458, 542]	[444, 556]
10	[942, 1058]	[923, 1077]

Pour l'étude de puissance, 10000 réalisations ont été considérées pour les deux mêmes longueurs de série. Afin de s'assurer de la même structure de corrélation pour la valeur initiale $(X_0, Y_0)'$ et du reste des observations, $(X_0, Y_0)'$ est générée selon une distribution normale centrée avec une matrice des variances et covariances obtenue avec l'algorithme de Ansley (1980).

4.5.2. Emplacement des valeurs aberrantes

Afin d'évaluer l'effet des valeurs aberrantes sur les diverses statistiques, nous avons considéré les trois scénarios suivants:

Scénario 1: On procède à aucune contamination.

Scénario 2: On soustrait de façon systématique la valeur 10 aux observations X_{26} et Y_{76} .

Pour $n = 200$, on ajoute en plus la valeur 10 à X_{101} et Y_{151} , et on soustrait 10 à X_{176} .

Scénario 3: On fait comme dans le scénario 2. Cependant, en plus, on ajoute à X_{51} et Y_{51} la valeur de 10. Pour $n = 200$, on soustrait à X_{126} et Y_{126} la valeur 10.

Des scénarios similaires se retrouvent dans Li (1988) et Li et Hui (1994). Le premier scénario permet d'évaluer la performance des diverses statistiques dans le contexte d'un processus gaussien VAR(1). Le second scénario permet d'évaluer la performance des procédures quand les valeurs aberrantes ne surviennent pas en même temps dans chacune des séries. Finalement, dans le troisième scénario, quelques valeurs aberrantes surviennent aux mêmes instants dans le temps.

4.5.3. Statistiques considérées

Les statistiques pondérées robustes sont dénotées $Q_{Rn}(k)$ et $Q_{Rn}^*(k)$, où k est un noyau. Nous considérons le noyau spectral quadratique (QS), le noyau de Bartlett (BAR), de Daniell (DAN), de Parzen (PAR) et finalement le noyau uniforme tronqué (TR). Les valeurs de m utilisées sont $m = 5, 8, 12$ pour $n = 100$ et $m = 5, 9, 15$ pour $n = 200$. Ce sont les valeurs utilisées par Hong (1996b), qui correspondent aux taux $\lfloor \log(n) \rfloor$, $\lfloor 3n^{0.2} \rfloor$ et $\lfloor 3n^{0.3} \rfloor$, où $\lfloor a \rfloor$ indique la partie entière de a . Nous avons choisi une fonction η de type Mallows, et les fonctions ψ considérées sont celles appartenant à la famille bicarré et à la famille de Huber. La constante d'ajustement pour le cas bicarré est 5.58 alors que pour le cas Huber la valeur est 1.65. Ces valeurs proviennent de Li et Hui (1994) et de Bustos et Yohai (1986). Les statistiques robustes de Li et Hui (1994) sont notées S_{RM} et S_{RM}^* . Les statistiques pondérées de Hong (1996b) sont notées $Q_n(k)$ et $Q_n(k)^*$, alors que les statistiques de Haugh (1976) sont notées S_M et S_M^* . Pour les statistiques de Haugh (1976) et de Li et Hui (1994), nous devons déterminer le nombre de délais M . Nous avons considéré $M = 6, 12, 18$.

Les méthodes d'estimation sont les méthodes des moindres carrés pour les statistiques usuelles non robustes et les tests robustes sont basés sur les estimateurs RA . Pour la comparaison des méthodes dans cette étude, nous avons considéré le cadre quelque peu idéalisé où nous ajustons les bons modèles AR(1) à chaque série. Pour les estimateurs RA avec fonction

η de type Mallows nous avons utilisé un algorithme itératif qui est décrit dans la section 2.2 de Bustos et Yohai (1986).

Toute la programmation a été faite en FORTRAN avec la librairie NAG.

4.5.4. Discussion des résultats sur l'étude de niveau

Chaque tableau contient les résultats pour $n = 100$ et $n = 200$ pour diverses valeurs de m ou de M .

On débute par discuter l'étude portant sur les niveaux, correspondant au cas où $\rho = 0$. Chaque groupe de trois lignes contient les résultats pour les trois scénarios. Au risque de fournir certaines répétitions dans l'analyse des résultats, nous procédons à l'analyse de chacun des tableaux individuellement, et nous allons dégager par la suite des remarques générales.

Les résultats du tableau 4.2 correspondent à une étude de la statistique Q_n . Quand il n'y a pas de valeurs aberrantes, les résultats montrent que les niveaux sont mieux contrôlés pour m grand. Aussi, des taux plus élevés donnent de meilleurs résultats. Cependant, pour les niveaux nominaux 1% et 5%, les niveaux observés sont à l'extérieur des zones NR1. Nous obtenons de meilleurs résultats avec le niveau nominal 10%, avec des niveaux observés dans ou très près des zones NR1. En présence de valeurs aberrantes, la statistique Q_n est inutilisable, avec des sous-rejets (scénario 2) et des sur-rejets (scénario 3) marqués.

Les résultats du tableau 4.3 correspondent à une étude de la statistique Q_n^* . Quand il n'y a pas de valeurs aberrantes, les résultats montrent que les niveaux sont mieux approximés que la statistique Q_n , particulièrement aux niveaux nominaux 1% et 5%. On note un comportement un peu différent du noyau de Bartlett. Pour le niveau nominal 1%, les niveaux sont à l'extérieur des zones NR1, mais les noyaux de Daniell et tronqué font raisonnablement bien. Le niveau nominal 5% est beaucoup mieux approximé avec Q_n^* comparativement à Q_n , avec des résultats dans les zones NR5 et NR1, pour m petit ou modéré. Le niveau nominal 10% est bien contrôlé, avec des niveaux observés à l'intérieur des zones NR5 et NR1 pour de petites valeurs de m . Cependant, en présence de valeurs aberrantes, la statistique Q_n^* est inutilisable, avec des sous-rejets (scénario 2) et des sur-rejets (scénario 3) marqués.

Les résultats du tableau 4.4 correspondent à la statistique de Haugh. Il est bien connu que les niveaux de S_n^* sont mieux contrôlés que ceux de S_n . Voir Haugh (1976). Ainsi, les niveaux pour S_n se détériorent quand M est grand, alors que S_n^* respecte de manière générale très bien les trois niveaux, avec des résultats dans les zones NR5. Cependant, en présence de valeurs aberrantes, ces statistiques sont inutilisables.

Les résultats du tableau 4.5 portent sur la statistique robuste Q_{Rn} en utilisant la fonction de Huber. Lorsqu'il n'y a pas de valeurs aberrantes, on constate que les niveaux sont mieux contrôlés pour m grands et un taux plus élevé donne de meilleurs résultats. Les niveaux observés sont à l'extérieur des zones NR1, pour les niveaux nominaux 1% et 5%. De meilleurs résultats sont obtenus au niveau 10%, avec des résultats dans les zones NR1 ou très près de ces dernières. En présence de valeurs aberrantes, les statistiques Q_{Rn} sont utilisables, avec des résultats sur les niveaux plus satisfaisants que ceux pour Q_n . Ainsi, dans le scénario 2, les résultats s'apparentent aux résultats obtenus sans valeurs aberrantes. On obtient de meilleurs résultats au niveau 10%. Dans le scénario 3, quoique les résultats sont plus satisfaisants que ceux obtenus avec la statistique Q_n , on observe des sur-rejets, avec des niveaux observés nettement à l'extérieur des zones NR1.

Dans le tableau 4.6, on considère la statistique robuste Q_{Rn}^* avec la fonction de Huber. Quand il n'y a pas de valeurs aberrantes, on observe que les niveaux sont mieux contrôlés que pour Q_{Rn} avec la fonction de Huber aux niveaux nominaux 1% et 5%. On note un comportement assez différent du noyau de Bartlett. Pour m grand, les nombres de rejets observés sont très près des zones NR1 aux niveaux 1% et 5%. On obtient des niveaux observés dans les zones NR5 ou NR1 pour m moyen ou grand, particulièrement pour les noyaux QS, DAN et PAR. Concernant les résultats dans le scénario 2, les résultats s'apparentent aux résultats obtenus sans valeurs aberrantes. Dans le scénario 3, quoique les résultats sont plus satisfaisants que ceux obtenus avec la statistique Q_n^* , on observe des sur-rejets, avec des niveaux observés nettement à l'extérieur des zones NR1.

On discute maintenant les résultats du tableau 4.7 portant sur les statistiques robustes de Li et Hui (1994) avec la fonction de Huber. Tout comme dans le cas non-robuste, les versions S_{Rn}^* se comportent mieux que S_{Rn} . Les niveaux sont bien approximés dans les scénarios 1 et 2, mais on observe un léger sur-rejet dans le scénario 3.

Dans le tableau 4.8, on considère la statistique robuste Q_{Rn} avec la fonction bicarré. Lorsque l'on regarde les résultats pour le premier scénario, on note que les niveaux sont mieux contrôlés quand m est grand, et que des taux plus élevés donnent de meilleurs résultats. Pour les niveaux nominaux 1% et 5%, les résultats sont à l'extérieur des zones NR1. Les meilleurs résultats sont obtenus pour le niveau nominal 10%, qui sont dans les zones NR1 ou tout près. Pour le scénario 2, les résultats s'apparentent aux résultats obtenus sans valeurs aberrantes. Pour le scénario 3, on note que l'on obtient de meilleurs résultats en utilisant la fonction

bicarré plutôt que la fonction de Huber, puisque le niveau est mieux contrôlé. Ainsi, les niveaux observés sont similaires au cas où il n'y a pas de valeurs aberrantes.

Les résultats du tableau 4.9 portent sur la statistique robuste Q_{Rn}^* avec la fonction bicarré. De manière générale, lorsque l'on compare avec Q_{Rn} avec fonction bicarré, on peut affirmer que les niveaux semblent mieux contrôlés avec Q_{Rn}^* et fonction bicarré. Lorsqu'il n'y a pas de valeurs aberrantes, on note un sur-rejet pour le niveau 1%. Cependant, les niveaux avec noyau Daniell et tronqué respectent raisonnablement les niveaux avec des résultats près des zones NR1. On note ici aussi un comportement différent du noyau de Bartlett. Le niveau nominal 5% est bien contrôlé pour m modéré et grand avec des résultats dans les zones NR5 et NR1. Le niveau nominal 10% est bien respecté en général. On note cependant qu'il semble que le noyau de Daniell demande une valeur de m plus petite, alors que les noyaux QS, DAN et PAR nécessitent des valeurs de m modérés. Les résultats sont cependant la plupart du temps dans les zones NR5 ou NR1. Pour le scénario 2, les résultats s'apparentent aux résultats obtenus sans valeur aberrante. Les niveaux semblent aussi bien contrôlés que dans le scénario 1. Concernant le scénario 3, contrairement à la fonction de Huber, nous obtenons des résultats qui s'apparentent aux scénario 1, et les niveaux sont en général aussi bien contrôlés que dans le scénario 1.

Dans le tableau 4.10, on considère les statistiques robustes de Li et Hui (1994) avec la fonction bicarré. Tout comme dans le cas non-robuste, les versions S_{Rn}^* se comportent mieux que S_{Rn} . Les niveaux sont bien approximés pour tous les scénarios. Li et Hui (1994) avaient obtenus de meilleurs résultats avec la fonction bicarré, et c'est à nouveau ce qui se produit dans notre étude.

Nous formulons maintenant des commentaires généraux sur l'étude de niveau. Quand il n'y a pas de valeurs aberrantes, il semble que les grandes valeurs de m fournissent de façon générale de meilleurs niveaux. Pour la statistique Q_n , le niveau nominal 10% est raisonnablement bien approximé, et les niveaux observés sont habituellement dans les zones de non-rejet NR1, où très près de ces dernières. On note une nette tendance à un sur-rejet pour les niveaux 1% et 5%, et les niveaux observés sont systématiquement à l'extérieur des zones NR1. Lorsque l'on regarde les résultats lorsque la statistique Q_n^* est considérée, le niveau 5% est beaucoup mieux approximé. Par exemple, pour $m = 12$, tous les niveaux observés sont dans les zones NR5 pour les noyaux QS et PAR. Pour le noyau DAN, il semble préférable selon notre étude de choisir une valeur de m moins grande. Dans l'étude de Hong (1996b), ce dernier note que Q_n et Q_n^* se comportent de manière similaire. Dans notre situation, il semble se dégager

que Q_n est plus approprié lorsque le niveau nominal est 10%, alors que Q_n^* semble mieux se comporter lorsque le niveau nominal est de 5%. Il est peut-être plus satisfaisant d'avoir des statistiques de test avec un bon comportement au niveau nominal 5%, qui est largement utilisé en pratique. De ce point de vue, Q_n^* semble préférable selon notre étude.

Avec la statistique de Hong, lorsque les valeurs aberrantes ne surviennent pas en même temps (scénario 2), un net sous-rejet est observé. Lorsque les valeurs aberrantes surviennent en même temps et que le noyau est autre que le noyau uniforme tronqué, on note un sur-rejet. Pour les statistiques de Haugh et de Hong avec le noyau uniforme tronqué, pour de grandes valeurs de m ou de M , le test a tendance à ne pas rejeter suffisamment (surtout perceptible pour $n = 100$). Ainsi, la présence de valeurs aberrantes peut perturber de façon notable les niveaux des tests usuels. Dans notre situation, avec le type de valeurs aberrantes considérées, les tests de Haugh et Hong usuels sont inutilisables, dans la mesure où les niveaux nominaux ne sont pas bien approximés.

On note que dans le contexte du scénario 1, les tests robustes avec les fonctions ψ de Huber ou bicarré se comportent de manière comparable au test de Hong. Les tests robustes de Li et Hui se comportent comme le test de Haugh. Cependant, pour les scénarios 2 et 3, la situation est différente quand on compare avec les statistiques de Haugh et de Hong. Pour le scénario 2, avec la fonction ψ de Huber ou bicarré, les niveaux se comparent aux niveaux du scénario 1. Les différences entre les fonctions ψ de Huber ou bicarré sont perceptibles pour le scénario 3. En effet, lorsque la fonction ψ de Huber est utilisée, on note une tendance à un sur-rejet, qui est perceptible aussi pour les statistiques de Li et Hui. Les statistiques robustes avec fonction bicarré ont un meilleur comportement.

4.5.5. Tableaux pour l'étude de niveau

TABLEAU 4.2. *Nombres de rejets (sous l'hypothèse nulle d'indépendance) basés sur 10000 réalisations en utilisant la fonction ψ donnant les fonctions de corrélation usuelles avec les estimateurs des moindres carrés et les statistiques de Hong Q_n données par (4.2). Pour chaque valeur de m , les trois rangées correspondent aux trois scénarios.*

	$\alpha = 0.01$					$\alpha = 0.05$					$\alpha = 0.10$				
$n = 100$															
m	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
5	322	332	322	302	219	722	724	728	716	637	1098	1100	1090	1091	1085
	39	51	36	34	12	134	164	134	128	82	263	318	251	238	176
	7625	8727	7564	7208	2051	8825	9446	8779	8581	4119	9303	9677	9281	9147	5409
8	254	284	258	244	197	679	701	679	671	625	1069	1092	1083	1064	1092
	25	33	23	19	6	95	103	99	84	42	185	215	182	168	98
	5298	7337	5270	4753	635	7207	8646	7161	6730	1820	8104	9156	8081	7715	2827
12	216	247	220	210	188	634	649	633	640	631	1081	1079	1072	1079	1104
	9	18	10	7	3	49	66	49	47	25	126	142	119	116	67
	2900	5374	2851	2379	148	4982	7167	4921	4380	563	6179	8071	6121	5606	1162
$n = 200$															
m	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
5	304	324	303	297	270	718	736	718	705	681	1076	1075	1083	1067	1074
	28	37	26	24	7	112	140	106	100	41	205	239	197	189	101
	9892	9979	9891	9843	7020	9982	9996	9980	9975	8705	9994	9999	9993	9991	9307
9	294	306	289	279	215	667	685	674	678	649	1065	1077	1065	1062	1098
	14	19	16	12	4	48	67	62	38	18	126	153	143	107	59
	9118	9770	9291	8814	2799	9707	9936	9781	9580	5223	9859	9983	9898	9789	6623
15	243	266	243	226	190	648	664	651	647	629	1104	1067	1095	1097	1050
	7	10	13	3	1	26	33	48	20	11	59	72	121	48	24
	6717	8803	8205	5779	525	8470	9548	9354	7771	1594	9088	9769	9677	8648	2579

TABLEAU 4.3. *Nombres de rejets (sous l'hypothèse nulle d'indépendance) basés sur 10000 réalisations en utilisant la fonction ψ donnant les fonctions de corrélation usuelles avec les estimateurs des moindres carrés et les statistiques de Hong Q_n^* données par (4.11). Pour chaque valeur de m , les trois rangées correspondent aux trois scénarios.*

		$\alpha = 0.01$					$\alpha = 0.05$					$\alpha = 0.10$					
$n = 100$	$m = 5$	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	
		292	354	279	274	288	680	753	657	671	835	1039	1161	1003	1030	1374	
		32	57	30	28	23	128	175	119	110	121	246	335	220	216	236	
		7479	8813	7321	7050	2529	8736	9480	8632	8501	4754	9272	9692	9185	9064	6095	
	$m = 8$	220	271	192	206	201	608	673	555	571	645	967	1041	907	939	1127	
		15	33	15	14	6	81	94	70	70	42	162	205	137	137	106	
		5001	7264	4796	4428	646	6952	8581	6722	6403	1876	7922	9108	7717	7473	2926	
	$m = 12$	160	207	135	150	118	507	575	432	480	473	884	946	753	871	892	
		6	15	5	4	2	39	57	31	34	17	95	124	70	74	46	
		2461	5091	2175	1950	89	4530	6902	4073	3774	407	5718	7843	5303	5066	872	
	$n = 200$	$m = 5$	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
			295	366	286	279	372	695	781	669	682	899	1048	1142	1038	1033	1428
24			43	24	23	18	105	156	95	94	73	196	262	181	180	174	
		9888	9983	9871	9831	7735	9981	9997	9977	9968	9101	9994	9999	9991	9991	9552	
$m = 9$		269	300	249	247	239	627	678	590	611	737	1011	1054	958	1007	1249	
		13	19	14	12	4	41	62	45	36	28	114	146	122	94	73	
		9033	9769	9175	8723	3098	9680	9936	9740	9536	5581	9843	9982	9870	9770	6961	
$m = 15$		194	242	173	183	157	572	614	513	556	546	966	991	848	952	966	
		6	9	8	3	1	22	28	31	17	10	48	64	74	37	22	
		6371	8704	7684	5391	433	8233	9498	9105	7467	1437	8966	9736	9559	8422	2407	

TABLEAU 4.4. *Nombres de rejets (sous l'hypothèse nulle d'indépendance) basés sur 10000 réalisations en utilisant la fonction ψ donnant les fonctions de corrélation usuelles avec les estimateurs des moindres carrés et les statistiques de Haugh S_M et S_M^* données par (4.1) et (4.10). Pour chaque valeur de M , les trois rangées correspondent aux trois scénarios.*

	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.10$	
$n = 100$						
M	S_M	S_M^*	S_M	S_M^*	S_M	S_M^*
6	51	79	361	460	782	944
	1	2	22	36	9	124
	448	507	2118	2266	3722	3892
12	25	85	263	488	617	1021
	0	2	7	19	25	61
	16	28	212	309	549	807
18	21	81	152	469	392	993
	0	0	3	10	9	41
	0	4	18	58	76	221
$n = 200$						
M	S_M	S_M^*	S_M	S_M^*	S_M	S_M^*
6	79	90	470	509	937	1015
	3	4	18	21	66	73
	3676	3748	7141	7211	8525	8591
12	65	92	404	532	814	1005
	0	0	5	9	20	36
	427	504	1984	2186	3543	3794
18	45	82	283	467	638	973
	0	0	3	8	13	22
	50	73	376	514	924	1170

TABLEAU 4.5. *Nombres de rejets (sous l'hypothèse nulle d'indépendance) basés sur 10000 réalisations en utilisant la fonction ψ de type Huber avec les estimateurs RA et les statistiques robustes Q_{Rn} données par (4.8). Pour chaque valeur de m , les trois rangées correspondent aux trois scénarios.*

		$\alpha = 0.01$					$\alpha = 0.05$					$\alpha = 0.10$				
$n = 100$																
m		QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
5		328	340	327	310	223	730	734	737	710	647	1112	1106	1110	1102	1083
		298	317	300	284	236	693	693	694	676	640	1071	1094	1069	1069	1075
		441	530	440	414	271	994	1061	1001	988	752	1458	1499	1461	1423	1251
8		259	296	263	248	209	669	704	671	660	626	1108	1099	1105	1101	1088
		251	277	249	241	208	687	679	680	676	612	1047	1067	1059	1049	1104
		351	425	351	335	239	883	986	890	857	723	1343	1423	1358	1312	1214
12		215	236	220	208	189	636	649	632	626	640	1096	1092	1092	1095	1101
		238	246	242	222	210	652	669	646	650	656	1086	1064	1077	1080	1085
		290	351	289	283	231	798	865	798	778	713	1290	1377	1298	1288	1220
$n = 200$																
m		QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
5		309	331	305	301	261	702	719	704	696	663	1086	1062	1097	1073	1066
		269	289	267	260	217	659	664	663	645	637	1044	1019	1051	1053	1039
		752	884	767	702	399	1423	1593	1423	1392	969	2002	2129	2009	1967	1504
9		286	305	285	281	208	675	687	666	681	624	1078	1076	1078	1067	1104
		239	244	243	238	207	635	644	640	641	620	1061	1033	1049	1061	1051
		529	657	536	495	307	1188	1365	1202	1140	814	1757	1940	1791	1730	1376
15		233	257	238	231	195	662	656	672	650	593	1080	1069	1088	1088	1060
		225	229	228	226	194	624	649	642	614	618	1039	1025	1047	1040	1063
		395	504	428	378	262	1000	1166	1057	966	756	1558	1724	1625	1529	1267

TABLEAU 4.6. *Nombres de rejets (sous l'hypothèse nulle d'indépendance) basés sur 10000 réalisations en utilisant la fonction ψ de type Huber avec les estimateurs RA et les statistiques robustes Q_{Rn}^* données par (4.9). Pour chaque valeur de m , les trois rangées correspondent aux trois scénarios.*

		$\alpha = 0.01$					$\alpha = 0.05$					$\alpha = 0.10$				
$n = 100$																
m		QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
5		307	357	286	287	301	680	776	657	652	842	1062	1141	1013	1042	1401
		275	358	268	258	311	655	737	625	637	828	1017	1145	978	1006	1339
		404	571	389	382	374	939	1101	894	918	988	1398	1551	1355	1352	1555
8		226	285	210	204	212	591	670	542	576	651	982	1054	893	963	1137
		216	261	198	204	216	599	650	552	588	636	929	1026	866	929	1152
		302	409	275	277	245	795	960	715	750	750	1230	1376	1142	1197	1271
12		171	198	140	157	136	502	566	425	483	470	882	971	757	846	891
		175	204	139	162	135	532	586	427	507	500	874	939	757	848	873
		229	298	190	205	166	629	767	539	606	535	1066	1235	933	1022	1006
$n = 200$																
m		QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
5		295	359	288	289	377	682	764	666	672	898	1062	1145	1034	1039	1442
		262	326	252	249	332	643	714	627	632	865	1023	1080	994	1018	1400
		728	940	733	678	569	1387	1673	1365	1353	1307	1955	2234	1924	1919	1954
9		266	303	246	250	234	624	677	585	625	709	1021	1065	956	994	1241
		219	241	212	215	245	600	636	564	581	696	1001	1026	935	969	1200
		493	649	477	460	374	1116	1348	1088	1073	927	1672	1916	1615	1605	1568
15		201	228	178	186	171	559	602	501	556	525	972	982	861	952	974
		189	202	173	186	167	533	578	482	519	557	903	951	821	894	967
		336	464	318	314	220	875	1088	827	839	687	1392	1622	1335	1341	1172

TABLEAU 4.7. *Nombres de rejets (sous l'hypothèse nulle d'indépendance) basés sur 10000 réalisations en utilisant la fonction ψ de type Huber avec les estimateurs RA et les statistiques de Li et Hui S_{RM} et S_{RM}^* données par (4.6) et (4.7). Pour chaque valeur de M , les trois rangées correspondent aux trois scénarios.*

		$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.10$	
$n = 100$							
M	S_{Rn}	S_{Rn}^*	S_{Rn}	S_{Rn}^*	S_{Rn}	S_{Rn}^*	
6	59	84	363	469	796	942	
	67	89	377	463	806	972	
	90	110	444	565	924	1093	
12	32	80	252	487	617	1016	
	46	94	280	510	647	1004	
	48	114	292	544	700	1128	
18	21	80	144	478	388	990	
	22	104	166	534	426	1057	
	29	108	198	576	466	1121	
$n = 200$							
M	S_{Rn}	S_{Rn}^*	S_{Rn}	S_{Rn}^*	S_{Rn}	S_{Rn}^*	
6	85	102	465	505	926	1001	
	81	92	433	480	872	953	
	133	159	655	711	1273	1352	
12	75	104	383	499	824	1028	
	75	105	367	496	823	1002	
	92	117	488	619	983	1226	
18	55	106	282	478	637	976	
	54	107	293	494	660	1030	
	73	138	405	622	816	1169	

TABLEAU 4.8. *Nombres de rejets (sous l'hypothèse nulle d'indépendance) basés sur 10000 réalisations en utilisant la fonction ψ de type bicarré avec les estimateurs RA et les statistiques robustes Q_{Rn} données par (4.8). Pour chaque valeur de m , les trois rangées correspondent aux trois scénarios.*

		$\alpha = 0.01$					$\alpha = 0.05$					$\alpha = 0.10$					
$n = 100$	$m = 5$	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	
		318	337	319	309	232	735	735	737	715	644	1119	1080	1105	1112	1081	
		314	326	318	306	234	727	708	731	727	665	1111	1116	1114	1100	1102	
		320	344	322	312	231	708	708	712	709	660	1080	1077	1086	1084	1086	
	$m = 8$	259	294	259	252	208	670	704	674	651	626	1092	1096	1084	1090	1094	
		266	280	270	260	208	687	710	685	683	639	1081	1083	1082	1081	1099	
		267	288	269	249	213	665	693	662	650	625	1061	1078	1060	1072	1051	
	$m = 12$	221	238	225	208	195	628	647	632	626	628	1091	1082	1090	1108	1085	
		222	253	223	221	206	649	675	651	653	612	1099	1092	1099	1100	1087	
		226	243	225	216	203	628	654	616	619	624	1071	1089	1057	1088	1086	
	$n = 200$	$m = 5$	316	337	316	316	257	708	726	715	706	664	1083	1063	1086	1082	1068
			298	334	293	288	254	709	710	713	694	655	1062	1061	1061	1062	1063
289			313	290	278	238	709	720	720	706	668	1105	1096	1111	1115	1088	
$m = 9$		282	308	282	275	206	678	684	665	664	634	1070	1077	1073	1076	1112	
		261	261	267	252	201	654	678	665	643	616	1045	1071	1060	1060	1057	
		252	269	255	243	203	688	711	691	671	637	1097	1112	1086	1097	1063	
$m = 15$		231	257	234	222	188	665	668	668	661	598	1089	1066	1088	1098	1063	
		226	236	221	222	206	649	647	647	644	616	1069	1072	1078	1067	1068	
		224	220	221	217	199	652	660	652	658	618	1087	1098	1093	1088	1042	

TABLEAU 4.9. *Nombres de rejets (sous l'hypothèse nulle d'indépendance) basés sur 10000 réalisations en utilisant la fonction ψ de type bicarré avec les estimateurs RA et les statistiques robustes Q_{Rn}^* données par (4.9). Pour chaque valeur de m , les trois rangées correspondent aux trois scénarios.*

		$\alpha = 0.01$					$\alpha = 0.05$					$\alpha = 0.10$					
$n = 100$	$m = 5$	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	
		297	351	284	287	311	692	769	664	668	828	1056	1138	1014	1037	1376	
		286	345	272	274	312	677	748	649	666	850	1059	1154	1008	1033	1363	
		300	372	284	282	314	681	740	651	661	822	1028	1117	996	1016	1372	
	$m = 8$	223	281	208	212	212	594	674	544	575	647	973	1062	896	956	1135	
		222	270	202	201	211	603	670	556	592	658	989	1047	904	958	1147	
		233	279	208	206	217	591	661	545	564	654	955	1040	890	946	1096	
	$m = 12$	170	205	143	164	141	499	580	438	483	477	888	953	747	851	895	
		172	199	140	162	151	521	590	437	501	461	908	965	766	870	869	
		164	208	145	153	134	509	565	436	483	474	869	960	745	846	855	
	$n = 200$	$m = 5$	307	372	295	302	371	688	779	687	682	901	1054	1142	1030	1052	1422
			280	363	272	274	373	688	762	677	663	900	1025	1123	994	1027	1404
281			356	274	268	362	695	786	678	681	933	1077	1151	1045	1075	1446	
$m = 9$		259	305	245	250	239	621	676	582	619	732	1001	1057	945	999	1265	
		244	260	224	228	236	614	671	575	592	715	988	1056	939	992	1222	
		233	267	220	218	246	641	701	609	623	717	1029	1100	970	1011	1222	
$m = 15$		203	230	185	189	168	572	607	489	556	531	966	990	854	944	970	
		194	215	173	180	170	560	589	491	543	554	947	990	836	921	977	
		183	208	163	176	176	547	595	478	537	556	976	1014	850	953	956	

TABLEAU 4.10. *Nombres de rejets (sous l'hypothèse nulle d'indépendance) basés sur 10000 réalisations en utilisant la fonction ψ de type bicarré avec les estimateurs RA et les statistiques de Li et Hui S_{RM} et S_{RM}^* données par les formules (4.6) et (4.7). Pour chaque valeur de M , les trois rangées correspondent aux trois scénarios.*

		$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$		
<i>n = 100</i>						
<i>M</i>	S_{Rn}	S_{Rn}^*	S_{Rn}	S_{Rn}^*	S_{Rn}	S_{Rn}^*
6	57	79	373	478	784	938
	58	77	385	462	796	962
	58	83	379	457	801	972
12	31	78	257	496	611	1010
	45	92	273	471	594	1006
	42	92	264	484	611	998
18	20	78	148	484	391	991
	24	85	157	476	387	993
	20	84	157	471	375	927
<i>n = 200</i>						
<i>M</i>	S_{Rn}	S_{Rn}^*	S_{Rn}	S_{Rn}^*	S_{Rn}	S_{Rn}^*
6	78	104	457	515	911	995
	77	90	455	509	912	1006
	79	97	447	495	944	1023
12	77	105	386	504	823	1030
	73	105	384	500	806	983
	67	87	365	499	790	1002
18	50	97	272	472	644	966
	46	91	309	481	644	1003
	42	100	290	496	654	978

4.5.6. Discussion des résultats sur l'étude de puissance

On retrouve les résultats pour chacun des trois scénarios. Les résultats sont basés sur les quantiles asymptotiques sauf ceux entre parenthèses dans le scénario 1 qui représentent les résultats basés sur les quantiles empiriques obtenus de l'étude sur le niveau.

Les tableaux 4.11 et 4.12 constituent une étude de puissance des tests de Hong. Les résultats de puissance ont un intérêt lorsque le niveau est raisonnablement bien contrôlé. Ainsi, les résultats obtenus dans les scénarios 2 et 3 sont plutôt d'un intérêt limité ici, puisque le test de Hong est durement affecté par les valeurs aberrantes considérées. Pour le scénario 1, on note un comportement similaire des statistiques Q_n et Q_n^* , quoique pour les noyaux QS, DAN et PAR la puissance semble être supérieure avec Q_n , alors que pour BAR c'est l'effet contraire pour m petit. Les noyaux autres que le noyau uniforme tronqué se comportent de manière similaire. Le noyau uniforme tronqué est celui fournissant la puissance la plus faible. Ainsi, un choix autre que le noyau uniforme tronqué permet des gains substantiels en terme de puissance. Le noyau de Bartlett semble plus puissant que les autres, ce qui n'entre pas en contradiction avec le résultat d'optimalité du noyau de Daniell, puisque le noyau BAR ne remplit pas les conditions de ce théorème. On note que m petit donne de meilleures propriétés de puissance. Aussi, les quantiles asymptotiques sont raisonnablement en accord avec les quantiles empiriques pour Q_n au niveau 10% et pour Q_n^* au niveau 5%, pour m modéré et grand. On constate à la lumière du tableau 4.13 que Q_n et Q_n^* sont plus puissants que S_n^* , lorsque l'on choisit un noyau autre que le noyau uniforme tronqué. Les résultats des scénarios 2 et 3 sont aussi de peu d'intérêt pour les tests de Haugh, puisque les niveaux et les puissances sont perturbés par la présence des valeurs aberrantes.

On étudie maintenant les tableaux 4.14, 4.15 et 4.16, qui contiennent les résultats de puissance avec les nouveaux tests et les tests de Li et Hui, utilisant la fonction ψ de Huber. On avait constaté que les niveaux semblaient mieux contrôlés pour les scénarios 1 et 2, alors qu'un sur-rejet semblait présent, en ce qui avait trait au scénario 3. On constate que les grandes puissances pour le scénario 3 sont occasionnées par les valeurs aberrantes survenant simultanément. Pour ce qui est du scénario 1, on constate que la puissance de Q_{Rn} est plus faible que celle de Q_n , et de même la puissance de Q_{Rn}^* est plus faible que celle de Q_n . C'est le prix à payer pour l'utilisation d'une statistique robuste lorsqu'il n'y a pas de valeurs aberrantes additives. On note cependant que ce prix ne semble pas trop élevé. On a vu que les niveaux étaient raisonnablement bien respectés pour le scénario 2. La présence de valeurs aberrantes semble diminuer quelque peu la puissance des tests. Cependant, à la lumière des résultats du

tableau 4.16, les versions robustes du test de Hong demeurent plus puissantes pour le scénario 2 que les tests de Li et Hui. Les statistiques robustes de Li et Hui avec fonction de Huber demeurent inutilisables pour le scénario 3, puisque les niveaux étaient mal approximés.

Les tableaux 4.17, 4.18 et 4.19 étudient la puissance mais avec la fonction ψ bicarré. Cette situation est particulièrement intéressante puisque les niveaux pouvaient être bien approximés pour les trois scénarios considérés. Tout comme dans le cas de la fonction de Huber, lorsque l'on regarde les résultats obtenus dans le scénario 1, on constate qu'il y a une perte de puissance en utilisant les tests robustes pondérés comparativement à l'utilisation des statistiques pondérées de Hong. Cependant, ce prix ne semble pas trop élevé. De même, les tests robustes de Li et Hui sont moins puissants que la statistique de Haugh. Pour m petit et modéré, il semble que la fonction de Huber donne de meilleurs résultats de puissance, alors que pour m grand il semble que la fonction bicarré fasse mieux. Les puissances observées pour les scénarios 2 et 3 sont assez similaires, tout en étant légèrement inférieures aux résultats obtenus dans le scénario 1. Lorsque l'on regarde le tableau 4.19, on peut conclure que les nouveaux tests semblent offrir une puissance avantageuse par rapport aux tests robustes de Li et Hui.

En conclusion à cette étude empirique, il semble que les meilleurs résultats ont été obtenus avec la fonction bicarré. Il semble que la présence de valeurs aberrantes occasionne une perte de puissance, mais nous obtenons des résultats beaucoup plus satisfaisants, comparativement à l'utilisation de la fonction d'autocovariance usuelle et de l'estimateur des moindres carrés. Dans le scénario 3, nous observons une puissance trop grande avec la fonction ψ de Huber, mais qui est possiblement relié au fait que les taux de rejets observés étaient trop grands par rapport aux niveaux nominaux. Les résultats avec la fonction ψ bicarré sont beaucoup plus satisfaisants, puisque la puissance se comporte de façon plus comparable au scénario 1.

4.5.7. Tableaux pour l'étude de puissance

TABLEAU 4.11. Nombres de rejets (sous l'hypothèse alternative) basés sur 10000 réalisations en utilisant la fonction ψ donnant les fonctions de corrélation usuelles avec les estimateurs des moindres carrés et les statistiques de Hong Q_n données par (4.2). Pour chaque valeur de m , les trois rangées correspondent aux trois scénarios.

		$\alpha = 0.01$					$\alpha = 0.05$					$\alpha = 0.10$				
$n = 100$																
m		QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
5		2736	3467	2731	2611	1233	3971	4677	3985	3854	2447	4817	5423	4816	4696	3312
		(1725)	(2204)	(1721)	(1601)	(698)	(3396)	(4033)	(3402)	(3288)	(2153)	(4636)	(5235)	(4636)	(4509)	(3164)
		238	433	232	219	49	629	884	601	560	176	954	1261	938	866	378
		9502	9824	9485	9367	5090	9829	9941	9817	9775	7270	9904	9976	9899	9881	8271
8		2023	2786	2029	1916	889	3349	4022	3365	3226	1969	4165	4880	4175	4066	2839
		(1320)	(1828)	(1297)	(1265)	(563)	(2869)	(3520)	(2871)	(2821)	(1667)	(4004)	(4726)	(3989)	(3951)	(2703)
		115	226	116	88	23	376	568	371	332	105	606	869	597	551	238
		8338	9418	8328	7946	2270	9330	9792	9313	9128	4473	9624	9886	9619	9497	5869
12		1494	2211	1513	1410	717	2820	3510	2828	2712	1722	3664	4350	3672	3587	2531
		(974)	(1463)	(971)	(944)	(415)	(2499)	(3069)	(2486)	(2407)	(1461)	(3527)	(4218)	(3545)	(3436)	(2382)
		49	103	47	45	9	207	360	194	163	59	400	583	388	359	145
		6255	8445	6200	5585	661	8039	9346	7993	7551	2070	8799	9628	8761	8435	3218
$n = 200$																
m		QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
5		5699	6582	5711	5515	3052	6919	7609	6943	6799	4669	7570	8137	7572	7450	5647
		(4120)	(5045)	(4130)	(4059)	(1938)	(6376)	(7145)	(6394)	(1208)	(4102)	(7428)	(8057)	(7432)	(7331)	(5521)
		451	749	437	391	68	962	1349	935	881	253	1400	1880	1362	1278	457
		10000	10000	9999	9998	9591	10000	10000	10000	10000	9904	10000	10000	10000	10000	9967
9		4370	5501	4403	4186	2035	5890	6791	5907	5743	3651	6710	7440	6720	6571	4622
		(3040)	(4221)	(3059)	(2925)	(1389)	(5362)	(6292)	(5376)	(5193)	(3221)	(6564)	(7356)	(6601)	(6431)	(4436)
		165	331	188	139	32	451	754	492	389	102	745	1140	820	640	207
		9960	9996	9971	9921	7214	9996	10000	9996	9992	8923	9997	10000	9999	9997	9410
15		3268	4403	3283	3122	1406	4867	5898	4881	4707	2837	5839	6752	5833	5673	3884
		(2235)	(3218)	(2239)	(2156)	(988)	(4394)	(5429)	(4375)	(4190)	(2520)	(5648)	(6637)	(5678)	(5512)	(3785)
		64	123	96	46	8	195	364	315	150	35	363	582	576	301	94
		9509	9926	9829	9238	2741	9869	9986	9964	9760	5279	9943	9997	9994	9887	6693

TABLEAU 4.12. Nombres de rejets (sous l'hypothèse alternative) basés sur 10000 réalisations en utilisant la fonction ψ donnant les fonctions de corrélation usuelles avec les estimateurs des moindres carrés et les statistiques de Hong Q_n^* données par (4.11). Pour chaque valeur de m , les trois rangées correspondent aux trois scénarios.

		$\alpha = 0.01$					$\alpha = 0.05$					$\alpha = 0.10$				
$n = 100$	m	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
	5	2618 (1725)	3564 (2204)	2562 (1721)	2492 (1601)	1510 (698)	3863 (3396)	4781 (4033)	3802 (3400)	3747 (3284)	2876 (2153)	4705 (4636)	5498 (5235)	4640 (4638)	4556 (4511)	3733 (3163)
		222	465	209	196	69	577	924	546	525	265	907	1312	849	808	505
		9463	9845	9414	9310	5707	9811	9948	9787	9750	7745	9898	9976	9891	9875	8671
	8	1861 (1319)	2721 (1828)	1773 (1294)	1720 (1265)	898 (563)	3156 (2869)	3964 (3520)	3020 (2871)	3013 (2821)	2022 (1668)	3964 (4006)	4797 (4728)	3804 (3990)	3846 (3951)	2899 (2702)
		93	216	81	74	23	324	545	276	268	109	540	837	496	488	261
		8154	9400	8013	7677	2286	9237	9782	9153	8992	4559	9571	9881	9511	9413	5949
	12	1264 (973)	1997 (1463)	1154 (971)	1155 (944)	535 (415)	2503 (2499)	3288 (3067)	2301 (2488)	2380 (2408)	1411 (1460)	3338 (3529)	4092 (4216)	3106 (3548)	3174 (3438)	2185 (2382)
		42	86	34	35	4	136	298	112	106	39	318	524	247	262	111
		5711	8227	5355	4998	470	7670	9241	7323	7084	1623	8544	9562	8252	8074	2692
$n = 200$	m	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
	5	5634 (4120)	6726 (5045)	5604 (4130)	5447 (4060)	3599 (1936)	6872 (6376)	7716 (7145)	6858 (6394)	6750 (6208)	5304 (4103)	7528 (7428)	8215 (8056)	7491 (7432)	7399 (7331)	6227 (5520)
		439	795	413	372	120	939	1422	890	856	377	1363	1971	1303	1252	659
		10000	10000	9999	9998	9749	10000	10000	10000	10000	9943	10000	10000	10000	10000	9987
	9	4252 (3040)	5481 (4219)	4171 (3059)	4040 (2926)	2220 (1387)	5769 (5362)	6769 (6292)	5691 (5376)	5587 (5194)	3872 (3218)	6585 (6567)	7423 (7356)	6515 (6602)	6453 (6429)	4926 (4438)
		150	327	155	124	35	416	746	430	357	120	696	1121	727	596	242
		9950	9996	9963	9912	7483	9995	10000	9996	9988	9072	9997	10000	9998	9997	9512
	15	3050 (2235)	4252 (3218)	2895 (2241)	2842 (2156)	1291 (988)	4598 (4393)	5749 (5429)	4410 (4377)	4393 (4186)	2662 (2522)	5586 (5648)	6619 (6636)	5360 (5680)	5408 (5515)	3691 (3786)
		50	108	71	37	6	163	323	241	120	30	316	548	427	249	81
		9424	9916	9729	9086	2459	9843	9985	9938	9686	5016	9928	9997	9978	9866	6471

TABLEAU 4.13. *Nombres de rejets (sous l'hypothèse alternative) basés sur 10000 réalisations en utilisant la fonction ψ donnant les fonctions de corrélation usuelles avec les estimateurs des moindres carrés et les statistiques de Haugh S_M et S_M^* données par (4.1) et (4.10). Pour chaque valeur de M , les trois rangées correspondent aux trois scénarios.*

	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.10$	
$n = 100$						
M	S_M	S_M^*	S_M	S_M^*	S_M	S_M^*
6	406 (639)	484 (581)	1541 (1899)	1750 (1845)	2611 (3024)	2846 (2949)
	6	10	72	98	222	272
	1868	2002	5076	5254	6855	7003
12	191 (415)	334 (403)	907 (1461)	1349 (1382)	1689 (2382)	2302 (2278)
	1	2	16	40	56	119
	131	214	920	1246	2038	2493
18	95 (354)	273 (319)	509 (1267)	1138 (1198)	1041 (2142)	2007 (2013)
	2	3	6	23	16	85
	12	39	140	308	387	779
$n = 200$						
M	S_M	S_M^*	S_M	S_M^*	S_M	S_M^*
6	1615 (1829)	1694 (1790)	3665 (3807)	3780 (3748)	5010 (5163)	5129 (5089)
	17	17	117	128	299	327
	8069	8111	9603	9612	9868	9869
12	880 (1158)	1036 (1093)	2371 (2664)	2687 (2583)	3571 (4036)	3913 (3902)
	1	4	28	42	93	120
	2676	2883	6097	6304	7749	7920
18	498 (877)	722 (814)	1689 (2360)	2152 (2238)	2710 (3551)	3357 (3419)
	2	7	10	16	27	46
	497	643	2153	2546	3735	4202

TABLEAU 4.14. Nombres de rejets (sous l'hypothèse alternative) basés sur 10000 réalisations en utilisant la fonction ψ de type Huber avec les estimateurs RA et les statistiques robustes Q_{Rn} données par (4.8).
Pour chaque valeur de m , les trois rangées correspondent aux trois scénarios.

		$\alpha = 0.01$					$\alpha = 0.05$					$\alpha = 0.10$				
$n = 100$																
m		QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
5		2552	3293	2569	2422	1169	3839	4498	3842	3715	2309	4642	5222	4677	4550	3194
		(1477)	(1957)	(1470)	(1402)	(660)	(3225)	(3890)	(3243)	(3130)	(2002)	(4428)	(5053)	(4441)	(4287)	(3033)
		2100	2686	2106	1988	999	3243	3853	3260	3128	1983	4054	4624	4050	3962	2732
		3679	4603	3700	3520	1695	5099	5875	5110	4953	3098	5889	6611	5900	5769	4042
8		1918	2596	1915	1798	849	3159	3880	3182	3077	1892	4042	4714	4046	3937	2749
		(1087)	(1588)	(1120)	(1052)	(483)	(2762)	(3421)	(2744)	(2646)	(1642)	(3861)	(4491)	(3848)	(3767)	(2616)
		1551	2150	1552	1471	753	2670	3294	2676	2576	1665	3507	4124	3499	3428	2438
		2771	3759	2794	2625	1249	4321	5162	4343	4169	2479	5198	5951	5195	5038	3476
12		1429	2053	1436	1341	673	2665	3355	2659	2589	1663	3570	4195	3575	3489	2461
		(880)	(1294)	(902)	(859)	(411)	(2351)	(2918)	(2324)	(2270)	(1417)	(3434)	(4013)	(3431)	(3316)	(2346)
		1216	1675	1223	1162	601	2253	2826	2237	2188	1475	3069	3659	3081	3008	2189
		2092	2987	2116	1997	985	3625	4496	3634	3502	2130	4583	5378	4595	4443	3019
$n = 200$																
m		QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
5		5440	6378	5452	5277	2854	6728	7403	6742	6575	4461	7392	7936	7403	7263	5437
		(3821)	(4883)	(3827)	(3731)	(1835)	(6142)	(6920)	(6157)	(6018)	(3947)	(7260)	(7856)	(7262)	(7136)	(5303)
		4370	5279	4374	4185	2123	5706	6434	5723	5575	3567	6424	7046	6432	6312	4482
		7160	7988	7173	6949	4165	8187	8721	8202	8053	5862	8683	9050	8685	8582	6787
9		4190	5300	4216	3991	1924	5654	6560	5675	5504	3444	6522	7266	6549	6370	4453
		(2840)	(3958)	(2840)	(2730)	(1350)	(5169)	(6074)	(5156)	(4965)	(3055)	(6390)	(7154)	(6385)	(6249)	(4275)
		3176	4185	3187	3044	1430	4635	5563	4648	4480	2807	5517	6327	5549	5393	3685
		5740	6946	5786	5536	2885	7170	8046	7208	6977	4623	7877	8565	7910	7739	5684
15		3047	4209	3061	2885	1301	4691	5690	4685	4513	2704	5615	6516	5623	5494	3710
		(2071)	(2949)	(2059)	(1978)	(859)	(4174)	(5169)	(4190)	(3997)	(2427)	(5427)	(6415)	(5456)	(5266)	(3618)
		2315	3221	2337	2177	994	3763	4649	3783	3631	2190	4664	5565	4697	4536	3111
		4427	5779	4537	4193	2012	6086	7175	6200	5913	3602	6960	7872	7051	6802	4740

TABLEAU 4.15. *Nombres de rejets (sous l'hypothèse alternative) basés sur 10000 réalisations en utilisant la fonction ψ de type Huber avec les estimateurs RA et les statistiques robustes Q_{Rn}^* données par (4.9). Pour chaque valeur de m , les trois rangées correspondent aux trois scénarios.*

	$\alpha = 0.01$					$\alpha = 0.05$					$\alpha = 0.10$				
$n = 100$															
m	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
5	2443 (1477)	3392 (1957)	2410 (1470)	2313 (1402)	1418 (660)	3727 (3223)	4571 (3890)	3658 (3242)	3594 (3130)	2724 (2001)	4541 (4427)	5324 (5053)	4457 (4441)	4396 (4289)	3621 (3033)
	2007	2765	1967	1892	1214	3135	3962	3068	2999	2337	3948	4717	3864	3848	3175
	3561	4709	3514	3380	2022	4982	5970	4930	4834	3543	5798	6667	5720	5650	4551
8	1740 (1087)	2541 (1588)	1678 (1118)	1626 (1052)	858 (483)	2976 (2761)	3814 (3421)	2866 (2746)	2855 (2645)	1947 (1642)	3821 (3863)	4641 (4488)	3660 (3849)	3705 (3767)	2821 (2616)
	1424	2092	1372	1355	769	2502	3213	2370	2375	1693	3300	4037	3162	3183	2504
	2566	3702	2459	2414	1267	4097	5092	3976	3927	2530	4970	5871	4826	4819	3556
12	1197 (888)	1879 (1294)	1100 (902)	1106 (860)	485 (411)	2358 (2351)	3129 (2918)	2129 (2325)	2199 (2270)	1350 (1417)	3215 (3438)	3965 (4014)	2951 (3428)	3065 (3318)	2112 (2346)
	1051	1542	952	977	462	1953	2615	1773	1850	1211	2737	3421	2497	2594	1863
	1821	2799	1677	1675	751	3256	4301	3021	3072	1808	4202	5138	3931	3998	2627
$n = 200$															
m	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
5	5390 (3821)	6498 (4882)	5377 (3827)	5216 (3730)	3400 (1835)	6682 (6139)	7489 (6919)	6639 (6157)	6516 (6017)	5082 (3945)	7339 (7260)	8018 (7854)	7317 (7262)	7211 (7136)	6031 (5304)
	4309	5407	4290	4136	2636	5649	6556	5627	5513	4137	6371	7144	6344	6247	5179
	7099	8092	7068	6880	4808	8148	8787	8138	8016	6469	8651	9098	8627	8547	7355
9	4076 (2839)	5282 (3956)	4000 (2841)	3855 (2731)	2108 (1350)	5532 (5168)	6540 (6074)	5459 (5156)	5370 (4964)	3671 (3055)	6430 (6393)	7250 (7155)	6290 (6385)	6231 (6247)	4737 (4274)
	3057	4168	3005	2925	1582	4508	5538	4426	4343	3005	5404	6298	5309	5254	3970
	5630	6927	5565	5400	3089	7055	8036	6999	6846	4873	7792	8557	7745	7620	5973
15	2826 (2071)	4060 (2946)	2695 (2062)	2650 (1978)	1193 (859)	4421 (4175)	5539 (5169)	4192 (4192)	4196 (3997)	2516 (2428)	5374 (5427)	6373 (6415)	5163 (5456)	5167 (5265)	3564 (3618)
	2102	3086	1963	1957	898	3525	4500	3363	3373	2034	4407	5385	4213	4253	2952
	4158	5614	4062	3915	1870	5859	7027	5747	5610	3418	6725	7750	6627	6550	4558

TABLEAU 4.16. Nombres de rejets (sous l'hypothèse alternative) basés sur 10000 réalisations en utilisant la fonction ψ de type Huber avec les estimateurs RA et les statistiques de Li et Hui S_{RM} et S_{RM}^* données par (4.6) et (4.7). Pour chaque valeur de M , les trois rangées correspondent aux trois scénarios.

	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.10$	
$n = 100$						
M	S_{Rn}	S_{Rn}^*	S_{Rn}	S_{Rn}^*	S_{Rn}	S_{Rn}^*
6	382 (598)	450 (531)	1470 (1788)	1656 (1744)	2511 (2929)	2776 (2866)
	333	391	1264	1432	2149	2398
	654	752	2034	2240	3240	3479
12	181 (411)	318 (370)	858 (1416)	1283 (1315)	1631 (2346)	2246 (2225)
	176	299	768	1168	1455	1998
	317	527	1198	1708	2098	2724
18	89 (365)	274 (312)	464 (1228)	1075 (1117)	1003 (2066)	1915 (1923)
	81	258	444	1023	937	1828
	163	448	736	1459	1392	2401
$n = 200$						
M	S_{Rn}	S_{Rn}^*	S_{Rn}	S_{Rn}^*	S_{Rn}	S_{Rn}^*
6	1481 (1592)	1550 (1550)	3473 (3640)	3585 (3568)	4797 (4971)	4902 (4901)
	1020	1084	2699	2817	3915	4028
	2388	2477	4787	4883	6122	6220
12	777 (966)	938 (926)	2273 (2634)	2528 (2536)	3429 (3810)	3769 (3705)
	541	657	1755	2021	2857	3153
	1288	1502	3215	3558	4524	4855
18	465 (744)	699 (688)	1563 (2251)	2058 (2122)	2581 (3393)	3239 (3274)
	358	539	1237	1650	2145	2724
	797	1100	2281	2798	3452	4044

TABLEAU 4.17. *Nombres de rejets (sous l'hypothèse alternative) basés sur 10000 réalisations en utilisant la fonction ψ de type bicarré avec les estimateurs RA et les statistiques robustes Q_{Rn} données par (4.8). Pour chaque valeur de m , les trois rangées correspondent aux trois scénarios.*

	$\alpha = 0.01$					$\alpha = 0.05$					$\alpha = 0.10$				
$n = 100$															
m	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
5	2539 (1493)	3294 (2001)	2561 (1506)	2421 (1382)	1185 (643)	3834 (3208)	4481 (3863)	3846 (3231)	3698 (3153)	2291 (1966)	4624 (4417)	5214 (5014)	4644 (4430)	4537 (4333)	3159 (3045)
	2344	3024	2355	2234	1107	3566	4177	3570	3450	2156	4355	4953	4365	4267	2982
	2446	3157	2456	2327	1134	3699	4362	3721	3562	2210	4503	5089	4521	4402	3052
8	1901 (1108)	2617 (1582)	1909 (1110)	1799 (1040)	833 (490)	3156 (2723)	3885 (3402)	3169 (2735)	3065 (2633)	1894 (1621)	4015 (3841)	4701 (4511)	4017 (3833)	3916 (3739)	2745 (2569)
	1764	2396	1778	1671	787	2938	3634	2937	2841	1827	3839	4411	3845	3726	2613
	1834	2510	1836	1721	842	3043	3761	3043	2946	1857	3892	4577	3914	3791	2648
12	1435 (915)	2058 (1252)	1446 (908)	1348 (857)	667 (426)	2661 (2339)	3341 (2929)	2659 (2311)	2563 (2235)	1655 (1399)	3559 (3382)	4170 (4026)	3557 (3392)	3463 (3308)	2461 (2301)
	1355	1872	1365	1288	617	2499	3113	2506	2424	1569	3380	3996	3388	3282	2346
	1363	1942	1382	1302	654	2560	3231	2559	2463	1596	3414	4077	3408	3357	2393
$n = 200$															
m	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
5	5435 (3810)	6360 (4783)	5456 (3824)	5260 (3720)	2857 (1809)	6696 (6112)	7393 (6902)	6713 (6137)	6550 (5987)	4453 (3933)	7374 (7256)	7917 (7835)	7392 (7271)	7271 (7162)	5437 (5287)
	4927	5871	4962	4763	2541	6257	6989	6282	6129	4070	6985	7599	7002	6858	5025
	5107	6054	5131	4933	2650	6408	7159	6424	6286	4156	7139	7752	7154	7013	5165
9	4163 (2854)	5278 (3898)	4178 (2856)	3968 (2766)	1951 (1332)	5650 (5104)	6550 (6053)	5670 (5147)	5481 (4957)	3409 (3083)	6498 (6385)	7257 (7168)	6506 (6402)	6358 (6242)	4432 (4199)
	3726	4740	3726	3561	1726	5199	6125	5207	5039	3156	6079	6857	6105	5909	4115
	3836	4943	3859	3665	1832	5345	6302	5366	5216	3285	6220	6997	6240	6083	4259
15	3061 (1995)	4176 (2950)	3070 (2030)	2881 (1883)	1308 (835)	4650 (4194)	5678 (5178)	4665 (4201)	4486 (4043)	2714 (2422)	5628 (5450)	6534 (6405)	5633 (5493)	5474 (5291)	3756 (3627)
	2723	3753	2750	2585	1174	4289	5217	4298	4123	2461	5171	6090	5177	5059	3441
	2845	3866	2891	2705	1230	4389	5403	4402	4257	2557	5342	6238	5375	5213	3531

TABLEAU 4.18. *Nombres de rejets (sous l'hypothèse alternative) basés sur 10000 réalisations en utilisant la fonction ψ de type bicarré avec les estimateurs RA et les statistiques robustes Q_{Rn}^* données par (4.9). Pour chaque valeur de m , les trois rangées correspondent aux trois scénarios.*

	$\alpha = 0.01$					$\alpha = 0.05$					$\alpha = 0.10$				
$n = 100$															
m	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
5	2438 (1494)	3390 (2001)	2397 (1506)	2311 (1382)	1439 (642)	3701 (3208)	4562 (3861)	3655 (3229)	3580 (3153)	2705 (1968)	4519 (4416)	5297 (5014)	4460 (4431)	4400 (4335)	3605 (3043)
	2245	3114	2199	2138	1373	3462	4248	3388	3339	2533	4240	5050	4157	4129	3433
	2343	3272	2310	2222	1385	3585	4446	3528	3439	2576	4386	5166	4304	4287	3547
8	1748 (1108)	2558 (1582)	1669 (1109)	1618 (1040)	844 (490)	2987 (2718)	3821 (3402)	2863 (2733)	2841 (2632)	1936 (1621)	3788 (3841)	4635 (4510)	3644 (3835)	3680 (3740)	2813 (2567)
	1636	2345	1575	1528	792	2731	3561	2621	2627	1869	3608	4350	3450	3489	2679
	1665	2443	1585	1556	847	2840	3693	2708	2708	1891	3675	4507	3529	3555	2712
12	1219 (915)	1867 (1252)	1115 (908)	1113 (858)	501 (426)	2336 (2338)	3124 (2926)	2116 (2311)	2205 (2235)	1347 (1398)	3176 (3377)	3949 (4026)	2941 (3395)	3034 (3308)	2115 (2302)
	1134	1723	1028	1029	466	2186	2901	2000	2074	1275	2997	3734	2751	2858	1998
	1175	1791	1083	1091	492	2233	3003	2019	2094	1318	3072	3835	2823	2931	2042
$n = 200$															
m	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR	QS	BAR	DAN	PAR	TR
5	5374 (3809)	6481 (4783)	5350 (3824)	5202 (3717)	3392 (1809)	6654 (6111)	7492 (6902)	6623 (6137)	6490 (5988)	5077 (3933)	7343 (7256)	7997 (7835)	7321 (7271)	7221 (7161)	6028 (5286)
	4872	6015	4853	4680	3031	6215	7095	6185	6057	4663	6938	7697	6922	6806	5658
	5056	6188	5043	4867	3181	6363	7271	6339	6232	4761	7100	7839	7071	6968	5817
9	4030 (2854)	5264 (3898)	3988 (2856)	3811 (2765)	2111 (1333)	5534 (5104)	6534 (6055)	5441 (5147)	5350 (4957)	3676 (3083)	6385 (6385)	7243 (7171)	6294 (6404)	6240 (6242)	4716 (4197)
	3602	4727	3521	3421	1874	5051	6113	4960	4900	3382	5955	6832	5866	5786	4372
	3730	4927	3646	3524	1983	5221	6281	5134	5084	3514	6120	6979	6015	5961	4518
15	2816 (1995)	4030 (2950)	2685 (2030)	2632 (1883)	1187 (835)	4396 (4193)	5499 (5178)	4177 (4202)	4182 (4039)	2510 (2422)	5363 (5451)	6374 (6404)	5140 (5493)	5167 (5291)	3563 (3627)
	2508	3577	2360	2338	1053	4024	5071	3811	3840	2296	4926	5939	4729	4759	3281
	2646	3709	2494	2442	1113	4152	5261	3939	3964	2390	5116	6087	4888	4932	3369

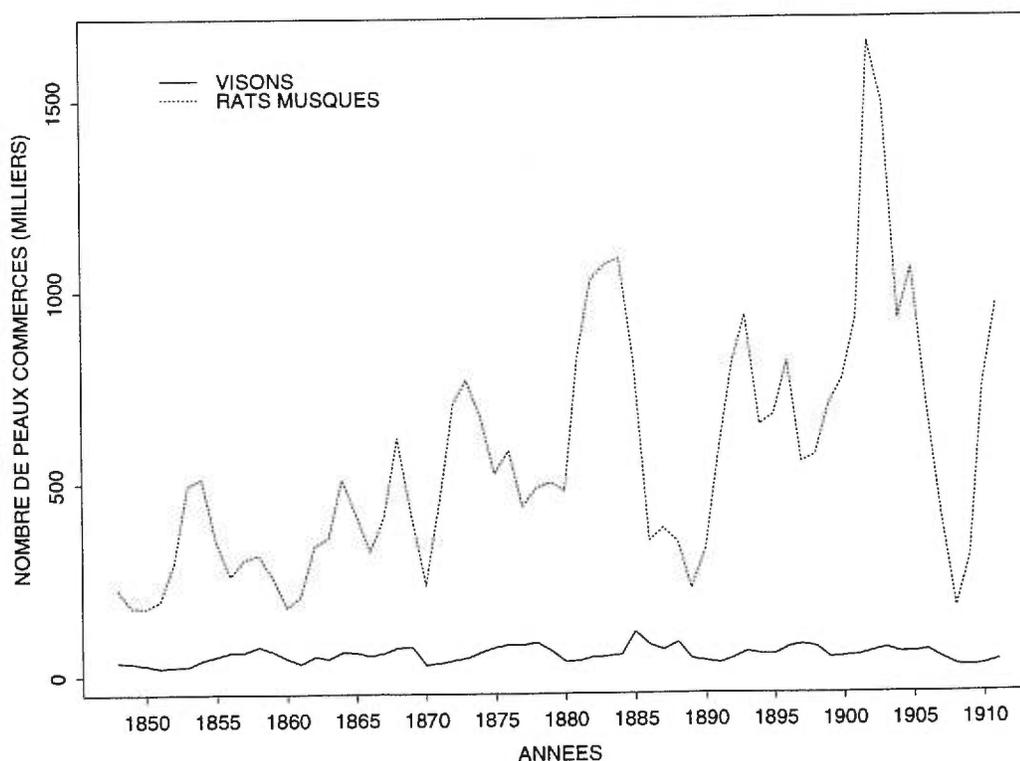
TABLEAU 4.19. *Nombres de rejets (sous l'hypothèse alternative) basés sur 10000 réalisations en utilisant la fonction ψ de type bicarré avec les estimateurs RA et les statistiques de Li et Hui S_{RM} et S_{RM}^* données par (4.6) et (4.7). Pour chaque valeur de M , les trois rangées correspondent aux trois scénarios.*

	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.10$	
$n = 100$						
M	S_{Rn}	S_{Rn}^*	S_{Rn}	S_{Rn}^*	S_{Rn}	S_{Rn}^*
6	371 (621)	451 (560)	1473 (1785)	1656 (1720)	2503 (2869)	2730 (2818)
	355	432	1382	1576	2381	2623
	377	465	1399	1584	2419	2661
12	183 (426)	322 (367)	876 (1399)	1291 (1310)	1632 (2301)	2230 (2206)
	179	301	804	1228	1541	2129
	176	310	825	1257	1577	2145
18	92 (369)	272 (315)	473 (1202)	1088 (1112)	998 (2068)	1916 (1926)
	80	251	447	1024	937	1834
	87	253	455	1037	951	1888
$n = 200$						
M	S_{Rn}	S_{Rn}^*	S_{Rn}	S_{Rn}^*	S_{Rn}	S_{Rn}^*
6	1482 (1581)	1560 (1549)	3448 (3576)	3555 (3510)	4785 (4956)	4886 (4895)
	1280	1357	3145	3258	4418	4531
	1405	1459	3253	3351	4524	4650
12	780 (981)	940 (908)	2258 (2631)	2545 (2538)	3424 (3824)	3769 (3708)
	662	798	2032	2286	3141	3485
	724	883	2116	2419	3317	3634
18	461 (766)	703 (724)	1573 (2281)	2051 (2121)	2625 (3419)	3223 (3280)
	422	631	1402	1848	2373	2962
	462	652	1477	1921	2447	3025

4.6. APPLICATION

Nous considérons dans cette section une application avec des données réelles. Les données considérées portent sur le nombre de peaux d'animaux commercées annuellement pour la période 1848-1911, pour le vison (*VISON*) et le rat musqué (*RATMUS*). Ces données sont basées sur des enregistrements des ventes de fourrure de la Compagnie de la Baie d'Hudson au Canada. Il est présumé que l'effort des trappeurs a été approximativement constant à chaque année, de telle manière que le nombre d'animaux capturés est une proportion constante de la population. L'interaction entre le vison et le rat musqué a été analysée à de nombreuses reprises. Voir par exemple Chan et Wallis (1978) et les références supplémentaires contenues dans cet article. Les données sont disponibles dans Brockwell et Davis (1991), séries H et I. L'étude de ces données pourrait provenir d'un intérêt pour une modélisation entre un prédateur (le vison) et sa proie (le rat musqué).

FIG. 4.2. Séries chronologiques portant sur le nombre de peaux d'animaux commercées annuellement pour la période 1848-1911, pour le vison et le rat musqué.



Le but de cette section est d'illustrer les méthodes élaborées dans les sections précédentes. En particulier, nous allons voir comment les analyses conventionnelles peuvent être affectées lorsque des valeurs aberrantes sont présentes dans les séries à l'étude.

4.6.1. Analyse préliminaire des données

Les données sont représentées dans la figure 4.2. Après un examen des séries, nous avons considéré la transformation logarithmique afin de stabiliser la variance de chacune de ces séries. Posons maintenant

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} \log(VISON)_t \\ \log(RATMUS)_t \end{pmatrix}.$$

Nous postulons que les deux séries chronologiques peuvent être décrites par des processus autorégressifs. Une étape préliminaire consiste à tester l'hypothèse que chaque processus possède une racine unitaire. Pour des analyses similaires, voir Fuller (1996, pp. 564-565) ou Brockwell et Davis (1996, pp. 192-195). Nous utilisons la méthode des moindres carrés. Nous ajustons séparément un processus autorégressif du neuvième ordre pour la série sur les rats musqués et un AR(2) pour la série sur les visons. L'équation ajustée pour X_t est donnée par

$$\hat{X}_t = 4.27 + 0.89X_{t-1} - 0.29X_{t-2},$$

(1.07) (0.13) (0.13)

alors que l'équation ajustée pour Y_t est donnée par

$$\begin{aligned} \hat{Y}_t = & 2.37 + 0.95Y_{t-1} - 0.57Y_{t-2} + 0.45Y_{t-3} - 0.31Y_{t-4} + 0.14Y_{t-5} - \\ & (1.53) (0.14) (0.20) (0.21) (0.24) (0.25) \\ & 0.34Y_{t-6} + 0.27Y_{t-7} - 0.15Y_{t-8} + 0.38Y_{t-9}. \\ & (0.25) (0.25) (0.23) (0.16) \end{aligned}$$

Nous sommes intéressés à tester pour une racine unitaire dans chacune des deux séries. Pour ce faire, nous appliquons le test de Dickey-Fuller, tel que décrit dans Brockwell et Davis (1996, section 6.3.1). Commençons par la série sur les visons. Posons $\nabla X_t = X_t - X_{t-1}$. En régressant ∇X_t sur 1, X_{t-1} et ∇X_{t-1} , nous trouvons que la statistique de test de Dickey-Fuller est -3.98, et nous rejettons l'hypothèse de racine unitaire, même au niveau 1%. Avec la série sur les rats musqués, la régression de ∇Y_t sur 1, Y_{t-1} , $\nabla Y_{t-1}, \dots, \nabla Y_{t-8}$ fournit une statistique de test

de Dickey-Fuller valant -1.50, et même au niveau 10% nous ne pouvons rejeter l'hypothèse nulle. Nous concluons ainsi à la présence d'une racine unitaire et choisissons de modéliser ∇Y_t par un processus AR(8). Notons qu'un examen visuel ainsi que l'analyse du corrélogramme de la série des rats musqués permettait de douter de la stationnarité. Aussi, un examen visuel semblait indiquer qu'il n'était pas nécessaire de différencier la série sur les visons.

Nous avons modélisé individuellement les séries X_t et ∇Y_t . Un examen des résidus des deux séries nous a permis de confirmer qu'un modèle AR(2) décrivait bien la série X_t et qu'un modèle AR(8) était satisfaisant pour la série ∇Y_t . La fonction S-plus `ar()` a été utilisée pour l'estimation. Incidemment, l'ordre retenu par la minimisation du critère AIC suggérait de modéliser ∇Y_t par un AR(8). Nous avons vérifié que les modèles résultants étaient stationnaires, en calculant les racines des polynômes autorégressifs.

4.6.2. Tests de non-corrélation entre les deux séries

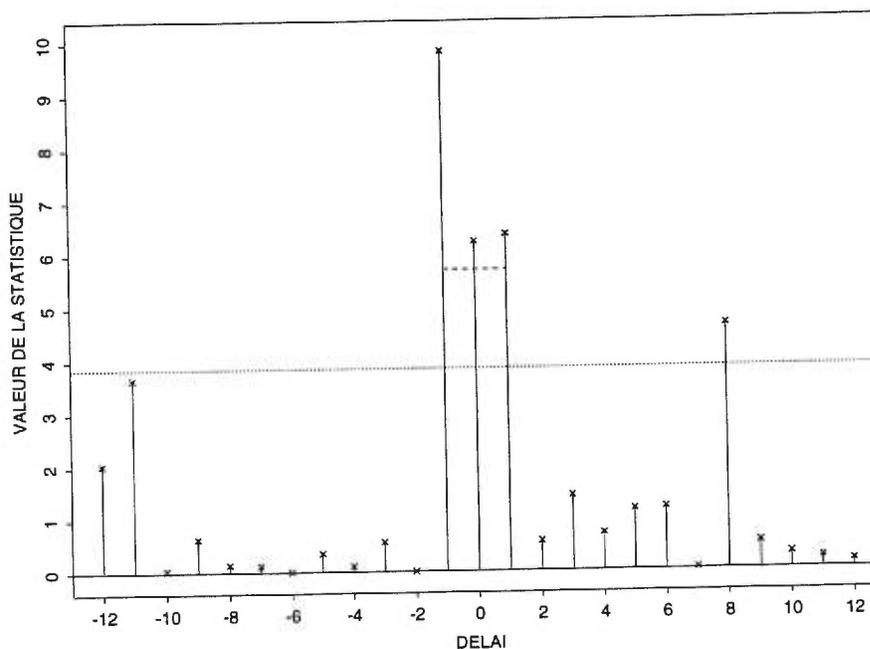
Pour tester la non-corrélation, on calcule la statistique de Hong Q_n^* dont les niveaux sont mieux contrôlés que Q_n dans notre étude empirique de la section 4.5. Nous avons choisi le noyau de Daniell et les valeurs $m = 3, 5, 10$. Nous obtenons respectivement $Q_n^* = 7.20, 6.34$ et $4.25 > 1.645$, et on rejette au niveau 5%. On rejette clairement l'hypothèse même au niveau 1%. (Dans cet exemple, avec $M = 6, 12$ on aurait également rejeté l'hypothèse d'indépendance avec la statistique de Haugh).

La figure 4.3 représente les statistiques $S^*(j)$, $|j| \leq 12$ données par l'équation (4.4) pour différents délais. On constate que des corrélations croisées significatives semblent présentes aux délais +1, 0 et -1. En fait, les valeurs des corrélations pour ces délais sont données par

délai	$r_{\hat{u}\hat{v}}(j)$
-1	-0.392
0	0.315
+1	0.315

où $r_{\hat{u}\hat{v}}(j)$ représente la corrélation croisée résiduelle entre la variable $\log(VISON)_t$ et la variable $\nabla \log(RATMUS)_{t-j}$. Ainsi, la figure 4.3 suggère également la dépendance entre les deux séries. En fait, Chan et Wallis (1978, p.175) concluent leurs analyses en affirmant qu'une croissance de rats musqués au temps t est habituellement suivie d'une augmentation du nombre de visons au temps $t + 1$, et qu'un grand nombre de visons au temps t est suivie d'une diminution du nombre de rats musqués au temps $t + 1$, confirmant un certain modèle de prédateur-proie.

FIG. 4.3. Tests à un délai basés sur les carrés des corrélations croisées entre les visons et les rats musqués. La ligne pointillée horizontale correspond à la valeur critique marginale au niveau $\alpha = 0.05$. La ligne pointillée supérieure correspond au test simultané de niveau $\alpha = 0.05$, basées sur les statistiques pour les délais $+1, 0, -1$.



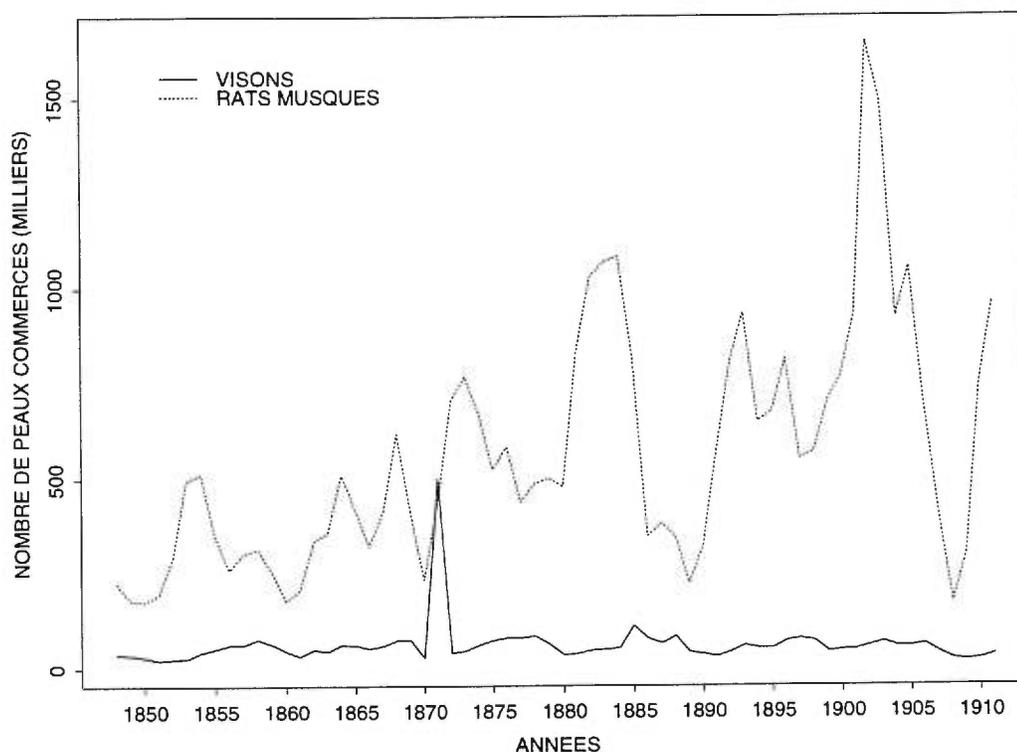
4.6.3. Tests de non-corrélation entre les deux séries avec une valeur aberrante

En guise d'illustration, nous remplaçons l'observation pour l'an 1871 dans la série des visons par 500000.

Les graphiques des deux séries se retrouvent dans la figure 4.4. Dans notre situation, un examen visuel permet de déceler une valeur aberrante additive, puisque la valeur aberrante est quelque peu extrême. Cependant, cette situation illustre bien notre point.

Nous ajustons encore une fois un AR(2) et un AR(8) aux séries X_t et ∇Y_t , respectivement. La valeur de la statistique de Hong avec le noyau de Daniell devient maintenant $Q_n^* = 1.51, 1.34$ et 0.96 , pour les mêmes valeurs $m = 3, 5, 10$ que précédemment. Ainsi, on ne parvient pas à rejeter la non-corrélation entre les deux séries au niveau 5%. Les tests basés sur les délais individuels sont représentés dans la figure 4.5, et on constate qu'ils sont assez affectés.

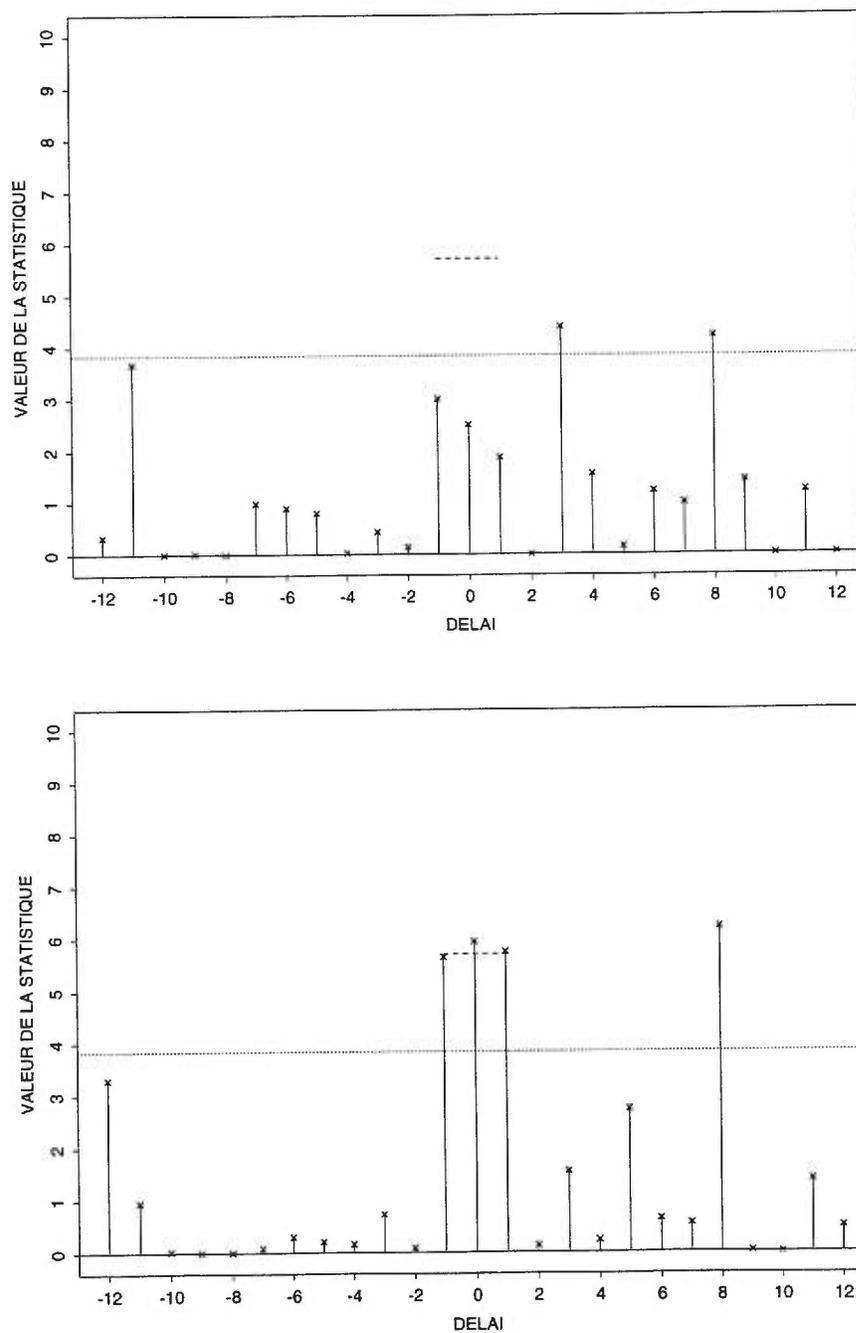
FIG. 4.4. Les deux variables à l'étude en présence d'une valeur aberrante.



Effectuons maintenant une analyse robuste. On considère ainsi la statistique Q_{Rn}^* , avec comme fonction ψ la fonction bicarré et avec le noyau de Daniell. Des modèles autorégressifs robustes sont calculés, avec la fonction `S-Plus ar.gm()`. Les estimateurs obtenus sont des GM-estimateurs. Quelques aspects des GM-estimateurs dans un contexte de séries chronologiques sont discutés dans Rousseeuw et Leroy (1987, pp. 278-279) pour les processus autorégressifs, dont la convergence en probabilité et la normalité asymptotique de ces estimateurs. La valeur de la statistique robustifiée de Hong devient alors $Q_{Rn}^* = 5.43, 4.60$ et 3.03 , pour $m = 3, 5$ et 10 , respectivement.

Ainsi, malgré la valeur aberrante, nous parvenons à conclure à la dépendance des deux séries. Le graphique de la figure 4.5 montre les tests basés sur les statistiques $S_R^*(j)$, $|j| \leq 12$, qui dépendent des corrélations croisées robustes à des délais individuels. On obtient des résultats plus satisfaisants que lorsque l'on utilise les corrélations croisées usuelles.

FIG. 4.5. Tests à un délai basés sur les carrés des corrélations croisées usuelles et robustes entre les visons et les rats musqués, avec une valeur aberrante. Le graphique du haut correspond aux corrélations croisées usuelles, alors que celui du bas correspond aux versions robustes. Les lignes pointillées s'interprètent de la même façon que précédemment.



4.7. PREUVE DU THÉORÈME

Nous montrons dans cette section le théorème 2. Posons Δ une constante positive bornée pouvant différer d'un endroit à l'autre. Posons aussi $k_{jm} = k(j/m)$ et $\eta_{t,s} = \eta(u_t/\sigma_u, v_s/\sigma_v)$. On montre alors le théorème suivant:

Étape 1. *Sous les hypothèses du théorème 2, $\tilde{Q}_{Rn} \rightarrow_L N(0, 1)$.*

DÉMONSTRATION. Posons $T_n = na^{-1} \sum_{j=-n+1}^{n-1} k_{jm}^2 \gamma_{uv}^2(j; \eta)$. On commence par décomposer $\gamma_{uv}^2(j; \eta)$, où l'on prend soin de distinguer les cas $j \geq 0$ et $j < 0$. Si $j \geq 0$, on montre que

$$\gamma_{uv}^2(j; \eta) = n^{-2} \sum_{t=j+1}^n \eta_{t,t-j}^2 + 2n^{-2} \sum_{t=j+2}^n \sum_{s=j+1}^{t-1} \eta_{t,t-j} \eta_{s,s-j},$$

alors que si $j < 0$, nous avons

$$\gamma_{uv}^2(j; \eta) = n^{-2} \sum_{t=j+1}^n \eta_{t-j,t}^2 + 2n^{-2} \sum_{t=j+2}^n \sum_{s=j+1}^{t-1} \eta_{t-j,t} \eta_{s-j,s}.$$

Ceci nous permet d'écrire T_n de la façon suivante:

$$\begin{aligned} T_n &= na^{-1} \sum_{j=0}^{n-1} k_{jm}^2 \gamma_{uv}^2(j; \eta) + na^{-1} \sum_{j=0}^{n-1} k_{jm}^2 \gamma_{uv}^2(-j; \eta), \\ &= (H_{1n} + H_{2n}) + (W_{1n} + W_{2n}), \\ &= H_n + W_n, \end{aligned}$$

où

$$\begin{aligned} H_{1n} &= \frac{1}{na} \sum_{j=0}^{n-1} k_{jm}^2 \left(\sum_{t=j+1}^n \eta_{t,t-j}^2 \right), \\ H_{2n} &= \frac{1}{na} \sum_{j=1}^{n-1} k_{jm}^2 \left(\sum_{t=j+1}^n \eta_{t-j,t}^2 \right), \\ W_{1n} &= \frac{2}{na} \sum_{j=0}^{n-2} k_{jm}^2 \left(\sum_{t=j+2}^n \sum_{s=j+1}^{t-1} \eta_{t,t-j} \eta_{s,s-j} \right), \\ W_{2n} &= \frac{2}{na} \sum_{j=1}^{n-2} k_{jm}^2 \left(\sum_{t=j+2}^n \sum_{s=j+1}^{t-1} \eta_{t-j,t} \eta_{s-j,s} \right). \end{aligned}$$

On calcule maintenant $E(H_n)$. Puisque $a = E(\eta_{s,t}^2)$, $\forall s, t$, on a alors que

$$\begin{aligned} E(H_{1n}) &= \frac{1}{an} \sum_{j=0}^{n-1} k_{jm}^2 \left[\sum_{t=j+1}^n E(\eta_{t,t-j}^2) \right], \\ &= \frac{1}{an} \sum_{j=0}^{n-1} k_{jm}^2 (n-j)a. \end{aligned}$$

De même, $E(H_{2n}) = \sum_{j=1}^{n-1} k_{jm}^2 (1-j/n)$, ce qui implique que $E(H_n) = M_n(k)$. On montre maintenant que $E(W_n) = 0$. Notons que pour $s < t$, $E(\eta_{t-j,t}\eta_{s-j,s}) = 0$ et $E(\eta_{t,t-j}\eta_{s,s-j}) = 0$, puisque $\eta(\cdot, \cdot)$ est une fonction impaire en chaque variable, que $\{u_t\}$ et $\{v_t\}$ sont indépendants de distributions symétriques. Ainsi, ceci nous permet d'affirmer que

$$\sum_{j=-n+1}^{n-1} k_{jm}^2 \gamma_{uv}^2(j; \eta) = O_p(m/n).$$

Afin d'établir l'étape 1, nous allons montrer deux choses:

$$\begin{aligned} (i) \quad & \frac{H_n - M_n(k)}{[2V_n(k)]^{1/2}} \rightarrow_P 0, \\ (ii) \quad & \frac{W_n}{[2V_n(k)]^{1/2}} \rightarrow_L N(0, 1). \end{aligned}$$

On commence par montrer (i). Le résultat sera établi si on montre la proposition suivante:

Proposition 1. $E[H_{jn} - E(H_{jn})]^2 = O(m^2/n)$, $j = 1, 2$.

DÉMONSTRATION. On montre le résultat pour $j = 1$. Le cas $j = 2$ se fait de la même manière.

On note que

$$H_{1n} - E(H_{1n}) = \sum_{j=0}^{n-1} k_{jm}^2 \left[n^{-1} \sum_{t=j+1}^n (\eta_{t,t-j}^2/a - 1) \right].$$

En utilisant l'inégalité de Minkovski,

$$E[H_{1n} - E(H_{1n})]^2 \leq \left\{ \sum_{j=0}^{n-1} k_{jm}^2 (E[n^{-1} \sum_{t=j+1}^n (\eta_{t,t-j}^2/a - 1)]^2)^{1/2} \right\}^2.$$

La proposition 1 sera établie si on montre que

$$E[n^{-1} \sum_{t=j+1}^n (\eta_{t,t-j}^2/a - 1)]^2 = O(n^{-1}), \text{ uniformément en } j.$$

Or

$$\begin{aligned} [n^{-1} \sum_{t=j+1}^n (\eta_{t,t-j}^2/a - 1)]^2 &= n^{-2} \sum_{t=j+1}^n (\eta_{t,t-j}^2/a - 1)^2 + \\ &2n^{-2} \sum_{t=j+2}^n \sum_{s=j+1}^{t-1} (\eta_{t,t-j}^2/a - 1)(\eta_{s,s-j}^2/a - 1), \end{aligned}$$

et donc

$$E[n^{-1} \sum_{t=j+1}^n (\eta_{t,t-j}^2/a - 1)]^2 = \frac{n-j}{n^2} E[\eta_{1,1}^2/a - 1]^2 = O(n^{-1}).$$

Puisque $E[(H_n - M_n(k))/(2V_n(k))^{1/2}] = 0$ et $\text{var}[(H_n - M_n(k))/(2V_n(k))^{1/2}] = \text{var}(H_n)/(2V_n(k)) = O(m/n)$, et puisque par hypothèse $m/n \rightarrow 0$, on a montré (i).

On s'applique maintenant à montrer (ii). On fait un changement de variables dans les indices on notant que $\sum_{j=1}^{n-2} \sum_{t=j+2}^n \sum_{s=j+1}^{t-1} = \sum_{t=3}^n \sum_{s=2}^{t-1} \sum_{j=1}^{s-1}$. Aussi $\sum_{j=0}^{n-2} \sum_{t=j+2}^n \sum_{s=j+1}^{t-1} = \sum_{t=2}^n \sum_{s=1}^{t-1} \sum_{j=0}^{s-1}$. Ceci nous permet d'écrire

$$W_n = 2a^{-1}\eta_{1,1}(n^{-1} \sum_{t=2}^n \eta_{t,t}) + W'_n,$$

où W'_n est défini par

$$\begin{aligned} W'_n &= n^{-1} \sum_{t=3}^n \sum_{s=2}^{t-1} \sum_{j=0}^{s-1} 2a^{-1}k_{jm}^2 \eta_{t,t-j} \eta_{s,s-j} + n^{-1} \sum_{t=3}^n \sum_{s=2}^{t-1} \sum_{j=1}^{s-1} 2a^{-1}k_{jm}^2 \eta_{t-j,t} \eta_{s-j,s}, \\ &= n^{-1} \sum_{t=3}^n (W_{1nt} + W_{2nt}), \\ &= n^{-1} \sum_{t=3}^n W_{nt}, \end{aligned}$$

avec les définitions suivantes pour W_{1nt} et W_{2nt} :

$$\begin{aligned} W_{1nt} &= \sum_{s=2}^{t-1} \sum_{j=0}^{s-1} 2a^{-1}k_{jm}^2 \eta_{t,t-j} \eta_{s,s-j}, \\ W_{2nt} &= \sum_{s=2}^{t-1} \sum_{j=1}^{s-1} 2a^{-1}k_{jm}^2 \eta_{t-j,t} \eta_{s-j,s}. \end{aligned}$$

On remarque que $2a^{-1}\eta_{1,1}(n^{-1} \sum_{t=2}^n \eta_{t,t}) = o_p(1)$ et peut être négligé par la suite. On note maintenant que (W_{nt}, \mathcal{F}_t) est une différence de martingales, où $\mathcal{F}_t = \sigma[(u_s, v_s), s \leq t]$. En effet, on vérifie que $E(W_{1nt} | \mathcal{F}_{t-1}) = 0$ puisque $E[\eta_{t,t-j} \eta_{s,s-j} | \mathcal{F}_{t-1}] = 0$. On fait de même

pour W_{2nt} . Afin de prouver (ii), on fait appel au théorème de Brown (1971), qui dans notre contexte affirme que $W_n/\sqrt{\text{var}(W_n)} \rightarrow_L N(0, 1)$ si on peut montrer les deux choses suivantes:

$$(A1) \quad \frac{n^{-2}}{\text{var}(W_n)} \sum_{t=3}^n E[W_{nt}^2 I(|W_{nt}| > \epsilon n[\text{var}(W_n)]^{1/2})] \rightarrow 0, \quad \forall \epsilon > 0,$$

$$(A2) \quad \frac{n^{-2}}{\text{var}(W_n)} \sum_{t=3}^n \check{W}_{nt}^2 \rightarrow_P 1,$$

où $I(A)$ est la fonction indicatrice de l'ensemble A et où $\check{W}_{nt}^2 = E(W_{nt}^2 | \mathcal{F}_{t-1})$. Commençons par montrer (A1). Il est suffisant de montrer la condition de Liapounov suivante:

$$\sigma^{-4}(n) n^{-4} \sum_{t=3}^n E(W_{nt}^4) \rightarrow 0,$$

où $\sigma^2(n) = 2V_n(k)$. En effet, la proposition suivante montre que $\text{var}(W'_n) = \sigma^2(n)$.

Proposition 2. $\text{var}(W'_n) = \sigma^2(n)$.

DÉMONSTRATION. $E(W_{nt}^2) = E(W_{1nt}^2) + E(W_{2nt}^2)$, puisque $E(W_{1nt}W_{2nt}) = 0$, ce qui se déduit du fait que

$$E[\eta_{t,t-j_1} \eta_{s_1,s_1-j_1} \eta_{t-j_2,t} \eta_{s_2-j_2,s_2}] = 0, \quad t > s_1, t > s_2.$$

On développe maintenant W_{1nt}^2 pour obtenir

$$\begin{aligned} W_{1nt}^2 &= 4a^{-2} \sum_{s=2}^{t-1} \left(\sum_{j=0}^{s-1} k_{jm}^2 \eta_{t,t-j} \eta_{s,s-j} \right)^2 + \\ &8a^{-2} \sum_{s_1=3}^{t-1} \sum_{s_2=2}^{s_1-1} \sum_{j_1=0}^{s_1-1} \sum_{j_2=0}^{s_2-1} k_{j_1 m}^2 k_{j_2 m}^2 \eta_{t,t-j_1} \eta_{s_1,s_1-j_1} \eta_{t,t-j_2} \eta_{s_2,s_2-j_2}. \end{aligned}$$

Ainsi $E(W_{1nt}^2) = 4a^{-2} \sum_{s=2}^{t-1} E[(\sum_{j=0}^{s-1} k_{jm}^2 \eta_{t,t-j} \eta_{s,s-j})^2]$, puisque

$$E(\eta_{t,t-j_1} \eta_{s_1,s_1-j_1} \eta_{t,t-j_2} \eta_{s_2,s_2-j_2}) = 0, \quad s_2 < s_1, s_1 < t, s_2 < t.$$

On développe le carré pour obtenir

$$\left(\sum_{j=0}^{s-1} k_{jm}^2 \eta_{t,t-j} \eta_{s,s-j} \right)^2 = \sum_{j=0}^{s-1} k_{jm}^4 \eta_{t,t-j}^2 \eta_{s,s-j}^2 + 2 \sum_{j_1=1}^{s-1} \sum_{j_2=0}^{j_1-1} k_{j_1 m}^2 k_{j_2 m}^2 \eta_{t,t-j_1} \eta_{s,s-j_1} \eta_{t,t-j_2} \eta_{s,s-j_2},$$

et ainsi $E(W_{1nt}^2) = 4 \sum_{s=2}^{t-1} \sum_{j=0}^{s-1} k_{jm}^4$, puisque

$$E(\eta_{t,t-j_1} \eta_{s,s-j_1} \eta_{t,t-j_2} \eta_{s,s-j_2}) = 0, \quad t > s, j_1 > j_2,$$

et aussi $E(\eta_{t,t-j}^2 \eta_{s,s-j}^2) = a^2$. On montre aussi que $E(W_{2nt}^2) = 4 \sum_{s=2}^{t-1} \sum_{j=1}^{s-1} k_{jm}^4$. Ceci implique que $E(W_{nt}^2) = 4 \sum_{s=2}^{t-1} \sum_{|j|=0}^{s-1} k_{jm}^4$. On fait alors des calculs similaires à Hong (1996a), et on montre que

$$\begin{aligned}
 \text{var}(W'_n) &= n^{-2} \sum_{t=3}^n E(W_{nt}^2), \\
 &= 4n^{-2} \sum_{t=3}^n \sum_{s=2}^{t-1} \sum_{|j|=0}^{s-1} k_{jm}^4, \\
 &= 2 \sum_{|j|=0}^{n-2} (1 - (|j| + 1)/n)(1 - |j|/n) k_{jm}^4, \\
 &= 2V_n(k).
 \end{aligned}$$

Ceci montre la proposition. □

Détaillons plus spécifiquement W_{1nt} et W_{2nt} de la façon suivante:

$$\begin{aligned}
 W_{1nt} &= 2a^{-1} \sum_{s=2}^{t-1} G_{t,s}^{(1)}, \\
 W_{2nt} &= 2a^{-1} \sum_{s=2}^{t-1} G_{t,s}^{(2)},
 \end{aligned}$$

où $G_{t,s}^{(1)} = \sum_{j=0}^{s-1} k_{jm}^2 \eta_{t,t-j} \eta_{s,s-j}$, $t > s$ et $G_{t,s}^{(2)} = \sum_{j=1}^{s-1} k_{jm}^2 \eta_{t-j,t} \eta_{s-j,s}$, $t > s$. □

On montre alors le lemme suivant.

Lemme 1. $E(G_{ts}^{(i)4}) = O(m^2)$, $t > s$, $i = 1, 2$.

DÉMONSTRATION. On va faire le résultat pour $G_{ts}^{(1)}$, l'autre cas étant similaire. Pour montrer le résultat, on commence par développer $G_{ts}^{(1)4}$. En calculant l'espérance, nous avons que

$$\begin{aligned}
E(G_{ts}^{(1)4}) &= \sum_{j=0}^{s-1} k_{jm}^8 E[(\eta_{t,t-j}\eta_{s,s-j})^4] + \\
&\quad \Delta \sum_{j_1 < j_2} k_{j_1 m}^4 k_{j_2 m}^4 E[(\eta_{t,t-j_1}\eta_{s,s-j_1})^2 (\eta_{t,t-j_2}\eta_{s,s-j_2})^2], \\
&\leq \Delta \left(\sum_{j=0}^{s-1} k_{jm}^8 + \sum_{j_1 < j_2} k_{j_1 m}^4 k_{j_2 m}^4 \right), \\
&\leq \Delta \left(\sum_{j=0}^{n-1} k_{jm}^4 \right)^2, \\
&= O(m^2).
\end{aligned}$$

Concernant la première égalité, les autres termes de $G_{ts}^{(1)4}$ sont d'espérance nulle, ce que l'on montre en conditionnant premièrement sur le processus $\{u_t\}$. La seconde inégalité découle du fait que les moments impliqués sont finis. \square

Le lemme précédent permet de montrer la proposition suivante:

Proposition 3. $E(W_{int}^4) = O(t^2 m^2)$, $i = 1, 2$.

DÉMONSTRATION. On fait les développements pour W_{1nt} , le terme W_{2nt} se faisant de manière comparable. Pour montrer le résultat, on commence par développer W_{1nt}^4 . En calculant l'espérance, on obtient alors

$$E(W_{1nt}^4) = 16a^{-4} \left\{ \sum_{s=2}^{t-1} E(G_{ts}^{(1)4}) + \Delta \sum_{s_1 < s_2} E(G_{ts_1}^{(1)2} G_{ts_2}^{(1)2}) \right\}.$$

Les autres termes dans le développement de W_{1nt}^4 sont d'espérance nulle, ce que l'on montre en conditionnant ces termes par rapport au processus $\{v_t\}$. On obtient alors

$$\begin{aligned}
E(W_{1nt}^4) &\leq \Delta \sum_{s_1=2}^{t-1} \sum_{s_2=2}^{t-1} E(G_{ts_1}^{(1)2} G_{ts_2}^{(1)2}), \\
&\leq \Delta \sum_{s_1=2}^{t-1} \sum_{s_2=2}^{t-1} [E(G_{ts_1}^{(1)4}) E(G_{ts_2}^{(1)4})]^{1/2}, \\
&\leq \Delta \left\{ \sum_{s=2}^{t-1} [E(G_{ts}^{(1)4})]^{1/2} \right\}^2, \\
&= O(t^2 m^2).
\end{aligned}$$

Ceci complète la preuve de la proposition. □

Ceci nous permet de montrer (A1) puisque

$$\sigma^{-4}(n) n^{-4} \sum_{t=3}^n E(W_{nt}^4) = O(n^{-1}).$$

On montre maintenant la partie (A2). On remarque en premier lieu que $\ddot{W}_{nt}^2 = \ddot{W}_{1nt}^2 + \ddot{W}_{2nt}^2$, où on pose $\ddot{W}_{1nt}^2 = 4a^{-2} E[(\sum_{s=2}^{t-1} G_{ts}^{(1)})^2 | \mathcal{F}_{t-1}]$ ainsi que $\ddot{W}_{2nt}^2 = 4a^{-2} E[(\sum_{s=2}^{t-1} G_{ts}^{(2)})^2 | \mathcal{F}_{t-1}]$. On va effectuer les détails pour \ddot{W}_{1nt}^2 . Le terme \ddot{W}_{2nt}^2 se traite de la même manière. On commence par décomposer \ddot{W}_{1nt}^2 de la manière suivante:

$$\ddot{W}_{1nt}^2 = E(\ddot{W}_{1nt}^2) + 4a^{-2} C_{1nt} + 4a^{-2} C_{2nt} + 4a^{-2} C_{3nt},$$

où

$$\begin{aligned}
E(\ddot{W}_{1nt}^2) &= 4 \sum_{s=2}^{t-1} \sum_{j=0}^{s-1} k_{jm}^4, \\
C_{1nt} &= \sum_{s=2}^{t-1} \sum_{j=0}^{s-1} k_{jm}^4 E[\eta_{s,s-j}^2 \eta_{t,t-j}^2 - a^2 | \mathcal{F}_{t-1}], \\
C_{2nt} &= 2 \sum_{s=2}^{t-1} \sum_{j_1=1}^{s-1} \sum_{j_2=0}^{j_1-1} k_{j_1 m}^2 k_{j_2 m}^2 E[\eta_{t,t-j_1} \eta_{t,t-j_2} \eta_{s,s-j_1} \eta_{s,s-j_2} | \mathcal{F}_{t-1}], \\
C_{3nt} &= 2 \sum_{s_1=3}^{t-1} \sum_{s_2=2}^{s_1-1} E[G_{ts_1}^{(1)} G_{ts_2}^{(1)} | \mathcal{F}_{t-1}].
\end{aligned}$$

Le lemme suivant sera utile pour la suite, et est une généralisation d'un résultat de Hong (1996a, p. 4).

Lemme 2. Pour $t_2 > s_2$, $t_1 > s_1$,

$$|E[E(G_{t_1 s_1}^{(1)} G_{t_1 s_2}^{(1)} | \mathcal{F}_{t_1-1}) E(G_{t_2 s_1}^{(1)} G_{t_2 s_2}^{(1)} | \mathcal{F}_{t_2-1})]| \leq \begin{cases} \Delta m^2 (m^{-1} \sum_{j=0}^{n-1} k_{jm}^2)^2, & \text{si } t_1 = t_2, \\ \Delta m (m^{-1} \sum_{j=0}^{n-1} k_{jm}^2)^2, & \text{si } t_1 < t_2. \end{cases}$$

DÉMONSTRATION. La démonstration est directe, se traitant par cas. Elle nécessite de nombreux développements qui sont omis ici. La preuve se fait de la même manière que le résultat correspondant dans Hong (1996a, p. 4). Cependant, les hypothèses sur la fonction η , le fait que $\{u_t\}$ et $\{v_t\}$ soient indépendants de distribution symétrique permet d'obtenir le résultat. \square

On montre maintenant la proposition suivante:

Proposition 4. $E(C_{1nt}^2) = O(t^2 m + tm^2)$, $E(C_{2nt}^2) = O(tm^3)$.

DÉMONSTRATION. On commence par montrer le résultat pour C_{1nt} . On écrit $C_{1nt} = \sum_{s=2}^{t-1} C_{1nst}$, où

$$C_{1nst} = \sum_{j=0}^{s-1} E[\eta_{s,s-j}^2 \eta_{t,t-j}^2 - a^2 | \mathcal{F}_{t-1}].$$

On obtient en développant

$$C_{1nt}^2 = \sum_{s=2}^{t-1} C_{1nst}^2 + 2 \sum_{s_1=3}^{t-1} \sum_{s_2=2}^{s_1-1} C_{1ns_1 t} C_{1ns_2 t}.$$

On étudie les deux morceaux séparément. Pour le premier terme, on montre facilement que $E(\sum_{s=2}^{t-1} C_{1nst}^2) = O(tm^2)$. L'étude du deuxième morceau permet de conclure que ce terme est d'ordre $O(t^2 m)$. Ainsi, $E(C_{1nt}^2) = O(t^2 m + tm^2)$. On montre maintenant le résultat pour C_{2nt} . On commence par remarquer en effectuant des changements de variables que l'on peut écrire alternativement C_{2nt} de la manière suivante:

$$\begin{aligned} C_{2nt} &= 2 \sum_{j_1=1}^{t-2} \sum_{j_2=j_1+1}^{t-2} \sum_{s=j_2+1}^{t-1} k_{j_1 m}^2 k_{j_2 m}^2 E[\eta_{t,t-j_1} \eta_{t,t-j_2} \eta_{s,s-j_1} \eta_{s,s-j_2} | \mathcal{F}_{t-1}], \\ &= 2 \sum_{j_1=1}^{t-2} k_{j_1 m}^2 C_{2nj_1 t}, \end{aligned}$$

où

$$C_{2nj_1 t} = \sum_{l_1=j_1+1}^{t-2} \sum_{s=l_1+1}^{t-1} k_{l_1 m}^2 E[\eta_{t,t-j_1} \eta_{t,t-j_2} \eta_{s,s-j_1} \eta_{s,s-j_2} | \mathcal{F}_{t-1}].$$

On développe alors C_{2nt}^2 , fournissant

$$C_{2nt}^2 = 4 \sum_{j_1=1}^{t-2} k_{j_1 m}^4 C_{2nj_1 t}^2 + 8 \sum_{j_1=2}^{t-2} \sum_{j_2=1}^{j_1-1} k_{j_1 m}^2 k_{j_2 m}^2 C_{2nj_1 t} C_{2nj_2 t}.$$

On traite alors les deux termes séparément. L'analyse du premier terme nous permet de montrer que $E(\sum_{j_1=1}^{t-2} k_{j_1 m}^4 C_{2nj_1 t}^2) = O(m^3 t)$ et l'espérance du second terme est nulle. Ainsi, $E(C_{2nt}^2) = O(m^3 t)$. \square

On peut maintenant montrer (A2). On veut montrer que $m^{-2} \text{var}(n^{-2} \sum_{t=3}^n \ddot{W}_{1nt}^2) \rightarrow 0$, et similairement pour \ddot{W}_{2nt} . Puisque

$$\begin{aligned} m^{-2} \text{var}(n^{-2} \sum_{t=3}^n \ddot{W}_{1nt}^2) &= m^{-2} E\{(n^{-2} \sum_{t=3}^n [\ddot{W}_{1nt} - E(\ddot{W}_{1nt})])^2\}, \\ &= m^{-2} E[(n^{-2} \sum_{i=1}^3 \sum_{t=3}^n 4a^{-2} C_{int})^2], \end{aligned}$$

il est donc suffisant de montrer

$$m^{-2} E[(n^{-2} \sum_{t=3}^n C_{int})^2] \rightarrow 0,$$

pour $i = 1, 2, 3$. Or en utilisant les résultats de la proposition précédente, on a que

$$\begin{aligned} m^{-2} E[n^{-2} \sum_{t=3}^n C_{1nt}]^2 &\leq m^{-2} n^{-4} \left\{ \sum_{t=3}^n E(C_{1nt}^2)^{1/2} \right\}^2, \\ &= O(m^{-1} + n^{-1}). \end{aligned}$$

Aussi,

$$\begin{aligned} m^{-2} E[n^{-2} \sum_{t=3}^n C_{2nt}]^2 &\leq m^{-2} n^{-4} \left\{ \sum_{t=3}^n E(C_{2nt}^2)^{1/2} \right\}^2, \\ &= O(m/n). \end{aligned}$$

Finalement,

$$m^{-2} E[n^{-2} \sum_{t=3}^n C_{3nt}]^2 = m^{-2} n^{-4} \sum_{t=3}^n E(C_{3nt}^2) + 2m^{-2} n^{-4} \sum_{t_2=3}^n \sum_{t_1=2}^{t_2-1} E(C_{3nt_1} C_{3nt_2}).$$

Or on vérifie que

$$E(C_{3nt_1} C_{3nt_2}) = 4 \sum_{s_2=3}^{t_1-1} \sum_{s_1=2}^{s_2-2} E[E(G_{t_1 s_1} G_{t_1 s_2} | \mathcal{F}_{t_1-1}) E(G_{t_2 s_1} G_{t_2 s_2} | \mathcal{F}_{t_2-1})]$$

Ainsi $E(C_{3nt}^2) \leq \Delta m^2 t^2$ et $E(C_{3nt_1} C_{3nt_2}) \leq \Delta m t_1^2$ par l'utilisation du lemme 2. Par conséquent, on a $m^{-2} E[n^{-2} \sum_{t=3}^n C_{3nt}]^2 = O(n^{-1} + m^{-1})$. Puisque $m/n \rightarrow 0$, $m \rightarrow \infty$, alors ceci établi (A2). Cela complète également la preuve de l'étape 1. \square

L'étape 1 visait à montrer que lorsque l'on considère \tilde{Q}_{Rn} qui est de la même forme que Q_{Rn} mais qui repose sur les processus innovations inobservés, alors $\tilde{Q}_{Rn} \rightarrow_L N(0, 1)$. Cette étape ne fait cependant pas intervenir les données et l'estimation des paramètres. C'est ce que nous allons faire dans l'étape 2. Plus précisément, nous allons établir que $\tilde{Q}_{Rn} - Q_{Rn} = o_p(1)$, ce qui nous permettra de conclure que $Q_{Rn} \rightarrow_L N(0, 1)$.

Ainsi l'étape 2 consiste à montrer le résultat suivant:

Étape 2.

$$\sum_{j=1-n}^{n-1} k_{jm}^2 [\gamma_{uv}^2(j; \eta) - \gamma_{\hat{u}\hat{v}}^2(j; \eta)] = o_p(\sqrt{m}/n).$$

En effet, on remarque que $V_n(k) = O(m)$. Or $a^2 - b^2 = (a - b)^2 + 2b(a - b)$ et donc $\gamma_{\hat{u}\hat{v}}^2(j; \eta) - \gamma_{uv}^2(j; \eta) = [\gamma_{\hat{u}\hat{v}}(j; \eta) - \gamma_{uv}(j; \eta)]^2 + 2\gamma_{uv}(j; \eta)[\gamma_{\hat{u}\hat{v}}(j; \eta) - \gamma_{uv}(j; \eta)]$. Il est donc suffisant de montrer les deux résultats suivants:

$$(B1) \quad \sum_{j=1-n}^{n-1} k_{jm}^2 [\gamma_{\hat{u}\hat{v}}(j; \eta) - \gamma_{uv}(j; \eta)]^2 = o_p(\sqrt{m}/n),$$

$$(B2) \quad \sum_{j=1-n}^{n-1} k_{jm}^2 \gamma_{uv}(j; \eta) [\gamma_{\hat{u}\hat{v}}(j; \eta) - \gamma_{uv}(j; \eta)] = o_p(\sqrt{m}/n).$$

On partitionne $\sum_{j=1-n}^{n-1} = \sum_{j < 0} + \sum_{j \geq 0}$ et on fait les détails pour $j \geq 0$, l'autre cas se traitant de manière similaire.

Afin d'établir le résultat, nous avons besoin de développements en séries de Taylor des résidus \hat{u}_t et \hat{v}_t . Ce genre de résultats a été établi pour les processus ARMA univariés dans Box et Pierce (1970) et pour les ARMA multidimensionnels dans Hosking (1980).

Proposition 5.

$$\begin{aligned} \hat{u}_t &= u_t + \sum_{r=1}^{p_1} \sum_{i=0}^{\infty} \Pi_{1i} (\hat{\phi}_{1r} - \phi_{1r}) X_{t-r-i} - \sum_{r=1}^{q_1} \sum_{i=0}^{\infty} \Pi_{1i} (\hat{\theta}_{1r} - \theta_{1r}) u_{t-r-i} + \lambda_{unt}, \\ \hat{v}_t &= v_t + \sum_{r=1}^{p_2} \sum_{i=0}^{\infty} \Pi_{2i} (\hat{\phi}_{2r} - \phi_{2r}) Y_{t-r-i} - \sum_{r=1}^{q_2} \sum_{i=0}^{\infty} \Pi_{2i} (\hat{\theta}_{2r} - \theta_{2r}) v_{t-r-i} + \lambda_{vnt}. \end{aligned}$$

Pour une preuve, voir par exemple pour la démarche le lemme 1 de Hosking (1980). Remarquons que Hosking (1980) se permet d'écrire $\lambda_{unt} = O_p(n^{-1})$ mais nous préférons

à ce stade conserver la dépendance sur t des restes λ_{unt} et λ_{vnt} . Cependant, la démarche de Hosking (1980) permet de conclure par exemple que $n^{-1} \sum_{t=1}^n \lambda_{unt} = O_p(n^{-1})$ et que $n^{-1} \sum_{t=1}^n \lambda_{vnt} = O_p(n^{-1})$. On a aussi que $n^{-1} \sum_{t=1}^n \lambda_{unt}^2 = O_p(n^{-2})$ et que $n^{-1} \sum_{t=1}^n \lambda_{vnt}^2 = O_p(n^{-2})$. Le lemme suivant nous permet d'avoir des résultats sur les ordres impliquant $u_t - \hat{u}_t$ et $v_t - \hat{v}_t$. Pour montrer le lemme suivant, on utilise la proposition 5.

Lemme 3. *Soit $\delta_{ut} = u_t - \hat{u}_t$ et $\delta_{vt} = v_t - \hat{v}_t$. On a alors que $n^{-1} \sum_{t=1}^n \delta_{ut}^2 = O_p(n^{-1})$ et $n^{-1} \sum_{t=1}^n \delta_{vt}^2 = O_p(n^{-1})$.*

Pour établir (B1) et (B2) nous aurons besoin de résultats sur $\gamma_{\hat{u}\hat{v}}(j; \eta) - \gamma_{uv}(j; \eta)$. Le développement en séries de Taylor de la fonction $\eta(\hat{u}_t, \hat{v}_t)$ autour du point (u_t, v_t) est utile à cette fin. Le résultat suivant se montre en effectuant un développement en séries de la fonction η , en sommant sur t et en divisant par n . Aussi il faut que $\sigma_u - \hat{\sigma}_u = O_p(n^{-1/2})$ et $\sigma_v - \hat{\sigma}_v = O_p(n^{-1/2})$. Notons que le résultat a été établi dans Li et Hui (1994, p. 106) pour le cas autorégressif, mais s'étant facilement au cas ARMA, lorsque les processus sont stationnaires et inversibles.

Proposition 6. *Soient $\eta_1(u, v) = \partial\eta(u, v)/\partial u$ et $\eta_2(u, v) = \partial\eta(u, v)/\partial v$. On a alors*

$$\begin{aligned} \gamma_{\hat{u}\hat{v}}(j; \eta) - \gamma_{uv}(j; \eta) &= -n^{-1} \sigma_u^{-1} \sum_{t=j+1}^n \eta_1(u_t/\sigma_u, v_{t-j}/\sigma_v) \delta_{ut} - \\ &\quad n^{-1} \sigma_v^{-1} \sum_{t=j+1}^n \eta_2(u_t/\sigma_u, v_{t-j}/\sigma_v) \delta_{v,t-j} + R_n. \end{aligned}$$

où $\delta_{ut} = u_t - \hat{u}_t$ et $\delta_{vt} = v_t - \hat{v}_t$. Le reste R_n est tel que $R_n = O_p(n^{-1})$.

On montre maintenant (B1). Pour ce faire, utilisant le fait que $(a_1 + a_2 + a_3)^2 \leq 4[a_1^2 + a_2^2 + a_3^2]$, nous obtenons en utilisant la proposition 6 que

$$\sum_{j=0}^{n-1} k_{jm}^2 [\gamma_{\hat{u}\hat{v}}(j; \eta) - \gamma_{uv}(j; \eta)]^2 \leq 4(\sigma_u^{-2} T_{1n} + \sigma_v^{-2} T_{2n} + T_{3n}),$$

où

$$\begin{aligned}
T_{1n} &= \sum_{j=0}^{n-1} k_{jm}^2 [n^{-1} \sum_{t=j+1}^n \eta_{1t,t-j} \delta_{ut}]^2, \\
T_{2n} &= \sum_{j=0}^{n-1} k_{jm}^2 [n^{-1} \sum_{t=j+1}^n \eta_{2t,t-j} \delta_{v,t-j}]^2, \\
T_{3n} &= \sum_{j=0}^{n-1} k_{jm}^2 R_n^2,
\end{aligned}$$

où $\eta_{1t,t-j} = \eta_1(u_t/\sigma_u, v_{t-j}/\sigma_v)$ et $\eta_{2t,t-j} = \eta_2(u_t/\sigma_u, v_{t-j}/\sigma_v)$.

On commence par étudier T_{1n} . Pour ce faire, on remplace δ_{ut} par la valeur obtenue dans la proposition 5. La démarche pour T_{2n} est la même.

On commence par écrire $n^{-1} \sum_{t=j+1}^n \eta_{1t,t-j} \delta_{ut}$ de la façon suivante:

$$\begin{aligned}
n^{-1} \sum_{t=j+1}^n \eta_{1t,t-j} \delta_{ut} &= - \sum_{r=1}^{p_1} (\hat{\phi}_{1r} - \phi_{1r}) \left\{ \sum_{i \geq 0} \Pi_{1i} [n^{-1} \sum_{t=j+1}^n \eta_{1t,t-j} X_{t-r-i}] \right\} + \\
&\quad \sum_{r=1}^{q_1} (\hat{\theta}_{1r} - \theta_{1r}) \left\{ \sum_{i \geq 0} \Pi_{1i} [n^{-1} \sum_{t=j+1}^n \eta_{1t,t-j} u_{t-r-i}] \right\} - \\
&\quad n^{-1} \sum_{t=j+1}^n \eta_{1t,t-j} \lambda_{unt}.
\end{aligned} \tag{4.12}$$

Ainsi on peut majorer $(n^{-1} \sum_{t=j+1}^n \eta_{1t,t-j} \delta_{ut})^2$ comme suit:

$$\begin{aligned}
(n^{-1} \sum_{t=j+1}^n \eta_{1t,t-j} \delta_{ut})^2 &\leq \Delta \left[\sum_{r=1}^{p_1} (\hat{\phi}_{1r} - \phi_{1r})^2 A_{rjn} + \right. \\
&\quad \left. \sum_{r=1}^{q_1} (\hat{\theta}_{1r} - \theta_{1r})^2 B_{rjn} + \right. \\
&\quad \left. (n^{-1} \sum_{t=j+1}^n \eta_{1t,t-j} \lambda_{unt})^2 \right],
\end{aligned}$$

où

$$\begin{aligned}
A_{rjn} &= \left(\sum_{i \geq 0} \Pi_{1i} [n^{-1} \sum_{t=j+1}^n \eta_{1t,t-j} X_{t-r-i}] \right)^2, \\
&= \left\{ n^{-1} \sum_{t=j+1}^n \eta_{1t,t-j} (\Pi_1(B) X_{t-r}) \right\}^2, \\
B_{rjn} &= \left(\sum_{i \geq 0} \Pi_{1i} [n^{-1} \sum_{t=j+1}^n \eta_{1t,t-j} u_{t-r-i}] \right)^2, \\
&= \left\{ n^{-1} \sum_{t=j+1}^n \eta_{1t,t-j} (\Pi_1(B) u_{t-r}) \right\}^2.
\end{aligned}$$

On a donc

$$\begin{aligned}
T_{1n} &\leq \Delta \left[\sum_{r=1}^{p_1} (\hat{\phi}_{1r} - \phi_{1r})^2 \left(\sum_{j=0}^{n-1} k_{jm}^2 A_{rjn} \right) + \right. \\
&\quad \left. \sum_{r=1}^{q_1} (\hat{\theta}_{1r} - \theta_{1r})^2 \left(\sum_{j=0}^{n-1} k_{jm}^2 B_{rjn} \right) + \right. \\
&\quad \left. \sum_{j=0}^{n-1} k_{jm}^2 \left(n^{-1} \sum_{t=j+1}^n \eta_{1t,t-j} \lambda_{unt} \right)^2 \right].
\end{aligned}$$

On montre que $\sum_{j=0}^{n-1} k_{jm}^2 A_{rjn} = O_p(m/n)$ et que $\sum_{j=0}^{n-1} k_{jm}^2 B_{rjn} = O_p(m/n)$. En effet on a que

$$\begin{aligned}
\left\{ n^{-1} \sum_{t=j+1}^n \eta_{1t,t-j} (\Pi_1(B) u_{t-r}) \right\}^2 &= n^{-2} \sum_{t=j+1}^n \eta_{1t,t-j}^2 (\Pi_1(B) X_{t-r})^2 + \\
&\quad 2n^{-2} \sum_{t=j+2}^n \sum_{s=j+1}^{t-1} \eta_{1t,t-j} \eta_{1s,s-j} (\Pi_1(B) X_{t-r}) (\Pi_1(B) X_{s-r}).
\end{aligned}$$

Puisque $E[\eta_{1t,t-j} \eta_{1s,s-j} (\Pi_1(B) X_{t-r}) (\Pi_1(B) X_{s-r})] = 0$, $t > s$, et que $E[\eta_{1t,t-j}^2 (\Pi_1(B) X_{t-r})^2]$ est fini, on montre donc que $\sum_{j=0}^{n-1} k_{jm}^2 A_{rjn} = O_p(m/n)$. On fait de la même manière pour B_{rjn} . On a également que

$$\left(n^{-1} \sum_{t=j+1}^n \eta_{1t,t-j} \lambda_{unt} \right)^2 \leq \left(n^{-1} \sum_{t=1}^n \eta_{1t,t-j}^2 \right) \left(n^{-1} \sum_{t=1}^n \lambda_{unt}^2 \right).$$

Ceci montre que $T_{1n} = O_p(m/n^2) = o_p(\sqrt{m}/n)$, puisque les estimateurs satisfont $\hat{\phi}_{1r} - \phi_{1r} = O_p(n^{-1/2})$ et $\hat{\theta}_{1r} - \theta_{1r} = O_p(n^{-1/2})$, que $n^{-1} \sum_{t=1}^n \lambda_{unt}^2 = O_p(n^{-2})$, et puisque $m/n \rightarrow 0$. Ceci

montre le résultat pour T_{1n} . On fait de la même manière pour T_{2n} . On a que $T_{3n} = O_p(m/n^2)$ puisque $R_n^2 = O_p(n^{-2})$. Ceci montre (B1).

On montre maintenant (B2). On remarque que l'on peut écrire

$$\sum_{j=0}^{n-1} k_{jm}^2 \gamma_{uv}(j; \eta) [\gamma_{\hat{u}\hat{v}}(j; \eta) - \gamma_{uv}(j; \eta)] = -\sigma_u^{-1} T_{4n} - \sigma_v^{-1} T_{5n} + T_{6n},$$

où

$$\begin{aligned} T_{4n} &= \sum_{j=0}^{n-1} k_{jm}^2 \gamma_{uv}(j; \eta) [n^{-1} \sum_{t=j+1}^n \eta_{1t, t-j} \delta_{ut}], \\ T_{5n} &= \sum_{j=0}^{n-1} k_{jm}^2 \gamma_{uv}(j; \eta) [n^{-1} \sum_{t=j+1}^n \eta_{2t, t-j} \delta_{v, t-j}], \\ T_{6n} &= \sum_{j=0}^{n-1} k_{jm}^2 \gamma_{uv}(j; \eta) R_n. \end{aligned}$$

En utilisant encore une fois l'équation (4.12), on peut écrire T_{4n} de la manière suivante:

$$\begin{aligned} T_{4n} &= -\sum_{r=1}^{p_1} (\hat{\phi}_{1r} - \phi_{1r}) C_{rn} \\ &\quad + \sum_{r=1}^{q_1} (\hat{\theta}_{1r} - \theta_{1r}) D_{rn} \\ &\quad - \sum_{j=0}^{n-1} k_{jm}^2 \gamma_{uv}(j; \eta) [n^{-1} \sum_{t=j+1}^n \eta_{1t, t-j} \lambda_{unt}], \end{aligned}$$

où

$$\begin{aligned} C_{rn} &= \sum_{j=0}^{n-1} k_{jm}^2 \gamma_{uv}(j; \eta) [n^{-1} \sum_{t=j+1}^n \eta_{1t, t-j} (\Pi_1(B) X_{t-r})], \\ D_{rn} &= \sum_{j=0}^{n-1} k_{jm}^2 \gamma_{uv}(j; \eta) [n^{-1} \sum_{t=j+1}^n \eta_{1t, t-j} (\Pi_1(B) u_{t-r})]. \end{aligned}$$

En utilisant l'inégalité de Minkovski, on a que

$$\begin{aligned} E|C_{rn}| &\leq \sum_{j=0}^{n-1} k_{jm}^2 E|\gamma_{uv}(j; \eta) [n^{-1} \sum_{t=j+1}^n \eta_{1t, t-j} (\Pi_1(B) X_{t-r})]| \\ &\leq \sum_{j=0}^{n-1} k_{jm}^2 [E(\gamma_{uv}^2(j; \eta))]^{1/2} [E(n^{-1} \sum_{t=j+1}^n \eta_{1t, t-j} (\Pi_1(B) X_{t-r}))^2]^{1/2}. \end{aligned}$$

On étudie maintenant le carré de $n^{-1} \sum_{t=j+1}^n \eta_{1t,t-j}(\Pi_1(B)X_{t-r})$. On a alors

$$\begin{aligned} (n^{-1} \sum_{t=j+1}^n \eta_{1t,t-j}(\Pi_1(B)X_{t-r}))^2 &= n^{-2} \sum_{t=j+1}^n \eta_{1t,t-j}^2(\Pi_1(B)X_{t-r})^2 + \\ &2n^{-2} \sum_{t=j+1}^n \sum_{s=j+2}^{t-1} (\eta_{1s,s-j} \eta_{1t,t-j})(\{\Pi_1(B)X_{s-r}\} \{\Pi_1(B)X_{t-r}\}). \end{aligned}$$

Ainsi, en prenant l'espérance, on montre que $E[n^{-1} \sum_{t=j+1}^n \eta_{1t,t-j}(\Pi_1(B)X_{t-r})]^2 = O(n^{-1})$ uniformément en j . On fait de même pour le terme impliquant D_{rn} . On montre aussi facilement que

$$\sum_{j=0}^{n-1} k_{jm}^2 \gamma_{uv}(j; \eta) [n^{-1} \sum_{t=j+1}^n \eta_{1t,t-j} \lambda_{unt}] = o_p(\sqrt{m}/n).$$

Ainsi, $T_{4n} = O_p(m/n^{3/2}) = o_p(\sqrt{m}/n)$ puisque $m/n \rightarrow 0$. On fait T_{5n} de la même manière. Ceci complète la preuve de la seconde étape.

Combinant les résultats de la première et de la seconde étape, nous avons montré le théorème.

BIBLIOGRAPHIE

- Abraham, B. et Ledolter, J. (1983), *Statistical Methods for Forecasting*, Wiley:New-York.
- Akaike, H. et Kitagawa, G. (1999), *The practice of time series analysis*, Springer-Verlag:Berlin.
- Ansley, C. F. (1980), 'Computation of the theoretical autocovariance function for a vector ARMA process', *Journal of Statistical Computation and Simulation* **12**, 15–24.
- Ben, M. G., Martinez, E. J. et Yohai, V. J. (1999), 'Robust estimation in vector autoregressive moving-average models', *Journal of Time Series Analysis* **20**, 381–399.
- Boudjellaba, H., Dufour, J.-M. et Roy, R. (1992), 'Testing causality between two vectors in multivariate autoregressive moving average models', *Journal of American Statistical Association* **87**, 1082–1090.
- Boudjellaba, H., Dufour, J.-M. et Roy, R. (1994), 'Simplified conditions for noncausality between vectors in multivariate ARMA models', *Journal of Econometrics* **63**, 271–287.
- Box, G. E. P. et Pierce, D. A. (1970), 'Distribution of residual autocorrelations in autoregressive-integrated moving average time series models', *Journal of American Statistical Association* **65**, 1509–1526.
- Brockwell, P. J. et Davis, R. A. (1991), *Time Series. Theory and Methods*, second edn, Springer-Verlag:Berlin.
- Brockwell, P. J. et Davis, R. A. (1996), *Introduction to Time Series and Forecasting*, Springer-Verlag:Berlin.
- Brown, B. M. (1971), 'Martingale central limit theorems', *Annals of Mathematical Statistics* **42**, 59–66.
- Bustos, O. H. et Yohai, V. J. (1986), 'Robust estimates for ARMA models', *Journal of American Statistical Association* **81**, 155–168.
- Chan, W.-Y. T. et Wallis, K. F. (1978), 'Multiple time series modelling: Another look at the mink-muskrat interaction', *Applied Statistics* **27**, 168–175.
- Derby, L. et Martin, R. D. (1979), 'Robust estimation of the first-order autoregressive parameter', *Journal of American Statistical Association* **74**, 140–146.
- El Himdi, K. et Roy, R. (1997), 'Tests for noncorrelation of two multivariate ARMA time series', *The Canadian Journal of Statistics* **25**, 233–256.
- Fox, A. J. (1972), 'Outliers in time series', *Journal of the Royal Statistical Society, B* **34**, 350–363.
- Fuller, W. A. (1996), *Introduction to Statistical Time Series*, second edn, Wiley:New-York.
- Granger, C. W. J. (1969), 'Investigating causal relations by econometric models and cross-spectral methods', *Econometrica* **37**, 424–438.
- Hallin, M., Jurečková, J., Pícek, J. et Zahaf, T. (1999), 'Nonparametric tests of independence of two autoregressive time series based on autoregression rank scores', *Journal of Statistical Planning and Inference* **75**, 319–330.
- Haugh, L. D. (1976), 'Checking the independence of two covariance-stationary time series: a univariate residual cross-correlation approach', *Journal of American Statistical Association* **71**, 378–385.

- Hong, Y. (1996a), 'A separate mathematical appendix for 'Testing for independence between two covariance stationary time series'.
- Hong, Y. (1996b), 'Testing for independence between two covariance stationary time series', *Biometrika* **83**, 615–625.
- Hosking, J. (1980), 'The multivariate portmanteau statistic', *Journal of American Statistical Association* **75**, 602–608.
- Judge, G. G., Hill, R. C., Grittiths, W. E., Lütkepohl, H. et Lee, T.-C. (1985), *The theory and practice of econometrics*, second edn, Wiley:New-York.
- Li, W. K. (1988), 'A goodness-of-fit test in robust time series modelling', *Biometrika* **75**, 355–361.
- Li, W. K. et Hui, Y. V. (1994), 'Robust residual cross correlation tests for lagged relations in time series', *Journal Statistical Computation and Simulation* **49**, 103–109.
- McLeod, A. I. (1979), 'Distribution of the residual cross-correlation in univariate ARMA time series models', *Journal of American Statistical Association* **74**, 849–855.
- Pierce, D. A. (1977), 'Relationships – and the lack thereof – between economic time series, with special reference to money and interest rates', *Journal of the American Statistical Association* **72**, 11–22. (C/R: p22-26).
- Pierce, D. A. et Haugh, L. D. (1977), 'Causality in temporal systems: characterizations and survey', *Journal of Econometrics* **5**, 265–293.
- Priestley, M. B. (1981a), *Univariate Series*, Vol. 1 of *Spectral Analysis and Time Series*, Academic:New-York.
- Priestley, M. B. (1981b), *Multivariate Series, Prediction and Control*, Vol. 2 of *Spectral Analysis and Time Series*, Academic:New-York.
- Rousseeuw, P. J. et Leroy, A. M. (1987), *Robust Regression and Outlier Detection*, New-York: Wiley.
- Wei, W. W. S. (1990), *Time Series Analysis. Univariate and Multivariate Methods*, Addison-Wesley:MA.

CONCLUSION

Cette thèse tente d'apporter des contributions en théorie des sondages et dans l'analyse des séries chronologiques. Nous allons résumer les contributions que nous jugeons les plus importantes, et également indiquer quelques avenues possibles de recherche.

CHAPITRE 1: ESTIMATEURS DE CALAGE ROBUSTES

Nous avons considéré dans ce chapitre le problème des valeurs aberrantes en théorie des sondages. Le cadre général adopté était les estimateurs de calage, qui permettait d'unifier les idées sur le calage, dans un contexte de robustesse. Les contributions intéressantes de cet article sont les suivantes:

1. Un lien a pu être effectué entre les estimateurs QR et les estimateurs de calage. En fait, les estimateurs QR sont des estimateurs de calage, et ainsi la classe des estimateurs RQR a pu être proposée, afin de restreindre les poids, pour qu'ils soient tous positifs par exemple.
2. Puisque les estimateurs QR et RQR sont des estimateurs de calage, leur forme a permis d'expliquer la provenance des poids w_k attachés aux unités y_k . Souvent, les poids de départ des estimateurs de calage sont les poids d'échantillonnage d_k . Dans notre situation, ce sont les poids r_k .
3. Un rapprochement entre les estimateurs QR et RQR et les estimateurs robustes proposés dans la littérature a été effectuée. Nous avons pu alors expliquer les différences et les particularités de nombreux estimateurs. Nous avons considéré la situation où aucune information auxiliaire est disponible, ainsi que lorsque nous disposons d'une telle information auxiliaire.
4. Nous avons pu proposer l'idée suivante: transformer en poids calés des poids implicites robustes judicieusement choisis. C'est l'idée possiblement la plus intéressante de l'article.
5. Dans l'étude empirique, les estimateurs robustes proposés se comparent plutôt bien avec les estimateurs robustes déjà existants. Dans le complément, nous avons étudié les estimateurs robustes proposés sous un plan d'échantillonnage Poisson séquentiel.

Des avenues intéressantes de recherche pourraient être les suivantes. Voir aussi la conclusion de l'essai 1.

1. Un estimateur robuste a été proposé, reposant sur un choix de (q_k, r_k) . Il serait intéressant de s'intéresser au choix optimal de ces constantes, surtout en ce qui a trait au choix optimal de r_k .
2. Les études empiriques de l'essai 1 et de l'annexe pourraient être approfondies. En particulier, l'influence des diverses constantes dans ces estimateurs pourrait être étudiée par simulation, sous différents plans d'échantillonnage.
3. Puisqu'en pratique nous avons plusieurs variables à l'étude, développer des estimateurs robustes pour les sondages tout usage est une autre avenue intéressante de recherche.

CHAPITRE 2: A NOTE ON JACKKNIFE VARIANCE ESTIMATION FOR THE GENERAL REGRESSION ESTIMATOR

Dans le deuxième chapitre, des formules explicites pour les estimateurs de variance jackknife en utilisant la technique des groupes aléatoires sont dégagées. En plus de l'analyse de la sur-estimation de la variance, la formule offre des avantages de calcul. Comme avenue de recherche, il serait intéressant d'établir un lien entre les formules obtenues et les pseudo-observations intervenants habituellement dans le calcul des estimateurs de variance jackknife.

CHAPITRE 3: TESTS DE CORRÉLATION SÉRIELLE DANS LES MODÈLES VARX

Dans le troisième essai, nous avons étudié des tests de corrélation sérielle dans les modèles multidimensionnels autorégressifs linéaires dynamiques avec variables explicatives (VARX). Une hypothèse importante assurant par exemple que les paramètres estimés par la méthode des moindres carrés soient convergents est l'indépendance des erreurs. Pour vérifier cette hypothèse, Hong (1996a) a développé une statistique de test dans les modèles ARX. Nous avons pu généraliser l'approche dans un contexte multidimensionnel. Comme avenue de recherche poursuivant le travail effectué dans le chapitre 3, nous pourrions considérer l'élaboration de tests de corrélation sérielle robuste dans les modèles ARX. En effet, la fonction d'autocovariance utilisée était la fonction d'autocovariance usuelle et les estimateurs considérés étaient les estimateurs des moindres carrés. Ainsi, la procédure proposée risque d'être très sensible à la présence de valeurs aberrantes. Ainsi, il faudrait développer des estimateurs robustes dans les modèles ARX suivant des techniques similaires à celles de Bustos et Yohai, qui sous certaines conditions devraient être convergents et asymptotiquement normaux. Par la suite,

avec la fonction d'autocovariance robuste de Bustos et Yohai, il devrait être possible de développer une statistique portmanteau pondérée à la manière de Hong. La puissance de la nouvelle statistique devrait être étudiée et comparée à la statistique de test de Li. Dans un second temps, il devrait être possible de développer une approche multidimensionnelle, ce qui est toujours intéressant puisqu'en pratique nous avons souvent à analyser plusieurs variables simultanément. Li et Hui (1989) ont considéré une telle approche généralisant l'approche de Li (1988), mais avec des statistiques reposant sur un nombre fixé de délais. Des travaux récents, illustrant bien l'importance du sujet, sur l'estimation robuste de modèles VARMA par Ben et al. (1999) devraient être utiles. Une approche multidimensionnelle telle que décrite ici devrait permettre d'obtenir des propriétés intéressantes de puissance.

CHAPITRE 4: TESTS D'INDÉPENDANCE ROBUSTES ENTRE DEUX SÉRIES CHRONOLOGIQUES

Dans le quatrième essai, le problème d'indépendance de deux séries a été considéré, dans un contexte où il pourrait y avoir des valeurs aberrantes. Puisque les statistiques de Haugh et Hong sont susceptibles d'être sensibles dans ce contexte, une statistique de test robuste a été proposée pour vérifier l'indépendance de deux séries qui généralise celle de Hong et celle de Li et Hui. Elle repose sur l'ensemble des délais et offre de meilleures propriétés de puissance. Comme avenue de recherche, il serait intéressant de considérer des tests d'indépendance robustes entre deux séries multidimensionnelles stationnaires ou non stationnaires, généralisant ainsi El Himdi et Roy (1997) et Pham, Roy et Cédras (2000).

BIBLIOGRAPHIE

- Ben, M. G., Martinez, E. J. et Yohai, V. J. (1999), 'Robust estimation in vector autoregressive moving-average models', *Journal of Time Series Analysis* **20**, 381–399.
- Box, G. E. P. et Pierce, D. A. (1970), 'Distribution of residual autocorrelations in autoregressive-integrated moving average time series models', *Journal of American Statistical Association* **65**, 1509–1526.
- Bustos, O. H. et Yohai, V. J. (1986), 'Robust estimates for ARMA models', *Journal of American Statistical Association* **81**, 155–168.
- Chambers, R. L. (1986), 'Outlier robust finite population estimation', *Journal of the American Statistical Association* **81**, 1063–1069.
- Chambers, R. L. et Kokic, P. N. (1993), An integrated approach for the treatment of outliers in sub-annual surveys, in 'Proceedings on the 49th Session, International Statistical Institute'.
- Deville, J.-C. et Särndal, C. E. (1992), 'Calibration estimators in survey sampling', *Journal of the American Statistical Association* **87**, 376–382.
- Donoho, D. L. et Huber, P. J. (1983), The notion of breakdown point, in P. J. Bickel, K. A. Doksum et J. L. Hodges, eds, 'A Festschrift for Erich Lehmann', Belmont, CA: Wadsworth.
- Duchesne, P. (1999), 'Robust calibration estimators', *Survey Methodology* **25**, 43–56.
- El Himdi, K. et Roy, R. (1997), 'Tests for noncorrelation of two multivariate ARMA time series', *The Canadian Journal of Statistics* **25**, 233–256.
- Fuller, W. A., Loughin, M. M. et Baker, H. D. (1994), 'Regression weighting in the presence of nonresponse with application to the 1987-1988 nationwide food consumption survey', *Survey Methodology* **20**, 75–85.
- Gambino, J. (1987), Dealing with outliers: A look at some methods used at statistics canada, Technical report, Business Survey Methods Division, Statistics Canada. Technical Report prepared for the Fifth Meeting of the Advisory Committee on Statistical Methods.
- Granger, C. W. J. (1969), 'Investigating causal relations by econometric models and cross-spectral methods', *Econometrica* **37**, 424–438.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. et Stahel, W. A. (1986), *Robust Statistics: The Approach Based on Influence Functions*, New-York: Wiley.
- Haugh, L. D. (1976), 'Checking the independance of two covariance-stationary time series: a univariate residual cross-correlation approach', *Journal of American Statistical Association* **71**, 378–385.
- Hidiroglou, M. A. et Srinath, K. P. (1981), 'Some estimators of the population total from simple random samples containing large units', *Journal of the American Statistical Association* **76**, 690–695.
- Hong, Y. (1996a), 'Consistent testing for serial correlation of unknown form', *Econometrica* **64**, 837–864.

- Hong, Y. (1996b), 'Testing for independence between two covariance stationary time series', *Biometrika* **83**, 615–625.
- Hosking, J. (1980), 'The multivariate portmanteau statistic', *Journal of American Statistical Association* **75**, 602–608.
- Kish, L. (1965), *Survey Sampling*, New-York: Wiley.
- Lee, H. (1991), Model-based estimators that are robust to outliers, in 'Proceedings of the 1991 Annual Research Conference'. U. S. Bureau of Census.
- Lee, H. (1995), Outliers in business surveys, in Cox, Binder, Chinnappa, Christiansono, Colledge et Kott, eds, 'Business Surveys Methods', New-York: Wiley.
- Lee, H., Ghangurbe, P. D., Mach, L. et Yung, W. (1992), 'Outliers in sample surveys'. Methodology Branch Working Paper BSMD-92-008E, Ottawa: Statistics Canada.
- Li, W. K. (1988), 'A goodness-of-fit test in robust time series modelling', *Biometrika* **75**, 355–361.
- Li, W. K. et Hui, Y. V. (1989), 'Robust multiple time series modelling', *Biometrika* **76**, 309–315.
- Li, W. K. et Hui, Y. V. (1994), 'Robust residual cross correlation tests for lagged relations in time series', *Journal Statistical Computation and Simulation* **49**, 103–109.
- McLeod, A. I. (1979), 'Distribution of the residual cross-correlation in univariate ARMA time series models', *Journal of American Statistical Association* **74**, 849–855.
- Paparoditis, E. (1999), 'Spectral density based goodness-of-fit tests for time series models'. À paraître dans *Scandinavian Journal of Statistics*.
- Paparoditis, E. (2000), 'On some power properties of goodness-of-fit tests in time series analysis'. À paraître dans a Festschrift by Chapman and Hall.
- Pham, D. T., Roy, R. et Cédras, L. (2000), 'Tests of non-correlation of two cointegrated ARMA time series'. Cahier de recherche CRM-2649, Centre de recherches mathématiques, Université de Montréal.
- Rousseeuw, P. J. et Leroy, A. M. (1987), *Robust Regression and Outlier Detection*, New-York: Wiley.
- Särndal, C. E. (1996), 'Efficient estimators with simple variance in unequal probability sampling', *Journal of the American Statistical Association* **91**, 1289–1300.
- Särndal, C. E., Swensson, B. et Wretman, J. H. (1989), 'The weighted residual technique for estimating the variance of the general regression estimator of the finite population total', *Biometrika* **76**, 527–537.
- Särndal, C. E., Swensson, B. et Wretman, J. H. (1992), *Model Assisted Survey Sampling*, New-York: Springer-Verlag.
- Simpson, D. G. et Chang, Y.-C. I. (1997), 'Reweighted approximate GM-estimators: asymptotics and residual-based graphics', *Journal of Statistical Planning and Inference* **57**, 273–293.
- Singh, A. C. et Mohl, C. A. (1996), 'Understanding calibration estimators in survey sampling', *Survey Methodology* **22**, 107–115.
- Stukel, D. M., Hidioglou, M. A. et Särndal, C. E. (1996), 'Variance estimation for calibration estimators: A comparison of jackknifing versus Taylor linearization', *Survey Methodology* **22**, 117–125.
- Wolter, K. M. (1985), *Introduction to Variance Estimation*, New-York: Springer-Verlag.
- Wright, R. L. (1983), 'Finite population sampling with multivariate auxiliary information', *Journal of the American Statistical Association* **78**, 879–884.