

**Université de Montréal**

**Sur les tests de type diagnostic dans la  
validation des hypothèses de bruit blanc et de  
non corrélation**

par

**Joël Sango**

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

28 septembre 2016

© Joël Sango, 2016

# Université de Montréal

Faculté des études supérieures

Cette thèse intitulée

## Sur les tests de type diagnostic dans la validation des hypothèses de bruit blanc et de non corrélation

présentée par

**Joël Sango**

a été évaluée par un jury composé des personnes suivantes :

*Manuel Morales*

---

(président-rapporteur)

*Pierre Duchesne*

---

(directeur de recherche)

*Maciej Augustyniak*

---

(membre du jury)

*Lynda Khalaf*

---

(examineur externe)

*Benoît Perron*

---

(représentant du doyen de la FAS)

Thèse acceptée le  
*28 septembre 2016*

---

## DÉDICACES

---

À mes parents, et à tous ceux que je ne nomme pas, mais qui se reconnaîtront.

## SOMMAIRE

---

Dans la modélisation statistique, nous sommes le plus souvent amené à supposer que le phénomène étudié est généré par une structure pouvant s'ajuster aux données observées. Cette structure fait apparaître une partie principale qui représente le mieux possible le phénomène étudié et qui devrait expliquer les données et une partie supposée négligeable appelée erreur ou innovation. Cette structure complexe est communément appelée un modèle, dont la forme peut être plus ou moins complexe. Afin de simplifier la structure, il est souvent supposé qu'elle repose sur un nombre fini de valeurs, appelées paramètres. Basé sur les données, ces paramètres sont estimés avec ce que l'on appelle des estimateurs. La qualité du modèle pour les données à notre disposition est également fonction des estimateurs et de leurs propriétés, par exemple, est-ce que les estimateurs sont raisonnablement proches des valeurs idéales, c'est-à-dire les vraies valeurs. Des questions d'importance portent sur la qualité de l'ajustement d'un modèle aux données, ce qui se fait par l'étude des propriétés probabilistes et statistiques du terme d'erreur. Aussi, l'étude des relations ou l'absence de ces dernières entre les phénomènes sous des hypothèses complexes sont aussi d'intérêt. Des approches possibles pour cerner ce genre de questions consistent dans l'utilisation des tests portemanteaux, dits également tests de diagnostic.

La thèse est présentée sous forme de trois projets.

Le premier projet est rédigé en langue anglaise. Il s'agit en fait d'un article actuellement soumis dans une revue avec comité de lecture. Dans ce projet, nous étudions le modèle vectoriel à erreurs multiplicatives (vMEM) pour lequel nous utilisons les propriétés des estimateurs des paramètres du modèle selon la méthode

des moments généralisés (GMM) afin d'établir la distribution asymptotique des autocovariances résiduelles. Ceci nous permet de proposer des nouveaux tests diagnostiques pour ce type de modèle. Sous l'hypothèse nulle d'adéquation du modèle, nous montrons que la statistique usuelle de Hosking-Ljung-Box converge vers une somme pondérée de lois de khi-carré indépendantes à un degré de liberté. Un test généralisé de Hosking-Ljung-Box est aussi obtenu en comparant la densité spectrale des résidus de l'estimation et celle présumée sous l'hypothèse nulle. Un avantage des tests spectraux est qu'ils nécessitent des estimateurs qui convergent à la vitesse  $n^{-1/2}$  où  $n$  est la taille de l'échantillon, et leur utilisation n'est pas restreinte à une technique particulière, comme par exemple la méthode des moments généralisés.

Dans le deuxième projet, nous établissons la distribution asymptotique sous l'hypothèse de faible dépendance des covariances croisées de deux processus stationnaires en covariance. La faible dépendance ici est définie en terme de l'effet limité d'une observation donnée sur les observations futures. Nous utilisons la notion de stabilité et le concept de contraction géométrique des moments. Ces conditions sont plus générales que celles de l'invariance des moments conditionnels d'ordre un à quatre utilisée jusque là par plusieurs auteurs. Un test statistique basé sur les covariances croisées et la matrice des variances et covariances de leur distribution asymptotique est alors proposé et sa distribution asymptotique établie. Dans l'implémentation du test, la matrice des variances et covariances des covariances croisées est estimée à l'aide d'une procédure autorégressive vectorielle robuste à l'autocorrélation et à l'hétéroscédasticité. Des simulations sont ensuite effectuées pour étudier les propriétés du test proposé.

Dans le troisième projet, nous considérons un modèle périodique multivarié et cointégré. La présence de cointégration entraîne l'existence de combinaisons linéaires périodiquement stationnaires des composantes du processus étudié. Le nombre de ces combinaisons linéaires linéairement indépendantes est appelé rang

de cointégration. Une méthode d'estimation en deux étapes est considérée. La première méthode est appelée estimation de plein rang. Dans cette approche, le rang de cointégration est ignoré. La seconde méthode est appelée estimation de rang réduit. Elle tient compte du rang de cointégration. Cette dernière est une approche non linéaire basée sur des itérations dont la valeur initiale est l'estimateur de plein rang. Les propriétés asymptotiques de ces estimateurs sont aussi établies. Afin de vérifier l'adéquation du modèle, des statistiques de test de type portemanteau sont considérées et leurs distributions asymptotiques sont étudiées. Des simulations sont par la suite présentées afin d'illustrer le comportement du test proposé.

#### **MOTS CLÉS :**

Modèles périodiques, estimation de plein rang, estimation de rang réduit, statistiques portemanteaux, faible dépendance, stabilité, contraction géométrique des moments, modèles vectoriels à erreurs multiplicatives, moyenne conditionnelle.

## SUMMARY

---

In statistical modeling, we assume that the phenomenon of interest is generated by a model that can be fitted to the observed data. The part of the phenomenon not explained by the model is called error or innovation. There are two parts in the model. The main part is supposed to explain the observed data, while the unexplained part which is supposed to be negligible is also called error or innovation. In order to simplify the structures, the model are often assumed to rely on a finite set of parameters. The quality of a model depends also on the parameter estimators and their properties. For example, are the estimators relatively close to the true parameters? Some questions also address the goodness-of-fit of the model to the observed data. This question is answered by studying the statistical and probabilistic properties of the innovations. On the other hand, it is also of interest to evaluate the presence or the absence of relationships between the observed data. Portmanteau or diagnostic type tests are useful to address such issue.

The thesis is presented in the form of three projects.

The first project is written in English as a scientific paper. It was recently submitted for publication. In that project, we study the class of vector multiplicative error models (vMEM). We use the properties of the Generalized Method of Moments to derive the asymptotic distribution of sample autocovariance function. This allows us to propose a new test statistic. Under the null hypothesis of adequacy, the asymptotic distributions of the popular Hosking-Ljung-Box (HLB) test statistics are found to converge in distribution to weighted sums of independent chi-squared random variables. A generalized HLB test statistic is motivated by

comparing a vector spectral density estimator of the residuals with the spectral density calculated under the null hypothesis.

In the second project, we derive the asymptotic distribution under weak dependence of cross covariances of covariance stationary processes. The weak dependence is defined in term of the limited effect of a given information on future observations. This recalls the notion of stability and geometric moment contraction. These conditions of weak dependence defined here are more general than the invariance of conditional moments used by many authors. A test statistic based on cross covariances is proposed and its asymptotic distribution is established. In the elaboration of the test statistics, the covariance matrix of the cross covariances is obtained from a vector autoregressive procedure robust to autocorrelation and heteroskedasticity. Simulations are also carried on to study the properties of the proposed test and also to compare it to existing tests.

In the third project, we consider a cointegrated periodic model. Periodic models are present in the domain of meteorology, hydrology and economics. When modelling many processes, it can happen that the processes are just driven by a common trend. This situation leads to spurious regressions when the series are integrated but have some linear combinations that are stationary. This is called cointegration. The number of stationary linear combinations that are linearly independent is called cointegration rank. So, to model the real relationship between the processes, it is necessary to take into account the cointegration rank. In the presence of periodic time series, it is called periodic cointegration. It occurs when time series are periodically integrated but have some linear combinations that are periodically stationary. A two step estimation method is considered. The first step is the full rank estimation method that ignores the cointegration rank. It provides initial estimators to the second step estimation which is the reduced rank estimation. It is non linear and iterative. Asymptotic properties of the estimators are also established. In order to check for model adequacy, portmanteau



type tests and their asymptotic distributions are also derived and their asymptotic distribution are studied. Simulation results are also presented to show the behaviour of the proposed test.

**Keywords.**

Periodic models, full rank estimation, reduced rank estimation, portmanteau test statistics, weak dependence, stability, geometric moment contraction, vector multiplicative error model, conditional mean.

# TABLE DES MATIÈRES

---

Dédicaces.....	iii
Sommaire.....	iv
Summary.....	vii
Liste des tableaux .....	xiv
Remerciements .....	xvi
<b>Chapitre 1. Préliminaires.....</b>	<b>1</b>
1.1. Racine unitaire et intégration .....	4
1.2. Cointégration.....	7
1.3. Les modèles périodiques .....	8
1.4. Les modèles non linéaires .....	10
1.5. La validation des modèles.....	11
1.6. Objectifs de recherche et contributions .....	13
<b>Chapitre 2. Evaluating vector multiplicative error models with the Hosking-Ljung-Box portmanteau test and kernel-based test statistics .....</b>	<b>15</b>
2.1. Introduction.....	17

2.2.	Some preliminaries and definition and estimation methods in vMEM models.....	21
2.2.1.	The vMEM time series models.....	21
2.2.2.	Estimation procedures of vMEM models.....	23
2.2.3.	Spectral density estimation of the innovation process.....	25
2.3.	The test statistics and their asymptotic distributions.....	26
2.3.1.	The Ng-Li-Yu test statistic and the Hosking-Ljung-Box portmanteau test.....	26
2.3.2.	Kernel-based test statistics.....	30
2.4.	Simulation experiments.....	34
2.4.1.	Models and tests statistics included in the Monte Carlo study..	34
2.4.2.	Discussion of the results.....	39
2.5.	An empirical application.....	43
2.6.	Discussion and conclusion.....	49
	Acknowledgement.....	50
	Appendix. Proof of Theorem 1.....	50
<b>Chapitre 3.</b>	<b>Tests de non corrélation sous faible dépendance entre deux processus stationnaires en covariance.....</b>	<b>55</b>
3.1.	Introduction.....	55
3.2.	Préliminaires.....	59
3.3.	Distribution limite des covariances croisées.....	62

3.4.	La statistique de test .....	66
3.5.	Estimation de la matrice des variances et covariances.....	67
3.6.	Simulations .....	70
3.7.	Conclusion.....	72
	Annexe.....	74
<b>Chapitre 4.</b>	<b>Sur l'estimation et la validation des modèles périodiques multivariés et cointégrés.....</b>	<b>82</b>
4.1.	Introduction .....	82
4.2.	Préliminaires .....	86
4.2.1.	Racine unitaire et intégration .....	89
4.3.	Modélisation de la cointégration dans les PVARs.....	90
4.3.1.	Formulation du modèle à correction d'erreurs .....	90
4.3.2.	Estimation de plein rang.....	96
4.3.3.	Estimation de rang réduit.....	110
4.3.3.1.	Distribution asymptotique de l'estimateur de rang réduit... ..	113
4.3.3.2.	Estimateurs initiaux et leur distribution limite .....	116
4.4.	Tests Diagnostiques sur le modèle estimé.....	119
4.4.1.	Distribution asymptotique des autocovariances résiduelles .....	119
4.4.2.	Simulations.....	123
4.5.	conclusion .....	126
	Annexes .....	128

<b>Conclusion.....</b>	<b>153</b>
<b>Bibliographie .....</b>	<b>158</b>

## LISTE DES TABLEAUX

---

- 2.1 Model parameters for the time series models  $DGP_0$  and  $DGP_1$  defined by (2.4.1) and (2.4.3), respectively. .... 34
- 2.2 Empirical levels (nominal levels 5% and 10%) of the tests statistics for checking the vMEM(1,1) model defined as (2.4.1). The following test statistics are included :  $T_n(k)$  and  $T_n^\dagger(k)$  for various kernels,  $HLB_n(m)$  with different decision rules,  $NLY_n(m)$  and  $NLY_n^-(m)$ . The lag order/smoothing parameters are  $m = m_n = [\log(n)], [3n^{0.2}], [3n^{0.3}]$ , with sample sizes  $n = 300, 400$  and  $500$ . .... 35
- 2.3 Exact powers (nominal levels 5% and 10%) of the tests statistics for checking a vMEM(1,1) model defined as (2.4.1) when the true model is  $DGP_1$ , based on asymptotic or approximate quantiles. The following test statistics are included :  $T_n(k)$  and  $T_n^\dagger(k)$  for various kernels,  $HLB_n(m)$  with different decision rules,  $NLY_n(m)$  and  $NLY_n^-(m)$ . The lag order/smoothing parameters are  $m = m_n = [\log(n)], [3n^{0.2}], [3n^{0.3}]$ , with  $n = 400, 500, 700$  and  $1000$ . .... 38
- 2.4 Exact powers (nominal levels 5% and 10%) of the tests statistics for checking a vMEM(1,1) model defined as (2.4.1) when the true model is  $DGP_1$ , based on empirical quantiles found in the level study. The following test statistics are included :  $T_n(k)$  and  $T_n^\dagger(k)$  for various kernels,  $HLB_n(m)$  with different decision rules,  $NLY_n(m)$  and  $NLY_n^-(m)$ . The lag order/smoothing parameters are  $m = m_n = [\log(n)], [3n^{0.2}], [3n^{0.3}]$ , with  $n = 400, 500, 700$  and  $1000$ . .... 41

2.5	Estimation results for modelling the variables (2.5.1) of the FTSE100 stock indice. Two models are estimated, the vMEM(1,1) model defined by (2.5.2) and the augmented model vMEM(2,1) (2.5.2). The estimators and the associated standard errors are given.....	46
2.6	Autocorrelations for lags one to five and asymptotic variances (in parentheses) from the adjustment of the vMEM(1,1) model for modelling the variables (2.5.1) of the FTSE100 stock index. ....	46
2.7	$P$ -values of the different tests statistics for the adjustment of the vMEM(2,1) model (2.5.3) for modelling the variables (2.5.1) of the FTSE100 stock index. ....	49
3.1	Niveau empirique des statistiques de test au seuil de 5%.....	71
3.2	Niveau empirique des statistiques de test au seuil de 10%.....	72
4.1	Valeurs des paramètres pour le premier modèle.....	124
4.2	Valeurs des paramètres pour le deuxième modèle.....	124
4.3	Niveaux empiriques en pourcentage pour la statistique portmanteau $\hat{Q}_H(\nu)$ définie en (4.4.6)et la version modifiée $\hat{Q}_H^*(\nu)$ définie en (4.4.7) pour le modèle 1. ....	125
4.4	Niveaux empiriques en pourcentage pour la statistique portmanteau $\hat{Q}_H(\nu)$ définie en (4.4.6)et la version modifiée $\hat{Q}_H^*(\nu)$ définie en (4.4.7) pour le modèle 2. ....	126

## REMERCIEMENTS

---

Je tiens à exprimer mes plus vifs remerciements au professeur Pierre Duchesne pour son attention et sa disponibilité malgré ses nombreuses charges. Les commentaires, les séances de discussion et les interactions ont eu un impact direct sur la forme et la qualité de ce travail de recherche. Sa compétence, sa rigueur scientifique et sa clairvoyance ont été pour moi une source d'inspiration et resteront des moteurs de mon travail de chercheur.

J'exprime tous mes remerciements à l'ensemble des membres de mon jury.

J'adresse toute ma gratitude aux professeurs, aux étudiants et au personnel non enseignant du département de mathématiques et de statistique de l'Université de Montréal qui m'ont aidé et soutenu que ce soit directement ou indirectement tout au long de ce travail.

Je remercie particulièrement le professeur Roch Roy pour son appui lors de mes études doctorales de même que le département de mathématiques et de statistique pour la mise à la disposition des étudiants d'un environnement propice à l'épanouissement intellectuel.

J'adresse toute mon affection à ma famille, et en particulier à ma maman qui par son soutien moral sa tendresse et son amour m'a aidé à comprendre que les problèmes de la vie ne se résolvent pas toujours par modèles de mathématique et de statistique. Mes remerciements s'adressent tout particulièrement à mon épouse Clarisse et mes deux fils Yvan et Steve pour m'avoir soutenu pendant les moments difficiles.



Une pensée pour terminer ces remerciements pour toi qui n'a pas vu l'aboutissement de mon travail mais je sais que tu aurais été très fier de ton fils.

# Chapitre 1

---

## PRÉLIMINAIRES

Quelque soit le domaine d'étude, une des principales préoccupations repose sur la compréhension de la dynamique des phénomènes d'intérêt. Dans notre contexte, les données étudiées sont observées à différents instants consécutifs dans le temps. Une série chronologique est une suite de données numériques, mesurées à des intervalles réguliers et représentant l'évolution d'une quantité spécifique dans le temps. L'étude des séries chronologiques a connu une évolution considérable ces dernières décennies et chaque fois, le souci est de se rapprocher du meilleur modèle possible, en incorporant les meilleures approches et techniques disponibles.

Cette modélisation comporte généralement trois étapes : l'identification du modèle, l'estimation des paramètres et la validation du modèle. L'identification du modèle consiste à déterminer le type de modèle qui s'ajuste le mieux aux données. Dans l'étude des séries chronologiques, il existe plusieurs types de modèles que nous pouvons tout simplement classer en modèles linéaires et non linéaires. Les trois étapes de modélisation reposent sur un présupposé particulièrement important, l'hypothèse de stationnarité du processus stochastique dont la série chronologique est une réalisation finie.

En fait, la stationnarité est une hypothèse qui simplifie grandement l'analyse des séries temporelles. En effet, il est alors possible d'introduire la fonction d'autocovariance et l'inférence statistique de ces autocovariances est également

possible. Telle qu'indiquée dans les définitions ci-dessous, la notion de stationnarité stipule que certaines caractéristiques du processus sous-jacent sont invariantes par rapport au temps.

**Définition 1.1.** *Le processus stochastique  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  de dimension  $d$ ,  $\mathbf{Y}_t = (Y_t(1), \dots, Y_t(d))^\top$  est dit strictement stationnaire (au sens fort) si et seulement si les distributions conjointes des vecteurs  $(\mathbf{Y}_{t_1}^\top, \mathbf{Y}_{t_2}^\top, \dots, \mathbf{Y}_{t_k}^\top)^\top$  et  $(\mathbf{Y}_{t_1+l}^\top, \mathbf{Y}_{t_2+l}^\top, \dots, \mathbf{Y}_{t_k+l}^\top)^\top$  sont identiques pour tout entier  $k$  positif et pour tout  $t_1, t_2, \dots, t_k, l \in \mathbb{Z}$ .*

Cette définition de la stationnarité repose sur l'invariance des distributions, mais l'hypothèse de stationnarité stricte est restrictive sur le plan pratique. Le concept de stationnarité le plus utile dans les applications est la stationnarité au sens large.

**Définition 1.2.** *Le processus stochastique  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  de dimension  $d$ ,  $\mathbf{Y}_t = (Y_t(1), \dots, Y_t(d))^\top$  est dit stationnaire au sens large ou faiblement stationnaire (ou stationnaire tout court) si et seulement si :*

- (i)  $E[Y_t^2(j)] < \infty, j = 1, \dots, d, \forall t,$
- (ii)  $E[\mathbf{Y}_t] = \boldsymbol{\mu}, \forall t,$
- (iii)  $E[(\mathbf{Y}_{t+k} - \boldsymbol{\mu})(\mathbf{Y}_t - \boldsymbol{\mu})^\top] = \boldsymbol{\Gamma}_k, \forall t, \forall k \in \mathbb{Z}.$

De même que le concept de stationnarité, la notion de bruit blanc est importante dans la modélisation des séries chronologiques.

**Définition 1.3.** *Le processus  $\{\boldsymbol{\varepsilon}_t, t \in \mathbb{Z}\}$  de dimension  $d$ ,  $\boldsymbol{\varepsilon}_t = (\varepsilon_t(1), \dots, \varepsilon_t(d))^\top$  est un bruit blanc vectoriel s'il satisfait les propriétés suivantes :*

- (i)  $E[\varepsilon_t^2(j)] < \infty, j = 1, \dots, d, \forall t,$
- (ii)  $E[\boldsymbol{\varepsilon}_t] = \mathbf{0}, \forall t,$
- (iii)  $E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s^\top] = \begin{cases} \boldsymbol{\Sigma}_\varepsilon, & \text{si } s = t, \\ \mathbf{0}, & \text{sinon.} \end{cases}$

Les bruits blancs sont certainement les exemples les plus simples de processus stationnaires.

Une des approches de modélisation les plus connues des modèles de série chronologique dans la littérature est celle de Box et Jenkins (1970). Ces derniers ont développé une classe de modèles permettant de modéliser les séries chronologiques univariées potentiellement intégrées et saisonnières. Il s'agit des modèles saisonniers autorégressifs intégrés moyennes mobiles (SARIMA). Cette méthodologie consiste à transformer des variables au préalable non stationnaires en processus autorégressifs moyennes mobiles (ARMA) stationnaires par l'application d'un filtre. Cela permet de comprendre la dynamique du phénomène et de faire des prévisions. L'extension aux séries chronologiques multivariées remonte au moins aux travaux de Quenouille (1957). Des travaux utiles sont ceux de Sims et al. (1990). Les processus vectoriels saisonniers autorégressifs et moyennes mobiles sont discutés dans Reinsel (1997).

Tel que définit par Reinsel (1997, p. 219), le processus stochastique  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$ , de dimension  $d$  est un processus multiplicatif saisonnier autorégressif moyenne mobile s'il est stationnaire et vérifie :

$$\Phi(B^s) \Pi(B) \mathbf{Y}_t = \Psi(B^s) \Theta(B) \boldsymbol{\varepsilon}_t, \quad (1.0.1)$$

où  $B$  est l'opérateur retard usuel,  $B^s$  l'opérateur retard saisonnier et  $s$  représente la saison. On rappelle que l'opérateur retard  $B$  satisfait  $B^0 = I$ ,  $B\mathbf{Y}_t = \mathbf{Y}_{t-1}$  et  $B^k\mathbf{Y}_t = \mathbf{Y}_{t-k}$ . Les polynômes non saisonniers autorégressifs et moyennes mobiles sont définis respectivement par :

$$\begin{aligned} \Pi(B) &= I_d - \Pi_1 B - \dots - \Pi_p B^p, \\ \Theta(B) &= I_d - \Theta_1 B - \dots - \Theta_q B^q, \end{aligned}$$

où  $\Pi_i$ ,  $i = 1, \dots, p$  et  $\Theta_j$ ,  $j = 1, \dots, q$  sont des matrices d'ordre  $d \times d$ . Les polynômes saisonniers autorégressifs et moyennes mobiles sont respectivement définis par :

$$\Phi(B^s) = I_d - \Phi_1 B^s - \dots - \Phi_P B^{sP},$$

$$\Psi(B^s) = I_d - \Psi_1 B^s - \dots - \Psi_Q B^{sQ},$$

où  $\Phi_i$ ,  $i = 1, \dots, P$  et  $\Psi_j$ ,  $j = 1, \dots, Q$  sont des matrices d'ordre  $d \times d$ . La matrice  $I_d$  est la matrice identité d'ordre  $d$ . Le processus  $\{\epsilon_t, t \in \mathbb{Z}\}$  représente les erreurs et correspond ici à un bruit blanc de matrice des variances et covariances  $E(\epsilon_t \epsilon_t^\top) = \Sigma_\epsilon$ . Le modèle (1.0.1) est appelé SVARMA( $p, q$ )  $\times$  ( $P, Q$ ) $_s$ . En absence de saisonnalité,  $\Phi(B^s) = I_d$ ,  $\Psi(B^s) = I_d$  et le modèle (1.0.1) devient un modèle VARMA( $p, q$ ). De même, dans le cas des processus univariés, le SVARMA( $p, q$ )  $\times$  ( $P, Q$ ) $_s$  peut être interprété comme un VARMA( $p + Ps, q + Qs$ ) dans lequel les coefficients satisfont des contraintes multiplicatives (Brockwell et Davis, 2002, p. 203).

Beaucoup d'approches ont été développées pour estimer les modèles VARMA, voir par exemple Lütkepohl (2005), Brockwell et Davis (1991), Hosking (1980) et Li et McLeod (1981). L'approche des modèles SARIMA représente une façon de modéliser la saisonnalité. Cette approche présuppose qu'en appliquant le filtre de différentiation un assez grand nombre de fois, le processus résultant peut être modélisé avec les présupposés de stationnarité. En particulier, l'autocovariance de délai  $h$  est indépendante de la saison et décroît vers zéro au fur et à mesure que  $h$  devient grand. Une critique d'une telle approche est que la corrélation entre janvier et février est la même qu'entre février et mars, par exemple. Dans la section 1.3, nous aborderons l'approche reposant sur les modèles périodiques afin de modéliser les séries chronologiques saisonnières.

## 1.1. RACINE UNITAIRE ET INTÉGRATION

Dans l'étude de la stationnarité des modèles ARMA, il est bien connu qu'un processus ARMA sera stationnaire causal si toutes les racines du polynôme autorégressif sont à l'extérieur du disque unité. Lorsque des racines sont sur le disque unité, l'étude des modèles ARIMA permet de régler la non-stationnarité par la différentiation, c'est-à-dire en appliquant de manière répétée le filtre  $(1 - B)$ .

Dans le cas multivarié, le résultat concernant les conditions de la stationnarité non anticipative portent sur l'étude du déterminant de la partie autorégressive

des modèles VARMA. Lorsque les racines sont à l'extérieur du disque unité, le processus est stationnaire et causal. Lorsque les racines sont sur le disque unité, on parle alors de racines unitaires.

Les notions de racines unitaires et d'intégration sont fortement liées. L'intégration à un ordre quelconque se définit à partir de l'intégration à un ordre inférieur, le point de départ étant l'intégration à l'ordre zéro. Beaucoup de définitions de l'intégration à l'ordre zéro sont connues dans la littérature. Les plus connues sont celles données par Engle et Granger (1987, p. 252), Engle et Granger (1991, p. 3), Banerjee et al. (1993, p. 7), Hendry (1995, p. 43) et Johansen (1995, pp. 34-35). Ces définitions sont équivalentes. Les définitions présentées ci-dessous proviennent de Johansen (1995, pp. 34-35). Dans la définition 1.4, on introduit précisément un processus multivarié linéaire, alors que la définition 1.5 porte sur l'intégration d'ordre zéro.

**Définition 1.4.** Soit  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  un processus stochastique qui satisfait  $\mathbf{Y}_t = \sum_{i=0}^{\infty} \mathbf{C}_i \boldsymbol{\varepsilon}_{t-i} = \mathbf{C}(B) \boldsymbol{\varepsilon}_t$ , où  $\mathbf{C}(B) = \sum_{i=0}^{\infty} \mathbf{C}_i B^i$  et  $\{\boldsymbol{\varepsilon}_t, t \in \mathbb{Z}\}$  est un bruit blanc tel que  $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top) = \boldsymbol{\Sigma}_\boldsymbol{\varepsilon}$ . Le processus  $\{\mathbf{Y}_t\}$  est un processus linéaire si  $\sum_{i=0}^{\infty} \mathbf{C}_i z^i$  est convergente pour  $|z| \leq 1 + \delta$ , pour un certain  $\delta > 0$ .

Une des propriétés des séries chronologiques qui permet d'étudier l'intégration de façon générale est l'intégration d'ordre zéro, que l'on notera  $I(0)$ .

**Définition 1.5.** Soit un processus stationnaire  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  satisfaisant  $\mathbf{Y}_t - E(\mathbf{Y}_t) = \sum_{i=0}^{\infty} \mathbf{C}_i \boldsymbol{\varepsilon}_{t-i}$ , avec  $\{\boldsymbol{\varepsilon}_t\}$  un bruit blanc. Le processus  $\{\mathbf{Y}_t\}$  est  $I(0)$  si  $\mathbf{C} = \sum_{i=0}^{\infty} \mathbf{C}_i \neq \mathbf{0}$ .

L'intégration d'ordre supérieur fait l'objet de la définition suivante. On la trouve aussi dans Johansen (1995, p. 35).

**Définition 1.6.** Une série chronologique  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  est dite intégrée d'ordre  $k$  ou  $I(k)$  si la série  $\{\Delta^k [\mathbf{Y}_t - E(\mathbf{Y}_t)], t \in \mathbb{Z}\}$  est  $I(0)$ . On note  $\Delta^k = (1 - B)^k$  où  $B$  est l'opérateur retard.

Tout processus stationnaire n'est pas forcément  $I(0)$ . Cependant, pour qu'un processus soit  $I(0)$ , il doit d'abord être stationnaire. Dans la suite, nous allons

adopter la terminologie de Banerjee et al. (1993) en assimilant stationnarité à  $I(0)$  dans la définition de l'intégration d'ordre supérieur.

Dans le cas des processus univariés, l'existence de la racine unitaire dans la partie autorégressive indique qu'il faut différencier la série pour aboutir à la stationnarité (Box et Jenkins (1970)). Si la racine unitaire est de multiplicité  $k$ , il faut différencier  $k$  fois. En plus de l'intégration usuelle, nous avons aussi l'intégration saisonnière.

**Définition 1.7.** *Une série chronologique univariée  $\{Y_t, t \in \mathbb{Z}\}$  est dite saisonnièrement intégrée d'ordre  $(l, L)$  si la série  $\{(1 - B^s)^L(1 - B)^l(Y_t - E(Y_t)), t \in \mathbb{Z}\}$  est stationnaire.*

Cette définition découle de celle des modèles SARIMA telle qu'exprimée par Box et Jenkins (1970). L'intégration saisonnière apparaît lorsque le polynôme saisonnier possède des racines unitaires. Elle fait intervenir des racines qui en module valent un mais qui ne sont pas égales à un. La racine unitaire valant précisément 1 correspond à la fréquence zéro, c'est-à-dire au cycle normal des observations successives, tandis que les autres racines unitaires correspondent à des fréquences plus élevées. Par exemple, dans le cas des données trimestrielles, nous avons le filtre saisonnier :

$$1 - B^4 = (1 - B)(1 + B)(1 + iB)(1 - iB),$$

où  $i = \sqrt{-1}$ . Ainsi, la racine  $-1$  correspond à deux cycles par an (semestriel) et les racines complexes à un cycle par an (annuel). La littérature sur la cointégration saisonnière est assez riche. Nous pouvons citer entre autre, Kunst (1990), Canova et Hansen (1995) et Balcombe (1999), pour ne citer que ces auteurs.

La méthodologie fondamentale pour aborder l'intégration saisonnière est celle développée par Hylleberg et al. (1990), plus couramment connue sous le nom de méthode HEGY. Cette méthode permet de distinguer entre l'intégration saisonnière et la saisonnalité exprimée de manière déterministe.

## 1.2. COINTÉGRATION

Il peut arriver que des combinaisons linéaires de série chronologiques univariées soient stationnaires alors que les séries individuelles ne le sont pas. On parle alors de cointégration. Un exemple important en économie est lorsque toutes les séries univariées sont intégrées d'ordre un mais que certaines combinaisons linéaires sont stationnaires. Les problématiques entourant la cointégration sont propres aux séries chronologiques multivariées. La définition suivante de la cointégration découle de Johansen (1995, p. 38).

**Définition 1.8.** Soit  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  un processus intégré d'ordre  $k$  au sens de la définition (1.6). Le processus  $\{\mathbf{Y}_t\}$  est dit cointégré d'ordre  $(k, b)$  (on note  $CI(k, b)$ ), de vecteur de cointégration  $\boldsymbol{\beta}$  où  $\boldsymbol{\beta} \neq \mathbf{0}$ , si  $\boldsymbol{\beta}^\top \mathbf{Y}_t$  est intégré d'ordre  $k - b$ . Ceci est noté  $I(k - b)$ .

Dans la définition 1.8, on note que les vecteurs de cointégration ne sont pas uniques. Le rang de cointégration représente le nombre de relations de cointégrations qui sont linéairement indépendantes.

Une classe utile de processus  $\{\mathbf{Y}_t\}$  est celle composée de processus qui sont  $CI(1, 1)$ . Dans un tel cas, les processus univariés sont  $I(1)$  et le vecteur de cointégration est tel que  $\{\boldsymbol{\beta}^\top \mathbf{Y}_t\}$  est  $I(0)$ . En particulier, le processus  $\{\boldsymbol{\beta}^\top \mathbf{Y}_t\}$  est stationnaire.

Déterminer le rang de cointégration est une étape importante pour modéliser les processus cointégrés. Un certain nombre d'approches sont connues dans la littérature. Parmi ces dernières, on trouve la méthode en deux étapes de Engle et Granger (1987), ainsi que la méthode de Philips et Ouliaris (1990). Ces méthodes se limitent à la détection d'une seule relation de cointégration. On retrouve également les méthodes de Johansen (1988) et de Sims et al. (1990) qui permettent de détecter plusieurs relations de cointégration. Dans notre étude, les rangs de cointégration sont supposés connus, mais se situeront dans la classe des processus vectoriels périodiques.



### 1.3. LES MODÈLES PÉRIODIQUES

Toutes les séries ne peuvent être filtrées pour atteindre la stationnarité, ceci parce que la structure d'autocorrélation pourrait fort bien dépendre de la saison. Dans les modèles de Box et Jenkins de type SARIMA et VARMA, les paramètres sont les mêmes pour toutes les saisons. Dans certaines situations, il est préférable de permettre de les laisser varier afin de mieux décrire les phénomènes de saisonnalité. Les modèles qui permettent de mieux décrire ce genre de comportement sont les modèles périodiques. Les modèles de séries chronologiques périodiques ont été appliqués en économie, en météorologie, en hydrologie et bien d'autres domaines. Pour les séries chronologiques périodiques univariées, on peut citer les travaux pionniers de Jones et Brelsford (1967), Pagano (1978), ainsi que ceux de Tiao et Grupe (1980). Dans les modèles multivariés, on retrouve par exemple Birchenhall et al. (1989) ou encore la monographie de Franses et Paap (2004).

Pour mieux représenter les processus périodiques, il est nécessaire de pouvoir identifier pour chaque observation la saison à laquelle elle correspond. L'index  $t$  dépend de la saison et s'écrit  $t = sT + \nu$ , où  $s$  représente le nombre total de saisons et  $\nu = 1, \dots, s$  une saison donnée. Un processus périodique fait alors apparaître une succession de cycles complets de saisons. Ces cycles complets sont indexés par  $T \in \mathbb{Z}$ . Un des exemples les plus courants de cycle complet de saisons est l'année. Dans ce cas, la notation  $t = sT + \nu$ , réfère à la saison  $\nu$  de l'année  $T + 1$ . Les concepts de stationnarité et de bruit blanc, pris dans le sens périodique, font l'objet des définitions 1.9 et 1.10.

**Définition 1.9.** *Le processus stochastique périodique  $\{\mathbf{Y}_t, t = sT + \nu, T \in \mathbb{Z}, \nu = 1, \dots, s\}$  de dimension  $d$ ,  $\mathbf{Y}_t = (Y_t(1), \dots, Y_t(d))^\top$  est dit périodiquement stationnaire de période  $s$  si et seulement si pour tout  $t = sT + \nu$ ,*

$$\begin{aligned}\boldsymbol{\mu}(\nu) &= E[\mathbf{Y}_t], \\ \boldsymbol{\Gamma}_k(\nu) &= E\left[\{\mathbf{Y}_{t+k} - \boldsymbol{\mu}(\nu + k)\}\{\mathbf{Y}_t - \boldsymbol{\mu}(\nu)\}^\top\right],\end{aligned}$$

*existent pour tout  $k \in \mathbb{Z}$  et dépendent seulement de  $k$  et de  $\nu$ .*

On définit maintenant le bruit blanc périodique.

**Définition 1.10.** *Le processus  $\{\boldsymbol{\varepsilon}_t, t = sT + \nu, T \in \mathbb{Z}, \nu = 1, \dots, s\}$  de dimension  $d$  et de période  $s$ ,  $\boldsymbol{\varepsilon}_t = (\varepsilon_t(1), \dots, \varepsilon_t(d))^\top$  est un bruit blanc périodique si :*

- (i)  $E[\varepsilon_t^2(j)] < \infty, j = 1, \dots, d, \forall t = sT + \nu,$
- (ii)  $E[\boldsymbol{\varepsilon}_t] = \mathbf{0}, \forall t = sT + \nu,$
- (iii)  $E[\boldsymbol{\varepsilon}_{t_1} \boldsymbol{\varepsilon}_{t_2}^\top] = \begin{cases} \boldsymbol{\Sigma}_\nu, & \text{si } t_1 = t_2 = sT + \nu, \\ \mathbf{0}, & \text{sinon.} \end{cases}$

De même que pour les modèles non périodiques, la modélisation ARMA peut aussi être faite pour les modèles périodiques, mais le filtre appliqué sera typiquement différent d'une saison à une autre. Dans le cas univarié, il est possible de considérer des modèles avec composantes moyennes mobiles que l'on appelle modèles de type ARMA périodique (PARMA). Il semble que l'approche dominante dans la lignée des modèles de Box et Jenkins repose sur l'utilisation de ces modèles PARMA. Nous introduisons dans ce qui suit les modèles vectoriels autorégressifs périodiques. Conceptuellement, c'est également possible de considérer une telle généralisation dans le cas multivarié. Cependant, même dans le cas non périodique, les modèles VARMA sont sujets à des problèmes d'identifiabilité. Ainsi, dans le but de simplifier la présentation, nous allons nous restreindre à l'étude des modèles autorégressifs périodiques. De plus, avec en tête les applications avec les données réelles, il semble plus vraisemblable que les modèles autorégressifs périodiques soient utilisés.

Le processus  $\{\mathbf{Y}_t, t = sT + \nu, T \in \mathbb{Z}, \nu = 1, \dots, s\}$  de dimension  $d$  et de période  $s$  est un processus autorégressif périodique multivarié (PVAR) s'il est périodiquement stationnaire et vérifie :

$$\mathbf{Y}_{sT+\nu} = \sum_{k=1}^{p(\nu)} \boldsymbol{\phi}_k(\nu) \mathbf{Y}_{sT+\nu-k} + \boldsymbol{\varepsilon}_{sT+\nu}, \quad (1.3.1)$$

où comme précédemment  $\nu = 1, \dots, s$  est la saison,  $s$  dénote le nombre total de saisons et  $p(\nu)$  correspond à l'ordre du polynôme autorégressif pour la saison  $\nu$ . Pour  $t = sT + \nu$ , le vecteur  $\mathbf{Y}_t$  est la réalisation du processus dans la saison  $\nu$  de l'année  $T + 1$ , où  $T = 1, \dots, N$ ,  $N \in \mathbb{N}^*$  étant le nombre total d'années. Les

coefficients  $\phi_k(\nu)$  sont des matrices d'ordre  $d \times d$ . Nous supposons que le processus d'erreur  $\{\boldsymbol{\varepsilon}_t, t \in \mathbb{Z}\}$  d'ordre  $d \times 1$  est un bruit blanc périodique de matrice des variances et covariances  $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top) = \boldsymbol{\Sigma}_\nu, t = sT + \nu$ .

On peut toujours considérer que  $p(\nu) \equiv p, \nu \in \{1, \dots, s\}$ , et de manière précise, l'ordre du processus être supposé égal à  $\max_{\nu=\{1, \dots, s\}} p(\nu)$ , ceci moyennant que certaines matrices de coefficients soient nulles. On parle alors d'un PVAR( $p$ ). Dans le modèle (1.3.1), la moyenne et la tendance ne sont pas prises en compte car elles sont supposées être au préalable retirées des séries. Dans ce qui suit, nous présumons alors que le processus est centré en zéro, à savoir  $E(\mathbf{Y}_t) = \mathbf{0}$  pour tous les indices  $t = sT + \nu$ .

#### 1.4. LES MODÈLES NON LINÉAIRES

Une approche possible dans les modèles non linéaires est de supposer que les observations et les innovations sont liées par une fonction non linéaire. Ces processus sont souvent caractérisés par la dépendance des moments par rapport au temps, l'asymétrie des cycles, les seuils et les ruptures. Les applications des séries chronologiques non linéaires touchent de nombreux domaines, dont l'hydrologie et l'ingénierie. Cependant, les séries chronologiques non linéaires ont trouvé de nombreuses applications en économétrie et dans l'analyse des séries chronologiques financières. Nous nous intéressons ici à la classe des modèles vectoriels à erreurs multiplicatives (vMEM). Présentés par Manganelli (2005) ainsi que par Cipollini et al. (2006), ces modèles constituent une généralisation au cas multivarié des modèles multiplicatifs à erreurs non négatives (MEM) introduits par Engle (2002). Les modèles vMEM modélisent les séries économiques prenant seulement des valeurs positives. Pour de tels modèles, il n'est pas toujours approprié de considérer le logarithme car il peut y avoir des valeurs nulles ou proches de zéro.

Soit  $\{\mathbf{X}_t, t \in \mathbb{Z}\}$  un processus à temps discrets dont les composantes sont définies sur  $\mathfrak{X} \subseteq ]0, +\infty[^d, d$  étant un entier strictement positif. On dit que le processus  $\{\mathbf{X}_t, t \in \mathbb{Z}\}$  admet la structure vMEM lorsqu'il peut s'écrire sous la

forme :

$$\mathbf{X}_t = \boldsymbol{\mu}_t \odot \boldsymbol{\varepsilon}_t, \quad (1.4.1)$$

où conditionnellement à  $\mathcal{F}_{t-1}$  qui représente l'information disponible jusqu'au temps  $t - 1$ ,  $\boldsymbol{\mu}_t$  ainsi défini sur  $]0, +\infty[^d$  satisfait :

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_t(\boldsymbol{\theta}, \mathcal{F}_{t-1}), \quad (1.4.2)$$

et  $\boldsymbol{\varepsilon}_t$  est une variable aléatoire définie sur  $]0, +\infty[^d$  avec :

$$\boldsymbol{\varepsilon}_t \sim D^+(\mathbf{1}_d, \boldsymbol{\Sigma}) \text{ pour tout } t \in \mathbb{Z}, \quad (1.4.3)$$

où  $D^+(\mathbf{1}_d, \boldsymbol{\Sigma})$  dénote une distribution avec  $E(\boldsymbol{\varepsilon}_t) = \mathbf{1}_d$  et  $\text{var}(\boldsymbol{\varepsilon}_t) = \boldsymbol{\Sigma}$ . Le symbole  $\mathbf{1}_d$  représente le vecteur d'ordre  $d$  composé uniquement de valeurs unitaires et  $\odot$  dénote l'opérateur de multiplication composante par composante des vecteurs et matrices. Cet opérateur est appelé produit de Hadamard. La matrice  $\boldsymbol{\Sigma}$  est carrée d'ordre  $d$ .

Les innovations  $\{\boldsymbol{\varepsilon}_t, t \in \mathbb{Z}\}$  sont des variables aléatoires à support positif. Elles sont supposées identiquement distribuées de moyenne  $\mathbf{1}_d$  et de matrice des variances et covariances  $\boldsymbol{\Sigma}$ . L'équation (1.4.3) permet d'avoir :

$$E(\mathbf{X}_t | \mathcal{F}_{t-1}) = \boldsymbol{\mu}_t, \quad (1.4.4)$$

$$\text{var}(\mathbf{X}_t | \mathcal{F}_{t-1}) = \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top \odot \boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\mu}_t) \boldsymbol{\Sigma} \text{diag}(\boldsymbol{\mu}_t). \quad (1.4.5)$$

Que ce soit dans les cas linéaires ou non linéaires, la validation de l'ajustement de ces modèles aux données consiste dans la plupart des cas à vérifier les hypothèses faites sur les innovations.

## 1.5. LA VALIDATION DES MODÈLES

Comme mentionné en début de ce chapitre, l'analyste est souvent amené à formuler un modèle afin de décrire le mieux possible les données. Il semble alors naturel de vérifier si les résidus qui sont en quelque sorte des « estimateurs » de l'innovation, sont compatibles avec les hypothèses de base du modèle.

De façon similaire, l'étude de relations entre deux (ou plus de deux) séries chronologiques, sous des hypothèses plus ou moins complexes, est également un champ d'application important. Là encore, des méthodes objectives sont nécessaires afin de quantifier les relations de non corrélation ou d'indépendance.

Traditionnellement, les tests diagnostiques sont particulièrement importants afin de vérifier avec des statistiques de test ce genre de considérations. On les retrouve notamment dans la méthodologie originale de Box et Jenkins. L'approche des tests portemanteaux repose sur l'étude de mesures d'autocorrélations résiduelles ou encore des corrélations croisées résiduelles. On peut alors considérer des mesures intégrées qui prennent en compte un certain nombre de délais.

Afin de vérifier l'absence de relations entre les séries chronologiques, au moins deux approches sont possibles. Dans la première, des modèles sont ajustés à chaque série, et des mesures reposant sur les corrélations croisées sont calculées. Cette approche provient de Haugh (1976). D'un point de vue théorique, cette approche est simple dans la mesure où les résidus devraient être relativement proches des séries iid. Ainsi, les propriétés d'indépendance peuvent être exploitées. D'un point de vue pratique, ce type d'approche nécessite que les bons modèles soient choisis et sont donc potentiellement sensibles à des problèmes de mauvaise spécification. De plus, en pratique, il est souvent d'intérêt d'émettre des jugements sur les séries elles-mêmes, et non sur les séries d'innovations.

La seconde approche est précisément non-paramétrique, dans le sens qu'aucun modèle ne doit être spécifié. Les tests reposent donc sur des mesures de corrélation entre les séries brutes, et chaque série pourrait être supposée stationnaire. La dépendance à l'intérieur de chaque série complexifie l'analyse, mais d'un point de vue pratique, les conclusions peuvent être statistiquement plus satisfaisantes.

Les développements de cette thèse s'inscrivent dans ce contexte, où des tests de type portemanteaux servent à valider des modèles ou encore à cerner l'absence de relations entre les séries. Dans ce qui suit, on expose les contributions les plus importantes de la thèse.

## 1.6. OBJECTIFS DE RECHERCHE ET CONTRIBUTIONS

Les objectifs principaux de la thèse portent sur la validation de modèles complexes avec des tests portemanteaux. Les modèles sont multivariés. Dans un cas, il est question de modèles non-linéaires vMEM qui sont validés avec des tests de type Hosking-Ljung-Box et aussi des tests justifiés par une approche spectrale. Dans un autre cas, il est question de l'étude de la non corrélation entre deux séries chronologiques qui pourraient cependant être dépendantes. Finalement, on considère aussi des modèles périodiques vectoriels et cointégrés. Nous décrivons brièvement les contributions originales de la thèse.

Dans le deuxième chapitre, nous introduisons le modèle vectoriel à erreurs multiplicatives pour lequel nous utilisons les propriétés de la méthode généralisée des moments pour estimer le modèle et établir la distribution asymptotique des estimateurs et des autocovariances échantillonnales. Ceci nous permet de proposer des nouveaux tests diagnostiques pour ce type de modèle. Sous l'hypothèse nulle d'adéquation du modèle, nous montrons que la statistique usuelle de Hosking-Ljung-Box converge vers une somme pondérée de lois de khi-carré indépendantes. Un test généralisé de Hosking-Ljung-Box est aussi obtenu en comparant la densité spectrale des résidus de l'estimation du modèle observé à la densité spectrale théorique obtenue sous l'hypothèse nulle d'adéquation du modèle. Les tests spectraux sont un peu plus flexibles quant aux propriétés des estimateurs qui doivent être seulement convergents.

Dans le troisième chapitre, nous obtenons la distribution asymptotique des autocovariances empiriques sous faible dépendance entre deux processus stationnaires en covariance. La faible dépendance ici se traduit par les conditions de stabilité de Wu (2005). Ces conditions sont plus générales que celles de Hannan (1976) qui stipule que les moments conditionnels sont constants. Ce résultat permet de construire une statistique de test de non corrélation qui nous permet d'évaluer la corrélation entre les deux processus dans le cas où la dépendance entre elles est faible pour satisfaire les conditions de stabilité décrites par Wu (2005). Dans l'évaluation de la statistique de test, la matrice des variances et covariances

de la distribution asymptotique est estimée par la procédure vectorielle autorégressive robuste à l'autocorrélation et l'hétéroscédasticité telle que présentée par Berk (1974) et Den Haan et Levin (1994). Cette approche est plus générale que celle supposant la nullité des cumulants d'ordre quatre présentée par Roy (1989).

Dans le quatrième chapitre, nous procédons à l'estimation et à l'étude des propriétés asymptotiques des paramètres d'un modèle périodique multivarié cointégré. Ce résultat généralise les travaux de Ahn et Reinsel (1990) et Yap et Reinsel (1995) qui ont fait un travail similaire dans le cas non périodique. En présence de cointégration nous sommes en présence d'un modèle à rang réduit dont l'estimation se fait en deux étapes. La première est l'estimation de plein rang et la deuxième est celle de rang réduit. Le but de la première est de trouver les valeurs initiales pour la deuxième. Le modèle de rang réduit est en effet non linéaire et s'estime de manière itérative. Par la suite, nous obtenons la distribution asymptotique des matrices d'autocorrélation résiduelles. Ce résultat permet de construire un test de type portemanteau dont la distribution est asymptotiquement khi-carrée. Ce résultat généralise celui d'Ursu et Duchesne (2009) qui l'ont fait dans le cas périodique mais non cointégré.

## Chapitre 2

---

# EVALUATING VECTOR MULTIPLICATIVE ERROR MODELS WITH THE HOSKING-LJUNG-BOX PORTMANTEAU TEST AND KERNEL-BASED TEST STATISTICS

**Abstract** Multivariate nonlinear time series models have experienced many developments for modelling data coming from financial applications. Several financial time series are realizations from nonnegative processes. An important class of models is composed of vector multiplicative error models (vMEM), which can describe contemporaneous correlations among innovations and the dynamic interdependencies among variables. Modelling and estimation issues have been addressed, but few diagnostic checking procedures are available. Here, new tests are proposed to check vMEM models. The asymptotic distributions of the popular Hosking-Ljung-Box (HLB) test statistics are found to converge in distribution to weighted sums of independent chi-squared random variables under the null hypothesis of adequacy. A generalized HLB test statistic is motivated by comparing a vector spectral density estimator of the residuals with the spectral density calculated under the null hypothesis. Under general conditions, that kind of approach leads to consistent procedures and to powerful measures of lack-of-fit. To improve the finite sample properties, the spectral test statistics rely on the power transformation of Chen and Deo (2004). Appealing properties of the spectral



procedures include the distribution-free property and the fact that they converge in distribution to convenient standard normal distributions under the null hypothesis. Simulation experiments are reported to appreciate the properties of the methods. An application using financial data previously analyzed by Cipollini (2013). illustrates the merits of our procedures.

*keywords* : Multivariate time series; nonnegative processes; nonlinear time series; portmanteau test statistic; spectral density.

## 2.1. INTRODUCTION

Several financial and economic phenomena can be described by nonnegative processes. Examples include measures of volume, number of trades, measures of volatility (such as absolute returns, daily ranges or realized volatilities), or durations between transactions. The Autoregressive Conditional Duration (ACD) model proposed by Engle and Russell (1998) represents a significant contribution in this area, since it proposes a model for the arrival time intervals between transactions. More fundamentally, it suggests a model for nonnegative stochastic processes, which are expected to be suitable under certain dependence structures (such as duration clustering). A related model is the log-ACD model of Bauwens and Giot (2000). Applications of ACD models include Engle and Russell (1998), Jasiak (1998), Engle (2000), Zhang and al. (2001), Bauwens and Veredas (2004), Manganelli (2005), Bauwens (2006) and Dionne and al. (2009). The survey article of Pacurar (2008) reviews the first decade of theoretical and empirical work on ACD modeling and it also discusses empirical studies using financial data. The ACD models have been generalized by Engle (2002), who introduced the class of multiplicative error models (MEM). These models are particularly appropriate in explaining stylized facts of financial time series, not only durations, such as persistence and clustering features of absolute returns or squared returns. Hautsch (2008) generalizes this approach by introducing a common latent dynamic factor serving as a subordinated process driving the individual components. Other examples of related time series models include the conditional autoregressive range (CARR) time series models of Chou (2005). Hautsch and Jeleskovic (2008) present an extensive review of MEM models, including applications to trading processes.

Manganelli (2005) and Engle and Gallo (2006) investigated vector time series composed of financial variables. If the innovation terms were uncorrelated, their models proved to be useful. In particular, the model parameters can be consistently estimated by fitting the model using an equation-by-equation procedure. However, uncorrelatedness of the innovations represents a strong and somewhat

restrictive assumption. In order to include dependence between the components of the innovation process, multivariate generalizations of MEM models have been investigated in Cipollini (2006) and Cipollini (2007). They introduced the so-called class of vector MEM (vMEM) time series models, in which joint modelization is made possible by considering element-by-element product of a vector of conditional expectations time a multivariate innovation process with positive support, composed of components having each a unit mean but a general covariance matrix.

In vMEM models, satisfactory statistical estimation methods afford simple explanations for the dynamic interdependencies among the variables. Thus, interactions among the measures of volatility in vMEM models may be studied and explained. As in classical time series models, the interpretation of coefficients reveals its practical importance, and for example zero coefficients may be viewed as absence of direct dynamic link between the variables of the dynamic system. However, to estimate efficiently vMEM models for realistic sample sizes from financial applications represents a challenging numerical problem. Two main approaches have been considered in the literature. In the first strategy, called the full parametric method, the maximum likelihood function is formulated, including the unknown parameters of the innovation distribution. Since the innovation process has a positive support, flexibility is achieved by considering copulas for the joint distribution of the innovation process. See Cipollini (2006) for a description of that approach. Algorithms have also been developed. In order to simplify the estimation procedure, an alternative approach consists in considering the semiparametric methodology developed in Cipollini (2013), which relies on the so-called Generalized Method of Moments (GMM). Model selection procedures are studied in Cipollini (2010), following a general-to-specific approach. In their method, over-parameterized models are first adjusted, and then using Wald-type test statistics, coefficients which are the least significant are gradually deleted with an iterative procedure.

Diagnostic checking the residuals appears to be an important part of any statistical modelling strategy. In vMEM modelling, the asymptotic distributions

of the residual autocovariances and autocorrelations are derived in Ng and al. (2016). When the asymptotic covariance matrices of the autocovariances or autocorrelations are non-singular, natural test statistics are proposed, which converge in distribution toward convenient chi-square distributions. To this date, no precise conditions seem available assuring that the asymptotic covariance matrices involved in their results are invertible. In fact, our data analysis suggests that numerical instabilities may occur in inverting the natural estimator of that matrix. To avoid that difficulty, the asymptotic distribution of the classical Hosking-Ljung-Box (HLB) test statistic can be derived. More precisely, as a corollary of a result of Ng and al. (2016), the asymptotic distribution of the HLB portmanteau test is obtained, which is a weighted sum of independent chi-square random variables. Interestingly, this approach is seen valid even if the asymptotic covariance matrix is arbitrarily close to a singular matrix, to the price of a slightly more complicated distribution. Using modern computer resources, such as the **R** package `CompQuadForm` described in Duchesne and Lafaye de Micheaux (2010), the critical values of weighted sum of independent chi-square random variables can be found rather easily.

By its nature, the asymptotic distribution of the HLB portmanteau test is obtained assuming uniform weighting and a fixed lag order. Under stationary assumptions, the autocovariances and autocorrelations of the residual process tend to zero quickly. In order to obtain more powerful procedures, this suggests that non-uniform weighting may be used, such that more weight would be attributed to low order lags. The lag order needs also to be carefully chosen. If it is selected too low, this may lead to inconsistent procedures. This leads us to adopt a spectral approach, where the test statistics are obtained by comparing a vector spectral density estimator of the residuals with the spectral density calculated under the null hypothesis. Kernel-based spectral density estimators are advocated, which rely on the choice of a kernel and a lag order or smoothing parameter. Thus a flexible weighting represents an advantage of the method. From the literature on diagnostic checking using spectral methods, these methods are consistent under general conditions. Since the lag order or smoothing parameter is allowed

to converge to infinity under a suitable rate, all the lags have eventually strictly positive weights, and consistency is obtained. Furthermore, the asymptotic distributions are model-free and distribution-free, as long as the model parameters are consistently estimated. Interestingly, using the truncated uniform kernel and assuming a fixed lag-order (with respect to the sample size), a generalized HLB test statistic is obtained. In many applications, residual autocorrelations with low lag orders should describe a large amount of the remaining dependence. By considering a more general weighting scheme, more powerful testing procedures are typically obtained by giving more weight to low order lags. The test statistics generalize proposals from Duchesne and Roy (2004), justified in vector autoregressive models with stationary exogenous variables.

From the results of Poulin and Duchesne (2008), the exact asymptotic distributions of the spectral tests are expected to be highly right-skewed. Using the power transformation of Chen and Deo (2004), the finite sample properties are improved. The bootstrap is an alternative method, which remains numerically intensive. This appears particularly true in our multivariate framework, given that in practical applications the sample sizes associated with financial time series can be quite large. The Chen-Deo power transformation is easy to implement and gives satisfactory results, both in term of empirical levels and powers. These claims are supported by the simulation results reported in this paper. We also discuss how to choose the lag order or smoothing parameter using graphical methods and sensibility analyzes.

The organization of the paper is as follows. In Section 2.2, some preliminaries are presented. The models and popular estimation procedures are described. In Section 2.3, two approaches to check vMEM time series models are developed. We obtain the asymptotic distribution of the HLB test statistic, which is a weighted sum of independent chi-square random variables. We also establish that the kernel-based test statistics have a convenient asymptotic normal distribution under the null hypothesis, and we develop the transformed kernel-based test statistics relying on the Chen-Deo power transformation. Some simulation results are reported in Section 2.4, where the exact levels are obtained and the empirical

powers are calculated. The proposed test statistics are compared with respect to levels and powers with many kernels to the test procedures of Ng and al. (2016). In Section 2.5, we present an application using financial data previously analyzed by Cipollini (2013). Some concluding remarks are offered in Section 2.6 and technical details are given in an Appendix.

## 2.2. SOME PRELIMINARIES AND DEFINITION AND ESTIMATION METHODS IN vMEM MODELS

We present in this section the models and some preliminaries. Throughout the paper, the symbols  $\xrightarrow{L}$  and  $\xrightarrow{P}$  stand for convergence in distribution and in probability, respectively. For a random variable  $X_n$ ,  $X_n \xrightarrow{P} 0$  is noted  $X_n = o_P(1)$  and  $X_n = O_P(1)$  means bounded in probability. Similar notations are used for random vectors. See Serfling (1980) for more details and properties. The Euclidean norm of a matrix  $\mathbf{A}$  is noted  $\|\mathbf{A}\| = \{\text{tr}(\mathbf{A}\mathbf{A}^\top)\}^{1/2}$ , where  $\text{tr}(\cdot)$  denotes the trace operator and  $\mathbf{A}^\top$  corresponds to the transpose matrix of  $\mathbf{A}$ .

### 2.2.1. The vMEM time series models

Let  $\mathbf{X} = \{\mathbf{X}_t, t \in \mathbb{Z}\}$  be a stationary and ergodic vector process, composed of the random vectors  $\mathbf{X}_t = (X_t(1), \dots, X_t(d))^\top$ . Each component of  $\mathbf{X}_t$  is assumed to be nonnegative, that is  $X_t(i) \geq 0$ ,  $i \in \{1, \dots, d\}$ . The semiparametric vector multiplicative error model (vMEM) is defined as :

$$\mathbf{X}_t = \boldsymbol{\mu}_t \odot \boldsymbol{\epsilon}_t, \quad (2.2.1)$$

where ' $\odot$ ' denotes the Hadamard product. For two  $n \times m$  matrices  $\mathbf{A} = (a_{ij})$  and  $\mathbf{B} = (b_{ij})$ ,  $\mathbf{A} \odot \mathbf{B}$  represents simply the element-by-element product  $\mathbf{C} = (c_{ij})$ , that is  $c_{ij} = a_{ij}b_{ij}$ ,  $i \in \{1, \dots, n\}$  and  $j \in \{1, \dots, m\}$ . It follows that the model (2.2.1) can equivalently be written as  $\boldsymbol{\mu}_t \odot \boldsymbol{\epsilon}_t = \text{diag}(\boldsymbol{\mu}_t)\boldsymbol{\epsilon}_t$ , where for a  $n \times 1$  vector  $\mathbf{a} = (a(1), \dots, a(n))^\top$ ,  $\text{diag}(\mathbf{a})$  denotes the  $n \times n$  matrix with the elements of  $\mathbf{a}$  on the main diagonal. The nonnegative innovation process  $\boldsymbol{\epsilon} = \{\boldsymbol{\epsilon}_t, t \in \mathbb{Z}\}$ , with  $\boldsymbol{\epsilon}_t = (\epsilon_t(1), \dots, \epsilon_t(d))^\top$ , is composed of independent and identically distributed

(iid) random vectors, such that  $E(\boldsymbol{\epsilon}_t) = \mathbf{1}_d$ , where  $\mathbf{1}_d$  corresponds to a  $d \times 1$  vector whose all components are equal to one, and a nonsingular covariance matrix  $\boldsymbol{\Sigma} = (\sigma_{ij})_{i,j \in \{1, \dots, d\}}$ . We assume that  $E\{\epsilon_t^4(i)\} < \infty$ ,  $i = 1, \dots, d$ . Let  $\mathcal{F}_{t-1}$  be the  $\sigma$ -algebra generated by  $\{\mathbf{X}_s, s \leq t-1\}$ . The  $d \times 1$  vector  $\boldsymbol{\mu}_t = (\mu_t(1), \dots, \mu_t(d))^\top$  represents the conditional expectation  $\boldsymbol{\mu}_t = E(\mathbf{X}_t | \mathcal{F}_{t-1})$ .

It is assumed that  $\boldsymbol{\mu}_t$  is function of a  $b \times 1$  vector of parameters, denoted  $\boldsymbol{\theta}$ , and we write  $\boldsymbol{\mu}_t = \boldsymbol{\mu}_t(\boldsymbol{\theta}; \mathcal{F}_{t-1})$ . It is also supposed that  $\boldsymbol{\mu}_t$  has continuous second-order derivatives, almost surely. Various specifications can be considered for the conditional mean  $\boldsymbol{\mu}_t$ . Following the vector GARCH literature, a possible model admits the simple structure :

$$\boldsymbol{\mu}_t = \boldsymbol{\alpha}_0 + \mathbf{A}\mathbf{X}_{t-1}, \quad (2.2.2)$$

where  $\boldsymbol{\alpha}_0$  corresponds to a  $d \times 1$  vector and  $\mathbf{A}$  is a  $d \times d$  matrix. This model is noted vMEM(1). The generalization to the vMEM( $p$ ) case is straightforward, adding the lagged terms  $\mathbf{X}_{t-i}$ ,  $i = 1 \dots, p$ . More generally, a vMEM( $p, q$ ) model can be expressed as :

$$\boldsymbol{\mu}_t = \boldsymbol{\alpha}_0 + \sum_{i=1}^p \mathbf{A}_i \mathbf{X}_{t-i} + \sum_{i=1}^q \mathbf{B}_i \boldsymbol{\mu}_{t-i}, \quad (2.2.3)$$

where  $\mathbf{A}_i$ ,  $i = 1, \dots, p$  and  $\mathbf{B}_i$ ,  $i = 1, \dots, q$ , represent  $d \times d$  matrices. That specification relies on lagged values of  $\boldsymbol{\mu}_t$ , which can be used to model significant persistence. A specification justified by financial considerations admits the following writing :

$$\boldsymbol{\mu}_t = \boldsymbol{\alpha}_0 + \sum_{i=1}^l \{\mathbf{A}_i \mathbf{X}_{t-i} + \mathbf{V}_i \mathbf{X}_{t-i}^{(-)} + \mathbf{B}_i \boldsymbol{\mu}_{t-i}\}, \quad (2.2.4)$$

where  $\mathbf{V}_i$ ,  $i = 1, \dots, l$  represent additional  $d \times d$  matrices. In model (2.2.4), the order  $l$  is the same for  $\mathbf{X}_t$ ,  $\mathbf{X}_t^{(-)}$  and  $\boldsymbol{\mu}_t$  and it corresponds to a maximal lag-order. The term  $\mathbf{V}_i \mathbf{X}_{t-i}^{(-)}$  aims to capture asymmetric effects associated with the sign of an observed variable. For example, Cipollini (2006) considers situations where  $\mathbf{X}_t^{(-)} = \mathbf{X}_t I(\{r_t < 0\})$ , where  $\{r_t, t \in \mathbb{Z}\}$  denotes the stationary stochastic process of returns and  $I(A)$  corresponds to the indicator function of the set  $A$ .

Other useful specifications are investigated in Manganelli (2005) and Hautsch (2008).

### 2.2.2. Estimation procedures of vMEM models

A vMEM model such as the vMEM( $p, q$ ) model defined by (2.2.3) may rely on  $d^2(p+q) + d$  independent parameters. Similar to vector autoregressive moving average models (VARMA), we can consider constraints, leading to structured vMEM models. Let

$$\gamma = \text{vec}(\alpha_0, \mathbf{A}_1, \dots, \mathbf{A}_p, \mathbf{B}_1, \dots, \mathbf{B}_q).$$

For a matrix  $\mathbf{A}$ ,  $\text{vec}(\mathbf{A})$  represents the vector obtained by stacking the columns of  $\mathbf{A}$ . Generally, constraints of the following form can be formulated :

$$\gamma = \mathbf{g}(\boldsymbol{\theta}),$$

where  $\mathbf{g}(\cdot)$  corresponds to a  $\mathbb{R}^b \rightarrow \mathbb{R}^{d^2(p+q)+d}$  function. A simple but useful case considers :

$$\gamma = \mathbf{R}\boldsymbol{\theta} + \mathbf{c},$$

where  $\mathbf{R}$  and  $\mathbf{c}$  correspond to a known matrix and vector, respectively. In practical applications, fixing coefficients to zero in the matrices  $\mathbf{A}_i$  and/or  $\mathbf{B}_i$  are useful linear constraints. See also the discussion of Reinsel (1997) in VARMA models. Note that the original specification of  $\boldsymbol{\mu}_t$  relies on a general vector  $\boldsymbol{\theta}$ , not on  $\gamma$  and the function  $\mathbf{g}(\cdot)$  can be nonlinear.

Estimation procedures for vMEM time series models can be parametric or semiparametric. A parametric approach relies on a specification of the full likelihood function, including parameters in the distribution of the error term. Due to the nonnegative support of the innovation process, the multivariate normal distribution does not correspond to a natural choice. A possible strategy uses copulas, where the marginal distributions and the copula function can be separately specified. See Nelsen (1999) for an introduction to copulas. This is the route taken by Cipollini (2007), who specified a normal copula function and Gamma



distributions for the marginal distributions. A popular method which avoid the formulation and the optimization of the full likelihood function over a large number of parameters is based on the generalized method of moments, the so-called GMM method. Consistency is ensured by the correct specification of the conditional expectation. That method involves  $b$  population moment conditions relying on the instruments  $\mathbf{G}_t$ . Thus the conditions are summarized in the relation :

$$E(\mathbf{G}_t \mathbf{u}_t) = \mathbf{0},$$

where  $\mathbf{u} = \{\mathbf{u}_t, t \in \mathbb{Z}\}$ , with  $\mathbf{u}_t = \mathbf{X}_t \odot \boldsymbol{\mu}_t - \mathbf{1}_d$  being a martingale difference sequence, and  $\mathbf{G}_t \equiv \mathbf{G}_t(\mathcal{F}_{t-1})$  relies on the information set  $\mathcal{F}_{t-1}$ . For two compatible matrices  $\mathbf{A} = (a_{ij})$  and  $\mathbf{B} = (b_{ij})$ ,  $b_{ij} \neq 0$  for all  $i, j$ , the operator ' $\odot$ ' represents the element-by-element division, that is  $\mathbf{A} \odot \mathbf{B} = (a_{ij}/b_{ij})$ . The sample moment conditions are :

$$n^{-1} \sum_{t=1}^n \mathbf{G}_t \mathbf{u}_t = \mathbf{0}.$$

An efficient choice of the instrument matrix leads to solve the following equation :

$$\mathbf{g}(\boldsymbol{\theta}) = n^{-1} \sum_{t=1}^n \mathbf{g}_t = \mathbf{0},$$

where :

$$\mathbf{g}_t(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{\mu}_t^\top}{\partial \boldsymbol{\theta}} \text{diag}^{-1}(\boldsymbol{\mu}_t) \boldsymbol{\Sigma}^{-1} (\boldsymbol{\epsilon}_t - \mathbf{1}_d),$$

and  $\partial \boldsymbol{\mu}_t^\top / \partial \boldsymbol{\theta}$  represents the  $p \times b$  matrix of partial derivatives. Since there are as many conditions as parameters (that is, a situation of exact identification), the GMM estimator  $\hat{\boldsymbol{\theta}}_{GMM}$  satisfies (and thus solve) the system asymptotically :  $\mathbf{g}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$ . In fact, it can be shown that  $n^{1/2}(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta}_0)$  converges to the multivariate normal distribution asymptotically, with the asymptotic covariance matrix  $\mathbf{V}^{-1}$ , where  $\mathbf{V}$  satisfies the relation :

$$\mathbf{V} = \lim_{n \rightarrow \infty} E \left\{ n^{-1} \sum_{t=1}^n \frac{\partial \boldsymbol{\mu}_t^\top}{\partial \boldsymbol{\theta}} \text{diag}^{-1}(\boldsymbol{\mu}_t) \boldsymbol{\Sigma}^{-1} \text{diag}^{-1}(\boldsymbol{\mu}_t) \frac{\partial \boldsymbol{\mu}_t}{\partial \boldsymbol{\theta}^\top} \right\}. \quad (2.2.5)$$

Under correct specification, these methods provide  $n^{1/2}$ -consistent estimators. That leads us to the following assumption on the estimation method.

**Assumption A.** Let the vMEM model be defined as (2.2.1). The vector of parameter  $\boldsymbol{\theta}_0$  is estimated by a  $n^{1/2}$ -consistent procedure, that is :

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \mathbf{O}_P(n^{-1/2}),$$

where  $\boldsymbol{\theta}_0$  denotes the true value of the model parameters in the conditional expectation  $\boldsymbol{\mu}_t(\boldsymbol{\theta}; \mathcal{F}_{t-1})$ .

The spectral density of a vector stationary process and its estimation with the kernel-based method are discussed in the next section.

### 2.2.3. Spectral density estimation of the innovation process

The spectral density  $\mathbf{f}_a(\omega)$ ,  $\omega \in [-\pi, \pi]$ , of an arbitrary zero-mean second order stationary process is defined by :

$$\mathbf{f}_a(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \boldsymbol{\Gamma}_a(h) \exp(-i\omega h), \quad \omega \in [-\pi, \pi],$$

where  $\boldsymbol{\Gamma}_a(h) = E(\mathbf{a}_t \mathbf{a}_{t-h}^\top)$ ,  $h \in \mathbb{Z}$ , denotes the theoretical lag- $h$  autocovariance and  $i = \sqrt{-1}$ . A sufficient condition assuring the existence of  $\mathbf{f}_a(\omega)$  is  $\sum_{h \geq 0} |\Gamma_{a,ij}(h)| < \infty$ ,  $i, j \in \{1, \dots, d\}$ , where  $\boldsymbol{\Gamma}_a(h) = (\Gamma_{a,ij}(h))_{i,j \in \{1, \dots, d\}}$ . See for example Fuller [49, pp. 170-171]. Once a vMEM model is estimated, the model residuals can be calculated. They are defined as :

$$\hat{\boldsymbol{\epsilon}}_t = \mathbf{X}_t \circ \boldsymbol{\mu}_t(\hat{\boldsymbol{\theta}}) \equiv \mathbf{X}_t \circ \hat{\boldsymbol{\mu}}_t, \quad t = 1, \dots, n. \quad (2.2.6)$$

It is convenient to introduce  $\hat{\mathbf{u}}_t = \hat{\boldsymbol{\epsilon}}_t - \mathbf{1}_d$ ,  $t = 1, \dots, n$ . A residual nonparametric kernel-based spectral density estimator is given by :

$$\hat{\mathbf{f}}_{\hat{\boldsymbol{\epsilon}}}(\omega) = \frac{1}{2\pi} \sum_{h=-n+1}^{n-1} k(h/m_n) \mathbf{C}_{\hat{\boldsymbol{\epsilon}}}(h) \exp(-i\omega h), \quad (2.2.7)$$

where the function  $k(\cdot)$  is a kernel function. It relies on the parameter  $m_n$ , which is a truncation point or a smoothing parameter, depending if the support of  $k(\cdot)$  is compact or unbounded. It satisfies the conditions  $m_n \rightarrow \infty$  but such that  $m_n/n \rightarrow 0$ . These assumptions on the rates are standard in spectral density estimation.

See for example Brockwell and Davis [22, Section 10.4]. The assumptions on the kernel function are summarized in Assumption B.

**Assumption B.** *The kernel function  $k : \mathbb{R} \rightarrow [-1, 1]$  is a symmetric function, such that  $k(0) = 1$  and  $\int_{-\infty}^{\infty} k^2(z)dz < \infty$ .*

The periodogram is retrieved by adopting the rectangular or truncated uniform kernel  $k_{TR}(z) = I(|z| \leq 1)$ . Other kernel functions frequently used in spectral density estimation are listed in Priestley [91, Section 6.2.3]. The lag- $h$  residual autocovariance  $\mathbf{C}_{\hat{\boldsymbol{\epsilon}}}(h)$  is defined as :

$$\mathbf{C}_{\hat{\boldsymbol{\epsilon}}}(h) = \begin{cases} n^{-1} \sum_{t=h+1}^n \{\hat{\boldsymbol{\epsilon}}_t - \text{ave}(\hat{\boldsymbol{\epsilon}})\} \{\hat{\boldsymbol{\epsilon}}_{t-h} - \text{ave}(\hat{\boldsymbol{\epsilon}})\}^\top, & h = 0, 1, \dots, n-1, \\ \mathbf{C}_{\hat{\boldsymbol{\epsilon}}}^\top(-h), & h = -1, \dots, -n+1, \end{cases}$$

where  $\text{ave}(\hat{\boldsymbol{\epsilon}}) = n^{-1} \sum_{t=1}^n \hat{\boldsymbol{\epsilon}}_t$ . We note  $\hat{\boldsymbol{\Sigma}} = \mathbf{C}_{\hat{\boldsymbol{\epsilon}}}(0)$ , which is consistent for  $\boldsymbol{\Sigma}$ .

Under Assumption A and supposing a correct specification of the model,  $\hat{\boldsymbol{\theta}}$  is seen  $n^{1/2}$ -consistent. Thus, an application of the ergodic theorem gives  $\text{ave}(\hat{\boldsymbol{\epsilon}}) \xrightarrow{P} \mathbf{1}_d$ .

Consequently, we also introduce :

$$\mathbf{C}_{\hat{\mathbf{u}}}(h) = \begin{cases} n^{-1} \sum_{t=h+1}^n \hat{\mathbf{u}}_t \hat{\mathbf{u}}_{t-h}^\top, & h = 0, 1, \dots, n-1, \\ \mathbf{C}_{\hat{\mathbf{u}}}^\top(-h), & h = -1, \dots, -n+1, \end{cases}$$

which will be useful in our mathematical developments. Diagnostic checking vMEM models is discussed in the next section.

## 2.3. THE TEST STATISTICS AND THEIR ASYMPTOTIC DISTRIBUTIONS

### 2.3.1. The Ng-Li-Yu test statistic and the Hosking-Ljung-Box portmanteau test

We now describe the test statistics for checking vMEM time series models. Suppose that a particular vMEM model is adjusted, say  $\boldsymbol{\mu}_t(\boldsymbol{\theta}; \mathcal{F}_{t-1})$ , for a certain  $\boldsymbol{\theta}$ . At this point, the analyst may wish to test if the model fits the data adequately.

One approach uses the following residual process :

$$\mathbf{e}_t = \mathbf{X}_t \circ \boldsymbol{\mu}_t(\boldsymbol{\theta}; \mathcal{F}_{t-1}). \quad (2.3.1)$$

Here,  $\{\mathbf{e}_t, t \in \mathbb{Z}\}$  denotes a stochastic process defined for each  $\boldsymbol{\theta}$ , where  $\boldsymbol{\theta}$  belongs to the parameter space. One may want to study the dependence structure of  $\{\mathbf{e}_t\}$ . In practice,  $\boldsymbol{\theta}$  is estimated by consistent procedures satisfying Assumption A, that is  $\text{plim}\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$  for a certain  $\boldsymbol{\theta}_0$ . If  $\boldsymbol{\mu}_t(\boldsymbol{\theta}_0; \mathcal{F}_{t-1}) = \boldsymbol{\mu}_{0t}(\mathcal{F}_{t-1})$ , almost surely, where  $\boldsymbol{\mu}_{0t}(\mathcal{F}_{t-1})$  denotes the real model, then  $\boldsymbol{\mu}_t(\boldsymbol{\theta}_0; \mathcal{F}_{t-1})$  is said appropriate for modelling  $\boldsymbol{\mu}_{0t}(\mathcal{F}_{t-1})$ . In this situation  $\mathbf{e}_t = \boldsymbol{\epsilon}_t$  and  $\{\mathbf{e}_t\}$  corresponds to an iid white noise process. In this case, the spectral density of  $\{\mathbf{e}_t\}$  reduces to a constant matrix :

$$\mathbf{f}_e(\omega) = \frac{\boldsymbol{\Gamma}_e(0)}{2\pi}, \quad \forall \omega \in [-\pi, \pi].$$

Alternatively, suppose that the model  $\boldsymbol{\mu}_t(\boldsymbol{\theta}; \mathcal{F}_{t-1})$  is inadequate in modelling the process  $\{\mathbf{X}_t, t \in \mathbb{Z}\}$ . Then, we have that  $P\{\boldsymbol{\mu}_t(\boldsymbol{\theta}; \mathcal{F}_{t-1}) = \boldsymbol{\mu}_{0t}(\mathcal{F}_{t-1})\} < 1$ , for all  $\boldsymbol{\theta}$ . Then it follows that  $\mathbf{e}_t \neq \boldsymbol{\epsilon}_t$ . The proposed test statistic will have power when the dependence structure of  $\{\mathbf{e}_t\}$  will be such that  $\boldsymbol{\Gamma}_e(h) \neq \mathbf{0}$  for a certain lag  $h$ . Under that kind of alternative hypothesis, the spectral density of  $\{\mathbf{e}_t\}$  will not be identically equal to a constant matrix.

Introduce the following  $d^2 \times m$  vectors :

$$\begin{aligned} \mathbf{c}_{\hat{\boldsymbol{\epsilon}}_m} &= (\mathbf{c}_{\hat{\boldsymbol{\epsilon}}}^\top(1), \dots, \mathbf{c}_{\hat{\boldsymbol{\epsilon}}}^\top(m))^\top, \\ \mathbf{c}_{\hat{\boldsymbol{u}}_m} &= (\mathbf{c}_{\hat{\boldsymbol{u}}}^\top(1), \dots, \mathbf{c}_{\hat{\boldsymbol{u}}}^\top(m))^\top, \end{aligned}$$

where  $\mathbf{c}_{\hat{\boldsymbol{\epsilon}}}(h) = \text{vec}\{\mathbf{C}_{\hat{\boldsymbol{\epsilon}}}(h)\}$  and  $\mathbf{c}_{\hat{\boldsymbol{u}}}(h) = \text{vec}\{\mathbf{C}_{\hat{\boldsymbol{u}}}(h)\}$ ,  $h = 1, \dots, m$ . The first ingredient consists in studying the asymptotic behavior of  $n^{1/2}\mathbf{c}_{\hat{\boldsymbol{u}}_m}$  under the null hypothesis of adequacy. It may be shown relatively easily that  $n^{1/2}\mathbf{c}_{\hat{\boldsymbol{u}}_m} = n^{1/2}\mathbf{c}_{\hat{\boldsymbol{\epsilon}}_m} + \mathbf{o}_P(1)$ , that is  $n^{1/2}\mathbf{c}_{\hat{\boldsymbol{u}}_m}$  and  $n^{1/2}\mathbf{c}_{\hat{\boldsymbol{\epsilon}}_m}$  share the same asymptotic distribution under the null hypothesis (that is to estimate the mean  $\mathbf{1}_d$  by  $\text{ave}(\hat{\boldsymbol{\epsilon}}_t)$  has no effect on the asymptotic distribution of  $n^{1/2}\mathbf{c}_{\hat{\boldsymbol{u}}_m}$ ). The asymptotic distribution of  $n^{1/2}\mathbf{c}_{\hat{\boldsymbol{u}}_m}$  has been first derived in Ng, Li, & Yu [82, Theorem 1]. The result is valid when the parameters are estimated using the GMM method; the asymptotic

limit is normal and is expressed as follows :

$$n^{1/2} \mathbf{c}_{\hat{\mathbf{u}}_m} \xrightarrow{L} \mathcal{N}_{d^2 m}(\mathbf{0}, \mathbf{\Omega}_m),$$

where  $\mathbf{\Omega}_m = \mathbf{I}_m \otimes \mathbf{\Sigma} \otimes \mathbf{\Sigma} - \mathbf{XV}^{-1}\mathbf{X}^\top$ ,  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_m)^\top$ , with  $\mathbf{X}_h$  satisfying the expression :

$$\mathbf{X}_h^\top = E \left[ \mathbf{u}_{t-h}^\top \otimes \left\{ (\partial \boldsymbol{\mu}_t^\top / \partial \boldsymbol{\theta}) \text{diag}^{-1}(\boldsymbol{\mu}_t) \right\} \right],$$

$h = 1, \dots, m$ , and  $\mathbf{V}$  corresponds to the asymptotic covariance matrix obtained under the GMM procedure. Assuming that the matrix  $\mathbf{\Omega}_m$  is non-singular, the portmanteau test proposed by Ng and al. (2016) enjoys under the null hypothesis a convenient chi-square distribution :

$$\text{NLY}_n(m) = n \mathbf{c}_{\hat{\boldsymbol{\epsilon}}_m}^\top \hat{\boldsymbol{\Omega}}_m^{-1} \mathbf{c}_{\hat{\boldsymbol{\epsilon}}_m} \xrightarrow{L} \chi_{d^2 m}^2, \quad (2.3.2)$$

where  $\hat{\boldsymbol{\Omega}}_m$  is a consistent estimator of  $\mathbf{\Omega}_m$ . How to construct such consistent estimator is discussed in detail in Ng and al. (2016). To the best of our knowledge, precise conditions which warrant that the matrices  $\hat{\boldsymbol{\Omega}}_m$  and  $\mathbf{\Omega}_m$  are nonsingular are not known. The application presented in Section 2.6 demonstrates that  $\hat{\boldsymbol{\Omega}}_m$  may be instable. The condition number, that is the ratio of the largest eigenvalue to the smallest eigenvalue may be quite large, suggesting that the matrix may be close of being singular. See Burden & Douglas Faires [23, Chap. 7] for details on ill-conditioned matrices. Following Duchesne and Francq (2008), the computation of the rank of the matrix  $\hat{\boldsymbol{\Omega}}_m$  may be recommended, and if  $\hat{\boldsymbol{\Omega}}_m$  is close to a singular matrix, a Moore-Penrose inverse may be used. Let  $\hat{\boldsymbol{\Omega}}_m^-$  be the Moore-Penrose inverse of  $\hat{\boldsymbol{\Omega}}_m$ . This leads us to the following test statistic :

$$\text{NLY}_n^-(m) = n \mathbf{c}_{\hat{\boldsymbol{\epsilon}}_m}^\top \hat{\boldsymbol{\Omega}}_m^- \mathbf{c}_{\hat{\boldsymbol{\epsilon}}_m}, \quad (2.3.3)$$

which under the null hypothesis converges in distribution toward :

$$\text{NLY}_n^-(m) \xrightarrow{L} \chi_{\text{rank}(\mathbf{\Omega}_m)}^2.$$

Other approaches could be used, based on regularized test statistics. See Dufour and Valery (2015). Alternatively, one may study the asymptotic distribution of the

Hosking-Ljung-Box statistic, which has been studied in linear time series models such as vector white noise, vector autoregressive and vector ARMA models (see, e.g., Chitturi (1974), Chitturi (1976) and Li and McLeod (1981)). It is defined as :

$$\text{HLB}_n(m) = n \sum_{h=1}^m \mathbf{c}_{\hat{\boldsymbol{\epsilon}}}(h) \left\{ \hat{\boldsymbol{\Sigma}}^{-1} \otimes \hat{\boldsymbol{\Sigma}}^{-1} \right\} \mathbf{c}_{\hat{\boldsymbol{\epsilon}}}(h), \quad (2.3.4)$$

whose asymptotic distribution is given in the next proposition.

**Proposition 1.** *Consider the vMEM time series model defined by (2.2.1), where the vector of parameters  $\boldsymbol{\theta}$  is estimated by the GMM model  $\hat{\boldsymbol{\theta}}_{\text{GMM}}$  with asymptotic covariance matrix  $\mathbf{V}^{-1}$ , where  $\mathbf{V}$  satisfies (2.2.5). Then, under the null hypothesis of adequacy, the Hosking-Ljung-Box test statistic defined by (2.3.4) converges asymptotically toward the following asymptotic distribution :*

$$\text{HLB}_n(m) \xrightarrow{L} \sum_{i=1}^{d^2 m} \xi_i \chi_i^2,$$

where the weights  $\xi_i$ ,  $i = 1, \dots, d^2 m$ , denote the eigenvalues of the matrix  $\mathbf{A}\boldsymbol{\Omega}_m$ , and  $\mathbf{A} = \mathbf{I}_m \otimes \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}$ . The random variables  $\chi_i^2$ ,  $i = 1, \dots, d^2 m$  correspond to independent chi-squared random variables with one degree of freedom.

*Proof :* Note that  $\hat{\boldsymbol{\Omega}}_m$  and  $\hat{\boldsymbol{\Sigma}}$  are consistent for  $\boldsymbol{\Omega}_m$  and  $\boldsymbol{\Sigma}$ , respectively. The asymptotic limit is found invoking a result from Box [19, Theorem 2.1].  $\square$

A serious advantage of that test statistic is that it avoids the computation of the inverse of  $\hat{\boldsymbol{\Omega}}_m$ . However, its asymptotic distribution is more complicated. As pointed out by Ng and al. (2016), it is no longer chi-squared. However, it is still tractable, using standard results on quadratic forms.

In practice  $\boldsymbol{\Omega}_m$  and its eigenvalues are unknown but they can be consistently estimated. The testing procedure based on  $\text{HLB}_n(m)$  at the  $\alpha$  level consists to reject the null hypothesis if  $\text{HLB}_n(m) > q_{1-\alpha}$ , where  $q_{1-\alpha}$  is the estimated quantile of  $\sum_{i=1}^{d^2 m} \hat{\xi}_i \chi_i^2$ , where  $\hat{\xi}_i$  denote the eigenvalues of  $\hat{\mathbf{A}}\hat{\boldsymbol{\Omega}}_m$ , where  $\hat{\mathbf{A}} = \mathbf{I}_m \otimes \hat{\boldsymbol{\Sigma}}^{-1} \otimes \hat{\boldsymbol{\Sigma}}^{-1}$ . Using modern statistical software, it is straightforward to find the quantiles of a finite weighted sum of independent chi-squared variables. For example, the **R** package `CompQuadForm` developed by Duchesne and Lafaye de Micheaux (2010) can serve that purpose.

Since  $\Sigma$  is assumed invertible and noting that  $\mathbf{A}\Omega_m = \mathbf{I}_{d^2m} - \mathbf{A}\mathbf{X}\mathbf{V}^{-1}\mathbf{X}^\top$ , it follows that the matrix  $\mathbf{A}\Omega_m$  has at least  $d^2m - \text{rank}(\mathbf{X})$  eigenvalues which are equal to unity. This suggests that if the eigenvalues different from one are (very) small, the distribution of  $\text{HLB}_n(m)$  can be approximated under certain circumstances by a  $\chi^2_{d^2m - \text{rank}(\mathbf{X})}$  distribution. Several  $\chi^2$  approximations will be considered in Section 2.4, compared to the one using the correct procedure relying on the quantiles from a finite weighted sum of independent chi-squared variables.

In diagnostic checking autoregressive conditional heteroscedastic (ARCH) models, Duchesne and Francq (2008) compared empirically test statistics similar in spirit to  $\text{HLB}_n(m)$  and  $\text{NLY}_n^-(m)$ . No clear winner has been found between these two statistical procedures. In general, the empirical powers were rather close. Consequently, each test statistic has its own merit.

### 2.3.2. Kernel-based test statistics

We now describe kernel-based test statistics which are consistent under general conditions for a large class of alternatives. Let  $\mathbf{f}_e(\omega)$  be the spectral density of  $\{\mathbf{e}_t\}$ , as defined by (2.3.1). Then under the null hypothesis of adequacy of a vMEM model, the spectral density of  $\{\mathbf{e}_t\}$  becomes a constant matrix :

$$\mathcal{H}_0 : \mathbf{f}_e(\omega) = \frac{\mathbf{\Gamma}_u(0)}{2\pi}, \quad \forall \omega \in [-\pi, \pi],$$

and under  $\mathcal{H}_0$  we have the equality  $\mathbf{\Gamma}_e(0) = \mathbf{\Gamma}_u(0)$ . When the model is misspecified, if  $\mathbf{\Gamma}_e(h) \neq \mathbf{0}$  for a certain lag, then the spectral density does not reduce to a constant in  $\omega$  :

$$\mathcal{H}_1 : \mathbf{f}_e(\omega) \neq \frac{\mathbf{\Gamma}_u(0)}{2\pi}.$$

For a given covariance matrix  $\mathbf{\Gamma}_u(0)$ , let us consider the following standardized distance measure :

$$Q^2(\mathbf{f}_1; \mathbf{f}_2) = 2\pi \int_{-\pi}^{\pi} \text{tr} \left[ \mathbf{\Gamma}_u^{-1}(0) \{\mathbf{f}_1(\omega) - \mathbf{f}_2(\omega)\}^* \mathbf{\Gamma}_u^{-1}(0) \{\mathbf{f}_1(\omega) - \mathbf{f}_2(\omega)\} \right] d\omega.$$

For a matrix  $\mathbf{A}$ ,  $\mathbf{A}^*$  denotes the transposed conjugate of the matrix  $\mathbf{A}$ , that is  $\mathbf{A}^* = \bar{\mathbf{A}}^\top$ . Under the null hypothesis,  $\mathbf{C}_u(0) = \mathbf{\Gamma}_u(0) + \mathbf{O}_P(n^{-1/2})$ . Thus,

arguments similar to those found in Duchesne & Roy [36, Proposition 1] yield :

$$Q^2(\hat{\mathbf{f}}_e; \mathbf{f}_0) = 2 \sum_{h=1}^{n-1} k^2(h/m_n) \text{tr} \left\{ \mathbf{C}_{\hat{\epsilon}}(h) \mathbf{C}_{\hat{\epsilon}}^{-1}(0) \mathbf{C}_{\hat{\epsilon}}^{\top}(h) \mathbf{C}_{\hat{\epsilon}}^{-1}(0) \right\} + o_P(m_n^{1/2}/n).$$

For checking vMEM models, a possible kernel-based test statistic is the following standardized measure :

$$T_n(k) = \frac{n \sum_{h=1}^{n-1} k^2(h/m_n) \text{tr} \left\{ \mathbf{C}_{\hat{\epsilon}}(h) \mathbf{C}_{\hat{\epsilon}}^{-1}(0) \mathbf{C}_{\hat{\epsilon}}^{\top}(h) \mathbf{C}_{\hat{\epsilon}}^{-1}(0) \right\} - d^2 M_n(k)}{\{2d^2 V_n(k)\}^{1/2}}, \quad (2.3.5)$$

where  $M_n(k) = \sum_{h=1}^{n-1} (1 - h/n) k^2(h/m_n)$  and  $V_n(k) = \sum_{h=1}^{n-1} (1 - h/n)(1 - (h+1)/n) k^4(h/m_n)$ . To study the asymptotic properties of  $T_n(k)$ , it is assumed that  $m_n \rightarrow \infty$  but such that  $m_n/n \rightarrow 0$ , which is a notable difference in the context of spectral density estimation.

**Example 1.** *Using the truncated uniform kernel  $k_{TR}(z) = I\{|z| \leq 1\}$ , simplifications occur :*

$$T_n(k_{TR}) = \frac{HLB_n(m_n) - d^2 m_n}{\{2d^2 m_n\}^{1/2}},$$

where

$$\begin{aligned} HLB_n(m_n) &= n \sum_{h=1}^{m_n} \mathbf{c}_{\hat{\epsilon}}^{\top}(h) \left\{ \hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1} \right\} \mathbf{c}_{\hat{\epsilon}}(h), \\ &= n \sum_{h=1}^{m_n} \text{tr} \left\{ \mathbf{C}_{\hat{\epsilon}}(h) \mathbf{C}_{\hat{\epsilon}}^{-1}(0) \mathbf{C}_{\hat{\epsilon}}^{\top}(h) \mathbf{C}_{\hat{\epsilon}}^{-1}(0) \right\}. \end{aligned}$$

The second equality comes from the relation

$\{\text{vec}(\mathbf{A})\}^{\top} (\mathbf{D} \otimes \mathbf{B}) \text{vec}(\mathbf{C}) = \text{tr}(\mathbf{A}^{\top} \mathbf{B} \mathbf{C} \mathbf{D}^{\top})$ . The test statistic  $HLB_n(m_n)$  affords a multivariate portmanteau test statistic, as the one studied in Hosking (1980). We call  $HLB_n(m_n)$  the Hosking-Ljung-Box test statistic, since this method yields the multivariate generalization of the popular Ljung-Box statistic used in diagnostic checking univariate time series models. See Li (2004) for a description of the portmanteau methods for various models. Intuitively, the standardizing factors  $d^2 m_n$  and  $2d^2 m_n$  correspond simply to the mean and variance terms of  $HLB_n(m_n)$ , respectively.

A first result establishes the asymptotic null distribution of  $T_n(k)$  when diagnostic checking the vMEM models.



**Theorem 1.** *Suppose that  $\{\mathbf{X}_t\}$  is a vMEM stochastic process as described in (2.2.1). Under Assumptions A and B, if  $m_n \rightarrow \infty$  such that  $m_n/n \rightarrow 0$ , the test statistic  $T_n(k)$  defined by (2.3.5) converges in distribution toward a standard normal distribution :*

$$T_n(k) \xrightarrow{L} \mathcal{N}(0,1).$$

From the results of Duchesne and Roy (2004), when the residuals are replaced by the innovations and that  $\{\mathbf{u}_t\}$  is an iid white noise process with finite fourth order moment, that is  $E\{u_t^4(i)\} < \infty$ ,  $i = 1, \dots, d$ , the following convergence result holds true :

$$\tilde{T}_n(k) = \frac{n \sum_{h=1}^{n-1} k^2 (h/m_n) \text{tr} \{ \mathbf{C}_u(h) \mathbf{C}_u^{-1}(0) \mathbf{C}_u^\top(h) \mathbf{C}_u^{-1}(0) \} - d^2 M_n(k)}{\{2d^2 V_n(k)\}^{1/2}} \xrightarrow{L} \mathcal{N}(0,1). \quad (2.3.6)$$

The asymptotic distribution of the test  $T_n(k)$  does not come directly from the results of Duchesne and Roy (2004), which are valid only for VAR models with exogenous variables. It is useful to introduce the weights  $k_{hn} = k(h/m_n)$ ,  $h = 1, \dots, n-1$ . The technical achievements in the proof of Theorem 1 are that the residuals defined by (2.2.6) of the vMEM model (2.2.1) satisfy the relations :

$$\begin{aligned} \sum_{h=1}^{n-1} k_{hn}^2 \text{tr} \{ \mathbf{C}_{\hat{\mathbf{u}}}(h) \mathbf{C}_{\hat{\mathbf{u}}}^{-1}(0) \mathbf{C}_{\hat{\mathbf{u}}}^\top(h) \mathbf{C}_{\hat{\mathbf{u}}}^{-1}(0) - \mathbf{C}_{\hat{\boldsymbol{\epsilon}}}(h) \mathbf{C}_{\hat{\boldsymbol{\epsilon}}}^{-1}(0) \mathbf{C}_{\hat{\boldsymbol{\epsilon}}}^\top(h) \mathbf{C}_{\hat{\boldsymbol{\epsilon}}}^{-1}(0) \} &= o_P(m_n^{1/2}, \mathfrak{B}17) \\ \sum_{h=1}^{n-1} k_{hn}^2 \text{tr} \{ \mathbf{C}_u(h) \mathbf{C}_u^{-1}(0) \mathbf{C}_u^\top(h) \mathbf{C}_u^{-1}(0) - \mathbf{C}_{\hat{\boldsymbol{\epsilon}}}(h) \mathbf{C}_{\hat{\boldsymbol{\epsilon}}}^{-1}(0) \mathbf{C}_{\hat{\boldsymbol{\epsilon}}}^\top(h) \mathbf{C}_{\hat{\boldsymbol{\epsilon}}}^{-1}(0) \} &= o_P(m_n^{1/2}, \mathfrak{B}18) \end{aligned}$$

From relations (2.3.7) and (2.3.8), a direct consequence is  $T_n(k) = \tilde{T}_n(k) + o_P(1)$  and the result follows. The detailed proof is provided in the Appendix. Interestingly,  $T_n(k)$  needs only  $n^{1/2}$ -consistency and is not limited to GMM estimation. The asymptotic distribution (2.3.6) is thus valid as long as Assumptions A and B are satisfied. However, the statistic  $T_n(k)$  relies on the random weighted sum  $\frac{n}{2} \hat{Q}(\hat{\mathbf{f}}_e; \mathbf{f}_0)$ , where :

$$\hat{Q}(\hat{\mathbf{f}}_e; \mathbf{f}_0) = 2 \sum_{h=1}^{n-1} k_{hn}^2 \text{tr} \left\{ \mathbf{C}_{\hat{\boldsymbol{\epsilon}}}(h) \mathbf{C}_{\hat{\boldsymbol{\epsilon}}}^{-1}(0) \mathbf{C}_{\hat{\boldsymbol{\epsilon}}}^\top(h) \mathbf{C}_{\hat{\boldsymbol{\epsilon}}}^{-1}(0) \right\}, \quad (2.3.9)$$

which is composed of positive terms. The exact distribution of  $T_n(k)$  is thus bounded below by the negative number  $-dM_n(k)/\{2V_n(k)\}^{1/2} = O(m_n^{1/2})$ , while the support of the limiting distribution is the real axis. Empirical studies suggest that the exact distribution of  $T_n(k)$  is in fact highly right-skewed, with significant over-rejection of the null hypothesis. See for example Hong (1996) or Duchesne and Roy (2004), among others. Chen and Deo (2004) studied this problem in a general framework. They noted that random variables which are positive linear combinations of positive independent random variables can have heavily right-skewed exact distributions. They proposed a power transformation which generalizes the Wilson-Hilferty cube root transformation for  $\chi^2$  random variables. Poulin and Duchesne (2008) derived the appropriate power transformation for test statistics such as  $T_n(k)$ . They found that the power transformation is invariant to the dimension. Note that they also considered the parametric bootstrap to find the critical values, which provided in passing excellent results in VAR/VARMA time series models. However, in the present context of financial data for multivariate non-linear time series models, the bootstrap remains numerically intensive, even with the modern computer resources. For example, in the time series described of Cipollini (2013), the sample sizes were always more than 3,200 observations. In practical applications, this suggests that the number of bootstrap replicates cannot be very large with large systems. Adding the inherent problem to estimate efficiently vMEM models (see Cipollini (2010)), alternative methods are desirable. This leads us to our proposal for diagnostic checking vMEM models with kernel methods, based on the Chen-Deo power transformation which is fast and easy to compute :

$$T_n^\dagger(k) = \frac{\left\{ \frac{n}{2} \hat{Q}(\hat{\mathbf{f}}_e; \mathbf{f}_0) \right\}^\beta - [\{M_n^\dagger(k)\}^\beta + \frac{1}{2}\beta(\beta-1)\{M_n^\dagger(k)\}^{\beta-2}\{2V_n^\dagger(k)\}]}{\beta\{M_n^\dagger(k)\}^{\beta-1}\{2V_n^\dagger(k)\}^{1/2}}, \quad (2.3.10)$$

where  $\hat{Q}(\hat{\mathbf{f}}_e; \mathbf{f}_0)$  is defined by (2.3.9). The standardization terms rely on  $M_n^\dagger(k) = \text{tr}(\mathbf{K}^\dagger \hat{\boldsymbol{\Omega}}_{n-1})$  and  $V_n^\dagger(k) = \text{tr}(\mathbf{K}^\dagger \hat{\boldsymbol{\Omega}}_{n-1} \mathbf{K}^\dagger \hat{\boldsymbol{\Omega}}_{n-1})$ , where  $\mathbf{K}^\dagger = \mathbf{K}^2 \otimes \hat{\boldsymbol{\Sigma}}^{-1} \otimes \hat{\boldsymbol{\Sigma}}^{-1}$  and  $\mathbf{K} = \text{diag}(k_{1n}, \dots, k_{n-1,n})$ . The Chen-Deo power transformation coefficient

$\beta$  relies on the chosen kernel function  $k(\cdot)$  and is given by :

$$\beta = 1 - \frac{2 \left( \sum_{h=1}^{n-1} k_{hn}^2 \right) \left( \sum_{h=1}^{n-1} k_{hn}^6 \right)}{3 \left( \sum_{h=1}^{n-1} k_{hn}^4 \right)^2}.$$

Using the  $\delta$ -method (see Serfling (1980)), we have the following convergence result under the null hypothesis :

$$T_n^\dagger(k) \xrightarrow{L} \mathcal{N}(0, 1).$$

In the next section, the test statistics proposed in this section are evaluated empirically and compared to the test statistic of Ng and al. (2016).

## 2.4. SIMULATION EXPERIMENTS

### 2.4.1. Models and tests statistics included in the Monte Carlo study

In the previous section, we have studied two classes of test statistics which should prove useful for diagnosing vMEM time series models. From a practical point of view, it is natural to inquire about the finite sample properties of these test statistics, in particular their exact levels and powers. It seems of interest to compare the Hosking-Ljung-Box test statistic defined by (2.3.4) and the kernel-based tests (2.3.5) and (2.3.10). To partially answer such questions, we report in this section simulation results. More precisely, we studied the exact levels and the powers of the test statistics for different time series lengths. The nominal levels 5% and 10 % have been adopted in the study, for several time series lengths. In ARMA modelling, the ARMA(1,1) model yields often a good starting point. Similarly, the vMEM(1,1) model represents a workhorse in vMEM modelling.

TABLE 2.1. Model parameters for the time series models  $DGP_0$  and  $DGP_1$  defined by (2.4.1) and (2.4.3), respectively.

$$\begin{aligned} \boldsymbol{\alpha}_0 &= \begin{pmatrix} 0.24 \\ 0.37 \end{pmatrix}, & \mathbf{A}_1 &= \begin{pmatrix} 0.40 & 0.14 \\ 0.14 & 0.40 \end{pmatrix}, & \mathbf{A}_2 &= \begin{pmatrix} 0.02 & 0.10 \\ 0.10 & 0.05 \end{pmatrix}, \\ \mathbf{B} &= \begin{pmatrix} 0.1 & 0.0 \\ 0.0 & 0.2 \end{pmatrix}, & \boldsymbol{\Phi} &= \begin{pmatrix} 1.00 & 0.35 \\ 0.35 & 1.00 \end{pmatrix}, \end{aligned}$$

TABLE 2.2. Empirical levels (nominal levels 5% and 10%) of the tests statistics for checking the vMEM(1,1) model defined as (2.4.1). The following test statistics are included :  $T_n(k)$  and  $T_n^\dagger(k)$  for various kernels,  $HLB_n(m)$  with different decision rules,  $NLY_n(m)$  and  $NLY_n^-(m)$ . The lag order/smoothing parameters are  $m = m_n = [\log(n)]$ ,  $[3n^{0.2}]$ ,  $[3n^{0.3}]$ , with sample sizes  $n = 300, 400$  and  $500$ .

	Original kernel-based test statistics					Kernel-based tests with Chen-Deo power transformation					Tests based on $\chi^2$ or weighted sums of $\chi^2$ distributions				
	TR	BAR	DAN	PAR	BARPR	TR	BAR	DAN	PAR	BARPR	$H_{n,WN}$	$H_{n,adj}$	$HLB_n$	$NLY_n$	$NLY_n^-$
$n = 300$															
	$\alpha = 5\%$														
5	0.010	0.000	0.000	0.002	0.000	0.058	0.046	0.046	0.040	0.044	0.006	0.046	0.056	0.212	0.212
9	0.012	0.002	0.004	0.004	0.004	0.036	0.040	0.040	0.048	0.044	0.008	0.044	0.036	0.182	0.182
16	0.016	0.006	0.008	0.008	0.008	0.030	0.034	0.028	0.034	0.032	0.012	0.038	0.030	0.126	0.126
	$\alpha = 10\%$														
5	0.024	0.002	0.002	0.002	0.002	0.108	0.072	0.074	0.070	0.074	0.014	0.096	0.108	0.262	0.262
9	0.024	0.002	0.006	0.006	0.006	0.072	0.076	0.074	0.086	0.078	0.024	0.104	0.072	0.220	0.220
16	0.036	0.008	0.012	0.012	0.012	0.060	0.098	0.082	0.080	0.088	0.030	0.090	0.060	0.134	0.134
$n = 400$															
	$\alpha = 5\%$														
5	0.012	0.000	0.000	0.002	0.000	0.050	0.040	0.042	0.042	0.040	0.008	0.044	0.050	0.184	0.184
9	0.022	0.002	0.002	0.002	0.002	0.050	0.040	0.042	0.042	0.042	0.014	0.046	0.050	0.226	0.226
18	0.034	0.006	0.012	0.012	0.010	0.052	0.050	0.048	0.054	0.054	0.032	0.030	0.052	0.134	0.134
	$\alpha = 10\%$														
5	0.024	0.000	0.002	0.002	0.002	0.106	0.074	0.076	0.086	0.082	0.020	0.086	0.106	0.234	0.234
9	0.040	0.002	0.004	0.008	0.002	0.110	0.098	0.096	0.108	0.106	0.036	0.088	0.110	0.270	0.270
18	0.054	0.014	0.026	0.030	0.028	0.086	0.096	0.086	0.102	0.102	0.052	0.080	0.086	0.154	0.154
$n = 500$															
	$\alpha = 5\%$														
6	0.008	0.002	0.002	0.002	0.002	0.068	0.058	0.052	0.058	0.056	0.008	0.050	0.066	0.172	0.172
10	0.020	0.006	0.006	0.008	0.006	0.048	0.048	0.048	0.056	0.052	0.012	0.048	0.048	0.210	0.210
19	0.016	0.006	0.008	0.008	0.008	0.040	0.052	0.040	0.040	0.046	0.010	0.066	0.040	0.148	0.148
	$\alpha = 10\%$														
6	0.020	0.002	0.002	0.006	0.006	0.114	0.096	0.092	0.096	0.096	0.020	0.116	0.114	0.230	0.230
10	0.032	0.008	0.008	0.008	0.008	0.090	0.116	0.104	0.110	0.116	0.024	0.112	0.090	0.260	0.260
19	0.048	0.012	0.016	0.024	0.018	0.080	0.094	0.080	0.092	0.094	0.042	0.102	0.080	0.178	0.178

To compare the exact distributions of the test statistics with their corresponding asymptotic (or approximate) distributions, the following bivariate data generating process (DGP) was used :

$$\text{DGP}_0 : \boldsymbol{\mu}_t = \boldsymbol{\alpha}_0 + \mathbf{A}_1 \mathbf{X}_{t-1} + \mathbf{B} \boldsymbol{\mu}_{t-1}, \quad (2.4.1)$$

where in the vMEM specification  $\mathbf{X}_t = \boldsymbol{\mu} \odot \boldsymbol{\epsilon}_t$ , the process  $\boldsymbol{\epsilon} = \{\boldsymbol{\epsilon}_t, t \in \mathbb{Z}\}$ , with  $\boldsymbol{\epsilon}_t = (\epsilon_t(1), \epsilon_t(2))^\top$ , was assumed an iid white noise such that the distribution of  $\boldsymbol{\epsilon}_t$  is constructed with a normal copula with correlation matrix  $\boldsymbol{\Phi}$ , and the marginal distributions were assumed of the Gamma type. More precisely,  $\epsilon_t(1) \sim \Gamma(8.33, 8.33)$  and  $\epsilon_t(2) \sim \Gamma(10.5, 10.5)$ . The model parameters are expressed in Table 2.1. The mathematical expectation of  $\mathbf{X}_t$  satisfies  $\boldsymbol{\mu}_0 = E(\mathbf{X}_t) = E(\boldsymbol{\mu}_t) = (0.85, 1.75)^\top$ . The structure of the matrix  $\mathbf{B}$  was supposed known and thus only the parameters on the main diagonal were estimated.

We describe the test statistics included in the study. First, we considered kernel-based test statistics. We considered five kernels : in addition to the truncated uniform kernel (TR), we included the Bartlett (BAR), Daniell (DAN), Parzen (PAR) and finally the Bartlett-Priestley (BARPR) kernels. The precise definitions of these kernels can be found in Duchesne and Roy (2004). We studied the original kernel-based versions defined by (2.3.5), and also the ones relying on the Chen-Deo power transformation, given by (2.3.10). Three rates were adopted :  $m_n = [\log(n)]$ ,  $m_n = [3n^{0.2}]$  and  $m_n = [3n^{0.3}]$ . Note that to allow fair comparisons, we considered  $m \equiv m_n$ . In the Tables, the results of the original kernel-based tests are regrouped under the heading ‘Original kernel-based test statistics’, while the ones based on the Chen-Deo power transformation appear in the column entitled ‘Kernel-based tests with Chen-Deo transformation’.

We also included test statistics based on quadratic forms with asymptotic or approximate  $\chi^2$  distributions or weighted sums of  $\chi^2$  distributions. The Hosking-Ljung-Box test statistic  $HLB_n(m)$  and the Ng-Li-Yu test statistic  $NLY_n(m)$  defined by (2.3.4) and (2.3.2), respectively, were included in the study. We calculated the empirical quantiles of  $HLB_n(m)$  using the Imhof algorithm included in the **R** package **CompQuadForm**. We investigated also the classical Hosking test statistic defined as :

$$H_n(m) = n \sum_{h=1}^m \frac{n}{n-h} \left\{ \mathbf{c}_{\hat{\epsilon}}^\top(h) \left( \hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1} \right) \mathbf{c}_{\hat{\epsilon}}(h) \right\}. \quad (2.4.2)$$

In that later case, the quantiles were determined using two approximate rules. The first rule assumes that the residuals can be treated as an iid white noise, and thus a usual white noise test relying on the quantile of the  $\chi_{d^2 m}^2$  distribution is performed. Given that the  $DGP_0$  is a structured vMEM(1,1) process and that the diagonal form of the matrix **B** in (2.4.1) is known, the second approximation rule uses the  $\chi_{d^2 m - d_\theta}^2$  distribution, where  $d_\theta$  is the number of estimated parameters (without the mean term, that is  $d_\theta = 6$ ). This procedure is valid in VAR models with linear constraints. See for example Lütkepohl [77, Chap.5]. This ad hoc rule adjusts the degrees of freedom due to the use of residuals and constrained estimation. We also included the modified Ng-Li-Yu test statistic  $NLY_n^-(m)$ , based

on the Moore-Penrose inverse  $\hat{\Omega}_m^-$ , in which the estimated rank of  $\hat{\Omega}_m$ , denoted  $r_m$ , is calculated. The decision rule is thus to reject the null hypothesis if the test is larger than the quantile  $\chi^2_{1-\alpha, r_m}$ . In the Tables, these tests are grouped in ‘Tests based on  $\chi^2$  or weighted sums of  $\chi^2$  distributions’. The results for the level study are presented in Table 2.2.

TABLE 2.3. Exact powers (nominal levels 5% and 10%) of the tests statistics for checking a vMEM(1,1) model defined as (2.4.1) when the true model is DGP<sub>1</sub>, based on asymptotic or approximate quantiles. The following test statistics are included :  $T_n(k)$  and  $T_n^+(k)$  for various kernels,  $HLB_n(m)$  with different decision rules,  $NLY_n(m)$  and  $NLY_n^-(m)$ . The lag order/smoothing parameters are  $m = m_n = [\log(n)], [3n^{0.2}], [3n^{0.3}]$ , with  $n = 400, 500, 700$  and 1000.

	Original kernel-based					Kernel-based tests with					Tests based on $\chi^2$ or				
	test statistics					Chen-Deo power transformation					weighted sums of $\chi^2$ distributions				
	TR	BAR	DAN	PAR	BARPR	TR	BAR	DAN	PAR	BARPR	$H_{n,WN}$	$H_{n,adj}$	$HLB_n$	$NLY_n$	$NLY_n^-$
$n = 400$	$\alpha = 5\%$														
5	0.176	0.182	0.186	0.192	0.186	0.344	0.534	0.524	0.524	0.520	0.156	0.350	0.344	0.456	0.456
9	0.144	0.192	0.204	0.200	0.204	0.262	0.472	0.440	0.428	0.438	0.130	0.226	0.262	0.358	0.358
18	0.132	0.174	0.168	0.168	0.166	0.182	0.362	0.314	0.312	0.328	0.118	0.192	0.182	0.246	0.246
	$\alpha = 10\%$														
5	0.232	0.226	0.240	0.240	0.236	0.456	0.614	0.592	0.590	0.604	0.224	0.484	0.456	0.502	0.502
9	0.208	0.246	0.252	0.238	0.248	0.380	0.572	0.530	0.516	0.542	0.190	0.324	0.380	0.414	0.414
18	0.200	0.226	0.224	0.224	0.224	0.270	0.484	0.408	0.406	0.442	0.184	0.278	0.268	0.270	0.270
$n = 500$	$\alpha = 5\%$														
6	0.184	0.168	0.180	0.184	0.184	0.300	0.556	0.532	0.520	0.532	0.168	0.388	0.300	0.416	0.416
10	0.148	0.188	0.192	0.188	0.192	0.236	0.476	0.424	0.404	0.424	0.132	0.314	0.232	0.356	0.356
19	0.148	0.176	0.180	0.168	0.176	0.168	0.340	0.276	0.276	0.284	0.132	0.218	0.168	0.288	0.292
	$\alpha = 10\%$														
6	0.228	0.240	0.248	0.260	0.256	0.416	0.660	0.636	0.628	0.632	0.216	0.502	0.412	0.500	0.500
10	0.188	0.256	0.260	0.244	0.256	0.352	0.572	0.516	0.524	0.532	0.172	0.408	0.352	0.392	0.392
19	0.180	0.220	0.208	0.204	0.208	0.228	0.452	0.372	0.360	0.388	0.172	0.336	0.228	0.320	0.324
$n = 700$	$\alpha = 5\%$														
6	0.280	0.370	0.400	0.385	0.400	0.455	0.665	0.630	0.625	0.645	0.260	0.480	0.450	0.510	0.510
11	0.205	0.360	0.360	0.335	0.355	0.370	0.600	0.560	0.555	0.565	0.190	0.372	0.370	0.440	0.440
21	0.200	0.315	0.290	0.285	0.290	0.250	0.505	0.425	0.415	0.440	0.185	0.304	0.250	0.355	0.355
	$\alpha = 10\%$														
6	0.365	0.425	0.430	0.440	0.425	0.615	0.780	0.755	0.745	0.770	0.355	0.604	0.615	0.575	0.575
11	0.300	0.435	0.415	0.400	0.420	0.450	0.690	0.645	0.640	0.650	0.295	0.496	0.450	0.510	0.510
21	0.270	0.375	0.340	0.335	0.345	0.335	0.625	0.530	0.530	0.545	0.255	0.412	0.335	0.375	0.375
$n = 1000$	$\alpha = 5\%$														
6	0.420	0.505	0.515	0.500	0.505	0.610	0.830	0.825	0.795	0.825	0.385	0.640	0.610	0.630	0.630
11	0.300	0.490	0.480	0.465	0.480	0.450	0.760	0.725	0.710	0.715	0.265	0.488	0.450	0.490	0.490
23	0.270	0.420	0.360	0.365	0.365	0.330	0.605	0.540	0.520	0.545	0.245	0.336	0.525	0.390	0.390
	$\alpha = 10\%$														
6	0.510	0.550	0.560	0.570	0.565	0.720	0.880	0.865	0.860	0.865	0.485	0.768	0.720	0.680	0.680
11	0.385	0.545	0.540	0.525	0.540	0.590	0.835	0.795	0.780	0.790	0.355	0.600	0.590	0.555	0.555
23	0.360	0.500	0.470	0.460	0.460	0.460	0.730	0.660	0.655	0.660	0.335	0.444	0.525	0.450	0.450

To study the power of the test statistics, we considered the following DGP :

$$\text{DGP}_1 : \boldsymbol{\mu}_t = \boldsymbol{\alpha}_0 + \mathbf{A}_1 \mathbf{X}_{t-1} + \mathbf{A}_2 \mathbf{X}_{t-2} + \mathbf{B} \boldsymbol{\mu}_{t-1}, \quad (2.4.3)$$

where the innovation process was the same that under  $DGP_0$ . The matrix  $\mathbf{A}_2$  is given in Table 2.1. The same structured vMEM(1,1) model has been estimated. We based our inferences using the quantiles coming from the asymptotic theory and also the ones relying on the approximate rules. The results are presented in Table 2.3. In order to adjust the results for the levels, we extracted the quantiles from the level study, and we calculated the empirical powers using these empirical quantiles, giving what we call size-adjusted powers. The results are presented in Table 2.4. For each sample size and under the scenarios  $DGP_0$  or  $DGP_1$ , a total of 1,000 replications has been considered for each experiment.

#### 2.4.2. Discussion of the results

First, we discuss the empirical results for the level study under  $DGP_0$ . They are presented in Table 2.2. In general, for all the methods, the approximation of the exact distribution by the asymptotic one improved with the sample size  $n$ . For the kernel-based test statistics  $T_n^\dagger(k)$ , the approximations were very satisfying at the 5% and 10% levels. Considerable under-rejection occurred for the original versions  $T_n(k)$ , giving conservative tests for all the kernels. As expected, the finite sample distributions of the original kernel-based test statistics were highly right-skewed, explaining these empirical findings. Thus the rest of the discussion on kernel methods concentrate on the test statistics based on the Chen-Deo power transformation.

As a function of the kernel  $k(\cdot)$ , the test statistics  $T_n^\dagger(k)$  performed similarly, except for the kernel  $k_{TR}(\cdot)$  which seemed superior for  $n = 300, 400$  and small values of  $m_n$ . For large  $n$  and higher rates of  $m_n$ , the observed differences were smaller. At the 5% significance level, all kernels (with a fixed value of  $m_n$ ) lead to rejection rates close to the nominal level when  $n = 300$ , with some under-rejection as  $m_n$  became larger. At  $n = 400, 500$ , all rejection rates were within two standard errors of 5%, that is in the interval [3.6, 6.3]. The same general conclusions have been observed at the 10% nominal level, where the empirical levels appeared very close to the nominal one for  $n = 400, 500$ .



The test statistic with the approximate decision rule  $H_{n,WN}(m_n)$  offered empirical levels relatively close to the nominal sizes, but they were generally inferior to those of  $H_{n,adj}(m_n)$  and  $HLB_n(m_n)$ . Thus  $H_{n,WN}(m_n)$  yielded a very conservative test for all combinations of  $m$  and  $n$ . In several cases, the approximation  $H_{n,adj}(m_n)$  seemed very satisfying, even if to use the quantiles from the weighted sum of  $\chi^2$  random variables gave in several cases better results. The asymptotic distribution  $H_{n,adj}(m_n)$  had at least  $4m - 6$  eigenvalues equal to one, which is rather large with respect to  $d_\theta = 6$  for the rates considered. For several cases we computed explicitly the eigenvalues :  $4m - 6$  were equal to one, while the others were small (e.g., for  $n = 500$ ,  $m \in \{5, 10, 19\}$ , the eigenvalues different from one were typically less than  $2 \times 10^{-1}$ ). This may partially explain the good performance of the approximate rule in our experiment.

TABLE 2.4. Exact powers (nominal levels 5% and 10%) of the tests statistics for checking a vMEM(1,1) model defined as (2.4.1) when the true model is DGP<sub>1</sub>, based on empirical quantiles found in the level study. The following test statistics are included :  $T_n(k)$  and  $T_n^+(k)$  for various kernels,  $HLB_n(m)$  with different decision rules,  $NLY_n(m)$  and  $NLY_n^-(m)$ . The lag order/smoothing parameters are  $m = m_n = [\log(n)], [3n^{0.2}], [3n^{0.3}]$ , with  $n = 400, 500, 700$  and 1000.

	Original kernel-based					Kernel-based tests with					Tests based on $\chi^2$ or				
	test statistics					Chen-Deo power transformation					weighted sums of $\chi^2$ distributions				
	TR	BAR	DAN	PAR	BARPR	TR	BAR	DAN	PAR	BARPR	$H_{n,WN}$	$H_{n,adj}$	$HLB_n$	$NLY_n$	$NLY_n^-$
$n = 400$	$\alpha = 5\%$														
5	0.348	0.578	0.563	0.570	0.568	0.388	0.560	0.542	0.548	0.552	0.345	0.345	0.348	0.165	0.165
9	0.315	0.525	0.505	0.485	0.485	0.250	0.483	0.422	0.453	0.461	0.315	0.315	0.315	0.143	0.143
18	0.238	0.413	0.388	0.358	0.380	0.224	0.352	0.314	0.284	0.282	0.235	0.235	0.238	0.188	0.188
	$\alpha = 10\%$														
5	0.445	0.728	0.713	0.675	0.718	0.451	0.623	0.568	0.599	0.598	0.443	0.443	0.445	0.288	0.288
9	0.445	0.623	0.570	0.550	0.573	0.320	0.592	0.560	0.545	0.579	0.443	0.443	0.445	0.220	0.220
18	0.328	0.510	0.480	0.445	0.473	0.282	0.461	0.411	0.398	0.410	0.330	0.330	0.328	0.268	0.268
$n = 500$	$\alpha = 5\%$														
6	0.360	0.698	0.660	0.653	0.665	0.423	0.601	0.585	0.571	0.582	0.353	0.353	0.360	0.230	0.228
10	0.305	0.605	0.543	0.523	0.560	0.291	0.538	0.490	0.492	0.493	0.303	0.303	0.305	0.135	0.135
19	0.210	0.453	0.403	0.373	0.390	0.271	0.412	0.341	0.343	0.341	0.208	0.208	0.210	0.143	0.140
	$\alpha = 10\%$														
6	0.480	0.793	0.783	0.773	0.783	0.482	0.680	0.640	0.636	0.656	0.470	0.470	0.480	0.323	0.320
10	0.415	0.708	0.658	0.605	0.645	0.394	0.638	0.602	0.582	0.590	0.415	0.415	0.415	0.225	0.223
19	0.300	0.550	0.503	0.515	0.505	0.354	0.525	0.466	0.445	0.449	0.293	0.293	0.300	0.240	0.238
$n = 700$	$\alpha = 5\%$														
6	0.496	0.772	0.768	0.728	0.760	0.525	0.705	0.699	0.676	0.679	0.488	0.488	0.496	0.232	0.232
11	0.416	0.696	0.644	0.624	0.640	0.339	0.626	0.591	0.580	0.568	0.412	0.412	0.416	0.144	0.144
21	0.240	0.552	0.500	0.468	0.488	0.323	0.471	0.432	0.371	0.410	0.240	0.240	0.240	0.128	0.128
	$\alpha = 10\%$														
6	0.584	0.816	0.800	0.772	0.800	0.580	0.770	0.784	0.781	0.784	0.584	0.584	0.584	0.364	0.364
11	0.532	0.744	0.720	0.704	0.712	0.388	0.712	0.693	0.656	0.691	0.532	0.532	0.532	0.200	0.200
21	0.404	0.624	0.600	0.576	0.584	0.390	0.590	0.547	0.554	0.554	0.400	0.400	0.404	0.240	0.240
$n = 1000$	$\alpha = 5\%$														
6	0.596	0.908	0.900	0.892	0.896	0.622	0.826	0.819	0.801	0.810	0.596	0.596	0.596	0.460	0.460
11	0.488	0.844	0.760	0.696	0.732	0.429	0.751	0.712	0.671	0.699	0.488	0.488	0.488	0.280	0.280
23	0.272	0.632	0.552	0.532	0.556	0.400	0.613	0.482	0.449	0.491	0.272	0.272	0.272	0.176	0.176
	$\alpha = 10\%$														
6	0.736	0.948	0.924	0.924	0.928	0.684	0.886	0.867	0.865	0.867	0.736	0.736	0.736	0.572	0.572
11	0.572	0.884	0.860	0.832	0.864	0.508	0.820	0.784	0.760	0.788	0.568	0.568	0.572	0.368	0.368
23	0.460	0.704	0.652	0.628	0.660	0.464	0.736	0.650	0.633	0.662	0.456	0.456	0.460	0.264	0.264

We found that the test statistics  $NLY_n(m)$  and  $NLY_n^-(m)$  were inferior. They gave similar results and they over-rejected strongly, particularly for small values of  $m$ . It seemed preferable to use the test statistics  $HLB_n(m_n)$  or even  $H_{n,adj}(m_n)$ ,

which offered better empirical levels, much closer to the nominal ones, at least in our experiments.

We now discuss the empirical powers. The results in Table 2.3 present the empirical powers using asymptotic or approximate quantiles, and Table 2.4 displays size-adjusted powers. For the original kernel-based tests, the results based on empirical and asymptotic quantiles differed considerably and these test statistics are thus not recommended for use. It appeared better to use the versions based on the Chen-Deo power transformation.

From Tables 2.3 and 2.4, the kernel-based tests calculated with the Chen-Deo power transformation, or the  $HLB_n$  test, the adjusted empirical powers were reasonably close to the empirical powers computed with the asymptotic quantiles. This is an interesting empirical finding, which suggests that the asymptotic distributions are close to the empirical ones and that the asymptotic quantiles of these methods are reliable in practical applications. That appears to be also the case for the test statistics  $H_{n,adj}$ . Due to the computation of the estimated quantiles from the weighted sum of  $\chi^2$  random variables, the test statistic  $HLB_n$  exhibited reasonable empirical levels and more empirical powers in Table 2.3 than in Table 2.4. In Table 2.4, the empirical powers of  $HLB_n$  and  $T_n(k_{TR})$  are identical, which is due to the fact that these methods are related by a linear transformation, see Example 1.

In addition, from Table 2.4, for a kernel and a value of  $m_n$ , the differences in size-adjusted powers between the original kernel-based test statistics and the ones relying in the Chen-Deo power transformation were generally unimportant. Thus for the kernel-based test we focus the rest of our power comparisons on the tests relying on the Chen-Deo power transformation.

Examining in more details the kernel-based test statistic with DAN, PAR and BARPR, we observed that they yielded very similar results. The truncated uniform kernel  $k_{TR}(\cdot)$  behave differently, and it was somewhat inferior. On the other hand, it seemed that the BAR kernel appeared slightly superior. These phenomena were perceptible with both the asymptotic and empirical quantiles.

From Table 2.4, substantial power improvements over the truncated uniform kernel or  $HLB_n$  occurred at several sample sizes and rates. Generally, for the chosen alternative, slower rates offered larger power. For example, at  $n = 500$  and  $m = 10$ ,  $T_n^\dagger(k_{BAR})$  appears almost twice more powerful than  $HLB_n(10)$  or  $T_n^\dagger(k_{TR})$ . However, it is useful to note that a large rate provided high power for  $HLB_n(10)$ , suggesting that in some cases it is still useful to consider truncated uniform weighting.

Since  $NLY_n$  and  $NLY_n^-$  over-rejected under the null hypothesis using the asymptotic quantiles, a false impression of high power was displayed for low values of  $m_n$ . In fact, the results from Table 2.4 reveal that these tests were less powerful than the proposed tests. For the cases investigated in this study, based on the empirical quantiles, the methods  $NLY_n$  and  $NLY_n^-$  were less powerful.

Given the nature of the alternative, putting more weight on smaller lags delivered powerful strategies. Thus, the kernel  $k_{TR}(\cdot)$  and the methods  $HLB_n(m)$ ,  $H_{n,adj}(m)$  and  $NLY_n(m)$  were generally the less powerful. The use of the kernel-based test statistics (with the Chen-Deo power transformation), relying on another kernel other than the truncated uniform one seems appropriate. In the next section, the proposed tests are used to validate models on real data. As seen on this section, the choice of  $m_n$  for the kernel-based methods may be delicate. Using graphical sensibility analyzes, we illustrate with a real application how the smoothing parameter may be chosen.

## 2.5. AN EMPIRICAL APPLICATION

To assess empirically the capacity of the diagnostic test statistics presented in the previous sections to check vMEM models, we will utilize data coming from the econometric literature. The sample consists of bivariate data on the stock index FTSE100. The data set describe absolute returns and realized volatilities between January 1996 and February 2009. After cleaning the data, the sample size is  $n = 2844$ . This bivariate sample has been analyzed by Cipollini (2013), as a stock index coming from the *Oxford Man Institute* (OMI) Realized Library,

see Shephard and Shephard (2010). Note that the data are available as part of the supplementary material of the paper of Cipollini (2013) on the *Journal of Applied Econometrics* website. The library contains daily returns and realized volatility kernels for ten stock indices, together with the range computed using daily measures (highs and lows) downloaded from Datastream. We present here only one representative stock. The variables are :

$$\begin{pmatrix} ret.abs_t \\ rv.vol_t \end{pmatrix} = \begin{pmatrix} 100|ret_t|\sqrt{252} \\ 100\sqrt{rv_t}\sqrt{252} \end{pmatrix}, \quad (2.5.1)$$

where  $ret.abs_t$  and  $rv_t$  denote the return and realized variance at time  $t$ , respectively. The variables have been defined as Cipollini (2013) and we also postulate the following structured vMEM(1,1) model :

$$\boldsymbol{\mu}_t = \boldsymbol{\alpha}_0 + \mathbf{A}\mathbf{X}_{t-1} + \mathbf{B}\boldsymbol{\mu}_{t-1}, \quad (2.5.2)$$

where

$$\boldsymbol{\alpha}_0 = \begin{pmatrix} \alpha_{01} \\ \alpha_{02} \end{pmatrix}, \quad \mathbf{A}_1 = \begin{pmatrix} a_{11,1} & a_{12,1} \\ a_{21,1} & a_{22,1} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} b_{11} & 0 \\ 0 & b_{22} \end{pmatrix}.$$

Using the **R** package `nleqslv`, the GMM estimators were calculated. The estimators of the model parameters with their associated standard errors are reported in Table 2.5.

We applied the test procedures  $T_n^\dagger(k_{TR})$ ,  $T_n^\dagger(k_{BAR})$ , as well as the test statistics  $HLB_n(m)$ ,  $NLY_n(m)$  and  $NLY_n^-(m)$ . We used only the kernel-based test statistics relying on the Chen-Deo power transformation, since from the simulation study of the previous section they displayed better empirical properties under the null hypothesis. Note that to use the truncated uniform and Bartlett kernels may be computationally efficient, given they have compact support. From the empirical study presented in Section 2.4, the Bartlett kernel appeared powerful, and the truncated uniform kernel remains a standard benchmark. All the  $P$ -values were highly significant, inferior to  $10^{-4}$ , suggesting an inadequate fit.

Here, the estimator of the  $\mathbf{\Omega}_m$  matrix displayed a large condition number. However, reasonable rules suggested that  $\hat{\mathbf{\Omega}}_m$  was invertible. Thus  $\text{NLY}_n$  and  $\text{NLY}_n^-$  delivered equal values.

We discuss the choice of  $m_n$ . In other contexts, data-driven choices of  $m_n$  have been suggested (see, e.g., Duchesne and Roy (2004)). In real applications, it may be preferable to plot the spectral density estimator for several lag orders, not just rely the testing inferences on a single punctual estimator. Using the Bartlett kernel, estimated spectral density estimators of the residuals are reported in Figure 2.1 for various smoothing parameters. We used the lag orders  $m_n = [\log(n)], [3n^{0.2}], [3n^{0.3}]$ , providing a visual sensibility analysis. From that Figure, rejection of the null hypothesis seemed mainly due to an inadequate fit of the second component of  $\mathbf{X}_t$ , since the estimated spectral density seemed significantly different from a constant. Incidentally, the sample autocorrelations of the residuals of the first component were compatible with a white noise, and we observed significant residual autocorrelations for the second component. Using the technical results of Ng and al. (2016) on the asymptotic distribution of the residual autocorrelations, the residual autocorrelations are reported in Table 2.6.

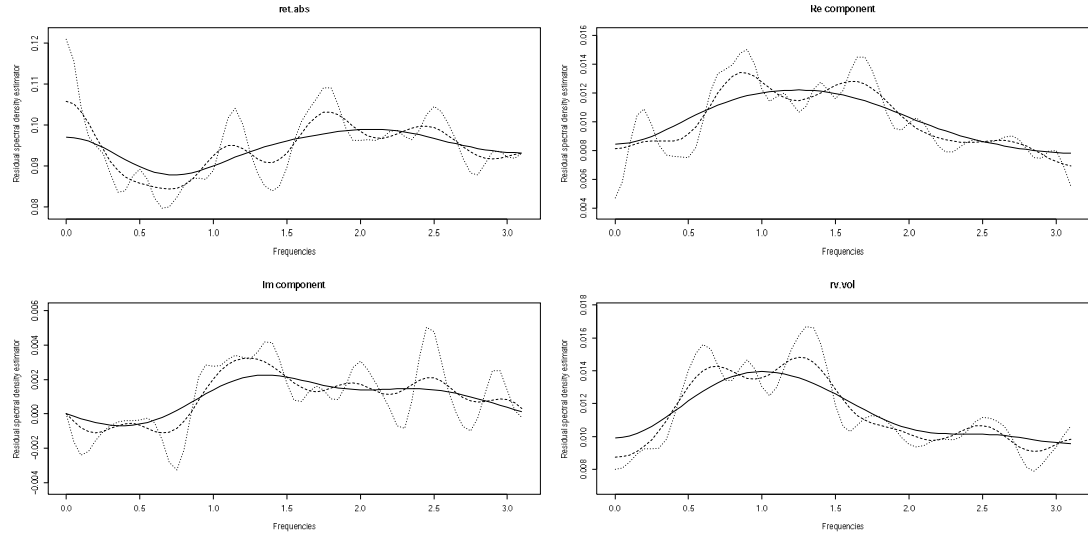
TABLE 2.5. Estimation results for modelling the variables (2.5.1) of the FTSE100 stock indice. Two models are estimated, the vMEM(1,1) model defined by (2.5.2) and the augmented model vMEM(2,1) (2.5.2). The estimators and the associated standard errors are given.

Parameters	vMEM(1,1) model		vMEM(2,1) model	
	Estimators	Standard Errors	Estimators	Standard Errors
$\hat{\alpha}_{01}$	0.038	0.139	0.054	0.136
$\hat{\alpha}_{02}$	0.250	0.050	0.104	0.027
$\hat{\alpha}_{11,1}$	-0.024	0.016	-0.048	0.019
$\hat{\alpha}_{21,1}$	0.386	0.005	0.044	0.006
$\hat{\alpha}_{12,1}$	0.416	0.056	0.401	0.070
$\hat{\alpha}_{22,1}$	0.316	0.018	0.401	0.022
$\hat{\alpha}_{11,2}$			0.041	0.020
$\hat{\alpha}_{21,2}$			-0.023	0.006
$\hat{\alpha}_{12,2}$			-0.014	0.100
$\hat{\alpha}_{22,2}$			-0.242	0.029
$\hat{b}_{11}$	0.641	0.047	0.651	0.069
$\hat{b}_{22}$	0.623	0.018	0.809	0.019
$\hat{\sigma}_{11}$	0.593		0.593	
$\hat{\sigma}_{12}$	0.063		0.064	
$\hat{\sigma}_{22}$	0.071		0.069	

TABLE 2.6. Autocorrelations for lags one to five and asymptotic variances (in parentheses) from the adjustment of the vMEM(1,1) model for modelling the variables (2.5.1) of the FTSE100 stock index.

lag	<i>ret.abs</i>	<i>rv.vol</i>
1	-0.016 (0.405)	0.067 (0.412)
2	-0.010 (0.772)	-0.070 (0.730)
3	0.032 (0.901)	-0.087 (0.927)
4	0.018 (0.949)	-0.045 (0.976)
5	0.020 (0.975)	0.003 (0.977)

FIGURE 2.1. Kernel-based spectral density estimator for the residuals of the adjustment of a vMEM(1,1) model to the bivariate time series corresponding to the stock index FTSE100. The Bartlett kernel is used for the lag orders  $m_n = \lceil \log(n) \rceil$  (solid line),  $\lceil 3n^{0.2} \rceil$  (dashed line), and  $\lceil 3n^{0.3} \rceil$  (dotted line).



For example, from Table 2.6, an approximate 95% confidence interval for the lag-1 residual autocorrelation associated to the variable  $rv.vol$  is  $0.067 \pm 1.96\sqrt{0.412/2844} = [0.04, 0.09]$ . Since the residual autocorrelations suggested unexplained dependence, we considered an augmented vMEM(2,1) model, adding the lagged term  $\mathbf{X}_{t-2}$  but keeping the diagonal structure of  $\mathbf{B}$  :

$$\boldsymbol{\mu}_t = \boldsymbol{\alpha}_0 + \mathbf{A}_1 \mathbf{X}_{t-1} + \mathbf{A}_2 \mathbf{X}_{t-2} + \mathbf{B} \boldsymbol{\mu}_{t-1}, \quad (2.5.3)$$

with additional coefficient matrix :

$$\mathbf{A}_2 = \begin{pmatrix} a_{11,2} & a_{12,2} \\ a_{21,2} & a_{22,2} \end{pmatrix}.$$

The estimation results are also displayed in Table 2.5. We applied the same test procedures. The  $P$ -values of the different tests for the two models are presented in Table 2.7.

The  $P$ -values of the kernel-based test statistics were larger than 5% for all  $m_n$  considered using the Bartlett kernel. This suggests that the residual lags at



low order lags were small. However, the test statistics  $HLB_n$  and  $T_n^\dagger(k_{TR})$  using uniform weighting suggested some dependence at higher lags. For the test statistic  $T_n^\dagger(k_{TR})$ , only  $m = 7$  was significant at the 5% level, offering a  $P$ -value of around 3.6%. However, that  $P$ -value was relatively large and not significant at the 1% level.

Note that the test  $NLY_n$  rejected strongly at all lags, which may be surprising. However, we noted in the simulation experiments of the previous section that  $NLY_n$  exhibited large over-rejection under the null, and low adjusted-size powers. We examined carefully the estimated covariance matrix. Calculation of the eigenvalues of  $\hat{\Omega}_m$  revealed some negative values, and also large condition numbers (e.g., when  $m = 32$ , the largest eigenvalue to the smallest positive eigenvalue was larger than 30,000), suggesting numerical instabilities. From a strictly numerical analysis point of view, the smallest positive eigenvalue was about  $10^{-5}$  when  $m = 32$ , which was still too large to be considered equal to zero.

We report the estimated spectral density estimator of the residuals of the vMEM(2,1) fit in Figure 2.2 using again the Bartlett kernel for the same lag orders  $m_n = [\log(n)], [3n^{0.2}], [3n^{0.3}]$ . It appeared that the spectral density estimator of the second component was much closer to a constant. An examen of the residual autocorrelation function revealed that for the second component one significant lag occurred at lag 10, which seems rather hard to explain and may be considered spurious. While the structured vMEM(2,1) can probably be improved, that fit was significantly better than the vMEM(1,1) adjustment.

FIGURE 2.2. Kernel-based spectral density estimator for the residuals of the adjustment of a vMEM(2,1) model to the bivariate time series corresponding to the stock index FTSE100. The Bartlett kernel is used for the lag orders  $m_n = \lfloor \log(n) \rfloor$  (solid line),  $\lfloor 3n^{0.2} \rfloor$  (dashed line), and  $\lfloor 3n^{0.3} \rfloor$  (dotted line).

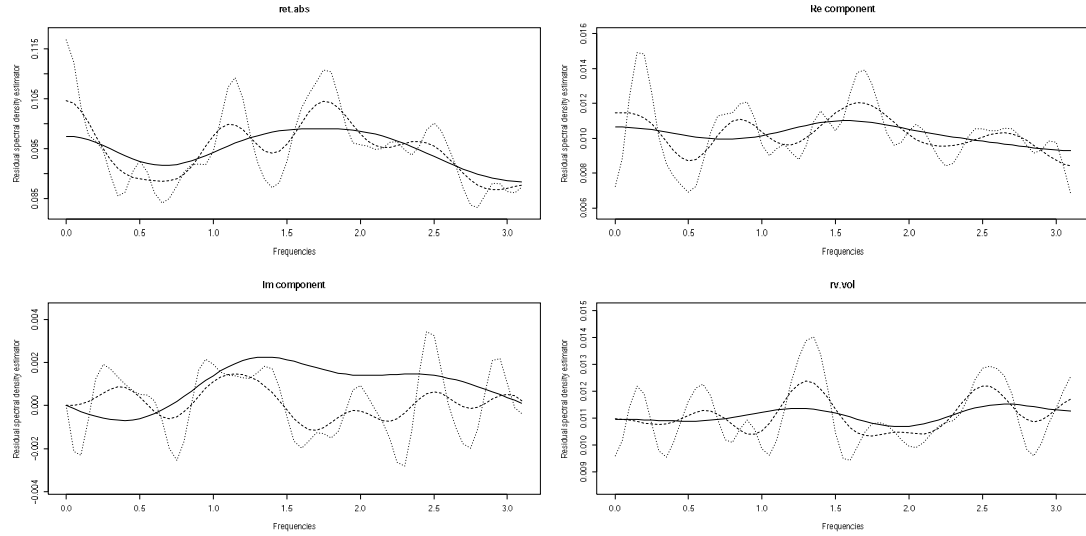


TABLE 2.7.  $P$ -values of the different tests statistics for the adjustment of the vMEM(2,1) model (2.5.3) for modelling the variables (2.5.1) of the FTSE100 stock index.

$m \equiv m_n$	$T_n^\dagger(k_{TR})$	$T_n^\dagger(k_{BAR})$	$HLB_n$	$NLY_n$
7	0.2635	0.4244	0.2141	0.0002
14	0.0359	0.2861	0.0066	0 <sup>+</sup>
32	0.0881	0.0710	0.0314	0 <sup>+</sup>

## 2.6. DISCUSSION AND CONCLUSION

In this article, new tests designed for diagnostic checking vMEM models have been proposed. The asymptotic distribution of the Hosking-Ljung-Box test statistic has been found to converge in distribution to weighted sums of independent chi-squared random variables under the null hypothesis of adequacy. A generalized HLB test statistic motivated by comparing a vector spectral density estimator of the residuals with the null spectral density lead to powerful measures of lack-of-fit. Simulation experiments have been reported to illustrate the empirical properties of the new methods. The Hosking-Ljung-Box has been compared to the original

Hosking test statistic. To use the quantiles from the estimated weighted sum of chi-square distribution provided good results. Somewhat surprisingly, to use the original Hosking test with an approximate rule based on a chi-square distribution with adjusted degrees of freedom gave also reasonable results in our experiments. This may be due to the large number of unit eigenvalues, and the positive but small remaining eigenvalues. We found that the kernel-based test statistics using the Chen-Deo power transformation offered empirical levels close to the nominal levels. Without the power transformation, that is using the original kernel-based methods, the test statistics were too conservative for practical use. The kernel-based test statistics (with a kernel different from the truncated uniform kernel) were often much more powerful than the Hosking-Ljung-Box test statistic. We found that the test statistics of Ng and al. (2016) over-rejected severely the null hypothesis. To consider a generalized inverse did not help, at least in our simulation experiments. Our methods were illustrated in an application, where we found a satisfactory model for data analyzed previously from the econometric literature. Overall, the kernel-based test statistic based on the Chen-Deo power transformation with the Bartlett kernel is recommended for use. The Hosking-Ljung-Box remains a good complement, given that it uses uniform weighting. It is hoped that the proposed methods will be useful for diagnosing vMEM models.

## ACKNOWLEDGEMENT

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

## APPENDIX. PROOF OF THEOREM 1

Under the null hypothesis, the process  $\{\mathbf{u}_t, t \in \mathbb{Z}\}$  is a vector white noise process composed of the random vectors  $\mathbf{u}_t = \mathbf{X}_t \odot \boldsymbol{\mu}_t - \mathbf{1}_d$ , with  $E(\mathbf{u}_t) = \mathbf{0}$  and  $\text{var}(\mathbf{u}_t) = \boldsymbol{\Sigma}$ . Let  $\boldsymbol{\Sigma}^{-1/2}$  be the Cholesky decomposition of the non-singular matrix  $\boldsymbol{\Sigma}^{-1}$  satisfying  $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\Sigma}^{-1/2})^\top$ . We let  $\mathbf{v}_t = \boldsymbol{\Sigma}^{-1/2}\mathbf{u}_t$  and  $\hat{\mathbf{v}}_t =$

$\Sigma^{-1/2} \left\{ \mathbf{X}_t \otimes \boldsymbol{\mu}_t(\hat{\boldsymbol{\theta}}) - \mathbf{1}_d \right\}$ . We also note  $\hat{\boldsymbol{\lambda}}_{nt} = \mathbf{u}_t - \hat{\mathbf{u}}_t$  and  $\hat{\boldsymbol{\gamma}}_{nt} = \Sigma^{-1/2} \hat{\boldsymbol{\lambda}}_{nt}$ . We recall the notation  $k_{hn} = k(h/m_n)$ ,  $h = 1, \dots, n-1$ .

Since  $\{\mathbf{u}_t\}$  is iid with finite fourth moments, the multivariate central limit theorem can be invoked to establish the asymptotic behavior of  $n^{1/2}(\mathbf{C}_u(0) - \Sigma)$ . In particular,  $\mathbf{C}_u(0) - \Sigma = \mathbf{O}_P(n^{-1/2})$ . Proceeding as in Duchesne and Roy (2004), it can be shown that  $\mathbf{C}_{\hat{u}}(0) - \Sigma = \mathbf{O}_P(n^{-1/2})$ . Furthermore, it is easily shown that to replace the mean  $\mathbf{1}_d$  by the sample residual mean gives the following relation :

$$\mathbf{C}_{\hat{\epsilon}}(0) - \mathbf{C}_{\hat{u}}(0) = \mathbf{O}_P(n^{-1}).$$

Thus it is sufficient to show the following relations :

$$\begin{aligned} \sum_{h=1}^{n-1} k_{hn}^2 \operatorname{tr} \{ \mathbf{C}_{\hat{u}}(h) \Sigma^{-1} \mathbf{C}_{\hat{u}}^\top(h) \Sigma^{-1} - \mathbf{C}_{\hat{\epsilon}}(h) \Sigma^{-1} \mathbf{C}_{\hat{\epsilon}}^\top(h) \Sigma^{-1} \} &= o_P(m_n^{1/2}/n), \\ \sum_{h=1}^{n-1} k_{hn}^2 \operatorname{tr} \{ \mathbf{C}_u(h) \Sigma^{-1} \mathbf{C}_u^\top(h) \Sigma^{-1} - \mathbf{C}_{\hat{\epsilon}}(h) \Sigma^{-1} \mathbf{C}_{\hat{\epsilon}}^\top(h) \Sigma^{-1} \} &= o_P(m_n^{1/2}/n). \end{aligned}$$

Since  $\mathbf{C}_{\hat{u}}(h) = \mathbf{C}_{\hat{\epsilon}}(h) + \mathbf{O}_P(n^{-1})$ , the first equality follows easily. This shows (2.3.7).

To establish the second expression and thus formula (2.3.8), consider

$$\sum_{h=1}^{n-1} k_{hn}^2 \left[ \operatorname{tr} \{ \mathbf{C}_{\hat{v}}(h) \mathbf{C}_{\hat{v}}^\top(h) \} - \operatorname{tr} \{ \mathbf{C}_v(h) \mathbf{C}_v^\top(h) \} \right],$$

where  $\mathbf{C}_{\hat{v}}(h) = n^{-1} \sum_{t=h+1}^n \hat{\mathbf{v}}_t \hat{\mathbf{v}}_{t-h}^\top$  and  $\mathbf{C}_v(h) = n^{-1} \sum_{t=h+1}^n \mathbf{v}_t \mathbf{v}_{t-h}^\top$ . For two compatible matrices  $\mathbf{A}$  and  $\mathbf{B}$ , direct algebra gives the decomposition  $\operatorname{tr}(\mathbf{A}\mathbf{A}^\top) - \operatorname{tr}(\mathbf{B}\mathbf{B}^\top) = \operatorname{tr}\{(\mathbf{A} - \mathbf{B})(\mathbf{A} - \mathbf{B})^\top\} + 2\operatorname{tr}\{\mathbf{B}(\mathbf{A} - \mathbf{B})^\top\}$ . Using that exact decomposition, it is sufficient to establish the following results :

$$\sum_{h=1}^{n-1} k_{hn}^2 \operatorname{tr} \left[ \{ \mathbf{C}_{\hat{v}}(h) - \mathbf{C}_v(h) \} \{ \mathbf{C}_{\hat{v}}(h) - \mathbf{C}_v(h) \}^\top \right] = O_P(n^{-1}), \quad (2.6.1)$$

$$\sum_{h=1}^{n-1} k_{hn}^2 \operatorname{tr} \left[ \mathbf{C}_v(h) \{ \mathbf{C}_{\hat{v}}(h) - \mathbf{C}_v(h) \}^\top \right] = o_P(m_n^{1/2}/n). \quad (2.6.2)$$

In order to show relation (2.6.1), we decompose further  $\mathbf{C}_{\hat{v}}(h) - \mathbf{C}_v(h)$  as :

$$\mathbf{C}_{\hat{v}}(h) - \mathbf{C}_v(h) = \mathbf{A}_1(h) + \mathbf{A}_2(h) + \mathbf{A}_3(h), \quad (2.6.3)$$

where  $\mathbf{A}_1(h) = -n^{-1} \sum_{t=h+1}^n \hat{\gamma}_{nt} \mathbf{v}_{t-h}^\top$ ,  $\mathbf{A}_2(h) = -n^{-1} \sum_{t=h+1}^n \mathbf{v}_t \hat{\gamma}_{n,t-h}^\top$ , and finally  $\mathbf{A}_3(h) = n^{-1} \sum_{t=h+1}^n \hat{\gamma}_{nt} \hat{\gamma}_{n,t-h}^\top$ . For three compatible matrices  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{C}$ , we have the following inequality :

$\text{tr}\{(\mathbf{A} + \mathbf{B} + \mathbf{C})(\mathbf{A} + \mathbf{B} + \mathbf{C})^\top\} \leq 4 \{ \text{tr}(\mathbf{A}\mathbf{A}^\top) + \text{tr}(\mathbf{B}\mathbf{B}^\top) + \text{tr}(\mathbf{C}\mathbf{C}^\top) \}$ . Thus a bound for expression (2.6.1) is :

$$\sum_{h=1}^{n-1} k_{hn}^2 \text{tr} \left[ \{ \mathbf{C}_{\hat{\mathbf{v}}}(h) - \mathbf{C}_{\mathbf{v}}(h) \} \{ \mathbf{C}_{\hat{\mathbf{v}}}(h) - \mathbf{C}_{\mathbf{v}}(h) \}^\top \right] \leq 4 \sum_{i=1}^3 E_{in},$$

where  $E_{in} = \sum_{h=1}^{n-1} k_{hn}^2 \text{tr}\{ \mathbf{A}_i(h) \mathbf{A}_i^\top(h) \}$ ,  $i \in \{1, 2, 3\}$ .

Let  $\boldsymbol{\mu}_t = \boldsymbol{\mu}_t(\boldsymbol{\theta}; \mathcal{F}_{t-1}) \equiv (\mu_t(1, \boldsymbol{\theta}), \dots, \mu_t(d, \boldsymbol{\theta}))^\top$  and introduce also  $\hat{\boldsymbol{\mu}}_t = \boldsymbol{\mu}_t(\hat{\boldsymbol{\theta}}; \mathcal{F}_{t-1}) \equiv (\mu_t(1, \hat{\boldsymbol{\theta}}), \dots, \mu_t(d, \hat{\boldsymbol{\theta}}))^\top$ . Note that in order to simplify the notations, we will tacitly drop  $\mathcal{F}_{t-1}$  in the components of  $\boldsymbol{\mu}_t$  and  $\hat{\boldsymbol{\mu}}_t$ .

For the component  $i$ , the Taylor expansion of the function  $\mu_t^{-1}(i; \hat{\boldsymbol{\theta}})$  satisfies :

$$\begin{aligned} \mu_t^{-1}(i; \hat{\boldsymbol{\theta}}) &= \mu_t^{-1}(i; \boldsymbol{\theta}_0) + \left\{ \frac{\partial \mu_t^{-1}(i; \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\}^\top (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \\ &\quad \frac{1}{2} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \left\{ \frac{\partial^2 \mu_t^{-1}(i; \bar{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right\} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0), \end{aligned}$$

where  $\partial \mu_t^{-1}(i; \boldsymbol{\theta}_0) / \partial \boldsymbol{\theta}$  is a  $p \times 1$  vector and we note  $\mathbf{H}_t(i; \bar{\boldsymbol{\theta}}_i) = \partial^2 \mu_t^{-1}(i; \bar{\boldsymbol{\theta}}_i) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$  the Hessian matrix. Simple algebra yields  $\partial \mu_t^{-1}(i, \boldsymbol{\theta}) / \partial \boldsymbol{\theta} = -\mu_t^{-2}(i, \boldsymbol{\theta}) \partial \mu_t(i, \boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ , see also Ng & al. [82, p. 95]. Collecting the results, we obtain :

$$\begin{aligned} \boldsymbol{\epsilon}_t(\hat{\boldsymbol{\theta}}) &= \boldsymbol{\epsilon}_t(\boldsymbol{\theta}_0) - \text{diag}(\boldsymbol{\epsilon}_t) \text{diag}^{-1}(\boldsymbol{\mu}_t) \left\{ \frac{\partial \boldsymbol{\mu}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top} \right\} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \\ &\quad \frac{1}{2} \left\{ \mathbf{I}_d \otimes (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \right\} \bar{\mathbf{H}}_t (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0), \end{aligned}$$

where  $\bar{\mathbf{H}}_t = (\mathbf{H}_t^\top(1; \bar{\boldsymbol{\theta}}_1), \dots, \mathbf{H}_t^\top(d; \bar{\boldsymbol{\theta}}_d))^\top$  is the  $d^2 \times d$  matrix composed of the  $d$  Hessian matrices coming from the  $d$  Taylor expansions.

Now we study  $E_{1n}$ . We write  $\hat{\gamma}_{nt}$  as :

$$\hat{\gamma}_{nt} = \boldsymbol{\Sigma}^{-1/2} \left[ \text{diag}(\boldsymbol{\epsilon}_t) \mathbf{W}_t(\boldsymbol{\theta}_0) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) - \frac{1}{2} \left\{ \mathbf{I}_d \otimes (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \right\} \bar{\mathbf{H}}_t (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \right],$$

where  $\mathbf{W}_t(\boldsymbol{\theta}_0) = \text{diag}^{-1}(\boldsymbol{\mu}_t) \partial \boldsymbol{\mu}_t / \partial \boldsymbol{\theta}^\top$ . It follows that  $\mathbf{A}_1(h) = \mathbf{A}_{11}(h) + \mathbf{A}_{12}(h)$ , where

$$\begin{aligned} \mathbf{A}_{11}(h) &= n^{-1} \sum_{t=h+1}^n \boldsymbol{\Sigma}^{-1/2} \text{diag}(\boldsymbol{\epsilon}_t) \mathbf{W}_t(\boldsymbol{\theta}_0) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \mathbf{v}_{t-h}^\top, \\ \mathbf{A}_{12}(h) &= -\frac{n^{-1}}{2} \sum_{t=h+1}^n \boldsymbol{\Sigma}^{-1/2} \left\{ \mathbf{I}_d \otimes (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \right\} \bar{\mathbf{H}}_t (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \mathbf{v}_{t-h}^\top. \end{aligned}$$

Using the inequality  $\text{tr} \{(\mathbf{A} + \mathbf{B})(\mathbf{A} + \mathbf{B})^\top\} \leq 2 \{ \text{tr}(\mathbf{A}\mathbf{A}^\top) + \text{tr}(\mathbf{B}\mathbf{B}^\top) \}$ , we find the following bound for  $E_{1n}$ :

$$E_{1n} \leq 2 \left[ \sum_{h=1}^{n-1} k_{hn}^2 \text{tr} \{ \mathbf{A}_{11}(h) \mathbf{A}_{11}^\top(h) \} + \sum_{h=1}^{n-1} k_{hn}^2 \text{tr} \{ \mathbf{A}_{12}(h) \mathbf{A}_{12}^\top(h) \} \right].$$

Nothing that  $\sum_{h=1}^{n-1} k_{hn}^2 \text{tr} \{ \mathbf{A}_{11}(h) \mathbf{A}_{11}^\top(h) \} = \sum_{h=1}^{n-1} k_{hn}^2 \|\mathbf{A}_{11}(h)\|^2$ , and using the properties of the Euclidian norm, it follows that :

$$\sum_{h=1}^{n-1} k_{hn}^2 \|\mathbf{A}_{11}(h)\|^2 \leq \|\boldsymbol{\Sigma}^{-1/2}\|^2 \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2 \left\{ \sum_{h=1}^{n-1} k_{hn}^2 \left( n^{-1} \sum_{t=h+1}^n a_{11th} \right)^2 \right\},$$

where  $a_{11th} = \|\text{diag}(\boldsymbol{\epsilon}_t) \mathbf{W}_t(\boldsymbol{\theta}_0)\| \|\mathbf{v}_{t-h}\|$ . We write  $n^{-1} \sum_{t=h+1}^n a_{11th} = n^{-1} \sum_{t=h+1}^n \dot{a}_{11th} + n^{-1} \sum_{t=h+1}^n E(a_{11th})$ , where  $\dot{a}_{11th} = a_{11th} - E(a_{11th})$ . Using  $(a + b)^2 \leq 2(a^2 + b^2)$ ,

$$\left( n^{-1} \sum_{t=h+1}^n a_{11th} \right)^2 \leq 2 \left[ \left( n^{-1} \sum_{t=h+1}^n \dot{a}_{11th} \right)^2 + \left\{ n^{-1} \sum_{t=h+1}^n E(a_{11th}) \right\}^2 \right]. \quad (2.6.4)$$

However, from Assumption A,  $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \mathbf{O}_P(n^{-1/2})$  and consequently  $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2 = \mathbf{O}_P(n^{-1})$ . Using the assumption of stationarity, it follows that  $\sum_{h=1}^{n-1} k_{hn}^2 \{E(a_{11th})\}^2 = \mathbf{O}(1)$  and

$$\sum_{h=1}^{n-1} k_{hn}^2 \left( n^{-1} \sum_{t=h+1}^n \dot{a}_{11th} \right)^2 = \mathbf{O}_P(m_n/n).$$

Thus  $\sum_{h=1}^n k_{hn}^2 \|\mathbf{A}_{11}(h)\|^2 = \mathbf{O}_P(n^{-1} + m_n/n^2)$ . Similarly,  $\sum_{h=1}^n k_{hn}^2 \|\mathbf{A}_{12}(h)\|^2 = \mathbf{O}_P(n^{-1})$ .

Using similar and simpler arguments, it can be shown that formulas  $E_{2n}$  and  $E_{3n}$  are such that  $E_{2n} = \sum_{h=1}^{n-1} k_{hn}^2 \text{tr} \{ \mathbf{A}_2(h) \mathbf{A}_2^\top(h) \} = \mathbf{O}_P(m_n/n^2)$  and  $E_{3n} = \sum_{h=1}^{n-1} k_{hn}^2 \text{tr} \{ \mathbf{A}_3(h) \mathbf{A}_3^\top(h) \} = \mathbf{O}_P(m_n/n^2)$ . This shows result (2.6.1).

We now study expression (2.6.2). Using the decomposition (2.6.3), we can write :

$$\sum_{h=1}^{n-1} k_{hn}^2 \text{tr} \left[ \mathbf{C}_v(h) \{ \mathbf{C}_{\hat{v}}(h) - \mathbf{C}_v(h) \}^\top \right] = \sum_{i=1}^3 F_{in},$$

where

$$\begin{aligned} F_{1n} &= \sum_{h=1}^{n-1} k_{hn}^2 \text{tr} \{ \mathbf{C}_v(h) \mathbf{A}_1^\top(h) \}, \\ F_{2n} &= \sum_{h=1}^{n-1} k_{hn}^2 \text{tr} \{ \mathbf{C}_v(h) \mathbf{A}_2^\top(h) \}, \\ F_{3n} &= \sum_{h=1}^{n-1} k_{hn}^2 \text{tr} \{ \mathbf{C}_v(h) \mathbf{A}_3^\top(h) \}. \end{aligned}$$

First we study  $F_{1n}$ . As noted previously,  $\mathbf{A}_1(h) = \mathbf{A}_{11}(h) + \mathbf{A}_{12}(h)$ , and we consider the following decomposition :  $F_{1n} = F_{11n} + F_{12n}$ , where :

$$\begin{aligned} F_{11n} &= \sum_{h=1}^{n-1} k_{hn}^2 \text{tr} \{ \mathbf{C}_v(h) \mathbf{A}_{11}^\top(h) \}, \\ F_{12n} &= \sum_{h=1}^{n-1} k_{hn}^2 \text{tr} \{ \mathbf{C}_v(h) \mathbf{A}_{12}^\top(h) \}. \end{aligned}$$

We do the details for  $F_{11n}$ . Using the Cauchy-Schwarz inequality, we find that :

$$|F_{11n}| \leq \sum_{h=1}^{n-1} k_{hn}^2 |\text{tr} \{ \mathbf{C}_v(h) \mathbf{A}_{11}^\top(h) \}| \leq \sum_{h=1}^{n-1} k_{hn}^2 \|\mathbf{C}_v(h)\| \|\mathbf{A}_{11}(h)\|.$$

However, uniformly in  $j$ ,  $E(\|\mathbf{C}_v(h)\|^2) = O(n^{-1})$ , and Assumption A yields  $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = O_P(n^{-1/2})$ . Since

$$\begin{aligned} \sum_{h=1}^{n-1} k_{hn}^2 |E(a_{11th})| &= O(1), \\ \sum_{h=1}^{n-1} k_{hn}^2 |n^{-1} \sum_{t=h+1}^n \dot{a}_{11th}| &= O_P(m_n/n^{1/2}), \end{aligned}$$

where  $a_{11th}$  and  $\dot{a}_{11th}$  are as in (2.6.4), it follows that  $F_{11n} = o_P(m_n^{1/2}/n)$ . Similarly,  $F_{12n} = o_P(m_n^{1/2}/n)$ . Thus  $F_{1n} = o_P(m_n^{1/2}/n)$ . The two other terms  $F_{2n}$  and  $F_{3n}$  are treated similarly. This shows the Theorem.  $\square$

# Chapitre 3

---

## TESTS DE NON CORRÉLATION SOUS FAIBLE DÉPENDANCE ENTRE DEUX PROCESSUS STATIONNAIRES EN COVARIANCE

### 3.1. INTRODUCTION

Plusieurs phénomènes physiques et économiques impliquent l'étude de plusieurs séries chronologiques. Par exemple, Shumway et Stoffer (2005) considèrent des applications portant sur la modélisation des populations de poissons mesurées mensuellement entre 1950 et 1987. Cette application tirée des sciences de l'environnement cherchait à comprendre les relations entre la population des poissons et un certain indice climatique. Comme autre exemple, Brockwell et Davis (1991, Chapitre 11) reprennent une série chronologique de Box et Jenkins (1970) portant sur une série de ventes expliquée par un indice économique.

Lorsque nous avons à modéliser deux séries chronologiques ou plus, une préoccupation légitime porte sur les relations pouvant exister entre elles. Dans les applications économiques, la mise en lumière des relations de causalité entre les séries chronologiques peut être particulièrement importante dans l'élaboration d'un système performant de prévisions.

Ainsi, avant de développer des modèles multivariés compliqués qui reposent typiquement sur un grand nombre de paramètres, il est utile de vérifier si les



séries chronologiques sont indépendantes ou minimalement non corrélées. Si des méthodes puissantes suggèrent la présence de relations plus ou moins complexes, alors dans un second temps il devient nécessaire de procéder à des analyses multivariées.

La première approche visant à tester l'indépendance entre deux séries chronologiques remonte aux travaux de Haugh (1976). Dans son approche, les séries chronologiques sont modélisées par des modèles autorégressifs moyennes mobiles (ARMA). Chacun des modèles est considéré correctement spécifié, et les résidus des vrais modèles sont obtenus. Les mesures de corrélation croisées résiduelles suivantes sont calculées :

$$r_{\hat{u}\hat{v}}(j) = \begin{cases} \frac{\sum_{t=j+1}^n \hat{u}_t \hat{v}_{t-j}}{\left\{ \sum_{t=1}^n \hat{u}_t^2 \sum_{t=1}^n \hat{v}_t^2 \right\}^{1/2}}, & 0 \leq j \leq n-1, \\ r_{\hat{u}\hat{v}}(-j), & -n+1 \leq j < 0, \end{cases}$$

où  $\hat{u}_t$ ,  $t = 1, \dots, n$  et  $\hat{v}_t$ ,  $t = 1, \dots, n$  dénotent les deux séries chronologiques résiduelles obtenues de la modélisation ARMA. Le test de Haugh est alors :

$$S_m(u) = n \sum_{j=-m}^m r_{\hat{u}\hat{v}}^2(j),$$

où la constante  $m$ , fixée par rapport à  $n$ , satisfait  $0 \leq m < n$  et est déterminée par l'analyste. La distribution asymptotique de  $S_m$  est  $\chi_m^2$  sous l'hypothèse nulle d'indépendance. Comme mentionné, un présupposé important est que chaque modèle ARMA est correctement spécifié. Si les ordres  $p$  et/ou  $q$  sont incorrectement choisis, alors on fait face à un problème de spécification qui peut avoir un effet sur les inférences. Il est à noter que le test de Haugh (1976) est en fait une version du test de Box et Pierce (1970) lorsque nous avons deux séries chronologiques. Le test de Box et Pierce (1970) a été étudié par Francq, Roy et Zakoïan (2005) lorsque le processus d'innovation est un bruit blanc faible. Motivé par une approche spectrale similaire à celle considérée dans le chapitre 2, Hong (1996) a

généralisé l'approche de Haugh (1976) en considérant la statistique suivante :

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

où

$$U_n(k; m) = \sum_{j=-n+1}^{n-1} \left(1 - \frac{|j|}{n}\right) k^2(j/m),$$

$$V_n(k; m) = \sum_{j=-n+1}^{n-1} \left(1 - \frac{|j|}{n}\right) \left(1 - \frac{|j|+1}{n}\right) k^4(j/m).$$

La fonction  $k(\cdot)$  représente le noyau qui est supposée être une fonction symétrique définie de  $\mathbb{R}$  vers  $[-1, 1]$  avec  $k(0) = 1$  et  $\int_{-\infty}^{\infty} k^2(t)dt < \infty$ .

L'approche de Hong (1996) nécessite également la formulation d'un modèle. Elle est cependant plus générale dans le sens qu'une procédure semi-paramétrique est adoptée. Dans son approche, chaque série est ajustée par un modèle autorégressif avec un ordre convergeant à l'infini à un certain taux. Duchesne et Roy (2004) ont généralisé l'approche de Hong (1996) dans une autre direction, considérant des techniques d'estimation robustes aux valeurs aberrantes et utilisant des mesures de corrélation robustifiées. Les méthodes de Haugh (1976) et de Hong (1996) ont été généralisées au cas multivarié par El Himdi et Roy (1997) et Bouhaddioui et Roy (2006a, 2006b), respectivement. Saidi (2007) quant à lui considère des tests spectraux dans les modèles multivariés et cointégrés. On retrouve également Pham, Roy et Cédras (2000) qui ont développé des tests portemanteaux entre deux processus cointégrés.

D'autres généralisations incluent Koch et Yang (1986) qui ont modifié le test de Haugh (1976) afin de prendre en considération certains motifs dans la dépendance des séries résiduelles. Cette approche a été généralisée et appliquée au cas multivarié dans Hallin et Saidi (2005).

Les approches précédentes utilisent toutes des mesures de corrélation croisées afin d'étudier l'absence de corrélation. En fait, ce ne sont des tests d'indépendance

que dans le cas où les phénomènes en cause sont gaussiens. Des tests d'indépendance entre deux ou plus de deux séries chronologiques sont développés dans Duchesne, Ghoudi et Rémillard (2012). Voir aussi Kim et Lee (2005).

Jusqu'à maintenant, l'hypothèse principale entre les séries est celle d'indépendance, même si plusieurs méthodes reposent sur des mesures de corrélation. Cependant, même si l'hypothèse nulle que l'on cherche à tester tente de cerner si les deux séries chronologiques en cause sont non corrélées, il serait souhaitable de permettre de la dépendance entre elles, même sous l'hypothèse nulle. Par exemple, dans les séries chronologiques financières, même si deux séries chronologiques sont non corrélées, il est possible que les séries résultent d'effets autorégressifs conditionnellement hétéroskédastiques (ARCH). De plus, il n'est pas impossible que soit présent de la causalité en variance, du *volatility spillover* ou du *risk spillover* (Hong, Liu et Wang (2009)). Dans le test de Haugh (1976), il est important de noter que la distribution du test est obtenue sous l'hypothèse nulle d'indépendance, ne permettant donc aucune forme de causalité en variance ou de dépendance entre les séries chronologiques.

Dans ce chapitre, en utilisant le concept de contraction géométrique des moments (CGM), nous allons présenter un cadre plus général, permettant de tester la non corrélation tout en n'excluant pas que les séries en cause soient tout de même dépendantes. Le concept de contraction géométrique des moments est considéré notamment dans Hsing et Wu (2004), Shao et Wu (2007), Wu et Shao (2004) et Shao (2011). Ici, la non corrélation est étudiée entre les séries brutes, contrairement aux approches précédentes qui présupposent des modèles plus ou moins généraux pour les données.

Le document est organisé de la façon suivante. Dans les trois premières sections, nous allons présenter le cadre général. Les sections 4 et 5 sont consacrées respectivement à la distribution limite des covariances croisées et à l'estimation de leur matrice des variances et covariances. La section 6 présente les simulations et la section 7 conclut.

Dans la suite du document, le symbole  $\xrightarrow{\mathcal{L}}$  signifie convergence en distribution. La notation  $X_n = o_p(1)$  signifie que la variable aléatoire  $X_n$  convergent en probabilité vers zéro tandis que  $X_n = O_p(1)$  veut dire que  $X_n$  est bornée en probabilité. Pour toute matrice  $\mathbf{A}$  de dimension  $d_1 \times d_2$ , la norme matricielle est notée  $\|\mathbf{A}\| = \sup_{|\mathbf{u}|_{d_2} \leq 1} |\mathbf{A}\mathbf{u}|_{d_1}$ , où  $|\cdot|_{d_1}$  et  $|\cdot|_{d_2}$  sont les normes euclidiennes des espaces  $\mathbb{R}^{d_1}$  et  $\mathbb{R}^{d_2}$ . Étant une norme matricielle, une propriété utile est  $\|\mathbf{A}\|^2 \leq \sum_{i,j} A_{i,j}^2$ , où  $A_{i,j}$ ,  $1 \leq i \leq d_1$  et  $1 \leq j \leq d_2$ , sont les composantes de la matrice  $\mathbf{A}$ . Pour tout vecteur  $\mathbf{Z}$  de dimension  $d$  et  $p \in \mathbb{N}^*$ , on dira que  $\mathbf{Z} \in \mathcal{L}^p$  si  $\|\mathbf{Z}\|_p = \{E(|\mathbf{Z}|_d^p)\}^{1/p} < \infty$ .

### 3.2. PRÉLIMINAIRES

Soient deux processus unidimensionnels  $\{X_t\}$  et  $\{Y_t\}$  stationnaires en covariance, de moyennes nulles et de variances  $\sigma_x^2$  et  $\sigma_y^2$ , respectivement.

Étant donné des réalisations aux instants  $t = 1, \dots, n$ , nous voulons tester la non corrélation entre  $\{X_t\}$  et  $\{Y_t\}$  en adoptant une approche non paramétrique, c'est-à-dire en n'ajustant aucun modèle à chaque série chronologique.

Définissons pour tout  $j \in \mathbb{Z}$ , les covariances croisées théoriques :

$$\gamma_{XY}(j) = \begin{cases} E(X_t Y_{t-j}), & j \geq 0, \\ \gamma_{YX}(-j), & j < 0. \end{cases} \quad (3.2.1)$$

et pour tout  $j$ ,  $-n+1 \leq j \leq n-1$  les covariances croisées échantillonnales :

$$\hat{\gamma}_{XY}(j) = \begin{cases} n^{-1} \sum_{t=j+1}^n X_t Y_{t-j}, & j \geq 0, \\ \hat{\gamma}_{YX}(-j), & j < 0. \end{cases} \quad (3.2.2)$$

Afin de tester la non corrélation entre les processus  $\{X_t\}$  et  $\{Y_t\}$ , nous allons dans un premier temps trouver la distribution limite du vecteur des covariances croisées  $\{\hat{\gamma}_{XY}(j)\}_{-m+1 \leq j \leq m-1}$ , pour  $m \in \mathbb{N}^*$ . Pour ce faire, nous avons besoin d'avoir une compréhension de la dynamique conjointe des deux séries. C'est dans ce sens qu'est définie l'hypothèse suivante.

**Hypothèse 3.1.** *Il existe une fonction mesurable  $\mathbf{G}$  et un processus multivarié, indépendant et identiquement distribué (iid)  $\{\boldsymbol{\xi}_t, t \in \mathbb{Z}\}$  de  $\mathbb{R}^2$  tel que  $\mathbf{Z}_t = (X_t, Y_t)^\top = \mathbf{G}(\dots, \boldsymbol{\xi}_{t-1}, \boldsymbol{\xi}_t)$  pour tout  $t$ .*

L'hypothèse 3.1 telle que mentionnée met en exergue la représentation des processus stochastique stationnaires. Beaucoup d'auteurs se sont penchés sur cette problématique, notamment Rosenblatt (1952) et Arjas et Lehtonen (1978). Il est connu que si le processus  $\{\mathbf{Z}_t\}$  est stationnaire et ergodique, la représentation comme dans l'hypothèse 3.1 est possible (Voir Borkar (1993) et Wu (2005)).

Brockwell et Davis (1991, pp. 404-405) considèrent la classe des processus linéaires

$$\mathbf{Z}_t = \sum_{j=-\infty}^{\infty} \mathbf{C}_j \boldsymbol{\xi}_{t-j},$$

où les composantes des matrices  $\mathbf{C}_j$ ,  $i \in \mathbb{Z}$  sont absolument sommables et  $\{\boldsymbol{\xi}_t\}$  est un bruit blanc faible. Cependant, ils réservent le nom de moyennes mobiles vectorielles (VMA) aux processus non-anticipatifs :

$$\mathbf{Z}_t = \sum_{j=0}^{\infty} \mathbf{C}_j \boldsymbol{\xi}_{t-j}. \quad (3.2.3)$$

Lorsque  $\{\boldsymbol{\xi}_t\}$  est un bruit blanc fort, on constate que les processus VMA( $\infty$ ) définis par 3.2.3 satisfont l'hypothèse 3.1.

Dans la présente représentation, le processus  $\{\boldsymbol{\xi}_t, t \in \mathbb{Z}\}$  est iid alors que dans celle de Wold, ces innovations sont seulement stationnaires au sens faible. Le fait qu'elles soient iid permet de construire plus facilement les théories asymptotiques. Dans la suite, nous allons rappeler les mesures de dépendance introduites par Wu (2005) et aussi la propriété de Contraction Géométrique des Moments (CGM).

Soit  $\{\boldsymbol{\xi}_t, t \in \mathbb{Z}\}$  un processus iid, posons  $\boldsymbol{\varepsilon}_t = (\dots, \boldsymbol{\xi}_{t-1}, \boldsymbol{\xi}_t)$  et  $\mathbf{Z}_t = \mathbf{G}(\boldsymbol{\varepsilon}_t)$  où  $\mathbf{G}$  est une fonction mesurable. Posons également :

$$\mathbf{Z}'_t = \mathbf{G}(\dots, \boldsymbol{\xi}_{-2}, \boldsymbol{\xi}_{-1}, \boldsymbol{\xi}'_0, \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_{t-1}, \boldsymbol{\xi}_t) = \mathbf{G}(\boldsymbol{\varepsilon}'_t), \quad (3.2.4)$$

où  $\boldsymbol{\varepsilon}'_t = (\dots, \boldsymbol{\xi}_{-2}, \boldsymbol{\xi}_{-1}, \boldsymbol{\xi}'_0, \dots, \boldsymbol{\xi}_{t-1}, \boldsymbol{\xi}_t)$ . Posons aussi pour tout  $k \in \mathbb{Z}$  :

$$\mathcal{F}_k = \sigma(\dots, \boldsymbol{\xi}_{k-1}, \boldsymbol{\xi}_k),$$

la  $\sigma$ -algèbre engendrée par  $(\dots, \boldsymbol{\xi}_{k-1}, \boldsymbol{\xi}_k)$  et  $\mathcal{P}_k(\cdot)$  l'opérateur de projection :

$$\mathcal{P}_k(\mathbf{Z}_t) = E(\mathbf{Z}_t | \mathcal{F}_k) - E(\mathbf{Z}_t | \mathcal{F}_{k-1}), \quad \forall t \in \mathbb{Z}.$$

**Définition 3.1.** *Pour tout  $p$  tel que  $2 < p \leq 4$  et pour tout  $t$ , les scalaires  $\delta_{p,t}$  et  $\pi_{p,t}$  définis respectivement par :*

$$\delta_{p,t} = \|\mathbf{Z}_t - \mathbf{Z}'_t\|_p, \quad (3.2.5)$$

et

$$\pi_{p,t} = \|\mathcal{P}_0(\mathbf{Z}_t)\|_p, \quad (3.2.6)$$

*représentent respectivement les mesures de dépendance physique et de dépendance prédictive.*

La notion de dépendance physique a été introduite par Wu (2005, p. 1450) dans le cas des processus univariés et généralisée par la suite aux processus multivariés par Liu et Lin (2009, p. 250) et Wu et Zhou (2011, p. 1399). L'expression (3.2.5) mesure le changement physique de  $\mathbf{Z}_t$  après la modification qui a été apportée à  $\boldsymbol{\varepsilon}_t$  tandis que l'expression (3.2.6) mesure l'effet de l'observation initiale sur les observations futures.

Le processus  $\{\mathbf{Z}_t\}$  est dit faiblement  $p$ -stable si :

$$\sum_{t=0}^{\infty} \delta_{p,t} < \infty,$$

alors qu'il sera dit fortement  $p$ -stable si :

$$\sum_{t=0}^{\infty} \pi_{p,t} < \infty.$$

En d'autres termes, cette dernière condition signifie que l'effet cumulé de l'observation initiale sur les observations futures est fini (Wu (2007, p. 13), Wu et Zhou (2011, p. 1399)).

La contraction géométrique des moments est aussi importante pour traduire la faible dépendance entre deux processus. La définition 3.2 permet de mieux appréhender en quoi consiste cette propriété. Ici, un autre couplage du processus

$\{\varepsilon_t\}$  semblable à celle de l'équation (3.2.4) consiste à poser :

$$\mathbf{Z}_t^* = \mathbf{G}(\dots \xi'_{-2}, \xi'_{-1}, \xi'_0, \xi_1, \dots, \xi_{t-1}, \xi_t),$$

ce qui nous mène à la définition suivante.

**Définition 3.2.** *Le processus  $\{\mathbf{Z}_t\}$  est dit CGM( $p$ ) s'il existe  $\alpha = \alpha(p) \in (0, 1)$  et  $C > 0$  tel que :*

$$\|\mathbf{Z}_t - \mathbf{Z}_t^*\|_p < C\alpha^t, \forall t. \quad (3.2.7)$$

Si  $\{\mathbf{Z}_t\}$  est CGM( $p$ ), alors  $\{\mathbf{Z}_t\}$  est  $p$ -stable (Wu (2009, p. 456), Wu et Zhou (2011, p. 1400)).

### 3.3. DISTRIBUTION LIMITE DES COVARIANCES CROISÉES

La distribution limite d'un vecteur de taille fixée des autocovariances échantillonnables est une étape importante de notre démarche. Basé sur ce résultat, un résultat nous permet d'obtenir la distribution asymptotique de notre statistique de test. Sous les conditions du théorème 3.1, on montre que la distribution asymptotique est normale. La preuve repose sur des arguments que l'on retrouve dans Wu (2011, Théorème 3). Notons que les résultats précédents étaient établis pour des autocovariances échantillonnables, alors que dans le cadre de notre travail, nous avons besoin d'un résultat pour les covariances croisées échantillonnables.

**Théorème 3.1.** *Soient  $m \in \mathbb{N}^*$ , sous les conditions de l'hypothèse 3.1, si  $X_t, Y_t \in L^4$  et que le processus  $\{\mathbf{Z}_t = (X_t, Y_t)^\top\}$  est CGM( $p$ ) pour  $p = 4$  alors :*

$$n^{1/2} \begin{pmatrix} \hat{\gamma}_{XY}(-m) - \gamma_{XY}(-m) \\ \vdots \\ \hat{\gamma}_{XY}(0) - \gamma_{XY}(0) \\ \vdots \\ \hat{\gamma}_{XY}(m) - \gamma_{XY}(m) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}_{m+1}(\mathbf{0}, \boldsymbol{\Sigma}_m), \quad (3.3.1)$$

avec,

$$\boldsymbol{\Sigma}_m = E(\mathbf{D}_0 \mathbf{D}_0^\top),$$

$$\begin{aligned}
D_0 &= \sum_{t=0}^{\infty} \mathcal{P}_0(\mathbf{W}_t) \in L^2, \\
\mathbf{W}_t &= (Y_t \mathbf{X}_{t,m}^\top, \mathbf{Y}_{t,m}^\top X_t)^\top, \\
\mathbf{Y}_{t,m} &= (Y_t, Y_{t-1}, \dots, Y_{t-m})^\top, \\
\mathbf{X}_{t,m} &= (X_{t-m}, \dots, X_{t-1})^\top.
\end{aligned}$$

*Preuve* : On commence par faire la preuve pour les termes  $\hat{\gamma}_{XY}(j)$  où  $j$  fixée et  $0 \leq j \leq m-1$ . On a :

$$\hat{\gamma}_{XY}(j) = n^{-1} \sum_{t=j+1}^n X_t Y_{t-j} = n^{-1} \sum_{t=j+1}^m X_t Y_{t-j} + n^{-1} \sum_{t=m+1}^n X_t Y_{t-j},$$

et donc

$$\begin{aligned}
n^{1/2} \hat{\gamma}_{XY}(j) &= n^{-1/2} \sum_{t=j+1}^m X_t Y_{t-j} + n^{-1/2} \sum_{t=m+1}^n X_t Y_{t-j}, \\
&= n^{-1/2} \sum_{t=m+1}^n X_t Y_{t-j} + o_P(1).
\end{aligned}$$

Nous avons pour  $q = p/2$ ,  $\|\mathcal{P}_0(X_t Y_{t-j})\|_q = \|E(X_t Y_{t-j} | \boldsymbol{\varepsilon}_0) - E(X_t Y_{t-j} | \boldsymbol{\varepsilon}_{-1})\|_q$ .

Puisque  $\boldsymbol{\varepsilon}'_0, \boldsymbol{\varepsilon}_j, j \in \mathbb{Z}$  sont iid,

$$E(X_t Y_{t-j} | \boldsymbol{\varepsilon}_{-1}) = E(X'_t Y'_{t-j} | \boldsymbol{\varepsilon}_{-1}) = E(X'_t Y'_{t-j} | \boldsymbol{\varepsilon}_0),$$

où  $(X'_t, Y'_t)^\top = \mathbf{G}(\boldsymbol{\varepsilon}'_t)$  où  $\boldsymbol{\varepsilon}'_t$  est donné dans la Définition 3.1.

Ainsi, d'après l'inégalité de Jensen et l'inégalité du triangle,

$$\begin{aligned}
\|\mathcal{P}_0(X_t Y_{t-j})\|_q &= \|(E(X_t Y_{t-j} - X'_t Y'_{t-j}) | \boldsymbol{\varepsilon}_0)\|_q \leq \|X_t Y_{t-j} - X'_t Y'_{t-j}\|_q, \\
&= \|(X_t - X'_t) Y_{t-j} + X'_t (Y_{t-j} - Y'_{t-j})\|_q, \\
&\leq \|(X_t - X'_t) Y_{t-j}\|_q + \|X'_t (Y_{t-j} - Y'_{t-j})\|_q, \\
&\leq \|X_t - X'_t\|_p \|Y_{t-j}\|_p + \|Y_{t-j} - Y'_{t-j}\|_p \|X'_t\|_p, \\
&\leq \|X_t - X'_t\|_p \|\mathbf{Z}_{t-j}\|_p + \|Y_{t-j} - Y'_{t-j}\|_p \|\mathbf{Z}'_t\|_p, \\
&\leq \left( \|X_t - X'_t\|_p + \|Y_{t-j} - Y'_{t-j}\|_p \right) \kappa_p,
\end{aligned}$$



où  $\kappa_p = \sup_{j \in \mathbb{Z}} (\|\mathbf{Z}_t\|_p)$ . Or,

$$\|X_t - X'_t\|_p \leq \|\mathbf{Z}_t - \mathbf{Z}'_t\|_p = \delta_{p,t},$$

et

$$\|Y_{t-j} - Y'_{t-j}\|_p \leq \|\mathbf{Z}_{t-j} - \mathbf{Z}'_{t-j}\|_p = \delta_{p,t-j}.$$

Ceci entraîne alors pour  $p = 4$  que :

$$\sum_{t=0}^{\infty} \|\mathcal{P}_0(X_t Y_{t-j})\|_q \leq \left( \sum_{t=0}^{\infty} \delta_{p,t} + \sum_{t=0}^{\infty} \delta_{p,t-j} \right) \kappa_p.$$

Pour  $t - j < 0$ ,  $\mathbf{Z}_{t-j} = \mathbf{Z}'_{t-j}$  et  $\delta_{p,t-j} = \|\mathbf{Z}_{t-j} - \mathbf{Z}'_{t-j}\|_p = 0$ . Il en découle que :

$$\sum_{t=0}^{\infty} \|\mathcal{P}_0(X_t Y_{t-j})\|_q \leq \left( \sum_{t=0}^{\infty} \delta_{p,t} + \sum_{t=0}^{\infty} \delta_{p,t-j} \right) \kappa_p = 2\kappa_p \sum_{t=0}^{\infty} \delta_{p,t} < \infty.$$

D'après Wu (2011, Théorème 3) :

$$\begin{aligned} n^{1/2} \{\hat{\gamma}_{XY}(j) - \gamma_{XY}(j)\} &= n^{-1/2} \sum_{t=m+1}^n \{X_t Y_{t-j} - E(X_t Y_{t-j})\} + o_P(1) \\ &\xrightarrow{\mathcal{L}} \mathcal{N}(0, E(D_{0j} D_{0j}^\top)), \end{aligned}$$

où  $D_{0j} = \sum_{t=0}^{\infty} \mathcal{P}_0(X_t Y_{t-j})$ . Le même raisonnement se fait pour les termes  $\hat{\gamma}_{YX}(j)$  où  $j$  fixée et  $1 \leq j \leq m - 1$ .

Pour généraliser le résultat au cas vectoriel, nous allons utiliser l'astuce de Cramer-Wold. Pour toute combinaison linéaire  $\sum_{l=1}^k \alpha_l \hat{\gamma}_{XY}(j_l)$  où les indices  $j_1, \dots, j_k$  sont compris entre  $-m + 1$  et  $m - 1$ , on a :

$$n^{1/2} \sum_{l=1}^k \alpha_l \hat{\gamma}_{XY}(j_l) = n^{-1/2} \sum_{t=m+1}^n \left( \sum_{l=1}^k \alpha_l X_{t-j'_l} Y_{t-j''_l} \right) + o_P(1),$$

avec  $j'_l = \frac{|j_l| - j_l}{2}$  et  $j''_l = \frac{|j_l| + j_l}{2}$ .

Il suffira de montrer que le processus  $\{\sum_{l=1}^k \alpha_l X_{t-j'_l} Y_{t-j''_l}, t \in \mathbb{Z}\}$  est fortement  $p$ -stable. Pour ce faire :

$$\|\mathcal{P}_0 \left( \sum_{l=1}^k \alpha_l X_{t-j'_l} Y_{t-j''_l} \right)\|_q = \left\| \sum_{l=1}^k \alpha_l \mathcal{P}_0 \left( X_{t-j'_l} Y_{t-j''_l} \right) \right\|_q,$$

$$\begin{aligned}
&\leq \sum_{l=1}^k |\alpha_l| \|\mathcal{P}_0 \left( X_{t-j'_l} Y_{t-j''_l} \right)\|_q, \\
&\leq \sum_{l=1}^k |\alpha_l| \left( \|X_{t-j'_l}\|_p \delta_{p,t-j''_l} + \|Y'_{t-j''_l}\|_p \delta_{p,t-j'_l} \right).
\end{aligned}$$

On obtient alors pour  $p = 4$  :

$$\begin{aligned}
\sum_{t=0}^{\infty} \|\mathcal{P}_0 \left( \sum_{l=1}^k \alpha_l X_{t-j'_l} Y_{t-j''_l} \right)\|_q &\leq \sum_{l=1}^k |\alpha_l| \left( \|X_{t-j'_l}\|_p \sum_{t=0}^{\infty} \delta_{p,t-j''_l} + \|Y'_{t-j''_l}\|_p \sum_{t=0}^{\infty} \delta_{p,t-j'_l} \right), \\
&< \infty.
\end{aligned}$$

Le même théorème s'applique donc et on obtient :

$$\sum_{l=1}^k \alpha_l \hat{\gamma}_{XY}(j_l) - \sum_{l=1}^k \alpha_l \gamma_{XY}(j_l) \xrightarrow{\mathcal{L}} \mathcal{N} \left( 0, E(D_{0K} D_{0K}^\top) \right),$$

avec

$$D_{0K} = \sum_{t=0}^{\infty} \mathcal{P}_0 \left( \sum_{l=1}^k \alpha_l X_{t-j'_l} Y_{t-j''_l} \right).$$

La distribution asymptotique d'un vecteur quelconque des  $\hat{\gamma}_{XY}(j)$  est alors déduite, et l'expression de  $\mathbf{D}_0$  est mieux comprise en observant que :

$$\begin{aligned}
n^{1/2} \begin{pmatrix} \hat{\gamma}_{XY}(-m) - \gamma_{XY}(-m) \\ \vdots \\ \hat{\gamma}_{XY}(0) - \gamma_{XY}(0) \\ \vdots \\ \hat{\gamma}_{XY}(m) - \gamma_{XY}(-m) \end{pmatrix} &= n^{1/2} \sum_{t=m+1}^n \begin{pmatrix} X_{t-m} Y_t - E(X_{t-m} Y_t) \\ \vdots \\ X_{t-1} Y_t - E(X_{t-1} Y_t) \\ X_t Y_t - E(X_t Y_t) \\ X_t Y_{t-1} - E(X_t Y_{t-1}) \\ \vdots \\ X_t Y_{t-m} - E(X_t Y_{t-m}) \end{pmatrix} + \mathbf{o}_P(1), \\
&= n^{-1/2} \sum_{t=m+1}^n \{ \mathbf{W}_t - E(\mathbf{W}_t) \} + \mathbf{o}_P(1).
\end{aligned}$$

La condition de  $p$ -stabilité englobe celle de Hannan (1976), à savoir que les moments conditionnels d'ordre 1 à 4 sont constants. En effet, si  $\{X_t\}$  et  $\{Y_t\}$  sont des processus MA( $\infty$ )  $X_t = \sum_{i=0}^{\infty} a_{1i} u_{1,t-i}$ ,  $Y_t = \sum_{i=0}^{\infty} a_{2i} u_{2,t-i}$ , et  $\mathcal{F}_{t-1}$  toute l'information

disponible avant la date  $t$  et qu'on suppose :

$$E(u_{i,t}|\mathcal{F}_{t-1}) = c_1,$$

$$E(u_{i,t}u_{j,t}|\mathcal{F}_{t-1}) = c_2,$$

$$E(u_{i,t}u_{j,t}u_{k,t}|\mathcal{F}_{t-1}) = c_3,$$

$$E(u_{i,t}u_{j,t}u_{k,t}u_{l,t}|\mathcal{F}_{t-1}) = c_4,$$

où  $c_1, c_2, c_3$  et  $c_4$  sont constants par rapport à  $t$  et pour tout  $i, j, k, l \in \{1, 2\}$ , alors le processus  $Z_j = \{X_t Y_{t+j}, t \in \mathbb{Z}\}$  vérifie  $\sum_{t=0}^{\infty} \pi_{p,t} < \infty$  pour tout  $0 \leq p \leq 4$ , où  $\pi_{p,t} = \sum_{t=0}^{\infty} \|\mathcal{P}_0(Z_{j,t})\|_p$ . Ceci est aussi expliqué dans la Remarque 6 de Wu et Min (2005). La preuve est présentée au Lemme 3.1 en annexe.

### 3.4. LA STATISTIQUE DE TEST

En posant :

$$\Gamma_m = \{\gamma_X(0)\mathbf{I}_{2m+1}\}^{-1} \Sigma_m \{\gamma_Y(0)\mathbf{I}_{2m+1}\}^{-1},$$

on a sous l'hypothèse de non corrélation :

$$n^{1/2} \begin{pmatrix} \hat{\rho}_{XY}(-m) \\ \vdots \\ \hat{\rho}_{XY}(0) \\ \vdots \\ \hat{\rho}_{XY}(m) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}_{m+1}(\mathbf{0}, \Gamma_m), \quad (3.4.1)$$

où :

$$\hat{\rho}_{XY}(j) = \frac{\hat{\gamma}_{XY}(j)}{\{\hat{\gamma}_X(0)\}^{1/2} \{\hat{\gamma}_Y(0)\}^{1/2}}, \quad -m \leq j \leq m.$$

En appliquant le produit scalaire aux deux membres de l'équation (3.4.1), la convergence en distribution est aussi vérifiée pour les formes quadratiques obtenues. Il en découle d'après Box (1954, Théorème 2.1) que :

$$n \sum_{j=-m}^m \hat{\rho}_{XY}^2(j) \xrightarrow{\mathcal{L}} \sum_{k=1}^{2m+1} \lambda_k \chi_{1,k}^2. \quad (3.4.2)$$

Les variables aléatoires  $\chi_{1,k}^2$ ,  $k = 1, \dots, 2m + 1$  sont des lois de khi-deux à un degré de liberté indépendantes. Les scalaires  $\lambda_k$ ,  $k = 1, \dots, 2m + 1$  sont les valeurs propres de la matrice  $\mathbf{\Gamma}_m$ . Ce résultat a été utilisé par Francq et al. (2005, p. 536). Si la matrice  $\mathbf{\Gamma}_m$  est inversible, alors :

$$n\hat{\boldsymbol{\rho}}_m^\top \mathbf{\Gamma}_m^{-1} \hat{\boldsymbol{\rho}}_m \xrightarrow{\mathcal{L}} \chi_{2m+1}^2, \quad (3.4.3)$$

avec  $\hat{\boldsymbol{\rho}}_m = (\hat{\rho}_{XY}(-m), \dots, \hat{\rho}_{XY}(0), \dots, \hat{\rho}_{XY}(m))^\top$ .

La matrice des variances et covariances  $\mathbf{\Gamma}_m$  n'étant pas toujours inversible (voir Francq et al. (2005, Remarque 3.2)), c'est la statistique de l'équation (3.4.2) qui sera utilisée.

### 3.5. ESTIMATION DE LA MATRICE DES VARIANCES ET COVARIANCES

Il a été vu que  $\boldsymbol{\Sigma}_m = E(\mathbf{D}_0 \mathbf{D}_0^\top)$  et  $\mathbf{D}_0 = \sum_{t=0}^{\infty} \mathcal{P}_0(\mathbf{W}_t) \in L^2$ . D'autre part,  $\mathbf{D}_0 = (D_{0j})_{-m \leq j \leq m}$  et pour  $-m \leq j_1, j_2 \leq m$  :

$$E(D_{0j_1} D_{0j_2}) = \frac{1}{4} \{E(D_{0j_1} + D_{0j_2})^2 - E(D_{0j_1} - D_{0j_2})^2\}.$$

Or :

$$D_{0j_1} + D_{0j_2} = \sum_{t=0}^{\infty} \mathcal{P}_0(X_{t-j_1'} Y_{t-j_1''} + X_{t-j_2'} Y_{t-j_2''}),$$

avec  $j_l' = \frac{|j_l| - j_l}{2}$  et  $j_l'' = \frac{|j_l| + j_l}{2}$ ,  $l = 1, 2$  et on a (voir Shao (2011, Lemme A.2)) :

$$E(D_{0j_1} + D_{0j_2})^2 = \sum_{k \in \mathbb{Z}} \text{cov}(X_{t-j_1'} Y_{t-j_1''} + X_{t-j_2'} Y_{t-j_2''}, X_{t-k-j_1'} Y_{t-k-j_1''} + X_{t-k-j_2'} Y_{t-k-j_2''}),$$

$$E(D_{0j_1} - D_{0j_2})^2 = \sum_{k \in \mathbb{Z}} \text{cov}(X_{t-j_1'} Y_{t-j_1''} - X_{t-j_2'} Y_{t-j_2''}, X_{t-k-j_1'} Y_{t-k-j_1''} - X_{t-k-j_2'} Y_{t-k-j_2''}).$$

On obtient donc après développements que :

$$\begin{aligned} E(D_{0j_1} D_{0j_2}) &= \frac{1}{2} \sum_{k \in \mathbb{Z}} \text{cov}(X_{t-j_1'} Y_{t-j_1''}, X_{t-k-j_2'} Y_{t-k-j_2''}) \\ &\quad + \frac{1}{2} \sum_{k \in \mathbb{Z}} \text{cov}(X_{t-j_2'} Y_{t-j_2''}, X_{t-k-j_1'} Y_{t-k-j_1''}). \end{aligned}$$

Si pour tout  $j_1, j_2$  tels que  $-m \leq j_1, j_2 \leq m$ , les processus,  $\{X_{t-j_1}'Y_{t-j_1}''\}$  et  $\{X_{t-j_2}'Y_{t-j_2}''\}$  sont stationnaires en covariance, alors pour tout  $k \in \mathbb{Z}$ ,

$$\text{cov}(X_{t-j_1}'Y_{t-j_1}'', X_{t-k-j_2}'Y_{t-k-j_2}'') = \text{cov}(X_{t-j_2}'Y_{t-j_2}'', X_{t+k-j_1}'Y_{t+k-j_1}''),$$

ce qui conduit en passant à la somme :

$$\sum_{k \in \mathbb{Z}} \text{cov}(X_{t-j_1}'Y_{t-j_1}'', X_{t-k-j_2}'Y_{t-k-j_2}'') = \sum_{k \in \mathbb{Z}} \text{cov}(X_{t-j_2}'Y_{t-j_2}'', X_{t-k-j_1}'Y_{t-k-j_1}'').$$

Ce qui mène au résultat suivant :

$$\begin{aligned} E(D_{0j_1}D_{0j_2}) &= \sum_{k \in \mathbb{Z}} \text{cov}(X_{t-j_1}'Y_{t-j_1}'', X_{t-k-j_2}'Y_{t-k-j_2}''), \\ &= \sum_{k \in \mathbb{Z}} \text{cov}(X_{t-j_2}'Y_{t-j_2}'', X_{t-k-j_1}'Y_{t-k-j_1}''). \end{aligned}$$

On obtient finalement sous cette condition que :

$$\Sigma_m = \sum_{k \in \mathbb{Z}} \text{cov}(\mathbf{W}_t, \mathbf{W}_{t+k}), \quad (3.5.1)$$

où

$$\mathbf{W}_t = \begin{pmatrix} X_{t-m}Y_t \\ \vdots \\ X_{t-1}Y_t \\ X_tY_t \\ X_tY_{t-1} \\ \vdots \\ X_tY_{t-m} \end{pmatrix}. \quad (3.5.2)$$

La décomposition de Wold pour le processus  $\{\mathbf{W}_t\}$  s'écrit :

$$\mathbf{W}_t = \mathbf{u}_t + \sum_{i=1}^{\infty} \mathbf{B}_i \mathbf{u}_{t-i}, \quad (3.5.3)$$

où  $\{\mathbf{u}_t\}$  est un bruit blanc faible de dimension  $2m + 1$ . Sous les conditions  $\sum_{i=1}^{\infty} \|\mathbf{B}_i\| < \infty$ ,  $\det \left( \mathbf{I}_{2m+1} + \sum_{i=1}^{\infty} \mathbf{B}_i z^i \right) \neq 0$  pour  $|z| \leq 1$  et  $\Sigma_{\mathbf{u}, \mathbf{u}} = \text{var}(\mathbf{u}_t)$  est

une matrice non singulière,  $\mathbf{W}_t$  admet la représentation  $AR(\infty)$  :

$$\mathbf{A}(B)\mathbf{W}_t = \mathbf{W}_t - \sum_{i=1}^{\infty} \mathbf{A}_i \mathbf{W}_{t-i} = \mathbf{u}_t, \quad (3.5.4)$$

avec  $\sum_{i=1}^{\infty} \|\mathbf{A}_i\| < \infty$  et  $\det\{\mathbf{A}(z)\} \neq 0$  si  $|z| \leq 1$ .

Il est d'usage à ce niveau de considérer  $2\pi\Sigma_m$  comme la densité spectrale à la fréquence zéro du processus  $\{\mathbf{W}_t\}$ . On obtient :

$$\Sigma_m = \{\mathbf{A}(1)\}^{-1} \Sigma_{\mathbf{u},\mathbf{u}} \{\mathbf{A}(1)^\top\}^{-1}. \quad (3.5.5)$$

Berk (1974), Francq et al. (2005), Den Haan et Levin (1994) ont montré que la matrice des variances et covariances  $\Sigma_m$  peut être estimée de manière convergente en ajustant un processus autorégressif d'ordre  $p = p(n)$ , tel que  $p \rightarrow \infty$  et  $p^3/n \rightarrow \infty$  lorsque  $n \rightarrow \infty$ . Nous allons montrer dans le cadre de ce document que ces mêmes résultats demeurent valables sous les conditions de l'hypothèse 3.2.

**Hypothèse 3.2.** *Les processus  $\{X_t\}$  et  $\{Y_t\}$  sont tels que  $\{\mathbf{W}_t\}$  défini en (3.5.2) admet la représentation (3.5.4) avec  $\|\mathbf{A}_i\| = o(i^{-2})$  quand  $i \rightarrow \infty$ , les racines de l'équation  $\det\{\mathbf{A}(z)\} = 0$  sont à l'extérieur du cercle unité et  $\Sigma_{\mathbf{u},\mathbf{u}} = \text{var}(\mathbf{u}_t)$  n'est pas singulière. De plus, le processus  $\{\mathbf{Z}_t = (X_t, Y_t)^\top\}$  est CGM( $k$ ) et est stationnaire d'ordre  $k$ , pour tout  $k$  tel que  $1 \leq k \leq 8$ .*

Le théorème suivant s'inscrit dans la même logique que Francq et al. (2005) et Den Haan et Levin (1994).

**Théorème 3.2.** *En notant  $\hat{\mathbf{A}}_{p,1}, \dots, \hat{\mathbf{A}}_{p,p}$  les coefficients de la régression multivariée par les moindres carrés ordinaires de  $\mathbf{W}_t$  sur  $\mathbf{W}_{t-1}, \dots, \mathbf{W}_{t-p}$  pour  $t = 1, \dots, n - m$ ,  $\hat{\mathbf{A}}_p(1) = \mathbf{I}_{2m+1} - \sum_{i=1}^p \hat{\mathbf{A}}_{p,i}$  et  $\hat{\Sigma}_{\mathbf{u}_p, \mathbf{u}_p}$ , l'estimateur de la matrice des variances et covariances des erreurs, nous avons :*

$$\hat{\Sigma}_m = \{\hat{\mathbf{A}}_p(1)\}^{-1} \hat{\Sigma}_{\mathbf{u}_p, \mathbf{u}_p} \{\hat{\mathbf{A}}_p(1)^\top\}^{-1} \rightarrow \{\mathbf{A}(1)\}^{-1} \Sigma_{\mathbf{u},\mathbf{u}} \{\mathbf{A}(1)^\top\}^{-1} \quad (3.5.6)$$

*en probabilité lorsque  $p = p(n) \rightarrow \infty$  et  $p^3/n \rightarrow \infty$  lorsque  $n \rightarrow \infty$ .*

*Preuve* : Il suffit de montrer que  $\hat{\mathbf{A}}_p(1)$  et  $\hat{\Sigma}_{\mathbf{u}_p, \mathbf{u}_p}$  convergent en probabilité respectivement vers  $\mathbf{A}_p(1)$  et  $\Sigma_{\mathbf{u}, \mathbf{u}}$  lorsque  $p = p(n) \rightarrow \infty$  et  $p^3/n \rightarrow \infty$ . La démarche est similaire à Francq et al. (2005, Théorème 5.2) et s'appuie sur les Lemmes 3.2 à 3.7 présentés en annexe.

### 3.6. SIMULATIONS

Dans la section précédente, nous avons présenté la procédure d'estimation de la matrice des variances et covariances de même que la statistique de test pour la non corrélation sous faible dépendance entre deux processus. Pour apprécier les propriétés à distance fini de cette procédure, nous allons présenter les résultats de simulation de Monte-Carlo destinés à étudier le niveau exact du test en question. Pour comparer la distribution exacte des statistiques de tests à ceux de la vraie distribution, le modèle suivant a été utilisé :

$$\begin{aligned} X_t &= (11/30)X_{t-1} - (1/30)X_{t-2} + u_t, \\ Y_t &= (1/2)Y_{t-1} + v_t. \end{aligned}$$

Les processus  $\{u_t\}$  et  $\{v_t\}$  sont donnés par :

$$\begin{aligned} u_t &= \xi_t + \xi_{t-1}^3, \\ v_t &= \xi_t^2 + \xi_{t-1}^2 - 2/3. \end{aligned}$$

Les processus  $\{\xi_t, t \in \mathbb{Z}\}$  sont iid et suivent une loi  $U(-1, 1)$ .

Les processus  $\{X_t\}$  et  $\{Y_t\}$  sont autorégressifs non corrélés admettant chacun une représentation moyenne mobile infinie avec des innovations dépendantes. Nous examinons la fréquence empirique de rejet de l'hypothèse nulle de non corrélation aux seuils de 5% et 10% , respectivement. Ceci a été fait pour des tailles d'échantillon  $n = 100, 200, 300, 500$  et 1000 et des ordres d'autocorrélation croisée  $m = 3, 5, 7, 10$  et 12. Pour chaque valeur du seuil, de taille d'échantillon et d'ordre d'autocorrélation, un total de 1000 répliquions ont été faites, et à chaque fois, la fréquence empirique de rejet de l'hypothèse nulle de non corrélation a été

reportée. Les statistiques suivantes font partie de l'étude :

$$Q_1(m) = n \sum_{j=-m}^m \hat{\rho}_{XY}^2(j), \quad (3.6.1)$$

$$Q_2(m) = n \sum_{j=-m}^m \frac{n}{n - |j|} \hat{\rho}_{XY}^2(j),$$

ont été considérées. Les quantiles empiriques ont été calculés en utilisant l'algorithme Imhof du package CompQuadForm développé par Duchesne et Lafaye de Micheaux (2010). Ces deux statistiques ont été comparées à la statistique du test de Haugh :

$$Q_H(m) = Q_2(m) = n \sum_{j=-m}^m \frac{n}{n - |j|} \hat{\rho}_{XY}^2(j). \quad (3.6.2)$$

Nous parlons de la démarche de Haugh quand nous supposons que cette statistique suit asymptotiquement une loi de khi-deux à  $2m + 1$  degrés de libertés.

Les résultats sont présentés dans les Tableaux 3.1 et 3.2.

TABLE 3.1. Niveau empirique des statistiques de test au seuil de 5%.

5%									
$n = 300$			$n = 500$			$n = 1000$			
$m$	$Q_1(m)$	$Q_2(m)$	$Q_H(m)$	$Q_1(m)$	$Q_2(m)$	$Q_H(m)$	$Q_1(m)$	$Q_2(m)$	$Q_H(m)$
3	0.040	0.040	0.104	0.062	0.062	0.116	0.050	0.050	0.070
5	0.062	0.062	0.128	0.052	0.052	0.174	0.058	0.058	0.112
7	0.044	0.046	0.182	0.054	0.056	0.256	0.042	0.042	0.136
10	0.062	0.066	0.272	0.050	0.054	0.428	0.046	0.046	0.168
12	0.048	0.050	0.336	0.058	0.064	0.558	0.066	0.058	0.178



TABLE 3.2. Niveau empirique des statistiques de test au seuil de 10%.

10%									
$n = 300$			$n = 500$			$n = 1000$			
$m$	$Q_1(m)$	$Q_2(m)$	$Q_H(m)$	$Q_1(m)$	$Q_2(m)$	$Q_H(m)$	$Q_1(m)$	$Q_2(m)$	$Q_H(m)$
3	0.120	0.120	0.198	0.108	0.108	0.172	0.114	0.114	0.104
5	0.086	0.086	0.252	0.112	0.114	0.206	0.126	0.126	0.166
7	0.098	0.100	0.354	0.104	0.104	0.268	0.084	0.086	0.218
10	0.090	0.092	0.550	0.092	0.096	0.380	0.088	0.092	0.246
12	0.078	0.086	0.654	0.082	0.088	0.426	0.086	0.092	0.306

Il apparaît dans les deux tableaux que les seuils théoriques des statistiques de test  $Q_1(m)$  et  $Q_2(m)$  définies à l'équation (3.6.1) sont proches des seuils théoriques. Par contre, celles du test de Haugh  $Q_H(m)$  définies en (3.6.2) sont plus éloignées du seuil théorique. De ce fait, utiliser les tests de Haugh pour juger de la non corrélation peut être trompeur car la distribution asymptotique n'est pas correcte quand il existe de la dépendance dans les séries chronologiques.

### 3.7. CONCLUSION

Dans ce document, il a été question de tester la non corrélation sous faible dépendance entre deux processus stationnaires. La faible dépendance a été considérée suivant l'approche des mesures de dépendance physiques et prédictives de Wu (2005). Cela a permis sous les conditions de stabilité d'obtenir la distribution asymptotique des vecteurs des covariances croisées. Ce résultat est plus général que celui qui a l'habitude d'être obtenu sous les hypothèses d'indépendance ou des conditions de Hannan (1976).

D'autre part, l'estimation de la matrice des variances et covariances des covariances croisées s'est appuyée sur les approches de Francq et al. (2005) et Den Haan et Levin (1974), s'inspirant de Berk (1974). Malgré la contrainte due à l'ajustement VAR d'un processus multidimensionnel, cette approche permet néanmoins de contourner la nullité des moments conditionnels d'ordre quatre. C'est cette dernière qui jusqu'ici a été utilisée dans la plupart des articles. La nouvelle approche

présente aussi un cadre dans lequel le test de non corrélation peut directement être appliqué sur les données brutes.

Les tableaux des simulations montrent que les seuils empiriques de 5% et 10% sont proches des seuils théoriques. La statistique de test construite se comporte mieux que celle de Haugh (1976) dont la distribution asymptotique est erronée en présence de dépendance.

## ANNEXE

**Lemme 3.1.** Soient  $X_t = \sum_{i=0}^{\infty} a_{1i}u_{1,t-i}$  et  $Y_t = \sum_{i=0}^{\infty} a_{2i}u_{2,t-i}$  deux processus linéaires tels que pour tout  $i, j, k, l \in \{1, 2\}$  :

$$E(u_{i,t}|\mathcal{F}_{t-1}) = c_1,$$

$$E(u_{i,t}u_{j,t}|\mathcal{F}_{t-1}) = c_2,$$

$$E(u_{i,t}u_{j,t}u_{k,t}|\mathcal{F}_{t-1}) = c_3,$$

$$E(u_{i,t}u_{j,t}u_{k,t}u_{l,t}|\mathcal{F}_{t-1}) = c_4,$$

où  $\mathcal{F}_t = \sigma(\dots, u_{1,t-1}, u_{2,t-1}, u_{1,t}, u_{2,t})$  est la  $\sigma$ -algèbre engendrée par :

$$(\dots, u_{1,t-1}, u_{2,t-1}, u_{1,t}, u_{2,t}),$$

alors le processus  $Z_j = \{X_t Y_{t+j}\}$  vérifie  $\sum_{t=0}^{\infty} \pi_{p,t} < \infty$  pour tout  $0 \leq p \leq 4$ , où

$$\pi_{p,t} = \sum_{t=0}^{\infty} \|\mathcal{P}_0(Z_{j,t})\|_p.$$

*Preuve :* Posons  $u_{1,t} = u_t$ ,  $u_{2,t} = v_t$ ,  $a_{1i} = a_i$  et  $a_{2i} = b_i$ . Pour tout  $i \geq 0$  et pour tout  $t$ , on a :

$$\begin{aligned} X_t Y_{t+j} &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} a_k b_l u_{t-k} v_{t-l+j}, \\ &= a_t b_{t+j} u_0 v_0 + a_t u_0 \sum_{\substack{l=0 \\ l \neq t+j}}^{\infty} b_l v_{t-l+j} + v_0 b_{t+j} \sum_{\substack{k=0 \\ k \neq t}}^{\infty} a_k u_{t-k} + \sum_{\substack{k,l=0 \\ k \neq t \\ l \neq t+j}}^{\infty} a_k b_l u_{t-k} v_{t-l+j}, \\ &= A_{t,j} + B_{t,j} + C_{t,j} + D_{t,j}. \end{aligned}$$

D'autre part,  $\mathcal{P}_0(X_t Y_{t+j}) = \mathcal{P}_0(A_{t,j}) + \mathcal{P}_0(B_{t,j}) + \mathcal{P}_0(C_{t,j}) + \mathcal{P}_0(D_{t,j})$ . De plus :

$$\begin{aligned} \mathcal{P}_0(A_{t,j}) &= E(A_{t,j}|\mathcal{F}_0) - E(A_{t,j}|\mathcal{F}_{-1}) = E(a_t b_{t-j} u_0 v_0|\mathcal{F}_0) - E(a_t b_{t-j} u_0 v_0|\mathcal{F}_{-1}), \\ &= a_t b_{t-j} u_0 v_0 - c_2 a_t b_{t-j} = a_t b_{t-j} (u_0 v_0 - c_2) = a_t b_{t-j} \mathcal{P}_0(u_0 v_0). \end{aligned}$$

$$\mathcal{P}_0(B_{t,j}) = \mathcal{P}_0(a_t u_0 \sum_{\substack{l=0 \\ l \neq t+j}}^{\infty} b_l v_{t-l+j}) = \sum_{\substack{l=0 \\ l \neq t+j}}^{\infty} a_t b_l \mathcal{P}_0(u_0 v_{t-l+j}),$$

$$\begin{aligned}
&= \sum_{t-l+j < 0} a_t b_l \mathcal{P}_0(u_0 v_{t-l+j}) + \sum_{t-l+j > 0} a_t b_l \mathcal{P}_0(u_0 v_{t-l+j}), \\
&= \sum_{t-l+j < 0} a_t b_l v_{t-l+j} \mathcal{P}_0(u_0) + \sum_{t-l+j > 0} a_t b_l c_1 \mathcal{P}_0(u_0),
\end{aligned}$$

car pour  $t - l + j > 0$ , on a :

$$\begin{aligned}
\mathcal{P}_0(u_0 v_{t-l+j}) &= E(u_0 v_{t-l+j} | \mathcal{F}_0) - E(u_0 v_{t-l+j} | \mathcal{F}_{-1}), \\
&= u_0 E(v_{t-l+j} | \mathcal{F}_0) - E(u_0 v_{t-l+j} | \mathcal{F}_{-1}), \\
&= u_0 E\{E(v_{t-l+j} | \mathcal{F}_0) | \mathcal{F}_{t-l+j-1}\} - E\{E(u_0 v_{t-l+j} | \mathcal{F}_{-1}) | \mathcal{F}_{t-l+j-1}\}, \\
&= u_0 E\{E(v_{t-l+j} | \mathcal{F}_{t-l+j-1}) | \mathcal{F}_0\} - E\{u_0 E(v_{t-l+j} | \mathcal{F}_{t-l+j-1}) | \mathcal{F}_{-1}\}, \\
&= u_0 E(c_1 | \mathcal{F}_0) - E(u_0 c_1 | \mathcal{F}_{-1}), \\
&= c_1 \{u_0 - E(u_0 | \mathcal{F}_{-1})\}, \\
&= c_1 \mathcal{P}_0(u_0).
\end{aligned}$$

$$\begin{aligned}
\mathcal{P}_0(C_{t,j}) &= \mathcal{P}_0 \left( v_0 b_{t+j} \sum_{\substack{k=0 \\ k \neq t}}^{\infty} a_k u_{t-k} \right) = \sum_{\substack{k=0 \\ k \neq t}}^{\infty} a_k b_{t+j} \mathcal{P}_0(v_0 u_{t-k}), \\
&= \sum_{t-k < 0} a_k b_{t+j} \mathcal{P}_0(v_0 u_{t-k}) + \sum_{t-k > 0} a_k b_{t+j} \mathcal{P}_0(v_0 u_{t-k}), \\
&= \sum_{t-k < 0} a_k b_{t+j} u_{t-k} \mathcal{P}_0(v_0) + \sum_{t-k > 0} c_1 a_k b_{t+j} \mathcal{P}_0(v_0),
\end{aligned}$$

car pour  $t - k > 0$  le même raisonnement s'applique que pour le cas  $t - l + j > 0$  ci-dessus. De plus,

$$\begin{aligned}
\mathcal{P}_0(D_{t,j}) &= \mathcal{P}_0 \left( \sum_{\substack{k \neq t \\ l \neq t+j}}^{\infty} a_k b_l u_{t-k} v_{t-l+j} \right), \\
&= \sum_{\substack{k \neq t \\ l \neq t+j}}^{\infty} a_k b_l \mathcal{P}_0(u_{t-k} v_{t-l+j}), \\
&= 0,
\end{aligned}$$

car si  $t - k > t - l + j > 0$ , alors :

$$\begin{aligned}
\mathcal{P}_0(u_{t-k}v_{t-l+j}) &= E(u_{t-k}v_{t-l+j}|\mathcal{F}_0) - E(u_{t-k}v_{t-l+j}|\mathcal{F}_{-1}), \\
&= E\{E(u_{t-k}v_{t-l+j}|\mathcal{F}_0)|\mathcal{F}_{t-k-1}\} - E\{E(u_{t-k}v_{t-l+j}|\mathcal{F}_{-1})|\mathcal{F}_{t-k-1}\}, \\
&= E\{v_{t-l+j}E(u_{t-k}|\mathcal{F}_{t-k-1})|\mathcal{F}_0\} - E\{v_{t-l+j}E(u_{t-k}|\mathcal{F}_{t-k-1})|\mathcal{F}_{-1}\}, \\
&= c_1\{E(v_{t-l+j}|\mathcal{F}_0) - E(v_{t-l+j}|\mathcal{F}_{-1})\} = c_1\mathcal{P}_0(v_{t-l+j}) = 0.
\end{aligned}$$

Si  $t - k < t - l + j < 0$ , alors :

$$\begin{aligned}
\mathcal{P}_0(u_{t-k}v_{t-l+j}) &= E(u_{t-k}v_{t-l+j}|\mathcal{F}_0) - E(u_{t-k}v_{t-l+j}|\mathcal{F}_{-1}), \\
&= u_{t-k}v_{t-l+j} - u_{t-k}v_{t-l+j}, \\
&= 0.
\end{aligned}$$

Le même raisonnement s'applique pour  $t - k > 0 > t - l + j$ ,  $t - l + j > 0 > t - k$  et  $t - l + j = t - k$ . On trouve alors que :

$$\begin{aligned}
\mathcal{P}_0(X_t Y_{t+j}) &= a_t b_{t-j} \mathcal{P}_0(u_0 v_0) + \sum_{t-l+j < 0} a_t b_l v_{t-l+j} \mathcal{P}_0(u_0) + \sum_{t-l+j > 0} a_t b_l c_1 \mathcal{P}_0(u_0) + \\
&\quad \sum_{t-k < 0} a_k b_{t+j} u_{t-k} \mathcal{P}_0(v_0) + \sum_{t-k > 0} c_1 a_k b_{t+j} \mathcal{P}_0(v_0).
\end{aligned}$$

Ceci entraîne que :

$$|E\{\mathcal{P}_0(X_t Y_{t+j})\}| < K_0 (|a_t| |b_{t-j}| + |a_t| + |b_{t+j}|),$$

où

$$\begin{aligned}
K_0 &= 4 \max\{\beta_1, \beta_2, \beta_3\}, \\
\beta_1 &= E(|\mathcal{P}_0(u_0 v_0)|^2), \\
\beta_2 &= E(|\mathcal{P}_0(u_0)|^2) \{E(v_t^2) + |c_1|\} \sum_{l \geq 0} |b_l|, \\
\beta_3 &= E(|\mathcal{P}_0(v_0)|^2) \{E(u_t^2) + |c_1|\} \sum_{k \geq 0} |a_k|.
\end{aligned}$$

Il en découle alors que  $\sum_{t=0}^{\infty} |E(\mathcal{P}_0(X_t Y_{t+j}))| < \infty$ .

**Lemme 3.2.** *Sous l'hypothèse 3.2, on a pour tout  $j_1, j_2$  tels que  $-m \leq j_1, j_2 \leq m$  :*

$$\sum_{k \in \mathbb{Z}} |\text{cov}(X_{t+j_1'} Y_{t+j_1''}, X_{t+k+j_2'} Y_{t+k+j_2''})| < \infty,$$

où  $j_l' = \frac{j_l + |j_l|}{2}$  et  $j_l'' = \frac{|j_l| - j_l}{2}$ ,  $l = 1, 2$ .

*Preuve :* La preuve est similaire à celle de Wu et Shao (2004, Proposition 2).

On a sous l'hypothèse de non corrélation et pour un  $k$  donné :

$$\begin{aligned} Q &= \text{cov}(X_{t+j_1'} Y_{t+j_1''}, X_{t+k+j_2'} Y_{t+k+j_2''}), \\ &= E(X_{t+j_1'} Y_{t+j_1''} X_{t+k+j_2'} Y_{t+k+j_2''}), \\ &= E(X_{j_1'-j_0} Y_{j_1''-j_0} X_{k+j_2'-j_0} Y_{k+j_2''-j_0}), \\ &= E(X_0 X_{m_1} X_{m_2} X_{m_3}). \end{aligned}$$

Ceci à cause de la stationnarité du processus  $\{X_{t+j_1'} Y_{t+j_1''} X_{t+k+j_2'} Y_{t+k+j_2''}\}$  telle que supposée,  $j_0 = \min(j_1', j_1'', k + j_2', k + j_2'')$  et  $0 \leq m_1 \leq m_2 \leq m_3$ .

Le quadruplet  $(X_0, X_{m_1}, X_{m_2}, X_{m_3})$  avec  $0 \leq m_1 \leq m_2 \leq m_3$  est une permutation possible de  $(X_{j_1'-j_0}, Y_{j_1''-j_0}, X_{k+j_2'-j_0}, Y_{k+j_2''-j_0})$  de telle sorte que les indices soient mis dans cet ordre croissant, la variable  $X$  peut prendre chacune des valeurs  $X_{j_1'-j_0}, Y_{j_1''-j_0}, X_{k+j_2'-j_0}$  ou  $Y_{k+j_2''-j_0}$ .

Posons  $m_0 = 0$  et  $Y_0 = X_{m_0} X_{m_1} X_{m_2}$ , on a alors :

$$Q = E(X_0 X_{m_1} X_{m_2} X_{m_3}) = E(Y_0 X_{m_3}) = E(Y_0 (X_{m_3} - X'_{m_3})) + E(Y_0 X'_{m_3}).$$

Or  $Y_0$  étant indépendant de  $X'_{m_3}$ ,  $E(Y_0 X'_{m_3}) = E(Y_0) E(X'_{m_3}) = 0$  car  $E(X'_{m_3}) = 0$ , ceci par définition des processus  $\{X_t\}$  et  $\{Y_t\}$ . D'après l'inégalité de Hölder et de Jensen :

$$|E\{Y_0 (X_{m_3} - X'_{m_3})\}| \leq \|X_{m_3} - X'_{m_3}\|_4 \|Y_0\|_{4/3}.$$

Puisque les processus  $\{X_t\}$  et  $\{Y_t\}$  appartiennent à  $L^4$ , on a :

$$\|Y_0\|_{4/3} = \|X_{m_0} X_{m_1} X_{m_2}\|_{4/3} < C_1,$$

où  $C_1$  ne dépend pas de  $m_0, m_1$  et  $m_2$ .

D'autre part, d'après la propriété CGM(4) du processus,  $(X_t, Y_t)^\top$ ,  $\|X_{m_3} - X'_{m_3}\|_4 \leq C_2 r^{m_3}$ . On obtient donc en fin de compte  $Q \leq C r^{m_3}$ , où  $C$  est une constante et  $0 < r < 1$ . En effet,  $m_3 = m_3(k) \rightarrow \infty$ , lorsque  $k \rightarrow \infty$ . On déduit alors que :

$$\begin{aligned} \sum_{k \in \mathbb{Z}} |\text{cov}(X_{t+j'_1} Y_{t+j''_1}, X_{t+k+j'_2} Y_{t+k+j''_2})| &\leq 2 \sum_{s=0}^{\infty} \sum_{(m_1, m_2, m_3) \in L(s)} |E(X_0 X_{m_1} X_{m_2} X_{m_3})|, \\ &\leq \sum_{s=0}^{\infty} C r^s < \infty, \end{aligned}$$

où  $C$  est une constante et  $L(s) = \{(m_1, m_2, m_3) \in \mathbb{Z} : \max_{1 \leq i \leq 3} |m_i| = s\}$ .

Considérons le problème de régression de  $\mathbf{W}_t$  sur  $\mathbf{W}_{t-1}, \dots, \mathbf{W}_{t-p}$  :

$$\mathbf{W}_t = \sum_{i=1}^p \mathbf{A}_{p,i} \mathbf{W}_{t-i} + \mathbf{u}_{p,t}, \quad \mathbf{u}_{p,t} \perp \{\mathbf{W}_{t-1}, \dots, \mathbf{W}_{t-p}\}.$$

Posons  $\underline{\mathbf{A}}_p = (\mathbf{A}_{p,1}, \dots, \mathbf{A}_{p,p})$  et  $\underline{\mathbf{W}}_{p,t} = (\mathbf{W}_{t-1}^\top, \dots, \mathbf{W}_{t-p}^\top)^\top$ . L'équation de régression devient :

$$\mathbf{W}_t = \underline{\mathbf{A}}_p \underline{\mathbf{W}}_{p,t} + \mathbf{u}_{p,t}. \quad (3.7.1)$$

En empilant toutes les observations de  $p+1$  à  $n$ , on obtient :

$$(\mathbf{W}_{p+1}, \dots, \mathbf{W}_n) = \underline{\mathbf{A}}_p (\underline{\mathbf{W}}_{p,p+1}, \dots, \underline{\mathbf{W}}_{p,n}) + (\mathbf{u}_{p,p+1}, \dots, \mathbf{u}_{p,n}).$$

L'estimateur des moindres carrés ordinaires de  $\underline{\mathbf{A}}_p$  est donc :

$$\hat{\underline{\mathbf{A}}}_p = \hat{\Sigma}_{\mathbf{W}, \underline{\mathbf{W}}_p} \hat{\Sigma}_{\underline{\mathbf{W}}_p, \underline{\mathbf{W}}_p}^{-1},$$

avec  $\hat{\Sigma}_{\mathbf{W}, \underline{\mathbf{W}}_p} = \frac{1}{n-p} \sum_{t=p+1}^n \mathbf{W}_t \underline{\mathbf{W}}_{p,t}^\top$  et  $\hat{\Sigma}_{\underline{\mathbf{W}}_p, \underline{\mathbf{W}}_p} = \frac{1}{n-p} \sum_{t=p+1}^n \underline{\mathbf{W}}_{p,t} \underline{\mathbf{W}}_{p,t}^\top$ .

L'estimateur des moindres carrés de la matrice des variances et covariances des erreurs est alors donné par :

$$\hat{\Sigma}_{\mathbf{u}_p, \mathbf{u}_p} = \frac{1}{n-p} \sum_{t=p+1}^n (\mathbf{W}_t - \hat{\underline{\mathbf{A}}}_p \underline{\mathbf{W}}_{p,t})(\mathbf{W}_t - \hat{\underline{\mathbf{A}}}_p \underline{\mathbf{W}}_{p,t})^\top.$$

**Lemme 3.3.** Posons  $\Sigma_{\underline{\mathbf{W}}_p, \underline{\mathbf{W}}_p} = E(\underline{\mathbf{W}}_{p,t} \underline{\mathbf{W}}_{p,t}^\top)$ ,  $\Sigma_{\mathbf{W}, \underline{\mathbf{W}}_p} = E(\mathbf{W}_t \underline{\mathbf{W}}_{p,t}^\top)$  alors :

$$\sup \max_{\{p \geq 1\}} \left( \|\Sigma_{\underline{\mathbf{W}}_p, \underline{\mathbf{W}}_p}\|, \|\Sigma_{\mathbf{W}, \underline{\mathbf{W}}_p}\|, \|\Sigma_{\underline{\mathbf{W}}_p, \underline{\mathbf{W}}_p}^{-1}\| \right) < \infty.$$

*Preuve :* Elle est similaire à Francq et al. (2005, Lemme A.2).

**Lemme 3.4.** Soit  $\eta_{t,j} = X_{t+j'} Y_{t+j''}$ , où  $j' = \frac{j+|j|}{2}$  et  $j'' = \frac{|j|-j}{2}$  si pour tout  $k, j_1, j_2, s \in \mathbb{Z}$ , le processus  $\{\eta_{t,j_1} \eta_{t+s,j_2} \eta_{t+k,j_1} \eta_{t+k+s,j_2}\}$  est stationnaire par rapport à  $t$ , et le processus  $\mathbf{Z}_t = (X_t, Y_t)^\top$  est CGM(8), alors, nous avons :

$$\sup_{\{s \in \mathbb{Z}\}} \sum_{k \in \mathbb{Z}} |E(\eta_{t,j_1} \eta_{t+s,j_2} \eta_{t+k,j_1} \eta_{t+k+s,j_2})| < \infty.$$

*Preuve :* La preuve est la même que celle du lemme 3.2 et s'appuie aussi sur Wu et Shao (2004, Proposition 2).

**Lemme 3.5.** Sous les hypothèses 3.1 et 3.2,  $p^{1/2} \|\hat{\Sigma}_{\underline{\mathbf{W}}_p, \underline{\mathbf{W}}_p} - \Sigma_{\underline{\mathbf{W}}_p, \underline{\mathbf{W}}_p}\|$ ,  $p^{1/2} \|\hat{\Sigma}_{\mathbf{W}, \mathbf{W}} - \Sigma_{\mathbf{W}, \mathbf{W}}\|$ , et  $p^{1/2} \|\hat{\Sigma}_{\mathbf{W}, \underline{\mathbf{W}}_p} - \Sigma_{\mathbf{W}, \underline{\mathbf{W}}_p}\|$  convergent en probabilité vers zéro, lorsque  $n \rightarrow \infty$  et  $p = o(n^{1/3})$ , avec :

$$\hat{\Sigma}_{\mathbf{W}, \mathbf{W}} = \frac{1}{n-p} \sum_{t=p+1}^n \mathbf{W}_t \mathbf{W}_t^\top,$$

et

$$\Sigma_{\mathbf{W}, \mathbf{W}} = E(\mathbf{W}_t \mathbf{W}_t^\top).$$

*Preuve :* Elle est identique à Francq et al. (2005), en constatant que, pour  $i_1, i_2$  tel que,  $1 \leq i_1, i_2 \leq p$  et  $l_1, l_2$  tels que  $0 \leq l_1, l_2 \leq 2m+1$ , l'élément de la ligne  $\{(i_1-1)(2m+1)+l_1\}$  et de la colonne  $\{(i_2-1)(2m+1)+l_2\}$  de est de la forme,  $\frac{1}{n-p} \sum_{t=p_0+1}^n \eta_t^{(4)}$  où  $i_0 = \min(i_1, i_2)$ ,  $\eta_t^{(4)} = X_{t-i_1+j_1'} Y_{t-i_1+j_1''} X_{t-i_2+j_2'} Y_{t-i_2+j_2''}$ ,  $j_1 = l_1 - m - 1$ ,  $j_2 = l_2 - m - 1$ ,  $j_l' = \frac{j_l+|j_l|}{2}$  et  $j_l'' = \frac{|j_l|-j_l}{2}$ ,  $l = 1, 2$ .

**Lemme 3.6.** Soit  $\underline{\mathbf{A}}_p^* = (\mathbf{A}_1, \dots, \mathbf{A}_p)$  où les  $\mathbf{A}_i$  sont définis dans l'équation (3.5.4) alors,

$$p^{1/2} \|\underline{\mathbf{A}}_p^* - \underline{\mathbf{A}}_p\| \rightarrow 0 \text{ lorsque } p \rightarrow \infty. \quad (3.7.2)$$

*Preuve :* On a :

$$\mathbf{W}_t = \underline{\mathbf{A}}_p \underline{\mathbf{W}}_{p,t} + \mathbf{u}_{p,t},$$



$$\begin{aligned}
&= \underline{\mathbf{A}}_p^* \underline{\mathbf{W}}_{p,t} + \sum_{i=p+1}^{\infty} \mathbf{A}_i \mathbf{W}_{t-i} + \mathbf{u}_t, \\
&= \underline{\mathbf{A}}_p^* \underline{\mathbf{W}}_{p,t} + \mathbf{u}_{p,t}^*.
\end{aligned}$$

En faisant la différence du premier et du dernier terme de droite, on a :

$$(\underline{\mathbf{A}}_p - \underline{\mathbf{A}}_p^*) \underline{\mathbf{W}}_{p,t} = \mathbf{u}_{p,t} - \mathbf{u}_{p,t}^*.$$

En multipliant les deux membres par  $\underline{\mathbf{W}}_{p,t}^\top$  et en prenant l'espérance, l'orthogonalité entre  $\mathbf{u}_{p,t}$  et  $\{\mathbf{W}_{t-1}, \dots, \mathbf{W}_{t-p}\}$  permet d'obtenir :

$$\underline{\mathbf{A}}_p - \underline{\mathbf{A}}_p^* = -\Sigma_{\mathbf{u}_{p,t}^*, \underline{\mathbf{W}}_p} \Sigma_{\underline{\mathbf{W}}_p, \underline{\mathbf{W}}_p}^{-1},$$

où  $\Sigma_{\mathbf{u}_{p,t}^*, \underline{\mathbf{W}}_p} = E(\mathbf{u}_{p,t}^* \underline{\mathbf{W}}_{p,t}^\top)$ . En considérant les résultats du lemme 3.3, il suffit de montrer que  $\|\Sigma_{\mathbf{u}_{p,t}^*, \underline{\mathbf{W}}_p}\|$  converge vers zéro lorsque  $p \rightarrow \infty$ .

On a :

$$\|\Sigma_{\mathbf{u}_{p,t}^*, \underline{\mathbf{W}}_p}\| = \left\| \sum_{i=p+1}^{\infty} \mathbf{A}_i E(\mathbf{W}_{t-i} \mathbf{W}_{p,t}^\top) \right\| \leq \left\| \sum_{k=1}^{\infty} \mathbf{A}_{r+k} E(\mathbf{W}_{t-r-k} \mathbf{W}_{p,t}^\top) \right\|.$$

Or, d'après les propriétés de la norme matricielle utilisée,

$$\begin{aligned}
\|E(\mathbf{W}_{t-r-k} \mathbf{W}_{p,t}^\top)\| &\leq (2m+1)p^{1/2} |E(X_{t+j_1'} Y_{t+j_1''}, X_{t+k+j_2'} Y_{t+k+j_2''})|, \\
&\leq \|X_t\|_4^2 \|Y_t\|_4^2.
\end{aligned}$$

On en déduit alors que :

$$\|\Sigma_{\mathbf{u}_{p,t}^*, \underline{\mathbf{W}}_p}\| \leq Cp^{1/2} \sum_{k=1}^{\infty} \|\mathbf{A}_{r+k}\|,$$

où  $C$  est une constante. Vu l'hypothèse suivant laquelle,  $\|\mathbf{A}_i\| = o(i^{-2})$ , on a  $p \sum_{k=1}^{\infty} \|\mathbf{A}_{r+k}\| = o(1)$  lorsque  $p \rightarrow \infty$ , ce qui donne le résultat.

**Lemme 3.7.** *Sous les hypothèses 3.1 et 3.2,*

$$\begin{aligned}
p^{1/2} \|\hat{\Sigma}_{\underline{\mathbf{W}}_p, \underline{\mathbf{W}}_p}^{-1} - \Sigma_{\underline{\mathbf{W}}_p, \underline{\mathbf{W}}_p}^{-1}\| &= o_P(1), \\
\|\hat{\underline{\mathbf{A}}}_p^* - \underline{\mathbf{A}}_p\| &= o_p(1),
\end{aligned}$$

lorsque  $n \rightarrow \infty$ ,  $p = o(n^{1/3})$  et  $p \rightarrow \infty$ .

*Preuve* : Avec les hypothèses 3.1 et 3.2, la démarche est similaire à Francq et al. (2005, lemmes A.6 et A.7).

# Chapitre 4

---

## SUR L'ESTIMATION ET LA VALIDATION DES MODÈLES PÉRIODIQUES MULTIVARIÉS ET COINTÉGRÉS

### 4.1. INTRODUCTION

Les phénomènes naturels présentant des caractéristiques qui varient dans le temps sont d'un intérêt particulier dans l'étude des données temporelles. Les exemples courants sont les processus saisonniers tels que la température, le débit des cours d'eau et le transport aérien. Une classe de modèles qui permet d'étudier ces phénomènes est celle des modèles périodiques. Dans ces modèles, les paramètres peuvent varier d'une saison à une autre alors que dans les modèles de Box et Jenkins de type SARIMA et VARMA, ils sont les mêmes pour toutes les saisons. Les modèles de séries chronologiques périodiques ont été appliqués en économie, en météorologie, en hydrologie et bien d'autres domaines.

Pour mieux représenter les processus périodiques, il est nécessaire de pouvoir identifier pour chaque observation la saison à laquelle elle correspond. Par conséquent, le paramètre d'indexation dépend de la saison. Ainsi, dans la définition du processus périodique  $\mathbf{Y} = \{\mathbf{Y}_t, t \in \mathbb{Z}\}$ , on pose  $t = sn + \nu$ , où  $s$  représente le nombre total de saisons et  $\nu = 1, \dots, s$  une saison donnée. Ces processus représentent ainsi une succession de cycles complets de saisons. Ces cycles complets sont indexés par  $n \in \mathbb{Z}$ . Un des exemples les plus courants de cycle complet de saisons est l'année. La notation  $t = sn + \nu$  réfère à saison  $\nu$  de l'année  $n + 1$ .

Certains auteurs se sont penchés sur l'estimation, la prédiction et l'inférence dans le cadre des séries chronologiques périodiques univariées. On peut citer les travaux de Jones et Brelsford (1967), Pagano (1978) et Tiao et Grupe (1980). Dans le contexte des modèles multivariés, on retrouve par exemple Birchenhall et al. (1989), Franses et Paap (2004) et Ursu et Duchesne (2009).

Ces études supposent que les séries univariées étudiées sont stationnaires au sens périodique. Ce n'est pas toujours le cas. Tout comme dans le cas non périodique, la non stationnarité peut être due à la présence de racines unitaires dans la représentation périodique autorégressive (PAR). On parle ici d'intégration au sens périodique. En présence de racines unitaires, il faut corriger la non stationnarité avant de procéder à l'estimation. Dans les modèles non périodiques univariés, ceci se fait en général par la différentiation dont l'ordre dépend de la multiplicité de la racine unitaire présente. Dans les modèles périodiques, le filtre peut s'avérer plus compliqué que la différentiation classique. On parle de quasidifférentiation, en plus de l'application du filtre classique. Le modèle obtenu après différentiation ou quasidifférentiation est le modèle à correction d'erreurs. Osborn (1988) et Franses (1994) ont discuté des arguments théoriques sur la présence et la conséquence des racines unitaires dans les modèles périodiques. Boswijk et Franses (1996) ont proposé un test de racine unitaire périodique et ont également analysé les effets de celle-ci à travers une représentation périodique autorégressive.

Toujours dans le cas périodique univarié, la présence de racines unitaires entraîne l'existence de combinaisons linéaires des séries des différentes saisons qui sont stationnaires au sens périodique alors que la série en elle-même ne l'est pas. Cette propriété est appelée cointégration. Franses (1994) et Franses et Paap (2004) ont discuté de la cointégration entre les saisons dans un modèle périodique autorégressif univarié en analysant le processus obtenu en empilant le vecteur des saisons. Boswijk et Franses (1996) et Boswijk et al. (1997) ont fait un travail similaire sans passer par le vecteur de saisons. La technique d'estimation préconisée était les moindres carrés non linéaires.

Il apparaît ici que la cointégration dans les modèles périodiques peut aussi apparaître à la fois dans le cas univarié et multivarié. En d'autres termes, dans

le cas multivarié, il pourrait exister des combinaisons linéaires périodiquement stationnaires entre les saisons des différentes séries et entre les séries pour une saison donnée. Ce qui fait de la cointégration un sujet très important dans la modélisation des modèles périodiques. Pour une saison donnée, le nombre de combinaisons linéaires de séries périodiquement stationnaires est appelé rang de cointégration dans la saison en question. La modélisation de la cointégration se fait habituellement en trois étapes, à savoir la représentation, l'estimation et l'inférence.

Dans les modèles non périodiques, la littérature sur la cointégration met en exergue un certain nombre d'approches de modélisation. De tout ceci, trois approches principales sont connues pour tester la cointégration. Les deux premières sont celles en deux étapes de Engle et Granger (1987) et de Philips et Ouliaris (1990). Ces deux approches ne permettent d'avoir qu'une seule relation de cointégration. Par la suite, nous avons celle de Johansen (1988), qui permet de détecter plusieurs relations de cointégration. L'estimation se fait d'habitude en deux étapes, à savoir l'estimation de plein rang qui ignore le rang de cointégration et l'estimation de rang réduit qui tient compte du rang de cointégration et s'appuie sur un modèle à correction d'erreurs. Ceci est connu plus couramment sous le nom de technique d'estimation de rang réduit.

Le modèle à correction d'erreurs, comme le nom l'indique, corrige le modèle naïf qui ignore la cointégration en faisant apparaître un modèle complet qui tient compte à la fois de l'intégration au sein de chaque série (stationnarité en différence) et de la cointégration entre les séries (stationnarité des combinaisons linéaires). Le rôle du modèle de plein rang est soit de construire le processus stationnaire basé sur la combinaison linéaire des séries (voir Engle et Granger (1987)) ou aussi de fournir des estimateurs initiaux pour l'estimation de rang réduit. Ahn et Reinsel (1990) de même que Yap et Reinsel (1995) ont estimé un modèle à correction d'erreurs à partir d'une procédure itérative dont la valeur initiale était obtenue à partir de l'estimateur de plein rang.

Pour les modèles périodiques multivariés, Kleibergen et Franses (1999) ont étudié la cointégration en imposant une structure dans les retards et en utilisant

la méthode généralisée des moments pour estimer les paramètres sous contrainte de déficience de rang. Boswijk et Franses (1995a) se sont penchés sur une problématique similaire dans le cadre de deux séries en faisant l'hypothèse d'exogénéité de la série multidimensionnelle, la variable dépendante étant considérée comme générée par un processus unidimensionnel. Franses (1995) a utilisé une approche similaire pour un processus bivarié.

Dans ce chapitre, nous procédons à l'estimation et à l'étude des propriétés asymptotiques des paramètres d'un modèle périodique multivarié cointégré à partir de la technique d'estimation de rang réduit. Ce résultat généralise les travaux de Ahn et Reinsel (1990) et Yap et Reinsel (1995) qui ont fait un travail similaire dans le cas non périodique. Nous déterminons par la suite la distribution asymptotique des matrices d'autocovariance résiduelles. Ce résultat permet de construire un test de type portemanteau dont la distribution est approximativement khi-carrée.

Le chapitre est organisé de la manière suivante. Dans la section 4.2, nous rappelons les notions fondamentales de stationnarité, de bruit blanc, de racine unitaire et de vecteur autorégressif au sens périodique. La section 4.3 se penche sur la modélisation de la cointégration dans les modèles périodiques multivariés en proposant la représentation de la cointégration, les modèles de plein rang et de rang réduit de même que l'estimation et l'inférence sur les paramètres obtenus. Ces résultats sont utilisés dans la section 4.4 pour construire un test diagnostique de type portemanteau à partir de la distribution asymptotique des autocovariances résiduelles. La conclusion est présentée dans la section 4.5.

Dans la suite du document, les symboles  $\xrightarrow{\mathcal{L}}$  et  $\xrightarrow{\mathcal{P}}$  signifient respectivement la convergence en distribution et en probabilité. Pour toute suite de valeurs réelles  $\{X_n\}$  et toute suite de termes strictement positifs  $\{a_n\}$ , la notation  $X_n = O(a_n)$  pour tout  $n$  veut dire que la suite  $\{X_n/a_n\}$  est bornée tandis que  $X_n = o(a_n)$  pour tout  $n$  veut dire que la suite  $\{X_n/a_n\}$  converge vers zéro. Pour toute suite de variables aléatoires  $\{Y_n\}$  et toute suite de termes strictement positifs  $b_n$ , la notation  $Y_n = O_p(b_n)$  pour tout  $n$  signifie que  $Y_n/b_n$  est bornée en probabilité tandis que  $Y_n = o_p(b_n)$  pour tout  $n$  veut dire que la suite  $\{Y_n/b_n\}$  converge en

probabilité vers zéro. Pour tout entier naturel  $d > 0$ , la matrice identité d'ordre  $d$  est notée  $\mathbf{I}_d$ . Les symboles  $L$  et  $B$  représentent respectivement l'opérateur retard saisonnier et annuel. De ce fait, pour la série périodique  $\{\mathbf{Y}_t, t = ns + \nu\}$ , où  $s$  est le nombre total de saisons et  $n$  l'année, on aura  $L\mathbf{Y}_t = \mathbf{Y}_{t-1} = \mathbf{Y}_{ns+\nu-1} = \mathbf{Y}_{\nu-1, n}$  et  $B\mathbf{Y}_t = B\mathbf{Y}_{ns+\nu} = \mathbf{Y}_{(n-1)s+\nu} = \mathbf{Y}_{\nu, n-1}$ .

## 4.2. PRÉLIMINAIRES

De même que pour les modèles non périodiques, la modélisation ARMA peut aussi être faite pour les modèles périodiques. Cependant, dans l'étude de la stationnarité, s'il est jugé que le processus est non stationnaire due à une racine unitaire, le filtre appliqué sera typiquement différent d'une saison à une autre. Il semble que l'approche dominante dans la lignée des modèles de Box et Jenkins repose sur l'utilisation des modèles ARMA périodiques (PARMA). Mais pour mieux l'appréhender, il est important de définir l'équivalent des notions fondamentales de la modélisation ARMA dans le cas périodique, à savoir la stationnarité et les bruits blancs.

**Définition 4.1.** *Le processus stochastique périodique  $\{\mathbf{Y}_t, t = sn + \nu, n \in \mathbb{Z}, \nu = 1, \dots, s\}$  de dimension  $d$ ,  $\mathbf{Y}_t = (Y_t(1), \dots, Y_t(d))^\top$  est dit périodiquement stationnaire de période  $s$  si et seulement si pour tout  $t = sn + \nu$ ,*

$$\begin{aligned} E(\mathbf{Y}_t) &= \boldsymbol{\mu}(\nu), \\ E\{(\mathbf{Y}_{t+k} - \boldsymbol{\mu}(\nu + k))(\mathbf{Y}_t - \boldsymbol{\mu}(\nu))^\top\} &= \boldsymbol{\Gamma}_k(\nu), \end{aligned}$$

*existent pour tout  $k \in \mathbb{Z}$  et dépendent seulement de  $k$  et de  $\nu$ .*

Nous avons aussi la définition des bruits blancs périodiques.

**Définition 4.2.** *Le processus vectoriel  $\{\boldsymbol{\varepsilon}_t, t = sn + \nu, n \in \mathbb{Z}, \nu = 1, \dots, s\}$  de dimension  $d$  et de période  $s$ , avec  $\boldsymbol{\varepsilon}_t = (\varepsilon_t(1), \dots, \varepsilon_t(d))^\top$ , est un bruit blanc périodique si les conditions suivantes sont satisfaites :*

- (i)  $E\{\varepsilon_t^2(j)\} < \infty, j = 1, \dots, d, \forall t = sn + \nu,$
- (ii)  $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}, \forall t = sn + \nu,$

$$(iii) \quad E(\boldsymbol{\varepsilon}_{t_1} \boldsymbol{\varepsilon}_{t_2}^\top) = \begin{cases} \boldsymbol{\Sigma}_\nu, & \text{si } t_1 = t_2 = sn + \nu, \\ \mathbf{0}, & \text{sinon.} \end{cases}$$

Nous introduisons dans ce qui suit les modèles vectoriels autorégressifs périodiques. Dans le cas univarié, il est possible de considérer des modèles avec composantes moyennes mobiles que l'on appelle modèles de type PARMA. Conceptuellement, c'est également possible de considérer une telle généralisation dans le cas multivarié. Cependant, même dans le cas non périodique, les modèles VARMA sont sujets à des problèmes d'identifiabilité. Ainsi, dans le but de simplifier les présentations, nous allons nous restreindre dans cette thèse à l'étude des modèles autorégressifs périodiques. D'un point de vue pratique, cette restriction n'est pas sévère compte tenu de l'inflation sérieuse du nombre de paramètres qu'implique les modèles périodiques multivariés.

Le processus  $\{\mathbf{Y}_t, t = sn + \nu, n \in \mathbb{Z}, \nu = 1, \dots, s\}$  de dimension  $d$  et de période  $s$  est un processus autorégressif périodique multivarié (PVAR) s'il est périodiquement stationnaire et vérifie :

$$\mathbf{Y}_{sn+\nu} = \sum_{k=1}^{p(\nu)} \boldsymbol{\phi}_k(\nu) \mathbf{Y}_{sn+\nu-k} + \boldsymbol{\varepsilon}_{sn+\nu}, \quad (4.2.1)$$

où  $\nu = 1, \dots, s$  est la saison,  $s$  le nombre total de saisons et  $p(\nu)$  l'ordre du polynôme autorégressif pour la saison  $\nu$ . Pour  $t = sn + \nu$ , le vecteur  $\mathbf{Y}_t$  est la réalisation du processus dans la saison  $\nu$  de l'année  $n+1$ , où  $n = 1, \dots, N$ ,  $N \in \mathbb{Z}$  étant le nombre total d'années. Les coefficients  $\boldsymbol{\phi}_k(\nu)$  sont des matrices d'ordre  $d \times d$ . Nous supposons que le processus d'erreur  $\{\boldsymbol{\varepsilon}_t, t \in \mathbb{Z}\}$  d'ordre  $d \times 1$  est un bruit blanc périodique de matrice des variances et covariances  $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top) = \boldsymbol{\Sigma}_\nu$ , où  $t = ns + \nu$ .

On peut toujours considérer que  $p(\nu) \equiv p$ ,  $\nu \in \{1, \dots, s\}$ , et de manière précise, l'ordre du processus être supposé égal à  $\max_{\nu=\{1, \dots, s\}} p(\nu)$ , ceci moyennant que certaines matrices de coefficients soient nulles. On parle alors d'un PVAR( $p$ ). Dans le modèle (4.2.1), la moyenne et la tendance ne sont pas prises en compte



car elles sont supposées être au préalable retirées des séries. Nous avons alors que le processus est de moyenne nulle, c'est-à-dire  $E(\mathbf{Y}_{sn+\nu}) = \mathbf{0}$ , pour tout  $t = ns + \nu$ .

Il apparaît clairement que ce processus n'est pas stationnaire au sens de la définition 1.2 car les corrélations et les variances dépendent de la saison. Cependant, il peut être périodiquement stationnaire dans le sens que  $\text{cov}(\mathbf{Y}_{i+n}, \mathbf{Y}_{j+n}) = \text{cov}(\mathbf{Y}_i, \mathbf{Y}_j)$ , pour tout entiers  $i, j$  et pour tout  $n$ . Sous cette condition, le processus est dit périodiquement corrélé, cyclostationnaire ou plus simplement périodiquement stationnaire. Dans un tel cas, une représentation alternative consistant à empiler les observations saisonnières pour chaque année est utilisée. Il s'agit du vecteur des saisons. Cette idée a été évoquée pour la première fois par Gladyshev (1961) et prise en compte par la suite par Tiao et Grupe (1980) et Osborn (1991) dans le cadre des modèles périodiques univariés. Cette représentation est aussi valable dans le cas multivarié. Pour ces modèles, on peut choisir d'empiler par saison ou par série univariée.

Soit  $\mathbf{Y}_{sn+\nu} = \mathbf{Y}_n(\nu) = (Y_{sn+\nu}(1), \dots, Y_{sn+\nu}(d))^\top$  le vecteur de dimension  $d$  représentant les observations de la série  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  pour la saison  $\nu$  de l'année  $n+1$ , et aussi  $\boldsymbol{\varepsilon}_n(\nu) = (\varepsilon_{sn+\nu}(1), \dots, \varepsilon_{sn+\nu}(d))^\top$ . Si nous posons  $\mathbf{Y}_n^* = (\mathbf{Y}_n^\top(1), \dots, \mathbf{Y}_n^\top(s))^\top$ ,  $\boldsymbol{\varepsilon}_n^* = (\boldsymbol{\varepsilon}_n^\top(1), \dots, \boldsymbol{\varepsilon}_n^\top(s))^\top$ , alors le modèle (4.2.1) admet la représentation :

$$\boldsymbol{\Phi}_0 \mathbf{Y}_n^* = \boldsymbol{\Phi}_1 \mathbf{Y}_{n-1}^* + \dots + \boldsymbol{\Phi}_P \mathbf{Y}_{n-P}^* + \boldsymbol{\varepsilon}_n^*. \quad (4.2.2)$$

Dans cette expression,  $P = \lfloor (p-1)/s \rfloor + 1$ , où  $\lfloor \cdot \rfloor$  représente la partie entière. Les coefficients  $\boldsymbol{\Phi}_k$ ,  $k = 1, \dots, P$  sont des matrices d'ordre  $(sd) \times (sd)$  dont les blocs  $\boldsymbol{\Phi}_k[i, j]$ ,  $1 \leq i, j \leq s$  sont des matrices carrées d'ordre  $d$  données par :

$$\boldsymbol{\Phi}_0[i, j] = \begin{cases} \mathbf{I}_d, & \text{si } i = j, \\ \mathbf{0}, & \text{si } j > i, \\ -\boldsymbol{\phi}_{i-j}(i), & \text{si } i < j, \end{cases}$$

$$\boldsymbol{\Phi}_k[i, j] = \boldsymbol{\phi}_{i+sk-j}(i) \text{ pour } k = 1, \dots, P.$$

Dans le modèle (4.2.2), les données ont été empilées par saison. L'un des avantages de cette représentation est le fait qu'elle permet d'étudier la stationnarité

suivant l'approche de Johansen (1988). Ainsi, le modèle (4.2.1) est périodiquement stationnaire de période  $s$  si le processus  $\{\mathbf{Y}_n^*, n \in \mathbb{Z}\}$  défini en (4.2.2) est stationnaire, c'est-à-dire si les racines de l'équation caractéristique :

$$\det(\Phi_0 - \Phi_1 z - \dots - \Phi_P z^P) = 0 \quad (4.2.3)$$

sont à l'extérieur du cercle unité.

L'estimation des modèles périodiques sous l'hypothèse de stationnarité au sens périodique a été faite par Pagano (1978), Li et Hui (1988) ainsi que dans Ursu et Duchesne (2009). Lorsque la condition de stationnarité n'est pas vérifiée, il devient nécessaire de faire intervenir d'autres techniques. L'une des raisons de violation de la stationnarité est la présence de racines unitaires dans le polynôme autorégressif de la représentation (4.2.2).

#### 4.2.1. Racine unitaire et intégration

Nous commençons par la définition de l'intégration dans le contexte des modèles périodiques.

**Définition 4.3.** *Un processus stochastique périodique univarié  $\{Y_t, t \in \mathbb{Z}\}$  de période  $s$  est périodiquement intégré d'ordre un (noté  $PI(1)$ ) s'il peut être rendu périodiquement stationnaire par l'application d'une filtre de différentiation périodique  $(1 - \alpha_\nu L)$ ,  $\nu = 1, \dots, s$ , où les coefficients,  $\alpha_1, \dots, \alpha_s$  satisfont la relation  $\prod_{\nu=1}^s \alpha_\nu = 1$ .*

Si tous les coefficients  $\alpha_\nu$ ,  $\nu = 1, \dots, s$  valent un, on se retrouve dans le cas typique des séries intégrées d'ordre un. La racine périodique englobe aussi la racine saisonnière, ce qui fait de ceci une approche assez générale pour aborder l'intégration. Les travaux sur l'intégration (simple, saisonnière, périodique) des séries univariées portent surtout sur les estimations et les tests. Pour les séries multivariées, la présence de racines unitaires engendre aussi des problèmes de cointégration entre les différentes séries. La notion de cointégration sera mieux comprise ici à partir de la définition périodique autorégressive et de la technique de modélisation qui sera proposée.

### 4.3. MODÉLISATION DE LA COINTÉGRATION DANS LES PVARs

#### 4.3.1. Formulation du modèle à correction d'erreurs

Soit  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  un processus stochastique de dimension  $d$ , où :

$$\mathbf{Y}_t = (Y_t(1), \dots, Y_t(d))^\top,$$

qui est donc un vecteur aléatoire de dimension  $d \times 1$  et  $t = sn + \nu$ .

Considérons le modèle périodique :

$$\mathbf{Y}_{sn+\nu} = \sum_{k=1}^p \phi_k(\nu) \mathbf{Y}_{sn+\nu-k} + \boldsymbol{\varepsilon}_{sn+\nu}, \quad (4.3.1)$$

où  $\phi_k(\nu)$ ,  $\nu = 1, \dots, s$ ,  $k = 1, \dots, p$  sont des matrices d'ordre  $d \times d$  et  $\boldsymbol{\varepsilon}_{sn+\nu} = (\varepsilon_{sn+\nu}(1), \dots, \varepsilon_{sn+\nu}(d))^\top$  est un vecteur de dimension  $d$  avec  $\nu = 1, \dots, s$  et  $n = 1, \dots, N$ . Pour  $t = sn + \nu$ , les notations  $\mathbf{Y}_t$ ,  $\mathbf{Y}_{sn+\nu}$  et  $\mathbf{Y}_{\nu,n}$  représentent la même quantité, à savoir la réalisation du processus  $\mathbf{Y}$  pour la saison  $\nu$  de l'année  $n + 1$ . Le modèle (4.3.1) peut se mettre sous la forme :

$$\phi_\nu(L) \mathbf{Y}_t = \boldsymbol{\varepsilon}_t, \quad (4.3.2)$$

où  $L$  est l'opérateur retard,  $\phi_\nu(L) = \mathbf{I}_d - \phi_1(\nu)L - \dots - \phi_p(\nu)L^p$ ,  $\mathbf{I}_d$  étant la matrice identité d'ordre  $d$ .

Le modèle (4.3.1) admet la représentation de vecteur des saisons ci-dessous telle que définie à l'équation (4.2.2). L'expression (4.2.2) peut aussi être écrite de la manière suivante :

$$\Phi(B) \mathbf{Y}_n^* = \boldsymbol{\varepsilon}_n^*, \quad (4.3.3)$$

où :

$$\begin{aligned} \Phi(B) &= \Phi_0 - \Phi_1 B - \dots - \Phi_P B^P, \\ \mathbf{Y}_n^* &= (\mathbf{Y}_n^\top(1), \dots, \mathbf{Y}_n^\top(s))^\top, \\ \boldsymbol{\varepsilon}_n^* &= (\boldsymbol{\varepsilon}_n^\top(1), \dots, \boldsymbol{\varepsilon}_n^\top(s))^\top. \end{aligned}$$

Le modèle à correction d'erreurs recherché sera mis sous la forme du vecteur des saisons. Pour y arriver, nous ferons les hypothèses suivantes :

**Hypothèse 4.1.** Soit  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  un processus stochastique de dimension  $d$ , où  $\mathbf{Y}_t = (Y_t(1), \dots, Y_t(d))^\top$  qui satisfait l'équation (4.3.2). On suppose que l'équation  $\det\{\Phi(z)\} = 0$  n'admet aucune racine à l'intérieur du cercle unité et qu'il existe des scalaires  $\alpha_j(\nu)$ ,  $\nu = 1, \dots, s$ , vérifiant la condition  $\prod_{\nu=1}^s \alpha_j(\nu) = 1$ ,  $j = 1, \dots, d$ , tel que les processus  $\{Y_t(j) - \alpha_j(\nu)Y_{t-1}(j)\}$ ,  $j = 1, \dots, d$ , sont périodiquement stationnaires.

L'hypothèse 4.1 est analogue aux conditions posées par del Barrio et Osborn (2008, section 2.2) pour définir la cointégration au sens périodique. Sous cette hypothèse, définissons la matrice diagonale d'ordre  $d$  :

$$\Theta(\nu) = \text{diag}(\alpha_1(\nu), \dots, \alpha_d(\nu)),$$

avec  $\Theta(\nu + ks) = \Theta(\nu)$ ,  $\forall k \in \mathbb{N}$ . La prochaine hypothèse indique des conditions sur le terme d'erreur.

**Hypothèse 4.2.** Pour tout  $\nu = 1, \dots, s$ , les vecteurs  $\boldsymbol{\varepsilon}_t$ ,  $t = sn + \nu$ ,  $n = 1, \dots, N$  sont indépendants et distribués de moyenne  $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$ , et de variance  $\text{cov}(\boldsymbol{\varepsilon}_t) = \Sigma_\nu$ . Nous supposons également que :

$$E(|\varepsilon_t(i)\varepsilon_t(j)\varepsilon_t(k)\varepsilon_t(l)|) < \infty, \forall i, j, k, l = 1, \dots, d, \forall t = ns + \nu,$$

et que les valeurs initiales,  $\mathbf{Y}_0, \dots, \mathbf{Y}_{1-p}$  sont toutes nulles.

L'hypothèse 4.1 suppose que composante à composante, le processus est périodiquement intégré d'ordre un. L'hypothèse 4.2 permet d'établir les propriétés asymptotiques des estimateurs de plein rang et de rang réduit. Un exemple de distribution satisfaisant l'hypothèse 4.2 est la loi normale multivariée. Nous avons la proposition suivante, démontrée dans l'annexe A.

**Proposition 4.1.** Le modèle (4.3.1) peut être mis sous la forme :

$$\mathbf{Y}_t - \Theta(\nu)\mathbf{Y}_{t-1} = \boldsymbol{\pi}(\nu)\mathbf{Y}_{t-1} + \sum_{k=1}^{p-1} \boldsymbol{\phi}_k^*(\nu)\{\mathbf{Y}_{t-k} - \Theta(\nu - k)\mathbf{Y}_{t-k-1}\} + \boldsymbol{\varepsilon}_t, \quad t = sn + \nu, \quad (4.3.4)$$

où :

$$\begin{aligned}\phi_k^*(\nu) &= \{\phi_{k+1}^*(\nu) - \phi_{k+1}(\nu)\}\{\Theta(\nu - k)\}^{-1}, \quad k = 1, \dots, p-1, \\ \boldsymbol{\pi}(\nu) &= \phi_1(\nu) - \Theta(\nu) - \phi_1^*(\nu).\end{aligned}$$

La proposition 4.1 est alors une extension du cas univarié au cas multivarié des résultats de Boswijk et Franses (1996, p. 225). Notre attention est portée ici vers les situations où pour tout  $\nu$ , les matrices  $\boldsymbol{\pi}(\nu)$  vérifient  $\det\{\boldsymbol{\pi}(\nu)\} = 0$ . Ainsi, toutes les matrices  $\boldsymbol{\pi}(\nu)$ ,  $\nu = 1, \dots, s$  sont singulières, et nous notons alors :

$$\text{rang}\{\boldsymbol{\pi}(\nu)\} = r_\nu, \quad \nu = 1, \dots, s,$$

avec les inégalités strictes  $r_\nu < d$  pour  $\nu = 1, \dots, s$ .

Cette situation crée une insuffisance de rang dans les matrices  $\boldsymbol{\pi}(\nu)$ ,  $\nu = 1, \dots, s$ . Le fait que  $\text{rang}\{\boldsymbol{\pi}(\nu)\} = r_\nu < d$  entraîne pour chaque saison  $\nu$  l'existence de  $r_\nu$  combinaisons linéaires de ces composantes qui soient périodiquement stationnaires. Ceci nous conduit à la définition de la cointégration au sens périodique.

**Définition 4.4.** *Le processus périodique  $\{\mathbf{Y}_t, t = ns + \nu\}$  d'ordre  $d$  est cointégré au sens périodique s'il existe des matrices  $\mathbf{B}_\nu$  d'ordre  $d \times r_\nu$  tels que pour tout  $\nu$ , le processus  $\{\mathbf{B}_\nu^\top \mathbf{Y}_{st+\nu}\}$  soit périodiquement stationnaire.*

La cointégration périodique est dite pleine s'il y a de la cointégration dans toutes les saisons avec  $\mathbf{B}_{\nu_1} \neq \mathbf{B}_{\nu_2}$  pour au moins deux saisons  $\nu_1$  et  $\nu_2$ . Elle est dite partielle s'il y a de la cointégration dans certaines saisons et pas dans d'autres. Les notions de cointégration périodique pleine et partielle ont été définies par Osborn (2002, Définition 4) et del Barrio et Osborn (2008, Définition 1). Après la définition générale, d'autres particularités sont aussi de très grands intérêts comme celle présentée dans la proposition suivante.

**Proposition 4.2.** *Sous l'hypothèse 4.1, s'il y a cointégration entre les séries dans une saison donnée, alors il y aura cointégration entre les séries dans toutes les autres saisons et le nombre de relations de cointégration encore appelé rang de cointégration sera le même pour chaque saison.*

Ceci a été évoqué par del Barrio et Osborn (2008, Lemme 1), dont une des conséquences stipule qu'il doit avoir le même nombre de relations de cointégration entre des séries  $PI(1)$ . En effet, pour chaque saison  $\nu$ , la matrice,  $\boldsymbol{\pi}(\nu)$  se factorise comme :

$$\boldsymbol{\pi}(\nu) = \boldsymbol{A}_\nu \boldsymbol{B}_\nu^\top,$$

où  $\boldsymbol{A}_\nu$  et  $\boldsymbol{B}_\nu$  sont de dimension  $d \times r_\nu$ ,  $r_\nu$  le rang de la cointégration et  $\boldsymbol{B}_\nu^\top$  le vecteur de cointégration dans la saison  $\nu$ . Par conséquent,  $\boldsymbol{B}_\nu^\top \boldsymbol{Y}_t = \boldsymbol{B}_\nu^\top \boldsymbol{Y}_{sn+\nu} = \boldsymbol{u}_{sn+\nu}$  qui est périodiquement stationnaire. D'autre part,  $\boldsymbol{Y}_{sn+\nu} - \boldsymbol{\Theta}(\nu) \boldsymbol{Y}_{sn+\nu-1} = \boldsymbol{e}_{sn+\nu-1}$ , où  $\boldsymbol{e}_{sn+\nu-1}$  est périodiquement stationnaire.

Ceci entraîne alors que :

$$\boldsymbol{B}_\nu^\top \{ \boldsymbol{\Theta}(\nu) \boldsymbol{Y}_{sn+\nu-1} + \boldsymbol{e}_{sn+\nu-1} \} = \boldsymbol{u}_{sn+\nu},$$

et aussi :

$$\boldsymbol{B}_\nu^\top \boldsymbol{\Theta}(\nu) \boldsymbol{Y}_{sn+\nu-1} = \boldsymbol{u}_{sn+\nu} - \boldsymbol{B}_\nu^\top \boldsymbol{e}_{sn+\nu-1},$$

qui est aussi périodiquement stationnaire. Il en résulte que le vecteur de cointégration dans la saison  $\nu - 1$  est  $\boldsymbol{B}_{\nu-1}^\top = \boldsymbol{B}_\nu^\top \boldsymbol{\Theta}(\nu)$  qui est de même rang que  $\boldsymbol{B}_\nu^\top$  car  $\boldsymbol{\Theta}(\nu)$  est inversible. Dans le cas où  $\alpha_j(\nu) = 1$  pour tout  $\nu = 1, \dots, s$  et pour tout  $j = 1, \dots, d$ , on a  $\boldsymbol{\Theta}(\nu) = \boldsymbol{I}_d$ , pour tout  $\nu$ , et le vecteur de cointégration est le même dans chaque saison, c'est-à-dire que  $\boldsymbol{B}_\nu = \boldsymbol{B}$ , indépendamment de  $\nu$ .

Dans le présent document, nous nous intéressons au cas où  $\alpha_j(\nu) = 1$  pour tout  $\nu = 1, \dots, s$  et pour tout  $j = 1, \dots, d$ . Ceci correspond au cas où toutes les séries sont  $I(1)$ . Cette situation correspond à la présence de cointégration non périodique entre des séries périodiques telle que mentionnée par Osborn (2002, Définition 5) et del Barrio et Osborn (2008, Lemme 1).

**Remarque 1.** Soit  $r$  le rang commun de cointégration. Si toutes les séries sont  $I(1)$  (ou aussi  $PI(1)$ ), elles admettent chacune  $s - 1$  relations de cointégration entre leur saisons. Il y a alors en tout  $d(s-1) + r$  relations de cointégration dans le vecteur des saisons et l'équation  $\det\{\boldsymbol{\Phi}(z)\} = 0$ , où  $\boldsymbol{\Phi}(z)$  est comme dans le modèle (4.3.3), admet une racine unitaire de multiplicité  $m = sd - (d(s-1) + r) = d - r$ .

Les autres racines sont supposées à l'extérieur du cercle unité. Ceci découle de l'hypothèse 4.1.

Dans le cas de deux séries toutes  $I(1)$  (ou aussi  $PI(1)$ ) et de quatre saisons par exemple, si l'équation  $\det\{\Phi(z)\} = 0$  admet une seule racine unitaire de multiplicité un, il y aura par conséquent sept relations de cointégration, dont trois entre les saisons de chaque série et une entre les deux séries dans chaque saison. Ce cas a été discuté par Franses et Paap (2004, p. 108) dans le cas d'un PVAR(1).

Notons qu'il peut aussi avoir de la cointégration pour deux séries différentes et des saisons différentes, par exemple entre la deuxième saison de la première série et la quatrième saison de la troisième série. Les deux types de cointégration déjà expliqués ci-dessus (entre les saisons d'une même série, et entre les séries d'une même saison) permettent de retrouver cette situation qui ne nous intéresse pas dans le cadre de cette étude. Notre attention est en effet portée sur la cointégration entre les séries dans une saison donnée, ce qui est en fait suffisant.

Des représentations utiles font l'objet des propositions 4.3 et 4.4.

**Proposition 4.3.** *Le modèle (4.3.4) admet la représentation de vecteur des saisons suivante :*

$$U_{0,n} = \begin{pmatrix} \pi(1) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \pi(s) \end{pmatrix} \begin{pmatrix} Y_{s,n-1} \\ Y_{1,n} \\ \vdots \\ Y_{s-1,n} \end{pmatrix} + \sum_{\substack{i=1 \\ i=sk+j}}^{p-1} \Gamma_i U_{j:(n-k)} + \begin{pmatrix} \varepsilon_{1,n} \\ \varepsilon_{2,n} \\ \vdots \\ \varepsilon_{s,n} \end{pmatrix}, \quad (4.3.5)$$

où

$$\begin{aligned} \Gamma_i &= \text{diag}(\phi_i^*(1), \dots, \phi_i^*(s)), \\ U_{j:n} &= (\mathbf{e}_{s-j+1,n-1}^\top, \dots, \mathbf{e}_{s,n-1}^\top, \mathbf{e}_{1,n}^\top, \dots, \mathbf{e}_{s-j,n}^\top)^\top, \\ \mathbf{e}_{\nu,n} &= \mathbf{e}_t = Y_t - \Theta(\nu) Y_{t-1}, \\ \pi(\nu) &= \phi_1(\nu) - \Theta(\nu) - \phi_1^*(\nu). \end{aligned}$$

*Preuve :* La preuve du résultat est présentée en Annexe B.

**Proposition 4.4.** *Le modèle (4.3.3) admet aussi la représentation de vecteur des saisons (4.3.5).*

*Preuve :* La démonstration est présentée dans l'Annexe C.

Le but de la proposition 4.4 est de montrer le lien entre le modèle correspondant au vecteur des saisons usuel et le modèle (4.3.5), afin de permettre de mieux saisir l'approche développée ici sur la base des approches classiques, c'est-à-dire en utilisant les résultats dans les modèles multivariés non périodiques. La proposition 4.5 donne une factorisation utile.

**Proposition 4.5.** *L'expression (4.3.5) peut être mise sous la forme :*

$$\Phi^*(B) \{I_{sd} - \Lambda(B)\} Y_n^* = \Pi(B) Y_n^* + \varepsilon_n^*, \quad (4.3.6)$$

où  $I_{sd}$  est la matrice identité d'ordre  $sd$ , et aussi :

$$\Lambda(B) = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \Theta(1)B \\ \Theta(2) & \mathbf{0} & \ddots & \ddots & \vdots \\ \mathbf{0} & \ddots & \ddots & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \Theta(s) & \mathbf{0} \end{pmatrix},$$

$$\Pi(B) = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \pi(1)B \\ \pi(2) & \mathbf{0} & \ddots & \ddots & \vdots \\ \mathbf{0} & \ddots & \ddots & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \pi(s) & \mathbf{0} \end{pmatrix},$$

et  $\Phi^*(B) = \sum_{k=0}^{P^*} \Phi_k^* B^k$  avec  $P^* = \lfloor (p-2)/s \rfloor + 1$ .

Pour tout  $k = 1, \dots, P^*$ ,  $\Phi_k^*$  est une matrice bloc d'ordre  $sd \times sd$ , dont les blocs notés  $\Phi_k^*[i, j]$ ,  $1 \leq i, j \leq s$ , sont des matrices d'ordre  $d \times d$  définies par :

$$\Phi_k^*[i, j] = \begin{cases} -\phi_{sk-j+i}^*(i), & 0 < sk - j + i \leq p - 1, k = 0, \dots, P^*, \\ I_d, & \text{si } k = 0 \text{ et } i = j, \\ \mathbf{0}, & \text{dans les autres cas.} \end{cases}$$



*Preuve* : La démonstration est présentée en Annexe D.

Par exemple, dans le cas où  $s = 4$ ,  $d = 2$  et  $p = 1$ , le modèle PAR(1) s'écrit :

$$\mathbf{Y}_t = \phi_1(\nu)\mathbf{Y}_{t-1} + \varepsilon_t, \quad t = 4n + \nu.$$

Le modèle à correction d'erreurs est donné par :

$$\begin{pmatrix} I_2 & \mathbf{0} & \mathbf{0} & -\Theta(1)B \\ -\Theta(2) & I_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\Theta(3) & I_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\Theta(4) & I_2 \end{pmatrix} \begin{pmatrix} \mathbf{Y}_{1,n} \\ \mathbf{Y}_{2,n} \\ \mathbf{Y}_{3,n} \\ \mathbf{Y}_{4,n} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \pi(1)B \\ \pi(2) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \pi(3) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \pi(4) & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{Y}_{1,n} \\ \mathbf{Y}_{2,n} \\ \mathbf{Y}_{3,n} \\ \mathbf{Y}_{4,n} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,n} \\ \varepsilon_{2,n} \\ \varepsilon_{3,n} \\ \varepsilon_{4,n} \end{pmatrix},$$

où pour tout  $\nu = 1, \dots, 4$ ,

$$\Theta(\nu) = \begin{pmatrix} \alpha_1(\nu) & 0 \\ 0 & \alpha_2(\nu) \end{pmatrix},$$

$$\pi(\nu) = \phi_1(\nu) - \Theta(\nu).$$

Cette représentation fournit le modèle à correction d'erreurs recherché et qui sera utilisée pour les estimations. L'équation (4.3.6) peut être considérée comme une extension des résultats et de l'approche générale de Boswijk et Franses (1996, p. 226) au cas multivarié.

### 4.3.2. Estimation de plein rang

L'estimation de plein rang consiste à estimer le modèle sans tenir compte de la déficience de rang dans les matrices  $\pi(\nu)$ ,  $\nu = 1, \dots, s$ . Elle sera faite dans le cas particulier où  $\alpha_j(\nu) = 1$ ,  $\forall j = 1, \dots, d$  et  $\nu = 1, \dots, s$ . Ceci correspond au cas où toutes les séries sont intégrées d'ordre un. L'idée ici repose sur le fait qu'il peut y avoir de la cointégration entre des processus périodiques intégrés d'ordre un. Ceci a été évoqué par Franses et Paap (2004, p. 109) de même que del Barrio et Osborn (2008, Lemme 1).

En posant dans l'équation (4.3.4) de la proposition 4.1 la relation  $\mathbf{W}_t = \mathbf{Y}_t - \mathbf{Y}_{t-1}$ , on obtient :

$$\mathbf{W}_t = \boldsymbol{\pi}(\nu)\mathbf{Y}_{t-1} + \sum_{k=1}^{p-1} \boldsymbol{\phi}_k^*(\nu)\mathbf{W}_{t-k} + \boldsymbol{\varepsilon}_t, \quad t = sn + \nu. \quad (4.3.7)$$

Posons maintenant,

$$\mathbf{X}_{t-1} = (\mathbf{Y}_{t-1}^\top, \mathbf{W}_{t-1}^\top, \dots, \mathbf{W}_{t-p+1}^\top)^\top,$$

et

$$\mathbf{F}_\nu = (\boldsymbol{\pi}(\nu), \boldsymbol{\phi}_1^*(\nu), \dots, \boldsymbol{\phi}_{p-1}^*(\nu)).$$

Le vecteur  $\mathbf{X}_{t-1}$  est d'ordre  $dp \times 1$  tandis que  $\mathbf{F}_\nu$  est d'ordre  $d \times dp$ . On a ainsi pour tout  $t = sn + \nu$ ,

$$\mathbf{W}_t = \mathbf{F}_\nu \mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t. \quad (4.3.8)$$

L'estimateur des moindres carrés ordinaires de  $\mathbf{F}_\nu$  dans le modèle (4.3.8) est donné pour tout  $\nu$  par :

$$\hat{\mathbf{F}}_\nu = \left( \sum_{n=1}^N \mathbf{W}_{sn+\nu} \mathbf{X}_{sn+(\nu-1)}^\top \right) \left( \sum_{n=1}^N \mathbf{X}_{sn+(\nu-1)} \mathbf{X}_{sn+(\nu-1)}^\top \right)^{-1}, \quad t = sn + \nu. \quad (4.3.9)$$

Dans la suite, nous allons déterminer les propriétés asymptotiques de l'estimateur de plein rang. Dans la proposition suivante, nous montrons que  $\{\mathbf{Y}_t\}$  peut être décomposé comme une somme de deux processus, l'un périodiquement stationnaire et un autre dont les incréments sont périodiquement stationnaires.

**Proposition 4.6.** *Si toutes les composantes du processus  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  sont  $I(1)$  telles que les premières différences soient périodiquement stationnaires, alors il peut se décomposer en deux processus. Un qui est périodiquement stationnaire et un autre dont les incréments sont périodiquement stationnaires.*

Des résultats similaires ont été utilisés par Ahn et Reinsel (1990), Yap et Reinsel (1995) et Pham, Roy et Cédras (2000) pour l'estimation à rang réduit dans le cadre des séries chronologiques non périodiques. Il est bien vrai que le fait de dire que les séries sont  $I(1)$  permet déjà d'avoir des incréments périodiquement stationnaires. Le but est d'exprimer la forme du processus périodiquement

stationnaire obtenu, afin de construire les statistiques de test. Nous prouvons maintenant la proposition 4.6.

*Preuve :* Posons  $m = d - r$ ,  $d$  étant le nombre total de séries,  $r$  le rang de la cointégration dans les saisons (voir Proposition 4.2) avec  $r < d$ . On a  $\pi(\nu) = \mathbf{A}_\nu \mathbf{B}^\top$ ,  $\mathbf{A}_\nu$  et  $\mathbf{B}$  étant des matrices de plein rang de dimension  $d \times r$ . La matrice  $\mathbf{B}^\top \mathbf{B}$  est donc inversible et pour toute matrice  $\mathbf{M}$  de  $d$  lignes, la matrice  $\mathbf{C} = [\mathbf{I}_d - \mathbf{B}(\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top] \mathbf{M}$  vérifie  $\mathbf{B}^\top \mathbf{C} = \mathbf{0}$ . Nous pouvons donc trouver une matrice  $\mathbf{P}_m$  d'ordre  $d \times m$  et de rang  $m$  telle que  $\pi(\nu) \mathbf{P}_m = \mathbf{0}$  pour tout  $\nu = 1, \dots, s$ . Soit alors  $\mathbf{P}_r$  d'ordre  $d \times r$  qui complète  $\mathbf{P}_m$  de manière à former une matrice inversible. Notons  $\begin{bmatrix} \mathbf{Q}_m \\ \mathbf{Q}_r \end{bmatrix}$  l'inverse de  $[\mathbf{P}_m, \mathbf{P}_r]$  tel que  $\mathbf{Q}_m$  soit d'ordre  $m \times d$  et  $\mathbf{Q}_r$  soit d'ordre  $r \times d$ . Nous avons  $\mathbf{Q}_r \mathbf{P}_m = \mathbf{0}$ ,  $\mathbf{Q}_m \mathbf{P}_r = \mathbf{0}$  et  $\mathbf{P}_m \mathbf{Q}_m + \mathbf{P}_r \mathbf{Q}_r = \mathbf{I}_d$ .

Posons :

$$\mathbf{P}_1 = \begin{pmatrix} \mathbf{0} & \mathbf{P}_m & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{P}_m \\ \mathbf{P}_m & \mathbf{0} & \dots & \dots & \mathbf{0} \end{pmatrix},$$

et

$$\mathbf{P}_2 = \begin{pmatrix} \mathbf{0} & \mathbf{P}_r & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{P}_r \\ \mathbf{P}_r & \mathbf{0} & \dots & \dots & \mathbf{0} \end{pmatrix}.$$

La matrice  $\mathbf{P}_1$  est d'ordre  $sd \times sm$  tandis que  $\mathbf{P}_2$  est d'ordre  $sd \times sr$ .

Définissons également :

$$\mathbf{Q}_1 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{Q}_m \\ \mathbf{Q}_m & \mathbf{0} & \ddots & \ddots & \vdots \\ \mathbf{0} & \ddots & \ddots & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{Q}_m & \mathbf{0} \end{pmatrix},$$

et

$$\mathbf{Q}_2 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{Q}_r \\ \mathbf{Q}_r & \mathbf{0} & \ddots & \ddots & \vdots \\ \mathbf{0} & \ddots & \ddots & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{Q}_r & \mathbf{0} \end{pmatrix}.$$

La matrice  $\mathbf{Q}_1$  est d'ordre  $ms \times sd$  et  $\mathbf{Q}_2$  est d'ordre  $rs \times sd$ . Les matrices  $\mathbf{Q}_1$  et  $\mathbf{Q}_2$  vérifient les relations suivantes :

$$\mathbf{P}_1 \mathbf{Q}_1 + \mathbf{P}_2 \mathbf{Q}_2 = \mathbf{I}_{sd},$$

$$\mathbf{Q}_1 \mathbf{P}_2 = \mathbf{0},$$

$$\mathbf{Q}_2 \mathbf{P}_1 = \mathbf{0},$$

$$\Pi(B) \mathbf{P}_1 = \mathbf{0}.$$

Le dernier résultat découle du fait que  $\pi(\nu) \mathbf{P}_m = \mathbf{0}$  pour tout  $\nu = 1, \dots, s$ . La matrice  $\mathbf{P} = [\mathbf{P}_1; \mathbf{P}_2]$  est inversible et son inverse est :

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{bmatrix}.$$

Définissons aussi :

$$\mathbf{\Lambda}_1(B) = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_m \\ \mathbf{I}_m B & \mathbf{0} & \ddots & \ddots & \vdots \\ \mathbf{0} & \ddots & \ddots & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{I}_m & \mathbf{0} \end{pmatrix},$$

et

$$\mathbf{\Lambda}_2(B) = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_r \\ \mathbf{I}_r B & \mathbf{0} & \ddots & \ddots & \vdots \\ \mathbf{0} & \ddots & \ddots & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{I}_r & \mathbf{0} \end{pmatrix}.$$

Les matrices  $\mathbf{\Lambda}_1(B)$  et  $\mathbf{\Lambda}_2(B)$  sont respectivement d'ordre  $sm \times sm$  et  $sr \times sr$ .

Les blocs sont respectivement des matrices d'ordre  $m \times m$  et  $r \times r$ .

D'autre part, on a :

$$\mathbf{P}_1 \mathbf{\Lambda}_1(B) \mathbf{Q}_1 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{P}_m \mathbf{Q}_m B \\ \mathbf{P}_m \mathbf{Q}_m & \mathbf{0} & \mathbf{0} & & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \mathbf{P}_m \mathbf{Q}_m & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{P}_m \mathbf{Q}_m & \mathbf{0} \end{pmatrix},$$

et

$$\mathbf{P}_2 \mathbf{\Lambda}_2(B) \mathbf{Q}_2 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{P}_r \mathbf{Q}_r B \\ \mathbf{P}_r \mathbf{Q}_r & \mathbf{0} & \mathbf{0} & & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \mathbf{P}_r \mathbf{Q}_r & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{P}_r \mathbf{Q}_r & \mathbf{0} \end{pmatrix}.$$

Puisque  $\mathbf{P}_m \mathbf{Q}_m + \mathbf{P}_r \mathbf{Q}_r = \mathbf{I}_d$ , il en ressort que :

$$\mathbf{P}_1 \mathbf{\Lambda}_1(B) \mathbf{Q}_1 + \mathbf{P}_2 \mathbf{\Lambda}_2(B) \mathbf{Q}_2 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_d \mathbf{B} \\ \mathbf{I}_d & \mathbf{0} & \ddots & \ddots & \vdots \\ \mathbf{0} & \ddots & \ddots & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{I}_d & \mathbf{0} \end{pmatrix} = \mathbf{\Lambda}(B),$$

où l'on rappelle que les séries sont toutes  $I(1)$ .

Donc  $\mathbf{I}_{sd} - \mathbf{\Lambda}(B) = \{\mathbf{I}_{sd} - \mathbf{P}_2 \mathbf{\Lambda}_2(B) \mathbf{Q}_2\} \{\mathbf{I}_{sd} - \mathbf{P}_1 \mathbf{\Lambda}_1(B) \mathbf{Q}_1\}$ , rappelant que  $\mathbf{Q}_2 \mathbf{P}_1 = \mathbf{0}$ . D'autre part, d'après l'équation (4.3.6), ainsi que la forme autorégressive (4.3.3),

$$\mathbf{\Phi}(B) = \mathbf{\Phi}^*(B) \{\mathbf{I}_{sd} - \mathbf{\Lambda}(B)\} - \mathbf{\Pi}(B).$$

En utilisant le fait que  $\mathbf{\Pi}(B) \mathbf{P}_1 = \mathbf{0}$ , on obtient alors :

$$\begin{aligned} \mathbf{\Phi}(B) &= \mathbf{\Phi}^*(B) \{\mathbf{I}_{sd} - \mathbf{P}_2 \mathbf{\Lambda}_2(B) \mathbf{Q}_2\} \{\mathbf{I}_{sd} - \mathbf{P}_1 \mathbf{\Lambda}_1(B) \mathbf{Q}_1\} - \mathbf{\Pi}(B), \\ &= \mathbf{\Phi}^*(B) \{\mathbf{I}_{sd} - \mathbf{P}_2 \mathbf{\Lambda}_2(B) \mathbf{Q}_2 - \mathbf{\Pi}(B)\} \{\mathbf{I}_{sd} - \mathbf{P}_1 \mathbf{\Lambda}_1(B) \mathbf{Q}_1\}, \\ &= \mathbf{\Phi}^{**}(B) \{\mathbf{I}_{sd} - \mathbf{P}_1 \mathbf{\Lambda}_1(B) \mathbf{Q}_1\} : \end{aligned}$$

avec  $\mathbf{\Phi}^{**}(B) = \mathbf{\Phi}^*(B) \{\mathbf{I}_{sd} - \mathbf{P}_2 \mathbf{\Lambda}_2(B) \mathbf{Q}_2 - \mathbf{\Pi}(B)\}$ .

Ceci permet alors d'avoir :

$$\det\{\mathbf{\Phi}(z)\} = \det\{\mathbf{\Phi}^{**}(z)\} \det\{\mathbf{I}_{sd} - \mathbf{P}_1 \mathbf{\Lambda}_1(z) \mathbf{Q}_1\}.$$

Or

$$\mathbf{I}_{sd} - \mathbf{P}_1 \mathbf{\Lambda}_1(z) \mathbf{Q}_1 = \mathbf{I}_{sd} - \mathbf{P}' \mathbf{\Omega}(z) \mathbf{P}'^{-1},$$

où :

$$\mathbf{P}' = \begin{pmatrix} \mathbf{0} & \mathbf{P}_m^* & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{P}_m^* \\ \mathbf{P}_m^* & \mathbf{0} & \dots & \dots & \mathbf{0} \end{pmatrix},$$

$$\mathbf{P}'^{-1} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{P}_m^{*-1} \\ \mathbf{P}_m^{*-1} & \mathbf{0} & \ddots & \ddots & \vdots \\ \mathbf{0} & \ddots & \ddots & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{P}_m^{*-1} & \mathbf{0} \end{pmatrix},$$

$$\Omega(z) = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{D}_m \\ \mathbf{D}_m z & \mathbf{0} & \ddots & \ddots & \vdots \\ \mathbf{0} & \ddots & \ddots & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{D}_m & \mathbf{0} \end{pmatrix},$$

avec  $\mathbf{D}_m = \begin{pmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$  qui est d'ordre  $d$ .

La matrice  $\mathbf{P}_m^*$  provient de la décomposition en éléments singuliers de la matrice  $\mathbf{P}_m \mathbf{Q}_m$  qui est d'ordre  $d \times d$  et idempotente de rang  $m$  car  $\mathbf{P}_m \mathbf{Q}_m + \mathbf{P}_r \mathbf{Q}_r = \mathbf{I}_d$ , et donc  $(\mathbf{P}_m \mathbf{Q}_m + \mathbf{P}_r \mathbf{Q}_r) \mathbf{P}_m \mathbf{Q}_m = \mathbf{P}_m \mathbf{Q}_m$  utilisant le fait que  $\mathbf{Q}_r \mathbf{P}_m = \mathbf{0}$ . Plus précisément,  $\mathbf{P}_m \mathbf{Q}_m = \mathbf{P}_m^* \mathbf{D}_m \mathbf{P}_m^{*-1}$  et on note aussi que  $\mathbf{D}_m \mathbf{D}_m = \mathbf{D}_m$ .

Puisque  $\mathbf{I}_{sd} - \Omega(z)$  est une matrice bloc ayant les blocs  $\mathbf{I}_d$  sur la diagonale, des opérations sur les colonnes permettent d'avoir une matrice triangulaire supérieure avec sur la diagonale le terme  $1 - z$  qui apparaît  $m$  fois et le terme  $un$  qui apparaît  $sd - m$  fois. Ceci entraîne que  $\det\{\mathbf{I}_{sd} - \Omega(z)\} = (1 - z)^m$ . De ce fait,  $\det\{\mathbf{I}_{sd} - \mathbf{P}_1 \mathbf{\Lambda}_1(z) \mathbf{Q}_1\} = (1 - z)^m$  et  $\det\{\Phi(z)\} = \det\{\Phi^{**}(z)\} (1 - z)^m$ .

Or, puisque  $un$  représente la seule racine du polynôme  $\det\{\Phi(z)\}$  de multiplicité  $m$  et que toutes les autres racines de l'équation  $\det\{\Phi(z)\} = 0$  sont à l'extérieur du disque unité, le polynôme  $\det\{\Phi^{**}(z)\}$  a donc toutes ses racines à l'extérieur du disque unité et le processus  $[\{\Phi^{**}(B)\}^{-1} \epsilon_n^*, n \in \mathbb{Z}]$  est stationnaire. Le modèle (4.3.6) devient alors :

$$\{\mathbf{I}_{sd} - \mathbf{P}_1 \mathbf{\Lambda}_1(B) \mathbf{Q}_1\} \mathbf{Y}_n^* = \{\Phi^{**}(B)\}^{-1} \epsilon_n^*. \quad (4.3.10)$$

En multipliant (4.3.10) de chaque côté par  $\mathbf{Q}_2$ , on obtient :

$$\mathbf{Q}_2 \mathbf{Y}_n^* = \mathbf{Q}_2 \{\Phi^{**}(B)\}^{-1} \boldsymbol{\varepsilon}_n^* = \mathbf{Y}_{2n}^*,$$

qui est stationnaire.

D'autre part, avec  $\mathbf{P} = [\mathbf{P}_1; \mathbf{P}_2]$  et  $\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{bmatrix}$ , nous avons :

$$\begin{aligned} \mathbf{Q}\Phi(B) &= \mathbf{Q}\Phi^{**}(B)(\mathbf{I}_{sd} - \mathbf{P}_1\Lambda_1(B)\mathbf{Q}_1) = \{\mathbf{Q}\Phi^{**}(B)\mathbf{P}\}\mathbf{Q}(\mathbf{I}_{sd} - \mathbf{P}_1\Lambda_1(B)\mathbf{Q}_1), \\ &= \{\mathbf{Q}\Phi^{**}(B)\mathbf{P}\} \begin{bmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{bmatrix} (\mathbf{I}_{sd} - \mathbf{P}_1\Lambda_1(B)\mathbf{Q}_1), \\ &= \{\mathbf{Q}\Phi^{**}(B)\mathbf{P}\} \begin{bmatrix} \mathbf{Q}_1 - \Lambda_1(B)\mathbf{Q}_1 \\ \mathbf{Q}_2 \end{bmatrix}. \end{aligned}$$

Il en ressort que :

$$\begin{bmatrix} \mathbf{Q}_1 \mathbf{Y}_n^* - \Lambda_1(B)\mathbf{Q}_1 \mathbf{Y}_n^* \\ \mathbf{Q}_2 \mathbf{Y}_n^* \end{bmatrix} = \{\mathbf{Q}\Phi^{**}(B)\mathbf{P}\}^{-1} \mathbf{Q}\boldsymbol{\varepsilon}_n^*,$$

qui est stationnaire.

Ainsi,  $\mathbf{Q}_1 \mathbf{Y}_n^* - \Lambda_1(B)\mathbf{Q}_1 \mathbf{Y}_n^*$  est également stationnaire. De manière plus précise,

$$\mathbf{Q}_1 \mathbf{Y}_n^* = \begin{pmatrix} \mathbf{Q}_m \mathbf{Y}_{s,n} - \mathbf{Q}_m \mathbf{Y}_{s-1,n} \\ \mathbf{Q}_m \mathbf{Y}_{1,n} - \mathbf{Q}_m \mathbf{Y}_{s,n-1} \\ \vdots \\ \mathbf{Q}_m \mathbf{Y}_{s-1,n} - \mathbf{Q}_m \mathbf{Y}_{s-2,n} \end{pmatrix}$$

et

$$\mathbf{Q}_2 \mathbf{Y}_n^* = \begin{pmatrix} \mathbf{Q}_r \mathbf{Y}_{s,n} \\ \mathbf{Q}_r \mathbf{Y}_{1,n} \\ \vdots \\ \mathbf{Q}_r \mathbf{Y}_{s-1,n} \end{pmatrix}.$$

Ceci constitue le résultat cherché car le processus du vecteur des saisons des processus  $\{\mathbf{Q}_m \mathbf{Y}_t - \mathbf{Q}_m \mathbf{Y}_{t-1}, t = ns + \nu\}$  et  $\{\mathbf{Q}_r \mathbf{Y}_t, t = ns + \nu\}$  étant stationnaire, les processus en question sont périodiquement stationnaires en général.



Ceci complète la preuve de la proposition 4.6. En effet,  $\mathbf{Y}_t$  s'écrivant comme  $(\mathbf{P}_1\mathbf{Q}_1 + \mathbf{P}_2\mathbf{Q}_2)\mathbf{Y}_n^* = \mathbf{P}_1\mathbf{Q}_1\mathbf{Y}_n^* + \mathbf{P}_2\mathbf{Q}_2\mathbf{Y}_n^*$ , nous avons établi la décomposition recherchée.

Nous établissons maintenant les propriétés de l'estimateur des moindres carrés (4.3.9). À cette fin, posons  $\mathbf{P}_a = [\mathbf{P}_m, \mathbf{P}_r]$  et  $\mathbf{Q}_a = [\mathbf{Q}_m^\top, \mathbf{Q}_r^\top]^\top$  où  $\mathbf{P}_m, \mathbf{P}_r, \mathbf{Q}_m$  et  $\mathbf{Q}_r$  sont définis dans la section précédente. Définissons :

$$\begin{aligned} \mathbf{Q}_a^* &= \text{diag}(\mathbf{Q}_a, \mathbf{I}_{d(p-1)}), \\ \mathbf{P}_a^* &= \text{diag}(\mathbf{P}_a, \mathbf{I}_{d(p-1)}), \\ \mathbf{D}_a &= \text{diag}(N\mathbf{I}_m, N^{1/2}\mathbf{I}_r), \\ \mathbf{D}_a^* &= \text{diag}(\mathbf{D}_a, N^{1/2}\mathbf{I}_{m(p-1)}). \end{aligned} \quad (4.3.11)$$

Posons également :

$$\begin{aligned} \mathbf{Z}_t &= \mathbf{Q}_a\mathbf{Y}_t = [\mathbf{Z}_{m:t}^\top, \mathbf{Z}_{r:t}^\top]^\top, \\ \mathbf{a}_t &= \mathbf{Q}_a\boldsymbol{\varepsilon}_t. \end{aligned} \quad (4.3.12)$$

On a :

$$\mathbf{X}_t^* = \mathbf{Q}_a^*\mathbf{X}_t = (\mathbf{Z}_{m:t}^\top, \mathbf{Z}_{r:t}^\top, \mathbf{W}_t^\top, \dots, \mathbf{W}_{t-p+2}^\top)^\top = (\mathbf{Z}_{m:t}^\top, \mathbf{U}_t^\top)^\top,$$

où  $\mathbf{U}_t = (\mathbf{Z}_{r:t}^\top, \mathbf{W}_t^\top, \dots, \mathbf{W}_{t-p+2}^\top)^\top$  est d'ordre  $(r + d(p-1)) \times 1$ . De plus,

$$\begin{aligned} \mathbf{Q}_a(\hat{\mathbf{F}}_\nu - \mathbf{F}_\nu^1)\mathbf{P}_a^* &= \mathbf{Q}_a \left[ \left( \sum_{n=1}^N \boldsymbol{\varepsilon}_{sn+\nu} \mathbf{X}_{sn+\nu-1}^\top \right) \left( \sum_{n=1}^N \mathbf{X}_{sn+\nu-1} \mathbf{X}_{sn+\nu-1}^\top \right)^{-1} \right] \mathbf{P}_a^*, \\ &= \left( \sum_{n=1}^N \mathbf{Q}_a \boldsymbol{\varepsilon}_{sn+\nu} \mathbf{X}_{sn+\nu-1}^\top \mathbf{Q}_a^* \right) \left( \sum_{n=1}^N \mathbf{Q}_a^* \mathbf{X}_{sn+\nu-1} \mathbf{X}_{sn+\nu-1}^\top \mathbf{Q}_a^* \right)^{-1}, \\ &= \left[ \left( \sum_{n=1}^N \mathbf{a}_{sn+\nu} \mathbf{X}_{sn+\nu-1}^* \right) \left( \sum_{n=1}^N \mathbf{X}_{sn+\nu-1}^* \mathbf{X}_{sn+\nu-1}^{*\top} \right)^{-1} \right]. \end{aligned}$$

Posons :

$$\mathbf{V}_\nu = \sum_{n=1}^N \mathbf{X}_{sn+\nu-1}^* \mathbf{X}_{sn+\nu-1}^{*\top} \quad (4.3.13)$$

et

$$\mathbf{R}_\nu = \sum_{n=1}^N \mathbf{a}_{sn+\nu} \mathbf{X}_{sn+\nu-1}^{*\top}. \quad (4.3.14)$$

On obtient alors :

$$\mathbf{X}_{sn+\nu-1}^* \mathbf{X}_{sn+\nu-1}^{*\top} = \begin{bmatrix} \mathbf{Z}_{m:(sn+\nu-1)} \mathbf{Z}_{m:(sn+\nu-1)}^\top & \mathbf{Z}_{m:(sn+\nu-1)} \mathbf{U}_{sn+\nu-1}^\top \\ \mathbf{U}_{sn+\nu-1} \mathbf{Z}_{m:(sn+\nu-1)}^\top & \mathbf{U}_{sn+\nu-1} \mathbf{U}_{sn+\nu-1}^\top \end{bmatrix},$$

ce qui entraîne que :

$$\mathbf{V}_\nu = \begin{bmatrix} \sum_{n=1}^N \mathbf{Z}_{m:(sn+\nu-1)} \mathbf{Z}_{m:(sn+\nu-1)}^\top & \sum_{n=1}^N \mathbf{Z}_{m:(sn+\nu-1)} \mathbf{U}_{sn+\nu-1}^\top \\ \sum_{n=1}^N \mathbf{U}_{sn+\nu-1} \mathbf{Z}_{m:(sn+\nu-1)}^\top & \sum_{n=1}^N \mathbf{U}_{sn+\nu-1} \mathbf{U}_{sn+\nu-1}^\top \end{bmatrix}. \quad (4.3.15)$$

On a également :

$$\mathbf{R}_\nu = \begin{bmatrix} \sum_{n=1}^N \mathbf{a}_{sn+\nu} \mathbf{Z}_{m:(sn+\nu-1)}^\top & \sum_{n=1}^N \mathbf{a}_{sn+\nu} \mathbf{U}_{sn+\nu-1}^\top \end{bmatrix}, \quad (4.3.16)$$

ce qui permet d'avoir :

$$\mathbf{Q}_a (\hat{\mathbf{F}}_\nu - \mathbf{F}_\nu) \mathbf{P}_a^* = \mathbf{R}_\nu \mathbf{V}_\nu^{-1}.$$

D'autre part, la décomposition nous permet d'avoir :

$$\begin{pmatrix} \mathbf{Q}_m \mathbf{Y}_{s,n} - \mathbf{Q}_m \mathbf{Y}_{s-1,n} \\ \mathbf{Q}_m \mathbf{Y}_{1,n} - \mathbf{Q}_m \mathbf{Y}_{s,n-1} \\ \mathbf{Q}_m \mathbf{Y}_{2,n} - \mathbf{Q}_m \mathbf{Y}_{1,n} \\ \vdots \\ \mathbf{Q}_m \mathbf{Y}_{s-1,n} - \mathbf{Q}_m \mathbf{Y}_{s-2,n} \end{pmatrix} = \mathbf{D}_{sm} \{ \mathbf{Q} \Phi^{**}(B) \mathbf{P} \}^{-1} \mathbf{A}_n,$$

avec  $\mathbf{A}_n = \mathbf{Q} \boldsymbol{\varepsilon}_n^*$ ,  $\mathbf{D}_{sm} = [\mathbf{I}_{sm}; \mathbf{0}]$ , d'ordre  $sd \times sd$ , et rappelons que  $\mathbf{I}_{sm}$  est la matrice identité d'ordre  $sm$ . La formule ci-dessus équivaut encore à :

$$\begin{pmatrix} \mathbf{Z}_{m:(sn+s)} - \mathbf{Z}_{m:(sn+s-1)} \\ \mathbf{Z}_{m:(sn+1)} - \mathbf{Z}_{m:((n-1)s+s)} \\ \mathbf{Z}_{m:(sn+2)} - \mathbf{Z}_{m:(sn+1)} \\ \vdots \\ \mathbf{Z}_{m:(sn+(s-1))} - \mathbf{Z}_{m:(sn+(s-1)-1)} \end{pmatrix} = \mathbf{D}_{sm} \{ \mathbf{Q} \Phi^{**}(B) \mathbf{P} \}^{-1} \mathbf{A}_n.$$

En écrivant  $D_{sm}\{\mathbf{Q}\Phi^{**}(B)\mathbf{P}\}^{-1} = \begin{pmatrix} \mathbf{L}_1(B) \\ \vdots \\ \mathbf{L}_\nu(B) \\ \vdots \\ \mathbf{L}_s(B) \end{pmatrix}$ , où les  $\mathbf{L}_\nu(\cdot)$ , sont tous d'ordre

$m \times sd$ , on obtient :

$$\begin{pmatrix} \mathbf{Z}_{m:(sn+s)} - \mathbf{Z}_{m:(sn+s-1)} \\ \mathbf{Z}_{m:(sn+1)} - \mathbf{Z}_{m:((n-1)s+s)} \\ \mathbf{Z}_{m:(sn+2)} - \mathbf{Z}_{m:(sn+1)} \\ \vdots \\ \mathbf{Z}_{m:(sn+(s-1))} - \mathbf{Z}_{m:(sn+(s-1)-1)} \end{pmatrix} = \begin{pmatrix} \mathbf{L}_1(B)\mathbf{A}_n \\ \vdots \\ \mathbf{L}_s(B)\mathbf{A}_n \end{pmatrix}.$$

Ce qui permet d'avoir pour tout  $\nu = 1, \dots, s$ , en additionnant les composantes des deux vecteurs ci-dessus en commençant par la valeur de l'indice  $sn + \nu$  :

$$\begin{aligned} \mathbf{Z}_{m:(sn+\nu)} - \mathbf{Z}_{m:((n-1)s+\nu)} &= \left\{ \sum_{k=s-\nu+1}^s \mathbf{L}_k(B) \right\} \mathbf{A}_n + \left\{ \sum_{k=1}^{s-\nu} \mathbf{L}_k(B) \right\} \mathbf{A}_{n-1}, \\ &= \left\{ \sum_{k=s-\nu+1}^s \mathbf{L}_k(B) + \sum_{k=1}^{s-\nu} \mathbf{L}_k(B)B \right\} \mathbf{A}_n, \\ &= \sum_{k=0}^{\infty} \boldsymbol{\psi}_{\nu,k} \mathbf{A}_{n-k}, \\ &= \left( \sum_{k=0}^{\infty} \boldsymbol{\psi}_{\nu,k} B^k \right) \mathbf{A}_n, \\ &= \boldsymbol{\psi}_\nu(B) \mathbf{A}_n. \end{aligned}$$

Ceci entraîne alors que :

$$\mathbf{Z}_{m:(sn+\nu)} = \sum_{k=0}^n \boldsymbol{\psi}_\nu(B) \mathbf{A}_k, \quad (4.3.17)$$

car  $\mathbf{Z}_{m:(-s+\nu)} = 0$  pour tout  $\nu$ .

**Remarque 2.** Pour tout  $\nu$ ,  $\sum_{k=0}^{\infty} k |\psi_{\nu,k}^{(ij)}| < \infty$  où  $\psi_{\nu,k}^{(ij)}$ ,  $1 \leq i \leq m$  et  $1 \leq j \leq sd$  sont les éléments de la matrices  $\boldsymbol{\psi}_{\nu,k}$ . Ces derniers étant les coefficients des polynômes  $\boldsymbol{\psi}_\nu$ ,  $\boldsymbol{\psi}_\nu(B) = \sum_{k=0}^{\infty} \boldsymbol{\psi}_{\nu,k} B^k$ . Cette propriété provient du fait que toutes

les racines du polynôme  $\{\mathbf{Q}\Phi^{**}(B)\mathbf{P}\}$  sont à l'extérieur du cercle unité et qu'il est de degré fini. Cela entraîne que les coefficients du polynôme  $\{\mathbf{Q}\Phi^{**}(B)\mathbf{P}\}^{-1}$  décroissent exponentiellement vers zéro (Hannan (1970, p. 11)). Cette propriété a été aussi évoquée par Ahn et Reinsel (1990, p. 815).

En constatant d'autre part que  $\mathbf{A}_n = \begin{pmatrix} \mathbf{Q}_1 \boldsymbol{\varepsilon}_n^* \\ \mathbf{Q}_2 \boldsymbol{\varepsilon}_n^* \end{pmatrix}$  où,

$$\mathbf{Q}_1 \boldsymbol{\varepsilon}_n^* = \begin{pmatrix} \mathbf{Q}_m \boldsymbol{\varepsilon}_{s,n} \\ \mathbf{Q}_m \boldsymbol{\varepsilon}_{1,n} \\ \vdots \\ \mathbf{Q}_m \boldsymbol{\varepsilon}_{s-1,n} \end{pmatrix},$$

et,

$$\mathbf{Q}_2 \boldsymbol{\varepsilon}_n^* = \begin{pmatrix} \mathbf{Q}_r \boldsymbol{\varepsilon}_{s,n} \\ \mathbf{Q}_r \boldsymbol{\varepsilon}_{1,n} \\ \vdots \\ \mathbf{Q}_r \boldsymbol{\varepsilon}_{s-1,n} \end{pmatrix},$$

nous avons  $\mathbf{a}_t = \begin{pmatrix} \mathbf{Q}_m \\ \mathbf{Q}_r \end{pmatrix} \boldsymbol{\varepsilon}_t = \mathbf{D}_\nu \mathbf{A}_n$ . La matrice  $\mathbf{D}_\nu$  permet d'extraire le sous-vecteur  $\mathbf{a}_t$  du vecteur  $\mathbf{A}_n$  de plus grande dimension. Dans le lemme suivant, les résultats asymptotiques sont exposés qui seront utiles dans l'étude de l'estimateur des moindres carrés (4.3.9).

**Lemme 4.1.** *Sous l'hypothèse 4.2 et la Remarque 2, nous avons les résultats suivants :*

$$\begin{aligned}
(i) \quad & N^{-2} \sum_{n=1}^N \mathbf{Z}_{m:(t-1)} \mathbf{Z}_{m:(t-1)}^\top \xrightarrow{\mathcal{L}} \boldsymbol{\psi}_{\nu-1}(1) \mathbf{Q} \boldsymbol{\Sigma}_\varepsilon^{1/2} \left\{ \int_0^1 \boldsymbol{w}(r) \boldsymbol{w}^\top(r) dr \right\} \boldsymbol{\Sigma}_\varepsilon^{1/2} \{\boldsymbol{\psi}_{\nu-1}(1) \mathbf{Q}\}^\top, \\
(ii) \quad & N^{-1} \sum_{n=1}^N \mathbf{a}_t \mathbf{Z}_{m:(t-1)}^\top \xrightarrow{\mathcal{L}} (\mathbf{D}_\nu \mathbf{Q}) \boldsymbol{\Sigma}_\varepsilon^{1/2} \left\{ \int_0^1 \boldsymbol{w}(r) d\boldsymbol{w}^\top(r) \right\}^\top \boldsymbol{\Sigma}_\varepsilon^{1/2} \{\boldsymbol{\psi}_{\nu-1}(1) \mathbf{Q}\}^\top \\
& + (\mathbf{D}_\nu \mathbf{Q}) \boldsymbol{\Sigma}_\varepsilon (\boldsymbol{\psi}_{\nu-1,0} \mathbf{Q})^\top, \\
(iii) \quad & N^{-3/2} \sum_{n=1}^N \mathbf{Z}_{m:(t-1)} \mathbf{U}_{t-1}^\top \xrightarrow{\mathcal{P}} \mathbf{0}, \\
(iv) \quad & N^{-1} \sum_{n=1}^N \mathbf{U}_{t-1} \mathbf{U}_{t-1}^\top \xrightarrow{\mathcal{P}} \boldsymbol{\Gamma}(\mathbf{U}_\nu) = \mathbf{D}_\nu^{**} \boldsymbol{\Gamma}_U^* (\mathbf{D}_\nu^{**})^\top, \\
(v) \quad & N^{-1/2} \sum_{n=1}^N \text{vec}(\mathbf{a}_t \mathbf{U}_{t-1}^\top) \xrightarrow{\mathcal{L}} \mathcal{N}_{dr+d^2(p-1)}(\mathbf{0}, \mathbf{D}_{\nu:2} \boldsymbol{\Gamma}(\mathbf{U}_{\nu:n}^*) \mathbf{D}_{\nu:2}^\top \otimes (\mathbf{D}_{\nu:1} \mathbf{Q}) \boldsymbol{\Sigma}^* (\mathbf{D}_{\nu:1} \mathbf{Q})^\top), \\
& \text{avec :}
\end{aligned}$$

$$\begin{aligned}
t &= ns + \nu, \\
\boldsymbol{\Gamma}_U^* &= E(\mathbf{U}_n^* \mathbf{U}_n^{*\top}), \\
\mathbf{U}_n^* &= (\mathbf{U}_{sn+1}^\top, \mathbf{U}_{sn+2}^\top, \dots, \mathbf{U}_{sn+s}^\top)^\top, \\
\boldsymbol{\Gamma}(\mathbf{U}_{\nu:n}^*) &= E(\mathbf{U}_{\nu:n}^* \mathbf{U}_{\nu:n}^{*\top}), \\
\mathbf{U}_{\nu:n}^* &= (\mathbf{U}_{sn+\nu}^\top, \dots, \mathbf{U}_{sn+s}^\top, \mathbf{U}_{s(n+1)+1}^\top, \dots, \mathbf{U}_{s(n+1)+\nu-1}^\top)^\top, \\
\mathbf{U}_t &= (\mathbf{Z}_{r:t}^\top, \mathbf{W}_t^\top, \dots, \mathbf{W}_{t-p+2}^\top)^\top, \\
\boldsymbol{\Sigma}^* &= E(\boldsymbol{\varepsilon}_{\nu:n}^* \boldsymbol{\varepsilon}_{\nu:n}^{*\top}), \\
\boldsymbol{\varepsilon}_{\nu:n}^* &= (\boldsymbol{\varepsilon}_{sn+\nu}^\top, \dots, \boldsymbol{\varepsilon}_{sn+s}^\top, \boldsymbol{\varepsilon}_{s(n+1)+1}^\top, \dots, \boldsymbol{\varepsilon}_{s(n+1)+\nu-1}^\top)^\top, \\
\boldsymbol{\Sigma}_\varepsilon &= E(\boldsymbol{\varepsilon}_n^* \boldsymbol{\varepsilon}_n^{*\top}) = \text{diag}(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_s),
\end{aligned}$$

où  $\boldsymbol{W}(r)$  un mouvement brownien standard de dimension  $sd$ . Le vecteur  $\mathbf{Z}_{r:t}$  est défini à l'équation (4.3.12) tandis que  $\mathbf{W}_t$  est défini dans l'équation (4.3.7). Les matrices  $\mathbf{D}_\nu$ ,  $\mathbf{D}_\nu^{**}$ ,  $\mathbf{D}_{\nu:1}$  et  $\mathbf{D}_{\nu:2}$  sont telles que :

$$\begin{aligned}
\mathbf{a}_t &= \mathbf{D}_\nu \mathbf{A}_n, \\
\boldsymbol{\varepsilon}_{ns+\nu} &= \mathbf{D}_{\nu:1} \boldsymbol{\varepsilon}_{\nu:n}^*, \\
\mathbf{U}_{sn+\nu} &= \mathbf{D}_\nu^{**} \mathbf{U}_n^*, \\
\mathbf{U}_{sn+\nu} &= \mathbf{D}_{\nu:2} \mathbf{U}_{\nu:n}^*.
\end{aligned}$$

*Preuve* : Elle est présentée en Annexe E.

Nous avons d'autre part :

$$\mathbf{Q}_a(\hat{\mathbf{F}}_\nu - \mathbf{F}_\nu)\mathbf{P}_a^*\mathbf{D}_a^* = \mathbf{R}_\nu\mathbf{D}_a^{*-1}(\mathbf{D}_a^{*-1}\mathbf{V}_\nu\mathbf{D}_a^{*-1})^{-1}, \quad (4.3.18)$$

où les matrices  $\mathbf{P}_a^*$  et  $\mathbf{D}_a^*$  sont données dans l'équation (4.3.11). Le résultat de l'équation (4.3.18) provient des formules (4.3.13), (4.3.14), (4.3.15) et (4.3.16). Or d'après le Lemme 4.1,

$$\mathbf{D}_a^{*-1}\mathbf{V}_\nu\mathbf{D}_a^{*-1} = \text{diag} \left( N^{-2} \sum_{n=1}^N \mathbf{Z}_{m:(t-1)} \mathbf{Z}_{m:(t-1)}^\top, N^{-1} \sum_{n=1}^N \mathbf{U}_{t-1} \mathbf{U}_{t-1}^\top \right),$$

où  $t = ns + \nu$ . Ceci permet d'avoir,

$$(\mathbf{D}_a^{*-1}\mathbf{V}_\nu\mathbf{D}_a^{*-1})^{-1} = \text{diag} \left( \left\{ N^{-2} \sum_{n=1}^N \mathbf{Z}_{m:(t-1)} \mathbf{Z}_{m:(t-1)}^\top \right\}^{-1}, \left\{ N^{-1} \sum_{n=1}^N \mathbf{U}_{t-1} \mathbf{U}_{t-1}^\top \right\}^{-1} \right),$$

où  $t = ns + \nu$ . D'autre part, on a aussi,

$$\mathbf{R}_\nu\mathbf{D}_a^{*-1} = \left[ N^{-1} \sum_{n=1}^N \mathbf{a}_t \mathbf{Z}_{m:(t-1)}^\top, N^{-1/2} \sum_{n=1}^N \mathbf{a}_t \mathbf{U}_{t-1}^\top \right].$$

Par conséquent, on a pour  $t = sn + \nu$  :

$$\mathbf{Q}_a(\hat{\mathbf{F}}_\nu - \mathbf{F}_\nu)\mathbf{P}_a^*\mathbf{D}_a^* = [\mathbf{H}_{1\nu}, \mathbf{H}_{2\nu}], \quad (4.3.19)$$

où

$$\begin{aligned} \mathbf{H}_{1\nu} &= \left( N^{-1} \sum_{n=1}^N \mathbf{a}_t \mathbf{Z}_{m:(t-1)}^\top \right) \left( N^{-2} \sum_{n=1}^N \mathbf{Z}_{m:(t-1)} \mathbf{Z}_{m:(t-1)}^\top \right)^{-1}, \\ \mathbf{H}_{2\nu} &= \left( N^{-1/2} \sum_{n=1}^N \mathbf{a}_t \mathbf{U}_{t-1}^\top \right) \left( N^{-1} \sum_{n=1}^N \mathbf{U}_{t-1} \mathbf{U}_{t-1}^\top \right)^{-1}. \end{aligned}$$

Le résultat principal de cette section est exposé, qui énonce le comportement asymptotique de l'estimateur des moindres carrés.

**Proposition 4.7.** *Soit  $\hat{\mathbf{F}}_\nu$  l'estimateur des moindres carrés ordinaires de  $\mathbf{F}_\nu$  tel que défini par l'équation (4.3.9) pour  $\nu = 1, \dots, s$ , alors :*

$$\mathbf{Q}_a(\hat{\mathbf{F}}_\nu - \mathbf{F}_\nu)\mathbf{P}_a^*\mathbf{D}_a^* \xrightarrow{\mathcal{L}} [\mathbf{M}_\nu, \mathbf{N}_\nu],$$

où  $\mathbf{M}_\nu = \mathbf{P}_{1:\nu} \mathbf{P}_{2:\nu}^{-1}$ , avec :

$$\begin{aligned} \mathbf{P}_{1:\nu} &= (\mathbf{D}_\nu \mathbf{Q}) \Sigma_\varepsilon^{1/2} \left\{ \int_0^1 \mathbf{W}(r) d\mathbf{W}^\top(r) \right\}^\top \Sigma_\varepsilon^{1/2} \{\psi_{\nu-1}(1) \mathbf{Q}\}^\top \\ &\quad + (\mathbf{D}_\nu \mathbf{Q}) \Sigma_\varepsilon (\psi_{\nu-1,0} \mathbf{Q})^\top, \\ \mathbf{P}_{2:\nu} &= \psi_{\nu-1}(1) \mathbf{Q} \Sigma_\varepsilon^{1/2} \left\{ \int_0^1 \mathbf{W}(r) \mathbf{W}^\top(r) dr \right\} \Sigma_\varepsilon^{1/2} \{\psi_{\nu-1}(1) \mathbf{Q}\}^\top, \end{aligned}$$

$\mathbf{W}(r)$  est un mouvement brownien standard de dimension  $sd$ ,  $\psi_{\nu-1,0}$  est le coefficient du polynôme  $\psi_{\nu-1}(z) = \sum_{k=0}^{\infty} \psi_{\nu-1,k} z^k$  correspondant à  $k=0$ ,  $\psi_{\nu-1}(1) = \sum_{k=0}^{\infty} \psi_{\nu-1,k}$ . Le polynôme  $\psi_{\nu-1}(\cdot)$  est défini à l'équation (4.3.17). Le vecteur  $\mathbf{N}_\nu$  est tel que  $\text{vec}(\mathbf{N}_\nu)$  soit distribué comme une loi normale :

$$\mathcal{N}_{rd+d^2(p-1)}(\mathbf{0}, \{\Gamma(\mathbf{U}_\nu)\}^{-1} \mathbf{D}_{\nu:2} \Gamma(\mathbf{U}_{\nu:n}^*) \mathbf{D}_{\nu:2}^\top \{\Gamma(\mathbf{U}_\nu)\}^{-1} \otimes (\mathbf{D}_{\nu:1} \mathbf{Q}) \Sigma^* (\mathbf{D}_{\nu:1} \mathbf{Q})^\top).$$

*Preuve* : Elle découle du Lemme 4.1.

### 4.3.3. Estimation de rang réduit

Dans cette section, la déficience de rang dans les matrices  $\boldsymbol{\pi}(\nu)$ ,  $\nu = 1, \dots, s$  est prise en compte. Rappelons que pour chaque saison, le modèle s'écrit :

$$\mathbf{W}_t = \mathbf{A}_\nu \mathbf{B}^\top \mathbf{Y}_{t-1} + \sum_{k=1}^{p-1} \phi_k^*(\nu) \mathbf{W}_{t-k} + \boldsymbol{\varepsilon}_t, \quad t = sn + \nu, \quad \mathbf{W}_t = \mathbf{Y}_t - \mathbf{Y}_{t-1}. \quad (4.3.20)$$

Ici,  $\boldsymbol{\pi}(\nu) = \mathbf{A}_\nu \mathbf{B}^\top$ ,  $\mathbf{A}_\nu$  et  $\mathbf{B}$  étant des matrices de plein rang de dimension  $d \times r$ ,  $r$  est le rang de la cointégration dans les saisons,  $m = d - r$ ,  $d$  étant le nombre total de séries,  $r < d$ . Ce modèle admet la représentation de vecteur des saisons de la formule (4.3.6) :

$$\Phi^*(B) \{I_{sd} - \Lambda(B)\} \mathbf{Y}_n^* = \Pi(B) \mathbf{Y}_n^* + \boldsymbol{\varepsilon}_n^*,$$

où

$$\boldsymbol{\varepsilon}_n^* = (\boldsymbol{\varepsilon}_{1,n}^\top, \dots, \boldsymbol{\varepsilon}_{s,n}^\top)^\top.$$

Nous avons pour chaque saison :

$$\boldsymbol{\varepsilon}_t = \boldsymbol{\varepsilon}_{\nu,n} = \mathbf{W}_t - \mathbf{A}_\nu \mathbf{B}^\top \mathbf{Y}_{t-1} - \sum_{k=1}^{p-1} \boldsymbol{\phi}_k^*(\nu) \mathbf{W}_{t-k}.$$

Sans perte de généralité, nous allons supposer que  $\mathbf{B}^\top$  est normalisée sous la forme,  $\mathbf{B}^\top = [\mathbf{I}_r, \mathbf{B}_0]$ , où  $\mathbf{B}_0$  est de dimension  $r \times m$ .

Définissons pour tout  $\nu$  :

$$\mathbf{K}_\nu = [\mathbf{A}_\nu, \boldsymbol{\phi}_1^*(\nu), \dots, \boldsymbol{\phi}_{p-1}^*(\nu)],$$

qui est une matrice d'ordre  $d \times (r + d(p-1))$ , et  $\boldsymbol{\sigma}_\nu = \text{vec}(\mathbf{K}_\nu)$  qui est un vecteur colonne d'ordre  $rd + d^2(p-1)$ . Posons  $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_1^\top, \dots, \boldsymbol{\sigma}_s^\top)^\top$  le vecteur colonne d'ordre  $sd(r + d(p-1))$ ,  $\boldsymbol{\beta}_0 = \text{vec}(\mathbf{B}_0)$ , le vecteur colonne d'ordre  $rm$  et  $\boldsymbol{\eta} = (\boldsymbol{\beta}_0^\top, \boldsymbol{\sigma}^\top)^\top$ , le vecteur colonne d'ordre  $rm + sd(r + d(p-1))$ . Le but ici est de trouver l'estimateur du maximum de vraisemblance de  $\boldsymbol{\eta}$ . On adopte le critère de vraisemblance conditionnelle.

La log-vraisemblance conditionnelle gaussienne du modèle s'écrit :

$$l = -\frac{Nsd}{2} - \frac{N}{2} \log \{ \det(\boldsymbol{\Sigma}_\varepsilon) \} - \frac{1}{2} \sum_{n=1}^N \boldsymbol{\varepsilon}_n^{*\top} \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{\varepsilon}_n^*,$$

où  $\boldsymbol{\Sigma}_\varepsilon = \text{diag}(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_s)$ ,  $N$  étant le nombre total d'années. Malgré le choix de la formule de la vraisemblance gaussienne, l'estimateur obtenu est convergent et asymptotiquement normal sous l'hypothèse 4.2. Ceci a été montré par Brockwell et Davis (1991, Théorème 10.8.1) pour l'estimation des paramètres d'un processus ARMA quelconque. Nous posons également :

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{\eta}^\top} &= - \sum_{n=1}^N \boldsymbol{\varepsilon}_n^{*\top} \boldsymbol{\Sigma}_\varepsilon^{-1} \frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\eta}^\top}, \\ \frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\eta}^\top} &= \left( \frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\beta}_0^\top}, \frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\sigma}^\top} \right), \\ \frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\sigma}^\top} &= \text{diag} \left( \frac{\partial \boldsymbol{\varepsilon}_{1,n}^*}{\partial \boldsymbol{\sigma}_1^\top}, \dots, \frac{\partial \boldsymbol{\varepsilon}_{s,n}^*}{\partial \boldsymbol{\sigma}_s^\top} \right), \\ \frac{\partial \boldsymbol{\varepsilon}_{\nu,n}}{\partial \boldsymbol{\sigma}_\nu^\top} &= -[(\mathbf{B}\mathbf{Y}_{t-1})^\top, \mathbf{W}_{t-1}^\top, \dots, \mathbf{W}_{t-p+1}^\top] \otimes \mathbf{I}_d, \end{aligned}$$



$$\begin{aligned}\frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\beta}_0^\top} &= \left( \left( \frac{\partial \boldsymbol{\varepsilon}_{1,n}}{\partial \boldsymbol{\beta}_0^\top} \right)^\top, \dots, \left( \frac{\partial \boldsymbol{\varepsilon}_{s,n}}{\partial \boldsymbol{\beta}_0^\top} \right)^\top \right)^\top, \\ \frac{\partial \boldsymbol{\varepsilon}_{\nu,n}}{\partial \boldsymbol{\beta}_0^\top} &= -(\mathbf{H}^\top \mathbf{Y}_{t-1})^\top \otimes \mathcal{A}_\nu, \\ \mathbf{H}^\top &= [\mathbf{0}, \mathbf{I}_m].\end{aligned}$$

Ainsi, les matrices  $\frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\sigma}^\top}$ ,  $\frac{\partial \boldsymbol{\varepsilon}_{\nu,n}}{\partial \boldsymbol{\sigma}_\nu^\top}$ ,  $\frac{\partial \boldsymbol{\varepsilon}_{\nu,n}}{\partial \boldsymbol{\beta}_0^\top}$  et  $\mathbf{H}$  sont respectivement d'ordre  $sd \times sd(r + d(p-1))$ ,  $d \times d(r + d(p-1))$  et finalement  $d \times rm$  et  $m \times d$ .

Il en résulte que :

$$\frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\beta}_0^\top} = \begin{pmatrix} -(\mathbf{H}^\top \mathbf{Y}_{s,n-1})^\top \otimes \mathcal{A}_1 \\ -(\mathbf{H}^\top \mathbf{Y}_{1,n})^\top \otimes \mathcal{A}_2 \\ \vdots \\ -(\mathbf{H}^\top \mathbf{Y}_{s-1,n})^\top \otimes \mathcal{A}_s \end{pmatrix}$$

est d'ordre  $ds \times rm$ . De plus :

$$\frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\eta}^\top} = \begin{bmatrix} -(\mathbf{H}^\top \mathbf{Y}_{s,n-1})^\top \otimes \mathcal{A}_1 & \frac{\partial \boldsymbol{\varepsilon}_{1,n}}{\partial \sigma_1^\top} & \mathbf{0} & \dots & \mathbf{0} \\ -(\mathbf{H}^\top \mathbf{Y}_{1,n})^\top \otimes \mathcal{A}_2 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ -(\mathbf{H}^\top \mathbf{Y}_{s-1,n})^\top \otimes \mathcal{A}_s & \mathbf{0} & \dots & \mathbf{0} & \frac{\partial \boldsymbol{\varepsilon}_{s,n}}{\partial \sigma_s^\top} \end{bmatrix} = -\mathbf{V}_n^*.$$

Pour  $\boldsymbol{\eta}_1$  une solution et  $\boldsymbol{\eta}_0$  donné, un développement de Taylor d'ordre un au voisinage de  $\boldsymbol{\eta}_0$  nous donne :

$$\left( \frac{\partial \boldsymbol{l}}{\partial \boldsymbol{\eta}^\top}(\boldsymbol{\eta}_1) \right)^\top = \left( \frac{\partial \boldsymbol{l}}{\partial \boldsymbol{\eta}^\top}(\boldsymbol{\eta}_0) \right)^\top + \frac{\partial}{\partial \boldsymbol{\eta}^\top} \left( \frac{\partial \boldsymbol{l}}{\partial \boldsymbol{\eta}^\top} \right)^\top (\boldsymbol{\eta}_0) (\boldsymbol{\eta}_1 - \boldsymbol{\eta}_0) + o(\|\boldsymbol{\eta}_1 - \boldsymbol{\eta}_0\|).$$

En faisant tendre  $\boldsymbol{\eta}_0$  vers  $\boldsymbol{\eta}_1$  et en posant  $\frac{\partial \boldsymbol{l}}{\partial \boldsymbol{\eta}^\top}(\boldsymbol{\eta}_1) = \mathbf{0}$ , on obtient :

$$\frac{\partial \boldsymbol{l}}{\partial \boldsymbol{\eta}^\top}(\boldsymbol{\eta}_1) = \mathbf{0} \Rightarrow \boldsymbol{\eta}_1 - \boldsymbol{\eta}_0 \simeq - \left( \frac{\partial}{\partial \boldsymbol{\eta}^\top} \left( \frac{\partial \boldsymbol{l}}{\partial \boldsymbol{\eta}^\top} \right)^\top (\boldsymbol{\eta}_0) \right)^{-1} \left( \frac{\partial \boldsymbol{l}}{\partial \boldsymbol{\eta}^\top}(\boldsymbol{\eta}_0) \right)^\top.$$

En posant  $\boldsymbol{\eta}_0 = \boldsymbol{\eta}_i$  et  $\boldsymbol{\eta}_1 = \boldsymbol{\eta}_{i+1}$ , on a d'après Ahn et Reinsel (1990, p. 817) et Yap et Reinsel (1995, p. 254) que l'estimateur de  $\boldsymbol{\eta}$  s'obtient par l'algorithme de

Newton Raphson :

$$\hat{\boldsymbol{\eta}}_{i+1} = \hat{\boldsymbol{\eta}}_i + \left[ \sum_{n=1}^N \mathbf{V}_n^{*\top} \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{V}_n^* \right]_{\hat{\boldsymbol{\eta}}_i}^{-1} \left[ \sum_{n=1}^N \mathbf{V}_n^{*\top} \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{\varepsilon}_n^* \right]_{\hat{\boldsymbol{\eta}}_i}. \quad (4.3.21)$$

La présence du terme quadratique dans l'équation 4.3.21 entraîne une convergence rapide de l'algorithme vers la solution recherchée.

#### 4.3.3.1. Distribution asymptotique de l'estimateur de rang réduit

Dans cette section nous établissons la distribution asymptotique de l'estimateur de rang réduit. Posons  $b = sd(r + d(p - 1))$  et  $\mathbf{D}^{**} = \text{diag}(N\mathbf{I}_{rm}, N^{1/2}\mathbf{I}_b)$ .

Nous avons que :

$$\mathbf{D}^{**\top} \mathbf{V}_n^{*\top} \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{V}_n^* \mathbf{D}^{**} = \begin{bmatrix} N^{-2} \left( \frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\beta}_0^\top} \right)^\top \boldsymbol{\Sigma}_\varepsilon^{-1} \frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\beta}_0^\top} & N^{-3/2} \left( \frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\beta}_0^\top} \right)^\top \boldsymbol{\Sigma}_\varepsilon^{-1} \frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\sigma}^\top} \\ N^{-3/2} \left( \frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\sigma}^\top} \right)^\top \boldsymbol{\Sigma}_\varepsilon^{-1} \frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\beta}_0^\top} & N^{-1} \left( \frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\sigma}^\top} \right)^\top \boldsymbol{\Sigma}_\varepsilon^{-1} \frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\sigma}^\top} \end{bmatrix}. \quad (4.3.22)$$

D'autre part, nous avons :

$$\left( \frac{\partial \boldsymbol{\varepsilon}_{\nu,n}}{\partial \boldsymbol{\beta}_0^\top} \right)^\top = -(\mathbf{H}^\top \mathbf{Y}_{sn+\nu-1}) \otimes \mathcal{A}_\nu^\top,$$

où

$$\begin{aligned} \mathbf{H}^\top \mathbf{Y}_{sn+\nu-1} &= \mathbf{H}^\top \mathbf{P}_a \mathbf{Z}_{sn+\nu-1} = [\mathbf{H}^\top \mathbf{P}_m, \mathbf{H}^\top \mathbf{P}_r] \begin{bmatrix} \mathbf{Z}_{m:(sn+\nu-1)} \\ \mathbf{Z}_{r:(sn+\nu-1)} \end{bmatrix}, \\ &= \mathbf{P}_{m1} \mathbf{Z}_{m:(sn+\nu-1)} + \mathbf{P}_{r1} \mathbf{Z}_{r:(sn+\nu-1)}, \end{aligned}$$

avec  $\mathbf{P}_{m1} = \mathbf{H}^\top \mathbf{P}_m$  et  $\mathbf{P}_{r1} = \mathbf{H}^\top \mathbf{P}_r$ . On retrouve alors que :

$$\left( \frac{\partial \boldsymbol{\varepsilon}_{\nu,n}}{\partial \boldsymbol{\beta}_0^\top} \right)^\top = -(\mathbf{P}_{m1} \mathbf{Z}_{m:(sn+\nu-1)} + \mathbf{P}_{r1} \mathbf{Z}_{r:(sn+\nu-1)}) \otimes \mathcal{A}_\nu^\top = \boldsymbol{\xi}_\nu \otimes \mathcal{A}_\nu^\top.$$

En passant à la somme et en multipliant par  $N^{-2}$ , on a pour  $t = sn + \nu$  :

$$N^{-2} \sum_{n=1}^N \left( \frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\beta}_0^\top} \right)^\top \boldsymbol{\Sigma}_\varepsilon^{-1} \frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\beta}_0^\top} = N^{-2} \sum_{n=1}^N (\boldsymbol{\xi}_\nu \boldsymbol{\xi}_\nu^\top \otimes \mathcal{A}_\nu^\top \boldsymbol{\Sigma}_\varepsilon^{-1} \mathcal{A}_\nu),$$

avec  $\boldsymbol{\xi}_\nu = \mathbf{P}_{m1} \mathbf{Z}_{m:(sn+\nu-1)} + \mathbf{P}_{r1} \mathbf{Z}_{r:(sn+\nu-1)}$ .

Or  $N^{-2} \sum_{n=1}^N \mathbf{Z}_{m:(t-1)} \mathbf{Z}_{r:(t-1)}^\top$  et  $N^{-2} \sum_{n=1}^N \mathbf{Z}_{r:(t-1)} \mathbf{Z}_{r:(t-1)}^\top$  sont  $\mathbf{o}_p(1)$ , car  $\mathbf{Z}_{m:t}$  représente le terme non stationnaire et  $\mathbf{Z}_{r:t}$  le terme stationnaire. Par le même principe qu'aux points (iii) et (iv) du lemme 4.1, on montre que :

$$\begin{aligned} N^{-3/2} \sum_{n=1}^N \mathbf{Z}_{m:(t-1)} \mathbf{Z}_{r:(t-1)}^\top &\xrightarrow{\mathcal{P}} \mathbf{0}, \\ N^{-1} \sum_{n=1}^N \mathbf{Z}_{r:(t-1)} \mathbf{Z}_{r:(t-1)}^\top &\xrightarrow{\mathcal{P}} \mathbf{\Gamma}_r, \end{aligned}$$

où  $\mathbf{\Gamma}_r = E(\mathbf{Z}_{r:(t-1)} \mathbf{Z}_{r:(t-1)}^\top)$ . Par conséquent,

$$\begin{aligned} N^{-2} \sum_{n=1}^N \left( \frac{\partial \varepsilon_n^*}{\partial \boldsymbol{\beta}_0^\top} \right)^\top \boldsymbol{\Sigma}_\varepsilon^{-1} \frac{\partial \varepsilon_n^*}{\partial \boldsymbol{\beta}_0^\top} &= N^{-2} \sum_{n=1}^N (\mathbf{P}_{m1} \mathbf{Z}_{m:(t-1)} \mathbf{Z}_{m:(t-1)}^\top \mathbf{P}_{m1}^\top \otimes \mathbf{A}_\nu^\top \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{A}_\nu) \\ &+ \mathbf{o}_p(1). \end{aligned}$$

D'autre part, on montre aussi en utilisant le même principe qu'au point (iii) du lemme 4.1 que  $N^{-3/2} \sum_{n=1}^N \left( \frac{\partial \varepsilon_n^*}{\partial \boldsymbol{\sigma}^\top} \right)^\top \boldsymbol{\Sigma}_\varepsilon^{-1} \frac{\partial \varepsilon_n^*}{\partial \boldsymbol{\beta}_0^\top} \xrightarrow{\mathcal{P}} \mathbf{0}$ . En remarquant aussi que,

$$\left( \frac{\partial \varepsilon_n^*}{\partial \boldsymbol{\beta}_0^\top} \right)^\top \boldsymbol{\Sigma}_\varepsilon^{-1} \frac{\partial \varepsilon_n^*}{\partial \boldsymbol{\beta}_0^\top} = \sum_{\nu=1}^s \left( \frac{\partial \varepsilon_{\nu,n}}{\partial \boldsymbol{\beta}_0^\top} \right)^\top \boldsymbol{\Sigma}_\varepsilon^{-1} \frac{\partial \varepsilon_{\nu,n}}{\partial \boldsymbol{\beta}_0^\top},$$

l'expression (4.3.22) devient :

$$\mathbf{D}^{**,-1} \left[ \sum_{n=1}^N \mathbf{V}_n^{*\top} \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{V}_n^* \right] \mathbf{D}^{**,-1} = \text{diag} [\boldsymbol{\Upsilon}_1, \boldsymbol{\Upsilon}_2] + \mathbf{o}_p(1), \quad (4.3.23)$$

où

$$\begin{aligned} \boldsymbol{\Upsilon}_1 &= N^{-2} \sum_{\nu=1}^s \sum_{n=1}^N (\mathbf{P}_{m1} \mathbf{Z}_{m:(t-1)} \mathbf{Z}_{m:(t-1)}^\top \mathbf{P}_{m1}^\top \otimes \mathbf{A}_\nu^\top \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{A}_\nu), \\ \boldsymbol{\Upsilon}_2 &= \text{diag} \left( N^{-1} \sum_{n=1}^N \left( \frac{\partial \varepsilon_{\nu,n}}{\partial \boldsymbol{\sigma}_\nu^\top} \right)^\top \boldsymbol{\Sigma}_\nu^{-1} \frac{\partial \varepsilon_{\nu,n}}{\partial \boldsymbol{\sigma}_\nu^\top}, \nu = 1, \dots, s \right), \end{aligned}$$

où  $s$  et  $\nu$  satisfont  $t = sn + \nu$ . L'expression (4.3.23) représente une matrice ayant  $s+1$  blocs sur la diagonale, dont un bloc pour le terme  $\boldsymbol{\beta}_0$  et un bloc pour chaque  $\boldsymbol{\sigma}_\nu$ ,  $\nu = 1, \dots, s$ . Toujours en se référant à l'expression de l'équation (4.3.21), nous allons maintenant faire un développement analogue à celui ci-dessus sur le terme

$\sum_{n=1}^N \mathbf{V}_n^{*\top} \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{\varepsilon}_n^*$ . Nous avons :

$$\sum_{n=1}^N \mathbf{V}_n^{*\top} \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{\varepsilon}_n^* = \begin{bmatrix} \sum_{n=1}^N \left( \frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\beta}_0^\top} \right)^\top \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{\varepsilon}_n^* \\ \sum_{n=1}^N \left( \frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\sigma}^\top} \right)^\top \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{\varepsilon}_n^* \end{bmatrix}.$$

Or,

$$\left( \frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\beta}_0^\top} \right)^\top \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{\varepsilon}_n^* = \sum_{\nu=1}^s \left( \frac{\partial \boldsymbol{\varepsilon}_{\nu,n}}{\partial \boldsymbol{\beta}_0^\top} \right)^\top \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{\varepsilon}_{\nu,n}.$$

Pour les mêmes raisons vues précédemment, on a :

$$N^{-1} \sum_{n=1}^N \left( \frac{\partial \boldsymbol{\varepsilon}_n^*}{\partial \boldsymbol{\beta}_0^\top} \right)^\top \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{\varepsilon}_n^* = N^{-1} \sum_{n=1}^N \left( \sum_{\nu=1}^s \mathbf{P}_{m1} \mathbf{Z}_{m:(sn+\nu-1)} \otimes \mathbf{A}_\nu^\top \boldsymbol{\Sigma}_\nu^{-1} \right) \boldsymbol{\varepsilon}_{\nu,n} + \mathbf{o}_p(1).$$

On obtient finalement :

$$\mathbf{D}^{** -1} \begin{bmatrix} \sum_{n=1}^N \mathbf{V}_n^{*\top} \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{\varepsilon}_n^* \end{bmatrix} = \begin{bmatrix} N^{-1} \sum_{n=1}^N \left( \sum_{\nu=1}^s \mathbf{P}_{m1} \mathbf{Z}_{m:(sn+\nu-1)} \otimes \mathbf{A}_\nu^\top \boldsymbol{\Sigma}_\nu^{-1} \right) \boldsymbol{\varepsilon}_{\nu,n} \\ N^{-1/2} \sum_{n=1}^N \left( \frac{\partial \boldsymbol{\varepsilon}_{1,n}}{\partial \boldsymbol{\sigma}_1^\top} \right)^\top \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\varepsilon}_{1,n} \\ \vdots \\ N^{-1/2} \sum_{n=1}^N \left( \frac{\partial \boldsymbol{\varepsilon}_{s,n}}{\partial \boldsymbol{\sigma}_s^\top} \right)^\top \boldsymbol{\Sigma}_s^{-1} \boldsymbol{\varepsilon}_{s,n} \end{bmatrix} + \mathbf{o}_p(1). \quad (4.3.24)$$

Ceci combiné avec l'équation (4.3.23) nous permet d'avoir :

$$N(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0) = \boldsymbol{\Xi}_N^{-1} \boldsymbol{\Delta}_N + \mathbf{o}_p(1), \quad (4.3.25)$$

où

$$\begin{aligned} \boldsymbol{\Xi}_N &= N^{-2} \sum_{\nu=1}^s \sum_{n=1}^N \mathbf{P}_{m1} \mathbf{Z}_{m:(sn+\nu-1)} \mathbf{Z}_{m:(sn+\nu-1)}^\top \mathbf{P}_{m1}^\top \otimes \mathbf{A}_\nu^\top \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{A}_\nu, \\ \boldsymbol{\Delta}_N &= \left( N^{-1} \sum_{n=1}^N \sum_{\nu=1}^s \mathbf{P}_{m1} \mathbf{Z}_{m:(sn+\nu-1)} \otimes \mathbf{A}_\nu^\top \boldsymbol{\Sigma}_\nu^{-1} \right) \boldsymbol{\varepsilon}_{\nu,n}. \end{aligned}$$

On a aussi :

$$\begin{aligned} N^{1/2}(\hat{\boldsymbol{\sigma}}_\nu - \boldsymbol{\sigma}_\nu) &= \left[ N^{-1} \sum_{n=1}^N \left( \frac{\partial \boldsymbol{\varepsilon}_{\nu,n}}{\partial \boldsymbol{\sigma}_\nu^\top} \right)^\top \boldsymbol{\Sigma}_\nu^{-1} \frac{\partial \boldsymbol{\varepsilon}_{\nu,n}}{\partial \boldsymbol{\sigma}_\nu^\top} \right]^{-1} \left( N^{-1/2} \sum_{n=1}^N \left( \frac{\partial \boldsymbol{\varepsilon}_{\nu,n}}{\partial \boldsymbol{\sigma}_\nu^\top} \right)^\top \boldsymbol{\Sigma}_\nu^{-1} \boldsymbol{\varepsilon}_{\nu,n} \right) \\ &+ \mathbf{o}_p(1). \end{aligned}$$

où,  $\frac{\partial \boldsymbol{\varepsilon}_{\nu,n}}{\partial \boldsymbol{\sigma}_\nu^\top} = -[(\mathbf{B}\mathbf{Y}_{t-1})^\top, \mathbf{W}_{t-1}^\top, \dots, \mathbf{W}_{t-p+1}^\top] \otimes \mathbf{I}_d = \mathbf{U}_{\nu,n}^*$ .

On a alors la proposition suivante, qui donne le comportement asymptotique de  $\hat{\beta}_0$  et de  $\hat{\sigma}_\nu$ .

**Proposition 4.8.** *Soit  $\hat{\eta} = (\hat{\beta}_0^\top, \hat{\sigma}^\top)^\top$ , l'estimateur de  $\eta$  obtenu en utilisant un estimateur initial consistant de  $\beta_0$ , on obtient alors les résultats asymptotiques suivants :*

$$\begin{aligned} N(\hat{\beta}_0 - \beta_0) &\xrightarrow{\mathcal{L}} \Xi^{-1}\Delta, \\ N^{1/2}(\hat{\sigma}_\nu - \sigma_\nu) &\xrightarrow{\mathcal{L}} \mathcal{N}_{rd+d^2(p-1)}(\mathbf{0}, \Sigma_{\nu,\varepsilon} \otimes \Sigma_{U^*;\nu}^{-1}), \end{aligned}$$

où

$$\begin{aligned} \Xi &= \sum_{\nu=1}^s \{P_{m1} M_{1:\nu} P_{m1}^\top \otimes \mathcal{A}_\nu^\top \Sigma_{\nu,\varepsilon}^{-1} \mathcal{A}_\nu\}, \\ \Delta &= \sum_{\nu=1}^s \text{vec}(\mathcal{A}_\nu^\top \Sigma_{\nu,\varepsilon}^{-1} P_a M_{2:\nu} P_{m1}^\top), \\ \Sigma_{U^*;\nu} &= \lim_{N \rightarrow \infty} \left\{ N^{-1} \sum_{n=1}^N (U_{\nu,n}^* \Sigma_{\nu,\varepsilon}^{-1} U_{\nu,n}^{*\top}) \right\}, \\ U_{\nu,n}^* &= -[(\mathbf{B}\mathbf{Y}_{t-1})^\top, \mathbf{W}_{t-1}^\top, \dots, \mathbf{W}_{t-p+1}^\top] \otimes \mathbf{I}_d, \end{aligned}$$

et :

$$\begin{aligned} \mathbf{M}_{1:\nu} &= \psi_{\nu-1}(1) \mathbf{Q} \Sigma_\varepsilon^{1/2} \left\{ \int_0^1 \mathbf{w}(r) \mathbf{w}^\top(r) dr \right\} \Sigma_\varepsilon^{1/2} (\psi_{\nu-1}(1) \mathbf{Q})^\top, \\ \mathbf{M}_{2:\nu} &= (\mathbf{D}_\nu \mathbf{Q}) \Sigma_\varepsilon^{1/2} \left\{ \int_0^1 \mathbf{w}(r) d\mathbf{W}^\top(r) \right\}^\top \Sigma_\varepsilon^{1/2} (\psi_{\nu-1}(1) \mathbf{Q})^\top \\ &\quad + (\mathbf{D}_\nu \mathbf{Q}) \Sigma_\varepsilon (\psi_{\nu-1,0} \mathbf{Q})^\top. \end{aligned}$$

*Preuve :* Elle découle du Lemme 4.1.

#### 4.3.3.2. Estimateurs initiaux et leur distribution limite

On a pour tout  $\nu$ ,  $\pi_\nu = \mathcal{A}_\nu \mathbf{B}^\top$ , avec  $\mathbf{B}^\top = [\mathbf{I}_r, \mathbf{B}_0]$ , après l'estimation de  $\pi_\nu$ , on trouve  $\mathbf{P}_{m:\nu}$  une matrice d'ordre  $d \times m$  telle que  $\pi_\nu \mathbf{P}_{m:\nu} = \mathbf{0}$ . La matrice  $\mathbf{P}_{m:\nu}$  peut encore s'écrire  $\mathbf{P}_{m:\nu} = [\mathbf{P}_{r1:\nu}^\top, \mathbf{P}_{m1:\nu}^\top]^\top$ , où  $\mathbf{P}_{r1:\nu}$  et  $\mathbf{P}_{m1:\nu}$  sont respectivement des matrices d'ordre  $r \times m$  et  $m \times m$ . On peut encore écrire  $\pi_\nu = [\hat{\mathbf{C}}_{1:\nu}, \hat{\mathbf{C}}_{2:\nu}]$  où

$\hat{\mathbf{C}}_{1:\nu}$  et  $\hat{\mathbf{C}}_{2:\nu}$  sont respectivement des matrices d'ordre  $d \times r$  et  $d \times m$ . On a :

$$\begin{aligned}
(\boldsymbol{\pi}_\nu \mathbf{P}_{m:\nu} = \mathbf{0}) &\Rightarrow \mathbf{A}_\nu \mathbf{B}^\top \mathbf{P}_{m:\nu} = \mathbf{0}, \\
&\Rightarrow \mathbf{B}^\top \mathbf{P}_{m:\nu} = \mathbf{0} \text{ parce que } \mathbf{A}_\nu \text{ est de plein rang,} \\
&\Rightarrow [\mathbf{I}_r, \mathbf{B}_0] \begin{bmatrix} \mathbf{P}_{r1:\nu} \\ \mathbf{P}_{m1:\nu} \end{bmatrix} = \mathbf{P}_{r1:\nu} + \mathbf{B}_0 \mathbf{P}_{m1:\nu} = \mathbf{0}, \\
&\Rightarrow \mathbf{B}_0 = -\mathbf{P}_{r1:\nu} \mathbf{P}_{m1:\nu}^{-1}.
\end{aligned}$$

De l'estimation de plein rang, on a d'après l'équation (4.3.19) que :

$$N \hat{\boldsymbol{\pi}}_\nu \mathbf{P}_m = \left( N^{-1} \sum_{n=1}^N \boldsymbol{\varepsilon}_t \mathbf{Z}_{m:(t-1)}^\top \right) \left( N^{-2} \sum_{n=1}^N \mathbf{Z}_{m:(t-1)} \mathbf{Z}_{m:(t-1)}^\top \right)^{-1} + o_p(1). \quad (4.3.26)$$

On peut encore écrire :

$$\begin{aligned}
N \hat{\boldsymbol{\pi}}_\nu \mathbf{P}_{m:\nu} \mathbf{P}_{m1:\nu}^{-1} &= N[\hat{\mathbf{C}}_{1:\nu}, \hat{\mathbf{C}}_{2:\nu}] \begin{bmatrix} \mathbf{P}_{r1:\nu} \\ \mathbf{P}_{m1:\nu} \end{bmatrix} \mathbf{P}_{m1:\nu}^{-1}, \\
&= N[\hat{\mathbf{C}}_{1:\nu}, \hat{\mathbf{C}}_{2:\nu}] \begin{bmatrix} \mathbf{P}_{r1:\nu} \mathbf{P}_{m1:\nu}^{-1} \\ \mathbf{I}_m \end{bmatrix}, \\
&= N \left( \hat{\mathbf{C}}_{1:\nu} \mathbf{P}_{r1:\nu} \mathbf{P}_{m1:\nu}^{-1} + \hat{\mathbf{C}}_{2:\nu} \right), \\
&= N \left( \hat{\mathbf{C}}_{2:\nu} - \hat{\mathbf{C}}_{1:\nu} \mathbf{B}_0 \right), \\
&= \left( N^{-1} \sum_{n=0}^{N-1} \boldsymbol{\varepsilon}_t \mathbf{Z}_{m:(t-1)}^\top \right) \left( N^{-2} \sum_{n=0}^{N-1} \mathbf{Z}_{m:(t-1)} \mathbf{Z}_{m:(t-1)}^\top \right)^{-1} \mathbf{P}_{m1:\nu}^{-1} \\
&\quad + o_p(1).
\end{aligned}$$

En multipliant les deux côtés par  $\hat{\Gamma}_\nu = \left( \hat{\mathbf{C}}_{1:\nu}^\top \hat{\boldsymbol{\Sigma}}_{\nu,\varepsilon}^{-1} \hat{\mathbf{C}}_{1:\nu} \right)^{-1} \hat{\mathbf{C}}_{1:\nu}^\top \hat{\boldsymbol{\Sigma}}_{\nu,\varepsilon}^{-1}$ . On obtient :

$$N \left( \hat{\Gamma}_\nu \hat{\mathbf{C}}_{2:\nu} - \mathbf{B}_0 \right) = \hat{\Gamma}_\nu \boldsymbol{\Upsilon}_3 \boldsymbol{\Upsilon}_4^{-1} + o_p(1), \quad (4.3.27)$$

où

$$\begin{aligned}
\boldsymbol{\Upsilon}_3 &= N^{-1} \sum_{n=0}^{N-1} \boldsymbol{\varepsilon}_t \mathbf{Z}_{m:(t-1)}^\top \mathbf{P}_{m1:\nu}^\top, \\
\boldsymbol{\Upsilon}_4 &= N^{-2} \mathbf{P}_{m1:\nu} \sum_{n=0}^{N-1} \mathbf{Z}_{m:(t-1)} \mathbf{Z}_{m:(t-1)}^\top \mathbf{P}_{m1:\nu}^\top,
\end{aligned}$$

où  $s$  et  $\nu$  satisfont  $t = ns + \nu$ .

L'estimateur  $\hat{\mathbf{B}}_\nu = \hat{\Gamma}_\nu \hat{\mathbf{C}}_{2:\nu}$ , est celui qui aurait été obtenu s'il n'y avait pas de périodicité. Il est similaire à celui obtenu par Yap et Reinsel (1995). Mais ici, nous cherchons un estimateur initial qui a les mêmes propriétés que l'estimateur de rang réduit.

L'équation (4.3.27) est équivalente à :

$$\begin{aligned} N \left( \hat{\mathbf{C}}_{1:\nu}^\top \hat{\Sigma}_{\nu,\varepsilon}^{-1} \hat{\mathbf{C}}_{1:\nu} \right) \left( \hat{\mathbf{B}}_\nu - \mathbf{B}_0 \right) & \left( N^{-2} \mathbf{P}_{m1:\nu} \sum_{n=1}^N \mathbf{Z}_{m:(t-1)} \mathbf{Z}_{m:(t-1)}^\top \mathbf{P}_{m1:\nu}^\top \right) \\ & = \hat{\mathbf{C}}_{1:\nu}^\top \hat{\Sigma}_{\nu,\varepsilon}^{-1} \left( N^{-1} \sum_{n=1}^N \varepsilon_t \mathbf{Z}_{m:(t-1)}^\top \mathbf{P}_{m1:\nu}^\top \right) + \mathbf{o}_p(1). \end{aligned}$$

Prenant la somme des deux côtés de l'égalité, on a :

$$\begin{aligned} N \sum_{\nu=1}^s \left( \hat{\mathbf{C}}_{1:\nu}^\top \hat{\Sigma}_{\nu,\varepsilon}^{-1} \hat{\mathbf{C}}_{1:\nu} \right) \left( \hat{\mathbf{B}}_\nu - \mathbf{B}_0 \right) & \left( N^{-2} \mathbf{P}_{m1:\nu} \sum_{n=1}^N \mathbf{Z}_{m:(t-1)} \mathbf{Z}_{m:(t-1)}^\top \mathbf{P}_{m1:\nu}^\top \right) \\ & = \sum_{\nu=1}^s \hat{\mathbf{C}}_{1:\nu}^\top \hat{\Sigma}_{\nu,\varepsilon}^{-1} \left( N^{-1} \sum_{n=1}^N \varepsilon_t \mathbf{Z}_{m:(t-1)}^\top \mathbf{P}_{m1:\nu}^\top \right) + \mathbf{o}_p(1). \end{aligned}$$

En appliquant l'opérateur vec des deux côtés de l'égalité ci-dessus et en définissant,

$$\begin{aligned} \mathbf{L}_\nu & = \left( N^{-2} \mathbf{P}_{m1:\nu} \sum_{n=1}^N \mathbf{Z}_{m:(t-1)} \mathbf{Z}_{m:(t-1)}^\top \mathbf{P}_{m1:\nu}^\top \right) \otimes \hat{\mathbf{C}}_{1:\nu}^\top \hat{\Sigma}_{\nu,\varepsilon}^{-1} \hat{\mathbf{C}}_{1:\nu}, \\ \mathbf{L}_{N:s} & = N^{-1} \sum_{\nu=1}^s \sum_{n=1}^N \left( \mathbf{P}_{m1:\nu} \mathbf{Z}_{m:(t-1)} \otimes \hat{\mathbf{C}}_{1:\nu}^\top \hat{\Sigma}_{\nu,\varepsilon}^{-1} \right) \varepsilon_t, \end{aligned}$$

on a :

$$N \left\{ \sum_{\nu=1}^s \mathbf{L}_{1:\nu} \text{vec} \left( \hat{\mathbf{B}}_\nu \right) - \text{vec} \left( \mathbf{B}_0 \right) \right\} = \left( \sum_{\nu=1}^s \mathbf{L}_{1:\nu} \right)^{-1} \mathbf{L}_{N:s} + \mathbf{o}_p(1).$$

On obtient alors un estimateur initial de  $\mathbf{B}_0$  donné par :

$$\tilde{\mathbf{B}}_0 = \sum_{\nu=1}^s \mathbf{L}_{1:\nu} \text{vec} \left( \hat{\mathbf{B}}_\nu \right),$$

qui vérifie  $N(\tilde{\beta}_0 - \beta_0) \xrightarrow{\mathcal{L}} \Xi^{-1}\Delta$  où les quantités  $\Xi$  et  $\Delta$  sont définis dans la proposition (4.8), ce qui signifie qu'il a la même distribution initiale que  $\hat{\beta}_0$ . Ce résultat sur l'hyper-convergence de l'estimateur initial a été établi dans le cas non périodique par Ahn et Reinsel (1990, Théorème 3). L'estimateur initial de  $\mathbf{A}_\nu$  est  $\hat{\mathbf{C}}_{1;\nu}$  pour tout  $\nu = 1, \dots, s$ . Dans la section suivante, nous proposons des tests diagnostiques sur le modèle estimé.

#### 4.4. TESTS DIAGNOSTIQUES SUR LE MODÈLE ESTIMÉ

##### 4.4.1. Distribution asymptotique des autocovariances résiduelles

L'approche développée ici est l'équivalent périodique de ce qui été fait par certains auteurs, notamment Brüggemann et al. (2004) dans le cas non périodique. Pour une saison donnée,  $\nu$ , considérons l'autocovariance échantillonnale :

$$\mathbf{C}(h; \nu) = N^{-1} \sum_{n=1}^N \varepsilon_{sn+\nu} \varepsilon_{sn+\nu-h}^\top, \quad h = 1, \dots, H. \quad (4.4.1)$$

Définissons aussi :

$$\mathbf{c}(h; \nu) = \text{vec}(\mathbf{C}(h; \nu)), \quad h = 1, \dots, H \text{ et } \mathbf{c}(\nu) = (\mathbf{C}(1; \nu)^\top, \dots, \mathbf{C}(H; \nu)^\top)^\top.$$

L'autocovariance résiduelle est donnée par :

$$\hat{\mathbf{C}}(h; \nu) = N^{-1} \sum_{n=1}^N \hat{\varepsilon}_t \hat{\varepsilon}_{t-h}^\top, \quad h = 1, \dots, H,$$

$$\text{où } t = sn + \nu \text{ et } \hat{\varepsilon}_t = \mathbf{W}_t - \hat{\mathbf{A}}_\nu \hat{\mathbf{B}}^\top - \sum_{k=1}^{p-1} \hat{\phi}_k^*(\nu) \mathbf{W}_{t-k}.$$

Établir la distribution asymptotique des autocovariances résiduelles est plus simple lorsque le vecteur de cointégration est supposé connu. Il faut alors établir le lien entre cette situation et le cas naturel où il est supposé inconnu. C'est l'objet du lemme suivant.

**Lemme 4.2.** *Soit,  $\tilde{\varepsilon}_t$  le résidu obtenu après l'estimation par les moindres carrés ordinaires du modèle (4.3.20) quand le vecteur de cointégration  $\mathbf{B}$  est supposé connu et  $\hat{\varepsilon}_t$  le résidu obtenu par la méthode de vraisemblance conditionnelle en*



supposant  $\mathbf{B}$  inconnu. Soient  $\tilde{\mathbf{C}}(h; \nu)$  et  $\hat{\mathbf{C}}(h; \nu)$  les autocovariances correspondantes pour  $\nu = 1, \dots, s$  et  $h = 1, \dots, H$ , alors :

$$\tilde{\mathbf{C}}(h; \nu) - \hat{\mathbf{C}}(h; \nu) = \mathcal{O}_p(N^{-1}).$$

*Preuve* : Elle est présentée en Annexe G.

Le lemme 4.2 implique que la distribution limite de  $N^{1/2}\tilde{\mathbf{C}}(h; \nu)$  est la même que celle de  $N^{1/2}\hat{\mathbf{C}}(h; \nu)$ . Dérivons alors la distribution limite de  $N^{1/2}\tilde{\mathbf{C}}(h; \nu)$ .

En supposant que  $\mathbf{B}$  est connu, définissons pour tout  $t = sn + \nu$  :

$$\begin{aligned} \tilde{\mathbf{X}}_t &= \left[ (\mathbf{B}^\top \mathbf{Y}_{t-1})^\top, \mathbf{W}_{t-1}^\top, \dots, \mathbf{W}_{t-p+1}^\top \right]^\top, \\ \mathbf{F}_{0:\nu} &= [\mathbf{A}_\nu, \phi_1^*(\nu), \dots, \phi_{p-1}^*(\nu)]. \end{aligned}$$

Nous avons alors pour  $t = sn + \nu$  :

$$\mathbf{W}_t = \tilde{\mathbf{X}}_t^\top \mathbf{F}_{0:\nu} + \varepsilon_t = \left( \tilde{\mathbf{X}}_t^\top \otimes \mathbf{I}_d \right) \mathbf{f}_\nu + \varepsilon_t, \quad (4.4.2)$$

où  $\mathbf{f}_\nu = \text{vec}(\mathbf{F}_{0:\nu})$ . Utilisant les mêmes arguments que dans Ursu et Duchesne (2009), le modèle est estimé par les moindres carrés et les autocovariances résiduelles vérifient la relation :

$$N^{1/2}\tilde{\mathbf{C}}(\nu) \xrightarrow{\mathcal{L}} \mathcal{N}_{d^2H}(\mathbf{0}, \mathbf{\Delta}(\nu)),$$

où,

$$\begin{aligned} \tilde{\mathbf{C}}(\nu) &= (\tilde{\mathbf{C}}(1; \nu)^\top, \dots, \tilde{\mathbf{C}}(H; \nu)^\top)^\top, \\ \mathbf{\Delta}(\nu) &= \mathbf{V}(\nu, H) \otimes \mathbf{\Sigma}_\nu - \mathbf{G}^\top(\nu) \mathbf{\Omega}^{-1}(\nu) \mathbf{G}(\nu) \otimes \mathbf{\Sigma}_\nu, \\ \mathbf{V}(\nu, H) &= \text{diag}(\mathbf{\Sigma}_{\nu-h}, h = 1, \dots, H), \\ \mathbf{\Sigma}_{\nu-h} &= \mathbf{\Sigma}_{s+\nu-h} \text{ si } \nu - h < 0, \\ \mathbf{G}(\nu) &= (\mathbf{G}(1; \nu), \dots, \mathbf{G}(H; \nu)), \\ \mathbf{G}(H; \nu) &= \mathbf{E}(\tilde{\mathbf{X}}_t \varepsilon_{t-h}^\top), \quad t = ns + \nu. \end{aligned}$$

Pour une meilleure dérivation de la distribution asymptotique du test portemanteau, nous allons considérer une représentation alternative du modèle de l'équation (4.3.20). C'est une forme autorégressive où les paramètres sont sujets à des contraintes linéaires (voir Brüggemann et al. (2004)).

**Proposition 4.9.** *Si on écrit  $\mathbf{Z}_t = \mathbf{P}_r \mathbf{Q}_r \mathbf{Y}_t + \mathbf{P}_m \mathbf{Q}_m \mathbf{W}_t$ , pour  $t = sn + \nu$ , le modèle (4.3.20) peut encore s'écrire :*

$$\mathbf{Z}_t = \sum_{k=1}^p \mathbf{D}_k(\nu) \mathbf{Z}_{t-k} + \varepsilon_t, \quad (4.4.3)$$

où :

$$\begin{aligned} \mathbf{W}_t &= \mathbf{Y}_t - \mathbf{Y}_{t-1}, \\ \mathbf{D}_1(\nu) &= -\{\mathbf{A}_\nu \mathbf{B}^\top + \boldsymbol{\phi}_1^*(\nu)\}, \\ \mathbf{D}_j(\nu) &= -\boldsymbol{\phi}_j^*(\nu) + \boldsymbol{\phi}_{j-1}^*(\nu) \mathbf{P}_r \mathbf{Q}_r, \quad j = 2, \dots, p-1, \\ \mathbf{D}_p(\nu) &= \boldsymbol{\phi}_{p-1}^*(\nu) \mathbf{P}_r \mathbf{Q}_r. \end{aligned}$$

*Preuve :* Elle est détaillée en Annexe F.

En notation matricielle, les contraintes s'écrivent :

$$\mathbf{D}_\nu^* = [\mathbf{D}_1(\nu), \dots, \mathbf{D}_p(\nu)] = [\mathbf{A}_\nu, \boldsymbol{\phi}_1^*(\nu), \dots, \boldsymbol{\phi}_{p-1}^*(\nu)] \mathbf{R} = \mathbf{F}_{0:\nu} \mathbf{R}, \quad (4.4.4)$$

où  $\mathbf{R}$  est une matrice d'ordre  $(d(p-1) + r) \times (dp)$  faite de blocs de dimension  $d \times d$  et  $r \times d$  tels que les différents blocs  $\mathbf{R}[i : j]$ ,  $i, j = 1, \dots, p$  soient définis par :

$$\begin{aligned} \mathbf{R}[1 : 1] &= \mathbf{B}^\top, \\ \mathbf{R}[i : i] &= \mathbf{P}_r \mathbf{Q}_r, \quad i = 2, \dots, p, \\ \mathbf{R}[i+1 : i] &= -\mathbf{I}_d, \quad i = 1, \dots, p-1, \\ \mathbf{R}[i : j] &= \mathbf{0}, \text{ dans les autres cas.} \end{aligned}$$

Par exemple pour  $p = 5$ , la matrice  $\mathbf{R}$  est :

$$\mathbf{R} = \begin{bmatrix} \mathbf{B}^\top & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{I}_d & \mathbf{P}_r \mathbf{Q}_r & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_d & \mathbf{P}_r \mathbf{Q}_r & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_d & \mathbf{P}_r \mathbf{Q}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_d & \mathbf{P}_r \mathbf{Q}_r \end{bmatrix}. \quad (4.4.5)$$

Dans le cas général, il est assez évident de constater que la matrice des contraintes  $\mathbf{R}$  est de rang  $d(p-1) + r$ . La transformation de l'équation (4.4.4) n'affecte pas les termes d'erreur  $\boldsymbol{\varepsilon}_t$  pour tout  $t$ . Ainsi, les estimateurs convergents de  $\mathbf{D}_\nu^*$  peuvent encore être trouvés en utilisant la relation  $\hat{\mathbf{D}}_\nu^* = \hat{\mathbf{F}}_{0,\nu} \mathbf{R}$ . Les résidus de l'équation (4.4.3) sont alors les mêmes que ceux de l'équation (4.4.2) (voir Brüggemann et al. (2004)).

Suivant Ursu et Duchesne (2009), la statistique portemanteau pour le test diagnostique dans le modèle (4.4.4) pour une saison donnée  $\nu$  est définie par :

$$\begin{aligned} \tilde{\mathcal{Q}}_H(\nu) &= N \tilde{\mathbf{c}}^\top(\nu) \left\{ \tilde{\mathbf{V}}^{-1}(\nu, H) \otimes \tilde{\boldsymbol{\Sigma}}_\nu^{-1} \right\} \tilde{\mathbf{c}}(\nu), \\ &= N \sum_{h=1}^H \text{tr} \left\{ \tilde{\mathbf{C}}^\top(h; \nu) \tilde{\boldsymbol{\Sigma}}_\nu^{-1} \tilde{\mathbf{C}}(h; \nu) \tilde{\boldsymbol{\Sigma}}_{\nu-h}^{-1} \right\}, \end{aligned}$$

qui est approximativement distribuée comme une loi  $\chi^2(d^2 H - K(\nu))$ , où  $K(\nu)$  est le rang de la matrice  $\mathbf{R}^\top \otimes \mathbf{I}_d$ . Ceci entraîne que  $K(\nu) = d^2(p-1) + rd$ .

D'après le Lemme 4.2, ce résultat obtenu quand la matrice de cointégration  $\mathbf{B}$  est connu reste valide quand cette matrice ne l'est pas. Il en résulte que cette distribution approximative est aussi satisfaite par la statistique :

$$\hat{\mathcal{Q}}_H(\nu) = N \sum_{h=1}^H \text{tr} \left\{ \hat{\mathbf{C}}^\top(h; \nu) \hat{\boldsymbol{\Sigma}}_\nu^{-1} \hat{\mathbf{C}}(h; \nu) \hat{\boldsymbol{\Sigma}}_{\nu-h}^{-1} \right\}. \quad (4.4.6)$$

Pour améliorer les propriétés sur petits échantillons, nous considérons aussi la version modifiée de cette statistique :

$$\hat{\mathcal{Q}}_H^*(\nu) = N \sum_{h=1}^H \frac{N}{N - \lfloor (h - \nu + s) / s \rfloor} \text{tr} \left\{ \hat{\mathbf{C}}^\top(h; \nu) \hat{\boldsymbol{\Sigma}}_\nu^{-1} \hat{\mathbf{C}}(h; \nu) \hat{\boldsymbol{\Sigma}}_{\nu-h}^{-1} \right\}. \quad (4.4.7)$$

Nous allons procéder à une étude de simulation afin d'étudier les propriétés des tests statistiques proposés.

#### 4.4.2. Simulations

Dans cette section, les propriétés de la statistique de test ont été étudiées avec des simulations. Le but est de comparer les quantiles empiriques à celles d'une distribution de  $\chi^2$ . Les données ont été générées suivant le processus ci-après :

$$\mathbf{W}_t = \mathbf{A}_\nu \mathbf{B}^\top \mathbf{Y}_{t-1} + \phi_1^*(\nu) \mathbf{W}_{t-1} + \boldsymbol{\varepsilon}_t,$$

où  $\mathbf{W}_t = \mathbf{Y}_t - \mathbf{Y}_{t-1}$ ,  $t = sn + \nu$ ,  $\nu = 1, \dots, s$ .

Dans les simulations, deux modèles ont été considérés, le premier correspond à  $d = 3$ ,  $p = 1$  et le second à  $d = 2$ ,  $p = 2$ . Dans les deux situations, le nombre de saisons a été maintenu à  $s = 4$ . Pour chacun des modèles, le coefficient d'ajustement  $\mathbf{A}_\nu$  et la matrice des variances et covariances  $\boldsymbol{\Sigma}_\nu$  ont été attribuées plusieurs valeurs mais tous les cas n'ont pas été présentés ici. Pour chaque saison, les innovations  $\boldsymbol{\varepsilon}_t$  ont été générées suivant une loi normale multivariée de moyenne nulle et de variance  $\boldsymbol{\Sigma}_\nu$ . Les paramètres ont été fixés de façon à assurer à la fois la périodicité, l'intégration de chaque processus et la cointégration entre les processus. Le nombre total de cycles de saisons a été fixé successivement à 200, 400. Pour chacune de ces valeurs, les données ont été générées 1000 fois et la régression de rang réduit a été ajustée au jeu de données telle que décrite dans la section 3. Les statistiques de portemanteau ont été calculées pour chaque jeu de données pour les valeurs de  $H = 5, 10, 15, 20, 25, 30, 35, 40, 45, 50$ . Pour chacune de ces valeurs, la fréquence empirique a été comparée à la fréquence nominale de 5% et 10%. Les paramètres utilisés dans le jeu de données se présentent comme suit :

Pour  $d = 3$ ,  $p = 1$ ,  $s = 4$ ,

$$\mathbf{A}_\nu = \begin{bmatrix} \alpha_{1,\nu} & 0 \\ 0 & 0 \\ 0 & -\alpha_{3,\nu} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix},$$

$$\Sigma_\nu = \begin{bmatrix} \omega_{1,\nu}^2 & \rho_\nu \omega_{1,\nu} \omega_{2,\nu} & \rho_\nu \omega_{1,\nu} \omega_{3,\nu} \\ \rho_\nu \omega_{2,\nu} \omega_{1,\nu} & \omega_{2,\nu}^2 & \rho_\nu \omega_{2,\nu} \omega_{3,\nu} \\ \rho_\nu \omega_{3,\nu} \omega_{1,\nu} & \rho_\nu \omega_{2,\nu} \omega_{3,\nu} & \omega_{3,\nu}^2 \end{bmatrix}.$$

TABLE 4.1. Valeurs des paramètres pour le premier modèle

	$\nu$			
	1	2	3	4
$\alpha_{1\nu}$	-0,64	0.45	-0.38	-0.44
$\alpha_{3\nu}$	-0.36	-0.34	-0.32	-0.55
$\rho_\nu$	0.75	0.65	0.85	0.71
$\omega_{1,\nu}$	0.80	0.80	0.80	0.80
$\omega_{2,\nu}$	0.90	0.90	0.90	0.90
$\omega_{3,\nu}$	0.75	0.75	0.75	0.75
$\omega_{4,\nu}$	0.95	0.95	0.95	0.95

Pour  $d = 2$ ,  $p = 2$ ,  $s = 4$ , Nous avons les paramètres suivants :

$$\alpha_\nu = \begin{bmatrix} \alpha_{1,\nu} \\ \alpha_{2,\nu} \end{bmatrix}, \quad \beta = \begin{bmatrix} 1 & 2 \end{bmatrix}, \quad \Sigma_\nu = \begin{bmatrix} \omega_{1,\nu}^2 & \rho_\nu \omega_{1,\nu} \omega_{2,\nu} \\ \rho_\nu \omega_{2,\nu} \omega_{1,\nu} & \omega_{2,\nu}^2 \end{bmatrix},$$

$$\phi_1^*(\nu) = \begin{bmatrix} \gamma_{0,\nu} & 0 \\ 0 & \gamma_{0,\nu} \end{bmatrix}.$$

TABLE 4.2. Valeurs des paramètres pour le deuxième modèle

	$\nu$			
	1	2	3	4
$\alpha_{1\nu}$	-0.25	0.35	-0.26	-0.15
$\alpha_{2\nu}$	-0.27	-0.23	-0.26	-0.18
$\rho_\nu$	0.75	0.65	0.85	0.71
$\omega_{1,\nu}$	0.80	0.80	0.80	0.80
$\omega_{2,\nu}$	0.90	0.9	0.90	0.90
$\gamma_{0,\nu}$	0.50	0.55	0.41	0.25

Les résultats des simulations sont présentés dans les tableaux 4.4 et 4.3 :

TABLE 4.3. Niveaux empiriques en pourcentage pour la statistique portmanteau  $\hat{Q}_H(\nu)$  définie en (4.4.6) et la version modifiée  $\hat{Q}_H^*(\nu)$  définie en (4.4.7) pour le modèle 1.

		$\alpha = 5\%, d = 3, p = 1$															
		$N = 200$								$N = 400$							
		$\hat{Q}_H(\nu)$				$\hat{Q}_H^*(\nu)$				$\hat{Q}_H(\nu)$				$\hat{Q}_H^*(\nu)$			
H \ / \ $\nu$		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
5		3.80	3.61	5.40	4.81	4.41	3.91	5.80	4.91	4.11	4.50	4.30	5.11	4.21	4.60	4.40	5.20
10		4.51	3.41	4.71	5.40	5.00	3.60	4.90	5.90	4.01	3.50	3.61	4.71	4.40	3.90	3.70	4.70
15		2.80	3.20	4.61	4.71	3.51	4.50	5.20	5.31	4.21	3.81	4.01	4.40	4.51	3.90	4.51	4.61
20		2.71	3.21	4.20	3.10	3.80	4.20	5.40	4.31	4.21	3.80	3.21	4.11	5.30	4.71	3.81	4.50
25		2.50	3.30	3.50	3.00	4.11	4.91	5.51	4.80	4.20	4.30	3.90	5.01	5.20	5.50	5.01	6.01
30		2.10	3.41	3.21	1.91	3.80	5.41	5.60	3.71	4.41	3.61	3.60	3.70	5.81	5.11	5.21	4.80
35		2.11	2.10	2.30	2.20	4.21	4.71	4.20	4.50	3.01	3.71	4.90	4.31	4.81	4.71	7.00	5.81
40		1.90	1.71	1.80	2.11	5.61	4.40	3.60	5.00	3.00	3.51	2.41	4.71	4.60	5.31	3.91	6.31
45		1.41	1.60	1.90	1.81	4.81	4.40	5.10	4.70	2.71	2.51	4.60	3.40	4.50	4.50	6.51	6.30
50		1.60	1.21	1.51	1.51	5.70	5.00	5.20	4.91	2.01	3.61	2.91	2.91	4.61	5.80	4.70	4.71
		$\alpha = 10\%, d = 3, p = 1$															
5		9.01	8.51	10.2	10.9	9.91	8.80	10.6	11.1	9.71	9.90	9.41	10.4	10.4	10.1	9.61	10.4
10		9.00	8.30	8.91	10.0	10.3	8.90	9.41	10.2	9.41	7.90	8.90	8.90	9.90	8.60	9.51	9.20
15		7.20	7.40	8.10	9.01	8.41	8.70	9.20	9.61	8.21	7.31	8.11	8.01	8.91	8.40	8.80	8.70
20		6.10	6.70	7.91	7.00	8.80	8.10	10.2	8.71	9.10	9.10	7.61	9.20	10.1	9.61	9.11	10.5
25		5.50	7.10	7.30	7.91	8.50	9.70	10.2	9.91	8.20	8.60	8.31	8.61	10.3	9.91	9.40	9.90
30		5.11	6.20	6.50	5.51	8.81	10.0	11.1	8.60	9.61	8.31	8.30	7.61	12.6	10.3	9.90	9.30
35		4.80	5.30	4.71	5.40	8.71	8.81	9.30	9.41	7.90	6.50	9.21	8.10	10.8	9.01	11.3	11.3
40		5.31	4.40	3.61	5.11	10.30	9.10	8.30	9.51	6.31	7.50	6.31	7.80	9.50	10.1	9.61	11.6
45		3.30	3.61	4.70	4.20	9.50	9.10	10.2	9.10	6.01	5.90	8.30	7.60	9.30	8.80	11.10	9.90
50		3.60	3.80	3.51	3.91	11.4	9.70	10.2	10.0	5.91	6.40	6.70	6.41	9.50	10.4	9.61	10.3

TABLE 4.4. Niveaux empiriques en pourcentage pour la statistique portmanteau  $\hat{Q}_H(\nu)$  définie en (4.4.6) et la version modifiée  $\hat{Q}_H^*(\nu)$  définie en (4.4.7) pour le modèle 2.

		$\alpha = 5\%, d = 2, p = 2$															
		$N = 200$								$N = 400$							
		$\hat{Q}_H(\nu)$				$\hat{Q}_H^*(\nu)$				$\hat{Q}_H(\nu)$				$\hat{Q}_H^*(\nu)$			
H \ / \ $\nu$		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
5		5.16	3.97	4.11	5.98	5.50	4.25	4.26	6.04	4.36	3.60	4.88	4.41	4.59	3.82	5.07	4.43
10		3.88	4.00	3.49	3.72	4.13	4.31	3.70	4.01	4.13	4.32	3.64	4.95	4.46	4.63	3.84	5.10
15		2.87	3.29	4.44	4.33	3.59	3.63	4.98	5.24	4.90	4.33	4.16	3.85	5.45	4.41	4.47	4.06
20		2.94	4.89	3.26	2.37	3.96	5.40	3.81	2.84	4.14	4.91	4.16	4.74	4.33	5.14	4.41	5.01
25		4.20	3.52	3.77	2.72	5.65	4.15	5.51	4.40	3.54	4.36	4.06	4.87	3.99	4.72	4.46	5.38
30		3.04	3.68	3.53	3.67	4.10	5.42	4.95	4.53	3.84	4.16	5.38	4.35	5.25	4.75	6.36	5.25
35		3.21	2.32	3.36	2.28	4.78	4.50	5.36	3.16	3.17	2.89	4.06	3.45	4.14	4.14	4.75	4.38
40		2.15	3.17	3.28	2.95	4.41	4.70	4.71	5.02	2.72	4.28	3.25	4.31	3.19	5.13	4.31	5.37
45		2.95	3.14	2.60	2.73	6.17	5.85	4.19	5.31	3.95	3.91	3.47	3.96	5.24	5.34	4.41	5.01
50		2.31	1.90	2.78	2.06	5.21	4.78	5.67	4.95	3.31	3.88	4.13	3.81	5.18	4.99	5.45	5.13
		$\alpha = 10\%, d = 2, p = 2$															
5		10.5	9.23	9.96	11.6	10.8	9.33	10.2	11.6	8.55	8.65	10.8	9.99	8.64	8.68	11.0	10.2
10		8.25	8.55	8.87	7.52	9.09	9.30	9.52	8.04	8.40	9.22	8.59	9.54	9.18	9.67	8.71	9.92
15		6.83	7.77	8.19	8.99	8.11	8.74	8.99	9.67	11.5	9.69	9.08	8.16	11.8	10.0	9.97	8.73
20		7.30	8.46	6.61	5.67	9.54	10.3	7.99	6.60	7.38	9.43	8.13	9.64	8.13	10.1	8.97	10.5
25		10.2	7.47	8.98	8.12	12.6	9.56	10.3	9.42	7.47	9.14	10.2	8.95	8.43	10.2	10.8	9.26
30		6.18	8.04	6.83	6.80	9.68	10.96	9.02	8.95	9.20	8.91	10.89	8.75	10.7	9.57	11.6	9.89
35		7.11	5.86	7.56	5.37	10.0	8.62	10.5	7.45	7.23	7.37	7.60	7.15	8.23	8.98	8.46	8.38
40		5.90	5.81	5.75	6.83	9.26	9.62	9.39	9.96	5.55	8.05	6.37	8.35	7.57	9.93	8.09	10.5
45		6.58	6.32	5.04	6.30	9.99	11.0	8.46	10.0	7.17	8.28	6.44	7.76	9.64	11.0	8.31	10.3
50		5.25	4.86	5.91	5.61	10.1	9.67	9.90	9.14	6.85	7.52	7.01	7.33	9.02	10.4	9.15	10.2

## 4.5. CONCLUSION

Dans ce chapitre, nous avons étudié les modèles de séries chronologiques multivariés périodiques et cointégrés. Plus précisément, il a été question d'estimer le modèle et de faire des tests de type diagnostics pour valider les résultats obtenus. Dans l'estimation, nous avons considéré le modèle de plein rang et le modèle de rang réduit. L'estimation du modèle de rang réduit est une procédure non linéaire

et itérative dont les valeurs initiales sont les estimateurs de plein rang. Les distributions asymptotiques de ces estimateurs ont été aussi obtenues, affichant le taux de convergence usuel de  $N^{-1}$  pour le vecteur de cointégration et  $N^{-1/2}$  pour les autres paramètres.

Pour diagnostiquer le modèle, un test statistique de type portemanteau a également été présenté. Ce test a été étudié par des simulations. D'après les résultats obtenus, ce test statistique a de très bonnes performances et peut s'avérer incontournable pour diagnostiquer des modèles périodiques cointégrés.



## ANNEXES

**Annexe A.** *Preuve de la Proposition 4.1*

Sans nuire à la généralité, la preuve sera faite pour  $p = 3$ . Pour  $t = sn + \nu$ , nous avons d'après l'équation (4.3.1) :

$$\begin{aligned}
Y_t &= \phi_1(\nu)Y_{t-1} + \phi_2(\nu)Y_{t-2} + \phi_3(\nu)Y_{t-3} + \varepsilon_t, \\
&= \phi_1(\nu)Y_{t-1} + \phi_2(\nu)Y_{t-2} + \phi_3(\nu)\{\Theta(\nu-2)\}^{-1}\{\Theta(\nu-2)Y_{t-3} - Y_{t-2} + Y_{t-2}\} \\
&\quad + \varepsilon_t, \\
&= \phi_1(\nu)Y_{t-1} + [\phi_2(\nu) + \phi_3(\nu)\{\Theta(\nu-2)\}^{-1}]Y_{t-2} \\
&\quad - \phi_3(\nu)\{\Theta(\nu-2)\}^{-1}\{Y_{t-2} - \Theta(\nu-2)Y_{t-3}\} + \varepsilon_t, \\
&= \phi_1(\nu)Y_{t-1} + \{\phi_2(\nu) - \phi_2^*(\nu)\}Y_{t-2} + \phi_2^*(\nu)\{Y_{t-2} - \Theta(\nu-2)Y_{t-3}\} + \varepsilon_t, \\
&= \phi_1(\nu)Y_{t-1} + \{\phi_2(\nu) - \phi_2^*(\nu)\}\{\Theta(\nu-1)\}^{-1}\{\Theta(\nu-1)Y_{t-2} - Y_{t-1} + Y_{t-1}\} \\
&\quad + \phi_2^*(\nu)\{Y_{t-2} - \Theta(\nu-2)Y_{t-3}\} + \varepsilon_t, \\
&= [\phi_1(\nu) + \{\phi_2(\nu) - \phi_2^*(\nu)\}\{\Theta(\nu-1)\}^{-1}]Y_{t-1} \\
&\quad - \{\phi_2(\nu) - \phi_2^*(\nu)\}\{\Theta(\nu-1)\}^{-1}\{Y_{t-1} - \Theta(\nu-1)Y_{t-2}\} \\
&\quad + \phi_2^*(\nu)\{Y_{t-2} - \Theta(\nu-2)Y_{t-3}\} + \varepsilon_t, \\
&= \{\phi_1(\nu) - \phi_1^*(\nu)\}Y_{t-1} + \phi_1^*(\nu)\{Y_{t-1} - \Theta(\nu-1)Y_{t-2}\} \\
&\quad + \phi_2^*(\nu)\{Y_{t-2} - \Theta(\nu-2)Y_{t-3}\} + \varepsilon_t.
\end{aligned}$$

L'expression ci-dessus mène à :

$$\begin{aligned}
Y_t - \Theta(\nu)Y_{t-1} &= \{\phi_1(\nu) - \Theta(\nu) - \phi_1^*(\nu)\}Y_{t-1} + \phi_1^*(\nu)\{Y_{t-1} - \Theta(\nu-1)Y_{t-2}\} \\
&\quad + \phi_2^*(\nu)\{Y_{t-2} - \Theta(\nu-2)Y_{t-3}\} + \varepsilon_t,
\end{aligned}$$

où

$$\begin{aligned}
\phi_2^*(\nu) &= -\phi_3(\nu)\{\Theta(\nu-2)\}^{-1}, \\
\phi_1^*(\nu) &= \{\phi_2(\nu) - \phi_2^*(\nu)\}\{\Theta(\nu-1)\}^{-1}, \\
\pi(\nu) &= \phi_1(\nu) - \Theta(\nu) - \phi_1^*(\nu).
\end{aligned}$$

Ce qui constitue le résultat recherché. La preuve dans le cas général suit exactement le même genre d'arguments.

**Annexe B. Preuve de la Proposition 4.3**

Sans nuire à la généralité, elle sera faite pour  $p = 7$  et  $s = 4$ .

$$\begin{pmatrix} e_{1,n} \\ e_{2,n} \\ e_{3,n} \\ e_{4,n} \end{pmatrix} = \begin{pmatrix} \pi(1)Y_{4,n-1} + \phi_1^*(1)e_{4,n-1} + \phi_2^*(1)e_{3,n-1} + \phi_3^*(1)e_{2,n-1} + \phi_4^*(1)e_{1,n-1} + \phi_5^*(1)e_{4,n-2} + \\ \pi(2)Y_{1,n} + \phi_1^*(2)e_{1,n} + \phi_2^*(2)e_{4,n-1} + \phi_3^*(2)e_{3,n-1} + \phi_4^*(2)e_{2,n-1} + \phi_5^*(2)e_{1,n-1} + \\ \pi(3)Y_{2,n} + \phi_1^*(3)e_{2,n} + \phi_2^*(3)e_{1,n} + \phi_3^*(3)e_{4,n-1} + \phi_4^*(3)e_{3,n-1} + \phi_5^*(3)e_{2,n-1} + \\ \pi(4)Y_{3,n} + \phi_1^*(4)e_{3,n} + \phi_2^*(4)e_{2,n} + \phi_3^*(4)e_{1,n} + \phi_4^*(4)e_{4,n-1} + \phi_5^*(4)e_{3,n-1} + \\ + \phi_6^*(1)e_{3,n-2} \\ + \phi_6^*(2)e_{4,n-2} \\ + \phi_6^*(3)e_{1,n-1} \\ + \phi_6^*(4)e_{2,n-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,n} \\ \varepsilon_{2,n} \\ \varepsilon_{3,n} \\ \varepsilon_{4,n} \end{pmatrix},$$

$$= \begin{pmatrix} \pi(1)Y_{4,n-1} \\ \pi(2)Y_{1,n} \\ \pi(3)Y_{2,n} \\ \pi(4)Y_{3,n} \end{pmatrix} + \begin{pmatrix} \phi_1^*(1)e_{4,n-1} \\ \phi_1^*(2)e_{1,n} \\ \phi_1^*(3)e_{2,n} \\ \phi_1^*(4)e_{3,n} \end{pmatrix} + \begin{pmatrix} \phi_2^*(1)e_{3,n-1} \\ \phi_2^*(2)e_{4,n-1} \\ \phi_2^*(3)e_{1,n} \\ \phi_2^*(4)e_{2,n} \end{pmatrix} + \dots + \begin{pmatrix} \phi_5^*(1)e_{4,n-2} \\ \phi_5^*(2)e_{1,n-1} \\ \phi_5^*(3)e_{2,n-1} \\ \phi_5^*(4)e_{3,n-1} \end{pmatrix} \\ + \begin{pmatrix} \phi_6^*(1)e_{3,n-2} \\ \phi_6^*(2)e_{4,n-2} \\ \phi_6^*(3)e_{1,n-1} \\ \phi_6^*(4)e_{2,n-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,n} \\ \varepsilon_{2,n} \\ \varepsilon_{3,n} \\ \varepsilon_{4,n} \end{pmatrix},$$

$$= \text{diag} \{ \pi(\nu) \} \begin{pmatrix} Y_{4,n-1} \\ Y_{1,n} \\ Y_{2,n} \\ Y_{3,n} \end{pmatrix} + \Gamma_1 \begin{pmatrix} e_{4,n-1} \\ e_{1,n} \\ e_{2,n} \\ e_{3,n} \end{pmatrix} + \Gamma_2 \begin{pmatrix} e_{3,n-1} \\ e_{4,n-1} \\ e_{1,n} \\ e_{2,n} \end{pmatrix} + \dots + \Gamma_5 \begin{pmatrix} e_{4,n-2} \\ e_{1,n-1} \\ e_{2,n-1} \\ e_{3,n-1} \end{pmatrix} \\ + \Gamma_6 \begin{pmatrix} e_{3,n-2} \\ e_{4,n-2} \\ e_{1,n-1} \\ e_{2,n-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,n} \\ \varepsilon_{2,n} \\ \varepsilon_{3,n} \\ \varepsilon_{4,n} \end{pmatrix},$$

$$\begin{aligned}
&= \text{diag} \{ \boldsymbol{\pi}(\nu) \} \begin{pmatrix} \mathbf{Y}_{4,n-1} \\ \mathbf{Y}_{1,n} \\ \mathbf{Y}_{2,n} \\ \mathbf{Y}_{3,n} \end{pmatrix} + \boldsymbol{\Gamma}_1 \mathbf{U}_{1:n} + \boldsymbol{\Gamma}_2 \mathbf{U}_{2:n} + \boldsymbol{\Gamma}_3 \mathbf{U}_{3:n} + \boldsymbol{\Gamma}_4 \mathbf{U}_{0:(n-1)} + \boldsymbol{\Gamma}_5 \mathbf{U}_{1:(n-1)} \\
&+ \boldsymbol{\Gamma}_6 \mathbf{U}_{2:(n-1)} + \begin{pmatrix} \boldsymbol{\varepsilon}_{1,n} \\ \boldsymbol{\varepsilon}_{2,n} \\ \boldsymbol{\varepsilon}_{3,n} \\ \boldsymbol{\varepsilon}_{4,n} \end{pmatrix}.
\end{aligned}$$

qui est le résultat cherché.

### Annexe C. Preuve de la proposition 4.4

La preuve sera faite pour  $p = 3$  et  $s = 4$ . Avec ces paramètres, le modèle (4.3.1) devient :

$$\mathbf{Y}_t = \phi_1(\nu) \mathbf{Y}_{t-1} + \phi_2(\nu) \mathbf{Y}_{t-2} + \phi_3(\nu) \mathbf{Y}_{t-3} + \boldsymbol{\varepsilon}_t,$$

qui admet la représentation en vecteur des saisons :

$$\boldsymbol{\Phi}_0 \mathbf{Y}_n^* = \boldsymbol{\Phi}_1 \mathbf{Y}_{n-1}^* + \boldsymbol{\varepsilon}_n^*,$$

avec :

$$\boldsymbol{\Phi}_0 = \begin{pmatrix} \mathbf{I}_d & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\phi_1(2) & \mathbf{I}_d & \mathbf{0} & \mathbf{0} \\ -\phi_2(3) & -\phi_1(3) & \mathbf{I}_d & \mathbf{0} \\ -\phi_3(4) & -\phi_2(4) & -\phi_1(4) & \mathbf{I}_d \end{pmatrix}$$

et

$$\boldsymbol{\Phi}_1 = \begin{pmatrix} \mathbf{0} & \phi_3(1) & \phi_2(1) & \phi_1(1) \\ \mathbf{0} & \mathbf{0} & \phi_3(2) & \phi_2(2) \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \phi_3(3) \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Définissons :

$$\Lambda_0 = \begin{pmatrix} I_d & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\Theta(2) & I_d & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\Theta(3) & I_d & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\Theta(4) & I_d \end{pmatrix}$$

et

$$\Lambda_1 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \Theta(1) \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

On a alors :

$$\Phi_0 Y_n^* = \Lambda_0 Y_n^* + (\Phi_0 - \Lambda_0) Y_n^* = \Phi_1 Y_{n-1}^* + \varepsilon_n^*.$$

ce qui entraîne,

$$\Lambda_0 Y_n^* = (\Lambda_0 - \Phi_0) Y_n^* + \Phi_1 Y_{n-1}^* + \varepsilon_n^*,$$

et

$$\Lambda_0 Y_n^* - \Lambda_1 Y_{n-1}^* = (\Lambda_0 - \Phi_0) Y_n^* + (\Phi_1 - \Lambda_1) Y_{n-1}^* + \varepsilon_n^*.$$

D'autre part,

$$\begin{aligned} \Lambda_0 - \Phi_0 &= \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \phi_1(2) - \Theta(2) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \phi_2(3) & \phi_1(3) - \Theta(3) & \mathbf{0} & \mathbf{0} \\ \phi_3(4) & \phi_2(4) & \phi_1(4) - \Theta(4) & \mathbf{0} \end{pmatrix}, \\ &= \mathbf{A}_1 + \mathbf{A}_2, \end{aligned}$$

où

$$\mathbf{A}_1 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \phi_1(2) - \Theta(2) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \phi_1(3) - \Theta(3) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \phi_1(4) - \Theta(4) & \mathbf{0} \end{pmatrix},$$

et

$$\mathbf{A}_2 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \phi_2(3) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \phi_3(4) & \phi_2(4) & \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Nous avons également :

$$\begin{aligned} \Phi_1 - \Lambda_1 &= \begin{pmatrix} \mathbf{0} & \phi_3(1) & \phi_2(1) & \phi_1(1) - \Theta(1) \\ \mathbf{0} & \mathbf{0} & \phi_3(2) & \phi_2(2) \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \phi_3(3) \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}, \\ &= \mathbf{A}_3 + \mathbf{A}_4, \end{aligned}$$

où

$$\mathbf{A}_3 = \begin{pmatrix} \mathbf{0} & \phi_3(1) & \phi_2(1) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \phi_3(2) & \phi_2(2) \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \phi_3(3) \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$$

et

$$\mathbf{A}_4 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \phi_1(1) - \Theta(1) \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

L'expression principale devient alors :

$$\begin{aligned} \Lambda_0 \mathbf{Y}_n^* - \Lambda_1 \mathbf{Y}_{n-1} &= (\mathbf{A}_1 + \mathbf{A}_2) \mathbf{Y}_n^* + (\mathbf{A}_3 + \mathbf{A}_4) \mathbf{Y}_{n-1}^* + \varepsilon_n^*, \\ &= (\mathbf{A}_1 \mathbf{Y}_n^* + \mathbf{A}_4 \mathbf{Y}_{n-1}^*) + (\mathbf{A}_2 \mathbf{Y}_n^* + \mathbf{A}_3 \mathbf{Y}_{n-1}^* + \varepsilon_n^*). \end{aligned}$$

On a encore :

$$\begin{aligned}
 A_2 Y_n^* + A_3 Y_{n-1}^* &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \phi_2(3) & 0 & 0 & 0 \\ \phi_3(4) & \phi_2(4) & 0 & 0 \end{pmatrix} Y_n^* \\
 &+ \begin{pmatrix} 0 & \phi_3(1) & \phi_2(1) & 0 \\ 0 & 0 & \phi_3(2) & \phi_2(2) \\ 0 & 0 & 0 & \phi_3(3) \\ 0 & 0 & 0 & 0 \end{pmatrix} Y_{n-1}^*, \\
 &= B_1 Y_n^* + B_2 Y_{n-1}^* + B_3 Y_n^* + B_4 Y_{n-1}^*,
 \end{aligned}$$

avec,

$$B_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \phi_2(3) & 0 & 0 & 0 \\ 0 & \phi_2(4) & 0 & 0 \end{pmatrix},$$

$$B_2 = \begin{pmatrix} 0 & 0 & \phi_2(1) & 0 \\ 0 & 0 & 0 & \phi_2(2) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$B_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \phi_3(4) & 0 & 0 & 0 \end{pmatrix},$$

et

$$B_4 = \begin{pmatrix} 0 & \phi_3(1) & 0 & 0 \\ 0 & 0 & \phi_3(2) & 0 \\ 0 & 0 & 0 & \phi_3(3) \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

On a également :

$$\begin{aligned}
 B_3 Y_n^* + B_4 Y_{n-1}^* &= \begin{pmatrix} \mathcal{M}_1 & 0 & 0 & 0 \\ 0 & \mathcal{M}_2 & 0 & 0 \\ 0 & 0 & \mathcal{M}_3 & 0 \\ 0 & 0 & 0 & \mathcal{M}_4 \end{pmatrix} \left[ \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -\Theta(2) & 0 & 0 & 0 \end{pmatrix} Y_n^* \right. \\
 &\quad \left. + \begin{pmatrix} 0 & -\Theta(3) & 0 & 0 \\ 0 & 0 & -\Theta(4) & 0 \\ 0 & 0 & 0 & -\Theta(1) \\ 0 & 0 & 0 & 0 \end{pmatrix} Y_{n-1}^* \right]
 \end{aligned}$$

où

$$\begin{aligned}
 \mathcal{M}_1 &= -\phi_3(1) \{\Theta(3)\}^{-1}, \\
 \mathcal{M}_2 &= -\phi_3(2) \{\Theta(4)\}^{-1}, \\
 \mathcal{M}_3 &= -\phi_3(3) \{\Theta(1)\}^{-1}, \\
 \mathcal{M}_4 &= -\phi_3(4) \{\Theta(2)\}^{-1}.
 \end{aligned}$$

qui peut encore s'écrire :

$$\begin{aligned}
 &\text{diag} [-\phi_3(\nu) \{\Theta(\nu-2)\}^{-1}] \left[ \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ I_d & 0 & 0 & 0 \\ -\Theta(2) & I_d & 0 & 0 \end{pmatrix} Y_n^* \right. \\
 &\quad \left. + \begin{pmatrix} 0 & -\Theta(3) & I_d & 0 \\ 0 & 0 & -\Theta(4) & I_d \\ 0 & 0 & 0 & -\Theta(1) \\ 0 & 0 & 0 & 0 \end{pmatrix} Y_{n-1}^* - C_1 Y_n^* - C_2 Y_{n-1}^* \right],
 \end{aligned}$$

où

$$\mathbf{C}_1 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_d & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_d & \mathbf{0} & \mathbf{0} \end{pmatrix},$$

et

$$\mathbf{C}_2 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{I}_d & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_d \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Nous avons :

$$\begin{aligned} \mathbf{A}_2 \mathbf{Y}_n^* + \mathbf{A}_3 \mathbf{Y}_{n-1}^* &= \mathbf{B}_1 \mathbf{Y}_n^* + \mathbf{B}_2 \mathbf{Y}_{n-1}^* \\ &+ \text{diag} [\phi_3(\nu) \{\Theta(\nu-2)\}^{-1}, \nu = 3, 4, 1, 2] (\mathbf{C}_1 \mathbf{Y}_n^* + \mathbf{C}_2 \mathbf{Y}_{n-1}^*) \\ &+ \Gamma_2 \mathbf{U}_{2:n}, \end{aligned}$$

ce qui est encore égal à :

$$\begin{aligned} &= \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \phi_2(3) + \phi_3(3) \{\Theta(1)\}^{-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \phi_2(4) + \phi_3(4) \{\Theta(2)\}^{-1} & \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{Y}_n^* \\ &+ \begin{pmatrix} \mathbf{0} & \mathbf{0} & \phi_2(1) + \phi_3(1) \{\Theta(3)\}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \phi_2(2) + \phi_3(2) \{\Theta(4)\}^{-1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{Y}_{n-1}^* \\ &+ \Gamma_2 \mathbf{U}_{2:n}, \end{aligned}$$

ce qui peut encore se mettre sous la forme,

$$\text{diag} \{\mathcal{M}(\nu)\} \left[ \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_d & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\Theta(2) & \mathbf{I}_d & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\Theta(3) & \mathbf{I}_d & \mathbf{0} \end{pmatrix} \mathbf{Y}_n^* \right]$$



$$+ \begin{pmatrix} \mathbf{0} & \mathbf{0} & -\Theta(4) & I_d \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\Theta(1) \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \left[ Y_{n-1}^* - C_3 Y_n - C_4 Y_{n-1}^* \right] + \Gamma_2 U_{2:n},$$

où,

$$\mathcal{M}(\nu) = -[\phi_2(\nu) + \phi_3(\nu) \{\Theta(\nu-2)\}^{-1}] \{\Theta(\nu-1)\}^{-1},$$

$$C_3 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ I_d & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_d & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_d & \mathbf{0} \end{pmatrix},$$

$$C_4 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & I_d \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Ceci mène à l'expression suivante :

$$A_2 Y_n^* + A_3 Y_{n-1}^* = \Gamma_1 U_{1:n} + \Gamma_2 U_{2:n} - [\text{diag} \{\phi_1^*(\nu)\}] (C_3 Y_n^* + C_4 Y_{n-1}^*).$$

En posant,

$$U_{0:n} = \Lambda_0 Y_n^* - \Lambda_1 Y_{n-1}^* = (A_1 Y_n^* + A_4 Y_{n-1}^*) + (A_2 Y_n^* + A_3 Y_{n-1}^*) + \varepsilon_n^*,$$

où

$$\mathbf{U}_{0:n} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \phi_1(2) - \Theta(2) - \phi_1^*(2) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \phi_1(3) - \Theta(3) - \phi_1^*(3) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \phi_1(4) - \Theta(4) - \phi_1^*(4) & \mathbf{0} \end{pmatrix} \mathbf{Y}_n^* \\ + \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \phi_1(1) - \Theta(1) - \phi_1^*(1) \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{Y}_{n-1}^*$$

$$+ \Gamma_1 \mathbf{U}_{1:n} + \Gamma_2 \mathbf{U}_{2:n} + \boldsymbol{\varepsilon}_n^*$$

On obtient en fin de compte le résultat cherché.

$$\mathbf{U}_{0:n} = \text{diag} \{ \pi(\nu), \nu = 1, 2, 3, 4 \} \begin{pmatrix} \mathbf{Y}_{4,n-1} \\ \mathbf{Y}_{1,n} \\ \mathbf{Y}_{2,n} \\ \mathbf{Y}_{3,n} \end{pmatrix} + \Gamma_1 \mathbf{U}_{1:n} + \Gamma_2 \mathbf{U}_{2:n} + \boldsymbol{\varepsilon}_n^*.$$

#### Annexe D. Preuve de la proposition 4.5

La démonstration sera faite pour  $p = 7$  et  $s = 4$ , elle découle de celle de la proposition 4.3, et plus précisément de l'expression suivante :

$$\begin{pmatrix} \mathbf{e}_{1,n} \\ \mathbf{e}_{2,n} \\ \mathbf{e}_{3,n} \\ \mathbf{e}_{4,n} \end{pmatrix} = \begin{pmatrix} \pi(1) \mathbf{Y}_{4,n-1} + \phi_1^*(1) \mathbf{e}_{4,n-1} + \phi_2^*(1) \mathbf{e}_{3,n-1} + \phi_3^*(1) \mathbf{e}_{2,n-1} + \phi_4^*(1) \mathbf{e}_{1,n-1} + \phi_5^*(1) \mathbf{e}_{4,n-2} + \\ \pi(2) \mathbf{Y}_{1,n} + \phi_1^*(2) \mathbf{e}_{1,n} + \phi_2^*(2) \mathbf{e}_{4,n-1} + \phi_3^*(2) \mathbf{e}_{3,n-1} + \phi_4^*(2) \mathbf{e}_{2,n-1} + \phi_5^*(2) \mathbf{e}_{1,n-1} + \\ \pi(3) \mathbf{Y}_{2,n} + \phi_1^*(3) \mathbf{e}_{2,n} + \phi_2^*(3) \mathbf{e}_{1,n} + \phi_3^*(3) \mathbf{e}_{4,n-1} + \phi_4^*(3) \mathbf{e}_{3,n-1} + \phi_5^*(3) \mathbf{e}_{2,n-1} + \\ \pi(4) \mathbf{Y}_{3,n} + \phi_1^*(4) \mathbf{e}_{3,n} + \phi_2^*(4) \mathbf{e}_{2,n} + \phi_3^*(4) \mathbf{e}_{1,n} + \phi_4^*(4) \mathbf{e}_{4,n-1} + \phi_5^*(4) \mathbf{e}_{3,n-1} + \\ + \phi_6^*(1) \mathbf{e}_{3,n-2} \\ + \phi_6^*(2) \mathbf{e}_{4,n-2} \\ + \phi_6^*(3) \mathbf{e}_{1,n-1} \\ + \phi_6^*(4) \mathbf{e}_{2,n-1} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_{1,n} \\ \boldsymbol{\varepsilon}_{2,n} \\ \boldsymbol{\varepsilon}_{3,n} \\ \boldsymbol{\varepsilon}_{4,n} \end{pmatrix}.$$

Il suffit d'aller dans cette expression et de regrouper suivant les vecteurs

$$\begin{pmatrix} \mathbf{e}_{1,n} \\ \mathbf{e}_{2,n} \\ \mathbf{e}_{3,n} \\ \mathbf{e}_{4,n} \end{pmatrix}, \quad \text{ceci pour } n = 1, \dots, N, \text{ en remarquant au passage que}$$

$$\begin{pmatrix} \mathbf{e}_{1,n} \\ \mathbf{e}_{2,n} \\ \mathbf{e}_{3,n} \\ \mathbf{e}_{4,n} \end{pmatrix} = (\mathbf{I}_{Sd} - \Lambda(B)) \mathbf{Y}_n^* \quad \text{et aussi que} \quad \begin{pmatrix} \pi(1) Y_{4,n-1} \\ \pi(2) Y_{1,n} \\ \pi(3) Y_{2,n} \\ \pi(4) Y_{3,n} \end{pmatrix} = \Pi(B) \mathbf{Y}_n^*,$$

cette expression devient alors :

$$\begin{pmatrix} \mathbf{e}_{1,n} \\ \mathbf{e}_{2,n} - \phi_1^*(2) \mathbf{e}_{1,n} \\ \mathbf{e}_{3,n} - \phi_1^*(3) \mathbf{e}_{2,n} - \phi_2^*(3) \mathbf{e}_{1,n} \\ \mathbf{e}_{4,n} - \phi_1^*(4) \mathbf{e}_{3,n} - \phi_2^*(4) \mathbf{e}_{2,n} - \phi_3^*(4) \mathbf{e}_{1,n} \end{pmatrix} = \begin{pmatrix} \phi_4^*(1) \mathbf{e}_{1,n-1} + \phi_3^*(1) \mathbf{e}_{2,n-1} + \phi_2^*(1) \mathbf{e}_{3,n-1} + \phi_1^*(1) \mathbf{e}_{4,n-1} \\ \phi_5^*(2) \mathbf{e}_{1,n-1} + \phi_4^*(2) \mathbf{e}_{2,n-1} + \phi_3^*(2) \mathbf{e}_{3,n-1} + \phi_2^*(2) \mathbf{e}_{4,n-1} \\ \phi_6^*(3) \mathbf{e}_{1,n-1} + \phi_5^*(3) \mathbf{e}_{2,n-1} + \phi_4^*(3) \mathbf{e}_{3,n-1} + \phi_3^*(3) \mathbf{e}_{4,n-1} \\ + \phi_6^*(4) \mathbf{e}_{2,n-1} + \phi_5^*(4) \mathbf{e}_{3,n-1} + \phi_4^*(4) \mathbf{e}_{4,n-1} \end{pmatrix} \\ - \begin{pmatrix} \phi_6^*(1) \mathbf{e}_{3,n-2} + \phi_5^*(1) \mathbf{e}_{4,n-2} \\ \phi_6^*(2) \mathbf{e}_{4,n-2} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \pi(1) Y_{4,n-1} \\ \pi(2) Y_{1,n} \\ \pi(3) Y_{2,n} \\ \pi(4) Y_{3,n} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,n} \\ \varepsilon_{2,n} \\ \varepsilon_{3,n} \\ \varepsilon_{4,n} \end{pmatrix},$$

qui peut se mettre sous la forme matricielle :

$$\begin{pmatrix} \mathbf{I}_d & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\phi_1^*(2) & \mathbf{I}_d & \mathbf{0} & \mathbf{0} \\ -\phi_2^*(3) & -\phi_1^*(3) & \mathbf{I}_d & \mathbf{0} \\ -\phi_3^*(4) & -\phi_2^*(4) & -\phi_1^*(4) & \mathbf{I}_d \end{pmatrix} \begin{pmatrix} \mathbf{e}_{1,n} \\ \mathbf{e}_{2,n} \\ \mathbf{e}_{3,n} \\ \mathbf{e}_{4,n} \end{pmatrix} = \begin{pmatrix} \phi_4^*(1) & \phi_3^*(1) & \phi_2^*(1) & \phi_1^*(1) \\ \phi_5^*(2) & \phi_4^*(2) & \phi_3^*(2) & \phi_2^*(2) \\ \phi_6^*(3) & \phi_5^*(3) & \phi_4^*(3) & \phi_3^*(3) \\ \mathbf{0} & \phi_6^*(4) & \phi_5^*(4) & \phi_4^*(4) \end{pmatrix} \begin{pmatrix} \mathbf{e}_{1,n-1} \\ \mathbf{e}_{2,n-1} \\ \mathbf{e}_{3,n-1} \\ \mathbf{e}_{4,n-1} \end{pmatrix} \\ - \begin{pmatrix} \mathbf{0} & \mathbf{0} & \phi_6^*(1) & \phi_5^*(1) \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \phi_6^*(2) \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{e}_{1,n-2} \\ \mathbf{e}_{2,n-2} \\ \mathbf{e}_{3,n-2} \\ \mathbf{e}_{4,n-2} \end{pmatrix} \\ = \begin{pmatrix} \pi(1) Y_{4,n-1} \\ \pi(2) Y_{1,n} \\ \pi(3) Y_{2,n} \\ \pi(4) Y_{3,n} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,n} \\ \varepsilon_{2,n} \\ \varepsilon_{3,n} \\ \varepsilon_{4,n} \end{pmatrix}.$$

On obtient alors le résultat cherché avec :

$$\phi_0^* = \begin{pmatrix} I_d & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\phi_1^*(2) & I_d & \mathbf{0} & \mathbf{0} \\ -\phi_2^*(3) & -\phi_1^*(3) & I_d & \mathbf{0} \\ -\phi_3^*(4) & -\phi_2^*(4) & -\phi_1^*(4) & I_d \end{pmatrix}, \phi_1^* = - \begin{pmatrix} \phi_4^*(1) & \phi_3^*(1) & \phi_2^*(1) & \phi_1^*(1) \\ \phi_5^*(2) & \phi_4^*(2) & \phi_3^*(2) & \phi_2^*(2) \\ \phi_6^*(3) & \phi_5^*(3) & \phi_4^*(3) & \phi_3^*(3) \\ \mathbf{0} & \phi_6^*(4) & \phi_5^*(4) & \phi_4^*(4) \end{pmatrix},$$

$$\phi_2^* = - \begin{pmatrix} \mathbf{0} & \mathbf{0} & \phi_6^*(1) & \phi_5^*(1) \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \phi_6^*(2) \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

**Annexe E.** *Preuve du lemme 4.1 i) Nous avons :*

$$\mathbf{Z}_{m:(sn+\nu-1)} = \sum_{k=0}^n \boldsymbol{\psi}_{\nu-1}(B) \mathbf{A}_k = \sum_{k=0}^n \boldsymbol{\psi}_{\nu-1}(B) \mathbf{Q} \boldsymbol{\varepsilon}_k^*, \text{ avec } \boldsymbol{\psi}_{\nu-1}(B) = \sum_{j=0}^{\infty} \boldsymbol{\psi}_{\nu-1,j} B^j.$$

D'autre part,  $\sum_{k=0}^n \boldsymbol{\psi}_{\nu-1}(B) \mathbf{Q} \boldsymbol{\varepsilon}_k^* = \boldsymbol{\psi}_{\nu-1}(1) \mathbf{Q} \sum_{l=1}^n \boldsymbol{\varepsilon}_l^* + \boldsymbol{\xi}_n - \boldsymbol{\xi}_0$ , où  $\boldsymbol{\xi}_n = \sum_{l=0}^{\infty} \boldsymbol{\omega}_l \boldsymbol{\varepsilon}_{n-l}^*$ ,  $\boldsymbol{\xi}_0 = \sum_{l=0}^{\infty} \boldsymbol{\sigma}_l \boldsymbol{\varepsilon}_{-l}^*$  pour  $\boldsymbol{\omega}_l = - \sum_{k=1}^{\infty} \boldsymbol{\psi}_{\nu-1,l+k} \mathbf{Q}$  et  $\boldsymbol{\sigma}_l = - \sum_{k=0}^{\infty} \boldsymbol{\psi}_{\nu-1,l+k} \mathbf{Q}$ . Les suites  $\{\boldsymbol{\omega}_l\}_{l=0}^{\infty}$  et  $\{\boldsymbol{\sigma}_l\}_{l=0}^{\infty}$  sont absolument sommables d'après la remarque 2.

Posons  $\boldsymbol{\zeta}_n = \sum_{l=1}^n \boldsymbol{\varepsilon}_l^*$ , nous avons :

$$\begin{aligned} N^{-2} \sum_{n=1}^N \mathbf{Z}_{m:(sn+\nu-1)} \mathbf{Z}_{m:(sn+\nu-1)}^\top &= \{\boldsymbol{\psi}_{\nu-1}(1) \mathbf{Q}\} \left( N^{-2} \sum_{n=1}^N \boldsymbol{\zeta}_n \boldsymbol{\zeta}_n^\top \right) \{\boldsymbol{\psi}_{\nu-1}(1) \mathbf{Q}\}^\top \\ &+ \{\boldsymbol{\psi}_{\nu-1}(1) \mathbf{Q}\} \left\{ N^{-2} \sum_{n=1}^N \boldsymbol{\zeta}_n (\boldsymbol{\xi}_n - \boldsymbol{\xi}_0)^\top \right\} \\ &+ \left\{ N^{-2} \sum_{n=1}^N (\boldsymbol{\xi}_n - \boldsymbol{\xi}_0) \boldsymbol{\zeta}_n^\top \right\} \{\boldsymbol{\psi}_{\nu-1}(1) \mathbf{Q}\}^\top \\ &+ \left\{ N^{-2} \sum_{n=1}^N (\boldsymbol{\xi}_n - \boldsymbol{\xi}_0) (\boldsymbol{\xi}_n - \boldsymbol{\xi}_0)^\top \right\}. \end{aligned}$$

D'après le Phillips et Durlauf (1986, Lemme 3.1) ou aussi Hamilton (1994, proposition 18.1),

$$N^{-2} \sum_{n=1}^N \boldsymbol{\zeta}_n \boldsymbol{\zeta}_n^\top \xrightarrow{\mathcal{L}} \boldsymbol{\Sigma}_\varepsilon^{1/2} \left\{ \int_0^1 \boldsymbol{W}(r) \boldsymbol{W}^\top(r) dr \right\} \boldsymbol{\Sigma}_\varepsilon^{1/2},$$

où  $\boldsymbol{W}(r)$  est un mouvement brownien standard d'ordre  $sd$ .

En utilisant la norme  $\|\mathbf{A}\| = [\text{tr}(\mathbf{A}^\top \mathbf{A})]^{1/2}$ , nous avons :

$$\begin{aligned}
E \left( \left\| N^{-2} \sum_{n=1}^N \xi_n \zeta_n^\top \right\| \right) &\leq N^{-2} E \left( \sum_{n=1}^N \|\xi_n\| \|\zeta_n^\top\| \right), \\
&\leq N^{-2} E \left\{ \left( \sum_{n=1}^N \|\xi_n\|^2 \right)^{1/2} \left( \sum_{n=1}^N \|\zeta_n^\top\|^2 \right)^{1/2} \right\}, \\
&\leq N^{-2} \left\{ E \left( \sum_{n=1}^N \|\xi_n\|^2 \right) \right\}^{1/2} \left\{ E \left( \sum_{n=1}^N \|\zeta_n^\top\|^2 \right) \right\}^{1/2},
\end{aligned}$$

or d'après Sims et al. (1990, Lemme 1), nous avons :

$$E \left( \sum_{n=1}^N \|\xi_n\|^2 \right) = O(N). \text{ D'après Phillips et Durlauf (1986, lemme 3.1),}$$

$$E \left( \sum_{n=1}^N \|\zeta_n^\top\|^2 \right) = O(N^2).$$

Par conséquent,

$$\begin{aligned}
E \left( \left\| N^{-2} \sum_{n=1}^N \xi_n \zeta_n^\top \right\| \right) &\leq N^{-2} (O(N))^{1/2} (O(N^2))^{1/2} = N^{-2} O(N^{1/2}) O(N), \\
&= N^{-2} O(N^{3/2}), \\
&= O(N^{-1/2}), \\
&= o(1).
\end{aligned}$$

En utilisant toujours le fait que  $E \left( \sum_{n=1}^N \|\xi_n\|^2 \right) = O(N)$  et  $E \left( \sum_{n=1}^N \|\zeta_n^\top\|^2 \right) = O(N^2)$ , nous montrons que :

$$\begin{aligned}
E \left( \left\| \sum_{n=1}^N \zeta_n (\xi_n - \xi_0)^\top \right\| \right) &= O(N^{3/2}), \\
E \left( \left\| \sum_{n=1}^N (\xi_n - \xi_0) (\xi_n - \xi_0)^\top \right\| \right) &= O(N).
\end{aligned}$$

Ceci permet d'avoir en fin de compte :

$$N^{-2} \sum_{n=1}^N \mathbf{Z}_{m:(t-1)} \mathbf{Z}_{m:(t-1)}^\top \xrightarrow{\mathcal{L}} \{\boldsymbol{\psi}_{\nu-1}(1)\mathbf{Q}\} \boldsymbol{\Sigma}_\varepsilon^{1/2} \left\{ \int_0^1 \boldsymbol{\omega}(r) \boldsymbol{\omega}^\top(r) dr \right\} \boldsymbol{\Sigma}_\varepsilon^{1/2} \{\boldsymbol{\psi}_{\nu-1}(1)\mathbf{Q}\}^\top.$$

où  $t = sn + \nu$

ii) Nous avons :

$$\mathbf{a}_{sn+\nu} = D_\nu \mathbf{Q} \boldsymbol{\varepsilon}_n^*,$$

et

$$\mathbf{Z}_{m:(sn+\nu-1)} = \sum_{k=0}^{n-1} \boldsymbol{\psi}_{\nu-1}(B) \mathbf{Q} \boldsymbol{\varepsilon}_{n-k}^* = \boldsymbol{\psi}_{\nu-1}(1) \mathbf{Q} \boldsymbol{\zeta}_n + \boldsymbol{\xi}_n - \boldsymbol{\xi}_0,$$

où  $\boldsymbol{\zeta}_n = \sum_{s=1}^n \boldsymbol{\varepsilon}_s^*$ .

Avec ceci, on obtient que :

$$\begin{aligned} N^{-1} \sum_{n=1}^N \mathbf{a}_{sn+\nu} \mathbf{Z}_{m:(sn+\nu-1)}^\top &= (\mathbf{D}_\nu \mathbf{Q}) \left\{ N^{-1} \sum_{n=1}^N \boldsymbol{\varepsilon}_n^* \boldsymbol{\zeta}_n^\top \right\} \{ \boldsymbol{\psi}_{\nu-1}(1) \mathbf{Q} \}^\top \\ &\quad + (\mathbf{D}_\nu \mathbf{Q}) N^{-1} \sum_{n=1}^N \boldsymbol{\varepsilon}_n^* (\boldsymbol{\xi}_n - \boldsymbol{\xi}_0)^\top. \end{aligned}$$

D'autre part,

$$\begin{aligned} N^{-1} \sum_{n=1}^N \boldsymbol{\varepsilon}_n^* \boldsymbol{\zeta}_n^\top &= N^{-1} \sum_{n=1}^N \boldsymbol{\varepsilon}_n^* (\boldsymbol{\zeta}_{n-1}^\top + \boldsymbol{\varepsilon}_n^{*\top}), \\ &= N^{-1} \sum_{n=1}^N \boldsymbol{\varepsilon}_n^* \boldsymbol{\zeta}_{n-1}^\top + N^{-1} \sum_{n=1}^N \boldsymbol{\varepsilon}_n^* \boldsymbol{\varepsilon}_n^{*\top}. \end{aligned}$$

D'après Philips et Durlauf (1986, Lemme 3.1),

$$N^{-1} \sum_{n=1}^N \boldsymbol{\varepsilon}_n^* \boldsymbol{\zeta}_{n-1}^\top \xrightarrow{\mathcal{L}} \boldsymbol{\Sigma}_\varepsilon^{1/2} \left\{ \int_0^1 \boldsymbol{w}(r) d\boldsymbol{w}^\top(r) \right\}^\top \boldsymbol{\Sigma}_\varepsilon^{1/2}.$$

D'après la loi des grands nombres,

$$N^{-1} \sum_{n=1}^N \boldsymbol{\varepsilon}_n^* \boldsymbol{\varepsilon}_n^{*\top} \xrightarrow{\mathcal{P}} \boldsymbol{\Sigma}_\varepsilon.$$

En fin de compte,

$$N^{-1} \sum_{n=1}^N \boldsymbol{\varepsilon}_n^* \boldsymbol{\zeta}_n^\top \xrightarrow{\mathcal{L}} \boldsymbol{\Sigma}_\varepsilon^{1/2} \left\{ \int_0^1 \boldsymbol{w}(r) d\boldsymbol{w}^\top(r) \right\}^\top \boldsymbol{\Sigma}_\varepsilon^{1/2} + \boldsymbol{\Sigma}_\varepsilon.$$

On a aussi :

$$\begin{aligned} N^{-1} \sum_{n=1}^N \boldsymbol{\varepsilon}_n^* \boldsymbol{\xi}_n^\top &= N^{-1} \sum_{n=1}^N \boldsymbol{\varepsilon}_n^* (\boldsymbol{\omega}_0 \boldsymbol{\varepsilon}_n^* + \boldsymbol{\xi}_{n-1}^*)^\top, \\ &= N^{-1} \sum_{n=1}^N \boldsymbol{\varepsilon}_n^* \boldsymbol{\varepsilon}_n^{*\top} \boldsymbol{\omega}_0^\top + N^{-1} \sum_{n=1}^N \boldsymbol{\varepsilon}_n^* \boldsymbol{\xi}_{n-1}^{*\top}, \end{aligned}$$

$$\text{où } \xi_n^* = \sum_{h=1}^{\infty} \omega_s \varepsilon_{n-h}^*.$$

D'autre part,  $\varepsilon_n^* \xi_n^{*\top} = \sum_{h=1}^{\infty} \omega_h \varepsilon_n^* \varepsilon_{n-h}^{*\top}$ . Étant donné que les cumulants d'ordre quatre du processus  $\{\varepsilon_n^*, n \in \mathbb{N}\}$  sont finis et que la série  $\sum_{s=0}^{\infty} \omega_s$  est absolument sommable, le processus  $\{\varepsilon_n^* \xi_n^{*\top}\}_{n=1}^N$  est une différence de martingale de variance finie par rapport à  $\sigma_n = \{\varepsilon_n^*, \varepsilon_{n-1}^*, \dots\}$ .

On en déduit que :

$$N^{-1} \sum_{n=1}^N \varepsilon_n^* \xi_n^{*\top} \xrightarrow{\mathcal{P}} \mathbf{0},$$

Voir Hamilton (1994, Exemple 7.11).

On obtient alors que :

$$\begin{aligned} N^{-1} \sum_{n=1}^N \mathbf{a}_{sn+\nu} \mathbf{Z}_{m:(sn+\nu-1)}^\top &\xrightarrow{\mathcal{L}} (D_\nu \mathbf{Q}) \Sigma_\varepsilon^{1/2} \left\{ \int_0^1 \mathbf{w}(r) d\mathbf{W}^\top(r) \right\}^\top \Sigma_\varepsilon^{1/2} \{\psi_{\nu-1}(1) \mathbf{Q}\}^\top, \\ &+ (D_\nu \mathbf{Q}) \Sigma_\varepsilon \{\psi_{\nu-1}(1) \mathbf{Q}\}^\top + (D_\nu \mathbf{Q}) \Sigma_\varepsilon \omega_0^\top \\ &= (D_\nu \mathbf{Q}) \Sigma_\varepsilon^{1/2} \left\{ \int_0^1 \mathbf{w}(r) d\mathbf{W}^\top(r) \right\}^\top \Sigma_\varepsilon^{1/2} \{\psi_{\nu-1}(1) \mathbf{Q}\}^\top \\ &+ (D_\nu \mathbf{Q}) \Sigma_\varepsilon (\psi_{\nu-1,0} \mathbf{Q})^\top. \end{aligned}$$

iii) On a  $\mathbf{U}_t = (\mathbf{Z}_{r:t}^\top, \mathbf{W}_t^\top, \dots, \mathbf{W}_{t-p+2}^\top)^\top$ . En multipliant l'équation (4.3.10) de part et d'autre par  $\mathbf{I}_{sd} - P_2 \Lambda_2(B) \mathbf{Q}_2$ , on obtient :

$$\{\mathbf{I}_{sd} - \Lambda(B)\} \mathbf{Y}_n^* = \{\mathbf{I}_{sd} - P_2 \Lambda_2(B) \mathbf{Q}_2\} \{\Phi^{**}(B)\}^{-1} \varepsilon_n^*,$$

ou de manière plus précise que :

$$\begin{pmatrix} Y_{1,n} - Y_{s,n-1} \\ Y_{2,n} - Y_{1,n} \\ Y_{3,n} - Y_{2,n} \\ \vdots \\ Y_{s,n} - Y_{s-1,n} \end{pmatrix} = \{\mathbf{I}_{sd} - P_2 \Lambda_2(B) \mathbf{Q}_2\} \{\Phi^{**}(B)\}^{-1} \varepsilon_n^*.$$

On obtient à partir de ceci que :

$$\mathbf{W}_t = \mathbf{Y}_t - \mathbf{Y}_{t-1} = \mathbf{Y}_{sn+\nu} - \mathbf{Y}_{sn+\nu-1},$$

$$\begin{aligned}
&= D_\nu^* \{I_{sd} - P_2 \Lambda_2(B) Q_2\} \{\Phi^{**}(B)\}^{-1} \epsilon_n^*, \\
&= G_\nu(B) \epsilon_n^*,
\end{aligned}$$

où  $D_\nu^*$  est une matrice d'ordre  $d \times sd$  constituée de  $s$  blocs matriciels disposés en lignes, les blocs valant la matrice identité d'ordre  $d$  à la position  $\nu$  et la matrice nulle ailleurs.

D'autre part, on a d'après l'équation (4.3.10) que :

$$Q_2 Y_n^* = Q_2 \{\Phi^{**}(B)\}^{-1} \epsilon_n^*.$$

Or,

$$Q_2 Y_n^* = \begin{pmatrix} Q_r Y_{s,n} \\ Q_r Y_{1,n} \\ \vdots \\ Q_r Y_{s-1,n} \end{pmatrix} = \begin{pmatrix} Z_{r:(sn+s)} \\ Z_{r:(sn+1)} \\ \vdots \\ Z_{r:(sn+s-1)} \end{pmatrix}.$$

Il apparaît à partir de ceci que pour  $t = sn + \nu$ ,  $Z_{r:t} = D_r^* Q_2 \{\Phi^{**}(B)\}^{-1} \epsilon_n^* = L_\nu(B) \epsilon_n^*$ , où  $D_r^*$  est une matrice d'ordre  $r \times rd$  constitué de  $s$  blocs de matrices d'ordre  $r \times r$  disposés en ligne. Les blocs valant la matrice identité d'ordre  $r$  à la position  $\nu + 1$  (modulo  $s$ ) et zéro ailleurs.

Nous avons alors :

$$U_t = \begin{pmatrix} Z_{r:t} \\ W_t \\ W_{t-1} \\ \vdots \\ W_{t-p+2} \end{pmatrix} = \begin{pmatrix} L_\nu(B) \\ G_\nu(B) \\ G_{\nu-1}(B) \\ \vdots \\ G_{\nu^*}(B) B^{p^*} \end{pmatrix} \epsilon_n^* = H_\nu(B) \epsilon_n^*,$$

où  $t - p + 2 = s(n - p^*) + \nu^*$ . D'autre part d'après la preuve (i), on a :

$$Z_{m:(sn+\nu-1)} = \psi_{\nu-1}(1) Q \zeta_n + \xi_n - \xi_0,$$

où  $\zeta_n = \sum_{s=1}^n \epsilon_n^*$ . et  $\xi_n - \xi_0 = \lambda_{\nu-1}(B) \epsilon_n^*$ .



Ceci donne :

$$N^{-1} \sum_{n=1}^N \mathbf{Z}_{m:(sn+\nu-1)} \mathbf{U}_{sn+\nu-1}^\top = \psi_{\nu-1}(1) \mathbf{Q} \left[ N^{-1} \sum_{n=1}^N \boldsymbol{\xi}_n \{ \mathbf{H}_{\nu-1}(B) \boldsymbol{\varepsilon}_n^* \}^\top \right] \\ + N^{-1} \sum_{n=1}^N \{ \boldsymbol{\lambda}_{\nu-1}(B) \boldsymbol{\varepsilon}_n^* \} \{ \mathbf{H}_{\nu-1}(B) \boldsymbol{\varepsilon}_n^* \}^\top,$$

Or d'après Sims et al. (1990, Lemme 1) et le fait que les processus  $\{ \mathbf{H}_{\nu-1}(B) \boldsymbol{\varepsilon}_n^* \}$  et  $\{ \boldsymbol{\lambda}_{\nu-1}(B) \boldsymbol{\varepsilon}_n^* \}$  sont absolument sommables (voir remarque 2),

$$N^{-1} \sum_{n=1}^N \boldsymbol{\xi}_n \{ \mathbf{H}_{\nu-1}(B) \boldsymbol{\varepsilon}_n^* \}^\top \xrightarrow{\mathcal{L}} \left\{ \int_0^1 \boldsymbol{\omega}(r) d\mathbf{W}^\top(r) \right\} \{ \mathbf{H}_{\nu-1}(1) \}^\top + \{ \mathbf{H}_{\nu-1}(1) \}^\top.$$

Toujours d'après Sims et al. (1990, Lemme 1),

$$N^{-1} \sum_{n=1}^N \{ \boldsymbol{\lambda}_{\nu-1}(B) \boldsymbol{\varepsilon}_n^* \} \{ \mathbf{H}_{\nu-1}(B) \boldsymbol{\varepsilon}_n^* \}^\top = \mathbf{O}_p(1).$$

Tout ceci nous permet alors de conclure que :

$$N^{-3/2} \sum_{n=1}^N \mathbf{Z}_{m:(sn+\nu-1)} \mathbf{U}_{sn+\nu-1}^\top = \mathbf{o}_p(1).$$

iv) Posons  $\mathbf{U}_n^* = (\mathbf{U}_{sn+1}^\top, \mathbf{U}_{sn+2}^\top, \dots, \mathbf{U}_{sn+s}^\top)^\top$ . D'après la démonstration du point (iii),  $\mathbf{U}_t = \mathbf{U}_{sn+\nu} = \mathbf{H}_\nu(B) \boldsymbol{\varepsilon}_n^*$ , ainsi,

$$\mathbf{U}_n^* = (\mathbf{H}_1(B)^\top, \mathbf{H}_2(B)^\top, \dots, \mathbf{H}_s(B)^\top)^\top \boldsymbol{\varepsilon}_n^* = \mathbf{H}(B) \boldsymbol{\varepsilon}_n^*.$$

Le processus  $\{ \mathbf{H}(B) \boldsymbol{\varepsilon}_n^* \}$  est absolument sommable d'après la remarque 2. Nous avons d'après Hamilton (1994, Proposition 10.2) que :

$$N^{-1} \sum_{n=1}^N \mathbf{U}_n^* \mathbf{U}_n^{*\top} \xrightarrow{\mathcal{P}} \boldsymbol{\Gamma}_U^* = \text{cov}(\mathbf{U}_n^* \mathbf{U}_n^{*\top}).$$

Par conséquent,

$$N^{-1} \sum_{n=1}^N \mathbf{U}_{sn+\nu-1} \mathbf{U}_{sn+\nu-1}^\top \xrightarrow{\mathcal{P}} \mathbf{D}_\nu^{**} \boldsymbol{\Gamma}_U^* (\mathbf{D}_\nu^{**})^\top,$$

où  $\mathbf{D}_\nu^{**}$  est une matrice d'ordre  $(r + d(p-1)) \times (rs + ds(p-1))$  constituée de  $s$  blocs disposés en lignes, les blocs valant, la matrice identité d'ordre  $rs + d(p-1)$  à la position  $\nu$  et la matrice nulle ailleurs.

v) Pour  $\nu_0$  donné, posons ici :

$$\mathbf{U}_{\nu_0, n}^* = \left( \mathbf{U}_{sn+\nu_0}^\top, \dots, \mathbf{U}_{sn+s}^\top, \mathbf{U}_{s(n+1)+1}^\top, \dots, \mathbf{U}_{s(n+1)+\nu_0-1}^\top \right)^\top$$

et

$$\boldsymbol{\varepsilon}_{\nu_0, n}^* = \left( \boldsymbol{\varepsilon}_{sn+\nu_0}^\top, \dots, \boldsymbol{\varepsilon}_{sn+s}^\top, \boldsymbol{\varepsilon}_{s(n+1)+1}^\top, \dots, \boldsymbol{\varepsilon}_{s(n+1)+\nu_0-1}^\top \right)^\top.$$

En effet, on commence à empiler ici à partir de  $\nu_0$  quelconque et pas forcément à partir de  $\nu_0 = 1$ .

On avait obtenu dans le cas  $\nu_0 = 1$  que  $\mathbf{a}_{sn+\nu_0} = \mathbf{D}_{\nu_0} \mathbf{Q} \boldsymbol{\varepsilon}_n^*$  et  $\mathbf{U}_t = \mathbf{H}_\nu(B) \boldsymbol{\varepsilon}_n^*$  de même, nous pouvons écrire  $\mathbf{a}_{sn+\nu_0} = \mathbf{D}_{\nu_0,1} \mathbf{Q} \boldsymbol{\varepsilon}_{\nu_0, n}^*$  et  $\mathbf{U}_t = \mathbf{H}_{\nu_0}^*(B) \boldsymbol{\varepsilon}_{\nu_0, n}^*$ . Cette écriture permet d'obtenir,  $\mathbf{U}_{t-1} = \mathbf{U}_{sn+\nu_0-1} = \mathbf{H}_{\nu_0-1}^*(B) \boldsymbol{\varepsilon}_{\nu_0, n-1}^*$ , qui ne dépendra que des valeurs retardées de  $\boldsymbol{\varepsilon}_{\nu_0, n}^*$ ,  $\mathbf{U}_{sn+\nu_0}$  peut aussi être extrait du grand vecteur  $\mathbf{U}_{\nu_0, n}^*$  en faisant  $\mathbf{U}_{sn+\nu_0} = \mathbf{D}_{\nu_0,2} \mathbf{U}_{\nu_0, n}^*$ . On obtient alors :

$$\begin{aligned} N^{-1/2} \sum_{n=1}^N \text{vec} \left( \mathbf{a}_{sn+\nu_0} \mathbf{U}_{sn+\nu_0-1}^\top \right) &= N^{-1/2} \sum_{n=1}^N \text{vec} \left( \mathbf{D}_{\nu_0,1} \mathbf{Q} \boldsymbol{\varepsilon}_{\nu_0, n}^* \mathbf{U}_{\nu_0, n-1}^{*\top} \mathbf{D}_{\nu_0,2}^\top \right) \\ &+ \mathbf{o}_p(1), \\ &= \{ \mathbf{D}_{\nu_0,2} \otimes (\mathbf{D}_{\nu_0,1} \mathbf{Q}) \} N^{-1/2} \sum_{n=1}^N \text{vec} \left( \boldsymbol{\varepsilon}_{\nu_0, n}^* \mathbf{U}_{\nu_0, n-1}^{*\top} \right) \\ &+ \mathbf{o}_p(1). \end{aligned}$$

D'autre part,  $\text{vec}(\boldsymbol{\varepsilon}_{\nu_0, n}^* \mathbf{U}_{\nu_0, n-1}^{*\top})$  est une différence de martingales par rapport à  $\{\boldsymbol{\varepsilon}_{\nu_0, n}^*, \boldsymbol{\varepsilon}_{\nu_0, n-1}^*, \dots\}$  (Hamilton 1994, proposition 7.9). D'après le théorème central limite des différences de martingales :

$$N^{-1/2} \sum_{n=1}^N \text{vec}(\boldsymbol{\varepsilon}_{\nu_0, n}^* \mathbf{U}_{\nu_0, n-1}^{*\top}) \xrightarrow{\mathcal{L}} \mathcal{N}_{rd+d^2(p-1)}(\mathbf{0}, \Gamma(\mathbf{U}_{\nu_0, n}^*) \otimes \boldsymbol{\Sigma}^*),$$

où  $\Gamma(\mathbf{U}_{\nu_0, n}^*) = \text{cov}(\mathbf{U}_{\nu_0, n}^*, \mathbf{U}_{\nu_0, n}^{*\top})$  et  $\boldsymbol{\Sigma}^* = E(\boldsymbol{\varepsilon}_{\nu_0, n}^* \boldsymbol{\varepsilon}_{\nu_0, n}^{*\top})$ .

Par conséquent,

$$N^{-1/2} \sum_{n=1}^N \text{vec} \left( \mathbf{a}_{sn+\nu_0} \mathbf{U}_{sn+\nu_0-1}^\top \right) \xrightarrow{\mathcal{L}} \mathcal{N}_{rd+d^2(p-1)}(\mathbf{0}, \mathbf{Q}).$$

où

$$\mathbf{Q} = \mathbf{D}_{\nu_0,2} \Gamma(\mathbf{U}_{\nu_0, n}^*) \mathbf{D}_{\nu_0,2}^\top \otimes (\mathbf{D}_{\nu_0,1} \mathbf{Q}) \boldsymbol{\Sigma}^* (\mathbf{D}_{\nu_0,1} \mathbf{Q})^\top$$

**Annexe F.** *Preuve de la Proposition 4.9*

Sans nuire à la généralité, elle sera faite pour  $p = 5$ . Pour ce faire, nous allons débiter de l'expression (4.4.3) pour retrouver celle de l'équation (4.3.20). Nous avons pour  $t = sn + \nu$  :

$$\begin{aligned}
\mathbf{Z}_t &= \mathbf{D}_1(\nu)\mathbf{Z}_{t-1} + \mathbf{D}_2(\nu)\mathbf{Z}_{t-2} + \mathbf{D}_3(\nu)\mathbf{Z}_{t-3} + \mathbf{D}_4(\nu)\mathbf{Z}_{t-4} + \mathbf{D}_5(\nu)\mathbf{Z}_{t-5} + \boldsymbol{\varepsilon}_t, \\
&= \{ \mathbf{A}_\nu \mathbf{B}^\top + \boldsymbol{\phi}_1^*(\nu) \} (\mathbf{P}_r \mathbf{Q}_r \mathbf{Y}_{t-1} + \mathbf{P}_m \mathbf{Q}_m \mathbf{W}_{t-1}) \\
&\quad + \{ \boldsymbol{\phi}_2^*(\nu) - \boldsymbol{\phi}_1^*(\nu) \mathbf{P}_r \mathbf{Q}_r \} (\mathbf{P}_r \mathbf{Q}_r \mathbf{Y}_{t-2} + \mathbf{P}_m \mathbf{Q}_m \mathbf{W}_{t-2}) \\
&\quad + \{ \boldsymbol{\phi}_3^*(\nu) - \boldsymbol{\phi}_2^*(\nu) \mathbf{P}_r \mathbf{Q}_r \} (\mathbf{P}_r \mathbf{Q}_r \mathbf{Y}_{t-3} + \mathbf{P}_m \mathbf{Q}_m \mathbf{W}_{t-3}) \\
&\quad + \{ \boldsymbol{\phi}_4^*(\nu) - \boldsymbol{\phi}_3^*(\nu) \mathbf{P}_r \mathbf{Q}_r \} (\mathbf{P}_r \mathbf{Q}_r \mathbf{Y}_{t-4} + \mathbf{P}_m \mathbf{Q}_m \mathbf{W}_{t-4}) \\
&\quad - \boldsymbol{\phi}_4^*(\nu) \mathbf{P}_r \mathbf{Q}_r (\mathbf{P}_r \mathbf{Q}_r \mathbf{Y}_{t-5} + \mathbf{P}_m \mathbf{Q}_m \mathbf{W}_{t-5}) + \boldsymbol{\varepsilon}_t,
\end{aligned}$$

qui peut se mettre sous la forme,

$$\begin{aligned}
\mathbf{Z}_t &= \mathbf{D}_1(\nu)\mathbf{Z}_{t-1} + \mathbf{D}_2(\nu)\mathbf{Z}_{t-2} + \mathbf{D}_3(\nu)\mathbf{Z}_{t-3} + \mathbf{D}_4(\nu)\mathbf{Z}_{t-4} + \mathbf{D}_5(\nu)\mathbf{Z}_{t-5} + \boldsymbol{\varepsilon}_t, \\
&= \{ \mathbf{A}_\nu \mathbf{B}^\top + \boldsymbol{\phi}_1^*(\nu) \} \{ \mathbf{P}_r \mathbf{Q}_r \mathbf{Y}_{t-1} + \mathbf{P}_m \mathbf{Q}_m (\mathbf{Y}_{t-1} - \mathbf{Y}_{t-2}) \} \\
&\quad + \{ \boldsymbol{\phi}_2^*(\nu) - \boldsymbol{\phi}_1^*(\nu) \mathbf{P}_r \mathbf{Q}_r \} \{ \mathbf{P}_r \mathbf{Q}_r \mathbf{Y}_{t-2} + \mathbf{P}_m \mathbf{Q}_m (\mathbf{Y}_{t-2} - \mathbf{Y}_{t-3}) \} \\
&\quad + \{ \boldsymbol{\phi}_3^*(\nu) - \boldsymbol{\phi}_2^*(\nu) \mathbf{P}_r \mathbf{Q}_r \} \{ \mathbf{P}_r \mathbf{Q}_r \mathbf{Y}_{t-3} + \mathbf{P}_m \mathbf{Q}_m (\mathbf{Y}_{t-3} - \mathbf{Y}_{t-4}) \} \\
&\quad + \{ \boldsymbol{\phi}_4^*(\nu) - \boldsymbol{\phi}_3^*(\nu) \mathbf{P}_r \mathbf{Q}_r \} \{ \mathbf{P}_r \mathbf{Q}_r \mathbf{Y}_{t-4} + \mathbf{P}_m \mathbf{Q}_m (\mathbf{Y}_{t-4} - \mathbf{Y}_{t-5}) \} \\
&\quad - \boldsymbol{\phi}_4^*(\nu) \mathbf{P}_r \mathbf{Q}_r \{ \mathbf{P}_r \mathbf{Q}_r \mathbf{Y}_{t-5} + \mathbf{P}_m \mathbf{Q}_m (\mathbf{Y}_{t-5} - \mathbf{Y}_{t-6}) \} + \boldsymbol{\varepsilon}_t.
\end{aligned}$$

En utilisant le fait que  $\mathbf{P}_r \mathbf{Q}_r + \mathbf{P}_m \mathbf{Q}_m = \mathbf{I}_d$ , on obtient :

$$\begin{aligned}
\mathbf{Z}_t &= \{ \mathbf{A}_\nu \mathbf{B}^\top + \boldsymbol{\phi}_1^*(\nu) \} \{ \mathbf{Y}_{t-1} - \mathbf{P}_m \mathbf{Q}_m \mathbf{Y}_{t-2} \} \\
&\quad + \{ \boldsymbol{\phi}_2^*(\nu) - \boldsymbol{\phi}_1^*(\nu) \mathbf{P}_r \mathbf{Q}_r \} \{ \mathbf{Y}_{t-2} - \mathbf{P}_m \mathbf{Q}_m \mathbf{Y}_{t-3} \} \\
&\quad + \{ \boldsymbol{\phi}_3^*(\nu) - \boldsymbol{\phi}_2^*(\nu) \mathbf{P}_r \mathbf{Q}_r \} \{ \mathbf{Y}_{t-3} - \mathbf{P}_m \mathbf{Q}_m \mathbf{Y}_{t-4} \} \\
&\quad + \{ \boldsymbol{\phi}_4^*(\nu) - \boldsymbol{\phi}_3^*(\nu) \mathbf{P}_r \mathbf{Q}_r \} \{ \mathbf{Y}_{t-4} - \mathbf{P}_m \mathbf{Q}_m \mathbf{Y}_{t-5} \} \\
&\quad - \boldsymbol{\phi}_4^*(\nu) \mathbf{P}_r \mathbf{Q}_r \{ \mathbf{Y}_{t-5} - \mathbf{P}_m \mathbf{Q}_m \mathbf{Y}_{t-6} \} + \boldsymbol{\varepsilon}_t.
\end{aligned}$$

En utilisant aussi le fait que  $Q_r P_m = \mathbf{0}$ , et  $B^\top P_m = \mathbf{0}$ , on obtient :

$$\begin{aligned}
Z_t &= A_\nu B^\top Y_{t-1} + \phi_1^*(\nu) Y_{t-1} - \phi_1^*(\nu) P_m Q_m Y_{t-2} \\
&\quad + \phi_2^*(\nu) Y_{t-2} - \phi_2^*(\nu) P_m Q_m Y_{t-3} - \phi_1^*(\nu) P_r Q_r Y_{t-2} \\
&\quad + \phi_3^*(\nu) Y_{t-3} - \phi_3^*(\nu) P_m Q_m Y_{t-4} - \phi_2^*(\nu) P_r Q_r Y_{t-3} \\
&\quad + \phi_4^*(\nu) Y_{t-3} - \phi_4^*(\nu) P_m Q_m Y_{t-5} - \phi_1^*(\nu) P_r Q_r Y_{t-4} \\
&\quad - \phi_4^*(\nu) P_r Q_r Y_{t-5} + \varepsilon_t.
\end{aligned}$$

En regroupant les expressions on obtient :

$$\begin{aligned}
Z_t &= A_\nu B^\top Y_{t-1} \\
&\quad + \phi_1^*(\nu) Y_{t-1} - \phi_1^*(\nu) (P_m Q_m + P_r Q_r) Y_{t-2} \\
&\quad + \phi_2^*(\nu) Y_{t-2} - \phi_2^*(\nu) (P_m Q_m + P_r Q_r) Y_{t-3} \\
&\quad + \phi_3^*(\nu) Y_{t-3} - \phi_3^*(\nu) (P_m Q_m + P_r Q_r) Y_{t-4} \\
&\quad + \phi_4^*(\nu) Y_{t-4} - \phi_4^*(\nu) (P_m Q_m + P_r Q_r) Y_{t-5} \\
&\quad + \varepsilon_t.
\end{aligned}$$

En utilisant une fois de plus le fait que  $P_r Q_r + P_m Q_m = I_d$ , on trouve le résultat cherché :

$$\begin{aligned}
Z_t &= A_\nu B^\top Y_{t-1} + \phi_1^*(\nu) (Y_{t-1} - Y_{t-2}) + \phi_2^*(\nu) (Y_{t-2} - Y_{t-3}) \\
&\quad + \phi_3^*(\nu) (Y_{t-3} - Y_{t-4}) + \phi_4^*(\nu) (Y_{t-4} - Y_{t-5}) + \varepsilon_t, \\
&= A_\nu B^\top Y_{t-1} + \phi_1^*(\nu) W_{t-1} + \phi_2^*(\nu) W_{t-2} \\
&\quad + \phi_3^*(\nu) W_{t-3} + \phi_4^*(\nu) W_{t-4}.
\end{aligned}$$

#### Annexe G. Preuve du Lemme 4.2

Ici, nous montrons que la démonstration de Brüggemann et al. (2004, Annexe A.1) s'applique aussi au cas des modèles périodiques. Nous avons d'après la

formule (4.3.20) :

$$W_t = \mathbf{A}_\nu \mathbf{B}^\top Y_{t-1} + \sum_{k=1}^{p-1} \phi_k^*(\nu) W_{t-k} + \varepsilon_t, \quad t = sn + \nu, \quad W_t = Y_t - Y_{t-1}, \quad (4.5.1)$$

et

$$\begin{aligned} \hat{\varepsilon}_t &= W_t - \hat{\mathbf{A}}_\nu \hat{\mathbf{B}}^\top Y_{t-1} - \sum_{k=1}^{p-1} \hat{\phi}_k^*(\nu) W_{t-k}, \\ &= W_t - \hat{\mathbf{A}}_\nu \hat{\mathbf{B}}^\top Y_{t-1} - \sum_{k=1}^{p-1} \hat{\phi}_k^*(\nu) W_{t-k} - \hat{\mathbf{A}}_\nu \mathbf{B}^\top Y_{t-1} + \hat{\mathbf{A}}_\nu \mathbf{B}^\top Y_{t-1}, \\ &= W_t - \left( \hat{\mathbf{A}}_\nu, \hat{\phi}_1^*(\nu), \dots, \hat{\phi}_{p-1}^*(\nu) \right) \begin{pmatrix} \mathbf{B}^\top Y_{t-1} \\ W_{t-1} \\ \vdots \\ W_{t-p+1} \end{pmatrix} - \hat{\mathbf{A}}_\nu (\hat{\mathbf{B}}^\top - \mathbf{B}^\top) Y_{t-1}, \\ &= W_t - \hat{K}_\nu G_t - \hat{\mathbf{A}}_\nu (\hat{\mathbf{B}}^\top - \mathbf{B}^\top) Y_{t-1}, \end{aligned}$$

où  $\hat{K}_\nu = \left( \hat{\mathbf{A}}_\nu, \hat{\phi}_1^*(\nu), \dots, \hat{\phi}_{p-1}^*(\nu) \right)$  et  $G_t = \left( (\mathbf{B}^\top Y_{t-1})^\top, W_{t-1}^\top, \dots, W_{t-p+1}^\top \right)^\top$ .

Considérons le modèle :

$$W_t = K_\nu G_t + e_t, \quad (4.5.2)$$

où  $K_\nu = \left( \mathbf{A}_\nu, \phi_1^*(\nu), \dots, \phi_{p-1}^*(\nu) \right)$  et  $G_t = \left( (\mathbf{B}^\top Y_{t-1})^\top, W_{t-1}^\top, \dots, W_{t-p+1}^\top \right)^\top$ . Notons  $\tilde{K}_\nu$  l'estimateur des moindres carrés ordinaires de ce modèle où l'on suppose  $\mathbf{B}$  connu et  $\tilde{\varepsilon}_t$  les résidus correspondants. On a :

$$\hat{\varepsilon}_t - \tilde{\varepsilon}_t = -\hat{\mathbf{A}}_\nu \hat{\mathbf{B}}^\top Y_{t-1} + \tilde{\mathbf{A}}_\nu \mathbf{B}^\top Y_{t-1} \quad (4.5.3)$$

$$\begin{aligned} &+ \sum_{k=1}^{p-1} \left\{ \tilde{\phi}_k^*(\nu) - \hat{\phi}_k^*(\nu) \right\} W_{t-k}, \\ &= -\hat{\mathbf{A}}_\nu \hat{\mathbf{B}}^\top Y_{t-1} + \tilde{\mathbf{A}}_\nu \mathbf{B}^\top Y_{t-1} - \hat{\mathbf{A}}_\nu \mathbf{B}^\top Y_{t-1} + \hat{\mathbf{A}}_\nu \mathbf{B}^\top Y_{t-1} \quad (4.5.4) \\ &+ \sum_{k=1}^{p-1} \left\{ \tilde{\phi}_k^*(\nu) - \hat{\phi}_k^*(\nu) \right\} W_{t-k}, \\ &= \left( \tilde{\mathbf{A}}_\nu - \hat{\mathbf{A}}_\nu \right) \mathbf{B}^\top Y_{t-1} + \sum_{k=1}^{p-1} \left\{ \tilde{\phi}_k^*(\nu) - \hat{\phi}_k^*(\nu) \right\} W_{t-k} \end{aligned}$$

$$\begin{aligned}
& - \hat{\mathcal{A}}_\nu \left( \hat{\mathbf{B}}^\top - \mathbf{B}^\top \right) Y_{t-1}, \\
& = (G_t \otimes I_d) \left( \tilde{\mathbf{k}}_\nu - \hat{\mathbf{k}}_\nu \right) - \hat{\mathcal{A}}_\nu \left( \hat{\mathbf{B}}^\top - \mathbf{B}^\top \right) Y_{t-1}, \\
& = H_t,
\end{aligned} \tag{4.5.5}$$

où  $\tilde{\mathbf{k}}_\nu = \text{vec}(\tilde{\mathbf{K}}_\nu)$  et  $\hat{\mathbf{k}}_\nu = \text{vec}(\hat{\mathbf{K}}_\nu)$ .

On peut alors écrire  $\hat{\boldsymbol{\varepsilon}}_t = \tilde{\boldsymbol{\varepsilon}}_t + H_t$ .

En utilisant les autocovariances, on a :

$$\begin{aligned}
\hat{\mathbf{C}}_h - \tilde{\mathbf{C}}_h & = N^{-1} \sum_{n=1}^N \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_{t-h}^\top - N^{-1} \sum_{n=1}^N \tilde{\boldsymbol{\varepsilon}}_t \tilde{\boldsymbol{\varepsilon}}_{t-h}^\top, \\
& = N^{-1} \sum_{n=1}^N (\tilde{\boldsymbol{\varepsilon}}_t + H_t) (\tilde{\boldsymbol{\varepsilon}}_{t-h} + H_{t-h})^\top - N^{-1} \sum_{n=1}^N \tilde{\boldsymbol{\varepsilon}}_t \tilde{\boldsymbol{\varepsilon}}_{t-h}^\top, \\
& = N^{-1} \sum_{n=1}^N \tilde{\boldsymbol{\varepsilon}}_t H_{t-h}^\top + N^{-1} \sum_{n=1}^N H_t \tilde{\boldsymbol{\varepsilon}}_{t-h}^\top + N^{-1} \sum_{n=1}^N H_t H_{t-h}^\top,
\end{aligned} \tag{4.5.6}$$

où  $t = sn + \nu$ . Il suffit de montrer que chacun des termes de la dernière ligne de l'équation ci-dessus est  $\mathbf{O}_p(N^{-1})$ .

Définissons  $\hat{\mathbf{G}}_t = \left( (\hat{\mathbf{B}}^\top Y_{t-1})^\top, W_{t-1}^\top, \dots, W_{t-p+1}^\top \right)^\top$ . On a d'autre part :

$$\begin{aligned}
\varepsilon_t & = W_t - \mathcal{A}_\nu \mathbf{B}^\top Y_{t-1} - \sum_{k=1}^{p-1} \phi_k^*(\nu) W_{t-k}, \\
& = W_t - \mathcal{A}_\nu \mathbf{B}^\top Y_{t-1} - \sum_{k=1}^{p-1} \phi_k^*(\nu) W_{t-k} - \mathcal{A}_\nu \hat{\mathbf{B}}^\top Y_{t-1} + \mathcal{A}_\nu \hat{\mathbf{B}}^\top Y_{t-1}, \\
& = W_t - (\mathcal{A}_\nu, \phi_1^*(\nu), \dots, \phi_{p-1}^*(\nu)) \begin{pmatrix} \hat{\mathbf{B}}^\top Y_{t-1} \\ W_{t-1} \\ \vdots \\ W_{t-p+1} \end{pmatrix} + \mathcal{A}_\nu (\hat{\mathbf{B}}^\top - \mathbf{B}^\top) Y_{t-1}, \\
& = W_t - K_\nu \hat{\mathbf{G}}_t + \mathcal{A}_\nu (\hat{\mathbf{B}}^\top - \mathbf{B}^\top) Y_{t-1}.
\end{aligned}$$

Alors,  $\hat{\mathbf{K}}_\nu = \left( \hat{\mathbf{A}}_\nu, \hat{\phi}_1^*(\nu), \dots, \hat{\phi}_{p-1}^*(\nu) \right)$  peut être obtenu par les moindres carrés ordinaires du modèle suivant :

$$\mathbf{W}_t = \mathbf{K}_\nu \hat{\mathbf{G}}_t + \mathbf{e}_t = \left( \hat{\mathbf{G}}_t^\top \otimes \mathbf{I}_d \right) \mathbf{k}_\nu + \mathbf{e}_t, \quad (4.5.7)$$

où  $\mathbf{k}_\nu = \text{vec}(\mathbf{K}_\nu)$ ,  $\mathbf{e}_t = \varepsilon_t - \hat{\mathbf{A}}_\nu \left( \hat{\mathbf{B}}^\top - \mathbf{B}^\top \right) \mathbf{Y}_{t-1}$ .

D'après la Proposition 4.8, on a  $\left( \hat{\mathbf{B}} - \mathbf{B} \right) = \mathbf{O}_p(N^{-1})$ . Ceci permet de déduire que les moyennes échantillonales de  $\hat{\mathbf{G}}_t$  et  $\mathbf{e}_t$  et  $\varepsilon_t$  sont d'ordre  $\mathbf{O}_p(N^{-1/2})$ . D'après l'équation (4.5.7), et les propriétés de l'estimateur des moindres carrés ordinaires, on a :

$$\begin{aligned} \hat{\mathbf{k}}_\nu &= \left\{ \left( \sum_{n=1}^N \hat{\mathbf{G}}_t \hat{\mathbf{G}}_t^\top \right)^{-1} \otimes \mathbf{I}_d \right\} \sum_{n=1}^N \left( \hat{\mathbf{G}}_t \otimes \mathbf{I}_d \right) \mathbf{W}_t + \mathbf{O}_p(N^{-1}), \quad (4.5.8) \\ &= \left\{ \left( \sum_{n=1}^N \hat{\mathbf{G}}_t \hat{\mathbf{G}}_t^\top \right)^{-1} \otimes \mathbf{I}_d \right\} \sum_{n=1}^N \left( \hat{\mathbf{G}}_t \otimes \mathbf{I}_d \right) \left\{ \left( \hat{\mathbf{G}}_t^\top \otimes \mathbf{I}_d \right) \mathbf{k}_\nu + \mathbf{e}_t \right\} + \mathbf{O}_p(N^{-1}), \\ &= \mathbf{k}_\nu + \left\{ \left( \sum_{n=1}^N \hat{\mathbf{G}}_t \hat{\mathbf{G}}_t^\top \right)^{-1} \otimes \mathbf{I}_d \right\} \sum_{n=1}^N \left( \hat{\mathbf{G}}_t \otimes \mathbf{I}_d \right) \mathbf{e}_t + \mathbf{O}_p(N^{-1}), \end{aligned}$$

où  $t = sn + \nu$ .

D'autre part,

$$\hat{\mathbf{G}}_t - \mathbf{G}_t = \left[ \mathbf{Y}_{t-1}^\top \left( \hat{\mathbf{B}} - \mathbf{B} \right), \mathbf{0}^\top, \dots, \mathbf{0}^\top \right]^\top, \quad (4.5.9)$$

et

$$\begin{aligned} \left( \hat{\mathbf{G}}_t \otimes \mathbf{I}_d \right) \mathbf{e}_t &= \left\{ \left( \mathbf{G}_t + \left[ \left( \hat{\mathbf{B}} - \mathbf{B} \right) \mathbf{Y}_{t-1}^\top, \mathbf{0}^\top, \dots, \mathbf{0}^\top \right]^\top \right) \otimes \mathbf{I}_d \right\} \\ &\quad \times \left\{ \varepsilon_t - \hat{\mathbf{A}}_\nu \left( \hat{\mathbf{B}} - \mathbf{B} \right) \mathbf{Y}_{t-1} \right\}, \\ &= \left( \mathbf{G}_t \otimes \mathbf{I}_d \right) \varepsilon_t - \left( \mathbf{G}_t \otimes \mathbf{I}_d \right) \hat{\mathbf{A}}_\nu \left( \hat{\mathbf{B}}^\top - \mathbf{B}^\top \right) \mathbf{Y}_{t-1} \\ &\quad + \left\{ \left[ \mathbf{Y}_{t-1}^\top \left( \hat{\mathbf{B}} - \mathbf{B} \right), \mathbf{0}^\top, \dots, \mathbf{0}^\top \right]^\top \otimes \mathbf{I}_d \right\} \varepsilon_t \\ &\quad - \left\{ \left[ \mathbf{Y}_{t-1}^\top \left( \hat{\mathbf{B}} - \mathbf{B} \right), \mathbf{0}^\top, \dots, \mathbf{0}^\top \right]^\top \otimes \mathbf{I}_d \right\} \left\{ \hat{\mathbf{A}}_\nu \left( \hat{\mathbf{B}}^\top - \mathbf{B}^\top \right) \mathbf{Y}_{t-1} \right\}. \end{aligned}$$

En considérant la somme membre à membre et en utilisant les résultats du Lemme 4.1, on a :

$$\sum_{n=1}^N (\hat{\mathbf{G}}_t \otimes \mathbf{I}_d) \mathbf{e}_t = \sum_{n=1}^N (\mathbf{G}_t \otimes \mathbf{I}_d) \boldsymbol{\varepsilon}_t + \mathbf{O}_p(1), \quad (4.5.10)$$

où  $t = sn + \nu$ .

D'autre part, en utilisant l'équation (4.5.9), on a :

$$\begin{aligned} \sum_{n=1}^N \hat{\mathbf{G}}_t \hat{\mathbf{G}}_t^\top &= \sum_{n=1}^N \left\{ \mathbf{G}_t + [\mathbf{Y}_{t-1}^\top (\hat{\mathbf{B}} - \mathbf{B}), \mathbf{0}^\top, \dots, \mathbf{0}^\top]^\top \right\} \\ &\quad \times \left\{ \mathbf{G}_t + [\mathbf{Y}_{t-1}^\top (\hat{\mathbf{B}} - \mathbf{B}), \mathbf{0}^\top, \dots, \mathbf{0}^\top]^\top \right\}^\top, \\ &= \sum_{n=1}^N \left\{ \mathbf{G}_t \mathbf{G}_t^\top + \mathbf{G}_t [\mathbf{Y}_{t-1}^\top (\hat{\mathbf{B}} - \mathbf{B}), \mathbf{0}^\top, \dots, \mathbf{0}^\top]^\top \right. \\ &\quad + \left. \mathbf{G}_t [\mathbf{Y}_{t-1}^\top (\hat{\mathbf{B}} - \mathbf{B}), \mathbf{0}^\top, \dots, \mathbf{0}^\top]^\top \right. \\ &\quad + \left. [\mathbf{Y}_{t-1}^\top (\hat{\mathbf{B}} - \mathbf{B}), \mathbf{0}^\top, \dots, \mathbf{0}^\top]^\top [\mathbf{Y}_{t-1}^\top (\hat{\mathbf{B}} - \mathbf{B}), \mathbf{0}^\top, \dots, \mathbf{0}^\top] \right\}, \end{aligned}$$

où  $t = sn + \nu$ .

En utilisant une fois de plus le Lemme 4.1 et la Proposition 4.8, on trouve que :

$$\sum_{n=1}^N \hat{\mathbf{G}}_t \hat{\mathbf{G}}_t^\top = \sum_{n=1}^N \mathbf{G}_t \mathbf{G}_t^\top + \mathbf{O}_p(1).$$

On obtient avec l'équation (4.5.7) :

$$\hat{\mathbf{k}}_\nu = \mathbf{k}_\nu + \left\{ \left( \sum_{n=1}^N \mathbf{G}_t \mathbf{G}_t^\top \right)^{-1} \otimes \mathbf{I}_d \right\} \left\{ \sum_{n=1}^N (\mathbf{G}_t \otimes \mathbf{I}_d) \boldsymbol{\varepsilon}_t \right\} + \mathbf{O}_p(N^{-1}). \quad (4.5.11)$$

En considérant l'équation, (4.5.2), on a également :

$$\tilde{\mathbf{k}}_\nu = \mathbf{k}_\nu + \left\{ \left( \sum_{n=1}^N \mathbf{G}_t \mathbf{G}_t^\top \right)^{-1} \otimes \mathbf{I}_d \right\} \sum_{n=1}^N (\mathbf{G}_t \otimes \mathbf{I}_d) \boldsymbol{\varepsilon}_t. \quad (4.5.12)$$

Les équations (4.5.11) et (4.5.12) permettent d'avoir :

$$\hat{\mathbf{k}}_\nu - \tilde{\mathbf{k}}_\nu = \mathbf{O}_p(N^{-1}). \quad (4.5.13)$$



*Le résultat est alors obtenu à partir des équations (4.5.6) et (4.5.3).*

# CONCLUSION

---

Dans cette thèse, nous apportons des contributions dans la modélisation et la validation des modèles récents de séries chronologiques multivariées. L'objectif principal était d'estimer et développer des tests de type diagnostique pour valider les hypothèses de bruit blanc et de non corrélation. Plus précisément, on s'est intéressé dans un premier temps à la vérification de la qualité de l'ajustement dans les modèles vectoriels à erreurs multiplicatives (vMEM). De plus, nous avons étudié la non corrélation sous faible dépendance entre deux séries chronologiques. Finalement, nous avons considéré l'estimation des paramètres et la vérification de l'adéquation dans les séries chronologiques périodiques cointégrées.

Dans la première partie, on s'est penché sur les modèles vMEMs. Ces modèles jouent un rôle très important dans la modélisation économique des données financières. Beaucoup de travaux ont été faits sur les modèles vMEMs, mais ces travaux se sont surtout limités sur l'estimation et des applications. Peu d'études ont été faites sur la validation de l'ajustement des modèles vMEMs. Dans ce projet, de nouveaux tests pour le diagnostic des modèles vMEMs ont été proposés. Deux approches ont été adoptées dont une dans le domaine du temps et l'autre dans le domaine des fréquences. Dans le domaine du temps, nous avons montré que sous l'hypothèse nulle d'adéquation, la statistique de Hosking-Ljung-Box converge vers une somme pondérée de lois de khi-carré à un degré de liberté indépendantes. Dans le domaine des fréquences, un test statistique a été obtenu en comparant la densité spectrale des résidus de l'estimation du modèle observé à la densité spectrale théorique sous l'hypothèse nulle d'adéquation. La transformation de Chen et Deo (2004) a par la suite été appliquée à ce test car la version

originale était très conservatrice. Les simulations ont montré dans le cas des deux approches de très bons comportements des nouveaux tests. Les tests spectraux (et plus précisément ceux ne se basant pas sur le noyau tronqué) se sont avérés plus puissants. Le test de Ng et al. (2016) basé sur l'inverse de la matrice des variances et covariances des autocovariances n'a pas donné de bons résultats. Même le fait de considérer l'inverse généralisé et en ajustant le degré de liberté de la distribution n'a pas amélioré les résultats. L'étude de la dynamique commune des deux composantes de l'indice boursier FTSE100 à l'aide d'un modèle vMEM a été faite afin d'illustrer les tests statistiques proposés. Parmi toutes ces statistiques, nous recommandons l'utilisation de la statistique de noyau basée sur la transformation de Chen et Deo (2004).

Dans la deuxième partie de la thèse, on s'est intéressé à la présence de dépendance dans les observations ou dans les erreurs lors de la modélisation. Les résultats de la plupart des études sur les tests d'adéquation ou de non corrélation dans la littérature ont été obtenus sous l'hypothèse d'indépendance ou de normalité, ou le cas échéant, sous des hypothèses se rapprochant de ces deux dernières telles que la nullité des cumulants d'ordre quatre ou aussi la constance des moments conditionnels jusqu'à l'ordre quatre (voir Hannan (1976)). Nous montrons ici que sous les conditions de stabilité définies par Wu (2005), la distribution asymptotique d'un vecteur des covariances croisées des séries en question est normale. Ces conditions permettent en effet de décrire des situations de dépendance faible qui stipule ici que l'effet cumulé d'une observation donnée sur les observations futures est fini. La matrice des variances et covariances de la distribution trouvée a été estimée par la procédure robuste à l'autocorrélation et à l'hétéroscédasticité (VARHAC) présentée par Den Haan et Levin (1974). Ceci permet de définir un test statistique dont la distribution asymptotique est une somme pondérée de lois de khi-carré indépendantes. Une étude de simulations permet de confirmer la performance de notre test par rapport au test de Haugh (1976) qui est le plus souvent utilisé pour tester la non corrélation entre deux processus.

Finalement, nous avons introduit les modèles périodiques multivariés et cointégrés en nous concentrant à la fois sur l'estimation et le diagnostic. Dans l'analyse

des séries chronologiques, la cointégration est une caractéristique liée à la présence de combinaisons linéaires stationnaires même si les séries elles-mêmes ne le sont pas. Le nombre de combinaisons linéaires stationnaires est appelé rang de cointégration. Cette situation crée une déficience de dimension dans l'espace vectoriel engendré par les processus en question. Ces combinaisons linéaires sont les tendances générales qui ne reflètent pas la vraie dynamique entre les séries. Il faut de ce fait définir et estimer un modèle à correction d'erreurs. C'est ce qui a été fait dans cette partie de la thèse. L'estimation a été faite en deux étapes. La première étape est l'estimation de plein rang qui ignore le rang de cointégration. Les paramètres servent de valeurs initiales à l'estimation de rang réduit. Ce dernier tient compte du rang de cointégration. À chaque étape, la distribution asymptotique des paramètres estimés est aussi déterminée. À notre connaissance, une approche similaire n'a pas encore été abordée dans la littérature pour estimer un modèle périodique multivarié et cointégré. Les résidus issus de l'estimation ont par la suite été utilisés pour construire un test de type portemanteau. La distribution asymptotique de ce test est une loi khi-carrée dont le nombre de degrés de liberté est ajusté suivant le nombre de paramètres estimés dans le modèle et le nombre de relations de cointégration. Des études de simulation ont montré que le test se comporte bien sur des échantillons de tailles relativement petites.

Comme première avenue de recherche, il serait intéressant de se pencher sur l'identification des modèles vMEMs. En effet, la définition des modèles vMEM fait intervenir trois processus à savoir le processus brut à modéliser, la moyenne conditionnelle et les erreurs. De ce fait, à cause de la non linéarité, l'estimation est très complexe et les algorithmes affichent souvent des problèmes de convergence. C'est pourquoi certains auteurs ont développé des modèles Log-vMEM ou des vMEM structurés dont l'interdépendance est modélisée par un facteur latent (voir Hautsch (2008)). Idéalement, il faudrait obtenir les conditions nécessaires et suffisantes de stabilité des paramètres du modèle vMEM sans changer la structure fondamentale du modèle. Il est espéré que ceci améliorerait la convergence des techniques d'estimation utilisées (maximum de vraisemblance avec les copules, méthode généralisée des moments, moindre carrés non linéaires). Toujours pour

garantir une meilleure estimation de ces modèles, Cipollini et al. (2007, Annexe 2) ont proposé des conditions sur les paramètres devant garantir la non négativité de la moyenne conditionnelle estimée. On peut s'intéresser non seulement à avoir les conditions sous lesquelles la moyenne conditionnelle est non négative, mais aussi à déterminer sous quelles conditions elle existe (de même que la forme de l'espace paramétrique) et comment l'estimer. D'autre part, une des préoccupations serait de trouver le meilleur modèle qui pourrait s'ajuster aux données. Il peut arriver d'ajuster des modèles vMEMs d'ordres différents sur les données, mais de ne pas savoir lequel des deux est meilleur suivant des critères fixés. Un de nos objectifs dans le futur est de définir et de justifier des critères tels que le AIC et le BIC devant permettre de sélectionner le meilleur modèle entre deux modèles ajustés.

S'agissant de la non corrélation sous faible dépendance, on peut s'intéresser dans le futur à une approche spectrale. La technique utilisée dans la thèse n'est qu'un cas particulier de l'approche spectrale lorsqu'on utilise un noyau tronqué. Le cas univarié de test de non corrélation sous faible dépendance a été présenté par Shao (2011). Ce travail a été fait pour les autocorrélations d'une même série et non sur les corrélations croisées entre deux séries différentes. Un autre projet intéressant serait l'extension dans le cas multivarié.

Dans le cas des modèles multivariés périodiques, ce qui compte être fait dans la suite c'est de développer un test de rang de cointégration. Dans le présent travail, nous avons supposé le rang de cointégration connu. Il aurait été préférable de faire un test comme celui de Johansen (1991) pour un rang de cointégration dans chaque saison ou un rang commun pour toutes les saisons. Certains auteurs ayant développés l'estimation de rang réduit dans le cas non périodique utilisent l'approche de Johansen (1991) pour tester le rang de cointégration. Ce test se base sur les valeurs propres ou sur la trace du polynôme autorégressif évalué au point unité. En exploitant une représentation adéquate des modèles périodiques (en considérant les polynômes autorégressifs par saison, ou en utilisant le polynôme de la représentation en vecteur de saisons), il devrait être possible d'adapter certaines méthodologies existantes à ce cas périodique.

Une autre extension est de prendre en considération les paramètres de quasi-différentiation à savoir ceux entrant dans la définition de l'intégration périodique d'ordre un en supposant qu'ils ne sont pas connus. Dans le cas univarié, ceci entraîne l'augmentation du nombre de paramètres. Des techniques peuvent être utilisées pour contourner ces difficultés comme une estimation en deux étapes ou aussi définir une estimation sous contrainte d'une structure existante dans les paramètres. Une approche similaire a été utilisée par Boswijk et Franses (1996) qui ont mis au point un test du rapport de vraisemblance pour évaluer la présence de racines unitaires dans un processus périodique intégré univarié en utilisant dans l'estimation une contrainte sur les coefficients de quasidifférentiation.

## Bibliographie

---

- [1] Ahn, S. K. & Reinsel, G. C. (1990). ‘Estimation for partially nonstationary multivariate autoregressive models’. *Journal of the American Statistical Association* **85**, 813–823.
- [2] Ansley, C. F. & Newbold, P. (1979). ‘On the finite sample distribution of residual autocorrelations in autoregressive-moving average models’. *Biometrika* **66**, 547–553.
- [3] Arjas, E. & Lehtonen, T. (1978). ‘Approximating many server queues by means of single server queues’. *Mathematics of Operation Research* **3**, 205–223.
- [4] Balcombe, K. (1999). ‘Seasonal unit root tests with structural break in deterministic seasonality’. *Oxford Bulletin of Economics and Statistics* **61**, 569–582.
- [5] Banerjee, A., Dolado, J. J., Galbraith, J. W. & Hendry, D. F. (1993). *Cointegration, Error Correction and the Econometric Analysis of Non-Stationary Data*. Oxford University Press, New York.
- [6] Bauwens, L. (2006). ‘Econometric analysis of intra-daily activity on Tokyo Stock Exchange’. *Monetary and Economic Studies* **24**, 1–23.
- [7] Bauwens, L. & Giot, P. (2000). ‘The logarithmic ACD model : An application to the bid-ask quote process of three NYSE stocks’. *Annales d’Économie et de Statistique* **60**, 117–149.
- [8] Bauwens, L. & Veredas, D. (2004). ‘The stochastic conditional duration model : A latent factor model for the analysis of financial durations’. *Journal of Econometrics* **119**, 381–412.
- [9] Berk, K. (1974). ‘Consistent autoregressive spectral estimates’. *The Annals of Statistics* **2**, 489–502.

- [10] Box, G. E. P. & Pierce D. A. (1970). 'Distribution of residual autocorrelations in autoregressive integrated moving average time series models'. *Journal of the American Statistical Association* **65**, 1509–1526.
- [11] Birchenall, C. R., Bladen-Ovell, A. P. L., Chui, D. R., Osborn, D. R. & Smith, J. P. (1989). 'A seasonal model of consumption'. *Economic Journal* **99**, 837–843.
- [12] Borkar, V. S. (1993). 'White-noise representations in stochastic realization theory'. *SIAM Journal on Control and Optimization* **31**, 1093–1102.
- [13] Boswijk, H. P. & Franses, P. H. (1996). 'Unit roots in periodic autoregressions'. *Journal of Time Series Analysis* **17**, 221–245.
- [14] Boswijk, H. P. & Franses, P. H. (1995a). 'Periodic cointegration : Representation and inference'. *The Review of Economics and Statistics* **77**, 436–454.
- [15] Boswijk, H. P. & Franses, P. H. (1995b). 'Testing for periodic integration'. *Economics Letters* **48**, 241–248.
- [16] Boswijk, H. P., Franses, P. H. & Haldrup, N. (1997). 'Multiple root in periodic autoregression'. *Journal of Econometrics* **80**, 167–193.
- [17] Bouhaddioui, C. & Roy, R. (2006a). 'On the distribution of the residual cross-correlations between two uncorrelated infinite order vector autoregressive series'. *Statistics and Probability Letters* **76**, 58–68.
- [18] Bouhaddioui, C. & Roy, R. (2006b). 'A generalized portmanteau test for independence of two infinite order vector autoregressive series'. *Journal of Time Series Analysis* **27**, 505–444.
- [19] Box, G. E. P. (1954). 'Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effects of inequality of variance in the one-way classification'. *The Annals of Mathematical Statistics* **25**, 290–302.
- [20] Box, G. E. P. & Jenkins, G. M. (1970). *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco.
- [21] Brockwell, P. J. & Davis, R. A. (2002). *Introduction to Time Series and Forecasting*. Springer-Verlag, New York.
- [22] Brockwell, P. J. & Davis, R. A. (1991). *Time Series : Theory and Methods*, 2ème édition. Springer-Verlag, New York.



- [23] Burden, R. L. & Douglas Faires, J. (2011). *Numerical Analysis*. Brooks/Cole, Boston.
- [24] Canova, F. & Hansen, B. E. (1995). ‘Are seasonal patterns constant over time? A test for seasonal stability’. *Journal of Business and Economic Statistics* **13**, 237–252.
- [25] Chen, W. W. & Deo, R. S. (2004). ‘Power transformations to induce normality and their applications’. *Journal of the Royal Statistical Society, Series B* **66**, 117–130.
- [26] Chitturi, R. V. (1976). ‘Distribution of multivariate white noise autocorrelations’, *Journal of the American Statistical Association* **71**, 223–226.
- [27] Chitturi, R. V. (1974). ‘Distribution of residual autocorrelations in multiple autoregressive schemes’, *Journal of the American Statistical Association* **69**, 928–934.
- [28] Chou, R. Y. (2005). ‘Forecasting financial volatilities with extreme values : The conditional autoregressive range (CARR) model’. *Journal of Money, Credit and Banking* **37**, 561–582.
- [29] Cipollini, F. & Gallo, G. M. (2010). ‘Automated variable selection in vector multiplicative error models’. *Computational Statistics & Data Analysis* **54**, 2470–2486.
- [30] Den Haan, W. & Levin, A. (1994). ‘Inferences from parametric and non-parametric covariance matrix estimation procedures’. *Manuscript, UCSD and Federal Reserve Board*.
- [31] del Barrio, C. T. & Osborn, D. R. (2008). ‘Cointegration for periodically integrated processes’. *Econometric Theory* **24**, 109–142.
- [32] Dionne, G., Duchesne, P. & Pacurar, M. (2009). ‘Intraday value at risk (IVAR) using tick-by-tick data with application to the Toronto Stock Exchange’. *Journal of Empirical Finance* **16**, 777–792.
- [33] Duchesne, P. & Francq, C. (2008). ‘On diagnostic checking time series models with portmanteau test statistics based on generalized inverses and  $\{2\}$ -inverses’. In : Brito, P. (Ed.), COMPSTAT 2008. Vol. 18 of Proceedings in Computational Statistics. Physica-Verlag, Heidelberg, pp. 143–154.

- [34] Duchesne, P., Ghoudi, K. & Rémillard, B. (2012). ‘On testing for independence between the innovations of several time series’. *The Canadian Journal of Statistics* **40**, 447–479.
- [35] Duchesne, P. & Lafaye de Micheaux, P. (2010). ‘Computing the distribution of quadratic forms : Further comparisons between the Liu-Tang-Zhang approximation and exact methods’. *Computational Statistics & Data Analysis* **54**, 858–862.
- [36] Duchesne, P. & Roy, R. (2004). ‘On consistent testing for serial correlation of unknown form in vector time series models’. *Journal of Multivariate Analysis* **89**, 148–180.
- [37] Dufour, J.-M. & Valéry, P. (2015). ‘Wald-type tests when rank conditions fail : a smooth regularization approach’. Technical Report, McGill University.
- [38] El Himdi, K. & Roy, R. (1997). ‘Tests for non correlation of two multivariate ARMA time series : A nonparametric approach’. *The Canadian Journal of Statistics* **25**, 233–256.
- [39] Engle, R. F. & Gallo, G. M. (2006). ‘A multiple indicators model for volatility using intra-daily data’. *Journal of Econometrics* **131**, 3–27.
- [40] Engle, R. F. (2002). ‘New frontiers for ARCH models’. *Journal of Applied Econometrics* **17**, 425–446.
- [41] Engle, R. F. (2000). ‘The econometrics of ultra-high frequency data’. *Econometrica* **68**, 1–22.
- [42] Engle, R. F. & Granger, C. W. J. (1991). *Long-Run Economic Relationships, Readings in Cointegration*. Oxford University Press, Oxford.
- [43] Engle, R. F. & Granger, C. W. J. (1987). ‘Cointegration and error correction : Representation, estimation and testing’. *Econometrica* **55**, 251–276.
- [44] Engle, R. F. & Russell, J. R. (1998). ‘Autoregressive conditional duration : A new model for irregularly spaced transaction data’. *Econometrica* **66**, 1127–1162.
- [45] Francq, C., Roy, R. & Zakoian, J. (2005). ‘Diagnostic checking in ARMA models with uncorrelated errors’. *Journal of the American Statistical Association* **100**, 532–544.

- [48] Franses, P. H. (1995). 'A vector of quarters representation for bivariate time series'. *Econometrics Reviews* **14**, 55–63.
- [47] Franses, P. H. (1994). 'A multivariate approach to modeling univariate seasonal time series'. *Journal of Econometrics* **63**, 133–151.
- [48] Franses, P. H. & Paap, R. (2004). *Periodic Time Series Models*. Oxford University Press, Oxford.
- [49] Fuller, W. A. (1996). *Introduction to Time Series*, 2ème édition. Wiley Series in Probability and Statistics. John Wiley & Sons, New York.
- [50] Gladyshev, E. G. (1961). 'Periodically correlated random sequences'. *Soviet Mathematics* **2**, 385–388.
- [51] Gouriéroux, C. & Montfort, A. (1995). *Séries Temporelles & Modèles Dynamiques*. Economica, Paris.
- [52] Hallin, M. & Saidi, A. (2005). 'Testing independence and causality between multivariate ARMA times series'. *Journal of Time Series Analysis* **23**, 86–106.
- [53] Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, New Jersey.
- [54] Hannan, E. J. (1976). 'The asymptotic distribution of serial covariances'. *The Annals of Statistics* **4**, 396–399.
- [55] Haugh, L. (1976). 'Checking the independence of two covariance-stationary time series : A univariate residual cross-correlation approach'. *Journal of the American Statistical Association* **71**, 378–385.
- [56] Hautsch, N. (2008). 'Capturing common components in high-frequency financial time series : A multivariate stochastic multiplicative error model'. *Journal of Economic Dynamics and Control* **32**, 3978–4015.
- [57] Hautsch, N. & Jeleskovic, V. (2008). 'Modelling high-frequency volatility and liquidity using multiplicative error models'. Technical Report SFB 649 Discussion Paper 2008-047, Humboldt-Universität zu Berlin, Germany.
- [58] Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford University Press, Oxford.
- [59] Hong, Y. (1996). 'Consistent testing for serial correlation of unknown form'. *Econometrica* **64**, 837–864.

- [60] Hong, Y., Liu, Y. & Wang, S. (2009). ‘Granger causality in risk and detection of extreme risk spillover between financial markets’. *Journal of Econometrics* **150**, 271–287.
- [61] Hosking, J. R. M. (1980). ‘The multivariate portmanteau statistic’. *Journal of the American Statistical Association* **75**, 602–608.
- [62] Hsing, T. & Wu, W. (2004). ‘On weighted U-statistics for stationary processes’. *The Annals of Probability* **32**, 1600–1631
- [63] Hylleberg, S., Engle, R. F. & Granger, C. J. W. (1990). ‘Seasonal integration and cointegration’. *Journal of Econometrics* **44**, 215–238.
- [64] Jasiak, J. (1998). ‘Persistence in intertrade durations’. *Finance* **19**, 166–195.
- [65] Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford.
- [66] Johansen, S. (1991). ‘Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models’. *Econometrica* **59**, 1551–1580.
- [67] Johansen, S. (1988). ‘Statistical analysis of cointegration vectors’. *Journal of Economic Dynamics and Control* **12**, 231–254.
- [68] Jones, R. H. & Brelsford, W. (1967). ‘Time series with periodic structure’. *Biometrika* **24**, 403–408.
- [69] Kim, E. & Lee, S. (2005). ‘A test for independence of two stationary infinite order autoregressive processes’. *Annals of the Institute of Statistical Mathematics* **57**, 105–127.
- [70] Kleibergen, F. & Franses, P. H. (1999). ‘Cointegration in a periodic vector autoregression’. Econometric Institute report 9906/A, Erasmus University, Rotterdam.
- [71] Koch, P. & Yang, S. S. (1986). ‘A method for testing the independence of two time series that accounts for a potential pattern in the cross-correlation function’. *Journal of the American Statistical Association* **81**, 553–544.
- [72] Kunst, R. M. (1990). ‘Seasonal cointegration in macroeconomic systems : Case studies for small and large european countries’. *Mimeo Institute for Advanced Studies*, Vienna.

- [73] Li, W. K. (2004). *Diagnostic Checks in Time Series*. Monographs on Statistics and Applied Probability 102. Chapman & Hall/CRC, New York.
- [74] Li, W. K. & Hui, Y. V. (1988). ‘An algorithm for the exact likelihood of periodic autoregressive moving-average models’. *Communications in Statistics-Simulation and Computation* **17**, 1483–1494.
- [75] Li, W. K. & McLeod, A. I. (1981). ‘Distribution of the residual autocorrelations in multivariate ARMA time series models’, *Journal of the Royal Statistical Society, Series B* **43**, 231–239.
- [76] Liu, W. & Lin, Z. (2009). ‘Strong approximation for a class of stationary processes’. *Stochastic Processes and their Applications* **119**, 249–280.
- [77] Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, New York.
- [78] Manganelli, S. (2005). ‘Duration, volume and volatility impact of trades’. *Journal of Financial Markets* **8**, 377–399.
- [79] McLeod, A. I. (1993). ‘Parsimony model adequacy, and periodic autocorrelation in time series forecasting’. *International Statistical Review* **61**, 387–393.
- [80] McLeod, A. I. (1979). ‘Distribution of the residual cross-correlation in univariate ARMA time series models’. *Journal of the American Statistical Association* **74**, 849–855.
- [81] Nelsen, R. B. (1999). *An Introduction to Copulas*. Lecture Notes in Statistics, Vol. 139, Springer-Verlag, Berlin.
- [82] Ng, F. C., Li, W. K. & Yu, L. H. (2016). ‘Diagnostic checking of the vector multiplicative error model’. *Computational Statistics & Data Analysis* **94**, 86–97.
- [83] Osborn, D. R. (1991). ‘The implications of periodically varying coefficients for seasonal time-series processes’. *Journal of Econometrics* **48**, 373–384.
- [osborn1988] Osborn, D. R. (1988). ‘Seasonality and habit persistence in a life-cycle model of consumption’. *Journal of Applied Econometrics* **3**, 255–266.
- [85] Pacurar, M. (2008). ‘Autoregressive conditional duration models in finance : A survey of the theoretical and empirical literature’. *Journal of Economic Surveys* **22**, 711–751.

- [86] Pagano, M. (1978). ‘On periodic and multiple autoregressions’. *The Annals of Statistics* **6**, 1310–1317.
- [87] Pham, T. P., Roy, R. & Cédras, L. (2000). ‘Tests for non correlation of two cointegrated ARMA time series’. *Centre de Recherche Mathématiques, Université de Montréal, Cahier de recherche 2649*.
- [88] Philips, P. C. B. & Durlauf, S. N. (1986). ‘Multiple time series with integrated variables’. *Review of Economic Studies* **53**, 473–496.
- [89] Philips, P. C. B. & Ouliaris, S. (1990). ‘Asymptotic properties of residual based tests for cointegration’. *Econometrica* **58**, 165–193.
- [90] Poulin, J. & Duchesne, P. (2008). ‘On the power transformation of kernel-based tests for serial correlation in vector time series : Some finite sample results and a comparison with the bootstrap’. *Computational Statistics & Data Analysis* **52**, 4432–4457.
- [91] Priestley, M. B. (1981). *Spectral Analysis and Time Series. Probability and Mathematical Statistics*. Academic Press, London.
- [92] Quenouille, M. H. (1957). *The Analysis of Multiple Time Series*. Griffin, London.
- [93] Reinsel, G. C. (1997). *Elements of Multivariate Time Series Analysis*, 2ème édition. Springer, New York.
- [94] Rosenblatt, M. (1952). ‘Remarks on a multivariate transformation’. *The Annals of Mathematical Statistics* **23**, 470–472.
- [95] Roy, R. (1989). ‘Asymptotic covariance structure of serial correlation in multivariate time series’. *Biometrika* **76**, 824–827.
- [96] Saidi, A. (2007). ‘Consistent test for non correlation of two cointegrated ARMA time series’. *The Canadian Journal of Statistics* **35**, 169–188.
- [97] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons Inc., New York.
- [98] Shao, X. (2011). ‘Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models’. *Econometric Theory* **27**, 312–343.

- [99] Shao, X. & Wu, W. B. (2007). ‘Asymptotic spectral theory for nonlinear time series’. *The Annals of Statistics* **35**, 1773–1801.
- [100] Shephard, N. & Sheppard, K. (2010). ‘Realising the future : Forecasting with high frequency based volatility (HEAVY) models’. *Journal of Applied Econometrics* **25**, 197–231.
- [101] Shumway, R. H. & Stoffer, D. S. (2011). *Time Series Analysis and Its Applications*, 3ème édition. Springer-Verlag, New York.
- [102] Sims, C. A., Stock, J. H. & Watson, M. W. (1990). ‘Inference in linear time series models with some unit roots’. *Econometrica* **58**, 113–144.
- [103] Tiao, G. C. & Grupe, M. R. (1980). ‘Hidden periodic autoregressive moving-average models in time series data’. *Biometrika* **67**, 365–373.
- [104] Ursu, E. & Duchesne, P. (2009). ‘Estimation and model adequacy checking for multivariate seasonal autoregressive time series with periodically varying parameters’. *Statistica Neerlandica* **63**, 183–212.
- [105] Wu, W. B. (2011). ‘Asymptotic theory for stationary processes’. *Statistics and Its Interface* **4**, 207–226.
- [106] Wu, W. B. (2009). ‘An asymptotic theory for sample covariances of Bernoulli shifts’. *Stochastic Processes and their Applications* **119**, 453–467.
- [107] Wu, W. B. (2007). ‘Strong invariance principles for dependent random variables’. *The Annals of Probability* **35**, 2294–2320.
- [108] Wu, W. B. (2005). ‘Nonlinear system theory : Another look at dependence’. *Proceedings of the National Academy of Sciences USA* **102**, 14150–14154.
- [109] Wu, W. B. & Min, W. (2005). ‘On linear processes with dependent innovations’. *Stochastic Processes and their Applications* **115**, 939–958.
- [110] Wu, W. B. & Shao, X. (2004). ‘Limit theorems for iterated random functions’. *Journal of Applied Probability* **41**, 425–436.
- [111] Wu, W. B. & Zhou, Z. (2011). ‘Gaussian approximations for non-stationary multiple time series’. *Statistica Sinica* **21**, 1397–1413.
- [112] Yap, S. F. & Reinsel, G. C. (1995). ‘Estimation and testing for unit roots in a partially nonstationary vector autoregressive moving average model’. *Journal of the American Statistical Association* **90**, 253–267.

- [113] Zhang, M. Y., Russell, J. & Tsay, R. (2001). 'A nonlinear autoregressive conditional duration model with applications to financial transaction data'. *Journal of Econometrics* **104**, 179–207.