

UNIVERSITÉ CHARLES DE GAULLE - LILLE III

THÈSE

présentée en vue de l'obtention du grade de

Docteur Ès-Sciences

Discipline : Mathématiques Appliquées aux Sciences
Économiques

présentée par

Yacouba BOUBACAR MAINASSARA

sous le titre

**Estimation, validation et identification
des modèles ARMA faibles multivariés**

COMPOSITION DU JURY

Directeurs de thèse	M. Christian FRANCO	Professeur	Université Lille 3
	M. Jean-Michel ZAKOÏAN	Professeur	Université Lille 3.
Rapporteurs	M. Sébastien LAURENT	Professeur	Université de Maastricht
	Mme Anne PHILIPPE	Professeure	Université de Nantes.
Présidente	Mme Laurence BROZE	Professeure	Université Lille 3.
Examineurs	M. Michel CARBON	Professeur	Université Rennes 2.
	M. Cristian PREDA	Professeur	Polytech'Lille.

Remerciements

Tout d'abord je tiens à remercier chaleureusement Messieurs les Professeurs Christian FRANCO et Jean-Michel ZAKOÏAN d'avoir accepté d'encadrer mon travail de recherche avec compétence et enthousiasme. Leurs qualités scientifiques et humaines, leurs suivis attentifs à tous les stades d'élaboration de cette thèse, leurs disponibilités à chaque fois que j'en avais eu recours, leurs soutiens et surtout leurs conseils très précieux qu'ils ont eu à me prodiguer m'ont permis de mener à bien cette thèse. Je tiens aussi à leurs exprimer ma profonde gratitude.

Je tiens à remercier Monsieur le Professeur Sébastien LAURENT et Madame la Professeure Anne PHILIPPE de l'honneur qu'ils me font d'avoir accepté d'être les rapporteurs de cette thèse et pour l'intérêt qu'ils ont porté à mon travail. Je remercie également Messieurs les Professeurs Michel CARBON et Cristian PREDA d'avoir accepté de siéger à ce jury.

Je tiens tout particulièrement à remercier chaleureusement Madame la Professeure Laurence BROZE d'avoir accepté d'être la présidente du jury de soutenance et pour tout le soutien qu'elle nous (mon épouse et moi) a apporté depuis mon master 2.

Un grand merci également à tous les membres du laboratoire EQUIPPE-GREMARS qui m'ont permis d'effectuer cette thèse dans une ambiance vraiment sereine et pour leur accueil très chaleureux. Je remercie vivement Madame Christiane FRANCO pour sa lecture très attentive de la partie en français de cette thèse. Je voudrais aussi adresser un grand merci à mes parents, à toute ma famille et à mes amis qui m'ont toujours soutenu moralement et matériellement pour poursuivre mes études jusqu'à ce niveau.

Je dédie ce travail à mon père décédé en juillet 2008, ma mère, mon fils et toute ma famille.

Enfin, je voudrais dire merci à tous ceux, famille ou amis, qui m'entourent et à l'ensemble des personnes présentes à cette soutenance.

Résumé

Dans cette thèse nous élargissons le champ d'application des modèles ARMA (Auto-Regressive Moving-Average) vectoriels en considérant des termes d'erreur non corrélés mais qui peuvent contenir des dépendances non linéaires. Ces modèles sont appelés des ARMA faibles vectoriels et permettent de traiter des processus qui peuvent avoir des dynamiques non linéaires très générales. Par opposition, nous appelons ARMA forts les modèles utilisés habituellement dans la littérature dans lesquels le terme d'erreur est supposé être un bruit iid. Les modèles ARMA faibles étant en particulier denses dans l'ensemble des processus stationnaires réguliers, ils sont bien plus généraux que les modèles ARMA forts. Le problème qui nous préoccupera sera l'analyse statistique des modèles ARMA faibles vectoriels. Plus précisément, nous étudions les problèmes d'estimation et de validation. Dans un premier temps, nous étudions les propriétés asymptotiques de l'estimateur du quasi-maximum de vraisemblance et de l'estimateur des moindres carrés. La matrice de variance asymptotique de ces estimateurs est de la forme "sandwich" $\Omega := J^{-1}IJ^{-1}$, et peut être très différente de la variance asymptotique $\Omega := 2J^{-1}$ obtenue dans le cas fort. Ensuite, nous accordons une attention particulière aux problèmes de validation. Dans un premier temps, en proposant des versions modifiées des tests de Wald, du multiplicateur de Lagrange et du rapport de vraisemblance pour tester des restrictions linéaires sur les paramètres de modèles ARMA faibles vectoriels. En second, nous nous intéressons aux tests fondés sur les résidus, qui ont pour objet de vérifier que les résidus des modèles estimés sont bien des estimations de bruits blancs. Plus particulièrement, nous nous intéressons aux tests portmanteau, aussi appelés tests d'autocorrélation. Nous montrons que la distribution asymptotique des autocorrélations résiduelles est normalement distribuée avec une matrice de covariance différente du cas fort (c'est-à-dire sous les hypothèses iid sur le bruit). Nous en déduisons le comportement asymptotique des statistiques portmanteau. Dans le cadre standard d'un ARMA fort, il est connu que la distribution asymptotique des tests portmanteau est correctement approximée par un chi-deux. Dans le cas général, nous montrons que cette distribution asymptotique est celle d'une somme pondérée de chi-deux. Cette distribution peut être très différente de l'approximation chi-deux usuelle du cas fort. Nous proposons donc des tests portmanteau modifiés pour tester l'adéquation de modèles ARMA faibles vectoriels. Enfin, nous nous sommes intéressés aux choix des modèles ARMA faibles vectoriels fondé sur la minimisation d'un critère d'information,

notamment celui introduit par Akaike (AIC). Avec ce critère, on tente de donner une approximation de la distance (souvent appelée information de Kullback-Leibler) entre la vraie loi des observations (inconnue) et la loi du modèle estimé. Nous verrons que le critère corrigé (AIC_c) dans le cadre des modèles ARMA faibles vectoriels peut, là aussi, être très différent du cas fort.

Mots-clés : AIC, autocorrelations résiduelles, estimateur HAC, estimateur spectral, information de Kullback-Leibler, modèles VARMA faibles, forme échelon, processus non linéaire, QMLE/LSE, représentation structurelle des modèles VARMA, test du Multiplicateur de Lagrange, tests portmanteau de Ljung-Box et Box-Pierce, test du rapport de vraisemblance, test de Wald.

Abstract

The goal of this thesis is to study the vector autoregressive moving-average (V)ARMA models with uncorrelated but non-independent error terms. These models are called weak VARMA by opposition to the standard VARMA models, also called strong VARMA models, in which the error terms are supposed to be iid. We relax the standard independence assumption, and even the martingale difference assumption, on the error term in order to be able to cover VARMA representations of general nonlinear models. The problems that are considered here concern the statistical analysis. More precisely, we concentrate on the estimation and validation steps. We study the asymptotic properties of the quasi-maximum likelihood (QMLE) and/or least squares estimators (LSE) of weak VARMA models. Conditions are given for the consistency and asymptotic normality of the QMLE/LSE. A particular attention is given to the estimation of the asymptotic variance matrix, which may be very different from that obtained in the standard framework. After identification and estimation of the vector autoregressive moving-average processes, the next important step in the VARMA modeling is the validation stage. The validity of the different steps of the traditional methodology of Box and Jenkins, identification, estimation and validation, depends on the noise properties. Several validation methods are studied. This validation stage is not only based on portmanteau tests, but also on the examination of the autocorrelation function of the residuals and on tests of linear restrictions on the parameters. Modified versions of the Wald, Lagrange Multiplier and Likelihood Ratio tests are proposed for testing linear restrictions on the parameters. We studied the joint distribution of the QMLE/LSE and of the noise empirical autocovariances. We then derive the asymptotic distribution of residual empirical autocovariances and autocorrelations under weak assumptions on the noise. We deduce the asymptotic distribution of the Ljung-Box (or Box-Pierce) portmanteau statistics for VARMA models with nonindependent innovations. In the standard framework (i.e. under the assumption of an iid noise), it is shown that the asymptotic distribution of the portmanteau tests is that of a weighted sum of independent chi-squared random variables. The asymptotic distribution can be quite different when the independence assumption is relaxed. Consequently, the usual chi-squared distribution does not provide an adequate approximation to the distribution of the Box-Pierce goodness-of fit portmanteau test. Hence we propose a method to adjust the critical values of the portmanteau tests. Finally, we considered the problem of

orders selection of weak VARMA models by means of information criteria. We propose a modified Akaike information criterion (AIC).

Keywords : AIC, Box-Pierce and Ljung-Box portmanteau tests, Discrepancy, Echelon form, Goodness-of-fit test, Kullback-Leibler information, Lagrange Multiplier test, Likelihood Ratio test, Nonlinear processes, Order selection, QMLE/LSE, Residual autocorrelation, Structural representation, Weak VARMA models, Wald test.

Table des matières

Résumé	ii
Abstract	iv
Table des matières	viii
Liste des tableaux	x
Table des figures	xi
1 Introduction	1
1.0.1 Concepts de base et contexte bibliographique	1
1.0.2 Représentations VARMA faibles de processus non linéaires	4
1.1 Résultats du chapitre 2	10
1.1.1 Estimation des paramètres	11
1.1.2 Estimation de la matrice de variance asymptotique	15
1.1.3 Tests sur les coefficients du modèle	17
1.2 Résultats du chapitre 3	21
1.2.1 Notations et paramétrisation des coefficients du modèle VARMA	22
1.2.2 Expression des dérivées des résidus du modèle VARMA	23
1.2.3 Expressions explicites des matrices J et I	24
1.2.4 Estimation de la matrice de variance asymptotique $\Omega := J^{-1}IJ^{-1}$	26
1.3 Résultats du chapitre 4	28
1.3.1 Modèle et paramétrisation des coefficients	29
1.3.2 Distribution asymptotique jointe de $\hat{\theta}_n$ et des autocovariances empiriques du bruit	30
1.3.3 Comportement asymptotique des autocovariances et autocorrélations résiduelles	31
1.3.4 Comportement asymptotique des statistiques portmanteau	32
1.3.5 Mise en oeuvre des tests portmanteau	35
1.4 Résultats du chapitre 5	37
1.4.1 Modèle et paramétrisation des coefficients	37
1.4.2 Définition du critère d'information	38

1.4.3	Contraste de Kullback-Leibler	39
1.4.4	Critère de sélection des ordres d'un modèle VARMA	40
1.5	Annexe	46
1.5.1	Stationnarité	46
1.5.2	Ergodicité	47
1.5.3	Accroissement de martingale	48
1.5.4	Mélange	49
	Références bibliographiques	51
2	Estimating weak structural VARMA models	55
2.1	Introduction	55
2.2	Model and assumptions	58
2.3	Quasi-maximum likelihood estimation	59
2.4	Estimating the asymptotic variance matrix	62
2.5	Testing linear restrictions on the parameter	63
2.6	Numerical illustrations	65
2.7	Appendix of technical proofs	71
2.8	Bibliography	76
2.9	Appendix of additional example	78
2.9.1	Verification of Assumption A8 on Example 2.1	78
2.9.2	Details on the proof of Theorem 2.1	79
2.9.3	Details on the proof of Theorem 2.2	82
3	Estimating the asymptotic variance of LSE of weak VARMA models	88
3.1	Introduction	88
3.2	Model and assumptions	90
3.3	Expression for the derivatives of the VARMA residuals	92
3.4	Explicit expressions for I and J	94
3.5	Estimating the asymptotic variance matrix	96
3.6	Technical proofs	97
3.7	References	110
3.8	Verification of expression of I and J on Examples	111
4	Multivariate portmanteau test for weak structural VARMA models	118
4.1	Introduction	119
4.2	Model and assumptions	120
4.3	Least Squares Estimation under non-iid innovations	121
4.4	Joint distribution of $\hat{\theta}_n$ and the noise empirical autocovariances	124
4.5	Asymptotic distribution of residual empirical autocovariances and auto-correlations	125

4.6	Limiting distribution of the portmanteau statistics	125
4.7	Examples	127
4.8	Implementation of the goodness-of-fit portmanteau tests	132
4.9	Numerical illustrations	133
4.9.1	Empirical size	134
4.9.2	Empirical power	138
4.10	Appendix	139
4.11	References	144
5	Model selection of weak VARMA models	152
5.1	Introduction	152
5.2	Model and assumptions	154
5.3	General multivariate linear regression model	157
5.4	Kullback-Leibler discrepancy	157
5.5	Criteria for VARMA order selection	158
5.5.1	Estimating the discrepancy	160
5.5.2	Other decomposition of the discrepancy	163
5.6	References	166
	Conclusion générale	168
	Perspectives	170

Liste des tableaux

2.1	Empirical size of standard and modified tests : relative frequencies (in %) of rejection of $H_0 : b_1(2, 2) = 0$. The number of replications is $N = 1000$	67
2.2	Empirical power of standard and modified tests : relative frequencies (in %) of rejection of $H_0 : b_1(2, 2) = 0$. The number of replications is $N = 1000$	68
4.1	Empirical size (in %) of the standard and modified versions of the LB test in the case of the strong VARMA(1, 1) model (4.13)-(4.14), with $\theta_0 = (0.225, -0.313, 0.750)$	135
4.2	Empirical size (in %) of the standard and modified versions of the LB test in the case of the strong VARMA(1, 1) model (4.13)-(4.14), with $\theta_0 = (0, 0, 0)$	135
4.3	Empirical size (in %) of the standard and modified versions of the LB test in the case of the weak VARMA(1, 1) model (4.13)-(4.15), with $\theta_0 = (0.225, -0.313, 0.750)$	136
4.4	Empirical size (in %) of the standard and modified versions of the LB test in the case of the weak VARMA(1, 1) model (4.13)-(4.15), with $\theta_0 = (0, 0, 0)$	136
4.5	Empirical size (in %) of the standard and modified versions of the LB test in the case of the weak VARMA(1, 1) model (4.13)-(4.16), with $\theta_0 = (0.225, -0.313, 0.750)$	137
4.6	Empirical size (in %) of the standard and modified versions of the LB test in the case of the weak VARMA(1, 1) model (4.13)-(4.17), with $\theta_0 = (0.225, -0.313, 0.750)$	137
4.7	Empirical size (in %) of the standard and modified versions of the LB test in the case of the weak VARMA(2, 2) model (4.18)-(4.15).	138
4.8	Empirical size (in %) of the standard and modified versions of the LB test in the case of the strong VARMA(2, 2) model (4.18)-(4.14).	139
4.9	Empirical size (in %) of the standard and modified versions of the LB test in the case of the strong VARMA(1, 1) model (4.13)-(4.14).	144
4.10	Empirical size (in %) of the standard and modified versions of the LB test in the case of the strong VARMA(1, 1) model (4.13)-(4.14).	145
4.11	Empirical size (in %) of the standard and modified versions of the LB test in the case of the weak VARMA(1, 1) model (4.13)-(4.15).	146

4.12	Empirical size (in %) of the standard and modified versions of the LB test in the case of the weak VARMA(1, 1) model (4.13)-(4.15).	147
4.13	Empirical size (in %) of the standard and modified versions of the LB test in the case of the weak VARMA(1, 1) model (4.13)-(4.16).	148
4.14	Empirical size (in %) of the standard and modified versions of the LB test in the case of the weak VARMA(1, 1) model (4.13)-(4.17).	149
4.15	Empirical size (in %) of the standard and modified versions of the LB test in the case of the weak VARMA(2, 2) model (4.18)-(4.15).	150
4.16	Empirical size (in %) of the standard and modified versions of the LB test in the case of the strong VARMA(2, 2) model (4.18)-(4.14).	151

Table des figures

- 2.1 QMLE of $N = 1,000$ independent simulations of the VARMA(1,1) model (2.9) with size $n = 2,000$ and unknown parameter $\vartheta_0 = ((a_1(2,2), b_1(2,1), b_1(2,2)) = (0.95, 2, 0)$, when the noise is strong (left panels) and when the noise is the weak noise (2.11) (right panels). Points (a)-(c), in the box-plots of the top panels, display the distribution of the estimation errors $\hat{\vartheta}(i) - \vartheta_0(i)$ for $i = 1, 2, 3$. The panels of the middle present the Q-Q plot of the estimates $\hat{\vartheta}(3) = \hat{b}_1(2,2)$ of the last parameter. The bottom panels display the distribution of the same estimates. The kernel density estimate is displayed in full line, and the centered Gaussian density with the same variance is plotted in dotted line. 69
- 2.2 Comparison of standard and modified estimates of the asymptotic variance Ω of the QMLE, on the simulated models presented in Figure 2.1. The diamond symbols represent the mean, over the $N = 1,000$ replications, of the standardized squared errors $n \{\hat{a}_1(2,2) - 0.95\}^2$ for (a) (0.02 in the strong and weak cases), $n \{\hat{b}_1(2,1) - 2\}^2$ for (b) (1.02 in the strong case and 1.01 in the strong case) and $n \{\hat{b}_1(2,2)\}^2$ for (c) (0.94 in the strong case and 0.43 in the weak case). 70

Chapitre 1

Introduction

Lorsque l'on dispose de plusieurs séries présentant des dépendances temporelles ou instantanées, il est utile de les étudier conjointement en les considérant comme les composantes d'un processus vectoriel (multivarié).

1.0.1 Concepts de base et contexte bibliographique

Dans la littérature économétrique et dans l'analyse des séries temporelles (chronologiques), la classe des modèles autorégressifs moyennes mobiles vectoriels (VARMA pour Vector AutoRegressive Moving Average) (voir Reinsel, 1997, Lütkepohl, 2005) et la sous-classe des modèles VAR (voir Lütkepohl, 1993) sont utilisées non seulement pour étudier les propriétés de chacune de ces séries, mais aussi pour décrire de possibles relations croisées entre les différentes séries chronologiques. Ces modèles VARMA occupent une place centrale pour la modélisation¹ des séries temporelles multivariées. Ils sont une extension naturelle des modèles ARMA qui constituent la classe la plus utilisée de modèles de séries temporelles univariées (voir Brockwell et Davis, 1991). Cette extension pose néanmoins des problèmes ardu, comme par exemple, l'identification et l'estimation des paramètres du modèle et suscite des axes de recherches spécifiques, comme la cointégration (voir Lütkepohl, 2005). Ces modèles sont généralement utilisés avec des hypothèses fortes sur le bruit qui en limitent la généralité. Ainsi, nous appelons VARMA forts les modèles standard dans lesquels le terme d'erreur est supposé

1. Opération par laquelle on établit un modèle d'un phénomène, afin d'en proposer une représentation qu'on peut interpréter, reproduire et simuler. Par ailleurs, modèle est synonyme de théorie, mais avec une connotation pratique : un modèle, c'est une théorie orientée vers l'action à laquelle elle doit servir. En clair, elle permet d'avoir un aperçu théorique d'une idée en vue d'un objectif concret.

être une suite indépendante et identiquement distribuée (*i.e.* iid), et nous parlons de modèles VARMA faibles quand les hypothèses sur le bruit sont moins restrictives. Nous parlons également de modèle semi-fort quand le bruit est supposé être une différence de martingale. Plusieurs formulations VARMA ont été introduites pour décrire différents types de séries temporelles multivariées. L'une des formulations les plus générales et les plus simples est le modèle VARMA structurel sans tendance suivant

$$A_{00}X_t - \sum_{i=1}^{p_0} A_{0i}X_{t-i} = B_{00}\epsilon_t - \sum_{j=1}^{q_0} B_{0j}\epsilon_{t-j}, \quad \forall t \in \mathbb{Z} \quad (1.1)$$

où $X_t = (X_{1t}, \dots, X_{dt})'$ (on notera par la suite $X = (X_t)$) est un processus vectoriel stationnaire au second ordre de dimension d , à valeurs réelles. Les paramètres A_{0i} , $i \in \{1, \dots, p_0\}$ et B_{0j} , $j \in \{1, \dots, q_0\}$ sont des matrices $d \times d$, et p_0 et q_0 sont des entiers appelés ordres. Le terme d'erreur $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{dt})'$ est un bruit blanc, c'est-à-dire une suite de variables aléatoires centrées ($E\epsilon_t = 0$), non corrélées, avec une matrice de covariance non singulière Σ_0 .

La représentation (1.1) est dite réduite lorsque l'on suppose $A_{00} = B_{00} = I_d$, et structurelle dans le cas contraire. Les modèles structurels permettent d'étudier des relations économiques instantanées. Les formes réduites sont plus pratiques d'un point de vue statistique, car elles donnent les prévisions de chaque composante en fonction des valeurs passées de l'ensemble des composantes.

Afin de donner une définition précise d'un modèle linéaire et non linéaire d'un processus, commençons par rappeler la décomposition de Wold (1938) qui affirme que si (Z_t) est un processus stationnaire, purement non déterministe, il peut être représenté comme une moyenne mobile infinie (*i.e.* $MA(\infty)$) (voir Brockwell et Davis, 1991, pour le cas univarié). Les arguments utilisés s'étendent directement au cas multivarié (voir Reinsel, 1997). Ainsi tout processus $X = (X_t)$ de dimension d vérifiant les conditions ci-dessus admet une représentation $MA(\infty)$ vectorielle

$$X_t = \sum_{\ell=0}^{\infty} \Psi_{\ell} \epsilon_{t-\ell}, \quad (1.2)$$

où $\sum_{\ell=0}^{\infty} \|\Psi_{\ell}\|^2 < \infty$. Le processus (ϵ_t) est l'innovation linéaire du processus X . Nous parlons de modèles linéaires lorsque le processus d'innovation (ϵ_t) est iid et de modèles non linéaires dans le cas contraire. Dans le cadre univarié Francq, Roy et Zakoïan (2005), Francq et Zakoïan (1998) ont montré qu'il existe des processus très variés admettant, à la fois, des représentations non linéaires et linéaires de type ARMA, pourvu que les hypothèses sur le bruit du modèle ARMA soient suffisamment peu restrictives. Relâcher cette hypothèse permet aux modèles VARMA faibles de couvrir une large classe de processus non linéaires.

Pour les applications économétriques, le cadre univarié est très restrictif. Les séries économiques présentent des interdépendances fortes rendant nécessaire l'étude simultanée de plusieurs séries. Le développement de la littérature sur la cointégration (fondée sur les travaux du prix Nobel d'économie Granger) atteste de l'importance de cette problématique. Parmi les modèles pouvant être ajustés au processus d'innovation linéaire, dont les autocorrélations sont nulles mais qui peut néanmoins présenter des dépendances temporelles, nous pouvons citer les modèles autorégressifs conditionnellement hétéroscédastiques (ARCH) introduits par Engle (1982) et leur extension GARCH (ARCH généralisés) due à Bollerslev (1986). Contrairement aux modèles ARMA, les extensions des modèles GARCH au cadre multivarié ne sont pas directes. Il existe dans la littérature différents types de modèles GARCH multivariés (voir Bauwens, Laurent et Rombouts, 2006), comme le modèle à corrélation conditionnelle constante proposé par Bollerslev (1988) et étendu par Jeantheau (1998). Les séries générées par ces modèles sont des différences de martingales et sont utilisés pour décrire des séries financières. Pour la modélisation de ces séries financières, d'autres modèles peuvent être utilisés, comme par exemple les modèles all-pass (Davis et Breidt, 2006) qui génèrent aussi des séries non corrélées mais qui ne sont pas en général des différences de martingales.

De nombreux outils statistiques ont été développés pour l'analyse, l'estimation et l'interprétation des modèles VARMA forts (voir *e.g.* Chitturi (1974), Dunsmuir et Hannan (1976), Hannan (1976), Hannan, Dunsmuir, et Deistler (1980), Hosking (1980, 1981a, 1981b), Kascha (2007), Li et McLeod (1981) et Reinsel, Basu et Yap (1992)). Le problème, qui nous préoccupera, sera l'analyse statistique des modèles VARMA faibles. Le principal but de cette thèse est d'étudier l'estimation et la validité d'outils statistiques dans le cadre des modèles VARMA faibles et d'en proposer des extensions. Les travaux consacrés à l'analyse statistique des modèles VARMA faibles sont pour l'heure essentiellement limités au cas univarié. Sous certaines hypothèses d'ergodicité et de mélange, la consistance et la normalité asymptotique de l'estimateur des moindres carrés ordinaires ont été montrés par Francq et Zakoïan (1998). Il se trouve que la précision de cet estimateur dépend de manière cruciale des hypothèses faites sur le bruit. Il est donc nécessaire d'adapter la méthodologie usuelle d'estimation de la matrice de variance asymptotique comme dans Francq et Zakoïan (2000). Parmi les autres résultats statistiques disponibles, signalons des travaux importants sur le comportement asymptotique des autocorrélations empiriques établis par Romano et Thombs (1996). Ces auteurs ont également introduits des processus non corrélés mais dépendants et qui peuvent être étendus au cas multivarié. Signalons aussi les travaux importants sur les gains d'efficacité des estimateurs de type GMM obtenus par Broze, Francq et Zakoïan (2001), sur l'utilisation des tests portmanteau par Francq, Roy et Zakoïan (2005), sur les tests de linéarité et l'estimation HAC de modèles ARMA faibles par Francq, et Zakoïan (2007) et sur l'utilisation d'une méthode d'estimation basée sur la régression proposée

par Hannan et Rissanen (1982) dans un cadre semi-fort. En analyse multivariée, des progrès importants ont été réalisés par Dufour et Pelletier (2005) qui ont proposé une généralisation de la méthode d'estimation basée sur la régression introduite par Hannan et Rissanen (1982) et étudié les propriétés asymptotiques de cet estimateur. Ils ont également proposé un critère d'information modifié convergent pour la sélection des modèles VARMA faibles

$$\log \det \tilde{\Sigma} + \dim(\gamma) \frac{(\log n)^{1+\delta}}{n}, \quad \delta > 0,$$

qui est aussi une extension de celui Hannan et Rissanen (1982). Par ailleurs, Francq et Raïssi (2007) ont étudié des tests portmanteau d'adéquation de modèles VAR faibles.

1.0.2 Représentations VARMA faibles de processus non linéaires

Cette thèse se situe dans la continuité des travaux de recherche cités précédemment. Parmi la grande diversité des modèles stochastiques de séries temporelles à temps discret, on distingue, et on oppose parfois, les modèles linéaires et les modèles non linéaires. En réalité ces deux classes de modèles ne sont pas incompatibles et peuvent même être complémentaires. Les fervents partisans des modèles non linéaires, ou de la prévision non paramétrique, reprochent souvent aux modèles linéaires ARMA (AutoRegressive Moving Average) d'être trop restrictifs, de ne convenir qu'à un petit nombre de séries. Ceci est surtout vrai si on suppose, comme on le fait habituellement, des hypothèses fortes sur le bruit qui intervient dans l'écriture ARMA. Des processus très variés admettent, à la fois, des représentations non linéaires et linéaires de type ARMA, pourvu que les hypothèses sur le bruit du modèle ARMA soient suffisamment peu restrictives. On parle alors de modèles ARMA faibles. L'étude des modèles ARMA faibles a des conséquences importantes pour l'analyse des séries économiques, pour lesquelles de nombreux travaux ont montrés que les modèles ARMA standard sont mal adaptés.

Interprétations des représentations linéaires faibles

Si dans (1.1), $\epsilon = (\epsilon_t)$ est un bruit blanc fort,² à savoir une suite de variables iid, alors on dit que $X = (X_t)$ est un modèle VARMA(p_0, q_0) structurel fort. Si l'on suppose que ϵ est une différence de martingale,³ alors on dit que X admet une représentation

2. Voir Remarque 1.13 de l'annexe pour la définition.

3. Voir Définition 1.11 de l'annexe.

VARMA(p_0, q_0) structurelle semi-forte. Et si on suppose simplement que ϵ est un bruit blanc faible,⁴ mais qu'il n'est pas forcément iid, ni même une différence de martingale, on dit que X admet une représentation VARMA(p_0, q_0) structurelle faible. Ainsi la distinction entre modèle VARMA(p_0, q_0) structurel fort, semi-fort ou faible n'est donc qu'une question d'hypothèse sur le bruit. En plus, du fait que la contrainte sur ϵ est moins forte pour un modèle structurel faible que pour un modèle structurel semi-fort ou fort, il est clair que la classe des modèles VARMA(p_0, q_0) structurels faibles est la plus large. Nous en déduisons la relation suivante

$$\{\text{VARMA forts}\} \subset \{\text{VARMA semi-forts}\} \subset \{\text{VARMA faibles}\}.$$

Soit $\mathcal{H}_X(t-1)$ l'espace de Hilbert engendré par les variables aléatoires X_{t-1}, X_{t-2}, \dots . Cet espace contient les combinaisons linéaires de la forme $\sum_{i=1}^m C_i X_{t-i}$ et leurs limites en moyenne quadratique. Notons $\mathcal{F}_X(t-1)$ la tribu engendrée par X_{t-1}, X_{t-2}, \dots . Nous appellerons $\mathcal{H}_X(t-1)$ le passé linéaire de X_t et $\mathcal{F}_X(t-1)$ sera appelé le passé de X_t . Il est parfois pratique d'écrire l'équation (1.1) sous la forme compacte

$$A(L)X_t = B(L)\epsilon_t$$

où L est l'opérateur retard (*i.e.* $X_{t-1} = LX_t$, $\epsilon_{t-k} = L^k \epsilon_t$ pour tout $t, k \in \mathbb{Z}$), $A(z) = A_0 - \sum_{i=1}^{p_0} A_i z^i$ est le polynôme VAR et $B(z) = B_0 - \sum_{j=1}^{q_0} B_j z^j$ est le polynôme MA vectoriel. Notons par $\det(\cdot)$, le déterminant d'une matrice carrée. Pour l'étude des séries chronologiques, la notion de stationnarité faible⁵ conduit à imposer des contraintes sur les racines des polynômes VAR et MA vectoriel. Si les racines de $\det A(z) = 0$ sont à l'extérieur du disque unité, alors X_t est une combinaison linéaire de $\epsilon_t, \epsilon_{t-1}, \dots$ (cette combinaison est une solution non anticipative ou causale de (1.1)). Si les racines de $\det B(z) = 0$ sont à l'extérieur du disque unité, alors ϵ_t peut s'écrire comme une combinaison linéaire de X_t, X_{t-1}, \dots (l'équation (1.1) est dite inversible). Les deux hypothèses précédentes sur les polynômes VAR et MA vectoriel seront regroupées dans l'hypothèse suivante

$$\mathbf{H}_1 : \quad \det A(z) \det B(z) = 0 \Rightarrow |z| > 1.$$

Sous l'hypothèse \mathbf{H}_1 , le passé (linéaire) de X coïncide donc avec le passé (linéaire) de ϵ :

$$\mathcal{F}_X(t-1) = \mathcal{F}_\epsilon(t-1), \quad \mathcal{H}_X(t-1) = \mathcal{H}_\epsilon(t-1).$$

Si en plus, ϵ est un bruit blanc faible, on en déduit que la meilleure prévision de X_t est une fonction linéaire du passé

$$\begin{aligned} A_{00} E \{X_t | \mathcal{H}_X(t-1)\} &= B_{00} E \{\epsilon_t | \mathcal{H}_\epsilon(t-1)\} + \sum_{i=1}^p A_{0i} X_{t-i} - \sum_{j=1}^q B_{0i} \epsilon_{t-j}, \\ &= \sum_{i=1}^p A_{0i} X_{t-i} - \sum_{j=1}^q B_{0i} \epsilon_{t-j}. \end{aligned}$$

4. Voir Définition 1.8 de l'annexe.

5. Voir Définition 1.7 de l'annexe.

Le bruit ϵ s'interprète donc comme une normalisation du processus des innovations linéaires de X :

$$\epsilon_t = B_{00}^{-1} A_{00} (X_t - E \{X_t | \mathcal{H}_X(t-1)\}).$$

Lorsque la forme VARMA est réduite, c'est-à-dire quand $A_{00} = B_{00} = I_d$, alors ϵ_t coïncide précisément avec l'innovation linéaire $X_t - E \{X_t | \mathcal{H}_X(t-1)\}$. Ainsi dans un modèle VARMA faible on suppose simplement que le terme d'erreur est l'innovation linéaire. Par contre, dans un modèle VARMA (semi-)fort on suppose que le terme d'erreur est l'innovation forte :

$$\epsilon_t = B_{00}^{-1} A_{00} (X_t - E \{X_t | \mathcal{F}_X(t-1)\}). \quad (1.3)$$

Ceci revient donc à dire que le meilleur prédicteur de X_t est une combinaison linéaire des valeurs passées, que l'on obtient en déroulant (1.1). C'est évidemment une hypothèse extrêmement restrictive car, au sens des moindres carrés, la meilleure prévision de X_t est souvent une fonction non linéaire du passé de X . L'hypothèse (1.3) est cependant d'usage courant pour le traitement statistique des modèles VARMA. Ceci permet d'utiliser la théorie asymptotique des martingales (Hall et Heyde, 1980) : lois des grands nombres, théorème de la limite centrale, . . . Ceci met néanmoins l'accent sur un défaut majeur du modèle VARMA (semi-)fort : il est peu plausible pour un certain nombre de séries temporelles car il est souvent peu réaliste de supposer a priori que le meilleur prédicteur est linéaire.

Approximation de la décomposition de Wold

Au moins comme approximation, sous certaines hypothèses de régularité, les modèles VARMA faibles sont assez généraux car, comme nous allons le voir dans cette partie, l'ensemble des processus VARMA faibles est dense dans l'ensemble des processus stationnaires au second ordre. Rappelons que, sous les hypothèses de (1.2), nous avons

$$X_t = \sum_{\ell=0}^{\infty} \Psi_{\ell} \epsilon_{t-\ell},$$

où (ϵ_t) est le processus des innovations linéaires du processus (X_t) . Ainsi pour approximer le processus (X_t) , considérons le processus suivant

$$X_t(q_0) = \Psi_0 \epsilon_t + \sum_{\ell=1}^{q_0} \Psi_{\ell} \epsilon_{t-\ell}.$$

Il est important de remarquer que (ϵ_t) est aussi le processus des innovations linéaires de $(X_t(q_0))$ si $\det \Psi(z)$ a ses racines à l'extérieur du disque unité, où $\Psi(z) = \Psi_0 + \sum_{i=1}^{q_0} \Psi_i z^i$. Si ce n'est pas le cas, il faut changer le bruit et les coefficients des matrices pour

obtenir la représentation explicite. Le processus $(X_t(q_0))$ est une $MA(q_0)$ vectorielle faible car $X_t(q_0)$ et $X_{t-h}(q_0)$ sont non corrélés pour $h > q_0$. Notons $\|\cdot\|_2$ la norme définie par $\|Z\|_2 = \sqrt{E\|Z\|^2}$, où Z est un vecteur aléatoire de dimension d . Comme $\sum_{\ell=0}^{\infty} \|\Psi_\ell\|^2 < \infty$, nous avons

$$\|X_t(q_0) - X_t\|_2^2 = E\|\epsilon_t\|^2 \sum_{i>q_0} \|\Psi_i\|^2 \rightarrow 0 \quad \text{quand } q_0 \rightarrow \infty.$$

Par conséquent l'ensemble des moyennes mobiles finies faibles est dense dans l'ensemble des processus stationnaires au second ordre et réguliers. La même propriété est évidemment vraie pour l'ensemble des modèles VARMA faibles. Trivialement, en suivant le principe de parcimonie qui nous fait préférer un modèle avec moins de paramètres, on utilise plutôt les modèles VARMA qui permettent parfois d'obtenir la même précision en utilisant moins de paramètres. Ainsi le modèle linéaire (1.2), qui regroupe les modèles VARMA et leurs limites, est très général si le terme d'erreur ϵ_t correspond à ce qui n'est pas prévu linéairement, il est par contre peu réaliste si le terme d'erreur est ce que l'on ne peut prévoir par aucune fonction du passé.

Agrégation temporelle des modèles VARMA faibles

La plupart des séries économiques, et plus particulièrement les séries financières, sont analysées à différentes fréquences (jour, semaine, mois, . . .). Le choix de la fréquence d'observation a souvent une importance cruciale quant aux propriétés de la série étudiée et, par suite au type de modèle adapté. Dans le cas des modèles GARCH, les travaux empiriques font généralement apparaître une persistance plus forte lorsque la fréquence augmente.

Du point de vue théorique, le problème de l'agrégation temporelle peut être posé de la manière suivante : étant donné un processus (X_t) et un entier m , quelles sont les propriétés du processus échantillonné (X_{mt}) (*i.e.* construit à partir de (X_t) en ne retenant que les dates multiples de m) ? Lorsque, pour tout entier m et pour tout modèle d'une classe donnée, admettant (X_t) comme solution, il existe un modèle de la même classe dont (X_{mt}) soit solution, cette classe est dite stable par agrégation temporelle.

Un exemple très simple de modèle stable par agrégation temporelle est évidemment le bruit blanc (fort ou faible) : la propriété d'indépendance (ou de non corrélation) subsiste lorsque l'on passe d'une fréquence donnée à une fréquence plus faible. Par contre, les modèles ARMA au sens fort ne sont généralement pas stables par agrégation temporelle. Ce n'est qu'en relâchant l'hypothèse d'indépendance du bruit (ARMA faibles)

qu'on obtient l'agrégation temporelle (voir Drost et Nijman (1993) pour le cas univarié de processus GARCH, Nijman et Enrique (1994) pour une extension au cas multivarié).

Exemples de processus ARMA forts vectoriels admettant une représentation faible

Certaines transformations de processus VARMA forts, comme par exemple des agrégations temporelles, des échantillonnages ou encore des transformations linéaires plus générales, admettent des représentations VARMA faibles. Les exemples suivants illustrent le cas où l'on considère des sous vecteurs du processus observé.

Exemple 1.1. Soit un processus (X_t) de dimension $d = d_1 + d_2$ qui satisfait un modèle VARMA(1,1) fort de la forme

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} 0_{d_1 \times d_1} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} X_{1t-1} \\ X_{2t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} - \begin{pmatrix} 0_{d_1 \times d_1} & 0_{d_1 \times d_2} \\ B_{21} & 0_{d_2 \times d_2} \end{pmatrix} \begin{pmatrix} \epsilon_{1t-1} \\ \epsilon_{2t-1} \end{pmatrix},$$

où X_{1t}, ϵ_{1t} (resp. X_{2t}, ϵ_{2t}) sont des vecteurs de dimension d_1 (resp. d_2) et les matrices A_{12}, A_{21}, A_{22} et B_{21} sont de tailles appropriées. Si l'on suppose que $\epsilon_t = (\epsilon'_{1t}, \epsilon'_{2t})'$ est un bruit blanc fort de matrice variance-covariance

$$\Omega_\epsilon = \begin{pmatrix} \Omega_{11} & 0_{d_1 \times d_2} \\ 0_{d_2 \times d_1} & \Omega_{22} \end{pmatrix},$$

alors X_{2t} satisfait le modèle VAR(2) suivant

$$X_{2t} = A_{22}X_{2t-1} + A_{21}A_{12}X_{2t-2} + \nu_t,$$

où $\nu_t = \epsilon_{2t} + (A_{21} - B_{21})\epsilon_{1t-1}$ est non corrélé. Cependant dès que l'on suppose qu'il existe une dépendance entre ϵ_{2t} et ϵ_{1t} , et que $A_{21} \neq B_{21}$, le processus (ν_t) n'est plus un bruit blanc fort.

Contribution de la thèse

L'objectif principal de la thèse est d'étudier dans quelle mesure les travaux cités précédemment sur l'analyse statistique des modèles ARMA peuvent s'étendre au cas multivarié faible. En particulier, nous étudions de quelle manière il convient d'adapter

les outils statistiques standard pour l'estimation et l'inférence des modèles VARMA faibles.

Le deuxième chapitre est consacré à établir les propriétés asymptotiques des estimateurs du quasi-maximum de vraisemblance (en anglais QMLE) et des moindres carrés (noté LSE pour Least Squares Estimator), afin de les comparer à celles du cas standard. Ensuite, nous accordons une attention particulière à l'estimation de la matrice de variance asymptotique de ces estimateurs. Cette matrice est de la forme "sandwich" $\Omega := J^{-1}IJ^{-1}$, et peut être très différente de la variance asymptotique standard, dont la forme est $\Omega := 2J^{-1}$. Enfin des versions modifiées des tests de Wald, du multiplicateur de Lagrange et du rapport de vraisemblance sont proposées pour tester des restrictions linéaires sur les paramètres libres du modèle.

Dans le troisième chapitre, nous considérons de façon plus détaillée le problème de l'estimation de la matrice de variance asymptotique Ω des estimateurs QML/LS d'un modèle VARMA sans faire l'hypothèse d'indépendance habituelle sur le bruit. Dans un premier temps, nous proposons une expression des dérivées des résidus en fonction des paramètres du modèle VARMA. Ceci nous permettra ensuite de donner une expression explicite des matrices I et J impliquées dans la variance asymptotique Ω , en fonction des paramètres des polynômes VAR et MA, et des moments d'ordre deux et quatre du bruit. Enfin nous en déduisons un estimateur de Ω , dont nous établissons la convergence.

Dans le quatrième chapitre, nous nous intéressons au problème de la validation des modèles VARMA faibles. Dans un premier temps, nous étudions la distribution asymptotique jointe de l'estimateur du QML et des autocovariances empiriques du bruit. Ceci nous permet d'obtenir les distributions asymptotiques des autocovariances et autocorrélations résiduelles. Ces autocorrélations résiduelles sont normalement distribuées avec une matrice de covariance différente du cas iid. Enfin, nous déduisons le comportement asymptotique des statistiques portmanteau. Dans le cadre standard (c'est-à-dire sous les hypothèses iid sur le bruit), il est connu que la distribution asymptotique des tests portmanteau est approximée par un chi-deux. Dans le cas général, nous montrons que cette distribution asymptotique est celle d'une somme pondérée de chi-deux. Cette distribution peut être très différente de l'approximation chi-deux usuelle du cas fort. Nous en déduisons des tests portmanteau modifiés pour tester l'adéquation de modèles VARMA faibles.

Le cinquième chapitre est consacré au problème très important de sélection de modèles en séries chronologiques. Depuis des années, ce champ de recherche suscite un intérêt croissant au sein de la communauté des économètres et des statisticiens. En attestent les nombreuses publications scientifiques dans ce domaine et les différents

champs d'application ouverts. Les praticiens se sont toujours trouvés confrontés au problème du choix d'un modèle adéquat leur permettant de décrire le mécanisme ayant généré les observations et de faire des prévisions. Ces questions concernent les différents domaines d'application des séries temporelles, tels que la communication, la climatologie, l'épidémiologie, les systèmes de contrôle, l'économétrie et la finance. Afin de résoudre ce problème de choix de modèles, nous proposons une modification du critère d'information de Akaike (AIC pour Akaike's Information Criterion). Ce AIC modifié est fondé sur une généralisation du critère AIC corrigé introduit par Tsai et Hurvich (1989).

Nous illustrons enfin nos résultats sur des études empiriques basées sur des expériences de Monte Carlo. Les logiciels de prévision qui utilisent la méthodologie de Box et Jenkins (identification, estimation et validation de modèles ARMA forts) ne conviennent pas aux ARMA faibles. Il nous faudra donc déterminer comment adapter les sorties de ces logiciels.

Le dernier chapitre propose des perspectives pour des développements futurs.

Nous présentons maintenant nos résultats de façon plus détaillée.

1.1 Résultats du chapitre 2

L'objectif de ce chapitre est d'étudier, dans un premier temps, les propriétés asymptotiques du QMLE et du LSE des paramètres d'un modèle VARMA sans faire l'hypothèse d'indépendance sur le bruit, contrairement à ce qui est fait habituellement pour l'inférence de ces modèles. Relâcher cette hypothèse permet aux modèles VARMA faibles de couvrir une large classe de processus non linéaires. Ensuite, nous accordons une attention particulière à l'estimation de la matrice de variance asymptotique de forme "sandwich" $\Omega := J^{-1}IJ^{-1}$. Nous établissons la convergence d'un estimateur de Ω . Enfin, des versions modifiées des tests de Wald, du multiplicateur de Lagrange et du rapport de vraisemblance sont proposées pour tester des restrictions linéaires sur les paramètres libres du modèle.

1.1.1 Estimation des paramètres

Pour l'estimation des paramètres, nous utiliserons la méthode du quasi-maximum de vraisemblance, qui est la méthode du maximum de vraisemblance gaussien lorsque l'hypothèse de bruit blanc gaussien est relâchée. Nous présentons l'estimation du modèle ARMA(p_0, q_0) structurel d -multivarié (1.1) dans lequel $X_t = (X_{1t}, \dots, X_{dt})'$ est un processus vectoriel stationnaire au second ordre, à valeurs réelles. Le terme d'erreur $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{dt})'$ est une suite de variables aléatoires centrées, non corrélées, avec une matrice de covariance non singulière Σ . Les paramètres sont les coefficients des matrices carrées d'ordre d suivantes : $A_i, i \in \{1, \dots, p_0\}, B_j, j \in \{1, \dots, q_0\}$ et Σ . Avant de détailler la procédure d'estimation et ses propriétés, nous étudions quelles conditions imposer aux matrices A_i, B_j et Σ afin d'assurer l'identifiabilité du modèle, c'est-à-dire l'unicité des $(p_0 + q_0 + 3)d^2$ paramètres du modèle. Nous supposons que ces matrices sont paramétrées suivant le vecteur des vraies valeurs des paramètres noté ϑ_0 . Nous notons $A_{0i} = A_i(\vartheta_0), i \in \{1, \dots, p_0\}, B_j = B_j(\vartheta_0), j \in \{1, \dots, q_0\}$ et $\Sigma_0 = \Sigma(\vartheta_0)$, où ϑ_0 appartient à l'espace compact des paramètres $\Theta \subset \mathbb{R}^{k_0}$, et k_0 est le nombre de paramètres inconnus qui est inférieur à $(p_0 + q_0 + 3)d^2$. Pour tout $\vartheta \in \Theta$, nous supposons aussi que les applications $\vartheta \mapsto A_i(\vartheta) i = 0, \dots, p_0, \vartheta \mapsto B_j(\vartheta) j = 0, \dots, q_0$ et $\vartheta \mapsto \Sigma(\vartheta)$ admettent des dérivées continues d'ordre 3.

Conditions d'identifiabilité

Pour la suite, nous notons $A_i(\vartheta), B_j(\vartheta)$ et $\Sigma(\vartheta)$ par A_i, B_j et Σ . Soit les polynômes $A_\vartheta(z) = A_0 - \sum_{i=1}^{p_0} A_i z^i$ et $B_\vartheta(z) = B_0 - \sum_{i=1}^{q_0} B_i z^i$. Dans le cadre de modèles VARMA(p_0, q_0) structurels, l'hypothèse \mathbf{H}_1 qui assure que les polynômes $A_{\vartheta_0}(z)$ et $B_{\vartheta_0}(z)$ n'ont pas de racine commune, ne suffit pas à garantir l'identifiabilité du paramètre, c'est-à-dire à garantir que seule la valeur $\vartheta = \vartheta_0$ satisfasse $A_\vartheta(L)X_t = B_\vartheta(L)\epsilon_t$. On suppose donc que

\mathbf{H}_2 : pour tout $\vartheta \in \Theta$ tel que $\vartheta \neq \vartheta_0$, soit les fonctions de transfert

$$A_0^{-1} B_0 B_\vartheta^{-1}(z) A_\vartheta(z) \neq A_{00}^{-1} B_{00} B_{\vartheta_0}^{-1}(z) A_{\vartheta_0}(z) \text{ pour un } z \in \mathbb{C}$$

ou

$$A_0^{-1} B_0 \Sigma B_0' A_0^{-1'} \neq A_{00}^{-1} B_{00} \Sigma_0 B_{00}' A_{00}^{-1'}.$$

Cette condition équivaut à l'existence d'un opérateur $U(L)$ tel que

$$A_\vartheta(L) = U(L)A_{\vartheta_0}(L), A_0 = U(1)A_{00}, B_\vartheta(L) = U(L)B_{\vartheta_0}(L) \text{ et } B_0 = U(1)B_{00}.$$

On dit que $U(L)$ est unimodulaire si $\det\{U(L)\}$ est une constante non nulle. Lorsque les seuls facteurs communs à deux polynômes $P(L)$ et $Q(L)$ sont unimodulaires, c'est-à-dire

si

$$P(L) = U(L)P_1(L), \quad Q(L) = U(L)Q_1(L) \implies \det\{U(L)\} = cste \neq 0,$$

on dit que $P(L)$ et $Q(L)$ sont coprimés à gauche (en anglais left-coprime). Divers types de conditions peuvent être imposées pour assurer l'identifiabilité (voir Reinsel, 1997, p. 37-40, ou Hannan (1971, 1976), Hannan et Deistler, 1988, section 2.7). Pour avoir identifiabilité, il est parfois nécessaire d'introduire des formes plus contraintes, comme par exemple la forme échelon. Pour des renseignements complémentaires sur les formes identifiables, et en particulier la forme échelon, on peut par exemple se référer à Hannan (1976), Hannan et Deistler (1976), Dufour J-M. et Pelletier D. (2005), Lütkepohl (1991, p. 246-247 et 289-297, 2005, p. 452-453 et 498-507) et Reinsel (1997).

Définition de l'estimateur du QML

Pour des raisons pratiques, nous écrivons le modèle sous sa forme réduite

$$X_t - \sum_{i=1}^{p_0} A_{00}^{-1} A_{0i} X_{t-i} = e_t - \sum_{j=1}^{q_0} A_{00}^{-1} B_{0j} B_{00}^{-1} A_{00} e_{t-j}, \quad e_t = A_{00}^{-1} B_{00} \epsilon_t.$$

On dispose d'une observation de longueur n , X_1, \dots, X_n . Pour $0 < t \leq n$ et pour tout $\vartheta \in \Theta$, les variables aléatoires $\tilde{e}_t(\vartheta)$ sont définies récursivement par

$$\tilde{e}_t(\vartheta) = X_t - \sum_{i=1}^{p_0} A_0^{-1} A_i X_{t-i} + \sum_{j=1}^{q_0} A_0^{-1} B_j B_0^{-1} A_0 \tilde{e}_{t-j}(\vartheta),$$

où les valeurs initiales inconnues sont remplacées par zéro : $\tilde{e}_0(\vartheta) = \dots = \tilde{e}_{1-q_0}(\vartheta) = X_0 = \dots = X_{1-p_0} = 0$. Nous montrons que, comme dans le cas univarié, ces valeurs initiales sont asymptotiquement négligeables et, en particulier, que $\tilde{e}_t(\vartheta_0) - e_t \rightarrow 0$ presque sûrement quand $t \rightarrow \infty$. Ainsi, le choix des valeurs initiales est sans effet sur les propriétés asymptotiques de l'estimateur. La quasi-vraisemblance gaussienne s'écrit

$$\tilde{L}_n(\vartheta) = \prod_{t=1}^n \frac{1}{(2\pi)^{d/2} \sqrt{\det \Sigma_e}} \exp \left\{ -\frac{1}{2} \tilde{e}_t'(\vartheta) \Sigma_e^{-1} \tilde{e}_t(\vartheta) \right\},$$

où $\Sigma_e = \Sigma_e(\vartheta) = A_0^{-1} B_0 \Sigma B_0' A_0^{-1}$. Un estimateur du QML de ϑ_0 est défini comme toute solution mesurable $\hat{\vartheta}_n$ de

$$\hat{\vartheta}_n = \arg \max_{\vartheta \in \Theta} \tilde{L}_n(\vartheta) = \arg \min_{\vartheta \in \Theta} \tilde{\ell}_n(\vartheta), \quad \tilde{\ell}_n(\vartheta) = \frac{-2}{n} \log \tilde{L}_n(\vartheta).$$

Propriétés asymptotiques de l'estimateur du QML/LS

Afin d'établir la convergence forte de l'estimateur du QML, il est utile d'approximer la suite $(\tilde{\epsilon}_t(\vartheta))$ par une suite stationnaire ergodique. C'est ainsi que nous faisons l'hypothèse d'ergodicité suivante

H₃ : Le processus (ϵ_t) est stationnaire et ergodique.

Nous pouvons maintenant énoncer le théorème de convergence forte suivant.

Théorème 1.1. (Convergence forte) Sous les hypothèses **H₁**, **H₂** et **H₃**, soit (X_t) une solution causale ou non anticipative de l'équation (1.1) et soit $(\hat{\vartheta}_n)$ une suite d'estimateurs du QML. Alors, nous avons presque sûrement $\hat{\vartheta}_n \rightarrow \vartheta_0$ quand $n \rightarrow \infty$.

Ainsi comme dans le cas où le processus des innovations est un bruit blanc fort, nous obtenons la consistance forte de l'estimateur $\hat{\vartheta}_n$ dans le cas faible.

Nous montrons ce théorème en appliquant le théorème 1.12⁶, les conditions d'identifiabilité et l'inégalité élémentaire $\text{Tr}(A^{-1}B) - \log \det(A^{-1}B) \geq \text{Tr}(A^{-1}A) - \log \det(A^{-1}A) = d$ pour toute matrice symétrique semi-définie positive de taille $d \times d$.

L'ergodicité ne suffit pas pour obtenir un théorème central limite (noté TCL dans la suite). Pour la normalité asymptotique du QMLE, nous avons donc besoin des hypothèses supplémentaires. Il est d'abord nécessaire de supposer que le paramètre ϑ_0 n'est pas situé sur le bord de l'espace compact des paramètres Θ .

H₄ : Nous avons $\vartheta_0 \in \overset{\circ}{\Theta}$, où $\overset{\circ}{\Theta}$ est l'intérieur de Θ .

Nous introduisons les coefficients de mélange fort d'un processus vectoriel $X = (X_t)$ définis par

$$\alpha_X(h) = \sup_{A \in \sigma(X_u, \leq t), B \in \sigma(X_u, \geq t+h)} |P(A \cap B) - P(A)P(B)|.$$

On dit que X est fortement mélangeant si $\alpha_X(h) \rightarrow 0$ quand $h \rightarrow \infty$. Nous considérons l'hypothèse suivante qui nous permet de contrôler la dépendance dans le temps du processus (ϵ_t) .

H₅ : Il existe un réel $\nu > 0$ tel que $E\|\epsilon_t\|^{4+2\nu} < \infty$ et les coefficients de mélange du processus (ϵ_t) vérifient $\sum_{k=0}^{\infty} \{\alpha_{\epsilon}(k)\}^{\frac{\nu}{2+\nu}} < \infty$.

Notons que d'après l'hypothèse **H₁** et l'équation (1.1), le processus (X_t) peut s'écrire comme une combinaison linéaire de $\epsilon_t, \epsilon_{t-1}, \dots$, ce qui implique, d'après **H₅** que le processus (X_t) est mélangeant. Cette hypothèse **H₅** est donc plus faible que l'hypothèse de bruit blanc fort de point de vue de la dépendance. D'autre part, nous faisons l'hypothèse d'existence de moments d'ordre 4^+ (supérieur à quatre), c'est-à-dire $E\|\epsilon_t\|^{4+2\nu} < \infty$

6. Théorème ergodique, voir annexe.

pour un réel $\nu > 0$, qui est légèrement plus forte que l'hypothèse d'existence de moments d'ordre 4 qui est faite dans le cas standard. Étant donné l'hypothèse \mathbf{H}_1 sur les racines des polynômes, il est clair que les hypothèses de moment $E\|\epsilon_t\|^{4+2\nu} < \infty$ et $E\|X_t\|^{4+2\nu} < \infty$ sont équivalentes. Par contre, les hypothèses de mélanges ne sont pas équivalentes.

Nous définissons la matrice des coefficients de la forme réduite du modèle par

$$M_{\vartheta_0} = [A_{00}^{-1}A_{01} : \cdots : A_{00}^{-1}A_{0p} : A_{00}^{-1}B_{01}B_{00}^{-1}A_{00} : \cdots : A_{00}^{-1}B_{0q}B_{00}^{-1}A_{00} : \Sigma_{\epsilon_0}].$$

Ainsi nous avons besoin d'une hypothèse qui spécifie comment cette matrice dépend du paramètre ϑ_0 . Soit \dot{M}_{ϑ_0} la matrice $\partial \text{vec}(M_{\vartheta}) / \partial \vartheta'$ appliquée en ϑ_0 .

\mathbf{H}_6 : La matrice \dot{M}_{ϑ_0} est de plein rang k_0 .

Nous pouvons maintenant énoncer le théorème suivant qui nous donne le comportement asymptotique de l'estimateur du QML.

Théorème 1.2. (Normalité asymptotique) Sous les hypothèses du Théorème 1.1, \mathbf{H}_4 , \mathbf{H}_5 et \mathbf{H}_6 , nous avons

$$\sqrt{n} \left(\hat{\vartheta}_n - \vartheta_0 \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Omega := J^{-1} I J^{-1}),$$

où $J = J(\vartheta_0)$ et $I = I(\vartheta_0)$, avec

$$J(\vartheta) = \lim_{n \rightarrow \infty} \frac{\partial^2}{\partial \vartheta \partial \vartheta'} \tilde{\ell}_n(\vartheta) \quad p.s., \quad I(\vartheta) = \lim_{n \rightarrow \infty} \text{Var} \frac{\partial}{\partial \vartheta} \tilde{\ell}_n(\vartheta).$$

La démonstration de ce théorème repose sur le théorème 1.13⁷ de Herrndorf (1984), le théorème 1.1 et l'inégalité⁸ de Markov (1968) pour les processus mélangeants.

Remarque 1.1. Dans le cas standard de modèles VARMA forts, i.e. quand l'hypothèse \mathbf{H}_3 est remplacée par celle que les termes d'erreur (ϵ_t) sont iid, nous avons $I = 2J$, ainsi $\Omega = 2J^{-1}$. Par contre dans le cas général, nous avons $I \neq 2J$. Comme conséquence, les logiciels utilisés pour ajuster les modèles VARMA forts ne fournissent pas une estimation correcte de la matrice Ω pour le cas de processus VARMA faibles. C'est aussi le même problème dans le cas univarié (voir Francq et Zakoïan (2007), et d'autres références).

Notons que, pour les modèles VARMA sous la forme réduite, il n'est pas restrictif de supposer que les coefficients $A_0, \dots, A_{p_0}, B_0, \dots, B_{q_0}$ sont fonctionnellement indépendants du coefficient Σ_{ϵ} . Ceci nous permet de faire l'hypothèse suivante

7. Voir TCL pour processus α -mélangeant en annexe

8. Voir inégalité de covariance en annexe

H₇ : Posons $\vartheta = (\vartheta^{(1)'}, \vartheta^{(2)'})'$, où $\vartheta^{(1)} \in \mathbb{R}^{k_1}$ dépend des coefficients A_0, \dots, A_{p_0} et B_0, \dots, B_{q_0} , et où $\vartheta^{(2)} = D \text{vec } \Sigma_e \in \mathbb{R}^{k_2}$ dépend uniquement de Σ_e , pour une matrice D de taille $k_2 \times d^2$, avec $k_1 + k_2 = k_0$.

Par abus de notation, nous écrivons $e_t(\vartheta) = e_t(\vartheta^{(1)})$. Le produit de Kronecker de deux matrices A et B est noté par $A \otimes B$ (noté dans la suite $A^{\otimes 2}$ quand $A = B$). Le théorème suivant montre que pour des modèles VARMA sous la forme réduite, les estimateurs du QML et du LS coïncident.

Théorème 1.3. Sous les hypothèses du Théorème 1.2 et **H₇**, le QMLE $\hat{\vartheta}_n = (\hat{\vartheta}_n^{(1)'}, \hat{\vartheta}_n^{(2)'})'$ peut être obtenu par

$$\hat{\vartheta}_n^{(2)} = D \text{vec } \hat{\Sigma}_e, \quad \hat{\Sigma}_e = \frac{1}{n} \sum_{t=1}^n \tilde{e}_t(\hat{\vartheta}_n^{(1)}) \tilde{e}_t'(\hat{\vartheta}_n^{(1)}),$$

et

$$\hat{\vartheta}_n^{(1)} = \arg \min_{\vartheta^{(1)}} \det \sum_{t=1}^n \tilde{e}_t(\vartheta^{(1)}) \tilde{e}_t'(\vartheta^{(1)}).$$

De plus, nous avons

$$J = \begin{pmatrix} J_{11} & 0 \\ 0 & J_{22} \end{pmatrix}, \quad \text{avec } J_{11} = 2E \left\{ \frac{\partial}{\partial \vartheta^{(1)}} e_t'(\vartheta_0^{(1)}) \right\} \Sigma_{e_0}^{-1} \left\{ \frac{\partial}{\partial \vartheta^{(1)'}} e_t(\vartheta_0^{(1)}) \right\}$$

et $J_{22} = D(\Sigma_{e_0}^{-1} \otimes \Sigma_{e_0}^{-1})D'$.

Remarque 1.2. Notons que la matrice J a la même expression dans les cas de modèles ARMA fort et faible (voir Lütkepohl (2005) page 480). Contrairement, à la matrice I qui est en général plus compliquée dans le cas faible que dans le cas fort.

1.1.2 Estimation de la matrice de variance asymptotique

Afin d'obtenir des intervalles de confiance ou de tester la significativité des coefficients VARMA faibles, il sera nécessaire de disposer d'un estimateur au moins faiblement consistant de la matrice de variance asymptotique Ω .

Estimation de la matrice J

La matrice J peut facilement être estimée par

$$\hat{J} = \begin{pmatrix} \hat{J}_{11} & 0 \\ 0 & \hat{J}_{22} \end{pmatrix}, \quad \hat{J}_{22} = D(\hat{\Sigma}_{e_0}^{-1} \otimes \hat{\Sigma}_{e_0}^{-1})D' \quad \text{et}$$

$$\hat{J}_{11} = \frac{2}{n} \sum_{t=1}^n \left\{ \frac{\partial}{\partial \vartheta^{(1)}} \tilde{e}'_t(\hat{\vartheta}_n^{(1)}) \right\} \hat{\Sigma}_{e0}^{-1} \left\{ \frac{\partial}{\partial \vartheta^{(1)}} \tilde{e}_t(\hat{\vartheta}_n^{(1)}) \right\}.$$

Dans le cas standard des VARMA forts $\hat{\Omega} = 2\hat{J}^{-1}$ est un estimateur fortement convergent de Ω . Dans le cas général des VARMA faibles, cet estimateur n'est pas convergent quand $I \neq 2J$.

Estimation de la matrice I

Nous avons besoin d'un estimateur convergent de la matrice I . Notons que

$$I = \text{var}_{as} \frac{1}{\sqrt{n}} \sum_{t=1}^n \Upsilon_t = \sum_{h=-\infty}^{+\infty} \text{cov}(\Upsilon_t, \Upsilon_{t-h}), \quad (1.4)$$

où

$$\Upsilon_t = \frac{\partial}{\partial \vartheta} \left\{ \log \det \Sigma_e + e'_t(\vartheta^{(1)}) \Sigma_e^{-1} e_t(\vartheta^{(1)}) \right\}_{\vartheta=\vartheta_0}.$$

L'existence de la somme des covariances dans (1.4) est une conséquence de l'hypothèse \mathbf{H}_5 et de l'inégalité de covariance de Davydov (1968). Cette matrice I est plus délicate à estimer. Néanmoins nous nous reposons sur deux méthodes suivantes pour l'estimer

- la méthode d'estimation non paramétrique du noyau aussi appelée HAC (Heteroscedasticity and Autocorrelation Consistent),
- la méthode d'estimation paramétrique de la densité spectrale (notée SP).

Définition de l'estimateur HAC de la matrice I

Dans la littérature économique, l'estimateur non paramétrique du noyau (aussi appelé HAC) est largement utilisé pour estimer une matrice de covariance de la forme de I . La technique que nous utilisons consiste à pondérer convenablement certains moments empiriques. Cette pondération se fait au moyen d'une suite de poids (ou fenêtre). Cette approche est semblable à celle de Andrews (1991), Newey et West (1994). Soit $\hat{\Upsilon}_t$ le vecteur obtenu en remplaçant ϑ_0 par $\hat{\vartheta}_n$ dans Υ_t . La matrice Ω est donc estimée par un estimateur "sandwich" de la forme

$$\hat{\Omega}^{\text{HAC}} = \hat{J}^{-1} \hat{I}^{\text{HAC}} \hat{J}^{-1}, \quad \hat{I}^{\text{HAC}} = \frac{1}{n} \sum_{t,s=1}^n \omega_{|t-s|} \hat{\Upsilon}_t \hat{\Upsilon}'_s,$$

où $\omega_0, \dots, \omega_{n-1}$ est une suite de poids (ou fenêtre).

Définition de l'estimateur SP de la matrice I

Une méthode alternative consiste à utiliser un modèle VAR paramétrique estimé de la densité spectrale du processus (Υ_t) . Cette approche a été étudiée par Berk (1974) (voir aussi den Haan et Levin, 1997). Ceci revient à interpréter $(2\pi)^{-1}I$ comme étant la densité spectrale évaluée à la fréquence 0 du processus stationnaire (Υ_t) (voir Brockwell et Davis, 1991, p. 459). Nous nous basons sur l'expression

$$I = \Phi^{-1}(1)\Sigma_u\Phi^{-1}(1)$$

quand (Υ_t) satisfait une représentation VAR(∞) de la forme

$$\Phi(L)\Upsilon_t := \Upsilon_t + \sum_{i=1}^{\infty} \Phi_i \Upsilon_{t-i} = u_t,$$

où u_t est un bruit blanc faible de matrice covariance Σ_u . Notons $\hat{\Phi}_{r,1}, \dots, \hat{\Phi}_{r,r}$ les coefficients de la régression des moindres carrés de $\hat{\Upsilon}_t$ sur $\hat{\Upsilon}_{t-1}, \dots, \hat{\Upsilon}_{t-r}$ et posons $\hat{\Phi}_r(z) = I_{k_0-s} + \sum_{i=1}^r \hat{\Phi}_{r,i} z^i$. Soit $\hat{u}_{r,t}$ les résidus de cette régression et $\hat{\Sigma}_{\hat{u}_r}$ la variance empirique de $\hat{u}_{r,1}, \dots, \hat{u}_{r,n}$. Nous énonçons maintenant le théorème suivant qui est une extension de celui de Francq, Roy et Zakoïan (2005).

Théorème 1.4. (Convergence faible de I) Sous les hypothèses du Théorème 1.3, nous supposons que le processus (Υ_t) admet une représentation VAR(∞) pour laquelle les racines de $\det \Phi(z) = 0$ sont à l'extérieur du disque unité, $\|\Phi_i\| = o(i^{-2})$, et que la matrice $\Sigma_u = \text{Var}(u_t)$ est non singulière. Nous supposons également que $\|\epsilon_t\|_{8+4\nu} < \infty$ et $\sum_{k=0}^{\infty} \{\alpha_{X,\epsilon}(k)\}^{\nu/(2+\nu)} < \infty$ pour un réel $\nu > 0$, où $\{\alpha_{X,\epsilon}(k)\}_{k \geq 0}$ désigne la suite des coefficients de mélange fort du processus $(X'_t, \epsilon'_t)'$. Alors, l'estimateur

$$\hat{I}^{\text{SP}} := \hat{\Phi}_r^{-1}(1)\hat{\Sigma}_{\hat{u}_r}\hat{\Phi}_r'^{-1}(1) \rightarrow I$$

en probabilité quand $r = r(n) \rightarrow \infty$ et $r^3/n \rightarrow 0$ quand $n \rightarrow \infty$.

La démonstration s'inspire de celle de Francq, Roy et Zakoïan (2005) et repose sur une série de lemmes.

1.1.3 Tests sur les coefficients du modèle

En dehors des contraintes imposées pour assurer l'identifiabilité du modèle VARMA(p_0, q_0), d'autres contraintes linéaires peuvent être testées sur les paramètres

ϑ_0 du modèle (en particulier $A_{0p} = 0$ ou $B_{0q} = 0$). Afin de tester s_0 contraintes linéaires sur les coefficients, nous définissons notre hypothèse nulle

$$H_0 : R_0 \vartheta_0 = \mathbf{r}_0,$$

où R_0 est une matrice $s_0 \times k_0$ connue de rang s_0 et \mathbf{r}_0 est aussi un vecteur connu de dimension s_0 . Pour effectuer ce test $H_0 : R_0 \vartheta_0 = \mathbf{r}_0$, il existe diverses approches asymptotiques dont les plus classiques sont la procédure de Wald (notée W), celle du multiplicateur de Lagrange (en anglais LM) encore appelée du score ou de Rao-score et la méthode du rapport de vraisemblance (notée LR pour likelihood ratio). Pour calculer nos différentes statistiques de test, nous considérons la matrice $\hat{\Omega} = \hat{J}^{-1} \hat{I} \hat{J}^{-1}$, où \hat{J} et \hat{I} sont des estimateurs consistants de J et I définis dans la section 1.1.2.

Statistique de Wald

Le principe est d'accepter l'hypothèse nulle, si l'estimateur non contraint $\hat{\vartheta}_n$ de ϑ_0 est suffisamment proche de zéro. Ainsi de la normalité asymptotique de $\hat{\vartheta}_0$, nous en déduisons que

$$\sqrt{n} \left(R_0 \hat{\vartheta}_n - \mathbf{r}_0 \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, R_0 \Omega R_0' := R_0 \left(J^{-1} I J^{-1} \right) R_0' \right),$$

quand $n \rightarrow \infty$. Sous les hypothèses du théorème 1.3 et l'hypothèse que I est inversible, la statistique de Wald modifiée est

$$\mathbf{W}_n = n \left(R_0 \hat{\vartheta}_n - \mathbf{r}_0 \right)' \left(R_0 \hat{\Omega} R_0' \right)^{-1} \left(R_0 \hat{\vartheta}_n - \mathbf{r}_0 \right).$$

Sous H_0 , cette statistique suit une distribution de $\chi_{s_0}^2$. Ainsi nous retrouvons la même formulation que dans le cas fort standard. Nous rejetons H_0 quand $\mathbf{W}_n > \chi_{s_0}^2(1 - \alpha)$ pour un niveau de risque asymptotique α .

Remarque 1.3. *Notons que le test de Wald nécessite la maximisation de la quasi-vraisemblance non contrainte, mais pas la maximisation de la quasi-vraisemblance contrainte.*

Statistique du LM

L'idée de ce test consiste à accepter l'hypothèse nulle, si le score contraint est proche de zéro. Soit $\hat{\vartheta}_n^c$ le QMLE contraint sous H_0 . Définissons le Lagrangien

$$\mathcal{L}(\vartheta, \lambda) = \tilde{\ell}_n(\vartheta) - \lambda'(R_0 \vartheta - \mathbf{r}_0),$$

où λ est le vecteur multiplicateur de Lagrange de dimension s_0 . Les conditions de premier ordre s'écrivent

$$\frac{\partial \tilde{\ell}_n}{\partial \vartheta}(\hat{\vartheta}_n^c) = R'_0 \hat{\lambda}, \quad R_0 \hat{\vartheta}_n^c = \mathbf{r}_0.$$

Notons $a \stackrel{c}{=} b$ pour signifier que $a = b + c$. En effectuant le développement limité de Taylor sous H_0 , nous avons

$$0 = \sqrt{n} \frac{\partial \tilde{\ell}_n(\hat{\vartheta}_n)}{\partial \vartheta} \stackrel{op(1)}{=} \sqrt{n} \frac{\partial \tilde{\ell}_n(\hat{\vartheta}_n^c)}{\partial \vartheta} - J \sqrt{n} (\hat{\vartheta}_n - \hat{\vartheta}_n^c).$$

Nous déduisons que

$$\sqrt{n}(R_0 \hat{\vartheta}_n - \mathbf{r}_0) = R_0 \sqrt{n}(\hat{\vartheta}_n - \hat{\vartheta}_n^c) \stackrel{op(1)}{=} R_0 J^{-1} \sqrt{n} \frac{\partial \tilde{\ell}_n(\hat{\vartheta}_n^c)}{\partial \vartheta} = R_0 J^{-1} R'_0 \sqrt{n} \hat{\lambda}.$$

Il en résulte, que sous l'hypothèse H_0 et les hypothèses précédentes la normalité asymptotique du vecteur multiplicateur de Lagrange,

$$\sqrt{n} \hat{\lambda} \xrightarrow{\mathcal{L}} \mathcal{N} \left\{ 0, (R_0 J^{-1} R'_0)^{-1} R_0 \Omega R'_0 (R_0 J^{-1} R'_0)^{-1} \right\}, \quad (1.5)$$

ainsi que la statistique du LM modifiée définie par

$$\begin{aligned} \mathbf{LM}_n &= n \hat{\lambda}' \left\{ (R_0 \hat{J}^{-1} R'_0)^{-1} R_0 \hat{\Omega} R'_0 (R_0 \hat{J}^{-1} R'_0)^{-1} \right\}^{-1} \hat{\lambda} \\ &= n \frac{\partial \tilde{\ell}_n}{\partial \vartheta'}(\hat{\vartheta}_n^c) \hat{J}^{-1} R'_0 \left(R_0 \hat{\Omega} R'_0 \right)^{-1} R_0 \hat{J}^{-1} \frac{\partial \tilde{\ell}_n}{\partial \vartheta}(\hat{\vartheta}_n^c). \end{aligned}$$

Rappelons que dans le cas de modèles VARMA forts $\hat{\Omega} = 2\hat{J}^{-1}$. Ce qui donne à la statistique du LM une forme plus conventionnelle $\mathbf{LM}_n^* = (n/2) \hat{\lambda}' R_0 \hat{J}^{-1} R'_0 \hat{\lambda}$. De la convergence (1.5) du vecteur multiplicateur de Lagrange, nous avons la distribution asymptotique de la statistique \mathbf{LM}_n qui suit une distribution $\chi_{s_0}^2$ sous H_0 . Pour un niveau de risque asymptotique α , l'hypothèse nulle est rejetée quand $\mathbf{LM}_n > \chi_{s_0}^2(1 - \alpha)$. Ceci reste valide dans le cas des modèles VARMA forts.

Remarque 1.4. *Notons que pour calculer la statistique du LM, nous remplaçons ϑ_0 par son estimation $\hat{\vartheta}_n^c$ (i.e. $\hat{\vartheta}_n$ sous H_0) dans l'expression de la variance asymptotique de $\sqrt{n} \hat{\lambda}$. Par conséquent, le test du LM nécessite uniquement la maximisation de la quasi-vraisemblance contrainte (i.e. la maximisation de la quasi-vraisemblance non contrainte sous H_0).*

Statistique du LR

Ce test est fondé sur la comparaison des valeurs maximales de la log-quasi-vraisemblance contrainte et non contrainte. L'hypothèse nulle est acceptée, si l'écart

entre les maxima contraint et non contraint de la log-quasi-vraisemblance est assez petit. Un développement limité de Taylor montre que

$$\sqrt{n} \left(\hat{\vartheta}_n - \hat{\vartheta}_n^c \right) \stackrel{op(1)}{=} -\sqrt{n} J^{-1} R_0' \hat{\lambda},$$

et que la statistique du LR satisfait

$$\mathbf{LR}_n := 2 \left\{ \log \tilde{L}_n(\hat{\vartheta}_n) - \log \tilde{L}_n(\hat{\vartheta}_n^c) \right\} \stackrel{op(1)}{=} \frac{n}{2} (\hat{\vartheta}_n - \hat{\vartheta}_n^c)' J (\hat{\vartheta}_n - \hat{\vartheta}_n^c) \stackrel{op(1)}{=} \mathbf{LM}_n^*.$$

En utilisant ces précédents calculs et des résultats standard sur les formes quadratiques de vecteurs (voir *e.g.* le lemme 17.1 de van der Vaart, 1998), nous trouvons que la statistique du LR (*i.e.* \mathbf{LR}_n) suit asymptotiquement une distribution $\sum_{i=1}^{s_0} \lambda_i Z_i^2$ où les Z_i sont iid $\mathcal{N}(0, 1)$ et les $\lambda_1, \dots, \lambda_{s_0}$ sont les valeurs propres de la matrice

$$\Sigma_{\mathbf{LR}} = J^{-1/2} S_{\mathbf{LR}} J^{-1/2}, \quad S_{\mathbf{LR}} = \frac{1}{2} R_0' (R_0 J^{-1} R_0')^{-1} R_0 \Omega R_0' (R_0 J^{-1} R_0')^{-1} R_0.$$

Notons que dans le cas de modèles VARMA forts, *i.e.* quand $\Omega = 2J^{-1}$, la matrice $\Sigma_{\mathbf{LR}} = J^{-1/2} R_0' (R_0 J^{-1} R_0')^{-1} R_0 J^{-1/2}$ est une matrice de projection. Les valeurs propres de cette matrice sont donc, soit égales à 0 ou 1, et le nombre de valeurs propres égales à 1 est $\text{Tr} J^{-1/2} R_0' (R_0 J^{-1} R_0')^{-1} R_0 J^{-1/2} = \text{Tr} I_{s_0} = s_0$. Ainsi, nous retrouvons le résultat bien connu que $\mathbf{LR}_n \sim \chi_{s_0}^2$ sous H_0 dans le cas des modèles VARMA forts. Pour les modèles VARMA faibles la distribution asymptotique est celle d'une forme quadratique de vecteurs gaussiens, qui peut être évaluée avec l'algorithme de Imhof (1961), qui a cependant l'inconvénient de nécessiter beaucoup de temps de calcul. Une alternative est d'utiliser une statistique transformée

$$\frac{n}{2} (\hat{\vartheta}_n - \hat{\vartheta}_n^c)' \hat{J} \hat{S}_{\mathbf{LR}}^- \hat{J} (\hat{\vartheta}_n - \hat{\vartheta}_n^c),$$

où $\hat{S}_{\mathbf{LR}}^-$ est l'inverse généralisée de $\hat{S}_{\mathbf{LR}}$. Cette statistique suit une distribution $\chi_{s_0}^2$ sous H_0 , quand \hat{J} et $\hat{S}_{\mathbf{LR}}$ sont des estimateurs faiblement convergents de J et $S_{\mathbf{LR}}$. L'estimateur $\hat{S}_{\mathbf{LR}}^-$ peut être obtenu à partir d'une décomposition des valeurs singulières d'un estimateur faiblement convergent $\hat{S}_{\mathbf{LR}}$ de $S_{\mathbf{LR}}$. Plus précisément, nous définissons la matrice diagonale $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_{k_0})$ où $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_{k_0}$ sont les valeurs propres de la matrice symétrique $\hat{S}_{\mathbf{LR}}$, et notons \hat{P} la matrice orthogonale telle que $\hat{S}_{\mathbf{LR}} = \hat{P} \hat{\Lambda} \hat{P}'$, nous posons

$$\hat{S}_{\mathbf{LR}}^- = \hat{P} \hat{\Lambda}^- \hat{P}', \quad \hat{\Lambda}^- = \text{diag}(\hat{\lambda}_1^{-1}, \dots, \hat{\lambda}_{s_0}^{-1}, 0, \dots, 0).$$

Alors, la matrice $\hat{S}_{\mathbf{LR}}^-$ converge faiblement vers la matrice $S_{\mathbf{LR}}^-$, laquelle satisfait $S_{\mathbf{LR}} S_{\mathbf{LR}}^- S_{\mathbf{LR}} = S_{\mathbf{LR}}$, puisque $S_{\mathbf{LR}}$ est de plein rang s_0 .

Remarque 1.5. Notons que le test du LR nécessite à la fois la maximisation de la quasi-vraisemblance contrainte et non contrainte, et en plus l'estimation de l'inverse généralisée de la matrice $S_{\mathbf{LR}}$ ou l'utilisation de l'algorithme de Imhof (1961).

Remarque 1.6. (Comparaison des tests) *Notons que le choix parmi ces trois principes peut être fondé sur les propriétés de ces tests qui sont difficiles à expliciter. Une alternative est de se reposer sur la simplicité de calcul de la statistique. Ainsi au vue des remarques 1.3, 1.4 et 1.5, les procédures de Wald et du multiplicateur de Lagrange se révèlent les plus simples. En général, la procédure du multiplicateur de Lagrange qui nécessite le calcul de l'estimateur contraint se révèle la plus simple.*

Enfin, nous illustrons nos résultats de ce chapitre 2 par une étude empirique basée sur des expériences de Monte Carlo réalisée sur de modèles VARMA(1, 1) bi-variés fort et faible sous la forme échelon dont le vecteur de paramètres à estimer $\vartheta = (a_1(2, 2), b_1(2, 1), b_1(2, 2))'$. L'innovation dans le cas du modèle fort est $\epsilon_t \sim \text{IID } \mathcal{N}(0, I_2)$ et dans le cas du modèle faible, l'innovation est

$$\epsilon_{i,t} = \eta_{i,t}(|\eta_{i,t-1}| + 1)^{-1}, \quad i = 1, 2 \quad \text{avec } \eta_t \sim \text{IID } \mathcal{N}(0, I_2).$$

Ce bruit blanc faible est une extension directe de celui défini par Romano et Thombs (1996) dans le cas univarié. Ces expériences montrent que les distributions de $\hat{a}_1(2, 2)$ et $\hat{b}_1(2, 1)$ sont similaires dans les cas fort et faible. Par contre le QMLE de $\hat{b}_1(2, 2)$ est plus précis dans le cas faible que dans le cas fort. Dans des simulations similaires avec comme bruit blanc faible $\epsilon_{i,t} = \eta_{i,t}\eta_{i,t-1}$ $i = 1, 2$, le QMLE de $\hat{b}_1(2, 2)$ est plus précis dans le cas fort que dans le cas faible. Ceci est en concordance avec les résultats de Romano et Thombs (1996) qui ont montré que, avec ce genre de bruit blanc faible la variance asymptotique des autocorrélations de l'échantillon peut être supérieure ou égale à 1 alors que dans le cas fort, cette variance asymptotique est 1. Ces expériences de Monte Carlo réalisées montrent aussi que les tests standard de Wald, du LM et du LR ne sont plus valides quand les innovations présentent des dépendances temporelles.

1.2 Résultats du chapitre 3

Dans ce chapitre, nous considérons le problème de l'estimation de la matrice de variance asymptotique des estimateurs QML/LS d'un modèle VARMA sans faire l'hypothèse d'indépendance habituelle sur le bruit. L'objectif est de proposer une méthode d'estimation qui est différente de celles utilisées dans le chapitre précédent (*i.e.* différente des méthodes du HAC et de la densité spectrale) de la matrice de variance asymptotique $\Omega := J^{-1}IJ^{-1}$. La méthode que nous proposons consiste à trouver des estimateurs des matrices J et I dans lesquels, les estimateurs des coefficients du modèle VARMA sont isolés de ceux dépendants de la distribution de l'innovation. Dans un premier temps, nous proposons une expression des dérivées des résidus en fonction des

résidus passés et des paramètres du modèle VARMA. Ceci nous permettra ensuite de donner une expression explicite des matrices I et J impliquées dans la variance asymptotique des estimateurs QML/LS, en fonction des paramètres des polynômes VAR et MA, et des moments d'ordre deux et quatre du bruit. Enfin nous en déduisons un estimateur de la matrice Ω . Nous établissons la convergence de cet estimateur.

1.2.1 Notations et paramétrisation des coefficients du modèle VARMA

Nous utilisons la méthode du quasi-maximum de vraisemblance déjà définie dans le chapitre précédent pour l'estimation des paramètres du modèle ARMA(p_0, q_0) structurel d -multivarié (1.1) dont le terme d'erreur $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{dt})'$ est une suite de variables aléatoires centrées, non corrélées, avec une matrice de covariance non singulière Σ . Les paramètres sont les coefficients des matrices carrées d'ordre d suivantes : $A_i, i \in \{1, \dots, p_0\}, B_j, j \in \{1, \dots, q_0\}$. Dans ce chapitre nous considérons la matrice Σ comme un paramètre de nuisance. Nous supposons que ces matrices sont paramétrées suivant le vecteur des vraies valeurs des paramètres noté θ_0 . Nous notons $A_{0i} = A_i(\theta_0), i \in \{1, \dots, p_0\}, B_j = B_j(\theta_0), j \in \{1, \dots, q_0\}$ et $\Sigma_0 = \Sigma(\theta_0)$, où θ_0 appartient à l'espace compact des paramètres $\Theta \subset \mathbb{R}^{k_0}$, et k_0 est le nombre de paramètres inconnus qui est inférieur à $(p_0 + q_0 + 2)d^2$ (*i.e.* le nombre de paramètres sans aucune contrainte). Pour tout $\theta \in \Theta$, nous supposons aussi que les applications $\theta \mapsto A_i(\theta) i = 0, \dots, p_0, \theta \mapsto B_j(\theta) j = 0, \dots, q_0$ et $\theta \mapsto \Sigma(\theta)$ admettent des dérivées continues d'ordre 3. Nous définissons $\text{vec}(\cdot)$ l'opérateur qui consiste à empiler en un vecteur les colonnes d'une matrice en partant de la première colonne jusqu'à la dernière. Sous l'hypothèse \mathbf{H}_1 , les matrices A_{00} et B_{00} sont inversibles. Ceci nous permet d'écrire le modèle réduit sous sa forme compacte

$$\mathbf{A}_\theta(L)X_t = \mathbf{B}_\theta(L)e_t(\theta),$$

où $\mathbf{A}_\theta(L) = I_d - \sum_{i=1}^{p_0} \mathbf{A}_i L^i$ est le polynôme VAR et $\mathbf{B}_\theta(L) = I_d - \sum_{i=1}^{q_0} \mathbf{B}_i L^i$ est le polynôme MA, avec $\mathbf{A}_i = A_0^{-1} A_i$ et $\mathbf{B}_i = A_0^{-1} B_i B_0^{-1} A_0$. Pour $\ell = 1, \dots, p_0$ et $\ell' = 1, \dots, q_0$, posons $\mathbf{A}_\ell = (\mathbf{a}_{ij, \ell})$, $\mathbf{B}_{\ell'} = (\mathbf{b}_{ij, \ell'})$, $\mathbf{a}_\ell = \text{vec}[\mathbf{A}_\ell]$ et $\mathbf{b}_{\ell'} = \text{vec}[\mathbf{B}_{\ell'}]$. Notons respectivement par

$$\mathbf{a} := (\mathbf{a}'_1, \dots, \mathbf{a}'_{p_0})' \quad \text{et} \quad \mathbf{b} := (\mathbf{b}'_1, \dots, \mathbf{b}'_{q_0})',$$

les coefficients des polynômes VAR et MA. Ainsi il n'est pas restrictif de réécrire $\theta = (\mathbf{a}', \mathbf{b}')$, où $\mathbf{a} \in \mathbb{R}^{k_1}$ dépend des matrices A_0, \dots, A_{p_0} et où $\mathbf{b} \in \mathbb{R}^{k_2}$ dépend de B_0, \dots, B_{q_0} , avec $k_1 + k_2 = k_0$. Pour $i, j = 1, \dots, d$, définissons les $(d \times d)$ opérateurs matriciels $M_{ij}(L)$ et $N_{ij}(L)$ par

$$M_{ij}(L) = \mathbf{B}_\theta^{-1}(L) E_{ij} \mathbf{A}_\theta^{-1}(L) \mathbf{B}_\theta(L) \quad \text{and} \quad N_{ij}(L) = \mathbf{B}_\theta^{-1}(L) E_{ij},$$

où E_{ij} est une matrice carrée d'ordre d prenant la valeur 1 à la position (i, j) et 0 ailleurs. Notons $\mathbf{A}_{ij,h}^*$ et $\mathbf{B}_{ij,h}^*$ les matrices carrées d'ordre d définies par

$$M_{ij}(z) = \sum_{h=0}^{\infty} \mathbf{A}_{ij,h}^* z^h, \quad N_{ij}(z) = \sum_{h=0}^{\infty} \mathbf{B}_{ij,h}^* z^h, \quad |z| \leq 1$$

pour $h \geq 0$. Par convention $\mathbf{A}_{ij,h}^* = \mathbf{B}_{ij,h}^* = 0$ quand $h < 0$. Soit $\mathbf{A}_h^* = [\mathbf{A}_{11,h}^* : \mathbf{A}_{21,h}^* : \cdots : \mathbf{A}_{dd,h}^*]$ et $\mathbf{B}_h^* = [\mathbf{B}_{11,h}^* : \mathbf{B}_{21,h}^* : \cdots : \mathbf{B}_{dd,h}^*]$ les matrices de tailles $d \times d^3$. Posons

$$\lambda_h(\theta) = [-\mathbf{A}_{h-1}^* : \cdots : -\mathbf{A}_{h-p_0}^* : \mathbf{B}_{h-1}^* : \cdots : \mathbf{B}_{h-q_0}^*], \quad (1.6)$$

la matrice de taille $d \times d^3(p_0 + q_0)$ constituée uniquement des paramètres polynômes VAR et MA. Cette matrice est bien définie car les coefficients \mathbf{A}_θ^{-1} et \mathbf{B}_θ^{-1} décroissent exponentiellement vers zéro.

1.2.2 Expression des dérivées des résidus du modèle VARMA

Dans le cas de modèles ARMA univariés, McLeod (1978) avait défini des dérivées résiduelles par

$$\frac{\partial e_t}{\partial \phi_i} = v_{t-i}, \quad i = 1, \dots, p_0 \quad \text{et} \quad \frac{\partial e_t}{\partial \beta_j} = u_{t-j}, \quad j = 1, \dots, q_0,$$

où ϕ_i et β_j sont respectivement les paramètres AR et MA. Définissons les polynômes $\phi_\theta(L) = 1 - \sum_{i=1}^{p_0} \phi_i L^i$ et $\varphi_\theta(L) = 1 - \sum_{i=1}^{q_0} \varphi_i L^i$. Notons ϕ_h^* et φ_h^* les coefficients définis par

$$\phi_\theta^{-1}(z) = \sum_{h=0}^{\infty} \phi_h^* z^h, \quad \varphi_\theta^{-1}(z) = \sum_{h=0}^{\infty} \varphi_h^* z^h, \quad |z| \leq 1$$

pour $h \geq 0$. Posons $\theta = (\theta_1, \dots, \theta_{p_0}, \theta_{p_0+1}, \dots, \theta_{p_0+q_0})'$, pour p_0 et q_0 différents de 0. Alors, on peut facilement représenter les dérivées résiduelles univariées par

$$\frac{\partial e_t(\theta)}{\partial \theta} = (v_{t-1}(\theta), \dots, v_{t-p_0}(\theta), u_{t-1}(\theta), \dots, u_{t-q_0}(\theta))',$$

où

$$v_t(\theta) = -\phi_\theta^{-1}(L)e_t(\theta) = -\sum_{h=0}^{\infty} \phi_h^* e_{t-h}(\theta), \quad u_t(\theta) = \varphi_\theta^{-1}(L)e_t(\theta) = \sum_{h=0}^{\infty} \varphi_h^* e_{t-h}(\theta)$$

avec l'innovation

$$e_t(\theta) = \varphi_\theta^{-1}(L)\phi_\theta(L)X_t.$$

Nous énonçons la proposition suivante qui donne une extension au cas multivarié des expressions de ces dérivées résiduelles.

Proposition 1.1. Sous les hypothèses \mathbf{H}_1 – \mathbf{H}_6 , nous avons

$$\frac{\partial e_t(\theta)}{\partial \theta'} = [V_{t-1}(\theta) : \cdots : V_{t-p_0}(\theta) : U_{t-1}(\theta) : \cdots : U_{t-q_0}(\theta)],$$

où

$$V_t(\theta) = - \sum_{h=0}^{\infty} \mathbf{A}_h^* (I_{d^2} \otimes e_{t-h}(\theta)) \quad \text{et} \quad U_t(\theta) = \sum_{h=0}^{\infty} \mathbf{B}_h^* (I_{d^2} \otimes e_{t-h}(\theta))$$

avec $\mathbf{A}_h^* = [\mathbf{A}_{11,h}^* : \mathbf{A}_{21,h}^* : \cdots : \mathbf{A}_{dd,h}^*]$ et $\mathbf{B}_h^* = [\mathbf{B}_{11,h}^* : \mathbf{B}_{21,h}^* : \cdots : \mathbf{B}_{dd,h}^*]$ les matrices de taille $d \times d^3$ définies immédiatement avant (1.6). De plus, pour $\theta = \theta_0$ nous avons

$$\frac{\partial e_t}{\partial \theta'} = \sum_{i \geq 1} \lambda_i (I_{d^2(p_0+q_0)} \otimes e_{t-i}),$$

où les matrices λ_i sont définies par (1.6).

La démonstration de cette proposition repose sur deux lemmes qui donnent les expressions des dérivées résiduelles par rapport aux paramètres VAR et MA.

1.2.3 Expressions explicites des matrices J et I

Nous donnons des expressions explicites des matrices I et J impliquées dans la variance asymptotique Ω du QMLE. Dans ces expressions, nous isolons tout ce qui est fonction du paramètre θ_0 du modèle VARMA, de ce qui est fonction de la distribution du bruit blanc faible e_t . Nous notons par

$$\mathcal{M} := E \left\{ (I_{d^2(p_0+q_0)} \otimes e_t')^{\otimes 2} \right\}$$

la matrice impliquant les moments d'ordre deux de l'innovation (e_t). Maintenant nous donnons l'expression de $J = J(\theta_0, \Sigma_{e_0})$, dans laquelle les termes dépendants de θ_0 (via les matrices λ_i) sont distingués de ceux dépendants des moments d'ordre deux du processus (e_t) (via la matrice \mathcal{M}) et des termes dépendants de la variance de l'innovation (via la matrice Σ_{e_0}).

Proposition 1.2. Sous les hypothèses de la proposition 1.1, nous avons

$$\text{vec } J = 2 \sum_{i \geq 1} \mathcal{M} \{ \lambda_i' \otimes \lambda_i' \} \text{vec } \Sigma_{e_0}^{-1},$$

où les matrices λ_i sont définies par (1.6).

La démonstration de cette proposition repose sur le théorème ergodique 1.12 et la proposition 1.1.

Rappelons que

$$I = \text{Var}_{as} \frac{1}{\sqrt{n}} \sum_{t=1}^n \Upsilon_t = \sum_{h=-\infty}^{+\infty} \text{Cov}(\Upsilon_t, \Upsilon_{t-h}), \quad (1.7)$$

où

$$\Upsilon_t = \frac{\partial}{\partial \theta} \left\{ \log \det \Sigma_e + e'_t(\theta) \Sigma_e^{-1} e_t(\theta) \right\}_{\theta=\theta_0}. \quad (1.8)$$

Comme pour la matrice J , nous décomposons I en termes impliquant le paramètre θ_0 du modèle VARMA et ceux impliquant la distribution des innovations e_t . Soit les matrices

$$\mathcal{M}_{ij,h} := E \left(\left\{ e'_{t-h} \otimes (I_{d^2(p_0+q_0)} \otimes e'_{t-j-h}) \right\} \otimes \left\{ e'_t \otimes (I_{d^2(p_0+q_0)} \otimes e'_{t-i}) \right\} \right).$$

Les termes dépendants des paramètres du modèle VARMA sont les matrices λ_i définies par (1.6). Les matrices

$$\Gamma(i, j) := \sum_{h=-\infty}^{+\infty} \mathcal{M}_{ij,h}$$

mettent en jeu les moments d'ordre quatre du bruit blanc faible e_t . Les termes dépendants de la variance de l'innovation sont dans la matrice Σ_{e_0} . Nous énonçons maintenant la proposition suivante pour la matrice $I = I(\theta_0, \Sigma_{e_0})$ qui est similaire à la proposition 1.2.

Proposition 1.3. Sous les hypothèses de la proposition 1.2, nous avons

$$\text{vec } I = 4 \sum_{i,j=1}^{+\infty} \Gamma(i, j) \left(\{I_d \otimes \lambda'_j\} \otimes \{I_d \otimes \lambda'_i\} \right) \text{vec} \left(\text{vec } \Sigma_{e_0}^{-1} \left\{ \text{vec } \Sigma_{e_0}^{-1} \right\}' \right),$$

où les matrices λ_i sont définies par (1.6).

La démonstration de cette proposition repose sur la proposition 1.1, ainsi que sur quelques formules de calculs sur le produit de Kronecker et l'opérateur vec .

Remarque 1.7. En considérant le cas univarié, i.e. quand $d = 1$, nous avons

$$\text{vec } J = 2 \sum_{i \geq 1} \{\lambda_i \otimes \lambda_i\}' \quad \text{et} \quad \text{vec } I = \frac{4}{\sigma^4} \sum_{i,j=1}^{+\infty} \gamma(i, j) \{\lambda_j \otimes \lambda_i\}',$$

où $\gamma(i, j) = \sum_{h=-\infty}^{+\infty} E(e_t e_{t-i} e_{t-h} e_{t-j-h})$ et les vecteurs $\lambda'_i \in \mathbb{R}^{p_0+q_0}$ sont définis par (1.6).

Remarque 1.8. *Toujours pour le cas univarié ($d = 1$), Francq, Roy and Zakoïan (2005) ont utilisé l'estimateur des moindres carrés et ont obtenu*

$$E \frac{\partial^2}{\partial \theta \partial \theta'} e_t^2(\theta_0) = 2 \sum_{i \geq 1} \sigma^2 \lambda_i \lambda_i' \quad \text{et} \quad \text{Var} \left\{ 2e_t(\theta_0) \frac{\partial e_t(\theta_0)}{\partial \theta} \right\} = 4 \sum_{i, j \geq 1} \gamma(i, j) \lambda_i \lambda_j'$$

où σ^2 est la variance du processus univarié e_t et les vecteurs $\lambda_i = (-\phi_{i-1}^*, \dots, -\phi_{i-p_0}^*, \varphi_{i-1}^*, \dots, \varphi_{i-q_0}^*)' \in \mathbb{R}^{p_0+q_0}$, avec la convention $\phi_i^* = \varphi_i^* = 0$ quand $i < 0$. En utilisant l'opérateur vec et la relation élémentaire $\text{vec}(aa') = a \otimes a'$, leur résultat donne

$$\text{vec } J = \text{vec } E \frac{\partial^2}{\partial \theta \partial \theta'} \ell_n(\theta_0) = \frac{1}{\sigma^2} \text{vec } E \frac{\partial^2}{\partial \theta \partial \theta'} e_t^2(\theta_0) = 2 \sum_{i \geq 1} \lambda_i \otimes \lambda_i \quad \text{et}$$

$$\text{vec } I = \text{vec } \text{Var} \left\{ \frac{\partial \ell_n(\theta_0)}{\partial \theta} \right\} = \frac{1}{\sigma^4} \text{vec } \text{Var} \left\{ 2e_t \frac{\partial e_t(\theta_0)}{\partial \theta} \right\} = \frac{4}{\sigma^4} \sum_{i, j \geq 1} \gamma(i, j) \lambda_i \otimes \lambda_j,$$

lesquelles sont les expressions données dans la remarque 1.7.

1.2.4 Estimation de la matrice de variance asymptotique $\Omega := J^{-1} I J^{-1}$

Étant donné que nous disposons des expressions explicites de I et J , nous nous intéressons à l'estimation de ces matrices. Soit $\hat{e}_t = \tilde{e}_t(\hat{\theta}_n)$ les résidus du QMLE quand $p_0 > 0$ et $q_0 > 0$, et soit $\hat{e}_t = e_t = X_t$ quand $p_0 = q_0 = 0$. Quand $p_0 + q_0 \neq 0$, nous avons $\hat{e}_t = 0$ pour $t \leq 0$ et $t > n$, et posons

$$\hat{e}_t = X_t - \sum_{i=1}^{p_0} A_0^{-1}(\hat{\theta}_n) A_i(\hat{\theta}_n) \hat{X}_{t-i} + \sum_{i=1}^{q_0} A_0^{-1}(\hat{\theta}_n) B_i(\hat{\theta}_n) B_0^{-1}(\hat{\theta}_n) A_0(\hat{\theta}_n) \hat{e}_{t-i},$$

pour $t = 1, \dots, n$, avec $\hat{X}_t = 0$ pour $t \leq 0$ et $\hat{X}_t = X_t$ pour $t \geq 1$. Soit $\hat{\Sigma}_{e_0} = n^{-1} \sum_{t=1}^n \hat{e}_t \hat{e}_t'$ un estimateur de Σ_{e_0} . La matrice \mathcal{M} impliquée dans l'expression de J peut facilement être estimée empiriquement par

$$\hat{\mathcal{M}}_n := \frac{1}{n} \sum_{t=1}^n \left\{ (I_{d^2(p+q)} \otimes \hat{e}_t')^{\otimes 2} \right\}.$$

En vue de la proposition 1.2, nous définissons un estimateur \hat{J}_n de J par

$$\text{vec } \hat{J}_n = \sum_{i \geq 1} \hat{\mathcal{M}}_n \left\{ \hat{\lambda}_i' \otimes \hat{\lambda}_i \right\} \text{vec } \hat{\Sigma}_{e_0}^{-1}.$$

Nous énonçons maintenant le théorème suivant qui montre la consistance forte de \hat{J}_n .

Théorème 1.5. (Convergence forte de \hat{J}_n) Sous les hypothèses de la proposition 1.3, nous avons

$$\hat{J}_n \rightarrow J \quad p.s \quad \text{quand} \quad n \rightarrow \infty.$$

La démonstration de ce théorème repose sur une série de trois lemmes, dans lesquels nous montrons les convergences presque sûres de $\hat{\lambda}_i$, de $\hat{\mathcal{M}}_n$ et de $\hat{\Sigma}_{e_0}^{-1}$.

Dans le cas de modèles VARMA forts standard $\hat{\Omega} = 2\hat{J}^{-1}$ est un estimateur fortement consistant de Ω . Dans le cadre général de modèles VARMA faibles, cet estimateur n'est pas consistant dès que $I \neq 2J$. Dans ce cas, nous avons besoin d'un estimateur de I . Cette matrice est délicate à estimer car son expression explicite fait intervenir une infinité de moments d'ordre quatre

$$\Gamma(i, j) = \sum_{h=-\infty}^{+\infty} \mathcal{M}_{ij,h}.$$

Afin d'estimer cette matrice, nous utilisons une technique qui consiste à pondérer convenablement certains moments empiriques (voir Andrews 1991, voir aussi Newey et West 1987). Cette pondération se fait au moyen d'une fonction de poids (ou fenêtre) et d'un paramètre de troncature. Soit

$$\mathcal{M}_{n\,ij,h} := \frac{1}{n} \sum_{t=1}^{n-|h|} (\{e'_{t-h} \otimes (I_{d^2(p_0+q_0)} \otimes e'_{t-j-h})\} \otimes \{e'_t \otimes (I_{d^2(p_0+q_0)} \otimes e'_{t-i})\}).$$

Pour estimer $\Gamma(i, j)$, nous considérons une suite de réels $(b_n)_{n \in \mathbb{N}^*}$ telle que

$$b_n \rightarrow 0 \quad \text{et} \quad nb_n^{\frac{10+4\nu}{\nu}} \rightarrow \infty \quad \text{quand} \quad n \rightarrow \infty, \quad (1.9)$$

et une fonction de poids $f : \mathbb{R} \rightarrow \mathbb{R}$ bornée, à support compact $[-a, a]$ et continue à l'origine avec $f(0) = 1$. Notons que sous les hypothèses précédentes, nous avons

$$b_n \sum_{|h| < n} |f(hb_n)| = O(1). \quad (1.10)$$

Soit

$$\hat{\mathcal{M}}_{n\,ij,h} := \frac{1}{n} \sum_{t=1}^{n-|h|} (\{\hat{e}'_{t-h} \otimes (I_{d^2(p_0+q_0)} \otimes \hat{e}'_{t-j-h})\} \otimes \{\hat{e}'_t \otimes (I_{d^2(p_0+q_0)} \otimes \hat{e}'_{t-i})\}).$$

Nous considérons la matrice

$$\hat{\Gamma}_n(i, j) := \sum_{h=-T_n}^{+T_n} f(hb_n) \hat{\mathcal{M}}_{n\,ij,h} \quad \text{et} \quad T_n = \left\lfloor \frac{a}{b_n} \right\rfloor,$$

où $[x]$ désigne la partie entière de x .

En vue de la proposition 1.3, nous définissons un estimateur \hat{I}_n de I par

$$\text{vec } \hat{I}_n = 4 \sum_{i,j=1}^{+\infty} \hat{\Gamma}_n(i,j) \left(\left\{ I_d \otimes \hat{\lambda}'_i \right\} \otimes \left\{ I_d \otimes \hat{\lambda}'_j \right\} \right) \text{vec} \left(\text{vec } \hat{\Sigma}_{\epsilon_0}^{-1} \left\{ \text{vec } \hat{\Sigma}_{\epsilon_0}^{-1} \right\}' \right).$$

Maintenant nous énonçons le théorème suivant qui montre la consistance faible de l'estimateur empirique \hat{I}_n .

Théorème 1.6. (Consistance faible de \hat{I}_n) Sous les hypothèses du théorème 1.5, nous avons

$$\hat{I}_n \rightarrow I \text{ en probabilité, quand } n \rightarrow \infty.$$

La démonstration de ce théorème repose sur une série de trois lemmes, dans lesquels nous montrons les convergences de $\hat{\lambda}_i$ et de $\hat{\Sigma}_{\epsilon_0}^{-1}$. Dans le dernier lemme, nous montrons la convergence de $\hat{\Gamma}_n(i,j)$ uniformément en i et j .

Les théorèmes 1.5 et 1.6 montrent alors que

$$\hat{\Omega}_n := \hat{J}_n^{-1} \hat{I}_n \hat{J}_n^{-1}$$

est un estimateur faiblement consistant de la matrice de covariance asymptotique $\Omega := J^{-1} I J^{-1}$.

1.3 Résultats du chapitre 4

Dans la modélisation des séries temporelles, la validité des différentes étapes de la méthodologie traditionnelle de Box et Jenkins, à savoir les étapes d'identification, d'estimation et de validation, dépend des propriétés du bruit blanc. Après l'identification et l'estimation du processus vectoriel autorégressif moyenne mobile, la prochaine étape importante dans la modélisation de modèles VARMA consiste à vérifier si le modèle estimé est compatible avec les données. Cette étape d'adéquation permet de valider ou d'invalider le choix des ordres du modèle. Ce choix est important pour la précision des prévisions linéaires et pour une bonne interprétation du modèle.

Pour les modèles VARMA(p_0, q_0), le choix de p_0 et q_0 est particulièrement important parce que le nombre $(p_0 + q_0 + 2)d^2$ de paramètres augmente rapidement avec p_0 et q_0 .

La sélection d'ordres trop grands a pour effet d'introduire des termes qui ne sont pas forcément pertinents dans le modèle et aussi d'engendrer des difficultés statistiques comme par exemple un trop grand nombre de paramètres à estimer, ce qui est susceptible d'engendrer une perte de précision de l'estimation des paramètres. Le praticien peut aussi choisir des ordres trop petits qui entraînent la perte d'une information qui peut être détectée par une corrélation des résidus ou encore une estimation non convergente des paramètres.

L'objectif de ce chapitre est d'étudier le comportement asymptotique du test portmanteau dans le cadre de modèles VARMA dont les termes d'erreur sont non corrélés mais peuvent contenir des dépendances non linéaires. Nous avons choisi d'étudier ce test car c'est l'outil le plus utilisé pour tester les ordres p_0 et q_0 . Ainsi pour des ordres p_0 et q_0 donnés, nous testons l'hypothèse nulle

$$H_0 : (X_t) \text{ satisfait une représentation } VARMA(p_0, q_0)$$

contre l'alternative

$$H_1 : (X_t) \text{ n'admet pas une représentation } VARMA, \\ \text{ou admet une représentation } VARMA(p, q) \text{ avec } p > p_0 \text{ ou } q > q_0.$$

Dans un premier temps, nous étudions la distribution asymptotique jointe du QMLE/LSE et des autocovariances empiriques du bruit. Ceci nous permet ensuite d'obtenir les distributions asymptotiques des autocovariances et autocorrélations résiduelles. Ces autocorrélations résiduelles sont normalement distribuées avec une matrice de covariance différente du cas iid. Enfin, nous déduisons le comportement asymptotique des statistiques portmanteau. Dans le cadre standard (c'est-à-dire sous les hypothèses iid sur le bruit), il est connu que la distribution asymptotique des tests portmanteau est approximée par un chi-deux. Dans le cas général, nous montrons que cette distribution asymptotique est celle d'une somme pondérée de chi-deux. Cette distribution peut être très différente de l'approximation chi-deux usuelle du cas fort. Nous en déduisons des tests portmanteau modifiés pour tester l'adéquation de modèles VARMA faibles.

1.3.1 Modèle et paramétrisation des coefficients

Pour une lecture indépendante de cette section, rappelons brièvement le cadre des chapitres précédents. Comme estimateur des paramètres du modèle ARMA(p_0, q_0) structurel d -multivarié (1.1), nous utilisons l'estimateur des moindres carrés défini dans le théorème 1.3. Le terme d'erreur $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{dt})'$ est une suite de variables aléatoires centrées, non corrélées, avec une matrice de covariance non singulière Σ . Les paramètres

sont les coefficients des matrices carrées d'ordre d suivantes : A_i , $i \in \{1, \dots, p_0\}$, B_j , $j \in \{1, \dots, q_0\}$. Dans ce chapitre aussi, nous considérons la matrice Σ comme un paramètre de nuisance. Nous supposons que ces matrices sont paramétrées suivant le vecteur des vraies valeurs des paramètres noté θ_0 . Nous notons $A_{0i} = A_i(\theta_0)$, $i \in \{1, \dots, p_0\}$, $B_j = B_j(\theta_0)$, $j \in \{1, \dots, q_0\}$ et $\Sigma_0 = \Sigma(\theta_0)$, où θ_0 appartient à l'espace compact des paramètres $\Theta \subset \mathbb{R}^{k_0}$, et k_0 est le nombre de paramètres inconnus qui est inférieur à $(p_0 + q_0 + 2)d^2$ (*i.e.* le nombre de paramètres sans aucune contrainte).

1.3.2 Distribution asymptotique jointe de $\hat{\theta}_n$ et des autocovariances empiriques du bruit

Définissons les autocovariances empiriques du bruit

$$\gamma(h) = \frac{1}{n} \sum_{t=h+1}^n e_t e_{t-h}' \quad \text{pour } 0 \leq h < n.$$

Notons que les $\gamma(h)$ ne sont pas des statistiques (sauf si $p_0 = q_0 = 0$) puisqu'elles dépendent des innovations théoriques $e_t = e_t(\theta_0)$. Nous considérons des vecteurs des m (pour $m \geq 1$) premières autocovariances empiriques du bruit

$$\gamma_m = (\{\text{vec}\gamma(1)\}', \dots, \{\text{vec}\gamma(m)\}')'$$

et soit

$$\Gamma(\ell, \ell') = \sum_{h=-\infty}^{\infty} E(\{e_{t-\ell} \otimes e_t\} \{e_{t-h-\ell'} \otimes e_{t-h}\}'),$$

pour $(\ell, \ell') \neq (0, 0)$. Afin d'assurer l'existence de ces matrices dans le cas de modèles ARMA univariés, Francq, Roy et Zakoïan (2004) ont montré dans leur lemme A.1 que $|\Gamma(\ell, \ell')| \leq K \max(\ell, \ell')$ pour une constante K . Ce résultat peut être directement étendu au cas de processus ARMA multivariés. Ainsi nous obtenons $\|\Gamma(\ell, \ell')\| \leq K \max(\ell, \ell')$ pour une constante K . La démonstration est similaire au cas univarié. Nous allons maintenant énoncer le théorème suivant qui est une extension directe du résultat donné dans Francq, Roy et Zakoïan (2004).

Théorème 1.7. Supposons $p_0 > 0$ ou $q_0 > 0$. Sous les hypothèses \mathbf{H}_1 – \mathbf{H}_6 , quand $n \rightarrow \infty$, nous avons

$$\sqrt{n}(\gamma_m, \hat{\theta}_n - \theta_0)' \xrightarrow{d} \mathcal{N}(0, \Xi) \quad \text{où} \quad \Xi = \begin{pmatrix} \Sigma_{\gamma_m} & \Sigma_{\gamma_m, \hat{\theta}_n} \\ \Sigma'_{\gamma_m, \hat{\theta}_n} & \Sigma_{\hat{\theta}_n} \end{pmatrix},$$

avec $\Sigma_{\gamma_m} = \{\Gamma(\ell, \ell')\}_{1 \leq \ell, \ell' \leq m}$, $\Sigma'_{\gamma_m, \hat{\theta}_n} = \text{Cov}(\sqrt{n}J^{-1}Y_n, \sqrt{n}\gamma_m)$ et $\Sigma_{\hat{\theta}_n} = \text{Var}_{as}(\sqrt{n}J^{-1}Y_n) = J^{-1}IJ^{-1}$.

La démonstration de ce théorème résulte des formules de dérivations matricielles standard, de l'inégalité de covariance obtenue par Davydov (1968), du théorème de la convergence de $\hat{\theta}_n$ et du théorème 1.13 (TCL).

1.3.3 Comportement asymptotique des autocovariances et autocorrélations résiduelles

Soit $\hat{e}_t = \tilde{e}_t(\hat{\theta}_n)$ les résidus du LSE quand $p_0 > 0$ et $q_0 > 0$, et soit $\hat{e}_t = e_t = X_t$ quand $p_0 = q_0 = 0$. Quand $p_0 + q_0 \neq 0$, nous avons $\hat{e}_t = 0$ pour $t \leq 0$ et $t > n$, et posons

$$\hat{e}_t = X_t - \sum_{i=1}^{p_0} A_0^{-1}(\hat{\theta}_n) A_i(\hat{\theta}_n) \hat{X}_{t-i} + \sum_{i=1}^{q_0} A_0^{-1}(\hat{\theta}_n) B_i(\hat{\theta}_n) B_0^{-1}(\hat{\theta}_n) A_0(\hat{\theta}_n) \hat{e}_{t-i},$$

pour $t = 1, \dots, n$, avec $\hat{X}_t = 0$ pour $t \leq 0$ et $\hat{X}_t = X_t$ pour $t \geq 1$. Définissons les autocovariances résiduelles

$$\hat{\Gamma}_e(h) = \frac{1}{n} \sum_{t=h+1}^n \hat{e}_t \hat{e}'_{t-h} \quad \text{pour } 0 \leq h < n.$$

Soit $\hat{\Sigma}_{e0} = \hat{\Gamma}_e(0) = n^{-1} \sum_{t=1}^n \hat{e}_t \hat{e}'_t$. Nous considérons des vecteurs des m (pour $m \geq 1$) premières autocovariances résiduelles

$$\hat{\Gamma}_m = \left(\left\{ \text{vec} \hat{\Gamma}_e(1) \right\}', \dots, \left\{ \text{vec} \hat{\Gamma}_e(m) \right\}' \right)'.$$

Soit les matrices diagonales

$$S_e = \text{Diag}(\sigma_e(1), \dots, \sigma_e(d)) \quad \text{et} \quad \hat{S}_e = \text{Diag}(\hat{\sigma}_e(1), \dots, \hat{\sigma}_e(d)),$$

où $\sigma_e^2(i)$ est la variance de la i -ème coordonnée de l'innovation e_t et $\hat{\sigma}_e^2(i)$ est son estimateur, avec $\sigma_e(i) = \sqrt{E e_{it}^2}$ et $\hat{\sigma}_e(i) = \sqrt{n^{-1} \sum_{t=1}^n \hat{e}_{it}^2}$. Nous définissons les autocorrélations théoriques et résiduelles de retard ℓ

$$R_e(\ell) = S_e^{-1} \Gamma_e(\ell) S_e^{-1} \quad \text{et} \quad \hat{R}_e(\ell) = \hat{S}_e^{-1} \hat{\Gamma}_e(\ell) \hat{S}_e^{-1},$$

avec $\Gamma_e(\ell) := E e_t e'_{t-\ell} = 0, \quad \forall \ell \neq 0$. Nous définissons aussi la matrice

$$\Phi_m = E \left\{ \left(\begin{array}{c} e_{t-1} \\ \vdots \\ e_{t-m} \end{array} \right) \otimes \frac{\partial e_t(\theta_0)}{\partial \theta'} \right\}. \quad (1.11)$$

Nous considérons des vecteurs des m (pour $m \geq 1$) premières autocorrélations résiduelles

$$\hat{\rho}_m = \left(\left\{ \text{vec} \hat{R}_e(1) \right\}', \dots, \left\{ \text{vec} \hat{R}_e(m) \right\}' \right)'.$$

Nous énonçons le résultat suivant qui donne le comportement asymptotique des autocovariances et autocorrélations résiduelles.

Théorème 1.8. Sous les hypothèses du théorème 1.7, nous avons

$$\sqrt{n}\hat{\Gamma}_m \Rightarrow \mathcal{N}(0, \Sigma_{\hat{\Gamma}_m}) \quad \text{et} \quad \sqrt{n}\hat{\rho}_m \Rightarrow \mathcal{N}(0, \Sigma_{\hat{\rho}_m}) \quad \text{où,}$$

$$\Sigma_{\hat{\Gamma}_m} = \Sigma_{\gamma_m} + \Phi_m \Sigma_{\hat{\theta}_n} \Phi_m' + \Phi_m \Sigma_{\hat{\theta}_n, \gamma_m} + \Sigma_{\hat{\theta}_n, \gamma_m}' \Phi_m' \quad (1.12)$$

$$\Sigma_{\hat{\rho}_m} = \{I_m \otimes (S_e \otimes S_e)^{-1}\} \Sigma_{\hat{\Gamma}_m} \{I_m \otimes (S_e \otimes S_e)^{-1}\} \quad (1.13)$$

et Φ_m est donné par (1.11).

Dans la démonstration de ce théorème, nous utilisons les autocovariances empiriques du bruit précédemment définies et le théorème 1.7. Le résultat (1.12) découle de la relation

$$\hat{\Gamma}_m := \left(\left\{ \text{vec} \hat{\Gamma}_e(1) \right\}', \dots, \left\{ \text{vec} \hat{\Gamma}_e(m) \right\}' \right)' = \gamma_m + \Phi_m (\hat{\theta}_n - \theta_0) + O_P(1/n),$$

que nous montrons en utilisant un développement limité de Taylor. Finalement nous obtenons le résultat (1.13) en calculant la variance asymptotique $\text{Var}(\sqrt{n}\hat{\rho}_m)$ de

$$\hat{\rho}_m = \{I_m \otimes (S_e \otimes S_e)^{-1}\} \hat{\Gamma}_m + O_P(n^{-1}).$$

1.3.4 Comportement asymptotique des statistiques portman-teau

Le test portmanteau a été introduit par Box et Pierce (1970) (noté **BP** dans la suite) afin de mesurer la qualité d'ajustement d'un modèle ARMA fort univarié. Ce test a été étendu au cas de modèles ARMA multivariés par Chitturi (1974). Comme dans le cas univarié, ce test est basé sur les résidus $\hat{\epsilon}_t$ résultants de l'estimation des paramètres du modèle (1.1). Hosking (1981a) a proposé plusieurs formes équivalentes à la statistique

de **BP** dont les plus basiques sont les suivantes

$$\begin{aligned}
P_m &= n \sum_{h=1}^m \text{Tr} \left(\hat{\Gamma}'_e(h) \hat{\Gamma}_e^{-1}(0) \hat{\Gamma}_e(h) \hat{\Gamma}_e^{-1}(0) \right) \\
&= n \sum_{h=1}^m \text{vec} \left(\hat{\Gamma}_e(h) \right)' \left(\hat{\Gamma}_e^{-1}(0) \otimes I_d \right) \text{vec} \left(\hat{\Gamma}_e^{-1}(0) \hat{\Gamma}_e(h) \right) \\
&= n \sum_{h=1}^m \text{vec} \left(\hat{\Gamma}_e(h) \right)' \left(\hat{\Gamma}_e^{-1}(0) \otimes I_d \right) \left(I_d \otimes \hat{\Gamma}_e^{-1}(0) \right) \text{vec} \left(\hat{\Gamma}_e(h) \right) \\
&= n \sum_{h=1}^m \text{vec} \left(\hat{\Gamma}_e(h) \right)' \left(\hat{\Gamma}_e^{-1}(0) \otimes \hat{\Gamma}_e^{-1}(0) \right) \text{vec} \left(\hat{\Gamma}_e(h) \right) \\
&= n \hat{\Gamma}'_m \left(I_m \otimes \left\{ \hat{\Gamma}_e^{-1}(0) \otimes \hat{\Gamma}_e^{-1}(0) \right\} \right) \hat{\Gamma}_m \\
&= n \hat{\rho}'_m \left(I_m \otimes \left\{ \hat{\Gamma}_e(0) \hat{\Gamma}_e^{-1}(0) \hat{\Gamma}_e(0) \right\} \otimes \left\{ \hat{\Gamma}_e(0) \hat{\Gamma}_e^{-1}(0) \hat{\Gamma}_e(0) \right\} \right) \hat{\rho}_m \\
&= n \hat{\rho}'_m \left(I_m \otimes \left\{ \hat{R}_e^{-1}(0) \otimes \hat{R}_e^{-1}(0) \right\} \right) \hat{\rho}_m.
\end{aligned}$$

Ces égalités sont obtenues à partir des relations élémentaires $\text{vec}(AB) = (I \otimes A) \text{vec} B$, $(A \otimes B)(C \otimes D) = AC \otimes BD$ et $\text{Tr}(ABC) = \text{vec}(A)'(C' \otimes I) \text{vec} B$.

Ljung et Box (1978) (**LB** dans la suite) ont proposé une version modifiée du test portmanteau univarié de **BP** qui est à nos jours l'un des outils de diagnostic le plus populaire dans la modélisation ARMA de séries chronologiques. D'une façon similaire à la statistique du test **LB** en univarié, Hosking (1980) a défini une version multivariée de cette statistique du test portmanteau

$$\tilde{P}_m = n^2 \sum_{h=1}^m (n-h)^{-1} \text{Tr} \left(\hat{\Gamma}'_e(h) \hat{\Gamma}_e^{-1}(0) \hat{\Gamma}_e(h) \hat{\Gamma}_e^{-1}(0) \right),$$

qui a de meilleurs résultats pour des échantillons de petites tailles que la statistique de **BP** quand les erreurs sont gaussiennes. Sous les hypothèses que le **DGP** (pour data generating process) suit un modèle VARMA(p_0, q_0) fort et que les racines des polynômes VAR et MA ne sont pas trop proches de la racine unité, la distribution asymptotique des statistiques P_m et \tilde{P}_m est généralement bien approximée par la distribution $\chi^2_{d^2m-k_0}$ ($d^2m > k_0$). Lorsque les innovations sont gaussiennes, Hosking (1980) a constaté que la distribution des échantillons finis de \tilde{P}_m est plus proche d'une distribution $\chi^2_{d^2(m-(p_0+q_0))}$ que celle de P_m . En vue du théorème 1.8 nous déduisons le résultat suivant, qui donne la loi asymptotique exacte de la statistique standard P_m . Nous verrons que cette distribution peut être très différente de celle d'une distribution $\chi^2_{d^2m-k_0}$ dans le cas de modèles VARMA(p_0, q_0).

Théorème 1.9. Sous les hypothèses du théorème 1.8, les statistiques P_m et \tilde{P}_m

convergent en loi, quand $n \rightarrow \infty$, vers

$$Z_m(\xi_m) = \sum_{i=1}^{d^2 m} \xi_{i,d^2 m} Z_i^2$$

où $\xi_m = (\xi_{1,d^2 m}, \dots, \xi_{d^2 m,d^2 m})'$ est le vecteur des valeurs propres de la matrice

$$\Omega_m = (I_m \otimes \Sigma_e^{-1/2} \otimes \Sigma_e^{-1/2}) \Sigma_{\hat{\Gamma}_m} (I_m \otimes \Sigma_e^{-1/2} \otimes \Sigma_e^{-1/2}),$$

et les Z_1, \dots, Z_m sont des variables indépendantes, centrées et réduites normalement distribuées.

Ainsi quand le processus d'erreurs est un bruit blanc faible, la distribution asymptotique des statistiques P_m et \tilde{P}_m est une somme pondérée de chi-deux.

En vue du théorème 1.9, la distribution asymptotique des statistiques du test port-manteau de **BP** et de **LB** dépend du paramètre de nuisance Σ_e , de la matrice Φ_m et des éléments de la matrice Ξ . Nous avons donc besoin d'estimateurs consistants de ces matrices inconnues. La matrice Σ_e peut être estimée à partir des résidus d'estimation par $\hat{\Sigma}_e = \hat{\Gamma}_e(0)$. La matrice Φ_m peut être estimée empiriquement par

$$\hat{\Phi}_m = \frac{1}{n} \sum_{t=1}^n \left\{ (\hat{e}'_{t-1}, \dots, \hat{e}'_{t-m})' \otimes \frac{\partial e_t(\theta_0)}{\partial \theta'} \right\}_{\theta_0 = \hat{\theta}_n}.$$

Afin d'estimer la matrice Ξ , nous utilisons la méthode d'estimation de la densité spectrale (déjà définie les chapitres précédents) du processus stationnaire $\Upsilon_t = (\Upsilon'_{1t}, \Upsilon'_{2t})'$, où $\Upsilon_{1t} = (e'_{t-1}, \dots, e'_{t-m})' \otimes e_t$ et $\Upsilon_{2t} = -2J^{-1}(\partial e'_t(\theta_0)/\partial \theta) \Sigma_{e0}^{-1} e_t(\theta_0)$. En interprétant $(2\pi)^{-1}\Xi$ comme étant la densité spectrale évaluée en zéro du processus stationnaire (Υ_t) , nous avons

$$\Xi = \Phi^{-1}(1) \Sigma_u \Phi^{-1}(1)$$

quand (Υ_t) satisfait une représentation VAR(∞) de la forme

$$\Phi(L)\Upsilon_t := \Upsilon_t + \sum_{i=1}^{\infty} \Phi_i \Upsilon_{t-i} = u_t, \quad (1.14)$$

où u_t est un bruit blanc faible de matrice de variance Σ_u . Puisque Υ_t est inconnu, nous posons $\hat{\Upsilon}_t$ le vecteur obtenu en remplaçant θ_0 par $\hat{\theta}_n$ dans l'expression de Υ_t . Nous définissons le polynôme $\hat{\Phi}_r(z) = I_{k_0+d^2 m} + \sum_{i=1}^r \hat{\Phi}_{r,i} z^i$, où $\hat{\Phi}_{r,1}, \dots, \hat{\Phi}_{r,r}$ sont les coefficients de la régression des moindres carrés de $\hat{\Upsilon}_t$ sur $\hat{\Upsilon}_{t-1}, \dots, \hat{\Upsilon}_{t-r}$. Notons $\hat{u}_{r,t}$ les résidus de cette régression et $\hat{\Sigma}_{\hat{u}_r}$ la variance empirique des résidus $\hat{u}_{r,1}, \dots, \hat{u}_{r,n}$. Nous établissons le théorème suivant qui est une extension du résultat de Francq, Roy et Zakoïan (2004).

Théorème 1.10. Sous les hypothèses du théorème 1.9, nous supposons que le processus (Y_t) admet une représentation VAR(∞) (1.14) dont les racines de $\det \Phi(z) = 0$ sont à l'extérieur du disque unité, $\|\Phi_i\| = o(i^{-2})$, et que la matrice $\Sigma_u = \text{Var}(u_t)$ est non singulière. De plus, nous supposons que $\|\epsilon_t\|_{8+4\nu} < \infty$ et $\sum_{k=0}^{\infty} \{\alpha_{X,\epsilon}(k)\}^{\nu/(2+\nu)} < \infty$ pour un réel $\nu > 0$, où $\{\alpha_{X,\epsilon}(k)\}_{k \geq 0}$ est une suite de coefficients de mélange fort du processus $(X'_t, \epsilon'_t)'$. Alors l'estimateur spectral de Ξ

$$\hat{\Xi}^{\text{SP}} := \hat{\Phi}_r^{-1}(1) \hat{\Sigma}_{\hat{u}_r} \hat{\Phi}'_r^{-1}(1) \rightarrow \Xi$$

en probabilité quand $r = r(n) \rightarrow \infty$ et $r^3/n \rightarrow 0$ quand $n \rightarrow \infty$.

Soit $\hat{\Omega}_m$ la matrice obtenu en remplaçant Ξ par son estimateur $\hat{\Xi}$ et Σ_e par $\hat{\Sigma}_e$ dans la matrice Ω_m . Nous définissons $\hat{\xi}_m = (\hat{\xi}_{1,d^2m}, \dots, \hat{\xi}_{d^2m,d^2m})'$ le vecteur des valeurs propres de $\hat{\Omega}_m$.

Pour un niveau asymptotique de risque α , le test de **LB** (resp. celui de **BP**) que nous proposons consiste à rejeter l'hypothèse H_0 (*i.e.* l'adéquation de modèles VARMA(p_0, q_0) faibles) quand

$$\tilde{P}_m > S_m(1 - \alpha) \quad (\text{resp.} \quad P_m > S_m(1 - \alpha))$$

où $S_m(1 - \alpha)$ est tel que $P \left\{ Z_m(\hat{\xi}_m) > S_m(1 - \alpha) \right\} = \alpha$.

1.3.5 Mise en oeuvre des tests portmanteau

Les observations dont nous disposons sont des vecteurs X_1, \dots, X_n de dimension d . Pour tester l'adéquation de modèles VARMA(p_0, q_0) faibles à ces données, nous proposons la méthode suivante pour implémenter les versions modifiées des tests portmanteau (**LB** et **BP**).

1. Calculer les estimateurs $\hat{A}_1, \dots, \hat{A}_{p_0}, \hat{B}_1, \dots, \hat{B}_{q_0}$ par QML.
2. Calculer les résidus de ce QMLE $\hat{e}_t = \tilde{e}_t(\hat{\theta}_n)$ quand $p_0 > 0$ où $q_0 > 0$, et soit $\hat{e}_t = e_t = X_t$ quand $p_0 = q_0 = 0$. Quand $p_0 + q_0 \neq 0$, nous avons $\hat{e}_t = 0$ pour $t \leq 0$ et $t > n$, et posons

$$\hat{e}_t = X_t - \sum_{i=1}^{p_0} A_0^{-1}(\hat{\theta}_n) A_i(\hat{\theta}_n) \hat{X}_{t-i} + \sum_{i=1}^{q_0} A_0^{-1}(\hat{\theta}_n) B_i(\hat{\theta}_n) B_0^{-1}(\hat{\theta}_n) A_0(\hat{\theta}_n) \hat{e}_{t-i},$$

pour $t = 1, \dots, n$, avec $\hat{X}_t = 0$ pour $t \leq 0$ et $\hat{X}_t = X_t$ pour $t \geq 1$.

3. Calculer les autocovariances résiduelles $\hat{\Gamma}_e(0) = \hat{\Sigma}_{e0}$ et $\hat{\Gamma}_e(h)$ pour $h = 1, \dots, m$ et $\hat{\Gamma}_m = \left(\left\{ \hat{\Gamma}_e(1) \right\}', \dots, \left\{ \hat{\Gamma}_e(m) \right\}' \right)'$.
4. Calculer la matrice $\hat{J} = 2n^{-1} \sum_{t=1}^n (\partial \hat{e}'_t / \partial \theta) \hat{\Sigma}_{e0}^{-1} (\partial \hat{e}_t / \partial \theta')$.
5. Calculer le processus $\hat{Y}_t = \left(\hat{Y}'_{1t}, \hat{Y}'_{2t} \right)'$, où $\hat{Y}_{1t} = (\hat{e}'_{t-1}, \dots, \hat{e}'_{t-m})' \otimes \hat{e}_t$ et $\hat{Y}_{2t} = -2\hat{J}^{-1} (\partial \hat{e}'_t / \partial \theta) \hat{\Sigma}_{e0}^{-1} \hat{e}_t$.
6. Ajuster le modèle $VAR(r)$

$$\hat{\Phi}_r(L) \hat{Y}_t := \left(I_{d^2m+k_0} + \sum_{i=1}^r \hat{\Phi}_{r,i}(L) \right) \hat{Y}_t = \hat{u}_{r,t}.$$

Notons que l'ordre r du modèle VAR peut être fixé ou sélectionné par un critère d'information comme celui de Akaike, le AIC.

7. Définir l'estimateur

$$\hat{\Xi}^{\text{SP}} := \hat{\Phi}_r^{-1}(1) \hat{\Sigma}_{\hat{u}_r} \hat{\Phi}_r^{-1}(1) = \begin{pmatrix} \hat{\Sigma}_{\gamma_m} & \hat{\Sigma}_{\gamma_m, \hat{\theta}_n} \\ \hat{\Sigma}'_{\gamma_m, \hat{\theta}_n} & \hat{\Sigma}_{\hat{\theta}_n} \end{pmatrix}, \quad \hat{\Sigma}_{\hat{u}_r} = \frac{1}{n} \sum_{t=1}^n \hat{u}_{r,t} \hat{u}'_{r,t}.$$

8. Définir l'estimateur

$$\hat{\Phi}_m = \frac{1}{n} \sum_{t=1}^n \left\{ (\hat{e}'_{t-1}, \dots, \hat{e}'_{t-m})' \otimes \frac{\partial e_t(\theta_0)}{\partial \theta'} \right\}_{\theta_0 = \hat{\theta}_n}.$$

9. Définir les estimateurs

$$\begin{aligned} \hat{\Sigma}_{\hat{\Gamma}_m} &= \hat{\Sigma}_{\gamma_m} + \hat{\Phi}_m \hat{\Sigma}_{\hat{\theta}_n} \hat{\Phi}'_m + \hat{\Phi}_m \hat{\Sigma}_{\hat{\theta}_n, \gamma_m} + \hat{\Sigma}'_{\hat{\theta}_n, \gamma_m} \hat{\Phi}'_m \\ \hat{\Sigma}_{\hat{\rho}_m} &= \left\{ I_m \otimes (\hat{S}_e \otimes \hat{S}_e)^{-1} \right\} \hat{\Sigma}_{\hat{\Gamma}_m} \left\{ I_m \otimes (\hat{S}_e \otimes \hat{S}_e)^{-1} \right\} \end{aligned}$$

10. Calculer les valeurs propres $\hat{\xi}_m = (\hat{\xi}_{1,d^2m}, \dots, \hat{\xi}_{d^2m,d^2m})'$ de la matrice

$$\hat{\Omega}_m = \left(I_m \otimes \hat{\Sigma}_{e0}^{-1/2} \otimes \hat{\Sigma}_{e0}^{-1/2} \right) \hat{\Sigma}_{\hat{\Gamma}_m} \left(I_m \otimes \hat{\Sigma}_{e0}^{-1/2} \otimes \hat{\Sigma}_{e0}^{-1/2} \right).$$

11. Calculer les statistiques des tests portmanteau de **BP** et de **LB**

$$\begin{aligned} P_m &= n \hat{\rho}'_m \left(I_m \otimes \left\{ \hat{R}_e^{-1}(0) \otimes \hat{R}_e^{-1}(0) \right\} \right) \hat{\rho}_m \quad \text{et} \\ \tilde{P}_m &= n^2 \sum_{h=1}^m \frac{1}{(n-h)} \text{Tr} \left(\hat{\Gamma}'_e(h) \hat{\Gamma}_e^{-1}(0) \hat{\Gamma}_e(h) \hat{\Gamma}_e^{-1}(0) \right). \end{aligned}$$

12. Évaluer les p -valeurs

$$P \left\{ Z_m(\hat{\xi}_m) > P_m \right\} \quad \text{et} \quad P \left\{ Z_m(\hat{\xi}_m) > \tilde{P}_m \right\}, \quad Z_m(\hat{\xi}_m) = \sum_{i=1}^{d^2m} \hat{\xi}_{i,d^2m} Z_i^2,$$

en utilisant l'algorithme de Imhof (1961). Le test de **BP** (resp. celui de **LB**) rejette l'adéquation de modèles VARMA(p_0, q_0) faibles quand la première (resp. la seconde) p -valeur est inférieure au niveau asymptotique α .

Les expériences de Monte Carlo réalisées dans ce chapitre sur des modèles VARMA bivariés montrent que les résultats des tests modifiés de **LB** et de **BP** sont meilleurs que ceux des tests standard de **LB** et de **BP**, en particulier quand les innovations présentent des dépendances et quand le nombre d'autocorrélations $m = 1, 2$ (*i.e.* $m \leq p_0 + q_0$). Il est donc préférable d'utiliser la distribution que nous proposons dans le théorème 1.9 plutôt que l'approximation chi-deux usuelle. Nous constatons aussi que, comme dans le cas où les erreurs sont gaussiennes (Hosking, 1980), la statistique de **LB** a de meilleurs résultats pour des échantillons de petites tailles que celle de **BP** pour des erreurs dépendantes.

1.4 Résultats du chapitre 5

Dans l'étape d'identification de la traditionnelle méthodologie de Box et Jenkins, un des problèmes les plus délicats est celui de la sélection d'un petit nombre de valeurs plausibles pour les ordres p_0 et q_0 du modèle VARMA.

Parmi les méthodes d'identification, les plus populaires sont celles basées sur l'optimisation d'un critère d'information. Le critère d'information de Akaike (noté AIC pour Akaike's information criterion) est le plus utilisé. L'objectif de ce chapitre est d'étudier le problème de sélection des ordres p_0 et q_0 de modèles VARMA dont les termes d'erreur sont non corrélés mais non nécessairement indépendants. Les fondements théoriques du critère AIC ne sont plus établis lorsque l'hypothèse de bruit iid est relâchée. Afin de remédier à ce problème, nous proposons un critère d'information de Akaike modifié (noté AIC_M).

1.4.1 Modèle et paramétrisation des coefficients

Nous utilisons la méthode du quasi-maximum de vraisemblance déjà définie dans le chapitre 2 pour l'estimation des paramètres du modèle ARMA(p_0, q_0) structurel d -multivarié (1.1) dont le terme d'erreur $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{dt})'$ est une suite de variables aléatoires centrées, non corrélées, avec une matrice de covariance non singulière Σ . Les paramètres sont les coefficients des matrices carrées d'ordre d suivantes : A_i , $i \in \{1, \dots, p_0\}$, B_j , $j \in \{1, \dots, q_0\}$ et Σ . Nous supposons que ces matrices sont paramétrées suivant le vecteur des vraies valeurs des paramètres noté θ_0 . Nous notons $A_{0i} = A_i(\theta_0)$, $i \in \{1, \dots, p_0\}$, $B_j = B_j(\theta_0)$, $j \in \{1, \dots, q_0\}$ et $\Sigma_0 = \Sigma(\theta_0)$, où θ_0 appartient à l'espace compact des paramètres $\Theta_{p_0, q_0} \subset \mathbb{R}^{k_0}$, et k_0 est le nombre de paramètres inconnus qui

est inférieur à $(p_0 + q_0 + 3)d^2$ (*i.e.* le nombre de paramètres sans aucune contrainte). Sous les hypothèses \mathbf{H}_1 – \mathbf{H}_6 , nous avons établi dans le chapitre 2 la consistance ($\hat{\theta}_n \rightarrow \theta_0$ *p.s.* quand $n \rightarrow \infty$) et la normalité asymptotique du QMLE :

$$\sqrt{n} (\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Omega := J^{-1} I J^{-1}), \quad (1.15)$$

où $J = J(\theta_0)$ et $I = I(\theta_0)$, avec

$$J(\theta) = \lim_{n \rightarrow \infty} \frac{2}{n} \frac{\partial^2}{\partial \theta \partial \theta'} \log L_n(\theta) \quad p.s$$

et

$$I(\theta) = \lim_{n \rightarrow \infty} \text{Var} \frac{2}{\sqrt{n}} \frac{\partial}{\partial \theta} \log L_n(\theta).$$

Dans le chapitre 3, sous les hypothèses \mathbf{H}_1 – \mathbf{H}_7 , nous avons obtenu des expressions explicites des matrices I_{11} et J_{11} , données par

$$\text{vec } J_{11} = 2 \sum_{i \geq 1} \mathcal{M} \{ \lambda'_i \otimes \lambda'_i \} \text{vec } \Sigma_{e_0}^{-1} \quad \text{et}$$

$$\text{vec } I_{11} = 4 \sum_{i,j=1}^{+\infty} \Gamma(i, j) (\{I_d \otimes \lambda'_j\} \otimes \{I_d \otimes \lambda'_i\}) \text{vec} \left(\text{vec } \Sigma_{e_0}^{-1} \{ \text{vec } \Sigma_{e_0}^{-1} \}' \right),$$

où la matrice

$$\mathcal{M} := E \left\{ (I_{d^2(p_0+q_0)} \otimes e'_t)^{\otimes 2} \right\},$$

les matrices λ_i dépendent du paramètre θ_0 et les matrices

$$\Gamma(i, j) = \sum_{h=-\infty}^{+\infty} E \left\{ \{e'_{t-h} \otimes (I_{d^2(p_0+q_0)} \otimes e'_{t-j-h})\} \otimes \{e'_t \otimes (I_{d^2(p_0+q_0)} \otimes e'_{t-i})\} \right\}$$

ont déjà été définies.

1.4.2 Définition du critère d'information

En statistique on est souvent confronté au problème de l'identification d'un modèle parmi m modèles, où les modèles candidats ont des paramètres θ_m de dimensions k_m . Le choix peut s'effectuer en estimant chaque modèle et en minimisant un critère de la forme

mesure de l'erreur d'ajustement + terme de pénalisation.

On mesure souvent cette erreur d'ajustement par la somme des carrés des résidus, ou encore par -2 fois la log-quasi-vraisemblance *i.e.* $-2 \log \tilde{L}_n(\hat{\theta}_m)$. Le terme de pénalisation est une fonction croissante de la dimension k_m du paramètre $\hat{\theta}_m$. Ce terme est

indispensable car la mesure de l'erreur d'ajustement est systématiquement minimale pour le modèle qui possède le plus grand nombre de paramètres, lorsque $\hat{\theta}_m$ minimise cette erreur d'ajustement et quand les modèles sont emboîtés. Le critère le plus connu est sans doute le AIC introduit par Akaike (1973) et défini de manière suivante

$$\text{AIC}(\text{modèle } m \text{ estimé}) = -2 \log \tilde{L}_n(\hat{\theta}_m) + \dim(\hat{\theta}_m).$$

Le critère AIC est fondé sur un estimateur de la divergence de Kullback-Leibler dont nous en parlerons ultérieurement. Tsai et Hurvich (1989, 1993) ont proposé une correction de biais du critère AIC pour des modèles de séries temporelles autorégressifs dans les cas univarié et multivarié sous les hypothèses que les innovations ϵ_t sont indépendantes et identiquement distribuées.

1.4.3 Contraste de Kullback-Leibler

Soit les observations $X = (X_1, \dots, X_n)$ de loi à densité f_0 inconnue par rapport à une mesure σ -finie μ . Considérons un modèle candidat m avec lequel les observations seraient à densité $f_m(\cdot, \theta_m)$, avec un paramètre θ_m de dimension k_m . L'écart entre le modèle candidat et le vrai modèle peut être mesuré par la divergence (ou information, ou entropie) de Kullback-Leibler

$$\Delta \{f_m(\cdot, \theta_m) | f_0\} = E_{f_0} \log \frac{f_0(X)}{f_m(X, \theta_m)} = E_{f_0} \log f_0(X) + \frac{1}{2} d \{f_m(\cdot, \theta_m) | f_0\},$$

où

$$d \{f_m(\cdot, \theta_m) | f_0\} = -2 E_{f_0} \log f_m(X, \theta_m) = -2 \int \{\log f_m(x, \theta_m)\} f_0(x) \mu(dx)$$

est souvent appelé le contraste de Kullback-Leibler (ou l'écart entre le modèle ajusté et le vrai modèle). Notons que $\Delta(f|f_0)$ n'est pas symétrique en f et f_0 , et s'interprète comme la capacité moyenne d'une observation de loi f_0 à se distinguer d'une observation de loi f . On peut aussi interpréter $\Delta(f|f_0)$ comme étant une perte d'information liée à l'utilisation du modèle f quand le vrai modèle est f_0 . En utilisant l'inégalité de Jensen, nous montrons que $\Delta \{f_m(\cdot, \theta_m) | f_0\}$ existe toujours dans $[0, \infty]$:

$$\begin{aligned} \Delta \{f_m(\cdot, \theta_m) | f_0\} &= - \int \log \frac{f_m(x, \theta_m)}{f_0(x)} f_0(x) \mu(dx) \\ &\geq - \log \int \frac{f_m(x, \theta_m)}{f_0(x)} f_0(x) \mu(dx) = 0, \end{aligned}$$

avec égalité si et seulement si les lois de $f_m(\cdot, \theta_m)$ et f_0 sont les mêmes *i.e.* $f_m(\cdot, \theta_m) = f_0$. Telle est la propriété principale de la divergence de Kullback-Leibler. Minimiser

$\Delta \{f_m(\cdot, \theta_m) | f_0\}$ revient à minimiser le contraste de Kullback-Leibler $d \{f_m(\cdot, \theta_m) | f_0\}$. Soit

$$\theta_{0,m} = \arg \inf_{\theta_m} d \{f_m(\cdot, \theta_m) | f_0\} = \arg \inf_{\theta_m} -2E \log f_m(X, \theta_m)$$

le paramètre optimal du modèle m (nous supposons qu'un tel paramètre $\theta_{0,m}$ existe). Nous estimons ce paramètre par le QMLE $\hat{\theta}_{n,m}$.

1.4.4 Critère de sélection des ordres d'un modèle VARMA

Soit

$$\begin{aligned} \tilde{\ell}_n(\theta) &= -\frac{2}{n} \log \tilde{L}_n(\theta) \\ &= \frac{1}{n} \sum_{t=1}^n \{d \log(2\pi) + \log \det \Sigma_e + \tilde{e}'_t(\theta) \Sigma_e^{-1} \tilde{e}_t(\theta)\}. \end{aligned}$$

Dans le lemme 7 de l'annexe du chapitre 2, nous avons montré que $\ell_n(\theta) = \tilde{\ell}_n(\theta) + o(1)$ p.s, où

$$\begin{aligned} \ell_n(\theta) &= -\frac{2}{n} \log L_n(\theta) \\ &= \frac{1}{n} \sum_{t=1}^n \{d \log(2\pi) + \log \det \Sigma_e + e'_t(\theta) \Sigma_e^{-1} e_t(\theta)\}, \end{aligned}$$

et où

$$e_t(\theta) = A_0^{-1} B_0 B_\theta^{-1}(L) A_\theta(L) X_t.$$

Nous avons aussi montré uniformément en $\theta \in \Theta$ que

$$\frac{\partial \ell_n(\theta)}{\partial \theta} = \frac{\partial \tilde{\ell}_n(\theta)}{\partial \theta} + o(1) \quad p.s.,$$

dans le lemme 9 de l'annexe du même chapitre. Notons que les mêmes égalités sont aussi vérifiées pour les dérivées secondes de $\tilde{\ell}_n$. Pour tout $\theta \in \Theta$, nous avons

$$-2 \log L_n(\theta) = nd \log(2\pi) + n \log \det \Sigma_e + \sum_{t=1}^n e'_t(\theta) \Sigma_e^{-1} e_t(\theta).$$

Dans la partie précédente consacrée au contraste de Kullback-Leibler, nous avons vu que minimiser l'information de Kullback-Leibler pour tout modèle candidat caractérisé par le paramètre θ , revient à minimiser le contraste $\Delta(\theta) = E \{-2 \log L_n(\theta)\}$. En omettant la constante $nd \log(2\pi)$, nous trouvons que

$$\Delta(\theta) = n \log \det \Sigma_e + n \text{Tr} (\Sigma_e^{-1} S(\theta)),$$

où $S(\theta) = E e_1(\theta) e'_1(\theta)$. Nous énonçons le lemme suivant.

Lemme 1.4. Pour tout $\theta \in \bigcup_{p,q \in \mathbb{N}} \Theta_{p,q}$, nous avons

$$\Delta(\theta) \geq \Delta(\theta_0).$$

Nous montrons ce lemme en utilisant le théorème 1.12 ergodique et l'inégalité élémentaire $\text{Tr}(A^{-1}B) - \log \det(A^{-1}B) \geq \text{Tr}(A^{-1}A) - \log \det(A^{-1}A) = d$ pour toute matrice symétrique semi-définie positive de taille $d \times d$.

En vue de ce lemme, l'application $\theta \mapsto \Delta(\theta)$ est minimale en $\theta = \theta_0$.

Estimation de la divergence moyenne

Soit $X = (X_1, \dots, X_n)$ les observations satisfaisant la représentation VARMA (1.1). Posons $\hat{\theta}_n$ le QMLE du paramètre θ du modèle VARMA candidat. Soit $\hat{e}_t = \tilde{e}_t(\hat{\theta}_n)$ les résidus du QMLE/LSE quand $p_0 > 0$ et $q_0 > 0$, et soit $\hat{e}_t = e_t = X_t$ quand $p_0 = q_0 = 0$. Quand $p + q \neq 0$, nous avons $\hat{e}_t = 0$ pour $t \leq 0$ et $t > n$, et posons

$$\hat{e}_t = X_t - \sum_{i=1}^{p_0} A_0^{-1}(\hat{\theta}_n) A_i(\hat{\theta}_n) \hat{X}_{t-i} + \sum_{i=1}^{q_0} A_0^{-1}(\hat{\theta}_n) B_i(\hat{\theta}_n) B_0^{-1}(\hat{\theta}_n) A_0(\hat{\theta}_n) \hat{e}_{t-i},$$

pour $t = 1, \dots, n$, avec $\hat{X}_t = 0$ pour $t \leq 0$ et $\hat{X}_t = X_t$ pour $t \geq 1$. En vue du Lemme 1.4, il est naturel de minimiser le contraste moyen $E\Delta(\hat{\theta}_n)$. L'écart $\{\Delta(\hat{\theta}_n) - \Delta(\theta_0)\}$ s'interprète comme une perte de précision globale moyenne quand on utilise le modèle estimé à la place du vrai modèle. Nous allons par la suite adapter le critère d'information de Akaike corrigé (noté AICc) introduit par Hurvich et Tsai (1989) pour obtenir un estimateur approximativement sans biais de $E\Delta(\hat{\theta}_n)$. En utilisant un développement limité de Taylor de $\partial \log L_n(\hat{\theta}_n) / \partial \theta^{(1)}$ au voisinage de $\theta_0^{(1)}$, il s'en suit que

$$\left(\hat{\theta}_n^{(1)} - \theta_0^{(1)}\right) \stackrel{op}{\underset{=}{\approx}} \frac{1}{\sqrt{n}} - \frac{2}{n} J_{11}^{-1} \frac{\partial \log L_n(\theta_0)}{\partial \theta^{(1)}} = -\frac{2}{n} \sum_{t=1}^n J_{11}^{-1} \frac{\partial e'_t(\theta_0)}{\partial \theta^{(1)}} \Sigma_{e_0}^{-1} e_t(\theta_0), \quad (1.16)$$

où $J_{11} = J_{11}(\theta_0)$ avec

$$J_{11}(\theta) = \lim_{n \rightarrow \infty} \frac{2}{n} \frac{\partial^2 \log L_n(\theta)}{\partial \theta^{(1)} \partial \theta^{(1)'}} \quad p.s.$$

Nous avons

$$E\Delta(\hat{\theta}_n) = En \log \det \hat{\Sigma}_e + nE\text{Tr} \left(\hat{\Sigma}_e^{-1} S(\hat{\theta}_n) \right), \quad (1.17)$$

où $\hat{\Sigma}_e = \Sigma_e(\hat{\theta}_n)$ est un estimateur de la matrice de variance des erreurs, avec $\Sigma_e(\theta) = n^{-1} \sum_{t=1}^n e_t(\theta) e'_t(\theta)$. Ensuite, le premier terme du second membre de l'équation (1.17) peut être estimée sans biais par $n \log \det \left\{ n^{-1} \sum_{t=1}^n e_t(\hat{\theta}_n) e'_t(\hat{\theta}_n) \right\}$. Par conséquent, nous

prenons uniquement en considération l'estimation du deuxième terme. En outre, en vue de (1.15), un développement de Taylor de $e_t(\theta)$ au voisinage de $\theta_0^{(1)}$ vérifie

$$e_t(\theta) = e_t(\theta_0) + \frac{\partial e_t(\theta_0)}{\partial \theta^{(1)'}} (\theta^{(1)} - \theta_0^{(1)}) + R_t, \quad (1.18)$$

où

$$R_t = \frac{1}{2} (\theta^{(1)} - \theta_0^{(1)})' \frac{\partial^2 e_t(\theta^*)}{\partial \theta^{(1)} \partial \theta^{(1)'}} (\theta^{(1)} - \theta_0^{(1)}) = O_P(\pi^2),$$

avec $\pi = \|\theta^{(1)} - \theta_0^{(1)}\|$ et θ^* est entre $\theta_0^{(1)}$ et $\theta^{(1)}$. Nous obtenons alors

$$\begin{aligned} S(\theta) &= S(\theta_0) + E \left\{ \frac{\partial e_t(\theta_0)}{\partial \theta^{(1)'}} (\theta^{(1)} - \theta_0^{(1)}) e_t'(\theta_0) \right\} + ER_t e_t'(\theta_0) \\ &\quad + E \left\{ e_t(\theta_0) (\theta^{(1)} - \theta_0^{(1)})' \frac{\partial e_t'(\theta_0)}{\partial \theta^{(1)}} \right\} + D(\theta^{(1)}) \\ &\quad + ER_t \left\{ (\theta^{(1)} - \theta_0^{(1)})' \frac{\partial e_t'(\theta_0)}{\partial \theta^{(1)}} \right\} + E e_t(\theta_0) R_t \\ &\quad + E \left\{ \frac{\partial e_t(\theta_0)}{\partial \theta^{(1)'}} (\theta^{(1)} - \theta_0^{(1)}) \right\} R_t + ER_t^2, \end{aligned}$$

où

$$D(\theta^{(1)}) = E \left\{ \frac{\partial e_t(\theta_0)}{\partial \theta^{(1)'}} (\theta^{(1)} - \theta_0^{(1)}) (\theta^{(1)} - \theta_0^{(1)})' \frac{\partial e_t'(\theta_0)}{\partial \theta^{(1)}} \right\}.$$

En utilisant les propriétés d'orthogonalité entre $e_t(\theta_0)$ et toute combinaison linéaire des valeurs passées de $e_t(\theta_0)$ (en particulier $\partial e_t(\theta_0)/\partial \theta'$ et $\partial^2 e_t(\theta_0)/\partial \theta \partial \theta'$), et le fait que les innovations sont centrées *i.e.* $E e_t(\theta_0) = 0$, nous avons

$$S(\theta) = S(\theta_0) + D(\theta^{(1)}) + O(\pi^4) = \Sigma_{e0} + D(\theta^{(1)}) + O(\pi^4),$$

où $\Sigma_{e0} = \Sigma_e(\theta_0)$. Nous pouvons donc écrire l'écart moyen (1.17) comme

$$\begin{aligned} E\Delta(\hat{\theta}_n) &= En \log \det \hat{\Sigma}_e + nE\text{Tr} \left(\hat{\Sigma}_e^{-1} \Sigma_{e0} \right) + nE\text{Tr} \left(\hat{\Sigma}_e^{-1} D(\hat{\theta}_n^{(1)}) \right) \\ &\quad + nE\text{Tr} \left(\hat{\Sigma}_e^{-1} \right) O_P \left(\frac{1}{n^2} \right). \end{aligned} \quad (1.19)$$

Comme dans une régression multivariée classique, nous avons la relation

$$\Sigma_{e0} \approx \frac{n}{n - d(p+q)} E \left\{ \hat{\Sigma}_e \right\} = \frac{dn}{dn - k_1} E \left\{ \hat{\Sigma}_e \right\}.$$

De cette précédente approximation et de la consistance de $\hat{\Sigma}_e$, nous déduisons

$$E \left\{ \hat{\Sigma}_e^{-1} \right\} \approx \left\{ E \hat{\Sigma}_e \right\}^{-1} \approx nd(nd - k_1)^{-1} \Sigma_{e0}^{-1}. \quad (1.20)$$

En utilisant des propriétés élémentaires sur la trace d'une matrice, nous avons

$$\begin{aligned} \text{Tr} \left\{ \Sigma_e^{-1}(\theta) D(\theta_n^{(1)}) \right\} &= \text{Tr} \left(\Sigma_e^{-1}(\theta) E \left\{ \frac{\partial e_t(\theta_0)}{\partial \theta^{(1)'}} (\theta^{(1)} - \theta_0^{(1)}) (\theta^{(1)} - \theta_0^{(1)})' \frac{\partial e_t'(\theta_0)}{\partial \theta^{(1)}} \right\} \right) \\ &= E \left(\text{Tr} \left\{ \frac{\partial e_t'(\theta_0)}{\partial \theta^{(1)}} \Sigma_e^{-1}(\theta) \frac{\partial e_t(\theta_0)}{\partial \theta^{(1)'}} (\theta^{(1)} - \theta_0^{(1)})' (\theta^{(1)} - \theta_0^{(1)}) \right\} \right) \\ &= \text{Tr} \left(E \left\{ \frac{\partial e_t'(\theta_0)}{\partial \theta^{(1)}} \Sigma_e^{-1}(\theta) \frac{\partial e_t(\theta_0)}{\partial \theta^{(1)'}} \right\} (\theta^{(1)} - \theta_0^{(1)})' (\theta^{(1)} - \theta_0^{(1)}) \right). \end{aligned}$$

Maintenant, en utilisant (1.15), (1.20) et cette dernière égalité en $\hat{\theta}_n$, nous avons

$$\begin{aligned} E \text{Tr} \left\{ \hat{\Sigma}_e^{-1} D(\hat{\theta}_n^{(1)}) \right\} &= \frac{1}{n} \text{Tr} \left(E \left\{ \frac{\partial e_t'(\theta_0)}{\partial \theta^{(1)}} \hat{\Sigma}_e^{-1} \frac{\partial e_t(\theta_0)}{\partial \theta^{(1)'}} \right\} \right. \\ &\quad \left. E n (\hat{\theta}_n^{(1)} - \theta_0^{(1)})' (\hat{\theta}_n^{(1)} - \theta_0^{(1)}) \right) \\ &= \frac{d}{nd - k_1} \text{Tr} \left(E \left\{ \frac{\partial e_t'(\theta_0)}{\partial \theta^{(1)}} \Sigma_{e0}^{-1} \frac{\partial e_t(\theta_0)}{\partial \theta^{(1)'}} \right\} J_{11}^{-1} I_{11} J_{11}^{-1} \right) \\ &= \frac{d}{2(nd - k_1)} \text{Tr} (I_{11} J_{11}^{-1}), \end{aligned}$$

où $J_{11} = 2E \left\{ \frac{\partial e_t'(\theta_0)}{\partial \theta^{(1)}} \Sigma_{e0}^{-1} \frac{\partial e_t(\theta_0)}{\partial \theta^{(1)'}} \right\}$ (Voir le Théorème 3 dans Boubacar Maïnassara et Francq, 2009). Par conséquent, en utilisant (1.20), nous avons

$$\begin{aligned} E \text{Tr} \left(\hat{\Sigma}_e^{-1} S(\hat{\theta}_n) \right) &= E \text{Tr} \left(\hat{\Sigma}_e^{-1} \Sigma_{e0} \right) + E \text{Tr} \left\{ \hat{\Sigma}_e^{-1} D(\hat{\theta}_n^{(1)}) \right\} \\ &\quad + E \left\{ \text{Tr} \left(\hat{\Sigma}_e^{-1} \right) O_P \left(\frac{1}{n^2} \right) \right\} \\ &= \frac{nd}{nd - k_1} \text{Tr} (\Sigma_{e0}^{-1} \Sigma_{e0}) + \frac{d}{2(nd - k_1)} \text{Tr} (I_{11} J_{11}^{-1}) + O \left(\frac{1}{n^2} \right) \\ &= \frac{nd^2}{nd - k_1} + \frac{d}{2(nd - k_1)} \text{Tr} (I_{11} J_{11}^{-1}) + O \left(\frac{1}{n^2} \right). \end{aligned}$$

Ainsi, en utilisant cette dernière égalité dans $E\Delta(\hat{\theta}_n)$, sous les hypothèses \mathbf{H}_1 – \mathbf{H}_7 nous obtenons un estimateur approximativement sans biais de $E\Delta(\hat{\theta}_n)$ (noté AIC_M pour AIC "modifié") donné par

$$\text{AIC}_M = n \log \det \hat{\Sigma}_e + \frac{n^2 d^2}{nd - k_1} + \frac{nd}{2(nd - k_1)} \text{Tr} \left(\hat{I}_{11,n} \hat{J}_{11,n}^{-1} \right),$$

où les vecteurs $\text{vec } \hat{I}_{11,n}$ et $\text{vec } \hat{J}_{11,n}$ sont des estimateurs des vecteurs $\text{vec } I_{11}$ et $\text{vec } J_{11}$ définis à la sous-section 1.4.1. En utilisant la relation $\text{Tr}(AB) = \text{vec}(A)' \text{vec}(B)$, nous avons

$$\text{AIC}_M = n \log \det \hat{\Sigma}_e + \frac{n^2 d^2}{nd - k_1} + \frac{nd}{2(nd - k_1)} \left(\text{vec } \hat{I}'_{11,n} \right)' \left(\text{vec } \hat{J}_{11,n}^{-1} \right). \quad (1.21)$$

Nous obtenons des estimateurs \hat{p} et \hat{q} des ordres p_0 et q_0 en minimisant le critère modifié (1.21).

Autre décomposition du contraste (écart)

Dans la section précédente le contraste (écart) minimal a été approximé par $-2E \log L_n(\hat{\theta}_n)$ (l'espérance est prise avec l'observation du vrai modèle X). Notons que l'étude de cette moyenne de l'écart est trop difficile en raison de la dépendance entre l'estimateur $\hat{\theta}_n$ et l'observation X . Une méthode alternative (légèrement différente de celle de la section précédente mais équivalente en interprétation) pour arriver à la quantité $E\Delta(\hat{\theta}_n)$ (écart moyen) consiste à considérer $\hat{\theta}_n$ comme étant le QMLE de θ fondé sur l'observation X et soit $Y = (Y_1, \dots, Y_n)$ l'observation indépendante de X et générée par le même modèle (1.1). Ensuite, nous approximons la distribution de (Y_t) par $L_n(Y, \hat{\theta}_n)$. Nous considérons donc l'écart moyen du modèle ajusté (modèle candidat Y) en $\hat{\theta}_n$. Ainsi, il est généralement plus facile de chercher un modèle qui minimise

$$C(\hat{\theta}_n) = -2E_Y \log L_n(\hat{\theta}_n), \quad (1.22)$$

où E_Y est l'espérance de l'observation Y du modèle candidat. Puisque $\hat{\theta}_n$ et Y sont indépendants, $C(\hat{\theta}_n)$ est la même quantité que l'écart moyen $E\Delta(\hat{\theta}_n)$. Le modèle minimisant (1.22) peut être interprété comme le modèle qui aura une meilleure approximation globale sur une copie indépendante d'observations de X , mais ce modèle peut ne pas être le meilleur pour les données à portée de main. Cette moyenne du contraste peut être décomposée comme

$$C(\hat{\theta}_n) = -2E_X \log L_n(\hat{\theta}_n) + a_1 + a_2,$$

où

$$a_1 = -2E_X \log L_n(\theta_0) + 2E_X \log L_n(\hat{\theta}_n)$$

et

$$a_2 = -2E_Y \log L_n(\hat{\theta}_n) + 2E_X \log L_n(\theta_0).$$

Le QMLE satisfait $\log L_n(\hat{\theta}_n) \geq \log L_n(\theta_0)$ presque sûrement (c'est-à-dire un ajustement anormal des données au modèle quand ce sont ces mêmes données qui ont ajusté le modèle), donc le terme a_1 s'interprète comme le sur-ajustement de ce QMLE. Notons que $E_X \log L_n(\theta_0) = E_Y \log L_n(\theta_0)$, donc le terme a_2 s'interprète comme le coût moyen d'utilisation (sur replications indépendantes de l'observation X) du modèle estimé à la place du vrai modèle. Nous discutons maintenant dans la proposition suivante des conditions de régularité dont nous aurons besoin afin que les termes a_1 et a_2 soient équivalents au nombre $\text{Tr}(I_{11}J_{11}^{-1})$.

Proposition 1.5. Sous les hypothèses \mathbf{H}_1 – \mathbf{H}_7 , quand $n \rightarrow \infty$, nous avons a_1 (sur-ajustement moyen) et a_2 (perte de précision sur replications indépendantes) sont tous deux égaux à $\text{Tr}(I_{11}J_{11}^{-1})$.

Remarque 1.9. Dans le cas de modèles VARMA forts c'est-à-dire quand l'hypothèse d'ergodicité \mathbf{H}_3 est remplacée par celle dont les termes d'erreurs sont iid, nous avons

$I_{11} = 2J_{11}$ et $\text{Tr}(I_{11}J_{11}^{-1}) = k_1$. Par conséquent, en vue de la proposition 1.5, nous obtenons que les termes a_1 (sur-ajustement moyen) et a_2 (perte de précision sur replications indépendantes) sont tous deux égaux à fois $k_1 = \dim(\theta_0^{(1)})$ (nombre de paramètres). Dans ce cas, un estimateur approximativement sans biais de $E\Delta(\hat{\theta}_n)$ prend la forme suivante

$$\begin{aligned} AIC_M &= n \log \det \hat{\Sigma}_e + \frac{n^2 d^2}{nd - k_1} + \frac{nd}{2(nd - k_1)} 2k_1 \\ &= n \log \det \hat{\Sigma}_e + \frac{nd}{nd - k_1} (nd + k_1) \\ &= n \log \det \hat{\Sigma}_e + nd + \frac{nd}{nd - k_1} 2k_1, \end{aligned} \quad (1.23)$$

qui illustre que les critères AIC standard et AIC_M ne diffèrent que de l'inclusion d'un terme de pénalité $nd/(nd - k_1)$. Ce facteur peut jouer un rôle important dans performance du AIC_M si k_1 est non négligeable par rapport à la taille de l'échantillon n . En particulier, ce facteur contribue à réduire le biais du AIC, lequel peut être important quand n n'est pas grand. Par conséquent, l'utilisation de cet estimateur amélioré de l'écart moyen devrait conduire à une amélioration des performances du AIC_M plus que le AIC en termes de choix du modèle.

Remarque 1.10. Sous certaines hypothèses de régularité, il est montré dans Findley (1993) que, les termes a_1 et a_2 sont tous les deux équivalents à k_1 . Dans ce cas, le critère AIC

$$AIC = -2 \log L_n(\hat{\theta}_n) + 2k_1 \quad (1.24)$$

est un estimateur approximativement sans biais du contraste $C(\hat{\theta}_n)$. La sélection des ordres du modèle est obtenue en minimisant (1.24) pour les modèles candidats.

Remarque 1.11. Pour une famille donnée de modèles candidats, nous préférons celui qui minimise $E\Delta(\hat{\theta}_n)$. Ainsi les ordres \hat{p} et \hat{q} du modèle sélectionné sont choisis dans l'ensemble des ordres qui minimise le critère d'information (1.21).

Remarque 1.12. En considérant le cas univarié $d = 1$, nous obtenons

$$AIC_M = n\hat{\sigma}_e^2 + \frac{n}{n - (p_0 + q_0)} \left(n + \frac{1}{\hat{\sigma}^4} \sum_{i,j,i'=1}^{+\infty} \hat{\gamma}(i,j) \left\{ \hat{\lambda}_j \hat{\lambda}_{i'}^{-1'} \otimes \hat{\lambda}_i \hat{\lambda}_{i'}^{-1'} \right\} \right),$$

où $\hat{\sigma}_e^2$ est la variance estimée du processus univarié (e_t) et où $\hat{\gamma}(i,j)$ sont des estimateurs des $\gamma(i,j) = \sum_{h=-\infty}^{+\infty} E(e_t e_{t-i} e_{t-h} e_{t-j-h})$. Les vecteurs $\lambda'_i \in \mathbb{R}^{p_0+q_0}$ sont les estimateurs des vecteurs définis par (1.6).

1.5 Annexe

Pour une lecture indépendante de l'ouvrage, cette annexe regroupe des résultats élémentaires de stationnarité, ergodicité et mélange. Nous reprenons pour cela l'annexe du livre de Francq et Zakoïan (2009), en énonçant certains résultats dans un cadre multivarié.

1.5.1 Stationnarité

La stationnarité joue un rôle majeur en séries temporelles car elle remplace de manière naturelle l'hypothèse d'observations iid en statistique standard. Garantissant que l'accroissement de la taille de l'échantillon s'accompagne d'une augmentation du même ordre de l'information, la stationnarité est à la base d'une théorie asymptotique générale. On considère la notion de stationnarité forte ou stricte et celle de la stationnarité faible ou au second ordre. Nous considérons un processus vectoriel $(X_t)_{t \in \mathbb{Z}}$ de dimension d , $X_t = (X_{1t}, \dots, X_{dt})'$.

Définition 1.6. Le processus vectoriel (X_t) est dit strictement stationnaire si les vecteurs $(X'_1, \dots, X'_t)'$ et $(X'_{1+h}, \dots, X'_{t+h})'$ ont la même loi jointe pour tout $h \in \mathbb{Z}$ et tout $t \in \mathbb{Z}$.

La définition de la stationnarité stricte s'énonce donc de la même manière en univarié et en multivarié. La stationnarité au second ordre se définit comme suit.

Définition 1.7. Le processus vectoriel (X_t) à valeurs réelles est dit stationnaire au second ordre si

- (i) $EX_{it}^2 < \infty \quad \forall t \in \mathbb{Z}, i \in \{1, \dots, d\}$,
- (ii) $EX_t = m \quad \forall t \in \mathbb{Z}$,
- (iii) $\text{Cov}(X_t, X_{t+h}) = E[(X_t - m)(X_{t+h} - m)'] = \Gamma(h) \quad \forall h, t \in \mathbb{Z}$.

La fonction $\Gamma(\cdot)$, à valeur dans l'espace des matrices $d \times d$ est appelée fonction d'auto-covariance de (X_t) .

Il est évident que $\Gamma_X(h) = \Gamma'_X(-h)$. En particulier $\Gamma_X(0) = \text{Var}(X_t)$ est une matrice symétrique. Cette définition de la stationnarité faible est moins exigeante que la précédente car elle n'impose de contraintes qu'aux deux premiers moments des variables (X_t) , mais contrairement à la stationnarité stricte, elle requiert l'existence de

ceux-ci. L'exemple le plus simple de processus stationnaire multivarié est le bruit blanc, défini comme une suite de variables centrées et non corrélées, de matrice de variance indépendante du temps.

Définition 1.8. Le processus vectoriel (ϵ_t) est appelé bruit blanc faible s'il vérifie

- (i) $E\epsilon_t = 0 \quad \forall t \in \mathbb{Z}$,
- (ii) $E\epsilon_t\epsilon_t' = \Sigma \quad \forall t \in \mathbb{Z}$,
- (iii) $\text{Cov}(\epsilon_t, \epsilon_{t+h}) = 0 \quad \forall h, t \in \mathbb{Z}, h \neq 0$,

où Σ , est une matrice $d \times d$ de variance-covariance non singulière.

Remarque 1.13. *Il est important de noter qu'aucune hypothèse d'indépendance n'est faite dans la définition du bruit blanc faible. Les variables aux différentes dates sont seulement non corrélées. Il est parfois nécessaire de remplacer l'hypothèse (iii) par l'hypothèse plus forte*

- (iii') *les variables ϵ_t sont des différences de martingale.*

On parle alors de bruit blanc semi-fort.

Il est parfois nécessaire de remplacer l'hypothèse (iii) par une hypothèse encore plus forte

- (iii'') *les variables ϵ_t et ϵ_{t-h} sont indépendantes.*

On parle alors de bruit blanc fort.

1.5.2 Ergodicité

On dit qu'une suite vectoriel stationnaire est ergodique si elle satisfait la loi forte des grands nombres.

Définition 1.9. Un processus vectoriel strictement stationnaire $(X_t)_{t \in \mathbb{Z}}$, à valeurs dans \mathbb{R}^d , est dit ergodique si et seulement si, pour tout borélien B de $\mathbb{R}^{d(h-1)}$ et tout entier h ,

$$\frac{1}{n} \sum_{t=1}^n I_B(X'_t, X'_{t+1}, \dots, X'_{t+h})' \rightarrow P \left[(X'_1, \dots, X'_{1+h})' \in B \right]$$

avec probabilité 1.

Certaines transformations de suites ergodiques restent ergodiques.

Théorème 1.11. Si $(X_t)_{t \in \mathbb{Z}}$ est un processus vectoriel strictement stationnaire ergodique de dimension d et si $(Y_t)_{t \in \mathbb{Z}}$ est défini par

$$Y_t = f(\dots, X'_{t-1}, X'_t, X'_{t+1}, \dots),$$

où f est une fonction mesurable de \mathbb{R}^∞ dans $\mathbb{R}^{\bar{d}}$, alors $(Y_t)_{t \in \mathbb{Z}}$ est également un processus vectoriel strictement stationnaire ergodique.

Théorème 1.12. Si $(X_t)_{t \in \mathbb{Z}}$ est un processus vectoriel strictement stationnaire et ergodique de dimension d , si f est une fonction mesurable de \mathbb{R}^∞ dans $\mathbb{R}^{\bar{d}}$ et si $E \|f(\dots, X'_{t-1}, X'_t, X'_{t+1}, \dots)\| < \infty$, alors

$$\frac{1}{n} \sum_{t=1}^n f(\dots, X'_{t-1}, X'_t, X'_{t+1}, \dots) \xrightarrow{p.s.} E f(\dots, X'_{t-1}, X'_t, X'_{t+1}, \dots).$$

1.5.3 Accroissement de martingale

Dans un jeu équitable de hasard pur (par exemple F et G jouent à pile ou face, F donne un euro à G quand la pièce fait pile, G donne un euro à F quand la pièce fait face), la fortune d'un joueur est une martingale.

Définition 1.10. Soit $(X_t)_{t \in \mathbb{N}}$ est un processus vectoriel de variables aléatoires réelles (v.a.r) et $(\mathcal{F}_t)_{t \in \mathbb{N}}$ une suite de tribus. La suite $(X_t, \mathcal{F}_t)_{t \in \mathbb{N}}$ est une martingale si et seulement si

1. $\mathcal{F}_t \subset \mathcal{F}_{t+1}$,
2. X_t est \mathcal{F}_t -mesurable,
3. $E \|X_t\| < \infty$,
4. $E(X_{t+1} | \mathcal{F}_t) = X_t$.

Quand on dit que $(X_t)_{t \in \mathbb{N}}$ est une martingale, on prend implicitement $\mathcal{F}_t = \sigma(X_u, u \leq t)$, c'est-à-dire la tribu engendrée par les valeurs passées et présentes.

Définition 1.11. Soit $(\eta_t)_{t \in \mathbb{N}}$ un processus vectoriel de v.a.r et $(\mathcal{F}_t)_{t \in \mathbb{N}}$ une suite de tribus. La suite $(\eta_t, \mathcal{F}_t)_{t \in \mathbb{N}}$ est une différence de martingale (ou une suite d'accroissement de martingale) si et seulement si

1. $\mathcal{F}_t \subset \mathcal{F}_{t+1}$,
2. η_t est \mathcal{F}_t -mesurable,
3. $E \|\eta_t\| < \infty$,
4. $E(\eta_{t+1} | \mathcal{F}_t) = 0$.

Remarque 1.14. Si la suite $(X_t, \mathcal{F}_t)_{t \in \mathbb{N}}$ est une martingale et si on pose $\eta_0 = X_0$, $\eta_t = X_t - X_{t-1}$, alors la suite $(\eta_t, \mathcal{F}_t)_{t \in \mathbb{N}}$ est une différence de martingale : $E(\eta_{t+1} | \mathcal{F}_t) = E(X_{t+1} | \mathcal{F}_t) - E(X_t | \mathcal{F}_t) = 0$.

Remarque 1.15. Si la suite $(\eta_t, \mathcal{F}_t)_{t \in \mathbb{N}}$ est une différence de martingale et si on pose $X_t = \sum_{n=0}^t \eta_n$, alors la suite $(X_t, \mathcal{F}_t)_{t \in \mathbb{N}}$ est une martingale : $E(X_{t+1} | \mathcal{F}_t) = E(\{X_t + \eta_{t+1}\} | \mathcal{F}_t) = X_t$.

1.5.4 Mélange

Définition de l' α -mélange

Dans un espace probabilisé $(\Omega, \mathcal{A}_0, P)$, considérons \mathcal{A} et \mathcal{B} deux sous-tribus de \mathcal{A}_0 . Nous définissons le coefficient de α -mélange (ou de mélange fort) entre ces deux tribus par

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |P(A \cap B) - P(A)P(B)|.$$

Il est évident que si \mathcal{A} et \mathcal{B} sont indépendantes alors $\alpha(\mathcal{A}, \mathcal{B}) = 0$. Si par contre $\mathcal{A} = \mathcal{B} = \{A, A^c, \emptyset, \Omega\}$ avec $P(A) = 1/2$ alors $\alpha(\mathcal{A}, \mathcal{B}) = 1/4$. Les coefficients de mélange fort d'un processus vectoriel $X = (X_t)$ sont définis par

$$\alpha_X(h) = \sup_t \alpha[\sigma(X_u, \leq t), \sigma(X_u, \geq t+h)].$$

Si X est stationnaire, on peut enlever le terme \sup_t . On dit que X est fortement mélangeant si $\alpha_X(h) \rightarrow 0$ quand $h \rightarrow \infty$.

Inégalité de covariance

Soient p, q et r trois nombres positifs tels que $p^{-1} + q^{-1} + r^{-1} = 1$. Davydov (1968) a montré l'inégalité de covariance suivante

$$|\text{Cov}(X, Y)| \leq K_0 \|X\|_p \|Y\|_q [\alpha\{\sigma(X), \sigma(Y)\}]^{1/r}, \quad (1.25)$$

où $\|X\|_p^p = EX^p$ et K_0 est une constante universelle. Dans le cas univarié, Davydov avait proposé $K_0 = 12$. Rio (1993) a obtenu une inégalité plus fine, faisant intervenir les fonctions quantiles de X et Y . Cette inégalité montre également que l'on peut prendre $K_0 = 4$ dans (1.25). Notons que (1.25) entraîne que la fonction d'autocovariance d'un processus stationnaire α -mélangeant (et possédant des moments suffisants) tend vers zéro. Ces résultats peuvent être directement étendus à un processus vectoriel.

Théorème central limite (TCL)

Herrndorf (1984) a montré le TCL suivant pour un processus α -mélangeant

Théorème 1.13. Soit $X = (X_t)$ un processus centré tel que

$$\sup_t \|X_t\|_{2+\nu} < \infty \quad \text{et} \quad \sum_{k=0}^{\infty} \{\alpha_X(k)\}^{\frac{\nu}{2+\nu}} < \infty \quad \text{pour un } \nu > 0.$$

Si $\sigma^2 = \lim_{n \rightarrow \infty} \text{Var} \left(n^{-1/2} \sum_{t=1}^n X_t \right)$ existe et est non nulle, alors

$$n^{-1/2} \sum_{t=1}^n X_t \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

Ce résultat aussi peut être directement étendu à un processus vectoriel.

Références bibliographiques

- Ahn, S. K.** (1988) Distribution for residual autocovariances in multivariate autoregressive models with structured parameterization. *Biometrika* 75, 590–93.
- Akaike, H.** (1973) Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory*, Eds. B. N. Petrov and F. Csáki, pp. 267–281. Budapest : Akadémia Kiado.
- Andrews, D.W.K.** (1991) Heteroskedasticity and autocorrélation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- Andrews, B., Davis, R. A. et Breidt, F.J.** (2006) Maximum likelihood estimation for all-pass time series models. *Journal of Multivariate Analysis* 97, 1638–59.
- Bauwens, L., Laurent, S. et Rombouts, J.V.K.** (2006) Multivariate GARCH models : a survey. *Journal of Applied Econometrics* 21, 79–109.
- Berk, K.N.** (1974) Consistent Autoregressive Spectral Estimates. *Annals of Statistics* 2, 489–502.
- Bollerslev, T.** (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.
- Bollerslev, T.** (1988) On the correlation structure for the generalized autoregressive conditional heteroskedastic process. *Journal of Time Series Analysis* 9, 121–131.
- Box, G. E. P. et Pierce, D. A.** (1970) Distribution of residual autocorrélations in autoregressive integrated moving average time series models. *Journal of the American Statistical Association* 65, 1509–26.
- Brockwell, P.J. et Davis, R.A.** (1991) *Time series : theory and methods*. Springer Verlag, New York.
- Broze, L., Francq, C. et Zakoïan, J-M.** (2001) Non redundancy of high order moment conditions for efficient GMM estimation of weak AR processes. *Economics Letters* 71, 317–322.
- Chitturi, R. V.** (1974) Distribution of residual autocorrélations in multiple autoregressive schemes. *Journal of the American Statistical Association* 69, 928–934.
- Davydov, Y. A.** (1968) Convergence of Distributions Generated by Stationary Stochastic Processes. *Theory of Probability and Applications* 13, 691–696.
- den Hann, W.J. et Levin, A.** (1997) A Practitioner’s Guide to Robust Covariance Matrix Estimation. In *Handbook of Statistics* 15, 291–341.
- Drost, F. C. et Nijman T. E.** (1993) Temporal Aggregation of GARCH Processes. *Econometrica*, 61, 4 909–927.

- Duchesne, P. et Roy, R.** (2004) On consistent testing for serial correlation of unknown form in vector time series models. *Journal of Multivariate Analysis* 89, 148–180.
- Dufour, J.-M. et Pelletier, D.** (2005) Practical methods for modelling weak VARMA processes : identification, estimation and specification with a macroeconomic application. *Technical report, Département de sciences économiques and CIREQ, Université de Montréal, Montréal, Canada.*
- Dunsmuir, W.T.M. et Hannan, E.J.** (1976) Vector linear time series models, *Advances in Applied Probability* 8, 339–364.
- Engle, R. F.** (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* 50, 987–1007.
- Fan, J. et Yao, Q.** (2003) *Nonlinear time series : Nonparametric and parametric methods.* Springer Verlag, New York.
- Findley, D.F.** (1993) The overfitting principles supporting AIC, Statistical Research Division Report RR 93/04, Bureau of the Census.
- Franccq, C. and Raïssi, H.** (2006) Multivariate Portmanteau Test for Autoregressive Models with Uncorrelated but Nonindependent Errors, *Journal of Time Series Analysis* 28, 454–470.
- Franccq, C., Roy, R. et Zakoïan, J.-M.** (2005) Diagnostic checking in ARMA Models with Uncorrelated Errors, *Journal of the American Statistical Association* 100, 532–544.
- Franccq, C. et Zakoïan, J.-M.** (1998) Estimating linear representations of nonlinear processes, *Journal of Statistical Planning and Inference* 68, 145–165.
- Franccq, C. et Zakoïan, J.-M.** (2000) Covariance matrix estimation of mixing weak ARMA models, *Journal of Statistical Planning and Inference* 83, 369–394.
- Franccq, C. et Zakoïan, J.-M.** (2005) Recent results for linear time series models with non independent innovations. In *Statistical Modeling and Analysis for Complex Data Problems*, Chap. 12 (eds P. DUCHESNE and B. RÉMILLARD). New York : Springer Verlag, 137–161.
- Franccq, C. et Zakoïan, J.-M.** (2007) HAC estimation and strong linearity testing in weak ARMA models, *Journal of Multivariate Analysis* 98, 114–144.
- Franccq, C. et Zakoïan, J.-M.** (2009) *Modèles GARCH : structure, inférence statistique et applications financières.* Collection Economie et Statistiques Avancées, Economica.
- Hall, P. et Heyde, C. C.** (1980) *Martingale limit theory and its application* Academic press, New York.
- Hannan, E.J.** (1976) The identification and parametrization of ARMAX and state space forms, *Econometrica* 44, 713–723.
- Hannan, E.J. et Deistler, M.** (1988) *The Statistical Theory of Linear Systems* John Wiley, New York.

- Hannan, E.J., Dunsmuir, W.T.M. et Deistler, M.** (1980) Estimation of vector ARMAX models, *Journal of Multivariate Analysis* 10, 275–295.
- Hannan, E.J. et Rissanen** (1982) Recursive estimation of mixed of Autoregressive Moving Average order, *Biometrika* 69, 81–94.
- Herrndorf, N.** (1984) A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables. *The Annals of Probability* 12, 141–153.
- Hosking, J.R.M.** (1980) The multivariate portmanteau statistic, *Journal of the American Statistical Association* 75, 602–608.
- Hosking, J.R.M.** (1981a) Equivalent forms of the multivariate portmanteau statistic, *Journal of the Royal Statistical Society B* 43, 261–262.
- Hosking, J.R.M.** (1981b) Lagrange-tests of multivariate time series models, *Journal of the Royal Statistical Society B* 43, 219–230.
- Hurvich, C. M. et Tsai, C-L.** (1989) Regression and time series model selection in small samples. *Biometrika* 76, 297–307.
- Hurvich, C. M. et Tsai, C-L.** (1993) A corrected Akaike information criterion for vector autoregressive model selection. *Journal of Time Series Analysis* 14, 271–279.
- Imhof, J.P.** (1961) Computing the distribution of quadratic forms in normal variables. *Biometrika* 48, 419–426.
- Jeantheau, T.** (1998) Strong consistency of estimators for multivariate ARCH models, *Econometric Theory* 14, 70–86.
- Kascha, C.** (2007) A Comparison of Estimation Methods for Vector Autoregressive Moving-Average Models. *ECO Working Papers, EUI ECO 2007/12*, <http://hdl.handle.net/1814/6921>.
- Li, W. K. et McLeod, A. I.** (1981) Distribution of the residual autocorrelations in multivariate ARMA time series models, *Journal of the Royal Statistical Society B* 43, 231–239.
- Lütkepohl, H.** (1993) *Introduction to multiple time series analysis*. Springer Verlag, Berlin.
- Lütkepohl, H.** (2005) *New introduction to multiple time series analysis*. Springer Verlag, Berlin.
- Lütkepohl, H. et Claessen, H.** (1997) Analysis of cointegrated VARMA processes. *Journal of Econometrics* 80, 223–239.
- Magnus, J.R. et H. Neudecker** (1988) *Matrix Differential Calculus with Application in Statistics and Econometrics*. New-York, Wiley.
- McLeod, A.I.** (1978) On the distribution of residual autocorrelations in Box-Jenkins models, *Journal of the Royal Statistical Society B* 40, 296–302.
- Newey, W.K., et West, K.D.** (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

- Nijman, T. E. et Sentana, E.** (1994) Marginalization and Contemporaneous Aggregation in Multivariate GARCH Processes. *Journal of Econometrics*, 71, 71-87.
- Nsiri, S., et Roy, R.** (1996) Identification of refined ARMA echelon form models for multivariate time series, *Journal of Multivariate Analysis* 56, 207-231
- Raïssi, H.** (2009) Autocorrelation based tests for vector error correction models with uncorrelated but nonindependent errors. A paraître dans *Test*.
- Reinsel, G.C.** (1993) *Elements of multivariate time series*. Springer Verlag, New York.
- Reinsel, G.C.** (1997) *Elements of multivariate time series Analysis*. Second edition. Springer Verlag, New York.
- Reinsel, G.C., Basu S. et Yap, S.F.** (1992) Maximum likelihood estimators in the multivariate Autoregressive Moving Average Model from a generalized least squares viewpoint, *Journal of Time Series Analysis* 13, 133-145.
- Rissanen, J. et Caines, P.E.** (1979) The strong consistency of maximum likelihood estimators for ARMA processes. *Annals of Statistics* 7, 297-315.
- Romano, J.L. et Thombs, L.A.** (1996) Inference for autocorrelations under weak assumptions, *Journal of the American Statistical Association* 91, 590-600.
- Schwarz, G.** (1978) Estimating the dimension of a model. *Annals of Statistics* 6, 461-464.
- Tong, H.** (1990) Non-linear time series : *A Dynamical System Approach*, Clarendon Press Oxford.
- van der Vaart, A.W.** (1998) Asymptotic statistics, *Cambridge University Press*, Cambridge.
- Wald, A.** (1949) Note on the consistency of the maximum likelihood estimate. *Annals of Mathematical Statistics* 20, 595-601.
- Wold, H.** (1938) A study in the analysis of stationary time series, Uppsala, Almqvist and Wiksell.

Chapitre 2

Estimating structural VARMA models with uncorrelated but non-independent error terms

Abstract The asymptotic properties of the quasi-maximum likelihood estimator (QMLE) of vector autoregressive moving-average (VARMA) models are derived under the assumption that the errors are uncorrelated but not necessarily independent. Relaxing the independence assumption considerably extends the range of application of the VARMA models, and allows to cover linear representations of general nonlinear processes. Conditions are given for the consistency and asymptotic normality of the QMLE. A particular attention is given to the estimation of the asymptotic variance matrix, which may be very different from that obtained in the standard framework. Modified versions of the Wald, Lagrange Multiplier and Likelihood Ratio tests are proposed for testing linear restrictions on the parameters.

Keywords : Echelon form, Lagrange Multiplier test, Likelihood Ratio test, Nonlinear processes, QMLE, Structural representation, VARMA models, Wald test.

2.1 Introduction

This paper is devoted to the problem of estimating VARMA representations of multivariate (nonlinear) processes.

In order to give a precise definition of a linear model and of a nonlinear process, first recall that by the Wold decomposition (see *e.g.* Brockwell and Davis, 1991, for

the univariate case, and Reinsel, 1997, in the multivariate framework) any zero-mean purely non deterministic d -dimensional stationary process (X_t) can be written in the form

$$X_t = \sum_{\ell=0}^{\infty} \Psi_{\ell} \epsilon_{t-\ell}, \quad (\epsilon_t) \sim \text{WN}(0, \Sigma) \quad (2.1)$$

where $\sum_{\ell} \|\Psi_{\ell}\|^2 < \infty$. The process (ϵ_t) is called the linear innovation process of the process $X = (X_t)$, and the notation $(\epsilon_t) \sim \text{WN}(0, \Sigma)$ signifies that (ϵ_t) is a *weak white noise*. A weak white noise is a stationary sequence of centered and uncorrelated random variables with common variance matrix Σ . By contrast, a *strong white noise*, denoted by $\text{IID}(0, \Sigma)$, is an independent and identically distributed (iid) sequence of random variables with mean 0 and variance Σ . A strong white noise is obviously a weak white noise, because independence entails uncorrelatedness, but the reverse is not true. Between weak and strong white noises, one can define a semi-strong white noise as a stationary martingale difference. An example of semi-strong white noise is the generalized autoregressive conditional heteroscedastic (GARCH) model. In the present paper, a process X is said to be *linear* when $(\epsilon_t) \sim \text{IID}(0, \Sigma)$ in (2.1), and is said to be *nonlinear* in the opposite case. With this definition, GARCH-type processes are considered as nonlinear. Leading examples of linear processes are the VARMA and the sub-class of the vector autoregressive (VAR) models with iid noise. Nonlinear models are becoming more and more employed because numerous real time series exhibit nonlinear dynamics, for instance conditional heteroscedasticity, which can not be generated by autoregressive moving-average (ARMA) models with iid noises.¹

The main issue with nonlinear models is that they are generally hard to identify and implement. This is why it is interesting to consider weak (V)ARMA models, that is ARMA models with weak white noises, such linear representations being universal approximations of the Wold decomposition (2.1). Linear and nonlinear processes also have exact weak ARMA representations because a same process may satisfy several models, and many important classes of nonlinear processes admit weak ARMA representations (see Francq, Roy and Zakoïan, 2005, and the references therein).

The estimation of autoregressive moving-average (ARMA) models is however much more difficult in the multivariate than in univariate case. A first difficulty is that non trivial constraints on the parameters must be imposed for identifiability of the parameters (see Reinsel, 1997, Lütkepohl, 2005). Secondly, the implementation of standard estimation methods (for instance the Gaussian quasi-maximum likelihood estimation) is not obvious because this requires a constrained high-dimensional optimization (see Lütkepohl, 2005, for a general reference and Kascha, 2007, for a numerical comparison

1. To cite few examples of nonlinear processes, let us mention the self-exciting threshold autoregressive (SETAR), the smooth transition autoregressive (STAR), the exponential autoregressive (EX-PAR), the bilinear, the random coefficient autoregressive (RCA), the functional autoregressive (FAR) (see Tong, 1990, and Fan and Yao, 2003, for references on these nonlinear time series models). All these nonlinear models have been initially proposed for univariate time series, but have multivariate extensions.

of alternative estimation methods of VARMA models). These technical difficulties certainly explain why VAR models are much more used than VARMA in applied works. This is also the reason why the asymptotic theory of weak ARMA model estimation is mainly limited to the univariate framework (see Francq and Zakoïan, 2005, for a review on weak ARMA models). Notable exceptions are Dufour and Pelletier (2005) who study the asymptotic properties of a generalization of the regression-based estimation method proposed by Hannan and Rissanen (1982) under weak assumptions on the innovation process, and Francq and Raïssi (2007) who study portmanteau tests for weak VAR models.

For the estimation of ARMA and VARMA models, the commonly used estimation method is the QMLE, which can also be viewed as a nonlinear least squares estimation (LSE). The asymptotic properties of the QMLE of VARMA models are well-known under the restrictive assumption that the errors ϵ_t are independent (see Lütkepohl, 2005). The asymptotic behavior of the QMLE has been studied in a much wider context by Dunsmuir and Hannan (1976) and Hannan and Deistler (1988) who proved consistency, under weak assumptions on the noise process and based on a spectral analysis. These authors also obtained asymptotic normality under a conditionally homoscedastic martingale difference assumption on the linear innovations. However, this assumption precludes most of the nonlinear models. Little is thus known when the martingale difference assumption is relaxed. Our aim in this paper is to consider a flexible VARMA specification covering the structural forms encountered in econometrics, and to relax the independence assumption, and even the martingale difference assumption, in order to be able to cover weak VARMA representations of general nonlinear models.

The paper is organized as follows. Section 2.2 presents the structural weak VARMA models that we consider here. Structural forms are employed in econometrics in order to introduce instantaneous relationships between economic variables. The identifiability issues are discussed. It is shown in Section 2.3 that the QMLE is strongly consistent when the weak white noise (ϵ_t) is ergodic, and that the QMLE is asymptotically normally distributed when (ϵ_t) satisfies mild mixing assumptions. The asymptotic variance of the QMLE may be very different in the weak and strong cases. Section 2.4 is devoted to the estimation of this covariance matrix. In Section 2.5 it is shown how the standard Wald, LM (Lagrange multiplier) and LR (likelihood ratio) tests must be adapted in the weak VARMA case in order to test for general linearity constraints. This section is also of interest in the univariate framework because, to our knowledge, these tests have not been studied for weak ARMA models. Numerical experiments are presented in Section 2.6. The proofs of the main results are collected in the appendix.

2.2 Model and assumptions

Consider a d -dimensional stationary process (X_t) satisfying a structural VARMA(p, q) representation of the form

$$A_{00}X_t - \sum_{i=1}^p A_{0i}X_{t-i} = B_{00}\epsilon_t - \sum_{i=1}^q B_{0i}\epsilon_{t-i}, \quad (\epsilon_t) \sim \text{WN}(0, \Sigma_0), \quad (2.2)$$

where Σ_0 is non singular and $t \in \mathbb{Z} = \{0, \pm 1, \dots\}$. The standard VARMA(p, q) form, which is sometimes called the reduced form, is obtained for $A_{00} = B_{00} = I_d$. The structural forms are mainly used in econometrics to identify structural economic shocks and to allow instantaneous relationships between economic variables. Of course, constraints are necessary for the identifiability of the $(p+q+3)d^2$ elements of the matrices involved in the VARMA equation (2.2). We thus assume that these matrices are parameterized by a vector ϑ_0 of lower dimension. We then write $A_{0i} = A_i(\vartheta_0)$ and $B_{0j} = B_j(\vartheta_0)$ for $i = 0, \dots, p$ and $j = 0, \dots, q$, and $\Sigma_0 = \Sigma(\vartheta_0)$, where ϑ_0 belongs to the parameter space $\Theta \subset \mathbb{R}^{k_0}$, and k_0 is the number of unknown parameters, which is typically much smaller than $(p+q+3)d^2$. The parametrization is often linear (see Example 2.1 below), and thus satisfies the following smoothness conditions.

A1 : The applications $\vartheta \mapsto A_i(\vartheta)$ $i = 0, \dots, p$, $\vartheta \mapsto B_j(\vartheta)$ $j = 0, \dots, q$ and $\vartheta \mapsto \Sigma(\vartheta)$ admit continuous third order derivatives for all $\vartheta \in \Theta$.

For simplicity we now write A_i, B_j and Σ instead of $A_i(\vartheta), B_j(\vartheta)$ and $\Sigma(\vartheta)$. Let $A_\vartheta(z) = A_0 - \sum_{i=1}^p A_i z^i$ and $B_\vartheta(z) = B_0 - \sum_{i=1}^q B_i z^i$. We assume that Θ corresponds to stable and invertible representations, namely

A2 : for all $\vartheta \in \Theta$, we have $\det A_\vartheta(z) \det B_\vartheta(z) \neq 0$ for all $|z| \leq 1$.

To show the strong consistency of the QMLE, we will use the following assumptions.

A3 : We have $\vartheta_0 \in \Theta$, where Θ is compact.

A4 : The process (ϵ_t) is stationary and ergodic.

A5 : For all $\vartheta \in \Theta$ such that $\vartheta \neq \vartheta_0$, either the transfer functions

$$A_0^{-1} B_0 B_\vartheta^{-1}(z) A_\vartheta(z) \neq A_{00}^{-1} B_{00} B_{\vartheta_0}^{-1}(z) A_{\vartheta_0}(z)$$

for some $z \in \mathbb{C}$, or

$$A_0^{-1} B_0 \Sigma B_0' A_0^{-1'} \neq A_{00}^{-1} B_{00} \Sigma_0 B_{00}' A_{00}^{-1'}.$$

Remark 2.1. The previous identifiability assumption is satisfied when the parameter space Θ is sufficiently constrained. Note that the last condition in **A5** can be dropped for the standard reduced forms in which $A_0 = B_0 = I_d$, but may be important for structural VARMA forms (see Example 2.1 below). The identifiability of VARMA processes has been studied in particular by Hannan (1976) who gave several procedures ensuring identifiability. In particular **A5** is satisfied when we impose $A_0 = B_0 = I_d$, **A2**, the common left divisors of $A_\vartheta(L)$ and $B_\vartheta(L)$ are unimodular (i.e. with nonzero constant determinant), and the matrix $[A_p : B_q]$ is of full rank.

The structural form (2.2) allows to handle seasonal models, instantaneous economic relationships, VARMA in the so-called echelon form representation, and many other constrained VARMA representations.

Example 2.1. Assume that income (Inc) and consumption (Cons) variables are related by the equations $\text{Inc}_t = c_1 + \alpha_{01} \text{Inc}_{t-1} + \alpha_{02} \text{Cons}_{t-1} + \epsilon_{1t}$ and $\text{Cons}_t = c_2 + \alpha_{03} \text{Inc}_t + \alpha_{04} \text{Inc}_{t-1} + \alpha_{05} \text{Cons}_{t-1} + \epsilon_{2t}$. In the stationary case the process $X_t = \{\text{Inc}_t - E(\text{Inc}_t), \text{Cons}_t - E(\text{Cons}_t)\}'$ satisfies a structural VAR(1) equation

$$\begin{pmatrix} 1 & 0 \\ -\alpha_{03} & 1 \end{pmatrix} X_t = \begin{pmatrix} \alpha_{01} & \alpha_{02} \\ \alpha_{04} & \alpha_{05} \end{pmatrix} X_{t-1} + \epsilon_t.$$

We also assume that the two components of ϵ_t correspond to uncorrelated structural economic shocks, with respective variances σ_{01}^2 and σ_{02}^2 . We thus have

$$\vartheta_0' = (\alpha_{01}, \alpha_{02}, \alpha_{03}, \alpha_{04}, \alpha_{05}, \sigma_{01}^2, \sigma_{02}^2).$$

Note that the identifiability condition **A5** is satisfied because for all $\vartheta = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \sigma_1^2, \sigma_2^2)' \neq \vartheta_0$ we have

$$I_2 - \begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_1\alpha_3 + \alpha_4 & \alpha_2\alpha_3 + \alpha_5 \end{pmatrix} z \neq I_2 - \begin{pmatrix} \alpha_{01} & \alpha_{02} \\ \alpha_{01}\alpha_{03} + \alpha_{04} & \alpha_{02}\alpha_{03} + \alpha_{05} \end{pmatrix} z$$

for some $z \in \mathbb{C}$, or

$$\begin{pmatrix} \sigma_1^2 & \sigma_1^2\alpha_3 \\ \sigma_1^2\alpha_3 & \sigma_1^2\alpha_3^2 + \sigma_2^2 \end{pmatrix} \neq \begin{pmatrix} \sigma_{01}^2 & \sigma_{01}^2\alpha_{03} \\ \sigma_{01}^2\alpha_{03} & \sigma_{01}^2\alpha_{03}^2 + \sigma_{02}^2 \end{pmatrix}.$$

2.3 Quasi-maximum likelihood estimation

Let X_1, \dots, X_n be observations of a process satisfying the VARMA representation (2.2). Note that from **A2** the matrices A_{00} and B_{00} are invertible. Introducing the innovation process

$$e_t = A_{00}^{-1} B_{00} \epsilon_t,$$

the structural representation $A_{\vartheta_0}(L)X_t = B_{\vartheta_0}(L)\epsilon_t$ can be rewritten as the reduced VARMA representation

$$X_t - \sum_{i=1}^p A_{00}^{-1} A_{0i} X_{t-i} = e_t - \sum_{i=1}^q A_{00}^{-1} B_{0i} B_{00}^{-1} A_{00} e_{t-i}. \quad (2.3)$$

For all $\vartheta \in \Theta$, we recursively define $\tilde{e}_t(\vartheta)$ for $t = 1, \dots, n$ by

$$\tilde{e}_t(\vartheta) = X_t - \sum_{i=1}^p A_0^{-1} A_i X_{t-i} + \sum_{i=1}^q A_0^{-1} B_i B_0^{-1} A_0 \tilde{e}_{t-i}(\vartheta),$$

with initial values $\tilde{e}_0(\vartheta) = \cdots = \tilde{e}_{1-q}(\vartheta) = X_0 = \cdots = X_{1-p} = 0$. It will be shown that these initial values are asymptotically negligible and, in particular, that $\tilde{e}_t(\vartheta_0) - e_t \rightarrow 0$ almost surely as $t \rightarrow \infty$. The Gaussian quasi-likelihood is given by

$$\tilde{L}_n(\vartheta) = \prod_{t=1}^n \frac{1}{(2\pi)^{d/2} \sqrt{\det \Sigma_e}} \exp \left\{ -\frac{1}{2} \tilde{e}'_t(\vartheta) \Sigma_e^{-1} \tilde{e}_t(\vartheta) \right\},$$

where

$$\Sigma_e = \Sigma_e(\vartheta) = A_0^{-1} B_0 \Sigma B_0' A_0^{-1'}$$

Note that the variance of e_t is $\Sigma_{e0} = \Sigma_e(\vartheta_0) = A_{00}^{-1} B_{00} \Sigma_0 B_{00}' A_{00}^{-1'}$. A quasi-maximum likelihood estimator (QMLE) is a measurable solution $\hat{\vartheta}_n$ of

$$\hat{\vartheta}_n = \arg \max_{\vartheta \in \Theta} \tilde{L}_n(\vartheta) = \arg \min_{\vartheta \in \Theta} \tilde{\ell}_n(\vartheta), \quad \tilde{\ell}_n(\vartheta) = \frac{-2}{n} \log \tilde{L}_n(\vartheta).$$

The following theorem shows that, for the consistency of the QMLE, the conventional assumption that the noise (ϵ_t) is an iid sequence can be replaced by the less restrictive ergodicity assumption **A4**. Dunsmuir and Hannan (1976) for VARMA in reduced form, and Hannan and Deistler (1988) for VARMAX models, obtained an equivalent result, using spectral analysis. For the proof, we do not use the spectral analysis techniques employed by the above-mentioned authors, but we follow the classical technique of Wald (1949), as was done by Rissanen and Caines (1979) to show the strong consistency of the Gaussian maximum likelihood estimator of VARMA models.

Theorem 2.1. *Let (X_t) be the causal solution of the VARMA equation (2.2) satisfying **A1–A5** and let $\hat{\vartheta}_n$ be a QMLE. Then $\hat{\vartheta}_n \rightarrow \vartheta_0$ a.s. as $n \rightarrow \infty$.*

For the asymptotic normality of the QMLE, it is necessary to assume that ϑ_0 is not on the boundary of the parameter space Θ .

A6 : We have $\vartheta_0 \in \overset{\circ}{\Theta}$, where $\overset{\circ}{\Theta}$ denotes the interior of Θ .

We now introduce mixing assumptions similar to those made by Francq and Zakoïan (1998), hereafter FZ. We denote by $\alpha_\epsilon(k)$, $k = 0, 1, \dots$, the strong mixing coefficients of the process (ϵ_t) .

A7 : We have $E\|\epsilon_t\|^{4+2\nu} < \infty$ and $\sum_{k=0}^{\infty} \{\alpha_\epsilon(k)\}^{\frac{\nu}{2+\nu}} < \infty$ for some $\nu > 0$. We define the matrix of the coefficients of the reduced form (2.3) by

$$M_{\vartheta_0} = [A_{00}^{-1} A_{01} : \cdots : A_{00}^{-1} A_{0p} : A_{00}^{-1} B_{01} B_{00}^{-1} A_{00} : \cdots : A_{00}^{-1} B_{0q} B_{00}^{-1} A_{00} : \Sigma_{e0}].$$

Now we need an assumption which specifies how this matrix depends on the parameter ϑ_0 . Let \dot{M}_{ϑ_0} be the matrix $\partial \text{vec}(M_\vartheta) / \partial \vartheta'$ evaluated at ϑ_0 .

A8 : The matrix \dot{M}_{ϑ_0} is of full rank k_0 .

One can easily verify that **A8** is satisfied in Example 2.1.

Theorem 2.2. *Under the assumptions of Theorem 2.1, and A6-A8, we have*

$$\sqrt{n} \left(\hat{\vartheta}_n - \vartheta_0 \right) \xrightarrow{L} \mathcal{N}(0, \Omega := J^{-1} I J^{-1}),$$

where $J = J(\vartheta_0)$ and $I = I(\vartheta_0)$, with

$$J(\vartheta) = \lim_{n \rightarrow \infty} \frac{\partial^2}{\partial \vartheta \partial \vartheta'} \tilde{\ell}_n(\vartheta) \quad a.s., \quad I(\vartheta) = \lim_{n \rightarrow \infty} \text{Var} \frac{\partial}{\partial \vartheta} \tilde{\ell}_n(\vartheta).$$

For VARMA models in reduced form, it is not very restrictive to assume that the coefficients $A_0, \dots, A_p, B_0, \dots, B_q$ are functionally independent of the coefficient Σ_e . Thus we can write $\vartheta = (\vartheta^{(1)'}, \vartheta^{(2)'})'$, where $\vartheta^{(1)} \in \mathbb{R}^{k_1}$ depends on A_0, \dots, A_p and B_0, \dots, B_q , and where $\vartheta^{(2)} \in \mathbb{R}^{k_2}$ depends on Σ_e , with $k_1 + k_2 = k_0$. With some abuse of notation, we will then write $e_t(\vartheta) = e_t(\vartheta^{(1)})$.

A9 : With the previous notation $\vartheta = (\vartheta^{(1)'}, \vartheta^{(2)'})'$, where $\vartheta^{(2)} = D \text{vec} \Sigma_e$ for some matrix D of size $k_2 \times d^2$.

The following theorem shows that for VARMA in reduced form, the QMLE and LSE coincide. We denote by $A \otimes B$ the Kronecker product of two matrices A and B .

Theorem 2.3. *Under the assumptions of Theorem 2.2 and A9 the QMLE $\hat{\vartheta}_n = (\hat{\vartheta}_n^{(1)'}, \hat{\vartheta}_n^{(2)'})'$ can be obtained from*

$$\hat{\vartheta}_n^{(2)} = D \text{vec} \hat{\Sigma}_e, \quad \hat{\Sigma}_e = \frac{1}{n} \sum_{t=1}^n \tilde{e}_t(\hat{\vartheta}_n^{(1)}) \tilde{e}_t'(\hat{\vartheta}_n^{(1)}),$$

and

$$\hat{\vartheta}_n^{(1)} = \arg \min_{\vartheta^{(1)}} \det \sum_{t=1}^n \tilde{e}_t(\vartheta^{(1)}) \tilde{e}_t'(\vartheta^{(1)}).$$

Moreover

$$J = \begin{pmatrix} J_{11} & 0 \\ 0 & J_{22} \end{pmatrix}, \quad \text{with } J_{11} = 2E \left\{ \frac{\partial}{\partial \vartheta^{(1)}} e_t'(\vartheta_0^{(1)}) \right\} \Sigma_{e_0}^{-1} \left\{ \frac{\partial}{\partial \vartheta^{(1)'}} e_t(\vartheta_0^{(1)}) \right\}$$

and $J_{22} = D(\Sigma_{e_0}^{-1} \otimes \Sigma_{e_0}^{-1})D'$.

Remark 2.2. *One can see that J has the same expression in the strong and weak ARMA cases (see Lütkepohl (2005) page 480). On the contrary, the matrix I is in general much more complicated in the weak case than in the strong case.*

Remark 2.3. *In the standard strong VARMA case, i.e. when A4 is replaced by the assumption that (ϵ_t) is iid, we have $I = 2J$, so that $\Omega = 2J^{-1}$. In the general case we have $I \neq 2J$. As a consequence the ready-made software used to fit VARMA do not provide a correct estimation of Ω for weak VARMA processes. The problem also holds in the univariate case (see Francq and Zakoïan, 2007, and the references therein).*

2.4 Estimating the asymptotic variance matrix

Theorem 2.2 can be used to obtain confidence intervals and significance tests for the parameters. The asymptotic variance Ω must however be estimated. The matrix J can easily be estimated by its empirical counterpart. For instance, under **A9**, one can take

$$\hat{J} = \begin{pmatrix} \hat{J}_{11} & 0 \\ 0 & \hat{J}_{22} \end{pmatrix}, \quad \hat{J}_{11} = \frac{2}{n} \sum_{t=1}^n \left\{ \frac{\partial}{\partial \vartheta^{(1)}} \tilde{e}'_t(\hat{\vartheta}_n^{(1)}) \right\} \hat{\Sigma}_e^{-1} \left\{ \frac{\partial}{\partial \vartheta^{(1)'}} \tilde{e}_t(\hat{\vartheta}_n^{(1)}) \right\},$$

and $\hat{J}_{22} = D(\hat{\Sigma}_e^{-1} \otimes \hat{\Sigma}_e^{-1})D'$. In the standard strong VARMA case $\hat{\Omega} = 2\hat{J}^{-1}$ is a strongly consistent estimator of Ω . In the general weak VARMA case this estimator is not consistent when $I \neq 2J$ (see Remark 2.3). So we need a consistent estimator of I . Note that

$$I = \text{Var}_{as} \frac{1}{\sqrt{n}} \sum_{t=1}^n \Upsilon_t = \sum_{h=-\infty}^{+\infty} \text{Cov}(\Upsilon_t, \Upsilon_{t-h}), \quad (2.4)$$

where

$$\Upsilon_t = \frac{\partial}{\partial \vartheta} \left\{ \log \det \Sigma_e + e'_t(\vartheta^{(1)}) \Sigma_e^{-1} e_t(\vartheta^{(1)}) \right\}_{\vartheta=\vartheta_0}. \quad (2.5)$$

In the econometric literature the nonparametric kernel estimator, also called heteroscedastic autocorrelation consistent (HAC) estimator (see Newey and West, 1987, or Andrews, 1991), is widely used to estimate covariance matrices of the form I . Let $\hat{\Upsilon}_t$ be the vector obtained by replacing ϑ_0 by $\hat{\vartheta}_n$ in Υ_t . The matrix Ω is then estimated by a "sandwich" estimator of the form

$$\hat{\Omega}^{\text{HAC}} = \hat{J}^{-1} \hat{I}^{\text{HAC}} \hat{J}^{-1}, \quad \hat{I}^{\text{HAC}} = \frac{1}{n} \sum_{t,s=1}^n \omega_{|t-s|} \hat{\Upsilon}_t \hat{\Upsilon}'_s,$$

where $\omega_0, \dots, \omega_{n-1}$ is a sequence of weights (see Andrews, 1991, and Newey and West, 1987, for the problem of the choice of weights).

Interpreting $(2\pi)^{-1}I$ as the spectral density of the stationary process (Υ_t) evaluated at frequency 0 (see Brockwell and Davis, 1991, p. 459), an alternative method consists in using a parametric AR estimate of the spectral density of (Υ_t) . This approach, which has been studied by Berk (1974) (see also den Haan and Levin, 1997), rests on the expression

$$I = \Phi^{-1}(1) \Sigma_u \Phi^{-1}(1)$$

when (Υ_t) satisfies an $\text{AR}(\infty)$ representation of the form

$$\Phi(L)\Upsilon_t := \Upsilon_t + \sum_{i=1}^{\infty} \Phi_i \Upsilon_{t-i} = u_t, \quad (2.6)$$

where u_t is a weak white noise with variance matrix Σ_u . Let $\hat{\Phi}_r(z) = I_{k_0} + \sum_{i=1}^r \hat{\Phi}_{r,i} z^i$, where $\hat{\Phi}_{r,1}, \dots, \hat{\Phi}_{r,r}$ denote the coefficients of the LS regression of $\hat{\Upsilon}_t$ on $\hat{\Upsilon}_{t-1}, \dots, \hat{\Upsilon}_{t-r}$.

Let $\hat{u}_{r,t}$ be the residuals of this regression, and let $\hat{\Sigma}_{\hat{u}_r}$ be the empirical variance of $\hat{u}_{r,1}, \dots, \hat{u}_{r,n}$.

We are now able to state the following theorem, which is an extension of a result given in Francq, Roy and Zakoïan (2005).

Theorem 2.4. *In addition to the assumptions of Theorem 2.2, assume that the process (Y_t) defined in (2.5) admits an $AR(\infty)$ representation (2.6) in which the roots of $\det \Phi(z) = 0$ are outside the unit disk, $\|\Phi_i\| = o(i^{-2})$, and $\Sigma_u = \text{Var}(u_t)$ is non-singular. Moreover we assume that $\|\epsilon_t\|_{8+4\nu} < \infty$ and $\sum_{k=0}^{\infty} \{\alpha_{X,\epsilon}(k)\}^{\nu/(2+\nu)} < \infty$ for some $\nu > 0$, where $\{\alpha_{X,\epsilon}(k)\}_{k \geq 0}$ denotes the sequence of the strong mixing coefficients of the process $(X'_t, \epsilon'_t)'$. Then the spectral estimator of I*

$$\hat{I}^{\text{SP}} := \hat{\Phi}_r^{-1}(1) \hat{\Sigma}_{\hat{u}_r} \hat{\Phi}_r'^{-1}(1) \rightarrow I$$

in probability when $r = r(n) \rightarrow \infty$ and $r^3/n \rightarrow 0$ as $n \rightarrow \infty$.

2.5 Testing linear restrictions on the parameter

It may be of interest to test s_0 linear constraints on the elements of ϑ_0 (in particular $A_{0p} = 0$ or $B_{0q} = 0$). We thus consider a null hypothesis of the form

$$H_0 : R_0 \vartheta_0 = \mathbf{r}_0$$

where R_0 is a known $s_0 \times k_0$ matrix of rank s_0 and \mathbf{r}_0 is a known s_0 -dimensional vector. The Wald, LM and LR principles are employed frequently for testing H_0 . The LM test is also called the score or Rao-score test. We now examine if these principles remain valid in the non standard framework of weak VARMA models.

Let $\hat{\Omega} = \hat{J}^{-1} \hat{I} \hat{J}^{-1}$, where \hat{J} and \hat{I} are consistent estimator of J and I , as defined in Section 2.4. Under the assumptions of Theorems 2.2 and 2.4, and the assumption that I is invertible, the Wald statistic

$$\mathbf{W}_n = n(R_0 \hat{\vartheta}_n - \mathbf{r}_0)' (R_0 \hat{\Omega} R_0')^{-1} (R_0 \hat{\vartheta}_n - \mathbf{r}_0)$$

asymptotically follows a $\chi_{s_0}^2$ distribution under H_0 . Therefore, the standard formulation of the Wald test remains valid. More precisely, at the asymptotic level α , the Wald test consists in rejecting H_0 when $\mathbf{W}_n > \chi_{s_0}^2(1 - \alpha)$. It is however important to note that a consistent estimator of the form $\hat{\Omega} = \hat{J}^{-1} \hat{I} \hat{J}^{-1}$ is required. The estimator $\hat{\Omega} = 2\hat{J}^{-1}$, which is routinely used in the time series softwares, is only valid in the strong VARMA case.

We now turn to the LM test. Let $\hat{\vartheta}_n^c$ be the restricted QMLE of the parameter under H_0 . Define the Lagrangean

$$\mathcal{L}(\vartheta, \lambda) = \tilde{\ell}_n(\vartheta) - \lambda'(R_0 \vartheta - \mathbf{r}_0),$$

where λ denotes a s_0 -dimensional vector of Lagrange multipliers. The first-order conditions yield

$$\frac{\partial \tilde{\ell}_n}{\partial \vartheta}(\hat{\vartheta}_n^c) = R_0' \hat{\lambda}, \quad R_0 \hat{\vartheta}_n^c = \mathbf{r}_0.$$

It will be convenient to write $a \stackrel{c}{=} b$ to signify $a = b + c$. A Taylor expansion gives under H_0

$$0 = \sqrt{n} \frac{\partial \tilde{\ell}_n(\hat{\vartheta}_n)}{\partial \vartheta} \stackrel{o_P(1)}{=} \sqrt{n} \frac{\partial \tilde{\ell}_n(\hat{\vartheta}_n^c)}{\partial \vartheta} - J \sqrt{n} (\hat{\vartheta}_n - \hat{\vartheta}_n^c).$$

We deduce that

$$\sqrt{n}(R_0 \hat{\vartheta}_n - \mathbf{r}_0) = R_0 \sqrt{n}(\hat{\vartheta}_n - \hat{\vartheta}_n^c) \stackrel{o_P(1)}{=} R_0 J^{-1} \sqrt{n} \frac{\partial \tilde{\ell}_n(\hat{\vartheta}_n^c)}{\partial \vartheta} = R_0 J^{-1} R_0' \sqrt{n} \hat{\lambda}.$$

Thus under H_0 and the previous assumptions,

$$\sqrt{n} \hat{\lambda} \xrightarrow{\mathcal{L}} \mathcal{N} \left\{ 0, (R_0 J^{-1} R_0')^{-1} R_0 \Omega R_0' (R_0 J^{-1} R_0')^{-1} \right\}, \quad (2.7)$$

so that the LM statistic is defined by

$$\begin{aligned} \mathbf{LM}_n &= n \hat{\lambda}' \left\{ (R_0 \hat{J}^{-1} R_0')^{-1} R_0 \hat{\Omega} R_0' (R_0 \hat{J}^{-1} R_0')^{-1} \right\}^{-1} \hat{\lambda} \\ &= n \frac{\partial \tilde{\ell}_n}{\partial \vartheta'}(\hat{\vartheta}_n^c) \hat{J}^{-1} R_0' \left(R_0 \hat{\Omega} R_0' \right)^{-1} R_0 \hat{J}^{-1} \frac{\partial \tilde{\ell}_n}{\partial \vartheta}(\hat{\vartheta}_n^c). \end{aligned}$$

Note that in the strong VARMA case, $\hat{\Omega} = 2\hat{J}^{-1}$ and the LM statistic takes the more conventional form $\mathbf{LM}_n^* = (n/2) \hat{\lambda}' R_0 \hat{J}^{-1} R_0' \hat{\lambda}$. In the general case, strong and weak as well, the convergence (2.7) implies that the asymptotic distribution of the \mathbf{LM}_n statistic is $\chi_{s_0}^2$ under H_0 . The null is therefore rejected when $\mathbf{LM}_n > \chi_{s_0}^2(1 - \alpha)$. Of course the conventional LM test with rejection region $\mathbf{LM}_n^* > \chi_{s_0}^2(1 - \alpha)$ is not asymptotically valid for general weak VARMA models.

Standard Taylor expansions show that

$$\sqrt{n}(\hat{\vartheta}_n - \hat{\vartheta}_n^c) \stackrel{o_P(1)}{=} -\sqrt{n} J^{-1} R_0' \hat{\lambda},$$

and that the LR statistic satisfies

$$\mathbf{LR}_n := 2 \left\{ \log \tilde{L}_n(\hat{\vartheta}_n) - \log \tilde{L}_n(\hat{\vartheta}_n^c) \right\} \stackrel{o_P(1)}{=} \frac{n}{2} (\hat{\vartheta}_n - \hat{\vartheta}_n^c)' J (\hat{\vartheta}_n - \hat{\vartheta}_n^c) \stackrel{o_P(1)}{=} \mathbf{LM}_n^*.$$

Using the previous computations and standard results on quadratic forms of normal vectors (see *e.g.* Lemma 17.1 in van der Vaart, 1998), we find that the \mathbf{LR}_n statistic is asymptotically distributed as $\sum_{i=1}^{s_0} \lambda_i Z_i^2$ where the Z_i 's are iid $\mathcal{N}(0, 1)$ and $\lambda_1, \dots, \lambda_{s_0}$ are the eigenvalues of

$$\Sigma_{\mathbf{LR}} = J^{-1/2} S_{\mathbf{LR}} J^{-1/2}, \quad S_{\mathbf{LR}} = \frac{1}{2} R_0' (R_0 J^{-1} R_0')^{-1} R_0 \Omega R_0' (R_0 J^{-1} R_0')^{-1} R_0.$$

Note that when $\Omega = 2J^{-1}$, the matrix $\Sigma_{\mathbf{LR}} = J^{-1/2}R'_0(R_0J^{-1}R'_0)^{-1}R_0J^{-1/2}$ is a projection matrix. Its eigenvalues are therefore equal to 0 and 1, and the number of eigenvalues equal to 1 is $\text{Tr } J^{-1/2}R'_0(R_0J^{-1}R'_0)^{-1}R_0J^{-1/2} = \text{Tr } I_{s_0} = s_0$. Therefore we retrieve the well-known result that $\mathbf{LR}_n \sim \chi_{s_0}^2$ under H_0 in the strong VARMA case. In the weak VARMA case, the asymptotic null distribution of \mathbf{LR}_n is complicated. It is possible to evaluate the distribution of a quadratic form of a Gaussian vector by means of the Imhof algorithm (Imhof, 1961), but the algorithm is time consuming. An alternative is to use the transformed statistic

$$\frac{n}{2}(\hat{\vartheta}_n - \hat{\vartheta}_n^c)' \hat{J} \hat{S}_{\mathbf{LR}}^- \hat{J} (\hat{\vartheta}_n - \hat{\vartheta}_n^c) \quad (2.8)$$

which follows a $\chi_{s_0}^2$ under H_0 , when \hat{J} and $\hat{S}_{\mathbf{LR}}^-$ are weakly consistent estimators of J and of a generalized inverse of $S_{\mathbf{LR}}$. The estimator $\hat{S}_{\mathbf{LR}}^-$ can be obtained from the singular value decomposition of any weakly consistent estimator $\hat{S}_{\mathbf{LR}}$ of $S_{\mathbf{LR}}$. More precisely, defining the diagonal matrix $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_{k_0})$ where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_{k_0}$ denote the eigenvalues of the symmetric matrix $\hat{S}_{\mathbf{LR}}$, and denoting by \hat{P} an orthonormal matrix such that $\hat{S}_{\mathbf{LR}} = \hat{P} \hat{\Lambda} \hat{P}'$, one can set

$$\hat{S}_{\mathbf{LR}}^- = \hat{P} \hat{\Lambda}^- \hat{P}', \quad \hat{\Lambda}^- = \text{diag}(\hat{\lambda}_1^{-1}, \dots, \hat{\lambda}_{s_0}^{-1}, 0, \dots, 0).$$

The matrix $\hat{S}_{\mathbf{LR}}^-$ then converges weakly to a matrix $S_{\mathbf{LR}}^-$ satisfying $S_{\mathbf{LR}} S_{\mathbf{LR}}^- S_{\mathbf{LR}} = S_{\mathbf{LR}}$, because $S_{\mathbf{LR}}$ has full rank s_0 .

2.6 Numerical illustrations

We first study numerically the behaviour of the QMLE for strong and weak VARMA models of the form

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & a_1(2, 2) \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ b_1(2, 1) & b_1(2, 2) \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix}, \quad (2.9)$$

where

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \sim \text{IID } \mathcal{N}(0, I_2), \quad (2.10)$$

in the strong case, and

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \eta_{1,t} (|\eta_{1,t-1}| + 1)^{-1} \\ \eta_{2,t} (|\eta_{2,t-1}| + 1)^{-1} \end{pmatrix}, \quad \text{with } \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix} \sim \text{IID } \mathcal{N}(0, I_2), \quad (2.11)$$

in the weak case. Model (2.9) is a VARMA(1,1) model in echelon form. The noise defined by (2.11) is a direct extension of a weak noise defined by Romano and Thombs (1996) in the univariate case. The numerical illustrations of this section are made with the free statistical software R (see <http://cran.r-project.org/>). We simulated $N = 1,000$

independent trajectories of size $n = 2,000$ of Model (2.9), first with the strong Gaussian noise (2.10), second with the weak noise (2.11). Figure 2.1 compares the distribution of the QMLE in the strong and weak noise cases. The distributions of $\hat{a}_1(2, 2)$ and $\hat{b}_1(2, 1)$ are similar in the two cases, whereas the QMLE of $\hat{b}_1(2, 2)$ is more accurate in the weak case than in the strong one. Similar simulation experiments, not reported here, reveal that the situation is opposite, that is the QMLE is more accurate in the strong case than in the weak case, when the weak noise is defined by $\epsilon_{i,t} = \eta_{i,t}\eta_{i,t-1}$ for $i = 1, 2$. This is in accordance with the results of Romano and Thombs (1996) who showed that, with similar noises, the asymptotic variance of the sample autocorrelations can be greater or less than 1 as well (1 is the asymptotic variance for strong white noises).

Figure 2.2 compares the standard estimator $\hat{\Omega} = 2\hat{J}^{-1}$ and the sandwich estimator $\hat{\hat{\Omega}} = \hat{J}^{-1}\hat{I}\hat{J}^{-1}$ of the QMLE asymptotic variance Ω . We used the spectral estimator $\hat{I} = \hat{I}^{\text{SP}}$ defined in Theorem 2.4, and the AR order r is automatically selected by AIC, using the function `VARselect()` of the `vars` R package. In the strong VARMA case we know that the two estimators are consistent. In view of the two top panels of Figure 2.2, it seems that the sandwich estimator is less accurate in the strong case. This is not surprising because the sandwich estimator is more robust, in the sense that this estimator continues to be consistent in the weak VARMA case, contrary to the standard estimator. It is clear that in the weak case $n\text{Var}\left\{\hat{b}_1(2, 2) - b_1(2, 2)\right\}^2$ is better estimated by $\hat{\hat{\Omega}}^{\text{SP}}(3, 3)$ (see the box-plot (c) of the right-bottom panel of Figure 2.2) than by $2\hat{J}^{-1}(3, 3)$ (box-plot (c) of the left-bottom panel). The failure of the standard estimator of Ω in the weak VARMA framework may have important consequences in terms of identification or hypothesis testing.

Table 2.1 displays the empirical sizes of the standard Wald, LM and LR tests, and that of the modified versions proposed in Section 2.5. For the nominal level $\alpha = 5\%$, the empirical size over the $N = 1,000$ independent replications should vary between the significant limits 3.6% and 6.4% with probability 95%. For the nominal level $\alpha = 1\%$, the significant limits are 0.3% and 1.7%, and for the nominal level $\alpha = 10\%$, they are 8.1% and 11.9%. When the relative rejection frequencies are outside the significant limits, they are displayed in bold type in Table 2.1. For the strong VARMA model I, all the relative rejection frequencies are inside the significant limits. For the weak VARMA model II, the relative rejection frequencies of the standard tests are definitely outside the significant limits. Thus the error of first kind is well controlled by all the tests in the strong case, but only by modified versions of the tests in the weak case. Table 2.2 shows that the powers of all the tests are very similar in the Model III case. The same is also true for the two modified tests in the Model IV case. The empirical powers of the standard tests are hardly interpretable for Model IV, because we have already seen in Table 2.1 that the standard versions of the tests do not well control the error of first kind in the weak VARMA framework.

From these simulation experiments and from the asymptotic theory, we draw the conclusion that the standard methodology, based on the QMLE, allows to fit VARMA

representations of a wide class of nonlinear multivariate time series. This standard methodology, including in particular the significance tests on the parameters, needs however to be adapted to take into account the possible lack of independence of the errors terms. In future works, we intent to study how the existing identification (see *e.g.* Nsiri and Roy, 1996) and diagnostic checking (see *e.g.* Duchesne and Roy, 2004) procedures should be adapted in the weak VARMA framework considered in the present paper.

TABLE 2.1 – Empirical size of standard and modified tests : relative frequencies (in %) of rejection of $H_0 : b_1(2, 2) = 0$. The number of replications is $N = 1000$.

Model	Length n	Level	Standard Test			Modified Test		
			Wald	LM	LR	Wald	LM	LR
I	$n = 500$	$\alpha = 1\%$	1.1	0.7	0.8	1.7	0.7	1.7
		$\alpha = 5\%$	5.0	4.5	5.1	6.0	5.2	6.0
		$\alpha = 10\%$	8.9	9.3	9.4	11.0	9.9	10.9
I	$n = 2,000$	$\alpha = 1\%$	0.7	0.8	0.7	1.0	0.6	1.0
		$\alpha = 5\%$	5.0	4.3	4.6	5.5	5.1	5.5
		$\alpha = 10\%$	9.2	8.6	8.8	10.0	9.0	10.2
II	$n = 500$	$\alpha = 1\%$	0.0	0.0	0.0	1.4	1.4	1.3
		$\alpha = 5\%$	0.6	0.5	0.6	6.2	6.5	6.1
		$\alpha = 10\%$	2.3	2.2	2.2	12.0	11.2	12.0
II	$n = 2,000$	$\alpha = 1\%$	0.0	0.0	0.0	0.9	0.7	0.9
		$\alpha = 5\%$	0.4	0.3	0.3	4.6	4.3	4.6
		$\alpha = 10\%$	1.3	1.3	1.3	9.2	9.8	9.2

I : Strong VARMA(1,1) model (2.9)-(2.10) with $\vartheta_0 = (0.95, 2, 0)$

II : Weak VARMA(1,1) model (2.9)-(2.11) with $\vartheta_0 = (0.95, 2, 0)$

TABLE 2.2 – Empirical power of standard and modified tests : relative frequencies (in %) of rejection of $H_0 : b_1(2, 2) = 0$. The number of replications is $N = 1000$.

Model	Length n	Level	Standard Test			Modified Test		
			Wald	LM	LR	Wald	LM	LR
III	$n = 500$	$\alpha = 1\%$	6.8	5.9	6.6	8.0	6.5	7.9
		$\alpha = 5\%$	20.5	19.4	20.4	21.6	20.1	21.7
		$\alpha = 10\%$	29.5	29.0	29.4	30.6	29.5	30.6
IV	$n = 500$	$\alpha = 1\%$	1.7	1.8	1.7	15.5	14.3	15.6
		$\alpha = 5\%$	11.4	9.4	10.1	35.1	34.0	35.0
		$\alpha = 10\%$	21.1	20.2	20.6	47.1	44.9	46.8

III : Strong VARMA(1,1) model (2.9)-(2.10) with $\vartheta_0 = (0.95, 2, 0.05)$

IV : Weak VARMA(1,1) model (2.9)-(2.11) with $\vartheta_0 = (0.95, 2, 0.05)$

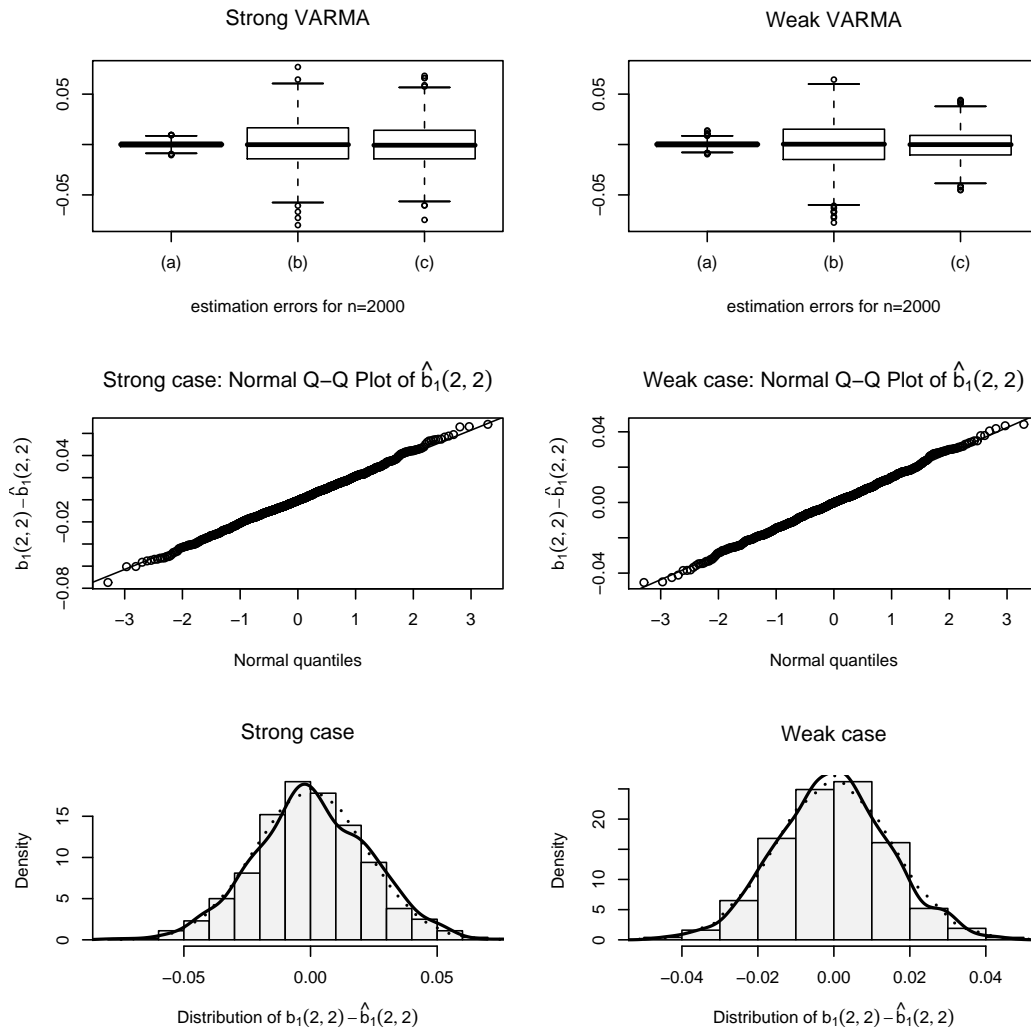


FIGURE 2.1 – QMLE of $N = 1,000$ independent simulations of the VARMA(1,1) model (2.9) with size $n = 2,000$ and unknown parameter $\vartheta_0 = ((a_1(2, 2), b_1(2, 1), b_1(2, 2)) = (0.95, 2, 0)$, when the noise is strong (left panels) and when the noise is the weak noise (2.11) (right panels). Points (a)-(c), in the box-plots of the top panels, display the distribution of the estimation errors $\hat{\vartheta}(i) - \vartheta_0(i)$ for $i = 1, 2, 3$. The panels of the middle present the Q-Q plot of the estimates $\hat{\vartheta}(3) = \hat{b}_1(2, 2)$ of the last parameter. The bottom panels display the distribution of the same estimates. The kernel density estimate is displayed in full line, and the centered Gaussian density with the same variance is plotted in dotted line.

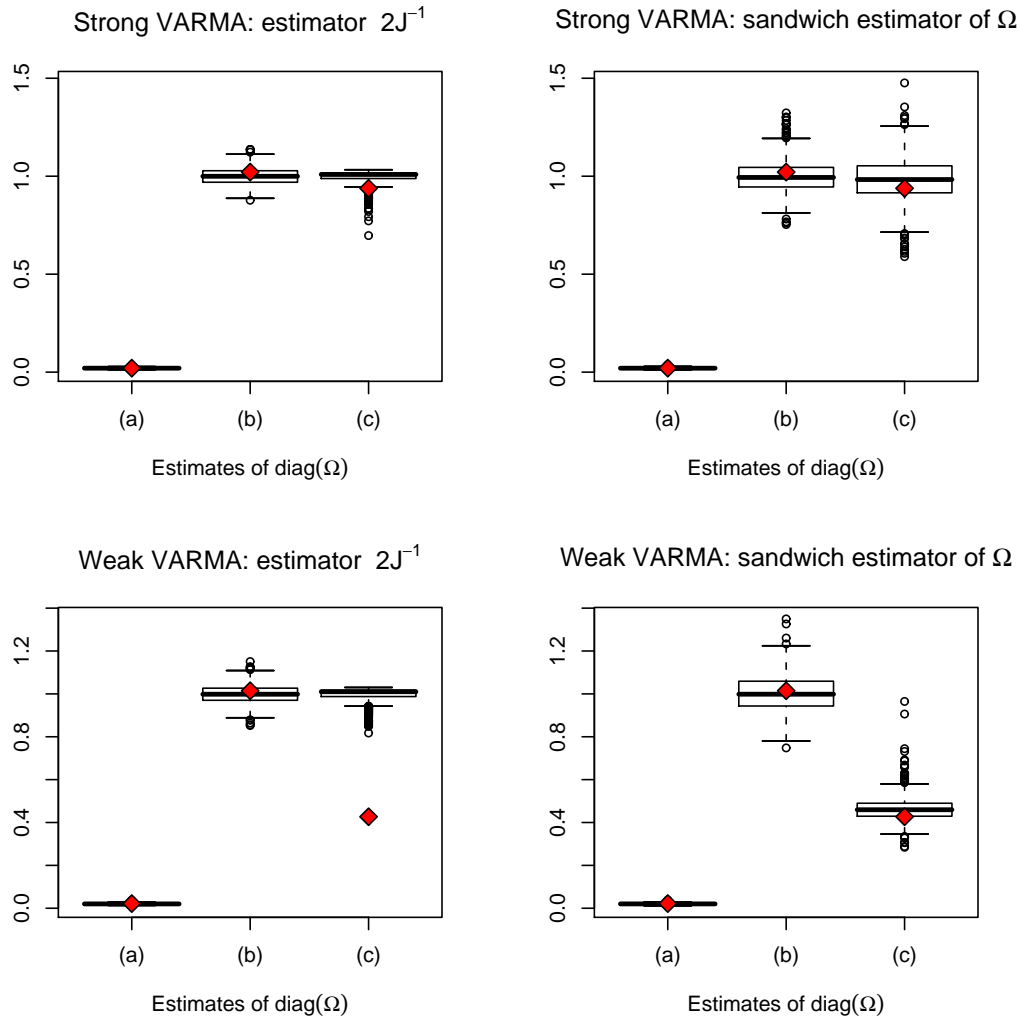


FIGURE 2.2 – Comparison of standard and modified estimates of the asymptotic variance Ω of the QMLE, on the simulated models presented in Figure 2.1. The diamond symbols represent the mean, over the $N = 1,000$ replications, of the standardized squared errors $n \{ \hat{a}_1(2, 2) - 0.95 \}^2$ for (a) (0.02 in the strong and weak cases), $n \{ \hat{b}_1(2, 1) - 2 \}^2$ for (b) (1.02 in the strong case and 1.01 in the strong case) and $n \{ \hat{b}_1(2, 2) \}^2$ for (c) (0.94 in the strong case and 0.43 in the weak case).

2.7 Appendix of technical proofs

We begin with a lemma useful to show the identifiability of ϑ_0 .

Lemma 1. Assume that Σ_0 is non singular and that **A5** holds true. If $A_0^{-1}B_0B_{\vartheta}^{-1}(L)A_{\vartheta}(L)X_t = A_{00}^{-1}B_{00}\epsilon_t$ with probability one and $A_0^{-1}B_0\Sigma B_0'A_0^{-1'} = A_{00}^{-1}B_{00}\Sigma_0B_{00}'A_{00}^{-1'}$, then $\vartheta = \vartheta_0$.

Proof : Let $\vartheta \neq \vartheta_0$. Assumption **A5** implies that either $A_0^{-1}B_0\Sigma B_0'A_0^{-1'} \neq A_{00}^{-1}B_{00}\Sigma_0B_{00}'A_{00}^{-1'}$ or there exist matrices C_i such that $C_{i_0} \neq 0$ for some $i_0 > 0$ and

$$A_0^{-1}B_0B_{\vartheta}^{-1}(z)A_{\vartheta}(z) - A_{00}^{-1}B_{00}B_{\vartheta_0}^{-1}(z)A_{\vartheta_0}(z) = \sum_{i=i_0}^{\infty} C_i z^i.$$

By contradiction, assume that $A_0^{-1}B_0B_{\vartheta}^{-1}(L)A_{\vartheta}(L)X_t = A_{00}^{-1}B_{00}\epsilon_t = A_{00}^{-1}B_{00}B_{\vartheta_0}^{-1}(L)A_{\vartheta_0}(L)X_t$ with probability one. This implies that there exists $\lambda \neq 0$ such that $\lambda'X_{t-i_0}$ is almost surely a linear combination of the components of X_{t-i} , $i > i_0$. By stationarity, it follows that $\lambda'X_t$ is almost surely a linear combination of the components of X_{t-i} , $i > 0$. Thus $\lambda'\epsilon_t = 0$ almost surely, which is impossible when the variance Σ_0 of ϵ_t is positive definite. \square

Proof of Theorem 2.1 : Note that, due to the initial conditions, $\{\tilde{e}_t(\vartheta)\}$ is not stationary, but can be approximated by the stationary ergodic process

$$e_t(\vartheta) = A_0^{-1}B_0B_{\vartheta}^{-1}(L)A_{\vartheta}(L)X_t. \quad (2.12)$$

From an extension of Lemma 1 in FZ, it is easy to show that $\sup_{\vartheta \in \Theta} \|\tilde{e}_t(\vartheta) - e_t(\vartheta)\| \rightarrow 0$ almost surely at an exponential rate, as $t \rightarrow \infty$. We thus have

$$\tilde{\ell}_n(\vartheta) \stackrel{op(1)}{=} \ell_n(\vartheta) := \frac{1}{n} \sum_{t=1}^n l_t(\vartheta) \quad \text{as } n \rightarrow \infty,$$

where

$$l_t(\vartheta) = d \log(2\pi) + \log \det \Sigma_e + e_t'(\vartheta) \Sigma_e^{-1} e_t(\vartheta).$$

Now the ergodic theorem shows that almost surely

$$\ell_n(\vartheta) \rightarrow d \log(2\pi) + Q(\vartheta),$$

where $Q(\vartheta) = \log \det \Sigma_e + E e_1'(\vartheta) \Sigma_e^{-1} e_1(\vartheta)$. We have

$$\begin{aligned} Q(\vartheta) &= E \{e_1(\vartheta_0)\}' \Sigma_e^{-1} \{e_1(\vartheta_0)\} + \log \det \Sigma_e \\ &\quad + E \{e_1(\vartheta) - e_1(\vartheta_0)\}' \Sigma_e^{-1} \{e_1(\vartheta) - e_1(\vartheta_0)\} \\ &\quad + 2E \{e_1(\vartheta) - e_1(\vartheta_0)\}' \Sigma_e^{-1} e_1(\vartheta_0). \end{aligned}$$

The last expectation is null because the linear innovation $e_t = e_t(\vartheta_0)$ is orthogonal to the linear past (*i.e.* to the Hilbert space H_{t-1} generated by linear combinations of the X_u for $u < t$), and because $\{e_t(\vartheta) - e_t(\vartheta_0)\}$ belongs to this linear past H_{t-1} . Moreover

$$\begin{aligned} Q(\vartheta_0) &= \log \det \Sigma_{e_0} + Ee'_1(\vartheta_0)\Sigma_{e_0}^{-1}e_1(\vartheta_0) \\ &= \log \det \Sigma_{e_0} + \text{Tr} \Sigma_{e_0}^{-1}Ee_1(\vartheta_0)e'_1(\vartheta_0) = \log \det \Sigma_{e_0} + d. \end{aligned}$$

Thus

$$Q(\vartheta) - Q(\vartheta_0) \geq \text{Tr} \Sigma_e^{-1}\Sigma_{e_0} - \log \det \Sigma_e^{-1}\Sigma_{e_0} - d \geq 0 \quad (2.13)$$

using the elementary inequality $\text{Tr}(A^{-1}B) - \log \det(A^{-1}B) \geq \text{Tr}(A^{-1}A) - \log \det(A^{-1}A) = d$ for all symmetric positive semi-definite matrices of order $d \times d$. At least one of the two inequalities in (2.13) is strict, unless if $e_1(\vartheta) = e_1(\vartheta_0)$ with probability 1 and $\Sigma_e = \Sigma_{e_0}$, which is equivalent to $\vartheta = \vartheta_0$ by Lemma 1. The rest of the proof relies on standard compactness arguments, as in Theorem 1 of FZ. \square

Proof of Theorem 2.2 : In view of Theorem 2.1 and A6, we have almost surely $\hat{\vartheta}_n \rightarrow \vartheta_0 \in \overset{\circ}{\Theta}$. Thus $\partial \tilde{\ell}_n(\hat{\vartheta}_n)/\partial \vartheta = 0$ for sufficiently large n , and a Taylor expansion gives

$$0 \stackrel{o_P(1)}{=} \sqrt{n} \frac{\partial \ell_n(\vartheta_0)}{\partial \vartheta} + \frac{\partial^2 \ell_n(\vartheta_0)}{\partial \vartheta \partial \vartheta'} \sqrt{n} (\hat{\vartheta}_n - \vartheta_0), \quad (2.14)$$

using arguments given in FZ (proof of Theorem 2). The proof then directly follows from Lemma 3 and Lemma 5 below. \square

We first state elementary derivative rules, which can be found in Appendix A.13 of Lütkepohl (1993).

Lemma 2. If $f(A)$ is a scalar function of a matrix A whose elements a_{ij} are function of a variable x , then

$$\frac{\partial f(A)}{\partial x} = \sum_{i,j} \frac{\partial f(A)}{\partial a_{ij}} \frac{\partial a_{ij}}{\partial x} = \text{Tr} \left\{ \frac{\partial f(A)}{\partial A'} \frac{\partial A}{\partial x} \right\}. \quad (2.15)$$

When A is invertible, we also have

$$\frac{\partial \log |\det(A)|}{\partial A'} = A^{-1} \quad (2.16)$$

$$\frac{\partial \text{Tr}(CA^{-1}B)}{\partial A'} = -A^{-1}BCA^{-1} \quad (2.17)$$

$$\frac{\partial \text{Tr}(CAB)}{\partial A'} = BC \quad (2.18)$$

Lemma 3. Under the assumptions of Theorem 2.2, almost surely

$$\frac{\partial^2 \ell_n(\vartheta_0)}{\partial \vartheta \partial \vartheta'} \rightarrow J,$$

where J is invertible.

Proof of Lemma 3 : Let $\vartheta = (\vartheta_1, \dots, \vartheta_{k_0})'$. In view of (2.15), (2.16) and (2.17), for all $i \in \{1, \dots, k_0\}$, we have

$$\frac{\partial l_t(\vartheta)}{\partial \vartheta_i} = \text{Tr} \left\{ \Sigma_e^{-1} \frac{\partial \Sigma_e}{\partial \vartheta_i} - \Sigma_e^{-1} e_t(\vartheta) e_t'(\vartheta) \Sigma_e^{-1} \frac{\partial \Sigma_e}{\partial \vartheta_i} \right\} + 2 \frac{\partial e_t'(\vartheta)}{\partial \vartheta_i} \Sigma_e^{-1} e_t(\vartheta). \quad (2.19)$$

Using the previous relations and (2.18), for all $i, j \in \{1, \dots, k_0\}$, we have

$$\begin{aligned} \frac{\partial^2 l_t(\vartheta)}{\partial \vartheta_i \partial \vartheta_j} &= \text{Tr} \left\{ \Sigma_e^{-1} \frac{\partial^2 \Sigma_e}{\partial \vartheta_i \partial \vartheta_j} - \Sigma_e^{-1} \frac{\partial \Sigma_e}{\partial \vartheta_i} \Sigma_e^{-1} \frac{\partial \Sigma_e}{\partial \vartheta_j} - \Sigma_e^{-1} e_t(\vartheta) e_t'(\vartheta) \Sigma_e^{-1} \frac{\partial^2 \Sigma_e}{\partial \vartheta_i \partial \vartheta_j} \right. \\ &\quad + \Sigma_e^{-1} \frac{\partial \Sigma_e}{\partial \vartheta_i} \Sigma_e^{-1} e_t(\vartheta) e_t'(\vartheta) \Sigma_e^{-1} \frac{\partial \Sigma_e}{\partial \vartheta_j} + \Sigma_e^{-1} e_t(\vartheta) e_t'(\vartheta) \Sigma_e^{-1} \frac{\partial \Sigma_e}{\partial \vartheta_i} \Sigma_e^{-1} \frac{\partial \Sigma_e}{\partial \vartheta_j} \\ &\quad \left. - \Sigma_e^{-1} \frac{\partial \Sigma_e}{\partial \vartheta_i} \Sigma_e^{-1} \frac{\partial e_t(\vartheta) e_t'(\vartheta)}{\partial \vartheta_j} \right\} + 2 \frac{\partial^2 e_t'(\vartheta)}{\partial \vartheta_i \partial \vartheta_j} \Sigma_e^{-1} e_t(\vartheta) \\ &\quad + 2 \frac{\partial e_t'(\vartheta)}{\partial \vartheta_i} \Sigma_e^{-1} \frac{\partial e_t(\vartheta)}{\partial \vartheta_j} - 2 \text{Tr} \left\{ \Sigma_e^{-1} e_t(\vartheta) \frac{\partial e_t'(\vartheta)}{\partial \vartheta_i} \Sigma_e^{-1} \frac{\partial \Sigma_e}{\partial \vartheta_j} \right\}. \end{aligned} \quad (2.20)$$

Using $E e_t e_t' = \Sigma_e$, $E e_t = 0$, the uncorrelatedness between e_t and the linear past H_{t-1} , $\partial e_t(\vartheta_0) / \partial \vartheta_i \in H_{t-1}$, and $\partial^2 e_t(\vartheta_0) / \partial \vartheta_i \partial \vartheta_j \in H_{t-1}$, we have

$$\begin{aligned} E \frac{\partial^2 l_t(\vartheta_0)}{\partial \vartheta_i \partial \vartheta_j} &= \text{Tr} \left\{ \Sigma_{e0}^{-1} \frac{\partial \Sigma_e(\vartheta_0)}{\partial \vartheta_i} \Sigma_{e0}^{-1} \frac{\partial \Sigma_e(\vartheta_0)}{\partial \vartheta_j} \right\} + 2E \frac{\partial e_t'(\vartheta_0)}{\partial \vartheta_i} \Sigma_{e0}^{-1} \frac{\partial e_t(\vartheta_0)}{\partial \vartheta_j} \\ &= J(i, j). \end{aligned} \quad (2.21)$$

The ergodic theorem and the next lemma conclude. \square

Lemma 4. Under the assumptions of Theorem 2.2, the matrix

$$J = E \frac{\partial^2 l_t(\vartheta_0)}{\partial \vartheta \partial \vartheta'}$$

is invertible.

Proof of Lemma 4 : In view of (2.21), we have $J = J_1 + J_2$, where

$$J_2 = 2E \frac{\partial e_t'(\vartheta_0)}{\partial \vartheta} \Sigma_{e0}^{-1} \frac{\partial e_t(\vartheta_0)}{\partial \vartheta'}$$

and

$$J_1(i, j) = \text{Tr} \left\{ \Sigma_{e0}^{-1/2} \frac{\partial \Sigma_e(\vartheta_0)}{\partial \vartheta_i} \Sigma_{e0}^{-1/2} \Sigma_{e0}^{-1/2} \frac{\partial \Sigma_e(\vartheta_0)}{\partial \vartheta_j} \Sigma_{e0}^{-1/2} \right\} = \mathbf{h}_i' \mathbf{h}_j,$$

with

$$\mathbf{h}_i = (\Sigma_{e0}^{-1/2} \otimes \Sigma_{e0}^{-1/2}) \mathbf{d}_i, \quad \mathbf{d}_i = \text{vec} \frac{\partial \Sigma_e(\vartheta_0)}{\partial \vartheta_i}.$$

In the previous derivations, we used the well-known relations $\text{Tr}(A'B) = (\text{vec}A)' \text{vec}B$ and $\text{vec}(ABC) = (C' \otimes A) \text{vec}B$. Note that the matrices J , J_1 and J_2 are semi-definite

positive. If J is singular, then there exists a vector $\mathbf{c} = (c_1, \dots, c_{k_0})' \neq 0$ such that $\mathbf{c}' J_1 \mathbf{c} = \mathbf{c}' J_2 \mathbf{c} = 0$. Since $\Sigma_{e_0}^{-1/2} \otimes \Sigma_{e_0}^{-1/2}$ and $\Sigma_{e_0}^{-1}$ are definite positive, we have $\mathbf{c}' J_1 \mathbf{c} = 0$ if and only if

$$\sum_{k=1}^{k_0} c_k \mathbf{d}_k = \sum_{k=1}^{k_0} c_k \text{vec} \frac{\partial \Sigma_e(\vartheta_0)}{\partial \vartheta_k} = 0 \quad (2.22)$$

and $\mathbf{c}' J_2 \mathbf{c} = 0$ if and only if $\sum_{k=1}^{k_0} c_k \frac{\partial e_t(\vartheta_0)}{\partial \vartheta_k} = 0$ a.s. Differentiating the two sides of the reduced form representation (2.3), the latter equation yields the VARMA($p-1, q-1$) equation $\sum_{i=1}^p A_i^* X_{t-i} = \sum_{j=1}^q B_j^* e_{t-j}$. The identifiability assumption **A5** excludes the existence of such a representation. Thus

$$A_i^* = \sum_{k=1}^{k_0} c_k \frac{\partial A_0^{-1} A_i}{\partial \vartheta_k}(\vartheta_0) = 0, \quad B_j^* = \sum_{k=1}^{k_0} c_k \frac{\partial A_0^{-1} B_j B_0^{-1} A_0}{\partial \vartheta_k}(\vartheta_0) = 0. \quad (2.23)$$

It can be seen that (2.22) and (2.23), for $i = 1, \dots, p$ and $j = 1, \dots, q$, are equivalent to $\dot{M}_{\vartheta_0} \mathbf{c} = 0$. We conclude from **A8**. \square

Lemma 5. Under the assumptions of Theorem 2.2,

$$\sqrt{n} \frac{\partial \ell_n(\vartheta_0)}{\partial \vartheta} \xrightarrow{\mathcal{L}} \mathcal{N}(0, I).$$

Proof of Lemma 5 : In view of (2.12), we have

$$\frac{\partial e_t(\vartheta_0)}{\partial \vartheta_i} = \sum_{\ell=1}^{\infty} d'_\ell e_{t-\ell}, \quad (2.24)$$

where the sequence of matrices $d_\ell = d_\ell(i)$ is such that $\|d_\ell\| \rightarrow 0$ at a geometric rate as $\ell \rightarrow \infty$. By (2.19), we have for all m

$$\begin{aligned} \frac{\partial \ell_t(\vartheta_0)}{\partial \vartheta_i} &= \text{Tr} \left[\Sigma_{e_0}^{-1} \{ I_d - e_t e'_t \Sigma_{e_0}^{-1} \} \frac{\partial \Sigma_e(\vartheta_0)}{\partial \vartheta_i} \right] + 2 \frac{\partial e'_t(\vartheta_0)}{\partial \vartheta_i} \Sigma_{e_0}^{-1} e_t \\ &= Y_{t,m,i} + Z_{t,m,i} \end{aligned}$$

where

$$\begin{aligned} Y_{t,m,i} &= \text{Tr} \left[\Sigma_{e_0}^{-1} \{ I_d - e_t e'_t \Sigma_{e_0}^{-1} \} \frac{\partial \Sigma_e(\vartheta_0)}{\partial \vartheta_i} \right] + 2 \sum_{\ell=1}^m e'_{t-\ell} d'_\ell \Sigma_{e_0}^{-1} e_t \\ Z_{t,m,i} &= 2 \sum_{\ell=m+1}^{\infty} e'_{t-\ell} d'_\ell \Sigma_{e_0}^{-1} e_t. \end{aligned}$$

Let $Y_{t,m} = (Y_{t,m,1}, \dots, Y_{t,m,k_0})'$ and $Z_{t,m} = (Z_{t,m,1}, \dots, Z_{t,m,k_0})'$. The processes $(Y_{t,m})_t$ and $(Z_{t,m})_t$ are stationary and centered. Moreover, under Assumption **A7** and m

fixed, the process $Y = (Y_{t,m})_t$ is strongly mixing, with mixing coefficients $\alpha_Y(h) \leq \alpha_\varepsilon(\max\{0, h - m\})$. Applying the central limit theorem (CLT) for mixing processes (see Herrndorf, 1984) we directly obtain

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n Y_{t,m} \xrightarrow{\mathcal{L}} \mathcal{N}(0, I_m), \quad I_m = \sum_{h=-\infty}^{\infty} \text{Cov}(Y_{t,m}, Y_{t-h,m}).$$

As in FZ Lemma 3, one can show that $I = \lim_{m \rightarrow \infty} I_m$ exists. Since $\|Z_{t,m}\|_2 \rightarrow 0$ at an exponential rate when $m \rightarrow \infty$, using the arguments given in FZ Lemma 4, one can show that

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left\{ \left\| n^{-1/2} \sum_{t=1}^n Z_{t,m} \right\| > \varepsilon \right\} = 0 \quad (2.25)$$

for every $\varepsilon > 0$. From a standard result (see *e.g.* Brockwell and Davis, 1991, Proposition 6.3.9), we deduce that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial l_t(\vartheta_0)}{\partial \vartheta} = \frac{1}{\sqrt{n}} \sum_{t=1}^n Y_{t,m} + \frac{1}{\sqrt{n}} \sum_{t=1}^n Z_{t,m} \xrightarrow{\mathcal{L}} \mathcal{N}(0, I),$$

which completes the proof. \square

Proof of Theorem 2.3 : Note that

$$\tilde{\ell}_n(\vartheta) = \frac{1}{n} \sum_{t=1}^n \tilde{l}_t(\vartheta), \quad \tilde{l}_t(\vartheta) = d \log(2\pi) + \log \det \Sigma_e + \tilde{e}'_t(\vartheta) \Sigma_e^{-1} \tilde{e}_t(\vartheta).$$

Under the assumption of the theorem, $\partial \tilde{e}'_t(\vartheta) / \partial \vartheta^{(2)} = 0$, and (2.19) yields

$$\frac{\partial \tilde{l}_t(\hat{\vartheta}_n)}{\partial \vartheta_i} = \text{Tr} \left[\hat{\Sigma}_e^{-1} \left\{ I_d - \tilde{e}_t(\hat{\vartheta}_n^{(1)}) \tilde{e}'_t(\hat{\vartheta}_n^{(1)}) \hat{\Sigma}_e^{-1} \right\} \frac{\partial \Sigma_e(\hat{\vartheta}_n)}{\partial \vartheta_i} \right]$$

for $i = k_1 + 1, \dots, k_0$, with $\hat{\Sigma}_e$ such that $\hat{\vartheta}_n^{(2)} = D \text{vec } \hat{\Sigma}_e$. Assumption **A6** entails that the first order condition $\partial \tilde{\ell}_n(\hat{\vartheta}_n) / \partial \vartheta^{(2)} = 0$ is satisfied for n large enough. We then have

$$\hat{\Sigma}_e = n^{-1} \sum_{t=1}^n \tilde{e}_t(\hat{\vartheta}_n^{(1)}) \tilde{e}'_t(\hat{\vartheta}_n^{(1)})$$

and

$$\tilde{\ell}_n(\hat{\vartheta}_n) = d \log(2\pi) + \log \det \hat{\Sigma}_e + d,$$

because

$$\frac{1}{n} \sum_{t=1}^n \tilde{e}'_t(\hat{\vartheta}_n^{(1)}) \hat{\Sigma}_e^{-1} \tilde{e}_t(\hat{\vartheta}_n^{(1)}) = \text{Tr} \left[\frac{1}{n} \sum_{t=1}^n \tilde{e}_t(\hat{\vartheta}_n^{(1)}) \tilde{e}'_t(\hat{\vartheta}_n^{(1)}) \hat{\Sigma}_e^{-1} \right] = d.$$

The conclusion follows. \square

2.8 Bibliography

- Andrews, D.W.K.** (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- Berk, K.N.** (1974) Consistent Autoregressive Spectral Estimates. *Annals of Statistics*, 2, 489–502.
- Brockwell, P.J. and Davis, R.A.** (1991) *Time series : theory and methods*. Springer Verlag, New York.
- den Hann, W.J. and Levin, A.** (1997) A Practitioner’s Guide to Robust Covariance Matrix Estimation. In *Handbook of Statistics* 15, Rao, C.R. and G.S. Madhala (eds), 291–341.
- Duchesne, P. and Roy, R.** (2004) On consistent testing for serial correlation of unknown form in vector time series models. *Journal of Multivariate Analysis* 89, 148–180.
- Dufour, J-M., and Pelletier, D.** (2005) Practical methods for modelling weak VARMA processes : identification, estimation and specification with a macroeconomic application. *Technical report, Département de sciences économiques and CIREQ, Université de Montréal, Montréal, Canada.*
- Dunsmuir, W.T.M. and Hannan, E.J.** (1976) Vector linear time series models, *Advances in Applied Probability* 8, 339–364.
- Fan, J. and Yao, Q.** (2003) *Nonlinear time series : Nonparametric and parametric methods*. Springer Verlag, New York.
- Francq, C. and Raïssi, H.** (2007) Multivariate Portmanteau Test for Autoregressive Models with Uncorrelated but Nonindependent Errors, *Journal of Time Series Analysis* 28, 454–470.
- Francq, C., Roy, R. and Zakoïan, J-M.** (2005) Diagnostic checking in ARMA Models with Uncorrelated Errors, *Journal of the American Statistical Association* 100, 532–544.
- Francq, C. and Zakoïan, J-M.** (1998) Estimating linear representations of nonlinear processes, *Journal of Statistical Planning and Inference* 68, 145–165.
- Francq, and Zakoïan, J-M.** (2005) Recent results for linear time series models with non independent innovations. In *Statistical Modeling and Analysis for Complex Data Problems*, Chap. 12 (eds P. DUCHESNE and B. RÉMILLARD). New York : Springer Verlag, 137–161.
- Francq, and Zakoïan, J-M.** (2007) HAC estimation and strong linearity testing in weak ARMA models, *Journal of Multivariate Analysis* 98, 114–144.
- Hannan, E.J.** (1976) The identification and parametrization of ARMAX and state space forms, *Econometrica* 44, 713–723.
- Hannan, E.J. and Deistler, M.** (1988) *The Statistical Theory of Linear Systems* John Wiley, New York.

- Hannan, E.J., Dunsmuir, W.T.M. and Deistler, M.** (1980) Estimation of vector ARMAX models, *Journal of Multivariate Analysis* 10, 275–295.
- Hannan, E.J. and Rissanen** (1982) Recursive estimation of mixed of Autoregressive Moving Average order, *Biometrika* 69, 81–94.
- Herrndorf, N.** (1984) A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables. *The Annals of Probability* 12, 141–153.
- Imhof, J.P.** (1961) Computing the distribution of quadratic forms in normal variables. *Biometrika* 48, 419–426.
- Kascha, C.** (2007) A Comparison of Estimation Methods for Vector Autoregressive Moving-Average Models. *ECO Working Papers, EUI ECO 2007/12*, <http://hdl.handle.net/1814/6921>.
- Lütkepohl, H.** (1993) *Introduction to multiple time series analysis*. Springer Verlag, Berlin.
- Lütkepohl, H.** (2005) *New introduction to multiple time series analysis*. Springer Verlag, Berlin.
- Newey, W.K., and West, K.D.** (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- Nsiri, S., and Roy, R.** (1996) Identification of refined ARMA echelon form models for multivariate time series. *Journal of Multivariate Analysis* 56, 207–231.
- Reinsel, G.C.** (1997) *Elements of multivariate time series Analysis*. Second edition. Springer Verlag, New York.
- Rissanen, J. and Caines, P.E.** (1979) The strong consistency of maximum likelihood estimators for ARMA processes. *Annals of Statistics* 7, 297–315.
- Romano, J.L. and Thombs, L.A.** (1996) Inference for autocorrelations under weak assumptions, *Journal of the American Statistical Association* 91, 590–600.
- Tong, H.** (1990) *Non-linear time series : A Dynamical System Approach*, Clarendon Press Oxford.
- van der Vaart, A.W.** (1998) *Asymptotic statistics*, Cambridge University Press, Cambridge.
- Wald, A.** (1949) Note on the consistency of the maximum likelihood estimate. *Annals of Mathematical Statistics* 20, 595–601.

2.9 Appendix of additional example

Example 2.2. Denoting by $a_{0i}(k, \ell)$ and $b_{0i}(k, \ell)$ the generic elements of the matrices A_{0i} and B_{0i} , the Kronecker indices are defined by $p_k = \max\{i : a_{0i}(k, \ell) \neq 0 \text{ or } b_{0i}(k, \ell) \neq 0 \text{ for some } \ell = 1, \dots, d\}$. To ensure relatively parsimonious parameterizations, one can specify an echelon form depending on the Kronecker indices (p_1, \dots, p_d) . The reader is referred to Lütkepohl (1993) for details about the echelon form. For instance, a 3-variate ARMA process with Kronecker indices $(1, 2, 0)$ admits the echelon form

$$\begin{aligned} & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \times & \times & 1 \end{pmatrix} X_t - \begin{pmatrix} \times & \times & 0 \\ 0 & \times & 0 \\ 0 & 0 & 0 \end{pmatrix} X_{t-1} - \begin{pmatrix} 0 & 0 & 0 \\ \times & \times & 0 \\ 0 & 0 & 0 \end{pmatrix} X_{t-2} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \times & \times & 1 \end{pmatrix} \epsilon_t - \begin{pmatrix} \times & \times & \times \\ \times & \times & \times \\ 0 & 0 & 0 \end{pmatrix} \epsilon_{t-1} - \begin{pmatrix} 0 & 0 & 0 \\ \times & \times & \times \\ 0 & 0 & 0 \end{pmatrix} \epsilon_{t-2} \end{aligned}$$

where \times denotes an unconstrained element. The variance of ϵ_t is defined by 6 additional parameters. This echelon form thus corresponds to a parametrization by a vector ϑ of size $k_0 = 24$.

2.9.1 Verification of Assumption A8 on Example 2.1

In this example, we have

$$M_{\vartheta_0} = \begin{pmatrix} \alpha_{01} & \alpha_{02} & \sigma_{01}^2 & \sigma_{01}^2 \alpha_{03} \\ \alpha_{01} \alpha_{03} + \alpha_{04} & \alpha_{02} \alpha_{03} + \alpha_{05} & \sigma_{01}^2 \alpha_{03} & \sigma_{01}^2 \alpha_{03}^2 + \sigma_{02}^2 \end{pmatrix}.$$

Thus

$$\dot{M}_{\vartheta_0} = \begin{pmatrix} 1 & \alpha_{03} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \alpha_{03} & 0 & 0 & 0 & 0 \\ 0 & \alpha_{01} & 0 & \alpha_{02} & 0 & \sigma_{01}^2 & \sigma_{01}^2 & 2\alpha_{03}\sigma_{01}^2 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \alpha_{03} & \alpha_{03} & \alpha_{03}^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

is of full rank $k_0 = 7$.

2.9.2 Details on the proof of Theorem 2.1

Lemma 6. Under the assumptions of Theorem 2.1, we have

$$\sup_{\vartheta \in \Theta} \|\tilde{e}_t(\vartheta) - e_t(\vartheta)\| \leq K\rho^t,$$

where ρ is a constant belonging to $[0, 1)$, and $K > 0$ is measurable with respect to the σ -field generated by $\{X_u, u \leq 0\}$.

Proof of Lemma 6 : We have

$$e_t(\vartheta) = X_t - \sum_{i=1}^p A_0^{-1} A_i X_{t-i} + \sum_{i=1}^q A_0^{-1} B_i B_0^{-1} A_0 e_{t-i}(\vartheta) \quad \forall t \in \mathbb{Z}, \quad (2.26)$$

and

$$\tilde{e}_t(\vartheta) = X_t - \sum_{i=1}^p A_0^{-1} A_i X_{t-i} + \sum_{i=1}^q A_0^{-1} B_i B_0^{-1} A_0 \tilde{e}_{t-i}(\vartheta) \quad t = 1, \dots, n \quad (2.27)$$

with the initial values $\tilde{e}_0(\vartheta) = \dots = \tilde{e}_{1-q}(\vartheta) = X_0 = \dots = X_{1-p} = 0$. Let

$$\underline{e}_t(\vartheta) = \begin{pmatrix} e_t(\vartheta) \\ e_{t-1}(\vartheta) \\ \vdots \\ e_{t-q+1}(\vartheta) \end{pmatrix}, \quad \underline{\tilde{e}}_t(\vartheta) = \begin{pmatrix} \tilde{e}_t(\vartheta) \\ \tilde{e}_{t-1}(\vartheta) \\ \vdots \\ \tilde{e}_{t-q+1}(\vartheta) \end{pmatrix}.$$

From (2.26) and (2.27), we have

$$\underline{e}_t(\vartheta) = \underline{b}_t + \mathbf{C} \underline{e}_{t-1}(\vartheta) \quad \forall t \in \mathbb{Z},$$

and

$$\underline{\tilde{e}}_t(\vartheta) = \underline{\tilde{b}}_t + \mathbf{C} \underline{\tilde{e}}_{t-1}(\vartheta) \quad t = 1, \dots, n,$$

where

$$\mathbf{C} = \begin{pmatrix} A_0^{-1} B_1 B_0^{-1} A_0 & \cdots & A_0^{-1} B_q B_0^{-1} A_0 \\ I_{(q-1)d} & & 0_{(q-1)d} \end{pmatrix}, \quad \underline{b}_t = \begin{pmatrix} X_t - \sum_{i=1}^p A_0^{-1} A_i X_{t-i} \\ 0_d \\ \vdots \\ 0_d \end{pmatrix},$$

$\underline{\tilde{b}}_t = \underline{b}_t$ for $t > p$, $\underline{\tilde{b}}_t = 0_{qd}$ for $t \leq 0$, and

$$\underline{\tilde{b}}_t = \begin{pmatrix} X_t - \sum_{i=1}^{t-1} A_0^{-1} A_i X_{t-i} \\ 0_d \\ \vdots \\ 0_d \end{pmatrix} \quad \text{for } t = 1, \dots, p.$$

Writing $\underline{d}_t(\vartheta) = \underline{e}_t(\vartheta) - \tilde{\underline{e}}_t(\vartheta)$, we obtain for $t > p$,

$$\begin{aligned} \underline{d}_t(\vartheta) &= \mathbf{C}\underline{d}_{t-1}(\vartheta) = \mathbf{C}^{t-p}\underline{d}_p(\vartheta) \\ &= \mathbf{C}^{t-p} \left\{ \left(\underline{b}_p - \tilde{\underline{b}}_p \right) + \mathbf{C} \left(\underline{b}_{p-1} - \tilde{\underline{b}}_{p-1} \right) + \cdots + \mathbf{C}^{p-1} \left(\underline{b}_1 - \tilde{\underline{b}}_1 \right) \right\} + \mathbf{C}^p \underline{b}_0. \end{aligned}$$

Note that \mathbf{C} is the companion matrix of the polynomial

$$\mathcal{P}(z) = I_d - \sum_{i=1}^q A_0^{-1} B_i B_0^{-1} A_0 z^i = A_0^{-1} B_\vartheta(z) B_0^{-1} A_0.$$

By **A2**, the zeroes of $\mathcal{P}(z)$ are of modulus strictly greater than one :

$$\mathcal{P}(z) = 0 \Rightarrow |z| > 1 \quad (2.28)$$

By a well-known result on companion matrices, (2.28) is equivalent to $\rho(\mathbf{C}) < 1$, where $\rho(\mathbf{C})$ denote the spectral radius of \mathbf{C} . By the compactness of Θ , we thus have

$$\sup_{\vartheta \in \Theta} \rho(\mathbf{C}) < 1.$$

We thus have

$$\sup_{\vartheta \in \Theta} \|\underline{d}_t(\vartheta)\| \leq K \rho^t,$$

where K and ρ are as in the statement of the lemma. The conclusion follows. \square

Lemma 7. Under the assumptions of Theorem 2.1, we have

$$\sup_{\vartheta \in \Theta} \left| \tilde{\ell}_n(\vartheta) - \ell_n(\vartheta) \right| = o(1)$$

almost surely.

Proof of Lemma 7 : We have

$$\tilde{\ell}_n(\vartheta) - \ell_n(\vartheta) = \frac{1}{n} \sum_{t=1}^n \{ \tilde{e}_t(\vartheta) - e_t(\vartheta) \}' \Sigma_e^{-1} \tilde{e}_t(\vartheta) + e_t(\vartheta)' \Sigma_e^{-1} \{ \tilde{e}_t(\vartheta) - e_t(\vartheta) \}.$$

In the proof of this lemma and in the rest on the paper, the letters K and ρ denote generic constants, whose values can be modified along the text, such that $K > 0$ and $0 < \rho < 1$. By Lemma 6,

$$\begin{aligned} \sup_{\vartheta \in \Theta} \left| \tilde{\ell}_n(\vartheta) - \ell_n(\vartheta) \right| &\leq \frac{K}{n} \sum_{t=1}^n \rho^t \left(\sup_{\vartheta \in \Theta} \|e_t(\vartheta)\| + \sup_{\vartheta \in \Theta} \|\tilde{e}_t(\vartheta)\| \right) \\ &\leq \frac{K}{n} \sum_{t=1}^n \rho^t \sup_{\vartheta \in \Theta} \|e_t(\vartheta)\| \end{aligned} \quad (2.29)$$

In view of (2.26), and using **A1** and the compactness of Θ , we have

$$e_t(\vartheta) = X_t + \sum_{i=1}^{\infty} C_i(\vartheta) X_{t-i}, \quad \sup_{\vartheta \in \Theta} \|C_i(\vartheta)\| \leq K \rho^i. \quad (2.30)$$

We thus have $E \sup_{\vartheta \in \Theta} \|e_t(\vartheta)\| < \infty$, and the Markov inequality entails

$$\sum_{t=1}^{\infty} P \left(\rho^t \sup_{\vartheta \in \Theta} \|e_t(\vartheta)\| > \varepsilon \right) \leq E \sup_{\vartheta \in \Theta} \|e_t(\vartheta)\| \sum_{t=1}^{\infty} \frac{\rho^t}{\varepsilon} < \infty.$$

By the Borel-Cantelli theorem, $\rho^t \sup_{\vartheta \in \Theta} \|e_t(\vartheta)\| \rightarrow 0$ almost surely as $t \rightarrow \infty$. The Cesàro theorem implies that the right-hand side of (2.29) converges to zero almost surely. \square

Lemma 8. Under the assumptions of Theorem 2.1, any $\vartheta \neq \vartheta_0$ has a neighborhood $V(\vartheta)$ such that

$$\liminf_{n \rightarrow \infty} \inf_{\vartheta^* \in V(\vartheta)} \tilde{\ell}_n(\vartheta^*) > El_1(\vartheta_0), \quad a.s. \quad (2.31)$$

Moreover for any neighborhood $V(\vartheta_0)$ of ϑ_0 we have

$$\limsup_{n \rightarrow \infty} \inf_{\vartheta^* \in V(\vartheta_0)} \tilde{\ell}_n(\vartheta^*) \leq El_1(\vartheta_0), \quad a.s. \quad (2.32)$$

Proof of Lemma 8 : For any $\vartheta \in \Theta$ and any positive integer k , let $V_k(\vartheta)$ be the open ball with center ϑ and radius $1/k$. Using Lemma 7, we have

$$\begin{aligned} \liminf_{n \rightarrow \infty} \inf_{\vartheta^* \in V_k(\vartheta) \cap \Theta} \tilde{\ell}_n(\vartheta^*) &\geq \liminf_{n \rightarrow \infty} \inf_{\vartheta^* \in V_k(\vartheta) \cap \Theta} \ell_n(\vartheta^*) - \limsup_{n \rightarrow \infty} \sup_{\vartheta \in \Theta} |\ell_n(\vartheta) - \tilde{\ell}_n(\vartheta)| \\ &\geq \liminf_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n \inf_{\vartheta^* \in V_k(\vartheta) \cap \Theta} l_t(\vartheta^*) \\ &= E \inf_{\vartheta^* \in V_k(\vartheta) \cap \Theta} l_1(\vartheta^*) \end{aligned}$$

For the last equality we applied the ergodic theorem to the ergodic stationary process $\{\inf_{\vartheta^* \in V_k(\vartheta) \cap \Theta} l_t(\vartheta^*)\}_t$. By the Beppo-Levi theorem, when k increases to ∞ , $E \inf_{\vartheta^* \in V_k(\vartheta) \cap \Theta} l_1(\vartheta^*)$ increases to $El_1(\vartheta)$. Because $El_1(\vartheta) = d \log(2\pi) + Q(\vartheta)$, the discussion which follows (2.13) entails $El_1(\vartheta) > El_1(\vartheta_0)$, and (2.31) follows.

To show (2.32), it suffices to remark that Lemma 7 and the ergodic theorem entail

$$\begin{aligned} \limsup_{n \rightarrow \infty} \inf_{\vartheta^* \in V(\vartheta) \cap \Theta} \tilde{\ell}_n(\vartheta^*) &\leq \limsup_{n \rightarrow \infty} \inf_{\vartheta^* \in V(\vartheta) \cap \Theta} \ell_n(\vartheta^*) + \limsup_{n \rightarrow \infty} \sup_{\vartheta \in \Theta} |\ell_n(\vartheta) - \tilde{\ell}_n(\vartheta)| \\ &\leq \limsup_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n l_t(\vartheta_0) \\ &= El_1(\vartheta_0). \end{aligned}$$

□

The proof of Theorem 2.1 is completed by the arguments of Wald (1949). More precisely, the compact set Θ is covered by a neighborhood $V(\vartheta_0)$ of ϑ_0 and a finite number of neighborhoods $V(\vartheta_1), \dots, V(\vartheta_k)$ satisfying (2.31) with ϑ replaced by ϑ_i , $i = 1, \dots, k$. In view of (2.31) and (2.32), we have almost surely

$$\inf_{\vartheta \in \Theta} \tilde{\ell}_n(\vartheta) = \min_{i=0,1,\dots,k} \inf_{\vartheta \in V(\vartheta_i) \cap \Theta} \tilde{\ell}_n(\vartheta) = \inf_{\vartheta \in V(\vartheta_0) \cap \Theta} \tilde{\ell}_n(\vartheta)$$

for n large enough. Since the neighborhood $V(\vartheta_0)$ can be chosen arbitrarily small, the conclusion follows.

2.9.3 Details on the proof of Theorem 2.2

Lemma 9. Under the assumptions of Theorem 2.2, we have

$$\sqrt{n} \sup_{\vartheta \in \Theta} \left\| \frac{\partial \tilde{\ell}_n(\vartheta)}{\partial \vartheta} - \frac{\partial \ell_n(\vartheta)}{\partial \vartheta} \right\| = o_P(1).$$

Proof of Lemma 9 : Similar to (2.30), Assumption A1 entails that, for $k = 1, \dots, k_0$,

$$\frac{\partial e_t(\vartheta)}{\partial \vartheta_k} = \sum_{i=1}^{\infty} C_i^{(k)}(\vartheta) X_{t-i}, \quad \frac{\partial \tilde{e}_t(\vartheta)}{\partial \vartheta_k} = \sum_{i=1}^{t-1} C_i^{(k)}(\vartheta) X_{t-i}, \quad (2.33)$$

$$C_i^{(k)}(\vartheta) = \frac{\partial C_i(\vartheta)}{\partial \vartheta_k}, \quad \sup_{\vartheta \in \Theta} \left\| C_i^{(k)}(\vartheta) \right\| \leq K \rho^i. \quad (2.34)$$

Similar to Lemma 6, we then have

$$\sup_{\vartheta \in \Theta} \left\| \frac{\partial \tilde{e}_t(\vartheta)}{\partial \vartheta'} - \frac{\partial e_t(\vartheta)}{\partial \vartheta'} \right\| \leq K \rho^t. \quad (2.35)$$

Using (2.19), we have

$$\sqrt{n} \left\{ \frac{\partial \tilde{\ell}_n(\vartheta)}{\partial \vartheta_k} - \frac{\partial \ell_n(\vartheta)}{\partial \vartheta_k} \right\} = a_1 + a_2,$$

with

$$\begin{aligned} a_1 &= \frac{2}{\sqrt{n}} \sum_{t=1}^n \left\{ \left(\frac{\partial e_t'(\vartheta)}{\partial \vartheta_k} - \frac{\partial \tilde{e}_t'(\vartheta)}{\partial \vartheta_k} \right) \Sigma_e^{-1} e_t(\vartheta) + \frac{\partial \tilde{e}_t'(\vartheta)}{\partial \vartheta_k} \Sigma_e^{-1} (e_t(\vartheta) - \tilde{e}_t(\vartheta)) \right\} \\ a_2 &= \text{Tr} \left(\Sigma_e^{-1} [\{\tilde{e}_t(\vartheta) - e_t(\vartheta)\} \tilde{e}_t'(\vartheta) + e_t(\vartheta) \{\tilde{e}_t(\vartheta) - e_t(\vartheta)\}'] \Sigma_e^{-1} \frac{\partial \Sigma_e}{\partial \vartheta_k} \right). \end{aligned}$$

From Lemma 6 and (2.35), it follows that

$$|a_1| + |a_2| \leq \frac{K}{\sqrt{n}} \sum_{t=1}^n b_t, \quad b_t = \rho^t \left\{ \sup_{\vartheta \in \Theta} \left\| \frac{\partial e_t(\vartheta)}{\partial \vartheta_k} \right\| + \sup_{\vartheta \in \Theta} \|e_t(\vartheta)\| + K\rho^t \right\}.$$

In view of (2.30) and (2.33)-(2.34), we have

$$E \sup_{\vartheta \in \Theta} \left\| \frac{\partial e_t(\vartheta)}{\partial \vartheta_k} \right\| < \infty, \quad E \sup_{\vartheta \in \Theta} \|e_t(\vartheta)\| < \infty.$$

By the Markov inequality, for all $\varepsilon > 0$ we thus have

$$P \left(\frac{K}{\sqrt{n}} \sum_{t=1}^n b_t > \varepsilon \right) \leq \frac{K}{\varepsilon \sqrt{n}} \sum_{t=1}^n \rho^t \rightarrow 0,$$

and the conclusion follows. \square

Lemma 10. Under the assumptions of Theorem 2.2, (2.14) holds, that is

$$0 \stackrel{o_P(1)}{=} \sqrt{n} \frac{\partial \ell_n(\vartheta_0)}{\partial \vartheta} + \frac{\partial^2 \ell_n(\vartheta_0)}{\partial \vartheta \partial \vartheta'} \sqrt{n} (\hat{\vartheta}_n - \vartheta_0).$$

Proof of Lemma 10 : By Lemma 9 and already given arguments, we have

$$0 = \sqrt{n} \frac{\tilde{\ell}_n(\hat{\vartheta}_n)}{\partial \vartheta} = \sqrt{n} \frac{\ell_n(\hat{\vartheta}_n)}{\partial \vartheta} + \sqrt{n} \frac{\tilde{\ell}_n(\hat{\vartheta}_n)}{\partial \vartheta} - \sqrt{n} \frac{\ell_n(\hat{\vartheta}_n)}{\partial \vartheta} \stackrel{o_P(1)}{=} \sqrt{n} \frac{\ell_n(\hat{\vartheta}_n)}{\partial \vartheta}.$$

Taylor expansions of the functions $\partial \ell_n(\cdot)/\partial \vartheta_i$ around ϑ_0 give

$$\sqrt{n} \frac{\ell_n(\hat{\vartheta}_n)}{\partial \vartheta} = \sqrt{n} \frac{\partial \ell_n(\vartheta_0)}{\partial \vartheta} + \left[\frac{\partial^2 \ell_n(\vartheta_i^*)}{\partial \vartheta_i \partial \vartheta_j} \right] \sqrt{n} (\hat{\vartheta}_n - \vartheta_0),$$

for some ϑ_i^* between $\hat{\vartheta}_n$ and ϑ_0 .

From (2.20) and Lemma 2, one can see that the random terms involved in

$$\frac{\partial^3 \ell_t(\vartheta)}{\partial \vartheta_{k_1} \partial \vartheta_{k_2} \partial \vartheta_{k_3}}$$

are linear combinations of the elements of the following matrices

$$\begin{aligned} & e_t(\vartheta) e_t'(\vartheta), \quad e_t(\vartheta) \frac{\partial e_t'(\vartheta)}{\partial \vartheta_{k_i}}, \quad e_t(\vartheta) \frac{\partial^2 e_t'(\vartheta)}{\partial \vartheta_{k_i} \partial \vartheta_{k_j}}, \\ & \frac{\partial e_t(\vartheta)}{\partial \vartheta_{k_i}} \frac{\partial e_t'(\vartheta)}{\partial \vartheta_{k_j}}, \quad \frac{\partial e_t(\vartheta)}{\partial \vartheta_{k_i}} \frac{\partial^2 e_t'(\vartheta)}{\partial \vartheta_{k_j} \partial \vartheta_{k_\ell}}, \quad e_t(\vartheta) \frac{\partial^3 e_t'(\vartheta)}{\partial \vartheta_{k_i} \partial \vartheta_{k_j} \partial \vartheta_{k_\ell}}. \end{aligned}$$

Since all these matrices have finite expectations uniformly in $\vartheta \in \Theta$, we have

$$E \sup_{\vartheta \in \Theta} \left| \frac{\partial^3 l_t(\vartheta)}{\partial \vartheta_{k_1} \partial \vartheta_{k_2} \partial \vartheta_{k_3}} \right| < \infty \quad (2.36)$$

for all $k_1, k_2, k_3 \in \{1, \dots, k_0\}$. Now a second Taylor expansion yields

$$\begin{aligned} \frac{\partial^2 \ell_n(\vartheta_i^*)}{\partial \vartheta_i \partial \vartheta_j} &= \frac{\partial^2 \ell_n(\vartheta_0)}{\partial \vartheta_i \partial \vartheta_j} + \frac{\partial}{\partial \vartheta'} \left\{ \frac{\partial^2 \ell_n(\vartheta^*)}{\partial \vartheta_i \partial \vartheta_j} \right\} (\vartheta_i^* - \vartheta_0) \\ &= \frac{\partial^2 \ell_n(\vartheta_0)}{\partial \vartheta_i \partial \vartheta_j} + o(1) \quad \text{a.s.} \end{aligned}$$

using the ergodic theorem, (2.36) and the almost sure convergence of ϑ_i^* to ϑ_0 . The proof is complete. \square

Now we need the following covariance inequality obtained by Davydov (1968).²

Lemma 11 (Davydov (1968)). Let p, q and r three positive numbers such that $p^{-1} + q^{-1} + r^{-1} = 1$. Davydov (1968) showed that

$$|\text{Cov}(X, Y)| \leq K_0 \|X\|_p \|Y\|_q [\alpha \{\sigma(X), \sigma(Y)\}]^{1/r}, \quad (2.37)$$

where $\|X\|_p^p = EX^p$, K_0 is an universal constant, and $\alpha \{\sigma(X), \sigma(Y)\}$ denotes the strong mixing coefficient between the σ -fields $\sigma(X)$ and $\sigma(Y)$ generated by the random variables X and Y , respectively.

Lemma 12. Under the assumptions of Theorem 2.2, (2.25) holds, that is

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left\{ \left\| n^{-1/2} \sum_{t=1}^n Z_{t,m} \right\| > \varepsilon \right\} = 0.$$

Proof of Lemma 12 : For $i = 1, \dots, k_0$, by stationarity we have

$$\begin{aligned} \text{var} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^n Z_{t,m,i} \right) &= \frac{1}{n} \sum_{t,s=1}^n \text{Cov}(Z_{t,m,i}, Z_{s,m,i}) \\ &= \frac{1}{n} \sum_{|h| < n} (n - |h|) \text{Cov}(Z_{t,m,i}, Z_{t-h,m,i}) \\ &\leq \sum_{h=-\infty}^{\infty} |\text{Cov}(Z_{t,m,i}, Z_{t-h,m,i})|. \end{aligned}$$

Because $\|d_\ell\| \leq K\rho^\ell$, we have

$$|Z_{t,m,i}| \leq K \sum_{\ell=m+1}^{\infty} \rho^\ell \|e_t\| \|e_{t-\ell}\|.$$

2. DAVYDOV, Y. A. (1968) Convergence of Distributions Generated by Stationary Stochastic Processes. *Theory of Probability and Applications* 13, 691–696.

Using also $E\|e_t\|^4 < \infty$, it follows from the Hölder inequality that

$$\sup_h |\text{Cov}(Z_{t,m,i}, Z_{t-h,m,i})| = \sup_h |EZ_{t,m,i}Z_{t-h,m,i}'| \leq K\rho^m. \quad (2.38)$$

Let $h > 0$ such that $[h/2] > m$. Write

$$Z_{t,m,i} = Z_{t,m,i}^{h^-} + Z_{t,m,i}^{h^+},$$

where

$$Z_{t,m,i}^{h^-} = 2 \sum_{\ell=m+1}^{[h/2]} e'_{t-\ell} d'_\ell \Sigma_{e0}^{-1} e_t, \quad Z_{t,m,i}^{h^+} = 2 \sum_{\ell=[h/2]+1}^{\infty} e'_{t-\ell} d'_\ell \Sigma_{e0}^{-1} e_t.$$

Note that $Z_{t,m,i}^{h^-}$ belongs to the σ -field generated by $\{e_t, e_{t-1}, \dots, e_{t-[h/2]}\}$ and that $Z_{t-h,m,i}$ belongs to the σ -field generated by $\{e_{t-h}, e_{t-h-1}, \dots\}$. Note also that $\alpha_\epsilon(\cdot) = \alpha_\epsilon(\cdot)$ and that, by **A7**, $E|Z_{t,m,i}^{h^-}|^{2+\nu} < \infty$ and $E|Z_{t-h,m,i}|^{2+\nu} < \infty$. Lemma **11** then entails that

$$\left| \text{Cov}(Z_{t,m,i}^{h^-}, Z_{t-h,m,i}) \right| \leq K\alpha_\epsilon^{\nu/(2+\nu)}([h/2]). \quad (2.39)$$

By the argument used to show (2.38), we also have

$$\left| \text{Cov}(Z_{t,m,i}^{h^+}, Z_{t-h,m,i}) \right| \leq K\rho^h \rho^m. \quad (2.40)$$

In view of (2.38), (2.39) and (2.40),

$$\sum_{h=0}^{\infty} |\text{Cov}(Z_{t,m,i}, Z_{t-h,m,i})| \leq K\rho^m + K \sum_{h=m}^{\infty} \alpha_\epsilon^{\nu/(2+\nu)}(h) \rightarrow 0$$

as $m \rightarrow \infty$ by **A7**. We have the same bound for $h < 0$. The conclusion follows from the Markov inequality. \square

Lemma 13. Let the assumptions of Theorem **2.2** be satisfied. For $\vartheta \in \Theta$, the matrix

$$I(\vartheta) = \lim_{n \rightarrow \infty} \text{Var} \frac{\partial}{\partial \vartheta} \ell_n(\vartheta),$$

exists.

Proof of Lemma 13 : Let

$$\dot{l}_t = \frac{\partial}{\partial \vartheta} l_t(\vartheta) = \left(\frac{\partial}{\partial \vartheta_1} l_t(\vartheta), \dots, \frac{\partial}{\partial \vartheta_{k_0}} l_t(\vartheta) \right)'. \quad (2.41)$$

By stationarity of (\dot{l}_t) , we have

$$\begin{aligned} \text{var} \left(\frac{\partial}{\partial \vartheta} \ell_n(\vartheta) \right) &= \text{var} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^n \dot{l}_t \right) = \frac{1}{n} \sum_{t,s=1}^n \text{Cov}(\dot{l}_t, \dot{l}_s) \\ &= \frac{1}{n} \sum_{h=-n+1}^{n-1} (n - |h|) \text{Cov}(\dot{l}_t, \dot{l}_{t-h}). \end{aligned}$$

By the dominated convergence theorem, it follows that $I(\vartheta)$ exists and is given by

$$I(\vartheta) = \sum_{h=-\infty}^{\infty} \text{Cov}(\dot{l}_t, \dot{l}_{t-h})$$

whenever

$$\sum_{h=-\infty}^{\infty} \left\| \text{Cov}(\dot{l}_t, \dot{l}_{t-h}) \right\| < \infty. \quad (2.42)$$

Recall that, for all $i \in \{1, \dots, k_0\}$ we have

$$\begin{aligned} \frac{\partial l_t(\vartheta)}{\partial \vartheta_i} &= \text{Tr} \left[\Sigma_e^{-1} \{I_d - e_t(\vartheta)e_t'(\vartheta)\Sigma_e^{-1}\} \frac{\partial \Sigma_e(\vartheta)}{\partial \vartheta_i} \right] + 2 \frac{\partial e_t'(\vartheta)}{\partial \vartheta_i} \Sigma_e^{-1} e_t(\vartheta) \\ &= \text{Tr} \left[\Sigma_e^{-1} \{I_d - e_t(\vartheta)e_t'(\vartheta)\Sigma_e^{-1}\} \frac{\partial \Sigma_e(\vartheta)}{\partial \vartheta_i} \right] + 2 \sum_{\ell=1}^{\infty} e_{t-\ell}'(\vartheta) d_{\ell}' \Sigma_e^{-1} e_t(\vartheta). \end{aligned}$$

In view of the last equality, the i -th component of \dot{l}_t is of the form

$$\dot{l}_{i t} = \sum_{\ell=0}^{\infty} \sum_{k_1, k_2=1}^d \alpha_{\ell}(i, k_1, k_2) e_{k_1 t}(\vartheta) e_{k_2 t-\ell}(\vartheta) \quad (2.43)$$

where

$$|\alpha_{\ell}(i, k_1, k_2)| \leq K \rho^{\ell}, \quad (2.44)$$

uniformly in $(i, k_1, k_2) \in \{1, \dots, k_0\} \times \{1, \dots, d\}^2$. In order to obtain (2.42), it suffices to show that

$$\sum_{h=-\infty}^{\infty} \left| \text{Cov}(\dot{l}_{i t}, \dot{l}_{j t-h}) \right| < \infty \quad (2.45)$$

for all $(i, j) \in \{1, \dots, k_0\}^2$. In view of (2.43) and (2.44), we have for $h \geq 0$

$$\left| \text{Cov}(\dot{l}_{i t}, \dot{l}_{j t-h}) \right| \leq c_1 + c_2,$$

where

$$c_1 = K \sum_{\ell=[h/2]+1}^{\infty} \sum_{\ell'=0}^{\infty} \sum_{k_1, \dots, k_4=1}^d \rho^{\ell+\ell'} \left| \text{Cov}(e_{k_1 t}(\vartheta) e_{k_2 t-\ell}(\vartheta), e_{k_3 t-h}(\vartheta) e_{k_4 t-\ell'-h}(\vartheta)) \right|$$

and

$$c_2 = K \sum_{\ell=0}^{[h/2]} \sum_{\ell'=0}^{\infty} \sum_{k_1, \dots, k_4=1}^d \rho^{\ell+\ell'} \left| \text{Cov}(e_{k_1 t}(\vartheta) e_{k_2 t-\ell}(\vartheta), e_{k_3 t-h}(\vartheta) e_{k_4 t-\ell'-h}(\vartheta)) \right|.$$

Because

$$\left| \text{Cov}(e_{k_1 t}(\vartheta) e_{k_2 t-\ell}(\vartheta), e_{k_3 t-h}(\vartheta) e_{k_4 t-\ell'-h}(\vartheta)) \right| \leq E \|e_t(\vartheta)\|^4 < \infty$$

by Assumption **A7**, we have

$$c_1 \leq K\rho^h.$$

Lemma 11 entails that

$$c_2 \leq K \sum_{\ell=0}^{[h/2]} \sum_{\ell'=0}^{\infty} \rho^{\ell+\ell'} \alpha_{\epsilon}^{\nu/(2+\nu)} ([h/2]) \leq K \alpha_{\epsilon}^{\nu/(2+\nu)} ([h/2]).$$

It follows that

$$\sum_{h=0}^{\infty} \left| \text{Cov} \left(i_{i,t}, i_{j,t-h} \right) \right| \leq K \sum_{h=0}^{\infty} \rho^h + K \sum_{h=0}^{\infty} \alpha_{\epsilon}^{\nu/(2+\nu)} ([h/2]) < \infty$$

by Assumption **A7**. The same bounds clearly holds for

$$\sum_{h=-\infty}^0 \left| \text{Cov} \left(i_{i,t}, i_{j,t-h} \right) \right|,$$

which shows (2.45) and completes the proof. \square

Chapitre 3

Estimating the asymptotic variance of LS estimator of VARMA models with uncorrelated but non-independent error terms

Abstract We consider the problems of estimating the asymptotic variance of the least squares (LS) and/ or quasi-maximum likelihood (QML) estimators of vector autoregressive moving-average (VARMA) models under the assumption that the errors are uncorrelated but not necessarily independent. We relax the standard independence assumption to extend the range of application of the VARMA models, and allow to cover linear representations of general nonlinear processes. We first give expressions for the derivatives of the VARMA residuals in terms of the parameters of the models. Secondly we give an explicit expression of the asymptotic variance of the QML/LS estimator, in terms of the VAR and MA polynomials, and of the second and fourth-order structure of the noise. We deduce a consistent estimator of the asymptotic variance of the QML/LS estimator.

Keywords : QMLE/LSE, residuals derivatives, Structural representation, weak VARMA models.

3.1 Introduction

The class of vector autoregressive moving-average (VARMA) models and the subclass of vector autoregressive (VAR) models are used in time series analysis and econometrics to describe not only the properties of the individual time series but also the

possible cross-relationships between the time series (see Reinsel, 1997, Lütkepohl, 2005, 1993). This paper is devoted to the problems of estimating the asymptotic variance of the least squares (LS) and/ or quasi-maximum likelihood (QML) estimators of VARMA models under the assumption that the errors are uncorrelated but not necessarily independent (*i.e.* weak VARMA models). A process $X = (X_t)$ is said to be *nonlinear* when the innovation process in the Wold decomposition (see *e.g.* Brockwell and Davis, 1991, for the univariate case, and Reinsel, 1997, in the multivariate framework) is uncorrelated but not necessarily independent, and is said to be *linear* in the opposite case (*i.e.* when the innovation process in the Wold decomposition is iid). We relax the standard independence assumption to extend the range of application of the VARMA models, and allow to cover linear representations of general *nonlinear* processes.

The estimation of weak VARMA models has been studied in Boubacar Mainassara and Francq (2009). They showed that the asymptotic variance of the usual estimators has the "sandwich" form $\Omega = J^{-1}IJ^{-1}$, which reduces to the standard form $\Omega = J^{-1}$ in the *linear* case. They used an empirical estimator of the matrix J . To obtain a consistent estimator of the matrix I , they used a nonparametric kernel estimator, also called heteroscedastic autocorrelation consistent (HAC) estimator (see Newey and West, 1987, or Andrews, 1991) and they also used an alternative method based on a parametric AR estimate of a spectral density at frequency 0 (see Brockwell and Davis, 1991, p. 459). The asymptotic theory of weak ARMA model is mainly limited to the univariate framework (see Francq and Zakoïan, 2005, for a review of this topic). In the multivariate analysis, notable exceptions are Dufour and Pelletier (2005) who study the asymptotic properties of a generalization of the regression-based estimation method proposed by Hannan and Rissanen (1982) under weak assumptions on the innovation process, Francq and Raïssi (2007) who study portmanteau tests for weak VAR models and Boubacar Mainassara (in a working paper, 2009) who studies portmanteau tests for weak VARMA models. The main goal of the present article is to complete the available results concerning the statistical analysis of weak VARMA models, by proposing another estimator of Ω , which allows to separate the effects due to the VARMA parameters from those due to the *nonlinear* structure of the noise. This topic has not been studied in the above-mentioned papers.

The paper is organized as follows. Section 3.2 presents the models that we consider here, and presents the results on the QMLE/LSE asymptotic distribution obtained by Boubacar Mainassara and Francq (2009). In Section 3.3 we give expressions for the derivatives of the VARMA residuals in terms of parameters of the models. Section 3.4 is devoted to find an explicit expression of the asymptotic variance of the QML/LS estimator, in terms of the VAR and MA polynomials, and of the second and fourth-order structure of the noise. In Section 3.5 we deduce a consistent estimator of this asymptotic variance. The proofs of the main results are collected in the appendix. We denote by $A \otimes B$ the Kronecker product of two matrices A and B (and by $A^{\otimes 2}$ when the matrix $A = B$), and by $\text{vec}A$ the vector obtained by stacking the columns of A . The reader is referred to Magnus and Neudecker (1988) for the properties of these operators.

Let 0_r be the null vector of \mathbb{R}^r , and let I_r be the $r \times r$ identity matrix.

3.2 Model and assumptions

Consider a d -dimensional stationary process (X_t) satisfying a structural VARMA(p, q) representation of the form

$$A_{00}X_t - \sum_{i=1}^p A_{0i}X_{t-i} = B_{00}\epsilon_t - \sum_{i=1}^q B_{0i}\epsilon_{t-i}, \quad \forall t \in \mathbb{Z} = \{0, \pm 1, \dots\}, \quad (3.1)$$

where ϵ_t is a white noise, namely a stationary sequence of centered and uncorrelated random variables with a non singular variance Σ_0 . The structural forms are mainly used in econometrics to introduce instantaneous relationships between economic variables. Of course, constraints are necessary for the identifiability of these representations. Let $[A_{00} \dots A_{0p} B_{00} \dots B_{0q}]$ be the $d \times (p+q+2)d$ matrix of all the coefficients, without any constraint. The matrix Σ_0 is considered as a nuisance parameter. The parameter of interest is denoted θ_0 , where θ_0 belongs to the parameter space $\Theta \subset \mathbb{R}^{k_0}$, and k_0 is the number of unknown parameters, which is typically much smaller than $(p+q+2)d^2$. The matrices $A_{00}, \dots, A_{0p}, B_{00}, \dots, B_{0q}$ involved in (3.1) are specified by θ_0 . More precisely, we write $A_{0i} = A_i(\theta_0)$ and $B_{0j} = B_j(\theta_0)$ for $i = 0, \dots, p$ and $j = 0, \dots, q$, and $\Sigma_0 = \Sigma(\theta_0)$. We need the following assumptions used by Boubacar Mainassara and Francq (2009) to ensure the consistence and the asymptotic normality of the quasi-maximum likelihood estimator (QMLE).

A1 : The applications $\theta \mapsto A_i(\theta)$ $i = 0, \dots, p$, $\theta \mapsto B_j(\theta)$ $j = 0, \dots, q$ and $\theta \mapsto \Sigma(\theta)$ admit continuous third order derivatives for all $\theta \in \Theta$.

For simplicity we now write A_i, B_j and Σ instead of $A_i(\theta), B_j(\theta)$ and $\Sigma(\theta)$. Let $A_\theta(z) = A_0 - \sum_{i=1}^p A_i z^i$ and $B_\theta(z) = B_0 - \sum_{i=1}^q B_i z^i$.

A2 : For all $\theta \in \Theta$, we have $\det A_\theta(z) \det B_\theta(z) \neq 0$ for all $|z| \leq 1$.

A3 : We have $\theta_0 \in \Theta$, where Θ is compact.

A4 : The process (ϵ_t) is stationary and ergodic.

A5 : For all $\theta \in \Theta$ such that $\theta \neq \theta_0$, either the transfer functions

$$A_0^{-1} B_0 B_\theta^{-1}(z) A_\theta(z) \neq A_{00}^{-1} B_{00} B_{\theta_0}^{-1}(z) A_{\theta_0}(z)$$

for some $z \in \mathbb{C}$, or

$$A_0^{-1} B_0 \Sigma B_0' A_0^{-1'} \neq A_{00}^{-1} B_{00} \Sigma_0 B_{00}' A_{00}^{-1'}.$$

A6 : We have $\theta_0 \in \overset{\circ}{\Theta}$, where $\overset{\circ}{\Theta}$ denotes the interior of Θ .

A7 : We have $E\|\epsilon_t\|^{4+2\nu} < \infty$ and $\sum_{k=0}^{\infty} \{\alpha_{\epsilon}(k)\}^{\frac{\nu}{2+\nu}} < \infty$ for some $\nu > 0$.

The reader is referred to Boubacar Mainassara and Francq (2009) for a discussion of these assumptions. Note that (ϵ_t) can be replaced by (X_t) in **A4**, because $X_t = A_{\theta_0}^{-1}(L)B_{\theta_0}(L)\epsilon_t$ and $\epsilon_t = B_{\theta_0}^{-1}(L)A_{\theta_0}(L)X_t$, where L stands for the backward operator. Note that from **A1** the matrices A_0 and B_0 are invertible. Introducing the innovation process $e_t = A_{00}^{-1}B_{00}\epsilon_t$, the structural representation $A_{\theta_0}(L)X_t = B_{\theta_0}(L)\epsilon_t$ can be rewritten as the reduced VARMA representation

$$X_t - \sum_{i=1}^p A_{00}^{-1}A_{0i}X_{t-i} = e_t - \sum_{i=1}^q A_{00}^{-1}B_{0i}B_{00}^{-1}A_{00}e_{t-i}.$$

We thus recursively define $\tilde{e}_t(\theta)$ for $t = 1, \dots, n$ by

$$\tilde{e}_t(\theta) = X_t - \sum_{i=1}^p A_0^{-1}A_iX_{t-i} + \sum_{i=1}^q A_0^{-1}B_iB_0^{-1}A_0\tilde{e}_{t-i}(\theta),$$

with initial values $\tilde{e}_0(\theta) = \dots = \tilde{e}_{1-q}(\theta) = X_0 = \dots = X_{1-p} = 0$. The gaussian quasi-likelihood is given by

$$L_n(\theta, \Sigma_e) = \prod_{t=1}^n \frac{1}{(2\pi)^{d/2} \sqrt{\det \Sigma_e}} \exp \left\{ -\frac{1}{2} \tilde{e}_t'(\theta) \Sigma_e^{-1} \tilde{e}_t(\theta) \right\}, \quad \Sigma_e = A_0^{-1}B_0\Sigma B_0' A_0^{-1'}.$$

A quasi-maximum likelihood estimator (QMLE) of θ and Σ_e is a measurable solution $(\hat{\theta}_n, \hat{\Sigma}_e)$ of

$$(\hat{\theta}_n, \hat{\Sigma}_e) = \arg \max_{\theta \in \Theta, \Sigma_e} L_n(\theta, \Sigma_e).$$

We now use the matrix M_{θ_0} of the coefficients of the reduced form to that made by Boubacar Mainassara and Francq (2009), where

$$M_{\theta_0} = [A_{00}^{-1}A_{01} : \dots : A_{00}^{-1}A_{0p} : A_{00}^{-1}B_{01}B_{00}^{-1}A_{00} : \dots : A_{00}^{-1}B_{0q}B_{00}^{-1}A_{00}].$$

Now we need an assumption which specifies how this matrix depends on the parameter θ_0 . Let \dot{M}_{θ_0} be the matrix $\partial \text{vec}(M_{\theta}) / \partial \theta'$ evaluated at θ_0 .

A8 : The matrix \dot{M}_{θ_0} is of full rank k_0 .

Under Assumptions **A1–A8**, Boubacar Mainassara and Francq (2009) have showed the consistency ($\hat{\theta}_n \rightarrow \theta_0$ *a.s.* as $n \rightarrow \infty$) and the asymptotic normality of the QMLE :

$$\sqrt{n} \left(\hat{\vartheta}_n - \vartheta_0 \right) \xrightarrow{L} \mathcal{N}(0, \Omega := J^{-1} I J^{-1}), \quad (3.2)$$

where $J = J(\theta_0, \Sigma_{e0})$ and $I = I(\theta_0, \Sigma_{e0})$, with

$$J(\theta, \Sigma_e) = \lim_{n \rightarrow \infty} \frac{2}{n} \frac{\partial^2}{\partial \theta \partial \theta'} \log L_n(\theta, \Sigma_e) \quad \text{a.s.}$$

and

$$I(\theta, \Sigma_e) = \lim_{n \rightarrow \infty} \text{Var} \frac{2}{\sqrt{n}} \frac{\partial}{\partial \theta} \log L_n(\theta, \Sigma_e). \quad (3.3)$$

3.3 Expression for the derivatives of the VARMA residuals

The reduced VARMA representation can be rewritten under the compact form

$$\mathbf{A}_\theta(L)X_t = \mathbf{B}_\theta(L)e_t(\theta),$$

where $\mathbf{A}_\theta(L) = I_d - \sum_{i=1}^p \mathbf{A}_i L^i$ and $\mathbf{B}_\theta(L) = I_d - \sum_{i=1}^q \mathbf{B}_i L^i$, with $\mathbf{A}_i = A_0^{-1} A_i$ and $\mathbf{B}_i = A_0^{-1} B_i B_0^{-1} A_0$. For $\ell = 1, \dots, p$ and $\ell' = 1, \dots, q$, let $\mathbf{A}_\ell = (\mathbf{a}_{ij,\ell})$, $\mathbf{B}_{\ell'} = (\mathbf{b}_{ij,\ell'})$, $\mathbf{a}_\ell = \text{vec}[\mathbf{A}_\ell]$ and $\mathbf{b}_{\ell'} = \text{vec}[\mathbf{B}_{\ell'}]$. We denote respectively by

$$\mathbf{a} := (\mathbf{a}'_1, \dots, \mathbf{a}'_p)' \quad \text{and} \quad \mathbf{b} := (\mathbf{b}'_1, \dots, \mathbf{b}'_q)',$$

the coefficients of the multivariate AR and MA parts. Thus we can rewrite $\theta = (\mathbf{a}', \mathbf{b}')'$, where $\mathbf{a} \in \mathbb{R}^{k_1}$ depends on A_0, \dots, A_p , and where $\mathbf{b} \in \mathbb{R}^{k_2}$ depends on B_0, \dots, B_q , with $k_1 + k_2 = k_0$. For $i, j = 1, \dots, d$, let $M_{ij}(L)$ and $N_{ij}(L)$ the $(d \times d)$ -matrix operators defined by

$$M_{ij}(L) = \mathbf{B}_\theta^{-1}(L)E_{ij}\mathbf{A}_\theta^{-1}(L)\mathbf{B}_\theta(L) \quad \text{and} \quad N_{ij}(L) = \mathbf{B}_\theta^{-1}(L)E_{ij},$$

where E_{ij} is the $d \times d$ matrix with 1 at position (i, j) and 0 elsewhere. We denote by $\mathbf{A}_{ij,h}^*$ and $\mathbf{B}_{ij,h}^*$ the $(d \times d)$ matrices defined by

$$M_{ij}(z) = \sum_{h=0}^{\infty} \mathbf{A}_{ij,h}^* z^h, \quad N_{ij}(z) = \sum_{h=0}^{\infty} \mathbf{B}_{ij,h}^* z^h, \quad |z| \leq 1$$

for $h \geq 0$. Take $\mathbf{A}_{ij,h}^* = \mathbf{B}_{ij,h}^* = 0$ when $h < 0$. Let respectively $\mathbf{A}_h^* = [\mathbf{A}_{11,h}^* : \mathbf{A}_{21,h}^* : \dots : \mathbf{A}_{dd,h}^*]$ and $\mathbf{B}_h^* = [\mathbf{B}_{11,h}^* : \mathbf{B}_{21,h}^* : \dots : \mathbf{B}_{dd,h}^*]$ the $d \times d^3$ matrices. We consider the white noise "empirical" autocovariances

$$\Gamma_e(h) = \frac{1}{n} \sum_{t=h+1}^n e_t e'_{t-h}, \quad \text{for } 0 \leq h < n.$$

For $k, k', m, m' = 1, \dots, \infty$, let

$$\Gamma(k, k') = \sum_{h=-\infty}^{\infty} E(\{e_{t-k} \otimes e_t\} \{e_{t-h-k'} \otimes e_{t-h}\}'),$$

and

$$\Gamma_{m,m'} = (\Gamma(k, k'))_{1 \leq k \leq m, 1 \leq k' \leq m'} = \sum_{h=-\infty}^{\infty} \text{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t-h})$$

where $\mathbf{Y}_t = (e'_{t-1}, \dots, e'_{t-m})' \otimes e_t$. Let the $d \times d^3(p+q)$ matrix

$$\lambda_h(\theta) = [-\mathbf{A}_{h-1}^* : \dots : -\mathbf{A}_{h-p}^* : \mathbf{B}_{h-1}^* : \dots : \mathbf{B}_{h-q}^*]. \quad (3.4)$$

This matrix is well defined because the coefficients of the series expansions of \mathbf{A}_θ^{-1} and \mathbf{B}_θ^{-1} decrease exponentially fast to zero. For the univariate ARMA model, Francq, Roy and Zakoïan (2004) have showed in Lemma A.1 that $|\Gamma(k, k')| \leq K \max(k, k')$ for some constant K . We can generalize this result for the multivariate ARMA model. Then we obtain $\|\Gamma(k, k')\| \leq K \max(k, k')$ for some constant K . The proof is similar to the univariate case. For univariate ARMA models, McLeod (1978) defined the noise derivatives by writing

$$\frac{\partial e_t}{\partial \phi_i} = v_{t-i}, \quad i = 1, \dots, p \quad \text{and} \quad \frac{\partial e_t}{\partial \beta_j} = u_{t-j}, \quad j = 1, \dots, q,$$

where ϕ_i and β_j are respectively the univariate AR and MA parameters. Let $\phi_\theta(L) = 1 - \sum_{i=1}^p \phi_i L^i$ and $\varphi_\theta(L) = 1 - \sum_{i=1}^q \varphi_i L^i$. We denote by ϕ_h^* and φ_h^* the coefficients defined by

$$\phi_\theta^{-1}(z) = \sum_{h=0}^{\infty} \phi_h^* z^h, \quad \varphi_\theta^{-1}(z) = \sum_{h=0}^{\infty} \varphi_h^* z^h, \quad |z| \leq 1$$

for $h \geq 0$. When p and q are not both equal to 0, let $\theta = (\theta_1, \dots, \theta_p, \theta_{p+1}, \dots, \theta_{p+q})'$. Then, McLeod (1978) showed that the univariate noise derivatives can be represented as

$$\frac{\partial e_t(\theta)}{\partial \theta} = (v_{t-1}(\theta), \dots, v_{t-p}(\theta), u_{t-1}(\theta), \dots, u_{t-q}(\theta))',$$

where

$$v_t(\theta) = -\phi_\theta^{-1}(L)e_t(\theta) = -\sum_{h=0}^{\infty} \phi_h^* e_{t-h}(\theta), \quad u_t(\theta) = \varphi_\theta^{-1}(L)e_t(\theta) = \sum_{h=0}^{\infty} \varphi_h^* e_{t-h}(\theta)$$

and

$$e_t(\theta) = \varphi_\theta^{-1}(L)\phi_\theta(L)X_t.$$

We can generalize these expansions for the multivariate case in the following proposition.

Proposition A 1. Under the assumptions **A1–A8**, we have

$$\frac{\partial e_t(\theta)}{\partial \theta'} = [V_{t-1}(\theta) : \dots : V_{t-p}(\theta) : U_{t-1}(\theta) : \dots : U_{t-q}(\theta)],$$

where

$$V_t(\theta) = -\sum_{h=0}^{\infty} \mathbf{A}_h^* (I_{d^2} \otimes e_{t-h}(\theta)) \quad \text{and} \quad U_t(\theta) = \sum_{h=0}^{\infty} \mathbf{B}_h^* (I_{d^2} \otimes e_{t-h}(\theta))$$

with the $d \times d^3$ matrices $\mathbf{A}_h^* = [\mathbf{A}_{11,h}^* : \mathbf{A}_{21,h}^* : \dots : \mathbf{A}_{dd,h}^*]$ and $\mathbf{B}_h^* = [\mathbf{B}_{11,h}^* : \mathbf{B}_{21,h}^* : \dots : \mathbf{B}_{dd,h}^*]$. Moreover, at $\theta = \theta_0$ we have

$$\frac{\partial e_t(\theta_0)}{\partial \theta'} = \sum_{i \geq 1} \lambda_i (I_{d^2(p+q)} \otimes e_{t-i}(\theta_0)),$$

with the λ_i 's are defined in (3.4).

3.4 Explicit expressions for I and J

We now give expressions for the matrices I and J involved in the asymptotic variance Ω of the QMLE. In these expressions, we separate terms depending from the parameter θ_0 from terms depending on the weak noise second-order structure. Let the matrix

$$\mathcal{M} := E \left\{ \left(I_{d^2(p+q)} \otimes e'_t \right)^{\otimes 2} \right\}$$

involving the second-order moments of (e_t) . For $(i_1, i_2) \in \{1, \dots, d^2(p+q)\} \times \{1, \dots, d^3(p+q)\}$, let $\mathcal{M}_{i_1 i_2}$ the (i_1, i_2) -th block of \mathcal{M} of size $d^2(p+q) \times d^3(p+q)$. Note that the block matrix \mathcal{M} is not block diagonal. For $j_1 \in \{1, \dots, d\}$, we have $\mathcal{M}_{i_1 i_2} \neq 0_{d^2(p+q) \times d^3(p+q)}$ if $i_2 = d(i_1 - 1) + j_1$ and $\mathcal{M}_{i_1 i_2} = 0_{d^2(p+q) \times d^3(p+q)}$ otherwise. For $(k, k') \in \{1, \dots, d^2(p+q)\} \times \{1, \dots, d^3(p+q)\}$ and $j_2 \in \{1, \dots, d\}$, the (k, k') -th element of $\mathcal{M}_{i_1 i_2}$ as of the form $\sigma_{j_1 j_2}^2$ if $k' = d(k - 1) + j_2$ and zero otherwise and where $\sigma_{j_1 j_2}^2 = E e_{j_1 t} e_{j_2 t} = \Sigma(j_1, j_2)$. We are now able to state the following proposition, which provides a form for $J = J(\theta_0, \Sigma_{e_0})$, in which the terms depending on θ_0 (through the matrices λ_i) are distinguished from the terms depending on the second-order moments of (e_t) (through the matrix \mathcal{M}) and the terms of the noise variance of the multivariate innovation process (through the matrix Σ_{e_0}).

Proposition A 2. Under Assumptions **A1–A8**, we have

$$\text{vec } J = 2 \sum_{i \geq 1} \mathcal{M} \{ \lambda'_i \otimes \lambda'_i \} \text{vec } \Sigma_{e_0}^{-1},$$

where the λ_i 's are defined by (3.4).

In view of (3.3), we have

$$I = \text{Var}_{as} \frac{1}{\sqrt{n}} \sum_{t=1}^n \Upsilon_t = \sum_{h=-\infty}^{+\infty} \text{Cov}(\Upsilon_t, \Upsilon_{t-h}), \quad (3.5)$$

where

$$\Upsilon_t = \frac{\partial}{\partial \theta} \left\{ \log \det \Sigma_e + e'_t(\theta) \Sigma_e^{-1} e_t(\theta) \right\}_{\theta=\theta_0}. \quad (3.6)$$

The existence of the sum of the right-hand side of (3.5) is a consequence of **A7** and of Davyvov's inequality (1968) (see *e.g.* Lemma 11 in Boubacar Mainassara and Francq, 2009). As the matrix J , we will decompose I in terms involving the VARMA parameter θ_0 and terms involving the distribution of the innovations e_t . Let the matrices

$$\mathcal{M}_{i,j,h} := E \left\{ \left(e'_{t-h} \otimes \left(I_{d^2(p+q)} \otimes e'_{t-j-h} \right) \right) \otimes \left(e'_t \otimes \left(I_{d^2(p+q)} \otimes e'_{t-i} \right) \right) \right\}.$$

The terms depending on the VARMA parameter are the matrices λ_i defined in (3.4) and let the matrices

$$\Gamma(i, j) = \sum_{h=-\infty}^{+\infty} \mathcal{M}_{i,j,h}$$

involving the fourth-order moments of e_t . The terms depending on the noise variance of the multivariate innovation process are in Σ_{e0} . For $(i_1, i_2) \in \{1, \dots, d^2(p+q)\} \times \{1, \dots, d^4(p+q)\}$, let $\mathcal{M}_{i_1 i_2, ij, h}$ be the (i_1, i_2) -th block of $\mathcal{M}_{ij, h}$ of size $d^2(p+q) \times d^4(p+q)$. Note that the block matrices $\mathcal{M}_{ij, h}$ are not block diagonal. For $j_1 \in \{1, \dots, d\}$, we have $\mathcal{M}_{i_1 i_2, ij, h} \neq 0_{d^2(p+q) \times d^4(p+q)}$ if $i_2 = d^2(i_1 - 1) - d(i_1 - 1) + j_1$ and $i_1 = d^3(p+q) + d^2(i_1 - 1) - d(i_1 - 1) + j_1$, and $\mathcal{M}_{i_1 i_2, ij, h} = 0_{d^2(p+q) \times d^4(p+q)}$ otherwise. For $(k, k') \in \{1, \dots, d^2(p+q)\} \times \{1, \dots, d^4(p+q)\}$ and $j_2 \in \{1, \dots, d\}$, the (k, k') -th element of $\mathcal{M}_{i_1 i_2, ij, h}$ is of the form $E e_{j_1 t-h} e_{j_1 t-j-h} e_{j_1 t} e_{j_2 t-i}$ when $k' = d^2(k-1) - d(k-1) + j_2$ and of the form $E e_{j_2 t-h} e_{j_2 t-j-h} e_{j_1 t} e_{j_2 t-i}$ when $k' = d^3(p+q) + d^2(k-1) - d(k-1) + j_2$, and zero otherwise. We now state an analog of Proposition 2 for $I = I(\theta_0, \Sigma_{e0})$.

Proposition A 3. Under Assumptions A1–A8, we have

$$\text{vec } I = 4 \sum_{i,j=1}^{+\infty} \Gamma(i, j) \left(\{I_d \otimes \lambda'_j\} \otimes \{I_d \otimes \lambda'_i\} \right) \text{vec} \left(\text{vec } \Sigma_{e0}^{-1} \{ \text{vec } \Sigma_{e0}^{-1} \}' \right),$$

where the λ_i 's are defined by (3.4).

Remark 3.1. Consider the univariate case $d = 1$. We obtain

$$\text{vec } J = 2 \sum_{i \geq 1} \{\lambda_i \otimes \lambda_i\}' \quad \text{and} \quad \text{vec } I = \frac{4}{\sigma^4} \sum_{i,j=1}^{+\infty} \gamma(i, j) \{\lambda_j \otimes \lambda_i\}',$$

where $\gamma(i, j) = \sum_{h=-\infty}^{+\infty} E(e_t e_{t-i} e_{t-h} e_{t-j-h})$ and $\lambda'_i \in \mathbb{R}^{p+q}$ are defined by (3.4).

Remark 3.2. Francq, Roy and Zakoïan (2005) considered the univariate case $d = 1$. In their paper, they used the least squares estimator and they obtained

$$E \frac{\partial^2}{\partial \theta \partial \theta'} e_t^2(\theta_0) = 2 \sum_{i \geq 1} \sigma^2 \lambda_i \lambda'_i \quad \text{and} \quad \text{Var} \left\{ 2e_t(\theta_0) \frac{\partial e_t(\theta_0)}{\partial \theta} \right\} = 4 \sum_{i,j \geq 1} \gamma(i, j) \lambda_i \lambda'_j$$

where σ^2 is the variance of the univariate process e_t and the vectors $\lambda_i = (-\phi_{i-1}^*, \dots, -\phi_{i-p}^*, \varphi_{i-1}^*, \dots, \varphi_{i-q}^*)' \in \mathbb{R}^{p+q}$, with the convention $\phi_i^* = \varphi_i^* = 0$ when $i < 0$. Using the vec operator and the elementary relation $\text{vec}(aa') = a \otimes a'$, their result writes

$$\text{vec } J = \text{vec } E \frac{\partial^2}{\partial \theta \partial \theta'} \ell_n(\theta_0) = \frac{1}{\sigma^2} \text{vec } E \frac{\partial^2}{\partial \theta \partial \theta'} e_t^2(\theta_0) = 2 \sum_{i \geq 1} \lambda_i \otimes \lambda_i \quad \text{and}$$

$$\text{vec } I = \text{vec } \text{Var} \left\{ \frac{\partial \ell_n(\theta_0)}{\partial \theta} \right\} = \frac{1}{\sigma^4} \text{vec } \text{Var} \left\{ 2e_t \frac{\partial e_t(\theta_0)}{\partial \theta} \right\} = \frac{4}{\sigma^4} \sum_{i,j \geq 1} \gamma(i, j) \lambda_i \otimes \lambda_j,$$

which are the expressions given in Remark 3.1.

3.5 Estimating the asymptotic variance matrix

In Section 3.4, we obtained explicit expressions for I and J . We now turn to the estimation of these matrices. Let $\hat{e}_t = \tilde{e}_t(\hat{\theta}_n)$ be the QMLE residuals when $p > 0$ or $q > 0$, and let $\hat{e}_t = e_t = X_t$ when $p = q = 0$. When $p + q \neq 0$, we have $\hat{e}_t = 0$ for $t \leq 0$ and $t > n$ and

$$\hat{e}_t = X_t - \sum_{i=1}^p A_0^{-1}(\hat{\theta}_n) A_i(\hat{\theta}_n) \hat{X}_{t-i} + \sum_{i=1}^q A_0^{-1}(\hat{\theta}_n) B_i(\hat{\theta}_n) B_0^{-1}(\hat{\theta}_n) A_0(\hat{\theta}_n) \hat{e}_{t-i},$$

for $t = 1, \dots, n$, with $\hat{X}_t = 0$ for $t \leq 0$ and $\hat{X}_t = X_t$ for $t \geq 1$. Let $\hat{\Sigma}_{e0} = n^{-1} \sum_{t=1}^n \hat{e}_t \hat{e}_t'$ be an estimator of Σ_{e0} . The matrix \mathcal{M} involved in the expression of J can easily be estimated by its empirical counterpart

$$\hat{\mathcal{M}}_n := \frac{1}{n} \sum_{t=1}^n \left\{ (I_{d^2(p+q)} \otimes \hat{e}_t')^{\otimes 2} \right\}.$$

In view of Proposition 2, we define an estimator \hat{J}_n of J by

$$\text{vec } \hat{J}_n = \sum_{i \geq 1} \hat{\mathcal{M}}_n \left\{ \hat{\lambda}'_i \otimes \hat{\lambda}'_i \right\} \text{vec } \hat{\Sigma}_{e0}^{-1}.$$

We are now able to state the following theorem, which shows the strong consistency of \hat{J}_n .

Theorem 3.1. *Under Assumptions A1–A8, we have*

$$\hat{J}_n \rightarrow J \quad \text{a.s. as } n \rightarrow \infty.$$

In the standard strong VARMA case $\hat{\Omega} = 2\hat{J}^{-1}$ is a strongly consistent estimator of Ω . In the general weak VARMA case this estimator is not consistent when $I \neq 2J$. So we need a consistent estimator of I . Let

$$\mathcal{M}_{n \, ij, h} := \frac{1}{n} \sum_{t=1}^{n-|h|} \left(\{e'_{t-h} \otimes (I_{d^2(p+q)} \otimes e'_{t-j-h})\} \otimes \{e'_t \otimes (I_{d^2(p+q)} \otimes e'_{t-i})\} \right).$$

To estimate $\Gamma(i, j)$ consider a sequence of real numbers $(b_n)_{n \in \mathbb{N}^*}$ such that

$$b_n \rightarrow 0 \quad \text{and} \quad n b_n^{\frac{10+4\nu}{\nu}} \rightarrow \infty \quad \text{as } n \rightarrow \infty, \quad (3.7)$$

and a weight function $f : \mathbb{R} \rightarrow \mathbb{R}$ which is bounded, with compact support $[-a, a]$ and continuous at the origin with $f(0) = 1$. Note that under the above assumptions, we have

$$b_n \sum_{|h| < n} |f(hb_n)| = O(1). \quad (3.8)$$

Let

$$\hat{\mathcal{M}}_{n\ ij,h} := \frac{1}{n} \sum_{t=1}^{n-|h|} \left(\{ \hat{e}'_{t-h} \otimes (I_{d^2(p+q)} \otimes \hat{e}'_{t-j-h}) \} \otimes \{ \hat{e}'_t \otimes (I_{d^2(p+q)} \otimes \hat{e}'_{t-i}) \} \right).$$

Consider the matrix

$$\hat{\Gamma}_n(i, j) := \sum_{h=-T_n}^{+T_n} f(hb_n) \hat{\mathcal{M}}_{n\ ij,h} \quad \text{and} \quad T_n = \left\lfloor \frac{a}{b_n} \right\rfloor,$$

where $[x]$ denotes the integer part of x . In view of Proposition 3, we define an estimator \hat{I}_n of I by

$$\text{vec } \hat{I}_n = 4 \sum_{i,j=1}^{+\infty} \hat{\Gamma}_n(i, j) \left(\{ I_d \otimes \hat{\lambda}'_i \} \otimes \{ I_d \otimes \hat{\lambda}'_j \} \right) \text{vec} \left(\text{vec } \hat{\Sigma}_{\epsilon_0}^{-1} \left\{ \text{vec } \hat{\Sigma}_{\epsilon_0}^{-1} \right\}' \right).$$

We are now able to state the following theorem, which shows the weak consistency of \hat{I}_n .

Theorem 3.2. *Under Assumptions A1–A8, we have*

$$\hat{I}_n \rightarrow I \quad \text{in probability as } n \rightarrow \infty.$$

Therefore Theorems 3.1 and 3.2 show that

$$\hat{\Omega}_n := \hat{J}_n^{-1} \hat{I}_n \hat{J}_n^{-1}$$

is a weakly estimator of the asymptotic covariance matrix $\Omega := J^{-1} I J^{-1}$.

3.6 Technical proofs

Proof of Proposition 1 : Because $\theta' = (\mathbf{a}', \mathbf{b}')$, Lemmas 3.3 and 3.4 below show that

$$\begin{aligned} \frac{\partial e_t(\theta)}{\partial \theta'} &= [V_{t-1}(\theta) : \dots : V_{t-p}(\theta) : U_{t-1}(\theta) : \dots : U_{t-q}(\theta)] \\ &= \sum_{h=0}^{\infty} [-\mathbf{A}_{h-1}^* (I_{d^2} \otimes e_{t-h}(\theta)) : \dots : -\mathbf{A}_{h-p}^* (I_{d^2} \otimes e_{t-h}(\theta)) : \\ &\quad \mathbf{B}_{h-1}^* (I_{d^2} \otimes e_{t-h}(\theta)) : \dots : \mathbf{B}_{h-q}^* (I_{d^2} \otimes e_{t-h}(\theta))] \\ &= \sum_{h=0}^{\infty} [-\mathbf{A}_{h-1}^* : \dots : -\mathbf{A}_{h-p}^* : \mathbf{B}_{h-1}^* : \dots : \mathbf{B}_{h-q}^*] \begin{bmatrix} I_{d^2 p} \otimes e_{t-h}(\theta) \\ I_{d^2 q} \otimes e_{t-h}(\theta) \end{bmatrix} \\ &= \sum_{h=1}^{\infty} \lambda_h(\theta) (I_{d^2(p+q)} \otimes e_{t-h}(\theta)). \end{aligned}$$

Hence, at $\theta = \theta_0$ we have

$$\frac{\partial e_t}{\partial \theta'} = \sum_{i \geq 1} \lambda_i (I_{d^2(p+q)} \otimes e_{t-i}).$$

It thus remains to prove the following two Lemmas. \square

Lemma 3.3. *We have*

$$\frac{\partial e_t(\theta)}{\partial \mathbf{a}'} = - \sum_{h=0}^{\infty} \mathbf{A}_{\theta,h}^* (I_{d^2 p} \otimes e_{t-h}(\theta)),$$

where $\mathbf{A}_{\theta,h}^* = [\mathbf{A}_{h-1}^* : \mathbf{A}_{h-2}^* : \dots : \mathbf{A}_{h-p}^*]$ is a $d \times d^3 p$ matrix.

Proof of Lemma 3.3. Differentiating the two terms of the following equality

$$\mathbf{A}_\theta(L)X_t = \mathbf{B}_\theta(L)e_t(\theta),$$

with respect to the AR coefficients, we obtain

$$\begin{aligned} \frac{\partial e_t(\theta)}{\partial \mathbf{a}_{ij,\ell}} &= -\mathbf{B}_\theta^{-1}(L)E_{ij}X_{t-\ell} = -\mathbf{B}_\theta^{-1}(L)E_{ij}\mathbf{A}_\theta^{-1}(L)\mathbf{B}_\theta(L)e_{t-\ell}(\theta) \\ &= -M_{ij}(L)e_{t-\ell}(\theta), \quad \ell = 1, \dots, p, \end{aligned}$$

where $E_{ij} = \partial \mathbf{A}_\ell / \partial \mathbf{a}_{ij,\ell}$ is the $d \times d$ matrix with 1 at position (i, j) and 0 elsewhere and

$$e_t(\theta) = \mathbf{B}_\theta^{-1}(L)\mathbf{A}_\theta(L)X_t.$$

Then we have

$$\frac{\partial e_t(\theta)}{\partial \mathbf{a}_{ij,\ell}} = - \sum_{h=0}^{\infty} \mathbf{A}_{ij,h}^* e_{t-\ell-h}(\theta). \quad (3.9)$$

Hence, for any \mathbf{a}_ℓ writing the multivariate noise derivatives

$$\begin{aligned} \frac{\partial e_t(\theta)}{\partial \mathbf{a}'_\ell} &= - \underbrace{[M_{11}(L)e_{t-\ell}(\theta) : M_{21}(L)e_{t-\ell}(\theta) : \dots : M_{dd}(L)e_{t-\ell}(\theta)]}_{d \times d^2} \\ &= - \underbrace{[M_{11}(L) : M_{21}(L) : \dots : M_{dd}(L)]}_{d \times d^3} \underbrace{\begin{bmatrix} e_{t-\ell}(\theta) & 0 & 0 & 0 \\ 0 & e_{t-\ell}(\theta) & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e_{t-\ell}(\theta) \end{bmatrix}}_{d^3 \times d^2} \\ &= - [M_{11}(L) : M_{21}(L) : \dots : M_{dd}(L)] (I_{d^2} \otimes e_{t-\ell}(\theta)) \\ &= -M(L) (I_{d^2} \otimes e_{t-\ell}(\theta)) = -M(L)\mathbf{e}_{t-\ell}(\theta) = V_{t-\ell}(\theta), \end{aligned}$$

where $M(L) = [M_{11}(L) : M_{21}(L) : \dots : M_{dd}(L)]$, $\mathbf{e}_{t-\ell}(\theta) = I_{d^2} \otimes e_{t-\ell}(\theta)$ and $V_{t-\ell}(\theta)$ are respectively the $d \times d^3$, $d^3 \times d^2$ and $d \times d^2$ matrices. Hence, we have

$$\begin{aligned} \frac{\partial e_t(\theta)}{\partial \mathbf{a}'_\ell} &= - \sum_{h=0}^{\infty} \mathbf{A}_h^* \mathbf{e}_{t-\ell-h}(\theta) = - \sum_{k=\ell=0}^{\infty} \mathbf{A}_{k-\ell}^* \mathbf{e}_{t-k}(\theta) = - \sum_{k=\ell}^{\infty} \mathbf{A}_{k-\ell}^* \mathbf{e}_{t-k}(\theta) \\ &= - \sum_{k=0}^{\infty} \mathbf{A}_{k-\ell}^* \mathbf{e}_{t-k}(\theta) = V_{t-\ell}(\theta), \end{aligned}$$

with $\mathbf{A}_{k-\ell}^* = 0$ when $k < \ell$. With these notations, we obtain

$$\begin{aligned} \frac{\partial e_t(\theta)}{\partial \mathbf{a}'} &= - \sum_{k=0}^{\infty} \underbrace{[\mathbf{A}_{k-1}^* \mathbf{e}_{t-k}(\theta) : \mathbf{A}_{k-2}^* \mathbf{e}_{t-k}(\theta) : \dots : \mathbf{A}_{k-p}^* \mathbf{e}_{t-k}(\theta)]}_{d \times d^2 p} \\ &= - \sum_{k=0}^{\infty} \underbrace{[\mathbf{A}_{k-1}^* : \mathbf{A}_{k-2}^* : \dots : \mathbf{A}_{k-p}^*]}_{d \times d^3 p} \underbrace{\begin{bmatrix} \mathbf{e}_{t-k}(\theta) & 0 & 0 & 0 \\ 0 & \mathbf{e}_{t-k}(\theta) & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{e}_{t-k}(\theta) \end{bmatrix}}_{d^3 p \times d^2 p} \\ &= - \sum_{k=0}^{\infty} [\mathbf{A}_{k-1}^* : \mathbf{A}_{k-2}^* : \dots : \mathbf{A}_{k-p}^*] (I_p \otimes \mathbf{e}_{t-k}(\theta)) \\ &= - \sum_{k=0}^{\infty} \mathbf{A}_{\theta,k}^* (I_{d^2 p} \otimes e_{t-k}(\theta)), \end{aligned}$$

where $\mathbf{A}_{\theta,k}^* = [\mathbf{A}_{k-1}^* : \mathbf{A}_{k-2}^* : \dots : \mathbf{A}_{k-p}^*]$ is the $d \times d^3 q$ matrix. The conclusion follows. \square

Lemma 3.4. *We have*

$$\frac{\partial e_t(\theta)}{\partial \mathbf{b}'} = \sum_{h=0}^{\infty} \mathbf{B}_{\theta,h}^* (I_{d^2 q} \otimes e_{t-h}(\theta)),$$

where $\mathbf{B}_{\theta,h}^* = [\mathbf{B}_{h-1}^* : \mathbf{B}_{h-2}^* : \dots : \mathbf{B}_{h-q}^*]$ is a $d \times d^3 q$ matrix.

Proof of Lemma 3.4. The proof is similar to that given in Lemma 3.3. Differentiating the two terms of the following equality

$$\mathbf{A}_\theta(L)X_t = \mathbf{B}_\theta(L)e_t(\theta),$$

with respect to the MA coefficients, we obtain

$$\begin{aligned} \frac{\partial e_t(\theta)}{\partial \mathbf{b}_{ij,\ell'}} &= \mathbf{B}_\theta^{-1}(L)E_{ij}\mathbf{B}_\theta^{-1}(L)\mathbf{A}_\theta(L)X_{t-\ell'} = \mathbf{B}_\theta^{-1}(L)E_{ij}e_{t-\ell'}(\theta) \\ &= N_{ij}(L)e_{t-\ell'}(\theta), \quad \ell' = 1, \dots, q, \end{aligned}$$

where $E_{ij} = \partial \mathbf{B}_{\ell'} / \partial \mathbf{b}_{ij, \ell'}$ is the $d \times d$ matrix with 1 at position (i, j) and 0 elsewhere. We then have

$$\frac{\partial e_t(\theta)}{\partial \mathbf{b}_{ij, \ell'}} = \sum_{h=0}^{\infty} \mathbf{B}_{ij, h}^* e_{t-\ell'-h}(\theta). \quad (3.10)$$

Similarly to Lemma 3.4, we have

$$\begin{aligned} \frac{\partial e_t(\theta)}{\partial \mathbf{b}'_{\ell'}} &= \underbrace{[N_{11}(L)e_{t-\ell'}(\theta) : N_{21}(L)e_{t-\ell'}(\theta) : \dots : N_{dd}(L)e_{t-\ell'}(\theta)]}_{d \times d^2} \\ &= N(L) (I_{d^2} \otimes e_{t-\ell'}(\theta)) = N(L) \mathbf{e}_{t-\ell'}(\theta) = U_{t-\ell'}(\theta), \end{aligned}$$

where $N(L) = [N_{11}(L) : N_{21}(L) : \dots : N_{dd}(L)]$ and $U_{t-\ell'}$ are respectively the $d \times d^3$ and $d \times d^2$ matrices. Then, we have

$$\frac{\partial e_t(\theta)}{\partial \mathbf{b}'_{\ell'}} = \sum_{h=0}^{\infty} \mathbf{B}_h^* \mathbf{e}_{t-\ell'-h}(\theta) = \sum_{k=0}^{\infty} \mathbf{B}_{k-\ell'}^* \mathbf{e}_{t-k}(\theta) = U_{t-\ell'}(\theta),$$

where $\mathbf{B}_{k-\ell'}^* = 0$ when $k < \ell'$. With these notations, we obtain

$$\begin{aligned} \frac{\partial e_t(\theta)}{\partial \mathbf{b}'} &= \sum_{k=0}^{\infty} \underbrace{[\mathbf{B}_{k-1}^* \mathbf{e}_{t-k}(\theta) : \mathbf{B}_{k-2}^* \mathbf{e}_{t-k}(\theta) : \dots : \mathbf{B}_{k-q}^* \mathbf{e}_{t-k}(\theta)]}_{d \times d^2 q} \\ &= \sum_{k=0}^{\infty} \mathbf{B}_{\theta, k}^* (I_{d^2 q} \otimes e_{t-k}(\theta)), \end{aligned}$$

where $\mathbf{B}_{\theta, k}^* = [\mathbf{B}_{k-1}^* : \mathbf{B}_{k-2}^* : \dots : \mathbf{B}_{k-q}^*]$ is the $d \times d^3 q$ matrix. The conclusion follows. \square

Proof of Proposition 2 : Let

$$\frac{\partial}{\partial \theta'} e_t(\theta_0) = \left(\frac{\partial}{\partial \theta_1} e_t(\theta_0), \dots, \frac{\partial}{\partial \theta_{k_0}} e_t(\theta_0) \right).$$

Let

$$\begin{aligned} \tilde{\ell}_n(\theta) &= -\frac{2}{n} \log \tilde{L}_n(\theta) \\ &= \frac{1}{n} \sum_{t=1}^n \{d \log(2\pi) + \log \det \Sigma_e + \tilde{e}'_t(\theta) \Sigma_e^{-1} \tilde{e}_t(\theta)\}. \end{aligned}$$

In Boubacar Mainassara and Francq (2009), it is shown that $\ell_n(\theta) = \tilde{\ell}_n(\theta) + o(1)$ a.s., where

$$\ell_n(\theta) = \frac{1}{n} \sum_{t=1}^n \{d \log(2\pi) + \log \det \Sigma_e + e'_t(\theta) \Sigma_e^{-1} e_t(\theta)\}.$$

It is also shown uniformly in $\theta \in \Theta$ that

$$\frac{\partial \ell_n(\theta)}{\partial \theta} = \frac{\partial \tilde{\ell}_n(\theta)}{\partial \theta} + o(1) \quad a.s.$$

The same equality holds for the second-order derivatives of $\tilde{\ell}_n$. We thus have

$$J = \lim_{n \rightarrow \infty} \mathbf{J}_n \quad a.s., \quad \text{where} \quad \mathbf{J}_n = \frac{\partial^2 \ell_n(\theta_0)}{\partial \theta \partial \theta'}.$$

Using well-known results on matrix derivatives, we have

$$\frac{\partial \ell_n(\theta_0)}{\partial \theta} = \frac{2}{n} \sum_{t=1}^n \left\{ \frac{\partial}{\partial \theta} e'_t(\theta_0) \right\} \Sigma_{e_0}^{-1} e_t(\theta_0). \quad (3.11)$$

In view of (3.11), we have

$$\begin{aligned} \mathbf{J}_n &= \frac{2}{n} \sum_{t=1}^n \left(\frac{\partial e'_t(\theta_0)}{\partial \theta} \Sigma_{e_0}^{-1} \frac{\partial e_t(\theta_0)}{\partial \theta'} + \frac{\partial^2 e'_t(\theta_0)}{\partial \theta \partial \theta'} \Sigma_{e_0}^{-1} e_t(\theta_0) \right) \\ &\rightarrow 2E \left\{ \frac{\partial^2 e'_t(\theta_0)}{\partial \theta \partial \theta'} \right\} \Sigma_{e_0}^{-1} e_t + 2E \left\{ \frac{\partial}{\partial \theta} e'_t(\theta_0) \right\} \Sigma_{e_0}^{-1} \left\{ \frac{\partial}{\partial \theta'} e_t(\theta_0) \right\}, \quad a.s. \end{aligned}$$

by the ergodic theorem. Using the orthogonality between e_t and any linear combination of the past values of e_t , we have $2E \left\{ \frac{\partial^2 e'_t(\theta_0)}{\partial \theta \partial \theta'} \right\} \Sigma_{e_0}^{-1} e_t = 0$. Thus we have

$$J = 2E \left\{ \frac{\partial}{\partial \theta} e'_t(\theta_0) \Sigma_{e_0}^{-1} \frac{\partial}{\partial \theta'} e_t(\theta_0) \right\}.$$

Using $\text{vec}(ABC) = (C' \otimes A) \text{vec}(B)$ with $C = I_{d^2(p+q)}$ and in view of Proposition 1, we obtain

$$\begin{aligned} \text{vec } J &= 2E \left\{ \frac{\partial e'_t(\theta_0)}{\partial \theta} \otimes \frac{\partial e'_t(\theta_0)}{\partial \theta} \right\} \text{vec } \Sigma_{e_0}^{-1} \\ &= 2 \sum_{i \geq 1} E \left\{ (I_{d^2(p+q)} \otimes e'_{t-i}) \lambda'_i \right\} \otimes \left\{ (I_{d^2(p+q)} \otimes e'_{t-i}) \lambda'_i \right\} \text{vec } \Sigma_{e_0}^{-1}. \end{aligned}$$

Using also $AC \otimes BD = (A \otimes B)(C \otimes D)$, we have

$$\begin{aligned} \text{vec } J &= 2 \sum_{i \geq 1} E \left\{ (I_{d^2(p+q)} \otimes e'_t) \otimes (I_{d^2(p+q)} \otimes e'_t) \right\} \{ \lambda'_i \otimes \lambda'_i \} \text{vec } \Sigma_{e_0}^{-1} \\ &= 2 \sum_{i \geq 1} E \left\{ (I_{d^2(p+q)} \otimes e'_t)^{\otimes 2} \right\} \lambda_i'^{\otimes 2} \text{vec } \Sigma_{e_0}^{-1} = 2 \sum_{i \geq 1} \mathcal{M} \lambda_i'^{\otimes 2} \text{vec } \Sigma_{e_0}^{-1}. \end{aligned}$$

The proof is complete. \square

Proof of Proposition 3 : In view of (3.6), let

$$\Upsilon_t = \frac{\partial}{\partial \theta} \mathcal{L}_t(\theta_0) = \left(\frac{\partial}{\partial \theta_1} \mathcal{L}_t(\theta), \dots, \frac{\partial}{\partial \theta_{k_0}} \mathcal{L}_t(\theta) \right)'_{\theta=\theta_0}$$

where

$$\mathcal{L}_t(\theta) = \log \det \Sigma_e + e'_t(\theta) \Sigma_e^{-1} e_t(\theta).$$

We have

$$\Upsilon_t = 2 \frac{\partial e'_t(\theta_0)}{\partial \theta} \Sigma_{e_0}^{-1} e_t(\theta_0) = 2 \left\{ e'_t(\theta_0) \otimes \frac{\partial e'_t(\theta_0)}{\partial \theta} \right\} \text{vec } \Sigma_{e_0}^{-1}.$$

In view of (3.5), we have

$$\begin{aligned} I &= \sum_{h=-\infty}^{+\infty} \text{Cov} \left(2 \left\{ \frac{\partial}{\partial \theta} e'_t(\theta_0) \right\} \Sigma_{e_0}^{-1} e_t, 2 \left\{ \frac{\partial}{\partial \theta} e'_{t-h}(\theta_0) \right\} \Sigma_{e_0}^{-1} e_{t-h} \right) \\ &= \sum_{h=-\infty}^{+\infty} \text{Cov} \left(2 \left\{ e'_t \otimes \frac{\partial e'_t}{\partial \theta} \right\} \text{vec } \Sigma_{e_0}^{-1}, 2 \left\{ e'_{t-h} \otimes \frac{\partial e'_{t-h}}{\partial \theta} \right\} \text{vec } \Sigma_{e_0}^{-1} \right) \\ &= 4 \sum_{h=-\infty}^{+\infty} E \left(\left\{ e'_t \otimes \frac{\partial e'_t}{\partial \theta} \right\} \text{vec } \Sigma_{e_0}^{-1} \left\{ \text{vec } \Sigma_{e_0}^{-1} \right\}' \left\{ e'_{t-h} \otimes \frac{\partial e'_{t-h}}{\partial \theta} \right\}' \right). \end{aligned}$$

Using the elementary relation $\text{vec}(ABC) = (C' \otimes A) \text{vec } B$, we have

$$\text{vec } I = 4 \sum_{h=-\infty}^{+\infty} E \left(\left\{ e'_{t-h} \otimes \frac{\partial e'_{t-h}}{\partial \theta} \right\} \otimes \left\{ e'_t \otimes \frac{\partial e'_t}{\partial \theta} \right\} \right) \text{vec} \left(\text{vec } \Sigma_{e_0}^{-1} \left\{ \text{vec } \Sigma_{e_0}^{-1} \right\}' \right).$$

By Proposition 1, we obtain

$$\begin{aligned} \text{vec } I &= 4 \sum_{h=-\infty}^{+\infty} \sum_{i,j=1}^{+\infty} E \left(\left\{ e'_{t-h} \otimes (I_{d^2(p+q)} \otimes e'_{t-j-h}) \lambda'_j \right\} \right. \\ &\quad \left. \otimes \left\{ e'_t \otimes (I_{d^2(p+q)} \otimes e'_{t-i}) \lambda'_i \right\} \right) \text{vec} \left(\text{vec } \Sigma_{e_0}^{-1} \left\{ \text{vec } \Sigma_{e_0}^{-1} \right\}' \right). \end{aligned}$$

Using $AC \otimes BD = (A \otimes B)(C \otimes D)$, we have

$$\begin{aligned} \text{vec } I &= 4 \sum_{h=-\infty}^{+\infty} \sum_{i,j=1}^{+\infty} E \left(\left\{ e'_{t-h} \otimes (I_{d^2(p+q)} \otimes e'_{t-j-h}) \right\} \left\{ I_d \otimes \lambda'_j \right\} \right. \\ &\quad \left. \otimes \left\{ e'_t \otimes (I_{d^2(p+q)} \otimes e'_{t-i}) \right\} \left\{ I_d \otimes \lambda'_i \right\} \right) \text{vec} \left(\text{vec } \Sigma_{e_0}^{-1} \left\{ \text{vec } \Sigma_{e_0}^{-1} \right\}' \right). \end{aligned}$$

Using also $AC \otimes BD = (A \otimes B)(C \otimes D)$, we have

$$\begin{aligned} \text{vec } I &= 4 \sum_{h=-\infty}^{+\infty} \sum_{i,j=1}^{+\infty} E \left(\left\{ e'_{t-h} \otimes (I_{d^2(p+q)} \otimes e'_{t-j-h}) \right\} \otimes \left\{ e'_t \otimes (I_{d^2(p+q)} \otimes e'_{t-i}) \right\} \right) \\ &\quad \left(\left\{ I_d \otimes \lambda'_j \right\} \otimes \left\{ I_d \otimes \lambda'_i \right\} \right) \text{vec} \left(\text{vec } \Sigma_{e_0}^{-1} \left\{ \text{vec } \Sigma_{e_0}^{-1} \right\}' \right) \\ &= 4 \sum_{i,j=1}^{+\infty} \Gamma(i, j) \left(\left\{ I_d \otimes \lambda'_j \right\} \otimes \left\{ I_d \otimes \lambda'_i \right\} \right) \text{vec} \left(\text{vec } \Sigma_{e_0}^{-1} \left\{ \text{vec } \Sigma_{e_0}^{-1} \right\}' \right), \end{aligned}$$

where

$$\Gamma(i, j) = \sum_{h=-\infty}^{+\infty} E \left(\{e'_{t-h} \otimes (I_{d^2(p+q)} \otimes e'_{t-j-h})\} \otimes \{e'_t \otimes (I_{d^2(p+q)} \otimes e'_{t-i})\} \right).$$

The proof is complete. \square

Proof of Remark 3.1. For $d = 1$, we have

$$\mathcal{M} := E \left\{ (I_{(p+q)} \times e_t)^{\otimes 2} \right\} = \sigma^2 I_{(p+q)^2},$$

where σ^2 is the variance of the univariate process. We also have

$$\begin{aligned} \Gamma(i, j) &= \sum_{h=-\infty}^{+\infty} E \left(\{e_{t-h} e_{t-j-h} I_{(p+q)}\} \otimes \{e_t e_{t-i} I_{(p+q)}\} \right) \\ &= \sum_{h=-\infty}^{+\infty} E (e_t e_{t-i} e_{t-h} e_{t-j-h}) I_{(p+q)^2} = \gamma(i, j) I_{(p+q)^2}. \end{aligned} \quad (3.12)$$

In view of Proposition 2, we have

$$\text{vec } J = 2 \sum_{i \geq 1} \mathcal{M} \{ \lambda'_i \otimes \lambda'_i \} \sigma^{-2}.$$

Replacing \mathcal{M} by $\sigma^2 I_{(p+q)^2}$ in $\text{vec } J$, we have

$$\text{vec } J = 2 \sum_{i \geq 1} \{ \lambda_i \otimes \lambda_i \}'.$$

Using (3.12) and in view of Proposition 3, we have

$$\text{vec } I = \frac{4}{\sigma^4} \sum_{i,j=1}^{+\infty} \Gamma(i, j) \{ \lambda'_j \otimes \lambda'_i \} = \frac{4}{\sigma^4} \sum_{i,j=1}^{+\infty} \gamma(i, j) \{ \lambda_j \otimes \lambda_i \}'.$$

The proof is complete. \square

Proof of Theorem 3.1. For any multiplicative norm, we have

$$\begin{aligned} \left\| \text{vec } J - \text{vec } \hat{J}_n \right\| &\leq 2 \sum_{i \geq 1} \left\{ \left\| \mathcal{M} - \hat{\mathcal{M}}_n \right\| \left\| \lambda_i'^{\otimes 2} \right\| \left\| \text{vec } (\Sigma_{e_0}^{-1}) \right\| \right. \\ &\quad + \left\| \hat{\mathcal{M}}_n \right\| \left\| \lambda_i'^{\otimes 2} - \hat{\lambda}_i'^{\otimes 2} \right\| \left\| \text{vec } \Sigma_{e_0}^{-1} \right\| \\ &\quad \left. + \left\| \hat{\mathcal{M}}_n \right\| \left\| \hat{\lambda}_i'^{\otimes 2} \right\| \left\| \text{vec } \left(\hat{\Sigma}_{e_0}^{-1} - \Sigma_{e_0}^{-1} \right) \right\| \right\}. \end{aligned}$$

The proof will thus follow from Lemmas 3.5, 3.6 and 3.8 below. \square

Lemma 3.5. *Under Assumptions A1–A8, we have*

$$\left\| \text{vec}(\hat{\lambda}_i - \lambda_i(\theta_0)) \right\| \leq K \rho^i \times o_{a.s.}(1) \quad a.s. \quad \text{as } n \rightarrow \infty,$$

where ρ is a constant belonging to $[0, 1[$, and $K > 0$.

Proof of Lemma 3.5. Boubacar Mainassara and Francq (2009) showed the strong consistency of $\hat{\theta}_n$ ($\hat{\theta}_n \rightarrow \theta_0$ a.s. as $n \rightarrow \infty$), which entails

$$\left\| \hat{\theta}_n - \theta_0 \right\| = o_{a.s.}(1). \quad (3.13)$$

We have $\mathbf{A}_{ij,h}^* = O(\rho^h)$ and $\mathbf{B}_{ij,h}^* = O(\rho^h)$ uniformly in $\theta \in \Theta$ for some $\rho \in [0, 1[$. In view of (3.4), we thus have $\sup_{\theta \in \Theta} \|\lambda_h(\theta)\| \leq K \rho^h$. Similarly for any $m \in \{1, \dots, k_0\}$, we have

$$\sup_{\theta \in \Theta} \left\| \frac{\partial \lambda_h(\theta)}{\partial \theta_m} \right\| \leq K \rho^h. \quad (3.14)$$

Using a Taylor expansion of $\text{vec } \hat{\lambda}_i$ about θ_0 , we obtain

$$\text{vec } \hat{\lambda}_i = \text{vec } \lambda_i + \frac{\partial \text{vec } \lambda_i(\theta_n^*)}{\partial \theta'} (\hat{\theta}_n - \theta_0),$$

where θ_n^* is between $\hat{\theta}_n$ and θ_0 . For any multiplicative norm, we have

$$\left\| \text{vec}(\hat{\lambda}_i - \lambda_i) \right\| \leq \left\| \frac{\partial \text{vec } \lambda_i(\theta_n^*)}{\partial \theta'} \right\| \left\| \hat{\theta}_n - \theta_0 \right\|.$$

In view of (3.13) and (3.14), the proof is complete. \square

Lemma 3.6. *Under Assumptions A1–A8, we have*

$$\hat{\mathcal{M}}_n \rightarrow \mathcal{M} \quad a.s. \quad \text{as } n \rightarrow \infty.$$

Proof of Lemma 3.6. We have

$$e_t(\theta) = X_t - \sum_{i=1}^p A_0^{-1} A_i X_{t-i} + \sum_{i=1}^q A_0^{-1} B_i B_0^{-1} A_0 e_{t-i}(\theta) \quad \forall t \in \mathbb{Z}. \quad (3.15)$$

For any $\theta \in \Theta$, let

$$\mathcal{M}_n(\theta) := \frac{1}{n} \sum_{t=1}^n \left\{ (I_{d^2(p+q)} \otimes e_t'(\theta))^{\otimes 2} \right\} \quad \text{and} \quad \mathcal{M}(\theta) := E \left\{ (I_{d^2(p+q)} \otimes e_t'(\theta))^{\otimes 2} \right\}.$$

Now the ergodic theorem shows that almost surely

$$\mathcal{M}_n(\theta) \rightarrow \mathcal{M}(\theta).$$

In view of (3.15), and using **A2** and the compactness of Θ , we have

$$e_t(\theta) = X_t + \sum_{i=1}^{\infty} C_i(\theta) X_{t-i}, \quad \sup_{\theta \in \Theta} \|C_i(\theta)\| \leq K\rho^i.$$

We thus have

$$E \sup_{\theta \in \Theta} \|e_t(\theta)\|^2 < \infty, \quad (3.16)$$

by Assumption **A7**. Now, we will consider the norm defined by $\|Z\|_2 = \sqrt{E \|Z\|^2}$, where Z is a d_1 random vector. In view of Proposition 1, (3.16) and $\sup_{\theta \in \Theta} \|\lambda_h(\theta)\| \leq K\rho^h$, we have

$$\left\| \sup_{\theta \in \Theta} \frac{\partial e_t(\theta)}{\partial \theta'} \right\|_2 \leq \sum_{i \geq 1} \sup_{\theta \in \Theta} \|\lambda_i(\theta)\| \times \left\| \sup_{\theta \in \Theta} \{I_{d^2(p+q)} \otimes e_t(\theta)\} \right\|_2 < \infty. \quad (3.17)$$

Let $e_t = (e_{1t}, \dots, e_{dt})'$. The non zero components of the vector $\text{vec } \mathcal{M}_n(\theta)$ are of the form $n^{-1} \sum_{t=1}^n e_{it}(\theta) e_{jt}(\theta)$, for $(i, j) \in \{1, \dots, d\}^2$. We deduce that the elements of the matrix $\partial \text{vec } \mathcal{M}_n(\theta) / \partial \theta'$ are linear combinations of

$$\frac{2}{n} \sum_{t=1}^n e_{it}(\theta) \frac{\partial e_{jt}(\theta)}{\partial \theta'}.$$

By the Cauchy-Schwartz inequality we have

$$\begin{aligned} \left\| \frac{2}{n} \sum_{t=1}^n \sup_{\theta \in \Theta} \left\{ e_{it}(\theta) \frac{\partial e_{jt}(\theta)}{\partial \theta'} \right\} \right\| &\leq \sqrt{\frac{2}{n} \sum_{t=1}^n \sup_{\theta \in \Theta} \{e_{it}^2(\theta)\} \times \frac{2}{n} \sum_{t=1}^n \left\| \sup_{\theta \in \Theta} \frac{\partial e_{jt}(\theta)}{\partial \theta'} \right\|^2} \\ &\leq 2 \sqrt{\frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \Theta} \|e_t(\theta)\|^2 \times \frac{1}{n} \sum_{t=1}^n \left\| \sup_{\theta \in \Theta} \frac{\partial e_t(\theta)}{\partial \theta'} \right\|^2}. \end{aligned}$$

The ergodic theorem shows that almost surely

$$\lim_{n \rightarrow \infty} \frac{2}{n} \sum_{t=1}^n \left\| \sup_{\theta \in \Theta} \left\{ e_{it}(\theta) \frac{\partial e_{jt}(\theta)}{\partial \theta'} \right\} \right\| \leq 2 \sqrt{E \sup_{\theta \in \Theta} \|e_t(\theta)\|^2 \times E \left\| \sup_{\theta \in \Theta} \frac{\partial e_t(\theta)}{\partial \theta'} \right\|^2}.$$

Now using (3.16) and (3.17), the right-hand side of the inequality is bounded. We then deduce that

$$\sup_{\theta \in \Theta} \left\| \frac{\partial \text{vec } \mathcal{M}_n(\theta)}{\partial \theta'} \right\| = O_{a.s.}(1). \quad (3.18)$$

A Taylor expansion of $\text{vec } \hat{\mathcal{M}}_n$ about θ_0 gives

$$\text{vec } \hat{\mathcal{M}}_n = \text{vec } \mathcal{M}_n + \frac{\partial \text{vec } \mathcal{M}_n(\theta_n^*)}{\partial \theta'} (\hat{\theta}_n - \theta_0),$$

where θ_n^* is between $\hat{\theta}_n$ and θ_0 . Using the strong consistency of $\hat{\theta}_n$ and (3.18), it is easily seen that

$$\text{vec } \hat{\mathcal{M}}_n = \text{vec } \mathcal{M}_n(\hat{\theta}_n) \rightarrow \text{vec } \mathcal{M}(\theta_0) = \text{vec } \mathcal{M}, \quad a.s.$$

The proof is complete. \square

Lemma 3.7. *Under Assumptions A1–A8, we have*

$$\hat{\Sigma}_{e0} \rightarrow \Sigma_{e0} \quad a.s., \quad \text{as } n \rightarrow \infty.$$

Proof of Lemma 3.7 : We have $\hat{\Sigma}_{e0} = \Sigma_n(\hat{\theta}_n)$ with $\Sigma_n(\theta) = n^{-1} \sum_{t=1}^n e_t(\theta)e_t'(\theta)$. By the ergodic theorem

$$\Sigma_n(\theta) \rightarrow \Sigma_e(\theta) := Ee_t(\theta)e_t'(\theta), \quad a.s.$$

Using the elementary relation $\text{vec}(aa') = a \otimes a$, where a is a vector, we have $\text{vec } \hat{\Sigma}_{e0} = n^{-1} \sum_{t=1}^n e_t(\hat{\theta}_n) \otimes e_t(\hat{\theta}_n)$ and $\text{vec } \Sigma_n(\theta_0) = n^{-1} \sum_{t=1}^n e_t(\theta_0) \otimes e_t(\theta_0)$. Using a Taylor expansion of $\text{vec } \hat{\Sigma}_{e0}$ around θ_0 and (3.2), we obtain

$$\text{vec } \hat{\Sigma}_{e0} = \text{vec } \Sigma_n(\theta_0) + \frac{1}{n} \sum_{t=1}^n \left\{ e_t \otimes \frac{\partial e_t}{\partial \theta'} + \frac{\partial e_t}{\partial \theta'} \otimes e_t \right\} (\hat{\theta}_n - \theta_0) + O_P\left(\frac{1}{n}\right).$$

Using the strong consistency of $\hat{\theta}_n$,

$$E \sup_{\theta \in \Theta} \|e_t(\theta)\|^2 < \infty \quad \text{and} \quad \left\| \sup_{\theta \in \Theta} \frac{\partial e_t(\theta)}{\partial \theta'} \right\|_2 < \infty,$$

it is easily seen that

$$\text{vec } \hat{\Sigma}_{e0} \rightarrow \text{vec } \Sigma_e(\theta_0) = \text{vec } \Sigma_{e0}, \quad a.s.$$

The proof is complete. \square

Lemma 3.8. *Under Assumptions A1–A8, we have*

$$\hat{\Sigma}_{e0}^{-1} \rightarrow \Sigma_{e0}^{-1} \quad a.s. \quad \text{as } n \rightarrow \infty.$$

Proof of Lemma 3.8 : For any multiplicative norm, we have

$$\left\| \hat{\Sigma}_{e0}^{-1} - \Sigma_{e0}^{-1} \right\| = \left\| -\hat{\Sigma}_{e0}^{-1} \left(\hat{\Sigma}_{e0} - \Sigma_{e0} \right) \Sigma_{e0}^{-1} \right\| \leq \left\| \hat{\Sigma}_{e0}^{-1} \right\| \left\| \hat{\Sigma}_{e0} - \Sigma_{e0} \right\| \left\| \Sigma_{e0}^{-1} \right\|.$$

In view of Lemma 3.7 and $\left\| \Sigma_{e0}^{-1} \right\| < \infty$ (because the matrix Σ_{e0} is nonsingular), we have

$$\left\| \hat{\Sigma}_{e0}^{-1} - \Sigma_{e0}^{-1} \right\| \rightarrow 0 \quad a.s.$$

The proof is complete. \square

Proof of Theorem 3.2. Let the matrices

$$\begin{aligned}\hat{\Lambda}_{ij} &= \left(\left\{ I_d \otimes \hat{\lambda}'_i \right\} \otimes \left\{ I_d \otimes \hat{\lambda}'_j \right\} \right), \quad \Lambda_{ij} = \left(\left\{ I_d \otimes \lambda'_i \right\} \otimes \left\{ I_d \otimes \lambda'_j \right\} \right), \\ \hat{\Delta}_{e0} &= \left(\text{vec } \hat{\Sigma}_{e0}^{-1} \left\{ \text{vec } \hat{\Sigma}_{e0}^{-1} \right\}' \right) \quad \text{and} \quad \Delta_{e0} = \left(\text{vec } \Sigma_{e0}^{-1} \left\{ \text{vec } \Sigma_{e0}^{-1} \right\}' \right).\end{aligned}$$

For any multiplicative norm, we have

$$\begin{aligned}\left\| \text{vec } I - \text{vec } \hat{I}_n \right\| &\leq 4 \sum_{i,j \geq 1} \left\{ \left\| \Gamma(i, j) - \hat{\Gamma}_n(i, j) \right\| \left\| \Lambda_{ij} \right\| \left\| \text{vec } \Delta_{e0} \right\| \right. \\ &\quad + \left\| \hat{\Gamma}_n(i, j) \right\| \left\| \Lambda_{ij} - \hat{\Lambda}_{ij} \right\| \left\| \text{vec } \Delta_{e0} \right\| \\ &\quad \left. + \left\| \hat{\Gamma}_n(i, j) \right\| \left\| \hat{\Lambda}_{ij} \right\| \left\| \text{vec } \left(\Delta_{e0} - \hat{\Delta}_{e0} \right) \right\| \right\}.\end{aligned}$$

Lemma 3.5 and $\|\lambda_i\| = O(\rho^i)$ entail

$$\begin{aligned}\left\| \text{vec } \hat{\Lambda}_{ij} - \text{vec } \Lambda_{ij} \right\| &\leq \left\| \text{vec } \left\{ I_d \otimes (\hat{\lambda}_i - \lambda_i) \otimes (I_d \otimes \hat{\lambda}_j) \right\} \right\| \\ &\quad + \left\| \text{vec } \left\{ (I_d \otimes \lambda_i) \otimes (I_d \otimes (\hat{\lambda}_j - \lambda_j)) \right\} \right\| \\ &\leq K \rho^{i+j} \times o_{a.s.}(1).\end{aligned}$$

We also have $\|\Lambda_{ij}\| = O(\rho^{i+j})$. In view of Lemma 3.8 and $\|\Sigma_{e0}^{-1}\| < \infty$, we have

$$\begin{aligned}\left\| \hat{\Delta}_{e0} - \Delta_{e0} \right\| &\leq \left\| \text{vec } \left(\hat{\Sigma}_{e0}^{-1} - \Sigma_{e0}^{-1} \right) \right\| \left\| \left\{ \text{vec } \hat{\Sigma}_{e0}^{-1} \right\}' \right\| \\ &\quad + \left\| \text{vec } \Sigma_{e0}^{-1} \right\| \left\| \left\{ \text{vec } \left(\hat{\Sigma}_{e0}^{-1} - \Sigma_{e0}^{-1} \right) \right\}' \right\| \rightarrow 0 \quad a.s.\end{aligned}$$

In the univariate case Francq, Roy and Zakoïan (2005) showed (see the proofs of their Lemmas A.1 and A.3) that $\sup_{\ell, \ell' > 0} |\Gamma(\ell, \ell')| < \infty$. This can be directly extended in the multivariate case. The non zero elements of $\Gamma(i, j)$ are of the form

$$\sum_{h=-\infty}^{\infty} E(e_{i_1 t} e_{i_2 t-i} e_{i_3 t-h} e_{i_4 t-h-j}), \quad \text{for } (i_1, i_2, i_3, i_4) \in \{1, \dots, d\}^4.$$

We thus have

$$\sup_{i,j \geq 1} \left| \sum_{h=-\infty}^{\infty} E(e_{i_1 t} e_{i_2 t-i} e_{i_3 t-h} e_{i_4 t-h-j}) \right| \leq \sup_{i,j \geq 1} \|\Gamma(i, j)\| < \infty. \quad (3.19)$$

We then deduce that $\sup_{i,j \geq 1} \|\Gamma(i, j)\| = O(1)$. The proof will thus follow from Lemma 3.9 below, in which we show the consistency of $\hat{\Gamma}_n(i, j)$ uniformly in i and j . \square

Lemma 3.9. *Under Assumptions A1–A8, we have*

$$\sup_{i,j} \left\| \hat{\Gamma}_n(i, j) - \Gamma(i, j) \right\| \rightarrow 0 \text{ in probability as } n \rightarrow \infty.$$

Proof of Lemma 3.9. For any $\theta \in \Theta$, let

$$\begin{aligned} \mathcal{M}_{n \ ij, h}(\theta) &:= \frac{1}{n} \sum_{t=1}^{n-|h|} \left(\{e'_{t-h}(\theta) \otimes (I_{d^2(p+q)} \otimes e'_{t-j-h}(\theta))\} \right. \\ &\quad \left. \otimes \{e'_t(\theta) \otimes (I_{d^2(p+q)} \otimes e'_{t-i}(\theta))\} \right). \end{aligned}$$

By the ergodic theorem, we have

$$\begin{aligned} \mathcal{M}_{n \ ij, h}(\theta) &\rightarrow \mathcal{M}_{ij, h}(\theta) := E \left(\{e'_{t-h}(\theta) \otimes (I_{d^2(p+q)} \otimes e'_{t-j-h}(\theta))\} \right. \\ &\quad \left. \otimes \{e'_t(\theta) \otimes (I_{d^2(p+q)} \otimes e'_{t-i}(\theta))\} \right) \text{ a.s.} \end{aligned}$$

A Taylor expansion of $\text{vec } \hat{\mathcal{M}}_{n \ ij, h}$ around θ_0 and (3.2) give

$$\text{vec } \hat{\mathcal{M}}_{n \ ij, h} = \text{vec } \mathcal{M}_{n \ ij, h} + \left\{ \frac{\partial \text{vec } \mathcal{M}_{n \ ij, h}}{\partial \theta'} \right\}_{\theta_0} (\hat{\theta}_n - \theta_0) + O_P \left(\frac{1}{n} \right).$$

In view of (3.19), we then deduce that

$$\limsup_{n \rightarrow \infty} \sup_{i, j \geq 1} \sup_{|h| < n} \left\| \frac{\partial \text{vec } \mathcal{M}_{n \ ij, h}}{\partial \theta'} \right\| < \infty, \quad \text{a.s.} \quad (3.20)$$

By the ergodic theorem, (3.2) and (3.20), for any multiplicative norm, we have

$$\begin{aligned} \sup_{i, j \geq 1} \sup_{|h| < n} \left\| \text{vec} \left(\hat{\mathcal{M}}_{n \ ij, h} - \mathcal{M}_{ij, h} \right) \right\| &\leq \limsup_{n \rightarrow \infty} \sup_{i, j \geq 1} \sup_{|h| < n} \left\| \frac{\partial \text{vec } \mathcal{M}_{n \ ij, h}}{\partial \theta'} \right\| \left\| \hat{\theta}_n - \theta_0 \right\| \\ &\quad + O_P \left(\frac{1}{n} \right) = O_P \left(\frac{1}{\sqrt{n}} \right). \end{aligned} \quad (3.21)$$

We have

$$\begin{aligned} \hat{\Gamma}_n(i, j) - \Gamma(i, j) &= \sum_{h=-T_n}^{+T_n} f(hb_n) \left(\hat{\mathcal{M}}_{n \ ij, h} - \mathcal{M}_{ij, h} \right) \\ &\quad + \sum_{h=-T_n}^{+T_n} \{f(hb_n) - 1\} \mathcal{M}_{ij, h} - \sum_{|h| > T_n} \mathcal{M}_{ij, h}. \end{aligned}$$

By the triangular inequality, for any multiplicative norm, we have

$$\sup_{i, j \geq 1} \left\| \hat{\Gamma}_n(i, j) - \Gamma(i, j) \right\| \leq g_1 + g_2 + g_3,$$

where

$$g_1 = \sup_{i,j \geq 1} \sup_{|h| < n} \left\| \hat{\mathcal{M}}_{n \ i j, h} - \mathcal{M}_{ij, h} \right\| \sum_{|h| \leq T_n} |f(hb_n)|,$$

$$g_2 = \sum_{|h| \leq T_n} |f(hb_n) - 1| \|\mathcal{M}_{ij, h}\| \quad \text{and} \quad g_3 = \sum_{|h| > T_n} \|\mathcal{M}_{ij, h}\|.$$

The non zero elements of $\mathcal{M}_{ij, h}$ are of the form $E(e_{i_1 t} e_{i_2 t-i} e_{i_3 t-h} e_{i_4 t-h-j})$, with $(i_1, i_2, i_3, i_4) \in \{1, \dots, d\}^4$. Now, using the covariance inequality obtained by Davydov (1968), it is easy to show that

$$\begin{aligned} |E(e_{i_1 t} e_{i_2 t-i} e_{i_3 t-h} e_{i_4 t-h-j})| &= |\text{Cov}(e_{i_1 t} e_{i_2 t-i}, e_{i_3 t-h} e_{i_4 t-h-j})| \\ &\leq K \alpha_\epsilon^{\nu/(2+\nu)}(h). \end{aligned}$$

We then deduce that

$$\|\mathcal{M}_{ij, h}\| \leq K \alpha_\epsilon^{\nu/(2+\nu)}(h). \quad (3.22)$$

In view of **A7**, we thus have $g_3 \rightarrow 0$ as $n \rightarrow \infty$. Let m be a fixed integer and write $g_2 \leq s_1 + s_2$, where

$$s_1 = \sum_{|h| \leq m} |f(hb_n) - 1| \|\mathcal{M}_{ij, h}\| \quad \text{and} \quad s_2 = \sum_{m < |h| \leq T_n} |f(hb_n) - 1| \|\mathcal{M}_{ij, h}\|.$$

For $|h| \leq m$, we have $hb_n \rightarrow 0$ as $n \rightarrow \infty$ and $f(hb_n) \rightarrow 1$, it follows that $s_1 \rightarrow 0$. If we choose m sufficiently large, s_2 becomes small, using (3.22) and the fact that $f(\cdot)$ is bounded. It follows that $g_2 \rightarrow 0$. In view of (3.8) and (3.21), we have

$$g_1 \leq \frac{1}{b_n} \sup_{i,j \geq 1} \sup_{|h| < n} \left\| \hat{\mathcal{M}}_{n \ i j, h} - \mathcal{M}_{ij, h} \right\| O(1) = O_P\left(\frac{1}{b_n \sqrt{n}}\right) = o_p(1),$$

since $nb_n^2 \rightarrow \infty$, in view of (3.7). The conclusion is complete. \square

3.7 References

- Andrews, D.W.K.** (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- Boubacar Mainassara, Y. and Francq, C.** (2009) Estimating structural VARMA models with uncorrelated but non-independent error terms. *Working Papers*, <http://perso.univ-lille3.fr/~cfrancq/pub.html>.
- Brockwell, P. J. and Davis, R. A.** (1991) *Time series : theory and methods*. Springer Verlag, New York.
- Davydov, Y. A.** (1968) Convergence of Distributions Generated by Stationary Stochastic Processes. *Theory of Probability and Applications* 13, 691–696.
- Dufour, J-M., and Pelletier, D.** (2005) Practical methods for modelling weak VARMA processes : identification, estimation and specification with a macroeconomic application. *Technical report, Département de sciences économiques and CIREQ, Université de Montréal, Montréal, Canada*.
- Francq, C. and Raïssi, H.** (2006) Multivariate Portmanteau Test for Autoregressive Models with Uncorrelated but Nonindependent Errors, *Journal of Time Series Analysis* 28, 454–470.
- Francq, C., Roy, R. and Zakoïan, J-M.** (2005) Diagnostic checking in ARMA Models with Uncorrelated Errors, *Journal of the American Statistical Association* 100, 532–544.
- Francq, and Zakoïan, J-M.** (2005) Recent results for linear time series models with non independent innovations. In *Statistical Modeling and Analysis for Complex Data Problems*, Chap. 12 (eds P. DUCHESNE and B. RÉMILLARD). New York : Springer Verlag, 137–161.
- Hannan, E. J. and Rissanen** (1982) Recursive estimation of mixed of Autoregressive Moving Average order, *Biometrika* 69, 81–94.
- Lütkepohl, H.** (1993) *Introduction to multiple time series analysis*. Springer Verlag, Berlin.
- Lütkepohl, H.** (2005) *New introduction to multiple time series analysis*. Springer Verlag, Berlin.
- Magnus, J.R. and H. Neudecker** (1988) *Matrix Differential Calculus with Application in Statistics and Econometrics*. New-York, Wiley.
- Newey, W.K., and West, K.D.** (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- Reinsel, G. C.** (1997) *Elements of multivariate time series Analysis*. Second edition. Springer Verlag, New York.

3.8 Verification of expression of I and J on Examples

The following examples illustrate how the matrices I and J depend on θ_0 and terms involving the distribution of the innovations e_t .

Example 3.1. Consider for instance a univariate ARMA(1,1) of the form $X_t = aX_{t-1} - be_{t-1} + \epsilon_t$, with variance σ^2 . Then, with our notations we have $\theta = (a, b)'$, $\theta_0 = (a_0, b_0)'$, $\mathbf{A}_i^* = a^i$, $\mathbf{B}_i^* = b^i$, $\lambda_i = (\mathbf{A}_{i-1}^*, \mathbf{B}_{i-1}^*) = (a^{i-1}, b^{i-1})$, $E\epsilon_t^2 = \sigma^2$, $\Gamma(i, j) = \sum_{h=-\infty}^{+\infty} E(\epsilon_t \epsilon_{t-i} \epsilon_{t-h} \epsilon_{t-j-h}) I_4 = \gamma(i, j) I_4$ and

$$\lambda_i \otimes \lambda_i = (a^{2(i-1)}, (ab)^{i-1}, (ab)^{i-1}, b^{2(i-1)}).$$

Thus, we have

$$\begin{aligned} \text{vec } J &= 2 \sum_{i \geq 1} \left(a_0^{2(i-1)}, (a_0 b_0)^{i-1}, (a_0 b_0)^{i-1}, b_0^{2(i-1)} \right)' \quad \text{and} \\ \text{vec } I &= \frac{4}{\sigma_0^4} \sum_{i, j=1}^{+\infty} \gamma(i, j) \left(a_0^{2(i-1)}, (a_0 b_0)^{i-1}, (a_0 b_0)^{i-1}, b_0^{2(i-1)} \right)', \end{aligned}$$

where σ_0 is the true value of σ . We then deduce that

$$J = 2 \begin{bmatrix} \frac{1}{1-a_0^2} & \frac{1}{1-a_0 b_0} \\ \frac{1}{1-a_0 b_0} & \frac{1}{1-b_0^2} \end{bmatrix} \quad \text{and} \quad I = \frac{4}{\sigma_0^4} \sum_{i, j=1}^{+\infty} \gamma(i, j) \begin{bmatrix} \frac{1}{1-a_0^2} & \frac{1}{1-a_0 b_0} \\ \frac{1}{1-a_0 b_0} & \frac{1}{1-b_0^2} \end{bmatrix}.$$

Example 3.2. Now we consider a bivariate VAR(1) of the form

$$X_t = \begin{pmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{pmatrix} X_{t-1} + e_t, \quad \Sigma_e = \begin{pmatrix} \sigma_{11}^2 & 0 \\ 0 & \sigma_{22}^2 \end{pmatrix}.$$

With our notations, we have $\theta = (\alpha_1, \alpha_2)'$, $\theta_0 = (\alpha_{01}, \alpha_{02})'$,

$$\mathbf{A}(L) = \begin{pmatrix} 1 - \alpha_1 L & 0 \\ 0 & 1 - \alpha_2 L \end{pmatrix}, \quad M_{ij}(z) = \sum_{h=0}^{\infty} E_{ij} \begin{pmatrix} \alpha_1^h & 0 \\ 0 & \alpha_2^h \end{pmatrix} z^h \quad i, j = 1, 2,$$

$$\mathbf{A}_{11, h}^* = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha_1^h & 0 \\ 0 & \alpha_2^h \end{pmatrix} = \begin{pmatrix} \alpha_1^h & 0 \\ 0 & 0 \end{pmatrix},$$

$$\mathbf{A}_{12, h}^* = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha_1^h & 0 \\ 0 & \alpha_2^h \end{pmatrix} = \begin{pmatrix} 0 & \alpha_2^h \\ 0 & 0 \end{pmatrix},$$

$$\mathbf{A}_{21, h}^* = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \alpha_1^h & 0 \\ 0 & \alpha_2^h \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ \alpha_1^h & 0 \end{pmatrix},$$

$$\mathbf{A}_{22, h}^* = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_1^h & 0 \\ 0 & \alpha_2^h \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \alpha_2^h \end{pmatrix}.$$

We also have

$$\mathbf{A}_h^* = [\mathbf{A}_{11,h}^* : \mathbf{A}_{21,h}^* : \mathbf{A}_{12,h}^* : \mathbf{A}_{22,h}^*] = \begin{bmatrix} \alpha_1^h & 0 & 0 & \alpha_2^h & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \alpha_1^h & 0 & 0 & \alpha_2^h \end{bmatrix},$$

$\lambda_h = [\mathbf{A}_{h-1}^*]$. Let $\mathbf{\Lambda}_h := \lambda_h \otimes \lambda_h$ a 4×64 matrix defined, with obvious notation, by

$$\begin{aligned} \mathbf{\Lambda}_h(1, 1) &= \mathbf{\Lambda}_h(2, 5) = \mathbf{\Lambda}_h(3, 33) = \mathbf{\Lambda}_h(4, 37) = \alpha_1^{2(h-1)}, \\ \mathbf{\Lambda}_h(1, 28) &= \mathbf{\Lambda}_h(2, 32) = \mathbf{\Lambda}_h(3, 60) = \mathbf{\Lambda}_h(4, 64) = \alpha_2^{2(h-1)}, \\ \mathbf{\Lambda}_h(1, 4) &= \mathbf{\Lambda}_h(1, 25) = \mathbf{\Lambda}_h(2, 8) = \mathbf{\Lambda}_h(2, 29) = \mathbf{\Lambda}_h(3, 36) \\ &= \mathbf{\Lambda}_h(3, 57) = \mathbf{\Lambda}_h(4, 40) = \mathbf{\Lambda}_h(4, 61) = (\alpha_1 \alpha_2)^{(h-1)} \end{aligned}$$

and 0 elsewhere. The matrix \mathcal{M} defined before Proposition 2 is a 16×64 matrix

$$\mathcal{M} = \begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} & \mathcal{M}_{13} & \mathcal{M}_{14} & \mathcal{M}_{15} & \mathcal{M}_{16} & \mathcal{M}_{17} & \mathcal{M}_{18} \\ \mathcal{M}_{21} & \mathcal{M}_{22} & \mathcal{M}_{23} & \mathcal{M}_{24} & \mathcal{M}_{25} & \mathcal{M}_{26} & \mathcal{M}_{27} & \mathcal{M}_{28} \\ \mathcal{M}_{31} & \mathcal{M}_{32} & \mathcal{M}_{33} & \mathcal{M}_{34} & \mathcal{M}_{35} & \mathcal{M}_{36} & \mathcal{M}_{37} & \mathcal{M}_{38} \\ \mathcal{M}_{41} & \mathcal{M}_{42} & \mathcal{M}_{43} & \mathcal{M}_{44} & \mathcal{M}_{45} & \mathcal{M}_{46} & \mathcal{M}_{47} & \mathcal{M}_{48} \end{bmatrix},$$

where for $(i_1, i_2) \in \{1, \dots, 4\} \times \{1, \dots, 8\}$ and $j_1 \in \{1, 2\}$, we have $\mathcal{M}_{i_1 i_2} \neq 0$ if $i_2 = 2(i_1 - 1) + j_1$ and $\mathcal{M}_{i_1 i_2} = 0$ otherwise. For $(k, k') \in \{1, \dots, 4\} \times \{1, \dots, 8\}$ and $j_2 \in \{1, 2\}$, the (k, k') -th element of $\mathcal{M}_{i_1 i_2}$ is of the form $\sigma_{j_1 j_2}^2$ if $k' = 2(k - 1) + j_2$ and 0 otherwise. Because $\sigma_{12}^2 = \sigma_{21}^2 = 0$, we thus have

$$\begin{aligned} \mathcal{M}_{11} = \mathcal{M}_{23} = \mathcal{M}_{35} = \mathcal{M}_{47} &= \begin{bmatrix} \sigma_{11}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{11}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{11}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{11}^2 & 0 \end{bmatrix}, \\ \mathcal{M}_{12} = \mathcal{M}_{24} = \mathcal{M}_{36} = \mathcal{M}_{48} &= \begin{bmatrix} 0 & \sigma_{22}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{22}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{22}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{22}^2 \end{bmatrix}, \end{aligned}$$

and all the other $\mathcal{M}_{i_1 i_2}$ matrices are equal to zero. We then have

$$\mathcal{M} \{ \lambda_h \otimes \lambda_h \}' = \begin{pmatrix} \sigma_{11}^2 \alpha_1^{2(h-1)} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \sigma_{11}^2 \alpha_1^{2(h-1)} & 0 & 0 \\ 0_{2 \times 4} & & & \\ \sigma_{22}^2 \alpha_2^{2(h-1)} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \sigma_{22}^2 \alpha_2^{2(h-1)} & 0 & 0 \\ 0 & 0 & \sigma_{11}^2 \alpha_1^{2(h-1)} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{11}^2 \alpha_1^{2(h-1)} \\ 0_{2 \times 4} & & & \\ 0 & 0 & \sigma_{22}^2 \alpha_2^{2(h-1)} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{22}^2 \alpha_2^{2(h-1)} \end{pmatrix},$$

and

$$\mathcal{M} \Lambda_h \text{vec} \Sigma_{\varepsilon 0}^{-1} = \left(\alpha_{01}^{2(h-1)}, 0_4', \sigma_{22}^2 \sigma_{11}^{-2} \alpha_{02}^{2(h-1)}, 0_4', \sigma_{11}^2 \sigma_{22}^{-2} \alpha_{01}^{2(h-1)}, 0_4', \alpha_{02}^{2(h-1)} \right)',$$

which entails that

$$\text{vec} J = 2 \left(\frac{1}{1 - \alpha_{01}^2}, 0_4', \frac{\sigma_{22}^2 \sigma_{11}^{-2}}{1 - \alpha_{02}^2}, 0_4', \frac{\sigma_{11}^2 \sigma_{22}^{-2}}{1 - \alpha_{01}^2}, 0_4', \frac{1}{1 - \alpha_{02}^2} \right)'.$$

Thus we have

$$J = 2 \begin{pmatrix} \frac{1}{1 - \alpha_{01}^2} & 0 & 0 & 0 \\ 0 & \frac{\sigma_{22}^2 \sigma_{11}^{-2}}{1 - \alpha_{02}^2} & 0 & 0 \\ 0 & 0 & \frac{\sigma_{11}^2 \sigma_{22}^{-2}}{1 - \alpha_{01}^2} & 0 \\ 0 & 0 & 0 & \frac{1}{1 - \alpha_{02}^2} \end{pmatrix}. \quad (3.23)$$

It remains to obtain the expression of I . The $\mathcal{M}_{ij,h}$ matrix introduced before Proposition 3 is a 16×256 matrix of the form

$$\mathcal{M}_{ij,h} = \begin{bmatrix} \mathcal{M}_{11,ij,h} & \mathcal{M}_{12,ij,h} & \dots & \mathcal{M}_{18,ij,h} & \vdots & \mathcal{M}_{19,ij,h} & \dots & \mathcal{M}_{16,ij,h} \\ \mathcal{M}_{21,ij,h} & \mathcal{M}_{22,ij,h} & \vdots & \mathcal{M}_{28,ij,h} & \vdots & \mathcal{M}_{29,ij,h} & \vdots & \mathcal{M}_{26,ij,h} \\ \mathcal{M}_{31,ij,h} & \mathcal{M}_{32,ij,h} & \vdots & \mathcal{M}_{38,ij,h} & \vdots & \mathcal{M}_{39,ij,h} & \vdots & \mathcal{M}_{36,ij,h} \\ \mathcal{M}_{41,ij,h} & \mathcal{M}_{42,ij,h} & \dots & \mathcal{M}_{18,ij,h} & \vdots & \mathcal{M}_{49,ij,h} & \dots & \mathcal{M}_{46,ij,h} \end{bmatrix},$$

where the $\mathcal{M}_{i_1 i_2, ij, h}$'s are 4×16 matrices. For $(i_1', i_2') \in \{1, \dots, 4\} \times \{1, \dots, 16\}$ and $j_1' \in \{1, 2\}$, we have $\mathcal{M}_{i_1' i_2', ij, h} \neq 0_{4 \times 16}$ if $i_2' = 2(i_1' - 1) + j_1'$ and $i_2' = 8 + 2(i_1' - 1) + j_1'$, and $\mathcal{M}_{i_1' i_2', ij, h} = 0_{4 \times 16}$ otherwise. For $j_2' \in \{1, 2\}$, we have $\mathcal{M}_{11, ij, h} = \mathcal{M}_{23, ij, h} = \mathcal{M}_{35, ij, h} = \mathcal{M}_{47, ij, h}$ where the (k, k') -th element is of the form $E e_{1t-h} e_{1t-j-h} e_{1t} e_{j_2' t-i}$

when $k' = 2(k - 1) + j'_2$ and of the form $Ee_{1t-h}e_{1t-j-h}e_{2t}e_{j'_2t-i}$ when $k' = 8 + 2(k - 1) + j'_2$. Similarly we have $\mathcal{M}_{12,ij,h} = \mathcal{M}_{24,ij,h} = \mathcal{M}_{36,ij,h} = \mathcal{M}_{48,ij,h}$ where the (k, k') -th element is of the form $Ee_{1t-h}e_{2t-j-h}e_{1t}e_{j'_2t-i}$ when $k' = 2(k - 1) + j'_2$ and of the form $Ee_{1t-h}e_{2t-j-h}e_{2t}e_{j'_2t-i}$ when $k' = 8 + 2(k - 1) + j'_2$. We also have $\mathcal{M}_{19,ij,h} = \mathcal{M}_{211,ij,h} = \mathcal{M}_{313,ij,h} = \mathcal{M}_{415,ij,h}$ where the (k, k') -th element is of the form $Ee_{2t-h}e_{1t-j-h}e_{1t}e_{j'_2t-i}$ when $k' = 2(k - 1) + j'_2$ and of the form $Ee_{2t-h}e_{1t-j-h}e_{2t}e_{j'_2t-i}$ when $k' = 8 + 2(k - 1) + j'_2$. We have $\mathcal{M}_{110,ij,h} = \mathcal{M}_{212,ij,h} = \mathcal{M}_{314,ij,h} = \mathcal{M}_{416,ij,h}$ where the (k, k') -th element is of the form $Ee_{2t-h}e_{2t-j-h}e_{1t}e_{j'_2t-i}$ when $k' = 2(k - 1) + j'_2$ and of the form $Ee_{2t-h}e_{2t-j-h}e_{2t}e_{j'_2t-i}$ when $k' = 8 + 2(k - 1) + j'_2$. We have the 16×4 matrix

$$I_2 \otimes \lambda'_h = \begin{pmatrix} \alpha_1^{h-1} & 0 & 0 & 0 \\ 0_{2 \times 4} & & & \\ \alpha_2^{h-1} & 0 & 0 & 0 \\ 0 & \alpha_1^{h-1} & 0 & 0 \\ 0_{2 \times 4} & & & \\ 0 & \alpha_2^{h-1} & 0 & 0 \\ 0 & 0 & \alpha_1^{h-1} & 0 \\ 0_{2 \times 4} & & & \\ 0 & 0 & \alpha_2^{h-1} & 0 \\ 0 & 0 & 0 & \alpha_1^{h-1} \\ 0_{2 \times 4} & & & \\ 0 & 0 & 0 & \alpha_2^{h-1} \end{pmatrix}$$

and let the 256×16 matrix $F_{ij} = (I_2 \otimes \lambda'_i) \otimes (I_2 \otimes \lambda'_j)$ defined, by

$$\begin{aligned}
 F_{ij}(1, 1) &= F_{ij}(5, 2) = F_{ij}(9, 3) = F_{ij}(13, 4) = F_{ij}(65, 5) = F_{ij}(69, 6) = F_{ij}(73, 7) \\
 &= F_{ij}(77, 8) = F_{ij}(129, 9) = F_{ij}(133, 10) = F_{ij}(137, 11) = F_{ij}(141, 12) \\
 &= F_{ij}(193, 13) = F_{ij}(197, 14) = F_{ij}(201, 15) = F_{ij}(205, 16) = \alpha_1^{i-1} \alpha_1^{j-1}, \\
 F_{ij}(52, 1) &= F_{ij}(56, 2) = F_{ij}(60, 3) = F_{ij}(64, 4) = F_{ij}(116, 5) = F_{ij}(120, 6) \\
 &= F_{ij}(124, 7) = F_{ij}(128, 8) = F_{ij}(180, 9) = F_{ij}(184, 10) = F_{ij}(188, 11) \\
 &= F_{ij}(192, 12) = F_{ij}(244, 13) = F_{ij}(248, 14) = F_{ij}(252, 15) \\
 &= F_{ij}(256, 16) = \alpha_2^{i-1} \alpha_2^{j-1},
 \end{aligned}$$

$$\begin{aligned}
 F_{ij}(4, 1) &= F_{ij}(8, 2) = F_{ij}(12, 3) = F_{ij}(16, 4) = F_{ij}(68, 5) = F_{ij}(72, 6) = F_{ij}(76, 7) \\
 &= F_{ij}(80, 8) = F_{ij}(132, 9) = F_{ij}(136, 10) = F_{ij}(140, 11) = F_{ij}(144, 12) \\
 &= F_{ij}(196, 13) = F_{ij}(200, 14) = F_{ij}(204, 15) = F_{ij}(208, 16) = \alpha_1^{i-1} \alpha_2^{j-1}, \\
 F_{ij}(49, 1) &= F_{ij}(53, 2) = F_{ij}(57, 3) = F_{ij}(61, 4) = F_{ij}(113, 5) = F_{ij}(117, 6) \\
 &= F_{ij}(121, 7) = F_{ij}(125, 8) = F_{ij}(177, 9) = F_{ij}(181, 10) = F_{ij}(185, 11) \\
 &= F_{ij}(189, 12) = F_{ij}(241, 13) = F_{ij}(245, 14) = F_{ij}(249, 15) \\
 &= F_{ij}(253, 16) = \alpha_2^{i-1} \alpha_1^{j-1},
 \end{aligned}$$

and 0 elsewhere. We also define the 16×16 matrix $\Lambda_{ij,h} = \mathcal{M}_{ij,h} F_{ij}$, with obvious notation, by

$$\begin{aligned} \Lambda_{ij,h}(1, 1) &= \Lambda_{ij,h}(3, 2) = \Lambda_{ij,h}(9, 5) = \Lambda_{ij,h}(11, 6) = E e_{1t-h} e_{1t-j-h} e_{1t} e_{1t-i} \alpha_{01}^{i-1} \alpha_{01}^{j-1}, \\ \Lambda_{ij,h}(1, 3) &= \Lambda_{ij,h}(3, 4) = \Lambda_{ij,h}(9, 7) = \Lambda_{ij,h}(11, 8) = E e_{1t-h} e_{1t-j-h} e_{2t} e_{1t-i} \alpha_{01}^{i-1} \alpha_{01}^{j-1}, \\ \Lambda_{ij,h}(2, 1) &= \Lambda_{ij,h}(4, 2) = \Lambda_{ij,h}(10, 5) = \Lambda_{ij,h}(12, 6) = E e_{1t-h} e_{1t-j-h} e_{1t} e_{2t-i} \alpha_{01}^{i-1} \alpha_{02}^{j-1}, \\ \Lambda_{ij,h}(2, 3) &= \Lambda_{ij,h}(4, 4) = \Lambda_{ij,h}(10, 7) = \Lambda_{ij,h}(12, 8) = E e_{1t-h} e_{1t-j-h} e_{2t} e_{2t-i} \alpha_{01}^{i-1} \alpha_{02}^{j-1}, \\ \Lambda_{ij,h}(1, 9) &= \Lambda_{ij,h}(3, 10) = \Lambda_{ij,h}(9, 13) = \Lambda_{ij,h}(11, 14) = E e_{2t-h} e_{1t-j-h} e_{1t} e_{1t-i} \alpha_{01}^{i-1} \alpha_{01}^{j-1}, \\ \Lambda_{ij,h}(1, 11) &= \Lambda_{ij,h}(3, 12) = \Lambda_{ij,h}(9, 15) = \Lambda_{ij,h}(11, 16) = E e_{2t-h} e_{1t-j-h} e_{2t} e_{1t-i} \alpha_{01}^{i-1} \alpha_{01}^{j-1}, \\ \Lambda_{ij,h}(2, 9) &= \Lambda_{ij,h}(4, 10) = \Lambda_{ij,h}(10, 13) = \Lambda_{ij,h}(12, 14) = E e_{2t-h} e_{1t-j-h} e_{1t} e_{2t-i} \alpha_{01}^{i-1} \alpha_{02}^{j-1}, \\ \Lambda_{ij,h}(2, 11) &= \Lambda_{ij,h}(4, 12) = \Lambda_{ij,h}(10, 15) = \Lambda_{ij,h}(12, 16) = E e_{2t-h} e_{1t-j-h} e_{2t} e_{2t-i} \alpha_{01}^{i-1} \alpha_{02}^{j-1}, \end{aligned}$$

$$\begin{aligned} \Lambda_{ij,h}(5, 1) &= \Lambda_{ij,h}(7, 2) = \Lambda_{ij,h}(13, 5) = \Lambda_{ij,h}(15, 6) = E e_{1t-h} e_{2t-j-h} e_{1t} e_{1t-i} \alpha_{02}^{i-1} \alpha_{01}^{j-1}, \\ \Lambda_{ij,h}(5, 3) &= \Lambda_{ij,h}(7, 4) = \Lambda_{ij,h}(13, 7) = \Lambda_{ij,h}(15, 8) = E e_{1t-h} e_{2t-j-h} e_{2t} e_{1t-i} \alpha_{02}^{i-1} \alpha_{01}^{j-1}, \\ \Lambda_{ij,h}(6, 1) &= \Lambda_{ij,h}(8, 2) = \Lambda_{ij,h}(14, 5) = \Lambda_{ij,h}(16, 6) = E e_{1t-h} e_{2t-j-h} e_{1t} e_{2t-i} \alpha_{02}^{i-1} \alpha_{02}^{j-1}, \\ \Lambda_{ij,h}(6, 3) &= \Lambda_{ij,h}(8, 4) = \Lambda_{ij,h}(14, 7) = \Lambda_{ij,h}(16, 8) = E e_{1t-h} e_{2t-j-h} e_{2t} e_{2t-i} \alpha_{02}^{i-1} \alpha_{02}^{j-1}, \\ \Lambda_{ij,h}(5, 9) &= \Lambda_{ij,h}(7, 10) = \Lambda_{ij,h}(13, 13) = \Lambda_{ij,h}(15, 14) = E e_{2t-h} e_{2t-j-h} e_{1t} e_{1t-i} \alpha_{02}^{i-1} \alpha_{01}^{j-1}, \\ \Lambda_{ij,h}(5, 11) &= \Lambda_{ij,h}(7, 12) = \Lambda_{ij,h}(13, 15) = \Lambda_{ij,h}(15, 16) = E e_{2t-h} e_{2t-j-h} e_{2t} e_{1t-i} \alpha_{02}^{i-1} \alpha_{01}^{j-1}, \\ \Lambda_{ij,h}(6, 9) &= \Lambda_{ij,h}(8, 10) = \Lambda_{ij,h}(14, 13) = \Lambda_{ij,h}(16, 14) = E e_{2t-h} e_{2t-j-h} e_{1t} e_{2t-i} \alpha_{02}^{i-1} \alpha_{02}^{j-1}, \\ \Lambda_{ij,h}(6, 11) &= \Lambda_{ij,h}(8, 12) = \Lambda_{ij,h}(14, 15) = \Lambda_{ij,h}(16, 16) = E e_{2t-h} e_{2t-j-h} e_{2t} e_{2t-i} \alpha_{02}^{i-1} \alpha_{02}^{j-1}, \end{aligned}$$

and 0 elsewhere. We then deduce that

$$\Lambda_{ij,h} \text{vec} \left(\text{vec} \Sigma_{e_0}^{-1} \{ \text{vec} \Sigma_{e_0}^{-1} \}' \right) = (\mathcal{I}_{ij,h}(1), \dots, \mathcal{I}_{ij,h}(16))',$$

where

$$\begin{aligned}
 \mathcal{I}_{ij,h}(1) &= Ee_{1t-h}e_{1t-j-h}e_{1t}e_{1t-i}\sigma_{11}^{-4}\alpha_{01}^{i-1}\alpha_{01}^{j-1}, \\
 \mathcal{I}_{ij,h}(2) &= Ee_{1t-h}e_{1t-j-h}e_{1t}e_{2t-i}\sigma_{11}^{-4}\alpha_{01}^{i-1}\alpha_{02}^{j-1}, \\
 \mathcal{I}_{ij,h}(3) &= Ee_{1t-h}e_{1t-j-h}e_{2t}e_{1t-i}\sigma_{11}^{-2}\sigma_{22}^{-2}\alpha_{01}^{i-1}\alpha_{01}^{j-1}, \\
 \mathcal{I}_{ij,h}(4) &= Ee_{1t-h}e_{1t-j-h}e_{2t}e_{2t-i}\sigma_{11}^{-2}\sigma_{22}^{-2}\alpha_{01}^{i-1}\alpha_{02}^{j-1}, \\
 \mathcal{I}_{ij,h}(5) &= Ee_{1t-h}e_{2t-j-h}e_{1t}e_{1t-i}\sigma_{11}^{-4}\alpha_{02}^{i-1}\alpha_{01}^{j-1}, \\
 \mathcal{I}_{ij,h}(6) &= Ee_{1t-h}e_{2t-j-h}e_{1t}e_{2t-i}\sigma_{11}^{-4}\alpha_{02}^{i-1}\alpha_{02}^{j-1}, \\
 \mathcal{I}_{ij,h}(7) &= Ee_{1t-h}e_{2t-j-h}e_{2t}e_{1t-i}\sigma_{11}^{-2}\sigma_{22}^{-2}\alpha_{02}^{i-1}\alpha_{01}^{j-1}, \\
 \mathcal{I}_{ij,h}(8) &= Ee_{1t-h}e_{2t-j-h}e_{2t}e_{2t-i}\sigma_{11}^{-2}\sigma_{22}^{-2}\alpha_{02}^{i-1}\alpha_{02}^{j-1}, \\
 \mathcal{I}_{ij,h}(9) &= Ee_{2t-h}e_{1t-j-h}e_{1t}e_{1t-i}\sigma_{11}^{-2}\sigma_{22}^{-2}\alpha_{01}^{i-1}\alpha_{01}^{j-1}, \\
 \mathcal{I}_{ij,h}(10) &= Ee_{2t-h}e_{1t-j-h}e_{1t}e_{2t-i}\sigma_{11}^{-2}\sigma_{22}^{-2}\alpha_{01}^{i-1}\alpha_{02}^{j-1}, \\
 \mathcal{I}_{ij,h}(11) &= Ee_{2t-h}e_{1t-j-h}e_{2t}e_{1t-i}\sigma_{22}^{-4}\alpha_{01}^{i-1}\alpha_{01}^{j-1}, \\
 \mathcal{I}_{ij,h}(12) &= Ee_{2t-h}e_{1t-j-h}e_{2t}e_{2t-i}\sigma_{22}^{-4}\alpha_{01}^{i-1}\alpha_{02}^{j-1}, \\
 \mathcal{I}_{ij,h}(13) &= Ee_{2t-h}e_{2t-j-h}e_{1t}e_{1t-i}\sigma_{11}^{-2}\sigma_{22}^{-2}\alpha_{02}^{i-1}\alpha_{01}^{j-1}, \\
 \mathcal{I}_{ij,h}(14) &= Ee_{2t-h}e_{2t-j-h}e_{1t}e_{2t-i}\sigma_{11}^{-2}\sigma_{22}^{-2}\alpha_{02}^{i-1}\alpha_{02}^{j-1}, \\
 \mathcal{I}_{ij,h}(15) &= Ee_{2t-h}e_{2t-j-h}e_{2t}e_{1t-i}\sigma_{22}^{-4}\alpha_{02}^{i-1}\alpha_{01}^{j-1}, \\
 \mathcal{I}_{ij,h}(16) &= Ee_{2t-h}e_{2t-j-h}e_{2t}e_{2t-i}\sigma_{22}^{-4}\alpha_{02}^{i-1}\alpha_{02}^{j-1},
 \end{aligned}$$

which entails that

$$\text{vec } I = 4 \sum_{i,j=1}^{+\infty} \sum_{h=-\infty}^{+\infty} (\mathcal{I}_{ij,h}(1), \dots, \mathcal{I}_{ij,h}(16))'.$$

Thus we have

$$I = 4 \sum_{i,j=1}^{+\infty} \sum_{h=-\infty}^{+\infty} \begin{pmatrix} \mathcal{I}_{ij,h}(1) & \mathcal{I}_{ij,h}(5) & \mathcal{I}_{ij,h}(9) & \mathcal{I}_{ij,h}(13) \\ \mathcal{I}_{ij,h}(2) & \mathcal{I}_{ij,h}(6) & \mathcal{I}_{ij,h}(10) & \mathcal{I}_{ij,h}(14) \\ \mathcal{I}_{ij,h}(3) & \mathcal{I}_{ij,h}(7) & \mathcal{I}_{ij,h}(11) & \mathcal{I}_{ij,h}(15) \\ \mathcal{I}_{ij,h}(4) & \mathcal{I}_{ij,h}(8) & \mathcal{I}_{ij,h}(12) & \mathcal{I}_{ij,h}(16) \end{pmatrix}.$$

In the particular case where e_t is a martingale difference, the expression of I simplifies. We then have

$$\mathcal{I}_{ij,h}(2) = \dots = \mathcal{I}_{ij,h}(5) = \mathcal{I}_{ij,h}(7) = \dots = \mathcal{I}_{ij,h}(10) = \mathcal{I}_{ij,h}(12) = \dots = \mathcal{I}_{ij,h}(15) = 0$$

and for $h = 0$, when $i = j$ we have

$$\begin{aligned}
 \mathcal{I}_{ii,0}(1) &= Ee_{1t}^2 e_{1t-i}^2 \sigma_{11}^{-4} \alpha_{01}^{2(i-1)}, & \mathcal{I}_{ii,0}(6) &= Ee_{1t}^2 e_{2t-i}^2 \sigma_{11}^{-4} \alpha_{02}^{2(i-1)}, \\
 \mathcal{I}_{ii,0}(11) &= Ee_{2t}^2 e_{1t-i}^2 \sigma_{22}^{-4} \alpha_{01}^{2(i-1)}, & \mathcal{I}_{ii,0}(16) &= Ee_{2t}^2 e_{2t-i}^2 \sigma_{22}^{-4} \alpha_{02}^{2(i-1)},
 \end{aligned}$$

and $\mathcal{I}_{ij,h}(\cdot) = 0$ for $h < 0$ or $h > 0$. We also have

$$I = 4 \sum_{i=1}^{+\infty} \begin{pmatrix} \mathcal{I}_{ii,0}(1) & 0 & 0 & 0 \\ 0 & \mathcal{I}_{ii,0}(6) & 0 & 0 \\ 0 & 0 & \mathcal{I}_{ii,0}(11) & 0 \\ 0 & 0 & 0 & \mathcal{I}_{ii,0}(16) \end{pmatrix}. \quad (3.24)$$

In view of (3.23) and (3.24), we have

$$\Omega = J^{-1} I J^{-1} = \sum_{i=1}^{+\infty} \text{Diag} \{ \Omega_i(1, 1), \Omega_i(2, 2), \Omega_i(3, 3), \Omega_i(4, 4) \},$$

where

$$\begin{aligned} \Omega_i(1, 1) &= (1 - \alpha_{01}^2)^2 \mathcal{I}_{ii,0}(1), & \Omega_i(2, 2) &= \left(\frac{1 - \alpha_{02}^2}{\sigma_{22}^2 \sigma_{11}^{-2}} \right)^2 \mathcal{I}_{ii,0}(6), \\ \Omega_i(3, 3) &= \left(\frac{1 - \alpha_{01}^2}{\sigma_{11}^2 \sigma_{22}^{-2}} \right)^2 \mathcal{I}_{ii,0}(11), & \Omega_i(4, 4) &= (1 - \alpha_{02}^2)^2 \mathcal{I}_{ii,0}(16). \end{aligned}$$

Chapitre 4

Multivariate portmanteau test for structural VARMA models with uncorrelated but non-independent error terms

Abstract We consider portmanteau tests for testing the adequacy of vector autoregressive moving-average (VARMA) models under the assumption that the errors are uncorrelated but not necessarily independent. We relax the standard independence assumption to extend the range of application of the VARMA models, and allow to cover linear representations of general nonlinear processes. We first study the joint distribution of the quasi-maximum likelihood estimator (QMLE) or the least squared estimator (LSE) and the noise empirical autocovariances. We then derive the asymptotic distribution of residual empirical autocovariances and autocorrelations under weak assumptions on the noise. We deduce the asymptotic distribution of the Ljung-Box (or Box-Pierce) portmanteau statistics for VARMA models with nonindependent innovations. In the standard framework (*i.e.* under iid assumptions on the noise), it is known that the asymptotic distribution of the portmanteau tests is that of a weighted sum of independent chi-squared random variables. The asymptotic distribution can be quite different when the independence assumption is relaxed. Consequently, the usual chi-squared distribution does not provide an adequate approximation to the distribution of the Box-Pierce goodness-of fit portmanteau test. Hence we propose a method to adjust the critical values of the portmanteau tests. Monte carlo experiments illustrate the finite sample performance of the modified portmanteau test.

Keywords : Goodness-of-fit test, QMLE/LSE, Box-Pierce and Ljung-Box portmanteau tests, residual autocorrelation, Structural representation, weak VARMA models.

4.1 Introduction

The vector autoregressive moving-average (VARMA) models are used in time series analysis and econometrics to represent multivariate time series (see Reinsel, 1997, Lütkepohl, 2005). These VARMA models are a natural extension of the univariate ARMA models, which constitute the most widely used class of univariate time series models (see *e.g.* Brockwell and Davis, 1991). The sub-class of vector autoregressive (VAR) models has been studied in the econometric literature (see also Lütkepohl, 1993). The validity of the different steps of the traditional methodology of Box and Jenkins, identification, estimation and validation, depends on the noises properties. After identification and estimation of the vector autoregressive moving-average processes, the next important step in the VARMA modeling consists in checking if the estimated model fits satisfactorily the data. This adequacy checking step allows to validate or invalidate the choice of the orders p and q . In VARMA(p, q) models, the choice of p and q is particularly important because the number of parameters, $(p + q + 2)d^2$, quickly increases with p and q , which entails statistical difficulties. Thus it is important to check the validity of a VARMA(p, q) model, for a given order p and q . This paper is devoted to the problem of the validation step of VARMA representations of multivariate processes. This validation stage is not only based on portmanteau tests, but also on the examination of the autocorrelation function of the residuals. Based on the residual empirical autocorrelations, Box and Pierce (1970) (**BP** hereafter) derived a goodness-of-fit test, the portmanteau test, for univariate strong ARMA models. Ljung and Box (1978) (**LB** hereafter) proposed a modified portmanteau test which is nowadays one of the most popular diagnostic checking tool in ARMA modeling of time series. The multivariate version of the **BP** portmanteau statistic was introduced by Chitturi (1974). We use this so-called portmanteau tests considered by Chitturi (1974) and Hosking (1980) for checking the overall significance of the residual autocorrelations of a VARMA(p, q) model (see also Hosking, 1981a,b; Li and McLeod, 1981; Ahn, 1988). Hosking (1981a) gave several equivalent forms of this statistic. The papers on the multivariate version of the portmanteau statistic are generally under the assumption that the errors ϵ_t are independent. This independence assumption is restrictive because it precludes conditional heteroscedasticity and/ or other forms of nonlinearity (see Francq and Zakoïan, 2005, for a review on weak univariate ARMA models). Relaxing this independence assumption allows to cover linear representations of general nonlinear processes and to extend the range of application of the VARMA models. VARMA models with nonindependent innovations (*i.e.* weak VARMA models) have been less studied than VARMA models with iid errors (*i.e.* strong VARMA models). The asymptotic theory of weak ARMA model validation is mainly limited to the univariate framework (see Francq and Zakoïan, 2005). In the multivariate analysis, notable exceptions are Dufour and Pelletier (2005) who study the choice of the order p and q of VARMA models under weak assumptions on the innovation process, Francq and Raïssi (2007) who study portmanteau tests for weak VAR models and Boubacar Mainassara and Francq (2009) who study the consistency and the asymptotic normality of the QMLE for weak VARMA model. The main

goal of the present article is to complete the available results concerning the statistical analysis of weak VARMA models by considering the adequacy problem, which have not been studied in the above-mentioned papers. We proceed to study the behaviour of the goodness-of fit portmanteau tests when the ϵ_t are not independent. We will see that the standard portmanteau tests can be quite misleading in the framework of non independent errors. A modified version of these tests is thus proposed.

The paper is organized as follows. Section 4.2 presents the structural weak VARMA models that we consider here. Structural forms are employed in econometrics in order to introduce instantaneous relationships between economic variables. Section 4.3 presents the results on the QMLE/LSE asymptotic distribution obtained by Boubacar Mainassara and Francq (2009) when (ϵ_t) satisfies mild mixing assumptions. Section 4.4 is devoted to the joint distribution of the QMLE/LSE and the noise empirical autocovariances. In Section 4.5 we derive the asymptotic distribution of residual empirical autocovariances and autocorrelations under weak assumptions on the noise. In Section 4.6 it is shown how the standard Ljung-Box (or Box-Pierce) portmanteau tests must be adapted in the case of VARMA models with nonindependent innovations. In Section 4.7 we give examples of multivariate weak white noises. Numerical experiments are presented in Section 4.9. The proofs of the main results are collected in the appendix.

We denote by $A \otimes B$ the Kronecker product of two matrices A and B , and by $\text{vec}A$ the vector obtained by stacking the columns of A . The reader is referred to Magnus and Neudecker (1988) for the properties of these operators. Let 0_r be the null vector of \mathbb{R}^r , and let I_r be the $r \times r$ identity matrix.

4.2 Model and assumptions

Consider a d -dimensional stationary process (X_t) satisfying a structural VARMA(p, q) representation of the form

$$A_{00}X_t - \sum_{i=1}^p A_{0i}X_{t-i} = B_{00}\epsilon_t - \sum_{i=1}^q B_{0i}\epsilon_{t-i}, \quad \forall t \in \mathbb{Z} = \{0, \pm 1, \dots\}, \quad (4.1)$$

where ϵ_t is a white noise, namely a stationary sequence of centered and uncorrelated random variables with a non singular variance Σ_0 . It is customary to say that (X_t) is a strong VARMA(p, q) model if (ϵ_t) is a strong white noise, that is, if it satisfies

A1 : (ϵ_t) is a sequence of independent and identically distributed (iid) random vectors, $E\epsilon_t = 0$ and $\text{Var}(\epsilon_t) = \Sigma_0$.

We say that (4.1) is a weak VARMA(p, q) model if (ϵ_t) is a weak white noise, that is, if it satisfies

A1' : $E\epsilon_t = 0$, $\text{Var}(\epsilon_t) = \Sigma_0$, and $\text{Cov}(\epsilon_t, \epsilon_{t-h}) = 0$ for all $t \in \mathbb{Z}$ and all $h \neq 0$.

Assumption **A1** is clearly stronger than **A1'**. The class of strong VARMA models is often considered too restrictive by practitioners. The standard VARMA(p, q) form, which is sometimes called the reduced form, is obtained for $A_{00} = B_{00} = I_d$. Let $[A_{00} \dots A_{0p} B_{00} \dots B_{0q}]$ be the $d \times (p + q + 2)d$ matrix of VAR and MA coefficients. The parameter $\theta_0 = \text{vec}[A_{00} \dots A_{0p} B_{00} \dots B_{0q}]$ belongs to the compact parameter space $\Theta \subset \mathbb{R}^{k_0}$, where $k_0 = (p + q + 2)d^2$ is the number of unknown parameters in VAR and MA parts.

It is important to note that, we cannot work with the structural representation (4.1) because it is not identified. The following assumption ensure the identification of the structural VARMA models.

A2 : For all $\theta \in \Theta$, $\theta \neq \theta_0$, we have $A_0^{-1} B_0 B_\theta^{-1}(L) A_\theta(L) X_t \neq A_{00}^{-1} B_{00} \epsilon_t$ with non zero probability, or $A_0^{-1} B_0 \Sigma B_0' A_0^{-1'} \neq A_{00}^{-1} B_{00} \Sigma_0 B_{00}' A_{00}^{-1'}$.

The previous identifiability assumption is satisfied when the parameter space Θ is sufficiently constrained.

For $\theta \in \Theta$, such that $\theta = \text{vec}[A_0 \dots A_p B_0 \dots B_q]$, write $A_\theta(z) = A_0 - \sum_{i=1}^p A_i z^i$ and $B_\theta(z) = B_0 - \sum_{i=1}^q B_i z^i$. We assume that Θ corresponds to stable and invertible representations, namely

A3 : for all $\theta \in \Theta$, we have $\det A_\theta(z) \det B_\theta(z) \neq 0$ for all $|z| \leq 1$.

Under the assumption **A3**, it is well known that the process (X_t) can be written as a MA(∞) of the form

$$X_t = \sum_{i=0}^{\infty} \psi_{0i} \epsilon_{t-i}, \quad \psi_{00} = A_{00}^{-1} B_{00}, \quad \sum_{i=0}^{\infty} \|\psi_{0i}\| < \infty. \quad (4.2)$$

A4 : Matrix Σ_0 is positive definite.

To show the strong consistency, we will use the following assumptions.

A5 : The process (ϵ_t) is stationary and ergodic.

Note that **A5** is entailed by **A1**, but not by **A1'**. Note that (ϵ_t) can be replaced by (X_t) in **A5**, because $X_t = A_{\theta_0}^{-1}(L) B_{\theta_0}(L) \epsilon_t$ and $\epsilon_t = B_{\theta_0}^{-1}(L) A_{\theta_0}(L) X_t$, where L stands for the backward operator.

4.3 Least Squares Estimation under non-iid innovations

Let X_1, \dots, X_n be observations of a process satisfying the VARMA representation (4.1). Let $\theta \in \Theta$ and $A_0 = A_0(\theta), \dots, A_p = A_p(\theta), B_0 = B_0(\theta), \dots, B_q = B_q(\theta), \Sigma = \Sigma(\theta)$ such that $\theta = \text{vec}[A_0 \dots A_p B_0 \dots B_q]$. Note that from **A3** the matrices A_0 and B_0 are invertible. Introducing the innovation process $e_t = A_{00}^{-1} B_{00} \epsilon_t$, the structural repre-

sentation $A_{\theta_0}(L)X_t = B_{\theta_0}(L)\epsilon_t$ can be rewritten as the reduced VARMA representation

$$X_t - \sum_{i=1}^p A_{00}^{-1} A_{0i} X_{t-i} = e_t - \sum_{i=1}^q A_{00}^{-1} B_{0i} B_{00}^{-1} A_{00} e_{t-i}. \quad (4.3)$$

Note that $e_t(\theta_0) = e_t$. For simplicity, we will omit the notation θ in all quantities taken at the true value, θ_0 . For all $\theta \in \Theta$, the assumption on the MA polynomial (from **A3**) implies that there exists a sequence of constants matrices $(C_i(\theta))$ such that $\sum_{i=1}^{\infty} \|C_i(\theta)\| < \infty$ and

$$e_t(\theta) = X_t - \sum_{i=1}^{\infty} C_i(\theta) X_{t-i}. \quad (4.4)$$

Given a realization X_1, X_2, \dots, X_n , the variable $e_t(\theta)$ can be approximated, for $0 < t \leq n$, by $\tilde{e}_t(\theta)$ defined recursively by

$$\tilde{e}_t(\theta) = X_t - \sum_{i=1}^p A_0^{-1} A_i X_{t-i} + \sum_{i=1}^q A_0^{-1} B_i B_0^{-1} A_0 \tilde{e}_{t-i}(\theta),$$

where the unknown initial values are set to zero : $\tilde{e}_0(\theta) = \dots = \tilde{e}_{1-q}(\theta) = X_0 = \dots = X_{1-p} = 0$. The gaussian quasi-likelihood is given by

$$L_n(\theta, \Sigma_e) = \prod_{t=1}^n \frac{1}{(2\pi)^{d/2} \sqrt{\det \Sigma_e}} \exp \left\{ -\frac{1}{2} \tilde{e}'_t(\theta) \Sigma_e^{-1} \tilde{e}_t(\theta) \right\}, \quad \Sigma_e = A_0^{-1} B_0 \Sigma B_0' A_0^{-1'}.$$

A quasi-maximum likelihood (QML) of θ and Σ_e are a measurable solution $(\hat{\theta}_n, \hat{\Sigma}_e)$ of

$$(\hat{\theta}_n, \hat{\Sigma}_e) = \arg \min_{\theta, \Sigma_e} \left\{ \log(\det \Sigma_e) + \frac{1}{n} \sum_{t=1}^n \tilde{e}_t(\theta) \Sigma_e^{-1} \tilde{e}'_t(\theta) \right\}.$$

Under the following additional assumptions, Boubacar Mainassara and Francq (2009) showed respectively in Theorem 1 and Theorem 2 the consistency and the asymptotic normality of the QML estimator of weak multivariate ARMA model.

Assume that θ_0 is not on the boundary of the parameter space Θ .

A6 : We have $\theta_0 \in \overset{\circ}{\Theta}$, where $\overset{\circ}{\Theta}$ denotes the interior of Θ .

We denote by $\alpha_\epsilon(k)$, $k = 0, 1, \dots$, the strong mixing coefficients of the process (ϵ_t) . The mixing coefficients of a stationary process $\epsilon = (\epsilon_t)$ are denoted by

$$\alpha_\epsilon(k) = \sup_{A \in \sigma(\epsilon_u, u \leq t), B \in \sigma(\epsilon_u, u \geq t+h)} |P(A \cap B) - P(A)P(B)|.$$

The reader is referred to Davidson (1994) for details about mixing assumptions.

A7 : We have $E\|\epsilon_t\|^{4+2\nu} < \infty$ and $\sum_{k=0}^{\infty} \{\alpha_\epsilon(k)\}^{\frac{\nu}{2+\nu}} < \infty$ for some $\nu > 0$.

One of the most popular estimation procedure is that of the least squares estimator (LSE) minimizing

$$\log \det \hat{\Sigma}_e = \log \det \left\{ \frac{1}{n} \sum_{t=1}^n \tilde{e}_t(\hat{\theta}) \tilde{e}'_t(\hat{\theta}) \right\},$$

or equivalently

$$\det \hat{\Sigma}_e = \det \left\{ \frac{1}{n} \sum_{t=1}^n \tilde{e}_t(\hat{\theta}) \tilde{e}_t'(\hat{\theta}) \right\}.$$

For the processes of the form (4.3), under **A1'**, **A2-A7**, it can be shown (see *e.g.* Boubacar Mainassara and Francq 2009), that the LS estimator of θ coincides with the gaussian quasi-maximum likelihood estimator (QMLE). More precisely, $\hat{\theta}_n$ satisfies, almost surely,

$$Q_n(\hat{\theta}_n) = \min_{\theta \in \Theta} Q_n(\theta),$$

where

$$Q_n(\theta) = \log \det \left\{ \frac{1}{n} \sum_{t=1}^n \tilde{e}_t(\theta) \tilde{e}_t'(\theta) \right\} \quad \text{or} \quad Q_n(\theta) = \det \left\{ \frac{1}{n} \sum_{t=1}^n \tilde{e}_t(\theta) \tilde{e}_t'(\theta) \right\}.$$

To obtain the consistency and asymptotic normality of the QMLE/LSE, it will be convenient to consider the functions

$$O_n(\theta) = \log \det \Sigma_n \quad \text{or} \quad O_n(\theta) = \det \Sigma_n,$$

where $\Sigma_n = \Sigma_n(\theta) = n^{-1} \sum_{t=1}^n e_t(\theta) e_t'(\theta)$ and $(e_t(\theta))$ is given by (4.4). Under **A1'**, **A2-A7** or **A1-A4** and **A6**, let $\hat{\theta}_n$ be the LS estimate of θ_0 by maximizing

$$O_n(\theta) = \log \det \left\{ \frac{1}{n} \sum_{t=1}^n e_t(\theta) e_t'(\theta) \right\}.$$

In the univariate case, Francq and Zakoïan (1998) showed the asymptotic normality of the LS estimator under mixing assumptions. This remains valid of the multivariate LS estimator. Then under the assumptions **A1'**, **A2-A7**, $\sqrt{n} (\hat{\theta}_n - \theta_0)$ is asymptotically normal with mean 0 and covariance matrix $\Sigma_{\hat{\theta}_n} := J^{-1} I J^{-1}$, where $J = J(\theta_0)$ and $I = I(\theta_0)$, with

$$J(\theta) = \lim_{n \rightarrow \infty} \frac{1}{n} \frac{\partial^2}{\partial \theta \partial \theta'} Q_n(\theta) \quad a.s.$$

and

$$I(\theta) = \lim_{n \rightarrow \infty} \text{Var} \frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} Q_n(\theta).$$

In the standard strong VARMA case, *i.e.* when **A5** is replaced by the assumption **A1** that (ϵ_t) is iid, we have $I = J$, so that $\Sigma_{\hat{\theta}_n} = J^{-1}$.

4.4 Joint distribution of $\hat{\theta}_n$ and the noise empirical autocovariances

Let $\hat{e}_t = \tilde{e}_t(\hat{\theta}_n)$ be the LS residuals when $p > 0$ or $q > 0$, and let $\hat{e}_t = e_t = X_t$ when $p = q = 0$. When $p + q \neq 0$, we have $\hat{e}_t = 0$ for $t \leq 0$ and $t > n$ and

$$\hat{e}_t = X_t - \sum_{i=1}^p A_0^{-1}(\hat{\theta}_n) A_i(\hat{\theta}_n) \hat{X}_{t-i} + \sum_{i=1}^q A_0^{-1}(\hat{\theta}_n) B_i(\hat{\theta}_n) B_0^{-1}(\hat{\theta}_n) A_0(\hat{\theta}_n) \hat{e}_{t-i},$$

for $t = 1, \dots, n$, with $\hat{X}_t = 0$ for $t \leq 0$ and $\hat{X}_t = X_t$ for $t \geq 1$. Let, $\hat{\Sigma}_{e0} = \hat{\Gamma}_e(0) = n^{-1} \sum_{t=1}^n \hat{e}_t \hat{e}_t'$. We denote by

$$\gamma(h) = \frac{1}{n} \sum_{t=h+1}^n e_t e_{t-h}' \quad \text{and} \quad \hat{\Gamma}_e(h) = \frac{1}{n} \sum_{t=h+1}^n \hat{e}_t \hat{e}_{t-h}'$$

the white noise "empirical" autocovariances and residual autocovariances. It should be noted that $\gamma(h)$ is not a statistic (unless if $p = q = 0$) because it depends on the unobserved innovations $e_t = e_t(\theta_0)$. For a fixed integer $m \geq 1$, let

$$\gamma_m = (\{\text{vec} \gamma(1)\}', \dots, \{\text{vec} \gamma(m)\}')'$$

and

$$\hat{\Gamma}_m = \left(\left\{ \text{vec} \hat{\Gamma}_e(1) \right\}', \dots, \left\{ \text{vec} \hat{\Gamma}_e(m) \right\}' \right)'$$

and let

$$\Gamma(\ell, \ell') = \sum_{h=-\infty}^{\infty} E \left(\{e_{t-\ell} \otimes e_t\} \{e_{t-h-\ell'} \otimes e_{t-h}\}' \right),$$

for $(\ell, \ell') \neq (0, 0)$. For the univariate ARMA model, Francq, Roy and Zakoïan (2005) have showed in Lemma A.1 that $|\Gamma(\ell, \ell')| \leq K \max(\ell, \ell')$ for some constant K , which is sufficient to ensure the existence of these matrices. We can generalize this result for the multivariate ARMA model. Then we obtain $\|\Gamma(\ell, \ell')\| \leq K \max(\ell, \ell')$ for some constant K . The proof is similar to the univariate case.

We are now able to state the following theorem, which is an extension of a result given in Francq, Roy and Zakoïan (2005).

Theorem 4.1. *Assume $p > 0$ or $q > 0$. Under Assumptions **A1'**-**A2-A7** or **A1-A4**, and **A6**, as $n \rightarrow \infty$, $\sqrt{n}(\gamma_m, \hat{\theta}_n - \theta_0)' \xrightarrow{d} \mathcal{N}(0, \Xi)$ where*

$$\Xi = \begin{pmatrix} \Sigma_{\gamma_m} & \Sigma_{\gamma_m, \hat{\theta}_n} \\ \Sigma'_{\gamma_m, \hat{\theta}_n} & \Sigma_{\hat{\theta}_n} \end{pmatrix},$$

with $\Sigma_{\gamma_m} = \{\Gamma(\ell, \ell')\}_{1 \leq \ell, \ell' \leq m}$, $\Sigma'_{\gamma_m, \hat{\theta}_n} = \text{Cov}(\sqrt{n} J^{-1} Y_n, \sqrt{n} \gamma_m)$ and $\Sigma_{\hat{\theta}_n} = \text{Var}_{as}(\sqrt{n} J^{-1} Y_n) = J^{-1} I J^{-1}$.

4.5 Asymptotic distribution of residual empirical autocovariances and autocorrelations

Let the diagonal matrices

$$S_e = \text{Diag}(\sigma_e(1), \dots, \sigma_e(d)) \quad \text{and} \quad \hat{S}_e = \text{Diag}(\hat{\sigma}_e(1), \dots, \hat{\sigma}_e(d)),$$

where $\sigma_e^2(i)$ is the variance of the i -th coordinate of e_t and $\hat{\sigma}_e^2(i)$ is its sample estimate, i.e. $\sigma_e(i) = \sqrt{Ee_{it}^2}$ and $\hat{\sigma}_e(i) = \sqrt{n^{-1} \sum_{t=1}^n \hat{e}_{it}^2}$. The theoretical and sample autocorrelations at lag ℓ are respectively defined by $R_e(\ell) = S_e^{-1} \Gamma_e(\ell) S_e^{-1}$ and $\hat{R}_e(\ell) = \hat{S}_e^{-1} \hat{\Gamma}_e(\ell) \hat{S}_e^{-1}$, with $\Gamma_e(\ell) := Ee_t e_{t-\ell}' = 0$ for all $\ell \neq 0$. Consider the vector of the first m sample autocorrelations

$$\hat{\rho}_m = \left(\left\{ \text{vec} \hat{R}_e(1) \right\}', \dots, \left\{ \text{vec} \hat{R}_e(m) \right\}' \right)'.$$

Theorem 4.2. *Under Assumptions A1-A4 and A6 or A1', A2-A7,*

$$\sqrt{n} \hat{\Gamma}_m \Rightarrow \mathcal{N}(0, \Sigma_{\hat{\Gamma}_m}) \quad \text{and} \quad \sqrt{n} \hat{\rho}_m \Rightarrow \mathcal{N}(0, \Sigma_{\hat{\rho}_m}) \quad \text{where,}$$

$$\Sigma_{\hat{\Gamma}_m} = \Sigma_{\gamma_m} + \Phi_m \Sigma_{\hat{\theta}_n} \Phi_m' + \Phi_m \Sigma_{\hat{\theta}_n, \gamma_m} + \Sigma_{\hat{\theta}_n, \gamma_m}' \Phi_m' \quad (4.5)$$

$$\Sigma_{\hat{\rho}_m} = \{I_m \otimes (S_e \otimes S_e)^{-1}\} \Sigma_{\hat{\Gamma}_m} \{I_m \otimes (S_e \otimes S_e)^{-1}\} \quad (4.6)$$

and Φ_m is given by (4.26) in the proof of this Theorem.

4.6 Limiting distribution of the portmanteau statistics

Box and Pierce (1970) (**BP** hereafter) derived a goodness-of-fit test, the portmanteau test, for univariate strong ARMA models. Ljung and Box (1978) (**LB** hereafter) proposed a modified portmanteau test which is nowadays one of the most popular diagnostic checking tool in ARMA modeling of time series. The multivariate version of the **BP** portmanteau statistic was introduced by Chitturi (1974). Hosking (1981a) gave several equivalent forms of this statistic. Basic forms are

$$\begin{aligned}
 P_m &= n \sum_{h=1}^m \text{Tr} \left(\hat{\Gamma}'_e(h) \hat{\Gamma}_e^{-1}(0) \hat{\Gamma}_e(h) \hat{\Gamma}_e^{-1}(0) \right) \\
 &= n \sum_{h=1}^m \text{vec} \left(\hat{\Gamma}_e(h) \right)' \left(\hat{\Gamma}_e^{-1}(0) \otimes I_d \right) \text{vec} \left(\hat{\Gamma}_e^{-1}(0) \hat{\Gamma}_e(h) \right) \\
 &= n \sum_{h=1}^m \text{vec} \left(\hat{\Gamma}_e(h) \right)' \left(\hat{\Gamma}_e^{-1}(0) \otimes I_d \right) \left(I_d \otimes \hat{\Gamma}_e^{-1}(0) \right) \text{vec} \left(\hat{\Gamma}_e(h) \right) \\
 &= n \sum_{h=1}^m \text{vec} \left(\hat{\Gamma}_e(h) \right)' \left(\hat{\Gamma}_e^{-1}(0) \otimes \hat{\Gamma}_e^{-1}(0) \right) \text{vec} \left(\hat{\Gamma}_e(h) \right) \\
 &= n \hat{\Gamma}'_m \left(I_m \otimes \left\{ \hat{\Gamma}_e^{-1}(0) \otimes \hat{\Gamma}_e^{-1}(0) \right\} \right) \hat{\Gamma}_m \\
 &= n \hat{\rho}'_m \left(I_m \otimes \left\{ \hat{\Gamma}_e(0) \hat{\Gamma}_e^{-1}(0) \hat{\Gamma}_e(0) \right\} \otimes \left\{ \hat{\Gamma}_e(0) \hat{\Gamma}_e^{-1}(0) \hat{\Gamma}_e(0) \right\} \right) \hat{\rho}_m \\
 &= n \hat{\rho}'_m \left(I_m \otimes \left\{ \hat{R}_e^{-1}(0) \otimes \hat{R}_e^{-1}(0) \right\} \right) \hat{\rho}_m.
 \end{aligned}$$

Where the equalities is obtained from the elementary relations $\text{vec}(AB) = (I \otimes A) \text{vec} B$, $(A \otimes B)(C \otimes D) = AC \otimes BD$ and $\text{Tr}(ABC) = \text{vec}(A)'(C' \otimes I) \text{vec} B$. Similarly to the univariate **LB** portmanteau statistic, Hosking (1980) defined the modified portmanteau statistic

$$\tilde{P}_m = n^2 \sum_{h=1}^m (n-h)^{-1} \text{Tr} \left(\hat{\Gamma}'_e(h) \hat{\Gamma}_e^{-1}(0) \hat{\Gamma}_e(h) \hat{\Gamma}_e^{-1}(0) \right).$$

These portmanteau statistics are generally used to test the null hypothesis

$$H_0 : (X_t) \text{ satisfies a VARMA}(p, q) \text{ representation}$$

against the alternative

$$H_1 : (X_t) \text{ does not admit a VARMA representation or admits a VARMA}(p', q') \text{ representation with } p' > p \text{ or } q' > q.$$

These portmanteau tests are very useful tools for checking the overall significance of the residual autocorrelations. Under the assumption that the data generating process (**DGP**) follows a strong VARMA(p, q) model, the asymptotic distribution of the statistics P_m and \tilde{P}_m is generally approximated by the $\chi^2_{d^2m-k_0}$ distribution ($d^2m > k_0$) (the degrees of freedom are obtained by subtracting the number of freely estimated VARMA coefficients from d^2m). When the innovations are gaussian, Hosking (1980) found that the finite-sample distribution of \tilde{P}_m is more nearly $\chi^2_{d^2(m-(p+q))}$ than that of P_m . From Theorem 4.2 we deduce the following result, which gives the exact asymptotic distribution of the standard portmanteau statistics P_m . We will see that the distribution may be very different from the $\chi^2_{d^2m-k_0}$ in the case of VARMA(p, q) models.

Theorem 4.3. *Under Assumptions A1-A4 and A6 or A1', A2-A7, the statistics P_m and \tilde{P}_m converge in distribution, as $n \rightarrow \infty$, to*

$$Z_m(\xi_m) = \sum_{i=1}^{d^2 m} \xi_{i,d^2 m} Z_i^2$$

where $\xi_m = (\xi_{1,d^2 m}, \dots, \xi_{d^2 m,d^2 m})'$ is the vector of the eigenvalues of the matrix

$$\Omega_m = (I_m \otimes \Sigma_e^{-1/2} \otimes \Sigma_e^{-1/2}) \Sigma_{\hat{\Gamma}_m} (I_m \otimes \Sigma_e^{-1/2} \otimes \Sigma_e^{-1/2}),$$

and Z_1, \dots, Z_m are independent $\mathcal{N}(0, 1)$ variables.

4.7 Examples

We now present examples of sequences of the uncorrelated but non independent random variables.

Example 4.1. *In the univariate case, Romano and Thombs (1996) built weak white noises (ϵ_t) by setting $\epsilon_t = \eta_t \eta_{t-1} \dots \eta_{t-k}$ where (η_t) is a strong white noise and $k \geq 1$. This approach has been extended by Francq and Raïssi (2006) to the d -multivariate by setting $\epsilon_t = B(\eta_t) \dots B(\eta_{t-k+1}) \eta_{t-k}$, where $\{\eta_t = (\eta_{1t}, \dots, \eta_{dt})'\}_t$ is a d -dimensional strong white noise, and $B(\eta_t) = \{B_{ij}(\eta_t)\}$ is a $d \times d$ random matrix whose elements are linear combinations of the components of η_t . It is obvious to see that $E\epsilon_t = 0$ and $\text{Cov}(\epsilon_t, \epsilon_{t-h}) = 0$ for all $h \neq 0$. Therefore ϵ_t is a white noise, but in general this noise is not a strong one. Indeed, assuming for simplicity that $k = 1$ and $B_{11}(\eta_t) = \eta_{1t}$ and $B_{1j}(\eta_t) = 0$ for all $j > 1$, we have*

$$\text{Cov}(\epsilon_{1t}^2, \epsilon_{1t-1}^2) = \{E\eta_{1t}^2\}^2 \text{Var}\{\eta_{1t}^2\} \neq 0,$$

which shows that the ϵ_t 's are not independent.

Example 4.2. *For $i = 1, \dots, d$ let $\epsilon_{it} = \prod_{m=0}^{2k} \eta_{it-m}$ be the i -th coordinate of the vector $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{dt})'$ where $(\eta_t) = (\eta_{1t}, \dots, \eta_{dt})'$ is a strong white noise and $k \geq 1$. This d -multivariate noise ϵ_t is the direct extension of the weak white noise defined by Romano and Thombs (1996) in the univariate case. To show the ϵ_t 's are not independent, we calculate $\text{Cov}(\epsilon_{it}^2, \epsilon_{it-2k}^2)$ and $\text{Cov}(\epsilon_{it}^2, \epsilon_{jt-2k}^2)$, for $j = 1, \dots, d$. For simplicity assume that the d -multivariate (η_t) is gaussian with $\text{Var}(\eta_{it}) = 1$ and $\text{Cov}(\eta_{it}, \eta_{jt}) = \rho^{|i-j|} \in (-1, 1)$. More generally take,*

$$\begin{aligned} \text{Cov}(\epsilon_{it}^2, \epsilon_{jt}^2) &= E\epsilon_{it}^2 \epsilon_{jt}^2 - \{E\epsilon_{it}^2\}^2 \{E\epsilon_{jt}^2\}^2 \\ &= E \prod_{m=0}^{2k} \eta_{it-m}^2 \eta_{jt-m}^2 - 1. \end{aligned}$$

Let, $\Sigma_{ij} = \begin{pmatrix} 1 & \rho^{|i-j|} \\ \rho^{|i-j|} & 1 \end{pmatrix}$ the covariance matrix of the vector $(\eta_{it}, \eta_{jt})'$. A joint probability density function of the bivariate random vector $(\eta_{it}, \eta_{jt})'$ with mean vector zero is of the form

$$f(x) = (2\pi)^{-1} |\Sigma_{ij}|^{-1/2} \exp[(-1/2)x' \Sigma_{ij}^{-1} x],$$

with $|\Sigma_{ij}| = 1 - \rho^{2|i-j|}$. If $|\rho| < 1$ the matrix Σ_{ij} is invertible and

$$\Sigma_{ij}^{-1} = \frac{1}{1 - \rho^{2|i-j|}} \times \begin{pmatrix} 1 & -\rho^{|i-j|} \\ -\rho^{|i-j|} & 1 \end{pmatrix}.$$

Then we have

$$f(x_1, x_2) = \frac{1}{2\pi \sqrt{1 - \rho^{2|i-j|}}} \exp \left\{ \frac{-1}{2(1 - \rho^{2|i-j|})} (x_1^2 + x_2^2 - 2\rho^{|i-j|} x_1 x_2) \right\}.$$

Letting $\beta = 1/2\pi \sqrt{1 - \rho^{2|i-j|}}$ and $\alpha = \sqrt{1 - \rho^{2|i-j|}}$, we calculate the integral

$$Ex_1^2 x_2^2 = \beta \int \int_{\mathbb{R}^2} x_1^2 x_2^2 e^{-\frac{1}{2} \left\{ \left(\frac{x_1 - \rho^{|i-j|} x_2}{\alpha} \right)^2 + x_2^2 \right\}} dx_1 dx_2.$$

Setting $u = (x_1 - \rho^{|i-j|} x_2) / \alpha$ and $s = x_2$, the Jacobian of this transformation is $|\mathbf{J}| = \alpha$. Then we have

$$\begin{aligned} Ex_1^2 x_2^2 &= \beta \alpha \int \int_{\mathbb{R}^2} (\alpha u + \rho^{|i-j|} s)^2 s^2 e^{-\frac{1}{2}(u^2 + s^2)} du ds, \\ &= \beta \alpha \int_{\mathbb{R}} s^2 e^{-\frac{1}{2}s^2} \left\{ \int_{\mathbb{R}} (\alpha^2 u^2 + 2\alpha \rho^{|i-j|} u s + \rho^{2|i-j|} s^2) e^{-\frac{1}{2}u^2} du \right\} ds. \end{aligned}$$

Integrating by parts and passing to polar coordinates, we have

$$\begin{aligned} Ex_1^2 x_2^2 &= \beta \alpha \int_{\mathbb{R}} \left(\alpha^2 \sqrt{2\pi} + \rho^{2|i-j|} \sqrt{2\pi} s^2 \right) s^2 e^{-\frac{1}{2}s^2} ds, \\ &= \beta \alpha (2\pi \alpha^2 + 6\pi \rho^{2|i-j|}) \\ &= 1 + 2\rho^{2|i-j|}. \end{aligned}$$

Thus, we have

$$\begin{aligned} Cov(\epsilon_{it}^2, \epsilon_{jt}^2) &= E\epsilon_{it}^2 \epsilon_{jt}^2 - \{E\epsilon_{it}^2\}^2 \{E\epsilon_{jt}^2\}^2 \\ &= E \prod_{m=0}^{2k} \eta_{it-m}^2 \eta_{jt-m}^2 - 1 \\ &= (1 + 2\rho^{2|i-j|})^{2k+1} - 1, \end{aligned}$$

then, we have

$$Cov(\epsilon_{it}^2, \epsilon_{jt-h}^2) = (1 + 2\rho^{2|i-j|})^{2k-h+1} - 1.$$

Thus

$$\text{Cov}(\epsilon_{it}^2, \epsilon_{j t-h}^2) = \begin{cases} 2\rho^{2|i-j|} & \text{if } h = 2k \\ 0 & \text{if } h \geq 2k + 1 \end{cases} \quad (4.7)$$

If $i = j$, we have

$$\text{Cov}(\epsilon_{it}^2, \epsilon_{it}^2) = 3^{2k+1} - 1$$

then, we have

$$\text{Cov}(\epsilon_{it}^2, \epsilon_{i t-h}^2) = 3^{2k-h+1} - 1.$$

Thus

$$\text{Cov}(\epsilon_{it}^2, \epsilon_{i t-h}^2) = \begin{cases} 3 - 1 & \text{if } h = 2k \\ 0 & \text{if } h \geq 2k + 1 \end{cases} \quad (4.8)$$

From (4.7) and (4.8), the conclusion follows that the ϵ_t 's are not independent.

Example 4.3. The example 4.2 may be complicated by switching successively the components at the i -th coordinate which are not pair lag with those at the $(i+1)$ -th coordinate of the vector $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{dt})'$ where $(\eta_t) = (\eta_{1t}, \dots, \eta_{dt})'$ is a strong white noise. For a d fixed, $i = 1, \dots, d-1$ and $k \geq 1$ let $\epsilon_{it} = \eta_{it} \prod_{m=1}^k \eta_{i+1 t-2m+1} \eta_{i t-2m}$ be the i -th coordinate of the vector ϵ_t and $\epsilon_{dt} = \eta_{dt} \prod_{m=1}^k \eta_{1 t-2m+1} \eta_{d t-2m}$ be the last coordinate. This d -multivariate noise ϵ_t is also an extension of the weak white noise defined by Romano and Thombs (1996) in the univariate case. For simplicity assume that the d -multivariate (η_t) is gaussian with $\text{Var}(\eta_{it}) = 1$, $\text{Cov}(\eta_{it}, \eta_{jt}) = \rho^{|i-j|} \in (-1, 1)$ and $E\eta_{it}^2 \eta_{jt}^2 = 1 + 2\rho^{2|i-j|}$, for $j = 1, \dots, d$. Then for $j = i$, we obtain

$$\begin{aligned} \text{Cov}(\epsilon_{it}^2, \epsilon_{i t-1}^2) &= E\epsilon_{it}^2 \epsilon_{i t-1}^2 - \{E\epsilon_{it}^2\}^2 \{E\epsilon_{i t-1}^2\}^2 \\ &= E\eta_{it}^2 \prod_{m=1}^k \eta_{i t-1}^2 \eta_{i+1 t-2m+1}^2 \eta_{i+1 t-2m}^2 \eta_{i t-2m}^2 \eta_{i t-2m-1}^2 - 1 \\ &= \tilde{\rho}^{2k} - 1 \end{aligned}$$

where $\tilde{\rho} = 1 + 2\rho^2$ and

$$\begin{aligned} \text{Cov}(\epsilon_{dt}^2, \epsilon_{d t-1}^2) &= E\epsilon_{dt}^2 \epsilon_{d t-1}^2 - \{E\epsilon_{dt}^2\}^2 \{E\epsilon_{d t-1}^2\}^2 \\ &= E\eta_{dt}^2 \prod_{m=1}^k \eta_{d t-1}^2 \eta_{1 t-2m+1}^2 \eta_{d t-2m}^2 \eta_{1 t-2m}^2 \eta_{d t-2m-1}^2 - 1 \\ &= (1 + 2\rho^{2|d-1|})^{2k} - 1. \end{aligned}$$

By induction, we have

$$\text{Cov}(\epsilon_{it}^2, \epsilon_{i,t-h}^2) = \tilde{\rho}^{2k-h+1} - 1 = \begin{cases} 2\rho^2 & \text{if } h = 2k \\ 0 & \text{if } h \geq 2k + 1 \end{cases} \quad (4.9)$$

and

$$\begin{aligned} \text{Cov}(\epsilon_{dt}^2, \epsilon_{d,t-h}^2) &= (1 + 2\rho^{2|d-1|})^{2k-h+1} - 1 \\ &= \begin{cases} 2\rho^{2|d-1|} & \text{if } h = 2k \\ 0 & \text{if } h \geq 2k + 1 \end{cases} . \end{aligned} \quad (4.10)$$

Similarly, for $i \neq j$ we have

$$\begin{aligned} \text{Cov}(\epsilon_{it}^2, \epsilon_{jt-1}^2) &= E\eta_{it}^2 \prod_{m=1}^k \eta_{jt-1}^2 \eta_{i+1,t-2m+1}^2 \eta_{j+1,t-2m}^2 \eta_{i,t-2m}^2 \eta_{j,t-2m-1}^2 - 1 \\ &= (1 + 2\rho^{2|i-j-1|})^{2k} - 1. \end{aligned}$$

By induction, we have

$$\begin{aligned} \text{Cov}(\epsilon_{it}^2, \epsilon_{j,t-h}^2) &= (1 + 2\rho^{2|i-j-1|})^{2k-h+1} - 1 \\ &= \begin{cases} 2\rho^{2|i-j-1|} & \text{if } h = 2k \\ 0 & \text{if } h \geq 2k + 1 \end{cases} . \end{aligned} \quad (4.11)$$

From (4.9), (4.10) and (4.11) the ϵ_t 's are not independent.

Example 4.4. The GARCH models constitute important examples of weak white noises in the univariate case. These models have numerous extensions to the multivariate framework (see Bauwens, Laurent and Rombouts (2006) for a review). Jeantheau (1998) has proposed the simplest extension of the multivariate GARCH with conditional constant correlation. In this model, the process (ϵ_t) verifies the following relation $\epsilon_t = H_t \eta_t$ where $\{\eta_t = (\eta_{1t}, \dots, \eta_{dt})'\}_t$ is an iid centered process with $\text{Var}\{\eta_{it}\} = 1$ and H_t is a diagonal matrix whose elements h_{iit} verify

$$\begin{pmatrix} h_{11t}^2 \\ \vdots \\ h_{ddt}^2 \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_d \end{pmatrix} + \sum_{i=1}^q A_i \begin{pmatrix} \epsilon_{1t-i}^2 \\ \vdots \\ \epsilon_{dt-i}^2 \end{pmatrix} + \sum_{j=1}^p B_j \begin{pmatrix} h_{11t-j}^2 \\ \vdots \\ h_{ddt-j}^2 \end{pmatrix}.$$

The elements of the matrices A_i and B_j , as well as the vector c_i , are supposed to be positive. In addition suppose that the stationarity conditions hold. For simplicity, consider the ARCH(1) case with A_1 such that $a_{dd} \neq 0$ and $a_{11} = \dots = a_{1,d-1} = 0$. Then it is easy to see that

$$\text{Cov}(\epsilon_{dt}^2, \epsilon_{d,t-1}^2) = \text{Cov}\{(c_d + a_{dd}\epsilon_{d,t-1}^2)\eta_{dt}^2, \epsilon_{d,t-1}^2\} = a_{dd} \text{Var}\{\epsilon_{dt}^2\} \neq 0,$$

which shows that the ϵ_t 's are not independent. Thus, the solution of the GARCH equation satisfies the assumption **A1'**, but does not satisfy **A1** in general.

It is seen in Theorem 4.3, that the asymptotic distribution of the **BP** and **LB** portmanteau tests depends of the nuisance parameters involving Σ_e , the matrix Φ_m and the elements of the matrix Ξ . We need an consistent estimator of the above unknown matrices. The matrix Σ_e can be consistently estimate by its sample estimate $\hat{\Sigma}_e = \hat{\Gamma}_e(0)$. The matrix Φ_m can be easily estimated by its empirical counterpart

$$\hat{\Phi}_m = \frac{1}{n} \sum_{t=1}^n \left\{ (\hat{e}'_{t-1}, \dots, \hat{e}'_{t-m})' \otimes \frac{\partial e_t(\theta_0)}{\partial \theta'} \right\}_{\theta_0 = \hat{\theta}_n}.$$

In the econometric literature the nonparametric kernel estimator, also called heteroskedastic autocorrelation consistent (HAC) estimator (see Newey and West, 1987, or Andrews, 1991), is widely used to estimate covariance matrices of the form Ξ . An alternative method consists in using a parametric AR estimate of the spectral density of $\Upsilon_t = (\Upsilon'_{1t}, \Upsilon'_{2t})'$, where $\Upsilon_{1t} = (e'_{t-1}, \dots, e'_{t-m})' \otimes e_t$ and $\Upsilon_{2t} = -2J^{-1}(\partial e'_t(\theta_0)/\partial \theta) \Sigma_{e0}^{-1} e_t(\theta_0)$. Interpreting $(2\pi)^{-1}\Xi$ as the spectral density of the stationary process (Υ_t) evaluated at frequency 0 (see Brockwell and Davis, 1991, p. 459). This approach, which has been studied by Berk (1974) (see also den Hann and Levin, 1997). So we have

$$\Xi = \Phi^{-1}(1)\Sigma_u\Phi^{-1}(1)$$

when (Υ_t) satisfies an $AR(\infty)$ representation of the form

$$\Phi(L)\Upsilon_t := \Upsilon_t + \sum_{i=1}^{\infty} \Phi_i \Upsilon_{t-i} = u_t, \quad (4.12)$$

where u_t is a weak white noise with variance matrix Σ_u . Since Υ_t is not observable, let $\hat{\Upsilon}_t$ be the vector obtained by replacing θ_0 by $\hat{\theta}_n$ in Υ_t . Let $\hat{\Phi}_r(z) = I_{k_0+d^2m} + \sum_{i=1}^r \hat{\Phi}_{r,i} z^i$, where $\hat{\Phi}_{r,1}, \dots, \hat{\Phi}_{r,r}$ denote the coefficients of the LS regression of $\hat{\Upsilon}_t$ on $\hat{\Upsilon}_{t-1}, \dots, \hat{\Upsilon}_{t-r}$. Let $\hat{u}_{r,t}$ be the residuals of this regression, and let $\hat{\Sigma}_{\hat{u}_r}$ be the empirical variance of $\hat{u}_{r,1}, \dots, \hat{u}_{r,n}$.

We are now able to state the following theorem, which is an extension of a result given in Francq, Roy and Zakoïan (2005).

Theorem 4.4. *In addition to the assumptions of Theorem 4.1, assume that the process (Υ_t) admits an $AR(\infty)$ representation (4.12) in which the roots of $\det \Phi(z) = 0$ are outside the unit disk, $\|\Phi_i\| = o(i^{-2})$, and $\Sigma_u = \text{Var}(u_t)$ is non-singular. Moreover we assume that $\|\epsilon_t\|_{8+4\nu} < \infty$ and $\sum_{k=0}^{\infty} \{\alpha_{X,\epsilon}(k)\}^{\nu/(2+\nu)} < \infty$ for some $\nu > 0$, where $\{\alpha_{X,\epsilon}(k)\}_{k \geq 0}$ denotes the sequence of the strong mixing coefficients of the process $(X'_t, \epsilon'_t)'$. Then the spectral estimator of Ξ*

$$\hat{\Xi}^{\text{SP}} := \hat{\Phi}_r^{-1}(1)\hat{\Sigma}_{\hat{u}_r}\hat{\Phi}_r'^{-1}(1) \rightarrow \Xi$$

in probability when $r = r(n) \rightarrow \infty$ and $r^3/n \rightarrow 0$ as $n \rightarrow \infty$.

Let $\hat{\Omega}_m$ be the matrix obtained by replacing Ξ by $\hat{\Xi}$ and Σ_e by $\hat{\Sigma}_e$ in Ω_m . Denote by $\hat{\xi}_m = (\hat{\xi}_{1,d^2m}, \dots, \hat{\xi}_{d^2m,d^2m})'$ the vector of the eigenvalues of $\hat{\Omega}_m$. At the asymptotic

level α , the **LB** test (resp. the **BP** test) consists in rejecting the adequacy of the weak VARMA(p, q) model when

$$\tilde{P}_m > S_m(1 - \alpha) \quad (\text{resp.} \quad P_m > S_m(1 - \alpha))$$

where $S_m(1 - \alpha)$ is such that $P \left\{ Z_m(\hat{\xi}_m) > S_m(1 - \alpha) \right\} = \alpha$.

4.8 Implementation of the goodness-of-fit portmanteau tests

Let X_1, \dots, X_n , be observations of a d -multivariate process. For testing the adequacy of weak VARMA(p, q) model, we use the following steps to implement the modified version of the portmanteau test.

1. Compute the estimates $\hat{A}_1, \dots, \hat{A}_p, \hat{B}_1, \dots, \hat{B}_q$ by QMLE.
2. Compute the QMLE residuals $\hat{e}_t = \tilde{e}_t(\hat{\theta}_n)$ when $p > 0$ or $q > 0$, and let $\hat{e}_t = e_t = X_t$ when $p = q = 0$. When $p + q \neq 0$, we have $\hat{e}_t = 0$ for $t \leq 0$ and $t > n$ and

$$\hat{e}_t = X_t - \sum_{i=1}^p A_0^{-1}(\hat{\theta}_n) A_i(\hat{\theta}_n) \hat{X}_{t-i} + \sum_{i=1}^q A_0^{-1}(\hat{\theta}_n) B_i(\hat{\theta}_n) B_0^{-1}(\hat{\theta}_n) A_0(\hat{\theta}_n) \hat{e}_{t-i},$$

for $t = 1, \dots, n$, with $\hat{X}_t = 0$ for $t \leq 0$ and $\hat{X}_t = X_t$ for $t \geq 1$.

3. Compute the residual autocovariances $\hat{\Gamma}_e(0) = \hat{\Sigma}_{e0}$ and $\hat{\Gamma}_e(h)$ for $h = 1, \dots, m$ and $\hat{\Gamma}_m = \left(\left\{ \hat{\Gamma}_e(1) \right\}', \dots, \left\{ \hat{\Gamma}_e(m) \right\}' \right)'$.
4. Compute the matrix $\hat{J} = 2n^{-1} \sum_{t=1}^n (\partial \hat{e}_t' / \partial \theta) \hat{\Sigma}_{e0}^{-1} (\partial \hat{e}_t / \partial \theta')$.
5. Compute $\hat{Y}_t = \left(\hat{Y}_{1t}', \hat{Y}_{2t}' \right)'$, where $\hat{Y}_{1t} = (\hat{e}_{t-1}', \dots, \hat{e}_{t-m}')' \otimes \hat{e}_t$ and $\hat{Y}_{2t} = -2\hat{J}^{-1} (\partial \hat{e}_t' / \partial \theta) \hat{\Sigma}_{e0}^{-1} \hat{e}_t$.
6. Fit the VAR(r) model

$$\hat{\Phi}_r(L) \hat{Y}_t := \left(I_{d^2 m + k_0} + \sum_{i=1}^r \hat{\Phi}_{r,i}(L) \right) \hat{Y}_t = \hat{u}_{r,t}.$$

The VAR order r can be fixed or selected by AIC information criteria.

7. Define the estimator

$$\hat{\Xi}^{\text{SP}} := \hat{\Phi}_r^{-1}(1) \hat{\Sigma}_{\hat{u}_r} \hat{\Phi}_r'^{-1}(1) = \begin{pmatrix} \hat{\Sigma}_{\gamma_m} & \hat{\Sigma}_{\gamma_m, \hat{\theta}_n} \\ \hat{\Sigma}'_{\gamma_m, \hat{\theta}_n} & \hat{\Sigma}_{\hat{\theta}_n} \end{pmatrix}, \quad \hat{\Sigma}_{\hat{u}_r} = \frac{1}{n} \sum_{t=1}^n \hat{u}_{r,t} \hat{u}_{r,t}'.$$

8. Define the estimator

$$\hat{\Phi}_m = \frac{1}{n} \sum_{t=1}^n \left\{ (\hat{e}'_{t-1}, \dots, \hat{e}'_{t-m})' \otimes \frac{\partial e_t(\theta_0)}{\partial \theta'} \right\}_{\theta_0 = \hat{\theta}_n}.$$

9. Define the estimators

$$\begin{aligned} \hat{\Sigma}_{\hat{\Gamma}_m} &= \hat{\Sigma}_{\gamma_m} + \hat{\Phi}_m \hat{\Sigma}_{\hat{\theta}_n} \hat{\Phi}'_m + \hat{\Phi}_m \hat{\Sigma}_{\hat{\theta}_n, \gamma_m} + \hat{\Sigma}'_{\hat{\theta}_n, \gamma_m} \hat{\Phi}'_m \\ \hat{\Sigma}_{\hat{\rho}_m} &= \left\{ I_m \otimes (\hat{S}_e \otimes \hat{S}_e)^{-1} \right\} \hat{\Sigma}_{\hat{\Gamma}_m} \left\{ I_m \otimes (\hat{S}_e \otimes \hat{S}_e)^{-1} \right\} \end{aligned}$$

10. Compute the eigenvalues $\hat{\xi}_m = (\hat{\xi}_{1, d^2 m}, \dots, \hat{\xi}_{d^2 m, d^2 m})'$ of the matrix

$$\hat{\Omega}_m = \left(I_m \otimes \hat{\Sigma}_{e0}^{-1/2} \otimes \hat{\Sigma}_{e0}^{-1/2} \right) \hat{\Sigma}_{\hat{\Gamma}_m} \left(I_m \otimes \hat{\Sigma}_{e0}^{-1/2} \otimes \hat{\Sigma}_{e0}^{-1/2} \right).$$

11. Compute the portmanteau statistics

$$\begin{aligned} P_m &= n \hat{\rho}'_m \left(I_m \otimes \left\{ \hat{R}_e^{-1}(0) \otimes \hat{R}_e^{-1}(0) \right\} \right) \hat{\rho}_m \quad \text{and} \\ \tilde{P}_m &= n^2 \sum_{h=1}^m \frac{1}{(n-h)} \text{Tr} \left(\hat{\Gamma}'_e(h) \hat{\Gamma}_e^{-1}(0) \hat{\Gamma}_e(h) \hat{\Gamma}_e^{-1}(0) \right). \end{aligned}$$

12. Evaluate the p -values

$$P \left\{ Z_m(\hat{\xi}_m) > P_m \right\} \quad \text{and} \quad P \left\{ Z_m(\hat{\xi}_m) > \tilde{P}_m \right\}, \quad Z_m(\hat{\xi}_m) = \sum_{i=1}^{d^2 m} \hat{\xi}_{i, d^2 m} Z_i^2,$$

using the Imhof algorithm (1961). The **BP** test (resp. the **LB** test) rejects the adequacy of the weak VARMA(p, q) model when the first (resp. the second) p -value is less than the asymptotic level α .

4.9 Numerical illustrations

In this section, by means of Monte Carlo experiments, we investigate the finite sample properties of the test introduced in this paper. For illustrative purpose, we only present the results of the modified and standard versions of the **LB** test. The results concerning the **BP** test are not presented here, because they are very close to those of the **LB** test. The numerical illustrations of this section are made with the softwares R (see <http://cran.r-project.org/>) and FORTRAN (to compute the p -values using the Imhof algorithm, 1961).

4.9.1 Empirical size

To generate the strong and weak VARMA models, we consider the bivariate model of the form

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & a_1(2, 2) \end{pmatrix} \begin{pmatrix} X_{1t-1} \\ X_{2t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ b_1(2, 1) & b_1(2, 2) \end{pmatrix} \begin{pmatrix} \epsilon_{1t-1} \\ \epsilon_{2t-1} \end{pmatrix}, \quad (4.13)$$

where $(a_1(2, 2), b_1(2, 1), b_1(2, 2)) = (0.225, -0.313, 0.750)$. This model is a VARMA(1,1) model in echelon form.

Strong VARMA case

We first consider the strong VARMA case. To generate this model, we assume that in (4.13) the innovation process (ϵ_t) is defined by

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \sim \text{IID } \mathcal{N}(0, I_2). \quad (4.14)$$

We simulated $N = 1,000$ independent trajectories of size $n = 500$, $n = 1,000$ and $n = 2,000$ of Model (4.13) with the strong Gaussian noise (4.14). For each of these N replications we estimated the coefficients $(a_1(2, 2), b_1(2, 1), b_1(2, 2))$ and we applied portmanteau tests to the residuals for different values of m .

For the standard **LB** test, the VARMA(1,1) model is rejected when the statistic \tilde{P}_m is greater than $\chi_{(4m-3)}^2(0.95)$, where m is the number of residual autocorrelations used in the **LB** statistic. This corresponds to a nominal asymptotic level $\alpha = 5\%$ in the standard case. We know that the asymptotic level of the standard **LB** test is indeed $\alpha = 5\%$ when $(a_1(2, 2), b_1(2, 1), b_1(2, 2)) = (0, 0, 0)$. Note however that, even when the noise is strong, the asymptotic level is not exactly $\alpha = 5\%$ when $(a_1(2, 2), b_1(2, 1), b_1(2, 2)) \neq (0, 0, 0)$.

For the modified **LB** test, the model is rejected when the statistic \tilde{P}_m is greater than $S_m(0.95)$ *i.e.* when the p -value $\left(P \left\{ Z_m(\hat{\xi}_m) > \tilde{P}_m \right\}\right)$ is less than the asymptotic level $\alpha = 0.05$. Let A and B be the (2×2) -matrices with non zero elements $a_1(2, 2)$, $b_1(2, 1)$ and $b_1(2, 2)$. When the roots of $\det(I_2 - Az) \det(I_2 - Bz) = 0$ are near the unit disk, the asymptotic distribution of \tilde{P}_m is likely to be far from its $\chi_{(4m-3)}^2$ approximation. Table 4.1 displays the relative rejection frequencies of the null hypothesis H_0 that the **DGP** follows an VARMA(1, 1) model, over the $N = 1,000$ independent replications. As expected the observed relative rejection frequency of the standard **LB** test is very far from the nominal $\alpha = 5\%$ when the number of autocorrelations used in the **LB** statistic is $m \leq p + q$. This is in accordance with the results in the literature on the

standard VARMA models. In particular, Hosking (1980) showed that the statistic \tilde{P}_m has approximately the chi-squared distribution $\chi_{d^2(m-(p+q))}^2$ without any identifiability constraint. Thus the error of first kind is well controlled by all the tests in the strong case, except for the standard **LB** test when $m \leq p + q$. We draw the somewhat surprising conclusion that, even in the strong VARMA case, the modified version may be preferable to the standard one.

TABLE 4.1 – Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the strong VARMA(1, 1) model (4.13)-(4.14), with $\theta_0 = (0.225, -0.313, 0.750)$.

n	$m = 1$			$m = 2$		
	500	1,000	2,000	500	1,000	2,000
modified LB	5.5	5.1	3.6	4.1	4.6	4.2
standard LB	22.0	21.3	21.7	7.1	7.9	7.5

n	$m = 3$			$m = 4$			$m = 6$		
	500	1,000	2,000	500	1,000	2,000	500	1,000	2,000
modified LB	4.2	4.4	3.6	3.0	3.9	4.2	3.3	3.9	4.1
standard LB	5.9	5.8	5.3	4.9	5.2	5.2	5.3	5.0	4.6

TABLE 4.2 – Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the strong VARMA(1, 1) model (4.13)-(4.14), with $\theta_0 = (0, 0, 0)$.

n	$m = 1$			$m = 2$			$m = 3$		
	500	1,000	2,000	500	1,000	2,000	500	1,000	2,000
modified LB	5.5	4.1	3.0	3.8	3.4	3.2	4.1	3.3	2.6
standard LB	18.3	18.7	16.9	6.0	6.2	4.8	5.2	4.5	3.5

Weak VARMA case

We now repeat the same experiments on different weak VARMA(1, 1) models. We first assume that in (4.13) the innovation process (ϵ_t) is an ARCH(1) (*i.e.* $p = 0, q = 1$) model

$$\begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} = \begin{pmatrix} h_{11t} & 0 \\ 0 & h_{22t} \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} \tag{4.15}$$

where

$$\begin{pmatrix} h_{11t}^2 \\ h_{22t}^2 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \epsilon_{1t-1}^2 \\ \epsilon_{2t-1}^2 \end{pmatrix},$$

with $c_1 = 0.3, c_2 = 0.2, a_{11} = 0.45, a_{21} = 0.4$ and $a_{22} = 0.25$. As expected, Table 4.3 shows that the standard **LB** test poorly performs to assess the adequacy of this weak

VARMA(1, 1) model. In view of the observed relative rejection frequency, the standard **LB** test rejects very often the true VARMA(1, 1). By contrast, the error of first kind is well controlled by the modified version of the **LB** test. We draw the conclusion that, at least for this particular weak VARMA model, the modified version is clearly preferable to the standard one.

TABLE 4.3 – Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the weak VARMA(1, 1) model (4.13)-(4.15), with $\theta_0 = (0.225, -0.313, 0.750)$.

n	$m = 1$			$m = 2$		
	500	1,000	2,000	500	1,000	2,000
modified LB	6.7	7.5	8.5	4.6	4.1	6.5
standard LB	48.3	50.0	50.3	33.1	36.5	39.4

n	$m = 3$			$m = 4$			$m = 6$		
	500	1,000	2,000	500	1,000	2,000	500	1,000	2,000
modified LB	4.4	4.4	5.4	2.8	4.1	5.5	4.0	3.3	4.9
standard LB	28.2	31.3	35.4	24.5	28.1	32.3	22.0	22.9	27.8

TABLE 4.4 – Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the weak VARMA(1, 1) model (4.13)-(4.15), with $\theta_0 = (0, 0, 0)$.

n	$m = 1$			$m = 2$		
	500	1,000	2,000	500	1,000	2,000
modified LB	6.6	6.8	7.4	3.1	4.5	5.4
standard LB	40.9	40.1	42.6	26.0	26.4	29.9

n	$m = 3$			$m = 4$		
	500	1,000	2,000	500	1,000	2,000
modified LB	2.4	4.2	3.2	1.3	4.7	4.0
standard LB	22.3	24.7	28.9	11.7	22.6	26.5

In two other sets of experiments, we assume that in (4.13) the innovation process (ϵ_t) is defined by

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \eta_{1,t}\eta_{2,t-1}\eta_{1,t-2} \\ \eta_{2,t}\eta_{1,t-1}\eta_{2,t-2} \end{pmatrix}, \quad \text{with} \quad \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix} \sim \text{IID } \mathcal{N}(0, I_2), \quad (4.16)$$

and then by

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \eta_{1,t}(|\eta_{1,t-1}| + 1)^{-1} \\ \eta_{2,t}(|\eta_{2,t-1}| + 1)^{-1} \end{pmatrix}, \quad \text{with} \quad \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix} \sim \text{IID } \mathcal{N}(0, I_2), \quad (4.17)$$

These noises are direct extensions of the weak noises defined by Romano and Thombs (1996) in the univariate case. Table 4.5 shows that, once again, the standard **LB** test

poorly performs to assess the adequacy of this weak VARMA(1, 1) model. In view of the observed relative rejection frequency, the standard **LB** test rejects very often the true VARMA(1, 1), as in Table 4.3. By contrast, the error of first kind is well controlled by the modified version of the **LB** test. We draw again the conclusion that, for this particular weak VARMA model, the modified version is clearly preferable to the standard one.

By contrast, Table 4.6 shows that the error of first kind is well controlled by all the tests in this particular weak VARMA model, except for the standard **LB** test when $m = 1$. On this particular example, the two versions of the **LB** test are almost equivalent when $m > 1$, but the modified version clearly outperforms the standard version when $m = 1$.

TABLE 4.5 – Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the weak VARMA(1, 1) model (4.13)-(4.16), with $\theta_0 = (0.225, -0.313, 0.750)$.

	$m = 1$			$m = 2$		
n	500	1,000	2,000	500	1,000	2,000
modified LB	2.4	2.4	4.1	4.0	3.3	3.2
standard LB	71.8	72.3	72.2	62.9	64.7	64.8

	$m = 3$			$m = 4$		
n	500	1,000	2,000	500	1,000	2,000
modified LB	2.6	2.7	2.4	2.6	1.4	1.7
standard LB	54.7	54.2	58.5	48.4	50.7	51.0

TABLE 4.6 – Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the weak VARMA(1, 1) model (4.13)-(4.17), with $\theta_0 = (0.225, -0.313, 0.750)$.

	$m = 1$			$m = 2$		
n	500	1,000	2,000	500	1,000	2,000
modified LB	5.0	4.4	4.9	4.9	3.6	5.1
standard LB	13.6	11.3	12.5	6.1	4.3	5.6

	$m = 3$			$m = 4$			$m = 6$		
n	500	1,000	2,000	500	1,000	2,000	500	1,000	2,000
modified LB	4.5	4.3	5.6	4.3	4.2	5.5	3.7	4.1	4.3
standard LB	5.2	4.0	5.8	4.6	4.1	5.1	3.9	4.0	3.8

4.9.2 Empirical power

In this part, we simulated $N = 1,000$ independent trajectories of size $n = 500$, $n = 1,000$ and $n = 2,000$ of a weak VARMA(2, 2) defined by

$$\begin{aligned} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} &= \begin{pmatrix} 0 & 0 \\ 0 & a_1(2, 2) \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & a_2(2, 2) \end{pmatrix} \begin{pmatrix} X_{1,t-2} \\ X_{2,t-2} \end{pmatrix} \\ &+ \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ b_1(2, 1) & b_1(2, 2) \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix} \\ &- \begin{pmatrix} 0 & 0 \\ b_2(2, 1) & b_2(2, 2) \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-2} \\ \epsilon_{2,t-2} \end{pmatrix}, \end{aligned} \tag{4.18}$$

where the innovation process (ϵ_t) is an ARCH(1) model given by (4.15) and where

$$\begin{aligned} &\{a_1(2, 2), a_2(2, 2), b_1(2, 1), b_1(2, 2), b_2(2, 1), b_2(2, 2)\} \\ &= (0.225, 0.061, -0.313, 0.750, -0.140, -0.160). \end{aligned}$$

For each of these N replications we fitted a VARMA(1, 1) model and performed standard and modified **LB** tests based on $m = 1, \dots, 4$ residual autocorrelations. The adequacy of the VARMA(1, 1) model is rejected when the p -value is less than 5%. Table 4.7 displays the relative rejection frequencies over the $N = 1,000$ independent replications. In this example, the standard and modified versions of the **LB** test have similar powers, except for $n = 500$. One could think that the modified version is slightly less powerful than the standard version. Actually, the comparison made in Table 4.7 is not fair because the actual level of the standard version is generally very greater than the 5% nominal level for this particular weak VARMA model (see Table 4.3).

TABLE 4.7 – Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the weak VARMA(2, 2) model (4.18)-(4.15).

n	$m = 1$			$m = 2$		
	500	1,000	2,000	500	1,000	2,000
modified LB	56.0	85.0	96.7	63.7	89.5	97.3
standard LB	98.2	100.0	100.0	97.1	99.9	100.0

n	$m = 3$			$m = 4$		
	500	1,000	2,000	500	1,000	2,000
modified LB	59.6	91.1	97.2	51.0	89.3	97.8
standard LB	97.5	100.0	100.0	97.0	100.0	100.0

Table 4.8 displays the relative rejection frequencies among the $N = 1,000$ independent replications. In this example, the standard and modified versions of the **LB** test have very similar powers.

TABLE 4.8 – Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the strong VARMA(2, 2) model (4.18)-(4.14).

n	$m = 1$			$m = 2$		
	500	1,000	2,000	500	1,000	2,000
modified LB	93.1	100.0	100.0	96.2	100.0	100.0
standard LB	99.4	100.0	100.0	97.4	100.0	100.0

n	$m = 3$			$m = 4$		
	500	1,000	2,000	500	1,000	2,000
modified LB	97.1	100.0	100.0	96.3	100.0	100.0
standard LB	97.7	100.0	100.0	97.4	100.0	100.0

As a general conclusion concerning the previous numerical experiments, one can say that the empirical sizes of the two versions are comparable, but the error of first kind is better controlled by the modified version than by the standard one. As expected, this latter feature holds for weak VARMA, but, more surprisingly, it is also true for strong VARMA models when m is small.

4.10 Appendix

For the proof of Theorem 4.1, we need respectively, the following lemmas on the standard matrices derivatives and on the covariance inequality obtained by Davydov (1968).

Lemma 4.5. *If $f(A)$ is a scalar function of a matrix A whose elements a_{ij} are function of a variable x , then*

$$\frac{\partial f(A)}{\partial x} = \sum_{i,j} \frac{\partial f(A)}{\partial a_{ij}} \frac{\partial a_{ij}}{\partial x} = \text{Tr} \left\{ \frac{\partial f(A)}{\partial A'} \frac{\partial A}{\partial x} \right\}. \quad (4.19)$$

When A is invertible, we also have

$$\frac{\partial \log |\det(A)|}{\partial A'} = A^{-1} \quad (4.20)$$

Lemma 4.6 (Davydov (1968)). *Let p, q and r three positive numbers such that $p^{-1} + q^{-1} + r^{-1} = 1$. Davydov (1968) showed that*

$$|\text{Cov}(X, Y)| \leq K_0 \|X\|_p \|Y\|_q [\alpha \{\sigma(X), \sigma(Y)\}]^{1/r}, \quad (4.21)$$

where $\|X\|_p^p = EX^p$, K_0 is an universal constant, and $\alpha \{\sigma(X), \sigma(Y)\}$ denotes the strong mixing coefficient between the σ -fields $\sigma(X)$ and $\sigma(Y)$ generated by the random variables X and Y , respectively.

Proof of Theorem 4.1. Recall that

$$Q_n(\theta) = \log \det \left\{ \frac{1}{n} \sum_{t=1}^n \tilde{e}_t(\theta) \tilde{e}_t'(\theta) \right\} \quad \text{and} \quad O_n(\theta) = \log \det \left\{ \frac{1}{n} \sum_{t=1}^n e_t(\theta) e_t'(\theta) \right\}.$$

In view of Theorem 1 in Boubacar Mainassara and Francq (2009) and **A6**, we have almost surely $\hat{\theta}_n \rightarrow \theta_0 \in \mathring{\Theta}$. Thus $\partial Q_n(\hat{\theta}_n)/\partial \theta = 0$ for sufficiently large n , and a standard Taylor expansion of the derivative of Q_n about θ_0 , taken at $\hat{\theta}_n$, yields

$$\begin{aligned} 0 &= \sqrt{n} \frac{\partial Q_n(\hat{\theta}_n)}{\partial \theta} = \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} + \frac{\partial^2 Q_n(\theta^*)}{\partial \theta \partial \theta'} \sqrt{n} (\hat{\theta}_n - \theta_0) \\ &= \sqrt{n} \frac{\partial O_n(\theta_0)}{\partial \theta} + \frac{\partial^2 O_n(\theta_0)}{\partial \theta \partial \theta'} \sqrt{n} (\hat{\theta}_n - \theta_0) + o_P(1), \end{aligned} \quad (4.22)$$

using arguments given in FZ (proof of Theorem 2), where θ^* is between θ_0 and θ_n . Thus, by standard arguments, we have from (4.22) :

$$\begin{aligned} \sqrt{n} (\hat{\theta}_n - \theta_0) &= -J^{-1} \sqrt{n} \frac{\partial O_n(\theta_0)}{\partial \theta} + o_P(1) \\ &= J^{-1} \sqrt{n} Y_n + o_P(1) \end{aligned}$$

where

$$Y_n = -\frac{\partial O_n(\theta_0)}{\partial \theta} = -\frac{\partial}{\partial \theta} \log \det \left\{ \frac{1}{n} \sum_{t=1}^n e_t(\theta_0) e_t'(\theta_0) \right\}.$$

Showing that the initial values are asymptotically negligible, and using (4.19) and (4.20), we have

$$\frac{\partial O_n(\theta)}{\partial \theta_i} = \text{Tr} \left\{ \frac{\partial \log |\Sigma_n|}{\partial \Sigma_n} \frac{\partial \Sigma_n}{\partial \theta_i} \right\} = \text{Tr} \left\{ \Sigma_n^{-1} \frac{\partial \Sigma_n}{\partial \theta_i} \right\}, \quad (4.23)$$

with

$$\frac{\partial \Sigma_n}{\partial \theta_i} = \frac{2}{n} \sum_{t=1}^n e_t(\theta) \frac{\partial e_t'(\theta)}{\partial \theta_i}.$$

Then, for $1 \leq i \leq k_0 = (p+q+2)d^2$, the i -th coordinate of the vector Y_n is of the form

$$Y_n^{(i)} = -\text{Tr} \left\{ \frac{2}{n} \sum_{t=1}^n \Sigma_{e0}^{-1} e_t(\theta_0) \frac{\partial e_t'(\theta_0)}{\partial \theta_i} \right\}, \quad \Sigma_{e0} = \Sigma_n(\theta_0).$$

It is easily shown that for $\ell, \ell' \geq 1$,

$$\begin{aligned} \text{Cov}(\sqrt{n} \text{vec } \gamma(\ell), \sqrt{n} \text{vec } \gamma(\ell')) &= \frac{1}{n} \sum_{t=\ell+1}^n \sum_{t'=\ell'+1}^n E(\{e_{t-\ell} \otimes e_t\} \{e_{t'-\ell'} \otimes e_{t'}\}') \\ &\rightarrow \Gamma(\ell, \ell') \quad \text{as } n \rightarrow \infty, \end{aligned}$$

Then, we have

$$\Sigma_{\gamma_m} = \{\Gamma(\ell, \ell')\}_{1 \leq \ell, \ell' \leq m}$$

$$\begin{aligned} \text{Cov}(\sqrt{n}J^{-1}Y_n, \sqrt{n} \text{vec } \gamma(\ell)) &= - \sum_{t=\ell+1}^n J^{-1}E \left(\frac{\partial O_n}{\partial \theta} \{e_{t-\ell} \otimes e_t\}' \right) \\ &\rightarrow - \sum_{h=-\infty}^{+\infty} 2J^{-1}E \left(\mathcal{E}_t \{e_{t-h-\ell} \otimes e_{t-h}\}' \right), \end{aligned}$$

where $\mathcal{E}_t = \left(\left(\text{Tr} \left\{ \Sigma_{e0}^{-1} e_t(\theta_0) \frac{\partial e_t'(\theta_0)}{\partial \theta_1} \right\} \right)', \dots, \left(\text{Tr} \left\{ \Sigma_{e0}^{-1} e_t(\theta_0) \frac{\partial e_t'(\theta_0)}{\partial \theta_{k_0}} \right\} \right)' \right)'$. Then, we have

$$\begin{aligned} \text{Cov}(\sqrt{n}J^{-1}Y_n, \sqrt{n}\gamma_m) &\rightarrow - \sum_{h=-\infty}^{+\infty} 2J^{-1}E \left\{ \mathcal{E}_t \left(\begin{pmatrix} e_{t-h-1} \\ \vdots \\ e_{t-h-m} \end{pmatrix} \otimes e_{t-h} \right)' \right\} \\ &= \Sigma'_{\gamma_m, \hat{\theta}_n} \end{aligned}$$

Applying the central limit theorem (CLT) for mixing processes (see Herrndorf, 1984) we directly obtain

$$\begin{aligned} \text{Var}_{as}(\sqrt{n}J^{-1}Y_n) &= J^{-1}IJ^{-1} \\ &= \Sigma_{\hat{\theta}_n} \end{aligned}$$

which shows the asymptotic covariance matrix of Theorem 4.1. It is clear that the existence of these matrices is ensured by the Davydov (1968) inequality (4.21) in Lemma 4.6. The proof is complete. \square

Proof of Theorem 4.2. Recall that

$$e_t(\theta) = X_t - \sum_{i=1}^{\infty} C_i(\theta)X_{t-i} = \mathbf{B}_{\theta}^{-1}(L)\mathbf{A}_{\theta}(L)X_t$$

where $\mathbf{A}_{\theta}(L) = I_d - \sum_{i=1}^p \mathbf{A}_i L^i$ and $\mathbf{B}_{\theta}(L) = I_d - \sum_{i=1}^q \mathbf{B}_i L^i$ with $\mathbf{A}_i = A_0^{-1}A_i$ and $\mathbf{B}_i = A_0^{-1}B_i B_0^{-1}A_0$. For $\ell = 1, \dots, p$ and $\ell' = 1, \dots, q$, let $\mathbf{A}_{\ell} = (\mathbf{a}_{ij, \ell})$ and $\mathbf{B}_{\ell'} = (\mathbf{b}_{ij, \ell'})$.

We define the matrices $\mathbf{A}_{ij, h}^*$ and $\mathbf{B}_{ij, h}^*$ by

$$\mathbf{B}_{\theta}^{-1}(z)E_{ij} = \sum_{h=0}^{\infty} \mathbf{A}_{ij, h}^* z^h, \quad \mathbf{B}_{\theta}^{-1}(z)E_{ij}\mathbf{B}_{\theta}^{-1}(z)\mathbf{A}_{\theta}(z) = \sum_{h=0}^{\infty} \mathbf{B}_{ij, h}^* z^h, \quad |z| \leq 1$$

for $h \geq 0$, where $E_{ij} = \partial \mathbf{A}_\ell / \partial \mathbf{a}_{ij,\ell} = \partial \mathbf{B}_{\ell'} / \partial \mathbf{b}_{ij,\ell'}$ is the $d \times d$ matrix with 1 at position (i, j) and 0 elsewhere. Take $\mathbf{A}_{ij,h}^* = \mathbf{B}_{ij,h}^* = 0$ when $h < 0$. For any $\mathbf{a}_{ij,\ell}$ and $\mathbf{b}_{ij,\ell'}$ writing respectively the multivariate noise derivatives

$$\frac{\partial e_t}{\partial \mathbf{a}_{ij,\ell}} = -\mathbf{B}_\theta^{-1}(L) E_{ij} X_{t-\ell} = -\sum_{h=0}^{\infty} \mathbf{A}_{ij,h}^* X_{t-h-\ell} \quad (4.24)$$

and

$$\frac{\partial e_t}{\partial \mathbf{b}_{ij,\ell'}} = \mathbf{B}_\theta^{-1}(L) E_{ij} \mathbf{B}_\theta^{-1}(L) \mathbf{A}_\theta(L) X_{t-\ell'} = \sum_{h=0}^{\infty} \mathbf{B}_{ij,h}^* X_{t-h-\ell'}. \quad (4.25)$$

On the other hand, considering $\hat{\Gamma}(h)$ and $\gamma(h)$ as values of the same function at the points $\hat{\theta}_n$ and θ_0 , a Taylor expansion about θ_0 gives

$$\begin{aligned} \text{vec } \hat{\Gamma}_e(h) &= \text{vec } \gamma(h) + \frac{1}{n} \sum_{t=h+1}^n \left\{ e_{t-h}(\theta) \otimes \frac{\partial e_t(\theta)}{\partial \theta'} \right. \\ &\quad \left. + \frac{\partial e_{t-h}(\theta)}{\partial \theta'} \otimes e_t(\theta) \right\}_{\theta=\theta_n^*} (\hat{\theta}_n - \theta_0) + O_P(1/n) \\ &= \text{vec } \gamma(h) + E \left(e_{t-h}(\theta_0) \otimes \frac{\partial e_t(\theta_0)}{\partial \theta'} \right) (\hat{\theta}_n - \theta_0) + O_P(1/n), \end{aligned}$$

where θ_n^* is between $\hat{\theta}_n$ and θ_0 . The last equality follows from the consistency of $\hat{\theta}_n$ and the fact that $(\partial e_{t-h} / \partial \theta')(\theta_0)$ is not correlated with e_t when $h \geq 0$. Then for $h = 1, \dots, m$,

$$\hat{\Gamma}_m := \left(\left\{ \text{vec } \hat{\Gamma}_e(1) \right\}', \dots, \left\{ \text{vec } \hat{\Gamma}_e(m) \right\}' \right)' = \gamma_m + \Phi_m (\hat{\theta}_n - \theta_0) + O_P(1/n),$$

where

$$\Phi_m = E \left\{ \left(\begin{array}{c} e_{t-1} \\ \vdots \\ e_{t-m} \end{array} \right) \otimes \frac{\partial e_t(\theta_0)}{\partial \theta'} \right\}. \quad (4.26)$$

In Φ_m , one can express $(\partial e_t / \partial \theta')(\theta_0)$ in terms of the multivariate derivatives (4.24) and (4.25). From Theorem 4.1, we have obtained the asymptotic joint distribution of γ_m and $\hat{\theta}_n - \theta_0$, which shows that the asymptotic distribution of $\sqrt{n} \hat{\Gamma}_m$, is normal, with mean zero and covariance matrix

$$\begin{aligned} \text{Var}_{as}(\sqrt{n} \hat{\Gamma}_m) &= \text{Var}_{as}(\sqrt{n} \gamma_m) + \Phi_m \text{Var}_{as}(\sqrt{n}(\hat{\theta}_n - \theta_0)) \Phi_m' \\ &\quad + \Phi_m \text{Cov}_{as}(\sqrt{n}(\hat{\theta}_n - \theta_0), \sqrt{n} \gamma_m) \\ &\quad + \text{Cov}_{as}(\sqrt{n} \gamma_m, \sqrt{n}(\hat{\theta}_n - \theta_0)) \Phi_m' \\ &= \Sigma_{\gamma_m} + \Phi_m \Sigma_{\hat{\theta}_n} \Phi_m' + \Phi_m \Sigma_{\hat{\theta}_n, \gamma_m} + \Sigma_{\hat{\theta}_n, \gamma_m}' \Phi_m'. \end{aligned}$$

From a Taylor expansion about θ_0 of $\text{vec } \hat{\Gamma}_e(0)$ we have, $\text{vec } \hat{\Gamma}_e(0) = \text{vec } \gamma(0) + O_P(n^{-1/2})$. Moreover, $\sqrt{n}(\text{vec } \gamma(0) - E \text{vec } \gamma(0)) = O_P(1)$ by the CLT for mixing processes. Thus

$\sqrt{n}(\hat{S}_e \otimes \hat{S}_e - S_e \otimes S_e) = O_P(1)$ and, using (4.5) and the ergodic theorem, we obtain

$$\begin{aligned}
 & n \left\{ \text{vec}(\hat{S}_e^{-1} \hat{\Gamma}_e(h) \hat{S}_e^{-1}) - \text{vec}(S_e^{-1} \hat{\Gamma}_e(h) S_e^{-1}) \right\} \\
 &= n \left\{ (\hat{S}_e^{-1} \otimes \hat{S}_e^{-1}) \text{vec} \hat{\Gamma}_e(h) - (S_e^{-1} \otimes S_e^{-1}) \text{vec} \hat{\Gamma}_e(h) \right\} \\
 &= n \left\{ (\hat{S}_e \otimes \hat{S}_e)^{-1} \text{vec} \hat{\Gamma}_e(h) - (S_e \otimes S_e)^{-1} \text{vec} \hat{\Gamma}_e(h) \right\} \\
 &= (\hat{S}_e \otimes \hat{S}_e)^{-1} \sqrt{n} (S_e \otimes S_e - \hat{S}_e \otimes \hat{S}_e) (S_e \otimes S_e)^{-1} \sqrt{n} \text{vec} \hat{\Gamma}_e(h) \\
 &= O_P(1).
 \end{aligned}$$

In the previous equalities we also use $\text{vec}(ABC) = (C' \otimes A) \text{vec}(B)$ and $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ when A and B are invertible. It follows that

$$\begin{aligned}
 \hat{\rho}_m &= \left(\left\{ \text{vec} \hat{R}_e(1) \right\}', \dots, \left\{ \text{vec} \hat{R}_e(m) \right\}' \right)' \\
 &= \left(\left\{ (\hat{S}_e \otimes \hat{S}_e)^{-1} \text{vec} \hat{\Gamma}_e(1) \right\}', \dots, \left\{ (\hat{S}_e \otimes \hat{S}_e)^{-1} \text{vec} \hat{\Gamma}_e(m) \right\}' \right)' \\
 &= \left\{ I_m \otimes (\hat{S}_e \otimes \hat{S}_e)^{-1} \right\} \hat{\Gamma}_m = \left\{ I_m \otimes (S_e \otimes S_e)^{-1} \right\} \hat{\Gamma}_m + O_P(n^{-1}).
 \end{aligned}$$

We now obtain (4.6) from (4.5). Hence, we have

$$\text{Var}(\sqrt{n} \hat{\rho}_m) = \left\{ I_m \otimes (S_e \otimes S_e)^{-1} \right\} \Sigma_{\hat{\Gamma}_m} \left\{ I_m \otimes (S_e \otimes S_e)^{-1} \right\}.$$

The proof is complete. \square

Proof of Theorem 4.4. The proof is similar to that given by Francq, Roy and Zakoïan (2005) for Theorem 5.1. \square

TABLE 4.9 – Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the strong VARMA(1,1) model (4.13)-(4.14).

Model	Length n	Level	Standard Test		Modified Test	
			$m = 1$	$m = 2$	$m = 1$	$m = 2$
I	500	$\alpha = 1\%$	5.7	1.9	1.1	0.9
		$\alpha = 5\%$	22.0	7.1	5.5	4.1
		$\alpha = 10\%$	35.8	14.7	11.1	9.0
I	1,000	$\alpha = 1\%$	6.1	1.5	1.6	0.7
		$\alpha = 5\%$	21.3	7.9	5.1	4.6
		$\alpha = 10\%$	35.5	15.0	10.5	9.6
I	2,000	$\alpha = 1\%$	4.2	1.9	0.9	0.8
		$\alpha = 5\%$	21.7	7.5	3.6	4.2
		$\alpha = 10\%$	36.2	15.5	8.6	9.2

I : Strong VARMA(1,1) model (4.13)-(4.14) with $\theta_0 = (0.225, -0.313, 0.750)$

Model	Length n	Level	Standard Test			Modified Test		
			$m = 3$	$m = 4$	$m = 6$	$m = 3$	$m = 4$	$m = 6$
I	500	$\alpha = 1\%$	1.1	0.7	1.0	0.6	0.4	0.3
		$\alpha = 5\%$	5.9	4.9	5.3	4.2	3.0	3.3
		$\alpha = 10\%$	11.5	10.3	11.1	8.7	8.2	9.0
I	1,000	$\alpha = 1\%$	1.1	0.8	0.7	0.3	0.3	0.5
		$\alpha = 5\%$	5.8	5.2	5.0	4.4	3.9	3.9
		$\alpha = 10\%$	12.0	10.1	9.4	9.3	8.7	8.6
I	2,000	$\alpha = 1\%$	0.6	1.2	0.8	0.4	0.8	0.7
		$\alpha = 5\%$	5.3	5.2	4.6	3.6	4.2	4.1
		$\alpha = 10\%$	10.7	11.0	10.0	8.3	9.2	9.3

I : Strong VARMA(1,1) model (4.13)-(4.14) with $\theta_0 = (0.225, -0.313, 0.750)$

4.11 References

- Ahn, S. K.** (1988) Distribution for residual autocovariances in multivariate autoregressive models with structured parameterization. *Biometrika* 75, 590–93.
- Andrews, D.W.K.** (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- Bauwens, L., Laurent, S. and Rombouts, J.V.K.** (2006) Multivariate GARCH models : a survey. *Journal of Applied Econometrics* 21, 79–109.
- Boubacar Mainassara, Y. and Francq, C.** (2009) Estimating structural VARMA models with uncorrelated but non-independent error terms. *Working Papers*, <http://perso.univ-lille3.fr/~cfrancq/pub.html>.

TABLE 4.10 – Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the strong VARMA(1,1) model (4.13)-(4.14).

Model	Length n	Level	Standard Test			Modified Test		
			$m = 1$	$m = 2$	$m = 3$	$m = 1$	$m = 2$	$m = 3$
I	500	$\alpha = 1\%$	4.9	1.0	0.7	1.1	0.5	0.3
		$\alpha = 5\%$	18.3	6.0	5.2	5.5	3.8	4.1
		$\alpha = 10\%$	31.5	11.1	9.7	11.1	7.2	7.7
I	1,000	$\alpha = 1\%$	4.8	0.7	0.6	0.9	0.4	0.5
		$\alpha = 5\%$	18.7	6.2	4.5	4.1	3.4	3.3
		$\alpha = 10\%$	29.2	10.7	9.9	9.4	7.1	7.3
I	2,000	$\alpha = 1\%$	3.8	1.2	0.8	0.8	0.7	0.6
		$\alpha = 5\%$	16.9	4.8	3.5	3.0	3.2	2.6
		$\alpha = 10\%$	30.1	11.4	8.6	7.2	6.8	5.7

I : Strong VARMA(1,1) model (4.13)-(4.14) with $\theta_0 = (0, 0, 0)$

- Box, G. E. P. and Pierce, D. A.** (1970) Distribution of residual autocorrelations in autoregressive integrated moving average time series models. *Journal of the American Statistical Association* 65, 1509–26.
- Brockwell, P. J. and Davis, R. A.** (1991) *Time series : theory and methods*. Springer Verlag, New York.
- Chitturi, R. V.** (1974) Distribution of residual autocorrelations in multiple autoregressive schemes. *Journal of the American Statistical Association* 69, 928–934.
- Davydov, Y. A.** (1968) Convergence of Distributions Generated by Stationary Stochastic Processes. *Theory of Probability and Applications* 13, 691–696.
- den Hann, W.J. and Levin, A.** (1997) A Practitioner’s Guide to Robust Covariance Matrix Estimation. In *Handbook of Statistics* 15, Rao, C.R. and G.S. Madhala (eds), 291–341.
- Dufour, J-M., and Pelletier, D.** (2005) Practical methods for modelling weak VARMA processes : identification, estimation and specification with a macroeconomic application. *Technical report, Département de sciences économiques and CIREQ, Université de Montréal, Montréal, Canada*.
- Franco, C. and Raïssi, H.** (2006) Multivariate Portmanteau Test for Autoregressive Models with Uncorrelated but Nonindependent Errors, *Journal of Time Series Analysis* 28, 454–470.
- Franco, C., Roy, R. and Zakoïan, J-M.** (2005) Diagnostic checking in ARMA Models with Uncorrelated Errors, *Journal of the American Statistical Association* 100, 532–544.
- Franco, and Zakoïan, J-M.** (1998) Estimating linear representations of nonlinear processes, *Journal of Statistical Planning and Inference* 68, 145–165.

TABLE 4.11 – Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the weak VARMA(1,1) model (4.13)-(4.15).

Model	Length n	Level	Standard Test		Modified Test	
			$m = 1$	$m = 2$	$m = 1$	$m = 2$
II	500	$\alpha = 1\%$	25.7	18.3	2.0	1.4
		$\alpha = 5\%$	48.3	33.1	6.7	4.6
		$\alpha = 10\%$	62.7	44.6	11.3	9.1
II	1,000	$\alpha = 1\%$	27.5	20.6	2.4	1.5
		$\alpha = 5\%$	50.0	36.5	7.5	4.1
		$\alpha = 10\%$	64.8	47.9	10.9	10.3
II	2,000	$\alpha = 1\%$	29.9	22.6	2.7	1.8
		$\alpha = 5\%$	50.3	39.4	8.5	6.5
		$\alpha = 10\%$	63.7	50.0	14.1	12.9

II : Weak VARMA(1,1) model (4.13)-(4.15) with $\theta_0 = (0.225, -0.313, 0.750)$

Model	Length n	Level	Standard Test			Modified Test		
			$m = 3$	$m = 4$	$m = 6$	$m = 3$	$m = 4$	$m = 6$
II	500	$\alpha = 1\%$	14.0	11.6	10.7	1.1	1.1	1.3
		$\alpha = 5\%$	28.2	24.5	22.0	4.4	2.8	4.0
		$\alpha = 10\%$	38.7	34.4	30.1	7.3	5.9	9.7
II	1,000	$\alpha = 1\%$	17.1	15.0	11.5	1.1	1.0	0.7
		$\alpha = 5\%$	31.3	28.1	22.9	4.4	4.1	3.3
		$\alpha = 10\%$	42.0	38.8	32.6	9.6	8.7	7.4
II	2,000	$\alpha = 1\%$	20.4	17.9	14.8	1.2	1.4	1.3
		$\alpha = 5\%$	35.4	32.3	27.8	5.4	5.5	4.9
		$\alpha = 10\%$	46.0	42.8	36.9	10.9	10.3	9.2

II : Weak VARMA(1,1) model (4.13)-(4.15) with $\theta_0 = (0.225, -0.313, 0.750)$

Francq, and Zakoïan, J-M. (2005) Recent results for linear time series models with non independent innovations. In *Statistical Modeling and Analysis for Complex Data Problems*, Chap. 12 (eds P. DUCHESNE and B. RÉMILLARD). New York : Springer Verlag, 137–161.

Herrndorf, N. (1984) A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables. *The Annals of Probability* 12, 141–153.

Hosking, J. R. M. (1980) The multivariate portmanteau statistic, *Journal of the American Statistical Association* 75, 602–608.

Hosking, J. R. M. (1981a) Equivalent forms of the multivariate portmanteau statistic, *Journal of the Royal Statistical Society B* 43, 261–262.

Hosking, J. R. M. (1981b) Lagrange-tests of multivariate time series models, *Journal of the Royal Statistical Society B* 43, 219–230.

TABLE 4.12 – Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the weak VARMA(1,1) model (4.13)-(4.15).

Model	Length n	Level	Standard Test		Modified Test	
			$m = 1$	$m = 2$	$m = 1$	$m = 2$
II	500	$\alpha = 1\%$	20.5	11.0	2.7	0.8
		$\alpha = 5\%$	40.9	26.0	6.6	3.1
		$\alpha = 10\%$	54.5	35.9	10.4	7.4
II	1,000	$\alpha = 1\%$	21.3	13.7	2.4	1.4
		$\alpha = 5\%$	40.1	26.4	6.8	4.5
		$\alpha = 10\%$	52.9	37.1	11.5	8.1
II	2,000	$\alpha = 1\%$	23.6	15.9	2.5	1.4
		$\alpha = 5\%$	42.6	29.9	7.4	5.4
		$\alpha = 10\%$	53.8	38.6	11.9	9.6

II : Weak VARMA(1,1) model (4.13)-(4.15) with $\theta_0 = (0, 0, 0)$

Model	Length n	Level	Standard Test		Modified Test	
			$m = 3$	$m = 4$	$m = 3$	$m = 4$
II	500	$\alpha = 1\%$	9.2	4.8	0.8	0.2
		$\alpha = 5\%$	22.3	11.7	2.4	1.3
		$\alpha = 10\%$	32.2	17.0	5.7	3.0
II	1,000	$\alpha = 1\%$	12.5	12.1	0.8	1.2
		$\alpha = 5\%$	24.7	22.6	4.2	4.7
		$\alpha = 10\%$	33.4	31.1	8.4	8.4
II	2,000	$\alpha = 1\%$	13.7	13.1	0.9	0.9
		$\alpha = 5\%$	28.9	26.5	3.2	4.0
		$\alpha = 10\%$	39.2	36.6	7.4	8.5

II : Weak VARMA(1,1) model (4.13)-(4.15) with $\theta_0 = (0, 0, 0)$

Jeantheau, T. (1998) Strong consistency of estimators for multivariate ARCH models, *Econometric Theory* 14, 70–86.

Li, W. K. and McLeod, A. I. (1981) Distribution of the residual autocorrelations in multivariate ARMA time series models, *Journal of the Royal Statistical Society B* 43, 231–239.

Lütkepohl, H. (1993) *Introduction to multiple time series analysis*. Springer Verlag, Berlin.

Lütkepohl, H. (2005) *New introduction to multiple time series analysis*. Springer Verlag, Berlin.

Magnus, J.R. and H. Neudecker (1988) *Matrix Differential Calculus with Application in Statistics and Econometrics*. New-York, Wiley.

TABLE 4.13 – Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the weak VARMA(1,1) model (4.13)-(4.16).

Model	Length n	Level	Standard Test		Modified Test	
			$m = 1$	$m = 2$	$m = 1$	$m = 2$
III	5,00	$\alpha = 1\%$	53.0	46.9	0.1	1.0
		$\alpha = 5\%$	71.8	62.9	2.4	4.0
		$\alpha = 10\%$	79.6	70.7	6.5	7.4
III	1,000	$\alpha = 1\%$	54.3	47.4	0.1	1.0
		$\alpha = 5\%$	72.3	64.7	2.4	3.3
		$\alpha = 10\%$	80.4	72.5	6.5	6.9
III	2,000	$\alpha = 1\%$	53.7	48.3	1.0	0.3
		$\alpha = 5\%$	72.2	64.8	4.1	3.2
		$\alpha = 10\%$	79.5	73.0	10.3	7.6
III : Weak VARMA(1,1) model (4.13)-(4.16) with $\theta_0 = (0.225, -0.313, 0.750)$						
Model	Length n	Level	Standard Test		Modified Test	
			$m = 3$	$m = 4$	$m = 3$	$m = 4$
III	500	$\alpha = 1\%$	36.5	31.4	0.5	0.5
		$\alpha = 5\%$	54.7	48.4	2.6	2.6
		$\alpha = 10\%$	63.9	57.4	5.6	5.4
III	1,000	$\alpha = 1\%$	38.8	34.8	0.4	0.3
		$\alpha = 5\%$	54.2	50.7	2.7	1.4
		$\alpha = 10\%$	64.3	59.5	5.7	3.4
III	2,000	$\alpha = 1\%$	40.0	35.1	0.1	0.1
		$\alpha = 5\%$	58.5	51.0	2.4	1.7
		$\alpha = 10\%$	66.5	60.9	6.5	5.2
III : Weak VARMA(1,1) model (4.13)-(4.16) with $\theta_0 = (0.225, -0.313, 0.750)$						

McLeod, A. I. (1978) On the distribution of residual autocorrelations in Box-Jenkins models, *Journal of the Royal Statistical Society B* 40, 296–302.

Newey, W.K., and West, K.D. (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

Reinsel, G. C. (1997) *Elements of multivariate time series Analysis*. Second edition. Springer Verlag, New York.

Romano, J. L. and Thombs, L. A. (1996) Inference for autocorrelations under weak assumptions, *Journal of the American Statistical Association* 91, 590–600.

TABLE 4.14 – Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the weak VARMA(1,1) model (4.13)-(4.17).

Model	Length n	Level	Standard Test		Modified Test	
			$m = 1$	$m = 2$	$m = 1$	$m = 2$
IV	500	$\alpha = 1\%$	2.2	1.4	0.5	0.7
		$\alpha = 5\%$	13.6	6.1	5.0	4.9
		$\alpha = 10\%$	25.6	11.4	11.4	9.5
IV	1,000	$\alpha = 1\%$	1.6	0.7	0.8	0.5
		$\alpha = 5\%$	11.3	4.3	4.4	3.6
		$\alpha = 10\%$	23.2	9.7	9.9	8.7
IV	2,000	$\alpha = 1\%$	1.6	1.5	0.8	1.2
		$\alpha = 5\%$	12.5	5.6	4.9	5.1
		$\alpha = 10\%$	26.6	11.5	10.0	10.1

IV : Weak VARMA(1,1) model (4.13)-(4.17) with $\theta_0 = (0.225, -0.313, 0.750)$

Model	Length n	Level	Standard Test			Modified Test		
			$m = 3$	$m = 4$	$m = 6$	$m = 3$	$m = 4$	$m = 6$
IV	500	$\alpha = 1\%$	0.7	0.7	0.6	0.3	0.2	0.6
		$\alpha = 5\%$	5.2	4.6	3.9	4.5	4.3	3.7
		$\alpha = 10\%$	9.9	8.8	9.2	9.7	9.2	8.4
IV	1,000	$\alpha = 1\%$	0.5	0.6	0.6	0.4	0.6	0.6
		$\alpha = 5\%$	4.0	4.1	4.0	4.3	4.2	4.1
		$\alpha = 10\%$	9.1	8.1	9.0	9.7	8.0	9.3
IV	2,000	$\alpha = 1\%$	1.4	0.8	0.6	1.6	0.8	1.7
		$\alpha = 5\%$	5.8	5.1	3.8	5.6	5.5	4.3
		$\alpha = 10\%$	10.4	11.4	8.4	10.8	9.1	8.8

IV : Weak VARMA(1,1) model (4.13)-(4.17) with $\theta_0 = (0.225, -0.313, 0.750)$

tel-00452032, version 1 - 1 Feb 2010

TABLE 4.15 – Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the weak VARMA(2, 2) model (4.18)-(4.15).

Model	Length n	Level	Standard Test		Modified Test	
			$m = 1$	$m = 2$	$m = 1$	$m = 2$
II	500	$\alpha = 1\%$	91.7	90.7	27.8	40.3
		$\alpha = 5\%$	98.2	97.1	56.0	63.7
		$\alpha = 10\%$	99.4	98.5	69.6	74.6
II	1,000	$\alpha = 1\%$	99.8	99.8	62.1	78.2
		$\alpha = 5\%$	100.0	99.9	85.0	89.5
		$\alpha = 10\%$	100.0	99.9	91.6	93.9
II	2,000	$\alpha = 1\%$	100.0	100.0	91.4	95.0
		$\alpha = 5\%$	100.0	100.0	96.7	97.3
		$\alpha = 10\%$	100.0	100.0	97.4	97.9

II : Weak VARMA(2,2) model (4.18)-(4.15) with $\theta_0 = (4.19)$

Model	Length n	Level	Standard Test		Modified Test	
			$m = 3$	$m = 4$	$m = 3$	$m = 4$
II	500	$\alpha = 1\%$	92.6	90.4	36.2	27.2
		$\alpha = 5\%$	97.5	97.0	59.6	51.0
		$\alpha = 10\%$	98.5	98.4	70.7	62.9
II	1,000	$\alpha = 1\%$	99.9	99.7	79.3	75.0
		$\alpha = 5\%$	100.0	100.0	91.1	89.3
		$\alpha = 10\%$	100.0	100.0	94.5	93.4
II	2,000	$\alpha = 1\%$	100.0	100.0	95.7	96.0
		$\alpha = 5\%$	100.0	100.0	97.2	97.8
		$\alpha = 10\%$	100.0	100.0	98.1	98.3

II : Weak VARMA(1,1) model (4.18)-(4.15) with $\theta_0 = (4.19)$

TABLE 4.16 – Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the strong VARMA(2, 2) model (4.18)-(4.14).

Model	Length n	Level	Standard Test		Modified Test	
			$m = 1$	$m = 2$	$m = 1$	$m = 2$
I	500	$\alpha = 1\%$	93.2	91.0	72.2	86.6
		$\alpha = 5\%$	99.4	97.4	93.1	96.2
		$\alpha = 10\%$	99.8	98.9	97.8	98.1
I	1,000	$\alpha = 1\%$	100.0	99.9	99.5	99.9
		$\alpha = 5\%$	100.0	100.0	100.0	100.0
		$\alpha = 10\%$	100.0	100.0	100.0	100.0
I	2,000	$\alpha = 1\%$	100.0	100.0	100.0	100.0
		$\alpha = 5\%$	100.0	100.0	100.0	100.0
		$\alpha = 10\%$	100.0	100.0	100.0	100.0

I : Strong VARMA(2,2) model (4.18)-(4.14) with $\theta_0 = (4.19)$

Model	Length n	Level	Standard Test		Modified Test	
			$m = 3$	$m = 4$	$m = 3$	$m = 4$
I	500	$\alpha = 1\%$	91.8	89.9	89.2	85.7
		$\alpha = 5\%$	97.7	97.4	97.1	96.3
		$\alpha = 10\%$	99.3	98.9	98.7	98.3
I	1,000	$\alpha = 1\%$	100.0	100.0	100.0	100.0
		$\alpha = 5\%$	100.0	100.0	100.0	100.0
		$\alpha = 10\%$	100.0	100.0	100.0	100.0
I	2,000	$\alpha = 1\%$	100.0	100.0	100.0	100.0
		$\alpha = 5\%$	100.0	100.0	100.0	100.0
		$\alpha = 10\%$	100.0	100.0	100.0	100.0

I : Strong VARMA(1,1) model (4.18)-(4.14) with $\theta_0 = (4.19)$

Chapitre 5

Model selection of weak VARMA models

Abstract We consider the problem of orders selections of vector autoregressive moving-average (VARMA) models under the assumption that the errors are uncorrelated but not necessarily independent. We relax the standard independence assumption to extend the range of application of the VARMA models, and allow to cover linear representations of general nonlinear processes. We propose a modified Akaike information criterion (AIC).

Keywords : AIC, discrepancy, Kullback-Leibler information, QMLE/LSE, order selection, structural representation, weak VARMA models.

5.1 Introduction

The class of vector autoregressive moving-average (VARMA) models and the subclass of vector autoregressive (VAR) models are used in time series analysis and econometrics to describe not only the properties of the individual time series but also the possible cross-relationships between the time series (see Reinsel, 1997, Lütkepohl, 2005, 1993). In a VARMA(p, q) models, the choice of p and q is particularly important because the number of parameters, $(p + q + 3)d^2$ where d is the number of the series, quickly increases with p and q , which entails statistical difficulties. If orders lower than the true orders of the VARMA(p, q) models are selected, the estimate of the parameters will not be consistent and if too high orders are selected, the accuracy of the estimation parameters is likely to be low. This paper is devoted to the problem of the choice (by minimizing an information criterion) of the orders of VARMA models under the assumption that the errors are uncorrelated but not necessarily independent.

Such models are called weak VARMA, by contrast to the strong VARMA models, that are the standard VARMA usually considered in the time series literature and in which the noise is assumed to be iid. Relaxing the independence assumption allows to extend the range of application of the VARMA models, and allows to cover linear representations of general nonlinear processes. The statistical inference of weak ARMA models is mainly limited to the univariate framework (see Francq and Zakoïan, 2005, for a review on weak ARMA models). In the multivariate analysis, important advances have been obtained by Dufour and Pelletier (2005) who study the asymptotic properties of a generalization of the regression-based estimation method proposed by Hannan and Rissanen (1982) under weak assumptions on the innovation process, Francq and Raïssi (2007) who study portmanteau tests for weak VAR models, Boubacar Mainassara and Francq (2009) who study the consistency and the asymptotic normality of the QMLE for weak VARMA models and Boubacar Mainassara (2009a, 2009b) who studies portmanteau tests for weak VARMA models and studies the estimation of the asymptotic variance of QML/LS estimator of weak VARMA models. Dufour and Pelletier (2005) have proposed a modified information criterion

$$\log \det \tilde{\Sigma} + \dim(\gamma) \frac{(\log n)^{1+\delta}}{n}, \quad \delta > 0,$$

which gives a consistent estimates of the orders p and q of a weak VARMA models. Their criterion is a generalization of the information criterion proposed by Hannan and Rissanen (1982).

The choice amongst the models is often made by minimizing an information criterion. The most popular criterion for model selection is the Akaike information criterion (AIC) proposed by Akaike (1973). The AIC was designed to be an approximately unbiased estimator of the expected Kullback-Leibler information of a fitted model. Tsai and Hurvich (1989, 1993) derived a bias correction to the AIC for univariate and multivariate autoregressive time series under the assumption that the errors ϵ_t are independent identically distributed (*i.e.* strong models). The main goal of our paper is to complete the above-mentioned results concerning the statistical analysis of weak VARMA models, by proposing a modified version of the AIC criterion.

The paper is organized as follows. Section 5.2 presents the models that we consider here and summarizes the results on the QMLE/LSE asymptotic distribution obtained by Boubacar Mainassara and Francq (2009). For self-containedness purposes, Section 5.3 recalls useful the results concerning the general multivariate linear regression model, and Section 5.4 presents the definition and mains properties of the Kullback-Leibler divergence. In Section 5.5, we present the AIC_M criterion which we minimize to choose the orders for a weak VARMA(p, q) models. This section is also of interest in the univariate framework because, to our knowledge, this model selection criterion has not been studied for weak ARMA models. We denote by $A \otimes B$ the Kronecker product of two matrices A and B (and by $A^{\otimes 2}$ when the matrix $A = B$), and by $\text{vec}A$ the vector obtained by stacking the columns of A . The reader is referred to Magnus and Neudecker (1988) for the properties of these operators. Let 0_r be the null vector of \mathbb{R}^r , and let I_r

be the $r \times r$ identity matrix.

5.2 Model and assumptions

Consider a d -dimensional stationary process (X_t) satisfying a structural VARMA(p_0, q_0) representation of the form

$$A_{00}X_t - \sum_{i=1}^{p_0} A_{0i}X_{t-i} = B_{00}\epsilon_t - \sum_{i=1}^{q_0} B_{0i}\epsilon_{t-i}, \quad \forall t \in \mathbb{Z} = \{0, \pm 1, \dots\}, \quad (5.1)$$

where ϵ_t is a white noise, namely a stationary sequence of centered and uncorrelated random variables with a non singular variance Σ_0 . The structural forms are mainly used in econometrics to introduce instantaneous relationships between economic variables. Of course, constraints are necessary for the identifiability of these representations. Let $[A_{00} \dots A_{0p_0} B_{00} \dots B_{0q_0} \Sigma_0]$ be the $d \times (p_0 + q_0 + 3)d$ matrix of all the coefficients, without any constraint. The parameter of interest is denoted θ_0 , where θ_0 belongs to the parameter space $\Theta_{p_0, q_0} \subset \mathbb{R}^{k_0}$, and k_0 is the number of unknown parameters, which is typically much smaller than $(p_0 + q_0 + 3)d^2$. The matrices $A_{00}, \dots, A_{0p_0}, B_{00}, \dots, B_{0q_0}$ involved in (5.1) and Σ_0 are specified by θ_0 . More precisely, we write $A_{0i} = A_i(\theta_0)$ and $B_{0j} = B_j(\theta_0)$ for $i = 0, \dots, p_0$ and $j = 0, \dots, q_0$, and $\Sigma_0 = \Sigma(\theta_0)$. We need the following assumptions used by Boubacar Mainassara and Francq (2009) to ensure the consistence and the asymptotic normality of the quasi-maximum likelihood estimator (QMLE).

A1 : The functions $\theta \mapsto A_i(\theta)$ $i = 0, \dots, p$, $\theta \mapsto B_j(\theta)$ $j = 0, \dots, q$ and $\theta \mapsto \Sigma(\theta)$ admit continuous third order derivatives for all $\theta \in \Theta_{p, q}$.

For simplicity we now write A_i, B_j and Σ instead of $A_i(\theta), B_j(\theta)$ and $\Sigma(\theta)$. Let $A_\theta(z) = A_0 - \sum_{i=1}^p A_i z^i$ and $B_\theta(z) = B_0 - \sum_{i=1}^q B_i z^i$.

A2 : For all $\theta \in \Theta_{p, q}$, we have $\det A_\theta(z) \det B_\theta(z) \neq 0$ for all $|z| \leq 1$.

A3 : We have $\theta_0 \in \Theta_{p_0, q_0}$, where Θ_{p_0, q_0} is compact.

A4 : The process (ϵ_t) is stationary and ergodic.

A5 : For all $\theta \in \Theta_{p, q}$ such that $\theta \neq \theta_0$, either the transfer functions

$$A_0^{-1} B_0 B_\theta^{-1}(z) A_\theta(z) \neq A_{00}^{-1} B_{00} B_{\theta_0}^{-1}(z) A_{\theta_0}(z)$$

for some $z \in \mathbb{C}$, or

$$A_0^{-1} B_0 \Sigma B_0' A_0^{-1'} \neq A_{00}^{-1} B_{00} \Sigma_0 B_{00}' A_{00}^{-1'}.$$

A6 : We have $\theta_0 \in \overset{\circ}{\Theta}_{p_0, q_0}$, where $\overset{\circ}{\Theta}_{p_0, q_0}$ denotes the interior of Θ_{p_0, q_0} .

A7 : We have $E\|\epsilon_t\|^{4+2\nu} < \infty$ and $\sum_{k=0}^{\infty} \{\alpha_\epsilon(k)\}^{\frac{\nu}{2+\nu}} < \infty$ for some $\nu > 0$.

The reader is referred to Boubacar Mainassara and Francq (2009) for a discussion of these assumptions. Note that (ϵ_t) can be replaced by (X_t) in **A4**, because $X_t = A_{\theta_0}^{-1}(L)B_{\theta_0}(L)\epsilon_t$ and $\epsilon_t = B_{\theta_0}^{-1}(L)A_{\theta_0}(L)X_t$, where L stands for the backward operator. Note that from **A1** the matrices A_0 and B_0 are invertible. Introducing the innovation process $e_t = A_{00}^{-1}B_{00}\epsilon_t$, the structural representation $A_{\theta_0}(L)X_t = B_{\theta_0}(L)\epsilon_t$ can be rewritten as the reduced VARMA representation

$$X_t - \sum_{i=1}^p A_{00}^{-1}A_{0i}X_{t-i} = e_t - \sum_{i=1}^q A_{00}^{-1}B_{0i}B_{00}^{-1}A_{00}e_{t-i}.$$

We thus recursively define $\tilde{e}_t(\theta)$ for $t = 1, \dots, n$ by

$$\tilde{e}_t(\theta) = X_t - \sum_{i=1}^p A_0^{-1}A_iX_{t-i} + \sum_{i=1}^q A_0^{-1}B_iB_0^{-1}A_0\tilde{e}_{t-i}(\theta),$$

with initial values $\tilde{e}_0(\theta) = \dots = \tilde{e}_{1-q}(\theta) = X_0 = \dots = X_{1-p} = 0$. The gaussian quasi-likelihood is given by

$$\tilde{L}_n(\theta) = \prod_{t=1}^n \frac{1}{(2\pi)^{d/2} \sqrt{\det \Sigma_e}} \exp \left\{ -\frac{1}{2} \tilde{e}_t'(\theta) \Sigma_e^{-1} \tilde{e}_t(\theta) \right\}, \quad \Sigma_e = A_0^{-1}B_0\Sigma B_0'A_0^{-1'}.$$

A quasi-maximum likelihood estimator (QMLE) of θ is a measurable solution $\hat{\theta}_n$ of

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} \tilde{L}_n(\theta).$$

We now use the matrix M_{θ_0} of the coefficients of the reduced form to that made by Boubacar Mainassara and Francq (2009), where

$$M_{\theta_0} = [A_{00}^{-1}A_{01} : \dots : A_{00}^{-1}A_{0p} : A_{00}^{-1}B_{01}B_{00}^{-1}A_{00} : \dots : A_{00}^{-1}B_{0q}B_{00}^{-1}A_{00} : \Sigma_{e0}].$$

Now we need an assumption which specifies how this matrix depends on the parameter θ_0 . Let \dot{M}_{θ_0} be the matrix $\partial \text{vec}(M_{\theta}) / \partial \theta'$ evaluated at θ_0 .

A8 : The matrix \dot{M}_{θ_0} is of full rank k_0 .

Under Assumptions **A1–A8**, Boubacar Mainassara and Francq (2009) showed the consistency ($\hat{\theta}_n \rightarrow \theta_0$ a.s. as $n \rightarrow \infty$) and the asymptotic normality of the QMLE :

$$\sqrt{n} \left(\hat{\theta}_n - \theta_0 \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Omega := J^{-1}IJ^{-1}), \quad (5.2)$$

where $J = J(\theta_0)$ and $I = I(\theta_0)$, with

$$J(\theta) = \lim_{n \rightarrow \infty} \frac{2}{n} \frac{\partial^2}{\partial \theta \partial \theta'} \log L_n(\theta) \quad a.s$$

and

$$I(\theta) = \lim_{n \rightarrow \infty} \text{Var} \frac{2}{\sqrt{n}} \frac{\partial}{\partial \theta} \log L_n(\theta).$$

Note that, for VARMA models in reduced form, it is not very restrictive to assume that the coefficients $A_0, \dots, A_p, B_0, \dots, B_q$ are functionally independent of the coefficient Σ_e . Thus we can write $\theta = (\theta^{(1)'}, \theta^{(2)'})'$, where $\theta^{(1)} \in \mathbb{R}^{k_1}$ depends on A_0, \dots, A_p and B_0, \dots, B_q , and where $\theta^{(2)} \in \mathbb{R}^{k_2}$ depends on Σ_e , with $k_1 + k_2 = k_0$. With some abuse of notation, we will then write $e_t(\theta) = e_t(\theta^{(1)})$.

A9 : With the previous notation $\theta = (\theta^{(1)'}, \theta^{(2)'})'$, where $\theta^{(2)} = D \text{vec } \Sigma_e$ for some matrix D of size $k_2 \times d^2$.

Let J_{11} and I_{11} be respectively the upper-left block of the matrices J and I , with appropriate size. Under Assumptions **A1–A9**, in a working paper, Boubacar Mainassara (2009) obtained explicit expressions of I_{11} and J_{11} , given by

$$\text{vec } J_{11} = 2 \sum_{i \geq 1} \mathcal{M} \{ \lambda'_i \otimes \lambda'_i \} \text{vec } \Sigma_{e0}^{-1} \quad \text{and}$$

$$\text{vec } I_{11} = 4 \sum_{i,j=1}^{+\infty} \Gamma(i, j) \left(\{ I_d \otimes \lambda'_j \} \otimes \{ I_d \otimes \lambda'_i \} \right) \text{vec} \left(\text{vec } \Sigma_{e0}^{-1} \{ \text{vec } \Sigma_{e0}^{-1} \}' \right),$$

where

$$\mathcal{M} := E \left\{ (I_{d^2(p+q)} \otimes e'_t)^{\otimes 2} \right\},$$

the matrices λ_i depend on θ_0 (see Boubacar Mainassara (2009) for the precise definition of these matrices) and

$$\Gamma(i, j) = \sum_{h=-\infty}^{+\infty} E \left(\{ e'_{t-h} \otimes (I_{d^2(p+q)} \otimes e'_{t-j-h}) \} \otimes \{ e'_t \otimes (I_{d^2(p+q)} \otimes e'_{t-i}) \} \right).$$

We first define an estimator \hat{J}_n of J by

$$\text{vec } \hat{J}_n = \sum_{i \geq 1} \hat{\mathcal{M}}_n \left\{ \hat{\lambda}'_i \otimes \hat{\lambda}'_i \right\} \text{vec } \hat{\Sigma}_{e0}^{-1}, \quad \text{where } \hat{\mathcal{M}}_n := \frac{1}{n} \sum_{t=1}^n \left\{ (I_{d^2(p+q)} \otimes \hat{e}'_t)^{\otimes 2} \right\}.$$

To estimate I consider a sequence of real numbers $(b_n)_{n \in \mathbb{N}^*}$ such that

$$b_n \rightarrow 0 \quad \text{and} \quad n b_n^{\frac{10+4\nu}{\nu}} \rightarrow \infty \quad \text{as } n \rightarrow \infty,$$

and a weight function $f : \mathbb{R} \rightarrow \mathbb{R}$ which is bounded, with compact support $[-a, a]$ and continuous at the origin with $f(0) = 1$. Let \hat{I}_n an estimator of I defined by

$$\text{vec } \hat{I}_n = 4 \sum_{i,j=1}^{+\infty} \hat{\Gamma}_n(i, j) \left(\{ I_d \otimes \hat{\lambda}'_i \} \otimes \{ I_d \otimes \hat{\lambda}'_j \} \right) \text{vec} \left(\text{vec } \hat{\Sigma}_{e0}^{-1} \{ \text{vec } \hat{\Sigma}_{e0}^{-1} \}' \right),$$

where

$$\hat{\Gamma}_n(i, j) := \sum_{h=-T_n}^{+T_n} f(h b_n) \hat{\mathcal{M}}_{n \, ij, h} \quad \text{and} \quad T_n = \left\lfloor \frac{a}{b_n} \right\rfloor,$$

where $[x]$ denotes the integer part of x , and where

$$\hat{\mathcal{M}}_{n \, ij, h} := \frac{1}{n} \sum_{t=1}^{n-|h|} \left(\{ \hat{e}'_{t-h} \otimes (I_{d^2(p+q)} \otimes \hat{e}'_{t-j-h}) \} \otimes \{ \hat{e}'_t \otimes (I_{d^2(p+q)} \otimes \hat{e}'_{t-i}) \} \right).$$

5.3 General multivariate linear regression model

Let $Z_t = (Z_{1t}, \dots, Z_{dt})'$ be a d -dimensional random vector of response variables, $X_t = (X_{1t}, \dots, X_{kt})'$ be a k -dimensional input variables and $B = (\beta_1, \dots, \beta_d)$ be a $k \times d$ matrix. We consider a multivariate linear model of the form $Z_{it} = X_t' \beta_i + \epsilon_{it}$, $i = 1, \dots, d$, or $Z_t' = X_t' B + \epsilon_t'$, $t = 1, \dots, n$, where the $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{dt})'$ are uncorrelated and identically distributed random vectors with variance $\Sigma = E \epsilon_t \epsilon_t'$. The i -th column of B (*i.e.* β_i) is the vector of regression coefficients for the i -th response variable. Now, given the n observations Z_1, \dots, Z_n and X_1, \dots, X_n , we define the $n \times d$ data matrix $\mathbf{Z} = (Z_1, \dots, Z_n)'$, the $n \times k$ matrix $\mathbf{X} = (X_1, \dots, X_n)'$ and the $n \times d$ matrix $\varepsilon = (\epsilon_1, \dots, \epsilon_n)'$. Then, we have the multivariate linear model $\mathbf{Z} = \mathbf{X}B + \varepsilon$. Now, it is well known that the QMLE of B is the same as the LSE and, hence, is given by

$$\hat{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}, \quad \text{that is,} \quad \hat{\beta}_i = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_i, \quad i = 1, \dots, d,$$

where $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{in})'$ is the i -th column of \mathbf{Z} . We also have

$$\hat{\varepsilon} := \mathbf{Z} - \mathbf{X}\hat{B} = M_{\mathbf{X}}\mathbf{Z} = \varepsilon - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon = M_{\mathbf{X}}\varepsilon,$$

where $M_{\mathbf{X}} = I_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is a projection matrix. The usual unbiased estimator of the error covariance matrix Σ is

$$\begin{aligned} \Sigma^* &= \frac{1}{n-k} \hat{\varepsilon}'\hat{\varepsilon} = \frac{1}{n-k} (\mathbf{Z} - \mathbf{X}\hat{B})'(\mathbf{Z} - \mathbf{X}\hat{B}) \\ &= \frac{1}{n-k} \sum_{t=1}^n (Z_t - \hat{B}'X_t)(Z_t - \hat{B}'X_t)' \end{aligned}$$

or $\Sigma^* = (n-k)^{-1} \sum_{t=1}^n \hat{\varepsilon}_t \hat{\varepsilon}_t'$, where the $\hat{\varepsilon}_t = Z_t - \hat{B}'X_t$ are the residual vectors. Note that the gaussian quasi-likelihood is given by

$$L_n(B, \Sigma; \mathbf{Z}) = \frac{1}{(2\pi)^{d/2} \sqrt{\det \Sigma_e}} \exp \left\{ -\frac{1}{2} \sum_{t=1}^n (Z_t - B'X_t)' \Sigma^{-1} (Z_t - B'X_t) \right\},$$

whose maximization shows that the QML estimators of B is equal to \hat{B} and that of Σ is $\hat{\Sigma} := n^{-1} \sum_{t=1}^n \hat{\varepsilon}_t \hat{\varepsilon}_t' = (n-k)n^{-1} \Sigma^*$. Because Σ^* is an unbiased estimator of the matrix Σ , by definition we have $E\{\Sigma^*\} = \Sigma$, we then deduce that

$$\frac{n}{n-k} E\{\hat{\Sigma}\} = \frac{1}{n-k} E\hat{\varepsilon}'\hat{\varepsilon} = \frac{1}{n-k} E\varepsilon' M_{\mathbf{X}} \varepsilon = \Sigma. \quad (5.3)$$

5.4 Kullback-Leibler discrepancy

Assume that, with respect to a σ -finite measure μ , the true density of the observations $X = (X_1, \dots, X_n)$ is f_0 , and that some candidate model m gives a density $f_m(\cdot, \theta_m)$

to the observations, where θ_m is a k_m -dimensional parameter. The discrepancy between the candidate and the true models can be measured by the Kullback-Leibler divergence (or information)

$$\Delta \{f_m(\cdot, \theta_m) | f_0\} = E_{f_0} \log \frac{f_0(X)}{f_m(X, \theta_m)} = E_{f_0} \log f_0(X) + \frac{1}{2} d \{f_m(\cdot, \theta_m) | f_0\},$$

where

$$d \{f_m(\cdot, \theta_m) | f_0\} = -2E_{f_0} \log f_m(X, \theta_m) = -2 \int \{\log f_m(x, \theta_m)\} f_0(x) \mu(dx)$$

is sometimes called the Kullback-Leibler contrast (or the discrepancy between the approximating and the true models). Using the Jensen inequality, we have

$$\begin{aligned} \Delta \{f_m(\cdot, \theta_m) | f_0\} &= - \int \log \frac{f_m(x, \theta_m)}{f_0(x)} f_0(x) \mu(dx) \\ &\geq - \log \int \frac{f_m(x, \theta_m)}{f_0(x)} f_0(x) \mu(dx) = 0, \end{aligned}$$

with equality if and only if $f_m(\cdot, \theta_m) = f_0$. This is the main property of the Kullback-Leibler divergence. Minimizing $\Delta \{f_m(\cdot, \theta_m) | f_0\}$ with respect to $f_m(\cdot, \theta_m)$ is equivalent to minimizing the contrast $d \{f_m(\cdot, \theta_m) | f_0\}$. Let

$$\theta_{0,m} = \arg \inf_{\theta_m} d \{f_m(\cdot, \theta_m) | f_0\} = \arg \inf_{\theta_m} -2E \log f_m(X, \theta_m)$$

be an optimal parameter for the model m corresponding to the density $f_m(\cdot, \theta_m)$ (assuming that such a parameter exists). We estimate this optimal parameter by QMLE $\hat{\theta}_{n,m}$.

5.5 Criteria for VARMA order selection

Let

$$\begin{aligned} \tilde{\ell}_n(\theta) &= -\frac{2}{n} \log \tilde{L}_n(\theta) \\ &= \frac{1}{n} \sum_{t=1}^n \{d \log(2\pi) + \log \det \Sigma_e + \tilde{e}'_t(\theta) \Sigma_e^{-1} \tilde{e}_t(\theta)\}. \end{aligned}$$

In Boubacar Mainassara and Francq (2009), it is shown that $\ell_n(\theta) = \tilde{\ell}_n(\theta) + o(1)$ a.s, where

$$\begin{aligned} \ell_n(\theta) &= -\frac{2}{n} \log L_n(\theta) \\ &= \frac{1}{n} \sum_{t=1}^n \{d \log(2\pi) + \log \det \Sigma_e + e'_t(\theta) \Sigma_e^{-1} e_t(\theta)\}, \end{aligned}$$

where

$$e_t(\theta) = A_0^{-1} B_0 B_\theta^{-1}(L) A_\theta(L) X_t.$$

It is also shown uniformly in $\theta \in \Theta_{p,q}$ that

$$\frac{\partial \ell_n(\theta)}{\partial \theta} = \frac{\partial \tilde{\ell}_n(\theta)}{\partial \theta} + o(1) \quad a.s.$$

The same equality holds for the second-order derivatives of $\tilde{\ell}_n$. For all $\theta \in \Theta_{p,q}$, we have

$$-2 \log L_n(\theta) = nd \log(2\pi) + n \log \det \Sigma_e + \sum_{t=1}^n e_t'(\theta) \Sigma_e^{-1} e_t(\theta).$$

In view of Section 5.4, minimizing the Kullback-Leibler information of any approximating (or candidate) model, characterized by the parameter vector θ , is equivalent to minimizing the contrast $\Delta(\theta) = E \{-2 \log L_n(\theta)\}$. Omitting the constant $nd \log(2\pi)$, we find that

$$\Delta(\theta) = n \log \det \Sigma_e + n \text{Tr} (\Sigma_e^{-1} S(\theta)),$$

where $S(\theta) = E e_1(\theta) e_1'(\theta)$. In view of the following Lemma, the function $\theta \mapsto \Delta(\theta)$ is minimal for $\theta = \theta_0$.

Lemma 5.1. *For all $\theta \in \bigcup_{p,q \in \mathbb{N}} \Theta_{p,q}$, we have*

$$\Delta(\theta) \geq \Delta(\theta_0).$$

Proof of Lemma 5.1 : We have

$$\begin{aligned} \Delta(\theta) &= n \log \det \Sigma_e + n \text{Tr} (\Sigma_e^{-1} \{ E e_1(\theta_0) e_1'(\theta_0) + 2 E e_1(\theta_0) \{ e_1(\theta) - e_1(\theta_0) \}' \\ &\quad + E (e_1(\theta) - e_1(\theta_0)) (e_1(\theta) - e_1(\theta_0))' \}). \end{aligned}$$

Now, using the fact that the linear innovation $e_t(\theta_0)$ is orthogonal to the linear past (*i.e.* to the Hilbert space H_{t-1} generated by the linear combinations of the X_u for $u < t$), it follows that $E e_1(\theta_0) \{ e_1(\theta) - e_1(\theta_0) \}' = 0$, since $\{ e_t(\theta) - e_t(\theta_0) \}$ belongs to the linear past H_{t-1} . We thus have

$$\begin{aligned} \Delta(\theta) &= n \log \det \Sigma_e + n \text{Tr} (\Sigma_e^{-1} \Sigma_{e0}) \\ &\quad + n \text{Tr} \{ \Sigma_e^{-1} E (e_1(\theta) - e_1(\theta_0)) (e_1(\theta) - e_1(\theta_0))' \}. \end{aligned}$$

Moreover

$$\begin{aligned} \Delta(\theta_0) &= n \log \det \Sigma_{e0} + n \text{Tr} (\Sigma_{e0}^{-1} S(\theta_0)) = n \log \det \Sigma_{e0} + n \text{Tr} (\Sigma_{e0}^{-1} \Sigma_{e0}) \\ &= n \log \det \Sigma_{e0} + nd. \end{aligned}$$

Thus, we obtain

$$\begin{aligned} \Delta(\theta) - \Delta(\theta_0) &= -n \log \det (\Sigma_e^{-1} \Sigma_{e0}) - nd + n \text{Tr} (\Sigma_e^{-1} \Sigma_{e0}) \\ &\quad + n \text{Tr} \{ \Sigma_e^{-1} E (e_1(\theta) - e_1(\theta_0)) (e_1(\theta) - e_1(\theta_0))' \} \\ &\geq -n \log \det (\Sigma_e^{-1} \Sigma_{e0}) - nd + n \text{Tr} (\Sigma_e^{-1} \Sigma_{e0}), \end{aligned}$$

with equality if and only if $e_1(\theta) = e_1(\theta_0)$ a.s. Using the elementary inequality $\text{Tr}(A^{-1}B) - \log \det(A^{-1}B) \geq \text{Tr}(A^{-1}A) - \log \det(A^{-1}A) = d$ for all symmetric positive semi-definite matrices of order $d \times d$, it is easy to see that $\Delta(\theta) - \Delta(\theta_0) \geq 0$. The proof is complete. \square

5.5.1 Estimating the discrepancy

Let $X = (X_1, \dots, X_n)$ be observation of a process satisfying the VARMA representation (5.1). Let $\hat{\theta}_n$ be the QMLE of the parameter θ of a candidate VARMA model. Let, $\hat{e}_t = \tilde{e}_t(\hat{\theta}_n)$ be the QMLE/LSE residuals when $p > 0$ or $q > 0$, and let $\hat{e}_t = e_t = X_t$ when $p = q = 0$. When $p + q \neq 0$, we have $\hat{e}_t = 0$ for $t \leq 0$ and $t > n$, and

$$\hat{e}_t = X_t - \sum_{i=1}^p A_0^{-1}(\hat{\theta}_n) A_i(\hat{\theta}_n) \hat{X}_{t-i} + \sum_{i=1}^q A_0^{-1}(\hat{\theta}_n) B_i(\hat{\theta}_n) B_0^{-1}(\hat{\theta}_n) A_0(\hat{\theta}_n) \hat{e}_{t-i},$$

for $t = 1, \dots, n$, with $\hat{X}_t = 0$ for $t \leq 0$ and $\hat{X}_t = X_t$ for $t \geq 1$.

In view of Lemma 5.1, it is natural to minimize an estimation of the theoretical criterion $E\Delta(\hat{\theta}_n)$. Note that $E\Delta(\hat{\theta}_n)$ can be interpreted as the average discrepancy when one uses the model of parameter $\hat{\theta}_n$. The Akaike information criterion (AIC) is an approximately unbiased estimator of $E\Delta(\hat{\theta}_n)$. We will adapt to weak VARMA models the corrected AIC version (AICc) introduced by Tsai and Hurvich (1989) for the univariate strong AR models. Under Assumptions **A1–A9**, an approximately unbiased estimator of $E\Delta(\hat{\theta}_n)$ is given by

$$\text{AIC}_M = n \log \det \hat{\Sigma}_e + \frac{n^2 d^2}{nd - k_1} + \frac{nd}{2(nd - k_1)} \left(\text{vec } \hat{I}'_{11,n} \right)' \left(\text{vec } \hat{J}_{11,n}^{-1} \right), \quad (5.4)$$

with $\text{vec } \hat{J}_{11,n}$ and $\text{vec } \hat{I}_{11,n}$ are defined in Section 5.2. The AIC_M stands for AIC "modified".

Justification of (5.4). Using a Taylor expansion of the functions $\partial \log L_n(\hat{\theta}_n) / \partial \theta^{(1)}$ around $\theta_0^{(1)}$, it follows that

$$\left(\hat{\theta}_n^{(1)} - \theta_0^{(1)} \right)^{o_p\left(\frac{1}{\sqrt{n}}\right)} - \frac{2}{n} J_{11}^{-1} \frac{\partial \log L_n(\theta_0)}{\partial \theta^{(1)}} = -\frac{2}{n} \sum_{t=1}^n J_{11}^{-1} \frac{\partial e'_t(\theta_0)}{\partial \theta^{(1)}} \Sigma_{e_0}^{-1} e_t(\theta_0), \quad (5.5)$$

where $a \stackrel{c}{=} b$ signifies $a = b + c$ and where $J_{11} = J_{11}(\theta_0)$ with

$$J_{11}(\theta) = \lim_{n \rightarrow \infty} \frac{2}{n} \frac{\partial^2 \log L_n(\theta)}{\partial \theta^{(1)} \partial \theta^{(1)'}} \quad a.s.$$

We have

$$E\Delta(\hat{\theta}_n) = En \log \det \hat{\Sigma}_e + nE\text{Tr} \left(\hat{\Sigma}_e^{-1} S(\hat{\theta}_n) \right), \quad (5.6)$$

where $\hat{\Sigma}_e = \Sigma_e(\hat{\theta}_n)$ is the estimated error variance matrix under $\hat{\theta}_n$, with $\Sigma_e(\theta) = n^{-1} \sum_{t=1}^n e_t(\theta)e_t'(\theta)$. Then the first term on the right-hand side of (5.6) can be estimated without bias by $n \log \det \left\{ n^{-1} \sum_{t=1}^n e_t(\hat{\theta}_n)e_t'(\hat{\theta}_n) \right\}$. Hence, only an estimate for the second term needs to be considered. Moreover, in view of (5.2), a Taylor expansion of $e_t(\theta)$ around $\theta_0^{(1)}$ yields

$$e_t(\theta) = e_t(\theta_0) + \frac{\partial e_t(\theta_0)}{\partial \theta^{(1)'}} (\theta^{(1)} - \theta_0^{(1)}) + R_t, \quad (5.7)$$

where

$$R_t = \frac{1}{2} (\theta^{(1)} - \theta_0^{(1)})' \frac{\partial^2 e_t(\theta^*)}{\partial \theta^{(1)} \partial \theta^{(1)'}} (\theta^{(1)} - \theta_0^{(1)}) = O_P(\pi^2),$$

with $\pi = \|\theta^{(1)} - \theta_0^{(1)}\|$ and θ^* is between $\theta_0^{(1)}$ and $\theta^{(1)}$. We then obtain

$$\begin{aligned} S(\theta) &= S(\theta_0) + E \left\{ \frac{\partial e_t(\theta_0)}{\partial \theta^{(1)'}} (\theta^{(1)} - \theta_0^{(1)}) e_t'(\theta_0) \right\} + ER_t e_t'(\theta_0) \\ &\quad + E \left\{ e_t(\theta_0) (\theta^{(1)} - \theta_0^{(1)})' \frac{\partial e_t'(\theta_0)}{\partial \theta^{(1)}} \right\} + D(\theta^{(1)}) \\ &\quad + ER_t \left\{ (\theta^{(1)} - \theta_0^{(1)})' \frac{\partial e_t'(\theta_0)}{\partial \theta^{(1)}} \right\} + E e_t(\theta_0) R_t \\ &\quad + E \left\{ \frac{\partial e_t(\theta_0)}{\partial \theta^{(1)'}} (\theta^{(1)} - \theta_0^{(1)}) \right\} R_t + ER_t^2, \end{aligned}$$

where

$$D(\theta^{(1)}) = E \left\{ \frac{\partial e_t(\theta_0)}{\partial \theta^{(1)'}} (\theta^{(1)} - \theta_0^{(1)}) (\theta^{(1)} - \theta_0^{(1)})' \frac{\partial e_t'(\theta_0)}{\partial \theta^{(1)}} \right\}.$$

Using the orthogonality between $e_t(\theta_0)$ and any linear combination of the past values of $e_t(\theta_0)$ (in particular $\partial e_t(\theta_0)/\partial \theta'$ and $\partial^2 e_t(\theta_0)/\partial \theta \partial \theta'$), and the fact that $E e_t(\theta_0) = 0$, we have

$$S(\theta) = S(\theta_0) + D(\theta^{(1)}) + O(\pi^4) = \Sigma_{e0} + D(\theta^{(1)}) + O(\pi^4),$$

where $\Sigma_{e0} = \Sigma_e(\theta_0)$. Thus, we can write the expected discrepancy quantity in (5.6) as

$$\begin{aligned} E\Delta(\hat{\theta}_n) &= En \log \det \hat{\Sigma}_e + nE\text{Tr} \left(\hat{\Sigma}_e^{-1} \Sigma_{e0} \right) + nE\text{Tr} \left(\hat{\Sigma}_e^{-1} D(\hat{\theta}_n^{(1)}) \right) \\ &\quad + nE \left\{ \text{Tr} \left(\hat{\Sigma}_e^{-1} \right) O_P \left(\frac{1}{n^2} \right) \right\}. \end{aligned} \quad (5.8)$$

As in the classical multivariate regression model, an analog of (5.3) is

$$\Sigma_{e0} \approx \frac{n}{n - d(p+q)} E \left\{ \hat{\Sigma}_e \right\} = \frac{dn}{dn - k_1} E \left\{ \hat{\Sigma}_e \right\}.$$

Thus, using the last approximation and from the consistency of $\hat{\Sigma}_e$, we obtain

$$E \left\{ \hat{\Sigma}_e^{-1} \right\} \approx \left\{ E \hat{\Sigma}_e \right\}^{-1} \approx nd(nd - k_1)^{-1} \Sigma_{e0}^{-1}. \quad (5.9)$$

Using the elementary property on the trace, we have

$$\begin{aligned} \text{Tr} \left\{ \Sigma_e^{-1}(\theta) D(\theta_n^{(1)}) \right\} &= \text{Tr} \left(\Sigma_e^{-1}(\theta) E \left\{ \frac{\partial e_t(\theta_0)}{\partial \theta^{(1)'}} (\theta^{(1)} - \theta_0^{(1)}) (\theta^{(1)} - \theta_0^{(1)})' \frac{\partial e_t'(\theta_0)}{\partial \theta^{(1)}} \right\} \right) \\ &= E \left(\text{Tr} \left\{ \frac{\partial e_t'(\theta_0)}{\partial \theta^{(1)}} \Sigma_e^{-1}(\theta) \frac{\partial e_t(\theta_0)}{\partial \theta^{(1)'}} (\theta^{(1)} - \theta_0^{(1)})' (\theta^{(1)} - \theta_0^{(1)}) \right\} \right) \\ &= \text{Tr} \left(E \left\{ \frac{\partial e_t'(\theta_0)}{\partial \theta^{(1)}} \Sigma_e^{-1}(\theta) \frac{\partial e_t(\theta_0)}{\partial \theta^{(1)'}} \right\} (\theta^{(1)} - \theta_0^{(1)})' (\theta^{(1)} - \theta_0^{(1)}) \right). \end{aligned}$$

Now, using (5.2), (5.9) and the last equality, we have

$$\begin{aligned} E \text{Tr} \left\{ \hat{\Sigma}_e^{-1} D(\hat{\theta}_n^{(1)}) \right\} &= \frac{1}{n} \text{Tr} \left(E \left\{ \frac{\partial e_t'(\theta_0)}{\partial \theta^{(1)}} \hat{\Sigma}_e^{-1} \frac{\partial e_t(\theta_0)}{\partial \theta^{(1)'}} \right\} \right. \\ &\quad \left. E n (\hat{\theta}_n^{(1)} - \theta_0^{(1)})' (\hat{\theta}_n^{(1)} - \theta_0^{(1)}) \right) \\ &= \frac{d}{nd - k_1} \text{Tr} \left(E \left\{ \frac{\partial e_t'(\theta_0)}{\partial \theta^{(1)}} \Sigma_{e0}^{-1} \frac{\partial e_t(\theta_0)}{\partial \theta^{(1)'}} \right\} J_{11}^{-1} I_{11} J_{11}^{-1} \right) \\ &= \frac{d}{2(nd - k_1)} \text{Tr} (I_{11} J_{11}^{-1}), \end{aligned}$$

where $J_{11} = 2E \left\{ \frac{\partial e_t'(\theta_0)}{\partial \theta^{(1)}} \Sigma_{e0}^{-1} \frac{\partial e_t(\theta_0)}{\partial \theta^{(1)'}} \right\}$ (see Theorem 3 in Boubacar Mainasara and Francq, 2009). Thus, using (5.9), we have

$$\begin{aligned} E \text{Tr} \left(\hat{\Sigma}_e^{-1} S(\hat{\theta}_n) \right) &= E \text{Tr} \left(\hat{\Sigma}_e^{-1} \Sigma_{e0} \right) + E \text{Tr} \left\{ \hat{\Sigma}_e^{-1} D(\hat{\theta}_n^{(1)}) \right\} \\ &\quad + E \left\{ \text{Tr} \left(\hat{\Sigma}_e^{-1} \right) O_P \left(\frac{1}{n^2} \right) \right\} \\ &= \frac{nd}{nd - k_1} \text{Tr} (\Sigma_{e0}^{-1} \Sigma_{e0}) + \frac{d}{2(nd - k_1)} \text{Tr} (I_{11} J_{11}^{-1}) + O \left(\frac{1}{n^2} \right) \\ &= \frac{nd^2}{nd - k_1} + \frac{d}{2(nd - k_1)} \text{Tr} (I_{11} J_{11}^{-1}) + O \left(\frac{1}{n^2} \right). \end{aligned}$$

Therefore, using the last expression in (5.8), we deduce an approximately unbiased estimator of $E\Delta(\hat{\theta}_n)$ given by

$$\text{AIC}_M = n \log \det \hat{\Sigma}_e + \frac{n^2 d^2}{nd - k_1} + \frac{nd}{2(nd - k_1)} \text{Tr} \left(\hat{I}_{11,n} \hat{J}_{11,n}^{-1} \right),$$

where $\text{vec} \hat{J}_{11,n}$ and $\text{vec} \hat{I}_{11,n}$ are defined in Section 5.2. Using $\text{Tr}(AB) = \text{vec}(A)' \text{vec}(B)$, we then obtain

$$\text{AIC}_M = n \log \det \hat{\Sigma}_e + \frac{n^2 d^2}{nd - k_1} + \frac{nd}{2(nd - k_1)} \left(\text{vec} \hat{I}_{11,n}' \right)' \left(\text{vec} \hat{J}_{11,n}^{-1} \right).$$

The justification is complete. \square

5.5.2 Other decomposition of the discrepancy

In Section 5.5.1, the minimal discrepancy (contrast) has been approximated by $-2E \log L_n(\hat{\theta}_n)$ (the expectation is taken under the true model X). Note that studying this average discrepancy is too difficult because of the dependance between $\hat{\theta}_n$ and X . An alternative slightly different but equivalent interpretation for arriving at the expected discrepancy quantity $E\Delta(\hat{\theta}_n)$, as a criterion for judging the quality of an approximating model, is obtained by supposing $\hat{\theta}_n$ be the QMLE of θ based on the observation X and let $Y = (Y_1, \dots, Y_n)$ be independent observation of a process satisfying the VARMA representation (5.1) (*i.e.* X and Y independent observations satisfying the same process). Then, we may be interested in approximating the distribution of (Y_t) by using $L_n(Y, \hat{\theta}_n)$. So we consider the discrepancy for the approximating model (model Y) that uses $\hat{\theta}_n$ and, thus, it is generally easier to search a model that minimizes

$$C(\hat{\theta}_n) = -2E_Y \log L_n(\hat{\theta}_n), \quad (5.10)$$

where E_Y denotes the expectation under the candidate model Y . Since $\hat{\theta}_n$ and Y are independent, $C(\hat{\theta}_n)$ is the same quantity as the expected discrepancy $E\Delta(\hat{\theta}_n)$. A model minimizing (5.10) can be interpreted as a model that will do globally the best job on an independent copy of X , but this model may not be the best for the data at hand. The average discrepancy can be decomposed into

$$C(\hat{\theta}_n) = -2E_X \log L_n(\hat{\theta}_n) + a_1 + a_2,$$

where

$$a_1 = -2E_X \log L_n(\theta_0) + 2E_X \log L_n(\hat{\theta}_n)$$

and

$$a_2 = -2E_Y \log L_n(\hat{\theta}_n) + 2E_X \log L_n(\theta_0).$$

The QMLE satisfies $\log L_n(\hat{\theta}_n) \geq \log L_n(\theta_0)$ almost surely, thus a_1 can be interpreted as the average over-adjustment (over-fitting) of this QMLE. Now, note that $E_X \log L_n(\theta_0) = E_Y \log L_n(\theta_0)$, thus a_2 can be interpreted as an average cost due to the use of the estimated parameter instead of the optimal parameter, when the model is applied to an independent replication of X . We now discuss the regularity conditions needed for a_1 and a_2 to be equivalent to $\text{Tr}(I_{11}J_{11}^{-1})$ in the following Proposition.

Proposition A 4. Under Assumptions **A1–A9**, a_1 and a_2 are both equivalent to $\text{Tr}(I_{11}J_{11}^{-1})$, as $n \rightarrow \infty$.

Proof of Proposition 4 : Using a Taylor expansion of the quasi log-likelihood, we obtain

$$-2 \log L_n(\theta_0) = -2 \log L_n(\hat{\theta}_n) + n(\hat{\theta}_n^{(1)} - \theta_0^{(1)})' J_{11}(\hat{\theta}_n^{(1)} - \theta_0^{(1)}) + o_P(1).$$

Taking the expectation (under the true model) of both sides, and in view of (5.2) we shown that

$$\begin{aligned} E_X n(\hat{\theta}_n^{(1)} - \theta_0^{(1)})' J_{11}(\hat{\theta}_n^{(1)} - \theta_0^{(1)}) &= \text{Tr} \left\{ J_{11} E_X n(\hat{\theta}_n^{(1)} - \theta_0^{(1)})' (\hat{\theta}_n^{(1)} - \theta_0^{(1)}) \right\} \\ &\rightarrow \text{Tr} (I_{11} J_{11}^{-1}), \end{aligned}$$

we then obtain $a_1 = \text{Tr} (I_{11} J_{11}^{-1}) + o(1)$. Now a Taylor expansion of the discrepancy yields

$$\begin{aligned} \Delta(\hat{\theta}_n) &= \Delta(\theta_0) + (\hat{\theta}_n^{(1)} - \theta_0^{(1)})' \frac{\partial \Delta(\theta)}{\partial \theta^{(1)}} \Big|_{\theta=\theta_0} \\ &\quad + \frac{1}{2} (\hat{\theta}_n^{(1)} - \theta_0^{(1)})' \frac{\partial^2 \Delta(\theta)}{\partial \theta^{(1)} \partial \theta^{(1)'}} \Big|_{\theta=\theta_0} (\hat{\theta}_n^{(1)} - \theta_0^{(1)}) + o_P(1) \\ &= \Delta(\theta_0) + n(\hat{\theta}_n^{(1)} - \theta_0^{(1)})' J_{11}(\hat{\theta}_n^{(1)} - \theta_0^{(1)}) + o_P(1), \end{aligned}$$

assuming that the discrepancy is smooth enough, and that we can take its derivatives under the expectation sign. We then deduce that

$$E_Y - 2 \log L_n(\hat{\theta}_n) = E_X \Delta(\hat{\theta}_n) = E_X \Delta(\theta_0) + \text{Tr} (I_{11} J_{11}^{-1}) + o(1),$$

which shows that a_2 is equivalent to a_1 . The proof is complete. \square

Remark 5.1. In the standard strong VARMA case, i.e. when **A4** is replaced by the assumption that (ϵ_t) is iid, we have $I_{11} = 2J_{11}$, so that $\text{Tr} (I_{11} J_{11}^{-1}) = k_1$. Therefore, in view of Proposition 4, a_1 and a_2 are both equivalent to $k_1 = \dim(\theta_0^{(1)})$. In this case, an approximately unbiased estimator of $E\Delta(\hat{\theta}_n)$ takes the following form

$$\begin{aligned} AIC_M &= n \log \det \hat{\Sigma}_e + \frac{n^2 d^2}{nd - k_1} + \frac{nd}{2(nd - k_1)} 2k_1 \\ &= n \log \det \hat{\Sigma}_e + \frac{nd}{nd - k_1} (nd + k_1) \\ &= n \log \det \hat{\Sigma}_e + nd + \frac{nd}{nd - k_1} 2k_1, \end{aligned} \tag{5.11}$$

which illustrates that the standard AIC and AIC_M differ only through the inclusion of the scale factor $nd/(nd - k_1)$ in the penalty term of AIC_M . This factor can play a substantial role in the performance of AIC_M if k_1 is non negligible fraction of the sample size n . In particular, this factor helps to reduce the bias of AIC, which may be substantial when n is not large. Consequently, use of this improved estimator of the discrepancy should lead to improved performance of AIC_M over AIC in terms of model selection.

Remark 5.2. Under some regularity assumptions given in Section 5.2, it is shown in Findley (1993) that, a_1 and a_2 are both equivalent to k_1 . In this case, the AIC formula

$$AIC = -2 \log L_n(\hat{\theta}_n) + 2k_1 \tag{5.12}$$

is an approximately unbiased estimate of the contrast $C(\hat{\theta}_n)$. Model selection is then obtained by minimizing (5.12) over the candidate models.

Remark 5.3. Given a collection of competing families of approximating models, the one that minimizes $E\Delta(\hat{\theta}_n)$ might be preferred. For model selection, we then choose \hat{p} and \hat{q} as the set which minimizes the information criterion (5.4).

Remark 5.4. Consider the univariate case $d = 1$. We then have

$$AIC_M = n\hat{\sigma}_e^2 + \frac{n}{n - (p + q)} \left(n + \frac{1}{\hat{\sigma}^4} \sum_{i,j,i'=1}^{+\infty} \hat{\gamma}(i,j) \left\{ \hat{\lambda}_j \hat{\lambda}_{i'}^{-1'} \otimes \hat{\lambda}_i \hat{\lambda}_{i'}^{-1'} \right\} \right),$$

where $\hat{\sigma}_e^2$ is the variance estimate of the univariate process and where $\hat{\gamma}(i,j)$ are the estimators of $\gamma(i,j) = \sum_{h=-\infty}^{+\infty} E(e_t e_{t-i} e_{t-h} e_{t-j-h})$ and $\hat{\lambda}'_i$ are the estimators of $\lambda'_i \in \mathbb{R}^{p+q}$ given in Section 5.2.

Proof of Remark 5.4 : For $d = 1$, we have

$$AIC_M = n\hat{\sigma}_e^2 + \frac{n}{n - (p + q)} \left\{ n + \frac{1}{2} \left(\text{vec } \hat{I}'_{11} \right)' \left(\text{vec } \hat{J}_{11}^{-1} \right) \right\}.$$

In view of Section 5.2, we obtained

$$\text{vec } \hat{J}_{11} = 2 \sum_{i' \geq 1} \left\{ \hat{\lambda}_{i'} \otimes \hat{\lambda}_{i'} \right\}' \quad \text{and} \quad \text{vec } \hat{I}_{11} = \frac{4}{\hat{\sigma}^4} \sum_{i,j=1}^{+\infty} \hat{\gamma}(i,j) \left\{ \hat{\lambda}_j \otimes \hat{\lambda}_i \right\}' ,$$

where $\hat{\gamma}(i,j)$ are the estimators of $\gamma(i,j) = \sum_{h=-\infty}^{+\infty} E(e_t e_{t-i} e_{t-h} e_{t-j-h})$ and $\hat{\lambda}'_i$ are the estimators of $\lambda'_i \in \mathbb{R}^{p+q}$ given in Section 5.2. Using the last expressions of $\text{vec } \hat{J}_{11}$ and $\text{vec } \hat{I}_{11}$, we then have

$$AIC_M = n\hat{\sigma}_e^2 + \frac{n}{n - (p + q)} \left(n + \frac{1}{\hat{\sigma}^4} \sum_{i,j,i'=1}^{+\infty} \hat{\gamma}(i,j) \left\{ \hat{\lambda}_j \otimes \hat{\lambda}_i \right\} \left\{ \hat{\lambda}_{i'} \otimes \hat{\lambda}_{i'} \right\}^{-1'} \right).$$

Using $(A \otimes B)(C \otimes D) = AC \otimes BD$, we have

$$AIC_M = n\hat{\sigma}_e^2 + \frac{n}{n - (p + q)} \left(n + \frac{1}{\hat{\sigma}^4} \sum_{i,j,i'=1}^{+\infty} \hat{\gamma}(i,j) \left\{ \hat{\lambda}_j \hat{\lambda}_{i'}^{-1'} \otimes \hat{\lambda}_i \hat{\lambda}_{i'}^{-1'} \right\} \right).$$

The proof is complete. \square

5.6 References

- Akaike, H.** (1973) Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory*, Eds. B. N. Petrov and F. Csáki, pp. 267–281. Budapest : Akadémia Kiado.
- Boubacar Mainassara, Y. and Francq, C.** (2009) Estimating structural VARMA models with uncorrelated but non-independent error terms. *Working Papers*, <http://perso.univ-lille3.fr/cfrancq/pub.html>.
- Brockwell, P. J. and Davis, R. A.** (1991) *Time series : theory and methods*. Springer Verlag, New York.
- Dufour, J-M., and Pelletier, D.** (2005) Practical methods for modelling weak VARMA processes : identification, estimation and specification with a macroeconomic application. *Technical report, Département de sciences économiques and CIREQ, Université de Montréal, Montréal, Canada*.
- Findley, D.F.** (1993) The overfitting principles supporting AIC, Statistical Research Division Report RR 93/04, Bureau of the Census.
- Francq, C. and Raïssi, H.** (2006) Multivariate Portmanteau Test for Autoregressive Models with Uncorrelated but Nonindependent Errors, *Journal of Time Series Analysis* 28, 454–470.
- Francq, C., Roy, R. and Zakoïan, J-M.** (2005) Diagnostic checking in ARMA Models with Uncorrelated Errors, *Journal of the American Statistical Association* 100, 532–544.
- Francq, and Zakoïan, J-M.** (1998) Estimating linear representations of nonlinear processes, *Journal of Statistical Planning and Inference* 68, 145–165.
- Francq, and Zakoïan, J-M.** (2000) Covariance matrix estimation of mixing weak ARMA models, *Journal of Statistical Planning and Inference* 83, 369–394.
- Francq, and Zakoïan, J-M.** (2005) Recent results for linear time series models with non independent innovations. In *Statistical Modeling and Analysis for Complex Data Problems*, Chap. 12 (eds P. DUCHESNE and B. RÉMILLARD). New York : Springer Verlag, 137–161.
- Francq, and Zakoïan, J-M.** (2007) HAC estimation and strong linearity testing in weak ARMA models, *Journal of Multivariate Analysis* 98, 114–144.
- Hannan, E. J. and Rissanen** (1982) Recursive estimation of mixed of Autoregressive Moving Average order, *Biometrika* 69, 81–94.
- Hurvich, C. M. and Tsai, C-L.** (1989) Regression and time series model selection in small samples. *Biometrika* 76, 297–307.
- Hurvich, C. M. and Tsai, C-L.** (1993) A corrected Akaike information criterion for vector autoregressive model selection. *Journal of Time Series Analysis* 14, 271–279.
- Lütkepohl, H.** (1993) *Introduction to multiple time series analysis*. Springer Verlag, Berlin.

- Lütkepohl, H.** (2005) *New introduction to multiple time series analysis*. Springer Verlag, Berlin.
- Magnus, J.R. and H. Neudecker** (1988) *Matrix Differential Calculus with Application in Statistics and Econometrics*. New-York, Wiley.
- Reinsel, G. C.** (1993) *Elements of multivariate time series*. Springer Verlag, New York.
- Reinsel, G. C.** (1997) *Elements of multivariate time series Analysis*. Second edition. Springer Verlag, New York.
- Romano, J. L. and Thombs, L. A.** (1996) Inference for autocorrelations under weak assumptions, *Journal of the American Statistical Association* 91, 590–600.

Conclusion générale

Dans cette thèse nous avons étudié des modèles VARMA faibles dont les termes d'erreur sont non corrélés mais non nécessairement indépendants. Ceci nous permet d'englober de nombreux processus non linéaires existant dans la littérature pour lesquels les hypothèses d'innovations iid ne sont plus valides.

Sous des hypothèses plus large (ergodicité et mélange), nous avons établi la convergence forte et la normalité asymptotique des estimateurs du QML/LS. Nous avons aussi estimé la matrice de variance asymptotique de ces estimateurs sous les mêmes hypothèses. Nous avons notamment montré que le comportement asymptotique des estimateurs peut être très différent du cas iid, si des dépendances existent entre les termes d'erreurs.

Nous avons proposé des versions modifiées des tests de Wald, du multiplicateur de Lagrange et du rapport de vraisemblance pour tester des restrictions linéaires sur les paramètres libres des modèles VARMA faibles. Nous avons également étudié la validité du test LR dans le cas de modèles VARMA forts standard.

Ensuite, nous avons abordé le problème de la validation des modèles VARMA faibles en proposant des tests portmanteau modifiés qui tiennent compte des dépendances des termes d'erreur. Dans un premier temps, nous avons étudié la distribution asymptotique jointe de l'estimateur du QML/LS et des autocovariances empiriques du bruit. Ceci nous permet ensuite d'obtenir les distributions asymptotiques des autocovariances et autocorrelations résiduelles, et aussi des statistiques portmanteau. Ces autocorrelations résiduelles sont normalement distribuées avec une matrice de covariance différente du cas iid. Dans ce cas il est aussi connu que la distribution asymptotique des tests portmanteau est approximée par un chi-deux. Dans le cas général, nous avons montré que cette distribution asymptotique est celle d'une somme pondérée de chi-deux. Cette distribution peut être très différente de l'approximation chi-deux usuelle du cas fort. Les tests portmanteau que nous avons proposé sont fondés sur une distribution exacte de la statistique de test. Enfin, nous avons adapté aux modèles VARMA des méthodes permettant d'obtenir des valeurs critiques de telle distribution.

Nous avons par ailleurs proposé un critère d'information de Akaike modifié pour faciliter les sélections des modèles VARMA faibles.

Signalons que les logiciels actuellement disponibles n'évaluent pas correctement la

précision des estimations des paramètres des représentations VARMA faibles, ce qui peut entraîner de graves erreurs de spécifications, notamment la sélection des modèles avec des ordres très grands. D'une manière générale nous avons justifié une pratique très courante, qui consiste à ajuster un modèle VARMA, même si les données dont on dispose présentent des non linéarités évidentes.

Perspectives

Le travail que nous avons développé dans cette thèse comporte de nombreuses extensions possibles. Parmi ces futures extensions nous citons :

Estimation des modèles VARMA faibles avec tendance déterministe

Une voie de recherche future pour le chapitre 2 est d'étendre l'étude des propriétés asymptotiques du QMLE/LSE, l'estimation de la matrice de variance asymptotique de ces estimateurs et les tests asymptotiques de Wald, du multiplicateur de Lagrange et du rapport de vraisemblance aux modèles VARMA avec une tendance déterministe, dans lesquels on suppose que les erreurs ne sont pas corrélées, mais non nécessairement indépendantes. Cependant, l'une des difficultés est que, en raison de la perte de stationnarité, les distributions asymptotiques du QMLE/LSE ne peuvent pas être calculées de la même manière que celles des processus stationnaires. La deuxième difficulté est liée au fait que les estimations des paramètres de ces modèles avec tendance déterministe ont différentes vitesses de convergence asymptotique. Par conséquent, il faudra des conditions supplémentaires afin d'établir la consistance et la normalité asymptotique de QMLE/LSE.

Sous l'hypothèse que les innovations linéaires sont une différence de martingale conditionnellement homoscédastique, Sims, Stock et Watson (1990) ont proposé une approche afin d'obtenir la distribution asymptotique de l'estimateur des paramètres dans le cas de modèles linéaires de séries temporelles en présence d'une ou de plusieurs racines unitaires.

Le comportement asymptotique de l'estimateur des moindres carrés ordinaires (MCO) d'un modèle de séries temporelles simple avec une tendance déterministe a également été étudié par Hamilton (1994) sous l'hypothèse précédente, c'est-à-dire en supposant que les erreurs sont une différence de martingale conditionnellement homoscédastique. Toutefois, ce genre d'hypothèse est très proche du cas iid et exclut la plupart

des modèles non linéaires (comme les modèles GARCH). C'est pourquoi il semble intéressant d'étudier l'effet de l'ajout de tendances déterministes à des VARMA faibles.

On peut aussi envisager d'étendre ce travail aux modèles VARMA saisonniers avec tendance déterministe. Le cas de tendances stochastiques est également très intéressant pour les applications, mais requiert des techniques complètement différentes.

Tests portmanteau de modèles VARMA faibles non stationnaires

Une voie de recherche future pour le chapitre 4 est d'étendre les tests portmanteau modifiés que nous avons proposés pour tester les ordres p_0 et q_0 des modèles VARMA(p_0, q_0) faibles cointégrés.

Sous des hypothèses que les termes d'erreurs sont iid, Lütkepoh et Claessen (1997) ont étudié la distribution asymptotique de l'estimateur du maximum de vraisemblance des paramètres d'un modèle VARMA sous la forme échelon. Raissi (2009) a récemment obtenu des résultats semblables dans le cas VAR faible. Nous chercherons à étendre ces résultats dans le cas des VARMA faibles.

Critère de sélection de modèles VARMA faibles

Le chapitre 5 pourra être complété de diverses manières. Tout d'abord, il conviendrait de développer des modifications de critères de style BIC (Bayesian Information Criterion, Schwarz 1978) adaptées au cas d'erreurs faibles. Nous savons que le critère BIC aboutit à une procédure convergente de sélection des ordres p et q d'un ARMA(p, q) univarié fort, et que ce n'est pas le cas avec le critère AIC. Qu'en est-il dans le cas VARMA faible? La question mérite d'être examinée.