



HAL
open science

Contribution à l'analyse statistique des modèles ARMA multivariés avec innovations linéaires non indépendantes

Yacouba Boubacar Maïnassara

► **To cite this version:**

Yacouba Boubacar Maïnassara. Contribution à l'analyse statistique des modèles ARMA multivariés avec innovations linéaires non indépendantes. Statistiques [math.ST]. Université de Franche-Comté (Besançon), 2017. tel-02292126

HAL Id: tel-02292126

<https://univ-fcomte.hal.science/tel-02292126v1>

Submitted on 19 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Université Bourgogne Franche-Comté
Laboratoire de Mathématiques de Besançon

Mémoire de synthèse des activités de recherche et d'encadrement présenté en vue de
l'obtention du diplôme

d'HABILITATION À DIRIGER DES RECHERCHES

Spécialité : Mathématiques Appliquées et Applications des Mathématiques

Contribution à l'analyse statistique des modèles ARMA multivariés avec innovations linéaires non indépendantes

présenté par

Yacouba BOUBACAR MAÏNASSARA

23 Novembre 2017

Présenté devant le jury composé de

M. Jean-Marc BARDET,	Université Paris 1-Panthéon-Sorbonne,	Rapporteur,
M. Clément DOMBRY,	Université de Bourgogne Franche-Comté,	Examineur,
M. Pierre DUCHESNE,	Université de Montréal,	Rapporteur,
M. Christian FRANCO,	CREST-ENSAE et Université Lille 3,	Examineur,
M. Guy MÉLARD,	Université Libre de Bruxelles,	Rapporteur,
Mme Anne PHILIPPE,	Université de Nantes,	Examineur,
Mme Delphine BLANKE,	Université d'Avignon,	Rapporteur,
M. Jean-Michel ZAKOÏAN,	CREST-ENSAE et Université Lille 3,	Examineur.

Table des matières

Table des matières	4
Remerciements	5
Liste des tableaux	7
Table des figures	9
Liste des travaux	11
Encadrement de thèses	13
Introduction générale	15
Contexte bibliographique et motivations	15
Contribution apportée	17
1 Contribution à l'estimation des modèles ARMA faibles multivariés	19
1.1 Propriétés asymptotiques des estimateurs du QMV	20
1.1.1 Définition de l'estimateur du QMV	21
1.1.2 Convergence presque sûr de l'estimateur du QMV	22
1.1.3 Normalité asymptotique de l'estimateur du QMV	24
1.2 Estimation de la matrice de variance asymptotique	29
1.2.1 Estimation de la matrice Ω proposée dans [A2, A5]	29
1.2.2 Estimation de la matrice Ω proposée dans [A9]	30
1.2.3 Estimation de la matrice Ω proposée dans [A7]	37
1.2.4 Approximation des matrices d'information par des sommes finies	41
1.2.5 Estimation des matrices d'information	42
1.3 Application numérique	42
1.4 Conclusion	45
2 Contribution à l'identification des modèles ARMA faibles multivariés	47
2.1 Définition du critère d'information	48
2.2 Critère de sélection non convergent des ordres	48
2.2.1 Critère AICc dans le cas d'erreurs faibles	48
2.2.2 Critère AIC dans le cas d'erreurs faibles	51
2.3 Critère de sélection convergent des ordres	53
2.3.1 Critères BIC et HQ modifiés pour VARMA faibles	53

2.3.2	Convergence des critères BIC_c et HQ_c	54
2.4	Illustrations numériques	56
2.5	Conclusion	59
3	Contribution à la validation des modèles ARMA faibles multivariés	61
3.1	Tests sur les coefficients du modèle	61
3.1.1	Statistique de Wald	62
3.1.2	Statistique du LM	62
3.1.3	Statistique du LR	63
3.2	Illustrations numériques	64
3.3	Tests sur les résidus du modèle	65
3.3.1	Comportement asymptotique des statistiques du test portmanteau multivarié	67
3.3.2	Comportement asymptotique des autocorrélations résiduelles auto-normalisées	70
3.3.3	Illustrations numériques	73
3.4	Conclusion	76
4	Contribution aux modèles GARCH multivariés	79
4.1	Modèle, conditions de stationnarité et d'identifiabilité	80
4.1.1	Présentation du modèle	80
4.1.2	Condition de stationnarité stricte	81
4.1.3	Conditions d'identifiabilité	82
4.2	Estimation du modèle	84
4.2.1	Quand la puissance est connue	84
4.2.2	Quand la puissance est inconnue	87
4.2.3	Illustrations numériques	88
4.3	Validation du modèle APGARCH quand la puissance est inconnue	93
4.3.1	Test portmanteau pour le modèle APGARCH quand la puissance est inconnue	94
4.3.2	Illustrations numériques	96
4.4	Conclusion	97
5	Contribution à l'estimation de modèles RCARMA faibles	99
5.1	Stationnarité du modèle RCARMA	100
5.2	Estimation du modèle RCARMA	101
6	Autre thématique étudiée	105
6.1	Loi univariée stable-Tweedie	106
6.2	Modèles NST	106
6.3	Conclusion	109
7	Encadrement, perspectives et responsabilités en recherche	111
7.1	Encadrement et co-encadrement d'étudiants	111
7.2	Travaux en cours et perspectives	112
7.2.1	Modèles GARCH multivariés à seuil en puissance	112
7.2.2	Modèles VARMA faibles stationnaires et/ou non stationnaires	112
7.2.3	Modèles à mémoires longues	113
7.2.4	Perspectives à long terme	114
7.3	Insertion dans l'équipe de recherche	114

Remerciements

Je tiens tout d'abord à exprimer ma profonde gratitude aux rapporteurs de cette habilitation, Jean-Marc BARDET, Pierre DUCHESNE, Guy MÉLARD et Delphine BLANKE de l'honneur qu'ils m'ont fait d'avoir accepté cette lourde tâche d'expertiser mon travail. Je remercie également : Clément DOMBRY, Christian FRANCO, Anne PHILIPPE et Jean-Michel ZAKOÏAN d'avoir accepté de siéger à ce jury.

Je tiens aussi à exprimer ma profonde gratitude à mes directeurs de thèse, Christian FRANCO et Jean-Michel ZAKOÏAN, de la confiance qu'ils m'ont témoigné en acceptant de m'initier à la recherche en mathématiques appliquées. J'ai été très heureux d'avoir pu travailler à leurs côtés et cela a été un véritable enrichissement pour moi. Depuis mes débuts jusqu'à aujourd'hui, ils m'ont toujours soutenu et ont toujours été disponibles à chaque fois que j'en avais eu besoin.

Un grand merci à l'ensemble de mes collaborateurs ; pour nos échanges et pour tout ce qu'ils m'ont apporté. Sans eux, ce mémoire n'aurait pas vu le jour. Particulièrement à Bruno pour nos nombreux moments de fou rire . . .

Un grand merci également à tous les membres du Laboratoire de Mathématiques de Besançon (LmB) qui m'ont accueilli, lors de mon arrivée, dans une ambiance agréable et chaleureuse et qui, collectivement font de ce laboratoire un cadre de travail calme et serein. En particulier, merci aux membres de l'équipe probabilités et statistique surtout pour leur soutien pendant mon mandat de responsable pédagogique du master 1 modélisation statistique de 2013 à 2016.

Je remercie vivement Bruno, Landy, Nabile et Pauline pour leur lecture très attentive de ce mémoire.

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 recherches jusqu'à ce niveau. Je remercie particulièrement mes frères (Moussa et Abdoul-Razak) et mon ami d'enfance Sani, qui m'ont fait un grand plaisir d'être présents à cette soutenance. Une grande pensée, en particulier, pour ma femme (Fati) et nos trois enfants (Boubacar, Yasmina et Yasser).

Je dédie ce travail à mon père décédé en juillet 2008, ma mère et toute ma famille. Enfin, je voudrais dire encore grand merci à tous ceux, famille ou amis, qui m'entourent et à l'ensemble des personnes présentes à cette soutenance.

Liste des tableaux

2.1	Pourcentage de sélection des ordres par les versions du critère AIC standard et modifié. . .	57
2.2	Pourcentage de sélection des ordres par les versions du critère BIC standard et modifié. . .	58
2.3	Pourcentage de sélection des ordres par les versions du critère HQ standard et modifié. . .	59
3.1	Pourcentages de rejet de l'hypothèse nulle, $H_0 : b_{22,1} = 0$, des versions standards et modifiées des tests.	65
3.2	Puissance des tests standards et modifiés : $H_0 : b_{22,1} = 0$	65
3.3	Pourcentage de rejet dans le cas VARMA(1, 1).	75
3.4	Pourcentage de rejet dans le cas VARMA(1, 1).	76
4.1	Distribution de l'estimateur du QMV de θ_0 quand la puissance est connue.	89
4.2	Distribution de l'estimateur du QMV de ν_0 quand la puissance est inconnue.	89
6.1	Récapitulatif des modèles NST k -dimensionnels, de paramètre puissance $p \geq 1$, de paramètre puissance généralisée $q = k + p - 1$, de paramètre de mesure de Lévy modifiée $\eta = [p(k - 1) + 1] / (p - 1)$ et de support de lois \mathbf{S}_p	107

Table des figures

1.1	Distributions des estimateurs du QMV du modèle (1.42) avec le bruit fort (1.43) (les trois blocs de gauche) et avec le bruit faible (1.44) (les trois blocs de droite). Les points (a)-(c), dans les deux blocs supérieurs, présentent les distributions des erreurs $\hat{\theta}(i) - \theta_0(i)$ pour $i = 1, 2, 3$. Les deux blocs du milieu présentent les Q-Q plots des estimateurs $\hat{\theta}(3) = \hat{b}_{22,1}$ et les deux derniers blocs inférieurs présentent les distributions de $\hat{\theta}(3)$. La densité estimée est tracée en trait plein et celle d'une loi gaussienne centrée de même variance en pointillés.	44
1.2	Comparaison des estimateurs standard et modifié de la matrice de variance asymptotique Ω des estimateurs du QMV du modèle (1.42) simulé présenté dans la figure 1.1. Le losange rouge représente la moyenne sur les $N = 1000$ répliquions indépendantes de $n\{\hat{a}_{22,1} - 0,95\}^2$ pour (a) (0,02 dans les cas fort et faible), $n\{\hat{b}_{21,1} - 2\}^2$ pour (b) (1,02 dans le cas fort et 1,01 dans le cas faible) et $n\{\hat{b}_{22,1}\}^2$ for (c) (0,94 dans le cas fort et 0,43 dans le cas faible).	45
4.1	Boîtes à moustaches des distributions de l'estimateur du QMV pour $n = 500$, quand la puissance est estimée.	90
4.2	Boîtes à moustaches des distributions de l'estimateur du QMV pour $n = 500$, quand la puissance est connue.	91
4.3	Boîtes à moustaches des distributions de l'estimateur du QMV pour $n = 5000$, quand la puissance est connue.	92
4.4	Boîtes à moustaches des distributions de l'estimateur du QMV pour $n = 5000$, quand la puissance est estimée.	93

Liste des travaux

La liste respecte l'ordre chronologique de publication des articles mais pas celui de leurs réalisations. Toutes les références du type [Axx] et [CRx] dans ce mémoire font référence à cette liste.

Articles parus ou à paraître

- [A1] Boubacar Maïnassara, Y. Tests portmanteau multivariés d'adéquation de modèles VARMA faibles. *C. R. Acad. Sci. Paris*, 348, 927–929, 2010. [BM10]
- [A2] Boubacar Maïnassara, Y. and Francq, C. Estimating structural VARMA models with uncorrelated but non-independent error terms. *Journal of Multivariate Analysis*, 102, 496–505, 2011. [BMF11]
- [A3] Boubacar Maïnassara, Y. Multivariate portmanteau test for structural VARMA models with uncorrelated but non-independent error terms. *Journal of Statistical Planning and Inference*, 141, 2961–2975, 2011. [BM11a]
- [A4] Boubacar Maïnassara, Y. Estimation des ordres de modèles ARMA faibles multivariés. *C. R. Acad. Sci. Paris*, 349, 695–698, 2011. [BM11c]
- [A5] Boubacar Maïnassara, Y. Estimation de la matrice de variance asymptotique des estimateurs du QMV de modèles ARMA faibles multivariés. *C. R. Acad. Sci. Paris*, 349, 817–820, 2011. [BM11b]
- [A6] Boubacar Maïnassara, Y. Selection of weak VARMA models by modified Akaike's information criteria. *Journal of Time Series Analysis*, 33, 121–130, 2012. [BM12]
- [A7] Boubacar Maïnassara, Y., Carbon, M. and Francq, C. Computing and estimating information matrices of weak ARMA models. *Computational Statistics and Data Analysis*, 56, 345–361, 2012. [BMCF12]
- [A8] Boubacar Maïnassara, Y. and Kokonendji, C. C. On normal stable Tweedie models and power-generalized variance functions of only one component. *TEST*, 23, 585–606, 2014. [BMK14]
- [A9] Boubacar Maïnassara, Y. Estimation of the variance of the quasi-maximum likelihood estimator of weak VARMA models. *Electronic Journal of Statistics*, 8, 2701–2740, 2014. [BM14]
- [A10] Boubacar Maïnassara, Y. and Raïssi, H. Semi-strong linearity testing in linear models with dependent but uncorrelated errors. *Statistics and Probability Letters*, 103, 110–115, 2015. [BMR15]
- [A11] Boubacar Maïnassara, Y. and Kokonendji, C. C. Modified Schwarz and Hannan-Quinn information criteria for weak VARMA models. *Statistical Inference for Stochastic Processes*, 19 (2), 199–217, 2016. [BMK16]

- [A12] Boubacar Maïnassara, Y. and Saussereau, B. (2016). Diagnostic checking in multivariate ARMA models with dependent errors using normalized residual autocorrelations. À paraître dans *Journal of the American Statistical Association*. [BMS17]

Articles soumis

- [A13] Boubacar Maïnassara, Y., Kadmiri, O. and Saussereau, B. (2017). Estimating Multivariate Asymmetric Power GARCH Models. *À soumettre*.
- [A14] Boubacar Maïnassara, Y., Kadmiri, O. and Saussereau, B. (2017). Portmanteau test for asymmetric power GARCH model when the power is unknown. *Soumis*.
- [A15] Boubacar Maïnassara, Y. and Rabehasaina, L. (2017). Estimating RCARMA models with uncorrelated but non-independent error terms. *Soumis*.
- [A16] Boubacar Maïnassara, Y. and Ilmi Amir, A. (2017). Goodness-of-fit tests for multiplicative seasonal ARMA models with uncorrelated but non-independent error terms. *Soumis*.

Articles en cours de rédaction

- [CR1] Boubacar Maïnassara, Y. and Saussereau, B. (2016). Estimating trend-stationary structural VARMA models with uncorrelated but non-independent error terms. *Preprint*.
- [CR2] Boubacar Maïnassara, Y., Estafa, Y. and Saussereau, B. (2016). Estimating FARIMA models with uncorrelated but non-independent error terms. *Preprint*.
- [CR3] Boubacar Maïnassara, Y., Kadmiri, O. and Saussereau, B. (2017). Portmanteau test for multivariate asymmetric power GARCH models. *Preprint*.
- [CR4] Boubacar Maïnassara, Y. and Thiam, B. (2017). Corner method for identifying weak ARMA models.
- [CR5] Boubacar Maïnassara, Y. and Ilmi Amir, A. (2017). Portmanteau tests for weak PARMA models.

Thèse de doctorat

- [T] Boubacar Maïnassara, Y. Estimation, validation et identification des modèles ARMA faibles multivariés. *Thèse de doctorat de l'Université Lille 3 (2009)*.

Les articles [A1, A2, A3, A4, A5, A6] sont issus de la thèse.

Encadrement de thèses

- 1) **Othman KADMIRI** [co-encadrement à 50% avec Sausseureau Bruno] (débuté en octobre 2014)
Thèse en cours mais la soutenance aura lieu avant juin 2018
Université : Franche-Comté
Sujet : Modèles GARCH multivariés
Bourse de l'Université de Franche-Comté.
- 2) **Abdoukarim ILMI AMIR** [encadrement à 100%] (débuté en octobre 2015)
Thèse en cours
Université : Franche-Comté
Sujet : Modèles ARMA faibles saisonniers et/ou périodiques
Bourse du gouvernement djiboutien
Cette thèse est à mi-temps, en effet : l'étudiant passe 3 à 4 mois maximum chaque année au Laboratoire de Mathématiques de Besançon pour travailler sur la thèse, et une fois rentré, il se consacre à sa charge d'enseignement à l'université de Djibouti. .
- 3) **Youssef ESSTAFI** [Co-encadrement à 50% avec Sausseureau Bruno] (débuté en octobre 2016)
Thèse en cours
Université : Franche-Comté
Sujet : Modèles FARIMA avec des erreurs dépendantes
Bourse ministérielle.

Introduction générale

Ce mémoire présente une synthèse de mes activités de recherche et d'encadrement depuis mon doctorat en 2009. Mes activités de recherche s'articulent principalement sur l'analyse statistique des séries temporelles et secondairement sur une famille, particulière, de modèles exponentiels appelée Normal Stable-Tweedie (NST).

L'accent a été mis de manière volontaire sur la thématique des séries temporelles qui donne son nom au mémoire. De plus, les thèmes actuellement étudiés dans le cadre des trois thèses que j'encadre et co-encadre sont sur cette thématique. Il me semble donc naturel dans le cadre d'une habilitation à diriger des recherches, de valoriser un travail d'encadrement doctoral.

On observe une suite x_1, \dots, x_n de vecteurs de \mathbb{R}^d indexés par le temps. Cette série chronologique (temporelle ou chronique) est dite univariée quand $d = 1$, et multivariée quand $d > 1$. Cette série est issue d'un phénomène aléatoire. On la considère comme une réalisation d'un processus stochastique $X = (X_t)_{t \in \mathbb{Z}}$. Le schéma de modélisation est

$$(\Omega, \mathcal{A}, \mathbb{P}) \quad \omega \in \Omega \mapsto X(\omega) = (\dots, x_1, \dots, x_n, \dots).$$

On dit que $X(\omega)$ est une observation, ou encore une trajectoire, du processus X . Lorsqu'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 multivarié (vectoriel). Je présente un bref état de l'art et quelques motivations sur cette thématique.

Contexte bibliographique et motivations

Les modèles AutoRégressifs (AR) ont été introduits par Yule (voir [Yul26]) en 1926 et complétés, en 1937, par Slutsky dans [Slu37] en introduisant les modèles moyennes mobiles (MA, Moving Average en anglais). Leur extension ARMA (AutoRegressive Moving Average, en anglais) est due à Wold (1938) dans ses travaux de thèse (voir [Wol54]). Cette classe de modèles ARMA se révèle très importante pour la prévision de processus stationnaires au second ordre. Ceux-ci découlent naturellement d'un résultat fondamental dû à Wold (1938) qui affirme que tout processus stationnaire au second ordre et purement non déterministe admet une représentation moyenne mobile infinie (voir [BD91], pour le cas univarié). Les arguments utilisés s'étendent directement au cas multivarié (voir [Rei97]). 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 co-intégration (fondée sur les travaux du prix Nobel d'économie Granger) atteste de l'importance de cette problématique.

Pour l'analyse des séries temporelles multivariées, la classe des modèles ARMA vectoriels (VARMA pour Vector ARMA) (voir [Rei97, Lüt05]) est utilisée 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 modélisation 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 [BD91]). Cette extension pose néanmoins des problèmes ardu, comme par exemple, l'identification et l'estimation des paramètres du modèle et suscite des axes de recherche spécifiques, comme la co-intégration (voir [Lüt05]). 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 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. Dans un modèles VARMA faible, on suppose simplement que le terme d'erreur est non corrélé. 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 [Eng82] et leur extension GARCH (ARCH généralisés) due à [Bol86] pour le cas univarié (voir [BLR06], pour une review sur les modèles GARCH multivariés), on peut également citer les modèles all-pass introduits par [BDT01, ADB06]. Ainsi la distinction entre modèle VARMA fort et faible n'est donc qu'une question d'hypothèse sur le bruit. Vu que la contrainte sur le terme d'erreur d'un modèle VARMA faible est moins forte que pour un modèle VARMA fort, il est clair que la classe des modèles VARMA faibles est la plus large. Nous en déduisons la relation suivante

$$\{\text{modèles VARMA forts}\} \subset \{\text{modèles VARMA faibles}\}.$$

De nombreux outils statistiques ont été développés pour l'estimation, l'analyse statistique et l'interprétation des modèles VARMA forts (voir par exemple [Chi74, Duf04, DJ14, DH76, Han76, HD88, HDD80, Hos80, Hos89, Hos81a, Hos81b, Kas12, LM81, RBY92]).

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 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. Dans le cadre univarié, [FZ98, FRZ05] 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. Certaines transformations de processus VARMA forts, comme par exemple des agrégations temporelles (voir [DN93]), des échantillonnages ou encore des transformations linéaires plus générales, admettent des représentations VARMA faibles. Par conséquent, relâcher cette hypothèse permettra aux modèles VARMA faibles de couvrir une large classe de processus non linéaires.

Contrairement aux modèles VARMA forts, les travaux consacrés à l'analyse statistique des modèles VARMA faibles sont globalement 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 [FZ98]. 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 (voir [FZ00]). Parmi les autres résultats statistiques disponibles, signalons des travaux importants sur le comportement asymptotique des autocorrélations empiriques établis par [LNS01, LNS02] et [RT96]. Ces der-

niers auteurs ont également introduit des processus non corrélés mais dépendants et qui peuvent être étendus au cas multivarié. Notons aussi les travaux importants sur les tests portmanteau par [FRZ05, Sha11, ZL15], sur les tests de linéarité et l'estimation HAC de modèles ARMA faibles par [FZ07], sur les processus dual et inverse de représentations ARMA faibles [EG10, EG15], sur l'identification des modèles ARMA faibles basée sur l'analyse des autocorrélations et autocorrélations inverses [FZ05b, EGF06]. Signalons aussi les travaux sur l'inférence statistique de processus dépendants par une approche d'auto-normalisation (voir [Sha0a, Sha0b, Sha12, Sha15]).

En analyse multivariée, des progrès importants ont été réalisés par [DR08] qui ont proposé une généralisation de la méthode d'estimation basée sur la régression introduite par [HR83] 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, qui est aussi une extension de celui introduit par [HR83]. Katayama a proposé une nouvelle statistique de test portmanteau dont la distribution asymptotique est celle d'un chi-deux (voir [Kat12]). On pourrait également citer les travaux existants sur les modèles VAR faibles, mais comme c'est une sous classe des modèles VARMA faibles, j'ai fait le choix (volontaire) de ne citer que les travaux basés sur la classe des VARMA faibles.

Nous parlerons de nos travaux réalisés sur les VARMA faibles sous forme de contribution dans la section suivante.

Contribution apportée

L'objectif principal de notre recherche est d'étudier dans quelle mesure les travaux cités précédemment sur l'analyse statistique des modèles ARMA forts peuvent s'étendre au cas multivarié faible.

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 fortement des propriétés du bruit blanc. Nous avons étudié toutes ces étapes dans le cadre de modèles VARMA faibles.

Le chapitre 1 de ce mémoire traite des propriétés asymptotiques des estimateurs du quasi-maximum de vraisemblance des paramètres de modèles VARMA faibles. Une attention particulière est ensuite accordée à l'estimation de la matrice de variance asymptotique de ces estimateurs qui est d'une forme "sandwich", et peut être très différente de la variance asymptotique standard. Nous prouvons la différence de précision qui peut exister entre les deux formes de manière théorique et à travers quelques illustrations numériques.

Le chapitre 2 est consacré au problème très important de sélection des ordres de modèles VARMA faibles. 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 des critères d'information modifiés en prouvant que les fondements théoriques des critères d'information standards ne sont plus établis quand l'hypothèse de bruit blanc fort est relâchée. Nous illustrons nos résultats sur des études empiriques basées sur des expériences de Monte Carlo.

Dans le chapitre 3, nous traitons le problème de la validation de modèles VARMA faibles. Dans un premier temps, nous proposons des versions modifiées des tests de Wald, du multiplicateur de Lagrange et du rapport de vraisemblance afin de pouvoir tester des restrictions linéaires sur les paramètres libres du modèle. Ensuite, nous traitons les tests fondés sur l'analyse des résidus de modèles VARMA faibles, plus particulièrement nous nous sommes intéressés au cas des tests portmanteau. Nous avons proposé de nouvelles statistiques permettant de tester l'adéquation de modèles VARMA faibles. Les résultats ont été illustrés par des expériences de Monte Carlo.

Dans le chapitre 4, nous traitons le problème de l'estimation et de la validation de modèles GARCH multivariés. C'est un travail développé dans le cadre du co-encadrement de la thèse de Othman Kadmiri commencée en octobre 2014.

Le chapitre 5 traite de l'estimation de modèles ARMA faibles à coefficients aléatoires (RCARMA, pour random coefficient ARMA). Je décris brièvement les premiers résultats obtenus sur la stationnarité et les propriétés asymptotiques des estimateurs de moindres carrés des paramètres de modèles RCARMA faibles.

Dans le chapitre 6, je décris brièvement les travaux indépendants de ma thématique principale décrite dans les chapitres 1 à 5. Nous introduisons la famille des lois Normal Stables-Tweedie (NST) k -dimensionnelles, $k > 1$. Ces NST sont constituées d'une marge unidimensionnelle stable-Tweedie positive et des $k - 1$ gaussiennes indépendantes de même variance égale à la valeur de la stable-Tweedie.

Le chapitre 7 détaille les trois co-encadrements et des perspectives pour des développements futurs.

Chapitre 1

Contribution à l'estimation des modèles ARMA faibles multivariés

Ce chapitre porte sur nos travaux liés à l'estimation des modèles VARMA faibles ainsi que sur l'étude des propriétés asymptotiques des estimateurs du quasi-maximum de vraisemblance (EQME). Nous présentons également nos travaux consacrés à l'estimation de la matrice de variance asymptotique de ces estimateurs, afin de les comparer à celle du cas standard. Les publications liées à ce chapitre sont : **[A2, A5, A7, A9]**

Pour l'estimation des paramètres du modèle ARMA ou VARMA, on utilise couramment la méthode du quasi-maximum de vraisemblance (QMV, pour la suite) qui est la méthode du maximum de vraisemblance gaussien lorsque l'hypothèse de bruit blanc gaussien est relâchée.

À la suite des travaux de [\[Wal64\]](#), l'estimation des modèles ARMA forts a donné lieu à de nombreux travaux. Par exemple : [\[Han73\]](#) a montré la convergence forte de divers estimateurs (essentiellement sous une hypothèse d'ergodicité sur le processus observé) et la normalité asymptotique sous une hypothèse de différence de martingale sur le bruit, [\[DH76\]](#) ont étendu ces travaux aux cas multivariés (voir aussi [\[HD88\]](#)), [\[RC79\]](#) ont établi une preuve de la convergence forte de l'estimateur du QMV pour les modèles VARMA forts. Les comportements asymptotiques des estimateurs du QMV de modèles VARMA forts ont été largement étudiés dans la littérature (voir [\[Lüt05\]](#)). Cependant ces résultats ne s'appliquent pas aux représentations ARMA ou VARMA faibles de processus non linéaires. En effet, les auteurs qui viennent d'être cités supposent toujours que le processus des innovations linéaires est une différence de martingales (au moins en ce qui concerne la normalité asymptotique). Dans le cas de modèles ARMA faibles univariés, la consistance et la normalité asymptotique de l'estimateur des moindres carrés ordinaires ont été montrés dans [\[FZ98\]](#) sous certaines hypothèses d'ergodicité et de mélange. Mais 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 co-intégration (fondée sur les travaux du prix Nobel d'économie Granger) atteste de l'importance de cette problématique. Dans **[A2]**, nous avons étudié les propriétés asymptotiques du QMV des paramètres d'un modèle VARMA faible sous des hypothèses essentielles d'ergodicité et de mélange en prouvant la convergence forte et la normalité asymptotique de l'estimateur du QMV. 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 [\[FZ00\]](#). C'est ce que nous avons fait dans **[A5, A9]** et pour un processus ARMA faible univarié, nous avons proposé dans **[A7]** une généralisation de la méthode

introduite dans [A9].

1.1 Propriétés asymptotiques des estimateurs du QMV

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 $X_t = (X_{1t}, \dots, X_{dt})'$ est un processus vectoriel stationnaire au second ordre de dimension d , d étant le nombre de séries considérées, à valeurs réelles. Les paramètres A_{0i} , $i \in \{0, \dots, p_0\}$ et B_{0j} , $j \in \{0, \dots, q_0\}$ sont des matrices $d \times d$, et p_0 et q_0 sont des entiers appelés ordres. Sans perte de généralité, nous supposons que les matrices A_{0p_0} et B_{0q_0} sont non nulles. Le terme d'erreur $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{dt})'$ est un bruit blanc vectoriel, c'est-à-dire une suite de variables aléatoires centrées ($\mathbb{E}(\epsilon_t) = 0$), non corrélées ($\text{Cov}(\epsilon_t, \epsilon_{t-h}) = 0$, $\forall h \neq 0$), avec une matrice de covariance non singulière $\Sigma_0 := \mathbb{E}(\epsilon_t \epsilon_t')$. Ces modèles VARMA sont généralement utilisés avec des hypothèses fortes sur le bruit qui en limitent la généralité. Ainsi, nous appelons VARMA forts les modèles standards dans lesquels le terme d'erreur est supposé être une suite indépendante et identiquement distribuée (iid), et nous parlons de modèles VARMA faibles quand les hypothèses sur le bruit sont moins restrictives. Dans ces modèles VARMA faibles, les termes d'erreur sont non corrélés mais peuvent contenir des dépendances non linéaires. Ceci permet de traiter des processus qui peuvent avoir des dynamiques non linéaires très générales. Relâcher cette hypothèse d'indépendance standard permet aux modèles VARMA faibles de couvrir une large classe de processus non linéaires.

Les paramètres sont les coefficients des matrices carrées d'ordre d suivantes : A_i , $i \in \{0, \dots, p_0\}$, B_j , $j \in \{0, \dots, q_0\}$ et Σ . Le vecteur des vraies valeurs des paramètres auquel on s'intéresse dans l'estimation est noté ϑ_0 . Notons par $\det(\cdot)$ ou $|\cdot|$ le déterminant d'une matrice carrée. Pour l'étude des séries chronologiques, la notion de stationnarité faible (au second ordre) conduit à imposer des contraintes sur les racines des polynômes $\det A_\vartheta(z)$ et $\det B_\vartheta(z)$ où $A_\vartheta(z) = A_0 - \sum_{i=1}^{p_0} A_i z^i$ et $B_\vartheta(z) = B_0 - \sum_{j=1}^{q_0} B_j z^j$. Supposons que

$$(H1) \det A_\vartheta(z) \det B_\vartheta(z) = 0 \Rightarrow |z| > 1.$$

L'intérêt principal de cette modélisation et des représentations obtenues en inversant successivement les polynômes $A_\vartheta(z)$ et $B_\vartheta(z)$ est de permettre de calculer les prévisions linéaires optimales du processus de manière beaucoup plus simple que sous la seule hypothèse de stationnarité au second ordre.

Bien que la modélisation linéaire standard des séries temporelles s'étende naturellement dans le cadre des modèles VARMA, cette extension pose néanmoins des problèmes ardu, comme par exemple l'identifiabilité. Ainsi nous allons imposer des conditions 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.

Dans le cadre du modèle VARMA (4.13), l'hypothèse (H1) qui assure que les polynômes $\det A_{\vartheta_0}(z)$ et $\det B_{\vartheta_0}(z)$ n'ont pas de racine commune dans le disque unité, ne suffit pas à garantir l'identifiabilité du paramètre, c'est-à-dire à garantir que seule la valeur $\vartheta = \vartheta_0$ satisfasse $A_{\vartheta}(L)X_t = B_{\vartheta}(L)\epsilon_t$ où L est l'opérateur retard (i.e. $X_{t-1} = LX_t$, $\epsilon_{t-k} = L^k\epsilon_t$ pour tout $t, k \in \mathbb{Z}$). On suppose donc que

(H2) : pour tout $\vartheta \in \Theta$ tel que $\vartheta \neq \vartheta_0$, les fonctions de transfert vérifient soit :

$$A_0^{-1}B_0B_{\vartheta}^{-1}(z)A_{\vartheta}(z) \neq A_{00}^{-1}B_{00}B_{\vartheta_0}^{-1}(z)A_{\vartheta_0}(z) \text{ pour un } z \in \mathbb{C}$$

ou soit

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

Divers types de conditions peuvent être imposés pour assurer l'identifiabilité (voir [Rei97] p. 37-40, ou [Han76], [HD88] section 2.7). Pour avoir l'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 [Han76, Lüt05, Rei97].

1.1.1 Définition de l'estimateur du QMV

La représentation (4.13) sera dite réduite lorsqu'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. Donc pour des raisons pratiques, nous écrivons le modèle (4.13) sous sa forme réduite

$$X_t - \sum_{i=1}^{p_0} A_{00}^{-1}A_{0i}X_{t-i} = e_t - \sum_{j=1}^{q_0} A_{00}^{-1}B_{0j}B_{00}^{-1}A_{00}e_{t-j}, \text{ où } e_t = A_{00}^{-1}B_{00}\epsilon_t. \quad (1.2)$$

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

$$\tilde{e}_t(\vartheta) = X_t - \sum_{i=1}^{p_0} A_0^{-1}A_iX_{t-i} + \sum_{j=1}^{q_0} A_0^{-1}B_jB_0^{-1}A_0\tilde{e}_{t-j}(\vartheta), \quad (1.3)$$

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 avons montré que, comme dans le cas univarié (voir [FZ98]), 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

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

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

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

Les résultats présentés ici sont dans [A2].

1.1.2 Convergence presque sûr de l'estimateur du QMV

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

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

Nous pouvons maintenant énoncer le théorème de convergence forte suivant.

Théorème 1. (Convergence forte) *Soit $(\hat{\vartheta}_n)$ une suite d'estimateurs du QMV de l'équation (4.13). Sous les hypothèses (H1), (H2) et (H3), 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. La démonstration de ce résultat est la même que celle établie pour le cas des modèles VARMA standard (voir [HD88] dans le cas VARMAX sous des hypothèses d'homoscédasticité sur le bruit), qui est basée sur celle proposée par [Wal49, Pfa69] pour prouver la consistance de l'estimateur du maximum de vraisemblance. Nous fournissons un schéma de cette preuve.

Démonstration. L'hypothèse (H1) sur les racines des polynômes VAR et MA entraîne que

$$e_t(\vartheta) = A_0^{-1} B_0 B_\vartheta^{-1}(L) A_\vartheta(L) X_t = X_t + \sum_{i=1}^{\infty} C_i(\vartheta) X_{t-i}, \quad \sup_{\vartheta \in \Theta} \|C_i(\vartheta)\| = O(\rho^i), \quad (1.4)$$

pour $\rho \in [0, 1[$. On montre d'abord que l'hypothèse d'identifiabilité (H2) entraîne pour tout $\vartheta \in \Theta$ et tout $t \in \mathbb{Z}$,

$$\epsilon_t(\vartheta) = \epsilon_t(\vartheta_0) \quad p.s. \Rightarrow \vartheta = \vartheta_0. \quad (1.5)$$

Ensuite, sous l'hypothèse (H1), on montre que les valeurs initiales sont asymptotiquement négligeables en montrant successivement que

$$\sup_{\vartheta \in \Theta} \|\tilde{\epsilon}_t(\vartheta) - e_t(\vartheta)\| = O(\rho^t), \quad (1.6)$$

$$\sup_{\vartheta \in \Theta} |\tilde{\ell}_n(\vartheta) - \ell_n(\vartheta)| = o(1), \quad p.s. \quad (1.7)$$

Au vu de (1.7), on obtient

$$\tilde{\ell}_n(\vartheta) = \ell_n(\vartheta) + o_{\mathbb{P}}(1) = \frac{1}{n} \sum_{t=1}^n l_t(\vartheta) \quad p.s. \quad n \rightarrow \infty,$$

où

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

En appliquant le théorème ergodique, on a presque sûrement

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

où $Q(\vartheta) = \log \det \Sigma_e + \mathbb{E} e_1'(\vartheta) \Sigma_e^{-1} e_1(\vartheta)$. Notons \mathcal{H}_{t-1} l'espace de Hilbert engendré par les variables aléatoires X_u pour $u < t$. Maintenant, il faut montrer que ce critère limite est minimisé en ϑ_0 :

$$Q(\vartheta) \geq Q(\vartheta_0), \quad \forall \vartheta \neq \vartheta_0 \in \Theta. \quad (1.8)$$

En effet, on a

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

Puisque $e_t = e_t(\vartheta_0)$ est orthogonale à \mathcal{H}_{t-1} et au vu de l'équation (1.4) : $\{e_t(\vartheta) - e_t(\vartheta_0)\} = \sum_{i=1}^{\infty} \{C_i(\vartheta) - C_i(\vartheta_0)\} X_{t-i}$ appartient à \mathcal{H}_{t-1} , on a $\mathbb{E} \{e_1(\vartheta) - e_1(\vartheta_0)\}' \Sigma_e^{-1} e_1(\vartheta_0) = 0$. En remarquant que

$$\begin{aligned} Q(\vartheta_0) &= \log \det \Sigma_{e0} + \mathbb{E} e_1'(\vartheta_0) \Sigma_{e0}^{-1} e_1(\vartheta_0) \\ &= \log \det \Sigma_{e0} + \text{Tr} (\Sigma_{e0}^{-1} \mathbb{E} e_1(\vartheta_0) e_1'(\vartheta_0)) = \log \det \Sigma_{e0} + d, \end{aligned}$$

on a alors

$$Q(\vartheta) - Q(\vartheta_0) \geq \text{Tr} \Sigma_e^{-1} \Sigma_{e0} - \log \det \Sigma_e^{-1} \Sigma_{e0} - d \geq 0 \quad (1.9)$$

en utilisant l'inégalité suivante valable pour toutes les matrices semi-définies positives d'ordre $d \times d$:

$$\text{Tr}(A^{-1}B) - \log \det(A^{-1}B) \geq \text{Tr}(A^{-1}A) - \log \det(A^{-1}A) = d.$$

Ainsi, au moins une des deux inégalités dans (1.9) est stricte, sauf si $e_1(\vartheta) = e_1(\vartheta_0)$ avec probabilité 1 et $\Sigma_e = \Sigma_{e0}$, ce qui est donc équivalent à $\vartheta = \vartheta_0$ d'après (1.5). Cela ne suffit pas pour conclure car, a priori, on ne sait pas si

$$\arg \min_{\vartheta} \lim_{n \rightarrow \infty} \tilde{\ell}_n(\vartheta) \text{ et } \lim_{n \rightarrow \infty} \arg \min_{\vartheta} \tilde{\ell}_n(\vartheta)$$

coïncident.

Pour tout $\vartheta \in \Theta$ et tout entier $k > 0$, on pose $V_k(\vartheta)$ la sphère de centre ϑ et de rayon $1/k$. En utilisant (1.7), on a

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

La dernière égalité est obtenue par application du théorème ergodique au processus stationnaire $\{\inf_{\vartheta^* \in V_k(\vartheta) \cap \Theta} \ell_t(\vartheta^*)\}_t$. Quand k tend vers $+\infty$, le théorème de Beppo-Levi montre que $\mathbb{E} \inf_{\vartheta^* \in V_k(\vartheta) \cap \Theta} \ell_1(\vartheta^*)$ tend vers $\mathbb{E} \ell_1(\vartheta)$. D'après (1.9), on obtient

$$\liminf_{n \rightarrow \infty} \inf_{\vartheta^* \in V(\vartheta)} \tilde{\ell}_n(\vartheta^*) > \mathbb{E} \ell_1(\vartheta), \quad p.s. \quad (1.10)$$

Remarquons également que (1.7), (1.8) et le théorème ergodique entraînent que

$$\begin{aligned} \limsup_{n \rightarrow \infty} \inf_{\vartheta^* \in V(\vartheta) \cap \Theta} \tilde{\ell}_n(\vartheta^*) &\leq \limsup_{n \rightarrow \infty} \inf_{\vartheta^* \in V(\vartheta) \cap \Theta} \ell_n(\vartheta^*) + \limsup_{n \rightarrow \infty} \sup_{\vartheta \in \Theta} |\ell_n(\vartheta) - \tilde{\ell}_n(\vartheta)| \\ &\leq \limsup_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n \ell_t(\vartheta_0) \\ &= \mathbb{E} \ell_1(\vartheta_0). \end{aligned}$$

De plus, pour tout voisinage $V(\vartheta_0)$ de ϑ_0 on a

$$\limsup_{n \rightarrow \infty} \inf_{\vartheta^* \in V(\vartheta_0)} \tilde{\ell}_n(\vartheta^*) \leq \mathbb{E} \ell_1(\vartheta_0), \quad p.s. \quad (1.11)$$

On conclut en utilisant les arguments standard de compacité (voir [Wal49]). Plus précisément, le compact Θ est recouvert par un voisinage $V(\vartheta_0)$ de ϑ_0 et la réunion d'un nombre fini de voisinages $V(\vartheta_1), \dots, V(\vartheta_k)$ satisfaisants (1.10) (par remplacement de ϑ par ϑ_i , $i = 1, \dots, k$ dans (1.10)). Il existe donc un sous-recouvrement de Θ par un nombre fini de ces ouverts tels que

$$\bigcup_{i=0}^k V(\vartheta_i) \supseteq \Theta.$$

D'après (1.10) et (1.11), on obtient presque sûrement

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

pour n assez grand. Par conséquent, presque sûrement, $\hat{\vartheta}_n$ est dans $V(\vartheta_0)$ pour n assez grand. Puisque le voisinage $V(\vartheta_0)$ peut être choisi arbitrairement petit, on obtient le résultat. \square

1.1.3 Normalité asymptotique de l'estimateur du QMV

L'ergodicité ne suffit pas pour obtenir un théorème central limite (noté TCL dans la suite). Pour la normalité asymptotique du QMV, nous avons donc besoin d'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

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

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 temporelle du processus (ϵ_t) .

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

Remarque 1.1. Notons que d'après l'hypothèse (H1) et l'équation (4.13), 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 (voir [Dav94]). 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 $\mathbb{E}\|\epsilon_t\|^{4+2\nu} < \infty$ 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 (H1) sur les racines des polynômes, il est clair que les hypothèses de moment $\mathbb{E}\|\epsilon_t\|^{4+2\nu} < \infty$ et $\mathbb{E}\|X_t\|^{4+2\nu} < \infty$ sont équivalentes. Par contre, les hypothèses de mélange ne sont pas équivalentes.

Nous définissons la matrice M_{ϑ_0} des coefficients de la forme réduite du modèle (1.2) par

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

Ainsi nous avons besoin d'une hypothèse (d'identifiabilité locale) 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 .

(H6) : La matrice \dot{M}_{ϑ_0} est de rang k_0 .

Remarque 1.2. Notons que (H6) est une hypothèse d'identifiabilité qui complète (H2), mais qu'aucune n'implique l'autre. L'hypothèse (H6) est locale alors que (H2) est une hypothèse globale pour tout $\vartheta \in \Theta$.

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

Théorème 2. (Normalité asymptotique) Sous les hypothèses du Théorème 1, (H4), (H5) et (H6), nous avons

$$\sqrt{n} \left(\hat{\vartheta}_n - \vartheta_0 \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Omega := J^{-1} I J^{-1}), \quad (1.12)$$

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

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

La démonstration de ce théorème repose sur le TCL pour processus α -mélangeant [Her84], le théorème 1 et l'inégalité de covariance de [Dav68] pour les processus mélangeants. Nous donnons un schéma de la preuve de ce théorème fondamental pour suite.

Démonstration. En utilisant des formules de dérivation élémentaire sur les traces (voir [Lüt05]), pour tout $i \in \{1, \dots, k_0\}$, on a

$$\frac{\partial \ell_t(\vartheta)}{\partial \vartheta_i} = \text{Tr} \left\{ \Sigma_e^{-1} \frac{\partial \Sigma_e}{\partial \vartheta_i} - \Sigma_e^{-1} e_t(\vartheta) e_t'(\vartheta) \Sigma_e^{-1} \frac{\partial \Sigma_e}{\partial \vartheta_i} \right\} + 2 \frac{\partial e_t'(\vartheta)}{\partial \vartheta_i} \Sigma_e^{-1} e_t(\vartheta). \quad (1.13)$$

L'hypothèse (H1) sur les racines des polynômes entraîne que, comme pour (1.4), les dérivées d'ordre $1, 2, \dots$ de $e_t(\vartheta)$ s'expriment comme des combinaisons linéaires de X_t, X_{t-1}, \dots dont les

coefficients tendent vers 0 à vitesse exponentielle, uniformément en $\vartheta \in \Theta$. Plus explicitement, pour $k = 1, \dots, k_0$, on a

$$\frac{\partial e_t(\vartheta)}{\partial \vartheta_k} = \sum_{i=1}^{\infty} C_i^{(k)}(\vartheta) X_{t-i}, \quad \frac{\partial \tilde{e}_t(\vartheta)}{\partial \vartheta_k} = \sum_{i=1}^{t-1} C_i^{(k)}(\vartheta) X_{t-i}, \quad (1.14)$$

$$C_i^{(k)}(\vartheta) = \frac{\partial C_i(\vartheta)}{\partial \vartheta_k}, \quad \sup_{\vartheta \in \Theta} \|C_i^{(k)}(\vartheta)\| = O(\rho^i). \quad (1.15)$$

Similairement à l'équation (1.6), on obtient alors

$$\sup_{\vartheta \in \Theta} \left\| \frac{\partial \tilde{e}_t(\vartheta)}{\partial \vartheta'} - \frac{\partial e_t(\vartheta)}{\partial \vartheta'} \right\| = O(\rho^t). \quad (1.16)$$

En utilisant (1.13), on obtient

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

où

$$a_1 = \frac{2}{\sqrt{n}} \sum_{t=1}^n \left\{ \left(\frac{\partial e_t'(\vartheta)}{\partial \vartheta_k} - \frac{\partial \tilde{e}_t'(\vartheta)}{\partial \vartheta_k} \right) \Sigma_e^{-1} e_t(\vartheta) + \frac{\partial \tilde{e}_t'(\vartheta)}{\partial \vartheta_k} \Sigma_e^{-1} (e_t(\vartheta) - \tilde{e}_t(\vartheta)) \right\}$$

$$a_2 = \text{Tr} \left(\Sigma_e^{-1} [\{\tilde{e}_t(\vartheta) - e_t(\vartheta)\} \tilde{e}_t'(\vartheta) + e_t(\vartheta) \{\tilde{e}_t(\vartheta) - e_t(\vartheta)\}'] \Sigma_e^{-1} \frac{\partial \Sigma_e}{\partial \vartheta_k} \right).$$

D'après (1.6) et (1.16), on a

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

D'après (1.4) et (1.14)-(1.15), on obtient

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

Par l'inégalité de Markov et pour tout $\varepsilon > 0$, on a

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

et donc

$$\sqrt{n} \sup_{\vartheta \in \Theta} \left\| \frac{\partial \tilde{\ell}_n(\vartheta)}{\partial \vartheta} - \frac{\partial \ell_n(\vartheta)}{\partial \vartheta} \right\| = o_{\mathbb{P}}(1). \quad (1.17)$$

Notons que les mêmes égalités sont aussi vérifiées pour les dérivées secondes de $\tilde{\ell}_n(\vartheta)$. Au vu du Théorème 1 et l'hypothèse (H4), on a presque sûrement $\hat{\vartheta}_n \rightarrow \vartheta_0 \in \overset{\circ}{\Theta}$. Ainsi $\partial \tilde{\ell}_n(\hat{\vartheta}_n)/\partial \vartheta = 0$ pour

n suffisamment grand. Un développement de Taylor autour de ϑ_0 donne

$$0 = \sqrt{n} \frac{\partial \ell_n(\vartheta_0)}{\partial \vartheta} + \frac{\partial^2 \ell_n(\vartheta_0)}{\partial \vartheta \partial \vartheta'} \sqrt{n} (\hat{\vartheta}_n - \vartheta_0) + o_{\mathbb{P}}(1). \quad (1.18)$$

L'hypothèse $(H6)$ permet de prouver que la matrice

$$J = \mathbb{E} \frac{\partial^2 \ell_n(\vartheta_0)}{\partial \vartheta \partial \vartheta'} \text{ est inversible.} \quad (1.19)$$

D'après (1.4), on a

$$\frac{\partial e_t(\vartheta_0)}{\partial \vartheta_i} = \sum_{\ell=1}^{\infty} d'_\ell e_{t-\ell}, \quad (1.20)$$

où la suite des matrices $d_\ell = d_\ell(i)$ est telle que $\|d_\ell\| \rightarrow 0$ à vitesse exponentielle quand $\ell \rightarrow \infty$. Au vu de (1.13), pour tout $m \geq 0$ et $i = 1, \dots, k_0$, on pose

$$\frac{\partial l_t(\vartheta_0)}{\partial \vartheta_i} = Y_{t,m,i} + Z_{t,m,i}$$

où

$$Y_{t,m,i} = \text{Tr} \left[\Sigma_{e_0}^{-1} \{ I_d - e_t e_t' \Sigma_{e_0}^{-1} \} \frac{\partial \Sigma_e(\vartheta_0)}{\partial \vartheta_i} \right] + 2 \sum_{\ell=1}^m e_{t-\ell}' d'_\ell \Sigma_{e_0}^{-1} e_t$$

$$Z_{t,m,i} = 2 \sum_{\ell=m+1}^{\infty} e_{t-\ell}' d'_\ell \Sigma_{e_0}^{-1} e_t.$$

Notons $Y_{t,m} = (Y_{t,m,1}, \dots, Y_{t,m,k_0})'$ et $Z_{t,m} = (Z_{t,m,1}, \dots, Z_{t,m,k_0})'$ des vecteurs de \mathbb{R}^{k_0} et remarquons que les processus $(Y_{t,m})_t$ et $(Z_{t,m})_t$ sont stationnaires et centrés. De plus, sous $(H5)$ et pour m fixé, le processus $Y = (Y_{t,m})_t$ est mélangeant (voir Théorème 14.1, page 210, dans [Dav94]) et satisfait $\alpha_Y(h) \leq \alpha_\varepsilon(\max\{0, h - m\})$. En appliquant le TCL pour des processus mélangeants de [Her84], on obtient

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

Comme dans [FZ98] (voir Lemme 3), on montre que la matrice $I = \lim_{m \rightarrow \infty} I_m$ (l'existence de I est due à une conséquence de l'hypothèse $(H5)$ et de l'inégalité de covariance de [Dav68]). Puisque $\|Z_{t,m}\|_2 \rightarrow 0$ à vitesse exponentielle quand $m \rightarrow \infty$, en utilisant les arguments donnés dans [FZ98] (voir Lemme 4), on montre que

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left\{ \left\| n^{-1/2} \sum_{t=1}^n Z_{t,m} \right\| > \varepsilon \right\} = 0 \quad (1.21)$$

pour $\varepsilon > 0$. En utilisant le résultat dans [BD91] (voir Proposition 6.3.9, pages 207-208), on déduit

que

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial l_t(\vartheta_0)}{\partial \vartheta} = \frac{1}{\sqrt{n}} \sum_{t=1}^n Y_{t,m} + \frac{1}{\sqrt{n}} \sum_{t=1}^n Z_{t,m} \xrightarrow{\mathcal{L}} \mathcal{N}(0, I). \quad (1.22)$$

La conclusion provient de (1.18), (1.19) et (1.22). \square

Remarque 1.3. (Différence de précision dans l'estimation des VARMA forts et faibles) Les logiciels développés pour estimer les modèles VARMA forts permettent aussi d'estimer de manière convergente les paramètres d'un modèle VARMA faible.

Dans le cas standard de modèles VARMA forts, i.e. quand l'hypothèse (H3) est remplacée par celle que les termes d'erreur (ϵ_t) sont iid, nous avons $I = 2J$, ainsi la matrice Ω se résume à $\Omega_s := 2J^{-1}$. Par contre dans le cas général, nous avons $I \neq 2J$ donc $\Omega_s \neq \Omega$. Comme conséquence, il est probable que les logiciels utilisés pour ajuster les modèles VARMA forts fournissent une estimation incorrecte de la matrice Ω (puisque'ils fourniront comme estimateur de Ω , un estimateur $\hat{\Omega}_s$ qui converge vers Ω_s) pour le cas de processus VARMA faibles. Donc on ne peut pas leur faire confiance pour tester, par exemple, la significativité des coefficients VARMA faibles. C'est aussi le même problème dans le cas univarié (voir [FZ07], et d'autres références).

Notons que, pour les modèles VARMA sous la forme réduite (1.2), il n'est nullement restrictif de supposer que les coefficients $A_0, \dots, A_{p_0}, B_0, \dots, B_{q_0}$ sont fonctionnellement indépendants du coefficient Σ_e . Ceci nous permet de faire l'hypothèse suivante

(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 duplication 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 (1.2), les estimateurs du QMV et du LS (least squares) coïncident.

Théorème 3. Sous les hypothèses du Théorème 2 et (H7), l'estimateur du QMV $\hat{\vartheta}_n = (\hat{\vartheta}_n^{(1)'}, \hat{\vartheta}_n^{(2)'})'$ 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

$$J = \begin{pmatrix} J_{11} & 0 \\ 0 & J_{22} \end{pmatrix}, \quad \text{avec } J_{11} = 2\mathbb{E} \left\{ \frac{\partial}{\partial \vartheta^{(1)}} e_t'(\vartheta_0^{(1)}) \right\} \Sigma_{e0}^{-1} \left\{ \frac{\partial}{\partial \vartheta^{(1)'}} e_t(\vartheta_0^{(1)}) \right\}$$

et $J_{22} = D(\Sigma_{e0}^{-1} \otimes \Sigma_{e0}^{-1})D'$.

Remarque 1.4. Notons que la matrice J a la même expression dans les cas de modèles VARMA fort et faible (voir [Lüt05] page 480, pour le cas VARMA fort), contrairement à la matrice I qui est en général plus compliquée dans le cas faible que dans le cas fort.

1.2 Estimation de la matrice de variance asymptotique

Les résultats présentés ici sont dans [A2, A5, A7, A9].

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 . Soit $\hat{e}_t = \tilde{e}_t(\hat{\theta}_n)$ les résidus du QMV 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

$$\hat{e}_t = X_t - \sum_{i=1}^{p_0} A_0^{-1}(\hat{\theta}_n) A_i(\hat{\theta}_n) \hat{X}_{t-i} + \sum_{i=1}^{q_0} A_0^{-1}(\hat{\theta}_n) B_i(\hat{\theta}_n) B_0^{-1}(\hat{\theta}_n) A_0(\hat{\theta}_n) \hat{e}_{t-i}, \quad (1.23)$$

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

1.2.1 Estimation de la matrice Ω proposée dans [A2, A5]

Soit $\hat{\Sigma}_{e0} = n^{-1} \sum_{t=1}^n \hat{e}_t \hat{e}_t'$ un estimateur de Σ_{e0} .

Estimation de la matrice J

La matrice J peut facilement être estimée par

$$\hat{J} = \begin{pmatrix} \hat{J}_{11} & 0 \\ 0 & \hat{J}_{22} \end{pmatrix}, \quad \hat{J}_{22} = D(\hat{\Sigma}_{e0}^{-1} \otimes \hat{\Sigma}_{e0}^{-1}) D' \quad \text{et}$$

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

Dans le cas standard des VARMA forts $\hat{\Omega}_s := 2\hat{J}^{-1}$ est un estimateur fortement convergent de Ω_s . Dans le cas général des VARMA faibles, cet estimateur n'est pas convergent quand $I \neq 2J$. Nous avons besoin d'un estimateur convergent de la matrice I .

Estimation de la matrice I

Rappelons que

$$I = \lim_{n \rightarrow +\infty} \text{Var} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^n \gamma_t \right) = \sum_{h=-\infty}^{+\infty} \text{cov}(\gamma_t, \gamma_{t-h}), \quad (1.24)$$

où

$$\gamma_t = \frac{\partial l_t(\vartheta_0)}{\partial \vartheta} = \frac{\partial}{\partial \vartheta} \left\{ \log \det \Sigma_e + e_t'(\vartheta^{(1)}) \Sigma_e^{-1} e_t(\vartheta^{(1)}) \right\}_{\vartheta=\vartheta_0}. \quad (1.25)$$

L'existence de la somme des covariances dans (1.24) est une conséquence de l'hypothèse (H5) et de l'inégalité de covariance de [Dav68]. Cette matrice I est plus délicate à estimer. Néanmoins, nous nous reposons sur les deux méthodes suivantes (proposées dans la littérature) pour l'estimer :

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

Définition de l'estimateur HAC de la matrice I

Dans la littérature économique, l'estimateur 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 [And91, NW87]. Soit $\hat{\gamma}_t$ le vecteur obtenu en remplaçant ϑ_0 par $\hat{\vartheta}_n$ dans γ_t . La matrice Ω est donc estimée par un estimateur "sandwich" de la forme

$$\hat{\Omega}^{\text{HAC}} = \hat{J}^{-1} \hat{\gamma}^{\text{HAC}} \hat{J}^{-1}, \quad \hat{\gamma}^{\text{HAC}} = \frac{1}{n} \sum_{t,s=1}^n \omega_{|t-s|} \hat{\gamma}_t \hat{\gamma}_s,$$

où $\omega_0, \dots, \omega_{n-1}$ est une suite de poids (ou fenêtre) (voir [And91, NW87] pour plus de détails sur le choix des poids).

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 (γ_t) . Cette approche a été étudiée par [Ber74] (voir aussi [dHL97]). Ceci revient à interpréter $(2\pi)^{-1}I$ comme étant la densité spectrale évaluée à la fréquence 0 du processus stationnaire (γ_t) (voir [BD91], p. 459). Nous nous basons sur l'expression

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

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

$$\Phi(L) \gamma_t := \gamma_t + \sum_{i=1}^{\infty} \phi_i \gamma_{t-i} = u_t,$$

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 $\hat{\gamma}_t$ sur $\hat{\gamma}_{t-1}, \dots, \hat{\gamma}_{t-r}$ et posons $\hat{\Phi}_r(z) = I_{k_0-s} + \sum_{i=1}^r \hat{\phi}_{r,i} z^i$. Soient $\hat{u}_{r,t}$ 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 [FRZ05].

Théorème 4. (Convergence faible de I) *Sous les hypothèses du Théorème 3, nous supposons que le processus (γ_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*

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

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 [FRZ05], qui repose sur une série de lemmes.

1.2.2 Estimation de la matrice Ω proposée dans [A9]

La méthode que nous proposons dans [A9] consiste à trouver des estimateurs des matrices J et I dans lesquels les estimateurs des coefficients du modèle VARMA (4.13) sont isolés de ceux

dépendant de la distribution de l'innovation. Nous considérons la matrice Σ_e comme un paramètre de nuisance. Dans ce cas, afin d'alléger nos notations, nous posons $\theta = \vartheta^{(1)}$.

La détermination des expressions explicites des matrices d'information de Fisher J et I dans un contexte général, en particulier pour des modèles non linéaires, est une tâche difficile (voir par exemple [DSG10] qui ont contourné cette difficulté à travers des simulations de Monte Carlo). Dans le cadre des processus linéaires (gaussiens), le problème du calcul des matrices d'information de Fisher et de leur inverse a été largement étudié. Diverses expressions de ces matrices ont été données par [Whi53, Whi63, Sid58, Dur59] et [BJ76]. Des algorithmes pour calculer ces matrices d'information ont été proposés par [McL84, KM90, KM04] et [GB06]. Par contre, pour le cas des modèles ARMA faibles, il existe très peu de travaux, à part ces exceptions (à notre connaissance bien sûr) : [FZ00, FZ07] et pour des modèles all-pass par [BDT01].

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 QMV, 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 dont nous établissons la convergence.

Expression des dérivées des résidus du modèle VARMA

Dans le cas de modèles ARMA univariés (*i.e.* quand $d = 1$), [McL78] avait défini des dérivées résiduelles par

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

où A_i et B_j sont respectivement les paramètres AR et MA, par convention $A_0 = B_0 = 1$. 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. Définissons les polynômes Notons A_h^* et B_h^* les coefficients définis par

$$A_\theta^{-1}(z) = \sum_{h=0}^{\infty} A_h^* z^h, \quad B_\theta^{-1}(z) = \sum_{h=0}^{\infty} B_h^* z^h, \quad |z| \leq 1$$

pour $h \geq 0$. 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

$$v_t(\theta) = -A_\theta^{-1}(L)e_t(\theta) = -\sum_{h=0}^{\infty} A_h^* e_{t-h}(\theta), \quad u_t(\theta) = B_\theta^{-1}(L)e_t(\theta) = \sum_{h=0}^{\infty} B_h^* e_{t-h}(\theta)$$

avec l'innovation

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

Pour s'inspirer du cas ARMA univarié, nous commençons par écrire le modèle réduit (1.2) 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 $\mathbf{A}_i = A_0^{-1} \mathbf{A}_i$ et $\mathbf{B}_i = A_0^{-1} \mathbf{B}_i B_0^{-1} A_0$. 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 $\mathbf{A}_h^* = [\mathbf{A}_{11,h}^* : \mathbf{A}_{21,h}^* : \dots : \mathbf{A}_{dd,h}^*]$ et $\mathbf{B}_h^* = [\mathbf{B}_{11,h}^* : \mathbf{B}_{21,h}^* : \dots : \mathbf{B}_{dd,h}^*]$ les matrices de tailles $d \times d^3$. Posons

$$\lambda_h(\theta) = [-\mathbf{A}_{h-1}^* : \dots : -\mathbf{A}_{h-p_0}^* : \mathbf{B}_{h-1}^* : \dots : \mathbf{B}_{h-q_0}^*], \quad (1.26)$$

la matrice de taille $d \times d^3(p_0 + q_0)$ constituée uniquement des paramètres des 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.

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

Proposition 1. *Sous les hypothèses (H1)–(H6), nous avons*

$$\frac{\partial \mathbf{e}_t(\theta)}{\partial \theta'} = [V_{t-1}(\theta) : \dots : V_{t-p_0}(\theta) : U_{t-1}(\theta) : \dots : U_{t-q_0}(\theta)],$$

où

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

avec $\mathbf{A}_h^* = [\mathbf{A}_{11,h}^* : \mathbf{A}_{21,h}^* : \dots : \mathbf{A}_{dd,h}^*]$ et $\mathbf{B}_h^* = [\mathbf{B}_{11,h}^* : \mathbf{B}_{21,h}^* : \dots : \mathbf{B}_{dd,h}^*]$ les matrices de taille $d \times d^3$ définies immédiatement avant (1.26). De plus, pour $\theta = \theta_0$ nous avons

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

où les matrices λ_i sont définies par (1.26).

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

Expressions explicites des matrices J_{11} et I_{11}

Nous donnons des expressions explicites des matrices I_{11} et J_{11} impliquées dans la variance asymptotique Ω_{11} . 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 . Notons

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

la matrice impliquant les moments d'ordre deux de l'innovation (e_t). Maintenant nous donnons l'expression de $J_{11} = J_{11}(\theta_0, \Sigma_{e0})$, dans laquelle les termes dépendant de θ_0 (via les matrices λ_i définies dans (1.26)) sont distingués de ceux dépendant des moments d'ordre deux du processus (e_t) (via la matrice \mathcal{M}) et des termes dépendant de la variance de l'innovation (via la matrice Σ_{e0}).

Proposition 2. *Sous les hypothèses de la proposition 1, nous avons*

$$\text{vec}(J_{11}) = 2 \sum_{i \geq 1} \mathcal{M} \{ \lambda_i' \otimes \lambda_i' \} \text{vec}(\Sigma_{e0}^{-1}),$$

où les matrices λ_i sont définies par (1.26).

La démonstration de cette proposition repose sur le théorème ergodique et la proposition 1.

Comme pour la matrice J_{11} , nous décomposons I_{11} en termes impliquant le paramètre θ_0 du modèle VARMA (1.2) et ceux impliquant la distribution des innovations e_t . Soient les matrices

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

Les termes dépendant des paramètres du modèle VARMA sont les matrices λ_i définies par (1.26). Les matrices

$$\Gamma(i,j) := \sum_{h=-\infty}^{+\infty} \mathcal{M}_{ij,h} \quad (1.27)$$

mettent en jeu les moments d'ordre quatre du bruit blanc faible e_t . Les termes dépendant de la variance de l'innovation sont dans la matrice Σ_{e0} . Nous énonçons maintenant la proposition suivante pour la matrice $I_{11} = I_{11}(\theta_0, \Sigma_{e0})$ qui est similaire à la proposition 2.

Proposition 3. *Sous les hypothèses de la proposition 2, nous avons*

$$\text{vec}(I_{11}) = 4 \sum_{i,j=1}^{+\infty} \Gamma(i,j) \left(\{ I_d \otimes \lambda_j' \} \otimes \{ I_d \otimes \lambda_i' \} \right) \text{vec} \left(\text{vec}(\Sigma_{e0}^{-1}) \{ \text{vec}(\Sigma_{e0}^{-1}) \}' \right),$$

où les matrices λ_i sont définies par (1.26).

La démonstration de cette proposition repose sur la proposition 1, ainsi que sur quelques formules de calcul sur le produit de Kronecker et l'opérateur vec .

Remarque 1.5. *En considérant le cas univarié, i.e. quand $d = 1$, nous avons*

$$\text{vec}(J_{11}) = 2 \sum_{i \geq 1} \{ \lambda_i \otimes \lambda_i \}' \quad \text{et} \quad \text{vec}(I_{11}) = \frac{4}{\sigma^4} \sum_{i,j=1}^{+\infty} \gamma(i,j) \{ \lambda_j \otimes \lambda_i \}',$$

où $\gamma(i, j) = \sum_{h=-\infty}^{+\infty} \mathbb{E}(e_t e_{t-i} e_{t-h} e_{t-j-h})$ et les vecteurs $\lambda'_i \in \mathbb{R}^{p_0+q_0}$ sont définis par (1.26).

Remarque 1.6. Toujours pour le cas univarié, [FRZ05] ont utilisé l'estimateur des moindres carrés et ont obtenu

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

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

$$\text{vec}(J_{11}) = 2 \sum_{i \geq 1} \lambda_i \otimes \lambda_i \quad \text{et} \quad \text{vec}(I_{11}) = \frac{4}{\sigma^4} \sum_{i, j \geq 1} \gamma(i, j) \lambda_i \otimes \lambda_j,$$

lesquelles sont les expressions données dans la remarque 1.5.

Définition des estimateurs de $\text{vec}(J_{11})$ et $\text{vec}(I_{11})$

Étant donné que nous disposons des expressions explicites de I_{11} et J_{11} , nous nous intéressons à l'estimation de ces matrices. 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 2, nous définissons un estimateur $\hat{J}_{11,n}$ de J_{11} par

$$\text{vec}(\hat{J}_{11,n}) = \sum_{i \geq 1} \hat{\mathcal{M}}_n \left\{ \hat{\lambda}'_i \otimes \hat{\lambda}'_i \right\} \text{vec}(\hat{\Sigma}_{e_0}^{-1}).$$

Nous énonçons maintenant le théorème suivant qui montre la consistance forte de $\hat{J}_{11,n}$.

Théorème 5. (Convergence forte de $\hat{J}_{11,n}$) *Sous les hypothèses de la proposition 3, nous avons*

$$\hat{J}_{11,n} \rightarrow J_{11} \quad \text{p.s. quand } n \rightarrow \infty.$$

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 $\hat{\Sigma}_{e_0}^{-1}$.

La matrice I_{11} est plus délicate à estimer car son expression explicite fait intervenir une infinité de moments d'ordre quatre à travers (1.27). Afin d'estimer cette matrice, nous utilisons une technique qui consiste à pondérer convenablement certains moments empiriques (voir [And91, NW87]). Cette pondération se fait au moyen d'une fonction de poids et d'un paramètre de troncature. Soit

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

Pour estimer (1.27), nous considérons une suite de réels $(b_n)_{n \in \mathbb{N}^*}$ telle que

$$b_n \rightarrow 0 \quad \text{et} \quad nb_n^{\frac{10+4\nu}{\nu}} \rightarrow \infty \quad \text{quand} \quad n \rightarrow \infty, \quad (1.28)$$

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

$$b_n \sum_{|h| < n} |f(hb_n)| = O(1). \quad (1.29)$$

Soit

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

Nous considérons la matrice

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

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

En vue de la proposition 3, nous définissons un estimateur $\hat{l}_{11, n}$ de l_{11} par

$$\text{vec}(\hat{l}_{11, n}) = 4 \sum_{i, j=1}^{+\infty} \hat{\mathbf{r}}_n(i, j) \left(\{ I_d \otimes \hat{\lambda}'_i \} \otimes \{ I_d \otimes \hat{\lambda}'_j \} \right) \text{vec} \left(\text{vec}(\hat{\Sigma}_{e_0}^{-1}) \{ \text{vec}(\hat{\Sigma}_{e_0}^{-1}) \}' \right).$$

Maintenant nous énonçons le théorème suivant qui montre la consistance faible de l'estimateur empirique $\hat{l}_{11, n}$.

Théorème 6. (Consistance faible de \hat{l}_n) *Sous les hypothèses du théorème 5, nous avons*

$$\hat{l}_{11, n} \rightarrow l_{11} \quad \text{en probabilité, quand} \quad n \rightarrow \infty.$$

Démonstration. Elle repose sur une série de trois lemmes, dans lesquels nous montrons les convergences de $\hat{\lambda}_i$ et de $\hat{\Sigma}_{e_0}^{-1}$. Nous montrons le dernier lemme, qui donne la convergence de $\hat{\mathbf{r}}_n(i, j)$ uniformément en i et j .

Pour tout $\theta \in \Theta$, on a

$$\begin{aligned} \mathcal{M}_{n \, ij, h}(\theta) &:= \frac{1}{n} \sum_{t=1}^{n-|h|} \left(\{ e'_{t-h}(\theta) \otimes (I_{d^2(p_0+q_0)} \otimes e'_{t-j-h}(\theta)) \} \otimes \{ e'_t(\theta) \otimes (I_{d^2(p_0+q_0)} \otimes e'_{t-i}(\theta)) \} \right) \\ &\rightarrow \mathcal{M}_{ij, h}(\theta) := \mathbb{E} \left(\{ e'_{t-h}(\theta) \otimes (I_{d^2(p_0+q_0)} \otimes e'_{t-j-h}(\theta)) \} \otimes \{ e'_t(\theta) \otimes (I_{d^2(p_0+q_0)} \otimes e'_{t-i}(\theta)) \} \right), \end{aligned}$$

presque sûrement, par application du théorème ergodique. Un développement de Taylor de $\text{vec} \hat{\mathcal{M}}_{n \, ij, h}$ autour de θ_0 et (1.12) donne

$$\text{vec} \hat{\mathcal{M}}_{n \, ij, h} = \text{vec} \mathcal{M}_{n \, ij, h} + \left\{ \frac{\partial \text{vec} \mathcal{M}_{n \, ij, h}(\theta)}{\partial \theta'} \right\}_{\theta=\theta_0} (\hat{\theta}_n - \theta_0) + O_{\mathbb{P}} \left(\frac{1}{n} \right).$$

Sous l'hypothèse que $\sup_{i,j \geq 1} \|\Gamma(i, j)\| = O(1)$, on déduit que

$$\lim_{n \rightarrow \infty} \sup_{i,j \geq 1} \sup_{|h| < n} \left\| \frac{\partial \text{vec} \mathcal{M}_{n, ij, h}}{\partial \theta'} \right\| < \infty, \quad p.s. \quad (1.30)$$

Par le théorème ergodique, (1.12) et (1.30), pour toute norme multiplicative, nous avons

$$\begin{aligned} \sup_{i,j \geq 1} \sup_{|h| < n} \left\| \text{vec} \left(\hat{\mathcal{M}}_{n, ij, h} - \mathcal{M}_{ij, h} \right) \right\| &\leq \lim_{n \rightarrow \infty} \sup_{i,j \geq 1} \sup_{|h| < n} \left\| \frac{\partial \text{vec} \mathcal{M}_{n, ij, h}}{\partial \theta'} \right\| \left\| \hat{\theta}_n - \theta_0 \right\| + O_{\mathbb{P}} \left(\frac{1}{n} \right) \\ &= O_{\mathbb{P}} \left(\frac{1}{\sqrt{n}} \right). \end{aligned} \quad (1.31)$$

Notons que

$$\hat{\Gamma}_n(i, j) - \Gamma(i, j) = \sum_{h=-T_n}^{+T_n} f(hb_n) \left(\hat{\mathcal{M}}_{n, ij, h} - \mathcal{M}_{ij, h} \right) + \sum_{h=-T_n}^{+T_n} \{f(hb_n) - 1\} \mathcal{M}_{ij, h} - \sum_{|h| > T_n} \mathcal{M}_{ij, h}.$$

Par l'inégalité triangulaire et pour toute norme multiplicative, nous avons

$$\sup_{i,j \geq 1} \left\| \hat{\Gamma}_n(i, j) - \Gamma(i, j) \right\| \leq g_1 + g_2 + g_3,$$

où

$$\begin{aligned} g_1 &= \sup_{i,j \geq 1} \sup_{|h| < n} \left\| \hat{\mathcal{M}}_{n, ij, h} - \mathcal{M}_{ij, h} \right\| \sum_{|h| \leq T_n} |f(hb_n)|, \\ g_2 &= \sum_{|h| \leq T_n} |f(hb_n) - 1| \|\mathcal{M}_{ij, h}\| \quad \text{et} \quad g_3 = \sum_{|h| > T_n} \|\mathcal{M}_{ij, h}\|. \end{aligned}$$

Les éléments non nuls de la matrice $\mathcal{M}_{ij, h}$ sont de la forme $\mathbb{E}(e_{i_1 t} e_{i_2 t-i} e_{i_3 t-h} e_{i_4 t-h-j})$, pour $(i_1, i_2, i_3, i_4) \in \{1, \dots, d\}^4$. En utilisant l'inégalité de [Dav68], on obtient

$$|\mathbb{E}(e_{i_1 t} e_{i_2 t-i} e_{i_3 t-h} e_{i_4 t-h-j})| = |\text{Cov}(e_{i_1 t} e_{i_2 t-i}, e_{i_3 t-h} e_{i_4 t-h-j})| \leq K \alpha_{\epsilon}^{\nu/(2+\nu)}(h)$$

et on en déduit que

$$\|\mathcal{M}_{ij, h}\| \leq K \alpha_{\epsilon}^{\nu/(2+\nu)}(h). \quad (1.32)$$

Ainsi, sous (H5), on a $g_3 \rightarrow 0$ quand $n \rightarrow \infty$, car c'est le reste d'une série convergente. Pour m entier fixé, on a $g_2 \leq s_1 + s_2$, où

$$s_1 = \sum_{|h| \leq m} |f(hb_n) - 1| \|\mathcal{M}_{ij, h}\| \quad \text{et} \quad s_2 = \sum_{m < |h| \leq T_n} |f(hb_n) - 1| \|\mathcal{M}_{ij, h}\|.$$

Pour $|h| \leq m$, on a $hb_n \rightarrow 0$ quand $n \rightarrow \infty$ et $f(hb_n) \rightarrow 1$, donc $s_1 \rightarrow 0$. Étant donné que $\sum_h \|\mathcal{M}_{ij, h}\| \leq \infty$ (voir (1.32)) et que $f(\cdot)$ est bornée, s_2 peut être rendu arbitrairement petit en choisissant m suffisamment grand. D'où $g_2 \rightarrow 0$. Au vu de (1.29) et (1.31), on a

$$g_1 \leq \frac{1}{b_n} \sup_{i,j \geq 1} \sup_{|h| < n} \left\| \hat{\mathcal{M}}_{n, ij, h} - \mathcal{M}_{ij, h} \right\| O(1) = O_{\mathbb{P}} \left(\frac{1}{b_n \sqrt{n}} \right) = o_{\mathbb{P}}(1),$$

puisque $nb_n^2 \rightarrow \infty$ (voir (1.28)). Ensuite, sous les hypothèse du théorème 2, on obtient

$$\sup_{i,j} \left\| \hat{\Gamma}_n(i,j) - \Gamma(i,j) \right\| \rightarrow 0 \text{ en probabilité quand } n \rightarrow \infty.$$

□

Les théorèmes 5 et 6 montrent alors que

$$\hat{\Omega}_{11,n} := \hat{J}_{11,n}^{-1} \hat{I}_{11,n} \hat{J}_{11,n}^{-1}$$

est un estimateur faiblement consistant de la matrice de covariance asymptotique Ω_{11} .

1.2.3 Estimation de la matrice Ω proposée dans [A7]

Dans [A7], nous avons proposé une généralisation de la méthode introduite dans [A9] pour un processus ARMA univarié (c'est-à-dire du modèle (4.13) pour $d = 1$). Concrètement, nous avons proposé des écritures explicites plus générales des matrices d'information I et J intervenant dans Ω . Ces matrices d'information I et J sont évaluées à n'importe quel point θ de l'espace compact des paramètres, pas nécessairement en la vraie valeur du paramètre θ_0 (contrairement à ce qui est fait dans [A5, A9] et dans toutes les références mentionnées jusque-là). Pour certaines applications (par exemple pour le calcul théorique des puissances des tests ou des pentes de Bahadur sous l'alternative), il est nécessaire d'évaluer ces matrices d'information en $\theta \neq \theta_0$. Nous avons également établi la convergence des estimateurs proposés.

Dans ce cas, nous obtenons une extension directe du théorème 2 en univarié.

Lemme 1. *Sous les hypothèses du théorème 2, pour $d = 1$, soit (X_t) une solution causale ou non anticipative de l'équation (4.13) et soit $(\hat{\theta}_n)$ une suite d'estimateurs du QMV. Alors nous avons*

$$\sqrt{n} (\hat{\theta}_n - \theta_0) \overset{d}{\rightsquigarrow} \mathcal{N}(0, \Omega = J^{-1} I J^{-1}) \text{ quand } n \rightarrow +\infty, \quad (1.33)$$

où $I = I_{\theta_0}$, $J = J_{\theta_0} = J_{\theta_0}^*$, avec

$$I_{\theta} = \frac{4}{\sigma_0^4} \sum_{h=-\infty}^{+\infty} \text{Cov} \left\{ \epsilon_t(\theta) \frac{\partial \epsilon_t(\theta)}{\partial \theta}, \epsilon_{t-h}(\theta) \frac{\partial \epsilon_{t-h}(\theta)}{\partial \theta'} \right\},$$

$$J_{\theta} = \frac{2}{\sigma_0^2} \mathbb{E} \frac{\partial \epsilon_t(\theta)}{\partial \theta} \frac{\partial \epsilon_t(\theta)}{\partial \theta'}, \quad J_{\theta}^* = \frac{2}{\sigma_0^2} \mathbb{E} \epsilon_t(\theta) \frac{\partial^2 \epsilon_t(\theta)}{\partial \theta \partial \theta'} + \frac{2}{\sigma_0^2} \mathbb{E} \frac{\partial \epsilon_t(\theta)}{\partial \theta} \frac{\partial \epsilon_t(\theta)}{\partial \theta'}.$$

Ces deux versions des matrices J_{θ} et J_{θ}^* sont facilement estimables par

$$\hat{J}_n = \frac{2}{\hat{\sigma}^2} \frac{1}{n} \sum_{t=1}^n \frac{\partial \tilde{\epsilon}_t(\hat{\theta}_n)}{\partial \theta} \frac{\partial \tilde{\epsilon}_t(\hat{\theta}_n)}{\partial \theta'}, \quad \hat{J}_n^* = \frac{2}{\hat{\sigma}^2} \frac{1}{n} \sum_{t=1}^n \tilde{\epsilon}_t(\hat{\theta}_n) \frac{\partial^2 \tilde{\epsilon}_t(\hat{\theta}_n)}{\partial \theta \partial \theta'} + \hat{J}_n, \quad (1.34)$$

et peuvent être considérées comme estimateur de la matrice J . Les matrices J_{θ} , J_{θ}^* et I_{θ} peuvent aussi être appelées des matrices d'information.

Remarque 1.7. *Quand l'hypothèse (H3) est remplacée par l'hypothèse que les termes d'erreur*

(ϵ_t) sont des différences de martingales, nous avons

$$I = I_{ss} := \frac{4}{\sigma_0^4} \mathbb{E} \epsilon_t^2 \frac{\partial \epsilon_t(\theta_0)}{\partial \theta} \frac{\partial \epsilon_t(\theta_0)}{\partial \theta'}.$$

Expression théorique des matrices d'information

Matrice d'information J_θ

En différenciant les deux membres de l'équation $A_\theta(L)X_t = B_\theta(L)\epsilon_t(\theta)$, pour $i, k = 1, \dots, p_0$ et $j, \ell = 1, \dots, q_0$, on obtient les dérivées résiduelles suivantes

$$\begin{aligned} \frac{\partial}{\partial A_i} \epsilon_t(\theta) &= -B_\theta^{-1}(L)X_{t-i} = -B_\theta^{-1}A_{\theta_0}^{-1}B_{\theta_0}(L)\epsilon_{t-i} := -\sum_{h=0}^{\infty} c_h^a \epsilon_{t-i-h} \\ \frac{\partial}{\partial B_j} \epsilon_t(\theta) &= -B_\theta^{-2}A_{\theta_0}^{-1}A_\theta B_{\theta_0}(L)\epsilon_{t-j} := -\sum_{h=0}^{\infty} c_h^b \epsilon_{t-j-h}, \\ \frac{\partial^2}{\partial B_j \partial A_i} \epsilon_t(\theta) &= B_\theta^{-2}A_{\theta_0}^{-1}B_{\theta_0}(L)\epsilon_{t-i-j} := \sum_{h=0}^{\infty} c_h^{ab} \epsilon_{t-i-j-h}, \\ \frac{\partial^2}{\partial B_j \partial B_\ell} \epsilon_t(\theta) &= 2B_\theta^{-3}A_{\theta_0}^{-1}A_\theta B_{\theta_0}(L)\epsilon_{t-j-\ell} := \sum_{h=0}^{\infty} c_h^{bb} \epsilon_{t-j-\ell-h}, \end{aligned}$$

et $\partial^2 \epsilon_t(\theta) / \partial A_i \partial A_k = 0$. De plus, on a

$$\epsilon_t(\theta) = B_\theta^{-1}A_{\theta_0}^{-1}A_\theta B_{\theta_0}(L)\epsilon_t := \sum_{h=0}^{\infty} c_h \epsilon_{t-h}.$$

Nous en déduisons le résultat suivant.

Proposition 4. *Les éléments des matrices J_θ et J_θ^* sont donnés par*

$$\begin{aligned} J_\theta(i, k) &= J_\theta^*(i, k) = 2 \sum_{s=0}^{\infty} c_{s+k-i}^a c_s^a, \\ J_\theta(p_0 + j, p_0 + \ell) &= 2 \sum_{s=0}^{\infty} c_{s+\ell-j}^b c_s^b, \\ J_\theta^*(p_0 + j, p_0 + \ell) &= 2 \sum_{s=0}^{\infty} c_{s+j+\ell} c_s^{bb} + J_\theta(p_0 + j, p_0 + \ell), \\ J_\theta(i, p_0 + \ell) &= 2 \sum_{s=\max\{0, i-\ell\}}^{\infty} c_{s+\ell-i}^a c_s^b, \\ J_\theta^*(i, p_0 + \ell) &= 2 \sum_{s=0}^{\infty} c_{s+i+\ell} c_s^{ab} + J_\theta(i, p_0 + \ell), \end{aligned}$$

pour $1 \leq i \leq k \leq p_0$ et $1 \leq j \leq \ell \leq q_0$.

Matrice d'information I_θ

Afin de trouver des expressions similaires pour I_θ , nous posons

$$\Gamma(m, m') = \sum_{h=-\infty}^{+\infty} \text{Cov}(\epsilon_t \epsilon_{t-m}, \epsilon_h \epsilon_{h-m'}). \quad (1.35)$$

Dans le cas fort (*i.e.* quand on suppose que les termes d'erreur sont iid), (1.35) se résume à

$$\Gamma(0, 0) = \mu_4 - \sigma^4, \quad \Gamma(m, m) = \Gamma(m, -m) = \sigma^4, \quad \Gamma(m', m'') = 0, \quad (1.36)$$

avec $\mu_4 = \mathbb{E}\epsilon_1^4$, $m \neq 0$ et $|m'| \neq |m''|$. Quand on suppose que les termes d'erreur sont des différences de martingales, on obtient également des simplifications sous l'hypothèse de symétrie suivante

$$\mathbb{E}\epsilon_{t_1} \epsilon_{t_2} \epsilon_{t_3} \epsilon_{t_4} = 0 \quad \text{quand} \quad t_1 \neq t_2, t_1 \neq t_3 \text{ et } t_1 \neq t_4. \quad (1.37)$$

Une hypothèse similaire à (1.37) est faite dans [FZ09] (voir Théorème 2) afin d'étudier le comportement asymptotique des autocorrélations empiriques d'un processus non linéaire. Dans [A7], nous avons montré que l'hypothèse (1.37) est satisfaite, en particulier, pour des modèles GARCH admettant des moments d'ordre quatre et dont les innovations sont de loi symétrique. En général, cette hypothèse (1.37) est vérifiée pour d'autres classes de modèles de différences de martingales. Dans ce cas, (1.35) donne

$$\Gamma(0, 0) = \sum_{h=-\infty}^{+\infty} \text{Cov}(\epsilon_t^2, \epsilon_{t-h}^2), \quad \Gamma(m, m) = \mathbb{E}\epsilon_t^2 \epsilon_{t-m}^2, \quad \Gamma(m', m'') = 0 \quad (1.38)$$

quand $m \neq 0$ et $|m'| \neq |m''|$. Nous vérifions cette condition (1.38) sur un GARCH(1, 1) dans l'exemple suivant.

Exemple 1.1. *Considérons le modèle GARCH(1, 1) suivant*

$$\begin{cases} \epsilon_t = \sqrt{h_t} \eta_t, & t = 1, 2, \dots \\ h_t = \omega + \alpha \epsilon_{t-1}^2 + \beta h_{t-1}, & (\eta_t) \sim \text{IID}(0, 1) \end{cases}$$

où $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$ et $\alpha^2 \mathbb{E}\eta_1^4 + \beta^2 + 2\alpha\beta < 1$ ¹. Posons la variable $\nu_t := \epsilon_t^2 - \sigma_t^2$ comme étant l'innovation du processus ϵ_t^2 , on obtient

$$\begin{aligned} \Gamma(0, 0) &= \mathbb{E}\nu_t^2 \frac{(1-\beta)^2}{(1-\alpha-\beta)^2}, \\ \Gamma(1, 1) &= \mathbb{E}\nu_t^2 \left(\alpha + \frac{\alpha^2(\alpha+\beta)}{1-(\alpha+\beta)^2} \right) + (\mathbb{E}\sigma_t^2)^2 \\ \Gamma(m, m) &= (\alpha+\beta)\Gamma(m-1, m-1) + \omega \mathbb{E}\sigma_t^2, \quad m > 1, \end{aligned}$$

1. Cette condition est nécessaire et suffisante pour l'existence d'une solution non anticipative et stationnaire à l'ordre quatre (voir *e.g.* Exemple 2.3 dans [FZ10]).

avec $\mathbb{E}\nu_t^2 = \mathbb{E}\eta_1^4 (\mathbb{E}\sigma_t^4 + 1 - 2\mathbb{E}\sigma_t^2)$,

$$\mathbb{E}\sigma_t^2 = \frac{\omega}{1 - \alpha - \beta}, \quad \mathbb{E}\sigma_t^4 = \frac{\omega^2(1 + \alpha + \beta)}{(1 - \alpha^2\mathbb{E}\eta_1^4 - \beta^2 - 2\alpha\beta)(1 - \alpha - \beta)}.$$

Proposition 5. Les éléments de la matrice l_θ sont donnés par

$$\begin{aligned} l_\theta(i, k) &= 4 \sum_{h_1, h_2, h_3, h_4=0}^{+\infty} c_{h_1} c_{h_2}^a c_{h_3} c_{h_4}^a \Gamma(h_2 + i - h_1, h_4 + k - h_3), \\ l_\theta(j, \ell) &= 4 \sum_{h_1, h_2, h_3, h_4=0}^{+\infty} c_{h_1} c_{h_2}^b c_{h_3} c_{h_4}^b \Gamma(h_2 + j - h_1, h_4 + \ell - h_3), \\ l_\theta(i, \ell) &= 4 \sum_{h_1, h_2, h_3, h_4=0}^{+\infty} c_{h_1} c_{h_2}^a c_{h_3} c_{h_4}^b \Gamma(h_2 + i - h_1, h_4 + \ell - h_3), \end{aligned}$$

pour $1 \leq i \leq k \leq p_0$ et $1 \leq j \leq \ell \leq q_0$.

Remarque 1.8. Notons que $c_0 = 1$ et quand $\theta = \theta_0$, on a $c_h = 0$ pour $h > 0$. L'expression de $l = l_{\theta_0}$ se résume à celle obtenue par [FRZ05]. Au vu de (1.36) et (1.38), on obtient également des simplifications dans le cas des modèles ARMA forts et des modèles ARMA dont l'innovation est une différence de martingales; en effet, la matrice l_θ s'obtient en sommant trois indices au lieu de quatre.

Exemples de calculs explicites et numériques des matrices J_θ and l_θ

Nous allons calculer les matrices d'information d'un modèle ARMA(1, 1) au point $\theta^* = (a, 0)'$ alors que le vrai paramètre est $\theta_0 = (0, b_0)'$ (i.e. que le processus générateur de données (noté dans suite, DGP pour data generating process) est une MA(1)). Avec nos notations, on a

$$\begin{aligned} \epsilon_t(\theta) &= \sum_{h=0}^{\infty} (-b)^h (X_{t-h} - aX_{t-h-1}) \\ &= \sum_{h=0}^{\infty} (-b)^h (\epsilon_{t-h} + (b_0 - a)\epsilon_{t-h-1} - ab_0\epsilon_{t-h-2}) \\ \frac{\partial \epsilon_t(\theta)}{\partial a} &= \sum_{h=0}^{\infty} (-b)^h (-a\epsilon_{t-h-1} - b_0\epsilon_{t-h-2}) \\ \frac{\partial \epsilon_t(\theta)}{\partial b} &= \sum_{h=1}^{\infty} -h(-b)^{h-1} (\epsilon_{t-h} + (b_0 - a)\epsilon_{t-h-1} - ab_0\epsilon_{t-h-2}). \end{aligned}$$

Appliqué en $\theta^* = (a, 0)'$, nous obtenons $\epsilon_t(\theta^*) = \epsilon_t + (b_0 - a)\epsilon_{t-1} - ab_0\epsilon_{t-2}$, $\partial \epsilon_t(\theta^*)/\partial a = -\epsilon_{t-1} - b_0\epsilon_{t-2}$, $\partial \epsilon_t(\theta^*)/\partial b = -\epsilon_{t-1} - (b_0 - a)\epsilon_{t-2} + ab_0\epsilon_{t-3}$, $\partial^2 \epsilon_t(\theta^*)/\partial a^2 = 0$, $\partial^2 \epsilon_t(\theta^*)/\partial a \partial b =$

$\epsilon_{t-2} + b_0\epsilon_{t-3}$ et $\partial^2\epsilon_t(\theta^*)/\partial b^2 = 2\epsilon_{t-2} + 2(b_0 - a)\epsilon_{t-3} - 2ab_0\epsilon_{t-4}$. Nous en déduisons alors

$$J_{\theta^*} = 2 \begin{pmatrix} 1 + b_0^2 & 1 + b_0(b_0 - a) \\ 1 + b_0(b_0 - a) & 1 + (b_0 - a)^2 + a^2b_0^2 \end{pmatrix},$$

$$J_{\theta^*}^* = J_{\theta^*} + 2 \begin{pmatrix} 0 & -ab_0 \\ -ab_0 & -2ab_0 \end{pmatrix}.$$

Maintenant, nous supposons que ϵ_t est un bruit blanc faible introduit par [RT96] et défini par

$$\epsilon_t = \eta_t \eta_{t-1} \cdots \eta_{t-k}, \quad (\eta_t) \sim \text{IID } \mathcal{N}(0, 1), \quad k \geq 0. \quad (1.39)$$

Il est impossible d'obtenir explicitement la matrice I_θ en $\theta = \theta^*$, mais au point $\theta = \theta_0$ nous obtenons les écritures explicites suivantes des matrices d'information avec ϵ_t défini par (1.39) :

$$J_{\theta_0} = J_{\theta_0}^* = 2 \begin{pmatrix} 1 & 1 \\ 1 & 1/(1 - b_0^2) \end{pmatrix},$$

$$I_{\theta_0} = 4 \times 3^k \begin{pmatrix} 1 & 1 \\ 1 & \frac{1 - (b_0^2/3)^{k+1}}{1 - b_0^2/3} + \frac{b_0^{2(k+1)}}{3^k(1 - b_0^2)} \end{pmatrix}.$$

Remarquons que $I_{\theta_0} = 2J_{\theta_0}$ dans le cas fort (*i.e.* quand $k = 0$ dans (1.39)). Pour des modèles plus complexes, ou en particulier au point $\theta^* \neq \theta_0$, l'évaluation de ces matrices d'information n'est pas faisable explicitement mais peut facilement être obtenue numériquement. Par exemple, dans ce cas avec $k = 3$, $\theta_0 = (0, 0.5)'$ et $\theta^* = (-0.4, -0.5)'$, on obtient

$$J_{\theta^*} = \begin{pmatrix} 2.33 & 4.33 \\ 4.33 & 11.25 \end{pmatrix}, \quad J_{\theta^*}^* = \begin{pmatrix} 2.33 & 6.33 \\ 6.33 & 17.65 \end{pmatrix}, \quad I_{\theta^*} = \begin{pmatrix} 1161.92 & 2177.66 \\ 2177.66 & 4187.63 \end{pmatrix}.$$

1.2.4 Approximation des matrices d'information par des sommes finies

Dans la pratique, les sommes infinies dans les matrices J_θ , J_θ^* et I_θ seront tronquées. Par exemple, la matrice I_θ qui est la plus compliquée des trois matrices peut être approchée par

$$I_\theta^M(i, k) = \sum_{h_1, h_2, h_3, h_4=0}^M c_{h_1} c_{h_2}^a c_{h_3} c_{h_4}^a \Gamma(h_2 + i - h_1, h_4 + k - h_3),$$

pour $1 \leq i \leq k \leq p$, et on définit de manière similaire les autres éléments de cette matrice. La proposition suivante fournit une valeur de M telle que I_θ^M soit égale I_θ pour une certaine tolérance ε . Nous utiliserons la norme matricielle définie par $\|A\| = \sum_{i,j} |A(i, j)|$.

Proposition 6. Notons $\bar{\rho}$ l'inverse du plus grand module des racines des polynômes A_{θ_0} et B_θ . Soit

$$\bar{\Gamma} = \max_{m, m' \geq 0} |\Gamma(m, m')|, \quad \bar{\pi} = \left(1 + \sum_{j=1}^{q_0} |B_{0j}| \right) \max_{i=0, \dots, p_0} |A_i|$$

où $|A_0| = 1$, et soit

$$K = \sqrt{2}(p_0 + q_0 + 1)\bar{\pi}^{-1/4} \left(\frac{-2(p_0 + 2q_0 - 1)}{\log \bar{\rho}} \right)^{(p_0 + 2q_0 - 1)} \bar{\rho}^{-0.5 - (p_0 + 2q_0 - 1)/\log \bar{\rho}}.$$

Pour tout $\varepsilon > 0$, si

$$M \geq M_\varepsilon := \frac{\log \sqrt{\varepsilon}(1 - \sqrt{\log \bar{\rho}})^2 / K^2}{\log \bar{\rho}}$$

alors

$$\left\| I_\theta - I_\theta^M \right\| \leq (p_0 + q_0)^2 \varepsilon.$$

1.2.5 Estimation des matrices d'information

Posons $J_{\theta,n}$ et $J_{\hat{\theta}_n}^*$ des estimateurs obtenus en remplaçant $\hat{\theta}_n$ par θ dans (1.34) de telle sorte que $\hat{J}_n = J_{\hat{\theta}_n,n}$ et $\hat{J}_n^* = J_{\hat{\theta}_n,n}^*$. Dans la proposition suivante, nous montrons qu'un estimateur de J_{θ^*} peut être (trivialement) déduit de celui θ^* .

Proposition 7. *Sous les hypothèses du lemme 1, quand $n \rightarrow \infty$,*

$$\text{si } \theta_n^* \rightarrow \theta^* \text{ p.s. alors } J_{\theta_n^*,n} \rightarrow J_{\theta^*} \text{ et } J_{\theta_n^*,n}^* \rightarrow J_{\theta^*}^* \text{ p.s.}$$

L'estimation de la variance de long terme I_θ est plus beaucoup plus délicate, comme mentionné précédemment. On peut l'estimer soit par la méthode 1 (HAC), soit par la méthode 2 (de la densité spectrale). Nous étendons ainsi le résultat du théorème 4 pour estimer I_θ par la méthode de la densité spectrale quand θ n'est pas nécessairement égal à θ_0 . Dans ce cas, le processus \mathcal{Y}_t défini par (4.3) est remplacé par le processus ∇_t défini par

$$\nabla_t = S_t - \mathbb{E}S_t, \quad S_t = \epsilon_t(\theta^*) \frac{\partial \epsilon_t(\theta^*)}{\partial \theta}. \quad (1.40)$$

Ensuite, en procédant de la même manière que dans la section 1.2, sous certaines conditions de régularité et si $\theta_n^* \rightarrow \theta^*$ p.s., nous obtenons

$$I_{\theta_n^*,n} := \hat{\Phi}_r^{-1}(1) \hat{\Sigma}_{\hat{u}_r} \hat{\Phi}_r'^{-1}(1) \rightarrow I_{\theta^*} \quad (1.41)$$

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

1.3 Application numérique

Dans cette section, nous présentons quelques résultats de simulations afin d'évaluer numériquement le comportement de l'estimateur du QMV ainsi que les différences de précisions entre les modèles VARMA fort et faible. Nous simulons $N = 1000$ trajectoires indépendantes, de taille $n = 2000$, du modèle VARMA(1, 1) bivarié suivant sous la forme échelon

$$\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & a_{22,1} \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ b_{21,1} & b_{22,1} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix}, \quad (1.42)$$

avec $(a_{22,1}, b_{21,1}, b_{22,1})' = (0.95, 2, 0)'$ et dont le processus des innovations $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t})'$ serait un bruit blanc fort ou faible. Dans un premier temps, nous considérons que (1.42) est un modèle

VARMA fort dont terme d'erreur est défini comme suit

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \sim \text{IID } \mathcal{N}(0, I_2), \quad (1.43)$$

et ensuite que $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t})'$ est donné par

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \eta_{1,t}(|\eta_{1,t-1}| + 1)^{-1} \\ \eta_{2,t}(|\eta_{2,t-1}| + 1)^{-1} \end{pmatrix}, \quad \text{avec } \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix} \sim \text{IID } \mathcal{N}(0, I_2), \quad (1.44)$$

dans le cas du modèle VARMA(1, 1) faible. Le bruit faible défini par (1.44) est une extension directe de celui introduit par [RT96] dans le cas univarié.

La figure 1.1 compare la distribution des estimateurs du QMV dans le cas des modèles VARMA(1, 1) fort et faible. On remarque que les distributions des estimateurs $\hat{a}_{22,1}$ et $\hat{b}_{21,1}$ sont similaires dans les deux cas (fort et faible) contrairement à celle de l'estimateur $\hat{b}_{22,1}$ qui est plus précise dans le cas faible que dans le cas fort. Par contre, dans d'autres situations on aboutit à une analyse inverse. En effet, en considérant ce bruit blanc faible $\epsilon_{i,t} = \eta_{i,t}\eta_{i,t-1}$ pour $i = 1, 2$, on trouve que la distribution de $\hat{b}_{22,1}$ est plus précise dans le cas fort que dans le cas faible. Ceci n'est pas surprenant et est d'ailleurs en concordance avec les résultats de [RT96], qui ont montré que pour des bruits blancs faibles similaires, la variance asymptotique des autocorrélations empiriques peut être plus grande ou plus petite que 1 alors que celle d'un bruit fort est égale à 1.

Enfin, la figure 1.2 compare les distributions des estimateurs de la variance asymptotique Ω : $\hat{\Omega}_S = 2\hat{J}^{-1}$ standard à celles des formes sandwich $\hat{\Omega} = \hat{J}^{-1}\hat{I}\hat{J}^{-1}$. Ici, nous utilisons $\hat{I} = \hat{I}^{\text{SP}}$ défini dans le théorème 10, où l'ordre r de la représentation VAR(∞) est automatiquement sélectionné par un critère AIC. Dans le cas de modèle VARMA forts, on sait que les deux estimateurs sont consistants. Les deux blocs supérieurs de la figure 1.2 montrent que l'estimateur sandwich est moins précis que l'estimateur standard $\hat{\Omega}_S = 2\hat{J}^{-1}$ dans le cas fort. Cela n'est pas surprenant car l'estimateur sandwich est plus robuste dans la mesure où il continue à être convergent dans le cas faible, contrairement à l'estimateur standard. Il est clair que dans ce cas VARMA faible, $n\text{Var}\{\hat{b}_{22,1} - b_{22,1}\}^2$ est mieux estimé par $\hat{\Omega}^{\text{SP}}(3, 3)$ que $\hat{\Omega}_S(3, 3) = 2\hat{J}^{-1}(3, 3)$ (voir les boîtes à moustaches (c) des deux blocs inférieurs de la figure 1.2). Cette différence de précision peut avoir des conséquences importantes sur l'identification et la validation des modèles VARMA faibles (voir les chapitres 2 et 3 pour plus de détails).

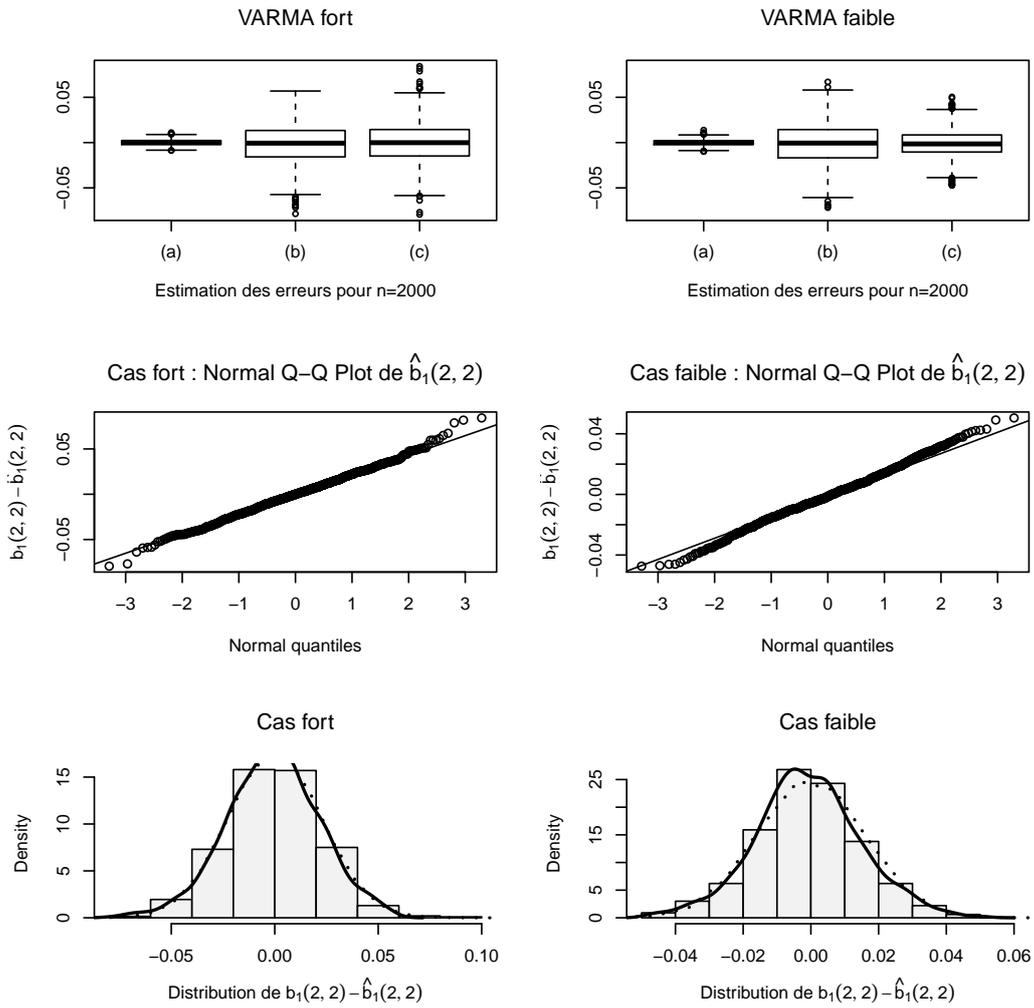


FIGURE 1.1 – Distributions des estimateurs du QMV du modèle (1.42) avec le bruit fort (1.43) (les trois blocs de gauche) et avec le bruit faible (1.44) (les trois blocs de droite). Les points (a)-(c), dans les deux blocs supérieurs, présentent les distributions des erreurs $\hat{\theta}(i) - \theta_0(i)$ pour $i = 1, 2, 3$. Les deux blocs du milieu présentent les Q-Q plots des estimateurs $\hat{\theta}(3) = \hat{b}_{22,1}$ et les deux derniers blocs inférieurs présentent les distributions de $\hat{\theta}(3)$. La densité estimée est tracée en trait plein et celle d'une loi gaussienne centrée de même variance en pointillés.

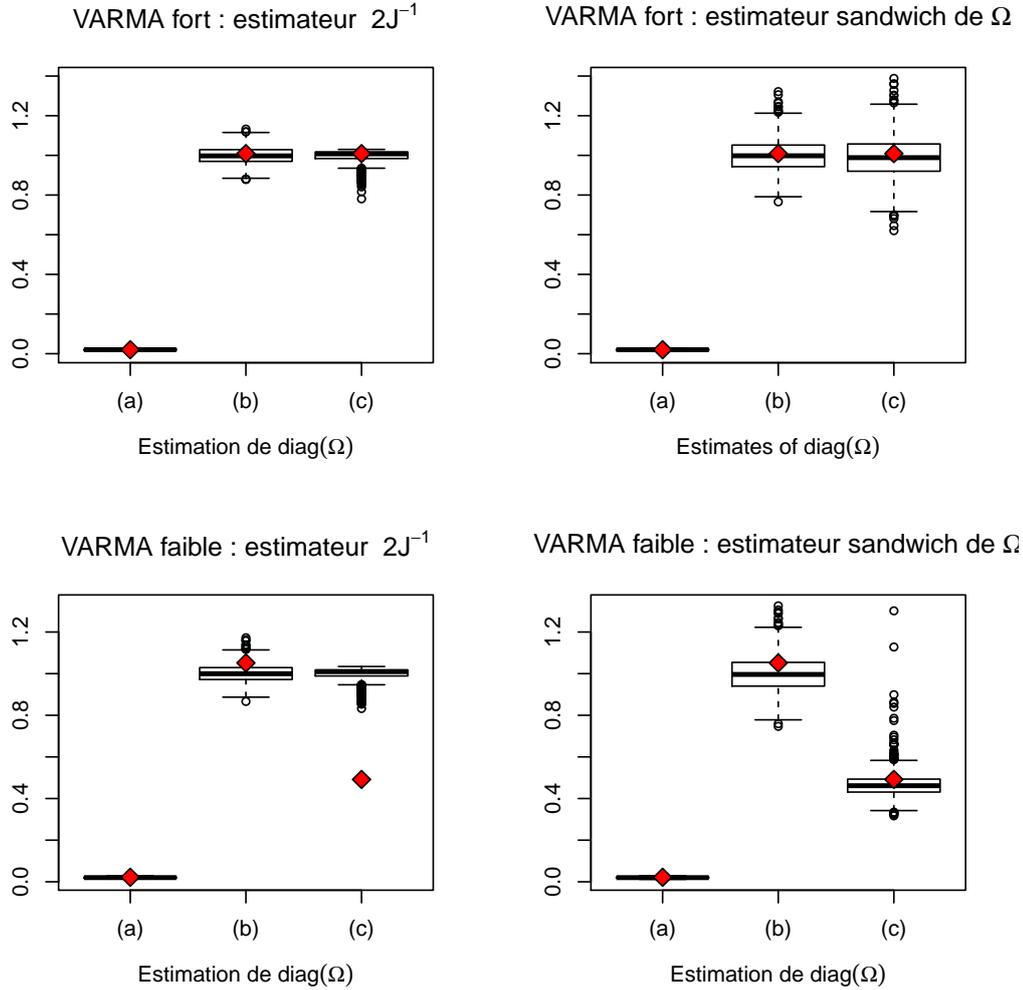


FIGURE 1.2 — Comparaison des estimateurs standard et modifié de la matrice de variance asymptotique Ω des estimateurs du QMV du modèle (1.42) simulé présenté dans la figure 1.1. Le losange rouge représente la moyenne sur les $N = 1000$ répliques indépendantes de $n\{\hat{a}_{22,1} - 0,95\}^2$ pour (a) (0,02 dans les cas fort et faible), $n\{\hat{b}_{21,1} - 2\}^2$ pour (b) (1,02 dans le cas fort et 1,01 dans le cas faible) et $n\{\hat{b}_{22,1}\}^2$ for (c) (0,94 dans le cas fort et 0,43 dans le cas faible).

1.4 Conclusion

Bien que les logiciels développés pour estimer les modèles VARMA forts permettent aussi d'estimer de manière convergente les paramètres d'un modèle VARMA faible, nous avons abordé le problème d'estimation des modèles VARMA faibles, en établissant les propriétés asymptotiques de leurs estimateurs du QMV et en montrant la différence de précision qu'il peut y avoir dans l'estimation (des VARMA forts et faibles). En effet, la matrice de variance asymptotique, $\Omega_s := 2J^{-1}$, des paramètres d'un modèle VARMA fort peut être très différente de celle des paramètres d'un modèle VARMA faible dès que $I \neq 2J$.

En conséquence, les logiciels actuellement utilisés pour ajuster les modèles VARMA forts

fournissent probablement une estimation incorrecte de la matrice de variance asymptotique Ω (puisqu'ils fourniront comme estimateur de Ω , un estimateur $\hat{\Omega}_s$ qui converge vers Ω_s) pour un processus VARMA faible. Cela peut entraîner de graves erreurs de spécification, notamment la sélection de modèles avec des ordres très grands. D'une manière générale, nous justifions une pratique très courante qui consiste à ajuster un modèle VARMA, même si les données dont on dispose présentent des non linéarités évidentes.

Le travail présenté dans ce chapitre serait donc fondamental pour l'inférence statistique des modèles VARMA faibles (voir les chapitres 2 et 3).

Chapitre 2

Contribution à l'identification des modèles ARMA faibles multivariés

Dans ce chapitre, on s'intéresse à la sélection des ordres p_0 et q_0 du modèle VARMA défini par (4.13) dont les termes d'erreur sont non corrélés mais non nécessairement indépendants.

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.

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 (voir Chapitre 3) ou encore une estimation non convergente des paramètres.

Parmi les méthodes d'identification, les plus populaires sont celles basées sur l'optimisation d'un critère d'information. L'objectif dans [A4, A6, A11] est d'étudier le problème de sélection des ordres p_0 et q_0 de modèles VARMA faibles. Les fondements théoriques des critères d'information ne sont plus établis lorsque l'hypothèse de bruit iid est relâchée. Afin de remédier à ce problème, nous proposons de nouveaux critères d'information modifiés qui tiennent compte des non linéarités du type de bruit à travers les matrices d'information de Fisher I et J . Nous avons proposé dans [A4, A6] un critère d'information de Akaike modifié (noté AIC_M) ainsi qu'une modification du critère AIC corrigé (AIC_c). Mais le critère AIC, ainsi que ses dérivés (standard et modifié proposés) ont l'inconvénient de fournir des estimations des ordres p_0 et q_0 non convergentes. C'est pourquoi dans [A11], nous développons des modifications de critères de style BIC (Bayesian Information Criterion) et HQ (Hannan Quinn) adaptées aux VARMA faibles. Ces nouveaux critères modifiés ont l'avantage de fournir des estimations convergentes des ordres p_0 et q_0 .

Nous utilisons la méthode du quasi-maximum de vraisemblance ainsi que la paramétrisation déjà définies dans le chapitre 1 pour l'estimation des paramètres du modèle VARMA(p, q) considéré.

Dans ce chapitre, nous considérons la matrice Σ_e comme un paramètre de nuisance, et afin d'alléger nos notations, nous posons $\theta = \vartheta^{(1)}$. Dans ce chapitre, nous remplacerons également dans la suite l'espace des paramètres Θ par $\Theta_{p,q}$.

Les publications liées à ce chapitre sont : [A4, A6, A11].

2.1 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 dimension 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.

2.2 Critère de sélection non convergent des ordres

Le critère d'information de Akaike (noté AIC pour Akaike's information criterion) introduit par [Aka73] est sans doute le plus connu, et est défini de manière suivante

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

Le critère AIC est fondé sur un estimateur de la divergence (ou de l'information) de Kullback-Leibler, qui mesure l'écart entre le modèle candidat et le vrai modèle. Mais il se révèle que le AIC a tendance à surestimer les ordres (à ne pas être efficace, surtout quand la dimension du vrai modèle est finie). Cela a donné lieu à de nombreuses publications afin d'améliorer ce critère (voir par exemple le critère FPE (final prediction error, en anglais) introduit par [Aka70]). Mais, pour la suite, nous insisterons plutôt sur les travaux de Tsai et Hurvich, qui ont montré que le AIC est un estimateur biaisé de l'information de Kullback-Leibler quand la taille n de l'échantillon est petite (voir [HT89]). Ils ont donc 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 (voir, respectivement, [HT89, HT93]). Leur critère AICc (AIC corrigé) se révèle efficace pour des échantillons de petites tailles et est, de surcroît, asymptotiquement égal au critère AIC. Nous allons dans un premier temps adapter ce critère AICc au cas des VARMA faibles.

2.2.1 Critère AICc dans le cas d'erreurs faibles

Rappelons que, d'après (1.7) et (1.17), nous avons respectivement

$$\begin{aligned} \ell_n(\theta) &= \tilde{\ell}_n(\theta) + o(1) \text{ p.s.}, \\ \frac{\partial \ell_n(\theta)}{\partial \theta} &= \frac{\partial \tilde{\ell}_n(\theta)}{\partial \theta} + o(1) \text{ p.s.} \end{aligned}$$

Notons que les mêmes égalités sont aussi vérifiées pour les dérivées secondes de $\tilde{\ell}_n$. Pour tout $\theta \in \Theta_{p,q}$, nous avons

$$-2 \log L_n(\theta) = nd \log(2\pi) + n \log \det \Sigma_e + \sum_{t=1}^n e_t'(\theta) \Sigma_e^{-1} e_t(\theta). \quad (2.1)$$

Notons que minimiser l'information de Kullback-Leibler pour tout modèle candidat caractérisé par le paramètre θ , revient à minimiser le contraste $\Delta(\theta) := \mathbb{E} \{-2 \log L_n(\theta)\}$. En omettant la constante $nd \log(2\pi)$ dans (2.1), nous trouvons que

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

où $S(\theta) = \mathbb{E} e_1(\theta) e_1'(\theta)$. Nous énonçons le lemme suivant qui montre que l'application $\theta \mapsto \Delta(\theta)$ est minimale en $\theta = \theta_0$.

Lemme 2. *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 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 $d \times d$.

Estimation de la divergence moyenne

Posons $\hat{\theta}_n$ l'estimateur du QMV du paramètre θ du modèle VARMA candidat. Soit $\hat{e}_t = \tilde{e}_t(\hat{\theta}_n)$ les résidus du QMV obtenus en remplaçant \hat{v}_n par $\hat{\theta}_n$ dans (1.23). En vue du Lemme 2, il est naturel de vouloir minimiser le contraste moyen, $\mathbb{E} \Delta(\hat{\theta}_n)$, qui s'interprète comme étant l'écart moyen quand on utilise le modèle de paramètre $\hat{\theta}_n$. Ainsi, 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. Dans [BM12], nous avons adapté le critère AICc afin d'obtenir un estimateur approximativement sans biais de $\mathbb{E} \Delta(\hat{\theta}_n)$ pour des modèles VARMA faibles. En utilisant un développement limité de Taylor de $\partial \log L_n(\hat{\theta}_n) / \partial \theta$ au voisinage de θ_0 , il s'ensuit que

$$\hat{\theta}_n - \theta_0 = -\frac{2}{n} J_{11}^{-1} \frac{\partial \log L_n(\theta_0)}{\partial \theta} + o_{\mathbb{P}} \left(\frac{1}{\sqrt{n}} \right) = -\frac{2}{n} \sum_{t=1}^n J_{11}^{-1} \frac{\partial e_t'(\theta_0)}{\partial \theta} \Sigma_{e0}^{-1} e_t(\theta_0) + o_{\mathbb{P}} \left(\frac{1}{\sqrt{n}} \right), \quad (2.2)$$

où $J_{11} = J_{11}(\theta_0)$ est définie dans le chapitre 1. Nous avons

$$\mathbb{E} \Delta(\hat{\theta}_n) = n \mathbb{E} \log \det \hat{\Sigma}_e + n \mathbb{E} \text{Tr}(\hat{\Sigma}_e^{-1} S(\hat{\theta}_n)), \quad (2.3)$$

où $\hat{\Sigma}_e = \Sigma_e(\hat{\theta}_n)$ est un estimateur de la matrice de variance des erreurs, avec $\Sigma_e(\theta) = n^{-1} \sum_{t=1}^n e_t(\theta) e_t'(\theta)$. Le premier terme du second membre de l'équation (2.3) peut être estimé 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 (1.12), un développement de

Taylor de $e_t(\theta)$ au voisinage de θ_0 vérifie

$$e_t(\theta) = e_t(\theta_0) + \frac{\partial e_t(\theta_0)}{\partial \theta'}(\theta - \theta_0) + R_t, \quad (2.4)$$

où

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

avec $\pi = \|\theta - \theta_0\|$ et θ^* est entre θ_0 et θ . Nous obtenons alors

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

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.* $\mathbb{E} e_t(\theta_0) = 0$, nous avons

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

où $\Sigma_{e0} = \Sigma_e(\theta_0)$. Nous pouvons donc réécrire l'écart moyen (2.3) comme suit

$$\mathbb{E} \Delta(\hat{\theta}_n) = n \mathbb{E} \log \det \hat{\Sigma}_e + n \mathbb{E} \text{Tr} \left(\hat{\Sigma}_e^{-1} \Sigma_{e0} \right) + n \mathbb{E} \text{Tr} \left(\hat{\Sigma}_e^{-1} D(\hat{\theta}_n) \right) + n \mathbb{E} \text{Tr} \left(\hat{\Sigma}_e^{-1} \right) O_{\mathbb{P}} \left(\frac{1}{n^2} \right). \quad (2.5)$$

Comme dans une régression multivariée classique, nous avons la relation

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

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

$$\mathbb{E} \left\{ \hat{\Sigma}_e^{-1} \right\} \approx \left\{ \mathbb{E} \hat{\Sigma}_e \right\}^{-1} \approx nd(nd - k_1)^{-1} \Sigma_{e0}^{-1}. \quad (2.6)$$

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

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

Maintenant, en utilisant (1.12), (2.6) et cette dernière égalité appliquée en $\hat{\theta}_n$, nous avons

$$\mathbb{E} \text{Tr} \left\{ \hat{\Sigma}_e^{-1} D(\hat{\theta}_n) \right\} = \frac{d}{2(nd - k_1)} \text{Tr} (I_{11} J_{11}^{-1}),$$

où $J_{11} = 2\mathbb{E} \left\{ \partial e_t'(\theta_0) / \partial \theta \Sigma_{e0}^{-1} \partial e_t(\theta_0) / \partial \theta' \right\}$. Par conséquent, en utilisant (2.6), nous avons

$$\mathbb{E} \text{Tr} \left(\hat{\Sigma}_e^{-1} S(\hat{\theta}_n) \right) = \frac{nd^2}{nd - k_1} + \frac{d}{2(nd - k_1)} \text{Tr} (I_{11} J_{11}^{-1}) + O\left(\frac{1}{n^2}\right).$$

Ainsi, en injectant cette dernière égalité dans (2.3), sous les hypothèses (H1)–(H6), nous obtenons un estimateur approximativement sans biais de $\mathbb{E}\Delta(\hat{\theta}_n)$ donné par

$$\text{AIC}_{cM} := n \log \det \hat{\Sigma}_e + \frac{n^2 d^2}{nd - k_1} + \frac{nd}{2(nd - k_1)} \text{Tr} \left(\hat{I}_{11,n} \hat{J}_{11,n}^{-1} \right), \quad (2.7)$$

où les matrices $\hat{I}_{11,n}$ et $\hat{J}_{11,n}$ sont des estimateurs des matrices I_{11} et J_{11} définis dans le chapitre 1.

Nous obtenons des estimateurs \hat{p} et \hat{q} des ordres p_0 et q_0 en minimisant le critère modifié (2.7).

Remarque 2.1. Dans le cas de modèles VARMA forts c'est-à-dire quand l'hypothèse d'ergodicité (H3) est remplacée par celle dont les termes d'erreurs sont iid, nous avons $I_{11} = 2J_{11}$ (voir Remarque 1.3). Par conséquent, nous avons $\text{Tr} (I_{11} J_{11}^{-1}) = 2k_1$. Dans ce cas, un estimateur approximativement sans biais de $\mathbb{E}\Delta(\hat{\theta}_n)$ prend la forme plus conventionnelle suivante

$$\text{AIC}_{cM} = n \log \det \hat{\Sigma}_e + nd + \frac{nd}{2(nd - k_1)} 2k_1 =: \text{AIC}_c,$$

qui illustre que les critères AIC standard et AIC_{cM} 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_{cM} 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 grande amélioration des performances du AIC_{cM} que le AIC en terme de choix du modèle.

2.2.2 Critère AIC dans le cas d'erreurs faibles

Les observations $X = (X_1, \dots, X_n)$ et $Y = (Y_1, \dots, Y_n)$ sont indépendantes et générées par le même modèle (4.13). Dans la section précédente le contraste (écart) minimal a été approché par $-2\mathbb{E} \log L_n(\hat{\theta}_n)$ (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 $\mathbb{E}\Delta(\hat{\theta}_n)$ (écart moyen) consiste à considérer $\hat{\theta}_n$ comme étant l'estimateur du QMV de θ fondé sur l'observation X et ensuite, nous approchons 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

$$C(\hat{\theta}_n) = -2\mathbb{E}_Y \log L_n(\hat{\theta}_n), \quad (2.8)$$

où \mathbb{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 $\mathbb{E}\Delta(\hat{\theta}_n)$. Le modèle minimisant (2.8) peut être in-

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. Comme dans [Fin93] pour le cas d'erreurs indépendantes, cette moyenne du contraste peut être décomposée comme suit

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

où

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

et

$$a_2 = -2\mathbb{E}_Y \log L_n(\hat{\theta}_n) + 2\mathbb{E}_X \log L_n(\theta_0).$$

L'estimateur QMV 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 cet estimateur du QMV. Notons que $\mathbb{E}_X \log L_n(\theta_0) = \mathbb{E}_Y \log L_n(\theta_0)$, donc le terme a_2 s'interprète comme le coût moyen d'utilisation (sur répliquations 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 $\text{Tr}(I_{11}J_{11}^{-1})/2$.

Proposition 8. *Sous les hypothèses (H1)–(H6), quand $n \rightarrow \infty$, nous avons, à un $o(1)$ près, a_1 et a_2 sont tous deux égaux à $\text{Tr}(I_{11}J_{11}^{-1})/2$.*

Démonstration. En faisant un développement (à l'ordre 2) de Taylor de la log –quasi-vraisemblance, nous obtenons

$$-2 \log L_n(\theta_0) = -2 \log L_n(\hat{\theta}_n) + \frac{n}{2}(\hat{\theta}_n - \theta_0)' J_{11}(\hat{\theta}_n - \theta_0) + o_{\mathbb{P}}(1).$$

Ainsi, en prenant l'espérance et au vu de (1.12), on montre que

$$\begin{aligned} \mathbb{E}_X n(\hat{\theta}_n - \theta_0)' J_{11}(\hat{\theta}_n - \theta_0) &= \text{Tr} \left\{ J_{11} n \mathbb{E}_X (\hat{\theta}_n - \theta_0)(\hat{\theta}_n - \theta_0)' \right\} \\ &\rightarrow \text{Tr}(I_{11}J_{11}^{-1}), \end{aligned}$$

et on en déduit que $a_1 = \text{Tr}(I_{11}J_{11}^{-1})/2 + o(1)$.

Le second résultat repose également sur un développement à l'ordre 2 du contraste. En prenant l'espérance et en supposant que $\Delta(\cdot)$ est intégrale (on pourra donc intervertir intégrale et dérivation), on a

$$\begin{aligned} \mathbb{E}_X \Delta(\hat{\theta}_n) &= \mathbb{E}_X \Delta(\theta_0) + (\hat{\theta}_n - \theta_0)' \frac{\partial \mathbb{E}_X \Delta(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} + \frac{1}{2} \mathbb{E}_X (\hat{\theta}_n - \theta_0)' \frac{\partial^2 \Delta(\theta)}{\partial \theta \partial \theta'} \Big|_{\theta=\theta_0} (\hat{\theta}_n - \theta_0) + o_{\mathbb{P}}(1) \\ &= \mathbb{E}_X \Delta(\theta_0) + \frac{n}{2} \mathbb{E}_X (\hat{\theta}_n - \theta_0)' J_{11}(\hat{\theta}_n - \theta_0) + o_{\mathbb{P}}(1) = \mathbb{E}_X \Delta(\theta_0) + \frac{1}{2} \text{Tr}(I_{11}J_{11}^{-1}) + o(1). \end{aligned}$$

On en déduit

$$-2\mathbb{E}_Y \log L_n(\hat{\theta}_n) = \mathbb{E}_X \Delta(\hat{\theta}_n) = \mathbb{E}_X \Delta(\theta_0) + \frac{1}{2} \text{Tr}(I_{11}J_{11}^{-1}) + o(1),$$

ce qui montre que a_2 est équivalent à a_1 . Cela complète la preuve. \square

Ainsi, au vu de la proposition 8, dans le cas des modèles VARMA faibles, le critère AIC prend la forme suivante

$$AIC_M := n \log \det \hat{\Sigma}_e + \text{Tr} \left(\hat{I}_{11,n} \hat{J}_{11,n}^{-1} \right) = -2 \log L_n(\hat{\theta}_n) + \text{Tr} \left(\hat{I}_{11,n} \hat{J}_{11,n}^{-1} \right). \quad (2.9)$$

La sélection des ordres du modèle (4.13) est obtenue en minimisant (2.9) pour les modèles candidats.

Remarque 2.2. Dans le cas de modèles VARMA forts, nous avons $\text{Tr} (I_{11} J_{11}^{-1}) = 2k_1$. Par conséquent, en vu de la proposition 8, nous obtenons également que $a_1 = a_2 = k_1 = \dim(\theta_0)$ et nous retrouvons le résultat de [Fin93]. Dans ce cas, l'équation (2.9) devient

$$AIC_M = -2 \log L_n(\hat{\theta}_n) + 2k_1 = AIC. \quad (2.10)$$

Remarque 2.3. Quand n tend vers $+\infty$, le critère AIC_{CM} (voir équation (2.7)) est équivalent au critère AIC_M (voir (2.9)). Par conséquent, pour une famille donnée de modèles candidats, nous préférons minimiser le critère modifié (2.7) pour estimer les ordres \hat{p} et \hat{q} du modèle (4.13).

2.3 Critère de sélection convergent des ordres

Un autre critère très utilisé est le SBIC (Schwarz Bayesian information criterion) introduit par [Sch78] et défini de manière suivante

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

Le terme de pénalisation tendant vers 0 moins rapidement, asymptotiquement le critère SBIC (encore appelé SIC ou BIC) sélectionnera des ordres inférieurs ou égaux à ceux sélectionnés par le critère AIC ou le critère AICc. Plus précisément, puisque $\ln(n) > 2$ dès que $n \geq 8$, SBIC pénalise davantage (le terme de log –vraisemblance quand n augmente) et sélectionne donc des modèles plus simples. La différence entre le critère AIC et le BIC se situe, tout simplement, dans la fonction de pénalisation. Comme pour le AIC, plusieurs modifications de ce critère existent.

Un troisième critère, également très utilisé, est le HQ (Hannan et Quinn) introduit par [HQ79] pour les processus autorégressifs (AR) et défini de manière suivante

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

Ce critère apparaît comme un compromis entre AIC et BIC. En effet, le critère HQ adoucit quelque peu la sévérité de la fonction de pénalisation du BIC relativement à l'augmentation de la taille de l'échantillon n tout en maintenant une forte convergence dans l'identification de l'ordre réel du modèle. Mais ces améliorations ont pour inconvénient de choisir le paramètre $c > 1$. Dans [A11], nous avons proposé des modifications des critères BIC et HQ adaptés aux cas des modèles VARMA faibles.

2.3.1 Critères BIC et HQ modifiés pour VARMA faibles

Comme pour les critères AIC et AICc, l'estimation du nombre de paramètres peut être très différente de celle obtenue dans le cadre standard (à savoir k_0) lorsque l'hypothèse d'indépendance sur le bruit est relâchée. Cela remet donc en cause les fondements théoriques des critères standards BIC et HQ.

Au vu de la proposition 8, nous proposons les critères BIC et HQ modifiés suivants pour des modèles VARMA faibles

$$\begin{aligned} \text{BIC}_M &:= -2 \log \tilde{\mathcal{L}}_n(\hat{\theta}_n) + \frac{1}{2} \text{Tr} \left(\hat{h}_{11,n} \hat{J}_{11,n}^{-1} \right) \log(n), \\ \text{HQ}_M &:= -2 \log \tilde{\mathcal{L}}_n(\hat{\theta}_n) + c \text{Tr} \left(\hat{h}_{11,n} \hat{J}_{11,n}^{-1} \right) \log \log(n), \quad c > 1; \end{aligned}$$

ou de manière équivalente comme

$$\text{BIC}_M := n \log \det \hat{\Sigma}_e + \frac{1}{2} \text{Tr} \left(\hat{h}_{11,n} \hat{J}_{11,n}^{-1} \right) \log(n), \quad (2.11)$$

$$\text{HQ}_M := n \log \det \hat{\Sigma}_e + c \text{Tr} \left(\hat{h}_{11,n} \hat{J}_{11,n}^{-1} \right) \log \log(n), \quad c > 1. \quad (2.12)$$

Ainsi, la sélection des ordres \hat{p} et \hat{q} du modèle (4.13) est obtenue en minimisant les critères d'information (2.11) ou (2.12).

Remarque 2.4. Au vu de la remarque 2.1, dans le cas des modèles VARMA forts, les critères (2.11) et (2.12) prennent les formes plus conventionnelles suivantes

$$\begin{aligned} \text{BIC}_M &= n \log \det \hat{\Sigma}_e + k_0 \log(n) = \text{BIC}, \\ \text{HQ}_M &= n \log \det \hat{\Sigma}_e + 2ck_0 \log \log(n) = \text{HQ}, \quad c > 1. \end{aligned}$$

Compte tenu des arguments ci-dessus, il serait important pour le praticien de disposer de résultats assez généraux pour l'utilisation des critères BIC et HQ ou de leurs modifications BIC_M et HQ_M . C'est ainsi que nous proposons les nouveaux critères suivants BIC_c et HQ_c (qui sont, respectivement, des extensions des critères BIC et HQ standard) dans lesquels k_0 (respectivement $2ck_0$) est remplacé par une fonction c_{k_0} qui croît avec k_0 . Les critères proposés sont

$$\text{BIC}_c := n \log \det \hat{\Sigma}_e + c_{k_0} \log(n), \quad (2.13)$$

$$\text{HQ}_c := n \log \det \hat{\Sigma}_e + c_{k_0} \log \log(n). \quad (2.14)$$

De plus, pour tout modèle candidat de paramètre θ_k avec $\dim(\theta_k) = k$ et au vu de (2.13) et (2.14), nous notons

$$\begin{aligned} \text{BIC}_c(k) &= n \log \det \hat{\Sigma}_e(k) + c_k \log(n), \\ \text{HQ}_c(k) &= n \log \det \hat{\Sigma}_e(k) + c_k \log \log(n), \end{aligned}$$

où $\hat{\Sigma}_e(k) = \Sigma_e(\hat{\theta}_{n,k})$ et $\hat{\theta}_{n,k}$ est l'estimateur du QMV du paramètre θ_k . Dans [A11], nous avons prouvé la consistance de ces nouveaux critères.

2.3.2 Convergence des critères BIC_c et HQ_c

Posons $k_1 = d^2(p_0 + q_0)$ le nombre de paramètres contenus dans le modèle (1.2) sans aucune contrainte. Nous définissons un modèle sur-paramétré comme celui qui a plus de paramètres que le vrai modèle. Plus précisément, les ordres de ce modèle vérifient $p' \geq p_0$ et $q' \geq q_0$, avec $p' > p_0$ ou $q' > q_0$ et nous notons $k'_1 = d^2(p' + q')$ le nombre de paramètres sans contrainte du modèle sur-paramétré.

Les deux propositions suivantes montrent que la probabilité asymptotique que le critère BIC_c (respectivement le critère HQ_c) surestime les ordres du modèle VARMA est nulle.

Proposition 9. *Soit (X_t) une solution causale de l'équation (4.13) satisfaisant les hypothèses (H1)–(H7). Le critère BIC_c surestime les ordres si $BIC_c(k'_1) < BIC_c(k_1)$. La probabilité asymptotique que le BIC_c choisisse des ordres plus grands que les vrais ordres est donnée par*

$$\lim_{n \rightarrow \infty} \mathbb{P} \{ BIC_c(k'_1) < BIC_c(k_1) \} = 0.$$

Proposition 10. *Sous les hypothèses de la proposition 9, la probabilité asymptotique que le HQ_c choisisse le modèle sur-paramétré est donnée par*

$$\lim_{n \rightarrow \infty} \mathbb{P} \{ HQ_c(k'_1) < HQ_c(k_1) \} = 0.$$

De manière analogue, nous définissons un modèle sous-paramétré (avec des ordres $p'' < p_0$ ou $q'' < q_0$) comme étant celui qui a moins de paramètres que le vrai modèle dont les vrais ordres sont p_0 et q_0 . Nous posons $k''_1 = d^2(p'' + q'')$ le nombre de paramètres sans contraintes du modèle sous-paramétré.

Les deux propositions suivantes montrent également que la probabilité asymptotique que le critère BIC_c (respectivement le critère HQ_c) sous-estime les ordres du modèle VARMA est nulle.

Proposition 11. *Soit (X_t) une solution causale de l'équation (4.13) satisfaisant les hypothèses (H1)–(H7). Le critère BIC_c sous-estime les ordres si $BIC_c(k''_1) \leq BIC_c(k_1)$. La probabilité asymptotique que le BIC_c choisisse des ordres plus petits que les vrais ordres est donnée par*

$$\lim_{n \rightarrow \infty} \mathbb{P} \{ BIC_c(k''_1) \leq BIC_c(k_1) \} = 0.$$

Proposition 12. *Sous les hypothèses de la proposition 11, la probabilité asymptotique que le HQ_c choisisse le modèle sous-paramétré est donnée par*

$$\lim_{n \rightarrow \infty} \mathbb{P} \{ HQ_c(k''_1) \leq HQ_c(k_1) \} = 0.$$

Les propositions 9 et 11 (respectivement les propositions 10 et 12) montrent que

$$(\hat{p}, \hat{q}) = \min_{p, q \in \Theta_{p, q}} BIC_c(k) \quad (\text{respectivement } (\hat{p}, \hat{q}) = \min_{p, q \in \Theta_{p, q}} HQ_c(k))$$

convergent vers les vrais ordres (p_0, q_0) . Contrairement au AIC_M où la probabilité asymptotique qu'il surestime les ordres du modèle VARMA est strictement positive. Pour des raisons de simplifications, posons $k'_1 = k_1 + \ell$.

Proposition 13. *Le critère AIC_M surestime les ordres si $AIC_M(k'_1) < AIC_M(k_1)$. Sous les hypothèses de la proposition 9, la probabilité que le AIC_M choisisse le modèle sur-paramétré est donnée par*

$$\mathbb{P} \{ AIC_M(k_1 + \ell) < AIC_M(k_1) \} \rightarrow \mathbb{P} \left\{ \chi_\ell^2 > \frac{2\ell + c_\ell}{2} \right\} > 0.$$

Cela montre la non convergence des critères AIC_M et AIC_{cM} . Par contre, la probabilité que ces critères sous-estiment les ordres est nulle

$$\lim_{n \rightarrow \infty} \mathbb{P} \{ AIC_M(k''_1) \leq AIC_M(k_1) \} = 0.$$

Remarque 2.5. Dans le cas des modèles VARMA forts, on a $c_\ell = 2\ell$ (voir Remarque 1.3). Dans ce cas, la probabilité que le AIC_M sélectionne le modèle sur-paramétré prend une forme plus conventionnelle

$$\mathbb{P}\{AIC_M(k_1 + \ell) < AIC_M(k_1)\} \rightarrow \mathbb{P}\{\chi_\ell^2 > 2\ell\} > 0.$$

2.4 Illustrations numériques

Dans cette section, nous présentons quelques résultats de simulations afin d'évaluer les performances des critères proposés pour des modèles VARMA faibles. Nous simulons $N = 1000$ trajectoires indépendantes, de tailles différentes, du modèle (1.42) (VARMA(1, 1) bivarié défini dans le chapitre 1) avec $\theta_0 = (a_{22,1}, b_{21,1}, b_{22,1})' = (0.225, 0.313, 0.750)'$ et dont le processus des innovations $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t})'$ suit un modèle GARCH(1, 1) bivarié, proposé par [Jea98], et défini par

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} = \begin{pmatrix} h_{11,t} & 0 \\ 0 & h_{22,t} \end{pmatrix} \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix}, \text{ où } \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix} \sim \text{IID } \mathcal{N}(0, \Sigma_\eta), \quad (2.15)$$

avec $\Sigma_\eta = \begin{pmatrix} 1.0 & 0.8 \\ 0.8 & 1.0 \end{pmatrix}$ et

$$\begin{pmatrix} h_{11,t}^2 \\ h_{22,t}^2 \end{pmatrix} = \begin{pmatrix} 0.93 \\ 0.82 \end{pmatrix} + \begin{pmatrix} 0.35 & 0.10 \\ 0.20 & 0.25 \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1}^2 \\ \epsilon_{2,t-1}^2 \end{pmatrix} + \begin{pmatrix} 0.40 & 0.15 \\ 0.10 & 0.40 \end{pmatrix} \begin{pmatrix} h_{11,t-1}^2 \\ h_{22,t-1}^2 \end{pmatrix}.$$

Les modèles candidats utilisés sont des modèles VARMA(p, q) bivariés sous la forme échelon suivante

$$\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \sum_{i=1}^p \begin{pmatrix} 0 & 0 \\ 0 & a_{22,i} \end{pmatrix} \begin{pmatrix} X_{1,t-i} \\ X_{2,t-i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} - \sum_{j=1}^q \begin{pmatrix} 0 & 0 \\ b_{21,j} & b_{22,j} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-j} \\ \epsilon_{2,t-j} \end{pmatrix}, \quad (2.16)$$

où les ordres $(p, q) \in \{1, 2, 3\}^2$. Plus précisément, nous avons 9 modèles candidats : VARMA(1, 1), VARMA(2, 2), VARMA(2, 1), VARMA(1, 2), VARMA(1, 3), VARMA(3, 1), VARMA(3, 2), VARMA(2, 3) and VARMA(3, 3). Pour chacune des N réplifications, nous estimons le paramètre

$$\theta_{p,q} = (a_{22,1}, \dots, a_{22,p}, b_{21,1}, \dots, b_{21,q}, b_{22,1}, \dots, b_{22,q})'$$

par la méthode du QMV définie dans le chapitre 1. Les versions standard et modifié des critères AIC, AICc, BIC et HQ sont utilisées afin de sélectionner les ordres des 9 modèles candidats. Les tables 2.1, 3.1 et 3.2 indiquent pour chaque modèle candidat le pourcentage de fois où il a été sélectionné parmi les N réplifications. Les ordres sélectionnés sont mis en gras dans les tables 2.1, 3.1 et 3.2.

Comme attendu, la table 2.1 montre qu'aucun critère AIC et AICc standard ne fonctionne pour ce modèle particulier (1.42)–(2.15) dans le sens où ils sélectionnent un modèle VARMA(3, 3) pour toutes les valeurs de n alors que le vrai modèle est un VARMA(1, 1). Par contre, le modèle VARMA(1, 1) faible est bien sélectionné par les versions modifiées, à l'exception du critère AIC_M quand $n \leq 500$ (qui sélectionne le modèle VARMA(3, 3) faible). Ceci est en concordance avec les résultats de [HT89, HT93] qui ont montré que pour des modèles VAR forts, le AIC est un estimateur biaisé de l'information de Kullback-Leibler en petit échantillon, et que ce biais tend à entraîner une sélection de modèle sur-paramétré contrairement au critère AICc. La même conclusion reste valable

pour les critères AIC_M et AIC_{cM} .

Dans la table 3.1, le modèle VARMA(1, 1) faible est fréquemment sélectionné par les versions standard et modifié du critère BIC, à l'exception du critère BIC standard qui sélectionne le modèle VARMA(3, 3) quand $n = 10000$. Comme attendu pour le BIC standard, quand n augmente, il n'est clairement pas performant et le pourcentage de l'ordre sélectionné décroît. Par contre, pour la version modifiée, le pourcentage de l'ordre sélectionné s'améliore quand n augmente, ce qui est en concordance avec la consistance du critère modifié.

Au vu des résultats de la table 3.2, on peut noter que : (i) bien que les deux versions de HQ sélectionnent le vrai modèle (à savoir le VARMA(1, 1)) quand $n \leq 500$, le critère HQ standard n'est clairement pas performant ; (ii) quand $n > 500$, le critère HQ standard surestime les ordres, par contre la version modifiée reste performante et sélectionne le vrai modèle VARMA(1, 1) simulé ; (iii) comme pour la version modifiée du BIC, le pourcentage de l'ordre sélectionné augmente avec n , ce qui illustre également la consistance du critère modifié.

TABLE 2.1 – Pourcentage de sélection des ordres par les versions du critère AIC standard et modifié.

Taille n		200							
Ordres	(1, 1)	(2, 2)	(2, 1)	(1, 2)	(3, 3)	(3, 2)	(3, 1)	(2, 3)	(1, 3)
AIC	21.9	5.7	2.9	4.3	39.8	4.5	3.5	12.5	4.9
AICc	23.6	5.9	3.2	4.7	37.5	4.2	3.6	12.0	5.3
AIC_M	25.4	4.2	2.4	3.9	35.8	5.5	3.9	12.7	6.2
AIC_{cM}	53.1	2.8	2.5	8.9	15.1	1.1	1.3	7.9	7.3
Taille n		500							
Ordres	(1, 1)	(2, 2)	(2, 1)	(1, 2)	(3, 3)	(3, 2)	(3, 1)	(2, 3)	(1, 3)
AIC	18.3	6.0	2.8	1.8	41.6	5.5	3.4	11.4	9.2
AICc	19.2	6.0	2.9	2.0	40.3	5.5	3.5	11.2	9.4
AIC_M	29.2	5.4	3.1	4.2	32.0	5.4	3.5	9.8	7.4
AIC_{cM}	48.4	3.3	3.0	7.8	17.5	1.4	1.6	6.6	10.4
Taille n		1,000							
Ordres	(1, 1)	(2, 2)	(2, 1)	(1, 2)	(3, 3)	(3, 2)	(3, 1)	(2, 3)	(1, 3)
AIC	11.5	5.2	2.5	3.3	50.0	5.6	2.9	11.8	7.2
AICc	11.8	5.3	2.6	3.2	49.7	5.6	3.0	11.8	7.0
AIC_M	31.2	6.8	4.5	4.0	29.4	4.9	3.9	9.0	6.3
AIC_{cM}	43.0	5.3	2.6	7.1	20.9	1.9	2.4	8.1	8.7
Taille n		2,000							
Ordres	(1, 1)	(2, 2)	(2, 1)	(1, 2)	(3, 3)	(3, 2)	(3, 1)	(2, 3)	(1, 3)
AIC	8.5	5.5	0.4	2.2	57.6	7.0	2.0	11.4	5.4
AICc	8.5	5.5	0.4	2.2	57.4	7.0	2.0	11.4	5.6
AIC_M	34.6	7.6	4.0	4.6	26.8	4.2	5.2	6.4	6.6
AIC_{cM}	41.0	5.3	2.6	7.1	22.9	1.9	2.4	8.1	8.7
Taille n		5,000							
Ordres	(1, 1)	(2, 2)	(2, 1)	(1, 2)	(3, 3)	(3, 2)	(3, 1)	(2, 3)	(1, 3)
AIC	5.7	3.7	0.6	1.3	66.6	4.6	1.7	10.9	4.9
AICc	5.7	3.7	0.6	1.3	66.6	4.6	1.7	10.9	4.9
AIC_M	40.2	6.6	5.8	3.8	23.0	3.8	3.7	5.3	7.8
AIC_{cM}	40.4	6.8	5.8	3.8	22.8	3.8	3.7	5.3	7.6

TABLE 2.2 – Pourcentage de sélection des ordres par les versions du critère BIC standard et modifié.

Taille n		200								
Ordres	(1, 1)	(2, 2)	(2, 1)	(1, 2)	(3, 3)	(3, 2)	(3, 1)	(2, 3)	(1, 3)	
BIC	61.6	3.6	5.5	3.5	14.7	1.7	3.2	5.1	1.1	
BIC _M	50.8	4.0	4.8	5.2	14.6	4.1	4.2	6.9	5.4	
Taille n		500								
Ordres	(1, 1)	(2, 2)	(2, 1)	(1, 2)	(3, 3)	(3, 2)	(3, 1)	(2, 3)	(1, 3)	
BIC	63.9	2.8	7.2	4.7	9.3	1.5	3.4	3.9	3.3	
BIC _M	68.8	2.5	6.8	5.2	6.4	2.0	2.9	2.8	2.6	
Taille n		1,000								
Ordres	(1, 1)	(2, 2)	(2, 1)	(1, 2)	(3, 3)	(3, 2)	(3, 1)	(2, 3)	(1, 3)	
BIC	54.5	3.8	7.4	4.2	13.8	2.4	5.7	4.8	3.4	
BIC _M	74.4	2.7	5.2	3.7	5.2	1.8	3.0	2.2	1.8	
Taille n		2,000								
Ordres	(1, 1)	(2, 2)	(2, 1)	(1, 2)	(3, 3)	(3, 2)	(3, 1)	(2, 3)	(1, 3)	
BIC	50.5	4.1	5.0	5.2	16.3	2.6	6.3	5.9	4.1	
BIC _M	82.7	1.5	3.2	3.3	2.8	0.8	2.5	1.6	1.6	
Taille n		5,000								
Ordres	(1, 1)	(2, 2)	(2, 1)	(1, 2)	(3, 3)	(3, 2)	(3, 1)	(2, 3)	(1, 3)	
BIC	38.1	5.6	6.7	3.8	26.1	3.0	5.9	4.6	6.2	
BIC _M	86.1	1.3	3.7	2.2	1.2	0.7	2.3	1.0	1.5	
Taille n		10,000								
Ordres	(1, 1)	(2, 2)	(2, 1)	(1, 2)	(3, 3)	(3, 2)	(3, 1)	(2, 3)	(1, 3)	
BIC	28.3	5.2	6.2	4.5	34.7	3.4	4.5	6.5	6.7	
BIC _M	88.4	0.8	3.8	2.1	1.7	0.7	1.5	0.7	0.3	

TABLE 2.3 – Pourcentage de sélection des ordres par les versions du critère HQ standard et modifié.

Taille n		200								
Ordres	(1, 1)	(2, 2)	(2, 1)	(1, 2)	(3, 3)	(3, 2)	(3, 1)	(2, 3)	(1, 3)	
HQ	40.4	5.5	4.3	4.5	26.9	2.6	4.1	8.5	3.2	
HQ _M	38.6	5.0	3.6	5.0	23.2	4.9	4.2	9.6	5.9	
Taille n		500								
Ordres	(1, 1)	(2, 2)	(2, 1)	(1, 2)	(3, 3)	(3, 2)	(3, 1)	(2, 3)	(1, 3)	
HQ	39.3	5.5	6.5	4.2	22.1	3.2	3.9	7.9	7.4	
HQ _M	55.2	3.7	5.4	5.2	14.4	3.6	3.1	4.7	4.7	
Taille n		1,000								
Ordres	(1, 1)	(2, 2)	(2, 1)	(1, 2)	(3, 3)	(3, 2)	(3, 1)	(2, 3)	(1, 3)	
HQ	30.2	5.0	5.8	3.6	31.4	4.4	6.3	7.3	6.0	
HQ _M	60.0	4.5	5.2	4.2	11.9	2.6	3.6	4.4	3.6	
Taille n		2,000								
Ordres	(1, 1)	(2, 2)	(2, 1)	(1, 2)	(3, 3)	(3, 2)	(3, 1)	(2, 3)	(1, 3)	
HQ	25.9	6.3	3.4	4.3	34.4	4.9	5.6	9.5	7.5	
HQ _M	68.7	3.6	5.3	4.6	7.2	1.6	3.4	2.8	2.8	
Taille n		5,000								
Ordres	(1, 1)	(2, 2)	(2, 1)	(1, 2)	(3, 3)	(3, 2)	(3, 1)	(2, 3)	(1, 3)	
HQ	16.3	6.4	3.3	3.1	47.6	4.6	3.5	8.3	6.9	
HQ _M	75.1	2.5	5.9	3.3	3.8	0.8	3.5	2.4	2.7	
Length n		10,000								
Order	(1, 1)	(2, 2)	(2, 1)	(1, 2)	(3, 3)	(3, 2)	(3, 1)	(2, 3)	(1, 3)	
HQ	11.3	4.3	2.1	2.9	57.8	4.3	1.8	9.1	6.4	
HQ _M	77.2	1.7	6.9	3.4	4.3	1.0	1.8	1.8	1.9	

2.5 Conclusion

Nous avons montré que l'utilisation des critères d'information standards n'est souvent pas justifiée quand l'hypothèse d'indépendance sur le bruit est relâchée. Au vu des illustrations numériques, l'utilisation de ces critères standards conduit à sélectionner des modèles VARMA faibles avec beaucoup de paramètres. Ces critères ont donc besoin d'être adaptés afin de tenir compte des possibles dépendances du terme d'erreur. C'est ainsi que nous avons proposé des critères modifiés et avons établi leurs propriétés asymptotiques.

Bien que le but de notre travail n'était pas d'établir une comparaison entre les différents critères, nous faisons néanmoins un commentaire sur les critères convergents. Dans la pratique, on ne connaît pas les vrais ordres; de plus, le fait qu'il soit possible que le vrai modèle ne soit pas contenu dans la classe de modèles considérés rend (ainsi) les arguments basés sur la seule convergence asymptotique insuffisants pour classer les critères les uns par rapport aux autres. Surtout dans certaines situations, le fait que le AIC_M ou AIC_{CM} choisisse systématiquement des modèles d'ordre plus élevé que ceux choisis par les critères BIC_M ou HQ_M peut se transformer en avantage dans les cas où le vrai modèle possède une dimension paramétrique infinie (par exemple).

Chapitre 3

Contribution à la validation des modèles ARMA faibles multivariés

Après les phases d'identification et d'estimation du processus VARMA, 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 p_0 et q_0 du modèle. Ce choix est important pour la précision des prévisions linéaires et pour une bonne interprétation du modèle. 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.

Pour la validation de ces modèles VARMA nous présentons deux étapes possibles :

- a) les tests sur les paramètres du modèle VARMA,
- b) les tests sur les résidus du modèle VARMA.

Dans ce chapitre, nous considérons également la matrice Σ_e comme un paramètre de nuisance et afin d'alléger nos notations, nous posons $\theta = \vartheta^{(1)}$.

Les publications liées à ce chapitre sont [A1, A3, A9, A12].

3.1 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 et en particulier $A_{0p} = 0$ ou $B_{0q} = 0$ pour la validation des ordres. Afin de tester s_0 contraintes linéaires sur les coefficients, nous définissons notre hypothèse nulle

$$H_0 : R_0 \theta_0 = r_0,$$

où R_0 est une matrice $s_0 \times k_0$ connue de rang s_0 et r_0 est aussi un vecteur connu de dimension s_0 . Pour effectuer ce test H_0 , il existe diverses approches asymptotiques mais dans [A9], nous nous sommes intéressés à la procédure de Wald (notée W dans la suite), à celle du multiplicateur de Lagrange (en anglais LM) encore appelée du score ou de Rao-score et à la méthode du rapport de vraisemblance (notée LR pour likelihood ratio). Pour calculer nos différentes statistiques de test,

nous considérons la matrice $\hat{\Omega}_{11} = \hat{J}_{11}^{-1} \hat{I}_{11}^{SP} \hat{J}_{11}^{-1}$, où \hat{J}_{11} et \hat{I}_{11}^{SP} sont des estimateurs consistants de J_{11} et I_{11} définis dans le chapitre 1. Comme il n'y a pas de risque de confusion dans ce chapitre, pour la suite nous posons $\Omega = \Omega_{11}$ et $\hat{\Omega} = \hat{\Omega}_{11}$.

3.1.1 Statistique de Wald

Le principe est d'accepter l'hypothèse nulle, si l'estimateur non contraint $\hat{\theta}_n$ de θ_0 est suffisamment proche de zéro. Ainsi de (1.12), nous déduisons que

$$\sqrt{n} \left(R_0 \hat{\theta}_n - r_0 \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, R_0 \Omega R_0' := R_0 \left(J_{11}^{-1} I_{11} J_{11}^{-1} \right) R_0' \right),$$

quand $n \rightarrow +\infty$. Sous les hypothèses du théorème 2 et l'hypothèse que I_{11} est inversible, nous obtenons la statistique de Wald modifiée définie par

$$\mathbf{W}_n = n(R_0 \hat{\theta}_n - r_0)' (R_0 \hat{\Omega} R_0')^{-1} (R_0 \hat{\theta}_n - r_0).$$

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

3.1.2 Statistique du LM

L'idée de ce test consiste à accepter l'hypothèse nulle, si le score contraint est proche de zéro. Soit $\hat{\theta}_n^c$ l'estimateur du QMV contraint sous H_0 . Nous définissons le Lagrangien

$$\mathcal{L}(\vartheta, \lambda) = \tilde{\ell}_n(\theta) - \lambda'(R_0 \theta - r_0),$$

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

$$\frac{\partial \tilde{\ell}_n}{\partial \theta}(\hat{\theta}_n^c) = R_0' \hat{\lambda}, \quad R_0 \hat{\theta}_n^c = r_0.$$

En effectuant le développement limité de Taylor sous H_0 , et au vu de (1.17), nous avons

$$0 = \sqrt{n} \frac{\partial \tilde{\ell}_n(\hat{\theta}_n)}{\partial \theta} = \sqrt{n} \frac{\partial \tilde{\ell}_n(\hat{\theta}_n^c)}{\partial \theta} - J_{11} \sqrt{n} (\hat{\theta}_n - \hat{\theta}_n^c) + o_{\mathbb{P}}(1).$$

Nous déduisons que

$$\sqrt{n}(R_0 \hat{\theta}_n - r_0) = R_0 \sqrt{n}(\hat{\theta}_n - \hat{\theta}_n^c) = R_0 J_{11}^{-1} \sqrt{n} \frac{\partial \tilde{\ell}_n(\hat{\theta}_n^c)}{\partial \theta} + o_{\mathbb{P}}(1) = R_0 J_{11}^{-1} R_0' \sqrt{n} \hat{\lambda} + o_{\mathbb{P}}(1).$$

Il en résulte, sous l'hypothèse H_0 et les hypothèses du théorème 2, la normalité asymptotique du vecteur multiplicateur de Lagrange

$$\sqrt{n} \hat{\lambda} \xrightarrow{\mathcal{L}} \mathcal{N} \left\{ 0, (R_0 J_{11}^{-1} R_0')^{-1} R_0 \Omega R_0' (R_0 J_{11}^{-1} R_0')^{-1} \right\}. \quad (3.1)$$

On obtient ainsi la statistique du LM modifiée définie par

$$\begin{aligned}\mathbf{LM}_n &= n\hat{\lambda}' \left\{ (R_0 \hat{J}_{11}^{-1} R_0')^{-1} R_0 \hat{\Omega} R_0' (R_0 \hat{J}_{11}^{-1} R_0')^{-1} \right\}^{-1} \hat{\lambda} \\ &= n \frac{\partial \tilde{\ell}_n}{\partial \theta'}(\hat{\theta}_n^c) \hat{J}_{11}^{-1} R_0' \left(R_0 \hat{\Omega} R_0' \right)^{-1} R_0 \hat{J}_{11}^{-1} \frac{\partial \tilde{\ell}_n}{\partial \theta}(\hat{\theta}_n^c).\end{aligned}$$

Notons que dans le cas de modèles VARMA forts $\hat{\Omega} = 2\hat{J}_{11}^{-1}$, ce qui donne à la statistique du LM une forme plus conventionnelle $\mathbf{LM}_n^* = (n/2)\hat{\lambda}' R_0 \hat{J}_{11}^{-1} R_0' \hat{\lambda}$. De la convergence (3.1) du vecteur multiplicateur de Lagrange, nous obtenons 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.

3.1.3 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

$$\sqrt{n} \left(\hat{\theta}_n - \hat{\theta}_n^c \right) = -\sqrt{n} J_{11}^{-1} R_0' \hat{\lambda} + o_{\mathbb{P}}(1),$$

et que la statistique du LR satisfait

$$\mathbf{LR}_n := 2 \left\{ \log \tilde{L}_n(\hat{\theta}_n) - \log \tilde{L}_n(\hat{\theta}_n^c) \right\} = \frac{n}{2} (\hat{\theta}_n - \hat{\theta}_n^c)' J_{11} (\hat{\theta}_n - \hat{\theta}_n^c) + o_{\mathbb{P}}(1) = \mathbf{LM}_n^* + o_{\mathbb{P}}(1).$$

En utilisant les calculs précédents et des résultats standards sur les formes quadratiques de vecteurs (voir [vdV98], Lemme 17.1), 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

$$\Sigma_{\mathbf{LR}} = J_{11}^{-1/2} S_{\mathbf{LR}} J_{11}^{-1/2}, \quad S_{\mathbf{LR}} = \frac{1}{2} R_0' (R_0 J_{11}^{-1} R_0')^{-1} R_0 \Omega R_0' (R_0 J_{11}^{-1} R_0')^{-1} R_0.$$

Notons que dans le cas de modèles VARMA forts, *i.e.* quand $\Omega = 2J_{11}^{-1}$, la matrice $\Sigma_{\mathbf{LR}} = J_{11}^{-1/2} R_0' (R_0 J_{11}^{-1} R_0')^{-1} R_0 J_{11}^{-1/2}$ se comporte comme une matrice de projection. Ses valeurs propres sont donc égales soit à 0 soit à 1, et le nombre de valeurs propres égales à 1 est donné par

$$\text{Tr} \left(J_{11}^{-1/2} R_0' (R_0 J_{11}^{-1} R_0')^{-1} R_0 J_{11}^{-1/2} \right) = \text{Tr}(I_{s_0}) = s_0.$$

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 [lmh61]. Une alternative est d'utiliser une statistique transformée

$$\frac{n}{2} (\hat{\theta}_n - \hat{\theta}_n^c)' \hat{J}_{11} \hat{S}_{\mathbf{LR}}^- \hat{J}_{11} (\hat{\theta}_n - \hat{\theta}_n^c),$$

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}_{11} et $\hat{S}_{\mathbf{LR}}$ sont des estimateurs faiblement convergents de J_{11} 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}_{LR} de S_{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}_{LR} , et notons \hat{P} la matrice orthogonale telle que $\hat{S}_{LR} = \hat{P}\hat{\Lambda}\hat{P}'$, nous posons

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

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

3.2 Illustrations numériques

Nous présentons quelques résultats de simulations afin d'évaluer les performances des tests proposés. Nous simulons $N = 1000$ trajectoires indépendantes, de tailles $n = 500$ et $n = 2000$, du modèle (1.42) (défini dans le chapitre 1) avec $\theta_0 = (a_{22,1}, b_{21,1}, b_{22,1})' = (0.95, 2, 0)'$ et dont le processus des innovations $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t})'$ est dans un premier temps le bruit fort (1.43), et puis le bruit blanc faible (1.44), tous deux définis dans le chapitre 1.

La table 3.1 donne les pourcentages de rejet de l'hypothèse nulle $H_0 : b_{22,1} = 0$ des différentes versions des tests de Wald, LM et du LR. Pour un niveau asymptotique $\alpha = 5\%$, les limites de significativité sont 3,6% et 6,4% avec une probabilité de 95%. Elles sont de 0,3% et 1,7% pour $\alpha = 1\%$, et enfin celles de 8,1% et 11,9% pour $\alpha = 10\%$. Les fréquences à l'extérieur de ces limites sont mises en caractères gras dans la table 3.1. Pour le modèle VARMA fort (1.42)–(1.43) (Modèle I), toutes les fréquences sont à l'intérieur des limites de significativité. Par contre, pour le modèle VARMA faible (1.42)–(1.44) (Modèle II), toutes les fréquences des tests standards sont résolument à l'extérieur des limites de significativité. En conclusion, l'erreur de première espèce est bien contrôlée par toutes les versions des tests dans le cas fort, mais seulement par les tests modifiés dans le cas faible.

Afin de comparer la puissance de ces tests, nous simulons $N = 1000$ trajectoires indépendantes, de taille $n = 500$, du même modèle (1.42) avec $\theta_0 = (a_{22,1}, b_{21,1}, b_{22,1})' = (0.95, 2, 0.05)'$ et dont le processus des innovations $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t})'$ est dans un premier temps le bruit fort (1.43), puis le bruit blanc faible (1.44), et nous testons l'hypothèse $H_0 : b_{22,1} = 0$. La table 3.2 montre que tous les tests ont une puissance similaire quand le bruit est fort (Modèle III). Les tests modifiés ont également une puissance similaire dans le cas faible (Modèle IV). Par contre, la puissance des tests standards est difficilement interprétable pour le modèle IV, car au vu des résultats de la table 3.1, les versions standard des tests ne contrôlent pas bien l'erreur de première espèce dans ce cas faible.

TABLE 3.1 – Pourcentages de rejet de l’hypothèse nulle, $H_0 : b_{22,1} = 0$, des versions standards et modifiées des tests.

Modèle	Taille n	Niveau	Tests standards			Tests modifiés		
			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 : Modèle VARMA(1, 1) fort (1.42)-(1.43) avec $\theta_0 = (0.95, 2, 0)$

II : Modèle VARMA(1, 1) faible (1.42)-(1.44) avec $\theta_0 = (0.95, 2, 0)$

TABLE 3.2 – Puissance des tests standards et modifiés : $H_0 : b_{22,1} = 0$.

Modèle	Taille n	Niveau	Tests standard			Tests modifiés		
			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 : Modèle VARMA(1, 1) fort (1.42)-(1.43) avec $\theta_0 = (0.95, 2, 0.05)$

IV : Modèle VARMA(1, 1) (1.42)-(1.44) avec $\theta_0 = (0.95, 2, 0.05)$

3.3 Tests sur les résidus du modèle

Ces tests ont pour objet de vérifier que les résidus (1.23) des modèles estimés sont bien des estimations de bruits blancs. Une invalidation de cette condition implique que le modèle est mal spécifié. Plus particulièrement, nous nous intéressons aux tests portmanteau, aussi appelés tests d'autocorrélation. Le test portmanteau a été introduit par [BP70] afin de mesurer la qualité d'ajustement d'un modèle ARMA fort univarié.

Les deux tests les plus utilisés, pour tester qu'une série est une réalisation d'un bruit blanc fort,

sont ceux de [BP70] et [LB78] définis respectivement par

$$Q_m^{\text{BP}} = n \sum_{h=1}^m \hat{\rho}^2(h) \text{ et } Q_m^{\text{LB}} = n(n+2) \sum_{h=1}^m \frac{\hat{\rho}^2(h)}{n-h}, \quad (3.2)$$

où $\hat{\rho}(\cdot)$ est la fonction d'autocorrélation résiduelle et m est un entier fixé. La statistique Q_m^{LB} a la même distribution asymptotique que Q_m^{BP} mais la statistique Q_m^{LB} est réputée avoir un meilleur comportement à distance finie. En s'inspirant des tests portmanteau univariés (3.2), [Chi74] et [Hos80] ont étendu ces tests au cas de modèles ARMA multivariés dont les versions multivariées des statistiques sont définies respectivement par

$$Q_m^{\text{C}} = n \sum_{h=1}^m \text{Tr} \left(\hat{f}'(h) \hat{f}^{-1}(0) \hat{f}(h) \hat{f}^{-1}(0) \right) \text{ et } Q_m^{\text{H}} = \sum_{h=1}^m \frac{n^2}{(n-h)} \text{Tr} \left(\hat{f}'(h) \hat{f}^{-1}(0) \hat{f}(h) \hat{f}^{-1}(0) \right), \quad (3.3)$$

où $\hat{f}(\cdot)$ (matrice carrée d'ordre d) est la fonction d'autocovariance résiduelle du processus multivarié. Sous les hypothèses d'un bruit blanc fort, pour un niveau asymptotique de risque α , la procédure standard consiste à rejeter l'hypothèse d'un modèle ARMA(p, q) ou VARMA(p, q) si les statistiques (3.2) ou (3.3) sont supérieures à un certain quantile d'ordre $1 - \alpha$ d'un chi-deux.

Dans le cas général de modèles VARMA dont les termes d'erreur sont non corrélés mais peuvent contenir des dépendances non linéaires, nous montrons dans [A1, A3] que la distribution asymptotique des statistiques (3.3) est celle d'une somme pondérée de chi-deux. Cette distribution peut être très différente de l'approximation chi-deux usuelle du cas fort. Dans [A12], nous proposons une méthode alternative à [A3] qui est basée sur une auto-normalisation des autocorrélations résiduelles. 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

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

Définissons les autocovariances empiriques et résiduelles du bruit par

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

Nous considérons les vecteurs des m (pour $m \geq 1$) premières autocovariances empiriques et résiduelles

$$\Gamma_m = (\{\text{vec} \Gamma_e(1)\}', \dots, \{\text{vec} \Gamma_e(m)\}')', \quad \hat{f}_m = \left(\{\text{vec} \hat{f}_e(1)\}', \dots, \{\text{vec} \hat{f}_e(m)\}' \right)'$$

Soient les matrices diagonales

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

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 $\sigma_e(i) = \sqrt{\mathbb{E}e_{it}^2}$ 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 $\Gamma_e(\ell) := \mathbb{E}e_t e_{t-\ell}' = 0$, $\forall \ell \neq 0$. Nous définissons aussi la matrice

$$\Phi_m = \mathbb{E} \left\{ \begin{pmatrix} e_{t-1} \\ \vdots \\ e_{t-m} \end{pmatrix} \otimes \frac{\partial e_t(\theta_0)}{\partial \theta'} \right\}. \quad (3.4)$$

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

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

3.3.1 Comportement asymptotique des statistiques du test portmanteau multivarié

Les travaux consacrés à la validation des modèles ARMA faibles sont pour l'heure limités au cas univarié. La distribution asymptotique des statistiques (3.2) a été établie dans [FRZ05] pour des modèles ARMA faibles univariés. Des tests portmanteau basés sur des méthodes de bootstrap, dont les distributions asymptotiques dépendent du DGP, ont été introduits par [ZL15] afin de tester l'adéquation de modèles ARMA faibles univariés. Dans le cas où le bruit blanc faible est directement observé, le comportement asymptotique des autocorrélations a été établi par [RT96]; [LNS01, LNS02] ont également abordé cette problématique en proposant une extension de la statistique Q_m^{BP} .

L'objectif de cette section est d'étudier le comportement asymptotique des statistiques du test portmanteau (3.3) dans le cadre de modèles VARMA dont les termes d'erreur sont non corrélés mais peuvent contenir des dépendances non linéaires. Dans un premier temps, nous étudions la distribution asymptotique jointe de l'estimateur, $\hat{\theta}_n$, du QMV et des autocovariances empiriques du bruit. Soit

$$\Gamma(\ell, \ell') = \sum_{h=-\infty}^{\infty} \mathbb{E} (\{e_{t-\ell} \otimes e_t\} \{e_{t-h-\ell'} \otimes e_{t-h}\}'), \quad \text{pour } (\ell, \ell') \neq (0, 0).$$

Nous énonçons le résultat suivant qui est une extension directe du résultat donné dans [FRZ05].

Théorème 7. *Supposons $p_0 > 0$ ou $q_0 > 0$. Sous les hypothèses (H1)–(H7), quand $n \rightarrow +\infty$, nous avons*

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

avec $\Sigma_{\Gamma_m} = \{\Gamma(\ell, \ell')\}_{1 \leq \ell, \ell' \leq m}$, $\Sigma'_{\hat{\theta}_n, \Gamma_m} = \lim_{n \rightarrow +\infty} \text{Cov}(\sqrt{n} J_{11}^{-1} \partial \ell_n(\theta_0, \Sigma_{e0}) / \partial \theta, \sqrt{n} \Gamma_m)$ et $\Sigma_{\hat{\theta}_n} = \Omega_{11} = J_{11}^{-1} h_{11} J_{11}^{-1}$.

La démonstration de ce théorème résulte des formules de dérivation matricielle standards, de

l'inégalité de covariance de [Dav68], et des théorèmes 1 et 2.

Ce résultat nous permet ensuite d'obtenir les distributions asymptotiques des autocovariances et autocorrélations résiduelles qui sont normalement distribuées avec une matrice de covariance différente du cas iid. Nous énonçons le résultat suivant qui donne le comportement asymptotique des autocovariances et autocorrélations résiduelles de modèles VARMA faibles.

Théorème 8. *Sous les hypothèses du théorème 7, nous avons*

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

$$\text{où } \Sigma_{\hat{\Gamma}_m} = \Sigma_{\gamma_m} + \Phi_m \Sigma_{\hat{\theta}_n} \Phi_m' + \Phi_m \Sigma_{\hat{\theta}_n, \gamma_m} + \Sigma_{\hat{\theta}_n, \gamma_m}' \Phi_m' \quad (3.5)$$

$$\Sigma_{\hat{\rho}_m} = \{I_m \otimes (S_e \otimes S_e)^{-1}\} \Sigma_{\hat{\Gamma}_m} \{I_m \otimes (S_e \otimes S_e)^{-1}\} \quad (3.6)$$

et Φ_m est donné par (3.4).

Nous donnons un schéma de la démonstration de ce résultat.

Démonstration. Pour la démonstration de ce théorème, nous utilisons les autocovariances empiriques du bruit précédemment définies et le théorème 7. Le résultat (3.5) découle de la relation

$$\hat{\Gamma}_m := \left(\left\{ \text{vec} \hat{\Gamma}_e(1) \right\}', \dots, \left\{ \text{vec} \hat{\Gamma}_e(m) \right\}' \right)' = \Gamma_m + \Phi_m (\hat{\theta}_n - \theta_0) + \mathcal{O}_{\mathbb{P}} \left(\frac{1}{n} \right), \quad (3.7)$$

que nous montrons en utilisant un développement limité de Taylor. Finalement nous obtenons le résultat (3.6) en calculant la variance asymptotique $\text{Var}(\sqrt{n}\hat{\rho}_m)$ avec

$$\hat{\rho}_m = \{I_m \otimes (S_e \otimes S_e)^{-1}\} \hat{\Gamma}_m + \mathcal{O}_{\mathbb{P}} \left(\frac{1}{n} \right). \quad (3.8)$$

□

Enfin, nous pouvons déduire le comportement asymptotique des statistiques portmanteau. Hosking (voir [Hos81a, Hos89]) a proposé plusieurs formes équivalentes à la statistique Q_m^c (voir équation (3.3)), dont les plus basiques sont les suivantes

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

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 $\text{Tr}(ABC) = \text{vec}(A)'(C' \otimes I)\text{vec}B$.

Sous les hypothèses que le DGP 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 Q_m^c et Q_m^h est généralement bien approchée par la distribution $\chi_{d^2m-k_0}^2$ ($d^2m > k_0$). En vue du théorème 8 nous déduisons le résultat suivant, qui donne la loi asymptotique exacte des statistiques (3.3). Nous verrons que cette distribution peut être très différente de celle d'une distribution $\chi_{d^2m-k_0}^2$ dans le cas de modèles VARMA(p_0, q_0) forts.

Théorème 9. *Sous les hypothèses du théorème 8, les statistiques Q_m^c et Q_m^h convergent en loi, quand $n \rightarrow +\infty$, vers*

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

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

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

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

Ainsi quand le processus d'erreur est un bruit blanc faible, la distribution asymptotique des statistiques Q_m^c et Q_m^h est une somme pondérée de chi-deux. Cette distribution peut être très différente de l'approximation chi-deux usuelle du cas fort. Nous en déduisons alors des tests portmanteau modifiés pour tester l'adéquation de modèles VARMA faibles.

En vue du théorème 9, la distribution asymptotique des statistiques (3.3) 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 $\hat{\Sigma}_e = \hat{f}_e(0)$. La matrice Φ_m peut également 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 au chapitre 1) du processus stationnaire $w_t = (w'_{1t}, w'_{2t})'$, où $w_{1t} = -2J_{11}^{-1}(\partial e'_t(\theta_0)/\partial \theta) \Sigma_{e_0}^{-1} e_t(\theta_0)$ et $w_{2t} = (e'_{t-1}, \dots, e'_{t-m})' \otimes e_t$.

En interprétant $(2\pi)^{-1}\Xi$ comme étant la densité spectrale évaluée en zéro du processus stationnaire (w_t) , nous avons

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

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

$$\Phi(L)w_t := w_t + \sum_{i=1}^{\infty} \phi_i w_{t-i} = u_t, \quad (3.9)$$

où u_t est un bruit blanc faible de matrice de variance Σ_u . Puisque w_t est inconnu, nous posons \hat{w}_t le vecteur obtenu en remplaçant θ_0 par $\hat{\theta}_n$ dans l'expression de w_t . Nous définissons le polynôme $\hat{\Phi}_r(z) = I_{k_0+d^2m} + \sum_{i=1}^r \hat{\Phi}_{r,i} z^i$, où $\hat{\Phi}_{r,1}, \dots, \hat{\Phi}_{r,r}$ sont les coefficients de la régression des moindres carrés de \hat{w}_t sur $\hat{w}_{t-1}, \dots, \hat{w}_{t-r}$. 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 [FRZ05].

Théorème 10. *Sous les hypothèses du théorème 9, nous supposons que le processus (w_t) admet une représentation $\text{VAR}(\infty)$ (3.9) 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 obtenue 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 , les tests portmanteau que nous proposons consiste à rejeter l'hypothèse H_0 (i.e. l'adéquation de modèles $\text{VARMA}(p_0, q_0)$ faibles) quand

$$Q^c > S_m(1 - \alpha) \quad (\text{respectivement } Q^h > S_m(1 - \alpha))$$

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

3.3.2 Comportement asymptotique des autocorrélations résiduelles auto-normalisées

Les tests portmanteau que nous avons proposés dans la section précédente sont certes fondés sur une distribution exacte des statistiques (3.3), mais sont beaucoup plus difficiles à implémenter car leurs valeurs critiques doivent être calculées à partir du DGP (des données), contrairement à celles des versions standards qui sont simplement données par des tables de χ^2 . L'autre difficulté majeure provient du fait que la distribution asymptotique des statistiques (3.3) dépend de paramètres inconnus (voir théorème 9) tels que : le paramètre de nuisance Σ_e , la matrice Φ_m et les éléments de la matrice Ξ , dont il faut trouver des estimateurs consistants. Comme déjà abordé pour l'estimation de la matrice I (voir le chapitre 1), la matrice Ξ est la plus délicate à estimer : soit par la méthode HAC, soit par celle de la densité spectrale. Cependant, on est encore confronté à de sérieuses difficultés quant au choix de la suite de poids (ou fenêtre) pour la méthode HAC ; le choix de l'ordre r de troncature de la représentation $\text{VAR}(\infty)$ est également crucial et difficile pour la méthode de la densité spectrale.

Dans [A12], nous proposons une méthode alternative où nous n'estimons pas une matrice de covariance asymptotique. Nous utiliserons une approche basée sur des techniques d'auto-normalisation (self-normalization, en anglais) pour construire une nouvelle statistique de test qui, sous l'hypothèse nulle, admet une distribution asymptotique qui ne dépend pas de paramètres inconnus. Cette technique d'auto-normalisation provient de [Lob01] et a déjà été étendue à d'autres cadres encore plus généraux par [KL06, Sha0a, Sha0b, Sha12]. On peut se référer à [Sha15] pour un état de l'art sur certains développements récents sur l'inférence statistique dans le domaine des séries temporelles utilisant cette approche d'auto-normalisation.

Soit $(B_K(r))_{r \geq 0}$ un mouvement Brownien K -dimensionnel commençant en 0. Pour $K \geq 1$, nous notons la variable aléatoire \mathcal{U}_K définie par

$$\mathcal{U}_K = B_K'(1) V_K^{-1} B_K(1) \tag{3.10}$$

où

$$V_K = \int_0^1 (B_K(r) - rB_K(1))(B_K(r) - rB_K(1))' dr. \quad (3.11)$$

Les valeurs critiques de \mathcal{U}_K ont été tabulées par Lobato dans [Lob01]. Nous notons Λ la matrice bloc dans $\mathbb{R}^{md^2 \times (k_0 + md^2)}$ définie par $\Lambda = (\Phi_m J_{11}^{-1} | I_{md^2})$. Au vu de (1.18) et (3.7), nous déduisons

$$\sqrt{n} \hat{f}_m = \frac{1}{\sqrt{n}} \sum_{t=1}^n \Lambda w_t + o_{\mathbb{P}}(1), \quad \text{où } w_t = \left(\left(-2 \frac{\partial e_t'}{\partial \vartheta} \Sigma_{e_0}^{-1} e_t \right)', (e_{t-1} \otimes e_t)', \dots, (e_{t-m} \otimes e_t)' \right)' \quad (3.12)$$

Contrairement à la section précédente, nous n'allons pas chercher à trouver la variance asymptotique du processus Λw_t mais plutôt appliquer un théorème central limite fonctionnel sur le processus $w := (w_t)_{t \geq 1}$ (comme dans le lemme 1 de [Lob01]). Enfin, nous définissons la matrice de normalisation $C_{md^2} \in \mathbb{R}^{md^2 \times md^2}$ par

$$C_{md^2} = \frac{1}{n^2} \sum_{t=1}^n S_t S_t' \quad \text{où } S_t = \sum_{j=1}^t (\Lambda w_j - \Gamma_m).$$

Afin d'assurer l'inversibilité de cette matrice C_{md^2} , nous aurons besoin de l'hypothèse (technique) suivante sur la distribution de ϵ_t .

(H8) Le processus (ϵ_t) a une densité strictement positive au voisinage de zéro.

Nous énonçons le résultat suivant, qui donne la distribution asymptotique des autocovariances et autocorrélations résiduelles renormalisées.

Théorème 11. *Supposons que $p_0 > 0$ ou $q_0 > 0$. Sous les hypothèses (H1)–(H8) et H_0 , nous avons*

$$n \hat{f}_m' C_{md^2}^{-1} \hat{f}_m \xrightarrow[n \rightarrow \infty]{d} \mathcal{U}_{md^2}. \quad (3.13)$$

De (3.8) et (3.13), nous déduisons

$$n \hat{\rho}_m' \{I_m \otimes (S_e \otimes S_e)\} C_{md^2}^{-1} \{I_m \otimes (S_e \otimes S_e)\} \hat{\rho}_m \xrightarrow[n \rightarrow \infty]{d} \mathcal{U}_{md^2}. \quad (3.14)$$

Nous donnons un schéma de la démonstration.

Démonstration. Nous notons $\mathbb{D}^k[0,1]$ l'espace de Skorokhod, qui est l'ensemble des fonctions càdlàg, de fonctions de \mathbb{R}^k à valeurs définies dans $[0,1]$. Cet espace est muni de la topologie de Skorokhod, nous notons également $\xrightarrow{\mathbb{D}^k}$ la convergence faible sur l'espace $\mathbb{D}^k[0,1]$.

Au vu de (3.12), on voit que le comportement asymptotique de \hat{f}_m dépend de celui de w_t . Soit $(B_m(r))_{r \geq 0}$ un mouvement Brownien m -dimensionnel et Ψ une matrice triangulaire inférieure dont les éléments diagonaux sont positifs. On montre d'abord que

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nr \rfloor} \Lambda w_j \xrightarrow[n \rightarrow \infty]{\mathbb{D}^m} \Psi B_m(r) \quad (3.15)$$

en utilisant la proposition 1, le théorème 7, des arguments similaires à ceux utilisés pour la

démonstration du théorème 2, et le théorème central limite fonctionnel de [Her84] pour processus α -mélangeants. En montrant (3.15), on satisfait l'hypothèse 1 de [Lob01] (voir équation (19)).

Ensuite, en suivant les arguments développés dans les sections 2 et 3 de [Lob01], le *continuous mapping theorem* sur l'espace de Skorokhod implique que

$$C_{md^2} \xrightarrow[n \rightarrow \infty]{d} \Psi V_{md^2} \Psi', \text{ où } V_{md^2} = \int_0^1 (B_{md^2}(r) - rB_{md^2}(1))(B_{md^2}(r) - rB_{md^2}(1))' dr. \quad (3.16)$$

De (3.12), en utilisant (3.15) et (3.16), on obtient

$$\begin{aligned} n \hat{\Gamma}'_m C_{md^2}^{-1} \hat{\Gamma}_m &= \frac{1}{n} \sum_{t=1}^n ((\Lambda w_t)' C_{md^2}^{-1} (\Lambda w_t)) \\ &\xrightarrow[n \rightarrow \infty]{d} (\Psi B_{md^2}(1))' (\Psi V_{md^2} \Psi')^{-1} (\Psi B_{md^2}(1)) = B'_{md^2}(1) V_{md^2}^{-1} B_{md^2}(1) =: \mathcal{U}_{md^2}, \end{aligned}$$

où la variable aléatoire \mathcal{U}_{md^2} est définie par (3.10). Par conséquent, on vient de prouver (3.13). On déduit (3.14) en utilisant (3.8) et (3.13), ce qui complète la preuve du théorème 11. \square

Remarque 3.1. Notons que quand $p = q = 0$, on n'aurait pas besoin d'estimer le paramètre θ_0 . Ainsi, au regard de la preuve, on remplace le vecteur w_t par

$$\tilde{w}_t = ((e_{t-1} \otimes e_t)', \dots, (e_{t-m} \otimes e_t)')$$

et la matrice Λ est remplacée par la matrice identité. Nous retrouvons alors le résultat donné par Lobato (voir Lemme 1 de [Lob01]).

Dans la pratique, le théorème 11 n'est pas exploitable car il comporte des quantités inconnues qu'il nous faut estimer. Nous notons $\hat{C}_{md^2} \in \mathbb{R}^{md^2 \times md^2}$ un estimateur de C_{md^2} qui est défini par

$$\hat{C}_{md^2} = \frac{1}{n^2} \sum_{t=1}^n \hat{S}_t \hat{S}_t' \text{ où } \hat{S}_t = \sum_{j=1}^t (\hat{\Lambda} \hat{w}_j - \hat{\Gamma}_m), \quad (3.17)$$

avec $\hat{\Lambda} = (\hat{\Phi}_m \hat{J}_{11,n}^{-1} | I_{md^2})$. Nous énonçons le suivant résultat qui est la version applicable du théorème 11.

Théorème 12. *Sous les hypothèse du théorème 11, nous avons*

$$n \hat{\Gamma}'_m \hat{C}_{md^2}^{-1} \hat{\Gamma}_m \xrightarrow[n \rightarrow \infty]{d} \mathcal{U}_{md^2}.$$

Le vecteur des autocorrélations résiduelles renormalisées satisfait alors

$$n \hat{\rho}'_m \left\{ I_m \otimes (\hat{S}_e \otimes \hat{S}_e) \right\} \hat{C}_{md^2}^{-1} \left\{ I_m \otimes (\hat{S}_e \otimes \hat{S}_e) \right\} \hat{\rho}_m \xrightarrow[n \rightarrow \infty]{d} \mathcal{U}_{md^2}$$

Ainsi, la nouvelle statistique que nous proposons pour tester l'adéquation de modèles VARMA faibles est la suivante

$$Q_m^{\text{SN}} := n \hat{\rho}'_m \left\{ I_m \otimes (\hat{S}_e \otimes \hat{S}_e) \right\} \hat{C}_{md^2}^{-1} \left\{ I_m \otimes (\hat{S}_e \otimes \hat{S}_e) \right\} \hat{\rho}_m. \quad (3.18)$$

En s'inspirant des statistiques (3.3) et (3.18), nous proposons la statistique suivante (3.19) qui est la version analogue de celle introduite par [Hos80]

$$\tilde{Q}_m^{\text{SN}} = n^2 \hat{\rho}'_m \left\{ I_m \otimes (\hat{S}_e \otimes \hat{S}_e) \right\} D_{n,md^2} \hat{C}_{md^2}^{-1} \left\{ I_m \otimes (\hat{S}_e \otimes \hat{S}_e) \right\} \hat{\rho}_m, \quad (3.19)$$

où $D_{n,md^2} \in \mathbb{R}^{md^2 \times md^2}$ est une matrice diagonale définie par $D_{n,md^2} = \text{Diag}(n/(n-1), \dots, n/(n-m))$. Les valeurs critiques des statistiques (3.18) et (3.19) sont tabulées dans [Lob01], ce qui rend finalement notre méthode plus proche de la méthode standard dans laquelle ces valeurs critiques sont déduites d'une table de χ^2 .

3.3.3 Illustrations numériques

Dans cette section, nous présentons quelques résultats de simulations afin d'évaluer la performance des tests portmanteau modifiés proposés. Nous introduisons les notations suivantes pour la suite

- LB_S et BP_S pour se référer aux tests de LB et BP utilisant les statistiques (3.3).
- LB_{SN} et BP_{SN} pour se référer aux tests auto-normalisés utilisant les statistiques (3.18) et (3.19)
- LB_{BM} et BP_{BM} pour se référer aux tests modifiés de LB et BP utilisant les statistiques (3.3) proposé dans [BM11a]

Nous considérons un modèle VARMA(1, 1) bivarié (voir [Rei97], chap. 3, exemple 3.2 p. 81) suivant

$$\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} a_{11,1} & a_{12,1} \\ a_{21,1} & a_{22,1} \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} - \begin{pmatrix} b_{11,1} & b_{12,1} \\ b_{21,1} & b_{22,1} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix} \quad (3.20)$$

où $\theta_0 = (a_{11,1}, a_{21,1}, a_{12,1}, a_{22,1}, b_{11,1}, b_{21,1}, b_{12,1}, b_{22,1})' = (1.2, 0.6, -0.5, 0.3, -0.6, 0.3, 0.3, 0.6)'$ et le terme d'erreur $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t})'$ serait un bruit blanc fort ou faible, qu'on définira ultérieurement. Remarquons que

$$\det \left\{ I_2 - z_1 \begin{pmatrix} 1.2 & -0.5 \\ 0.6 & 0.3 \end{pmatrix} \right\} = 0 \text{ et } \det \left\{ I_2 - z_2 \begin{pmatrix} -0.6 & 0.3 \\ 0.3 & 0.6 \end{pmatrix} \right\} = 0$$

pour $z_1 = 1.136 \pm 0.473i$ (de module $|z_1| = 1.23$) et pour $z_2 = \pm(0.45)^{-1/2}$ (de module $|z_2| = 1,49$). Le modèle (3.20) peut être vu comme un modèle VARMA(1, 1) sous forme échelon (noté, $\text{ARMA}_E(1, 1)$), contrairement au modèle (1.42) qui est un $\text{ARMA}_E(0, 1)$ (se référer à [Lüt05] pour plus de détails sur cette forme échelon, chap. 12, définition 12.2 p. 453).

Soit $\eta = ((\eta_{1,t}, \eta_{2,t})')_{t \geq 1}$ une suite de variable aléatoire telle que

$$\begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix} \stackrel{\text{loi}}{=} \mathcal{N}(0, I_2).$$

Dans un premier temps, nous considérons le cas de modèle VARMA fort en supposant que le terme d'erreur $\epsilon = ((\epsilon_{1,t}, \epsilon_{2,t})')_{t \geq 1}$ dans (3.20) est défini comme suit

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \stackrel{\text{loi}}{=} \mathcal{N}(0, I_2). \quad (3.21)$$

Considérons le modèle ARCH(1) suivant proposé par [Jea98]

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} = \begin{pmatrix} h_{11,t} & 0 \\ 0 & h_{22,t} \end{pmatrix} \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix} \quad (3.22)$$

où

$$\begin{pmatrix} h_{11,t}^2 \\ h_{22,t}^2 \end{pmatrix} = \begin{pmatrix} 0.3 \\ 0.2 \end{pmatrix} + \begin{pmatrix} 0.45 & 0.00 \\ 0.40 & 0.25 \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1}^2 \\ \epsilon_{2,t-1}^2 \end{pmatrix}.$$

Pour le premier modèle VARMA faible considéré, nous supposons que dans (3.20) le terme d'erreur ϵ est un ARCH(1) défini par (3.22). Ensuite, nous considérons également trois autres modèles VARMA faibles en supposant que dans (3.20) le terme d'erreur ϵ est donné respectivement par

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \eta_{1,t}\eta_{2,t-1}\eta_{1,t-2} \\ \eta_{2,t}\eta_{1,t-1}\eta_{2,t-2} \end{pmatrix}, \quad (3.23)$$

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \eta_{1,t}^2\eta_{2,t-1}\eta_{1,t-2} \\ \eta_{2,t}^2\eta_{1,t-1}\eta_{2,t-2} \end{pmatrix}, \quad (3.24)$$

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \eta_{1,t}(|\eta_{1,t-1}| + 1)^{-1} \\ \eta_{2,t}(|\eta_{2,t-1}| + 1)^{-1} \end{pmatrix}. \quad (3.25)$$

Ces trois derniers bruits faibles sont des extensions directes des bruits faibles univariés définis par [RT96].

Nous simulons $N = 1000$ trajectoires indépendantes de taille $n = 10000$ du modèle (3.20). La même série est divisée en trois séries de tailles $n = 500$, $n = 2000$ et $n = 10000$. Pour chacune des N répliques, nous utilisons la méthode du QMV pour estimer le paramètre θ_0 . Ensuite, nous appliquons les tests portmanteau sur les résidus pour différentes valeurs de $m \in \{1, \dots, 5\}$. Le niveau asymptotique des tests est $\alpha = 5\%$.

Pour les tests standards, l'hypothèse nulle est rejetée quand les statistiques (3.3) sont supérieures à $\chi_{(4m-8)}^2(0, 95)$. Pour les tests BP_{SN} et LB_{SN} , l'hypothèse nulle est rejetée quand les statistiques (3.18) et (3.19) sont supérieures à $\mathcal{U}_{4m}(0, 95)$, où les valeurs critiques $\mathcal{U}_K(0, 95)$ sont tabulées dans [Lob01] pour $K = 1, \dots, 20$.

Les tables 3.3 et 3.4 présentent les pourcentages de rejet de l'hypothèse nulle H_0 que le DGP suit un modèle VARMA(1, 1) sur toutes les N répliques indépendantes. Au vu des résultats des tables 3.3 et 3.4, les tests que nous proposons contrôlent mieux l'erreur de première espèce (à part quelques exceptions pour le test auto-normalisé quand $n \leq 2000$) que les tests standards surtout quand les innovations présentent des dépendances. Plus surprenant encore, même pour le modèle VARMA fort, les tests proposés sont meilleurs surtout quand m est petit. Contrairement aux tests proposés, les tests standards ne sont pas applicables quand $m \leq 2$.

S'agissant des puissances, tous les tests ont des puissances similaires à l'exception des tests LB_{SN} et BP_{SN} qui sont moins puissants quand $n = 500$.

En conclusion, les tests que nous proposons sont préférables aux tests standards surtout quand le nombre d'autocorrélations utilisé est petit.

TABLE 3.3 – Pourcentage de rejet dans le cas VARMA(1, 1).

Modèle	Taille n	Lag m	LB _{SN}	BP _{SN}	LB _{BM}	BP _{BM}	LB _S	BP _S
I	$n = 500$	1	4.8	4.9	4.2	4.0	n.a.	n.a.
		2	5.2	5.2	3.4	3.3	n.a.	n.a.
		3	4.0	4.1	4.4	4.2	20.6	19.9
		4	3.5	3.6	4.6	4.3	11.6	11.5
		5	3.5	3.7	3.9	3.8	8.2	7.8
I	$n = 2,000$	1	4.6	4.6	2.9	2.9	n.a.	n.a.
		2	4.8	4.8	3.2	3.1	n.a.	n.a.
		3	4.3	4.3	3.9	3.9	19.6	19.4
		4	5.1	5.1	3.5	3.4	10.5	10.4
		5	4.8	4.8	4.3	4.3	8.3	8.1
I	$n = 10,000$	1	5.7	5.7	3.7	3.7	n.a.	n.a.
		2	5.4	5.4	4.7	4.7	n.a.	n.a.
		3	5.3	5.3	4.4	4.4	19.0	19.0
		4	5.7	5.7	4.5	4.5	10.6	10.6
		5	6.1	6.1	4.4	4.3	8.5	8.5
II	$n = 500$	1	3.9	3.9	5.3	5.1	n.a.	n.a.
		2	2.9	2.9	5.5	5.2	n.a.	n.a.
		3	2.1	2.1	2.9	2.7	34.9	34.0
		4	1.1	1.2	2.8	2.7	24.5	23.7
		5	0.9	0.9	2.3	2.0	18.0	17.7
II	$n = 2,000$	1	3.9	3.9	6.1	5.9	n.a.	n.a.
		2	5.0	5.0	7.0	6.9	n.a.	n.a.
		3	3.3	3.3	5.4	5.3	43.9	43.8
		4	3.1	3.1	5.5	5.4	32.1	32.0
		5	3.0	3.0	4.9	4.8	26.1	25.9
II	$n = 10,000$	1	4.4	4.4	5.3	5.3	n.a.	n.a.
		2	4.9	4.9	6.1	6.1	n.a.	n.a.
		3	3.9	3.9	5.6	5.6	50.0	49.9
		4	4.2	4.2	5.4	5.4	37.2	37.1
		5	3.3	3.3	5.6	5.6	28.4	28.3
III	$n = 500$	1	3.4	3.4	6.6	6.5	n.a.	n.a.
		2	1.6	1.6	4.6	4.5	n.a.	n.a.
		3	0.3	0.3	2.1	2.1	50.7	50.1
		4	0.1	0.1	1.3	1.3	37.7	36.9
		5	0.0	0.0	1.1	1.1	29.7	28.9
III	$n = 2,000$	1	4.8	4.8	6.3	6.3	n.a.	n.a.
		2	3.7	3.7	5.4	5.4	n.a.	n.a.
		3	2.7	2.7	5.0	5.0	56.5	56.4
		4	2.5	2.5	3.6	3.5	41.8	41.3
		5	1.6	1.6	2.4	2.4	34.4	34.3
III	$n = 10,000$	1	5.1	5.1	6.2	6.2	n.a.	n.a.
		2	5.2	5.2	6.0	6.0	n.a.	n.a.
		3	4.0	4.0	5.8	5.8	58.6	58.5
		4	4.3	4.3	5.6	5.6	44.9	44.9
		5	4.2	4.2	5.7	5.6	38.2	38.2

I : Modèle VARMA(1, 1) fort (3.20)-(3.21) ; II : Modèle VARMA(1, 1) faible (3.20)-(3.22).
 III : Modèle VARMA(1, 1) faible (3.20)-(3.23).

TABLE 3.4 – Pourcentage de rejet dans le cas VARMA(1, 1).

Modèle	Taille n	Lag m	LB _{SN}	BP _{SN}	LB _{BM}	BP _{BM}	LB _S	BP _S
IV	$n = 500$	1	2.4	2.4	9.5	9.5	n.a.	n.a.
		2	0.7	0.8	6.5	6.5	n.a.	n.a.
		3	0.0	0.0	5.2	5.2	52.0	51.5
		4	0.2	0.2	5.1	5.1	40.5	40.1
		5	0.2	0.2	4.5	4.5	37.3	36.4
IV	$n = 2,000$	1	3.1	3.1	6.0	6.0	n.a.	n.a.
		2	2.5	2.5	4.8	4.8	n.a.	n.a.
		3	2.5	2.5	3.2	3.2	61.8	61.5
		4	2.0	2.0	3.0	3.0	50.2	50.0
		5	0.4	0.4	1.8	1.7	43.9	43.8
IV	$n = 10,000$	1	4.7	4.7	4.1	4.1	n.a.	n.a.
		2	4.4	4.4	5.6	5.6	n.a.	n.a.
		3	3.8	3.8	5.2	5.2	70.4	70.4
		4	3.5	3.5	4.1	4.1	57.3	57.3
		5	3.1	3.1	4.2	4.2	50.3	50.3
V	$n = 500$	1	4.0	4.0	3.6	3.6	n.a.	n.a.
		2	4.5	4.6	3.8	3.8	n.a.	n.a.
		3	3.9	4.1	4.2	4.2	15.1	14.7
		4	3.6	3.8	4.4	4.0	9.6	9.5
		5	3.1	3.2	3.7	3.4	7.8	7.0
V	$n = 2,000$	1	5.5	5.5	3.7	3.7	n.a.	n.a.
		2	5.4	5.5	4.4	4.4	n.a.	n.a.
		3	4.5	4.6	3.8	3.8	16.6	16.4
		4	4.4	4.4	4.0	4.0	9.5	9.3
		5	4.3	4.3	4.3	4.3	7.9	7.8
V	$n = 10,000$	1	4.3	4.3	3.6	3.6	n.a.	n.a.
		2	4.2	4.2	3.2	3.2	n.a.	n.a.
		3	4.7	4.7	3.9	3.9	15.3	15.3
		4	4.4	4.4	4.1	4.1	9.1	9.1
		5	4.8	4.8	3.9	3.9	6.8	6.8

IV : Modèle VARMA(1, 1) faible (3.20)-(3.24).

V : Modèle VARMA(1, 1) faible (3.20)-(3.25).

3.4 Conclusion

Nous avons abordé le problème de la validation des modèles VARMA faibles en proposant des tests modifiés qui tiennent compte des dépendances des termes d'erreur.

Dans un premier temps, 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. Ensuite, nous avons proposé des statistiques de tests modifiées permettant de vérifier que les résidus des modèles VARMA faibles sont des estimations de bruit blanc tout en établissant leur comportement asymptotique.

Les expériences de Monte Carlo réalisées sur des modèles VARMA bivariés montrent que les résultats des tests proposés sont meilleurs que ceux des tests standards, en particulier quand les innovations présentent des dépendances.

Chapitre 4

Contribution aux modèles GARCH multivariés

Depuis la mise en évidence par Mandelbrot en 1963 d'un certain nombre de caractéristiques (régularités statistiques ou encore faits stylisés) communes aux séries financières (voir [Man63]), leur modélisation a connu de grandes évolutions dans ces quarante dernières années. Leurs principales propriétés sont compatibles avec celles d'un bruit blanc faible. Ces caractéristiques rendent particulièrement complexe la modélisation des séries financières. Cette complexité n'est pas seulement due à la grande variété des séries utilisées (prix d'action, taux d'intérêt, taux de change, ...), à l'importance de la fréquence d'observation (seconde, minute, heure, jour, ...) ou à la disponibilité d'échantillons de très grande taille. Elle tient surtout à l'existence de ces faits stylisés difficiles à reproduire à partir de modèles stochastiques. Ces régularités empiriques, vérifiées et complétées depuis par de nombreux auteurs (voir par exemple [Fam65, Bla76]), apparaissent plus ou moins nettement en fonction de la fréquence d'observation de la série et de sa nature. Elles ont ensuite été enrichies par [Bla76] par l'introduction d'une caractéristique dite d'effet de levier (qui nous intéressera dans la suite), en remarquant une asymétrie de l'impact des valeurs passées positives et négatives sur la volatilité de la date courante. Plus précisément, les valeurs négatives (baisses des cours) tendent à provoquer plus de turbulences que les valeurs positives (hausses des cours) de même ampleur.

Les spécificités illustrent la difficulté de modéliser les séries financières. Les formulations classiques (de type ARMA forts) sont inappropriées car centrées sur la structure d'autocovariance des processus. Ainsi, plusieurs modèles ont été introduits dans la littérature économétrique afin de prendre en compte les propriétés particulières des séries financières. Mais dans ce chapitre, nous nous intéressons aux modèles ARCH introduits par [Eng82] et à leurs extensions GARCH dues à [Bol86], qui ont connu un grand succès dans la littérature. Cela est dû notamment à leur spécification de la variance conditionnelle (volatilité), qui ne dépend que du module des valeurs passées : l'effet sur la volatilité de la date présente des innovations passées positives et négatives est donc identique. Cette symétrie des modèles GARCH standard est en contradiction avec plusieurs études sur les séries d'action qui mettent en évidence une corrélation négative entre le carré des innovations de la date présente et les innovations passées (si la distribution conditionnelle était symétrique, cette corrélation serait nulle). Pour pallier ce problème, diverses paramétrisations de la variance conditionnelle ont été proposées. Nelson (voir [Nel91a]) a introduit le modèle GARCH exponentiel (EGARCH). Mais l'une des formulations la plus naturelle est d'introduire l'asymétrie en spécifiant la variance conditionnelle en fonction des composantes positives et négatives des inno-

vations passées (voir par exemple [GJR93] pour le modèle GJR-GARCH et [RZ93] pour le modèle TGARCH (Threshold GARCH, en anglais)). La prolifération des modèles GARCH a amené certains auteurs à proposer des familles de modèles GARCH regroupant plusieurs modèles prenant en compte l'effet de levier. Motivés par les transformations de Box-Cox, [HB92] ont introduit le modèle ARCH en puissance (PARCH, power-transformed ARCH en anglais). Dans ce modèle, l'écart-type conditionnel (également appelé volatilité) élevé à une puissance δ est exprimé comme fonction linéaire de ses valeurs passées ainsi que des valeurs passées du processus observé, élevées à la même puissance (voir également [DGE93]). Plus récemment, [HK04] ont étendu le modèle PARCH à la classe des modèles asymétriques APARCH (the power transformed asymmetric (threshold) ARCH en anglais); [PWT08] ont introduit le modèle APARCH généralisé (APGARCH, pour generalized the APARCH).

Pour les applications économétriques, le cadre univarié est très restrictif. Mais, contrairement aux modèles ARMA, les extensions des modèles GARCH au cadre multivarié (MGARCH, pour multivariate GARCH, dans la suite) ne sont pas directes. Il existe dans la littérature différents types de modèles MGARCH (voir par exemple [Nel91b, EK95, Eng02, MCHL08]). On pourra se référer à [BLR06, BHL12] pour un état de l'art récent sur les modèles MGARCH. Parmi les nombreuses spécifications des modèles MGARCH, les plus populaires semblent être les modèles MGARCH à corrélation conditionnelle constante (notés dans la suite CCC-GARCH) proposés par [Bol90] et étendus par [Jea98].

Plusieurs extensions des modèles GARCH ont été introduits dans la littérature pour tenir compte de l'effet de levier mais ces extensions restent très peu nombreuses pour le cas multivarié. À notre connaissance, [MHC09] étendent le modèle GARCH-GJR (voir [GJR93]) aux modèles CCC-GARCH asymétrique (voir également [MCHL08]). Récemment, [FZ12] ont proposé un modèle CCC-AGARCH (asymmetric CCC-GARCH) qui contient le modèle CCC-GARCH introduit par [Bol90] ainsi que celui de [Jea98]. On pourra se référer également [SGnE16] pour des modèles log-GARCH-X univariés et multivariés. L'objectif de ce chapitre est de proposer une famille de modèles MGARCH asymétriques en puissance et d'en étudier les propriétés probabilistes et statistiques. Les publications liées à ce chapitre sont : [A13, A14].

4.1 Modèle, conditions de stationnarité et d'identifiabilité

Posons m le nombre de séries considérées et introduisons les notations suivantes : $X^d = (X_1^{d_1}, \dots, X_m^{d_m})'$ avec $\underline{d}, \underline{X} \in \mathbb{R}^m$, $x^+ = \max(0, x)$ et $x^- = \min(0, x)$.

4.1.1 Présentation du modèle

Le processus $\underline{\varepsilon}_t$ à valeurs dans \mathbb{R}^m , $\underline{\varepsilon}_t = (\varepsilon_{1,t}, \dots, \varepsilon_{m,t})'$, est dit CCC-APGARCH(p, q) (c'est-à-dire un CCC-AGARCH(p, q) en puissance $\underline{\delta}_0$ ou $\underline{\delta}_0$ -CCC-AGARCH(p, q)) s'il vérifie

$$\begin{cases} \underline{\varepsilon}_t = H_t^{1/2} \eta_t, \\ H_t = D_t R_0 D_t, \quad D_t = \text{diag}(\sqrt{h_{1,t}}, \dots, \sqrt{h_{m,t}}), \\ h_t^{\delta_0/2} = \underline{\omega}_0 + \sum_{i=1}^q \left\{ A_{0i}^+(\underline{\varepsilon}_{t-i}^+)^{\delta_0/2} + A_{0i}^-(\underline{\varepsilon}_{t-i}^-)^{\delta_0/2} \right\} + \sum_{j=1}^p B_{0j} h_{t-j}^{\delta_0/2}, \end{cases} \quad (4.1)$$

où $\underline{h}_t = (h_{1,t}, \dots, h_{m,t})'$ et

$$\underline{\varepsilon}_t^+ = (\{\varepsilon_{1,t}^+\}^2, \dots, \{\varepsilon_{m,t}^+\}^2)' \quad \underline{\varepsilon}_t^- = (\{-\varepsilon_{1,t}^-\}^2, \dots, \{-\varepsilon_{m,t}^-\}^2)',$$

avec $\underline{\omega}_0$ et $\underline{\delta}_0$ des vecteurs de taille $m \times 1$ à coefficients strictement positifs, les A_{0i}^+ , A_{0i}^- et B_{0j} sont des matrices de dimension $m \times m$ à coefficients positifs et R_0 est une matrice de corrélation. Le processus des innovations $(\eta_t)_t$ est un vecteur de taille $m \times 1$ vérifiant l'hypothèse suivante :

A0 : (η_t) est une suite de variables aléatoires iid, centrées et de matrice de variance unité.

Sous **A0**, la matrice H_t s'interprète comme la variance conditionnelle (volatilité) du processus $\underline{\varepsilon}_t$.

Remarque 4.1. La puissance $\underline{\delta}_0$ dans la représentation (4.1) est considérée vectorielle (et non scalaire) car on a constaté que dans la pratique, les séries univariées n'ont pas forcément la même puissance.

La représentation (4.1) englobe plusieurs modèles MGARCH. Par exemple, en supposant que $A_{0i}^+ = A_{0i}^-$, le modèle (4.1) peut être vu comme une extension multivariée et une généralisation du modèle PARCH introduit par [HB92] (que nous noterons CCC-PGARCH dans la suite). De plus, si on fixe le vecteur puissance $\underline{\delta}_0 = (2, \dots, 2)'$, on obtient le modèle CCC-GARCH(p, q) proposé par [Jea98]; et en supposant que $A_{0i}^+ \neq A_{0i}^-$, on retrouve le modèle CCC-AGARCH(p, q) introduit par [FZ12]. En fixant aussi $m = 1$, qui correspond au cas univarié, on retrouve le modèle APGARCH introduit par [PWT08].

4.1.2 Condition de stationnarité stricte

La démarche pour établir la condition suffisante de stationnarité stricte du modèle $\underline{\delta}_0$ -CCC-AGARCH(p, q) est similaire à celle donnée par [FZ12]. Nous commençons par réécrire la première équation de la représentation (4.1) comme suit

$$\underline{\varepsilon}_t = D_t \tilde{\eta}_t, \quad \text{où } \tilde{\eta}_t = (\tilde{\eta}_{1,t}, \dots, \tilde{\eta}_{m,t})' = R_0^{1/2} \eta_t. \quad (4.2)$$

Posons

$$\gamma_t^{\pm, (\delta_0)} = \text{diag} \left((\tilde{\eta}_{1,t}^{\pm})^{\delta_{0,1}}, \dots, (\tilde{\eta}_{m,t}^{\pm})^{\delta_{0,m}} \right) \quad \text{où } \delta_0 = (\delta_{0,1}, \dots, \delta_{0,m})'.$$

En multipliant par $(\gamma_t^{\pm, (\delta_0)})$ la troisième équation de (4.1), on a

$$(\underline{\varepsilon}_t^{\pm})^{\delta_0/2} = (\gamma_t^{\pm, (\delta_0)}) \underline{h}_t^{\delta_0/2}, \quad (4.3)$$

ce qui s'écrit (matriciellement)

$$\underline{z}_t = \underline{b}_t + C_t \underline{z}_{t-1},$$

où

$$\underline{z}_t = \left(\left\{ (\underline{\varepsilon}_t^+)^{\frac{\delta_0}{2}} \right\}', \dots, \left\{ (\underline{\varepsilon}_{t-q+1}^+)^{\frac{\delta_0}{2}} \right\}', \left\{ (\underline{\varepsilon}_t^-)^{\frac{\delta_0}{2}} \right\}', \dots, \left\{ (\underline{\varepsilon}_{t-q+1}^-)^{\frac{\delta_0}{2}} \right\}', \left\{ \underline{h}_t^{\frac{\delta_0}{2}} \right\}', \dots, \left\{ \underline{h}_{t-p+1}^{\frac{\delta_0}{2}} \right\}' \right)',$$

$$\underline{b}_t = \left(\left\{ (\gamma_t^{+, (\delta_0)}) \underline{\omega}_0 \right\}', 0'_{m(q-1)}, \left\{ (\gamma_t^{-, (\delta_0)}) \underline{\omega}_0 \right\}', 0'_{m(q-1)}, \{\underline{\omega}_0\}', 0'_{m(p-1)} \right)'$$

et

$$C_t = \begin{pmatrix} \gamma_t^{+,(\delta_0)} A_{01:q}^+ & \gamma_t^{+,(\delta_0)} A_{01:q}^- & \gamma_t^{+,(\delta_0)} B_{01:p} \\ I_{m(q-1)} & 0_{m(q-1) \times m(p+q+1)} & \\ \gamma_t^{-,(\delta_0)} A_{01:q}^+ & \gamma_t^{-,(\delta_0)} A_{01:q}^- & \gamma_t^{-,(\delta_0)} B_{01:p} \\ 0_{m(q-1) \times mq} & I_{m(q-1)} & 0_{m(q-1) \times m(p-1)} \\ A_{01:q}^+ & A_{01:q}^- & B_{01:p} \\ 0_{m(p+q+1) \times m(q-1)} & & I_{m(p-1)} \end{pmatrix} \quad (4.4)$$

une matrice de taille $(p+2q)m \times (p+2q)m$ où les matrices $A_{01:q}^+ := (A_{01}^+ \dots A_{0q}^+)$, $A_{01:q}^- := (A_{01}^- \dots A_{0q}^-)$ et $B_{01:p} := (B_{01} \dots B_{0p})$ sont de taille $m \times pm$. Soit $\gamma(\mathbf{C}_0)$, le coefficient (ou le plus grand exposant) de Lyapunov de la suite $\mathbf{C}_0 = \{C_t, t \in \mathbb{Z}\}$, défini par

$$\gamma(\mathbf{C}_0) := \lim_{t \rightarrow +\infty} \frac{1}{t} \mathbb{E} [\log \|C_t C_{t-1} \dots C_1\|] = \inf_{t \geq 1} \frac{1}{t} \mathbb{E} [\log \|C_t C_{t-1} \dots C_1\|].$$

Nous énonçons le résultat suivant.

Théorème 13. (*Stationnarité stricte*)

Une condition nécessaire et suffisante d'existence d'un processus strictement stationnaire, solution non anticipative du modèle (4.1) est que $\gamma(\mathbf{C}_0) < 0$. Lorsque $\gamma(\mathbf{C}_0) < 0$, cette solution stationnaire non anticipative est unique et ergodique.

Démonstration. Elle est similaire à celle donnée par [FZ12]. □

Les corollaires suivants sont des conséquences de la condition nécessaire de stationnarité stricte. Notons $\rho(A)$ le rayon spectral d'une matrice carrée A .

Corollaire 1. (*Condition nécessaire de stationnarité stricte simple à vérifier*)

Soit le polynôme à coefficients matriciels défini par $\mathcal{B}_0(z) = I_m - zB_{01} - \dots - z^p B_{0p}$ pour $z \in \mathbb{C}$ et la matrice

$$B_0 = \begin{pmatrix} B_{01:p} \\ I_{(p-1)m} & 0_{(p-1)m \times 1} \end{pmatrix}$$

Alors si $\gamma(\mathbf{C}_0) < 0$, on a les propriétés équivalentes suivantes :

- (i) Les racines de $\det(\mathcal{B}_0(z))$ sont de modules à l'extérieur du disque unité,
- (ii) $\rho(B_0) < 1$.

Corollaire 2.

On suppose que $\gamma(\mathbf{C}_0) < 0$. Soit ε_t la solution strictement stationnaire et non anticipative du modèle (4.1). Il existe $s > 0$ tel que $\mathbb{E} \|\varepsilon_t^{\delta_0/2}\|^s < \infty$ et $\mathbb{E} \|\varepsilon_t^{\delta_0/2}\|^s < \infty$.

4.1.3 Conditions d'identifiabilité

Avant l'estimation, nous étudions les conditions à imposer sur les matrices A_{0i}^+ , A_{0i}^- et B_{0j} afin d'assurer l'unicité des paramètres du modèle (4.1). Le vecteur des paramètres est noté

$$\nu := (\underline{\omega}', \alpha_1^+, \dots, \alpha_q^+, \alpha_1^-, \dots, \alpha_q^-, \beta_1', \dots, \beta_p', \underline{\tau}', \rho)'$$

où α_i^+ et α_i^- sont définis par $\alpha_i^\pm = \text{vec}(A_i^\pm)$ pour $i = 1, \dots, q$, $\beta_j = \text{vec}(B_j)$ pour $j = 1, \dots, p$ et $\rho = (\rho_{21}, \dots, \rho_{m1}, \rho_{32}, \dots, \rho_{m2}, \dots, \rho_{mm-1})'$ où les ρ_{ij} sont les composantes de la matrice de corrélation R . Le paramètre ν appartient à un espace des paramètres noté

$$\Delta \subset]0, +\infty[^m \times]0, \infty[^{m^2(2q+p)} \times]0, +\infty[^m \times]-1, 1[^{m(m-1)/2}.$$

La vraie valeur inconnue du paramètre est notée

$$\nu_0 := (\underline{\omega}'_0, \alpha_{01}^{+'}, \dots, \alpha_{0m}^{+'}, \alpha_{01}^{-'}, \dots, \alpha_{0m}^{-'}, \beta_{01}', \dots, \beta_{0m}', \underline{\delta}'_0, \rho'_0)'$$

Nous notons pour une matrice quelconque, $A := A(\nu)$, $\forall \nu \in \Delta$ et $A_0 := A(\nu_0)$.

Soit $\mathcal{A}^+(L) = \sum_{i=1}^q A_i^+ L^i$, $\mathcal{A}^-(L) = \sum_{i=1}^q A_i^- L^i$ et $\mathcal{B}(L) = I_m - \sum_{j=1}^p B_j L^j$. Par convention $\mathcal{A}^\pm = 0$ si $q = 0$ et $\mathcal{B}(L) = I_m$ si $p = 0$. Si les racines de $\det(\mathcal{B}_0(L)) = 0$ sont de modules strictement supérieurs à 1, de la forme compacte : $\mathcal{B}_0(L) \underline{h}_t^{\delta_0/2} = \underline{\omega}_0 + \mathcal{A}_0^+(L) (\underline{\varepsilon}_t^+)^{\delta_0/2} + \mathcal{A}_0^-(L) (\underline{\varepsilon}_t^-)^{\delta_0/2}$, on a

$$\underline{h}_t^{\delta_0/2} = \mathcal{B}_0(1)^{-1} \underline{\omega}_0 + \mathcal{B}_0(L)^{-1} \mathcal{A}_0^+(L) (\underline{\varepsilon}_t^+)^{\delta_0/2} + \mathcal{B}_0(L)^{-1} \mathcal{A}_0^-(L) (\underline{\varepsilon}_t^-)^{\delta_0/2}. \quad (4.5)$$

Dans le cas multivarié, supposer que les polynômes $\mathcal{A}_0^+(L)$, $\mathcal{A}_0^-(L)$ et $\mathcal{B}_0(L)$ n'ont pas de racine commune ne suffit pas à garantir l'identifiabilité du paramètre, car cela ne garantit pas qu'il n'existe pas un autre triplet $(\mathcal{A}_0^+, \mathcal{A}_0^-, \mathcal{B}_0)$ tel que

$$\mathcal{B}^{-1} \mathcal{A}^+ = \mathcal{B}_0^{-1} \mathcal{A}_0^+ \quad \text{et} \quad \mathcal{B}^{-1} \mathcal{A}^- = \mathcal{B}_0^{-1} \mathcal{A}_0^-. \quad (4.6)$$

Le paramètre θ_0 est dit identifiable si (4.5) n'est pas vérifiée lorsque ν_0 est remplacé par $\nu \neq \nu_0$ appartenant à Δ . La condition (4.6) équivaut à l'existence d'un opérateur $U(B)$ tel que

$$\mathcal{A}^+(L) = U(L) \mathcal{A}_0^+(L), \quad \mathcal{A}^-(L) = U(L) \mathcal{A}_0^-(L) \quad \text{et} \quad \mathcal{B}(L) = U(L) \mathcal{B}_0(L).$$

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 (left-coprime, en anglais). Divers types de conditions peuvent être imposées pour assurer l'identifiabilité (voir [Rei97], p. 37-40, ou [Han76] ou [HD88], section 2.7). Afin d'obtenir une condition peu restrictive, nous procédons comme dans [FZ12]. Nous définissons pour chaque colonne i des opérateurs matriciels \mathcal{A}^+ , \mathcal{A}^- et \mathcal{B} , leurs degrés maximaux respectifs, $q_i^+(\nu)$, $q_i^-(\nu)$ et $p_i(\nu)$. Supposons que les valeurs maximales de ces ordres soient imposées, c'est-à-dire

$$\forall \nu \in \Delta, \forall i = 1, \dots, m, \quad q_i^+(\nu) \leq q_i^+, \quad q_i^-(\nu) \leq q_i^-, \quad \text{and} \quad p_i(\nu) \leq p_i \quad (4.7)$$

où $q_i^+ \leq q$, $q_i^- \leq q$ et $p_i \leq p$ sont des entiers fixés. Nous notons $a_{q_i^+}^+(i)$ le vecteur colonne des coefficients $L^{q_i^+}$, $a_{q_i^-}^-(i)$ le vecteur colonne des coefficients $L^{q_i^-}$, dans la i^e colonne de \mathcal{A}_0^+ (respectivement de \mathcal{A}_0^-) et $b_{p_i}(i)$ le vecteur colonne des coefficients L^{p_i} dans la i^e colonne de \mathcal{B}_0 .

Proposition 14. (Condition d'identifiabilité simple)

Si les polynômes de matrices $\mathcal{A}_0^+(L)$, $\mathcal{A}_0^-(L)$ et $\mathcal{B}_0(L)$ sont coprimés à gauche, $\mathcal{A}_0^+(1) + \mathcal{A}_0^-(1) \neq 0$

et si la matrice

$$M(\mathcal{A}_0^+(L), \mathcal{A}_0^-(L), \mathcal{B}_0(L)) = \left[a_{q_1^+}^+(1) \dots a_{q_m^+}^+(m) a_{q_1^-}^-(1) \dots a_{q_m^-}^-(m) b_{p_1}(1) \dots b_{p_m}(m) \right]$$

est de plein rang m , sous les contraintes (4.7) avec $q_i^+ = q_i^+(\nu_0)$, $q_i^- = q_i^-(\nu_0)$ et $p_i = p_i(\nu_0)$ pour toute valeur de i , alors

$$\begin{cases} \mathcal{B}(L)^{-1} \mathcal{A}^+(L) = \mathcal{B}_0(L)^{-1} \mathcal{A}_0^+(L) \\ \mathcal{B}(L)^{-1} \mathcal{A}^-(L) = \mathcal{B}_0(L)^{-1} \mathcal{A}_0^-(L) \end{cases} \Rightarrow (\mathcal{A}^+, \mathcal{A}^-, \mathcal{B}) = (\mathcal{A}_0^+, \mathcal{A}_0^-, \mathcal{B}_0).$$

Démonstration. La démonstration est identique à celle donnée par [FZ12] dans le cas du modèle CCC-AGARCH. \square

4.2 Estimation du modèle

Pour l'estimation des modèles GARCH ou MGARCH, la méthode couramment utilisée est la celle du QMV. La théorie asymptotique des modèles MGARCH est bien connue dans la littérature. Par exemple, [Jea98] a donné des conditions assez générales pour établir la convergence forte de l'estimateur du QMV de modèles MGARCH, [CL03, HP09] ont montré la consistance et la normalité asymptotique des paramètres du modèle BEKK (acronyme de Baba, Engle, Kraft and Kroner). Les propriétés asymptotiques d'un modèle VARMA-MGARCH ont été établies par [LM03]. On pourra voir également [BW09] qui ont étudié les propriétés asymptotiques de l'estimateur du QMV pour une classe très générale de processus multivariés causaux contenant des modèles asymétriques. Plus récemment, [FZ16] ont établi les propriétés asymptotiques de modèles MGARCH estimables équation par équation (voir aussi [FS17] pour la classe des modèle log – MGARCH – X à corrélations conditionnelles dynamiques).

Nous présentons alors l'estimation du modèle (4.1) par la méthode du QMV.

4.2.1 Quand la puissance est connue

Dans cette section, nous obtenons le comportement asymptotique de l'estimateur du QMV des paramètres du modèle (4.1) quand la puissance $\underline{\delta}_0$ est connue. Ces estimateurs peuvent être appliqués aux modèles particuliers mentionnés dans la section présentation du modèle (CCC-GARCH, CCC-AGARCH, CCC-PGARCH, ...). Pour simplifier les notations, dans cette section, nous posons $\underline{\delta}_0 = \underline{\delta}$ (comme la puissance est connue).

Les paramètres sont les coefficients des matrices ω_0 , A_{0j}^+ , A_{0j}^- et B_{0j} , ainsi que les coefficients de la partie triangulaire inférieure (hors diagonale) de la matrice de corrélation R_0 . Le nombre de paramètres à estimer est donc

$$s_0 = m + m^2(p + 2q) + \frac{m(m-1)}{2}.$$

Le vecteur des paramètres est noté :

$$\theta := (\underline{\omega}', \alpha_1^{+'}, \dots, \alpha_m^{+'}, \alpha_1^{-'}, \dots, \alpha_m^{-'}, \beta_1', \dots, \beta_m', \rho')',$$

et le paramètre θ appartient à l'espace des paramètres

$$\Theta \subset]0, +\infty[^m \times [0, \infty[^{m^2(2q+p)} \times]-1, 1[^{m(m-1)/2}.$$

La vraie valeur du paramètre est notée :

$$\theta_0 := (\underline{\omega}'_0, \alpha_{01}^{+'}, \dots, \alpha_{0m}^{+'}, \alpha_{01}^{-'}, \dots, \alpha_{0m}^{-'}, \beta_{01}', \dots, \beta_{0m}', \rho'_0)'$$

Soit $\underline{\varepsilon}_t = (\varepsilon_{t1}, \dots, \varepsilon_{tm})$ une réalisation de longueur n de l'unique solution strictement stationnaire non anticipative ($\underline{\varepsilon}_t$) du modèle (4.1). Conditionnellement aux valeurs initiales $\underline{\varepsilon}_0, \dots, \underline{\varepsilon}_{1-q}, \tilde{h}_0, \dots, \tilde{h}_{1-p}$ positives ou nulles, la quasi-vraisemblance gaussienne s'écrit

$$L_n(\theta) = L_n(\theta; \underline{\varepsilon}_1, \dots, \underline{\varepsilon}_n) = \prod_{t=1}^n \frac{1}{(2\pi)^{m/2} |\tilde{H}_t|^{1/2}} \exp\left(-\frac{1}{2} \underline{\varepsilon}'_t \tilde{H}_t^{-1} \underline{\varepsilon}_t\right),$$

où les \tilde{H}_t sont définis récursivement, pour $t \geq 1$, par

$$\begin{cases} \tilde{H}_t = \tilde{D}_t R \tilde{D}_t, & \tilde{D}_t = \text{diag}\left(\sqrt{\tilde{h}_{1,t}}, \dots, \sqrt{\tilde{h}_{m,t}}\right) \\ \tilde{h}_t^{\delta/2} := \tilde{h}_t^{\delta/2}(\theta) = \underline{\omega} + \sum_{i=1}^q A_i^+ (\underline{\varepsilon}_{t-i}^+)^{\delta/2} + A_i^- (\underline{\varepsilon}_{t-i}^-)^{\delta/2} + \sum_{j=1}^p B_j \tilde{h}_{t-j}^{\delta/2}. \end{cases}$$

Un estimateur du QMV de θ est défini comme toute solution mesurable $\hat{\theta}_n$ de

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} L_n(\theta) = \arg \min_{\theta \in \Theta} \tilde{\mathcal{L}}_n(\theta), \quad (4.8)$$

où

$$\tilde{\mathcal{L}}_n(\theta) = \frac{1}{n} \sum_{t=1}^n \tilde{l}_t, \quad \tilde{l}_t = \tilde{l}_t(\theta) = \underline{\varepsilon}'_t \tilde{H}_t^{-1} \underline{\varepsilon}_t + \log |\tilde{H}_t|.$$

Afin d'établir la convergence forte de l'estimateur (4.14), nous faisons les hypothèses suivantes :

- A1** : $\theta_0 \in \Theta$ et Θ est compact,
- A2** : $\gamma(\mathbf{C}_0) < 0$ et $\forall \theta \in \Theta, \det(\mathcal{B}(z)) = 0 \Rightarrow |z| > 1$,
- A3** : Pour $i = 1, \dots, m$ la distribution de $\tilde{\eta}_{it}$ est non concentrée sur 2 points et $\mathbb{P}(\tilde{\eta}_{it} > 0) \in (0, 1)$.
- A4** : Si $p > 0, \mathcal{A}_0^+(1) + \mathcal{A}_0^-(1) \neq 0, \mathcal{A}_0^+(z), \mathcal{A}_0^-(z)$ et $\mathcal{B}_0(z)$ sont coprimés à gauche et la matrice $M(\mathcal{A}_0^+, \mathcal{A}_0^-, \mathcal{B}_0)$ est de plein rang m .

Si l'espace Θ est contraint par (4.7), l'hypothèse **A4** peut être remplacée par la condition plus générale **A4'** : **A4** avec $M(\mathcal{A}_0^+, \mathcal{A}_0^-, \mathcal{B}_0)$ remplacée par $[A_{0q}^+ A_{0q}^- B_{0p}]$.

- A5** : R est une matrice de corrélation définie positive pour tout $\theta \in \Theta$.

L'hypothèse **A2** implique qu'il existe une solution, $(h_t)_t = (h_t(\theta))_t$, strictement stationnaire non anticipative et ergodique de

$$h_t^{\delta/2} := h_t^{\delta/2}(\theta) = \underline{\omega} + \sum_{i=1}^q A_i^+ (\underline{\varepsilon}_{t-i}^+)^{\delta/2} + A_i^- (\underline{\varepsilon}_{t-i}^-)^{\delta/2} + \sum_{j=1}^p B_j h_{t-j}^{\delta/2}. \quad (4.9)$$

Notons $D_t = D_t(\theta) = \text{diag}(\sqrt{h_{1,t}}, \dots, \sqrt{h_{m,t}})$ et $H_t = H_t(\theta) = D_t(\theta)RD_t(\theta)$, et définissons

$$\mathcal{L}_n(\theta) = \frac{1}{n} \sum_{t=1}^n l_t, \quad l_t = l_t(\theta) = \varepsilon_t' H_t^{-1} \varepsilon_t + \log |H_t|.$$

Nous énonçons le théorème de la convergence forte.

Théorème 14. (*Convergence forte*)

Soit $(\hat{\theta}_n)$ une suite d'estimateurs du QMV vérifiant (4.14). Alors, sous **A0**, ..., **A5**, on a

$$\hat{\theta}_n \longrightarrow \theta_0, \quad \text{presque sûrement quand } n \rightarrow +\infty.$$

Démonstration. Nous donnons un schéma de la démonstration. On montre les étapes suivantes :

- (i) $\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} |\mathcal{L}_n(\theta) - \tilde{\mathcal{L}}_n(\theta)| = 0$, presque sûrement.
- (ii) $(\exists t \in \mathbb{Z} \text{ tel que } \underline{h}_t^{\delta/2}(\theta) = \underline{h}_t^{\delta/2}(\theta_0) \text{ presque sûrement et } R = R_0) \Rightarrow \theta = \theta_0$.
- (iii) $\mathbb{E}_{\theta_0} |l_t(\theta_0)| < \infty$, et si $\theta \neq \theta_0$, $\mathbb{E}_{\theta_0} [l_t(\theta)] > \mathbb{E}_{\theta_0} [l_t(\theta_0)]$.
- (iv) Pour tout $\theta \neq \theta_0$, il existe un voisinage $V(\theta)$ tel que

$$\liminf_{n \rightarrow \infty} \inf_{\theta^* \in V(\theta)} \tilde{\mathcal{L}}_n(\theta^*) > \mathbb{E}_{\theta_0} l_1(\theta_0), \quad \text{presque sûrement.}$$

□

Pour la normalité asymptotique, nous avons besoin d'hypothèses supplémentaires suivantes :

A6 : $\theta_0 \in \overset{\circ}{\Theta}$, où $\overset{\circ}{\Theta}$ est l'intérieur de Θ .

A7 : $\mathbb{E} \|\eta_t \eta_t'\|^2 < \infty$.

Théorème 15. (*Normalité asymptotique*)

Sous les hypothèses **A0**, ..., **A7**, nous avons

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

où J est une matrice définie positive et I est une matrice semi-définie positive, définies par

$$I := I(\theta_0) = \mathbb{E} \left[\frac{\partial l_t(\theta_0)}{\partial \theta} \frac{\partial l_t(\theta_0)}{\partial \theta'} \right], \quad J := J(\theta_0) = \mathbb{E} \left[\frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta'} \right].$$

Démonstration. Voici un schéma de la preuve qui repose sur le développement de Taylor

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = - \left(\frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \tilde{l}_t}{\partial \theta_i \partial \theta_j}(\theta_{ij}^*) \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial \tilde{l}_t}{\partial \theta}(\theta_0) \right),$$

où les θ_{ij}^* sont entre $\hat{\theta}_n$ et θ_0 . Ensuite, on montre que

$$\frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \tilde{l}_t}{\partial \theta_i \partial \theta_j}(\theta_{ij}^*) \longrightarrow J(i, j) \text{ en probabilité,} \quad (4.10)$$

et que

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial \tilde{l}_t}{\partial \theta}(\theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I). \quad (4.11)$$

□

Remarque 4.2. Dans le cas univarié, c'est-à-dire quand $m = 1$, on a $I = (\mathbb{E}\eta_t^4 - 1)J$ où

$$J = \frac{4}{\underline{\delta}_0^2} \mathbb{E} \left(\frac{1}{\underline{h}_t^{\underline{\delta}_0}(\theta_0)} \frac{\partial \underline{h}_t^{\underline{\delta}_0/2}(\theta_0)}{\partial \theta} \frac{\partial \underline{h}_t^{\underline{\delta}_0/2}(\theta_0)}{\partial \theta'} \right).$$

Ainsi, on retrouve les résultats obtenus par [HZ11].

4.2.2 Quand la puissance est inconnue

Dans cette section, nous considérons que la puissance $\underline{\delta}_0$ est inconnue et est conjointement estimée avec les autres paramètres. L'extension au cas où $\underline{\delta}_0$ est estimé est loin d'être évidente car, comme mentionné par [HZ11], l'identification de ce paramètre pose des difficultés théoriques et éventuellement numériques. Pour identifier $\underline{\delta}_0$, nous imposons une condition un peu plus forte que l'hypothèse **A3** :

A8 : η_t a une densité strictement positive au voisinage de zéro.

Dans cette section, le nombre de paramètres à estimer est

$$s_0 = 2m + m^2(2q + p) + \frac{m(m-1)}{2}.$$

Pour définir l'estimateur du QMV du paramètre ν , nous remplaçons H_t par \mathcal{H}_t dans les expressions du critère défini par (4.14). Conditionnellement aux valeurs initiales positives, nous définissons récursivement $\tilde{\mathcal{H}}_t$, pour $t \geq 1$,

$$\begin{cases} \tilde{\mathcal{H}}_t = \tilde{D}_t R \tilde{D}_t, & \tilde{D}_t = \text{diag} \left(\sqrt{\tilde{h}_{1,t}}, \dots, \sqrt{\tilde{h}_{m,t}} \right) \\ \tilde{h}_t := \tilde{h}_t(\nu) = \left(\underline{\omega} + \sum_{i=1}^q A_i^+ (\underline{\varepsilon}_{t-i}^+)^{\tau/2} + A_i^- (\underline{\varepsilon}_{t-i}^-)^{\tau/2} + \sum_{j=1}^p B_j \tilde{h}_{t-j}^{\tau/2} \right)^{2/\tau}. \end{cases}$$

L'estimateur du QMV de ν est défini comme toute solution mesurable $\hat{\nu}_n$ de

$$\hat{\nu}_n = \arg \min_{\nu \in \Delta} \tilde{\mathcal{L}}_n(\nu), \quad (4.12)$$

où

$$\tilde{\mathcal{L}}_n(\nu) = \frac{1}{n} \sum_{t=1}^n \tilde{l}_t, \quad \tilde{l}_t = \tilde{l}_t(\nu) = \underline{\varepsilon}_t' \tilde{\mathcal{H}}_t^{-1} \underline{\varepsilon}_t + \log |\tilde{\mathcal{H}}_t|.$$

Pour établir les propriétés asymptotiques de l'estimateur, nous avons besoin de certaines hypothèses similaires à celles de la section 4.2.1. Nous allons donc faire les hypothèses **A1**, ..., **A6** dans lesquelles le paramètre θ est remplacé par ν , et supposer que l'espace des paramètres Θ est remplacé par Δ .

Théorème 16. (*Consistance forte*)

Soit $(\hat{\nu}_n)$ une suite d'estimateurs du QMV vérifiant (4.14). Alors, sous les hypothèses **A0**, ..., **A5**, nous avons

$$\hat{\nu}_n \longrightarrow \nu_0, \quad \text{presque sûrement quand } n \rightarrow +\infty.$$

Théorème 17. (*Normalité asymptotique*)

Sous les hypothèses **A0**, ..., **A8**, nous avons

$$\sqrt{n}(\hat{\nu}_n - \nu_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, J^{-1}IJ^{-1}),$$

où J est une matrice définie positive et I est une matrice semi définie positive, définies par

$$I := I(\nu_0) = \mathbb{E} \left[\frac{\partial l_t(\nu_0)}{\partial \nu} \frac{\partial l_t(\nu_0)}{\partial \nu'} \right], \quad J := J(\nu_0) = \mathbb{E} \left[\frac{\partial^2 l_t(\nu_0)}{\partial \nu \partial \nu'} \right].$$

Remarque 4.3. Quand $m = 1$, nous avons $I = (\mathbb{E}\eta_t^4 - 1)J$ avec

$$J = \mathbb{E} \left(\frac{\partial \log h_t^2(\nu_0)}{\partial \nu} \frac{\partial \log h_t^2(\nu_0)}{\partial \nu'} \right).$$

On retrouve les résultats obtenus par [HZ11] (voir théorème 3.1).

4.2.3 Illustrations numériques

Dans cette section, nous supposons que $\eta_t \sim \mathcal{N}(0, I_2)$. Nous simulons $N = 100$ trajectoires indépendantes de tailles $n = 500$ et $n = 5000$ d'un modèle ARCH bivarié, c'est-à-dire le modèle (4.1) dont les ordres sont $q = 1$ et $p = 0$ avec $\underline{\delta} = (2, 2)'$. Les deux premières lignes des tables 6.1 et 4.2 donnent les vraies valeurs des paramètres ayant permis de générer les trajectoires. La table 6.1 donne le biais et le RMSE de l'estimation quand la puissance est connue. Au vu de cette table, on remarque que le biais et le RMSE décroissent quand n augmente. Les figures 4.2 et 4.3 résument les distributions de l'estimateur du QMV des simulations réalisées et confirment que la précision de l'estimation s'améliore quand n augmente. Dans le cas où $\underline{\delta}$ est inconnue, les résultats sont très satisfaisants, bien que dans le cas univarié [HZ11] mettent en évidence des difficultés numériques liées à l'estimation de $\underline{\delta}$ (voir leur Section 4.1). On aboutit finalement à une analyse similaire au cas dans lequel la puissance est connue (voir 4.2, figures 4.1 et 4.4).

4.2. ESTIMATION DU MODÈLE CCC-APGARCH

TABLE 4.1 – Distribution de l'estimateur du QMV de θ_0 quand la puissance est connue.

Taille	Vraies val.	$\underline{\omega}_0$	A_0^+		A_0^-		ρ_0
		1 1	0.25 0.05	0.05 0.25	0.5 0.5	0.5 0.5	
$n = 500$	Biais	-0.00498 -0.00090	0.00180 0.00683	0.00789 -0.02083	-0.00525 -0.01505	0.00485 0.00122	-0.00471
	RMSE	0.10714 0.12073	0.08290 0.04308	0.04520 0.07730	0.12666 0.12738	0.12379 0.15095	0.03774
$n = 5000$	Biais	0.00105 0.00044	-0.00063 -0.00175	-0.00024 0.00188	0.00346 0.00175	-0.00324 0.00121	0.00010
	RMSE	0.03526 0.03745	0.02129 0.01437	0.01464 0.02220	0.04219 0.03883	0.03950 0.03890	0.01143

TABLE 4.2 – Distribution de l'estimateur du QMV de ν_0 quand la puissance est inconnue.

Taille	Vraies val.	$\underline{\omega}_0$	A_0^+		A_0^-		$\underline{\delta}_0$	ρ_0
		1 1	0.25 0.05	0.05 0.25	0.5 0.5	0.5 0.5	2 2	
$n = 500$	Biais	0.12600 0.11629	-0.03211 0.00937	0.00370 -0.02054	-0.00853 0.04720	-0.00155 0.00177	0.27015 0.33354	-0.00112
	RMSE	0.45385 0.48839	0.08449 0.07917	0.06808 0.09275	0.14385 0.26265	0.27142 0.16417	1.13662 1.14144	0.03772
$n = 5000$	Biais	0.00921 0.00072	-0.00490 -0.00052	0.00082 0.00183	-0.00526 0.01204	-0.00574 0.00196	0.00000 0.00000	-0.00013
	RMSE	0.05023 0.04825	0.02434 0.01935	0.01783 0.02661	0.04278 0.08433	0.07725 0.04402	0.16761 0.16004	0.01067

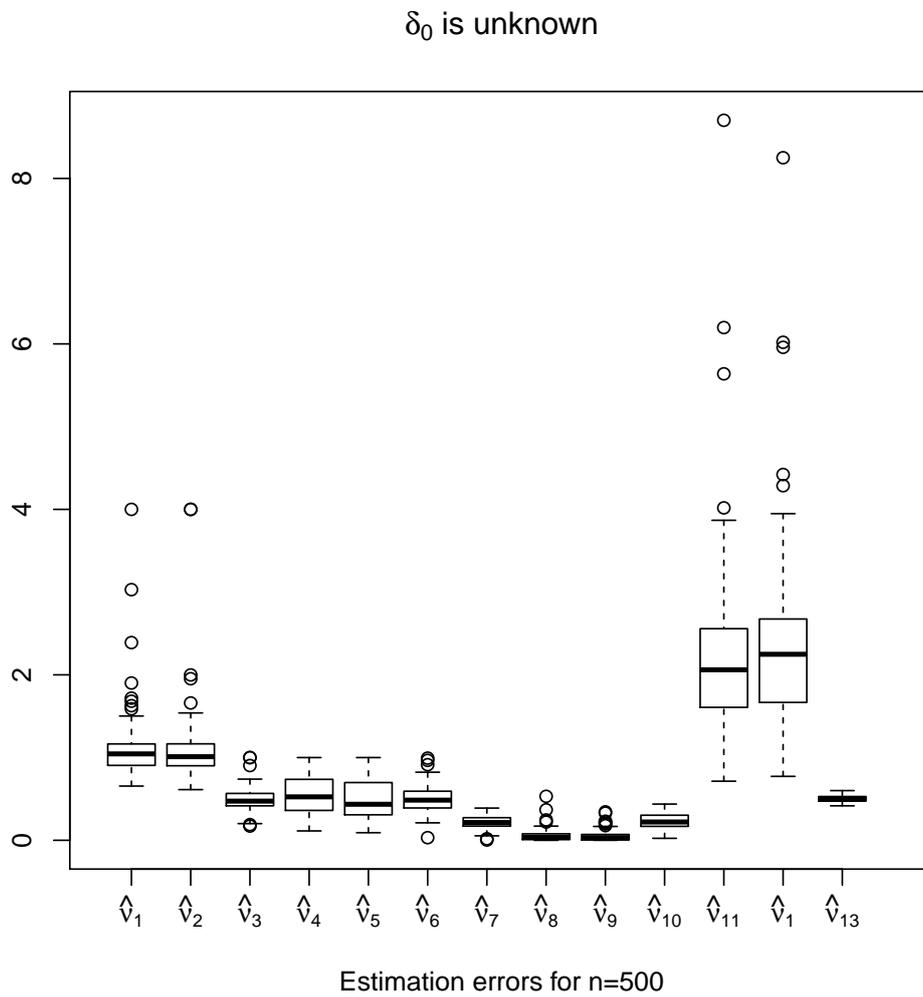


FIGURE 4.1 – Boîtes à moustaches des distributions de l'estimateur du QMV pour $n = 500$, quand la puissance est estimée.

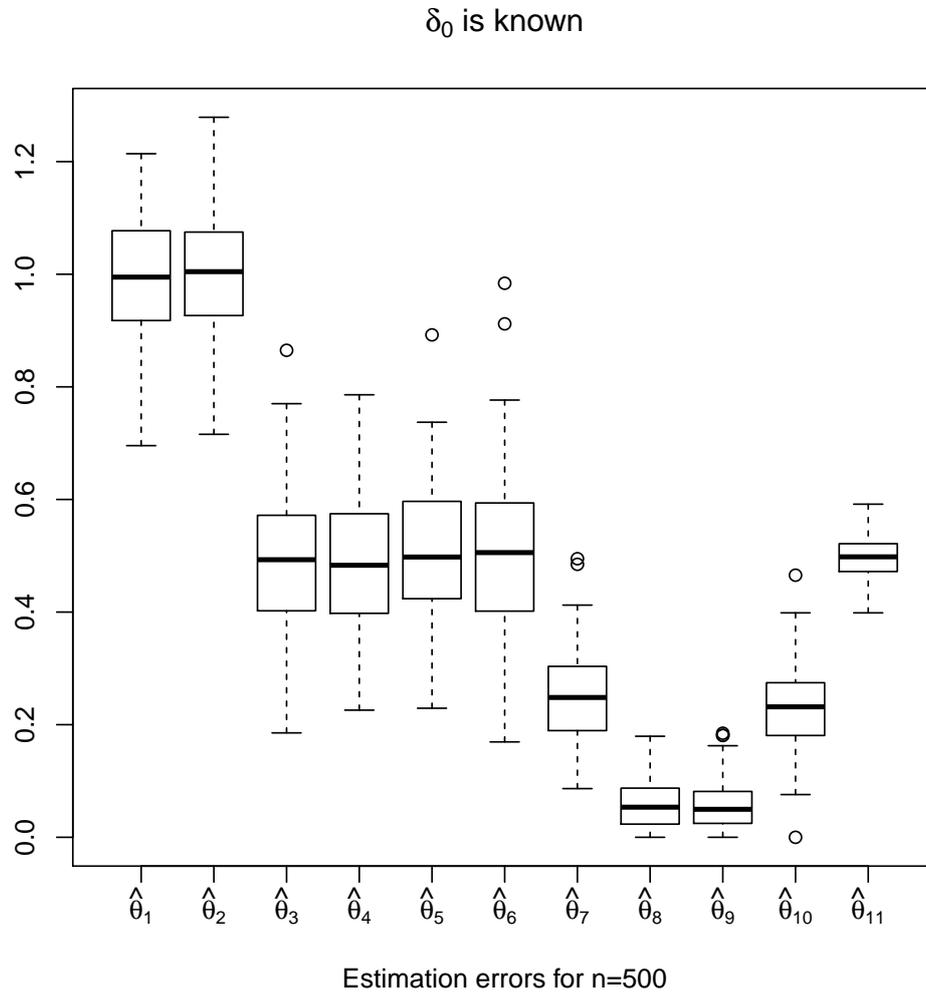


FIGURE 4.2 – Boîtes à moustaches des distributions de l'estimateur du QMV pour $n = 500$, quand la puissance est connue.

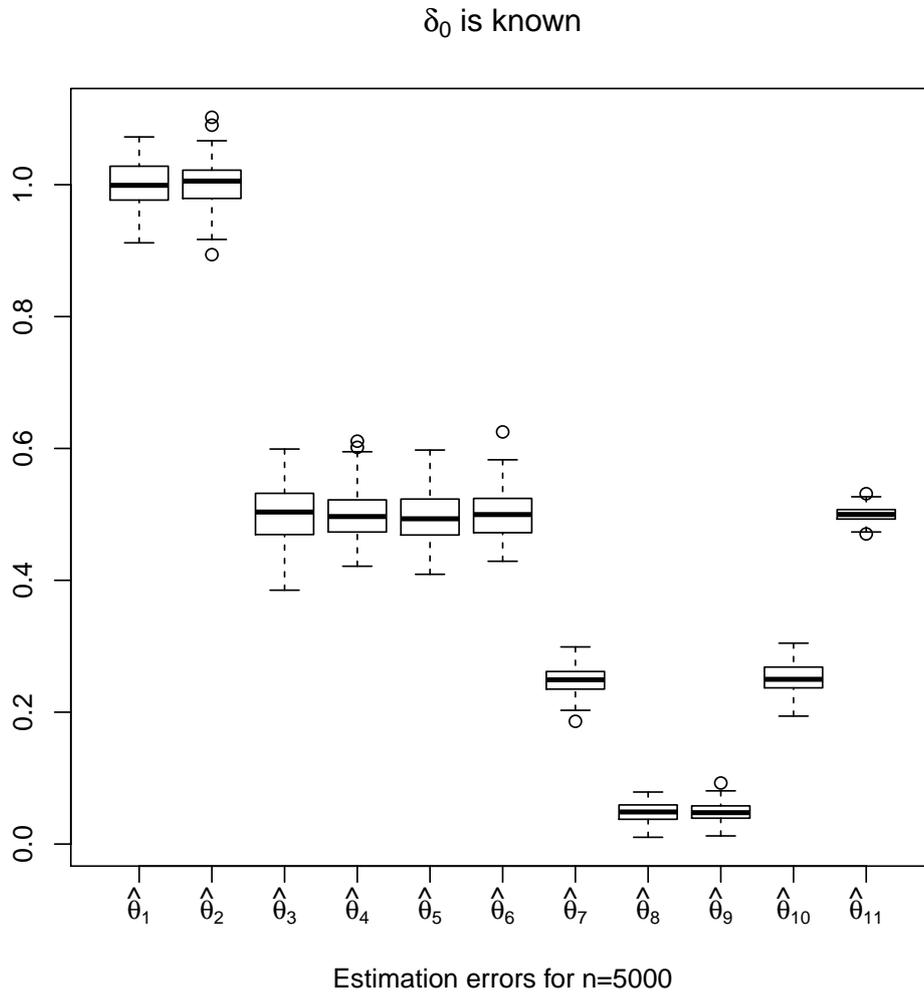


FIGURE 4.3 – Boîtes à moustaches des distributions de l'estimateur du QMV pour $n = 5000$, quand la puissance est connue.

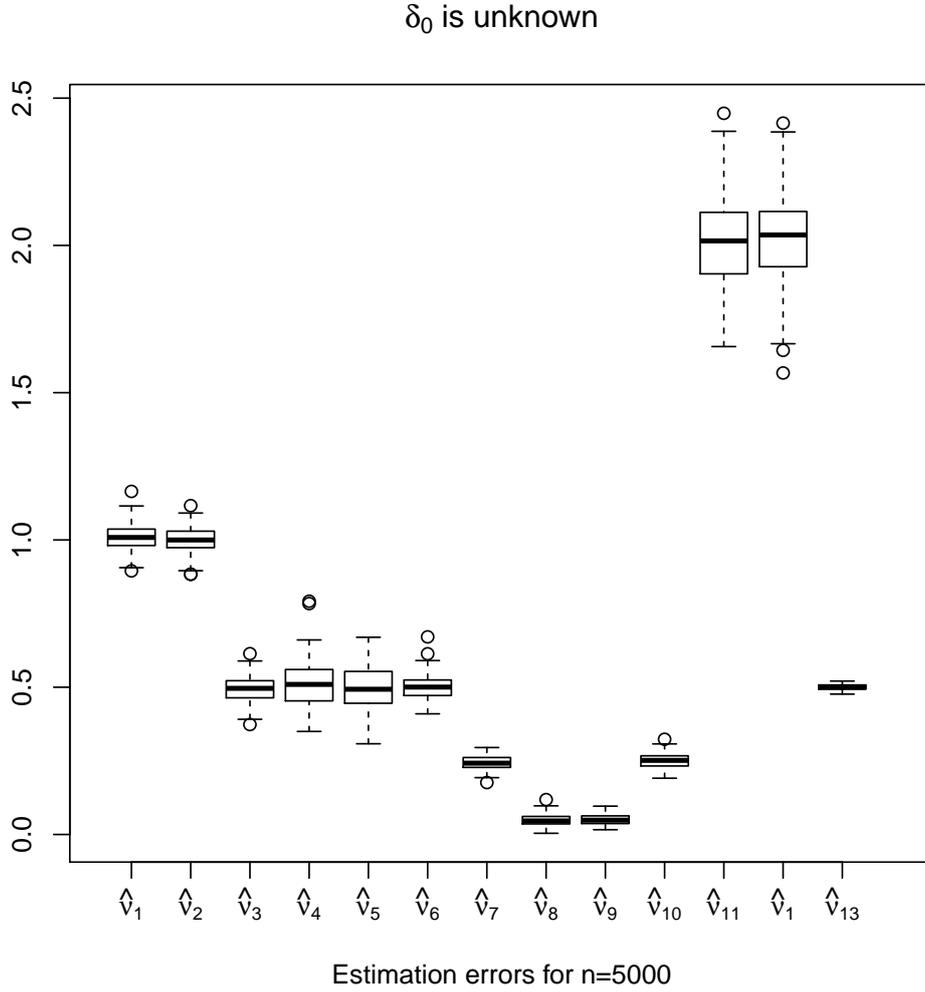


FIGURE 4.4 – Boîtes à moustaches des distributions de l'estimateur du QMV pour $n = 5000$, quand la puissance est estimée.

4.3 Validation du modèle APGARCH quand la puissance est inconnue

Dans cette section, nous considérons le modèle APGARCH introduit par [DGE93] et défini par

$$\begin{cases} \varepsilon_t = \zeta_t \eta_t \\ \zeta_t^\delta = \omega_0 + \sum_{i=1}^q \alpha_{0i}^+ (\varepsilon_{t-i}^+)^{\delta} + \alpha_{0i}^- (-\varepsilon_{t-i}^-)^{\delta} + \sum_{j=1}^p \beta_{0j} \zeta_{t-j}^{\delta} \end{cases} \quad (4.13)$$

Le vecteur inconnu des vraies valeurs du paramètre est noté $\nu_0 := (\theta_0', \delta)'$, où

$$\theta_0 = (\omega_0, \alpha_{01}^+, \dots, \alpha_{0q}^+, \alpha_{01}^-, \dots, \alpha_{0q}^-, \beta_{01}, \dots, \beta_{0p})'$$

et le vecteur des paramètres $\nu := (\theta', \tau)' = (\omega, \alpha_1^+, \dots, \alpha_q^+, \alpha_1^-, \dots, \alpha_q^-, \beta_1, \dots, \beta_p, \tau)'$.

Comme nous l'avons déjà mentionné au chapitre 3, cette étape de validation permet de valider ou d'invalider le choix des ordres p et q . Pour tester si une série univariée est une réalisation d'un bruit blanc fort, les tests portmanteau standards sont ceux définis par (3.2). Nous avons montré dans le chapitre 3, pour des ordres p et q fixés, que les distributions asymptotiques de ces statistiques (3.2) n'est plus une loi de χ_m^2 lorsque la série n'est pas un bruit blanc fort. Depuis les travaux de [LB78] et [McL78], les tests portmanteau sont devenus un outil important dans l'analyse des séries temporelles et en particulier pour tester l'adéquation de modèles ARMA(p, q). On pourra se référer également à [Li04], qui est un livre de référence sur les tests portmanteau. Ces tests standards, basés sur les autocorrélations, sont sans importance pour les processus GARCH qui sont des différences de martingales et donc non corrélés. Par conséquent, pour tester la nullité simultanée des m premières autocorrélations du modèle APGARCH (4.13), il nous faudrait des statistiques portmanteau plus robustes. Pour les GARCH standards, [LM94] et [LL97] ont proposé des tests portmanteau basés sur le carré des autocorrélations résiduelles. Les propriétés asymptotiques des statistiques de tests sous une forme quadratique des carrés des autocorrélations résiduelles ont été étudiées par [BHK03]. Un test de spécification pour un modèle GARCH(1, 1) basé sur une statistique de test de type Cramér-Von Mises a été proposé par [LKN15]. Tout récemment, [FWZ16] ont proposé un test portmanteau pour des modèles log-GARCH et EGARCH et [CF11] ont étudié le cas du modèle APARCH (4.13) quand la puissance δ est connue et ont proposé un test portmanteau dans ce cas. Aucun des travaux cités ci-dessus ne couvrent le cas où la puissance δ est inconnue et est estimée simultanément avec les autres paramètres. Dans [A14], nous considérons ce cas afin de compléter les travaux de [CF11].

4.3.1 Test portmanteau pour le modèle APGARCH quand la puissance est inconnue

L'espace des paramètres est noté $\Delta \subseteq]0, +\infty[\times]0, +\infty[^{2q+p} \times]0, +\infty[$. Soit $\varepsilon_1, \dots, \varepsilon_n$ une réalisation de longueur n de l'unique solution strictement stationnaire non anticipative (ε_t) du modèle (4.13). Conditionnellement aux valeurs initiales $\varepsilon_0, \dots, \varepsilon_{1-q}, \zeta_0 \geq 0, \dots, \zeta_{1-p} \geq 0$, on définit récursivement

$$\tilde{\zeta}_t = \tilde{\zeta}_t(\vartheta) = \left(\omega + \sum_{i=1}^q \alpha_i^+ (\varepsilon_{t-i}^+)^{\tau} + \alpha_i^- (-\varepsilon_{t-i}^-)^{\tau} + \sum_{j=1}^p \beta_j \tilde{\zeta}_{t-j}^{\tau} \right)^{1/\tau}, \text{ pour } t \geq 1.$$

L'estimateur du QMV de ν_0 du modèle (4.13) est défini comme toute solution mesurable $\hat{\nu}_n$ de

$$\hat{\nu}_n = \arg \min_{\vartheta \in \Delta} \frac{1}{n} \sum_{t=1}^n \tilde{I}_t(\vartheta), \quad \tilde{I}_t(\vartheta) = \frac{\varepsilon_t^2}{\tilde{\zeta}_t^2} + \log(\tilde{\zeta}_t^2). \quad (4.14)$$

Sous les hypothèses **A0**, **A4**, **A6**, **A7** et **A8**, la consistance et la normalité asymptotique de l'estimateur du QMV du modèle (4.13) ont été établies par [HZ11] quand la puissance δ est inconnue.

Proposition 15. (Hamadeh and Zakoïan, [HZ11]).

Soit $\hat{\nu}_n$ une suite d'estimateurs du QMV vérifiant (4.14). Sous les hypothèses précédentes, on a

$\hat{\nu}_n \rightarrow \nu_0$ p.s. quand $n \rightarrow \infty$. De plus, en posant $\kappa_\eta = \mathbb{E}\eta_t^4$, on a

$$\sqrt{n}(\hat{\nu}_n - \nu_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, (\kappa_\eta - 1)J^{-1}),$$

où

$$J := \mathbb{E}_{\vartheta_0} \left[\frac{\partial^2 l_t(\vartheta_0)}{\partial \vartheta \partial \vartheta'} \right] = 4\mathbb{E}_{\vartheta_0} \left[\frac{\partial \log(\zeta_t)(\vartheta_0)}{\partial \vartheta} \frac{\partial \log(\zeta_t)(\vartheta_0)}{\partial \vartheta'} \right], \text{ avec } l_t(\vartheta) = \frac{\varepsilon_t^2}{\zeta_t^2} + \log(\zeta_t^2)$$

Afin d'établir la distribution asymptotique de la statistique du test portmanteau, nous aurons besoin de cette hypothèse technique suivante sur la loi de η_t , qui permet d'assurer l'inversibilité de la matrice de normalisation dans la statistique de test.

A9 : η_t prend plus de 4 valeurs positives et plus de 4 valeurs négatives.

Remarque 4.4. Quand la puissance δ est connue, cette hypothèse se résume à : η_t prend plus de 3 valeurs positives et plus de 3 valeurs négatives, comme dans [CF11].

Pour tester l'hypothèse nulle

$$\mathbf{H}_0 : \text{le processus } (\varepsilon_t) \text{ satisfait le modèle (4.13),}$$

nous définissons de manière classique les autocovariances des carrés des résidus par

$$\hat{r}_h = \frac{1}{n} \sum_{t=|h|+1}^n (\hat{\eta}_t^2 - 1)(\hat{\eta}_{t-|h|}^2 - 1), \quad \hat{\eta}_t^2 = \frac{\varepsilon_t^2}{\hat{\zeta}_t^2},$$

pour $|h| < n$ et où $\hat{\zeta}_t = \tilde{\zeta}_t(\hat{\nu}_n)$. Pour un entier $m \geq 1$ fixé, nous définissons le vecteur des m premières autocovariances par

$$\hat{\mathbf{r}}_m = (\hat{r}_1, \dots, \hat{r}_m), \text{ tel que } 1 \leq m < n.$$

Nous aurons besoin des estimateurs de κ_η et J impliqués dans la normalité asymptotique de $\hat{\nu}_n$. Les quantités κ_η et J peuvent être estimées empiriquement par

$$\hat{\kappa}_\eta = \frac{1}{n} \sum_{t=1}^n \frac{\varepsilon_t^4}{\tilde{\zeta}_t^4} \quad \text{et} \quad \hat{J} = \frac{4}{n} \sum_{t=1}^n \frac{1}{\tilde{\zeta}_t^2(\hat{\nu}_n)} \frac{\partial \tilde{\zeta}_t(\hat{\nu}_n)}{\partial \nu} \frac{\partial \tilde{\zeta}_t(\hat{\nu}_n)}{\partial \nu'}.$$

Pour tout $\nu \in \Delta$, les dérivées dans l'expression de \hat{J} peuvent être calculées pour $t > 0$ de manière récursive par

$$\begin{aligned} \frac{\partial \tilde{\zeta}_t(\nu)}{\partial \theta} &= \frac{1}{\tau} \tilde{\zeta}_t^{1-\tau} \underline{\zeta}_t(\nu) + \sum_{j=1}^p \beta_j \left(\frac{\tilde{\zeta}_t(\nu)}{\tilde{\zeta}_{t-j}(\nu)} \right)^{1-\tau} \frac{\partial \tilde{\zeta}_{t-j}(\nu)}{\partial \theta}, \\ \frac{\partial \tilde{\zeta}_t(\nu)}{\partial \tau} &= \left\{ -\frac{1}{\tau^2} \log(\tilde{\zeta}_t^\tau(\nu)) + \frac{1}{\tau} \left[\sum_{i=1}^q \alpha_i^+ \log(\varepsilon_{t-i}^+) (\varepsilon_{t-i}^+)^{\tau} + \alpha_i^- \log(-\varepsilon_{t-i}^-) (-\varepsilon_{t-i}^-)^{\tau} \right. \right. \\ &\quad \left. \left. + \sum_{j=1}^p \beta_j \left(\log(\tilde{\zeta}_{t-j}(\nu)) + \tau \frac{1}{\tilde{\zeta}_{t-j}^\tau(\nu)} \frac{\partial \tilde{\zeta}_{t-j}(\nu)}{\partial \tau} \right) \tilde{\zeta}_{t-j} \right] \frac{1}{\tilde{\zeta}_{t-j}^\tau(\nu)} \right\} \tilde{\zeta}_t(\nu). \end{aligned}$$

avec les valeurs initiales $\partial \tilde{\zeta}_t(\nu)/\partial \nu = 0$, pour tout $t = 0, \dots, 1 - p$ et où le vecteur

$$\tilde{\zeta}_t(\nu) = (1, (\varepsilon_{t-1}^+)^{\tau}, \dots, (\varepsilon_{t-q}^+)^{\tau}, (-\varepsilon_{t-1}^-)^{\tau}, \dots, (-\varepsilon_{t-q}^-)^{\tau}, \tilde{\zeta}_{t-1}^{\tau}, \dots, \tilde{\zeta}_{t-p}^{\tau})'. \quad (4.15)$$

Par convention, on pose $\log(\varepsilon_t^+) = 0$ si $\varepsilon_t \leq 0$ et respectivement $\log(-\varepsilon_t^-) = 0$ si $\varepsilon_t \geq 0$. Nous définissons la matrice \hat{C}_m de taille $m \times (2q + p + 2)$, dont la $(h, k) - i^e$ composante est définie par

$$\hat{C}_m(h, k) = -\frac{2}{n} \sum_{t=h+1}^n (\hat{\eta}_{t-h}^2 - 1) \frac{1}{\tilde{\zeta}_t(\hat{\nu}_n)} \frac{\partial \tilde{\zeta}_t(\hat{\nu}_n)}{\partial \nu_k}, \quad (4.16)$$

pour $1 \leq h \leq m$ et $1 \leq k \leq 2q + p + 2$. Le résultat suivant est établi dans le cas où la puissance δ est estimée conjointement avec les autres paramètres.

Théorème 18.

*Sous les hypothèses **A0–A4**, **A6–A9** et **H₀**, la matrice $\hat{D} := (\hat{\kappa}_{\eta} - 1)^2 I_m - (\hat{\kappa}_{\eta} - 1) \hat{C}_m \hat{J}^{-1} \hat{C}_m$ converge en probabilité vers la matrice définie $D := (\kappa_{\eta} - 1)^2 I_m - (\kappa_{\eta} - 1) C_m J^{-1} C_m'$ et nous obtenons*

$$n \hat{\mathbf{r}}_m' \hat{D}^{-1} \hat{\mathbf{r}}_m \xrightarrow{\mathcal{L}} \chi_m^2.$$

Démonstration. Voici un schéma de la preuve en 4 étapes :

- (i) prouver qu'asymptotiquement, l'influence des valeurs initiales est négligeable sur la statistique $\hat{\mathbf{r}}_m$,
- (ii) établir la distribution asymptotique de $\sqrt{n} \hat{\mathbf{r}}_m$,
- (iii) prouver l'inversibilité de la matrice D ,
- (iv) et enfin établir la convergence vers la matrice D .

□

On rejette ainsi l'adéquation du modèle APGARARCH(p, q) définie par (4.13), quand la puissance δ est inconnue, pour un niveau asymptotique α quand

$$n \hat{\mathbf{r}}_m' \hat{D}^{-1} \hat{\mathbf{r}}_m > \chi_m^2(1 - \alpha), \quad (4.17)$$

où $\chi_m^2(1 - \alpha)$ représente le quantile d'ordre $(1 - \alpha)$ d'une distribution de chi-deux à m degrés de liberté.

4.3.2 Illustrations numériques

Nous présentons quelques expériences de Monte Carlo permettant d'illustrer numériquement la performance du test proposé. Nous simulons $N = 1000$ trajectoires indépendantes de taille $n = 5000$ du modèle APGARARCH(p, q) (4.13) avec des ordres $(p, q) \in \{0, 1, 2\} \times \{1, 2, 3\}$. La distribution de η_t est une loi de Student à 9 degrés de liberté, standardisée afin d'obtenir une variance égale à 1. Le vecteur de paramètre utilisé pour simuler le modèle APGARARCH(1, 1) est $\nu_{0,11} := (0.04, 0.02, 0.13, 0.85, 1)'$. Concernant les autres modèles, les vecteurs de paramètres utilisés pour générer les trajectoires sont notés $\nu_{0,pq}$ (ils sont proches de $\nu_{0,11}$) et ont été choisis de telle sorte que les conditions d'existence des moments d'ordre quatre soient satisfaites.

Pour chacune de ces N répliquions et pour chaque modèle APGARARCH(p, q) considéré, nous estimons le paramètre $\nu_{0,pq}$ par QMV et nous appliquons le test portmanteau proposé sur le

carré des résidus pour différentes valeurs de m . Pour un niveau asymptotique $\alpha = 5\%$, l'intervalle de significativité est $[3, 6\%, 6, 4\%]$ avec une probabilité de 95% et est $[3, 2\%, 6, 9\%]$ avec une probabilité de 99%.

Le tableau de gauche de la table 1 donne le pourcentages de rejet des ordres p et q du modèle APGARCH(p, q) considéré, quand la puissance δ est estimée conjointement avec les autres paramètres. On remarque que, globalement, tous les pourcentages sont à l'intérieur des intervalles de significativité avec une probabilité 95% et 99%. Le test proposé contrôle bien l'erreur de première espèce. Le tableau de droite montre la puissance du test quand l'hypothèse nulle est un modèle APGARCH(2, 2), avec $\nu_{0,22} = (1, 0.1, 0.05, 0.2, 0.1, 0.40, 0.1, 1)'$. Le test fait bien la distinction entre les modèles considérés.

Tests de bruit APGARCH								Puissance empirique					
niveau	(p, q)	m						m					
		2	4	6	8	10	12	2	4	6	8	10	12
$\alpha = 5\%$	(0, 1)	4.5	4.7	4.2	5.9	6.3	5.5	99.6	99.8	99.8	99.8	100.0	99.3
	(1, 1)	4.7	4.6	3.8	6.9	7.4	6.0	16.8	10.5	10.2	10.4	10.0	8.0
	(1, 2)	4.3	4.1	5.3	5.5	5.9	5.2	5.4	5.2	5.7	6.6	6.3	5.3
	(1, 3)	5.5	6.5	6.2	6.3	6.2	8.0	15.8	18.5	16.5	17.1	18.3	14.9
	(2, 1)	4.7	6.2	4.7	6.3	6.1	6.5	35.7	28.3	25.6	25.3	24.1	22.0
	(2, 2)	6.4	6.4	6.0	6.3	6.3	5.3	6.4	6.4	6.0	6.3	6.3	5.3

Table 1 :

Tableau de gauche : Pourcentages de rejet des modèles APGARCH(p, q) considérés.

Tableau de droite : Poucentages de rejet du modèle APGARCH(2, 2).

La table 2 présente les pourcentages de rejet du modèle APGARCH(1, 1) quand on fait varier la puissance $\delta \in [0.5, 3]$, qui est toujours estimée. Le paramètre θ_0 utilisé pour simuler les différentes trajectoires est $\theta_0 = (0.04, 0.02, 0.13, 0.85)'$. Nous aboutissons à la même conclusion que précédemment, c'est-à-dire que le test contrôle bien l'erreur de première espèce.

level	(δ)	m						level	(δ)	m					
		2	4	6	8	10	12			2	4	6	8	10	12
$\alpha = 1\%$	0.5	1.0	1.2	1.1	1.2	2.3	2.4	$\alpha = 5\%$	0.5	4.7	5.6	4.8	6.0	6.0	6.4
	1	1.3	1.3	1.4	1.7	1.3	1.7		1	3.2	4.4	5.8	5.4	5.4	5.1
	1.5	1.4	1.2	1.0	1.7	1.3	1.7		1.5	4.4	4.2	5.5	6.0	4.6	5.1
	2	1.8	1.4	1.0	1.4	1.8	1.7		2	5.2	4.7	5.5	5.1	5.7	7.4
	2.5	1.1	1.1	0.9	1.8	1.8	1.1		2.5	3.7	4.9	4.9	4.3	5.0	5.3
	3	1.3	1.9	1.2	1.7	1.7	1.6		3	3.8	3.4	4.8	5.4	4.9	6.6

Table 2 :

Pourcentage de rejet du modèle APGARCH(1, 1) pour différentes valeurs de δ .

4.4 Conclusion

Nous avons établi les propriétés asymptotiques des estimateurs du QMV de modèles CCC-APGARCH, quand le vecteur puissance est connu ou inconnu, sans faire d'hypothèse d'existence de moments. En effet, nous avons uniquement besoin de la stationnarité stricte pour laquelle

nous fournissons une condition nécessaire et suffisante explicite. Nous avons également abordé le problème de la validation en dimension 1, en proposant un test portmanteau robuste quand la puissance est estimée conjointement avec les autres paramètres. Les résultats obtenus ont été illustrés numériquement.

Chapitre 5

Contribution à l'estimation de modèles RCARMA faibles

La classe des modèles de séries temporelles à coefficients aléatoires a connu de grandes évolutions depuis les travaux de [Ham88, Ham89] et [NQ82]. Une grande partie des travaux dans la littérature était consacrée aux modèles ARMA à coefficients récurrents mais non périodiques (voir [AM98, AM06, BF03, Dah97] ainsi que [FR97, Ham90, HS94]). Les propriétés asymptotiques des modèles ARMA à changement de régime markovien ont été largement étudiées dans la littérature statistique (cf par exemple [BMR99, FR98, FZ01, FZ02, KK15] ou [Ham94]). On pourra se référer à [AM97, BL01] pour la classe de modèles ARMA univariés à coefficients périodiques.

Dans [A15] nous étudions les modèles ARMA(p, q) à coefficients aléatoires (RCARMA, pour random coefficient ARMA) de la forme :

$$X_t - \sum_{i=1}^p a_i^0(\Delta_t) X_{t-i} = \epsilon_t - \sum_{j=1}^q b_j^0(\Delta_t) \epsilon_{t-j}, \quad t \in \mathbb{Z}, \quad (5.1)$$

où $(\Delta_t)_{t \in \mathbb{Z}}$ est un processus stationnaire observé à valeurs discrètes dans un ensemble fini \mathcal{S} de dimension $\text{Card}(\mathcal{S}) = K$. Le processus des innovations linéaires $\epsilon := (\epsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc faible. Nous supposons que le processus (Δ_t) est indépendant de (ϵ_t) .

La représentation (5.1) sera dite RCARMA fort si ϵ est un bruit blanc fort et sera dite RCARMA dans le cas contraire.

Le paramètre d'intérêt est noté $\theta_0 := (a_i^0(s), b_j^0(s), i = 1, \dots, p, j = 1, \dots, q, s \in \mathcal{S})$ et appartient à l'intérieur de l'espace compact des paramètres

$$\Theta \subset \left\{ (a_i(s), b_j(s), i = 1, \dots, p, j = 1, \dots, q, s \in \mathcal{S}) \in \mathbb{R}^{(p+q) \times K} \right\}.$$

La variance σ^2 est considérée comme un paramètre de nuisance.

La représentation (5.1) a été étudiée par [FG04a, FG04b]. Ces auteurs ont donné des conditions pour la convergence et la normalité asymptotique pour une suite d'estimateur des moindres carrés et, également pour une suite d'estimateur des moindres carrés quasi-généralisés, sous l'hypothèse d'indépendance du terme d'erreur. Dans la pratique, cette hypothèse d'indépendance est souvent très restrictive. Par conséquent, nous relâchons cette hypothèse d'indépendance afin d'élargir le champs d'application des modèles RCARMA.

Nous introduisons les notations suivantes afin de faire apparaître le paramètre θ_0 dans la

représentation (5.1). Pour tout $\theta = (a_i(s), b_j(s), i = 1, \dots, p, j = 1, \dots, q, s \in \mathcal{S}) \in \Theta$, nous posons $\underline{a}_i := (a_i(s), s \in \mathcal{S}), i = 1, \dots, p$ et $\underline{b}_j := (b_j(s), s \in \mathcal{S}), j = 1, \dots, q$. Soit $\mathbf{e}(s)$ un vecteur ligne de taille $1 \times K$ tel que la i^{e} composante est $1_{[s=i]}$. Remarquons alors que $\forall t \in \mathbb{Z}$

$$a_i(\Delta_t) = \langle \mathbf{e}(\Delta_t), \underline{a}_i \rangle := g_i^a(\Delta_t, \theta), \quad b_j(\Delta_t) = \langle \mathbf{e}(\Delta_t), \underline{b}_j \rangle := g_j^b(\Delta_t, \theta), \quad i = 1, \dots, p, j = 1, \dots, q,$$

où $\langle \cdot, \cdot \rangle$ désigne le produit scalaire. Alors (5.1) devient

$$X_t - \sum_{i=1}^p g_i^a(\Delta_t, \theta) X_{t-i} = \epsilon_t - \sum_{j=1}^q g_j^b(\Delta_t, \theta) \epsilon_{t-j}. \quad (5.2)$$

Notons en plus, que les $g_i^a(s, \theta)$ et $g_j^b(s, \theta)$ sont linéaires en θ pour tout i, j et s .

5.1 Stationnarité du modèle RCARMA

Afin d'étudier l'existence de solutions stationnaires de (5.2), nous allons nous inspirer de [FZ01]. Nous commençons par introduire les matrices

$$A(s) := \begin{pmatrix} g_1^a(s, \theta) & \cdots & \cdots & g_p^a(s, \theta) \\ & & & 0 \\ & & & \vdots \\ & I_{p-1} & & 0 \end{pmatrix}, \quad B(s, \theta) := \begin{pmatrix} g_1^b(s, \theta) & \cdots & \cdots & g_q^b(s, \theta) \\ & & & 0 \\ & & & \vdots \\ & I_{q-1} & & 0 \end{pmatrix}$$

pour tout $s \in \mathcal{S}, \theta \in \Theta$ et, on pose

$$\epsilon_t(\theta) - \sum_{j=1}^q g_j^b(\Delta_t, \theta) \epsilon_{t-j}(\theta) = X_t - \sum_{i=1}^p g_i^a(\Delta_t, \theta) X_{t-i}, \quad \forall t \in \mathbb{Z}. \quad (5.3)$$

Nous introduisons les notations suivantes :

$$\begin{aligned} \mathbf{e}_1 &:= (1, 0, \dots, 0) \in \mathbb{R}^{p+q}, \\ \mathbf{e}_{p+1} &:= (\mathbf{e}_{p+1,i})_{i=1, \dots, p+q}, \quad \mathbf{e}_{p+1,1} = 1, \quad \mathbf{e}_{p+1,i} = 1_{[i=p+1]}, \quad i = 2, \dots, p+q, \\ M &:= (m_{ij})_{i,j=1, \dots, p+q}, \quad m_{ij} = 1_{[i=q+1, j=1 \text{ ou } i=1, j=1]}, \end{aligned}$$

$$\Phi(s, \theta) := \left(\begin{array}{c|ccc} B(s, \theta) & g_1^a(s, \theta) & \cdots & g_p^a(s, \theta) \\ \hline & 0 & \cdots & 0 \\ & & & \vdots \\ 0 & & I_{p-1} & 0 \end{array} \right), \quad s \in \mathcal{S}, \theta \in \Theta,$$

$$\psi(s) := \left(\begin{array}{c|ccc} & g_1^b(s, \theta_0) & \cdots & g_q^b(s, \theta_0) \\ \hline A(s) & & 0 & \\ \hline & 0 & \cdots & 0 \\ 0 & & I_{q-1} & \vdots \\ & & & 0 \end{array} \right), \quad s \in \mathcal{S}.$$

Nous énonçons le résultat suivant donnant la condition suffisante de stationnarité.

Proposition 16. *En supposant que*

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \ln \mathbb{E} \left(\sup_{\theta \in \Theta} \left\| \prod_{i=1}^t \phi(\Delta_i, \theta) \right\|^8 \right) < 0, \quad \limsup_{t \rightarrow \infty} \frac{1}{t} \ln \mathbb{E} \left(\left\| \prod_{i=1}^t \psi(\Delta_i) \right\|^8 \right) < 0, \quad (5.4)$$

alors pour tout $t \in \mathbb{Z}$ et $\theta \in \Theta$, l'unique solution stationnaire de (5.3) est donnée par

$$\epsilon_t(\theta) = \sum_{i=0}^{\infty} c_i(\theta, \Delta_t, \dots, \Delta_{t-i+1}) \epsilon_{t-i}, \quad \text{où} \quad (5.5)$$

$$c_i(\theta, \Delta_t, \dots, \Delta_{t-i+1}) = \sum_{k=0}^i e_1 \prod_{j=0}^{k-1} \phi(\Delta_{t-j}, \theta) M \prod_{j'=k}^{i-1} \psi(\Delta_{t-j'}) e'_{p+1}, \quad (5.6)$$

avec la convention $\prod_i^j = 1$ si $i > j$. De plus, pour chaque $t \in \mathbb{Z}$, la suite $(c_i(\theta, \Delta_t, \dots, \Delta_{t-i+1}))_{i \in \mathbb{N}}$ est unique dans l'ensemble

$$\left\{ (d_i)_{i \in \mathbb{N}} \text{ indépendante de } (\epsilon_t)_{t \in \mathbb{Z}} \text{ telle que } \mathbb{E} \left(\sum_{i=0}^{\infty} d_i^2 \right) < +\infty \right\}$$

des suites de variable aléatoire.

La décomposition (5.5) peut être vue comme une généralisation de la décomposition de Wold (voir le théorème 5.7.1 p.187 de [BD91]).

Remarque 5.1. *La condition (5.4) permet l'existence de régimes explosifs, qui est compatible avec la stricte stationnarité.*

5.2 Estimation du modèle RCARMA

En disposant d'observation $(X_1, \Delta_1), (X_2, \Delta_2), \dots, (X_n, \Delta_n)$ de longueur n , pour $0 < t \leq n$ et pour tout $\theta \in \Theta$, nous définissons récursivement les variables aléatoires $(e_t(\theta))$ par

$$e_t(\theta) - \sum_{j=1}^q g_j^b(\Delta_t, \theta) e_{t-j}(\theta) = \tilde{X}_t - \sum_{i=1}^p g_i^a(\Delta_t, \theta) \tilde{X}_{t-i}, \quad (5.7)$$

où les valeurs initiales inconnues sont remplacées par zéro et $\tilde{X}_t = X_t 1_{[t \geq 1]}$, $\forall t \in \mathbb{Z}$. Nous énonçons les résultats suivants.

Corollaire 3. *Le processus $(e_t(\theta))_{t \in \mathbb{Z}}$ défini par (5.7) peut être réécrit comme*

$$e_t(\theta) = \sum_{i=0}^{\infty} c_i^e(t, \theta, \Delta_t, \dots, \Delta_{t-i+1}) \epsilon_{t-i}, \quad t \geq p+1, \text{ où} \quad (5.8)$$

$$c_i^e(t, \theta, \Delta_t, \dots, \Delta_{t-i+1}) = \sum_{k=0}^{\min(t-1, i)} e_1 \prod_{j=0}^{k-1} \Phi(\Delta_{t-j}, \theta) M \prod_{j'=k}^{i-1} \Psi(\Delta_{t-j'}) e'_{p+1}. \quad (5.9)$$

Lemme 3. *Les coefficients aléatoires $c_i(\theta, \Delta_t, \dots, \Delta_{t-i+1})$, $i \in \mathbb{Z}$ et $t \in \mathbb{Z}$, vérifient les propriétés suivantes :*

- $\theta \mapsto c_i(\theta, \Delta_t, \dots, \Delta_{t-i+1})$, $\theta \mapsto \nabla [c_i(\theta, \Delta_t, \dots, \Delta_{t-i+1})]^2$ et $\theta \mapsto \nabla^2 [c_i(\theta, \Delta_t, \dots, \Delta_{t-i+1})]^2$ sont des fonctions polynômiales p.s.,
- on a

$$\begin{aligned} \limsup_{i \rightarrow \infty} \frac{1}{i} \ln \mathbb{E} \left(\sup_{\theta \in \Theta} [c_i(\theta, \Delta_i, \dots, \Delta_1)]^4 \right) &< 0, \\ \limsup_{i \rightarrow \infty} \frac{1}{i} \ln \mathbb{E} \left(\sup_{\theta \in \Theta} \left\| \nabla^j [c_i(\theta, \Delta_i, \dots, \Delta_1)] \right\|^4 \right) &< 0, \quad j = 2, 3. \end{aligned} \quad (5.10)$$

De plus, les coefficients $c_i^e(t, \theta, \Delta_{t-1}, \dots, \Delta_{t-i})$, $i \in \mathbb{Z}$, $t \geq 0$, satisfont

$$\begin{aligned} \limsup_{i \rightarrow \infty} \frac{1}{i} \ln \sup_{t \geq 0} \mathbb{E} \left(\sup_{\theta \in \Theta} [c_i^e(t, \theta, \Delta_t, \dots, \Delta_{t-i+1})]^4 \right) &< 0, \\ \limsup_{i \rightarrow \infty} \frac{1}{i} \ln \sup_{t \geq 0} \mathbb{E} \left(\sup_{\theta \in \Theta} \left\| \nabla^j [c_i^e(t, \theta, \Delta_t, \dots, \Delta_{t-i+1})] \right\|^4 \right) &< 0, \quad j = 2, 3. \end{aligned} \quad (5.11)$$

On dit que $\hat{\theta}_n$ est un estimateur des moindres carrés si (p.s.)

$$Q_n(\hat{\theta}_n) = \min_{\theta \in \Theta} Q_n(\theta) \text{ où } Q_n(\theta) = \frac{1}{n} \sum_{t=1}^n e_t^2(\theta).$$

Pour établir les propriétés asymptotiques de cet estimateur, nous avons besoin des hypothèses suivantes.

- (H1) Le processus $(\epsilon_t)_{t \in \mathbb{Z}}$ est stationnaire et ergodique.
- (H2) Pour un réel $\nu > 0$, les processus $(\epsilon_t)_{t \in \mathbb{Z}}$ et $(\Delta_t)_{t \in \mathbb{Z}}$ vérifient $\sum_{h=0}^{\infty} \alpha_\epsilon(h)^{\frac{\nu}{\nu+2}} < +\infty$ et $\sum_{h=0}^{\infty} \alpha_\Delta(h)^{\frac{\nu}{\nu+2}} < +\infty$.
- (H3) Le processus $(\epsilon_t)_{t \in \mathbb{Z}}$ satisfait également $\mathbb{E}[|\epsilon_t|^{\nu+4}] < +\infty$.

La proposition suivante que $Q_n(\theta)$ et $O_n(\theta) := n^{-1} \sum_{t=1}^n \epsilon_t^2(\theta)$ ont la même distribution asymptotique.

Proposition 17. *Pour tout $\alpha \in (0, 1)$, nous avons*

1. $\sup_{\theta \in \Theta} |Q_n(\theta) - O_n(\theta)|$ converge vers 0 p.s., et $n^\alpha \|\sup_{\theta \in \Theta} |Q_n(\theta) - O_n(\theta)|\|_1$ tend vers 0 quand $n \rightarrow \infty$,
2. $\sup_{\theta \in \Theta} \|\nabla(Q_n(\theta) - O_n(\theta))\|$ et $\sup_{\theta \in \Theta} \|\nabla^j(Q_n(\theta) - O_n(\theta))\|$, pour $j = 2, 3$ convergent vers 0 p.s.,
3. $n^\alpha \|\sup_{\theta \in \Theta} |\nabla(Q_n - O_n)(\theta)|\|_1 \rightarrow 0$ quand $n \rightarrow \infty$.

Nous énonçons dans les théorèmes suivants la consistance et la normalité asymptotique de l'estimateur des moindres carrés.

Théorème 19 (Consistance). *Soit $\hat{\theta}_n$ une suite d'estimateur des moindres carrés. Sous l'hypothèse (H1), nous avons p.s. $\hat{\theta}_n \rightarrow \theta_0$ quand $n \rightarrow \infty$.*

Théorème 20 (Normalité asymptotique). *Sous les hypothèses (H1), (H2) et (H3), nous avons*

$$\sqrt{n} \left(\hat{\theta}_n - \theta_0 \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \Omega := J^{-1} I J^{-1} \right), \quad n \rightarrow +\infty, \quad (5.12)$$

où

$$J := J(\theta_0) = 2\mathbb{E} \left(\frac{\partial \epsilon_t(\theta_0)}{\partial \theta} \frac{\partial \epsilon_t(\theta_0)}{\partial \theta'} \right) \text{ et } I := I(\theta_0) = 4 \sum_{k=-\infty}^{\infty} \mathbb{E} \left(\epsilon_t(\theta_0) \epsilon_{t-k}(\theta_0) \frac{\partial \epsilon_t(\theta_0)}{\partial \theta} \frac{\partial \epsilon_{t-k}(\theta_0)}{\partial \theta'} \right).$$

Pour estimer ses matrices d'information, on utilise les mêmes démarches introduites dans le chapitre 1.

Chapitre 6

Autre thématique étudiée

Je décris brièvement dans ce chapitre des travaux indépendants de ma thématique principale décrite dans les chapitres 1 à 5.

La variance généralisée est le déterminant de la matrice de covariance. Elle est une mesure de variabilité à l'intérieur d'une population multivariée. Très bien connue sous l'hypothèse gaussienne (voir par exemple [Jør13]), la variance généralisée demeure peu maniable dans la jungle de la non-normalité multivariée.

Les familles exponentielles naturelles (FENs dans la suite) constituent une part importante de la statistique théorique. Depuis plusieurs décennies, elles sont largement étudiées et l'ouvrage de [BN78] en donne une introduction détaillée. Les familles exponentielles de dispersion ont aussi été largement étudiées (voir l'ouvrage de [KBJ00] ainsi que [Jr97]). Dans le cadre des modèles exponentiels naturels, ou de dispersion, la matrice de covariance exprimée en fonction de la moyenne $\mathbf{m} = (m_1, \dots, m_k)^\top$ s'appelle la fonction variance, dont la plus courante est de la forme quadratique $\mathbf{V}_F(\mathbf{m}) = \mathbf{A}(\mathbf{m}, \mathbf{m}) + \mathbf{B}(\mathbf{m}) + \mathbf{C}$, où $\mathbf{A}(\mathbf{m}, \mathbf{m})$, $\mathbf{B}(\mathbf{m})$ et \mathbf{C} sont respectivement des matrices réelles symétriques bilinéaire, linéaire et constante en $\mathbf{m} \in \mathbf{M}_F \subseteq \mathbb{R}^k$ (voir [Let89]). Le cas particulier des fonctions variance quadratiques simples avec $\mathbf{A}(\mathbf{m}, \mathbf{m}) = a \cdot \mathbf{m}\mathbf{m}^\top$, pour $a \in \mathbb{R}$, a été complètement étudié (voir par exemple [Cas96]). Cependant, leurs fonctions variances généralisées $\det(\mathbf{V}_F(\mathbf{m}))$ ne sont explorées qu'en partie.

Un cas très particulier de ces familles, générant un modèle de dispersion exponentielle, est celui des modèles Tweedie, historiquement introduits par [Twe84] et étudiés ensuite par [Jr97], qui sont une généralisation des lois α -stables. Leur fonction variance est bien spécifique et est donnée par $\mathbf{V}_F(\mathbf{m}) = \mathbf{m}^p$ avec $p \in]-\infty, 0] \cup [1, +\infty[$. Les lois Tweedie interviennent dans un nombre important de champs d'application, comme l'écologie, les sciences actuarielles, le contrôle qualité, ou encore la physique.

Nous introduisons dans [A8] une expression de la fonction variance généralisée qui n'est autre qu'une puissance réelle d'une seule composante m_j du vecteur moyenne \mathbf{m} . Pour y aboutir, nous construisons d'abord les modèles appropriés à partir des lois univariées stables-Tweedie ayant des fonctions variance puissances, puis nous explicitons les fonctions variance correspondantes, lesquelles présentent une nouvelle forme pour les modèles exponentiels. Enfin, nous proposons des estimateurs sans biais de variance minimum pour les variances généralisées de cette famille que nous appellerons *normale stable Tweedie* (NST).

6.1 Loi univariée stable-Tweedie

Une loi univariée stable-Tweedie positive est définie par

$$\mu_{\alpha,t}(dx) = \begin{cases} t^x(x!)^{-1} \exp(-t) \mathbb{I}_{x \in \mathbb{N}} = \mu_{-\infty,t}(x) \delta_x & \text{si } \alpha = -\infty \\ \delta_0(dx) + \frac{dx}{x} \sum_{j=1}^{\infty} \frac{t^j(\alpha-1)^j}{j! \alpha^j \Gamma(-\alpha j) [(1-\alpha)x]^{\alpha j}} \mathbb{I}_{x>0} = \mu_{\alpha,t}(x) dx & \text{si } \alpha < 0 \\ x^{t-1} [\Gamma(t)]^{-1} \exp(-t) \mathbb{I}_{x>0} dx = \mu_{0,t}(x) dx & \text{si } \alpha = 0 \\ \frac{1}{\pi x} \sum_{j=1}^{\infty} \frac{t^j \Gamma(1+\alpha j) \sin(-j\pi\alpha)}{j! \alpha^j (\alpha-1)^{-j} [(1-\alpha)x]^{\alpha j}} \mathbb{I}_{x>0} dx = \mu_{\alpha,t}(x) dx & \text{si } 0 < \alpha < 1, \end{cases}$$

où $\alpha \in [-\infty, 1[$ est, le paramètre d'indice, $t > 0$ est un paramètre temporel (du processus de Lévy associé) ou d'échelle, $\Gamma(\cdot)$ représente la fonction gamma classique, et \mathbb{I}_A désigne la fonction indicatrice de l'ensemble A . Nous introduisons le paramètre puissance p , défini par

$$(p-1)(1-\alpha) = 1 \quad (6.1)$$

ce qui est équivalent à $p = p(\alpha) = 1 + 1/(1-\alpha)$ ou $\alpha = \alpha(p) = 1 + 1/(1-p)$, et la fonction variance unité (avec $t = 1$) s'écrit simplement : $V_p(m) = m^p$, $p \in [1, +\infty[$, $m \in M_p \subseteq]0, +\infty[$. La transformée de Laplace de $\mu_{\alpha,t}$ peut être écrite comme suit

$$L_{\mu_{\alpha,t}}(\theta_0) = \int_{\mathbb{R}} \exp(\theta_0 x) \mu_{\alpha,t}(dx) = (L_{\mu_{\alpha,1}}(\theta_0))^t =: (L_{\mu_p}(\theta_0))^t \quad (6.2)$$

avec $\mu_{\alpha,1} = \mu_p$ et $\theta_0 \in \Theta(\mu_{\alpha,1})$ tel que

$$\Theta(\mu_{\alpha,1}) = \begin{cases} \mathbb{R} & \text{for } p = 0, 1 \\ [0, \infty) & \text{for } p < 0 \text{ or } 0 < p < 1 \\ (-\infty, 0) & \text{for } 1 < p \leq 2 \text{ or } p = \infty \\ (-\infty, 0] & \text{for } 2 < p < \infty. \end{cases} \quad (6.3)$$

La fonction cumulante associée $K_{\mu_{\alpha,t}}(\theta_0) = \log L_{\mu_{\alpha,t}}(\theta_0) = tK_{\mu_{\alpha,1}}(\theta_0) = tK_{\mu_p}(\theta_0)$ est simplement donnée, via (6.1), par

$$K_{\mu_{\alpha,1}}(\theta_0) = \begin{cases} \exp(\theta_0) & \text{for } p = 1 \\ -\log(-\theta_0) & \text{for } p = 2 \\ [(\alpha-1)/\alpha] [\theta_0/(\alpha-1)]^\alpha & \text{for } p \neq 1, 2. \end{cases} \quad (6.4)$$

6.2 Modèles NST

Maintenant, nous introduisons la famille des mesures positives σ -finies $\nu_{\alpha,t}$ sur \mathbb{R}^k et génératrices des NST de la manière suivante : pour $\mathbf{x} = (x_1, \dots, x_k)^\top$, on définit

$$\nu_{\alpha,t}(d\mathbf{x}) = \mu_{\alpha,t}(dx_1) \prod_{j=2}^k \mu_{2,x_1}(dx_j), \quad (6.5)$$

où μ_{2,x_1} est la loi normale (univariée) centrée et de variance x_1 . Pour $p = p(\alpha) \geq 1$ (i.e. $\alpha = \alpha(p) \in [-\infty, 1[$) et $t > 0$, le modèle exponentiel multivarié engendré par $\nu_{\alpha,t}$ de (6.5)

est l'ensemble $\mathbf{G}_{p,t} = \mathbf{G}(\nu_{\alpha,t}) = \{\mathbf{P}(\theta; \alpha, t); \theta = (\theta_1, \dots, \theta_k)^\top \in \Theta_{\alpha,t}\}$ des lois de probabilité

$$\mathbf{P}(\theta; \alpha, t)(d\mathbf{x}) = \exp\left\{\theta^\top \mathbf{x} - \mathbf{K}_{\nu_{\alpha,t}}(\theta)\right\} \nu_{\alpha,t}(d\mathbf{x}).$$

La fonction cumulante $\mathbf{K}_{\nu_{\alpha,t}}(\theta) = \log \int_{\mathbb{R}^k} \exp(\theta^\top \mathbf{x}) \nu_{\alpha,t}(d\mathbf{x})$ est finie pour tout θ dans le domaine canonique $\Theta_{\alpha,t} := \Theta(\nu_{\alpha,t})$. On montre alors que

$$\mathbf{K}_{\nu_{\alpha,t}}(\theta) = tK_{\mu_{\alpha,1}}\left(\theta_1 + \frac{1}{2} \sum_{j=2}^k \theta_j^2\right) \quad (6.6)$$

avec $\Theta_{\alpha,1} = \left\{\theta \in \mathbb{R}^k; \left(\theta_1 + \sum_{j=2}^k \theta_j^2/2\right) \in \Theta(\mu_{\alpha,1})\right\}$. Par le vecteur moyenne $\mathbf{m} := \mathbf{K}'_{\nu_{\alpha,t}}(\theta)$, la fonction variance généralisée peut être obtenue par $\det \mathbf{K}''_{\nu_{\alpha,t}}(\theta(\mathbf{m}))$, où $\mathbf{K}'_{\nu_{\alpha,t}}$ et $\mathbf{K}''_{\nu_{\alpha,t}}$ sont respectivement des dérivées première et seconde.

La table 6.1 résume les résultats et les différentes sous familles de NST dont une partie est déjà apparue dans la littérature (e.g. financière, actuarielle, bayésienne).

TABLE 6.1 – Récapitulatif des modèles NST k -dimensionnels, de paramètre puissance $p \geq 1$, de paramètre puissance généralisée $q = k + p - 1$, de paramètre de mesure de Lévy modifiée $\eta = [p(k-1) + 1]/(p-1)$ et de support de lois \mathbf{S}_p .

Type(s)	$p = p(\alpha)$	$q = q(p, k)$	$\eta = \eta(p, k)$	\mathbf{S}_p
Normal Poisson	$p = 1$	$q = k$	$\eta = +\infty$	$\mathbb{N} \times \mathbb{R}^{k-1}$
Normal Poisson composé	$1 < p < 2$	$k < q < k + 1$	$\eta > 2k - 1$	$\mathbb{R}_+ \times \mathbb{R}^{k-1}$
Normal gamma décentré	$p = 1,5$	$q = k + 0.5$	$\eta = 3k - 1$	$\mathbb{R}_+ \times \mathbb{R}^{k-1}$
Normal gamma	$p = 2$	$q = k + 1$	$\eta = 2k - 1$	$\mathbb{R}_+^* \times \mathbb{R}^{k-1}$
Normal stable positif	$p > 2$	$q > k + 1$	$0 < \eta < 2k - 1$	$\mathbb{R}_+^* \times \mathbb{R}^{k-1}$
Normal gaussien inverse	$p = 3$	$q = k + 2$	$\eta = 1,5k - 1$	$\mathbb{R}_+^* \times \mathbb{R}^{k-1}$

Dans tout ce qui suit nous fixons la dimension de l'espace égale à $k \in \{2, 3, \dots\}$. Notons par $(\mathbf{e}_i)_{i=1, \dots, k}$ une base orthonormale de \mathbb{R}^k et par $\mathbf{Diag}(d_1, \dots, d_k)$ la matrice diagonale d'ordre $k \times k$. Pour $\mathbf{a} = (a_1, \dots, a_k)^\top \in \mathbb{R}^k$ et $\mathbf{b} = (b_1, \dots, b_k)^\top \in \mathbb{R}^k$, nous utilisons les notations $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^\top \mathbf{b}$ et $\mathbf{a} \otimes \mathbf{b} = \mathbf{a} \mathbf{b}^\top$ pour désigner le scalaire $\sum_{j=1}^k a_j b_j$ et la matrice symétrique $(a_j b_j)_{j,j=1, \dots, k}$ respectivement. Avec ces notations, pour un réel $\gamma \neq 0$ et \mathbf{A} une matrice $k \times k$ alors la représentation suivante du déterminant sera utile :

$$\det \begin{pmatrix} \gamma & \mathbf{b}^\top \\ \mathbf{a} & \mathbf{A} \end{pmatrix} = \gamma \det(\mathbf{A} - \gamma^{-1} \mathbf{a} \otimes \mathbf{b}). \quad (6.7)$$

Nous énonçons alors les résultats relatifs aux modèles NST.

Théorème 21. Soit $p = p(\alpha) \geq 1$ et $t > 0$. Alors la fonction variance généralisée de $\mathbf{G}_{p,t} = \mathbf{G}(\nu_{\alpha,t})$ engendré par $\nu_{\alpha,t}$ de (6.5) est

$$\det \mathbf{V}_{\mathbf{G}_{p,t}}(\mathbf{m}) = t^{1-p} (\langle \mathbf{e}_1, \mathbf{m} \rangle)^{k+p-1} = t^{1-p} m_1^{k+p-1},$$

pour $\mathbf{m} \in \mathbf{M}_{\mathbf{G}_{p,t}} =]0, +\infty[\times \mathbb{R}^{k-1}$.

Démonstration. Pour la démonstration, on utilise (6.7) et la proposition suivante, qui découle de (6.6). \square

Proposition 18. Soit $p = p(\alpha) \geq 1$ et $t > 0$. Alors la fonction variance de $\mathbf{G}_{p,t} = \mathbf{G}(\nu_{\alpha,t})$ est

$$\mathbf{V}_{\mathbf{G}_{p,t}}(\mathbf{m}) = t^{1-p} (\langle \mathbf{e}_1, \mathbf{m} \rangle)^{p-2} \cdot \mathbf{m} \otimes \mathbf{m} + \langle \mathbf{e}_1, \mathbf{m} \rangle \cdot \mathbf{Diag}(0, 1, \dots, 1), \quad (6.8)$$

pour tout $\mathbf{m} = (m_1, \dots, m_k)^\top$ dans $\mathbf{M}_{\mathbf{G}_{p,t}} =]0, +\infty[\times \mathbb{R}^{k-1}$.

Puisque toutes les lois de $\mathbf{G}_{p,t} = \mathbf{G}(\nu_{\alpha,t})$ sont indéfiniment divisibles, le résultat suivant fournit les mesures de Lévy modifiées $\rho(\nu_{\alpha,t})$ satisfaisant $\mathbf{K}_{\rho(\nu_{\alpha,t})}(\theta) = \log \det \mathbf{K}_{\nu_{\alpha,t}}''(\theta)$.

Théorème 22. Soit $\nu_{\alpha(p),t}$ une mesure génératrice (6.5) de NST pour $p = p(\alpha) \geq 1$ et $t > 0$ donnés. On note $\eta(p, k) = [p(k-1) + 1] / (p-1) \in]0, +\infty[$ le paramètre de mesure de Lévy modifiée. Alors

$$\rho(\nu_{\alpha(p),t}) = \begin{cases} t^k [1 - \alpha(p)]^{\eta(p,k)} \cdot \nu_{0,\eta(p,k)} & \text{pour } p > 1 \\ t^k \cdot \nu_{-\infty,k} & \text{pour } p = 1. \end{cases}$$

Voici le comportement de $\rho(\nu_{\alpha(p),t})$ en terme de fonction variance laquelle ne dépend plus du paramètre temporel $t > 0$.

Corollaire 4. Soit $\mathbf{G}'_p = \mathbf{G}(\rho(\nu_{\alpha(p),t}))$ avec $p \geq 1$, $t > 0$ et $\eta(p, k)$ du Théorème 22. Alors

$$\mathbf{V}_{\mathbf{G}'_p}(\bar{\mathbf{m}}) = \begin{cases} \frac{1}{\eta(p,k)} \cdot \bar{\mathbf{m}} \otimes \bar{\mathbf{m}} + \langle \mathbf{e}_1, \bar{\mathbf{m}} \rangle \cdot \mathbf{Diag}(0, 1, \dots, 1) & \text{pour } p > 1 \\ (\langle \mathbf{e}_1, \bar{\mathbf{m}} \rangle)^{-1} \cdot \bar{\mathbf{m}} \otimes \bar{\mathbf{m}} + \langle \mathbf{e}_1, \bar{\mathbf{m}} \rangle \cdot \mathbf{Diag}(0, 1, \dots, 1) & \text{pour } p = 1, \end{cases}$$

avec $\bar{\mathbf{m}} \in]0, +\infty[\times \mathbb{R}^{k-1}$.

La caractérisation des $\nu_{\alpha(p),t}$ par leur mesure de Lévy modifiée dans le Théorème 22 fournit des solutions à l'équation de Monge-Ampère suivante :

$$\det \mathbf{K}''(\theta) = \exp\{a\mathbf{K}_{\nu_{0,\eta}}(\theta) + b\mathbf{K}_{\nu_{-\infty,k}}(\theta) + c\}, \quad (6.9)$$

où \mathbf{K} est une fonction cumulante à déterminer pour des valeurs fixées de a , b et c dans \mathbb{R} .

Enfin, nous terminons par deux méthodes d'estimation des variances généralisées dans les modèles NST. Soit $\mathbf{X}_1, \dots, \mathbf{X}_n$ des vecteurs aléatoires iid. de loi $\mathbf{P}(\mathbf{m}, \mathbf{G}_{p,t}) \in \mathbf{G}(\nu_{\alpha,t})$ dans la famille NST, i.e. pour $p = p(\alpha) \geq 1$ et $t > 0$ fixés. On note $\bar{\mathbf{X}} = (\mathbf{X}_1 + \dots + \mathbf{X}_n) / n = (\bar{X}_1, \dots, \bar{X}_k)^\top$ le vecteur moyenne empirique dont la première composante est \bar{X}_1 , alors l'estimateur du maximum de vraisemblance de la variance généralisée $\det \mathbf{V}_{\mathbf{G}_{p,t}}(\mathbf{m})$ est facilement obtenu, via Théorème 21 pour des modèles exponentiels, par

$$T_{n;k;p,t} = \det \mathbf{V}_{\mathbf{G}_{p,t}}(\bar{\mathbf{X}}) = t^{1-p} (\bar{X}_1)^{k+p-1}.$$

Pour tout $p \geq 1$, $T_{n;k;p,t}$ est un estimateur biaisé de $\det \mathbf{V}_{\mathbf{G}_{p,t}}(\mathbf{m}) = t^{1-p} (\langle \mathbf{e}_1, \mathbf{m} \rangle)^{k+p-1} = t^{1-p} m_1^{k+p-1}$ avec un risque quadratique difficilement calculable en général.

Pour contourner cette absence de bonnes propriétés de $T_{n;k;p,t}$ à taille finie, nous cherchons directement l'estimateur *sans biais de variance minimum* $U_{n;k;p,t}$ de $\det \mathbf{V}_{\mathbf{G}_{p,t}}(\mathbf{m})$. Ce dernier est

obtenu pour tout entier $n > k$:

$$U_{n;k;p,t} = C_{n;k;p,t}(n\bar{\mathbf{X}}),$$

où $C_{n;k;p,t} : \mathbb{R}^k \rightarrow]0, +\infty[$ vérifie $\nu_{n;k;p,t}(d\mathbf{x}) = C_{n;k;p,t}(\mathbf{x})\nu_{\alpha,nt}(d\mathbf{x})$ de telle sorte que $\nu_{n;k;p,t}$ soit la mesure transportée de

$$\frac{1}{(k+1)!} \left(\det \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_{k+1} \end{bmatrix} \right)^2 \nu_{\alpha,t}(d\mathbf{x}_1) \cdots \nu_{\alpha,t}(d\mathbf{x}_n)$$

par l'application $(\mathbf{x}_1, \dots, \mathbf{x}_n) \mapsto \mathbf{x}_1 + \cdots + \mathbf{x}_n$. Le dernier résultat montre le chemin pour exprimer $C_{n;k;p,t}(\cdot)$ dont nous avons besoin pour calculer numériquement l'estimation issue de $U_{n;k;p,t}$; il montre également que cet estimateur ne dépend que de la première composante univariée stable-Tweedie.

Théorème 23. Soit $\mathbf{X}_1, \dots, \mathbf{X}_n$ des vecteurs aléatoires i.i.d. de loi $\mathbf{P}(\mathbf{m}, \mathbf{G}_{p,t}) \in \mathbf{G}(\nu_{\alpha,t})$ de NST, i.e. $p = p(\alpha) \geq 1$, $t > 0$ et $\eta(p, k) \in]0, +\infty[$ associé à la mesure de Lévy modifiée $\rho(\nu_{\alpha(p),t})$ par le Théorème 22. Alors

$$C_{n;k;p,t}(\mathbf{x}) = \frac{\nu_{\alpha(p),nt} * \rho(\nu_{\alpha(p),t})(d\mathbf{x})}{\nu_{\alpha(p),nt}(d\mathbf{x})}.$$

En particulier, on a : $C_{n;k;p,t}(\mathbf{x}) =$

$$\left\{ \begin{array}{ll} (te^{-1})^k [1 + k(nt)^{-1}]^{x_1} & \text{pour } p = 1 \\ t^k \Gamma(nt) [\Gamma(nt + 2k - 1)]^{-1} x_1^{2k-1} & \text{pour } p = 2 \\ \frac{t^k 2^{1-3k/2}}{\Gamma(3k/2-1)} x_1^{3/2} \exp \left\{ \left[(nt)^2 + \sum_{j=2}^k x_j^2 \right] (2x_1)^{-1} + 1 - 3k/2 \right\} & \text{pour } p = 3 \\ \times \int_0^{x_1} y_1^{1.5k-2} (x_1 - y_1)^{-3/2} \exp \left[-(nt)^2 / 2(x_1 - y_1) \right] dy_1 & \\ t^k [1 - \alpha(p)]^{\eta(p,k)} \exp \left[(2x_1)^{-1} \sum_{j=2}^k x_j^2 \right] [\mu_{\alpha(p),nt}(x_1)]^{-1} & \text{pour } p \neq 1, 2, 3. \\ \times \int_0^{x_1} \mu_{\alpha(p),nt}(x_1 - y_1) \mu_{0,\eta(p,k)}(y_1) dy_1 & \end{array} \right.$$

6.3 Conclusion

Nous avons décrit (voir Table 6.1) la nouvelle classe des modèles k -variés NST, de fonctions variances puissances simples p , et de fonctions variances généralisées puissances univariées $q = q(p, k)$. Les NST ont des propriétés intéressantes pour la pratique et admettant une certaine flexibilité à l'usage.

Chapitre 7

Encadrement, perspectives et responsabilités en recherche

Nous résumons dans ce chapitre quelques travaux en cours ou des projets de recherche à explorer qui sont dans la continuité de ceux présentés dans ce mémoire. Nous donnons également quelques détails sur des encadrements d'étudiants dans le cadre de mes recherches, ainsi que d'autres aspects, tels que l'implication dans des projets et diverses responsabilités liées à la recherche.

7.1 Encadrement et co-encadrement d'étudiants

J'encadre régulièrement des mémoires de recherche, des projets et des stages d'étudiants en Master 1 et 2 modélisation statistique. J'encade et co-encade également les thèses suivantes :

- 1) **Co-encadrement à 50%** (depuis octobre 2014), avec Saussereau Bruno, de la thèse d'université d'Othman KADMIRI sur les modèles GARCH multivariés à seuil en puissance. Bourse de l'Université de Franche-Comté. Le chapitre 4 présenté dans ce mémoire est issu de ce co-encadrement.
- 2) **Encadrement à 100%** (depuis octobre 2015) de la thèse d'université de Abdoukarim ILMI AMIR sur les modèles ARMA faibles saisonniers et/ou périodiques. Dans [A15], nous avons proposé des tests d'adéquation de modèles ARMA faibles saisonniers. Dans [CR5], nous envisageons de compléter les travaux sur les modèles PARMA périodiques, dont les propriétés asymptotiques des estimateurs ont été déjà établies par [FRS11], en abordant le problème de la validation de ces modèles afin de comparer nos résultats à ceux existant sur les PARMA forts dans la littérature (voir par exemple [McL94, McL95, UD09]). Ensuite, le travail serait d'étendre les travaux de [UD09] au cas mixte des modèles SPARMA (seasonal PARMA) faibles. Cette thèse est à mi-temps, en effet l'étudiant passe 3 à 4 mois maximum chaque année au Laboratoire de Mathématiques de Besançon pour travailler sur la thèse, et une fois rentré, il se consacre à sa charge d'enseignement à l'université de Djibouti. Bourse du gouvernement djiboutien.
- 3) **Co-encadrement à 50%** (depuis octobre 2016), avec Saussereau Bruno, de la thèse d'université de Youssef ESSTAFI sur les modèles FARIMA avec des erreurs dépendantes. Nous en donnerons plus de détails dans la section suivante sur les travaux en cours.

7.2 Travaux en cours et perspectives

Le problème qui nous préoccupera à court terme sera de développer l'analyse statistique pour les modèles ARMA et VARMA faibles stationnaires et/ou non stationnaires et/ou à mémoire longue. À long terme, il s'agira d'étudier des modèles ARMA et VARMA faibles en temps continu.

7.2.1 Modèles GARCH multivariés à seuil en puissance

Dans la continuité de la thèse d'Othman Kadmiri, nous souhaitons étendre dans [CR3] le test de validation univarié (pour $m = 1$) proposé dans [A14] au cadre des modèles CCC-APGARCH 4.1, quand le vecteur des puissances δ_0 est soit connu soit inconnu. Pour l'instant, nous nous sommes limités au cas de modèles GARCH à corrélations conditionnelles constantes. Il serait plus intéressant de considérer une corrélation R_t qui évolue dans le temps, contrairement à ce que nous avons fait dans [A13, CR3]. Nous envisageons donc d'étendre [A13, CR3] à des modèles MGARCH asymétriques à corrélations conditionnelles dynamiques.

7.2.2 Modèles VARMA faibles stationnaires et/ou non stationnaires

Estimation des modèles VARMA faibles avec tendance déterministe

Dans [CR1], nous essayons d'étendre l'étude des propriétés asymptotiques de l'estimateur du QMV établi dans [A2], l'estimation de la matrice de variance asymptotique de ces estimateurs, ainsi que les tests asymptotiques de Wald, du multiplicateur de Lagrange et du rapport de vraisemblance (introduits dans [A7, A9]) 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 de l'estimateur du QMV ne peuvent pas être calculées de la même manière que celles des processus stationnaires (voir [A2]). La deuxième difficulté est liée au fait que les estimations des paramètres de ces modèles avec tendance déterministe ont des vitesses de convergence asymptotique différentes. Par conséquent, il nous faudra des conditions supplémentaires afin d'établir la consistance et la normalité asymptotique de l'EQMV. Plus précisément, le théorème central limite pour des processus α -mélangeants de [Her84] ne s'appliquera plus, car les composantes du processus vectoriel n'ont pas forcément la même loi; il nous faudrait un TCL pour un tableau triangulaire de processus α -mélangeants comme celui introduit par [FZ05a].

Dans la continuité de la thèse de Abdoulkarim Ilmi Amir, on envisage d'étendre ce travail aux modèles VARMA faibles saisonniers et/ou périodiques avec ou sans tendance déterministe. Le cas des tendances stochastiques est également très intéressant à aborder pour les applications.

Tests d'adéquation de modèles VARMA faibles co-intégrés

Considérons des séries temporelles multivariées non stationnaires. La théorie de la co-intégration postule que, dans certains cas, il est possible de former à partir de combinaisons linéaires indépendantes de ces séries temporelles multivariées non stationnaires, des processus de dimension réduite mais stationnaires. Le nombre de combinaisons linéaires indépendantes est appelé le rang de co-intégration. Ces combinaisons linéaires sont les relations de co-intégration ou relations de long terme. L'objectif serait d'étudier le rang et les relations de co-intégration dans le cadre des VARMA faibles. Une voie de recherche future est d'étudier l'estimation de modèles VARMA faibles

co-intégrés. Puis d'étendre les tests portmanteau modifiés que nous avons proposés dans [A1, A3, A12] pour tester les ordres p et q des modèles VARMA(p, q) faibles co-intégrés.

Autres méthodes de sélection des ordres de modèles VARMA faibles

En dehors des méthodes basées sur la minimisation d'un critère d'information (comme dans [A4, A6, A11]), il existe aussi d'autres méthodes d'identification de modèles ARMA basés sur l'examen des corrélogrammes, par exemple. Dans [CR4], nous nous intéressons à une méthode d'identification dite méthode du coin, afin d'identifier les ordres de modèles ARMA(p, q) et VARMA(p, q) faibles. Cette méthode est basée sur l'examen du comportement des résidus, où les ordres p et q du modèle sont caractérisés par un coin de 0 dans un tableau (d'où le nom de la méthode). Les éléments de ce tableau sont des autocorrélations obtenues de façon récursive.

Ensuite, on envisage d'aborder d'autres méthodes d'identification des ordres de modèles ARMA(p, q) et VARMA(p, q) faibles, cette fois basées sur l'examen des corrélogrammes et corrélogrammes partiels comme dans [FZ05b] ou basées sur l'analyse des autocorrélations inverses comme dans [EGF06].

7.2.3 Modèles à mémoires longues

Tous les travaux que nous avons développés jusqu'à présent sont des modèles de séries temporelles à temps discret avec erreurs faibles, dont la série des covariances est convergente. Ces types de modèles sont souvent appelés modèles à mémoire courte.

La question qui mériterait d'être examinée est : qu'en est-il dans le cas des modèles ARMA faibles à mémoires longues (c'est-à-dire quand la somme des covariances est divergente) ?

Nous tentons d'apporter les premières réponses à cette question dans le cadre de la thèse de Youssef Esstafa débutée en octobre 2016. Dans un premier temps, nous nous intéressons à l'estimation. Ensuite, nous accorderons une attention particulière à la validation et à l'identification de ces modèles ARMA faibles à mémoires longues.

Estimation de modèles FARIMA faibles

Les processus à mémoire longue occupent une place de plus en plus importante dans la littérature des séries temporelles. Pour modéliser le comportement de mémoire longue, plusieurs modèles peuvent être utilisés. Les processus FARIMA (Fractionally AutoRegressive Integrated Moving Average) sont parmi les modèles les plus connus et les plus utilisés. Ils sont une généralisation des modèles ARIMA (AutoRegressive Integrated Moving Average), pour lesquels le paramètre de différenciation est un entier. Les processus FARIMA permettent au paramètre de différenciation de prendre des valeurs réelles.

On dit qu'un processus stationnaire au second ordre $(X_t)_{t \in \mathbb{Z}}$, à valeurs réelles, admet une représentation FARIMA(p, d_0, q) faible si pour tout $t \in \mathbb{Z}$,

$$a(L)(1-L)^{d_0}X_t = b(L)\epsilon_t,$$

où $(\epsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc faible. Les polynômes $a(L) = 1 + \sum_{i=1}^p a_i L^i$ et $b(L) = 1 + \sum_{j=1}^q b_j L^j$ (avec $p, q \in \mathbb{N}$ et $a_1, \dots, a_p, b_1, \dots, b_q \in \mathbb{R}$) ont leurs racines en dehors du disque unité et n'ont aucun zéro en commun, et $0 < d_0 < 1/2$. L'opérateur de différenciation fractionnaire, $(1-L)^{d_0}$, est défini en utilisant la formule du binôme généralisée.

Dans [CR2], nous établissons la convergence forte et la normalité asymptotique des paramètres des modèles FARIMA faibles. Une attention particulière a été consacrée à l'estimation de la matrice de variance asymptotique, qui peut être là aussi très différente de celle obtenue dans le cas des FARIMA forts.

Validation et identification de modèles FARIMA faibles

Nous envisageons de développer des tests de validation pour les modèles FARIMA faibles. Ces tests sont 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 proposerons donc des tests portmanteau modifiés pour tester l'adéquation de modèles FARIMA faibles. Ce travail est en cours de finalisation.

Ensuite, nous nous intéresserons aux choix des ordres des modèles FARIMA faibles. Ce choix sera fondé sur la minimisation d'un critère d'information.

7.2.4 Perspectives à long terme

Tous les travaux que nous avons développés jusqu'à présent se limitent, souvent, à des modèles stationnaires de séries temporelles à temps discret avec erreurs faibles. Nous envisageons d'étendre ces travaux au cadre non stationnaire.

Les modèles en temps continu occupent une place centrale en finance mathématique. La plupart des développements théoriques concernant la valorisation de produits dérivés sont effectués à partir de processus en temps continu obtenus comme solutions d'équations de diffusion. Une seconde voie de recherche serait donc d'étendre l'étude de ces processus ARMA et VARMA faibles cette fois en temps continu afin d'établir des relations entre les deux types de modélisation (en temps discret et continu).

7.3 Insertion dans l'équipe de recherche

Depuis mon arrivée au Laboratoire de Mathématiques de Besançon (UMR CNRS 6623) en septembre 2011, je fais partie de l'équipe Probabilités et Statistique.

J'organise et j'ai co-organisé le séminaire de mon équipe depuis septembre 2012. J'ai fait plusieurs présentations dans les séminaires de mon équipe, et j'ai également eu l'occasion de faire un exposé lors des journées de notre laboratoire en janvier 2012. Depuis 2012, je donne un cours de recherche spécialisé en statistique en Master 2 Modélisation Statistique et j'encadre régulièrement des mémoires de recherche, des projets et des stages d'étudiants du Master 1 et 2 modélisation statistique. J'ai également assuré la responsabilité pédagogique du Master 1 modélisation statistique de 2013 à 2016.

Organisation d'événements

J'ai également co-organisé plusieurs événements dans le cadre des trimestres du LMB : deux conférences en séries temporelles (35 participants chacune) et un mini cours en séries temporelles à valeurs entières (40 participants).

J'ai été examinateur de la thèse de doctorat de mathématiques appliquées de Clément Marsilli et également membre invité du jury de la thèse de doctorat de mathématiques appliquées de Basad Al-Sarray.

Financements

- 1) Projet bonus qualité recherche (BQR) de 2012-2014, financé par l'Université de Franche-Comté, d'un montant de 5000 euros.
- 2) Porteur du projet région de 2017 à 2020, réunissant la majorité des membres de l'équipe de probabilités et statistique du LMB, financé par la région Bourgogne-Franche-Comté d'un montant de 32000 euros. Ce projet impliquerait également d'autres chercheurs extérieurs à l'Université de Franche-Comté.

Bibliographie

- [ADB06] Beth Andrews, Richard A. Davis, and F. Jay Breidt. Maximum likelihood estimation for all-pass time series models. *J. Multivariate Anal.*, 97(7) :1638–1659, 2006.
- [Aka70] Hirotugu Akaike. Statistical predictor identification. *Ann. Inst. Statist. Math.*, 22 :203–217, 1970.
- [Aka73] H. Akaike. Information theory and an extension of the maximum likelihood principle. pages 267–281, 1973.
- [AM97] Paul L. Anderson and Mark M. Meerschaert. Periodic moving averages of random variables with regularly varying tails. *Ann. Statist.*, 25(2) :771–785, 1997.
- [AM98] Rajae Azrak and Guy Mélard. The exact quasi-likelihood of time-dependent ARMA models. *J. Statist. Plann. Inference*, 68(1) :31–45, 1998.
- [AM06] Rajae Azrak and Guy Mélard. Asymptotic properties of quasi-maximum likelihood estimators for ARMA models with time-dependent coefficients. *Stat. Inference Stoch. Process.*, 9(3) :279–330, 2006.
- [And91] Donald W. K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3) :817–858, 1991.
- [BD91] Peter J. Brockwell and Richard A. Davis. *Time series : theory and methods*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 1991.
- [BDT01] F. Jay Breidt, Richard A. Davis, and A. Alexandre Trindade. Least absolute deviation estimation for all-pass time series models. *Ann. Statist.*, 29(4) :919–946, 2001.
- [Ber74] Kenneth N. Berk. Consistent autoregressive spectral estimates. *Ann. Statist.*, 2 :489–502, 1974. Collection of articles dedicated to Jerzy Neyman on his 80th birthday.
- [BF03] Abdelouahab Bibi and Christian Francq. Consistent and asymptotically normal estimators for cyclically time-dependent linear models. *Ann. Inst. Statist. Math.*, 55(1) :41–68, 2003.
- [BHK03] István Berkes, Lajos Horváth, and Piotr Kokoszka. Asymptotics for GARCH squared residual correlations. *Econometric Theory*, 19 :515–540, 2003.
- [BHL12] Luc Bauwens, Christian M Hafner, and Sébastien Laurent. *Handbook of volatility models and their applications*, volume 3. John Wiley & Sons, 2012.
- [BJ76] George E. P. Box and Gwilym M. Jenkins. *Time series analysis : forecasting and control*. Holden-Day, San Francisco, Calif.-Düsseldorf-Johannesburg, revised edition, 1976. Holden-Day Series in Time Series Analysis.
- [BL01] I. V. Basawa and Robert Lund. Large sample properties of parameter estimates for periodic ARMA models. *J. Time Ser. Anal.*, 22(6) :651–663, 2001.

- [Bla76] F Black. Studies of stock price volatility changes. In proceedings from the ASA. *Business and Economic Statistics Section*, pages 177–181, 1976.
- [BLR06] Luc Bauwens, Sébastien Laurent, and Jeroen V. K. Rombouts. Multivariate GARCH models : a survey. *J. Appl. Econometrics*, 21(1) :79–109, 2006.
- [BM10] Yacouba Boubacar Mainassara. Tests portmanteau multivariés d'adéquation de modèles VARMA faibles. *C. R. Math. Acad. Sci. Paris*, 348(15-16) :927–929, 2010.
- [BM11a] Y. Boubacar Mainassara. Multivariate portmanteau test for structural VARMA models with uncorrelated but non-independent error terms. *J. Statist. Plann. Inference*, 141(8) :2961–2975, 2011.
- [BM11b] Yacouba Boubacar Maïnassara. Estimation de la matrice de variance asymptotique des estimateurs du QMV de modèles ARMA faibles multivariés. *C. R. Math. Acad. Sci. Paris*, 349(13-14) :817–820, 2011.
- [BM11c] Yacouba Boubacar Mainassara. Estimation des ordres de modèles ARMA faibles multivariés. *C. R. Math. Acad. Sci. Paris*, 349(11-12) :695–698, 2011.
- [BM12] Y. Boubacar Maïnassara. Selection of weak VARMA models by modified Akaike's information criteria. *J. Time Series Anal.*, 33(1) :121–130, 2012.
- [BM14] Yacouba Boubacar Maïnassara. Estimation of the variance of the quasi-maximum likelihood estimator of weak VARMA models. *Electron. J. Stat.*, 8(2) :2701–2740, 2014.
- [BMCF12] Y. Boubacar Mainassara, M. Carbon, and C. Francq. Computing and estimating information matrices of weak ARMA models. *Comput. Statist. Data Anal.*, 56(2) :345–361, 2012.
- [BMF11] Y. Boubacar Mainassara and C. Francq. Estimating structural VARMA models with uncorrelated but non-independent error terms. *J. Multivariate Anal.*, 102(3) :496–505, 2011.
- [BMK14] Yacouba Boubacar Maïnassara and Célestin C. Kokonendji. On normal stable Tweedie models and power-generalized variance functions of only one component. *TEST*, 23(3) :585–606, 2014.
- [BMK16] Yacouba Boubacar Maïnassara and Célestin C. Kokonendji. Modified Schwarz and Hannan-Quinn information criteria for weak VARMA models. *Stat. Inference Stoch. Process.*, 19(2) :199–217, 2016.
- [BMR99] M. Billio, A. Monfort, and C. P. Robert. Bayesian estimation of switching ARMA models. *J. Econometrics*, 93(2) :229–255, 1999.
- [BMR15] Yacouba Boubacar Maïnassara and Hamdi Raïssi. Semi-strong linearity testing in linear models with dependent but uncorrelated errors. *Statist. Probab. Lett.*, 103 :110–115, 2015.
- [BMS17] Yacouba Boubacar Maïnassara and Bruno Sausseureau. Diagnostic checking in multivariate arma models with dependent errors using normalized residual autocorrelations. *J. Amer. Statist. Assoc.*, ?? :??–??, 2017.
- [BN78] Ole Barndorff-Nielsen. *Information and exponential families in statistical theory*. John Wiley & Sons, Ltd., Chichester, 1978. Wiley Series in Probability and Mathematical Statistics.

- [Bol86] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31 :307–327, 1986.
- [Bol90] Tim Bollerslev. Modelling the coherence in short-run nominal exchange rates : a multivariate generalized arch model. *The review of economics and statistics*, pages 498–505, 1990.
- [BP70] G. E. P. Box and David A. Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Amer. Statist. Assoc.*, 65 :1509–1526, 1970.
- [BW09] Jean-Marc Bardet and Olivier Wintenberger. Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. *Ann. Statist.*, 37(5B) :2730–2759, 2009.
- [Cas96] M. Casalis. The $2d + 4$ simple quadratic natural exponential families on \mathbf{R}^d . *Ann. Statist.*, 24(4) :1828–1854, 1996.
- [CF11] Michel Carbon and Christian Francq. Portmanteau goodness-of-fit test for asymmetric power GARCH models. *Austrian Journal of Statistics*, 40 :55–64, 2011.
- [Chi74] Ratnam V. Chitturi. Distribution of residual autocorrelations in multiple autoregressive schemes. *J. Amer. Statist. Assoc.*, 69 :928–934, 1974.
- [CL03] F. Comte and O. Lieberman. Asymptotic theory for multivariate GARCH processes. *J. Multivariate Anal.*, 84(1) :61–84, 2003.
- [Dah97] R. Dahlhaus. Fitting time series models to nonstationary processes. *Ann. Statist.*, 25(1) :1–37, 1997.
- [Dav68] Ju. A. Davydov. Convergence of distributions generated by stationary stochastic processes. *Theor. of Proba. and Appl.*, 13(2) :691–696, 1968.
- [Dav94] James Davidson. *Stochastic limit theory*. Advanced Texts in Econometrics. The Clarendon Press, Oxford University Press, New York, 1994. An introduction for econometricians.
- [DGE93] Zhuanxin Ding, Clive WJ Granger, and Robert F Engle. A long memory property of stock market returns and a new model. *Journal of empirical finance*, 1 :83–106, 1993.
- [DH76] W. Dunsmuir and E. J. Hannan. Vector linear time series models. *Advances in Appl. Probability*, 8(2) :339–364, 1976.
- [dHL97] Wouter J. den Haan and Andrew T. Levin. A practitioner’s guide to robust covariance matrix estimation. In *Robust inference*, volume 15 of *Handbook of Statist.*, pages 299–342. North-Holland, Amsterdam, 1997.
- [DJ14] Jean-Marie Dufour and Tarek Jouini. Asymptotic distributions for quasi-efficient estimators in echelon VARMA models. *Comput. Statist. Data Anal.*, 73 :69–86, 2014.
- [DN93] Feike C. Drost and Theo E. Nijman. Temporal aggregation of GARCH processes. *Econometrica*, 61(4) :909–927, 1993.
- [DR08] Pierre Duchesne and Roch Roy. Practical methods for modelling weak varma processes : identification, estimation and specification with a macroeconomic application. *Discussion Paper, McGill University (Department of Economics), CIREQ and CIRANO*, 2008.

- [DSG10] Sonjoy Das, James C. Spall, and Roger Ghanem. Efficient Monte Carlo computation of Fisher information matrix using prior information. *Comput. Statist. Data Anal.*, 54(2) :272–289, 2010.
- [Duf04] Denis Dufour, Jean-Marie. et Pelletier. On consistent testing for serial correlation of unknown form in vector time series models. *J. Multivariate Anal.*, 89(1) :148–180, 2004.
- [Dur59] James Durbin. Efficient estimation of parameters in moving-average models. *Biometrika*, 46(3/4) :306–316, 1959.
- [EG10] Ahmed El Ghini. Processus dual et inverse d'un ARMA et application à la réversibilité temporelle. *C. R. Math. Acad. Sci. Paris*, 348(1-2) :85–88, 2010.
- [EG15] Ahmed El Ghini. Probabilistic properties of parametric dual and inverse time series models generated by ARMA models. *Comm. Statist. Theory Methods*, 44(21) :4651–4661, 2015.
- [EGF06] Ahmed El Ghini and Christian Francq. Asymptotic relative efficiency of goodness-of-fit tests based on inverse and ordinary autocorrelations. *J. Time Ser. Anal.*, 27(6) :843–855, 2006.
- [EK95] Robert F. Engle and Kenneth F. Kroner. Multivariate simultaneous generalized arch. *Econometric Theory*, 11(1) :122–150, 1995.
- [Eng82] Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica : Journal of the Econometric Society*, pages 987–1007, 1982.
- [Eng02] Robert Engle. Dynamic conditional correlation : a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *J. Bus. Econom. Statist.*, 20(3) :339–350, 2002.
- [Fam65] Eugene F Fama. The behavior of stock-market prices. *The journal of Business*, 38(1) :34–105, 1965.
- [FG04a] Christian Francq and Antony Gautier. Estimation of time-varying ARMA models with Markovian changes in regime. *Statist. Probab. Lett.*, 70(4) :243–251, 2004.
- [FG04b] Christian Francq and Antony Gautier. Large sample properties of parameter least squares estimates for time-varying ARMA models. *J. Time Ser. Anal.*, 25(5) :765–783, 2004.
- [Fin93] David F. Findley. The overfitting principles supporting aic. 1993.
- [FR97] Christian Francq and Michel Roussignol. On white noises driven by hidden Markov chains. *J. Time Ser. Anal.*, 18(6) :553–578, 1997.
- [FR98] Christian Francq and Michel Roussignol. Ergodicity of autoregressive processes with Markov-switching and consistency of the maximum-likelihood estimator. *Statistics*, 32(2) :151–173, 1998.
- [FRS11] Christian Francq, Roch Roy, and Abdessamad Saidi. Asymptotic properties of weighted least squares estimation in weak PARMA models. *J. Time Series Anal.*, 32(6) :699–723, 2011.
- [FRZ05] Christian Francq, Roch Roy, and Jean-Michel Zakoïan. Diagnostic checking in ARMA models with uncorrelated errors. *J. Amer. Statist. Assoc.*, 100(470) :532–544, 2005.

- [FS17] Christian Francq and Genaro Sucarrat. An equation-by-equation estimator of a multivariate log-GARCH-X model of financial returns. *J. Multivariate Anal.*, 153 :16–32, 2017.
- [FWZ16] Christian Francq, Olivier Wintenberger, and Jean-Michel Zakoïan. Goodness-of-fit tests for Log-GARCH and EGARCH models. *TEST*, pages 1–25, 2016.
- [FZ98] Christian Francq and Jean-Michel Zakoïan. Estimating linear representations of nonlinear processes. *J. Statist. Plann. Inference*, 68(1) :145–165, 1998.
- [FZ00] Christian Francq and Jean-Michel Zakoïan. Covariance matrix estimation for estimators of mixing weak ARMA models. *J. Statist. Plann. Inference*, 83(2) :369–394, 2000.
- [FZ01] Christian Francq and Jean-Michel Zakoïan. Stationarity of multivariate markov-switching arma models. *J. Econometrics*, 102(2) :339–364, 2001.
- [FZ02] Christian Francq and Jean-Michel Zakoïan. Autocovariance structure of powers of switching-regime ARMA processes. *ESAIM Probab. Statist.*, 6 :259–270, 2002. New directions in time series analysis (Luminy, 2001).
- [FZ05a] Christian Francq and Jean-Michel Zakoïan. A central limit theorem for mixing triangular arrays of variables whose dependence is allowed to grow with the sample size. *Econometric Theory*, 21(6) :1165–1171, 2005.
- [FZ05b] Christian Francq and Jean-Michel Zakoïan. Recent results for linear time series models with non independent innovations. In *Statistical modeling and analysis for complex data problems*, volume 1 of *GERAD 25th Anniv. Ser.*, pages 241–265. Springer, New York, 2005.
- [FZ07] Christian Francq and Jean-Michel Zakoïan. HAC estimation and strong linearity testing in weak ARMA models. *J. Multivariate Anal.*, 98(1) :114–144, 2007.
- [FZ09] Christian Francq and Jean-Michel Zakoïan. Bartlett's formula for a general class of nonlinear processes. *J. Time Series Anal.*, 30(4) :449–465, 2009.
- [FZ10] Christian Francq and Jean-Michel Zakoïan. *GARCH Models : Structure, Statistical Inference and Financial Applications*. Wiley, 2010.
- [FZ12] Christian Francq and Jean-Michel Zakoïan. QML estimation of a class of multivariate asymmetric GARCH models. *Econometric Theory*, 28 :179–206, 2012.
- [FZ16] Christian Francq and Jean-Michel Zakoïan. Estimating multivariate volatility models equation by equation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 78(3) :613–635, 2016.
- [GB06] E. J. Godolphin and S. R. Bane. On the evaluation of the information matrix for multiplicative seasonal time-series models. *J. Time Ser. Anal.*, 27(2) :167–190, 2006.
- [GJR93] Lawrence R Glosten, Ravi Jagannathan, and David E Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, 48 :1779–1801, 1993.
- [Ham88] James D. Hamilton. Rational-expectations econometric analysis of changes in regime : an investigation of the term structure of interest rates. *J. Econom. Dynam. Control*, 12(2-3) :385–423, 1988. Economic time series with random walk and other nonstationary components.
- [Ham89] James D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2) :357–384, 1989.

- [Ham90] James D. Hamilton. Analysis of time series subject to changes in regime. *J. Econometrics*, 45(1-2) :39–70, 1990.
- [Ham94] James D. Hamilton. *Time series analysis*. Princeton University Press, Princeton, NJ, 1994.
- [Han73] E. J. Hannan. The asymptotic theory of linear time-series models. *J. Appl. Probability*, 10 :130–145, corrections, *ibid.* 10 (1973), 913, 1973.
- [Han76] E. J. Hannan. The identification and parametrization of ARMAX and state space forms. *Econometrica*, 44(4) :713–723, 1976.
- [HB92] M. L. Higgins and A. K. Bera. A class of nonlinear ARCH models. *Internat. Econom. Rev.*, 33(1) :137–158, 1992.
- [HD88] E. J. Hannan and Manfred Deistler. *The statistical theory of linear systems*. Wiley Series in Probability and Mathematical Statistics : Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1988.
- [HDD80] E. J. Hannan, W. T. M. Dunsmuir, and M. Deistler. Estimation of vector ARMAX models. *J. Multivariate Anal.*, 10(3) :275–295, 1980.
- [Her84] Norbert Herrndorf. A functional central limit theorem for weakly dependent sequences of random variables. *Ann. Probab.*, 12(1) :141–153, 1984.
- [HK04] S. Y. Hwang and Tae Yoon Kim. Power transformation and threshold modeling for ARCH innovations with applications to tests for ARCH structure. *Stochastic Process. Appl.*, 110(2) :295–314, 2004.
- [Hos80] J. R. M. Hosking. The multivariate portmanteau statistic. *J. Amer. Statist. Assoc.*, 75(371) :602–608, 1980.
- [Hos81a] J. R. M. Hosking. Equivalent forms of the multivariate portmanteau statistic. *J. Roy. Statist. Soc. Ser. B*, 43(2) :261–262, 1981.
- [Hos81b] J. R. M. Hosking. Lagrange-multiplier tests of multivariate time-series models. *J. Roy. Statist. Soc. Ser. B*, 43(2) :219–230, 1981.
- [Hos89] J. R. M. Hosking. Corrigendum : “Equivalent forms of the multivariate portmanteau statistic” [*J. Roy. Statist. Soc. Ser. B* 43 (1981), no. 2, 261–262 ; MR0626774 (82h :62153)]. *J. Roy. Statist. Soc. Ser. B*, 51(2) :303, 1989.
- [HP09] Christian M. Hafner and Arie Preminger. On asymptotic theory for multivariate GARCH models. *J. Multivariate Anal.*, 100(9) :2044–2054, 2009.
- [HQ79] E. J. Hannan and B. G. Quinn. The determination of the order of an autoregression. *J. Roy. Statist. Soc. Ser. B*, 41(2) :190–195, 1979.
- [HR83] E. J. Hannan and J. Rissanen. Errata : “Recursive estimation of mixed autoregressive-moving average order” [*Biometrika* 69 (1982), no. 1, 81–94 ; MR0655673 (84e :62136)]. *Biometrika*, 70(1) :303, 1983.
- [HS94] James D Hamilton and Raul Susmel. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of econometrics*, 64(1) :307–333, 1994.
- [HT89] Clifford M. Hurvich and Chih-Ling Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2) :297–307, 1989.
- [HT93] Clifford M. Hurvich and Chih-Ling Tsai. A corrected Akaike information criterion for vector autoregressive model selection. *J. Time Ser. Anal.*, 14(3) :271–279, 1993.

- [HZ11] Tawfik Hamadeh and Jean-Michel Zakoïan. Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH processes. *Journal of Statistical Planning and Inference*, 141 :488–507, 2011.
- [Imh61] J. P. Imhof. Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48 :419–426, 1961.
- [Jea98] Thierry Jeantheau. Strong consistency of estimators for multivariate ARCH models. *Econometric Theory*, 14(1) :70–86, 1998.
- [Jør13] Bent Jørgensen. Construction of multivariate dispersion models. *Braz. J. Probab. Stat.*, 27(3) :285–309, 2013.
- [Jr97] Bent Jørgensen. *The theory of dispersion models*, volume 76 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1997.
- [Kas12] Christian Kascha. A comparison of estimation methods for vector autoregressive moving-average models. *Econometric Rev.*, 31(3) :297–324, 2012.
- [Kat12] Naoya Katayama. Chi-squared portmanteau tests for structural VARMA models with uncorrelated errors. *J. Time Series Anal.*, 33(6) :863–872, 2012.
- [KBJ00] Samuel Kotz, N. Balakrishnan, and Norman L. Johnson. *Continuous multivariate distributions. Vol. 1*. Wiley Series in Probability and Statistics : Applied Probability and Statistics. Wiley-Interscience, New York, second edition, 2000. Models and applications.
- [KK15] Chang-Jin Kim and Jaeho Kim. Bayesian inference in regime-switching ARMA models with absorbing states : the dynamics of the ex-ante real interest rate under regime shifts. *J. Bus. Econom. Statist.*, 33(4) :566–578, 2015.
- [KL06] Chung-Ming Kuan and Wei-Ming Lee. Robust M tests without consistent estimation of the asymptotic covariance matrix. *J. Amer. Statist. Assoc.*, 101(475) :1264–1275, 2006.
- [KM90] André Klein and Guy Mélard. Fisher’s information matrix for seasonal autoregressive-moving average models. *J. Time Ser. Anal.*, 11(3) :231–237, 1990.
- [KM04] André Klein and Guy Mélard. An algorithm for computing the asymptotic Fisher information matrix for seasonal SISO models. *J. Time Ser. Anal.*, 25(5) :627–648, 2004.
- [LB78] G. M. Ljung and G. E. P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2) :pp. 297–303, 1978.
- [Let89] Gérard Letac. Le problème de la classification des familles exponentielles naturelles de \mathbb{R}^d ayant une fonction variance quadratique. *Probability measures on groups IX*, pages 192–216, 1989.
- [Li04] Wai Keung Li. *Diagnostic checks in time series*. Boca Raton, Florida : Chapman and Hall, 2004.
- [LKN15] Anne Leucht, Jens-Peter Kreiss, and Michael H Neumann. A model specification test for GARCH(1,1) processes. *Scandinavian Journal of Statistics*, 42 :1167–1193, 2015.
- [LL97] Shiqing Ling and Wai Keung Li. On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *Journal of the American Statistical Association*, 92 :1184–1194, 1997.
- [LM81] W. K. Li and A. I. McLeod. Distribution of the residual autocorrelations in multivariate ARMA time series models. *J. Roy. Statist. Soc. Ser. B*, 43(2) :231–239, 1981.

- [LM94] Wai Keung Li and TK Mak. On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *Journal of Time Series Analysis*, 15 :627–636, 1994.
- [LM03] Shiqing Ling and Michael McAleer. Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory*, 19(2) :280–310, 2003.
- [LNS01] Ignacio N. Lobato, J. C. Nankervis, and N. E. Savin. Testing for autocorrelation using a modified box- pierce q test. *Inter. Econ. Review*, 42(1) :187–205, 2001.
- [LNS02] Ignacio N. Lobato, J. C. Nankervis, and N. E. Savin. Testing for zero autocorrelation in the presence of statistical dependence. *Econ. Theory*, 18(3) :730–743, 2002.
- [Lob01] Ignacio N. Lobato. Testing that a dependent process is uncorrelated. *J. Amer. Statist. Assoc.*, 96(455) :1066–1076, 2001.
- [Lüt05] Helmut Lütkepohl. *New introduction to multiple time series analysis*. Springer-Verlag, Berlin, 2005.
- [Man63] Benoit Mandelbrot. The variation of certain speculative prices. *The journal of business*, 36(4) :394–419, 1963.
- [MCHL08] Michael McAleer, Felix Chan, Suhejla Hoti, and Offer Lieberman. Generalized autoregressive conditional correlation. *Econometric Theory*, 24(6) :1554–1583, 2008.
- [McL78] AI McLeod. On the distribution of residual autocorrelations in Box-Jenkins models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 296–302, 1978.
- [McL84] AI McLeod. Duality and other properties of multiplicative seasonal autoregressive-moving average models. *Biometrika*, 71(1) :207–211, 1984.
- [McL94] A. I. McLeod. Diagnostic checking of periodic autoregression models with application. *J. Time Ser. Anal.*, 15(2) :221–233, 1994.
- [McL95] A. I. McLeod. Addendum : “Diagnostic checking of periodic autoregression models with application” [J. Time Ser. Anal. **15** (1994), no. 2, 221–233]. *J. Time Ser. Anal.*, 16(6) :647–648, 1995.
- [MHC09] Michael McAleer, Suhejla Hoti, and Felix Chan. Structure and asymptotic theory for multivariate asymmetric conditional volatility. *Econometric Rev.*, 28(5) :422–440, 2009.
- [Nel91a] Daniel B Nelson. Conditional heteroskedasticity in asset returns : A new approach. *Econometrica : Journal of the Econometric Society*, pages 347–370, 1991.
- [Nel91b] Daniel B. Nelson. Conditional heteroskedasticity in asset returns : a new approach. *Econometrica*, 59(2) :347–370, 1991.
- [NQ82] Des F. Nicholls and Barry G. Quinn. *Random coefficient autoregressive models : an introduction*, volume 11 of *Lecture Notes in Statistics*. Springer-Verlag, New York-Berlin, 1982. Lecture Notes in Physics, 151.
- [NW87] Whitney K. Newey and Kenneth D. West. A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3) :703–708, 1987.
- [Pfa69] J Pfanzagl. On the measurability and consistency of minimum contrast estimates. *Metrika*, 14(1) :249–272, 1969.
- [PWT08] Jiazhu Pan, Hui Wang, and Howell Tong. Estimation and tests for power-transformed and threshold GARCH models. *J. Econometrics*, 142(1) :352–378, 2008.

- [RBY92] Gregory C. Reinsel, Sabyasachi Basu, and Sook Fwe Yap. Maximum likelihood estimators in the multivariate autoregressive moving-average model from a generalized least squares viewpoint. *J. Time Ser. Anal.*, 13(2) :133–145, 1992.
- [RC79] J. Rissanen and P. E. Caines. The strong consistency of maximum likelihood estimators for ARMA processes. *Ann. Statist.*, 7(2) :297–315, 1979.
- [Rei97] Gregory C. Reinsel. *Elements of multivariate time series analysis*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 1997.
- [RT96] Joseph P. Romano and Lori A. Thombs. Inference for autocorrelations under weak assumptions. *J. Amer. Statist. Assoc.*, 91(434) :590–600, 1996.
- [RZ93] Roger Rabemananjara and Jean-Michel Zakoïan. Threshold ARCH models and asymmetries in volatility. *Journal of Applied Econometrics*, 8 :31–49, 1993.
- [Sch78] Gideon Schwarz. Estimating the dimension of a model. *Ann. Statist.*, 6, 1978.
- [SGnE16] Genaro Sucarrat, Steffen Grønneberg, and Alvaro Escibano. Estimation and inference in univariate and multivariate log-GARCH-X models when the conditional density is unknown. *Comput. Statist. Data Anal.*, 100 :582–594, 2016.
- [Sha0a] Xiaofeng Shao. A self-normalized approach to confidence interval construction in time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(3) :343–366, 2010a.
- [Sha0b] Xiaofeng Shao. Corrigendum : A self-normalized approach to confidence interval construction in time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(5) :695–696, 2010b.
- [Sha11] Xiaofeng Shao. Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models. *Econometric Theory*, 27(2) :312–343, 2011.
- [Sha12] Xiaofeng Shao. Parametric inference in stationary time series models with dependent errors. *Scand. J. Stat.*, 39(4) :772–783, 2012.
- [Sha15] Xiaofeng Shao. Self-normalization for time series : a review of recent developments. *J. Amer. Statist. Assoc.*, 110(512) :1797–1817, 2015.
- [Sid58] MM Siddiqui. On the inversion of the sample covariance matrix in a stationary autoregressive process. *The Annals of Mathematical Statistics*, 29(2) :585–588, 1958.
- [Slu37] Eugen Slutsky. The summation of random causes as the source of cyclic processes. *Econometrica : Journal of the Econometric Society*, pages 105–146, 1937.
- [Twe84] M. C. K. Tweedie. An index which distinguishes between some important exponential families. In *Statistics : applications and new directions (Calcutta, 1981)*, pages 579–604. Indian Statist. Inst., Calcutta, 1984.
- [UD09] Eugen Ursu and Pierre Duchesne. On modelling and diagnostic checking of vector periodic autoregressive time series models. *J. Time Series Anal.*, 30(1) :70–96, 2009.
- [vdV98] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.
- [Wal49] Abraham Wald. Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statistics*, 20 :595–601, 1949.
- [Wal64] A. M. Walker. Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series. *J. Austral. Math. Soc.*, 4 :363–384, 1964.

- [Whi53] P. Whittle. Estimation and information in stationary time series. *Ark. Mat.*, 2 :423–434, 1953.
- [Whi63] Peter Whittle. On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. *Biometrika*, 50(1-2) :129–134, 1963.
- [Wol54] Herman Wold. *A study in the analysis of stationary time series*. Almqvist and Wiksell, Stockholm, 1954. 2d ed, With an appendix by Peter Whittle.
- [Yul26] G Udney Yule. Why do we sometimes get nonsense-correlations between time-series?—a study in sampling and the nature of time-series. *Journal of the royal statistical society*, 89(1) :1–63, 1926.
- [ZL15] Ke Zhu and Wai Keung Li. A bootstrapped spectral test for adequacy in weak ARMA models. *J. Econometrics*, 187(1) :113–130, 2015.