Modèles de séries temporelles à mémoire longue avec innovations dépendantes

Thèse de doctorat de
l'Université Bourgogne Franche-Comté

École doctorale CARNOT-PASTEUR

présentée et soutenue publiquement le 8 novembre 2019
en vue de l'obtention du grade de

Docteur de l'Université Bourgogne Franche-Comté
(mention Mathématiques Appliquées et Applications des Mathématiques)

par

Youssef ESSTAFA

Composition du jury :

Rapporteurs :	M. Jean-Marc BARDET	Université Paris 1 Panthéon-Sorbonne
	Mme Anne PHILIPPE	Université de Nantes
Président :	M. Clément DOMBRY	Université Bourgogne Franche-Comté
Examinateurs :	M. Alexandre BROUSTE	Le Mans Université
	M. Christian FRANCQ	Université Lille 3 et CREST
	M. Christophe HURLIN	Université d'Orléans
Directeurs de thèse :	M. Yacouba BOUBACAR MAÏNASSARA	Université Bourgogne Franche-Comté
	M. Bruno SAUSSEREAU	Université Bourgogne Franche-Comté

laboratoiredemathématiquesdebesançon
UNIVERSITÉ DE FRANCHE-COMTÉ • CNRS • UMR 6623

Remerciements

En préambule à ce travail, je souhaiterais adresser mes remerciements les plus sincères aux personnes qui m'ont apporté leur aide et qui ont contribué à l'élaboration de cette thèse.

Je tiens tout d'abord à remercier très chaleureusement mes directeurs de thèse Monsieur Yacouba BOUBACAR MAÏNASSARA et Monsieur Bruno SAUSSEREAU pour la confiance qu'ils m'ont accordée en acceptant d'encadrer ce travail avec enthousiasme, rigueur et bienveillance. Leurs qualités scientifiques exceptionnelles associées à leur incroyable gentillesse représentent la clef de ce travail. Je leur suis infiniment reconnaissant pour les conseils avisés qu'ils m'ont prodigué tout au long de mon cursus universitaire, leurs suivis attentifs à tous les niveaux d'élaboration de cette thèse, leur soutien infaillible, leur disponibilité dès que nécessaire et tous les moments de partages qui ont enrichi mon parcours.

Je tiens à remercier Monsieur le Professeur Jean-Marc BARDET et Madame la Professeure Anne PHILIPPE de l'honneur qu'ils me font d'avoir accepté d'être les rapporteurs de cette thèse et de l'intérêt porté à l'égard de mon travail. Je leur exprime toute ma gratitude pour le temps précieux qu'ils ont consacré à la lecture de cette thèse et à l'élaboration de leurs rapports.

Je remercie également Messieurs les Professeurs Alexandre BROUSTE, Clément DOMBRY, Christian FRANCQ et Christophe HURLIN d'avoir accepté de faire partie de mon jury de thèse.

Un grand merci à Madame la Professeure Anne PHILIPPE et Monsieur Ulrich RAZAFISON d'avoir constitué mon comité de suivi de thèse pendant ces trois dernières années.

J'adresse mes remerciements à l'ensemble des membres du Laboratoire de Mathématiques de Besançon qui m'ont permis de m'intégrer rapidement et de réaliser mes projets dans une ambiance très sereine et conviviale.

Mes derniers remerciements et non les moindres, s'adressent à mes parents, à toute ma famille et à mes amis pour leur soutien constant et leur présence sans faille. Un dernier grand merci très particulier à ma compagne Manou pour ses encouragements, son soutien inconditionnel et sa patience tout au long de ce travail.

Je dédie ce travail à la mémoire de mon papa décédé le 22 avril 2017 et à ma chère maman ...

Résumé

Dans cette thèse nous considérons, dans un premier temps, le problème de l'analyse statistique des modèles FARIMA (Fractionally AutoRegressive Integrated Moving-Average) induits par un bruit blanc non corrélé mais qui peut contenir des dépendances non linéaires très générales. Ces modèles sont appelés FARIMA faibles et permettent de modéliser des processus à mémoire longue présentant des dynamiques non linéaires, de structures souvent non-identifiées, très générales. Relâcher l'hypothèse d'indépendance sur le terme d'erreur, une hypothèse habituellement imposée dans la littérature, permet aux modèles FARIMA faibles d'élargir considérablement leurs champs d'application en couvrant une large classe de processus à mémoire longue non linéaires. Les modèles FARIMA faibles sont denses dans l'ensemble des processus stationnaires purement non déterministes, la classe formée par ces modèles englobe donc celle des processus FARIMA avec un bruit indépendant et identiquement distribué (iid). Nous appelons par la suite FARIMA forts les modèles dans lesquels le terme d'erreur est supposé être un bruit iid.

Nous établissons les procédures d'estimation et de validation des modèles FARIMA faibles. Nous montrons, sous des hypothèses faibles de régularités sur le bruit, que l'estimateur des moindres carrés des paramètres des modèles FARIMA(p, d, q) faibles est fortement convergent et asymptotiquement normal. La matrice de variance asymptotique de l'estimateur des moindres carrés des modèles FARIMA(p, d, q) faibles est de la forme "sandwich" $\Omega := J^{-1}IJ^{-1}$. Cette matrice peut être très différente de la variance asymptotique $\Omega := 2J^{-1}$ obtenue dans le cas fort (*i.e.* dans le cas où le bruit est supposé iid). Nous proposons, par deux méthodes différentes, un estimateur convergent de cette matrice. Une méthode alternative basée sur une approche d'auto-normalisation est également proposée pour construire des intervalles de confiance des paramètres des modèles FARIMA(p, d, q) faibles. Cette technique nous permet de contourner le problème de l'estimation de la matrice de variance asymptotique de l'estimateur des moindres carrés.

Nous accordons ensuite une attention particulière au problème de la validation des modèles FARIMA(p, d, q) faibles. Nous montrons que les autocorrélations résiduelles ont une distribution asymptotique normale de matrice de covariance différente de celle obtenue dans le cadre

des FARIMA forts. Cela nous permet de déduire la loi asymptotique exacte des statistiques portmanteau et de proposer ainsi des versions modifiées des tests portmanteau standards de Box-Pierce et Ljung-Box. Il est connu que la distribution asymptotique des tests portmanteau est correctement approximée par un khi-deux lorsque le terme d'erreur est supposé iid. Dans le cas général, nous montrons que cette distribution asymptotique est celle d'une somme pondérée de khi-deux. Elle peut être très différente de l'approximation khi-deux usuelle du cas fort. Nous adoptons la même approche d'auto-normalisation utilisée pour la construction des intervalles de confiance des paramètres des modèles FARIMA faibles pour tester l'adéquation des modèles FARIMA(p, d, q) faibles. Cette méthode a l'avantage de contourner le problème de l'estimation de la matrice de variance asymptotique du vecteur joint de l'estimateur des moindres carrés et des autocovariances empiriques du bruit.

Dans un second temps, nous traitons dans cette thèse le problème de l'estimation des modèles autorégressifs d'ordre 1 induits par un bruit gaussien fractionnaire d'indice de Hurst H supposé connu. Nous étudions, plus précisément, la convergence et la normalité asymptotique de l'estimateur des moindres carrés généralisés du paramètre autorégressif de ces modèles.

Mots-clés : Processus à mémoire longue, processus non linéaires, modèles FARIMA faibles, estimateur des moindres carrés, convergence forte, normalité asymptotique, estimateur spectral, approche d'auto-normalisation, autocorrélations résiduelles, tests portmanteau de Ljung-Box et Box-Pierce, mouvement brownien fractionnaire, bruit gaussien fractionnaire, modèles autorégressifs fractionnaires, estimateur des moindres carrés généralisés.

Abstract

We first consider, in this thesis, the problem of statistical analysis of FARIMA (Fractionally AutoRegressive Integrated Moving-Average) models endowed with uncorrelated but non-independent error terms. These models are called weak FARIMA and can be used to fit long-memory processes with general nonlinear dynamics. Relaxing the independence assumption on the noise, which is a standard assumption usually imposed in the literature, allows weak FARIMA models to cover a large class of nonlinear long-memory processes. The weak FARIMA models are dense in the set of purely non-deterministic stationary processes, the class of these models encompasses that of FARIMA processes with an independent and identically distributed noise (iid). We call thereafter strong FARIMA models the models in which the error term is assumed to be an iid innovations.

We establish procedures for estimating and validating weak FARIMA models. We show, under weak assumptions on the noise, that the least squares estimator of the parameters of weak FARIMA(p, d, q) models is strongly consistent and asymptotically normal. The asymptotic variance matrix of the least squares estimator of weak FARIMA(p, d, q) models has the "sandwich" form $\Omega := J^{-1}IJ^{-1}$. This matrix can be very different from the asymptotic variance $\Omega := 2J^{-1}$ obtained in the strong case (*i.e.* in the case where the noise is assumed to be iid). We propose, by two different methods, a convergent estimator of this matrix. An alternative method based on a self-normalization approach is also proposed to construct confidence intervals for the parameters of weak FARIMA(p, d, q) models.

We then pay particular attention to the problem of validation of weak FARIMA(p, d, q) models. We show that the residual autocorrelations have a normal asymptotic distribution with a covariance matrix different from that one obtained in the strong FARIMA case. This allows us to deduce the exact asymptotic distribution of portmanteau statistics and thus to propose modified versions of portmanteau tests. It is well known that the asymptotic distribution of portmanteau tests is correctly approximated by a chi-squared distribution when the error term is assumed to be iid. In the general case, we show that this asymptotic distribution is a mixture of chi-squared distributions. It can be very different from the usual chi-squared approximation of the strong case. We adopt the same self-normalization approach used for constructing the confidence inter-

vals of weak FARIMA model parameters to test the adequacy of weak FARIMA(p, d, q) models. This method has the advantage of avoiding the problem of estimating the asymptotic variance matrix of the joint vector of the least squares estimator and the empirical autocovariances of the noise.

Secondly, we deal in this thesis with the problem of estimating autoregressive models of order 1 endowed with fractional Gaussian noise when the Hurst parameter H is assumed to be known. We study, more precisely, the convergence and the asymptotic normality of the generalized least squares estimator of the autoregressive parameter of these models.

Keywords : Long-memory processes, nonlinear processes, weak FARIMA models, least squares estimator, consistency, asymptotic normality, spectral estimator, self-normalization approach, residual autocorrelations, portmanteau tests of Ljung-Box and Box-Pierce, fractional Brownian motion, fractional Gaussian noise, fractional autoregressive models, generalized least squares estimator.

Table des matières

Remerciements i

Résumé iii

Abstract v

Liste des tableaux 3

Table des figures 7

1 Introduction 8
- 1.1 Quelques définitions des différents types de mémoire . . . 11
- 1.2 Concepts de base et contexte bibliographique . . . 16
- 1.3 Contribution de la thèse . . . 20
- 1.4 Un retour sur quelques propriétés principales des modèles FARIMA . . . 22
- 1.5 Résultats du chapitre 2 . . . 25
  - 1.5.1 Définition de l'estimateur des moindres carrés . . . 26
  - 1.5.2 Propriétés asymptotiques de l'estimateur des moindres carrés . . . 27
  - 1.5.3 Estimation de la matrice de variance asymptotique . . . 29
  - 1.5.4 Construction des intervalles de confiance des paramètres des modèles FARIMA faibles par auto-normalisation . . . 32
- 1.6 Résultats du chapitre 3 . . . 35
  - 1.6.1 La distribution asymptotique des autocorrélations résiduelles . . . 37
  - 1.6.2 Version modifiée du test portmanteau . . . 39
  - 1.6.3 La distribution asymptotique auto-normalisée des autocorrélations résiduelles . . . 42
- 1.7 Résultats du chapitre 4 . . . 45
  - 1.7.1 Estimation des moindres carrés généralisés des modèles FrAR d'ordre 1 lorsque l'exposant de Hurst est connu . . . 45
  - 1.7.2 Propriétés asymptotiques de l'estimateur des moindres carrés généralisés 48

Bibliographie . . . 53

2 Estimating FARIMA models with uncorrelated but non-independent error terms **54**
2.1 Introduction . . . 55
2.2 Least squares estimation . . . 58
2.2.1 Notations and assumptions . . . 58
2.2.2 Asymptotic properties . . . 60
2.3 Estimating the asymptotic variance matrix . . . 61
2.3.1 Estimation of the asymptotic matrix $I(\theta_0)$ . . . 62
2.3.2 A self-normalized approach to confidence interval construction in weak FARIMA models . . . 63
2.4 Numerical illustrations . . . 66
2.4.1 Simulation studies and empirical sizes for confidence intervals . . . 66
2.4.2 Application to real data . . . 68
2.5 Conclusion . . . 68
2.6 Proofs . . . 69
2.6.1 Preliminary results . . . 69
2.6.2 Proof of Theorem 2.1 . . . 73
2.6.3 Proof of Theorem 2.2 . . . 74
2.6.4 Proof of the convergence of the variance matrix estimator . . . 89
2.6.5 Invertibility of the normalization matrix $P_{p+q+1,n}$ . . . 103
2.6.6 Proof of Theorem 2.4 . . . 104
2.6.7 Proof of Theorem 2.5 . . . 107
2.7 Figures and tables . . . 109
2.8 Details on the proof of Theorem 2.1 . . . 121
Bibliographie . . . 129

3 Diagnostic checking in FARIMA models with uncorrelated but non-independent error terms **130**
3.1 Introduction . . . 131
3.2 Assumptions and estimation procedure . . . 133
3.3 Diagnostic checking in weak FARIMA models . . . 136
3.3.1 Asymptotic distribution of the residual autocorrelations . . . 136
3.3.2 Modified version of the portmanteau test . . . 139
3.3.3 Self-normalized asymptotic distribution of the residual autocorrelations 141
3.3.4 Example of explicit calculation of $\Sigma_{\hat{\rho}_m}$ and C_m . . . 143
3.4 Numerical illustrations . . . 148
3.4.1 Simulation studies and empirical sizes . . . 148
3.4.2 Empirical power . . . 149
3.4.3 Illustrative example . . . 150
3.5 Proofs . . . 152
3.5.1 Preliminary results . . . 152
3.5.2 Proof of Proposition 3.1 . . . 154
3.5.3 Proof of Theorem 3.1 . . . 156

3.5.4 Proof of Proposition 3.2 . . . 165
3.5.5 Proof of Theorem 3.4 . . . 166
3.5.6 Proof of Theorem 3.5 . . . 170
3.6 Figures and tables . . . 171
3.7 Supplementary material: additional Monte Carlo experiments . . . 194
3.8 Another one . . . 201
Bibliographie . . . 205

4 Generalized least squares estimation of fractional autoregressive models 206
4.1 Introduction . . . 207
4.2 Generalized least squares estimation of FrAR models of order 1 when the Hurst exponent is known . . . 209
4.2.1 Statement of the problem and notations . . . 209
4.2.2 Asymptotic properties . . . 211
4.3 Numerical illustrations . . . 211
4.3.1 Simulation studies . . . 212
4.4 Proofs . . . 212
4.4.1 Preliminary results . . . 212
4.4.2 Proof of Theorem 4.1 . . . 216
4.4.3 Proof of Theorem 4.2 . . . 228
4.5 Figures . . . 234
Bibliographie . . . 241

Conclusion générale et perspectives x

Liste des tableaux

2.1 Empirical size of standard and modified confidence interval : relative frequencies (in %) of rejection. Modified SN stands for the self-normalized approach. In Modified we use the sandwich estimator of the asymptotic variance Ω of the LSE while in Standard we use $\hat{\Omega}_S$. The number of replications is $N = 1000$. 115

2.2 Fitting a FARIMA$(1, d, 1)$ model to the squares of the 4 daily returns considered. The corresponding $p-$values are reported in parentheses. The last column presents the estimated residual variance. 119

2.3 Modified confidence interval at the asymptotic level $\alpha = 5\%$ for the parameters estimated in Table 2.2. Modified SN stands for the self-normalized approach while Modified corresponds to the confidence interval obtained by using the sandwich estimator of the asymptotic variance Ω of the LSE. 120

3.1 Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a strong FARIMA$(1, d_0, 1)$ defined by (3.34) with $\theta_0 = (0.9, 0.2, d_0)$. The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$. 173

3.2 Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(1, d_0, 1)$ defined by (3.34) with $\theta_0 = (0.9, 0.2, d_0)$ and where $\omega = 0.4$, $\alpha_1 = 0.3$ and $\beta_1 = 0.3$ in (3.20). The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$. 174

3.3 Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(1, d_0, 1)$ defined by (3.34)–(3.35) with $\theta_0 = (0.9, 0.2, d_0)$. The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$. 175

3.4 Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a strong FARIMA$(0, d_0, 0)$ defined by (3.34) with $\theta_0 = (0, 0, d_0)$. The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$. 176

3.5 Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(0, d_0, 0)$ defined by (3.34) with $\theta_0 = (0, 0, d_0)$ and where $\omega = 0.4$, $\alpha_1 = 0.3$ and $\beta_1 = 0.3$ in (3.20). The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$. 177

3.6 Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of weak FARIMA$(0, d_0, 0)$ defined by (3.34)–(3.35) with $\theta_0 = (0, 0, d_0)$. The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$. 178

3.7 Empirical power (in %) of the modified and standard versions of the LB and BP tests in the case of a strong FARIMA$(1, d_0, 2)$ defined by (3.36) with $\theta_0 = (0.9, 1, -0.2, d_0)$. The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$. 179

3.8 Empirical power (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(1, d_0, 2)$ defined by (3.36) with $\theta_0 = (0.9, 1, -0.2, d_0)$ and where $\omega = 0.4$, $\alpha_1 = 0.3$ and $\beta_1 = 0.3$ in (3.20). The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$. 180

3.9 Empirical power (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(1, d_0, 2)$ defined by (3.36)–(3.35) with $\theta_0 = (0.9, 1, -0.2, d_0)$. The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$. 181

3.10 Empirical power (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(1, d_0, 2)$ defined by (3.36) with $\theta_0 = (0., 0.2, 0., d_0)$ and where $\omega = 0.4$, $\alpha_1 = 0.3$ and $\beta_1 = 0.3$ in (3.20). The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$. 182

3.11 Empirical power (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(1, d_0, 2)$ defined by (3.36)–(3.35) with $\theta_0 = (0., 0.2, 0., d_0)$. The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$. 183

3.12 Modified and standard versions of portmanteau tests to check the null hypothesis that the S&P 500 squared returns follow a FARIMA$(1, 0.2338, 1)$ model (3.37). 184

3.13 Modified and standard versions of portmanteau tests to check the null hypothesis that the Nikkei squared returns follow a FARIMA$(0, 0.2132, 0)$ model as in (3.37) with $a = b = 0$. 185

3.14 Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a strong FARIMA$(1, d_0, 1)$ defined by (3.34) with $\theta_0 = (0.9, 0.2, d_0)$. The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$. 195

3.15 Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(1, d_0, 1)$ defined by (3.34) with $\theta_0 = (0.9, 0.2, d_0)$ and where $\omega = 0.4$, $\alpha_1 = 0.3$ and $\beta_1 = 0.3$ in (3.20). The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$. 196

3.16 Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(1, d_0, 1)$ defined by (3.34) with $\theta_0 = (0.9, 0.2, d_0)$ and where $\omega = 0.04$, $\alpha_1 = 0.12$ and $\beta_1 = 0.85$ in (3.20). The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$. 197

3.17 Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a strong FARIMA$(0, d_0, 0)$ defined by (3.34) with $\theta_0 = (0, 0, d_0)$. The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$. 198

3.18 Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(0, d_0, 0)$ defined by (3.34) with $\theta_0 = (0, 0, d_0)$ with $\omega = 0.4$, $\alpha_1 = 0.3$ and $\beta_1 = 0.3$ in (3.35). The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$. 199

3.19 Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(0, d_0, 0)$ defined by (3.34)–(3.35) with $\theta_0 = (0, 0, d_0)$. The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$. 200

3.20 Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(1, d_0, 1)$ defined by (3.34) with $\theta_0 = (0.9, 0.2, d_0)$ and where $\omega = 0.04$, $\alpha_1 = 0.12$ and $\beta_1 = 0.85$ in (3.20). The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$. 202

Table des figures

2.1 LSE of $N = 1,000$ independent simulations of the FARIMA$(1, d, 1)$ model (2.19) with size $n = 2,000$ and unknown parameter $\theta_0 = (a, b, d) = (-0.7, -0.2, 0.4)$, when the noise is strong (left panel), when the noise is semi-strong (2.20) (middle panel) and when the noise is weak of the form (2.21) (right panel). Points (a)-(c), in the box-plots, display the distribution of the estimation error $\hat{\theta}_n(i) - \theta_0(i)$ for $i = 1, 2, 3$. 109

2.2 LSE of $N = 1,000$ independent simulations of the FARIMA$(1, d, 1)$ model (2.19) with size $n = 2,000$ and unknown parameter $\theta_0 = (a, b, d) = (-0.7, -0.2, 0.4)$. The top panels present respectively, from left to right, the Q-Q plot of the estimates $\hat{a}_n$, $\hat{b}_n$ and $\hat{d}_n$ of a, b and d in the strong case. Similarly the middle and the bottom panels present respectively, from left to right, the Q-Q plot of the estimates $\hat{a}_n$, $\hat{b}_n$ and $\hat{d}_n$ of a, b and d in the semi-strong and weak cases. 110

2.3 LSE of $N = 1,000$ independent simulations of the FARIMA$(1, d, 1)$ model (2.19) with size $n = 2,000$ and unknown parameter $\theta_0 = (a, b, d) = (-0.7, -0.2, 0.4)$. The top panels present respectively, from left to right, the distribution of the estimates $\hat{a}_n$, $\hat{b}_n$ and $\hat{d}_n$ of a, b and d in the strong case. Similarly the middle and the bottom panels present respectively, from left to right, the distribution of the estimates $\hat{a}_n$, $\hat{b}_n$ and $\hat{d}_n$ of a, b and d in the semi-strong and weak cases. The kernel density estimate is displayed in full line, and the centered Gaussian density with the same variance is plotted in dotted line. 111

2.4 Comparison of standard and modified estimates of the asymptotic variance Ω of the LSE, on the simulated models presented in Figure 2.1. The diamond symbols represent the mean, over $N = 1,000$ replications, of the standardized errors $n(\hat{a}_n + 0.7)^2$ for (a) (1.90 in the strong case and 4.32 (resp. 1.80) in the semi-strong case (resp. in the weak case)), $n(\hat{b}_n + 0.2)^2$ for (b) (5.81 in the strong case and 11.33 (resp. 8.88) in the semi-strong case (resp. in the weak case)) and $n(\hat{d}_n - 0.4)^2$ for (c) (1.28 in the strong case and 2.65 (resp. 1.40) in the semi-strong case (resp. in the weak case)). 112

2.5 A zoom of the left-middle and left-bottom panels of Figure 2.4 113

2.6 A zoom of the right-middle and right-bottom panels of Figure 2.4 114

2.7 Closing prices of the four stock market indices from the starting date of each index to February 14, 2019. 116
2.8 Returns of the four stock market indices from the starting date of each index to February 14, 2019. 117
2.9 Sample autocorrelations of squared returns of the four stock market indices. . 118

3.1 Returns and the sample autocorrelations of squared returns of the S&P 500. . 171
3.2 Autocorrelation of the FARIMA$(1, 0.2338, 1)$ residuals for the squares of the S&P 500 returns. The horizontal dotted lines (blue color) correspond to the 5% significant limits obtained under the strong FARIMA assumption. The solid lines (red color) and dashed lines (green color) correspond also to the 5% significant limits under the weak FARIMA assumption. The full lines correspond to the asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.1. The dashed lines (green color) correspond to the self-normalized asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.5. 172
3.3 Returns and the sample autocorrelations of squared returns of the Nikkei. . . 186
3.4 Autocorrelation of the FARIMA$(0, 0.2132, 0)$ residuals for the squares of the Nikkei returns. The horizontal dotted lines (blue color) correspond to the 5% significant limits obtained under the strong FARIMA assumption. The solid lines (red color) and dashed lines (green color) correspond also to the 5% significant limits under the weak FARIMA assumption. The full lines correspond to the asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.1. The dashed lines (green color) correspond to the self-normalized asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.5. 187
3.5 Autocorrelation of a realization of size $n = 2,000$ for a weak FARIMA$(1, 0.01, 1)$ model (3.34)–(3.20) with $\theta_0 = (0.9, 0.2, 0.01)$ and where $\omega = 0.04$, $\alpha_1 = 0.13$ and $\beta_1 = 0.88$. The horizontal dotted lines (blue color) correspond to the 5% significant limits obtained under the strong FARIMA assumption. The solid lines (red color) and dashed lines (green color) correspond also to the 5% significant limits under the weak FARIMA assumption. The full lines correspond to the asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.1. The dashed lines (green color) correspond to the self-normalized asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.5. 188

3.6 Autocorrelation of a realization of size $n = 2,000$ for a weak FARIMA$(1, 0.25, 1)$ model (3.34)–(3.20) with $\theta_0 = (0.9, 0.2, 0.25)$ and where $\omega = 0.04$, $\alpha_1 = 0.13$ and $\beta_1 = 0.88$. The horizontal dotted lines (blue color) correspond to the 5% significant limits obtained under the strong FARIMA assumption. The solid lines (red color) and dashed lines (green color) correspond also to the 5% significant limits under the weak FARIMA assumption. The full lines correspond to the asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.1. The dashed lines (green color) correspond to the self-normalized asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.5. 189

3.7 Autocorrelation of a realization of size $n = 2,000$ for a weak FARIMA$(1, 0.49, 1)$ model (3.34)–(3.20) with $\theta_0 = (0.9, 0.2, 0.49)$ and where $\omega = 0.04$, $\alpha_1 = 0.13$ and $\beta_1 = 0.88$. The horizontal dotted lines (blue color) correspond to the 5% significant limits obtained under the strong FARIMA assumption. The solid lines (red color) and dashed lines (green color) correspond also to the 5% significant limits under the weak FARIMA assumption. The full lines correspond to the asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.1. The dashed lines (green color) correspond to the self-normalized asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.5. 190

3.8 Autocorrelation of a realization of size $n = 2,000$ for a weak FARIMA$(0, 0.01, 0)$ model (3.34)–(3.20) with $\theta_0 = (0, 0, 0.01)$ and where $\omega = 0.04$, $\alpha_1 = 0.13$ and $\beta_1 = 0.88$. The horizontal dotted lines (blue color) correspond to the 5% significant limits obtained under the strong FARIMA assumption. The solid lines (red color) and dashed lines (green color) correspond also to the 5% significant limits under the weak FARIMA assumption. The full lines correspond to the asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.1. The dashed lines (green color) correspond to the self-normalized asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.5. 191

3.9 Autocorrelation of a realization of size $n = 2,000$ for a weak FARIMA$(0, 0.25, 0)$ model (3.34)–(3.20) with $\theta_0 = (0, 0, 0.25)$ and where $\omega = 0.04$, $\alpha_1 = 0.13$ and $\beta_1 = 0.88$. The horizontal dotted lines (blue color) correspond to the 5% significant limits obtained under the strong FARIMA assumption. The solid lines (red color) and dashed lines (green color) correspond also to the 5% significant limits under the weak FARIMA assumption. The full lines correspond to the asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.1. The dashed lines (green color) correspond to the self-normalized asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.5. 192

3.10 Autocorrelation of a realization of size $n = 2,000$ for a weak FARIMA$(0, 0.49, 0)$ model (3.34)–(3.20) with $\theta_0 = (0, 0, 0.49)$ and where $\omega = 0.04$, $\alpha_1 = 0.13$ and $\beta_1 = 0.88$. The horizontal dotted lines (blue color) correspond to the 5% significant limits obtained under the strong FARIMA assumption. The solid lines (red color) and dashed lines (green color) correspond also to the 5% significant limits under the weak FARIMA assumption. The full lines correspond to the asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.1. The dashed lines (green color) correspond to the self-normalized asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.5. 193

4.1 Distribution of the generalized least squares estimator of $N = 2,000$ independent simulations of the fractional autoregressive model (4.11) with size $n = 30$ and unknown parameter $a = 0.3$. The Hurst index is assumed to be equal to 0.7. . 234

4.2 Distribution of the generalized least squares estimator of $N = 2,000$ independent simulations of the fractional autoregressive model (4.11) with size $n = 500$ and unknown parameter $a = 0.3$. The Hurst index is assumed to be equal to 0.7. . 235

4.3 Distribution of the generalized least squares estimator of $N = 2,000$ independent simulations of the fractional autoregressive model (4.11) with size $n = 2,000$ and unknown parameter $a = 0.3$. The Hurst index is assumed to be equal to 0.7. 236

4.4 Box plot describing the distribution of the variance estimates of the generalized least squares estimator of $N = 2,000$ independent simulations of the fractional autoregressive model (4.11) with size $n = 30$ and unknown parameter $a = 0.3$. The Hurst index is assumed to be equal to 0.7. The diamond symbol represent the mean of the standardized errors $n(\hat{a}_n - 0.3)^2$, on the simulated models presented in Figure 4.1. 237

4.5 Box plot describing the distribution of the variance estimates of the generalized least squares estimator of $N = 2,000$ independent simulations of the fractional autoregressive model (4.11) with size $n = 500$ and unknown parameter $a = 0.3$. The Hurst index is assumed to be equal to 0.7. The diamond symbol represent the mean of the standardized errors $n(\hat{a}_n - 0.3)^2$, on the simulated models presented in Figure 4.2. 238

4.6 Box plot describing the distribution of the variance estimates of the generalized least squares estimator of $N = 2,000$ independent simulations of the fractional autoregressive model (4.11) with size $n = 2,000$ and unknown parameter $a = 0.3$. The Hurst index is assumed to be equal to 0.7. The diamond symbol represent the mean of the standardized errors $n(\hat{a}_n - 0.3)^2$, on the simulated models presented in Figure 4.3. 239

Chapitre 1

Introduction

Sommaire

1.1 Quelques définitions des différents types de mémoire 11
1.2 Concepts de base et contexte bibliographique 16
1.3 Contribution de la thèse 20
1.4 Un retour sur quelques propriétés principales des modèles FARIMA 22
1.5 Résultats du chapitre 2 25
1.5.1 Définition de l'estimateur des moindres carrés 26
1.5.2 Propriétés asymptotiques de l'estimateur des moindres carrés . . 27
1.5.3 Estimation de la matrice de variance asymptotique 29
1.5.4 Construction des intervalles de confiance des paramètres des modèles FARIMA faibles par auto-normalisation 32
1.6 Résultats du chapitre 3 35
1.6.1 La distribution asymptotique des autocorrélations résiduelles . . 37
1.6.2 Version modifiée du test portmanteau 39
1.6.3 La distribution asymptotique auto-normalisée des autocorrélations résiduelles 42
1.7 Résultats du chapitre 4 45
1.7.1 Estimation des moindres carrés généralisés des modèles FrAR d'ordre 1 lorsque l'exposant de Hurst est connu 45
1.7.2 Propriétés asymptotiques de l'estimateur des moindres carrés généralisés 48
Bibliographie 53

Les séries temporelles présentent souvent une dépendance linéaire [1] et/ou non linéaire [2]. Dans la plupart des cas l'hypothèse d'indépendance des séries temporelles s'avère très restrictive et ne reflète pas la véritable structure des autocorrélations des séries. Généralement, cette hypothèse est uniquement utilisée pour approximer la structure des corrélations des séries chronologiques. D'importantes corrélations pour de faibles retards peuvent parfois être détectées et des processus à mémoire courte peuvent suffire à modéliser la structure de dépendance des séries. Les processus ARMA [3] (voir Box and Jenkins [1970] et Francq and Zakoïan [1998] notamment) ou VARMA [4] pour le cadre multivarié (voir Lütkepohl [2005] et Boubacar Maïnassara [2009]) sont des exemples de processus à mémoire courte.

Cependant, dans de nombreuses disciplines scientifiques et de nombreux domaines appliqués, notamment l'hydrologie, la climatologie, l'économie, la finance et l'informatique, les autocorrélations de certaines séries temporelles diminuent très lentement et sont de somme extrêmement élevée. Ce phénomène peut être dû à plusieurs facteurs, en particulier la non-stationnarité et la dépendance de long terme (dans le cadre stationnaire). Dans ce travail, nous nous intéressons uniquement au deuxième facteur.

Contrairement aux processus à mémoire courte pour lesquels la fonction d'autocovariance tend rapidement vers 0 (éventuellement avec une vitesse de convergence exponentielle), les processus présentant une dépendance de long terme ou tout simplement les processus à mémoire longue sont caractérisés (au sens de la covariance) par une décroissance très lente (plus lente que la décroissance exponentielle) de leurs fonctions d'autocovariances. Cela implique qu'il existe une dépendance non-négligeable entre le présent et tous les points du passé du processus qu'il faut tenir en considération.

La notion de mémoire longue s'est initialement développée avec les travaux de Hurst [1951] pour la conception d'un système de contrôle du Nil avec une série de barrages et de réservoirs. Hurst avait montré que les modèles traditionnellement utilisés en hydrologie [5] sous-estiment très largement la complexité des fluctuations hydrologiques. Plus précisément, il avait trouvé que le rapport entre la capacité d'un réservoir d'eau et l'écart-type de N décharges successives d'un fleuve est proportionnel à N^{H}, où H est une constante comprise entre 0 et 1 appelée exposant

1. La structure de dépendance linéaire est représentée par les fonctions d'autocovariance et d'autocorrélation. Un processus stochastique présente une dépendances linéaire si sa fonction d'autocovariance est non nulle (voir Granger and Joyeux [1980], Hosking [1981], Fox and Taqqu [1986], Dahlhaus [1989], Palma [2007], Beran et al. [2013], Gromykov et al. [2018] entre autres).

2. Les séries temporelles non linéaires présentent des caractéristiques qui ne peuvent pas être modélisées par les processus linéaires : variance qui varie dans le temps, cycles asymétriques, dépendance dans les moments d'ordre supérieur à 1, seuils et ruptures (voir par exemple Tong [1990], Francq and Zakoïan [1998], Francq et al. [2005], Bauwens et al. [2006], Fan and Yao [2008], Francq and Zakoïan [2010], Boubacar Mainassara [2011], Boubacar Mainassara and Francq [2011], Shao [2011], Boubacar Mainassara et al. [2012], Shao [2012], Boubacar Maïnassara [2014], Boubacar Maïnassara and Saussereau [2018]).

3. L'acronyme ARMA désigne *AutoRegressive Moving-Average*, ce qui signifie en français "AutoRégressifs Moyennes-Mobiles".

4. L'acronyme VARMA désigne *Vector AutoRegressive Moving-Average*, ce qui signifie en français "Auto-Régressifs Moyennes-Mobiles Vectoriels".

5. Ces modèles supposent que les niveaux d'eau successifs d'une rivière sont indépendants ou faiblement dépendants au cours du temps. Autrement dit, ces modèles considèrent que les niveaux d'eau successifs d'une rivière sont de mémoire nulle ou de mémoire courte.

de Hurst.

Cette découverte a mis en évidence le décalage entre la théorie empirique standard et la pratique pour certaines séries temporelles caractérisées par une structure de dépendance particulière. Cette constatation fut nommée "phénomène de Hurst".

Le phénomène de Hurst est resté mystérieux (d'un point de vue mathématique) jusqu'aux travaux de Mandelbrot en 1960. Ce dernier a introduit la notion d'autosimilarité stochastique [6] (voir Mandelbrot [1965]), un concept qui a permis de formaliser les bases des processus qui sont ensuite devenus parmi les modèles les plus utilisés dans le domaine de la mémoire longue : le mouvement brownien fractionnaire [7] et ses accroissements stationnaires, le bruit gaussien fractionnaire à temps discret [8].

Le mouvement brownien fractionnaire a été développé pour la première fois par Mandelbrot et Van Ness en 1968 (voir Mandelbrot and Van Ness [1968]). Ce processus dépend d'un paramètre permettant de contrôler la dépendance des séries temporelles. Autrement dit, la valeur prise par le paramètre nous permet de déduire la nature de dépendance qui caractérise la série temporelle. Ce paramètre, qui n'est d'autre que le paramètre de Hurst H introduit précédemment, permet donc de classifier les séries temporelles en fonction de leur structure de dépendance : mémoire courte ou nulle, mémoire longue positive et anti-persistance. Un retour sur ces classes de dépendances sera présenté par la suite.

Étant donné le caractère discret de l'espace temps de plusieurs séries temporelles (par exemple les séries économiques et physiques), Mandelbrot and Wallis [1969a,b,c] ont introduit un analogue en temps discret du mouvement brownien fractionnaire en temps continu, appelé bruit gaussien fractionnaire. Ce processus gaussien stationnaire est défini comme étant l'accroissement du mouvement brownien fractionnaire (voir la note de bas de page 8), il permet également de reproduire cet effet de mémoire longue.

Dans l'optique de modéliser la mémoire longue, Granger and Joyeux [1980] avaient proposé une nouvelle classe de modèles constituée des processus FARIMA [9] $(0,d,0)$. La mémoire longue dans ce cas est contrôlée explicitement par l'opérateur fractionnaire intervenant dans la définition du processus.

En 1981, afin de décrire les fluctuations à court terme des séries temporelles à mémoire longue, Hosking [1981] a généralisé le modèle proposé par Granger et Joyeux en introduisant la classe des modèles FARIMA(p,d,q). Cette classe présente quant à elle une généralisation

6. Un processus stochastique $(X_t)_{t\in\mathbb{R}}$ est dit autosimilaire de paramètre d'autosimilarité $0 < H < 1$ (ou H-autosimilaire) si pour tout $c > 0$, la distribution de $(X_{ct})_{t\in\mathbb{R}}$ est la même que celle de $(c^H X_t)_{t\in\mathbb{R}}$.

7. Le mouvement brownien fractionnaire d'exposant de Hurst $0 < H < 1$, noté $(B_H(t))_{t\in\mathbb{R}}$, est l'unique processus gaussien centré, nul en zéro et continu dont la covariance est donnée par :

$$\mathbb{E}\left[B_H(t)B_H(s)\right] = \frac{\sigma_H^2}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right),$$

où $\sigma_H^2 = \mathrm{Var}\{B_H(1)\}$.

8. Un bruit gaussien fractionnaire $(X_t)_{t\in\mathbb{Z}}$ est le processus des accroissements du mouvement brownien fractionnaire $(B_H(t))_{t\in\mathbb{R}}$, i.e. pour tout $t \in \mathbb{Z}$, $X_t = B_H(t+1) - B_H(t)$.

9. L'acronyme FARIMA désigne *Fractionally AutoRegressive Integrated Moving-Average*, ce qui signifie en français "AutoRégressifs Moyennes-Mobiles Fractionnairement Intégrés".

des modèles ARIMA [10] standards dans lesquels l'exposant de différenciation d est un entier. Les processus FARIMA(p, d, q) ont donc l'avantage de modéliser conjointement le comportement de mémoire longue des séries temporelles et leur dynamique de court terme au travers d'un paramètre d'intégration fractionnaire d et des paramètres autorégressifs et moyennes-mobiles.

1.1 Quelques définitions des différents types de mémoire

Comme indiqué précédemment, les séries temporelles peuvent être classées selon les caractéristiques de leur structure de dépendance. En effet, considérons $(X_t)_{t\in\mathbb{Z}}$ un processus stochastique stationnaire au second-ordre de fonction d'autocovariance $\gamma_X(h) := \text{Cov}(X_t, X_{t-h})$ $(h \in \mathbb{Z})$ et de densité spectrale $f_X(\omega) = (2\pi)^{-1}\sum_{h=-\infty}^{+\infty}\gamma_X(h)\exp(-ih\omega)$ $(\omega \in [-\pi, \pi]$ et $i^2 = -1)$. Le processus $(X_t)_{t\in\mathbb{Z}}$ est dit (a) à mémoire longue, (b) à mémoire courte ou (c) anti-persistant si $f_X(\omega)$ (a) diverge vers $+\infty$, (b) converge vers une constante finie, ou (c) tend vers 0 respectivement quand $|\omega| \to 0$.

Comme $\sum_{h=-\infty}^{+\infty}\gamma_X(h) = 2\pi f_X(0)$, cette dernière définition heuristique est équivalente (dans un sens que nous préciserons ci-dessous) à : $(X_t)_{t\in\mathbb{Z}}$ est (a) à mémoire longue, (b) à mémoire courte ou (c) anti-persistant si (a) $\sum_{h=-\infty}^{+\infty}\gamma_X(h) = +\infty$, (b) $0 < \sum_{h=-\infty}^{+\infty}\gamma_X(h) < +\infty$ ou (c) $\sum_{h=-\infty}^{+\infty}\gamma_X(h) = 0$ respectivement.

Afin d'identifier d'une manière formelle les différents types de structures de dépendances, la définition la plus utilisée dans ce contexte est la suivante (voir Beran et al. [2013]) :

Définition 1.1. *Soit $(X_t)_{t\in\mathbb{Z}}$ un processus stochastique stationnaire au second-ordre de fonction d'autocovariance $\gamma_X(\cdot)$ et de densité spectrale $f_X(\cdot)$ définie pour tout $\omega \in [-\pi, \pi]$ par :*

$$f_X(\omega) = \frac{1}{2\pi}\sum_{h=-\infty}^{+\infty}\gamma_X(h)\exp(-ih\omega).$$

Alors $(X_t)_{t\in\mathbb{Z}}$ est dit (a) à mémoire longue, (b) à mémoire intermédiaire, (c) à mémoire courte ou (d) anti-persistant si

$$f_X(\omega) = Q_f(\omega)|\omega|^{-2d},$$

où $Q_f(\cdot)$ est une fonction positive symétrique variant lentement [11] à l'origine, et (a) $d \in]0, 1/2[$,

10. L'acronyme ARIMA désigne *AutoRegressive Integrated Moving-Average*, ce qui signifie en français "AutoRégressifs Moyennes-Mobiles Intégrés".

11. En principe, toutes les notions habituelles de "variation lente" peuvent être utilisées dans la définition de $Q_f(\cdot)$. Les plus courantes sont les définitions de Karamata et Zygmund données ci-dessous :

Une fonction $Q :]c, +\infty[\to \mathbb{R}$ $(c \geq 0)$ est dite à variation lente à l'infini au sens de Karamata si, pour x suffisamment grand, elle est positive (et mesurable) et si pour tout $r > 0$,

$$\lim_{x\to+\infty}\frac{Q(rx)}{Q(x)} = 1.$$

La fonction est dite à variation lente à l'infini au sens de Zygmund si, pour x suffisamment grand, elle est

(b) $d = 0$ *et* $\lim_{\omega\to 0} Q_f(\omega) = +\infty$, *(c)* $d = 0$ *et* $\lim_{\omega\to 0} Q_f(\omega) = K_f \in]0, +\infty[$ *ou (d)* $d \in]-1/2, 0[$ *respectivement.*

Dans le théorème suivant (voir Beran et al. [2013]), nous présentons l'équivalence entre le comportement de la densité spectrale à l'origine et la décroissance asymptotique de la fonction d'autocovariance.

Théorème 1.1. *Soient* $\gamma(\cdot)$ *et* $f(\cdot)$ *respectivement la fonction d'autocovariance et la densité spectrale d'un processus stationnaire au second-ordre. Alors, nous avons les implications suivantes :*

— *Si*

$$f(\omega) = Q_f(\omega)|\omega|^{-2d} \quad (0 < \omega < \pi),$$

où $d \in]-1/2, 1/2[\setminus\{0\}$ *et* $Q_f(\cdot)$ *est à variation lente à l'origine au sens de Zygmund (voir 11) et à variation bornée*[12] *sur* $]a, \pi[$ *pour tout* $a > 0$, *alors*

$$\lim_{h\to+\infty} \frac{\gamma(h)}{Q_\gamma(h)|h|^{2d-1}} = 1,$$

où

$$Q_\gamma(h) = 2Q_f\left(\frac{1}{h}\right)\Gamma(1-2d)\sin(\pi d),$$

avec $\Gamma(\cdot)$ *la fonction gamma.*

— *Si*

$$\gamma(h) = Q_\gamma(h)|h|^{2d-1},$$

où $Q_\gamma(\cdot)$ *est à variation lente à l'infini au sens de Zygmund et soit* $d \in]0, 1/2[$ *ou* $d \in]-1/2, 0[$ *et* $\sum_{h=-\infty}^{+\infty} \gamma(h) = 0$, *alors*

$$\lim_{\omega\to 0} \frac{f(\omega)}{Q_f(\omega)|\omega|^{-2d}} = 1,$$

avec

$$Q_f(\omega) = \frac{Q_\gamma\left(1/\omega\right)\Gamma(2d)\cos(\pi d)}{\pi}.$$

positive et pour tout $\delta > 0$, il existe $x_0(\delta) > 0$ tel que pour tout $x > x_0(\delta)$, les fonctions $g_1(x) = x^\delta Q(x)$ et $g_2(x) = x^{-\delta} Q(x)$ sont monotones.

De même, une fonction Q est dite à variation lente à l'origine si la fonction $Q^* : x \to Q(x^{-1})$ varie lentement à l'infini.

12. Une fonction f définie sur un ensemble E totalement ordonné et à valeurs dans un espace métrique (F, λ) est dite à variation bornée si

$$\sup_{P=(x_0,x_1,\ldots,x_n)\in\mathcal{P}} \sum_{i=1}^{n} \lambda(f(x_{i-1}), f(x_i)),$$

où $\mathcal{P}$ est l'ensemble de toutes les subdivisions $P = (x_0, x_1, \ldots, x_n)$ d'un intervalle quelconque de E, est finie.

La première implication du théorème 1.1 nous permet, sous certaines conditions, de faire la transition du domaine fréquentiel (avec le comportement de la densité spectrale au voisinage de 0) au domaine temporel (avec le comportement au voisinage de $+\infty$ de la fonction d'autocovariance). Ce résultat nous conduit à une caractérisation plus précise des différents types de structures de dépendances en fonction de la somme des autocovariances. En effet :

Corollaire 1.1. *Si*

$$f(\omega) = Q_f(\omega)|\omega|^{-2d} \quad (0 < \omega < \pi),$$

où $d \in]-1/2, 1/2[$ *et* $Q_f(\cdot)$ *est à variation lente à l'origine au sens de Zygmund et à variation bornée sur* $]a, \pi[$ *pour tout* $a > 0$*, alors, en utilisant le fait que les fonctions à variation lente sont dominées par les fonctions puissances, nous obtenons que :*

— *Pour* $-1/2 < d < 0$,

$$\sum_{h=-\infty}^{+\infty} \gamma(h) = 2\pi f(0) = 0.$$

— *Pour* $0 < d < 1/2$,

$$\sum_{h=-\infty}^{+\infty} \gamma(h) = 2\pi \lim_{\omega \to 0} f(\omega) = +\infty.$$

— *Pour* $d = 0$,

$$0 < \sum_{h=-\infty}^{+\infty} \gamma(h) = 2\pi f(0) = 2\pi K_f < \infty$$

si $0 < \lim_{\omega \to 0} Q_f(\omega) = K_f < \infty$ *et*

$$\sum_{h=-\infty}^{+\infty} \gamma(h) = 2\pi \lim_{\omega \to 0} f(\omega) = +\infty$$

si $\lim_{\omega \to 0} Q_f(\omega) = +\infty$.

D'après ce dernier corollaire, il est clair que la caractérisation de la dépendance linéaire par la densité spectrale est plus élégante que celle via la fonction d'autocovariance. Cette conclusion est justifiée par le fait que l'équation $f(\omega) = Q_f(\omega)|\omega|^{-2d}$ permet d'englober les quatre cas de dépendance (la mémoire longue, la mémoire intermédiaire, la mémoire courte et l'antipersistance).

Dans l'exemple suivant, nous revenons sur les différentes structures de dépendance qu'un bruit gaussien fractionnaire peut présenter. Nous donnerons plus de détails sur le comportement asymptotique de la fonction d'autocovariance de ce processus.

Exemple 1.1. *Soit* $(B_{\mathsf{H}}(t))_{t\in\mathbb{R}}$ *un mouvement brownien fractionnaire d'indice de Hurst* $\mathsf{H} \in]0, 1[$ *(voir la note de bas de page 7) et notons par* $(\epsilon_t^{\mathsf{H}})_{t\in\mathbb{Z}}$ *le processus à temps discret des accroissements de* $(B_{\mathsf{H}}(t))_{t\in\mathbb{R}}$ *(i.e. pour tout* $t \in \mathbb{Z}$, $\epsilon_t^{\mathsf{H}} = B_{\mathsf{H}}(t+1) - B_{\mathsf{H}}(t)$*). Le processus*

$(\epsilon_t^{\mathsf{H}})_{t\in\mathbb{Z}}$ *est appelé bruit gaussien fractionnaire. La fonction d'autocovariance du bruit gaussien fractionnaire* $(\epsilon_t^{\mathsf{H}})_{t\in\mathbb{Z}}$ *est donnée pour tout* $k \in \mathbb{Z}$ *par :*

$$\gamma_{\epsilon^{\mathsf{H}}}(k) := \mathrm{Cov}(\epsilon_1^{\mathsf{H}}, \epsilon_{1+k}^{\mathsf{H}}) = \frac{\sigma_{\mathsf{H}}^2}{2}\left(|k+1|^{2\mathsf{H}} + |k-1|^{2\mathsf{H}} - 2|k|^{2\mathsf{H}}\right), \tag{1.1}$$

où $\sigma_{\mathsf{H}}^2 = \mathrm{Var}(B_{\mathsf{H}}(1))$. *Cette structure spécifique de covariance est une conséquence immédiate de l'autosimilarité de* $(B_{\mathsf{H}}(t))_{t\in\mathbb{R}}$. *En effet, en utilisant cette propriété de conservation de loi par changement temporel (voir la note de bas de page 6), nous obtenons que, pour tout* $t \geq s > 0$,

$$\mathrm{Cov}\left\{B_{\mathsf{H}}(t), B_{\mathsf{H}}(t)\right\} = \mathrm{Var}\left\{B_{\mathsf{H}}(t)\right\} = \mathrm{Var}\left\{t^{\mathsf{H}} B_{\mathsf{H}}(1)\right\} = t^{2\mathsf{H}}\sigma_{\mathsf{H}}^2$$

et

$$\mathrm{Var}\left\{B_{\mathsf{H}}(t) - B_{\mathsf{H}}(s)\right\} = \mathrm{Var}\left\{B_{\mathsf{H}}(t-s)\right\} = \sigma_{\mathsf{H}}^2 (t-s)^{2\mathsf{H}}.$$

Comme $\mathrm{Var}\{B_{\mathsf{H}}(t) - B_{\mathsf{H}}(s)\} = \mathrm{Var}\{B_{\mathsf{H}}(t)\} + \mathrm{Var}\{B_{\mathsf{H}}(s)\} - 2\mathrm{Cov}\{B_{\mathsf{H}}(t), B_{\mathsf{H}}(s)\}$, *nous en déduisons que pour tout* $t, s \in \mathbb{R}$,

$$\mathrm{Cov}\{B_{\mathsf{H}}(t), B_{\mathsf{H}}(s)\} = \frac{\sigma_{\mathsf{H}}^2}{2}\left(|t|^{2\mathsf{H}} + |s|^{2\mathsf{H}} - |t-s|^{2\mathsf{H}}\right).$$

Cette dernière équation (due à l'autosimilarité de $(B_{\mathsf{H}}(t))_{t\in\mathbb{R}}$*) nous permet de déduire, par un simple calcul de covariance, le résultat donné dans (1.1).*

Remarquons que dans le cas où $\mathsf{H} = 1/2$, *toutes les autocovariances du processus* $(\epsilon_t^{\mathsf{H}})_{t\in\mathbb{Z}}$ *sont nulles sauf pour le retard 0. Vu le caractère gaussien du bruit* $(\epsilon_t^{\mathsf{H}})_{t\in\mathbb{Z}}$, *cette décorrélation implique l'indépendance du processus. Le bruit gaussien fractionnaire est donc sans mémoire (ou de mémoire nulle) lorsque le paramètre de Hurst* H *est égal à* $1/2$.

Supposons maintenant que $\mathsf{H} \in]0, 1/2[\cup]1/2, 1[$ *dans toute la suite de l'exemple.*

Un développement de Taylor de la fonction $\ell : x \to (1-x)^{2\mathsf{H}} - 2 + (1+x)^{2\mathsf{H}}$ *au point 0 implique que pour* k *suffisamment grand,*

$$\gamma_{\epsilon^{\mathsf{H}}}(k) = \frac{\sigma_{\mathsf{H}}^2}{2} k^{2\mathsf{H}} \ell(1/k) = \sigma_{\mathsf{H}}^2 \mathsf{H}(2\mathsf{H}-1) k^{2\mathsf{H}-2} + o(k^{2\mathsf{H}-2}).$$

En prenant $d = \mathsf{H} - 1/2$ *et* $Q_\gamma(\cdot) = \sigma_{\mathsf{H}}^2 \mathsf{H}(2\mathsf{H}-1)$, *le deuxième point du théorème 1.1 implique que la densité spectrale* $f_{\epsilon^{\mathsf{H}}}$ *du processus* $(\epsilon_t^{\mathsf{H}})_{t\in\mathbb{Z}}$ *vérifie, pour* $\omega \to 0$,

$$f_{\epsilon^{\mathsf{H}}}(\omega) \sim \frac{\sigma_{\mathsf{H}}^2 \mathsf{H}(2\mathsf{H}-1)\Gamma(2\mathsf{H}-1)\sin(\pi\mathsf{H})}{\pi} |\omega|^{1-2\mathsf{H}} = \frac{\sigma_{\mathsf{H}}^2 \Gamma(2\mathsf{H}+1)\sin(\pi\mathsf{H})}{2\pi} |\omega|^{1-2\mathsf{H}}.$$

Ainsi, la définition 1.1 nous permet de conclure que :

- *lorsque* $0 < \mathsf{H} < 1/2$, *le processus* $(\epsilon_t^{\mathsf{H}})_{t\in\mathbb{Z}}$ *est anti-persistant.*
- *quand* $1/2 < \mathsf{H} < 1$, *le bruit gaussien fractionnaire* $(\epsilon_t^{\mathsf{H}})_{t\in\mathbb{Z}}$ *présente une longue dépendance linéaire.*

Nous illustrons dans l'exemple suivant le cas de la dépendance intermédiaire. Par définition, ce type de dépendance se situe entre la dépendance de long terme et celle de court terme (voir 1.1).

Exemple 1.2. *Soit $(X_t)_{t\in\mathbb{Z}}$ un processus stochastique stationnaire au second-ordre de densité spectrale $f_X(\cdot)$ définie par :*

$$f_X(\omega) = \begin{cases} (1/2\pi)\ln|\pi/\omega| & si\ \omega \in [-\pi,\pi] \\ 0 & sinon \end{cases} = Q_f(\omega).$$

La fonction $Q_f(\cdot)$ est symétrique positive variant lentement à l'origine. De plus, $Q_f(\omega) \to +\infty$ quand $\omega \to 0$. Le processus $(X_t)_{t\in\mathbb{Z}}$ est donc à dépendance intermédiaire d'après la définition 1.1. La fonction d'autocovariance $\gamma_X(\cdot)$ de $(X_t)_{t\in\mathbb{Z}}$ est définie pour tout $k \in \mathbb{Z}$ par :

$$\gamma_X(k) = \int_{-\pi}^{\pi} f_X(\omega) e^{ik\omega} d\omega.$$

En utilisant la parité de $f_X(\cdot)$, la variance du processus $(X_t)_{t\in\mathbb{Z}}$ est donc donnée par :

$$\mathrm{Var}(X_t) = \gamma_X(0) = \int_{-\pi}^{\pi} f_X(\omega)d\omega = 2\int_0^{\pi} f_X(\omega)d\omega = \frac{1}{\pi}\left\{\pi\ln(\pi) - \int_0^{\pi}\ln(\omega)d\omega\right\} = 1.$$

Cette dernière égalité indique que la fonction d'autocovariance du processus $(X_t)_{t\in\mathbb{Z}}$ est celle de ses autocorrélations . Quand $k > 0$, la parité de $f_X(\cdot)$ implique que

$$\begin{aligned}\gamma_X(k) &= \int_0^{\pi} f_X(\omega)\left\{e^{ik\omega} + e^{-ik\omega}\right\} d\omega = 2\int_0^{\pi} f_X(\omega)\cos(k\omega)d\omega \\ &= \frac{1}{\pi}\int_0^{\pi}\cos(k\omega)\left\{\ln(\pi) - \ln(\omega)\right\} d\omega = -\frac{1}{\pi}\int_0^{\pi}\cos(k\omega)\ln(\omega)d\omega \\ &= \frac{1}{k\pi}\int_0^{\pi}\frac{\sin(k\omega)}{\omega}d\omega = \frac{1}{k\pi}\mathrm{Si}(k\pi), \end{aligned} \tag{1.2}$$

où $\mathrm{Si}(\cdot)$ est la fonction sinus intégral. Lorsque $k \to +\infty$, nous obtenons l'intégrale de Dirichlet

$$\lim_{k\to+\infty}\mathrm{Si}(k\pi) = \int_0^{+\infty}\frac{\sin(\omega)}{\omega}d\omega = \frac{\pi}{2}.$$

Ainsi, de (1.2) et en utilisant ce dernier résultat, la fonction d'autocovariance du processus $(X_t)_{t\in\mathbb{Z}}$ vérifie l'équivalence suivante :

$$\gamma_X(k) \sim \frac{1}{2k} \quad (k \to +\infty).$$

Par conséquent, quand $n \to +\infty$, nous déduisons que

$$\sum_{k=-(n-1)}^{n-1}\gamma_X(k) \sim \ln(n).$$

1.2 Concepts de base et contexte bibliographique

La modélisation simultanée de la dynamique à court terme et de la dépendance de long terme a permis aux processus FARIMA(p, d, q) d'occuper une place centrale dans les classes des modèles à mémoire longue et des modèles à mémoire courte. En effet, lorsque le paramètre de différentiation fractionnaire d est égal à 0, nous retrouvons les modèles ARMA standards à mémoire courte. Ces processus ont aussi l'avantage d'être facilement applicables (en économie par exemple) dans la mesure où ils constituent une extension directe des processus ARIMA. Ces modèles sont généralement utilisés avec des hypothèses fortes sur le bruit qui en limitent la généralité. Ainsi, nous appelons FARIMA forts les modèles standards dans lesquels le terme d'erreur est supposé être une suite indépendante et identiquement distribuée (*i.e.* iid), et nous parlons de modèles FARIMA faibles quand les hypothèses sur le bruit sont moins restrictives. Nous parlons également de modèles semi-forts quand le bruit est supposé être une différence de martingale [13]. La distinction entre modèles FARIMA forts, semi-forts ou faibles n'est donc qu'une question d'hypothèses sur le bruit. Il est important de signaler que les contraintes sur le bruit dans un modèle FARIMA faible sont moins restrictives que celles exigées dans le cadre semi-fort ou fort. De plus, en exploitant le fait qu'un processus centré, intégrable et indépendant est une différence de martingale par rapport à la filtration naturelle [14], nous en déduisons les inclusions suivantes :

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

Relâcher l'hypothèse d'indépendance et celle considérant le bruit comme une différence de martingale permettent d'élargir le champs d'application des modèles FARIMA faibles en couvrant une large classe de modèles non linéaires.

Afin de définir rigoureusement les notions de linéarité et de non linéarité d'un modèle, commençons par rappeler la décomposition de Wold [1938] qui affirme que tout processus

13. Soient $(X_t)_{t\in\mathbb{N}}$ une suite de variables aléatoires réelles et $(\mathcal{F}_t)_{t\in\mathbb{N}}$ une filtration (une suite croissante de tribus). On dit que $(X_t)_{t\in\mathbb{N}}$ est une martingale par rapport à $(\mathcal{F}_t)_{t\in\mathbb{N}}$ si :

1. X_t est $\mathcal{F}_t$-mesurable pour tout entier t (on dit que $(X_t)_{t\in\mathbb{N}}$ est adaptée à la filtration $(\mathcal{F}_t)_{t\in\mathbb{N}}$),
2. X_t est intégrable pour tout entier t (*i.e.* $\mathbb{E}|X_t| < \infty$, $\forall t \in \mathbb{N}$),
3. $\mathbb{E}[X_{t+1}|\mathcal{F}_t] = X_t$.

Soient $(Y_t)_{t\in\mathbb{N}}$ une suite de variables aléatoires réelles et $(\mathcal{F}_t)_{t\in\mathbb{N}}$ une filtration. On dit que la suite $(Y_t, \mathcal{F}_t)_{t\in\mathbb{N}}$ est une différence de martingale si :

1. $(Y_t)_{t\in\mathbb{N}}$ est adaptée à la filtration $(\mathcal{F}_t)_{t\in\mathbb{N}}$,
2. Y_t est intégrable pour tout entier t,
3. $\mathbb{E}[Y_{t+1}|\mathcal{F}_t] = 0$.

14. Soit $(X_t)_{t\in\mathbb{N}}$ une suite de variables aléatoires. On dit que $(\mathcal{F}_t)_{t\in\mathbb{N}}$ définie pour tout $t \in \mathbb{N}$ par $\mathcal{F}_t = \sigma(X_0, \ldots, X_n)$, est la filtration naturelle de la suite $(X_t)_{t\in\mathbb{N}}$.

stationnaire au second-ordre est la somme d'un processus régulier [15] et d'un processus singulier [16] orthogonaux entre eux. Cette décomposition est unique (voir Palma [2007]).

D'après le théorème de la décomposition de Wold, un processus $(X_t)_{t\in\mathbb{Z}}$ stationnaire au second-ordre et purement non déterministe peut être exprimé pour tout $t \in \mathbb{Z}$ sous la forme :

$$X_t = \sum_{i=0}^{+\infty} \psi_i \epsilon_{t-i}, \tag{1.3}$$

où $\psi_0 = 1$, $\sum_{i=0}^{+\infty} \psi_i^2 < \infty$ et $(\epsilon_t)_{t\in\mathbb{Z}}$ est un bruit blanc faible de variance $\sigma_\epsilon^2 \in]0, +\infty[$ (voir Brockwell and Davis [1991]).
Rappelons qu'un bruit blanc faible est un processus non autocorrélé d'espérance nulle et de variance finie. Il est important de noter qu'un bruit blanc faible est un processus qui peut contenir (par définition) des dépendances non linéaires très générales. Aucune hypothèse d'indépendance n'est imposée dans la définition du bruit blanc faible. Les variables aux différentes dates sont seulement non corrélées.

Nous parlons également d'un bruit blanc fort lorsque le processus vérifie toutes les contraintes intervenant dans la définition d'un bruit blanc faible et lorsqu'il est supposé de plus indépendant et identiquement distribué. Nous utilisons l'appellation bruit blanc semi-fort pour désigner que, additionnellement aux hypothèses standards définissant le bruit blanc faible, le bruit est une différence de martingale.

Dans (1.3), le processus $(\epsilon_t)_{t\in\mathbb{Z}}$ est l'innovation linéaire du processus $(X_t)_{t\in\mathbb{Z}}$. Nous parlons de modèles linéaires lorsque le processus d'innovation $(\epsilon_t)_{t\in\mathbb{Z}}$ est iid et de modèles non linéaires dans le cas contraire. Dans le cadre des modèles à mémoire courte, Francq et al. [2005] et Francq and Zakoïan [1998] ont montré qu'il existe des processus très variés admettant, à la fois, des représentations non linéaires et linéaires de type ARMA, pourvu que les hypothèses sur le bruit du modèle ARMA soient suffisamment peu restrictives.

Il est courant que les séries temporelles présentent des dépendances non linéaires difficilement identifiables. L'utilisation dans ce cas des modèles FARIMA forts ou semi-forts peut conduire à des imprécisions assez conséquentes dans la modélisation, et donc l'obtention de résultats totalement aberrants.

Afin de modéliser les séries temporelles non linéaires et à mémoire longue sans se poser la question sur la structure de non linéarité correspondante, les processus FARIMA faibles semblent

15. Soit $(X_t)_{t\in\mathbb{Z}}$ un processus stationnaire au second-ordre et considérons, pour tout $t \in \mathbb{Z}$, $\mathcal{H}_t^X = \text{vect}\{X_s : s < t\}$ le passé linéaire de X_t. Le processus $(X_t)_{t\in\mathbb{Z}}$ est dit régulier ou purement non déterministe si et seulement si
$$\mathcal{H}_{-\infty}^X = \{0\}.$$

16. Un processus $(X_t)_{t\in\mathbb{Z}}$ stationnaire au second-ordre est dit singulier ou déterministe si et seulement si, pour tout $t \in \mathbb{Z}$,
$$X_t \in \mathcal{H}_{-\infty}^X.$$

être une solution raisonnable et justifiée dans ce contexte. En effet, comme mentionné précédemment, ces modèles ont l'avantage de tenir en compte conjointement l'effet de mémoire longue et celui issu de la dynamique de mémoire courte. De plus, le fait d'utiliser que les hypothèses standards définissant le bruit sans imposer d'autres contraintes sur la structure de dépendance des erreurs permet à ces modèles d'ajuster des séries temporelles avec des dynamiques non linéaires très générales.

Un processus $(X_t)_{t\in\mathbb{Z}}$ centré et stationnaire au second-ordre, à valeurs réelles, admet une représentation FARIMA(p, d_0, q) faible si, pour tout $t \in \mathbb{Z}$,

$$a(L)(1-L)^{d_0}X_t = b(L)\epsilon_t, \tag{1.4}$$

où $d_0 \in]-1/2, 1/2[$ est le paramètre de différentiation fractionnaire (dit aussi paramètre de longue mémoire), L est l'opérateur retard (*i.e.* pour tout $t \in \mathbb{Z}$ et pour tout entier k, $L^k X_t = X_{t-k}$) et $a(L) = 1 - \sum_{i=1}^{p} a_i L^i$ et $b(L) = 1 - \sum_{i=1}^{q} b_i L^i$ sont respectivement l'opérateur autorégressif et moyennes-mobiles [17]. L'innovation linéaire $(\epsilon_t)_{t\in\mathbb{Z}}$ est supposée être un processus stationnaire au second-ordre vérifiant l'hypothèse suivante :

(H0) : $\mathbb{E}[\epsilon_t] = 0$, $\mathrm{Var}(\epsilon_t) = \sigma_\epsilon^2 \in]0, +\infty[$ et pour tout $t \in \mathbb{Z}$ et $h \neq 0$, $\mathrm{Cov}(\epsilon_t, \epsilon_{t-h}) = 0$.

Le problème, qui nous préoccupera dans un premier temps, sera l'analyse statistique des modèles FARIMA faibles. Dans la littérature, la théorie asymptotique de l'estimation et de la validation des modèles FARIMA est principalement limitée aux modèles FARIMA forts et semi-forts. Cette restriction est expliquée par le fait que les modèles non linéaires sont généralement difficiles à identifier et à implémenter.

L'estimateur du maximum de vraisemblance de Whittle dans le domaine fréquentiel est couramment utilisé pour estimer les paramètres des modèles FARIMA (voir par exemple Dahlhaus [1989], Fox and Taqqu [1986], Taqqu and Teverovsky [1997], Giraitis and Surgailis [1990]). L'étude des propriétés asymptotiques de cet estimateur est suffisamment développée dans le cas où les erreurs sont supposées indépendantes et identiquement distribuées et dans le cadre où le bruit est considéré être une différence de martingale (voir Beran [1995], Beran et al. [2013], Palma [2007], Baillie et al. [1996], Ling and Li [1997], Hauser and Kunst [1998], entre autres). Tous ces travaux se restreignent au cas d'absence de dépendance non linéaire ou, au mieux, au cas où la dépendance a une structure bien identifiable. Par exemple, dans la modélisation des séries temporelles financières, afin de capturer l'hétéroscédasticité conditionnelle, il est courant que les innovations des modèles FARIMA sont supposées posséder une structure GARCH [18] (voir, par exemple, Baillie et al. [1996], Hauser and Kunst [1998]). Les modèles GARCH, introduits par Bollerslev [1986], constituent une généralisation des modèles autorégressifs conditionnellement hétéroscédastiques (ARCH) introduits par Engle [1982]. Ces processus

17. Les opérateurs $a(L)$ et $b(L)$ représentent le comportement de mémoire courte du modèle. Par convention $a_0 = b_0 = 1$.

18. L'acronyme GARCH désigne *Generalized AutoRegressive Conditional Heteroskedasticity*, ce qui signifie en français "AutoRégressifs Conditionnellement Hétéroscédastiques Généralisés".

sont des différences de martingales. Ils occupent une place centrale dans la classe des modèles conditionnellement hétéroscédastiques à mémoire courte (voir Francq and Zakoïan [2010]).

Il ne fait donc aucun doute qu'il est important de disposer d'une procédure d'inférence pour les paramètres des modèles FARIMA lorsque les erreurs sont sujettes à une dépendance non linéaire inconnue très générale. Nous connaissons peu de choses lorsque l'hypothèse de différence de martingale est relâchée.

La théorie asymptotique de l'estimation des modèles FARIMA faibles a été très peu étudiée. À notre connaissance, Shao [2010b, 2012] sont les seuls travaux dans ce contexte. Sous de faibles hypothèses sur le bruit, l'auteur a obtenu la normalité asymptotique de l'estimateur de Whittle (voir Whittle [1953]) sans pour autant dresser le problème d'inférence des modèles FARIMA faibles. Cette limitation est justifiée par le fait que, dans le cadre des modèles FARIMA induits par un bruit blanc faible présentant des dépendances non linéaires très générales et non communes, la matrice de variance asymptotique de l'estimateur de Whittle dépend de l'intégrale de la densité spectrale des cumulants joints d'ordre quatre du bruit.

En utilisant des méthodes non-paramétriques qui dépendent entièrement d'un paramètre appelé fenêtre, il est possible de construire une estimation de cette intégrale, mais il n'existe aucune indication sur le choix de cette fenêtre dans les procédures d'estimation (voir Shao [2012], Taniguchi [1982], Keenan [1987], Chiu [1988] pour plus de détails). Cette difficulté est causée par la structure de dépendance du bruit. En effet, dans le cas des modèles FARIMA forts, un estimateur consistent de la matrice de covariance asymptotique est obtenu indépendamment du choix de la fenêtre. Lorsque le bruit présente une dynamique non linéaire inhabituelle, aucune formule explicite pour un estimateur convergent de la matrice de variance asymptotique ne semble être fournie dans la littérature (voir Shao [2012]).

Afin de valider ou d'invalider le choix des ordres p et q des modèles FARIMA(p, d, q) forts, Li and McLeod [1986] ont proposé une version modifiée des tests portmanteau[19] pour vérifier la significativité globale des autocorrélations résiduelles des modèles FARIMA(p, d, q). La théorie asymptotique de la statistique de test de Box and Pierce [1970] et celle des versions modifiées proposées par Ljung and Box [1978] (pour le cas des modèles ARMA forts) et Li and McLeod [1986] (pour la validation des modèles FARIMA forts) a été établie sous l'hypothèse iid sur le bruit. Ces résultats peuvent être invalides lorsque le processus d'innovation est non corrélé mais dépendant (voir par exemple Romano and Thombs [1996], Francq et al. [2005], Boubacar Maïnassara and Saussereau [2018], Zhu and Li [2015], Lobato et al. [2001], Lobato et al. [2002]).

De même que l'étape de l'estimation, la validation des modèles FARIMA a été essentiellement développée dans le cadre fort et semi-fort. Shao [2011] a récemment étudié l'adéquation de modèles de séries temporelles à mémoire longue avec des erreurs ayant une structure d'une

19. Ces tests ont été introduits par Box, G. E. P. et Pierce, D. A. en 1970 (voir Box and Pierce [1970]) pour vérifier la significativité des modèles ARMA(p, q) forts. Ces tests sont basés sur les autocorrélations empiriques résiduelles. Ils permettent précisément de valider ou d'invalider le choix des ordres p et q selon si les autocorrélations des résidus sont proches (au sens statistique) de 0 ou non.

différence de martingale et présentant une hétéroscédasticité conditionnelle non-paramétrique. Cet auteur a également généralisé, sous de faibles hypothèses sur le bruit, le test [20] proposé par Hong [1996]. Comme mentionné par l'auteur, son travail n'est valable que lorsque le paramètre de mémoire est dans $[0, 1/4[$ (voir la remarque 3.3 dans Shao [2011]). Toujours dans le contexte semi-fort, Ling and Li [1997] ont étudié la théorie asymptotique de la statistique de test de Box and Pierce [1970] pour les modèles FARIMA-GARCH en considérant une forme paramétrique pour le modèle GARCH.

À notre connaissance, il n'existe aucune méthodologie permettant de vérifier l'adéquation des modèles FARIMA lorsque les erreurs correspondantes présentent une hétéroscédasticité conditionnelle inconnue. Nous pensons que cela est dû à la difficulté qui se pose lorsqu'il faut estimer la matrice de covariance asymptotique.

1.3 Contribution de la thèse

L'objectif principal de la thèse est d'étendre, dans un premier temps, les procédures standards d'estimation et d'inférence des modèles ARMA à mémoire courte induits par des bruits à dépendances non linéaires (voir Francq and Zakoïan [1998], Francq and Zakoïan [2000], Francq and Zakoïan [2005], Francq et al. [2005], Boubacar Mainassara et al. [2012]) au cas des modèles FARIMA faibles. Dans un second temps, une autre partie de la thèse a été consacrée au problème de la validation des modèles FARIMA avec un bruit non corrélé mais non indépendant. Dans un deuxième temps, le problème de l'estimation des modèles autorégressifs avec un bruit gaussien fractionnaire a été abordé.

Dans le deuxième chapitre, nous établissons les propriétés asymptotiques de l'estimateur des moindres carrés (noté LSE pour Least Squares Estimator) $\hat{\theta}_n$ des paramètres des modèles FARIMA(p, d, q) faibles. Nous accordons ensuite une attention particulière au problème de l'estimation de la matrice de variance asymptotique de cet estimateur. Cette matrice est de la forme "sandwich" $\Omega := J^{-1}(\theta_0)I(\theta_0)J^{-1}(\theta_0)$, et peut être très différente de la variance asymptotique du LSE dans le cadre standard (*i.e.* dans le cas des modèles FARIMA forts) dont l'expression est donnée par $\Omega_S := 2J^{-1}(\theta_0)$. Nous explicitons les dérivées des résidus en fonction des paramètres du modèle FARIMA(p, d, q). Cela nous permettra après de donner les expressions des matrices $I(\theta_0)$ et $J(\theta_0)$ impliquées dans la définition de la variance asymptotique Ω en fonction du paramètre de mémoire longue, des paramètres autorégressifs et moyennes-mobiles et des moments d'ordre deux et quatre du bruit. Nous en déduisons un estimateur de Ω, dont nous établissons la convergence.

Nous conclurons ce chapitre par la construction d'intervalles de confiance des paramètres des modèles FARIMA(p, d, q) faibles en procédant par une approche d'auto-normalisation. Cette technique est basée sur l'introduction d'une matrice dépendante du bruit $P_{p+q+1,n}$, dite d'auto-

20. La statistique de test dans ce cas est définie comme étant la distance quadratique de l'estimateur à noyaux de la densité spectrale normalisée (voir Priestley [1981]) du bruit et de la valeur exacte sous l'hypothèse nulle de cette dernière.

normalisation, et sur l'utilisation de la convergence faible de $\|\sqrt{n}P_{p+q+1,n}^{-1/2}(\hat{\theta}_n - \theta_0)\|_{\mathbb{R}^n}^2$. Nous illustrons enfin nos résultats sur des études empiriques basées sur des expériences de Monte Carlo. Les logiciels de prévision qui utilisent la méthodologie de Box et Jenkins (identification, estimation et validation de modèles FARIMA forts) ne conviennent pas aux modèles FARIMA faibles. Un travail d'adaptation des sorties de ces logiciels a été réalisé.

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

Une deuxième méthode alternative basée sur une approche d'auto-normalisation est proposée dans ce chapitre pour tester l'adéquation des modèles FARIMA faibles. L'utilisation de cette technique a pour principal objectif d'éviter l'estimation de la matrice de variance asymptotique du vecteur joint des autocorrélations résiduelles et de l'estimateur des moindres carrés. Les valeurs critiques de la statistique de test introduite sont tabulées dans Lobato [2001]. Cette méthode est en quelque sorte proche de la méthode standard où la loi de la statistique de test de portmanteau est approximée par un khi-deux. Nous établissons dans cette partie du travail l'extension des résultats de Boubacar Maïnassara and Saussereau [2018] sur la validation des modèles VARMA faibles en utilisant la même approche d'auto-normalisation au cas des modèles FARIMA faibles.

Les résultats théoriques de ce chapitre sont illustrés par des simulations et des applications aux données réelles.

Dans le quatrième chapitre, nous traitons le problème de l'estimation des modèles autorégressifs induits par un bruit gaussien fractionnaire (voir la note de bas de page 8) lorsque l'exposant de Hurst est supposé connu. Le caractère de longue mémoire du processus est contrôlé dans ce cas implicitement par le bruit. La dynamique de mémoire courte est, quant à elle, modélisée par la partie autorégressive du modèle. Nous établissons la convergence et la normalité asymptotique de l'estimateur des moindres carrés généralisés des paramètres du modèle et nous illustrons enfin nos résultats sur des études empiriques basées sur des expériences de Monte Carlo.

Le dernier chapitre propose des perspectives pour des développements futurs.

1.4 Un retour sur quelques propriétés principales des modèles FARIMA

Avant d'énoncer nos premiers principaux résultats sur le comportement asymptotique de l'estimateur des moindres carrés des modèles FARIMA faibles, nous faisons d'abord un retour sur quelques propriétés fondamentales des processus FARIMA pour plus de familiarisation. Ces propriétés sont communes entre les trois classes des modèles FARIMA évoquées ci-dessus. Nous commençons par rappeler les conditions d'obtention d'une solution stationnaire du processus défini dans (1.4) et nous revenons sur l'unicité, l'inversibilité et la causalité de cette solution [21].

Nous explicitons dans le théorème suivant les conditions permettant l'obtention d'une solution stationnaire unique, causale et inversible du processus défini dans (1.4) :

Théorème 1.2. *Soit $(X_t)_{t\in\mathbb{Z}}$ un processus stochastique admettant une représentation FARIMA comme dans (1.4). Supposons que les polynômes $a(\cdot)$ et $b(\cdot)$ correspondant respectivement à la partie autorégressive et moyennes-mobiles ne partagent aucune racine commune et que $d_0 \in]-1/2, 1/2[$.*

1. *Si $a(z) \neq 0$ pour $|z| = 1$, alors il existe une unique solution stationnaire purement non-déterministe $(X_t)_{t\in\mathbb{Z}}$ de (1.4) donnée pour tout $t \in \mathbb{Z}$ par*

$$X_t = \sum_{j=-\infty}^{+\infty} \psi_j \epsilon_{t-j},$$

   où $\psi(z) = (1-z)^{-d_0} b(z)/a(z)$.
2. *La solution $(X_t)_{t\in\mathbb{Z}}$ est causale si et seulement si les racines de $a(\cdot)$ sont en dehors du disque unité fermé $\{z : |z| \leq 1\}$.*
3. *La solution $(X_t)_{t\in\mathbb{Z}}$ est inversible si et seulement si les racines de $b(\cdot)$ sont en dehors du disque unité fermé $\{z : |z| \leq 1\}$.*
4. *Si la solution $(X_t)_{t\in\mathbb{Z}}$ est causale et inversible (i.e. si $a(z) \neq 0$ pour $|z| \leq 1$ et $b(z) \neq 0$ pour $|z| \leq 1$) alors sa fonction d'autocorrélation $\rho_X(\cdot)$ et sa densité spectrale $f_X(\cdot)$*

21.
— Le processus $(X_t)_{t\in\mathbb{Z}}$ défini dans (1.4) est dit inversible s'il existe $\pi_0, \pi_1, \ldots$ tels que $\sum_{j=0}^{+\infty} |\pi_j| < \infty$ et pour tout $t \in \mathbb{Z}$,

$$\epsilon_t = \pi(L) X_t,$$

où $\pi(z) = \sum_{j=0}^{+\infty} \pi_j z^j$ et L est l'opérateur retard.
— le processus $(X_t)_{t\in\mathbb{Z}}$ dans (1.4) est dit causal (ou plus précisément une fonction causale de $(\epsilon_t)_{t\in\mathbb{Z}}$) s'il existe $\psi_0, \psi_1, \ldots$ tels que $\sum_{j=0}^{+\infty} |\psi_j| < \infty$ et pour tout $t \in \mathbb{Z}$,

$$X_t = \psi(L) \epsilon_t,$$

où $\psi(z) = \sum_{j=0}^{+\infty} \psi_j z^j$.

satisfont, pour $d_0 \neq 0$,

$$\rho_X(h) \sim Ch^{2d_0-1} \text{ quand } h \to +\infty,$$

où $C \neq 0$, *et*

$$f_X(\omega) = \frac{\sigma_\epsilon^2}{2\pi}\frac{|b(e^{-i\omega})|^2}{|a(e^{-i\omega})|^2}|1-e^{-i\omega}|^{-2d_0} \sim \frac{\sigma_\epsilon^2}{2\pi}\left\{\frac{b(1)}{a(1)}\right\}^2 \omega^{-2d_0} \text{ quand } \omega \to 0.$$

Les détails de la preuve de ce théorème sont présentés dans Brockwell and Davis [1991].

Le résultat énoncé dans le quatrième point du théorème 1.2 confirme, lorsque le paramètre de mémoire d_0 appartient à $]0, 1/2[$, le caractère de mémoire longue des processus FARIMA(p, d_0, q). Cette dépendance linéaire est justifiée par la décroissance très lente de la fonction d'autocovariance du processus, les autocorrélations restent très significatives de telle sorte que leur série est absolument divergente. Pour détecter ce phénomène, nous étudions généralement le comportement de la fonction d'autocovariance au voisinage de l'infini ou, d'une manière équivalente (suivant le sens que nous avons préciser dans le théorème 1.1), celui de la densité spectrale à l'origine. Le choix des opérateurs modélisant les perturbations de mémoire courte n'influence en rien le caractère de mémoire longue du processus. La partie utilisée pour ajuster le caractère de mémoire courte du processus se limite aux autocorrélations à faible retard. Nous pouvons donc se restreindre au cas des modèles FARIMA$(0, d_0, 0)$ pour illustrer la présence de la forte dépendance linéaire des processus FARIMA(p, d_0, q) généraux.

D'après (1.4), un processus $(X_t)_{t\in\mathbb{Z}}$ centré et stationnaire au second-ordre, à valeurs réelles, admet une représentation FARIMA$(0, d_0, 0)$ faible si, pour tout $t \in \mathbb{Z}$,

$$(1-L)^{d_0} X_t = \epsilon_t, \tag{1.5}$$

où $d_0 \in]-1/2, 1/2[$ et $(\epsilon_t)_{t\in\mathbb{Z}}$ est un bruit blanc faible de variance σ_ϵ^2.

L'opérateur de différentiation fractionnaire $(1-L)^{d_0}$ est donné, en utilisant la formule du binôme généralisée, par

$$(1-L)^{d_0} = \sum_{j=0}^{+\infty} \alpha_j(d_0) L^j, \tag{1.6}$$

où, pour tout $j \in \mathbb{N}$,

$$\alpha_j(d_0) = \frac{\Gamma(j-d_0)}{\Gamma(j+1)\Gamma(-d_0)}.$$

La fonction d'autocovariance $\gamma_X(\cdot)$ de $(X_t)_{t\in\mathbb{Z}}$ est définie pour tout $k \in \mathbb{Z}$ par :

$$\gamma_X(k) = \int_{-\pi}^{\pi} f_X(\omega) e^{ik\omega} d\omega, \tag{1.7}$$

où

$$f_X(\omega) = \frac{1}{2\pi}\sum_{k=-\infty}^{+\infty} \gamma_X(k) e^{-ik\omega}.$$

En utilisant les deux équations (1.5) et (1.6), nous obtenons que

$$\begin{aligned}
f_X(\omega) &= \frac{1}{2\pi}\sum_{k=-\infty}^{+\infty}\mathrm{Cov}(X_t, X_{t-k})\mathrm{e}^{-ik\omega} \\
&= \frac{1}{2\pi}\sum_{k=-\infty}^{+\infty}\mathrm{Cov}\left\{(1-L)^{-d_0}\epsilon_t, (1-L)^{-d_0}\epsilon_{t-k}\right\}\mathrm{e}^{-ik\omega} \\
&= \frac{1}{2\pi}\sum_{k=-\infty}^{+\infty}\sum_{j_1=0}^{+\infty}\sum_{j_2=0}^{+\infty}\alpha_{j_1}(-d_0)\alpha_{j_2}(-d_0)\mathrm{Cov}(\epsilon_{t-j_1}, \epsilon_{t-k-j_2})\mathrm{e}^{-ik\omega} \\
&= \frac{\sigma_\epsilon^2}{2\pi}\sum_{j_1=0}^{+\infty}\sum_{j_2=0}^{+\infty}\alpha_{j_1}(-d_0)\alpha_{j_2}(-d_0)\mathrm{e}^{i(j_2-j_1)\omega} \\
&= \frac{\sigma_\epsilon^2}{2\pi}\left|\sum_{j_1=0}^{+\infty}\alpha_{j_1}(-d_0)\mathrm{e}^{-ij_1\omega}\right|^2 = \frac{\sigma_\epsilon^2}{2\pi}\left|1-\mathrm{e}^{-i\omega}\right|^{-2d_0} \\
&= \frac{2^{-2d_0-1}\sigma_\epsilon^2}{\pi}\left|\sin\left(\frac{\omega}{2}\right)\right|^{-2d_0}.
\end{aligned}$$

Ainsi, en prenant en considération la parité de la fonction $\omega \to |\sin(\omega/2)|^{-2d_0}$, l'équation (1.7) devient

$$\gamma_X(k) = \frac{2^{-2d_0-1}\sigma_\epsilon^2}{\pi}\int_{-\pi}^{\pi}\left|\sin\left(\frac{\omega}{2}\right)\right|^{-2d_0}\cos(k\omega)\mathrm{d}\omega.$$

Remarquons que la fonction $\omega \to |\sin(\omega/2)|^{-2d_0}\cos(k\omega)$ est 2π-périodique. La fonction d'autocovariance $\gamma_X(\cdot)$ du processus $(X_t)_{t\in\mathbb{Z}}$ peut donc se réécrire pour tout entier relatif k sous la forme :

$$\begin{aligned}
\gamma_X(k) &= \frac{2^{-2d_0-1}\sigma_\epsilon^2}{\pi}\left\{\int_{-\pi}^{0}\left|\sin\left(\frac{\omega}{2}\right)\right|^{-2d_0}\cos(k\omega)\mathrm{d}\omega + \int_{0}^{2\pi}\left|\sin\left(\frac{\omega}{2}\right)\right|^{-2d_0}\cos(k\omega)\mathrm{d}\omega\right. \\
&\qquad \left. - \int_{\pi}^{2\pi}\left|\sin\left(\frac{\omega}{2}\right)\right|^{-2d_0}\cos(k\omega)\mathrm{d}\omega\right\} \\
&= \frac{2^{-2d_0-1}\sigma_\epsilon^2}{\pi}\int_{0}^{2\pi}\left|\sin\left(\frac{\omega}{2}\right)\right|^{-2d_0}\cos(k\omega)\mathrm{d}\omega.
\end{aligned}$$

Gradshteyn and Ryzhik [2007] ont montré que, pour tout $\nu > 0$ et pour tout $a \in \mathbb{R}$,

$$\int_0^{\pi}\{\sin(\omega)\}^{\nu-1}\cos(a\omega)\mathrm{d}\omega = \frac{\pi\cos(a\pi/2)\Gamma(\nu+1)2^{1-\nu}}{\nu\Gamma\{(\nu+a+1)/2\}\Gamma\{(\nu-a+1)/2\}}.$$

Nous en déduisons alors que

$$\gamma_X(k) = \frac{\sigma_\epsilon^2\cos(k\pi)\Gamma(2-2d_0)}{(1-2d_0)\Gamma(1-d_0+k)\Gamma(1-d_0-k)}.$$

Nous utilisons la formule des compléments pour obtenir que

$$\Gamma(1-d_0-k) = \Gamma(1-d_0)\Gamma(d_0)/(-1)^k\Gamma(k+d_0).$$

De plus, comme $\Gamma(2-2d_0) = (1-2d_0)\Gamma(1-2d_0)$, la fonction d'autocovariance $\gamma_X(\cdot)$ peut être réécrite explicitement sous la forme :

$$\gamma_X(k) = \sigma_\epsilon^2 \frac{\Gamma(1-2d_0)}{\Gamma(1-d_0)\Gamma(d_0)} \frac{\Gamma(k+d_0)}{\Gamma(1+k-d_0)}.$$

Cela implique que pour tout $t \in \mathbb{Z}$,

$$\mathrm{Var}(X_t) = \sigma_\epsilon^2 \frac{\Gamma(1-2d_0)}{\{\Gamma(1-d_0)\}^2}.$$

Finalement, la fonction d'autocorrélation $\rho_X(\cdot)$ du processus $(X_t)_{t\in\mathbb{Z}}$ est donnée pour tout retard $k \in \mathbb{Z}$ par :

$$\rho_X(k) = \frac{\Gamma(1-d_0)}{\Gamma(d_0)} \frac{\Gamma(k+d_0)}{\Gamma(1+k-d_0)}.$$

En utilisant la formule asymptotique de Stirling pour la fonction gamma, nous montrons que

$$\gamma_X(k) \sim \sigma_\epsilon^2 \frac{\Gamma(1-2d_0)\mathrm{e}^{1-2d_0}}{\Gamma(1-d_0)\Gamma(d_0)} k^{2d_0-1} \tag{1.8}$$

quand $k \to +\infty$.
La fonction constante $k \to \sigma_\epsilon^2\Gamma(1-2d_0)\mathrm{e}^{1-2d_0}/\Gamma(1-d_0)\Gamma(d_0)$ est évidemment à variation lente à l'infini au sens de Zygmund. Le processus $(X_t)_{t\in\mathbb{Z}}$ défini dans (1.5) est, d'après le théorème 1.1, à mémoire longue lorsque le paramètre de mémoire $d_0 \in]0, 1/2[$.

Tous ces résultats peuvent être généralisés au cas des modèles FARIMA(p, d_0, q). Nous pouvons montrer que l'introduction des opérateurs modélisant la dynamique de courte mémoire ne change pas le comportement asymptotique de la fonction d'autocovariance du processus. Le type de mémoire du processus est totalement déterminé par la valeur de d_0. Palma [2007] avait donné l'expression de la fonction d'autocovariance des modèles FARIMA(p, d_0, q).

Nous présentons maintenant nos principaux résultats sur l'estimation et la validation des modèles FARIMA faibles et sur l'estimation des processus autorégressifs induits par un bruit gaussien fractionnaire de façon plus détaillée :

1.5 Résultats du chapitre 2

Nous développons dans ce chapitre, dans un premier temps, les propriétés asymptotiques de l'estimateur des moindres carrés des paramètres des modèles FARIMA(p, d_0, q) faibles. Nous

estimons simultanément le paramètre de mémoire et les paramètres autorégressifs et moyennes-mobiles. Nous considérons ensuite le problème de l'estimation de la matrice de variance asymptotique de l'estimateur des moindres carrés des modèles FARIMA(p, d_0, q) induits par un bruit blanc faible. Nous proposons un estimateur convergent de cette matrice. Enfin, nous construisons les intervalles de confiance des paramètres des modèles FARIMA(p, d_0, q) faibles en utilisant une approche d'auto-normalisation. Cette approche nous permet de contourner le problème de l'estimation de la matrice de variance asymptotique de l'estimateur des moindres carrés.

1.5.1 Définition de l'estimateur des moindres carrés

Soit $(X_t)_{t\in\mathbb{Z}}$ le processus défini dans (1.4). Supposons que le paramètre de mémoire $d_0 \in]0, 1/2[$[22]. Afin d'assurer l'existence et l'unicité d'une solution stationnaire, inversible et causale de (1.4), nous supposons que les polynômes $a(\cdot)$ et $b(\cdot)$ ont leurs racines en dehors du disque unité (voir la dernière section 1.4 pour plus de détails). Il est d'usage courant pour le traitement statistique des modèles FARIMA de considérer que $\theta_0 = (a_1, a_2, \ldots, a_p, b_1, b_2, \ldots, b_q, d_0)'$ est l'unique paramètre vérifiant l'équation (1.4). Nous garantissons cette condition d'identifiabilité du paramètre par le fait de considérer que le polynôme autorégressif $a(\cdot)$ et le polynôme correspondant à la partie moyennes-mobiles $b(\cdot)$ ne partagent aucun facteur en commun.
Ces hypothèses standards sont souvent utilisées pour dresser l'inférence des modèles FARIMA (voir par exemple Dahlhaus [1989], Fox and Taqqu [1986] et Palma [2007]).

Soit Θ^* l'espace des paramètres

$$\begin{aligned}\Theta^* := \{&\theta = (\theta_1, \theta_2, ..., \theta_{p+q})' \in \mathbb{R}^{p+q};\ a_\theta(z) = 1 - \theta_1 z - \cdots - \theta_p z^p \\ &\text{et } b_\theta(z) = 1 - \theta_{p+1} z - \cdots - \theta_{p+q} z^q \text{ ont leurs zéros en dehors du disque unité} \\ &\text{et n'ont aucun facteur en commun}\}.\end{aligned}$$

Notons par Θ le produit cartésien $\Theta^* \times [d_1, d_2]$, où $[d_1, d_2] \subset]0, 1/2[$ avec $d_1 \leq d_0 \leq d_2$.
Le vrai paramètre (inconnu) $\theta_0 = (a_1, a_2, \ldots, a_p, b_1, b_2, \ldots, b_q, d_0)'$ est supposé appartenir à l'espace des paramètres Θ.
Pour tout $\theta \in \Theta$, soit $(\epsilon_t(\theta))_{t\in\mathbb{Z}}$ le processus stationnaire au second ordre solution de

$$\epsilon_t(\theta) = \sum_{j\geq 0} \alpha_j(d) X_{t-j} - \sum_{i=1}^{p} \theta_i \sum_{j\geq 0} \alpha_j(d) X_{t-i-j} + \sum_{j=1}^{q} \theta_{p+j} \epsilon_{t-j}(\theta). \tag{1.9}$$

En remplaçant θ par θ_0 dans (1.9), nous obtenons, presque sûrement, le bruit blanc faible introduit dans (1.4).
Étant donné une observation de longueur n, $X_1, X_2, \ldots, X_n$, les variables aléatoires $\epsilon_t(\theta)$ peuvent être approximées, pour $0 < t \leq n$, par $\tilde{\epsilon}_t(\theta)$ définies récursivement comme solutions

22. Nous considérons le cas où $(X_t)_{t\in\mathbb{Z}}$ est stationnaire et à mémoire longue (voir la section 1.4).

de

$$\tilde{\epsilon}_t(\theta) = \sum_{j=0}^{t-1} \alpha_j(d) X_{t-j} - \sum_{i=1}^{p} \theta_i \sum_{j=0}^{t-i-1} \alpha_j(d) X_{t-i-j} + \sum_{j=1}^{q} \theta_{p+j} \tilde{\epsilon}_{t-j}(\theta), \tag{1.10}$$

avec $\tilde{\epsilon}_t(\theta) = X_t = 0$ si $t \leq 0$.

Nous avons montré que ces valeurs initiales sont asymptotiquement négligeables uniformément en θ. L'écart $\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)$ tend vers 0 presque sûrement quand $t \to +\infty$ et ceci pour tout $\theta \in \Theta$ (voir le lemme 2.4 ci-après). Le choix des valeurs initiales est sans effet sur les propriétés asymptotiques de l'estimateur du paramètre θ_0.

Notons maintenant par Θ_δ^* l'espace compact

$$\Theta_\delta^* = \left\{ \theta \in \mathbb{R}^{p+q};\ \text{les racines des polynômes } a_\theta(z) \text{ et } b_\theta(z) \text{ ont un module } \geq 1+\delta \right\}.$$

Nous définissons comme précédemment l'ensemble Θ_δ comme le produit cartésien de Θ_δ^* par $[d_1, d_2]$, *i.e.* $\Theta_\delta = \Theta_\delta^* \times [d_1, d_2]$, où δ est une constante strictement positive choisie de telle sorte que θ_0 appartient à Θ_δ.

La variable aléatoire $\widehat{\theta}_n$ est dite estimateur des moindres carrés de θ_0 si elle satisfait, presque sûrement,

$$\hat{\theta}_n = \underset{\theta \in \Theta_\delta}{\operatorname{argmin}}\ Q_n(\theta), \ \text{où}\ Q_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \tilde{\epsilon}_t^2(\theta).$$

1.5.2 Propriétés asymptotiques de l'estimateur des moindres carrés

Le premier résultat principal est la convergence forte de l'estimateur des moindres carrés, cette propriété asymptotique est prouvée sous l'hypothèse de régularité sur le processus $(\epsilon_t)_{t\in\mathbb{Z}}$ suivante :

(H1) : Le processus $(\epsilon_t)_{t\in\mathbb{Z}}$ est strictement stationnaire et ergodique.

Le résultat est donné par le théorème suivant :

Théorème 1.3. *(Convergence forte) Supposons que $(\epsilon_t)_{t\in\mathbb{Z}}$ vérifie (1.4). Soit $(\hat{\theta}_n)_n$ une suite d'estimateurs des moindres carrés. Sous* **(H1)**, *nous avons*

$$\hat{\theta}_n \xrightarrow[n\to+\infty]{a.s.} \theta_0.$$

Ainsi, comme dans le cas où le processus des innovations est un bruit blanc fort, nous obtenons la consistance forte de l'estimateur $\hat{\theta}_n$ des paramètres des processus FARIMA(p, d_0, q) à mémoire longue dans le cas faible. Cette généralisation est établie sans hypothèses supplémentaires.

Pour les modèles FARIMA faibles, l'ergodicité ne suffit pas pour établir un théorème central limite. Afin d'obtenir la normalité asymptotique de l'estimateur des moindres carrés, il est

nécessaire de contrôler cette dépendance non linéaire générale inconnue que présente le bruit. Pour cela, l'hypothèse de la sommabilité des cumulants joints d'ordre 4 du processus $(\epsilon_t)_{t\in\mathbb{Z}}$ est souvent utilisée dans ce contexte (voir par exemple Shao [2010b] et Shao [2010c] dans le cadre des modèles FARIMA ou même Francq and Zakoïan [2007] et Zhu and Li [2015] dans le contexte ARMA à mémoire courte).

Les coefficients de mélange fort $\{\alpha_\epsilon(h)\}_{h\in\mathbb{N}}$ du processus $(\epsilon_t)_{t\in\mathbb{Z}}$ présentent un moyen couramment utilisé pour contrôler la dépendance probabiliste non linéaire des processus. Ils sont définis pour tout entier h par :

$$\alpha_\epsilon(h) = \sup_{A\in\mathcal{F}_{-\infty}^{t}, B\in\mathcal{F}_{t+h}^{\infty}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|,$$

où $\mathcal{F}_{-\infty}^{t} = \sigma(\epsilon_u, u \le t)$ et $\mathcal{F}_{t+h}^{\infty} = \sigma(\epsilon_u, u \ge t+h)$.

Doukhan and León [1989] ont montré que l'hypothèse

(H2) : Il existe un entier τ tel que pour $\nu \in]0,1]$, nous avons $\mathbb{E}|\epsilon_t|^{\tau+\nu} < \infty$ et $\sum_{h=0}^{\infty}(h+1)^{k-2}\{\alpha_\epsilon(h)\}^{\nu/(k+\nu)} < \infty$ pour $k = 1,\dots,\tau$,

implique la sommabilité des cumulants joints d'ordre τ du processus $(\epsilon_t)_{t\in\mathbb{Z}}$. Plus précisément, ils ont prouvé que **(H2)** implique l'hypothèse faible suivante :

(H2') : Il existe un entier $\tau \ge 2$ tel que $C_\tau := \sum_{i_1,\dots,i_{\tau-1}\in\mathbb{Z}} |\text{cum}(\epsilon_0, \epsilon_{i_1}, \dots, \epsilon_{i_{\tau-1}})| < \infty$.

Les coefficients de mélange fort sont généralement utilisés pour étudier l'inférence des modèles statistiques présentant une dépendance non linéaire dont la structure est non identifiée (voir par exemple Francq and Zakoïan [1998] et Boubacar Mainassara and Francq [2011]).

Il est néanmoins important de noter que la vitesse de convergence des coefficients de mélange imposée dans **(H2)** est plus importante que celle exigée pour prouver les propriétés asymptotiques des estimateurs des processus ARMA ou VARMA à mémoire courte. Ceci est justifié par l'effet apporté par l'opérateur fractionnaire. Les coefficients des représentations AR et MA infinies de (1.4) ont une décroissance en puissance, contrairement à la décroissance exponentielle de leurs analogues dans le cadre des modèles ARMA par exemple (voir la sous-section 2.6.1 pour plus de détails et commentaires).

Avant d'énoncer le résultat portant sur la normalité asymptotique de l'estimateur des moindres carrés des paramètres du processus défini dans (1.4), nous introduisons les notations suivantes :

$$O_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2(\theta), \tag{1.11}$$

où $(\epsilon_t(\theta))_{t\in\mathbb{Z}}$ est le processus donné dans (1.9),

$$I(\theta) = \lim_{n\to\infty} Var\left\{\sqrt{n}\frac{\partial}{\partial\theta}O_n(\theta)\right\} \text{ et } J(\theta) = \lim_{n\to\infty}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}O_n(\theta)\right] \text{ p.s.}$$

L'existence de ces deux dernières matrices est établie par la suite.

Note deuxième principal résultat est donné par le théorème suivant :

Théorème 1.4. *(Normalité asymptotique). Supposons que $(\epsilon_t)_{t\in\mathbb{Z}}$ vérifie (1.4) et que θ_0 est à l'intérieur de Θ. Soit $(\hat{\theta}_n)_n$ une suite d'estimateurs des moindres carrés. Sous* **(H1)** *et* **(H2)** *avec $\tau = 4$, la suite $(\sqrt{n}(\hat{\theta}_n - \theta_0))_n$ a une distribution asymptotique normale centrée et de matrice de variance $\Omega := J^{-1}(\theta_0)I(\theta_0)J^{-1}(\theta_0)$.*

Nous montrons dans la suite que la matrice $J(\theta_0)$ peut se réécrire sous la forme :

$$J(\theta_0) = 2\mathbb{E}\left[\frac{\partial}{\partial\theta}\epsilon_t(\theta_0)\frac{\partial}{\partial\theta'}\epsilon_t(\theta_0)\right].$$

Ceci permet de confirmer que $J(\theta_0)$ a la même expression dans les deux cas des modèles FARIMA forts et faibles (voir le théorème 1 de Beran [1995]).

La matrice $I(\theta_0)$ a, quant à elle, une expression en général plus compliquée dans le cas des modèles FARIMA faibles que dans le contexte fort.

Remarque 1.1. *Dans le cas standard des modèles FARIMA forts, i.e. quand $(\epsilon_t)_{t\in\mathbb{Z}}$ est considéré comme une suite iid de variables aléatoires, nous avons $I(\theta_0) = 2\sigma_\epsilon^2 J(\theta_0)$. Ainsi la matrice de covariance asymptotique est réduite à $\Omega_S := 2\sigma_\epsilon^2 J^{-1}(\theta_0)$. Généralement, quand le bruit est non indépendant, cette simplification ne peut pas avoir lieu et nous avons $I(\theta_0) \neq 2\sigma_\epsilon^2 J(\theta_0)$. La vraie variance asymptotique $\Omega = J^{-1}(\theta_0)I(\theta_0)J^{-1}(\theta_0)$ obtenue dans le cadre des modèles FARIMA faibles peut être très différente de Ω_S.*

Les logiciels actuellement disponibles pour l'analyse des séries temporelles utilisent l'estimation standard empirique de Ω_S. Ainsi, ces logiciels ne peuvent pas fournir une estimation correcte de la matrice Ω. Ce problème a été aussi remarqué dans le cas des modèles ARMA faible (voir Francq and Zakoïan [2007]). Il est donc intéressant de trouver un estimateur convergent et robuste[23] *de la matrice Ω.*

1.5.3 Estimation de la matrice de variance asymptotique

Afin d'obtenir des intervalles de confiance ou de tester la significativité des coefficients FARIMA faibles, il sera nécessaire de disposer d'un estimateur convergent de la matrice de

23. La robustesse ici signifie que l'estimateur reste convergent pour les autres classes des modèles FARIMA.

variance asymptotique Ω. La matrice $J(\theta_0)$ peut facilement être estimée empiriquement par la matrice carrée d'ordre $p+q+1$ définie par :

$$\hat{J}_n = \frac{2}{n}\sum_{t=1}^{n}\left\{\frac{\partial}{\partial\theta}\tilde{\epsilon}_t\left(\hat{\theta}_n\right)\right\}\left\{\frac{\partial}{\partial\theta'}\tilde{\epsilon}_t\left(\hat{\theta}_n\right)\right\}. \tag{1.12}$$

La convergence forte de $\hat{J}_n$ vers $J(\theta_0)$ est classique (voir le lemme 2.8 dans la sous-section 2.6.3 pour plus de détails).

D'après la remarque 1.1, la matrice $\hat{\Omega}_S := 2\hat{\sigma}_\epsilon^2\hat{J}_n^{-1}$ avec $\hat{\sigma}_\epsilon^2 = Q_n(\hat{\theta}_n)$ est un estimateur fortement convergent de Ω_S dans le cas standard des modèles FARIMA forts.

Dans le cas général des modèles FARIMA faibles, cet estimateur ne converge pas lorsque $I(\theta_0) \neq 2\sigma_\epsilon^2 J(\theta_0)$. Nous avons donc besoin d'un estimateur de $I(\theta_0)$.

Estimation de la matrice asymptotique $I(\theta_0)$

Soit, pour tout $t \in \mathbb{Z}$,

$$H_t(\theta_0) = 2\epsilon_t(\theta_0)\frac{\partial}{\partial\theta}\epsilon_t(\theta_0) = \left(2\epsilon_t(\theta_0)\frac{\partial}{\partial\theta_1}\epsilon_t(\theta_0), \ldots, 2\epsilon_t(\theta_0)\frac{\partial}{\partial\theta_{p+q+1}}\epsilon_t(\theta_0)\right)'. \tag{1.13}$$

Nous montrons dans la preuve du lemme 2.9 que

$$I(\theta_0) = \lim_{n\to\infty}\mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}H_t(\theta_0)\right) = \sum_{h=-\infty}^{\infty}\mathrm{Cov}\left(H_t(\theta_0), H_{t-h}(\theta_0)\right).$$

En suivant les mêmes arguments développés dans Boubacar Mainassara et al. [2012], la matrice $I(\theta_0)$ peut être estimée en utilisant l'approche de Berk (voir Berk [1974]). Plus précisément, en interprétant $I(\theta_0)/2\pi$ comme la densité spectrale du processus stationnaire $(H_t(\theta_0))_{t\in\mathbb{Z}}$ évaluée à la fréquence 0, nous pouvons estimer la densité spectrale de $(H_t(\theta_0))_{t\in\mathbb{Z}}$ en utilisant une approche basée sur une autorégression paramétrique. Nous obtenons ainsi un estimateur de la matrice $I(\theta_0)$.

Pour tout $\theta \in \Theta$, $H_t(\theta)$ est une fonction mesurable de $\{\epsilon_s, s \leq t\}$. Le processus stationnaire $(H_t(\theta_0))_{t\in\mathbb{Z}}$ admet la décomposition de Wold (voir la section 1.2) suivante :

$$H_t(\theta_0) = u_t + \sum_{k=1}^{\infty}\psi_k u_{t-k},$$

où $(u_t)_{t\in\mathbb{Z}}$ est un bruit blanc faible $(p+q+1)$-dimensionnel de matrice de variance Σ_u.

Supposons que Σ_u est non singulière, que $\sum_{k=1}^{\infty}\|\psi_k\| < \infty$ et que $\det(I_{p+q+1}+\sum_{k=1}^{\infty}\psi_k z^k) \neq 0$ quand $|z| \leq 1$. Alors $(H_t(\theta_0))_{t\in\mathbb{Z}}$ admet une représentation AR(∞) faible de la forme :

$$\Phi(L)H_t(\theta_0) := H_t(\theta_0) - \sum_{k=1}^{\infty}\Phi_k H_{t-k}(\theta_0) = u_t, \tag{1.14}$$

tels que $\sum_{k=1}^{\infty} \|\Phi_k\| < \infty$ et $\det\{\Phi(z)\} \neq 0$ pour tout $|z| \leq 1$. Il est prouvé dans Boubacar Maïnassara [2009] et Lütkepohl [2005] qu'il existe $K > 0$ et $0 < \rho < 1$ tel que

$$\|\Phi_k\| \leq K\,\rho^k.$$

Grâce aux remarques précédentes, l'estimation de $I(\theta_0)$ est donc basée sur l'expression

$$I(\theta_0) = \Phi^{-1}(1)\Sigma_u\Phi^{-1}(1).$$

Considérons la régression de $H_t(\theta_0)$ sur $H_{t-1}(\theta_0), \ldots, H_{t-r}(\theta_0)$ définie par :

$$H_t(\theta_0) = \sum_{k=1}^{r} \Phi_{r,k} H_{t-k}(\theta_0) + u_{r,t}, \tag{1.15}$$

où $u_{r,t}$ est non corrélé avec $H_{t-1}(\theta_0), \ldots, H_{t-r}(\theta_0)$. Comme $H_t(\theta_0)$ est non observable, nous introduisons $\hat{H}_t \in \mathbb{R}^{p+q+1}$ obtenue en remplaçant dans (1.13) $\epsilon_t(\theta_0)$ par $\tilde{\epsilon}_t(\theta_0)$ et θ_0 par $\hat{\theta}_n$:

$$\hat{H}_t = 2\tilde{\epsilon}_t(\hat{\theta}_n)\frac{\partial}{\partial\theta}\tilde{\epsilon}_t(\hat{\theta}_n)\ .$$

Notons par $\hat{\Phi}_r(z) = I_{p+q+1} - \sum_{k=1}^{r} \hat{\Phi}_{r,k} z^k$, où $\hat{\Phi}_{r,1}, \ldots, \hat{\Phi}_{r,r}$ les coefficients de la régression linéaire de $\hat{H}_t$ sur $\hat{H}_{t-1}, \ldots, \hat{H}_{t-r}$. Soient $\hat{u}_{r,t}$ les résidus de cette régression et $\hat{\Sigma}_{\hat{u}_r}$ la variance empirique (définie dans (1.16) ci-dessous) de $\hat{u}_{r,1}, \ldots, \hat{u}_{r,r}$. Les estimateurs des moindres carrés de $\underline{\Phi}_r = (\Phi_{r,1}, \ldots, \Phi_{r,r})$ et de $\Sigma_{u_r} = \mathrm{Var}(u_{r,t})$ sont donnés par :

$$\underline{\hat{\Phi}}_r = \hat{\Sigma}_{\hat{H},\underline{\hat{H}}_r}\hat{\Sigma}_{\underline{\hat{H}}_r}^{-1} \quad \text{et} \quad \hat{\Sigma}_{\hat{u}_r} = \frac{1}{n}\sum_{t=1}^{n}\left(\hat{H}_t - \underline{\hat{\Phi}}_r\underline{\hat{H}}_{r,t}\right)\left(\hat{H}_t - \underline{\hat{\Phi}}_r\underline{\hat{H}}_{r,t}\right)', \tag{1.16}$$

où

$$\underline{\hat{H}}_{r,t} = (\hat{H}'_{t-1}, \ldots, \hat{H}'_{t-r})', \quad \hat{\Sigma}_{\hat{H},\underline{\hat{H}}_r} = \frac{1}{n}\sum_{t=1}^{n}\hat{H}_t\underline{\hat{H}}'_{r,t} \text{ et } \hat{\Sigma}_{\underline{\hat{H}}_r} = \frac{1}{n}\sum_{t=1}^{n}\underline{\hat{H}}_{r,t}\underline{\hat{H}}'_{r,t},$$

avec par convention $\hat{H}_t = 0$ quand $t \leq 0$. Nous supposons que $\hat{\Sigma}_{\underline{\hat{H}}_r}$ est inversible (ce qui est vrai asymptotiquement).

Dans le cas des processus linéaires avec des innovations indépendantes, Berk (voir Berk [1974]) avait montré que la densité spectrale peut être estimée en ajustant des modèles autorégressifs d'ordre $r = r(n)$, avec r qui tend vers l'infini et r^3/n qui tend vers 0 quand n tend vers l'infini. La difficulté ici est que le processus $(H_t(\theta_0))_{t\in\mathbb{Z}}$ est multivarié et non observé. Boubacar Mainassara et al. [2012] et Boubacar Mainassara and Francq [2011] ont montré que le résultat de Berk [1974] reste valide pour les processus multivariés avec des innovations non indépendantes. Nous étendons les résultats de Boubacar Mainassara et al. [2012] aux modèles FARIMA faibles à mémoire longue.

Le résultat de l'étude asymptotique de l'estimateur de $I(\theta_0)$ en utilisant l'approche de la densité spectrale est donné dans le théorème suivant :

Théorème 1.5. *Supposons que* $(\epsilon_t)_{t\in\mathbb{Z}}$ *vérifie* (1.4) *et que* θ_0 *est à l'intérieur de* Θ*. Supposons de plus que le processus* $(H_t(\theta_0))_{t\in\mathbb{Z}}$ *défini dans* (1.13) *admet une représentation* $AR(\infty)$ *faible comme dans* (1.14)*. Sous* **(H1)** *et* **(H2')** *avec* $\tau = 8$*, l'estimateur spectral*

$$\hat{I}_n^{\mathrm{SP}} := \hat{\Phi}_r^{-1}(1)\hat{\Sigma}_{\hat{u}_r}\hat{\Phi}_r'^{-1}(1) \xrightarrow[n\to+\infty]{\mathbb{P}} I(\theta_0) = \Phi^{-1}(1)\Sigma_u\Phi^{-1}(1),$$

où r *dépend de* n *et satisfait* $\lim_{n\to+\infty} r^5(n)/n^{1-2(d_2-d_1)} = 0$ *(rappelons que* $d_0 \in [d_1,d_2] \subset]0,1/2[$*).*

La preuve de ce théorème est donnée dans la sous-section 2.6.4.

1.5.4 Construction des intervalles de confiance des paramètres des modèles FARIMA faibles par auto-normalisation

Nous avons vu précédemment que l'obtention des intervalles de confiance des paramètres des modèles FARIMA faibles est contrainte à la contruction d'un estimateur convergent de la matrice de variance $I(\theta_0)$ (voir les théorèmes 1.4 et 1.5). L'approche paramétrique basée sur une estimation autorégressive de la densité spectrale de $(H_t(\theta_0))_{t\in\mathbb{Z}}$ présente l'inconvénient de choisir le paramètre de troncature r dans (1.15). Ce choix de l'ordre de troncature est souvent crucial et difficile. L'objectif de cette partie du travail est de contourner cette difficulté.

Cette sous-section est également intéressante car, à notre connaissance, elle n'a pas été étudiée pour les modèles FARIMA faibles. Une exception notable est Shao [2012] qui a étudié ce problème dans un cas de mémoire courte (voir l'hypothèse 1 dans Shao [2012] qui implique que le processus X présente une dépendance de court terme).

Nous proposons une méthode alternative pour obtenir des intervalles de confiance pour les modèles FARIMA faibles en évitant l'estimation de la matrice de covariance asymptotique $I(\theta_0)$. Cette méthode est basée sur une approche d'auto-normalisation utilisée pour construire une statistique qui dépend du vrai paramètre θ_0 et dont la distribution asymptotique dépend seulement des ordres p et q de la partie autorégressive et moyennes-mobiles du modèle (voir le théorème 1 de Shao [2012] pour une référence dans le cas des ARMA faibles).

L'idée vient de Lobato [2001] et a déjà été étendue par Boubacar Maïnassara and Saussereau [2018], Kuan and Lee [2006], Shao [2010a,c, 2012] à des cadres plus généraux. Voir aussi Shao [2015] pour un aperçu de certains développements récents concernant l'inférence de données de séries chronologiques en utilisant l'approche auto-normalisée.

Expliquons brièvement l'idée de l'auto-normalisation.

Le développement de Taylor de la fonction $\partial Q_n(\cdot)/\partial\theta$ autour de θ_0, nous permet d'obtenir,

sous l'hypothèse de l'appartenance de θ_0 à l'intérieur de Θ, que

$$0 = \sqrt{n}\frac{\partial}{\partial\theta}Q_n(\hat{\theta}_n) = \sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0) + \left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}Q_n\left(\theta^*_{n,i,j}\right)\right]\sqrt{n}\left(\hat{\theta}_n - \theta_0\right), \tag{1.17}$$

où les $\theta^*_{n,i,j}$ sont entre $\hat{\theta}_n$ et θ_0.

En utilisant le fait que

$$\sqrt{n}\left(\frac{\partial}{\partial\theta}O_n(\theta_0) - \frac{\partial}{\partial\theta}Q_n(\theta_0)\right) = \sqrt{n}\frac{\partial}{\partial\theta}O_n(\theta_0) + \left\{\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}Q_n(\theta^*_{n,i,j})\right] - J(\theta_0) + J(\theta_0)\right\}\sqrt{n}(\hat{\theta}_n - \theta_0),$$

l'équation (1.17) implique que

$$\sqrt{n}\frac{\partial}{\partial\theta}O_n(\theta_0) + J(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0) = o_{\mathbb{P}}(1). \tag{1.18}$$

Ceci est dû aux propriétés techniques suivantes :

- la convergence en probabilité de $\sqrt{n}\partial Q_n(\theta_0)/\partial\theta$ vers $\sqrt{n}\partial O_n(\theta_0)/\partial\theta$ (voir le lemme 2.5 ci-après),
- la convergence presque-sûre de $[\partial^2 Q_n(\theta^*_{n,i,j})/\partial\theta_i\partial\theta_j]$ vers $J(\theta_0)$ (voir le lemme 2.8 ci-après),
- la suite $(\sqrt{n}(\hat{\theta}_n - \theta_0))_n$ est une suite tendue (voir le théorème 1.4) et
- l'existence et l'inversibilité de la matrice $J(\theta_0)$ (voir le lemme 2.6 ci-après).

Ainsi, de (1.18), nous obtenons que

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t + o_{\mathbb{P}}(1),$$

où

$$U_t = -J^{-1}(\theta_0)H_t(\theta_0),$$

avec $(H_t(\theta_0))_{t\in\mathbb{Z}}$ le processus défini dans (1.13).

À ce stade, la méthode classique qui consiste à estimer la matrice de covariance asymptotique $I(\theta_0)$ ne sera pas considérée. Nous essayons plutôt d'appliquer le lemme 1 dans Lobato [2001]. Il faut donc vérifier qu'un théorème central limite fonctionnel est valable pour le processus $U := (U_t)_{t\geq 1}$. Pour cela, nous définissons la matrice de normalisation $P_{p+q+1,n}$ de $\mathbb{R}^{(p+q+1)\times(p+q+1)}$ par :

$$P_{p+q+1,n} = \frac{1}{n^2}\sum_{t=1}^{n}\left(\sum_{j=1}^{t}(U_j - \bar{U}_n)\right)\left(\sum_{j=1}^{t}(U_j - \bar{U}_n)\right)', \tag{1.19}$$

où $\bar{U}_n = (1/n)\sum_{i=1}^{n} U_i$.

L'inversibilité de cette matrice est assurée lorsque le processus $(\epsilon_t)_{t\in\mathbb{Z}}$ admet une densité strictement positive sur un voisinage de zéro.

Proposition 1.1. *Supposons que $(\epsilon_t)_{t\in\mathbb{Z}}$ a une densité strictement positive sur un voisinage de zéro. Sous les hypothèses du théorème 1.4, la matrice $P_{p+q+1,n}$ est presque-sûrement non singulière.*

La preuve de cette proposition est donnée dans la sous-section 2.6.5.

Soit $(B_m(r))_{r\geq 0}$ un mouvement brownien standard m-dimensionnel. Pour $m \geq 1$, notons par $\mathcal{U}_m$ la variable aléatoire définie par :

$$\mathcal{U}_m = B_m'(1) V_m^{-1} B_m(1), \tag{1.20}$$

où

$$V_m = \int_0^1 (B_m(r) - rB_m(1))\,(B_m(r) - rB_m(1))'\, dr. \tag{1.21}$$

Les valeurs critiques de $\mathcal{U}_m$ ont été tabulées par Lobato [2001].

Le théorème suivant énonce la distribution asymptotique auto-normalisée du vecteur aléatoire $\sqrt{n}(\hat{\theta}_n - \theta_0)$.

Théorème 1.6. *Supposons que $(\epsilon_t)_{t\in\mathbb{Z}}$ a une densité strictement positive sur un voisinage de zéro. Sous les hypothèses du théorème 1.4,*

$$n(\hat{\theta}_n - \theta_0)' P_{p+q+1,n}^{-1}(\hat{\theta}_n - \theta_0) \xrightarrow[n\to+\infty]{\text{en loi}} \mathcal{U}_{p+q+1}.$$

La preuve de ce théorème est donnée dans la sous-section 2.6.6.

Le théorème 1.6 est inutilisable dans la pratique car la matrice de normalisation $P_{p+q+1,n}$ est non observable. Ce problème sera corrigé lorsque nous remplaçons la matrice $P_{p+q+1,n}$ par sa partie empirique ou observable

$$\hat{P}_{p+q+1,n} = \frac{1}{n^2}\sum_{t=1}^{n}\left(\sum_{j=1}^{t}(\hat{U}_j - \tfrac{1}{n}\textstyle\sum_{k=1}^{n}\hat{U}_k)\right)\left(\sum_{j=1}^{t}(\hat{U}_j - \tfrac{1}{n}\textstyle\sum_{k=1}^{n}\hat{U}_k)\right)' \quad \text{où } \hat{U}_j = -\hat{J}_n^{-1}\hat{H}_j. \tag{1.22}$$

La quantité ci-dessus est observable et nous sommes donc en mesure d'énoncer la version applicable du théorème 1.6.

Théorème 1.7. *Supposons que $(\epsilon_t)_{t\in\mathbb{Z}}$ a une densité strictement positive sur un voisinage de zéro. Sous les hypothèses du théorème 1.4,*

$$n(\hat{\theta}_n - \theta_0)' \hat{P}_{p+q+1,n}^{-1}(\hat{\theta}_n - \theta_0) \xrightarrow[n\to+\infty]{\text{en loi}} \mathcal{U}_{p+q+1}.$$

La preuve de ce théorème est donnée dans la sous-section 2.6.7.

Au seuil asymptotique α, une région de confiance à $100(1-\alpha)\%$ des éléments de θ_0 est donc donnée par l'ensemble des valeurs du vecteur θ satisfaisant l'inégalité suivante :

$$n(\hat{\theta}_n - \theta)' \hat{P}^{-1}_{p+q+1,n}(\hat{\theta}_n - \theta) \leq \mathcal{U}_{p+q+1,\alpha},$$

où $\mathcal{U}_{p+q+1,\alpha}$ est le quantile d'ordre $1-\alpha$ de la distribution de $\mathcal{U}_{p+q+1}$.

Corollaire 1.2. *Pour tout $1 \leq i \leq p+q+1$, une région de confiance à $100(1-\alpha)\%$ de $\theta_0(i)$ est donnée par l'ensemble suivant :*

$$\left\{ x \in \mathbb{R} \; ; \; n\big(\hat{\theta}_n(i) - x\big)^2 \hat{P}^{-1}_{p+q+1,n}(i,i) \leq \mathcal{U}_{1,\alpha} \right\},$$

où $\mathcal{U}_{1,\alpha}$ est le quantile d'ordre $1-\alpha$ de la distribution de $\mathcal{U}_1$.

La preuve de ce corollaire est similaire à celle du théorème 1.7 quand nous nous restreignons au cas unidimensionnel.

1.6 Résultats du chapitre 3

Après la phase de l'estimation, la prochaine étape importante dans la modélisation des modèles FARIMA faibles consiste à vérifier si le modèle estimé ajuste correctement les données. Cette étape d'adéquation permet de valider ou d'invalider le choix des ordres du modèle. Ce choix est important pour la précision des prévisions linéaires et pour une bonne interprétation du modèle.

Le nombre des paramètres dans un modèle FARIMA(p, d_0, q) augmente rapidement avec p et q. Il est donc nécessaire d'avoir une procédure permettant de confirmer ou non le choix de ces ordres. D'une part, la sélection d'ordres trop grands a pour conséquence l'introduction de certains termes qui ne sont pas forcément pertinents dans le modèle. Elle peut aussi entraîner 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.

D'autre part, la sélection d'ordres trop petits entraîne la perte d'une information qui peut être détectée par une corrélation des résidus ou encore une estimation non convergente des paramètres.

Dans ce chapitre, nous étudions dans un premier temps le comportement asymptotique des statistiques portmanteau dans le cadre de modèles FARIMA(p, d_0, q) induits par un bruit non corrélé mais qui peut contenir des dépendances non linéaires très générales. Ceci permet de développer des versions modifiées des tests portmanteau pour la validation des modèles FARIMA(p, d_0, q) faibles. Nous proposons dans un second temps une deuxième méthode alternative basée sur une approche d'auto-normalisation pour tester l'adéquation des modèles FARIMA(p, d_0, q) faibles. Cette méthode a l'avantage de contourner le problème de l'estimation de la matrice de variance asymptotique du vecteur joint de l'estimateur des moindres carrés

et des autocovariances empiriques du bruit.

Dans cette section, nous gardons les notations introduites dans la section précédente sur les principaux résultats asymptotiques de l'estimateur des moindres carrés des modèles FARIMA faibles (voir la section 1.5). Tous les résultats de la section 1.5 seront réutilisés dans cette section sans faire de rappel.

Pour $t \geq 1$, notons par $\hat{e}_t = \tilde{\epsilon}_t(\hat{\theta}_n)$ les résidus des moindres carrés. En utilisant (1.10), il est clair que $\hat{e}_t = 0$ pour $t \leq 0$ et $t > n$. De (1.4), nous déduisons que

$$\hat{e}_t = \sum_{j=0}^{t-1} \alpha_j(\hat{d})\hat{X}_{t-j} - \sum_{i=1}^{p} \hat{\theta}_i \sum_{j=0}^{t-i-1} \alpha_j(\hat{d})\hat{X}_{t-i-j} + \sum_{j=1}^{q} \hat{\theta}_{p+j}\hat{e}_{t-j},$$

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

Pour un entier $m \geq 1$ fixé, considérons le vecteur des autocovariances résiduelles

$$\hat{\gamma}_m = (\hat{\gamma}(1), \ldots, \hat{\gamma}(m))' \text{ où } \hat{\gamma}(h) = \frac{1}{n} \sum_{t=h+1}^{n} \hat{e}_t \, \hat{e}_{t-h} \text{ pour } 0 \leq h < n.$$

Dans la suite nous aurons aussi besoin du vecteur des m premières autocorrélations résiduelles

$$\hat{\rho}_m = (\hat{\rho}(1), \ldots, \hat{\rho}(m))' \text{ où } \hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0).$$

Depuis les travaux de Box and Pierce [1970] et de Ljung and Box [1978], les tests portmanteau sont devenus des outils populaires permettant de tester l'adéquation des modèles ARMA à mémoire courte. En se basant sur les autocorrélations empiriques résiduelles du bruit, leurs statistiques de test sont définies respectivement par

$$Q_m^{\text{BP}} = n \sum_{h=1}^{m} \hat{\rho}^2(h) \quad \text{et} \quad Q_m^{\text{LB}} = n(n+2) \sum_{h=1}^{m} \frac{\hat{\rho}^2(h)}{n-h}. \tag{1.23}$$

Ces statistiques sont souvent utilisées pour tester l'hypothèse nulle

$(\mathcal{H}\mathbf{0})$: $(X_t)_{t\in\mathbb{Z}}$ satisfait une représentation FARIMA(p, d_0, q) ;

contre l'alternative

$(\mathcal{H}\mathbf{1})$: $(X_t)_{t\in\mathbb{Z}}$ n'admet pas de représentation FARIMA ou admet une représentation FARIMA(p', d_0, q') avec $p' > p$ ou $q' > q$.

Ces tests sont des outils très utiles pour vérifier la signification globale des autocorrélations résiduelles.

1.6.1 La distribution asymptotique des autocorrélations résiduelles

Tout d'abord, l'hypothèse **(H2')** [24] sur la sommabilité des cumulants joints du processus $(\epsilon_t)_{t\in\mathbb{Z}}$ impliquera la normalité asymptotique des autocovariances "empiriques"

$$\gamma_m = (\gamma(1), \ldots, \gamma(m))' \text{ où } \gamma(h) = \frac{1}{n}\sum_{t=h+1}^{n} \epsilon_t\, \epsilon_{t-h} \text{ pour } 0 \leq h < n. \tag{1.24}$$

Il est à noter que les $\gamma(h)$ ne sont pas des statistiques calculables car elles dépendent des innovations non observées $\epsilon_t = \epsilon_t(\theta_0)$. Elles sont introduites pour faciliter les preuves techniques de nos principaux résultats.

Soit Ψ_m la matrice de $\mathbb{R}^{m\times(p+q+1)}$ définie par :

$$\Psi_m = \mathbb{E}\left\{ \begin{pmatrix} \epsilon_{t-1} \\ \vdots \\ \epsilon_{t-m} \end{pmatrix} \frac{\partial \epsilon_t}{\partial \theta'} \right\}. \tag{1.25}$$

Par un développement de Taylor de $\sqrt{n}\hat{\gamma}_m$, nous montrons que (voir la section 3.5.3)

$$\sqrt{n}\hat{\gamma}_m = \sqrt{n}\gamma_m + \Psi_m\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) + o_{\mathbb{P}}(1), \tag{1.26}$$

où Ψ_m est la matrice donnée dans (1.25). Nous prouvons également que (voir encore la section 3.5.3)

$$\sqrt{n}\hat{\rho}_m = \sqrt{n}\frac{\hat{\gamma}_m}{\sigma_\epsilon^2} + o_{\mathbb{P}}(1). \tag{1.27}$$

Ainsi, de (1.27), la distribution asymptotique des autocorrélations résiduelles $\sqrt{n}\hat{\rho}_m$ dépend de la distribution de $\hat{\gamma}_m$. D'après l'équation (1.26), l'étude de la distribution asymptotique jointe du vecteur $\sqrt{n}(\hat{\theta}_n' - \theta_0', \gamma_m')'$ nous permettra de déduire la loi asymptotique des autocovariances résiduelles $\sqrt{n}\hat{\gamma}_m$.

Boubacar Maïnassara et al. [2019] ont montré que $\hat{\theta}_n \to \theta_0 \in \mathring{\Theta}$ presque sûrement quand $n \to +\infty$. Comme $\partial Q_n(\hat{\theta}_n)/\partial\theta = 0$, un développement de Taylor de $\partial Q_n(\cdot)/\partial\theta$ autour de θ_0 implique que

$$\sqrt{n}\frac{\partial}{\partial\theta}O_n(\theta_0) + J(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0) = o_{\mathbb{P}}(1). \tag{1.28}$$

L'équation (1.28) est prouvée dans Boubacar Maïnassara et al. [2019] (voir la preuve du théorème 2). Par conséquent, de (1.28), nous avons

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -\frac{2}{\sqrt{n}}\sum_{t=1}^{n} J^{-1}(\theta_0)\epsilon_t(\theta_0)\frac{\partial\epsilon_t(\theta_0)}{\partial\theta} + o_{\mathbb{P}}(1). \tag{1.29}$$

24. Cette hypothèse a été introduite dans la section 1.5 pour établir la normalité asymptotique de l'estimateur des moindres carrés des modèles FARIMA faibles.

Pour $m, m' \in \mathbb{N}^*$, soit $\Gamma_{m,m'} = [\Gamma(\ell,\ell')]_{1\leq\ell\leq m, 1\leq\ell'\leq m'}$ la matrice définie pour tout $1 \leq \ell \leq m$ et pour tout $1 \leq \ell' \leq m'$ par

$$\Gamma(\ell,\ell') = \sum_{h=-\infty}^{\infty} \mathbb{E}\left[\epsilon_t \epsilon_{t-\ell} \epsilon_{t-h} \epsilon_{t-h-\ell'}\right].$$

L'existence de $\Gamma(\ell,\ell')$ est prouvée dans le lemme 3.3.

Proposition 1.2. *Supposons que $(\epsilon_t)_{t\in\mathbb{Z}}$ vérifie (1.4) et que θ_0 est à l'intérieur de Θ. Soit $(\hat{\theta}_n)_n$ une suite d'estimateurs des moindres carrés. Sous* **(H1)** *et* **(H2)** *avec $\tau = 4$, le vecteur aléatoire*

$$\sqrt{n}\left(\left(\hat{\theta}_n - \theta_0\right)', \gamma_m'\right)'$$

a une distribution asymptotique normale centrée et de matrice de variance

$$\Xi = \begin{pmatrix} \Sigma_{\hat{\theta}} & \Sigma_{\hat{\theta},\gamma_m} \\ \Sigma'_{\hat{\theta},\gamma_m} & \Gamma_{m,m} \end{pmatrix} = \sum_{h=-\infty}^{\infty} \mathbb{E}\left[U_t U_{t-h}'\right],$$

où

$$U_t = \begin{pmatrix} U_{1t} \\ U_{2t} \end{pmatrix} = \begin{pmatrix} -2J^{-1}(\theta_0)\epsilon_t(\theta_0)\frac{\partial}{\partial\theta}\epsilon_t(\theta_0) \\ (\epsilon_{t-1},\ldots,\epsilon_{t-m})'\epsilon_t \end{pmatrix}. \tag{1.30}$$

L'équation (1.30) est un résultat immédiat de (1.24) et (1.29).

La preuve de cette proposition est donnée dans la sous-section 3.5.2.

Le théorème suivant, qui est une extension du résultat donné dans Francq et al. [2005], fournit la distribution asymptotique des autocovariances et autocorrélations résiduelles des modèles FARIMA faibles :

Théorème 1.8. *Sous les hypothèses de la proposition 1.2, nous avons*

$$\sqrt{n}\hat{\gamma}_m \xrightarrow[n\to+\infty]{\text{en loi}} \mathcal{N}(0, \Sigma_{\hat{\gamma}_m}) \quad \text{où} \quad \Sigma_{\hat{\gamma}_m} = \Gamma_{m,m} + \Psi_m \Sigma_{\hat{\theta}} \Psi_m' + \Psi_m \Sigma_{\hat{\theta},\gamma_m} + \Sigma'_{\hat{\theta},\gamma_m} \Psi_m' \tag{1.31}$$

et

$$\sqrt{n}\hat{\rho}_m \xrightarrow[n\to+\infty]{\text{en loi}} \mathcal{N}(0, \Sigma_{\hat{\rho}_m}) \quad \text{où} \quad \Sigma_{\hat{\rho}_m} = \frac{1}{\sigma_\epsilon^4}\Sigma_{\hat{\gamma}_m}. \tag{1.32}$$

La preuve détaillée de ce théorème est présentée dans la sous-section 3.5.3.

Remarque 1.2. *Pour un modèle FARIMA(p, d_0, q), il est clair que d'après le théorème 1.8, la distribution asymptotique des autocorrélations résiduelles dépend uniquement de la distribution du bruit à travers les quantités $\Gamma(\ell,\ell')$ (qui dépendent quant à elles des moments d'ordre quatre du bruit). Il est aussi important de noter que cette distribution asymptotique dépend de la normalité asymptotique de l'estimateur des moindres carrés des modèles FARIMA(p, d_0, q) seulement à travers la matrice $\Sigma_{\hat{\theta}}$.*

Remarque 1.3. *Dans le cas des modèles FARIMA forts standards,* i.e. *quand le bruit* $(\epsilon_t)_{t\in\mathbb{Z}}$ *est considéré iid, Boubacar Maïnassara et al. [2019] (qui est le chapitre 2 de ce travail) ont montré dans la remarque 2 que* $I(\theta_0) = 2\sigma_\epsilon^2 J(\theta_0)$. *Ainsi la matrice de covariance asymptotique est réduite à* $\Sigma_{\hat\theta} = 2\sigma_\epsilon^2 J^{-1}(\theta_0)$. *Dans le cas fort, nous pouvons aussi montrer que* $\Gamma(\ell,\ell') = 0$ *quand* $\ell \neq \ell'$ *et que* $\Gamma(\ell,\ell) = \sigma_\epsilon^4$ *quand* $h = 0$. *Ainsi* $\Gamma_{m,m}$ *est réduite dans ce cas à* $\Gamma_{m,m} = \sigma_\epsilon^4 I_m$, *où* I_m *est la matrice identité de* $\mathbb{R}^{m\times m}$. *Comme* $\Sigma_{\hat\theta} = 2\sigma_\epsilon^2 J^{-1}(\theta_0)$, *nous obtenons que*

$$\begin{aligned}
\Sigma_{\hat\theta,\gamma_m} &= -2\sum_{h=-\infty}^{\infty} \mathbb{E}\left\{\epsilon_t J^{-1}(\theta_0)\frac{\partial\epsilon_t(\theta_0)}{\partial\theta}\right\}\left\{\begin{pmatrix}\epsilon_{t-1-h}\\ \vdots \\ \epsilon_{t-m-h}\end{pmatrix}\epsilon_{t-h}\right\}' \\
&= -\left(2\sigma_\epsilon^2 J^{-1}(\theta_0)\right)\left\{\mathbb{E}\left[\begin{pmatrix}\epsilon_{t-1}\\ \vdots \\ \epsilon_{t-m}\end{pmatrix}\frac{\partial\epsilon_t(\theta_0)}{\partial\theta'}\right]\right\}' \\
&= -\Sigma_{\hat\theta}\Psi_m'.
\end{aligned}$$

Notons respectivement par $\Sigma^{\mathrm{S}}_{\hat\gamma_m}$ *et* $\Sigma^{\mathrm{S}}_{\hat\rho_m}$ *les matrices de variances asymptotiques obtenues dans* (1.31) *et* (1.32) *pour le cas des modèles FARIMA forts. Nous pouvons facilement déduire que dans ce cas*

$$\Sigma^{\mathrm{S}}_{\hat\gamma_m} = \sigma_\epsilon^4 I_m - 2\sigma_\epsilon^2 \Psi_m J^{-1}(\theta_0)\Psi_m' \quad \text{et} \quad \Sigma^{\mathrm{S}}_{\hat\rho_m} = I_m - \frac{2}{\sigma_\epsilon^2}\Psi_m J^{-1}(\theta_0)\Psi_m',$$

qui sont les matrices obtenues par Li and McLeod [1986].

Pour valider un modèle FARIMA(p, d_0, q), la technique la plus basique dans ce contexte consiste à examiner la fonction d'autocorrélation des résidus. Le théorème 1.8 peut être utilisé pour obtenir des limites de signification asymptotiques pour les autocorrélations résiduelles. Cependant, les matrices de variances asymptotiques $\Sigma_{\hat\gamma_m}$ et $\Sigma_{\hat\rho_m}$ dépendent des matrices inconnues Ξ, Ψ_m et du scalaire strictement positif σ_ϵ^2 qui doivent être estimés.

1.6.2 Version modifiée du test portmanteau

Le résultat suivant donne la distribution limite des statistiques portmanteau standards (1.23) sous des hypothèses générales sur le processus des innovations du modèle FARIMA(p, d_0, q) ajusté. Ce résultat est une conséquence du théorème 1.8.

Théorème 1.9. *Sous les hypothèses du théorème 1.8 et sous* $(\mathcal{H}\mathbf{0})$, *les statistiques* Q_m^{BP} *et* Q_m^{LB} *introduites dans* (1.23) *convergent en distribution, quand* $n \to \infty$, *vers*

$$Z_m(\xi_m) = \sum_{k=1}^{m} \xi_{k,m} Z_k^2,$$

où $\xi_m = (\xi_{1,m},\ldots,\xi_{m,m})'$ *est le vecteur des valeurs propres de la matrice* $\Sigma_{\hat\rho_m} = \sigma_\epsilon^{-4}\Sigma_{\hat\gamma_m}$ *et* $Z_1,\ldots,Z_m$ *sont des variables aléatoires indépendantes et identiquement distribuées de loi* $\mathcal{N}(0,1)$.

Il est possible d'évaluer la distribution d'une forme quadratique d'un vecteur gaussien à l'aide de l'algorithme d'Imhof (voir Imhof [1961]).

Remarque 1.4. *D'après la remarque 1.3, quand m est grand,* $\Sigma^{\mathrm{S}}_{\hat{\rho}_m} \simeq I_m - 2\sigma_\epsilon^{-2}\Psi_m J^{-1}(\theta_0)\Psi_m'$ *est proche d'une matrice de projection. Ces valeurs propres sont donc égales à 0 et 1. Le nombre de valeurs propres égales à 1 est* $\mathrm{Tr}(I_m - 2\sigma_\epsilon^{-2}\Psi_m J^{-1}(\theta_0)\Psi_m') = \mathrm{Tr}(I_{m-(p+q+1)}) = m-(p+q+1)$*, donc* $p+q+1$ *valeurs propres égales à 0. La notation* $\mathrm{Tr}(\cdot)$ *désigne l'opérateur trace. Nous retrouvons ainsi le résultat bien connu obtenu par Li and McLeod [1986]. Plus précisément, sous l'hypothèse nulle* $(\mathcal{H}\mathbf{0})$ *et dans le cas des FARIMA forts, les distributions asymptotiques des statistiques* Q_m^{BP} *et* Q_m^{LB} *sont approximées par un* $\mathcal{X}^2_{m-(p+q+1)}$*, où* $m > p+q+1$ *et* $\mathcal{X}^2_k$ *est la distribution de khi-deux à k degrés de liberté. Le théorème 1.9 montre que cette approximation n'est plus valable dans le cadre de modèles FARIMA*(p, d_0, q) *faibles et que les distributions asymptotiques des statistiques* Q_m^{BP} *et* Q_m^{LB} *sous* $(\mathcal{H}\mathbf{0})$ *sont plus compliquées.*

La distribution limite $Z_m(\xi_m)$ dépend du paramètre de nuisance σ_ϵ^2, de la matrice Ψ_m et des éléments de Ξ. Il est donc nécessaire, lorsque les hypothèses sur le bruit sont faibles, de fournir un estimateur convergent de la matrice de covariance asymptotique $\Sigma_{\hat{\rho}_m}$ pour pouvoir appliquer le théorème 1.9 en pratique.

La matrice Ψ_m de $\mathbb{R}^{m\times(p+q+1)}$ et la variance du bruit σ_ϵ^2 peuvent être estimées empiriquement. Ainsi, nous pouvons utiliser

$$\hat{\Psi}_m = \frac{1}{n}\sum_{t=1}^{n}\left\{(\hat{e}_{t-1},\ldots,\hat{e}_{t-m})'\,\frac{\partial \hat{e}_t}{\partial \theta'}\right\} \quad \text{et} \quad \hat{\sigma}_\epsilon^2 = \hat{\gamma}(0) = \frac{1}{n}\sum_{t=1}^{n}\hat{e}_t^2.$$

Concernant la matrice Ξ, un estimateur convergent est obtenu en utilisant l'approche basée sur l'estimation de la densité spectrale par une autorégression paramétrique (voir par exemple Berk [1974], Boubacar Maïnassara et al. [2019], Boubacar Mainassara et al. [2012] et den Haan and Levin [1997] pour une explication plus complète de cette méthode). De (1.30), la matrice $\Xi/2\pi$ peut être interprétée comme la densité spectrale du processus stationnaire $(U_t)_{t\in\mathbb{Z}} = ((U_{1t}', U_{2t}')')_{t\in\mathbb{Z}}$ évaluée à la fréquence 0 (voir p. 459 de Brockwell and Davis [1991]). Cette estimation est donc basée sur l'expression suivante :

$$\Xi = A^{-1}(1)\Sigma_v A'^{-1}(1)$$

quand $(U_t)_{t\in\mathbb{Z}}$ satisfait une représentation AR(∞) de la forme

$$A(L)U_t := U_t - \sum_{k=1}^{\infty} A_k U_{t-k} = v_t, \tag{1.33}$$

tels que $\sum_{k=1}^{\infty}\|A_k\| < \infty$ et $\det\{A(z)\} \neq 0$ pour tout $|z| \leq 1$. Le processus $(v_t)_{t\in\mathbb{Z}}$ est un bruit blanc faible $(p+q+1+m)$-dimensionnel dont la matrice de variance est notée par Σ_v. Il est prouvé dans Boubacar Maïnassara [2009], Lütkepohl [2005] qu'il existe $K > 0$ et $0 < \rho < 1$ tel que

$$\|A_k\| \leq K\,\rho^k. \tag{1.34}$$

Puisque U_t est non observable, nous introduisons $\hat{U}_t \in \mathbb{R}^{p+q+1+m}$ obtenu en remplaçant $\epsilon_t(\theta_0)$ par $\tilde{\epsilon}_t(\hat{\theta}_n)$ et $J(\theta_0)$ par sa partie empirique ou observable $\hat{J}_n$ dans (1.30). Soit $\hat{\Delta}_r(z) = I_{p+q+1+m} - \sum_{k=1}^{r} \hat{\Delta}_{r,k} z^k$, où $\hat{\Delta}_{r,1}, \ldots, \hat{\Delta}_{r,r}$ dénotent les coefficients de la régression linéaire de $\hat{U}_t$ sur $\hat{U}_{t-1}, \ldots, \hat{U}_{t-r}$. Notons par $\hat{v}_{r,t}$ les résidus de cette régression et soit $\hat{\Sigma}_{\hat{v}_r}$ la variance empirique de $\hat{v}_{r,1}, \ldots, \hat{v}_{r,n}$. Nous sommes maintenant en mesure d'énoncer le théorème 1.10 qui est une extension du résultat donné dans Boubacar Mainassara et al. [2012].

Théorème 1.10. *Supposons que $(\epsilon_t)_{t\in\mathbb{Z}}$ vérifie (1.4) et que θ_0 est à l'intérieur de Θ. Supposons de plus que le processus $(U_t)_{t\in\mathbb{Z}}$ défini dans (1.30) admet une représentation AR(∞) faible comme dans (1.33) pour laquelle la matrice $\Sigma_v = \mathrm{Var}(v_t)$ est non-singulière. Sous* **(H1)** *et* **(H2')** *avec $\tau = 8$, l'estimateur spectral de Ξ satisfait*

$$\hat{\Xi}_n^{\mathrm{SP}} := \hat{\Delta}_r^{-1}(1)\hat{\Sigma}_{\hat{v}_r}\hat{\Delta}_r'^{-1}(1) \xrightarrow[n\to+\infty]{\mathbb{P}} \Xi = \Delta^{-1}(1)\Sigma_v\Delta^{-1}(1)$$

où r dépend de n et vérifie $\lim_{n\to+\infty} r^5(n)/n^{1-2(d_2-d_1)} = 0$ (rappelons que $d_0 \in [d_1,d_2] \subset]0,1/2[$).

La preuve de ce théorème est similaire à celle du théorème 3 dans Boubacar Maïnassara et al. [2019].

Nous pouvons ainsi définir les versions modifiées des tests portemanteau de Box-Pierce (BP) et de Ljung-Box (LB). Quand le terme d'erreur $(\epsilon_t)_{t\in\mathbb{Z}}$ est un bruit blanc fort et quand le nombre m des autocorrélations résiduelles est suffisamment grand, les versions standards des tests portemanteau sont des outils pratiques pour détecter si les ordres p et q d'un modèle FARIMA(p, d_0, q) sont bien choisis (voir la remarque 1.4). Ces versions donnent en général des résultats totalement incorrects lorsque le bruit présente des dépendances non linéaires.

Les versions modifiées, que nous nous apprêtons à introduire, visent à détecter si les ordres p et q d'un modèle FARIMA(p, d_0, q) faible sont bien choisis. Ces tests restent également valides asymptotiquement pour les modèles FARIMA(p, d_0, q) forts, même pour un m petit. Les versions modifiées des tests portemanteau seront notées $\mathrm{BP_W}$ et $\mathrm{LB_W}$, l'indice W fait référence au terme faible.

Notons par $\hat{\Sigma}_{\hat{\rho}_m}$ la matrice obtenue en remplaçant Ξ par $\hat{\Xi}$ et σ_ϵ^2 par $\hat{\sigma}_\epsilon^2$ dans $\Sigma_{\hat{\rho}_m}$. Soit $\hat{\xi}_m = (\hat{\xi}_{1,m}, \ldots, \hat{\xi}_{m,m})'$ le vecteur des valeurs propres de $\hat{\Sigma}_{\hat{\rho}_m}$. Au seuil asymptotique α, le test $\mathrm{BP_W}$ (respectivement le test $\mathrm{LB_W}$) consiste à rejeter l'hypothèse nulle d'un modèle FARIMA(p, d_0, q) faible (l'adéquation d'un modèle FARIMA(p, d_0, q) faible) quand

$$Q_m^{\mathrm{BP}} > S_m(1-\alpha) \quad (\text{respectivement} \quad Q_m^{\mathrm{LB}} > S_m(1-\alpha)),$$

où $S_m(1-\alpha)$ est tel que $\mathbb{P}(Z_m(\hat{\xi}_m) > S_m(1-\alpha)) = \alpha$. Nous soulignons le fait que les versions modifiées des tests portemanteau de Box-Pierce et de Ljung-Box sont plus difficiles à mettre en œuvre. En effet, les valeurs critiques de la loi limite de leurs statistiques doivent être calculées à partir des données. Les valeurs critiques de la méthode standard sont, quand à elles, simplement

déduites d'une table χ^2. Les p-valeurs

$$\mathbb{P}\left\{Z_m(\hat{\xi}_m) > Q_m^{\text{BP}}\right\} \quad \text{et} \quad \mathbb{P}\left\{Z_m(\hat{\xi}_m) > Q_m^{\text{LB}}\right\}, \text{avec } Z_m(\hat{\xi}_m) = \sum_{i=1}^{m} \hat{\xi}_{i,m} Z_i^2,$$

sont évaluées par l'algorithme de Imhof (voir Imhof [1961]).

Une deuxième méthode évitant l'estimation de la matrice de variance asymptotique du vecteur joint de l'estimateur des moindres carrés et des autocovariances empiriques est proposée dans la prochaine sous-section.

1.6.3 La distribution asymptotique auto-normalisée des autocorrélations résiduelles

Au vu du théorème 1.9, la distribution asymptotique des statistiques de test définies dans (1.23) est une somme pondérée de khi-deux. Nous avons montré que les coefficients de pondération sont les valeurs propres de la matrice de covariance asymptotique $\Sigma_{\hat{\rho}_m}$, du vecteur des autocorrélations résiduelles, obtenue dans le théorème 1.8. Nous avons aussi souligné que cette matrice de variance asymptotique dépend des matrices inconnues Ξ et Ψ_m, et de la variance du bruit σ_ϵ^2. L'applicabilité en pratique du résultat obtenu dans le théorème 1.8 a été donc contrainte à l'estimation de ces paramètres inconnus.

Nous avons vu précédemment que l'obtention d'un estimateur convergent de la matrice de covariance asymptotique $\Sigma_{\hat{\rho}_m}$ du vecteur des autocorrélations résiduelles nécessite le développement d'un estimateur au moins faiblement consistent de la matrice Ξ. Cette matrice, contrairement à Ψ_m et au scalaire σ_ϵ^2, ne peut pas être estimée empiriquement. La solution que nous avons proposé était d'estimer la densité spectrale du processus stationnaire $(U_t)_{t\in\mathbb{Z}}$ par une autorégression paramétrique. Ceci nous a permis par la suite de fournir un estimateur convergent de la matrice Ξ qui n'est rien d'autre que 2π la densité spectrale du processus $(U_t)_{t\in\mathbb{Z}}$ évaluée à la fréquence 0 (voir le théorème 1.10).
Cependant, cette approche pose le problème du choix du paramètre de troncature. En effet, cette méthode est basée sur une représentation autorégressive infinie du processus stationnaire $(U_t)_{t\in\mathbb{Z}}$ (voir (1.33)). Le choix de l'ordre de troncature est donc crucial et difficile.

Dans cette sous-section, nous proposons une méthode alternative où nous n'estimons pas la matrice de covariance asymptotique de la distribution jointe du vecteur de l'estimateur des moindres carrés et des autocovariances empiriques. Cette méthode est une extension des résultats obtenus par Boubacar Maïnassara and Saussereau [2018]. Elle est basée sur une approche d'auto-normalisation pour construire une statistique de test dont la distribution asymptotique dépend uniquement des ordres p et q de la partie autorégressive et moyennes-mobiles du modèle. Cette approche a été étudiée par Boubacar Maïnassara and Saussereau [2018] dans le cas des modèles ARMA faibles. Boubacar Maïnassara and Saussereau [2018] ont proposé des versions modifiées des statistiques portemanteau standards de Box-Pierce et de Ljung-Box.

En plus du fait qu'elle permet de contourner le problème de l'estimation de la matrice de variance asymptotique Ξ, cette technique nous épargne le calcul (à partir des données) des valeurs critiques de la distribution limite des statistiques de tests. En effet, les valeurs critiques de la loi limite de la statistique de test proposée par la méthode basée sur l'approche d'auto-normalisation sont tabulées dans Lobato [2001]. En quelques sortes, cette méthode est finalement plus proche de la méthode standard dans laquelle les valeurs critiques sont simplement déduites d'une table $\mathcal{X}^2$.

Notons par Λ la matrice par blocs de $\mathbb{R}^{m\times(p+q+1+m)}$ définie par $\Lambda = (\Psi_m | I_m)$. De (1.26) et (1.29), nous déduisons que

$$\sqrt{n}\hat{\gamma}_m = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Lambda U_t + o_{\mathbb{P}}(1).$$

À ce stade, la méthode classique qui consiste à estimer la matrice de covariance asymptotique Ξ ne sera pas considérée. Nous essayons plutôt d'appliquer le lemme 1 dans Lobato [2001]. Il faut donc vérifier qu'un théorème central limite fonctionnel est valable pour le processus $U := (U_t)_{t\geq 1}$. Pour cela, nous définissons la matrice de normalisation C_m de $\mathbb{R}^{m\times m}$ par

$$C_m = \frac{1}{n^2}\sum_{t=1}^{n} S_t S_t',$$

où

$$S_t = \sum_{j=1}^{t}(\Lambda U_j - \gamma_m).$$

L'inversibilité de cette matrice est assurée lorsque le processus $(\epsilon_t)_{t\in\mathbb{Z}}$ admet une densité strictement positive sur un voisinage de zéro.

Proposition 1.3. *Supposons que $(\epsilon_t)_{t\in\mathbb{Z}}$ a une densité strictement positive sur un voisinage de zéro. Sous les hypothèses du théorème 1.8, la matrice C_m est presque-sûrement non singulière.*

La preuve de cette proposition est donnée dans la sous-section 3.5.4.

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

$$\mathcal{U}_K = B_K'(1) V_K^{-1} B_K(1), \tag{1.35}$$

où

$$V_K = \int_0^1 (B_K(r) - rB_K(1))(B_K(r) - rB_K(1))' \, dr. \tag{1.36}$$

Les valeurs critiques de $\mathcal{U}_K$ ont été tabulées par Lobato [2001].

Le théorème suivant énonce la distribution asymptotique auto-normalisée des autocovariances et autocorrélations résiduelles.

Théorème 1.11. *Supposons que $(\epsilon_t)_{t\in\mathbb{Z}}$ a une densité strictement positive sur un voisinage de zéro. Sous les hypothèses du théorème 1.8,*

$$n\hat{\gamma}_m' C_m^{-1}\hat{\gamma}_m \xrightarrow[n\to+\infty]{\text{en loi}} \mathcal{U}_m \text{ et } n\sigma_\epsilon^4 \hat{\rho}_m' C_m^{-1}\hat{\rho}_m \xrightarrow[n\to+\infty]{\text{en loi}} \mathcal{U}_m.$$

La preuve de ce théorème est donnée dans la sous-section 3.5.5.

Le théorème 1.11 est inutilisable dans la pratique car la matrice de normalisation C_m et le paramètre de nuisance σ_ϵ^2 ne sont pas observables. Ce problème sera corrigé lorsque nous remplaçons la matrice C_m et la variance du bruit σ_ϵ^2 par leurs parties empiriques ou observables. Ainsi, nous notons

$$\hat{C}_m = \frac{1}{n^2}\sum_{t=1}^{n} \hat{S}_t \hat{S}_t',$$

où

$$\hat{S}_t = \sum_{j=1}^{t} \left(\hat{\Lambda}\hat{U}_j - \hat{\gamma}_m\right),$$

avec $\hat{\Lambda} = (\hat{\Psi}_m | I_m)$. Le vecteur $\hat{U}_t$ et la variable $\hat{\sigma}_\epsilon^2$ sont définis dans la sous-section 1.6.2.

Les quantités ci-dessus sont observables et nous sommes donc en mesure d'énoncer la version applicable du théorème 1.11.

Théorème 1.12. *Supposons que $(\epsilon_t)_{t\in\mathbb{Z}}$ a une densité strictement positive sur un voisinage de zéro. Sous les hypothèses du théorème 1.11,*

$$n\hat{\gamma}_m' \hat{C}_m^{-1}\hat{\gamma}_m \xrightarrow[n\to+\infty]{\text{en loi}} \mathcal{U}_m \text{ et } Q_m^{\text{SN}} = n\hat{\sigma}_\epsilon^4 \hat{\rho}_m' \hat{C}_m^{-1}\hat{\rho}_m \xrightarrow[n\to+\infty]{\text{en loi}} \mathcal{U}_m.$$

La preuve de ce théorème est donnée dans la sous-section 3.5.6.

Ainsi, en se basant sur ce dernier théorème, nous proposons la version modifiée de la statistique de Ljung-Box suivante :

$$\tilde{Q}_m^{\text{SN}} = n\hat{\sigma}_\epsilon^4 \hat{\rho}_m' D_{n,m}^{1/2} \hat{C}_m^{-1} D_{n,m}^{1/2} \hat{\rho}_m,$$

où $D_{n,m} \in \mathbb{R}^{m\times m}$ est une matrice diagonale avec $(n+2)/(n-1), \ldots, (n+2)/(n-m)$ sur la diagonale.

Ces versions modifiées des tests portmanteau standards seront notées par BP_{SN} et LB_{SN}, l'indice SN fait référence au terme auto-normalisé.

Au seuil asymptotique α, le test BP_{SN} (respectivement le test LB_{SN}) consiste à rejeter l'hypothèse nulle d'un modèle FARIMA(p, d_0, q) faible (l'adéquation d'un modèle FARIMA(p, d_0, q) faible) quand

$$Q_m^{\text{SN}} > \mathcal{U}_{m,1-\alpha} \quad (\text{respectivement} \quad Q_m^{\text{SN}} > \mathcal{U}_{m,1-\alpha}),$$

où $\mathcal{U}_{m,1-\alpha}$ est tel que $\mathbb{P}(\mathcal{U}_m > \mathcal{U}_{m,1-\alpha}) = \alpha$.

1.7 Résultats du chapitre 4

L'objectif principal de ce chapitre est l'estimation des modèles autorégressifs d'ordre 1 induits par un bruit gaussien fractionnaire (voir 8) d'indice de Hurst H (supposé connu). Le comportement de mémoire longue dans ce cas est contrôlé implicitement par le bruit (voir l'exemple 1.1 de la section 1.1). La dynamique de mémoire courte est, quant à elle, modélisée par la partie autorégressive du modèle. Nous considérons le problème de l'estimation des moindres carrés généralisés des modèles autorégressifs d'ordre 1 induits par un bruit gaussien fractionnaire. Plus précisément, nous étudions la convergence et la normalité asymptotique de l'estimateur des moindres carrés généralisés du paramètre autorégressif du processus stationnaire défini, pour tout $t \in \mathbb{Z}$, par

$$X_t = a_0 X_{t-1} + \epsilon_t^{\mathrm{H}}, \tag{1.37}$$

où $(\epsilon_t^{\mathrm{H}})_{t\in\mathbb{Z}}$ est un bruit gaussien fractionnaire d'indice de Hurst H. Le paramètre autorégressif a_0 et le paramètre de Hurst H, associé au bruit gaussien fractionnaire $(\epsilon_t^{\mathrm{H}})_{t\in\mathbb{Z}}$, sont supposés satisfaire l'hypothèse standard suivante :

(H0) : $a_0 \in]-1,1[$ et $\mathrm{H} \in]0,1[$.

La condition $a_0 \in]-1,1[$ permet d'assurer l'existence, l'unicité et la causalité d'une solution stationnaire de (1.37) (voir le théorème 1.2).

Nous présentons dans la sous-section suivante le cadre et l'énoncé du problème. Les principaux résultats sur l'étude asymptotique de l'estimateur des moindres carrés généralisés des modèles autorégressifs d'ordre 1 induits par un bruit gaussien fractionnaire (ou simplement des modèles autorégressifs fractionnaires d'ordre 1 (FrAR(1)) seront énoncés ultérieurement dans une autre sous-section.

1.7.1 Estimation des moindres carrés généralisés des modèles FrAR d'ordre 1 lorsque l'exposant de Hurst est connu

Pour tout $a \in]-1,1[$, notons par $(\epsilon_t^{\mathrm{H}}(a))_{t\in\mathbb{Z}}$ le bruit gaussien fractionnaire défini comme solution de

$$\epsilon_t^{\mathrm{H}}(a) = X_t - aX_{t-1}, \quad \forall t \in \mathbb{Z}. \tag{1.38}$$

Remarquons que la distribution de $\epsilon_t^{\mathrm{H}}(a)$ ne dépend pas de a. Notons aussi que pour tout $t \in \mathbb{Z}$, $\epsilon_t^{\mathrm{H}}(a_0) = \epsilon_t^{\mathrm{H}}$ presque-sûrement. Étant donné une réalisation $X_1, \ldots, X_{n+1}$ de longueur $n+1$, $\epsilon_t^{\mathrm{H}}(a)$ peut être approximée, pour $0 < t \leq n+1$, par $e_t^{\mathrm{H}}(a)$ définie récursivement par

$$e_t^{\mathrm{H}}(a) = X_t - aX_{t-1}, \tag{1.39}$$

avec $e_t^{\mathrm{H}}(a) = X_t = 0$ si $t \leq 0$. Il est clair que ces valeurs initiales sont asymptotiquement négligeables et ceci uniformément en a. En effet, $\epsilon_t^{\mathrm{H}}(a) - e_t^{\mathrm{H}}(a) \to 0$ quand $t \to +\infty$ pour

tout $a \in]-1,1[$. Le choix de ces valeurs initiales n'a donc aucune influence sur les propriétés asymptotiques de l'estimateur du paramètre autorégressif du modèle.

De (1.39), nous déduisons que

$$\begin{pmatrix} X_2 \\ \vdots \\ X_{n+1} \end{pmatrix} = a \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} + \begin{pmatrix} e_2^{\mathsf{H}}(a) \\ \vdots \\ e_{n+1}^{\mathsf{H}}(a) \end{pmatrix}. \tag{1.40}$$

Par construction, le vecteur aléatoire $(e_2^{\mathsf{H}}(a), \ldots, e_{n+1}^{\mathsf{H}}(a))^{\mathsf{T}}$ donné dans (1.40) est égal au vecteur $(\epsilon_2^{\mathsf{H}}(a), \ldots \epsilon_{n+1}^{\mathsf{H}}(a))^{\mathsf{T}}$, où $(\epsilon_t^{\mathsf{H}}(a))_{t\in\mathbb{Z}}$ est le processus des erreurs théoriques introduit dans (1.38). En exploitant les propriétés du bruit gaussien fractionnaire $(\epsilon_t^{\mathsf{H}}(a))_{t\in\mathbb{Z}}$, nous pouvons facilement montrer que le vecteur aléatoire $(e_2^{\mathsf{H}}(a), \ldots, e_{n+1}^{\mathsf{H}}(a))^{\mathsf{T}}$ est gaussien centré de matrice de covariance $\Omega_{n,\mathsf{H}} := [\gamma_{\epsilon^{\mathsf{H}}}(j-i)]_{1\leq i,j\leq n} = [\mathrm{Cov}(\epsilon_i^{\mathsf{H}}, \epsilon_j^{\mathsf{H}})]_{1\leq i,j\leq n}$.

Remarque 1.5. *La matrice de covariance $\Omega_{n,\mathsf{H}}$ dépend uniquement du paramètre de Hurst* H. *Comme $\Omega_{n,\mathsf{H}}$ est réelle symétrique, elle est diagonalisable à l'aide d'une matrice de passage orthogonale. Plus explicitement, il existe $P_{n,\mathsf{H}}$ réelle orthogonale telle que $D_{n,\mathsf{H}} = P_{n,\mathsf{H}}^{\mathsf{T}} \Omega_{n,\mathsf{H}} P_{n,\mathsf{H}}$ est une matrice diagonale. En prenant en considération le fait que la matrice $\Omega_{n,\mathsf{H}}$ est définie positive (voir la proposition 4.1 ci-après), nous obtenons que $\Omega_{n,\mathsf{H}}^{1/2} := P_{n,\mathsf{H}} D_{n,\mathsf{H}}^{1/2} P_{n,\mathsf{H}}^{\mathsf{T}}$ où $D_{n,\mathsf{H}}^{1/2} = \mathrm{diag}((\lambda_{1,\mathsf{H}}^{(n)})^{1/2}, \ldots, (\lambda_{n,\mathsf{H}}^{(n)})^{1/2})$ et les $\lambda_{i,\mathsf{H}}^{(n)}$ sont les valeurs propres de $\Omega_{n,\mathsf{H}}$. La matrice $\Omega_{n,\mathsf{H}}^{1/2}$ est une racine carrée de $\Omega_{n,\mathsf{H}}$ puisqu'elle vérifie $\Omega_{n,\mathsf{H}}^{1/2} \Omega_{n,\mathsf{H}}^{1/2} = \Omega_{n,\mathsf{H}}$.*

Remarque 1.6. *L'estimateur des moindres carrés ordinaires du paramètre autorégressif a_0 est biaisé. Pour voir cela, notons par $\hat{a}_n^{(\mathrm{OLS})}$ l'estimateur des moindres carrés standard de a_0. Il est bien connu que*

$$\hat{a}_n^{(\mathrm{OLS})} = \frac{\sum_{t=1}^n X_t X_{t-1}}{\sum_{t=1}^n X_{t-1}^2}.$$

Grâce à (1.37), nous avons

$$\hat{a}_n^{(\mathrm{OLS})} - a_0 = \frac{\sum_{t=1}^n \epsilon_t^{\mathsf{H}} X_{t-1}}{\sum_{t=1}^n X_{t-1}^2}.$$

L'ergodicité de $(\epsilon_t^{\mathsf{H}})_{t\in\mathbb{Z}}$ et la stationnarité de $(X_t)_{t\in\mathbb{Z}}$ impliquent que

$$\hat{a}_n^{(\mathrm{OLS})} - a_0 \xrightarrow[n\to\infty]{\text{p.s.}} \frac{\mathbb{E}\left[\epsilon_2^{\mathsf{H}} X_1\right]}{\mathrm{Var}\,(X_1)}.$$

En utilisant l'inversibilité du polynôme autorégressif $a_0(z) = 1 - a_0 z$ lorsque le paramètre a_0 vérifie **(H0)** *et la structure de la fonction d'autocovariance du processus $(\epsilon_t^{\mathsf{H}})_{t\in\mathbb{Z}}$ (voir l'exemple 1.1 de la section 1.1), nous obtenons que*

$$\mathbb{E}\left[\epsilon_2^{\mathsf{H}} X_1\right] = \frac{a_0 - 2}{2} + \frac{(a_0-1)^2}{2} \sum_{j\geq 2} a_0^{j-2} j^{2\mathsf{H}}.$$

Puisque la fonction $\mathsf{H} \in]0,1[\to \sum_{j\geq 2} a_0^{j-2} j^{2\mathsf{H}}$ *est bijective, nous déduisons que* $\mathbb{E}[\epsilon_2^{\mathsf{H}} X_1] = 0$ *si et seulement si* $\mathsf{H} = 1/2$. *Ceci nous ramène au cas standard des modèles autorégressifs induits par un bruit indépendant et identiquement distribué dont la loi est distribution normale centrée réduite.*

À cause de ce problème de biais de l'estimateur des moindres carrés ordinaires du paramètre a_0 dans le modèle (1.37) (problème dû à la corrélation du bruit), nous considérons ici l'estimation des moindres carrés généralisés. Cette technique consiste à transformer le modèle (1.39) en un autre modèle autorégressif induit par un bruit blanc fort standard (*i.e.*, une suite indépendante et identiquement distribuée de variables aléatoires de loi $\mathcal{N}(0,1)$) tout en gardant le même paramètre autorégressif inconnu a_0. Nous étudions ensuite l'estimation des moindres carrés ordinaires du nouveau modèle obtenu.

Au vu de la remarque 1.5, le vecteur $\Omega_{n,\mathsf{H}}^{-1/2}(e_2^{\mathsf{H}}(a),\ldots,e_{n+1}^{\mathsf{H}}(a))^{\mathsf{T}}$ est gaussien centré réduit. Ainsi, une transformation naturelle qui peut être proposée pour obtenir ce modèle autorégressif intermédiaire standard consiste à considérer une représentation vectorielle du modèle (1.39) et de multiplier tous les termes de l'équation obtenue par la matrice appropriée $\Omega_{n,\mathsf{H}}^{-1/2}$. Plus précisément, soit

$$Y_{n,\mathsf{H}} = \Omega_{n,\mathsf{H}}^{-1/2}\begin{pmatrix} X_2 \\ \vdots \\ X_{n+1}\end{pmatrix},\ Z_{n,\mathsf{H}} = \Omega_{n,\mathsf{H}}^{-1/2}\begin{pmatrix} X_1 \\ \vdots \\ X_{n}\end{pmatrix} \text{ et } U_{n,\mathsf{H}}(a) = \Omega_{n,\mathsf{H}}^{-1/2}\begin{pmatrix} e_2^{\mathsf{H}}(a) \\ \vdots \\ e_{n+1}^{\mathsf{H}}(a)\end{pmatrix}. \tag{1.41}$$

Alors, de (1.39), nous avons

$$Y_{n,\mathsf{H}} = aZ_{n,\mathsf{H}} + U_{n,\mathsf{H}}(a). \tag{1.42}$$

Le modèle (1.42) est un modèle autorégressif standard théorique associé au modèle de départ (1.37). L'avantage de ce modèle est la structure de son bruit, le terme d'erreur du modèle (1.42) est un vecteur gaussien centré réduit. Par conséquent, pour estimer le paramètre inconnu a_0, nous utilisons ce modèle et nous appliquons la procédure d'estimation des moindres carrés ordinaires pour obtenir un estimateur convergent et asymptotiquement normal.

La variable aléatoire $\hat{a}_n$ est dite estimateur des moindres carrés généralisés si elle satisfait, presque sûrement,

$$\hat{a}_n = \underset{a\in]-1,1[}{\operatorname{argmin}} \frac{1}{n}\left\|U_{n,\mathsf{H}}(a)\right\|_{\mathbb{R}^n}^2. \tag{1.43}$$

La résolution du problème d'optimisation présenté dans (1.43) permet d'obtenir la formule explicite de $\hat{a}_n$. Nous pouvons facilement prouver que

$$\hat{a}_n = \frac{Z_{n,\mathsf{H}}^{\mathsf{T}} Y_{n,\mathsf{H}}}{\left\|Z_{n,\mathsf{H}}\right\|_{\mathbb{R}^n}^2}. \tag{1.44}$$

De plus, au vu de cette dernière égalité et de l'équation (1.42), nous déduisons que

$$\hat{a}_n - a_0 = \frac{Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)}{\|Z_{n,\mathrm{H}}\|_{\mathbb{R}^n}^2}. \tag{1.45}$$

1.7.2 Propriétés asymptotiques de l'estimateur des moindres carrés généralisés

Les propriétés asymptotiques de l'estimateur des moindres carrés généralisés du paramètre a_0 du modèle autorégressif fractionnaire d'ordre 1 introduit dans (1.37) sont énoncées, lorsque le paramètre de Hurst H est supposé connu, dans les deux théorèmes suivants.

Théorème 1.13. *(Convergence). Supposons que $(X_t)_{t\in\mathbb{Z}}$ satisfait (1.37). Sous **(H0)**, nous avons*

$$\hat{a}_n \xrightarrow[n\to+\infty]{\mathbb{P}} a_0,$$

où $(\hat{a}_n)_{n\geq 1}$ est une suite d'estimateurs des moindres carrés généralisés comme définis dans (1.44).

La preuve de ce théorème est donnée dans la sous-section 4.4.2.

Théorème 1.14. *(Normalité asymptotique). Supposons que $(X_t)_{t\in\mathbb{Z}}$ satisfait (1.37). Sous **(H0)**, la suite des variables aléatoires $(\sqrt{n}(\hat{a}_n - a_0))_{n\geq 1}$ a une distribution asymptotique centrée et de variance $1 - a_0^2$.*

La preuve de ce théorème est présentée en détails dans la sous-section 4.4.3.

Bibliographie

Baillie, R. T., Chung, C.-F., and Tieslau, M. A. (1996). Analysing inflation by the fractionally integrated ARFIMA-GARCH model. *Journal of applied econometrics*, 11(1) :23–40.

Bauwens, L., Laurent, S., and Rombouts, J. V. K. (2006). Multivariate GARCH models : a survey. *J. Appl. Econometrics*, 21(1) :79–109.

Beran, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *J. Roy. Statist. Soc. Ser. B*, 57(4) :659–672.

Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2013). *Long-memory processes*. Springer, Heidelberg. Probabilistic properties and statistical methods.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *Ann. Statist.*, 2 :489–502. Collection of articles dedicated to Jerzy Neyman on his 80th birthday.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3) :307 – 327.

Boubacar Maïnassara, Y. (2009). *Estimation, validation et identification des modèles ARMA faibles multivariés*. PHD thesis of Université Lille 3.

Boubacar Mainassara, Y. (2011). Multivariate portmanteau test for structural VARMA models with uncorrelated but non-independent error terms. *J. Statist. Plann. Inference*, 141(8) :2961–2975.

Boubacar Maïnassara, Y. (2014). Estimation of the variance of the quasi-maximum likelihood estimator of weak VARMA models. *Electron. J. Stat.*, 8(2) :2701–2740.

Boubacar Mainassara, Y., Carbon, M., and Francq, C. (2012). Computing and estimating information matrices of weak ARMA models. *Comput. Statist. Data Anal.*, 56(2) :345–361.

Boubacar Maïnassara, Y., Esstafa, Y., and Saussereau, B. (2019). Estimating FARIMA models with uncorrelated but non-independent error terms. *Working document*.

Boubacar Mainassara, Y. and Francq, C. (2011). Estimating structural VARMA models with uncorrelated but non-independent error terms. *J. Multivariate Anal.*, 102(3) :496–505.

Boubacar Maïnassara, Y. and Saussereau, B. (2018). Diagnostic checking in multivariate arma models with dependent errors using normalized residual autocorrelations. *J. Amer. Statist. Assoc.*, 113(524) :1813–1827.

Box, G. E. P. and Jenkins, G. M. (1970). *Times series analysis. Forecasting and control*. Holden-Day, San Francisco, Calif.-London-Amsterdam.

Box, G. E. P. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Amer. Statist. Assoc.*, 65 :1509–1526.

Brockwell, P. J. and Davis, R. A. (1991). *Time series : theory and methods*. Springer Series in Statistics. Springer-Verlag, New York, second edition.

Chiu, S.-T. (1988). Weighted least squares estimators on the frequency domain for the parameters of a time series. *Ann. Statist.*, 16(3) :1315–1326.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.*, 17(4) :1749–1766.

den Haan, W. J. and Levin, A. T. (1997). A practitioner's guide to robust covariance matrix estimation. In *Robust inference*, volume 15 of *Handbook of Statist.*, pages 299–342. North-Holland, Amsterdam.

Doukhan, P. and León, J. (1989). Cumulants for stationary mixing random sequences and applications to empirical spectral density. *Probab. Math. Stat*, 10 :11–26.

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingrom inflation. *Econometrica*, 50 :391–407.

Fan, J. and Yao, Q. (2008). *Nonlinear time series : nonparametric and parametric methods*. Springer Science & Business Media.

Fox, R. and Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.*, 14(2) :517–532.

Francq, C., Roy, R., and Zakoïan, J.-M. (2005). Diagnostic checking in ARMA models with uncorrelated errors. *J. Amer. Statist. Assoc.*, 100(470) :532–544.

Francq, C. and Zakoïan, J.-M. (1998). Estimating linear representations of nonlinear processes. *J. Statist. Plann. Inference*, 68(1) :145–165.

Francq, C. and Zakoïan, J.-M. (2000). Covariance matrix estimation for estimators of mixing weak ARMA models. *J. Statist. Plann. Inference*, 83(2) :369–394.

Francq, C. and Zakoïan, J.-M. (2005). Recent results for linear time series models with non independent innovations. In *Statistical modeling and analysis for complex data problems*, volume 1 of *GERAD 25th Anniv. Ser.*, pages 241–265. Springer, New York.

Francq, C. and Zakoïan, J.-M. (2007). HAC estimation and strong linearity testing in weak ARMA models. *J. Multivariate Anal.*, 98(1) :114–144.

Francq, C. and Zakoïan, J.-M. (2010). *GARCH Models : Structure, Statistical Inference and Financial Applications*. Wiley.

Giraitis, L. and Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probab. Theory Related Fields*, 86(1) :87–104.

Gradshteyn, I. S. and Ryzhik, I. M. (2007). *Table of integrals, series, and products*. Elsevier/Academic Press, Amsterdam, seventh edition. Translated from the Russian, Translation edited and with a preface by Alan Jeffrey and Daniel Zwillinger, With one CD-ROM (Windows, Macintosh and UNIX).

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal.*, 1(1) :15–29.

Gromykov, G., Haye, M. O., and Philippe, A. (2018). A frequency-domain test for long range dependence. *Statistical Inference for Stochastic Processes*, 21(3) :513–526.

Hauser, M. and Kunst, R. (1998). Fractionally integrated models with arch errors : with an application to the swiss 1-month euromarket interest rate. *Review of Quantitative Finance and Accounting*, 10(1) :95–113.

Hong, Y. (1996). Consistent testing for serial correlation of unknown form. *Econometrica*, 64(4) :837–864.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68(1) :165–176.

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng.*, 116 :770–799.

Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48 :419–426.

Keenan, D. M. (1987). Limiting behavior of functionals of higher-order sample cumulant spectra. *Ann. Statist.*, 15(1) :134–151.

Kuan, C.-M. and Lee, W.-M. (2006). Robust M tests without consistent estimation of the asymptotic covariance matrix. *J. Amer. Statist. Assoc.*, 101(475) :1264–1275.

Li, W. K. and McLeod, A. I. (1986). Fractional time series modelling. *Biometrika*, 73(1) :217–221.

Ling, S. and Li, W. K. (1997). On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *J. Amer. Statist. Assoc.*, 92(439) :1184–1194.

Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2) :pp. 297–303.

Lobato, I. N. (2001). Testing that a dependent process is uncorrelated. *J. Amer. Statist. Assoc.*, 96(455) :1066–1076.

Lobato, I. N., Nankervis, J. C., and Savin, N. E. (2001). Testing for autocorrelation using a modified box- pierce q test. *Inter. Econ. Review*, 42(1) :187–205.

Lobato, I. N., Nankervis, J. C., and Savin, N. E. (2002). Testing for zero autocorrelation in the presence of statistical dependence. *Econ. Theory*, 18(3) :730–743.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer-Verlag, Berlin.

Mandelbrot, B. B. (1965). Une classe de processus stochastiques homothétiques à soi ; application à la loi climatologique de H. E. Hurst. 260 :3274–3277.

Mandelbrot, B. B. and Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, 10 :422–437.

Mandelbrot, B. B. and Wallis, J. R. (1969a). Computer experiments with fractional gaussian noises : Part 1, averages and variances. *Water Resources Research*, 5(1) :228–241.

Mandelbrot, B. B. and Wallis, J. R. (1969b). Computer experiments with fractional gaussian noises : Part 2, rescaled ranges and spectra. *Water Resources Research*, 5(1) :242–259.

Mandelbrot, B. B. and Wallis, J. R. (1969c). Computer experiments with fractional gaussian noises : Part 3, mathematical appendix. *Water Resources Research*, 5(1) :260–267.

Palma, W. (2007). *Long-memory time series*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ. Theory and methods.

Priestley, M. B. (1981). *Spectral analysis and time series*. Academic Press London ; New York.

Romano, J. P. and Thombs, L. A. (1996). Inference for autocorrelations under weak assumptions. *J. Amer. Statist. Assoc.*, 91(434) :590–600.

Shao, X. (2010a). Corrigendum : A self-normalized approach to confidence interval construction in time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(5) :695–696.

Shao, X. (2010b). Nonstationarity-extended Whittle estimation. *Econometric Theory*, 26(4) :1060–1087.

Shao, X. (2010c). A self-normalized approach to confidence interval construction in time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(3) :343–366.

Shao, X. (2011). Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models. *Econometric Theory*, 27(2) :312–343.

Shao, X. (2012). Parametric inference in stationary time series models with dependent errors. *Scand. J. Stat.*, 39(4) :772–783.

Shao, X. (2015). Self-normalization for time series : a review of recent developments. *J. Amer. Statist. Assoc.*, 110(512) :1797–1817.

Taniguchi, M. (1982). On estimation of the integrals of the fourth order cumulant spectral density. *Biometrika*, 69(1) :117–122.

Taqqu, M. S. and Teverovsky, V. (1997). Robustness of Whittle-type estimators for time series with long-range dependence. *Comm. Statist. Stochastic Models*, 13(4) :723–757. Heavy tails and highly volatile phenomena.

Tong, H. (1990). *Non-linear time series : a dynamical system approach*. Oxford University Press.

Whittle, P. (1953). Estimation and information in stationary time series. *Ark. Mat.*, 2 :423–434.

Wold, H. (1938). *A study in the analysis of stationary time series*. PhD thesis, Almqvist & Wiksell.

Zhu, K. and Li, W. K. (2015). A bootstrapped spectral test for adequacy in weak ARMA models. *J. Econometrics*, 187(1) :113–130.

Chapter 2

Estimating FARIMA models with uncorrelated but non-independent error terms

Contents

2.1 Introduction **55**
2.2 Least squares estimation **58**
2.2.1 Notations and assumptions 58
2.2.2 Asymptotic properties 60
2.3 Estimating the asymptotic variance matrix **61**
2.3.1 Estimation of the asymptotic matrix $I(\theta_0)$ 62
2.3.2 A self-normalized approach to confidence interval construction in weak FARIMA models 63
2.4 Numerical illustrations **66**
2.4.1 Simulation studies and empirical sizes for confidence intervals . . 66
2.4.2 Application to real data 68
2.5 Conclusion **68**
2.6 Proofs **69**
2.6.1 Preliminary results 69
2.6.2 Proof of Theorem 2.1 73
2.6.3 Proof of Theorem 2.2 74
2.6.4 Proof of the convergence of the variance matrix estimator 89
2.6.5 Invertibility of the normalization matrix $P_{p+q+1,n}$ 103
2.6.6 Proof of Theorem 2.4 104
2.6.7 Proof of Theorem 2.5 107
2.7 Figures and tables **109**
2.8 Details on the proof of Theorem 2.1 . . . . . **121**
Bibliographie **129**

Abstract: In this paper we derive the asymptotic properties of the least squares estimator (LSE) of fractionally integrated autoregressive moving-average (FARIMA) models under the assumption that the errors are uncorrelated but not necessarily independent nor martingale differences. We relax considerably the independence and even the martingale difference assumptions on the innovation process to extend the range of application of the FARIMA models. We propose a consistent estimator of the asymptotic covariance matrix of the LSE which may be very different from that obtained in the standard framework. A self-normalized approach to confidence interval construction for weak FARIMA model parameters is also presented. All our results are done under a mixing assumption on the noise. Finally, some simulation studies and an application to the daily returns of stock market indices are presented to corroborate our theoretical work.
AMS 2000 subject classifications: Primary 62M10; secondary 91B84.
Keywords and phrases: Nonlinear processes; FARIMA models; Least-squares estimator; Consistency; Asymptotic normality; Spectral density estimation; Self-normalization; Cumulants.

2.1 Introduction

Long memory processes takes a large part in the literature of time series (see for instance Granger and Joyeux [1980], Fox and Taqqu [1986], Dahlhaus [1989], Hosking [1981], Beran et al. [2013], Palma [2007], among others). They also play an important role in many scientific disciplines and applied fields such as hydrology, climatology, economics, finance, to name a few. To model the long memory phenomenon, a widely used model is the fractional autoregressive integrated moving average (FARIMA, for short) model. Consider a second order centered stationary process $X := (X_t)_{t\in\mathbb{Z}}$ satisfying a FARIMA(p, d_0, q) representation of the form

$$a(L)(1-L)^{d_0}X_t = b(L)\epsilon_t, \tag{2.1}$$

where $d_0 \in]0, 1/2[$ is the long memory parameter, L stands for the back-shift operator and $a(L) = 1 - \sum_{i=1}^{p} a_i L^i$ is the autoregressive (AR for short) operator and $b(L) = 1 - \sum_{i=1}^{q} b_i L^i$ is the moving average (MA for short) operator (by convention $a_0 = b_0 = 1$). The operators a and b represent the short memory part of the model. The linear innovation process $\epsilon := (\epsilon_t)_{t\in\mathbb{Z}}$ is assumed to be a stationary sequence satisfies

(A0): $\mathbb{E}[\epsilon_t] = 0$, $\mathrm{Var}(\epsilon_t) = \sigma_\epsilon^2$ and $\mathrm{Cov}(\epsilon_t, \epsilon_{t+h}) = 0$ for all $t \in \mathbb{Z}$ and all $h \neq 0$.

Under the above assumptions the process ϵ is called a weak white noise. Different sub-classes of FARIMA models can be distinguished depending on the noise assumptions. It is customary to say that X is a strong FARIMA(p, d_0, q) representation and we will do this henceforth if in (2.1) ϵ is a strong white noise, namely an independent and identically distributed (iid for short) sequence of random variables with mean 0 and common variance. A strong white noise is obviously a weak white noise because independence entails uncorrelatedness. Of course the converse is not true. Between weak and strong noises, one can say that ϵ is a semi-strong white noise if ϵ is

a stationary martingale difference, namely a sequence such that $\mathbb{E}(\epsilon_t|\epsilon_{t-1}, \epsilon_{t-2}, \dots) = 0$. An example of semi-strong white noise is the generalized autoregressive conditional heteroscedastic (GARCH) model (see Francq and Zakoïan [2010]). If ϵ is a semi-strong white noise in (2.1), X is called a semi-strong FARIMA(p, d_0, q). If no additional assumption is made on ϵ, that is if ϵ is only a weak white noise (not necessarily iid, nor a martingale difference), the representation (2.1) is called a weak FARIMA(p, d_0, q). It is clear from these definitions that the following inclusions hold:

$$\{\text{strong FARIMA}(p, d_0, q)\} \subset \{\text{semi-strong FARIMA}(p, d_0, q)\} \subset \{\text{weak FARIMA}(p, d_0, q)\}.$$

Nonlinear models are becoming more and more employed because numerous real time series exhibit nonlinear dynamics. For instance conditional heteroscedasticity can not be generated by FARIMA models with iid noises.[1] As mentioned by Francq and Zakoïan [1998, 2005] in the case of ARMA models, many important classes of nonlinear processes admit weak ARMA representations in which the linear innovation is not a martingale difference. The main issue with nonlinear models is that they are generally hard to identify and implement. These technical difficulties certainly explain the reason why the asymptotic theory of FARIMA model estimation is mainly limited to the strong or semi-strong FARIMA model.

Now we present some of the main works about FARIMA model estimation when the noise is strong or semi-strong. For the estimation of long-range dependent process, the commonly used estimation method is based on the Whittle frequency domain maximum likelihood estimator (MLE) (see for instance Dahlhaus [1989], Fox and Taqqu [1986], Taqqu and Teverovsky [1997], Giraitis and Surgailis [1990]). The asymptotic properties of the MLE of FARIMA models are well-known under the restrictive assumption that the errors ϵ_t are independent or martingale difference (see Beran [1995], Beran et al. [2013], Palma [2007], Baillie et al. [1996], Ling and Li [1997], Hauser and Kunst [1998], among others). All the works mentioned above assume either strong or semi-strong innovations. In the modeling of financial time series, for example, the GARCH assumption on the errors is often used (see for instance Baillie et al. [1996], Hauser and Kunst [1998]) to capture the conditional heteroscedasticity. There is no doubt that it is important to have a soundness inference procedure for the parameter in the FARIMA model when the (possibly dependent) error is subject to unknown conditional heteroscedasticity. Little is thus known when the martingale difference assumption is relaxed. Our aim in this paper is to consider a flexible FARIMA specification and to relax the independence assumption (and even the martingale difference assumption) in order to be able to cover weak FARIMA representations of general nonlinear models. This is why it is interesting to consider weak FARIMA models.

A very few works deal with the asymptotic behavior of the MLE of weak FARIMA models. To our knowledge, Shao [2010b, 2012] are the only papers on this subject. Under weak assumptions on the noise process, the author has obtained the asymptotic normality of the Whittle estimator (see Whittle [1953]). Nevertheless, the inference problem is not addressed. This is due to the

1. To cite few examples of nonlinear processes, let us mention the self-exciting threshold autoregressive (SETAR), the smooth transition autoregressive (STAR), the exponential autoregressive (EXPAR), the bilinear, the random coefficient autoregressive (RCA), the functional autoregressive (FAR) (see Tong [1990] and Fan and Yao [2008] for references on these nonlinear time series models).

fact that the asymptotic covariance matrix of the Whittle estimator involves the integral of the fourth-order cumulant spectra of the dependent errors ϵ_t. Using non-parametric bandwidth-dependent methods, one build an estimation of this integral but there is no guidance on the choice of the bandwidth in the estimation procedures (see Chiu [1988], Keenan [1987], Shao [2012], Taniguchi [1982] for further details). The difficulty is caused by the dependence in ϵ_t. Indeed, for strong noise, a bandwidth-free consistent estimator of the asymptotic covariance matrix is available. When ϵ_t is dependent, no explicit formula for a consistent estimator of the asymptotic variance matrix seems to be provided in the literature (see Shao [2012]).

In this work we propose to adopt for weak FARIMA models the estimation procedure developed in Francq and Zakoïan [1998] so we use the least squares estimator (LSE for short). We show that a strongly mixing property and the existence of moments are sufficient to obtain a consistent and asymptotically normally distributed least squares estimator for the parameters of a weak FARIMA representation. For technical reasons, we often use an assumption on the summability of cumulants. This can be a consequence of a mixing and moments assumptions (see Doukhan and León [1989], for more details). These kind of hypotheses enable us to circumvent the problem of the lack of speed of convergence (due to the long-range dependence) in the infinite AR or MA representations. We fix this gap by proposing rather sharp estimations of the infinite AR and MA representations in the presence of long-range dependence (see Subsection 2.6.1 for details).

In our opinion there are three major contributions in this work. The first one is to show that the estimation procedure developed in Francq and Zakoïan [1998] can be extended to weak FARIMA models. This goal is achieved thanks to Theorem 2.1 and Theorem 2.2 in which the consistency and the asymptotic normality are stated. The second one is to provide an answer to the open problem raised by Shao [2012] (see also Shao [2010b]) on the asymptotic covariance matrix estimation. We propose in our work a weakly consistent estimator of the asymptotic variance matrix (see Theorem 2.3). Thanks to this estimation of the asymptotic variance matrix, we can construct a confidence region for the estimation of the parameters. Finally another method to construct such confidence region is achieved thanks to an alternative method using a self normalization procedure (see Theorem 2.5).

The paper is organized as follows. Section 2.2 shown that the least squares estimator for the parameters of a weak FARIMA model is strongly consistent when the weak white noise (ϵ_t) is ergodic and stationary, and that the LSE is asymptotically normally distributed when (ϵ_t) satisfies mixing assumptions. The asymptotic variance of the LSE may be very different in the weak and strong cases. Section 2.3 is devoted to the estimation of this covariance matrix. We also propose a self-normalization-based approach to constructing a confidence region for the parameters of weak FARIMA models which avoids to estimate the asymptotic covariance matrix. We gather in Section 2.7 all our figures and tables. These simulation studies and illustrative applications on real data are presented and discussed in Section 2.4. The proofs of the main results are collected in Section 2.6.

In all this work, we shall use the matrix norm defined by $\|A\| = \sup_{\|x\|\leq 1} \|Ax\| = \rho^{1/2}(A'A)$, when A is a $\mathbb{R}^{k_1\times k_2}$ matrix, $\|x\|^2 = x'x$ is the Euclidean norm of the vector $x \in \mathbb{R}^{k_2}$, and $\rho(\cdot)$ denotes the spectral radius.

2.2 Least squares estimation

In this section we present the parametrization and the assumptions that are used in the sequel. Then we state the asymptotic properties of the LSE of weak FARIMA models.

2.2.1 Notations and assumptions

We make the following standard assumption on the roots of the AR and MA polynomials in (2.1).

(A1): The polynomials $a(z)$ and $b(z)$ have all their roots outside of the unit disk with no common factors.

Let Θ^* be the parameter space

$$\Theta^* := \Big\{ (\theta_1, \theta_2, ..., \theta_{p+q}) \in \mathbb{R}^{p+q}, \text{ where } a_\theta(z) = 1 - \sum_{i=1}^{p} \theta_i z^i, \text{ and } b_\theta(z) = 1 - \sum_{j=1}^{q} \theta_{p+j} z^j \text{ have all their zeros outside the unit disk and have no zero in common} \Big\}.$$

Denote by Θ the cartesian product $\Theta^* \times [d_1, d_2]$, where $[d_1, d_2] \subset]0, 1/2[$ with $d_1 \leq d_0 \leq d_2$. The unknown parameter of interest $\theta_0 = (a_1, a_2, \ldots, a_p, b_1, b_2, \ldots, b_q, d_0)$ is supposed to belong to the parameter space Θ.

The fractional difference operator $(1-L)^{d_0}$ is defined, using the generalized binomial series, by

$$(1-L)^{d_0} = \sum_{j \geq 0} \alpha_j(d_0) L^j,$$

where for all $j \geq 0$, $\alpha_j(d_0) = \Gamma(j - d_0) / \{\Gamma(j+1)\Gamma(-d_0)\}$ and $\Gamma(\cdot)$ is the Gamma function. Using the Stirling formula we obtain that for large j, $\alpha_j(d_0) \sim j^{-d_0 - 1} / \Gamma(-d_0)$ (one refers to Beran et al. [2013] for further details).

For all $\theta \in \Theta$ we define $(\epsilon_t(\theta))_{t \in \mathbb{Z}}$ as the second order stationary process which is the solution of

$$\epsilon_t(\theta) = \sum_{j \geq 0} \alpha_j(d) X_{t-j} - \sum_{i=1}^{p} \theta_i \sum_{j \geq 0} \alpha_j(d) X_{t-i-j} + \sum_{j=1}^{q} \theta_{p+j} \epsilon_{t-j}(\theta). \tag{2.2}$$

Observe that, for all $t \in \mathbb{Z}$, $\epsilon_t(\theta_0) = \epsilon_t$ a.s. Given a realization $X_1, \ldots, X_n$ of length n, $\epsilon_t(\theta)$ can be approximated, for $0 < t \leq n$, by $\tilde{\epsilon}_t(\theta)$ defined recursively by

$$\tilde{\epsilon}_t(\theta) = \sum_{j=0}^{t-1} \alpha_j(d) X_{t-j} - \sum_{i=1}^{p} \theta_i \sum_{j=0}^{t-i-1} \alpha_j(d) X_{t-i-j} + \sum_{j=1}^{q} \theta_{p+j} \tilde{\epsilon}_{t-j}(\theta), \tag{2.3}$$

with $\tilde{\epsilon}_t(\theta) = X_t = 0$ if $t \leq 0$. It will be shown that these initial values are asymptotically negligible uniformly in θ and, in particular, that $\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta) \to 0$ almost surely as $t \to \infty$ (see

Lemma 2.4 hereafter). Thus the choice of the initial values has no influence on the asymptotic properties of the model parameters estimator.

Let Θ_δ^* denote the compact set

$$\Theta_\delta^* = \left\{\theta \in \mathbb{R}^{p+q};\ \text{the roots of the polynomials } a_\theta(z) \text{ and } b_\theta(z) \text{ have modulus } \geq 1+\delta\right\}.$$

We define the set Θ_δ as the cartesian product of Θ_δ^* by $[d_1, d_2]$, i.e. $\Theta_\delta = \Theta_\delta^* \times [d_1, d_2]$, where δ is a strictly positive constant chosen such that θ_0 belongs to Θ_δ.

The random variable $\hat{\theta}_n$ is called least squares estimator if it satisfies, almost surely,

$$\hat{\theta}_n = \underset{\theta\in\Theta_\delta}{\operatorname{argmin}}\ Q_n(\theta),\ \text{where}\ Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\tilde{\epsilon}_t^2(\theta). \tag{2.4}$$

Our main results are proven under the following assumptions:

(A2): The process $(\epsilon_t)_{t\in\mathbb{Z}}$ is strictly stationary and ergodic.

The strong consistency of the least squares estimator will be proved under the three above assumptions (**(A0)**, **(A1)** and **(A2)**). For the asymptotic normality of the LSE, additional assumptions are required. It is necessary to assume that θ_0 is not on the boundary of the parameter space Θ.

(A3): We have $\theta_0 \in \overset{\circ}{\Theta}$, where $\overset{\circ}{\Theta}$ denotes the interior of Θ.

The stationary process ϵ is not supposed to be an independent sequence. So one needs to control its dependency by means of its strong mixing coefficients $\{\alpha_\epsilon(h)\}_{h\in\mathbb{N}}$ defined by

$$\alpha_\epsilon(h) = \sup_{A\in\mathcal{F}_{-\infty}^t, B\in\mathcal{F}_{t+h}^\infty} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|,$$

where $\mathcal{F}_{-\infty}^t = \sigma(\epsilon_u, u \leq t)$ and $\mathcal{F}_{t+h}^\infty = \sigma(\epsilon_u, u \geq t+h)$.

We shall need an integrability assumption on the moments of the noise ϵ and a summability condition on the strong mixing coefficients $(\alpha_\epsilon(k))_{k\geq 0}$.

(A4): There exists an integer τ such that for some $\nu \in]0,1]$, we have $\mathbb{E}|\epsilon_t|^{\tau+\nu} < \infty$ and $\sum_{h=0}^\infty (h+1)^{k-2}\{\alpha_\epsilon(h)\}^{\frac{\nu}{k+\nu}} < \infty$ for $k = 1,\ldots,\tau$.

Note that **(A4)** implies the following weak assumption on the joint cumulants of the innovation process ϵ (see Doukhan and León [1989], for more details).

(A4'): There exists an integer $\tau \geq 2$ such that $C_\tau := \sum_{i_1,\ldots,i_{\tau-1}\in\mathbb{Z}} |\mathrm{cum}(\epsilon_0, \epsilon_{i_1}, \ldots, \epsilon_{i_{\tau-1}})| < \infty$.

In the above expression, $\text{cum}(\epsilon_0, \epsilon_{i_1}, \ldots, \epsilon_{i_{\tau-1}})$ denotes the $\tau-$th order cumulant of the stationary process. Due to the fact that the ϵ_t's are centered, we notice that for fixed (i, j, k)

$$\text{cum}(\epsilon_0, \epsilon_i, \epsilon_j, \epsilon_k) = \mathbb{E}[\epsilon_0\epsilon_i\epsilon_j\epsilon_k] - \mathbb{E}[\epsilon_0\epsilon_i]\mathbb{E}[\epsilon_j\epsilon_k] - \mathbb{E}[\epsilon_0\epsilon_j]\mathbb{E}[\epsilon_i\epsilon_k] - \mathbb{E}[\epsilon_0\epsilon_k]\mathbb{E}[\epsilon_i\epsilon_j].$$

Assumption **(A4)** is a usual technical hypothesis which is useful when one prove the asymptotic normality (see Francq and Zakoïan [1998] for example). Let us notice however that we impose a stronger convergence speed for the mixing coefficients than in the works on weak ARMA processes. This is due to the fact that the coefficients in the AR or MA representation of $\epsilon_t(\theta)$ have no more exponential decay because of the fractional operator (see Subsection 2.6.1 for details and comments).

As mentioned before, Hypothesis **(A4)** implies **(A4')** which is also a technical assumption usually used in the fractionally integrated ARMA processes framework (see for instance Shao [2010c]) or even in an ARMA context (see Francq and Zakoïan [2007], Zhu and Li [2015]). One remarks that in Shao [2010b], the author emphasized that a geometric moment contraction implies **(A4')**. This provides an alternative to strong mixing assumptions but, to our knowledge, there is no relation between this two kinds of hypotheses.

2.2.2 Asymptotic properties

The asymptotic properties of the LSE of the weak FARIMA model are stated in the following two theorems.

Theorem 2.1. *(Consistency). Assume that $(\epsilon_t)_{t\in\mathbb{Z}}$ satisfies (2.1) and belonging to $\mathbb{L}^2$. Let $(\hat{\theta}_n)_n$ be a sequence of least squares estimators. Under Assumptions **(A0)**, **(A1)** and **(A2)**, we have*

$$\hat{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta_0.$$

The proof of this theorem is given in Subsection 2.6.2.

In order to state our asymptotic normality result, we define the function

$$O_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2(\theta), \tag{2.5}$$

where the sequence $(\epsilon_t(\theta))_{t\in\mathbb{Z}}$ is given by (2.2). We consider the following information matrices

$$I(\theta) = \lim_{n\to\infty} Var\left\{\sqrt{n}\frac{\partial}{\partial\theta}O_n(\theta)\right\} \text{ and } J(\theta) = \lim_{n\to\infty}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}O_n(\theta)\right] \text{ a.s.}$$

The existence of these matrices are proved when one demonstrates the following result.

Theorem 2.2. *(Asymptotic normality). We assume that $(\epsilon_t)_{t\in\mathbb{Z}}$ satisfies (2.1). Under **(A0)**-**(A3)** and Assumption **(A4)** with $\tau = 4$, the sequence $(\sqrt{n}(\hat{\theta}_n - \theta_0))_{n\geq 1}$ has a limiting centered normal distribution with covariance matrix $\Omega := J^{-1}(\theta_0)I(\theta_0)J^{-1}(\theta_0)$.*

The proof of this theorem is given in Subsection 2.6.3.

Remark 2.1. *Hereafter (see more precisely (2.52)), we will be able to prove that*

$$J(\theta_0) = 2\mathbb{E}\left[\frac{\partial}{\partial\theta}\epsilon_t(\theta_0)\frac{\partial}{\partial\theta'}\epsilon_t(\theta_0)\right] .$$

Thus the matrix $J(\theta_0)$ has the same expression in the strong and weak FARIMA cases (see Theorem 1 of Beran [1995]). On the contrary, the matrix $I(\theta_0)$ is in general much more complicated in the weak case than in the strong case.

Remark 2.2. *In the standard strong FARIMA case,* i.e. *when* **(A2)** *is replaced by the assumption that (ϵ_t) is iid, we have $I(\theta_0) = 2\sigma_\epsilon^2 J(\theta_0)$. Thus the asymptotic covariance matrix is then reduced as $\Omega_S := 2\sigma_\epsilon^2 J^{-1}(\theta_0)$. Generally, when the noise is not an independent sequence, this simplification can not be made and we have $I(\theta_0) \neq 2\sigma_\epsilon^2 J(\theta_0)$. The true asymptotic covariance matrix $\Omega = J^{-1}(\theta_0)I(\theta_0)J^{-1}(\theta_0)$ obtained in the weak FARIMA framework can be very different from Ω_S. As a consequence, for the statistical inference on the parameter, the ready-made softwares used to fit FARIMA do not provide a correct estimation of Ω for weak FARIMA processes because the standard time series analysis softwares use empirical estimators of Ω_S. The problem also holds in the weak ARMA case (see Francq and Zakoïan [2007] and the references therein).This is why it is interesting to find an estimator of Ω which is consistent for both weak and (semi-)strong FARIMA cases.*

Based on the above remark, the next subsection deals with two different methods in order find an estimator of Ω.

2.3 Estimating the asymptotic variance matrix

For statistical inference problem, the asymptotic variance Ω has to be estimated. In particular Theorem 2.2 can be used to obtain confidence intervals and significance tests for the parameters.

First of all, the matrix $J(\theta_0)$ can be estimated empirically by the square matrix $\hat{J}_n$ of order $p+q+1$ defined by:

$$\hat{J}_n = \frac{2}{n}\sum_{t=1}^{n}\left\{\frac{\partial}{\partial\theta}\tilde{\epsilon}_t\left(\hat{\theta}_n\right)\right\}\left\{\frac{\partial}{\partial\theta'}\tilde{\epsilon}_t\left(\hat{\theta}_n\right)\right\}. \tag{2.6}$$

The convergence of $\hat{J}_n$ to $J(\theta_0)$ is classical (see Lemma 2.8 in Subsection 2.6.3 for details).

In the standard strong FARIMA case, in view of Remark 2.2, we have $\hat{\Omega}_S := 2\hat{\sigma}_\epsilon^2\hat{J}_n^{-1}$ with $\hat{\sigma}_\epsilon^2 = Q_n(\hat{\theta}_n)$. Thus $\hat{\Omega}_S$ is a strongly consistent estimator of Ω_S. In the general weak FARIMA case, this estimator is not consistent when $I(\theta_0) \neq 2\sigma_\epsilon^2 J(\theta_0)$. So we need a consistent estimator of $I(\theta_0)$.

2.3.1 Estimation of the asymptotic matrix $I(\theta_0)$

For all $t \in \mathbb{Z}$, let

$$H_t(\theta_0) = 2\epsilon_t(\theta_0)\frac{\partial}{\partial\theta}\epsilon_t(\theta_0) = \left(2\epsilon_t(\theta_0)\frac{\partial}{\partial\theta_1}\epsilon_t(\theta_0), \ldots, 2\epsilon_t(\theta_0)\frac{\partial}{\partial\theta_{p+q+1}}\epsilon_t(\theta_0)\right)' . \tag{2.7}$$

We shall see in the proof of Lemma 2.9 that

$$I(\theta_0) = \lim_{n\to\infty} \mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} H_t(\theta_0)\right) = \sum_{h=-\infty}^{\infty} \mathrm{Cov}\left(H_t(\theta_0), H_{t-h}(\theta_0)\right).$$

Following the arguments developed in Boubacar Mainassara et al. [2012], the matrix $I(\theta_0)$ can be estimated using Berk's approach (see Berk [1974]). More precisely, by interpreting $I(\theta_0)/2\pi$ as the spectral density of the stationary process $(H_t(\theta_0))_{t\in\mathbb{Z}}$ evaluated at frequency 0, we can use a parametric autoregressive estimate of the spectral density of $(H_t(\theta_0))_{t\in\mathbb{Z}}$ in order to estimate the matrix $I(\theta_0)$.

For any $\theta \in \Theta$, $H_t(\theta)$ is a measurable function of $\{\epsilon_s, s \leq t\}$. The stationary process $(H_t(\theta_0))_{t\in\mathbb{Z}}$ admits the following Wold decomposition $H_t(\theta_0) = u_t + \sum_{k=1}^{\infty}\psi_k u_{t-k}$, where $(u_t)_{t\in\mathbb{Z}}$ is a $(p+q+1)-$variate weak white noise with variance matrix Σ_u.

Assume that Σ_u is non-singular, that $\sum_{k=1}^{\infty}\|\psi_k\| < \infty$, and that $\det(I_{p+q+1}+\sum_{k=1}^{\infty}\psi_k z^k) \neq 0$ when $|z| \leq 1$. Then $(H_t(\theta_0))_{t\in\mathbb{Z}}$ admits a weak multivariate AR(∞) representation of the form

$$\Phi(L)H_t(\theta_0) := H_t(\theta_0) - \sum_{k=1}^{\infty}\Phi_k H_{t-k}(\theta_0) = u_t, \tag{2.8}$$

such that $\sum_{k=1}^{\infty}\|\Phi_k\| < \infty$ and $\det\{\Phi(z)\} \neq 0$ for all $|z| \leq 1$. It is proved in Boubacar Maïnassara [2009], Lütkepohl [2005] that one may find a constant K and $0 < \rho < 1$ such that

$$\|\Phi_k\| \leq K\,\rho^k. \tag{2.9}$$

Thanks to the previous remarks, the estimation of $I(\theta_0)$ is therefore based on the following expression

$$I(\theta_0) = \Phi^{-1}(1)\Sigma_u\Phi'^{-1}(1).$$

Consider the regression of $H_t(\theta_0)$ on $H_{t-1}(\theta_0), \ldots, H_{t-r}(\theta_0)$ defined by

$$H_t(\theta_0) = \sum_{k=1}^{r}\Phi_{r,k}H_{t-k}(\theta_0) + u_{r,t}, \tag{2.10}$$

where $u_{r,t}$ is uncorrelated with $H_{t-1}(\theta_0), \ldots, H_{t-r}(\theta_0)$. Since $H_t(\theta_0)$ is not observable, we introduce $\hat{H}_t \in \mathbb{R}^{p+q+1}$ obtained by replacing $\epsilon_t(\theta_0)$ by $\tilde{\epsilon}_t(\theta_0)$ and θ_0 by $\hat{\theta}_n$ in (2.7):

$$\hat{H}_t = 2\tilde{\epsilon}_t(\hat{\theta}_n)\frac{\partial}{\partial\theta}\tilde{\epsilon}_t(\hat{\theta}_n) . \tag{2.11}$$

Let $\hat{\Phi}_r(z) = I_{p+q+1} - \sum_{k=1}^{r} \hat{\Phi}_{r,k} z^k$, where $\hat{\Phi}_{r,1}, \ldots, \hat{\Phi}_{r,r}$ denote the coefficients of the LS regression of $\hat{H}_t$ on $\hat{H}_{t-1}, \ldots, \hat{H}_{t-r}$. Let $\hat{u}_{r,t}$ be the residuals of this regression and let $\hat{\Sigma}_{\hat{u}_r}$ be the empirical variance (defined in (2.12) below) of $\hat{u}_{r,1}, \ldots, \hat{u}_{r,r}$. The LSE of $\underline{\Phi}_r = (\Phi_{r,1}, \ldots, \Phi_{r,r})$ and $\Sigma_{u_r} = \mathrm{Var}(u_{r,t})$ are given by

$$\underline{\hat{\Phi}}_r = \hat{\Sigma}_{\hat{H},\underline{\hat{H}}_r} \hat{\Sigma}_{\underline{\hat{H}}_r}^{-1} \quad \text{and} \quad \hat{\Sigma}_{\hat{u}_r} = \frac{1}{n} \sum_{t=1}^{n} \left(\hat{H}_t - \underline{\hat{\Phi}}_r \underline{\hat{H}}_{r,t} \right) \left(\hat{H}_t - \underline{\hat{\Phi}}_r \underline{\hat{H}}_{r,t} \right)', \tag{2.12}$$

where

$$\underline{\hat{H}}_{r,t} = (\hat{H}'_{t-1}, \ldots, \hat{H}'_{t-r})', \quad \hat{\Sigma}_{\hat{H},\underline{\hat{H}}_r} = \frac{1}{n} \sum_{t=1}^{n} \hat{H}_t \underline{\hat{H}}'_{r,t} \text{ and } \hat{\Sigma}_{\underline{\hat{H}}_r} = \frac{1}{n} \sum_{t=1}^{n} \underline{\hat{H}}_{r,t} \underline{\hat{H}}'_{r,t},$$

with by convention $\hat{H}_t = 0$ when $t \leq 0$. We assume that $\hat{\Sigma}_{\underline{\hat{H}}_r}$ is non-singular (which holds true asymptotically).

In the case of linear processes with independent innovations, Berk (see Berk [1974]) has shown that the spectral density can be consistently estimated by fitting autoregressive models of order $r = r(n)$, whenever r tends to infinity and r^3/n tends to 0 as n tends to infinity. There are differences with Berk [1974]: $(H_t(\theta_0))_{t\in\mathbb{Z}}$ is multivariate, is not directly observed and is replaced by $(\hat{H}_t)_{t\in\mathbb{Z}}$. It is shown that this result remains valid for the multivariate linear process $(H_t(\theta_0))_{t\in\mathbb{Z}}$ with non-independent innovations (see Boubacar Mainassara et al. [2012], Boubacar Mainassara and Francq [2011], for references in weak (multivariate) ARMA models). We will extend the results of Boubacar Mainassara et al. [2012] to weak FARIMA models.

The asymptotic study of the estimator of $I(\theta_0)$ using the spectral density method is given in the following theorem.

Theorem 2.3. *We assume* ***(A0)-(A3)*** *and Assumption* **(A4')** *with* $\tau = 8$. *In addition, we assume that the process* $(H_t(\theta_0))_{t\in\mathbb{Z}}$ *defined in* (2.7) *admits a multivariate* $AR(\infty)$ *representation* (2.8). *Then, the spectral estimator of* $I(\theta_0)$

$$\hat{I}_n^{\mathrm{SP}} := \hat{\Phi}_r^{-1}(1) \hat{\Sigma}_{\hat{u}_r} \hat{\Phi}_r'^{-1}(1) \xrightarrow[n\to\infty]{\mathbb{P}} I(\theta_0) = \Phi^{-1}(1) \Sigma_u \Phi'^{-1}(1)$$

where r *depends on* n *and satisfies* $\lim_{n\to\infty} r^5(n)/n^{1-2(d_2-d_1)} = 0$ *(remind that* $d_0 \in [d_1,d_2] \subset]0,1/2[$*).*

The proof of this theorem is given in Subsection 2.6.4.

A second method to estimate the asymptotic matrix (or rather avoiding estimate it) is proposed in the next subsection.

2.3.2 A self-normalized approach to confidence interval construction in weak FARIMA models

We have seen previously that we may obtain confidence intervals for weak FARIMA model parameters as soon as we can construct a convergent estimator of the variance matrix $I(\theta_0)$ (see

Theorems 2.2 and 2.3). The parametric approach based on an autoregressive estimate of the spectral density of $(H_t(\theta_0))_{t\in\mathbb{Z}}$ that we used before has the drawback of choosing the truncation parameter r in (2.10). This choice of the order truncation is often crucial and difficult. So the aim of this section is to avoid such a difficulty.

This section is also of interest because, to our knowledge, it has not been studied for weak FARIMA models. Notable exception is Shao [2012] who studied this problem in a short memory case (see Assumption 1 in Shao [2012] that implies that the process X is short-range dependent).

We propose an alternative method to obtain confidence intervals for weak FARIMA models by avoiding the estimation of the asymptotic covariance matrix $I(\theta_0)$. It is based on a self-normalization approach used to build a statistic which depends on the true parameter θ_0 and which is asymptotically distribution-free (see Theorem 1 of Shao [2012] for a reference in weak ARMA case). The idea comes from Lobato [2001] and has been already extended by Boubacar Maïnassara and Saussereau [2018], Kuan and Lee [2006], Shao [2010a,c, 2012] to more general frameworks. See also Shao [2015] for a review on some recent developments on the inference of time series data using the self-normalized approach.

Let us briefly explain the idea of the self-normalization.

By a Taylor expansion of the function $\partial Q_n(\cdot)/\partial\theta$ around θ_0, under **(A3)**, we have

$$0=\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\hat{\theta}_n)=\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0)+\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}Q_n\left(\theta^*_{n,i,j}\right)\right]\sqrt{n}\left(\hat{\theta}_n-\theta_0\right), \tag{2.13}$$

where the $\theta^*_{n,i,j}$'s are between $\hat{\theta}_n$ and θ_0. Using the following equation

$$\begin{aligned}\sqrt{n}\left(\frac{\partial}{\partial\theta}O_n(\theta_0)-\frac{\partial}{\partial\theta}Q_n(\theta_0)\right)&=\sqrt{n}\frac{\partial}{\partial\theta}O_n(\theta_0)\\&\quad+\left\{\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}Q_n(\theta^*_{n,i,j})\right]-J(\theta_0)+J(\theta_0)\right\}\sqrt{n}(\hat{\theta}_n-\theta_0),\end{aligned}$$

we shall be able to prove that (2.13) implies that

$$\sqrt{n}\frac{\partial}{\partial\theta}O_n(\theta_0)+J(\theta_0)\sqrt{n}(\hat{\theta}_n-\theta_0)=o_{\mathbb{P}}(1). \tag{2.14}$$

This is due to the following technical properties:

— the convergence in probability of $\sqrt{n}\partial Q_n(\theta_0)/\partial\theta$ to $\sqrt{n}\partial O_n(\theta_0)/\partial\theta$ (see Lemma 2.5 hereafter),

— the almost-sure convergence of $[\partial^2 Q_n(\theta^*_{n,i,j})/\partial\theta_i\partial\theta_j]$ to $J(\theta_0)$ (see Lemma 2.8 hereafter),

— the tightness of the sequence $(\sqrt{n}(\hat{\theta}_n-\theta_0))_n$ (see Theorem 2.2) and

— the existence and invertibility of the matrix $J(\theta_0)$ (see Lemma 2.6 hereafter).

Thus we obtain from (2.14) that

$$\sqrt{n}(\hat{\theta}_n-\theta_0)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}U_t+o_{\mathbb{P}}(1),$$

where (remind (2.7))

$$U_t = -J^{-1}(\theta_0)H_t(\theta_0).$$

At this stage, we do not rely on the classical method that would consist in estimating the asymptotic covariance matrix $I(\theta_0)$. We rather try to apply Lemma 1 in Lobato [2001]. So we need to check that a functional central limit theorem holds for the process $U := (U_t)_{t\geq 1}$. For that sake, we define the normalization matrix $P_{p+q+1,n}$ of $\mathbb{R}^{(p+q+1)\times(p+q+1)}$ by

$$P_{p+q+1,n} = \frac{1}{n^2}\sum_{t=1}^{n}\left(\sum_{j=1}^{t}(U_j - \bar{U}_n)\right)\left(\sum_{j=1}^{t}(U_j - \bar{U}_n)\right)', \tag{2.15}$$

where $\bar{U}_n = (1/n)\sum_{i=1}^{n} U_i$. To ensure the invertibility of the normalization matrix $P_{p+q+1,n}$ (it is the result stated in the next proposition), we need the following technical assumption on the distribution of ϵ_t.

(A5): The process $(\epsilon_t)_{t\in\mathbb{Z}}$ has a positive density on some neighborhood of zero.

Proposition 2.1. *Under the assumptions of Theorem 2.2 and* **(A5)**, *the matrix $P_{p+q+1,n}$ is almost surely non singular.*

The proof of this proposition is given in Subsection 2.6.5.

Let $(B_m(r))_{r\geq 0}$ be a m-dimensional Brownian motion starting from 0. For $m \geq 1$, we denote by $\mathcal{U}_m$ the random variable defined by:

$$\mathcal{U}_m = B_m'(1)V_m^{-1}B_m(1), \tag{2.16}$$

where

$$V_m = \int_0^1 (B_m(r) - rB_m(1))\,(B_m(r) - rB_m(1))'\,dr. \tag{2.17}$$

The critical values of $\mathcal{U}_m$ have been tabulated by Lobato [2001].

The following theorem states the self-normalized asymptotic distribution of the random vector $\sqrt{n}(\hat{\theta}_n - \theta_0)$.

Theorem 2.4. *Under the assumptions of Theorem 2.2 and* ***(A5)***, *we have*

$$n(\hat{\theta}_n - \theta_0)'P_{p+q+1,n}^{-1}(\hat{\theta}_n - \theta_0) \xrightarrow[n\to\infty]{\text{in law}} \mathcal{U}_{p+q+1}.$$

The proof of this theorem is given in Subsection 2.6.6.

Of course, the above theorem is useless for practical purpose because the normalization matrix $P_{p+q+1,n}$ is not observable. This gap will be fixed below when one replaces the matrix $P_{p+q+1,n}$ by its empirical or observable counterpart

$$\hat{P}_{p+q+1,n} = \frac{1}{n^2}\sum_{t=1}^{n}\left(\sum_{j=1}^{t}(\hat{U}_j - \tfrac{1}{n}\textstyle\sum_{k=1}^{n}\hat{U}_k)\right)\left(\sum_{j=1}^{t}(\hat{U}_j - \tfrac{1}{n}\textstyle\sum_{k=1}^{n}\hat{U}_k)\right)' \text{ where } \hat{U}_j = -\hat{J}_n^{-1}\hat{H}_j. \tag{2.18}$$

The above quantity is observable and we are able to state our Theorem which is the applicable version of Theorem 2.4.

Theorem 2.5. *Under the assumptions of Theorem 2.2 and **(A5)**, we have*

$$n(\hat{\theta}_n - \theta_0)' \hat{P}^{-1}_{p+q+1,n}(\hat{\theta}_n - \theta_0) \xrightarrow[n\to\infty]{\text{in law}} \mathcal{U}_{p+q+1}.$$

The proof of this theorem is given in Subsection 2.6.7.

At the asymptotic level α, a joint $100(1-\alpha)\%$ confidence region for the elements of θ_0 is then given by the set of values of the vector θ which satisfy the following inequality:

$$n(\hat{\theta}_n - \theta)' \hat{P}^{-1}_{p+q+1,n}(\hat{\theta}_n - \theta) \leq \mathcal{U}_{p+q+1,\alpha},$$

where $\mathcal{U}_{p+q+1,\alpha}$ is the quantile of order $1-\alpha$ for the distribution of $\mathcal{U}_{p+q+1}$.

Corollary 2.1. *For any $1 \leq i \leq p+q+1$, a $100(1-\alpha)\%$ confidence region for $\theta_0(i)$ is given by the following set:*

$$\left\{ x \in \mathbb{R} \; ; \; n\big(\hat{\theta}_n(i) - x\big)^2 \hat{P}^{-1}_{p+q+1,n}(i,i) \leq \mathcal{U}_{1,\alpha} \right\},$$

where $\mathcal{U}_{1,\alpha}$ denotes the quantile of order $1-\alpha$ of the distribution for $\mathcal{U}_1$.

The proof of this corollary is similar to that of Theorem 2.5 when one restricts ourselves to a one dimensional case.

2.4 Numerical illustrations

In this section, we investigate the finite sample properties of the asymptotic results that we introduced in this work. For that sake we use Monte Carlo experiments. The numerical illustrations of this section are made with the open source statistical software R (see R Development Core Team, 2017) or (see http://cran.r-project.org/).

2.4.1 Simulation studies and empirical sizes for confidence intervals

We study numerically the behavior of the LSE for FARIMA models of the form

$$(1-L)^d (X_t - aX_{t-1}) = \epsilon_t - b\epsilon_{t-1}, \tag{2.19}$$

where the unknown parameter is taken as $\theta_0 = (a, b, d) = (-0.7, -0.2, 0.4)$. First we assume that in (2.19) the innovation process $(\epsilon_t)_{t\in\mathbb{Z}}$ is an iid centered Gaussian process with common variance 1 which corresponds to the strong FARIMA case. In two other experiments we consider that in (2.19) the innovation processes $(\epsilon_t)_{t\in\mathbb{Z}}$ are defined respectively by

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = 0.04 + 0.12\epsilon_{t-1}^2 + 0.85\sigma_{t-1}^2 \end{cases} \tag{2.20}$$

and

$$\epsilon_t = \eta_t^2 \eta_{t-1}, \tag{2.21}$$

where $(\eta_t)_{t\geq 1}$ is a sequence of iid centered Gaussian random variables with variance 1. Note that the innovation process in (2.21) is not a martingale difference whereas it is the case of the noise defined in (2.20).

We simulated $N = 1,000$ independent trajectories of size $n = 2,000$ of Model (2.19) in the three following case: the strong Gaussian noise, the semi-strong noise (2.20) and the weak noise (2.21).

Figure 2.1, Figure 2.2 and Figure 2.3 compare the distribution of the LSE in these three contexts. The distributions of $\hat{d}_n$ are similar in the three cases whereas the LSE $\hat{a}_n$ of a is more accurate in the weak case than in the strong and semi-strong cases. The distributions of $\hat{b}_n$ are more accurate in the strong case than in the weak case. Remark that in the weak case the distributions of $\hat{b}_n$ are more accurate to the semi-strong ones.

Figure 2.4 compares standard estimator $\hat{\Omega}_S = 2\hat{\sigma}_\epsilon^2 \hat{J}_n^{-1}$ and the sandwich estimator $\hat{\Omega} = \hat{J}_n^{-1}\hat{I}_n^{SP}\hat{J}_n^{-1}$ of the LSE asymptotic variance Ω. We used the spectral estimator $\hat{I}_n^{SP}$ defined in Theorem 2.3. The multivariate AR order r (see (2.10)) is automatically selected by AIC (we use the function `VARselect()` of the **vars** R package). In the strong FARIMA case we know that the two estimators are consistent. In view of the two upper subfigures of Figure 2.4, it seems that the sandwich estimator is less accurate in the strong case. This is not surprising because the sandwich estimator is more robust, in the sense that this estimator remains consistent in the semi-strong and weak FARIMA cases, contrary to the standard estimator (see the middle and bottom subfigures of Figure 2.4). Figure 2.5 (resp. Figure 2.6) presents a zoom of the left(right)-middle and left(right)-bottom panels of Figure 2.4. It is clear that in the semi-strong or weak case $n(\hat{a}_n - a)^2$, $n(\hat{b}_n - b)^2$ and $n(\hat{d}_n - d)^2$ are, respectively, better estimated by $\hat{J}_n^{-1}\hat{I}_n^{SP}\hat{J}_n^{-1}(1,1)$, $\hat{J}_n^{-1}\hat{I}_n^{SP}\hat{J}_n^{-1}(2,2)$ and $\hat{J}_n^{-1}\hat{I}_n^{SP}\hat{J}_n^{-1}(3,3)$ (see Figure 2.6) than by $2\hat{\sigma}_\epsilon^2\hat{J}_n^{-1}(1,1)$, $2\hat{\sigma}_\epsilon^2\hat{J}_n^{-1}(2,2)$ and $2\hat{\sigma}_\epsilon^2\hat{J}_n^{-1}(3,3)$ (see Figure 2.5). The failure of the standard estimator of Ω in the weak FARIMA framework may have important consequences in terms of identification or hypothesis testing and validation.

Now we are interested in standard confidence interval and the modified versions proposed in Subsections 2.3.1 and 2.3.2. Table 2.1 displays the empirical sizes in the three previous different FARIMA cases. For the nominal level $\alpha = 5\%$, the empirical size over the $N = 1,000$ independent replications should vary between the significant limits 3.6% and 6.4% with probability 95%. For the nominal level $\alpha = 1\%$, the significant limits are 0.3% and 1.7%, and for the nominal level $\alpha = 10\%$, they are 8.1% and 11.9%. When the relative rejection frequencies are outside the significant limits, they are displayed in bold type in Table 2.1. For the strong FARIMA model, all the relative rejection frequencies are inside the significant limits for n large. For the semi-strong FARIMA model, the relative rejection frequencies of the standard confidence interval are definitely outside the significant limits, contrary to the modified versions proposed. For the weak FARIMA model, only the standard confidence interval of $\hat{b}_n$ is outside the significant limits when n increases. As a conclusion, Table 2.1 confirms the comments done concerning Figure 2.4.

2.4.2 Application to real data

We now consider an application to the daily returns of four stock market indices (CAC, DAX, Nikkei and S&P 500). The returns are defined by $r_t = \log(p_t/p_{t-1})$ where p_t denotes the price index of the stock market indices at time t. The observations cover the period from the starting date of each index to February 14, 2019. Figure 2.7 (resp. Figure 2.8) plots the closing prices (resp. the returns) of the four stock market indices. Figure 2.9 shows that the squared returns $(r_t^2)_{t\geq 1}$ are generally strongly autocorrelated.

In Financial Econometrics the returns are often assumed to be martingale increments. The squares of the returns have often second-order moments close to those of an ARMA$(1,1)$ which is compatible with a GARCH$(1,1)$ model for the returns (see Francq and Zakoïan [2010]). A long-range memory property of the stock market returns series was also largely investigated by Ding et al. [1993] (see also Beran et al. [2013], Palma [2007] and Baillie et al. [1996]). The squared returns $(r_t^2)_{t\geq 1}$ have significant positive autocorrelations at least up to lag 100 (see Figure 2.9) which confirm the claim that stock market returns have long-term memory (see Ding et al. [1993]). In particular the returns $(r_t)_{t\geq 1}$ process is characterized by substantially more correlation between absolute or squared returns than between the returns themselves.

Now we focus on the dynamics of the squared returns and we fit a FARIMA$(1,d,1)$ model to the squares of the 4 daily returns. Denoting by $(X_t)_{t\geq 1}$ the mean corrected series of the squared returns, we adjust the following model

$$(1-L)^d (X_t - aX_{t-1}) = \epsilon_t - b\epsilon_{t-1}.$$

Table 2.2 displays the LSE of the parameter $\theta = (a,b,d)$ of each squared of daily returns. The $p-$values of the corresponding LSE, $\hat{\theta}_n = (\hat{a}_n, \hat{b}_n, \hat{d}_n)$ are given in parentheses. The last column presents the estimated residual variance. Note that for all series, the estimated coefficients $|\hat{a}_n|$ and $|\hat{b}_n|$ are smaller than one and this is in accordance with our Assumption **(A1)**. We also observe that for all series the estimated long-range dependence coefficients $\hat{d}_n$ are significant for any reasonable asymptotic level and are inside $]0, 0.5[$. We thus think that the assumption **(A3)** is satisfied and thus our asymptotic normality theorem can be applied. Table 2.3 then presents for each serie the modified confidence interval at the asymptotic level $\alpha = 5\%$ for the parameters estimated in Table 2.2.

2.5 Conclusion

Taking into account the possible lack of independence of the error terms, we show in this paper that we can fit FARIMA representations of a wide class of nonlinear long memory times series. This is possible thanks to our theoretical results and it is illustrated in our real cases and simulations studies.

This standard methodology (when the noise is supposed to be iid), in particular the significance tests on the parameters, needs however to be adapted to take into account the possible lack of independence of the errors terms. A first step has been done thanks to our results on the confidence intervals. In future works, we intent to study how the existing identification

(see Boubacar Maïnassara [2012], Boubacar Maïnassara and Kokonendji [2016]) and diagnostic checking (see Boubacar Maïnassara and Saussereau [2018], Francq et al. [2005]) procedures should be adapted in the presence of long-range dependence framework and dependent noise.

2.6 Proofs

In all our proofs, K is a strictly positive constant that may vary from line to line.

2.6.1 Preliminary results

In this subsection, we shall give some results on estimations of the coefficient of formal power series that will arise in our study. Some of them are well know on some others are new to our knowledge. We will make some precise comments hereafter.

We begin by recalling the following properties on power series. If for $|z| \leq R$, the power series $f(z) = \sum_{i\geq 0} a_i z^i$ and $g(z) = \sum_{i\geq 0} b_i z^i$ are well defined, then one has $(fg)(z) = \sum_{i\geq 0} c_i z^i$ is also well defined for $|z| \leq R$ with the sequence $(c_i)_{i\geq 0}$ which is given by $c = a * b$ where $*$ denotes the convolution product between a and b defined by $c_i = \sum_{k=0}^{i} a_k b_{i-k} = \sum_{k=0}^{i} a_{i-k} b_i$. We will make use of the Young inequality that states that if the sequence $a \in \ell^p$ and $b \in \ell^q$ and such that $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$ with $1 \leq p, q, r \leq \infty$, then

$$\|a * b\|_{\ell^r} \leq \|a\|_{\ell^p} \times \|b\|_{\ell^q}.$$

Now we come back to the power series that arise in our context. Remind that for the true value of the parameter,

$$a_{\theta_0}(L)(1-L)^{d_0} X_t = b_{\theta_0}(L)\epsilon_t. \tag{2.22}$$

Thanks to the assumptions on the moving average polynomials b_θ and the autoregressive polynomials a_θ, the power series a_θ^{-1} and b_θ^{-1} are well defined.

Thus the functions $\epsilon_t(\theta)$ defined in (2.2) can be written as

$$\epsilon_t(\theta) = b_\theta^{-1}(L)a_\theta(L)(1-L)^d X_t \tag{2.23}$$

$$= b_\theta^{-1}(L)a_\theta(L)(1-L)^{d-d_0} a_{\theta_0}^{-1}(L)b_{\theta_0}(L)\epsilon_t \tag{2.24}$$

and if we denote $\gamma(\theta) = (\gamma_i(\theta))_{i\geq 0}$ the sequence of coefficients of the power series $b_\theta^{-1}(z)a_\theta(z)(1-z)^d$ (which is absolutely convergent for at least for $|z| \leq 1$), we may write for all $t \in \mathbb{Z}$:

$$\epsilon_t(\theta) = \sum_{i\geq 0} \gamma_i(\theta) X_{t-i}, \tag{2.25}$$

In the same way, by (2.23) one has

$$X_t = (1-L)^{-d} a_\theta^{-1}(L)b_\theta(L)\epsilon_t(\theta)$$

and if we denote $\eta(\theta) = (\eta_i(\theta))_{i\geq 0}$ the coefficients of the power series $(1-z)^{-d}a_\theta^{-1}(z)b_\theta(z)$ one has

$$X_t = \sum_{i\geq 0} \eta_i(\theta)\epsilon_{t-i}(\theta) \ . \tag{2.26}$$

We strength the fact that $\gamma_0(\theta) = \eta_0(\theta) = 1$ for all θ.

For large j, Hallin et al. [1999] have shown that uniformly in θ the sequences $\gamma(\theta)$ and $\eta(\theta)$ satisfy

$$\frac{\partial^k \gamma_j(\theta)}{\partial\theta_{i_1}\cdots\partial\theta_{i_k}} = \mathrm{O}\left(j^{-1-d}\left\{\log(j)\right\}^k\right), \text{ for } k = 0,1,2,3, \tag{2.27}$$

and

$$\frac{\partial^k \eta_j(\theta)}{\partial\theta_{i_1}\cdots\partial\theta_{i_k}} = \mathrm{O}\left(j^{-1+d}\left\{\log(j)\right\}^k\right), \text{ for } k = 0,1,2,3. \tag{2.28}$$

Note that, in view of (2.25), (2.26) and (2.27), for all $\theta \in \Theta_\delta$, $\epsilon_t(\theta)$ belongs to $\mathbb{L}^2$, that $(\epsilon_t(\theta))_{t\in\mathbb{Z}}$ is an ergodic sequence and that, for all $t \in \mathbb{Z}$, the function $\epsilon_t(\cdot)$ is a continuous function.

One difficulty that has to be addressed is that (2.25) includes the infinite past $(X_{t-i})_{i\geq 0}$ whereas only a finite number of observations $(X_t)_{1\leq t\leq n}$ are available to compute the estimators defined in (2.4). The simplest solution is truncation which amounts to setting all unobserved values equal to zero. Thus, for all $\theta \in \Theta$ and $1 \leq t \leq n$ one defines

$$\tilde{\epsilon}_t(\theta) = \sum_{i=0}^{t-1} \gamma_i(\theta)X_{t-i} = \sum_{i\geq 0} \gamma_i^t(\theta)X_{t-i} \tag{2.29}$$

where the truncated sequence $\gamma^t(\theta) = (\gamma_i^t(\theta))_{i\geq 0}$ is defined by

$$\gamma_i^t(\theta) = \begin{cases} \gamma_i(\theta) & \text{if } \ 0 \leq i \leq t-1 \ , \\ 0 & \text{otherwise.} \end{cases}$$

Since our assumptions are made on the noise in (2.1), it will be useful to express the random variables $\epsilon_t(\theta)$ and its partial derivatives with respect to θ, as a function of $(\epsilon_{t-i})_{i\geq 0}$.

From (2.24), there exists a sequence $\lambda(\theta) = (\lambda_i(\theta))_{i\geq 0}$ such that

$$\epsilon_t(\theta) = \sum_{i=0}^{\infty} \lambda_i(\theta)\,\epsilon_{t-i} \tag{2.30}$$

where the sequence $\lambda(\theta)$ is given by the sequence of the coefficients of the power series $b_\theta^{-1}(z)a_\theta(z)(1-z)^{d-d_0}a_{\theta_0}^{-1}(z)b_{\theta_0}(z)$. Consequently $\lambda(\theta) = \gamma(\theta) * \eta(\theta_0)$ or, equivalently,

$$\lambda_i(\theta) = \sum_{j=0}^{i} \gamma_j(\theta)\eta_{i-j}(\theta_0). \tag{2.31}$$

We proceed in the same way as regard to the derivatives of $\epsilon_t(\theta)$. More precisely, for any $\theta \in \Theta$, $t \in \mathbb{Z}$ and $1 \leq k, l \leq p+q+1$ there exists sequences $\dot{\lambda}_k(\theta) = (\dot{\lambda}_{i,k}(\theta))_{i\geq 1}$ and $\ddot{\lambda}_{k,l}(\theta) = (\ddot{\lambda}_{i,k,l}(\theta))_{i\geq 1}$ such that

$$\frac{\partial \epsilon_t(\theta)}{\partial \theta_k} = \sum_{i=1}^{\infty} \dot{\lambda}_{i,k}(\theta)\, \epsilon_{t-i} \tag{2.32}$$

$$\frac{\partial^2 \epsilon_t(\theta)}{\partial \theta_k \partial \theta_l} = \sum_{i=1}^{\infty} \ddot{\lambda}_{i,k,l}(\theta)\, \epsilon_{t-i}. \tag{2.33}$$

Of course it holds that $\dot{\lambda}_k(\theta) = \frac{\partial \gamma(\theta)}{\partial \theta_k} * \eta(\theta_0)$ and $\ddot{\lambda}_{k,l}(\theta) = \frac{\partial^2 \gamma(\theta)}{\partial \theta_k \partial \theta_l} * \eta(\theta_0)$.

Similarly we have

$$\tilde{\epsilon}_t(\theta) = \sum_{i=0}^{\infty} \lambda_i^t(\theta)\, \epsilon_{t-i}, \tag{2.34}$$

$$\frac{\partial \tilde{\epsilon}_t(\theta)}{\partial \theta_k} = \sum_{i=1}^{\infty} \dot{\lambda}_{i,k}^t(\theta)\, \epsilon_{t-i} \tag{2.35}$$

$$\frac{\partial^2 \tilde{\epsilon}_t(\theta)}{\partial \theta_k \partial \theta_l} = \sum_{i=1}^{\infty} \ddot{\lambda}_{i,k,l}^t(\theta)\, \epsilon_{t-i}, \tag{2.36}$$

where $\lambda^t(\theta) = \gamma^t(\theta) * \eta(\theta_0)$, $\dot{\lambda}_k^t(\theta) = \frac{\partial \gamma^t(\theta)}{\partial \theta_k} * \eta(\theta_0)$ and $\ddot{\lambda}_{k,l}^t(\theta) = \frac{\partial^2 \gamma^t(\theta)}{\partial \theta_k \partial \theta_l} * \eta(\theta_0)$.

In order to handle the truncation error $\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)$, one needs informations on the sequence $\lambda(\theta) - \lambda^t(\theta)$. This is the purpose on the following lemma.

Lemma 2.1. *For $2 \leq r \leq \infty$, $1 \leq k \leq p+q+1$ and $\theta \in \Theta$, we have*

$$\| \lambda(\theta) - \lambda^t(\theta) \|_{\ell^r} = O\left(t^{-1+\frac{1}{r}-(d-d_0)}\right)$$

and

$$\| \dot{\lambda}_k(\theta) - \dot{\lambda}_k^t(\theta) \|_{\ell^r} = O\left(t^{-1+\frac{1}{r}-(d-d_0)}\right).$$

Proof. We have

$$\lambda(\theta) - \lambda^t(\theta) = \left(\gamma(\theta) - \gamma^t(\theta)\right) * \eta(\theta_0).$$

In view of (2.28), the sequence $\eta(\theta_0)$ belongs to ℓ^q for any $q > 1/(1-d_0)$. Young's inequality for convolution yields that for all $r \geq 2$

$$\| \lambda(\theta) - \lambda^t(\theta) \|_{\ell^r} \leq \| \gamma(\theta) - \gamma^t(\theta) \|_{\ell^p} \| \eta(\theta_0) \|_{\ell^q} \tag{2.37}$$

with $q = (1-(d_0+\beta))^{-1} > 1/(1-d_0)$ and $p = r/(1+r(d_0+\beta))$, for some $\beta > 0$ sufficiently small. Thus there exists K such that $\| \eta(\theta_0) \|_{\ell^q} \leq K$. Since for any $j \geq 0$,

$$\gamma_j(\theta) - \gamma_j^t(\theta) = \begin{cases} 0 & \text{if } 0 \leq j \leq t-1 \\ \gamma_j(\theta) & \text{otherwise} \end{cases}$$

we obtain using (2.27) that

$$\begin{aligned}
\| \lambda(\theta) - \lambda^t(\theta) \|_{\ell^r} &\leq K \left(\sum_{k=0}^{\infty} |\gamma_k(\theta) - \gamma_k^t(\theta)|^p \right)^{1/p} \\
&\leq K \left(\sum_{k=t}^{\infty} |\gamma_k(\theta)|^p \right)^{1/p} \\
&\leq K \left(\sum_{k=t}^{\infty} \frac{1}{k^{p+pd}} \right)^{1/p} \\
&\leq K \left(\int_t^{\infty} \frac{1}{x^{p+pd}} \mathrm{d}x \right)^{1/p} \\
&\leq K t^{-1-d+\frac{1}{p}} \\
&\leq K t^{-1+\frac{1}{r}-(d-d_0)+\beta},
\end{aligned}$$

where the constant K varies from line to line. The conclusion follows by tending β to 0.

The second point of the lemma is shown in the same way as the first. This is because from (2.27), the coefficient $\partial\gamma_j(\theta)/\partial\theta_k = \mathrm{O}(j^{-1-d+\zeta})$ for any small enough $\zeta > 0$. The proof of the lemma is then complete. □

Remark 2.3. *Taking $r = \infty$ in the above lemma implies that the sequence $\dot{\lambda}_k(\theta_0) - \dot{\lambda^t}_k(\theta_0)$ is bounded and more precicely there exists K such that*

$$\sup_{j \geq 1} \left| \dot{\lambda}_{j,k}(\theta_0) - \dot{\lambda^t}_{j,k}(\theta_0) \right| \leq \frac{K}{t} \tag{2.38}$$

for any t and any $1 \leq k \leq p+q+1$.

One shall also need the following lemmas.

Lemma 2.2. *For any $2 \leq r \leq \infty$, $1 \leq k \leq p+q+1$ and $\theta \in \Theta$, there exists a constant K such that we have*

$$\| \dot{\lambda}_k^t(\theta) \|_{\ell^r} \leq K.$$

Proof. The proof follows the same arguments than the proof of Lemma 2.1. □

Lemma 2.3. *There exists a constant K such that we have*

$$\left| \dot{\lambda}_{i,k}(\theta_0) \right| \leq \frac{K}{i}. \tag{2.39}$$

Proof. For $1 \leq k \leq p+q+1$, the sequence $\dot{\lambda}_k(\theta) = (\dot{\lambda}_{i,k}(\theta))_{i \geq 1}$ is in fact the sequence of the coefficients in the power series of

$$\frac{\partial}{\partial\theta_k} \left(b_\theta^{-1}(z) a_\theta(z) (1-z)^{d-d_0} a_{\theta_0}^{-1}(z) b_{\theta_0}(z) \right) .$$

Thus $\dot{\lambda}_{i,k}(\theta_0)$ is the $i-$th coefficient taken in $\theta = \theta_0$. There are three cases.

◇ $k = 1, \dots, p$:
Since

$$\frac{\partial}{\partial \theta_k}\left(b_\theta^{-1}(z)a_\theta(z)(1-z)^{d-d_0}a_{\theta_0}^{-1}(z)b_{\theta_0}(z)\right) = -b_\theta^{-1}(z)z^k(1-z)^{d-d_0}a_{\theta_0}^{-1}(z)b_{\theta_0}(z)\ ,$$

we deduce that $\dot{\lambda}_{i,k}(\theta_0)$ is the $i-$th coefficient of $z^k a_{\theta_0}^{-1}(z)$ which satisfies $\dot{\lambda}_{i,k}(\theta_0) \leq K\rho^i$ for some $0\rho < 1$ (see Francq and Zakoïan [1998] for example).

◇ $k = p+1, \dots, p+q$:
We have

$$\frac{\partial}{\partial \theta_k}\left(b_\theta^{-1}(z)a_\theta(z)(1-z)^{d-d_0}a_{\theta_0}(z)b_{\theta_0}(z)\right) = \left(\frac{\partial}{\partial \theta_k}b_\theta^{-1}(z)\right)a_\theta(z)(1-z)^{d-d_0}a_{\theta_0}^{-1}(z)b_{\theta_0}(z)$$

and consequently $\dot{\lambda}_{i,k}(\theta_0)$ is the $i-$th coefficient of $(\frac{\partial}{\partial \theta_k}b_{\theta_0}^{-1}(z))b_{\theta_0}(z)$ which also satisfies $\dot{\lambda}_{i,k}(\theta_0) \leq K\rho^i$ (see Francq and Zakoïan [1998]).
The last case will not be a consequence of the usual works on ARMA processes.

◇ $k = p+q+1$:
In this case, $\theta_k = d$ and so we have

$$\frac{\partial}{\partial \theta_k}\left(b_\theta^{-1}(z)a_\theta(z)(1-z)^{d-d_0}a_{\theta_0}^{-1}(z)b_{\theta_0}(z)\right) = b_\theta^{-1}(z)a_\theta(z)\ln(1-z)(1-z)^{d-d_0}a_{\theta_0}^{-1}(z)b_{\theta_0}(z)$$

and consequently $\dot{\lambda}_{i,k}(\theta_0)$ is the $i-$th coefficient of $\ln(1-z)$ which is equal to $-1/i$.

The three above cases imply the expected result. □

2.6.2 Proof of Theorem 2.1

We can follow from line to line the proof of Theorem 1 in Francq and Zakoïan [1998]. The only difference relies on the following Lemma in which it is stated that the choice of the initial values has no influence on the estimation. Its proof is completely different from the one done in Francq and Zakoïan [1998] because we do not have the same speed of convergence.

Lemma 2.4. *Under the assumptions of Theorem 2.1, we have almost surely*

$$\lim_{t\to\infty}\sup_{\theta\in\Theta_\delta}|\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)| = 0. \tag{2.40}$$

Proof. From (2.25) and (2.29), for all $\theta \in \Theta_\delta$ and all $t \in \mathbb{Z}$, we have

$$\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta) = \sum_{j\geq 0}\gamma_j(\theta)X_{t-j} - \sum_{j=0}^{t-1}\gamma_j(\theta)X_{t-j} = \sum_{j\geq t}\gamma_j(\theta)X_{t-j} = \sum_{k\geq 0}\gamma_{t+k}(\theta)X_{-k}.$$

Recall that for any sequence $(Y_n)_n$ of random variables it holds that

$$Y_n \xrightarrow[n\to\infty]{\text{a.s.}} Y \Leftrightarrow \sup_{k\geq n}|Y_k - Y| \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Hence $\sup_{\theta\in\Theta_\delta}|\epsilon_t(\theta)-\tilde{\epsilon}_t(\theta)|$ converges almost surely to 0 as soon as

$$\sup_{k\geq t}\sup_{\theta\in\Theta_\delta}|\epsilon_k(\theta)-\tilde{\epsilon}_k(\theta)|$$

converges in probability to 0. In view of (2.27), for all $\beta>0$ and for large t we have

$$\begin{aligned}
\mathbb{P}\left(\sup_{k\geq t}\sup_{\theta\in\Theta_\delta}|\epsilon_k(\theta)-\tilde{\epsilon}_k(\theta)|>\beta\right) &= \mathbb{P}\left(\sup_{k\geq t}\sup_{\theta\in\Theta_\delta}\left|\sum_{j\geq 0}\gamma_{k+j}(\theta)X_{-j}\right|>\beta\right)\\
&\leq \mathbb{P}\left(\sum_{j\geq 0}\sup_{k\geq t}\sup_{\theta\in\Theta_\delta}|\gamma_{k+j}(\theta)|\,|X_{-j}|>\beta\right)\\
&\leq \frac{K}{\beta}\left(\sup_{t\in\mathbb{Z}}\mathbb{E}\,|X_t|\right)\sum_{j\geq 0}\left(\frac{1}{t+j}\right)^{1+d_1},\\
&\leq \frac{K}{\beta d_1}(t-1)^{-d_1}\underset{t\to\infty}{\longrightarrow}0
\end{aligned}$$

and (2.40) is proved.

□

2.6.3 Proof of Theorem 2.2

By a Taylor expansion of the function $\partial Q_n(\cdot)/\partial\theta$ around θ_0 and under **(A3)**, we have

$$0=\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\hat{\theta}_n)=\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0)+\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}Q_n\left(\theta^*_{n,i,j}\right)\right]\sqrt{n}\left(\hat{\theta}_n-\theta_0\right), \tag{2.41}$$

where the $\theta^*_{n,i,j}$'s are between $\hat{\theta}_n$ and θ_0. The equation (2.41) can be rewritten in the form:

$$\sqrt{n}\frac{\partial}{\partial\theta}O_n(\theta_0)-\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0)=\sqrt{n}\frac{\partial}{\partial\theta}O_n(\theta_0)+\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}Q_n\left(\theta^*_{n,i,j}\right)\right]\sqrt{n}\left(\hat{\theta}_n-\theta_0\right). \tag{2.42}$$

Under the assumptions of Theorem 2.2, it will be shown respectively in Lemma 2.5 and Lemma 2.8 that

$$\sqrt{n}\frac{\partial}{\partial\theta}O_n(\theta_0)-\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0)=o_{\mathbb{P}}(1),$$

and

$$\lim_{n\to\infty}\left\{\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}Q_n\left(\theta^*_{n,i,j}\right)\right]-J(\theta_0)\right\}=0\quad\text{a.s.}$$

As a consequence, the asymptotic normality of $\sqrt{n}(\hat{\theta}_n-\theta_0)$ will be a consequence of the one of $\sqrt{n}\partial/\partial\theta O_n(\theta_0)$.

Lemma 2.5. *For $1\leq k\leq p+q+1$, under the assumptions of Theorem 2.2, we have*

$$\sqrt{n}\left(\frac{\partial}{\partial\theta_k}Q_n(\theta_0)-\frac{\partial}{\partial\theta_k}O_n(\theta_0)\right)=o_{\mathbb{P}}(1). \tag{2.43}$$

Proof. Throughout this proof, $\theta = (\theta_1, ..., \theta_{p+q}, d)' \in \Theta_\delta$ is such that $d_0 < d \leq d_2$ where d_2 is the upper bound of the support of the long-range dependence parameter d_0.

The proof is quite long so we divide it in several steps.

◇ **Step 1: preliminaries** For $1 \leq k \leq p+q+1$ we have

$$\begin{aligned}
\sqrt{n}\frac{\partial}{\partial\theta_k}Q_n(\theta_0) &= \frac{2}{\sqrt{n}}\sum_{t=1}^{n}\tilde{\epsilon}_t(\theta_0)\frac{\partial}{\partial\theta_k}\tilde{\epsilon}_t(\theta_0) \\
&= \frac{2}{\sqrt{n}}\sum_{t=1}^{n}\left(\tilde{\epsilon}_t(\theta_0)-\tilde{\epsilon}_t(\theta)\right)\frac{\partial}{\partial\theta_k}\tilde{\epsilon}_t(\theta_0) + \frac{2}{\sqrt{n}}\sum_{t=1}^{n}\left(\tilde{\epsilon}_t(\theta)-\epsilon_t(\theta)\right)\frac{\partial}{\partial\theta_k}\tilde{\epsilon}_t(\theta_0) \\
&\quad + \frac{2}{\sqrt{n}}\sum_{t=1}^{n}\left(\epsilon_t(\theta)-\epsilon_t(\theta_0)\right)\frac{\partial}{\partial\theta_k}\tilde{\epsilon}_t(\theta_0) + \frac{2}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t(\theta_0)\left(\frac{\partial}{\partial\theta_k}\tilde{\epsilon}_t(\theta_0)-\frac{\partial}{\partial\theta_k}\epsilon_t(\theta_0)\right) \\
&\quad + \frac{2}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t(\theta_0)\frac{\partial}{\partial\theta_k}\epsilon_t(\theta_0) \\
&= \Delta_{n,1}^k(\theta) + \Delta_{n,2}^k(\theta) + \Delta_{n,3}^k(\theta) + \Delta_{n,4}^k(\theta_0) + \sqrt{n}\frac{\partial}{\partial\theta_k}O_n(\theta_0), \qquad (2.44)
\end{aligned}$$

where

$$\begin{aligned}
\Delta_{n,1}^k(\theta) &= \frac{2}{\sqrt{n}}\sum_{t=1}^{n}\left(\tilde{\epsilon}_t(\theta_0)-\tilde{\epsilon}_t(\theta)\right)\frac{\partial}{\partial\theta_k}\tilde{\epsilon}_t(\theta_0), \\
\Delta_{n,2}^k(\theta) &= \frac{2}{\sqrt{n}}\sum_{t=1}^{n}\left(\tilde{\epsilon}_t(\theta)-\epsilon_t(\theta)\right)\frac{\partial}{\partial\theta_k}\tilde{\epsilon}_t(\theta_0), \\
\Delta_{n,3}^k(\theta) &= \frac{2}{\sqrt{n}}\sum_{t=1}^{n}\left(\epsilon_t(\theta)-\epsilon_t(\theta_0)\right)\frac{\partial}{\partial\theta_k}\tilde{\epsilon}_t(\theta_0) \\
\Delta_{n,4}^k(\theta_0) &= \frac{2}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t(\theta_0)\left(\frac{\partial}{\partial\theta_k}\tilde{\epsilon}_t(\theta_0)-\frac{\partial}{\partial\theta_k}\epsilon_t(\theta_0)\right).
\end{aligned}$$

Using (2.30) and (2.34), the fourth term $\Delta_{n,4}^k(\theta_0)$ can be rewritten in the form:

$$\Delta_{n,4}^k(\theta_0) = \frac{2}{\sqrt{n}}\sum_{t=1}^{n}\sum_{j=1}^{\infty}\left\{\dot{\lambda}_{j,k}^{t}(\theta_0) - \dot{\lambda}_{j,k}(\theta_0)\right\}\epsilon_t\epsilon_{t-j}. \qquad (2.45)$$

Therefore, if we prove that the three sequences of random variables $(\Delta_{n,1}^k(\theta) + \Delta_{n,3}^k(\theta))_n$, $(\Delta_{n,2}^k(\theta))_n$ and $(\Delta_{n,4}^k(\theta_0))_n$ converge in probability towards 0, then (2.43) will be true.

◇ **Step 2: convergence in probability of $(\Delta_{n,4}^k(\theta_0))_n$ to 0**

For simplicity, in this step we denote in the sequel by $\dot{\lambda}_{j,k}$ the coefficient $\dot{\lambda}_{j,k}(\theta_0)$ and by $\dot{\lambda}_{j,k}^{t}$

the coefficient $\dot{\lambda}_{j,k}^{t}(\theta_0)$. Let $\varrho(\cdot,\cdot)$ be the function defined for $1 \leq t, s \leq n$ by

$$\varrho(t,s) = \sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\left\{\dot{\lambda}_{j_1,k} - \dot{\lambda}_{j_1,k}^{t}\right\}\left\{\dot{\lambda}_{j_2,k} - \dot{\lambda}_{j_2,k}^{s}\right\}\mathbb{E}\left[\epsilon_t\epsilon_{t-j_1}\epsilon_s\epsilon_{s-j_2}\right].$$

For all $\beta > 0$, using the symmetry of the function $\varrho(t,s)$, we obtain that

$$\begin{aligned}
\mathbb{P}\left(\left|\Delta_{n,4}^{k}(\theta_0)\right| \geq \beta\right) &\leq \frac{4}{n\beta^2}\mathbb{E}\left[\left(\sum_{t=1}^{n}\sum_{j=1}^{\infty}\left\{\dot{\lambda}_{j,k} - \dot{\lambda}_{j,k}^{t}\right\}\epsilon_t\epsilon_{t-j}\right)^2\right] \\
&\leq \frac{4}{n\beta^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\left\{\dot{\lambda}_{j_1,k} - \dot{\lambda}_{j_1,k}^{t}\right\}\left\{\dot{\lambda}_{j_2,k} - \dot{\lambda}_{j_2,k}^{s}\right\}\mathbb{E}\left[\epsilon_t\epsilon_{t-j_1}\epsilon_s\epsilon_{s-j_2}\right] \\
&\leq \frac{8}{n\beta^2}\sum_{t=1}^{n}\sum_{s=1}^{t}\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\left\{\dot{\lambda}_{j_1,k} - \dot{\lambda}_{j_1,k}^{t}\right\}\left\{\dot{\lambda}_{j_2,k} - \dot{\lambda}_{j_2,k}^{s}\right\}\mathbb{E}\left[\epsilon_t\epsilon_{t-j_1}\epsilon_s\epsilon_{s-j_2}\right].
\end{aligned}$$

By the stationarity of $(\epsilon_t)_{t\in\mathbb{Z}}$ which is assumed in **(A2)**, we have

$$\begin{aligned}
\mathbb{E}\left[\epsilon_t\epsilon_{t-j_1}\epsilon_s\epsilon_{s-j_2}\right] = \operatorname{cum}\left(\epsilon_0, \epsilon_{-j_1}, \epsilon_{s-t}, \epsilon_{s-t-j_2}\right) &+ \mathbb{E}\left[\epsilon_0\epsilon_{-j_1}\right]\mathbb{E}\left[\epsilon_{s-t}\epsilon_{s-t-j_2}\right] \\
&+ \mathbb{E}\left[\epsilon_0\epsilon_{s-t}\right]\mathbb{E}\left[\epsilon_{-j_1}\epsilon_{s-t-j_2}\right] + \mathbb{E}\left[\epsilon_0\epsilon_{s-t-j_2}\right]\mathbb{E}\left[\epsilon_{-j_1}\epsilon_{s-t}\right].
\end{aligned}$$

Since the noise is not correlated, we deduce that $\mathbb{E}\left[\epsilon_0\epsilon_{-j_1}\right] = 0$ and $\mathbb{E}\left[\epsilon_0\epsilon_{s-t-j_2}\right] = 0$ for $1 \leq j_1, j_2$ and $s \leq t$. Consequently we obtain

$$\begin{aligned}
&\mathbb{P}\left(\left|\Delta_{n,4}^{k}(\theta_0)\right| \geq \beta\right) \\
&\quad\leq \frac{8}{n\beta^2}\sum_{t=1}^{n}\sum_{s=1}^{t}\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\sup_{j_1\geq 1}\left|\dot{\lambda}_{j_1,k} - \dot{\lambda}_{j_1,k}^{t}\right|\left|\dot{\lambda}_{j_2,k} - \dot{\lambda}_{j_2,k}^{s}\right|\left|\operatorname{cum}\left(\epsilon_0, \epsilon_{-j_1}, \epsilon_{s-t}, \epsilon_{s-t-j_2}\right)\right| \\
&\qquad+ \frac{8}{n\beta^2}\sum_{t=1}^{n}\sum_{s=1}^{t}\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\left|\dot{\lambda}_{j_1,k} - \dot{\lambda}_{j_1,k}^{t}\right|\left|\dot{\lambda}_{j_2,k} - \dot{\lambda}_{j_2,k}^{s}\right|\left|\mathbb{E}\left[\epsilon_0\epsilon_{s-t}\right]\mathbb{E}\left[\epsilon_{-j_1}\epsilon_{s-t-j_2}\right]\right|. \qquad (2.46)
\end{aligned}$$

If

$$\sum_{s=1}^{t}\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\sup_{j_1\geq 1}\left|\dot{\lambda}_{j_1,k} - \dot{\lambda}_{j_1,k}^{t}\right|\left|\dot{\lambda}_{j_2,k} - \dot{\lambda}_{j_2,k}^{s}\right|\left|\operatorname{cum}\left(\epsilon_0, \epsilon_{-j_1}, \epsilon_{s-t}, \epsilon_{s-t-j_2}\right)\right| \xrightarrow[t\to\infty]{} 0, \qquad (2.47)$$

Cesàro's Lemma implies that the first term in the right hand side of (2.46) tends to 0. Thanks to Lemma 2.1 applied with $r = \infty$ (or see Remark 2.3) and Assumption **(A4')** with $\tau = 4$, we

obtain that

$$\begin{aligned}\sum_{s=1}^{t}\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\sup_{j_1\geq 1}\left|\dot{\lambda}_{j_1,k}-\dot{\lambda}_{j_1,k}^{t}\right|\left|\dot{\lambda}_{j_2,k}-\dot{\lambda}_{j_2,k}^{s}\right|\left|\mathrm{cum}\left(\epsilon_0,\epsilon_{-j_1},\epsilon_{s-t},\epsilon_{s-t-j_2}\right)\right| \\ \leq \frac{K}{t}\sum_{s=1}^{t}\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\left|\mathrm{cum}\left(\epsilon_0,\epsilon_{-j_1},\epsilon_{s-t},\epsilon_{s-t-j_2}\right)\right| \\ \leq \frac{K}{t}\sum_{s=-\infty}^{\infty}\sum_{j_1=-\infty}^{\infty}\sum_{j_2=-\infty}^{\infty}\left|\mathrm{cum}\left(\epsilon_0,\epsilon_s,\epsilon_{j_1},\epsilon_{j_2}\right)\right| \xrightarrow[t\to\infty]{} 0\ ,\end{aligned}$$

hence (2.47) holds true. Concerning the second term of right hand side of the inequality (2.46), we have

$$\begin{aligned}\frac{8}{n\beta^2}\sum_{t=1}^{n}\sum_{s=1}^{t}\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\left|\dot{\lambda}_{j_1,k}-\dot{\lambda}_{j_1,k}^{t}\right|\left|\dot{\lambda}_{j_2,k}-\dot{\lambda}_{j_2,k}^{s}\right|\left|\mathbb{E}\left[\epsilon_0\epsilon_{s-t}\right]\mathbb{E}\left[\epsilon_{-j_1}\epsilon_{s-t-j_2}\right]\right| \\ = \frac{8\sigma_\epsilon^2}{n\beta^2}\sum_{t=1}^{n}\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\left|\dot{\lambda}_{j_1,k}-\dot{\lambda}_{j_1,k}^{t}\right|\left|\dot{\lambda}_{j_2,k}-\dot{\lambda}_{j_2,k}^{t}\right|\left|\mathbb{E}\left[\epsilon_{-j_1}\epsilon_{-j_2}\right]\right| \\ = \frac{8\sigma_\epsilon^4}{n\beta^2}\sum_{t=1}^{n}\sum_{j_1=1}^{\infty}\left|\dot{\lambda}_{j_1,k}-\dot{\lambda}_{j_1,k}^{t}\right|^2 \\ = \frac{8\sigma_\epsilon^4}{n\beta^2}\sum_{t=1}^{n}\left\|\dot{\lambda}_{k}-\dot{\lambda}_{k}^{t}\right\|_{\ell^2}^2 \\ \leq \frac{K}{\beta^2}\frac{1}{n}\sum_{t=1}^{n}\frac{1}{t}\xrightarrow[n\to\infty]{} 0,\end{aligned}$$

where we have used the fact that the noise is not correlated, Lemma 2.1 with $r = 2$ and Cesàro's Lemma. This ends Step 2.

◇ **Step 3: $(\Delta_{n,2}^k(\theta))_n$ converges in probability to 0**
For all $\beta > 0$, we have

$$\mathbb{P}\left(\left|\Delta_{n,2}^k(\theta)\right|\geq\beta\right)\leq\frac{2}{\beta\sqrt{n}}\sum_{t=1}^{n}\left\|\tilde{\epsilon}_t(\theta)-\epsilon_t(\theta)\right\|_{\mathbb{L}^2}\left\|\frac{\partial}{\partial\theta_k}\tilde{\epsilon}_t(\theta_0)\right\|_{\mathbb{L}^2}.$$

First, using Lemma 2.2, we have

$$\begin{aligned}\left\|\frac{\partial}{\partial\theta_k}\tilde{\epsilon}_t(\theta_0)\right\|_{\mathbb{L}^2}^2 &= \mathbb{E}\left[\left(\sum_{i=1}^{\infty}\dot{\lambda}_{i,k}^{t}\left(\theta_0\right)\epsilon_{t-i}\right)^2\right] \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\dot{\lambda}_{i,k}^{t}\left(\theta_0\right)\dot{\lambda}_{j,k}^{t}\left(\theta_0\right)\mathbb{E}\left[\epsilon_{t-i}\epsilon_{t-j}\right]\end{aligned}$$

$$= \sigma_\epsilon^2 \sum_{i=1}^{\infty} \left\{ \dot{\lambda}_{i,k}^{t}(\theta_0) \right\}^2$$
$$\leq K. \tag{2.48}$$

In view of (2.30), (2.34) and (2.48), we may write

$$\begin{aligned}
\mathbb{P}\left(\left|\Delta_{n,2}^k(\theta)\right| \geq \beta\right) &\leq \frac{K}{\beta\sqrt{n}} \sum_{t=1}^{n} \left(\mathbb{E}\left[\left(\tilde{\epsilon}_t(\theta) - \epsilon_t(\theta)\right)^2 \right] \right)^{1/2} \\
&\leq \frac{K}{\beta\sqrt{n}} \sum_{t=1}^{n} \left(\sum_{i\geq 0} \sum_{j\geq 0} \left(\lambda_i^t(\theta) - \lambda_i(\theta)\right) \left(\lambda_j^t(\theta) - \lambda_j(\theta)\right) \mathbb{E}\left[\epsilon_{t-i}\epsilon_{t-j}\right] \right)^{1/2} \\
&\leq \frac{\sigma_\epsilon K}{\beta\sqrt{n}} \sum_{t=1}^{n} \left(\sum_{i\geq 0} \left(\lambda_i^t(\theta) - \lambda_i(\theta)\right)^2 \right)^{1/2} \\
&\leq \frac{\sigma_\epsilon K}{\beta\sqrt{n}} \sum_{t=1}^{n} \left\| \lambda(\theta) - \lambda^t(\theta) \right\|_{\ell^2}.
\end{aligned}$$

We use Lemma 2.1, the fact that $d > d_0$ and the fractional version of Cesàro's Lemma [2], we obtain

$$\mathbb{P}\left(\left|\Delta_{n,2}^k(\theta)\right| \geq \beta\right) \leq \frac{\sigma_\epsilon K}{\beta} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{1}{t^{1/2+(d-d_0)}} \xrightarrow[n\to\infty]{} 0.$$

This proves the expected convergence in probability.

◇ **Step 4: Convergence of** $(\Delta_{n,1}^k(\theta) + \Delta_{n,3}^k(\theta))_n$
Note now that, for all $n \geq 1$, we have

$$\Delta_{n,1}^k(\theta) + \Delta_{n,3}^k(\theta) = \frac{2}{\sqrt{n}} \sum_{t=1}^{n} \left\{ \left(\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)\right) - \left(\epsilon_t(\theta_0) - \tilde{\epsilon}_t(\theta_0)\right) \right\} \frac{\partial}{\partial\theta_k} \tilde{\epsilon}_t(\theta_0).$$

By the mean value theorem, there exists $0 < c_\omega < 1$ such that

$$\left| \left(\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)\right) - \left(\epsilon_t(\theta_0) - \tilde{\epsilon}_t(\theta_0)\right) \right| \leq \left\| \frac{\partial(\epsilon_t - \tilde{\epsilon}_t)}{\partial\theta}((1-c_\omega)\theta + c_\omega\theta_0) \right\|_{\mathbb{R}^{p+q+1}} \|\theta - \theta_0\|_{\mathbb{R}^{p+q+1}}. \tag{2.49}$$

2. Recall that the fractional version of Cesàro's Lemma states that for $(h_t)_t$ a sequence of positive real numbers, $\kappa > 0$ and $c \geq 0$ we have

$$\lim_{t\to\infty} h_t t^{1-\kappa} = |\kappa|\, c \Rightarrow \lim_{n\to\infty} \frac{1}{n^\kappa} \sum_{t=0}^{n} h_t = c.$$

Following the same method than in the previous step we obtain

$$\begin{aligned}
&\mathbb{E}\Big((\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)) - (\epsilon_t(\theta_0) - \tilde{\epsilon}_t(\theta_0)) \Big)^2 \\
&\qquad \leq \|\theta - \theta_0\|^2_{\mathbb{R}^{p+q+1}} \sum_{k=1}^{p+q+1} \mathbb{E}\left[\left| \frac{\partial(\epsilon_t - \tilde{\epsilon}_t)}{\partial\theta_k}((1-c_\omega)\theta + c_\omega\theta_0) \right|^2 \right] \\
&\qquad \leq \|\theta - \theta_0\|^2_{\mathbb{R}^{p+q+1}} \sum_{k=1}^{p+q+1} \sup_\theta \mathbb{E}\left[\left| \frac{\partial(\epsilon_t - \tilde{\epsilon}_t)}{\partial\theta_k}(\theta) \right|^2 \right] \\
&\qquad \leq \|\theta - \theta_0\|^2_{\mathbb{R}^{p+q+1}} \sum_{k=1}^{p+q+1} \sigma_\epsilon^2 \sup_\theta \left\| (\dot{\lambda}_k - \dot{\lambda}_k^{\,t})(\theta) \right\|^2_{\ell^2} \\
&\qquad \leq K \|\theta - \theta_0\|^2_{\mathbb{R}^{p+q+1}} \sup_{d; d_0 \leq d \leq d_2} \left(\frac{1}{t^{1/2+(d-d_0)}} \right)^2 \\
&\qquad \leq K \|\theta - \theta_0\|^2_{\mathbb{R}^{p+q+1}} \frac{1}{t}, \qquad (2.50)
\end{aligned}$$

where we have used the fact that the function

$$\theta \mapsto \mathbb{E}\left[\left| \frac{\partial(\epsilon_t - \tilde{\epsilon}_t)}{\partial\theta_k}(\theta) \right|^2 \right]$$

is bounded and continuous. By (2.50) and (2.48), it follows that

$$\mathbb{P}\left(\left| \Delta^k_{n,1}(\theta) + \Delta^k_{n,3}(\theta) \right| \geq \beta \right) \leq \frac{K}{\beta} \|\theta - \theta_0\|_{\mathbb{R}^{p+q+1}} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{1}{t^{1/2}}$$

and the fractional version of Cesàro's Lemma implies

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \Delta^k_{n,1}(\theta) + \Delta^k_{n,3}(\theta) \right| \geq \beta \right) \leq \frac{K}{\beta} \|\theta - \theta_0\|_{\mathbb{R}^{p+q+1}}. \qquad (2.51)$$

◇ **Step 5: end of the proof** For any $\varepsilon > 0$, we choose θ such that $\frac{K}{\beta} \|\theta - \theta_0\|_{\mathbb{R}^{p+q+1}} \leq \varepsilon$. Then, from (2.51), there exists n_0 such that for all $n \geq n_0$,

$$\mathbb{P}\left(\left| \Delta^k_{n,1}(\theta) + \Delta^k_{n,3}(\theta) \right| \geq \beta \right) \leq \varepsilon.$$

By Step 2 and 3, one also has for $n \geq n_0$

$$\mathbb{P}\left(\left| \Delta^k_{n,2}(\theta) + \Delta^k_{n,4}(\theta_0) \right| \geq \beta \right) \leq \varepsilon.$$

Therefore, for all $n \geq n_0$,

$$\begin{aligned}
\mathbb{P}\left(\left| \sqrt{n}\frac{\partial}{\partial\theta_k} Q_n(\theta_0) - \sqrt{n}\frac{\partial}{\partial\theta_k} O_n(\theta_0) \right| \geq 2\beta \right) &\leq \mathbb{P}\left(\left| \Delta^k_{n,1}(\theta) + \Delta^k_{n,3}(\theta) \right| \geq \beta \right) \\
&\quad + \mathbb{P}\left(\left| \Delta^k_{n,2}(\theta) + \Delta^k_{n,4}(\theta_0) \right| \geq \beta \right) \\
&\leq \varepsilon
\end{aligned}$$

and the expected convergence in probability is proved. □

We show in the following lemma the existence and invertibility of $J(\theta_0)$.

Lemma 2.6. *Under Assumptions of Theorem 2.2, the matrix*

$$J(\theta_0) = \lim_{n\to\infty}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}O_n(\theta_0)\right]$$

exists almost surely and is invertible.

Proof. For all $1 \leq i,j \leq p+q+1$, we have

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}O_n(\theta_0) = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\epsilon_t^2(\theta_0) = \frac{2}{n}\sum_{t=1}^{n}\left\{\frac{\partial}{\partial\theta_i}\epsilon_t(\theta_0)\frac{\partial}{\partial\theta_j}\epsilon_t(\theta_0) + \epsilon_t(\theta_0)\frac{\partial^2}{\partial\theta_i\partial\theta_j}\epsilon_t(\theta_0)\right\}.$$

Note that in view of (2.25), (2.26) and (2.27), the first and second order derivatives of $\epsilon_t(\cdot)$ belong to $\mathbb{L}^2$. By using the ergodicity of $(\epsilon_t)_{t\in\mathbb{Z}}$ assumed in Assumption **(A2)**, we deduce that

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}O_n(\theta_0) \xrightarrow[n\to\infty]{\text{a.s.}} 2\mathbb{E}\left[\frac{\partial}{\partial\theta_i}\epsilon_t(\theta_0)\frac{\partial}{\partial\theta_j}\epsilon_t(\theta_0)\right] + 2\mathbb{E}\left[\epsilon_t(\theta_0)\frac{\partial^2}{\partial\theta_i\partial\theta_j}\epsilon_t(\theta_0)\right].$$

By (2.30) and (2.33), ϵ_t and $\partial\epsilon_t(\theta_0)/\partial\theta$ are non correlated as well as ϵ_t and $\partial^2\epsilon_t(\theta_0)/\partial\theta\partial\theta$. Thus we have

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}O_n(\theta_0) \xrightarrow[n\to\infty]{\text{a.s.}} J(\theta_0)(i,j) := 2\mathbb{E}\left[\frac{\partial}{\partial\theta_i}\epsilon_t(\theta_0)\frac{\partial}{\partial\theta_j}\epsilon_t(\theta_0)\right]. \tag{2.52}$$

From (2.30) and (2.39) we obtain that

$$\begin{aligned}\mathbb{E}\left[\frac{\partial}{\partial\theta_i}\epsilon_t(\theta_0)\frac{\partial}{\partial\theta_j}\epsilon_t(\theta_0)\right] &= \mathbb{E}\left[\left(\sum_{k_1\geq 1}\dot{\lambda}_{k_1,i}(\theta_0)\,\epsilon_{t-k_1}\right)\left(\sum_{k_2\geq 1}\dot{\lambda}_{k_2,j}(\theta_0)\,\epsilon_{t-k_2}\right)\right]\\ &= \sum_{k_1\geq 1}\sum_{k_2\geq 1}\dot{\lambda}_{k_1,i}(\theta_0)\,\dot{\lambda}_{k_2,j}(\theta_0)\,\mathbb{E}\left[\epsilon_{t-k_1}\epsilon_{t-k_2}\right]\\ &\leq K\,\sigma_\epsilon^2\sum_{k_1\geq 1}\left(\frac{1}{k_1}\right)^2 < \infty.\end{aligned}$$

Therefore $J(\theta_0)$ exists almost surely.

If the matrix $J(\theta_0)$ is not invertible, there exists some real constants $c_1,\ldots,c_{p+q+1}$ not all equal to zero such that

$$\mathbf{c}'J(\theta_0)\mathbf{c} = \sum_{i=1}^{p+q+1}\sum_{j=1}^{p+q+1}c_jJ(\theta_0)(j,i)c_i = 0,$$

where $\mathbf{c} = (c_1,\ldots,c_{p+q+1})'$. In view of (2.52) we obtain that

$$\sum_{i=1}^{p+q+1}\sum_{j=1}^{p+q+1}\mathbb{E}\left[\left(c_j\frac{\partial\epsilon_t(\theta_0)}{\partial\theta_j}\right)\left(c_i\frac{\partial\epsilon_t(\theta_0)}{\partial\theta_i}\right)\right] = \mathbb{E}\left[\left(\sum_{k=1}^{p+q+1}c_k\frac{\partial\epsilon_t(\theta_0)}{\partial\theta_k}\right)^2\right] = 0,$$

which implies that

$$\sum_{k=1}^{p+q+1} c_k \frac{\partial \epsilon_t(\theta_0)}{\partial \theta_k} = 0 \quad \text{a.s. or equivalenty} \quad \mathbf{c}' \frac{\partial \epsilon_t(\theta_0)}{\partial \theta} = 0 \quad \text{a.s.} \tag{2.53}$$

Differentiating the equation (2.1), we obtain that

$$\mathbf{c}' \frac{\partial}{\partial \theta} \left\{ a_{\theta_0}(L)(1-L)^{d_0} \right\} X_t = \mathbf{c}' \left\{ \frac{\partial}{\partial \theta} b_{\theta_0}(L) \right\} \epsilon_t + b_{\theta_0}(L) \mathbf{c}' \frac{\partial}{\partial \theta} \epsilon_t(\theta_0).$$

and by (2.53) we may write that

$$\mathbf{c}' \left(\frac{\partial}{\partial \theta} \left\{ a_{\theta_0}(L)(1-L)^{d_0} \right\} X_t - \left\{ \frac{\partial}{\partial \theta} b_{\theta_0}(L) \right\} \epsilon_t \right) = 0 \quad \text{a.s.}$$

It follows that (2.1) can therefore be rewritten in the form:

$$\left(a_{\theta_0}(L)(1-L)^{d_0} + \mathbf{c}' \frac{\partial}{\partial \theta} \left\{ a_{\theta_0}(L)(1-L)^{d_0} \right\} \right) X_t = \left(b_{\theta_0}(L) + \mathbf{c}' \frac{\partial}{\partial \theta} b_{\theta_0}(L) \right) \epsilon_t, \quad \text{a.s.}$$

Under Assumption **(A1)** the representation in (2.1) is unique (see Hosking [1981]) so

$$\mathbf{c}' \frac{\partial}{\partial \theta} \left\{ a_{\theta_0}(L)(1-L)^{d_0} \right\} = 0 \text{ and} \tag{2.54}$$

$$\mathbf{c}' \frac{\partial}{\partial \theta} b_{\theta_0}(L) = 0. \tag{2.55}$$

First, (2.55) implies that

$$\sum_{k=p+1}^{p+q} c_k \frac{\partial}{\partial \theta_k} b_{\theta_0}(L) = \sum_{k=p+1}^{p+q} -c_k L^k = 0$$

and thus $c_k = 0$ for $p+1 \leq k \leq p+q$.

Similarly, (2.54) yields that

$$\sum_{k=1}^{p} c_k \frac{\partial}{\partial \theta_k} a_{\theta_0}(L)(1-L)^{d_0} + c_{p+q+1} a_{\theta_0}(L) \frac{\partial (1-L)^d}{\partial d}(d_0) = 0 \ .$$

Since $\partial(1-L)^d/\partial d = (1-L)^d \ln(1-L)$, it follows that

$$-\sum_{k=1}^{p} c_k L^k + c_{p+q+1} \sum_{k \geq 0} e_k L^k = 0 \ ,$$

where the sequence $(e_k)_{k \geq 1}$ is given by the coefficients of the power series $a_{\theta_0}(L) \ln(1-L)$. Since $e_0 = 0$ and $e_1 = -1$, we obtain that

$$\begin{aligned} c_1 &= -c_{p+q+1} \\ c_k &= e_k c_{p+q+1} \text{ for } k = 2, \ldots, p \\ 0 &= e_k c_{p+q+1} \text{ for } k \geq p+1. \end{aligned}$$

Since the polynomial a_{θ_0} is not the null polynomial, this implies that $c_{p+q+1} = 0$ and then c_k for $1 \leq k \leq p$. Thus $\mathbf{c} = 0$ which leads us to a contradiction. Hence $J(\theta_0)$ is invertible.

□

Lemma 2.7. *For any $1 \leq i, j \leq p+q+1$ and under the assumptions of Theorem 2.1, we have almost surely*

$$\lim_{t\to\infty} \sup_{\theta\in\Theta_\delta} \left| \frac{\partial}{\partial\theta_i}\epsilon_t(\theta) - \frac{\partial}{\partial\theta_i}\tilde{\epsilon}_t(\theta) \right| = 0 \quad \textit{and} \quad \lim_{t\to\infty} \sup_{\theta\in\Theta_\delta} \left| \frac{\partial^2}{\partial\theta_i\partial\theta_j}\epsilon_t(\theta) - \frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde{\epsilon}_t(\theta) \right| = 0. \quad (2.56)$$

Proof. The proof uses the same arguments that the proof of Lemma 2.4 so it is omitted.

□

Lemma 2.8. *For any $1 \leq i, j \leq p+q+1$ and under the assumptions of Theorem 2.1, we have almost surely*

$$\lim_{n\to\infty} \left\{ \left[\frac{\partial^2}{\partial\theta_i\partial\theta_j} Q_n\left(\theta^*_{n,i,j}\right) \right] - J(\theta_0) \right\} = 0 \quad (2.57)$$

where $\theta^*_{n,i,j}$ is defined in (2.41).

Proof. For any $\theta \in \Theta_\delta$, let

$$J_n(\theta) = \frac{\partial^2}{\partial\theta\partial\theta'} Q_n(\theta) = \frac{2}{n}\sum_{t=1}^{n} \left\{ \frac{\partial}{\partial\theta}\tilde{\epsilon}_t(\theta) \right\} \left\{ \frac{\partial}{\partial\theta'}\tilde{\epsilon}_t(\theta) \right\} + \frac{2}{n}\sum_{t=1}^{n} \tilde{\epsilon}_t(\theta) \frac{\partial^2}{\partial\theta\partial\theta'}\tilde{\epsilon}_t(\theta),$$

and

$$J^*_n(\theta) = \frac{\partial^2}{\partial\theta\partial\theta'} O_n(\theta) = \frac{2}{n}\sum_{t=1}^{n} \left\{ \frac{\partial}{\partial\theta}\epsilon_t(\theta) \right\} \left\{ \frac{\partial}{\partial\theta'}\epsilon_t(\theta) \right\} + \frac{2}{n}\sum_{t=1}^{n} \epsilon_t(\theta) \frac{\partial^2}{\partial\theta\partial\theta'}\epsilon_t(\theta).$$

We have

$$\begin{aligned} \left| \frac{\partial^2}{\partial\theta_i\partial\theta_j} Q_n\left(\theta^*_{n,i,j}\right) - J(\theta_0)(i,j) \right| &\leq \left| J_n(\theta^*_{n,i,j})(i,j) - J^*_n(\theta^*_{n,i,j})(i,j) \right| \\ &\quad + \left| J^*_n(\theta^*_{n,i,j})(i,j) - J^*_n(\theta_0)(i,j) \right| + \left| J^*_n(\theta_0)(i,j) - J(\theta_0)(i,j) \right|. \end{aligned} \quad (2.58)$$

So it is enough to show that the three terms in the right hand side of (2.58) tend almost-surely to 0 when n tends to infinity. Following the same arguments as the proof of Lemma 2.6 and applying the ergodic theorem, we obtain that

$$J^*_n(\theta_0) \xrightarrow[n\to\infty]{\text{a.s.}} 2\mathbb{E}\left[\frac{\partial}{\partial\theta}\epsilon_t(\theta_0) \frac{\partial}{\partial\theta'}\epsilon_t(\theta_0) \right] = J(\theta_0).$$

Let us now show that the term $|J_n^*(\theta_{n,i,j}^*)(i,j) - J_n^*(\theta_0)(i,j)|$ converges almost-surely to 0. In view of (2.25) and (2.27), we have

$$\begin{aligned}
&\sup_{\theta\in\Theta_\delta}\left\|\frac{\partial}{\partial\theta}\left(\frac{\partial}{\partial\theta_i}\epsilon_t(\theta)\frac{\partial}{\partial\theta_j}\epsilon_t(\theta)\right)\right\| \\
&\quad= \sup_{\theta\in\Theta_\delta}\left\|\frac{\partial}{\partial\theta}\left\{\left(\sum_{k_1\geq 1}\frac{\partial}{\partial\theta_i}\gamma_{k_1}(\theta)X_{t-k_1}\right)\left(\sum_{k_2\geq 1}\frac{\partial}{\partial\theta_j}\gamma_{k_2}(\theta)X_{t-k_2}\right)\right\}\right\| \\
&\quad= \sup_{\theta\in\Theta_\delta}\left\|\frac{\partial}{\partial\theta}\left(\sum_{k_1,k_2\geq 1}\frac{\partial}{\partial\theta_i}\gamma_{k_1}(\theta)\frac{\partial}{\partial\theta_j}\gamma_{k_2}(\theta)X_{t-k_1}X_{t-k_2}\right)\right\| \\
&\quad\leq \sup_{\theta\in\Theta_\delta}\left\|\sum_{k_1,k_2\geq 1}\left(\frac{\partial}{\partial\theta}\frac{\partial}{\partial\theta_i}\gamma_{k_1}(\theta)\right)\frac{\partial}{\partial\theta_j}\gamma_{k_2}(\theta)X_{t-k_1}X_{t-k_2}\right\| \\
&\qquad+ \sup_{\theta\in\Theta_\delta}\left\|\sum_{k_1,k_2\geq 1}\frac{\partial}{\partial\theta_i}\gamma_{k_1}(\theta)\left(\frac{\partial}{\partial\theta}\frac{\partial}{\partial\theta_j}\gamma_{k_2}(\theta)\right)X_{t-k_1}X_{t-k_2}\right\| \\
&\quad\leq \sum_{k_1,k_2\geq 1}\sup_{\theta\in\Theta_\delta}\left\|\frac{\partial}{\partial\theta}\frac{\partial}{\partial\theta_i}\gamma_{k_1}(\theta)\right\|\sup_{\theta\in\Theta_\delta}\left\|\frac{\partial}{\partial\theta_j}\gamma_{k_2}(\theta)\right\||X_{t-k_1}||X_{t-k_2}| \\
&\qquad+ \sum_{k_1,k_2\geq 1}\sup_{\theta\in\Theta_\delta}\left\|\frac{\partial}{\partial\theta_i}\gamma_{k_1}(\theta)\right\|\sup_{\theta\in\Theta_\delta}\left\|\frac{\partial}{\partial\theta}\frac{\partial}{\partial\theta_j}\gamma_{k_2}(\theta)\right\||X_{t-k_1}||X_{t-k_2}| \\
&\quad\leq K\sum_{k_1,k_2\geq 1}(\log(k_1))^2\,k_1^{-1-d_1}\log(k_2)k_2^{-1-d_1}\,|X_{t-k_1}||X_{t-k_2}| \\
&\qquad+ K\sum_{k_1,k_2\geq 1}\log(k_1)k_1^{-1-d_1}(\log(k_2))^2\,k_2^{-1-d_1}\,|X_{t-k_1}||X_{t-k_2}|\,.
\end{aligned}$$

Consequently we obtain

$$\begin{aligned}
\mathbb{E}_{\theta_0}\left[\sup_{\theta\in\Theta_\delta}\left\|\frac{\partial}{\partial\theta}\left(\frac{\partial}{\partial\theta_i}\epsilon_t(\theta)\frac{\partial}{\partial\theta_j}\epsilon_t(\theta)\right)\right\|\right] &\leq K\sum_{k_1,k_2\geq 1}(\log(k_1))^2\,k_1^{-1-d_1}\log(k_2)k_2^{-1-d_1}\sup_{t\in\mathbb{Z}}\mathbb{E}_{\theta_0}|X_t|^2 \\
&\quad+ K\sum_{k_1,k_2\geq 1}\log(k_1)k_1^{-1-d_1}(\log(k_2))^2\,k_2^{-1-d_1}\sup_{t\in\mathbb{Z}}\mathbb{E}_{\theta_0}|X_t|^2 \\
&\leq K. \qquad (2.59)
\end{aligned}$$

Following the same approach used in (2.59), we have

$$\mathbb{E}_{\theta_0}\left[\sup_{\theta\in\Theta_\delta}\left\|\frac{\partial}{\partial\theta}\left\{\epsilon_t(\theta)\frac{\partial^2}{\partial\theta_i\partial\theta_j}\epsilon_t(\theta)\right\}\right\|\right] < \infty. \qquad (2.60)$$

A Taylor expansion implies that there exists a random variable $\theta_{n,i,j}^{**}$'s between $\theta_{n,i,j}^*$ and θ_0 such

that

$$\begin{aligned}\left|J_n^*(\theta_{n,i,j}^*)(i,j)-J_n^*(\theta_0)(i,j)\right| &= \left|\frac{\partial}{\partial\theta}J_n^*(\theta_{n,i,j}^{**})(i,j)\cdot(\theta_{n,i,j}^*-\theta_0)\right| \\ &\le \sup_{\theta\in\Theta_\delta}\left\|\frac{\partial}{\partial\theta}J_n^*(\theta)(i,j)\right\|\left\|\theta_{n,i,j}^*-\theta_0\right\| \\ &\le \frac{2}{n}\sum_{t=1}^n\sup_{\theta\in\Theta_\delta}\left\|\frac{\partial}{\partial\theta}\left(\frac{\partial}{\partial\theta_i}\epsilon_t(\theta)\frac{\partial}{\partial\theta_j}\epsilon_t(\theta)\right)\right\|\left\|\theta_{n,i,j}^*-\theta_0\right\| \\ &\quad+\frac{2}{n}\sum_{t=1}^n\sup_{\theta\in\Theta_\delta}\left\|\frac{\partial}{\partial\theta}\left\{\epsilon_t(\theta)\frac{\partial^2}{\partial\theta_i\partial\theta_j}\epsilon_t(\theta)\right\}\right\|\left\|\theta_{n,i,j}^*-\theta_0\right\|.\end{aligned}$$

By Theorem 2.1, the ergodic theorem, (2.59) and (2.60) imply that $\lim_{n\to\infty}|J_n^*(\theta_{n,i,j}^*)(i,j)-J_n^*(\theta_0)(i,j)|=0$ a.s.

To prove the almost-sure convergence of the first term of the right hand side of (2.58) it suffices to show that

$$\frac{1}{n}\sum_{t=1}^n\sup_{\theta\in\Theta_\delta}\left|\frac{\partial}{\partial\theta_i}\epsilon_t(\theta)\frac{\partial}{\partial\theta_j}\epsilon_t(\theta)-\frac{\partial}{\partial\theta_i}\tilde\epsilon_t(\theta)\frac{\partial}{\partial\theta_j}\tilde\epsilon_t(\theta)\right|$$

and

$$\frac{1}{n}\sum_{t=1}^n\sup_{\theta\in\Theta_\delta}\left|\epsilon_t(\theta)\frac{\partial^2}{\partial\theta_i\partial\theta_j}\epsilon_t(\theta)-\tilde\epsilon_t(\theta)\frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde\epsilon_t(\theta)\right|$$

converge almost-surely to 0. On one hand we have

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^n\sup_{\theta\in\Theta_\delta}\left|\frac{\partial}{\partial\theta_i}\epsilon_t(\theta)\frac{\partial}{\partial\theta_j}\epsilon_t(\theta)-\frac{\partial}{\partial\theta_i}\tilde\epsilon_t(\theta)\frac{\partial}{\partial\theta_j}\tilde\epsilon_t(\theta)\right| \\ &\quad\le\frac{1}{n}\sum_{t=1}^n\left\{\sup_{\theta\in\Theta_\delta}\left|\frac{\partial}{\partial\theta_i}\epsilon_t(\theta)-\frac{\partial}{\partial\theta_i}\tilde\epsilon_t(\theta)\right|\sup_{\theta\in\Theta_\delta}\left|\frac{\partial}{\partial\theta_j}\epsilon_t(\theta)\right|\right. \\ &\qquad\left.+\sup_{\theta\in\Theta_\delta}\left|\frac{\partial}{\partial\theta_i}\tilde\epsilon_t(\theta)\right|\sup_{\theta\in\Theta_\delta}\left|\frac{\partial}{\partial\theta_j}\tilde\epsilon_t(\theta)-\frac{\partial}{\partial\theta_j}\epsilon_t(\theta)\right|\right\} \\ &\quad\le\left(\frac{1}{n}\sum_{t=1}^n\left(\sup_{\theta\in\Theta_\delta}\left|\frac{\partial}{\partial\theta_i}\epsilon_t(\theta)-\frac{\partial}{\partial\theta_i}\tilde\epsilon_t(\theta)\right|\right)^2\right)^{1/2}\left(\frac{1}{n}\sum_{t=1}^n\left(\sup_{\theta\in\Theta_\delta}\left|\frac{\partial}{\partial\theta_j}\epsilon_t(\theta)\right|\right)^2\right)^{1/2} \\ &\qquad+\left(\frac{1}{n}\sum_{t=1}^n\left(\sup_{\theta\in\Theta_\delta}\left|\frac{\partial}{\partial\theta_i}\tilde\epsilon_t(\theta)\right|\right)^2\right)^{1/2}\left(\frac{1}{n}\sum_{t=1}^n\left(\sup_{\theta\in\Theta_\delta}\left|\frac{\partial}{\partial\theta_j}\tilde\epsilon_t(\theta)-\frac{\partial}{\partial\theta_j}\epsilon_t(\theta)\right|\right)^2\right)^{1/2}.\end{aligned}$$

In view of (2.25) and (2.27) it follows that

$$\mathbb{E}_{\theta_0}\left[\left(\sup_{\theta\in\Theta_\delta}\left|\frac{\partial}{\partial\theta_j}\epsilon_t(\theta)\right|\right)^2\right]\le\sup_{t\in\mathbb{Z}}\mathbb{E}_{\theta_0}|X_t|^2\left(\sum_{k_1\ge1}\log(k_1)k_1^{-1-d_1}\right)^2<\infty.$$

Similar commutations imply that

$$\mathbb{E}_{\theta_0}\left[\left(\sup_{\theta\in\Theta_\delta}\left|\frac{\partial}{\partial\theta_j}\tilde{\epsilon}_t(\theta)\right|\right)^2\right]<\infty.$$

Cesàro's Lemma, Lemma (2.7) and the ergodic theorem yield

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta_\delta}\left|\frac{\partial}{\partial\theta_i}\epsilon_t(\theta)\frac{\partial}{\partial\theta_j}\epsilon_t(\theta)-\frac{\partial}{\partial\theta_i}\tilde{\epsilon}_t(\theta)\frac{\partial}{\partial\theta_j}\tilde{\epsilon}_t(\theta)\right|\xrightarrow[n\to\infty]{\text{a.s.}}0.$$

On the other hand, one similarly may prove that

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta_\delta}\left|\epsilon_t(\theta)\frac{\partial^2}{\partial\theta_i\partial\theta_j}\epsilon_t(\theta)-\tilde{\epsilon}_t(\theta)\frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde{\epsilon}_t(\theta)\right|\xrightarrow[n\to\infty]{\text{a.s.}}0.$$

Thus

$$\sup_{\theta\in\Theta_\delta}\|J_n(\theta)-J_n^*(\theta)\|\xrightarrow[n\to\infty]{\text{a.s.}}0$$

and the lemma is proved. □

The following lemma states the existence of the matrix $I(\theta_0)$.

Lemma 2.9. *Under the assumptions of Theorem 2.2, the matrix*

$$I(\theta_0)=\lim_{n\to\infty}Var\left\{\sqrt{n}\frac{\partial}{\partial\theta}O_n(\theta_0)\right\}$$

exists.

Proof. By the stationarity of $(H_t(\theta_0))_{t\in\mathbb{Z}}$ (remind that this process is defined in (2.7)), we have

$$\begin{aligned}Var\left\{\sqrt{n}\frac{\partial}{\partial\theta}O_n(\theta_0)\right\}&=Var\left\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}H_t(\theta_0)\right\}\\&=\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}Cov\left\{H_t(\theta_0),H_s(\theta_0)\right\}\\&=\frac{1}{n}\sum_{h=-n+1}^{n-1}(n-|h|)\,Cov\left\{H_t(\theta_0),H_{t-h}(\theta_0)\right\}.\end{aligned}$$

By the dominated convergence theorem, the matrix $I(\theta_0)$ exists and is given by

$$I(\theta_0)=\sum_{h=-\infty}^{\infty}Cov\left\{H_t(\theta_0),H_{t-h}(\theta_0)\right\}$$

whenever

$$\sum_{h=-\infty}^{\infty} \| Cov\left\{H_t(\theta_0), H_{t-h}(\theta_0)\right\} \| < \infty. \tag{2.61}$$

For $s \in \mathbb{Z}$ and $1 \leq k \leq p+q+1$, we denote $H_{s,k}(\theta_0) = 2\epsilon_s(\theta_0)\frac{\partial}{\partial\theta_k}\epsilon_s(\theta_0)$ the $k-$th entry of $H_s(\theta_0)$. In view of (2.30) we have

$$\begin{aligned}
|Cov\left\{H_{t,i}(\theta_0), H_{t-h,j}(\theta_0)\right\}| &= 4\left|Cov\left(\sum_{k_1\geq 1}\dot{\lambda}_{k_1,i}(\theta_0)\,\epsilon_t\epsilon_{t-k_1}, \sum_{k_2\geq 1}\dot{\lambda}_{k_2,j}(\theta_0)\,\epsilon_{t-h}\epsilon_{t-h-k_2}\right)\right| \\
&\leq 4\sum_{k_1\geq 1}\sum_{k_2\geq 1}\left|\dot{\lambda}_{k_1,i}(\theta_0)\right|\left|\dot{\lambda}_{k_2,j}(\theta_0)\right| \left|\mathbb{E}\left[\epsilon_t\epsilon_{t-k_1}\epsilon_{t-h}\epsilon_{t-h-k_2}\right]\right| \\
&\leq \sum_{k_1\geq 1}\sum_{k_2\geq 1}\frac{K}{k_1k_2}\left|\mathbb{E}\left[\epsilon_t\epsilon_{t-k_1}\epsilon_{t-h}\epsilon_{t-h-k_2}\right]\right|
\end{aligned}$$

where we have used Lemma 2.3. It follows that

$$\begin{aligned}
\sum_{h=-\infty}^{\infty}|Cov\left\{H_{t,i}(\theta_0), H_{t-h,j}(\theta_0)\right\}| \leq &\sum_{h\in\mathbb{Z}\setminus\{0\}}\sum_{k_1\geq 1}\sum_{k_2\geq 1}\frac{K}{k_1k_2}\left|\text{cum}\left(\epsilon_t, \epsilon_{t-k_1}, \epsilon_{t-h}, \epsilon_{t-h-k_2}\right)\right| \\
&+\sum_{k_1\geq 1}\sum_{k_2\geq 1}\frac{K}{k_1k_2}\left|\mathbb{E}\left[\epsilon_t\epsilon_{t-k_1}\epsilon_t\epsilon_{t-k_2}\right]\right| .
\end{aligned}$$

Thanks to the stationarity of $(\epsilon_t)_{t\in\mathbb{Z}}$ and Assumption **(A4')** with $\tau = 4$ we deduce that

$$\begin{aligned}
\sum_{h=-\infty}^{\infty}|Cov\left\{H_{t,i}(\theta_0), H_{t-h,j}(\theta_0)\right\}| \leq &\sum_{h\in\mathbb{Z}\setminus\{0\}}\sum_{k_1\geq 1}\sum_{k_2\geq 1}\frac{K}{k_1k_2}\left|\text{cum}\left(\epsilon_0, \epsilon_{-k_1}, \epsilon_{-h}, \epsilon_{-h-k_2}\right)\right| \\
&+\sum_{k_1\geq 1}\sum_{k_2\geq 1}\frac{K}{k_1k_2}\left|\mathbb{E}\left[\epsilon_0\epsilon_{-k_1}\epsilon_0\epsilon_{-k_2}\right]\right| \\
\leq K\sum_{h,k,l\in\mathbb{Z}}&\left|\text{cum}\left(\epsilon_0, \epsilon_k, \epsilon_h, \epsilon_l\right)\right| \\
&+\sum_{k_1\geq 1}\sum_{k_2\geq 1}\frac{K}{k_1k_2}\Big(\left|\text{cum}\left(\epsilon_0, \epsilon_{-k_1}, \epsilon_0, \epsilon_{-k_2}\right)\right| \\
&\qquad + \sigma_\epsilon^2\left|\mathbb{E}\left[\epsilon_{-k_1}\epsilon_{-k_2}\right]\right|\Big) \\
\leq K\sum_{h,k,l\in\mathbb{Z}}&\left|\text{cum}\left(\epsilon_0, \epsilon_k, \epsilon_h, \epsilon_l\right)\right| + K\sigma_\epsilon^4\sum_{k_1\geq 1}\left(\frac{1}{k_1}\right)^2 \leq K
\end{aligned}$$

and we obtain the expected result. □

Lemma 2.10. *Under Assumptions of Theorem 2.2, the random vector* $\sqrt{n}(\partial/\partial\theta)O_n(\theta_0)$ *has a limiting normal distribution with mean* 0 *and covariance matrix* $I(\theta_0)$.

Proof. Observe that for any $t \in \mathbb{Z}$

$$\mathbb{E}\left[\epsilon_t \frac{\partial}{\partial\theta}\epsilon_t(\theta_0)\right] = 0 \tag{2.62}$$

because $\partial\epsilon_t(\theta_0)/\partial\theta$ belongs to the Hilbert space $\mathbf{H}_\epsilon(t-1)$, linearly generated by the family $(\epsilon_s)_{s\leq t-1}$. Therefore we have

$$\lim_{n\to\infty}\mathbb{E}\left[\sqrt{n}\frac{\partial}{\partial\theta}O_n(\theta_0)\right] = \lim_{n\to\infty}\frac{2}{\sqrt{n}}\sum_{t=1}^{n}\mathbb{E}\left[\epsilon_t\frac{\partial}{\partial\theta}\epsilon_t(\theta_0)\right] = 0.$$

For $i \geq 1$, we denote $\Lambda_i(\theta_0) = (\dot{\lambda}_{i,1}(\theta_0), \ldots, \dot{\lambda}_{i,p+q+1}(\theta_0))'$ and we introduce for $r \geq 1$

$$H_{t,r}(\theta_0) = 2\sum_{i=1}^{r}\Lambda_i(\theta_0)\,\epsilon_t\epsilon_{t-i} \quad \text{and} \quad G_{t,r}(\theta_0) = 2\sum_{i\geq r+1}\Lambda_k(\theta_0)\,\epsilon_t\epsilon_{t-i}.$$

From (2.30) we have

$$\sqrt{n}\frac{\partial}{\partial\theta}O_n(\theta_0) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}H_{t,r}(\theta_0) + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}G_{t,r}(\theta_0).$$

Since $H_{t,r}(\theta_0)$ is a function of finite number of values of the process $(\epsilon_t)_{t\in\mathbb{Z}}$, the stationary process $(H_{t,r}(\theta_0))_{t\in\mathbb{Z}}$ satisfies a mixing property (see Theorem 14.1 in Davidson [1994], p. 210) of the form **(A4)**. The central limit theorem for strongly mixing processes (see Herrndorf [1984]) implies that $(1/\sqrt{n})\sum_{t=1}^{n}H_{t,r}(\theta_0)$ has a limiting $\mathcal{N}(0, I_r(\theta_0))$ distribution with

$$I_r(\theta_0) = \lim_{n\to\infty} Var\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}H_{t,r}(\theta_0)\right).$$

Since $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}H_{t,r}(\theta_0)$ and $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}H_t(\theta_0)$ have zero expectation, we shall have

$$\lim_{r\to\infty} Var\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}H_{t,r}(\theta_0)\right) = Var\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}H_t(\theta_0)\right) = Var\left\{\sqrt{n}\frac{\partial}{\partial\theta}O_n(\theta_0)\right\},$$

as soon as

$$\lim_{r\to\infty}\mathbb{E}\left[\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}H_t(\theta_0) - \frac{1}{\sqrt{n}}\sum_{t=1}^{n}H_{t,r}(\theta_0)\right\|^2\right] = 0. \tag{2.63}$$

As a consequence we will have $\lim_{r\to\infty} I_r(\theta_0) = I(\theta_0)$. The limit in (2.63) is obtained as follows:

$$\begin{aligned}\mathbb{E}\left[\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} H_t(\theta_0) - \frac{1}{\sqrt{n}}\sum_{t=1}^{n} H_{t,r}(\theta_0)\right\|^2_{\mathbb{R}^{p+q+1}}\right] &= \mathbb{E}\left[\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} G_{t,r}(\theta_0)\right\|^2_{\mathbb{R}^{p+q+1}}\right] \\ &\leq \frac{4}{n}\sum_{l=1}^{p+q+1}\mathbb{E}\left[\left(\sum_{t=1}^{n}\sum_{k\geq r+1}\dot{\lambda}_{k,l}(\theta_0)\,\epsilon_{t-k}\epsilon_t\right)^2\right] \\ &\leq \frac{4}{n}\sum_{l=1}^{p+q+1}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{k\geq r+1}\sum_{j\geq r+1}\left|\dot{\lambda}_{k,l}(\theta_0)\right|\left|\dot{\lambda}_{j,l}(\theta_0)\right|\left|\mathbb{E}\left[\epsilon_{t-k}\epsilon_t\epsilon_{s-j}\epsilon_s\right]\right|,\end{aligned}$$

We use successively the stationarity, Lemma 2.3 and Assumption **(A4')** with $\tau = 4$ in order to obtain that

$$\begin{aligned}&\mathbb{E}\left[\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} H_t(\theta_0) - \frac{1}{\sqrt{n}}\sum_{t=1}^{n} H_{t,r}(\theta_0)\right\|^2_{\mathbb{R}^{p+q+1}}\right] \\ &\quad\leq \frac{4}{n}\sum_{l=1}^{p+q+1}\sum_{h=1-n}^{n-1}\sum_{k\geq r+1}\sum_{j\geq r+1}\left|\dot{\lambda}_{k,l}(\theta_0)\right|\left|\dot{\lambda}_{j,l}(\theta_0)\right|(n-|h|)\left|\mathbb{E}\left[\epsilon_{t-k}\epsilon_t\epsilon_{t-h-j}\epsilon_{t-h}\right]\right| \\ &\quad\leq 4\sum_{l=1}^{p+q+1}\sum_{h=-\infty}^{\infty}\sum_{k\geq r+1}\sum_{j\geq r+1}\left|\dot{\lambda}_{k,l}(\theta_0)\right|\left|\dot{\lambda}_{j,l}(\theta_0)\right|\left|\mathbb{E}\left[\epsilon_{t-k}\epsilon_t\epsilon_{t-h-j}\epsilon_{t-h}\right]\right| \\ &\quad\leq \frac{K}{(r+1)^2}\sum_{h\neq 0}\sum_{k\geq r+1}\sum_{j=-\infty}^{\infty}\left|\mathrm{cum}\left(\epsilon_0,\epsilon_{-k},\epsilon_{-j},\epsilon_{-h}\right)\right| \\ &\qquad+\frac{K}{(r+1)^2}\sum_{k\geq r+1}\sum_{j\geq r+1}\left|\mathrm{cum}\left(\epsilon_0,\epsilon_{-k},\epsilon_{-j},\epsilon_0\right)\right| + K\sigma_\epsilon^4\sum_{k\geq r+1}\left(\frac{1}{k}\right)^2\end{aligned}$$

and we obtain the convergence stated in (2.63) when $r\to\infty$.

Using Theorem 7.7.1 and Corollary 7.7.1 of Anderson (see Anderson [1971] pages 425-426), the Lemma is proved once we have, uniformly in n,

$$Var\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} G_{t,r}(\theta_0)\right) \xrightarrow[r\to\infty]{} 0\ .$$

Arguing as before we may write

$$\begin{aligned}\left[Var\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} G_{t,r}(\theta_0)\right)\right]_{ij} &= \left[Var\left(\frac{2}{\sqrt{n}}\sum_{t=1}^{n}\sum_{k\geq r+1}\Lambda_k(\theta_0)\epsilon_{t-k}\epsilon_t\right)\right]_{ij} \\ &= \frac{4}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{k_1\geq r+1}\sum_{k_2\geq r+1}\dot{\lambda}_{k_1,i}(\theta_0)\,\dot{\lambda}_{k_2,j}(\theta_0)\,\mathbb{E}\left[\epsilon_{t-k_1}\epsilon_t\epsilon_{s-k_2}\epsilon_s\right]\end{aligned}$$

$$\leq 4 \sum_{h=-\infty}^{\infty} \sum_{k_1,k_2 \geq r+1} \left| \dot{\lambda}_{k_1,i}(\theta_0)\, \dot{\lambda}_{k_2,j}(\theta_0) \right| \left| \mathbb{E}\left[\epsilon_{t-k_1} \epsilon_t \epsilon_{t-h-k_2} \epsilon_{t-h} \right] \right|.$$

and we obtain that

$$\sup_n Var\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} G_{t,r}(\theta_0) \right) \xrightarrow[r\to\infty]{} 0, \tag{2.64}$$

which completes the proof. □

No we can end this quite long proof of the asymptotic normality result.

Proof of Theorem 2.2

In view of Lemma 2.5, the equation (2.42) can be rewritten in the form:

$$o_{\mathbb{P}}(1) = \sqrt{n} \frac{\partial}{\partial \theta} O_n(\theta_0) + \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} Q_n\left(\theta^*_{n,i,j}\right) \right] \sqrt{n} \left(\hat{\theta}_n - \theta_0 \right).$$

From Lemma 2.10 $\sqrt{n}(\widehat{\theta}_n - \theta_0)(\partial^2/\partial\theta_i\partial\theta_j)Q_n(\theta^*_{n,i,j})$ converges in distribution to $\mathcal{N}(0, I(\theta_0))$. Using Lemma 2.8 and Slutsky's theorem we deduce that

$$\left(\left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} Q_n\left(\theta^*_{n,i,j}\right) \right], \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} Q_n\left(\theta^*_{n,i,j}\right) \right] \sqrt{n}(\hat{\theta}_n - \theta_0) \right)$$

converges in distribution to $(J(\theta_0), Z)$ with $\mathbb{P}_Z = \mathcal{N}(0, I)$. Consider now the function $h : \mathbb{R}^{(p+q+1)\times(p+q+1)} \times \mathbb{R}^{p+q+1} \to \mathbb{R}^{p+q+1}$ that maps (A, X) to $A^{-1}X$. If D_h denotes the set of discontinuity points of h, we have $\mathbb{P}((J(\theta_0), Z) \in D_h) = 0$. By the continuous mapping theorem

$$h\left(\left[(\partial^2/\partial\theta_i\partial\theta_j) Q_n(\theta^*_{n,i,j}) \right], \left[(\partial^2/\partial\theta_i\partial\theta_j) Q_n(\theta^*_{n,i,j}) \right] \sqrt{n}(\hat{\theta}_n - \theta_0) \right)$$

converges in distribution to $h(J(\theta_0), Z)$ and thus $\sqrt{n}(\hat{\theta}_n - \theta_0)$ has a limiting normal distribution with mean 0 and covariance matrix $J^{-1}(\theta_0) I(\theta_0) J^{-1}(\theta_0)$. The proof of Theorem 2.2 is then completed.

2.6.4 Proof of the convergence of the variance matrix estimator

We show in this section the convergence in probability of $\hat{\Omega} := \hat{J}_n^{-1} \hat{I}_n^{SP} \hat{J}_n^{-1}$ to Ω, which is an adaptation of the arguments used in Boubacar Mainassara et al. [2012].

Using the same approach as that followed in Lemma 2.8, we show that $\hat{J}_n$ converges almost surely to J. We give below the proof of the convergence in probability of the estimator $\hat{I}_n^{SP}$, obtained using the approach of the spectral density, towards I.

We recall that the matrix norm used is given by $\|A\| = \sup_{\|x\|\leq 1} \|Ax\| = \rho^{1/2}(A'A)$, when A is a $\mathbb{R}^{k_1\times k_2}$ matrix, $\|x\|^2 = x'x$ is the Euclidean norm of the vector $x \in \mathbb{R}^{k_2}$, and $\rho(\cdot)$ denotes the spectral radius. This norm satisfies

$$\|A\|^2 \leq \sum_{i=1}^{k_1}\sum_{j=1}^{k_2} a_{i,j}^2, \tag{2.65}$$

with $a_{i,j}$ the entries of $A \in \mathbb{R}^{k_1\times k_2}$. The choice of the norm is crucial for the following results to hold (with e.g. the Euclidean norm, this result is not valid).

We denote

$$\Sigma_{H,\underline{H}_r} = \mathbb{E}H_t\underline{H}_{r,t}', \quad \Sigma_H = \mathbb{E}H_tH_t', \quad \Sigma_{\underline{H}_r} = \mathbb{E}\underline{H}_{r,t}\underline{H}_{r,t}'$$

where $H_t := H_t(\theta_0)$ is definied in (2.7) and $\underline{H}_{r,t} = (H_{t-1}', \ldots, H_{t-r}')'$. For any $n \geq 1$, we have

$$\begin{aligned}
\hat{I}_n^{\mathrm{SP}} &= \hat{\Phi}_r^{-1}(1)\hat{\Sigma}_{\hat{u}_r}\hat{\Phi}_r'^{-1}(1) \\
&= \left(\hat{\Phi}_r^{-1}(1) - \Phi^{-1}(1)\right)\hat{\Sigma}_{\hat{u}_r}\hat{\Phi}_r'^{-1}(1) + \Phi^{-1}(1)\left(\hat{\Sigma}_{\hat{u}_r} - \Sigma_u\right)\hat{\Phi}_r'^{-1}(1) \\
&\qquad + \Phi^{-1}(1)\Sigma_u\left(\hat{\Phi}_r'^{-1}(1) - \Phi'^{-1}(1)\right) + \Phi^{-1}(1)\Sigma_u\Phi'^{-1}(1).
\end{aligned}$$

We then obtain

$$\begin{aligned}
\left\|\hat{I}_n^{\mathrm{SP}} - I(\theta_0)\right\| &\leq \left\|\hat{\Phi}_r^{-1}(1) - \Phi^{-1}(1)\right\|\left\|\hat{\Sigma}_{\hat{u}_r}\right\|\left\|\hat{\Phi}_r'^{-1}(1)\right\| + \left\|\Phi^{-1}(1)\right\|\left\|\hat{\Sigma}_{\hat{u}_r} - \Sigma_u\right\|\left\|\hat{\Phi}_r'^{-1}(1)\right\| \\
&\qquad + \left\|\Phi^{-1}(1)\right\|\|\Sigma_u\|\left\|\hat{\Phi}_r'^{-1}(1) - \Phi'^{-1}(1)\right\| \\
&\leq \left\|\hat{\Phi}_r^{-1}(1) - \Phi^{-1}(1)\right\|\left(\left\|\hat{\Sigma}_{\hat{u}_r}\right\|\left\|\hat{\Phi}_r'^{-1}(1)\right\| + \left\|\Phi^{-1}(1)\right\|\|\Sigma_u\|\right) \\
&\qquad + \left\|\hat{\Sigma}_{\hat{u}_r} - \Sigma_u\right\|\left\|\hat{\Phi}_r'^{-1}(1)\right\|\left\|\Phi^{-1}(1)\right\| \\
&\leq \left\|\hat{\Phi}_r^{-1}(1)\right\|\left\|\Phi(1) - \hat{\Phi}_r(1)\right\|\left\|\Phi^{-1}(1)\right\|\left(\left\|\hat{\Sigma}_{\hat{u}_r}\right\|\left\|\hat{\Phi}_r'^{-1}(1)\right\| + \left\|\Phi^{-1}(1)\right\|\|\Sigma_u\|\right) \\
&\qquad + \left\|\hat{\Sigma}_{\hat{u}_r} - \Sigma_u\right\|\left\|\hat{\Phi}_r'^{-1}(1)\right\|\left\|\Phi^{-1}(1)\right\|.
\end{aligned} \tag{2.66}$$

In view of (2.66), to prove the convergence in probability of $\hat{I}_n^{\mathrm{SP}}$ to $I(\theta_0)$, it suffices to show that $\hat{\Phi}_r(1) \to \Phi(1)$ and $\hat{\Sigma}_{\hat{u}_r} \to \Sigma_u$ in probability. Let the $r\times 1$ vector $1_r = (1,\ldots,1)'$ and the $r(p+q+1)\times(p+q+1)$ matrix $\mathbf{E}_r = I_{p+q+1}\otimes 1_r$, where $\otimes$ denotes the matrix Kronecker product and I_m the $m\times m$ identity matrix. Write $\underline{\Phi}_r^* = (\Phi_1,\ldots,\Phi_r)$ where the Φ_i's are defined

by (2.8). We have

$$\begin{aligned}\left\|\hat{\Phi}_r(1)-\Phi(1)\right\| &= \left\|\sum_{k=1}^{r}\hat{\Phi}_{r,k}-\sum_{k=1}^{r}\Phi_{r,k}+\sum_{k=1}^{r}\Phi_{r,k}-\sum_{k=1}^{\infty}\Phi_k\right\| \\ &\leq \left\|\sum_{k=1}^{r}\left(\hat{\Phi}_{r,k}-\Phi_{r,k}\right)\right\|+\left\|\sum_{k=1}^{r}\left(\Phi_{r,k}-\Phi_k\right)\right\|+\left\|\sum_{k=r+1}^{\infty}\Phi_k\right\| \\ &\leq \left\|\left(\underline{\hat{\Phi}}_r-\underline{\Phi}_r\right)\mathbf{E}_r\right\|+\left\|\left(\underline{\Phi}_r^*-\underline{\Phi}_r\right)\mathbf{E}_r\right\|+\left\|\sum_{k=r+1}^{\infty}\Phi_k\right\| \\ &\leq \sqrt{p+q+1}\sqrt{r}\left(\left\|\underline{\hat{\Phi}}_r-\underline{\Phi}_r\right\|+\left\|\underline{\Phi}_r^*-\underline{\Phi}_r\right\|\right)+\left\|\sum_{k=r+1}^{\infty}\Phi_k\right\|. \end{aligned} \tag{2.67}$$

Under the assumptions of Theorem 2.3 (see (2.9)) we have

$$\left\|\sum_{k=r+1}^{\infty}\Phi_k\right\| \leq \sum_{k=r+1}^{\infty}\|\Phi_k\| \leq K\sum_{k=r+1}^{\infty}\rho^k \xrightarrow[n\to\infty]{\text{a.s.}} 0.$$

Therefore it is enough to show that $\sqrt{r}\|\underline{\hat{\Phi}}_r-\underline{\Phi}_r\|$ and $\sqrt{r}\|\underline{\Phi}_r^*-\underline{\Phi}_r\|$ converge in probability towards 0 in order to obtain the convergence in probability of $\hat{\Phi}_r(1)$ towards $\Phi(1)$. From (2.10) we have

$$H_t(\theta_0)=\underline{\Phi}_r\underline{H}_{r,t}(\theta_0)+u_{r,t}, \tag{2.68}$$

and thus

$$\Sigma_{u_r}=\mathrm{Var}(u_{r,t})=\mathbb{E}\left[u_{r,t}\left(H_t(\theta_0)-\underline{\Phi}_r\underline{H}_{r,t}(\theta_0)\right)'\right].$$

The vector $u_{r,t}$ is orthogonal to $\underline{H}_{r,t}(\theta_0)$. Therefore

$$\begin{aligned}\mathrm{Var}(u_{r,t}) &= \mathbb{E}\left[\left(H_t(\theta_0)-\underline{\Phi}_r\underline{H}_{r,t}(\theta_0)\right)H_t'(\theta_0)\right] \\ &= \Sigma_H-\underline{\Phi}_r\Sigma'_{H,\underline{H}_r}.\end{aligned}$$

Consequently the least squares estimator of Σ_{u_r} can be rewritten in the form:

$$\hat{\Sigma}_{\hat{u}_r}=\hat{\Sigma}_{\hat{H}}-\underline{\hat{\Phi}}_r\hat{\Sigma}'_{\hat{H},\underline{\hat{H}}_r}, \tag{2.69}$$

where

$$\hat{\Sigma}_{\hat{H}}=\frac{1}{n}\sum_{t=1}^{n}\hat{H}_t\hat{H}_t'. \tag{2.70}$$

Similar arguments combined with (2.8) yield

$$\begin{aligned}\Sigma_u = \mathbb{E}\left[u_t u_t^{'}\right] &= \mathbb{E}\left[u_t H_t^{'}(\theta_0)\right] \\ &= \mathbb{E}\left[H_t(\theta_0)H_t^{'}(\theta_0)\right] - \sum_{k=1}^{r}\Phi_k\mathbb{E}\left[H_{t-k}(\theta_0)H_t^{'}(\theta_0)\right] - \sum_{k=r+1}^{\infty}\Phi_k\mathbb{E}\left[H_{t-k}(\theta_0)H_t^{'}(\theta_0)\right] \\ &= \Sigma_H - \underline{\Phi}_r^{*}\Sigma_{H,\underline{H}_r}^{'} - \sum_{k=r+1}^{\infty}\Phi_k\mathbb{E}\left[H_{t-k}(\theta_0)H_t^{'}(\theta_0)\right].\end{aligned}$$

By (2.69) we obtain

$$\begin{aligned}\left\|\hat{\Sigma}_{\hat{u}_r} - \Sigma_u\right\| &= \left\|\hat{\Sigma}_{\hat{H}} - \underline{\hat{\Phi}}_r\hat{\Sigma}_{\hat{H},\underline{\hat{H}}_r}^{'} - \Sigma_H + \underline{\Phi}_r^{*}\Sigma_{H,\underline{H}_r}^{'} + \sum_{k=r+1}^{\infty}\Phi_k\mathbb{E}\left[H_{t-k}(\theta_0)H_t^{'}(\theta_0)\right]\right\| \\ &= \left\|\hat{\Sigma}_{\hat{H}} - \Sigma_H - \left(\underline{\hat{\Phi}}_r - \underline{\Phi}_r^{*}\right)\hat{\Sigma}_{\hat{H},\underline{\hat{H}}_r}^{'} - \underline{\Phi}_r^{*}\left(\hat{\Sigma}_{\hat{H},\underline{\hat{H}}_r}^{'} - \Sigma_{H,\underline{H}_r}^{'}\right)\right. \\ &\qquad \left. + \sum_{k=r+1}^{\infty}\Phi_k\mathbb{E}\left[H_{t-k}(\theta_0)H_t^{'}(\theta_0)\right]\right\| \\ &\leq \left\|\hat{\Sigma}_{\hat{H}} - \Sigma_H\right\| + \left\|\left(\underline{\hat{\Phi}}_r - \underline{\Phi}_r^{*}\right)\left(\hat{\Sigma}_{\hat{H},\underline{\hat{H}}_r}^{'} - \Sigma_{H,\underline{H}_r}^{'}\right)\right\| + \left\|\left(\underline{\hat{\Phi}}_r - \underline{\Phi}_r^{*}\right)\Sigma_{H,\underline{H}_r}^{'}\right\| \\ &\quad + \left\|\underline{\Phi}_r^{*}\left(\hat{\Sigma}_{\hat{H},\underline{\hat{H}}_r}^{'} - \Sigma_{H,\underline{H}_r}^{'}\right)\right\| + \left\|\sum_{k=r+1}^{\infty}\Phi_k\mathbb{E}\left[H_{t-k}(\theta_0)H_t^{'}(\theta_0)\right]\right\|. \qquad (2.71)\end{aligned}$$

From Lemma 2.9 and hypotheses of Theorem 2.3 (see (2.9)) we deduce that

$$\begin{aligned}\left\|\sum_{k=r+1}^{\infty}\Phi_k\mathbb{E}\left[H_{t-k}(\theta_0)H_t^{'}(\theta_0)\right]\right\| &\leq \sum_{k=r+1}^{\infty}\|\Phi_k\|\left\|\mathbb{E}\left[H_{t-k}(\theta_0)H_t^{'}(\theta_0)\right]\right\| \\ &\leq K\sum_{k=r+1}^{\infty}\rho^k \xrightarrow[n\to\infty]{\text{a.s.}} 0.\end{aligned}$$

Observe also that

$$\|\underline{\Phi}_r^{*}\|^2 \leq \sum_{k\geq 1}\mathrm{Tr}\left(\Phi_k\Phi_k^{'}\right) < \infty.$$

Therefore the convergence $\hat{\Sigma}_{\hat{u}_r}$ to Σ_u will be a consequence of the four following properties:

- $\|\hat{\Sigma}_{\hat{H}} - \Sigma_H\| = o_{\mathbb{P}}(1)$,
- $\mathbb{P} - \lim_{n\to\infty}\|\underline{\hat{\Phi}}_r - \underline{\Phi}_r^{*}\| = 0$,
- $\mathbb{P} - \lim_{n\to\infty}\|\hat{\Sigma}_{\hat{H},\underline{\hat{H}}_r}^{'} - \Sigma_{H,\underline{H}_r}^{'}\| = 0$ and
- $\|\Sigma_{H,\underline{H}_r}^{'}\| = O(1)$.

The above properties will be proved thanks several lemmas that are stated and proved hereafter. This ends the proof of Theorem 2.3. For this, consider the following lemmas:

Lemma 2.11. *Under the assumptions of Theorem 2.3, we have*

$$\sup_{r\geq 1}\max\left\{\left\|\Sigma_{H,\underline{H}_r}\right\|,\left\|\Sigma_{\underline{H}_r}\right\|,\left\|\Sigma_{\underline{H}_r}^{-1}\right\|\right\}<\infty.$$

Proof. See Lemma 1 in the supplementary material of Boubacar Mainassara et al. [2012]. □

Lemma 2.12. *Under the assumptions of Theorem 2.3 there exists a finite positive constant K such that, for $1\leq r_1,r_2\leq r$ and $1\leq m_1,m_2\leq p+q+1$ we have*

$$\sup_{t\in\mathbb{Z}}\sum_{h=-\infty}^{\infty}\left|\mathrm{Cov}\left\{H_{t-r_1,m_1}(\theta_0)H_{t-r_2,m_2}(\theta_0),H_{t-r_1-h,m_1}(\theta_0)H_{t-r_2-h,m_2}(\theta_0)\right\}\right|<K.$$

Proof. We denote in the sequel by $\dot{\lambda}_{j,k}$ the coefficient $\dot{\lambda}_{j,k}(\theta_0)$ defined in (2.31).

Using the fact that the process $(H_t(\theta_0))_{t\in\mathbb{Z}}$ is centered and taking into consideration the strict stationarity of $(\epsilon_t)_{t\in\mathbb{Z}}$ we obtain that for any $t\in\mathbb{Z}$

$$\begin{aligned}
&\sum_{h=-\infty}^{\infty}\left|\mathrm{Cov}\big(H_{t-r_1,m_1}(\theta_0)H_{t-r_2,m_2}(\theta_0),H_{t-r_1-h,m_1}(\theta_0)H_{t-r_2-h,m_2}(\theta_0)\big)\right|\\
&\quad=\sum_{h=-\infty}^{\infty}\Big|\mathbb{E}\left[H_{t-r_1,m_1}(\theta_0)H_{t-r_2,m_2}(\theta_0)H_{t-r_1-h,m_1}(\theta_0)H_{t-r_2-h,m_2}(\theta_0)\right]\\
&\qquad\qquad-\mathbb{E}\left[H_{t-r_1,m_1}(\theta_0)H_{t-r_2,m_2}(\theta_0)\right]\mathbb{E}\left[H_{t-r_1-h,m_1}(\theta_0)H_{t-r_2-h,m_2}(\theta_0)\right]\Big|\\
&\quad\leq\sum_{h=-\infty}^{\infty}\left|\mathrm{cum}\big(H_{t-r_1,m_1}(\theta_0),H_{t-r_2,m_2}(\theta_0),H_{t-r_1-h,m_1}(\theta_0),H_{t-r_2-h,m_2}(\theta_0)\big)\right|\\
&\qquad\qquad+\sum_{h=-\infty}^{\infty}\left|\mathbb{E}\left[H_{t-r_1,m_1}(\theta_0)H_{t-r_1-h,m_1}(\theta_0)\right]\right|\left|\mathbb{E}\left[H_{t-r_2,m_2}(\theta_0)H_{t-r_2-h,m_2}(\theta_0)\right]\right|\\
&\qquad\qquad+\sum_{h=-\infty}^{\infty}\left|\mathbb{E}\left[H_{t-r_1,m_1}(\theta_0)H_{t-r_2-h,m_2}(\theta_0)\right]\right|\left|\mathbb{E}\left[H_{t-r_2,m_2}(\theta_0)H_{t-r_1-h,m_1}(\theta_0)\right]\right|\\
&\quad\leq\sum_{h=-\infty}^{\infty}\sum_{i_1,j_1,k_1,\ell_1\geq 1}\left|\dot{\lambda}_{i_1,m_1}\dot{\lambda}_{j_1,m_2}\dot{\lambda}_{k_1,m_1}\dot{\lambda}_{\ell_1,m_2}\right|\\
&\qquad\times\left|\mathrm{cum}\left(\epsilon_0\epsilon_{-i_1},\epsilon_{r_1-r_2}\epsilon_{r_1-r_2-j_1},\epsilon_{-h}\epsilon_{-h-k_1},\epsilon_{r_1-r_2-h}\epsilon_{r_1-r_2-h-\ell_1}\right)\right|\\
&\qquad\qquad+T^{(1)}_{r1,m_1,r_2,m_2}+T^{(2)}_{r1,m_1,r_2,m_2},
\end{aligned}$$

where

$$T^{(1)}_{r1,m_1,r_2,m_2}=\sum_{h=-\infty}^{\infty}\left|\mathbb{E}\left[H_{t-r_1,m_1}(\theta_0)H_{t-r_1-h,m_1}(\theta_0)\right]\right|\left|\mathbb{E}\left[H_{t-r_2,m_2}(\theta_0)H_{t-r_2-h,m_2}(\theta_0)\right]\right|$$

and

$$T^{(2)}_{r1,m_1,r_2,m_2} = \sum_{h=-\infty}^{\infty} \left|\mathbb{E}\left[H_{t-r_1,m_1}(\theta_0)H_{t-r_2-h,m_2}(\theta_0)\right]\right| \left|\mathbb{E}\left[H_{t-r_2,m_2}(\theta_0)H_{t-r_1-h,m_1}(\theta_0)\right]\right|.$$

Thanks to Lemma 2.3 one may use the product theorem for the joint cumulants (Brillinger [1981]) as in the proof of Lemma A.3. in Shao [2011] in order to obtain that

$$\sum_{h=-\infty}^{\infty} \sum_{i_1,j_1,k_1,\ell_1\geq 1} \left|\dot{\lambda}_{i_1,m_1}\dot{\lambda}_{j_1,m_2}\dot{\lambda}_{k_1,m_1}\dot{\lambda}_{\ell_1,m_2}\right| \left|\mathrm{cum}\left(\epsilon_0\epsilon_{-i_1}, \epsilon_{r_1-r_2}\epsilon_{r_1-r_2-j_1}, \epsilon_{-h}\epsilon_{-h-k_1}, \epsilon_{r_1-r_2-h}\epsilon_{r_1-r_2-h-\ell_1}\right)\right|$$
$$< \infty.$$

where we have used the absolute summability of the k-th ($k = 2, \ldots, 8$) cumulants assumed in **(A4')** with $\tau = 8$.

Observe now that

$$\begin{aligned} T^{(1)}_{r1,m_1,r_2,m_2} &= \sum_{h=-\infty}^{\infty} \left|\mathbb{E}\left[H_{t-r_1,m_1}(\theta_0)H_{t-r_1-h,m_1}(\theta_0)\right]\right| \left|\mathbb{E}\left[H_{t-r_2,m_2}(\theta_0)H_{t-r_2-h,m_2}(\theta_0)\right]\right| \\ &\leq \sup_{h\in\mathbb{Z}} \left|\mathbb{E}\left[H_{t-r_1,m_1}(\theta_0)H_{t-r_1-h,m_1}(\theta_0)\right]\right| \sum_{h=-\infty}^{\infty} \left|\mathbb{E}\left[H_{t-r_2,m_2}(\theta_0)H_{t-r_2-h,m_2}(\theta_0)\right]\right|. \end{aligned}$$

For any $h \in \mathbb{Z}$, from (2.30) we have

$$\begin{aligned} \left|\mathbb{E}\left[H_{t-r_1,m_1}(\theta_0)H_{t-r_1-h,m_1}(\theta_0)\right]\right| &\leq \sum_{i,j\geq 1} \left|\dot{\lambda}_{i,m_1}\right| \left|\dot{\lambda}_{j,m_1}\right| \left|\mathrm{cum}\left(\epsilon_0, \epsilon_{-i}, \epsilon_{-h}, \epsilon_{-h-j}\right)\right| \\ &\quad + \sum_{i,j\geq 1} \left|\dot{\lambda}_{i,m_1}\right| \left|\dot{\lambda}_{j,m_1}\right| \Big\{ \left|\mathbb{E}\left[\epsilon_0\epsilon_{-i}\right]\mathbb{E}\left[\epsilon_{-h}\epsilon_{-h-j}\right]\right| \\ &\qquad + \left|\mathbb{E}\left[\epsilon_0\epsilon_{-h}\right]\mathbb{E}\left[\epsilon_{-i}\epsilon_{-h-j}\right]\right| + \left|\mathbb{E}\left[\epsilon_0\epsilon_{-h-j}\right]\mathbb{E}\left[\epsilon_{-i}\epsilon_{-h}\right]\right| \Big\} \\ &\leq \sum_{i,j\geq 1} \left|\mathrm{cum}\left(\epsilon_0, \epsilon_{-i}, \epsilon_{-h}, \epsilon_{-h-j}\right)\right| + \sigma_\epsilon^4 \sum_{i\geq 1} \left|\dot{\lambda}_{i,m_1}\right|^2. \end{aligned}$$

Under Assumption **(A4')** with $\tau = 4$ and in view of Lemma 2.3 we may write that

$$\sup_{h\in\mathbb{Z}} \left|\mathbb{E}\left[H_{t-r_1,m_1}(\theta_0)H_{t-r_1-h,m_1}(\theta_0)\right]\right| \leq \sup_{h\in\mathbb{Z}} \sum_{i,j\geq 1} \left|\mathrm{cum}\left(\epsilon_0, \epsilon_{-i}, \epsilon_{-h}, \epsilon_{-h-j}\right)\right| + \sigma_\epsilon^4 \sum_{i\geq 1} \left|\dot{\lambda}_{i,m_1}\right|^2 < \infty.$$

Similarly, we obtain

$$\begin{aligned} \sum_{h=-\infty}^{\infty} \left|\mathbb{E}\left[H_{t-r_2,m_2}(\theta_0)H_{t-r_2-h,m_2}(\theta_0)\right]\right| &\leq \sum_{h=-\infty}^{\infty} \sum_{i,j\geq 1} \left|\mathrm{cum}\left(\epsilon_0, \epsilon_{-i}, \epsilon_{-h}, \epsilon_{-h-j}\right)\right| + \sigma_\epsilon^4 \sum_{i\geq 1} \left|\dot{\lambda}_{i,m_1}\right|^2 \\ &< \infty. \end{aligned}$$

Consequently $T^{(1)}_{r1,m_1,r_2,m_2} < \infty$ and the same approach yields that $T^{(2)}_{r1,m_1,r_2,m_2} < \infty$ and the lemma is proved. □

Let $\hat{\Sigma}_{\underline{H}_r}$, $\hat{\Sigma}_H$ and $\hat{\Sigma}_{H,\underline{H}_r}$ be the matrices obtained by replacing $\hat{H}_t$ by $H_t(\theta_0)$ in $\hat{\Sigma}_{\underline{\hat{H}}_r}$, $\hat{\Sigma}_{\hat{H}}$ and $\hat{\Sigma}_{\hat{H},\underline{\hat{H}}_r}$.

Lemma 2.13. *Under the assumptions of Theorem 2.3, $\sqrt{r}\|\hat{\Sigma}_{\underline{H}_r} - \Sigma_{\underline{H}_r}\|$, $\sqrt{r}\|\hat{\Sigma}_{H,\underline{H}_r} - \Sigma_{H,\underline{H}_r}\|$ and $\sqrt{r}\|\hat{\Sigma}_H - \Sigma_H\|$ tend to zero in probability as $n \to \infty$ when $r = o(n^{1/3})$.*

Proof. For $1 \leq m_1, m_2 \leq p+q+1$ and $1 \leq r_1, r_2 \leq r$, the $(\{(r_1-1)(p+q+1)+m_1\}, \{(r_2-1)(p+q+1)+m_2\})$−th element of $\hat{\Sigma}_{\underline{H}_r}$ is given by:

$$\frac{1}{n}\sum_{t=1}^{n} H_{t-r_1,m_1}(\theta_0)H_{t-r_2,m_2}(\theta_0).$$

For all $\beta > 0$, we use (2.65) and we obtain

$$\begin{aligned}
\mathbb{P}\left(\sqrt{r}\left\|\hat{\Sigma}_{\underline{H}_r} - \Sigma_{\underline{H}_r}\right\| \geq \beta\right) &\leq \frac{r}{\beta^2}\mathbb{E}\left\|\hat{\Sigma}_{\underline{H}_r} - \Sigma_{\underline{H}_r}\right\|^2 \\
&\leq \frac{r}{\beta^2}\mathbb{E}\left\|\frac{1}{n}\sum_{t=1}^{n}\underline{H}_{r,t}\underline{H}'_{r,t} - \mathbb{E}\left[\underline{H}_{r,t}\underline{H}'_{r,t}\right]\right\|^2 \\
&\leq \frac{r}{\beta^2}\sum_{r_1=1}^{r}\sum_{r_2=1}^{r}\sum_{m_1=1}^{p+q+1}\sum_{m_2=1}^{p+q+1}\mathbb{E}\left(\frac{1}{n}\sum_{t=1}^{n}H_{t-r_1,m_1}(\theta_0)H_{t-r_2,m_2}(\theta_0)\right. \\
&\qquad \left. - \mathbb{E}\left[H_{t-r_1,m_1}(\theta_0)H_{t-r_2,m_2}(\theta_0)\right]\right)^2.
\end{aligned}$$

The stationarity of the process $(H_{t-r_1,m_1}(\theta_0)H_{t-r_2,m_2}(\theta_0))_t$ and Lemma 2.12 imply

$$\begin{aligned}
&\mathbb{P}\left(\sqrt{r}\left\|\hat{\Sigma}_{\underline{H}_r} - \Sigma_{\underline{H}_r}\right\| \geq \beta\right) \\
&\leq \frac{r}{\beta^2}\sum_{r_1=1}^{r}\sum_{r_2=1}^{r}\sum_{m_1=1}^{p+q+1}\sum_{m_2=1}^{p+q+1}\mathrm{Var}\left(\frac{1}{n}\sum_{t=1}^{n}H_{t-r_1,m_1}(\theta_0)H_{t-r_2,m_2}(\theta_0)\right) \\
&\leq \frac{r}{(n\beta)^2}\sum_{r_1=1}^{r}\sum_{r_2=1}^{r}\sum_{m_1=1}^{p+q+1}\sum_{m_2=1}^{p+q+1}\sum_{t=1}^{n}\sum_{s=1}^{n}\mathrm{Cov}\left(H_{t-r_1,m_1}(\theta_0)H_{t-r_2,m_2}(\theta_0), H_{s-r_1,m_1}(\theta_0)H_{s-r_2,m_2}(\theta_0)\right) \\
&\leq \frac{r}{(n\beta)^2}\sum_{r_1=1}^{r}\sum_{r_2=1}^{r}\sum_{m_1=1}^{p+q+1}\sum_{m_2=1}^{p+q+1}\sum_{h=1-n}^{n-1}(n-|h|) \\
&\qquad \times \mathrm{Cov}\left(H_{t-r_1,m_1}(\theta_0)H_{t-r_2,m_2}(\theta_0), H_{t-h-r_1,m_1}(\theta_0)H_{t-h-r_2,m_2}(\theta_0)\right) \\
&\leq \frac{r}{n\beta^2}\sum_{r_1=1}^{r}\sum_{r_2=1}^{r}\sum_{m_1=1}^{p+q+1}\sum_{m_2=1}^{p+q+1}
\end{aligned}$$

$$\sup_{t\in\mathbb{Z}} \sum_{h=-\infty}^{\infty} |\mathrm{Cov}\left(H_{t-r_1,m_1}(\theta_0)H_{t-r_2,m_2}(\theta_0), H_{t-h-r_1,m_1}(\theta_0)H_{t-h-r_2,m_2}(\theta_0)\right)|$$
$$\leq \frac{C(p+q+1)^2 r^3}{n\beta^2}.$$

Consequently we have

$$\begin{aligned}\mathbb{E}\left[r\left\|\hat{\Sigma}_H - \Sigma_H\right\|^2\right] &\leq \mathbb{E}\left[r\left\|\hat{\Sigma}_{H,\underline{H}_r} - \Sigma_{H,\underline{H}_r}\right\|^2\right] \\ &\leq \mathbb{E}\left[r\left\|\hat{\Sigma}_{\underline{H}_r} - \Sigma_{\underline{H}_r}\right\|^2\right] \\ &\leq \frac{C(p+q+1)^2 r^3}{n} \xrightarrow[n\to\infty]{} 0\end{aligned}$$

when $r = o(n^{1/3})$. The conclusion follows. □

We show in the following lemma that the previous lemma remains valid when we replace $H_t(\theta_0)$ by $\hat{H}_t$.

Lemma 2.14. *Under the assumptions of Theorem 2.3, $\sqrt{r}\|\hat{\Sigma}_{\underline{\hat{H}}_r} - \Sigma_{\underline{H}_r}\|$, $\sqrt{r}\|\hat{\Sigma}_{\hat{H},\underline{\hat{H}}_r} - \Sigma_{H,\underline{H}_r}\|$ and $\sqrt{r}\|\hat{\Sigma}_{\hat{H}} - \Sigma_H\|$ tend to zero in probability as $n \to \infty$ when $r = o(n^{(1-2(d_2-d_1))/5})$.*

Proof. As mentioned in the end of the proof of the previous lemma, we only have to deal with the term $\sqrt{r}\|\hat{\Sigma}_{\underline{\hat{H}}_r} - \Sigma_{\underline{H}_r}\|$.

We denote $\hat{\Sigma}_{\underline{H}_{r,n}}$ the matrix obtained by replacing $\tilde{\epsilon}_t(\hat{\theta}_n)$ by $\epsilon_t(\hat{\theta}_n)$ in $\hat{\Sigma}_{\underline{\hat{H}}_r}$. We have

$$\sqrt{r}\left\|\hat{\Sigma}_{\underline{\hat{H}}_r} - \Sigma_{\underline{H}_r}\right\| \leq \sqrt{r}\left\|\hat{\Sigma}_{\underline{\hat{H}}_r} - \hat{\Sigma}_{\underline{H}_{r,n}}\right\| + \sqrt{r}\left\|\hat{\Sigma}_{\underline{H}_{r,n}} - \hat{\Sigma}_{\underline{H}_r}\right\| + \sqrt{r}\left\|\hat{\Sigma}_{\underline{H}_r} - \Sigma_{\underline{H}_r}\right\|.$$

By Lemma 2.13, the term $\sqrt{r}\|\hat{\Sigma}_{\underline{H}_r} - \Sigma_{\underline{H}_r}\|$ converges in probability. The lemma will be proved as soon as we show that

$$\sqrt{r}\left\|\hat{\Sigma}_{\underline{\hat{H}}_r} - \hat{\Sigma}_{\underline{H}_{r,n}}\right\| = o_{\mathbb{P}}(1) \quad \text{and} \tag{2.72}$$
$$\sqrt{r}\left\|\hat{\Sigma}_{\underline{H}_{r,n}} - \hat{\Sigma}_{\underline{H}_r}\right\| = o_{\mathbb{P}}(1), \tag{2.73}$$

when $r = o(n^{(1-2(d_2-d_1))/5})$. This is done in two separate steps.

Step 1: proof of (2.72).
For all $\beta > 0$, we have

$$\begin{aligned}
&\mathbb{P}\left(\sqrt{r}\left\|\hat{\Sigma}_{\underline{\hat{H}}_r} - \hat{\Sigma}_{\underline{H}_{r,n}}\right\| \geq \beta\right) \\
&\leq \frac{\sqrt{r}}{\beta}\mathbb{E}\left\|\hat{\Sigma}_{\underline{\hat{H}}_r} - \hat{\Sigma}_{\underline{H}_{r,n}}\right\| \\
&\leq \frac{\sqrt{r}}{\beta}\mathbb{E}\left\|\frac{1}{n}\sum_{t=1}^{n}\underline{\hat{H}}_{r,t}\underline{\hat{H}}'_{r,t} - \frac{1}{n}\sum_{t=1}^{n}\underline{H}^{(n)}_{r,t}\underline{H}^{(n)'}_{r,t}\right\| \\
&\leq \frac{K\sqrt{r}}{\beta}\sum_{r_1=1}^{r}\sum_{r_2=1}^{r}\sum_{m_1=1}^{p+q+1}\sum_{m_2=1}^{p+q+1}\mathbb{E}\left|\frac{1}{n}\sum_{t=1}^{n}\hat{H}_{t-r_1,m_1}\hat{H}_{t-r_2,m_2} - \frac{1}{n}\sum_{t=1}^{n}H^{(n)}_{t-r_1,m_1}H^{(n)}_{t-r_2,m_2}\right|,
\end{aligned}$$

where

$$H^{(n)}_{t,m} = 2\epsilon_t(\hat{\theta}_n)\frac{\partial}{\partial\theta_m}\epsilon_t(\hat{\theta}_n) \quad \text{and} \quad \underline{H}^{(n)}_{r,t} = \left(H^{(n)'}_{t-1},\ldots,H^{(n)'}_{t-r}\right)'.$$

It is follow that

$$\begin{aligned}
\mathbb{P}\left(\sqrt{r}\left\|\hat{\Sigma}_{\underline{\hat{H}}_r} - \hat{\Sigma}_{\underline{H}_{r,n}}\right\| \geq \beta\right) \leq \frac{4K\sqrt{r}}{n\beta}\sum_{r_1=1}^{r}\sum_{r_2=1}^{r}\sum_{m_1=1}^{p+q+1}\sum_{m_2=1}^{p+q+1} \\
\mathbb{E}\left|\sum_{t=1}^{n}\tilde{\epsilon}_{t-r_1}(\hat{\theta}_n)\frac{\partial}{\partial\theta_{m_1}}\tilde{\epsilon}_{t-r_1}(\hat{\theta}_n)\tilde{\epsilon}_{t-r_2}(\hat{\theta}_n)\frac{\partial}{\partial\theta_{m_2}}\tilde{\epsilon}_{t-r_2}(\hat{\theta}_n)\right. \\
\left. - \epsilon_{t-r_1}(\hat{\theta}_n)\frac{\partial}{\partial\theta_{m_1}}\epsilon_{t-r_1}(\hat{\theta}_n)\epsilon_{t-r_2}(\hat{\theta}_n)\frac{\partial}{\partial\theta_{m_2}}\epsilon_{t-r_2}(\hat{\theta}_n)\right|. \qquad (2.74)
\end{aligned}$$

Observe now that

$$\begin{aligned}
&\tilde{\epsilon}_{t-r_1}(\hat{\theta}_n)\frac{\partial}{\partial\theta_{m_1}}\tilde{\epsilon}_{t-r_1}(\hat{\theta}_n)\tilde{\epsilon}_{t-r_2}(\hat{\theta}_n)\frac{\partial}{\partial\theta_{m_2}}\tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) - \epsilon_{t-r_1}(\hat{\theta}_n)\frac{\partial}{\partial\theta_{m_1}}\epsilon_{t-r_1}(\hat{\theta}_n)\epsilon_{t-r_2}(\hat{\theta}_n)\frac{\partial}{\partial\theta_{m_2}}\epsilon_{t-r_2}(\hat{\theta}_n) \\
&= \left(\tilde{\epsilon}_{t-r_1}(\hat{\theta}_n) - \epsilon_{t-r_1}(\hat{\theta}_n)\right)\frac{\partial}{\partial\theta_{m_1}}\tilde{\epsilon}_{t-r_1}(\hat{\theta}_n)\tilde{\epsilon}_{t-r_2}(\hat{\theta}_n)\frac{\partial}{\partial\theta_{m_2}}\tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) \\
&\quad + \epsilon_{t-r_1}(\hat{\theta}_n)\left(\frac{\partial}{\partial\theta_{m_1}}\tilde{\epsilon}_{t-r_1}(\hat{\theta}_n) - \frac{\partial}{\partial\theta_{m_1}}\epsilon_{t-r_1}(\hat{\theta}_n)\right)\tilde{\epsilon}_{t-r_2}(\hat{\theta}_n)\frac{\partial}{\partial\theta_{m_2}}\tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) \\
&\quad + \epsilon_{t-r_1}(\hat{\theta}_n)\frac{\partial}{\partial\theta_{m_1}}\epsilon_{t-r_1}(\hat{\theta}_n)\left(\tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) - \epsilon_{t-r_2}(\hat{\theta}_n)\right)\frac{\partial}{\partial\theta_{m_2}}\tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) \\
&\quad + \epsilon_{t-r_1}(\hat{\theta}_n)\frac{\partial}{\partial\theta_{m_1}}\epsilon_{t-r_1}(\hat{\theta}_n)\epsilon_{t-r_2}(\hat{\theta}_n)\left(\frac{\partial}{\partial\theta_{m_2}}\tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) - \frac{\partial}{\partial\theta_{m_2}}\epsilon_{t-r_2}(\hat{\theta}_n)\right).
\end{aligned}$$

We replace the above identity in (2.74) and we obtain by Hölder's inequality that

$$\mathbb{P}\left(\sqrt{r}\left\|\hat{\Sigma}_{\underline{\hat{H}}_r} - \hat{\Sigma}_{\underline{H}_{r,n}}\right\| \geq \beta\right) \leq \frac{4K\sqrt{r}}{n\beta}\sum_{r_1=1}^{r}\sum_{r_2=1}^{r}\sum_{m_1=1}^{p+q+1}\sum_{m_2=1}^{p+q+1}(T_{n,1}+T_{n,2}+T_{n,3}+T_{n,4}) \quad (2.75)$$

where

$$T_{n,1} = \sum_{t=1}^{n} \left\| \tilde{\epsilon}_{t-r_1}(\hat{\theta}_n) - \epsilon_{t-r_1}(\hat{\theta}_n) \right\|_{\mathbb{L}^2} \left\| \frac{\partial}{\partial \theta_{m_1}} \tilde{\epsilon}_{t-r_1}(\hat{\theta}_n) \right\|_{\mathbb{L}^6} \left\| \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) \right\|_{\mathbb{L}^6} \left\| \frac{\partial}{\partial \theta_{m_2}} \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) \right\|_{\mathbb{L}^6},$$

$$T_{n,2} = \sum_{t=1}^{n} \left\| \epsilon_{t-r_1}(\hat{\theta}_n) \right\|_{\mathbb{L}^6} \left\| \frac{\partial}{\partial \theta_{m_1}} \tilde{\epsilon}_{t-r_1}(\hat{\theta}_n) - \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\hat{\theta}_n) \right\|_{\mathbb{L}^2} \left\| \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) \right\|_{\mathbb{L}^6} \left\| \frac{\partial}{\partial \theta_{m_2}} \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) \right\|_{\mathbb{L}^6},$$

$$T_{n,3} = \sum_{t=1}^{n} \left\| \epsilon_{t-r_1}(\hat{\theta}_n) \right\|_{\mathbb{L}^6} \left\| \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\hat{\theta}_n) \right\|_{\mathbb{L}^6} \left\| \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) - \epsilon_{t-r_2}(\hat{\theta}_n) \right\|_{\mathbb{L}^2} \left\| \frac{\partial}{\partial \theta_{m_2}} \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) \right\|_{\mathbb{L}^6},$$

$$T_{n,4} = \sum_{t=1}^{n} \left\| \epsilon_{t-r_1}(\hat{\theta}_n) \right\|_{\mathbb{L}^6} \left\| \frac{\partial}{\partial \theta_{m_1}} \epsilon_{t-r_1}(\hat{\theta}_n) \right\|_{\mathbb{L}^6} \left\| \epsilon_{t-r_2}(\hat{\theta}_n) \right\|_{\mathbb{L}^6} \left\| \frac{\partial}{\partial \theta_{m_2}} \tilde{\epsilon}_{t-r_2}(\hat{\theta}_n) - \frac{\partial}{\partial \theta_{m_2}} \epsilon_{t-r_2}(\hat{\theta}_n) \right\|_{\mathbb{L}^2}.$$

For all $\theta \in \Theta_\delta$ and $t \in \mathbb{Z}$, in view of (2.30) and Lemma 2.1, we have

$$\begin{aligned}
\left\| \tilde{\epsilon}_t(\hat{\theta}_n) - \epsilon_t(\hat{\theta}_n) \right\|_{\mathbb{L}^2} &= \left(\mathbb{E}\left[\left\{ \sum_{j \geq 0} \left(\lambda_j^t(\hat{\theta}_n) - \lambda_j(\hat{\theta}_n) \right) \epsilon_{t-j} \right\}^2 \right] \right)^{1/2} \\
&\leq \sup_{\theta \in \Theta_\delta} \left(\mathbb{E}\left[\left\{ \sum_{j \geq 0} \left(\lambda_j^t(\theta) - \lambda_j(\theta) \right) \epsilon_{t-j} \right\}^2 \right] \right)^{1/2} \\
&\leq \sigma_\epsilon \sup_{\theta \in \Theta_\delta} \| \lambda(\theta) - \lambda^t(\theta) \|_{\ell^2} \\
&\leq K \frac{1}{t^{1/2-(d_2-d_1)}}.
\end{aligned}$$

It is not difficult to prove that $\tilde{\epsilon}_t(\theta)$ and $\partial \tilde{\epsilon}_t(\theta)/\partial\theta$ belong to $\mathbb{L}^6$. The fact that $\epsilon_t(\theta)$ and $\partial \epsilon_t(\theta)/\partial\theta$ have moment of order 6 can be proved using the same method than in Lemma 2.12 using the absolute summability of the k-th ($k = 2, \ldots, 8$) cumulants assumed in **(A4')** with $\tau = 8$. We deduce that

$$\begin{aligned}
T_{n,1} \leq K \sum_{t=1}^{n} \left\| \tilde{\epsilon}_{t-r_1}(\hat{\theta}_n) - \epsilon_{t-r_1}(\hat{\theta}_n) \right\|_{\mathbb{L}^2} &\leq K \sum_{t=1-r}^{0} \left\| \epsilon_t(\hat{\theta}_n) \right\|_{\mathbb{L}^2} + K \sum_{t=1}^{n} \left\| \tilde{\epsilon}_t(\hat{\theta}_n) - \epsilon_t(\hat{\theta}_n) \right\|_{\mathbb{L}^2} \\
&\leq K \left(r + \sum_{t=1}^{n} \frac{1}{t^{1/2-(d_2-d_1)}} \right).
\end{aligned}$$

Then we obtain

$$T_{n,1} \leq K \left(r + n^{1/2+(d_2-d_1)} \right). \tag{2.76}$$

The same calculations hold for the terms $T_{n,2}$, $T_{n,3}$ and $T_{n,4}$. Thus

$$T_{n,1} + T_{n,2} + T_{n,3} + T_{n,4} \leq K \left(r + n^{1/2+(d_2-d_1)} \right) \tag{2.77}$$

and reporting this estimation in (2.75) implies that

$$\mathbb{P}\left(\sqrt{r}\left\|\hat{\Sigma}_{\underline{\hat{H}}_r}-\hat{\Sigma}_{\underline{H}_{r,n}}\right\|\geq\beta\right)\leq\frac{Kr^{5/2}(p+q+1)^2}{n\beta}\left(r+n^{1/2+(d_2-d_1)}\right)$$
$$\leq K\left(\frac{r^{7/2}}{n}+\frac{r^{5/2}}{n^{1/2-(d_2-d_1)}}\right).$$

Since $2/7>(1-2(d_2-d_1))/5$, the sequence $\sqrt{r}\left\|\hat{\Sigma}_{\underline{\hat{H}}_r}-\hat{\Sigma}_{\underline{H}_{r,n}}\right\|$ converges in probability to 0 as $n\to\infty$ when $r=r(n)=o(n^{(1-2(d_2-d_1))/5})$.

Step 2: proof of (2.73).
First we follow the same approach than in the previous step. We have

$$\left\|\hat{\Sigma}_{\underline{H}_{r,n}}-\hat{\Sigma}_{\underline{H}_r}\right\|^2=\left\|\frac{1}{n}\sum_{t=1}^{n}\underline{H}^{(n)}_{r,t}\underline{H}^{(n)\prime}_{r,t}-\frac{1}{n}\sum_{t=1}^{n}\underline{H}_{r,t}\underline{H}'_{r,t}\right\|^2$$
$$\leq\sum_{r_1=1}^{r}\sum_{r_2=1}^{r}\sum_{m_1=1}^{p+q+1}\sum_{m_2=1}^{p+q+1}\left(\frac{1}{n}\sum_{t=1}^{n}H^{(n)}_{t-r_1,m_1}H^{(n)}_{t-r_2,m_2}-\frac{1}{n}\sum_{t=1}^{n}H_{t-r_1,m_1}H_{t-r_2,m_2}\right)^2$$
$$\leq 16\sum_{r_1=1}^{r}\sum_{r_2=1}^{r}\sum_{m_1=1}^{p+q+1}\sum_{m_2=1}^{p+q+1}\left(\frac{1}{n}\sum_{t=1}^{n}\epsilon_{t-r_1}(\hat{\theta}_n)\frac{\partial}{\partial\theta_{m_1}}\epsilon_{t-r_1}(\hat{\theta}_n)\epsilon_{t-r_2}(\hat{\theta}_n)\frac{\partial}{\partial\theta_{m_2}}\epsilon_{t-r_2}(\hat{\theta}_n)\right.$$
$$\left.-\epsilon_{t-r_1}(\theta_0)\frac{\partial}{\partial\theta_{m_1}}\epsilon_{t-r_1}(\theta_0)\epsilon_{t-r_2}(\theta_0)\frac{\partial}{\partial\theta_{m_2}}\epsilon_{t-r_2}(\theta_0)\right)^2.$$

Since

$$\epsilon_{t-r_1}(\hat{\theta}_n)\frac{\partial}{\partial\theta_{m_1}}\epsilon_{t-r_1}(\hat{\theta}_n)\epsilon_{t-r_2}(\hat{\theta}_n)\frac{\partial}{\partial\theta_{m_2}}\epsilon_{t-r_2}(\hat{\theta}_n)-\epsilon_{t-r_1}(\theta_0)\frac{\partial}{\partial\theta_{m_1}}\epsilon_{t-r_1}(\theta_0)\epsilon_{t-r_2}(\theta_0)\frac{\partial}{\partial\theta_{m_2}}\epsilon_{t-r_2}(\theta_0)$$
$$=\left(\epsilon_{t-r_1}(\hat{\theta}_n)-\epsilon_{t-r_1}(\theta_0)\right)\frac{\partial}{\partial\theta_{m_1}}\epsilon_{t-r_1}(\hat{\theta}_n)\epsilon_{t-r_2}(\hat{\theta}_n)\frac{\partial}{\partial\theta_{m_2}}\epsilon_{t-r_2}(\hat{\theta}_n)$$
$$+\epsilon_{t-r_1}(\theta_0)\left(\frac{\partial}{\partial\theta_{m_1}}\epsilon_{t-r_1}(\hat{\theta}_n)-\frac{\partial}{\partial\theta_{m_1}}\epsilon_{t-r_1}(\theta_0)\right)\epsilon_{t-r_2}(\hat{\theta}_n)\frac{\partial}{\partial\theta_{m_2}}\epsilon_{t-r_2}(\hat{\theta}_n)$$
$$+\epsilon_{t-r_1}(\theta_0)\frac{\partial}{\partial\theta_{m_1}}\epsilon_{t-r_1}(\theta_0)\left(\epsilon_{t-r_2}(\hat{\theta}_n)-\epsilon_{t-r_2}(\theta_0)\right)\frac{\partial}{\partial\theta_{m_2}}\epsilon_{t-r_2}(\hat{\theta}_n)$$
$$+\epsilon_{t-r_1}(\theta_0)\frac{\partial}{\partial\theta_{m_1}}\epsilon_{t-r_1}(\theta_0)\epsilon_{t-r_2}(\theta_0)\left(\frac{\partial}{\partial\theta_{m_2}}\epsilon_{t-r_2}(\hat{\theta}_n)-\frac{\partial}{\partial\theta_{m_2}}\epsilon_{t-r_2}(\theta_0)\right),$$

one has

$$\left\|\hat{\Sigma}_{\underline{H}_{r,n}}-\hat{\Sigma}_{\underline{H}_r}\right\|^2\leq 16\sum_{r_1=1}^{r}\sum_{r_2=1}^{r}\sum_{m_1=1}^{p+q+1}\sum_{m_2=1}^{p+q+1}(U_{n,1}+U_{n,2}+U_{n,3}+U_{n,4})^2 \qquad (2.78)$$

where

$$U_{n,1} = \frac{1}{n}\sum_{t=1}^{n} \left|\epsilon_{t-r_1}(\hat{\theta}_n) - \epsilon_{t-r_1}(\theta_0)\right| \left|\frac{\partial}{\partial\theta_{m_1}}\epsilon_{t-r_1}(\hat{\theta}_n)\right| \left|\epsilon_{t-r_2}(\hat{\theta}_n)\right| \left|\frac{\partial}{\partial\theta_{m_2}}\epsilon_{t-r_2}(\hat{\theta}_n)\right|,$$

$$U_{n,2} = \frac{1}{n}\sum_{t=1}^{n} \left|\epsilon_{t-r_1}(\theta_0)\right| \left|\frac{\partial}{\partial\theta_{m_1}}\epsilon_{t-r_1}(\hat{\theta}_n) - \frac{\partial}{\partial\theta_{m_1}}\epsilon_{t-r_1}(\theta_0)\right| \left|\epsilon_{t-r_2}(\hat{\theta}_n)\right| \left|\frac{\partial}{\partial\theta_{m_2}}\epsilon_{t-r_2}(\hat{\theta}_n)\right|,$$

$$U_{n,3} = \frac{1}{n}\sum_{t=1}^{n} \left|\epsilon_{t-r_1}(\theta_0)\right| \left|\frac{\partial}{\partial\theta_{m_1}}\epsilon_{t-r_1}(\theta_0)\right| \left|\epsilon_{t-r_2}(\hat{\theta}_n) - \epsilon_{t-r_2}(\theta_0)\right| \left|\frac{\partial}{\partial\theta_{m_2}}\epsilon_{t-r_2}(\hat{\theta}_n)\right|$$

$$U_{n,4} = \frac{1}{n}\sum_{t=1}^{n} \left|\epsilon_{t-r_1}(\theta_0)\right| \left|\frac{\partial}{\partial\theta_{m_1}}\epsilon_{t-r_1}(\theta_0)\right| \left|\epsilon_{t-r_2}(\theta_0)\right| \left|\frac{\partial}{\partial\theta_{m_2}}\epsilon_{t-r_2}(\hat{\theta}_n) - \frac{\partial}{\partial\theta_{m_2}}\epsilon_{t-r_2}(\theta_0)\right|.$$

Taylor expansions around θ_0 yield that there exists $\underline{\theta}$ and $\overline{\theta}$ between $\hat{\theta}_n$ and θ_0 such that

$$\left|\epsilon_t(\hat{\theta}_n) - \epsilon_t(\theta_0)\right| \leq w_t \left\|\hat{\theta}_n - \theta_0\right\| \text{ and}$$
$$\left|\frac{\partial}{\partial\theta_m}\epsilon_t(\hat{\theta}_n) - \frac{\partial}{\partial\theta_m}\epsilon_t(\theta_0)\right| \leq q_t \left\|\hat{\theta}_n - \theta_0\right\|$$

with $w_t = \left\|\partial\epsilon_t(\underline{\theta})/\partial\theta'\right\|$ and $q_t = \left\|\partial^2\epsilon_t(\overline{\theta})/\partial\theta'\partial\theta_m\right\|$. Using the fact that

$$\mathbb{E}\left|w_{t-r_1}\frac{\partial}{\partial\theta_{m_1}}\epsilon_{t-r_1}(\hat{\theta}_n)\epsilon_{t-r_2}(\hat{\theta}_n)\frac{\partial}{\partial\theta_{m_2}}\epsilon_{t-r_2}(\hat{\theta}_n)\right| < \infty$$

and that $(\sqrt{n}(\hat{\theta}_n - \theta_0))_n$ is a tight sequence (which implies that $\|\hat{\theta}_n - \theta_0\| = \mathrm{O}_{\mathbb{P}}(1/\sqrt{n})$), we deduce that

$$U_{n,1} = \mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right).$$

The same arguments are valid for $U_{n,2}$, $U_{n,3}$ and $U_{n,4}$. Consequently $U_{n,1} + U_{n,2} + U_{n,3} + U_{n,4} = \mathrm{O}_{\mathbb{P}}(1/\sqrt{n})$ and (2.78) yields

$$\left\|\hat{\Sigma}_{\underline{H}_{r,n}} - \hat{\Sigma}_{\underline{H}_r}\right\|^2 = \mathrm{O}_{\mathbb{P}}\left(\frac{r^2}{n}\right).$$

When $r = \mathrm{o}(n^{1/3})$ we finally obtain $\sqrt{r}\|\hat{\Sigma}_{\underline{H}_{r,n}} - \hat{\Sigma}_{\underline{H}_r}\| = \mathrm{o}_{\mathbb{P}}(1)$.

□

Lemma 2.15. *Under the assumptions of Theorem 2.3, we have*

$$\sqrt{r}\,\|\underline{\Phi}_r^* - \underline{\Phi}_r\| = \mathrm{o}_{\mathbb{P}}(1) \quad \text{as } r \to \infty.$$

Proof. Recall that by (2.8) and (2.68) we have

$$H_t(\theta_0) = \underline{\Phi}_r \underline{H}_{r,t} + u_{r,t} = \underline{\Phi}_r^* \underline{H}_{r,t} + \sum_{k=r+1}^{\infty} \Phi_k H_{t-k}(\theta_0) + u_t := \underline{\Phi}_r^* \underline{H}_{r,t} + u_{r,t}^*.$$

By the orthogonality conditions in (2.8) and (2.68), one has

$$\begin{aligned}\Sigma_{u_r^*, \underline{H}_r} := \mathbb{E}\left[u_{r,t}^* \underline{H}_{r,t}'\right] &= \mathbb{E}\left[\left(H_t(\theta_0) - \underline{\Phi}_r^* \underline{H}_{r,t}\right) \underline{H}_{r,t}'\right] = \mathbb{E}\left[\left(\underline{\Phi}_r \underline{H}_{r,t} + u_{r,t} - \underline{\Phi}_r^* \underline{H}_{r,t}\right) \underline{H}_{r,t}'\right] \\ &= \left(\underline{\Phi}_r - \underline{\Phi}_r^*\right) \Sigma_{\underline{H}_r},\end{aligned}$$

and consequently

$$\underline{\Phi}_r^* - \underline{\Phi}_r = -\Sigma_{u_r^*, \underline{H}_r} \Sigma_{\underline{H}_r}^{-1}. \tag{2.79}$$

Using Lemma 2.11 and Lemma 2.12, (2.79) implies that

$$\begin{aligned}\mathbb{P}\left(\sqrt{r}\,\|\underline{\Phi}_r^* - \underline{\Phi}_r\| \geq \beta\right) &\leq \frac{\sqrt{r}}{\beta} \left\|\Sigma_{u_r^*, \underline{H}_r}\right\| \left\|\Sigma_{\underline{H}_r}^{-1}\right\| \\ &\leq \frac{K\sqrt{r}}{\beta} \left\|\mathbb{E}\left[\left(\sum_{k \geq r+1} \Phi_k H_{t-k}(\theta_0) + u_t\right) \underline{H}_{r,t}'\right]\right\| \\ &\leq \frac{K\sqrt{r}}{\beta} \sum_{k \geq r+1} \|\Phi_k\| \left\|\mathbb{E}\left[H_{t-k}(\theta_0) \underline{H}_{r,t}'\right]\right\| \\ &\leq \frac{K\sqrt{r}}{\beta} \sum_{\ell \geq 1} \|\Phi_{\ell+r}\| \left\|\mathbb{E}\left[H_{t-\ell-r}(\theta_0)\left(H_{t-1}'(\theta_0), \ldots, H_{t-r}'(\theta_0)\right)\right]\right\| \\ &\leq \frac{K\sqrt{r}}{\beta} \sum_{\ell \geq 1} \|\Phi_{\ell+r}\| \left(\sum_{j=1}^{p+q+1} \sum_{k=1}^{p+q+1} \sum_{r_1=1}^{r} \left|\mathbb{E}\left[H_{t-r-\ell,j}(\theta_0) H_{t-r_1,k}(\theta_0)\right]\right|^2\right)^{1/2} \\ &\leq \frac{K\sqrt{r}}{\beta} \sum_{\ell \geq 1} \|\Phi_{\ell+r}\| \left(\sum_{j=1}^{p+q+1} \sum_{k=1}^{p+q+1} \sum_{r_1=1}^{r} \mathbb{E}\left[H_{t-r-\ell,j}^2(\theta_0)\right] \mathbb{E}\left[H_{t-r_1,k}^2(\theta_0)\right]\right)^{1/2} \\ &\leq \frac{K(p+q+1)r}{\beta} \sum_{\ell \geq 1} \|\Phi_{\ell+r}\|.\end{aligned}$$

By (2.9), $r \sum_{\ell \geq 1} \|\Phi_{\ell+r}\| = o(1)$ as $r \to \infty$. The proof of the lemma follows. □

Lemma 2.16. *Under the assumptions of Theorem 2.3, we have*

$$\sqrt{r}\left\|\hat{\Sigma}_{\underline{\hat{H}}_r}^{-1} - \Sigma_{\underline{H}_r}^{-1}\right\| = o_{\mathbb{P}}(1)$$

as $n \to \infty$ when $r = o(n^{(1-2(d_2-d_1))/5})$ and $r \to \infty$.

Proof. We have

$$\left\|\hat{\Sigma}_{\underline{\hat{H}}_r}^{-1}-\Sigma_{\underline{H}_r}^{-1}\right\| \leq \left(\left\|\hat{\Sigma}_{\underline{\hat{H}}_r}^{-1}-\Sigma_{\underline{H}_r}^{-1}\right\|+\left\|\Sigma_{\underline{H}_r}^{-1}\right\|\right)\left\|\Sigma_{\underline{H}_r}-\hat{\Sigma}_{\underline{\hat{H}}_r}\right\|\left\|\Sigma_{\underline{H}_r}^{-1}\right\|,$$

and by induction we obtain

$$\left\|\hat{\Sigma}_{\underline{\hat{H}}_r}^{-1}-\Sigma_{\underline{H}_r}^{-1}\right\| \leq \left\|\Sigma_{\underline{H}_r}^{-1}\right\|\sum_{k=1}^{\infty}\left\|\Sigma_{\underline{H}_r}-\hat{\Sigma}_{\underline{\hat{H}}_r}\right\|^k\left\|\Sigma_{\underline{H}_r}^{-1}\right\|^k.$$

We have

$$\begin{aligned}
&\mathbb{P}\left(\sqrt{r}\|\hat{\Sigma}_{\underline{\hat{H}}_r}^{-1}-\Sigma_{\underline{H}_r}^{-1}\| > \beta\right)\\
&\quad\leq \mathbb{P}\left(\sqrt{r}\left\|\Sigma_{\underline{H}_r}^{-1}\right\|\sum_{k=1}^{\infty}\left\|\Sigma_{\underline{H}_r}-\hat{\Sigma}_{\underline{\hat{H}}_r}\right\|^k\left\|\Sigma_{\underline{H}_r}^{-1}\right\|^k > \beta\right)\\
&\quad\leq \mathbb{P}\left(\sqrt{r}\left\|\Sigma_{\underline{H}_r}^{-1}\right\|\sum_{k=1}^{\infty}\left\|\Sigma_{\underline{H}_r}-\hat{\Sigma}_{\underline{\hat{H}}_r}\right\|^k\left\|\Sigma_{\underline{H}_r}^{-1}\right\|^k > \beta \text{ and } \left\|\Sigma_{\underline{H}_r}-\hat{\Sigma}_{\underline{\hat{H}}_r}\right\|\left\|\Sigma_{\underline{H}_r}^{-1}\right\| < 1\right)\\
&\quad+\mathbb{P}\left(\sqrt{r}\left\|\Sigma_{\underline{H}_r}^{-1}\right\|\sum_{k=1}^{\infty}\left\|\Sigma_{\underline{H}_r}-\hat{\Sigma}_{\underline{\hat{H}}_r}\right\|^k\left\|\Sigma_{\underline{H}_r}^{-1}\right\|^k > \beta \text{ and } \left\|\Sigma_{\underline{H}_r}-\hat{\Sigma}_{\underline{\hat{H}}_r}\right\|\left\|\Sigma_{\underline{H}_r}^{-1}\right\| \geq 1\right)\\
&\quad\leq \mathbb{P}\left(\sqrt{r}\frac{\left\|\Sigma_{\underline{H}_r}^{-1}\right\|^2\left\|\Sigma_{\underline{H}_r}-\hat{\Sigma}_{\underline{\hat{H}}_r}\right\|}{1-\left\|\Sigma_{\underline{H}_r}-\hat{\Sigma}_{\underline{\hat{H}}_r}\right\|\left\|\Sigma_{\underline{H}_r}^{-1}\right\|} > \beta\right)+\mathbb{P}\left(\sqrt{r}\left\|\Sigma_{\underline{H}_r}-\hat{\Sigma}_{\underline{\hat{H}}_r}\right\|\left\|\Sigma_{\underline{H}_r}^{-1}\right\| \geq 1\right)\\
&\quad\leq \mathbb{P}\left(\sqrt{r}\left\|\Sigma_{\underline{H}_r}-\hat{\Sigma}_{\underline{\hat{H}}_r}\right\| > \frac{\beta}{\left\|\Sigma_{\underline{H}_r}^{-1}\right\|^2+\beta r^{-1/2}\left\|\Sigma_{\underline{H}_r}^{-1}\right\|}\right)\\
&\qquad\qquad+\mathbb{P}\left(\sqrt{r}\left\|\Sigma_{\underline{H}_r}-\hat{\Sigma}_{\underline{\hat{H}}_r}\right\| \geq \left\|\Sigma_{\underline{H}_r}^{-1}\right\|^{-1}\right).
\end{aligned}$$

Lemmas 2.11 and 2.14 imply the result. □

Lemma 2.17. *Under the assumptions of Theorem 2.3, we have*

$$\sqrt{r}\left\|\hat{\underline{\Phi}}_r-\underline{\Phi}_r\right\| = o_{\mathbb{P}}(1) \quad \text{as } r\to\infty \text{ and } r = o(n^{(1-2(d_2-d_1))/5}).$$

Proof. Lemmas 2.11 and 2.16 yield

$$\left\|\hat{\Sigma}_{\underline{\hat{H}}_r}^{-1}\right\| \leq \left\|\hat{\Sigma}_{\underline{\hat{H}}_r}^{-1}-\Sigma_{\underline{H}_r}^{-1}\right\|+\left\|\Sigma_{\underline{H}_r}^{-1}\right\| = O_{\mathbb{P}}(1). \tag{2.80}$$

By (2.68), we have

$$0 = \mathbb{E}\left[u_{r,t}\underline{H}_{r,t}'\right] = \mathbb{E}\left[\left(H_t(\theta_0)-\underline{\Phi}_r\underline{H}_{r,t}\right)\underline{H}_{r,t}'\right] = \Sigma_{H,\underline{H}_r}-\underline{\Phi}_r\Sigma_{\underline{H}_r},$$

and so we have $\underline{\Phi}_r = \Sigma_{H,\underline{H}_r}\Sigma_{\underline{H}_r}^{-1}$. Lemmas 2.11, 2.14, 2.16 and (2.80) imply

$$\begin{aligned}\sqrt{r}\left\|\hat{\underline{\Phi}}_r - \underline{\Phi}_r\right\| &= \sqrt{r}\left\|\hat{\Sigma}_{\hat{H},\hat{\underline{H}}_r}\hat{\Sigma}_{\hat{\underline{H}}_r}^{-1} - \Sigma_{H,\underline{H}_r}\Sigma_{\underline{H}_r}^{-1}\right\| \\ &= \sqrt{r}\left\|\left(\hat{\Sigma}_{\hat{H},\hat{\underline{H}}_r} - \Sigma_{H,\underline{H}_r}\right)\hat{\Sigma}_{\hat{\underline{H}}_r}^{-1} + \Sigma_{H,\underline{H}_r}\left(\hat{\Sigma}_{\hat{\underline{H}}_r}^{-1} - \Sigma_{\underline{H}_r}^{-1}\right)\right\| \\ &= o_{\mathbb{P}}(1),\end{aligned}$$

and the lemma is proved. □

Proof of Theorem 2.3

Since by Lemma 2.14 we have $\|\hat{\Sigma}_{\hat{H}} - \Sigma_H\| = o_{\mathbb{P}}(r^{-1/2}) = o_{\mathbb{P}}(1)$ and $\|\hat{\Sigma}_{\hat{H},\hat{\underline{H}}_r} - \Sigma_{H,\underline{H}_r}\| = o_{\mathbb{P}}(r^{-1/2}) = o_{\mathbb{P}}(1)$, and by Lemma 2.15 $\|\hat{\underline{\Phi}}_r - \underline{\Phi}_r^*\| = o_{\mathbb{P}}(r^{-1/2}) = o_{\mathbb{P}}(1)$, Theorem 2.3 is then proved.

2.6.5 Invertibility of the normalization matrix $P_{p+q+1,n}$

The following proofs are quite technical and are adaptations of the arguments used in Boubacar Maïnassara and Saussereau [2018].

To prove Proposition 2.1, we need to introduce the following notation.

We denote S_t the vector of $\mathbb{R}^{p+q+1}$ defined by:

$$S_t = \sum_{j=1}^{t} U_j = \sum_{j=1}^{t} -2J^{-1}H_j = -2J^{-1}\sum_{j=1}^{t}\epsilon_j\frac{\partial}{\partial\theta}\epsilon_j(\theta_0),$$

and $S_t(i)$ is its $i-$th component. We have

$$S_{t-1}(i) = S_t(i) - U_t(i). \tag{2.81}$$

If the matrix $P_{p+q+1,n}$ is not invertible, there exists some real constants $d_1,\dots,d_{p+q+1}$ not all equal to zero, such that $\mathbf{d}'P_{p+q+1,n}\mathbf{d} = 0$, where $\mathbf{d} = (d_1,\dots,d_{p+q+1})'$. Thus we may write that $\sum_{i=1}^{p+q+1}\sum_{j=1}^{p+q+1} d_jP_{p+q+1,n}(j,i)d_i = 0$ or equivalently

$$\frac{1}{n^2}\sum_{t=1}^{n}\sum_{i=1}^{p+q+1}\sum_{j=1}^{p+q+1} d_j\left(\sum_{k=1}^{t}(U_k(j) - \bar{U}_n(j))\right)\left(\sum_{k=1}^{t}(U_k(i) - \bar{U}_n(i))\right)d_i = 0.$$

Then

$$\sum_{t=1}^{n}\left(\sum_{i=1}^{p+q+1} d_i\left(\sum_{k=1}^{t}(U_k(i) - \bar{U}_n(i))\right)\right)^2 = 0,$$

which implies that for all $t \geq 1$

$$\sum_{i=1}^{p+q+1} d_i \left(\sum_{k=1}^{t} (U_k(i) - \bar{U}_n(i)) \right) = \sum_{i=1}^{p+q+1} d_i \left(S_t(i) - \frac{t}{n} S_n(i) \right) = 0.$$

So we have

$$\frac{1}{t} \sum_{i=1}^{p+q+1} d_i S_t(i) = \sum_{i=1}^{p+q+1} d_i \left(\frac{1}{n} S_n(i) \right). \tag{2.82}$$

We apply the ergodic theorem and we use the orthogonality of ϵ_t and $(\partial/\partial\theta)\epsilon_t(\theta_0)$ in order to obtain that

$$\sum_{i=1}^{p+q+1} d_i \left(\frac{1}{n} \sum_{k=1}^{n} U_k(i) \right) \xrightarrow[n\to\infty]{\text{a.s.}} \sum_{i=1}^{p+q+1} d_i \mathbb{E}\left[U_k(i)\right] = -2 \sum_{i,j=1}^{p+q+1} d_i J^{-1}(i,j) \mathbb{E}\left[\epsilon_k \frac{\partial \epsilon_k}{\partial \theta_j} \right] = 0 \ .$$

Reporting this convergence in (2.82) implies that $\sum_{i=1}^{p+q+1} d_i S_t(i) = 0$ a.s. for all $t \geq 1$. By (2.81), we deduce that

$$\sum_{i=1}^{p+q+1} d_i U_t(i) = -2 \sum_{i=1}^{p+q+1} d_i \sum_{j=1}^{p+q+1} J^{-1}(i,j) \left(\epsilon_t \frac{\partial \epsilon_t}{\partial \theta_j} \right) = 0, \quad \text{a.s.}$$

Thanks to Assumption **(A5)**, $(\epsilon_t)_{t\in\mathbb{Z}}$ has a positive density in some neighborhood of zero and then $\epsilon_t \neq 0$ almost-surely. So we would have $\mathbf{d}' J^{-1} \frac{\partial \epsilon_t}{\partial \theta} = 0$ a.s. Now we can follow the same arguments that we developed in the proof of the invertibility of J (see Proof of Lemma 2.6 and more precisely (2.53)) and this leads us to a contradiction. We deduce that the matrix $P_{p+q+1,n}$ is non singular.

2.6.6 Proof of Theorem 2.4

The arguments follows the one Boubacar Maïnassara and Saussereau [2018] in a simpler context.

Recall that the Skorohod space $\mathbb{D}^\ell[0,1]$ is the set of $\mathbb{R}^\ell-$valued functions on [0,1] which are right-continuous and has left limits everywhere. It is endowed with the Skorohod topology and the weak convergence on $\mathbb{D}^\ell[0,1]$ is mentioned by $\xrightarrow{\mathbb{D}^\ell}$. The integer part of x will be denoted by $\lfloor x \rfloor$.

The goal at first is to show that there exists a lower triangular matrix T with nonnegative diagonal entries such that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} U_t \xrightarrow[n\to\infty]{\mathbb{D}^{p+q+1}} (TT')^{1/2} B_{p+q+1}(r), \tag{2.83}$$

where $(B_{p+q+1}(r))_{r\geq 0}$ is a $(p+q+1)-$dimensional standard Brownian motion. Using (2.30), U_t can be rewritten as

$$U_t = \left(-2 \left\{ \sum_{i=1}^{\infty} \dot{\lambda}_{i,1}(\theta_0)\, \epsilon_t \epsilon_{t-i}, \ldots, \sum_{i=1}^{\infty} \dot{\lambda}_{i,p+q+1}(\theta_0)\, \epsilon_t \epsilon_{t-i} \right\} J^{-1'} \right)'.$$

The non-correlation between ϵ_t's implies that the process $(U_t)_{t\in\mathbb{Z}}$ is centered. In order to apply the functional central limit theorem for strongly mixing process, we need to identify the asymptotic covariance matrix in the classical central limit theorem for the sequence $(U_t)_{t\in\mathbb{Z}}$. It is proved in Theorem 2.2 that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t \xrightarrow[n\to\infty]{\text{in law}} \mathcal{N}\left(0, \Omega =: 2\pi f_U(0)\right),$$

where $f_U(0)$ is the spectral density of the stationary process $(U_t)_{t\in\mathbb{Z}}$ evaluated at frequency 0. The existence of the matrix Ω has already been discussed (see the proofs of lemmas 2.6 and 2.9).

Since the matrix Ω is positive definite, it can be factorized as $\Omega = TT'$ where the $(p+q+1)\times(p+q+1)$ lower triangular matrix T has nonnegative diagonal entries. Therefore, we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (TT')^{-1/2} U_t \xrightarrow[n\to\infty]{\text{in law}} \mathcal{N}\left(0, I_{p+q+1}\right),$$

where $(TT')^{-1/2}$ is the Moore-Penrose inverse (see Magnus and Neudecker [1999], p. 36) of $(TT')^{1/2}$ and I_{p+q+1} is the identity matrix of order $p+q+1$.

As in the proof of the asymptotic normality of $(\sqrt{n}(\hat{\theta}_n - \theta_0))_n$, the distribution of $n^{-1/2}\sum_{t=1}^{n} U_t$ when n tends to infinity is obtained by introducing the random vector U_t^k defined for any strictly positive integer k by

$$U_t^k = \left(-2 \left\{ \sum_{i=1}^{k} \dot{\lambda}_{i,1}(\theta_0)\, \epsilon_t \epsilon_{t-i}, \ldots, \sum_{i=1}^{k} \dot{\lambda}_{i,p+q+1}(\theta_0)\, \epsilon_t \epsilon_{t-i} \right\} J^{-1'} \right)'.$$

Since U_t^k depends on a finite number of values of the noise-process $(\epsilon_t)_{t\in\mathbb{Z}}$, it also satisfies a mixing property (see Theorem 14.1 in Davidson [1994], p. 210). The central limit theorem for strongly mixing process of Herrndorf [1984] shows that its asymptotic distribution is normal with zero mean and variance matrix Ω_k that converges when k tends to infinity to Ω (see the proof of Lemma 2.10):

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t^k \xrightarrow[n\to\infty]{\text{in law}} \mathcal{N}\left(0, \Omega_k\right).$$

The above arguments also apply to matrix Ω_k with some matrix T_k which is defined analogously as T. Consequently, we obtain

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (T_k T_k')^{-1/2} U_t^k \xrightarrow[n\to\infty]{\text{in law}} \mathcal{N}(0, I_{p+q+1}).$$

Now we are able to apply the functional central limit theorem for strongly mixing process of Herrndorf [1984] and we obtain that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr\rfloor}(T_k T_k^{'})^{-1/2}U_t^k \xrightarrow[n\to\infty]{\mathbb{D}^{p+q+1}} B_{p+q+1}(r).$$

Since

$$(TT^{'})^{-1/2}U_t^k = \left((TT^{'})^{-1/2} - (T_k T_k^{'})^{-1/2}\right)U_t^k + (T_k T_k^{'})^{-1/2}U_t^k,$$

we may use the same approach as in the proof of Lemma 2.10 in order to prove that

$$n^{-1/2}\sum_{t=1}^{n}((TT^{'})^{-1/2} - (T_k T_k^{'})^{-1/2})U_t^k$$

converge in distribution to 0. Consequently we obtain that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr\rfloor}(TT^{'})^{-1/2}U_t^k \xrightarrow[n\to\infty]{\mathbb{D}^{p+q+1}} B_{p+q+1}(r).$$

In order to conclude that (2.83) is true, it remains to observe that uniformly with respect to n it holds that

$$\tilde{Y}_n^k(r) := \frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr\rfloor}(TT^{'})^{-1/2}\tilde{Z}_t^k \xrightarrow[n\to\infty]{\mathbb{D}^{p+q+1}} 0, \tag{2.84}$$

where

$$\tilde{Z}_t^k = \left(-2\left\{\sum_{i=k+1}^{\infty}\dot{\lambda}_{i,1}(\theta_0)\,\epsilon_t\epsilon_{t-i},\ldots,\sum_{i=k+1}^{\infty}\dot{\lambda}_{i,p+q+1}(\theta_0)\,\epsilon_t\epsilon_{t-i}\right\}J^{-1'}\right)^{'}.$$

By (2.64)

$$\sup_n \mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\tilde{Z}_t^k\right) \xrightarrow[n\to\infty]{} 0$$

and since $\lfloor nr\rfloor \le n$,

$$\sup_{0\le r\le 1}\sup_n\left\{\left\|\tilde{Y}_n^k(r)\right\|\right\} \xrightarrow[n\to\infty]{} 0.$$

Thus (2.84) is true and the proof of (2.83) is achieved.

By (2.83) we deduce that

$$\frac{1}{\sqrt{n}}\left(\sum_{j=1}^{\lfloor nr\rfloor}(U_j - \bar{U}_n)\right) \xrightarrow[n\to\infty]{\mathbb{D}^{p+q+1}} (TT^{'})^{1/2}\left(B_{p+q+1}(r) - rB_{p+q+1}(1)\right). \tag{2.85}$$

One remarks that the continuous mapping theorem on the Skorohod space yields

$$P_{p+q+1,n} \xrightarrow[n\to\infty]{\text{in law}} (TT')^{1/2}\left[\int_0^1 \{B_{p+q+1}(r) - rB_{p+q+1}(1)\}\{B_{p+q+1}(r) - rB_{p+q+1}(1)\}' \,\mathrm{dr}\right](TT')^{1/2}$$
$$= (TT')^{1/2} V_{p+q+1} (TT')^{1/2}.$$

Using (2.83), (2.85) and the continuous mapping theorem on the Skorohod space, one finally obtains

$$n(\hat{\theta}_n - \theta_0)' P_{p+q+1,n}^{-1}(\hat{\theta}_n - \theta_0)$$
$$\xrightarrow[n\to\infty]{\mathbb{D}^{p+q+1}} \left\{(TT')^{1/2} B_{p+q+1}(1)\right\}' \left\{(TT')^{1/2} V_{p+q+1} (TT')^{1/2}\right\}^{-1} \left\{(TT')^{1/2} B_{p+q+1}(1)\right\}$$
$$= B'_{p+q+1}(1) V_{p+q+1}^{-1} B_{p+q+1}(1) := \mathcal{U}_{p+q+1}.$$

The proof of Theorem 2.4 is then complete.

2.6.7 Proof of Theorem 2.5

In view of (2.15) and (2.18), we write $\hat{P}_{p+q+1,n} = P_{p+q+1,n} + Q_{p+q+1,n}$ where

$$Q_{p+q+1,n} = (J(\theta_0)^{-1} - \hat{J}_n^{-1}) \frac{1}{n^2} \sum_{t=1}^{n} \left(\sum_{j=1}^{t} (H_j - \tfrac{1}{n}\textstyle\sum_{k=1}^{n} H_k)\right)\left(\sum_{j=1}^{t} (H_j - \tfrac{1}{n}\textstyle\sum_{k=1}^{n} H_k)\right)'$$
$$+ \hat{J}_n^{-1} \frac{1}{n^2} \sum_{t=1}^{n} \left\{ \left(\sum_{j=1}^{t} (H_j - \tfrac{1}{n}\textstyle\sum_{k=1}^{n} H_k)\right)\left(\sum_{j=1}^{t} (H_j - \tfrac{1}{n}\textstyle\sum_{k=1}^{n} H_k)\right)' \right.$$
$$\left. - \left(\sum_{j=1}^{t} (\hat{H}_j - \tfrac{1}{n}\textstyle\sum_{k=1}^{n} \hat{H}_k)\right)\left(\sum_{j=1}^{t} (\hat{H}_j - \tfrac{1}{n}\textstyle\sum_{k=1}^{n} \hat{H}_k)\right)' \right\}.$$

Using the same approach as in Lemma 2.8, $\hat{J}_n$ converges almost surely to J. Thus we deduce that the first term in the right hand side of the above equation tends to zero in probability.

The second term is a sum composed of the following terms

$$q_{s,t}^{i,j,k,l} = \epsilon_s(\theta_0)\epsilon_t(\theta_0) \frac{\partial \epsilon_s(\theta_0)}{\partial \theta_i} \frac{\partial \epsilon_t(\theta_0)}{\partial \theta_j} - \tilde{\epsilon}_s(\hat{\theta}_n)\tilde{\epsilon}_t(\hat{\theta}_n) \frac{\partial \tilde{\epsilon}_s(\hat{\theta}_n)}{\partial \theta_k} \frac{\partial \tilde{\epsilon}_t(\hat{\theta}_n)}{\partial \theta_l}\ .$$

Using similar arguments done before (see for example the use of Taylor's expansion in Subsection 2.6.4, we have $q_{s,t}^{i,j,k,l} = o_{\mathbb{P}}(1)$ as n goes to infinity and thus $Q_{p+q+1,n} = o_{\mathbb{P}}(1)$. So one may find a matrix $Q^*_{p+q+1,n}$ that tends to the null matrix in probability and such that

$$n\left(\hat{\theta}_n - \theta_0\right)' \hat{P}_{p+q+1,n}^{-1} \left(\hat{\theta}_n - \theta_0\right) = n\left(\hat{\theta}_n - \theta_0\right)' (P_{p+q+1,n} + Q_{p+q+1,n})^{-1} \left(\hat{\theta}_n - \theta_0\right)$$
$$= n\left(\hat{\theta}_n - \theta_0\right)' P_{p+q+1,n}^{-1} \left(\hat{\theta}_n - \theta_0\right)$$
$$+ n\left(\hat{\theta}_n - \theta_0\right)' Q^*_{p+q+1,n} \left(\hat{\theta}_n - \theta_0\right).$$

Thanks to the arguments developed in the proof of Theorem 2.4, $n(\hat{\theta}_n - \theta_0)' P^{-1}_{p+q+1,n}(\hat{\theta}_n - \theta_0)$ converges in distribution. So $n(\hat{\theta}_n - \theta_0)' Q^{*}_{p+q+1,n}(\hat{\theta}_n - \theta_0)$ tends to zero in distribution, hence in probability. Then $n(\hat{\theta}_n - \theta_0)' \hat{P}^{-1}_{p+q+1,n}(\hat{\theta}_n - \theta_0)$ and $n(\hat{\theta}_n - \theta_0)' P^{-1}_{p+q+1,n}(\hat{\theta}_n - \theta_0)$ have the same limit in distribution and the result is proved.

2.7 Figures and tables

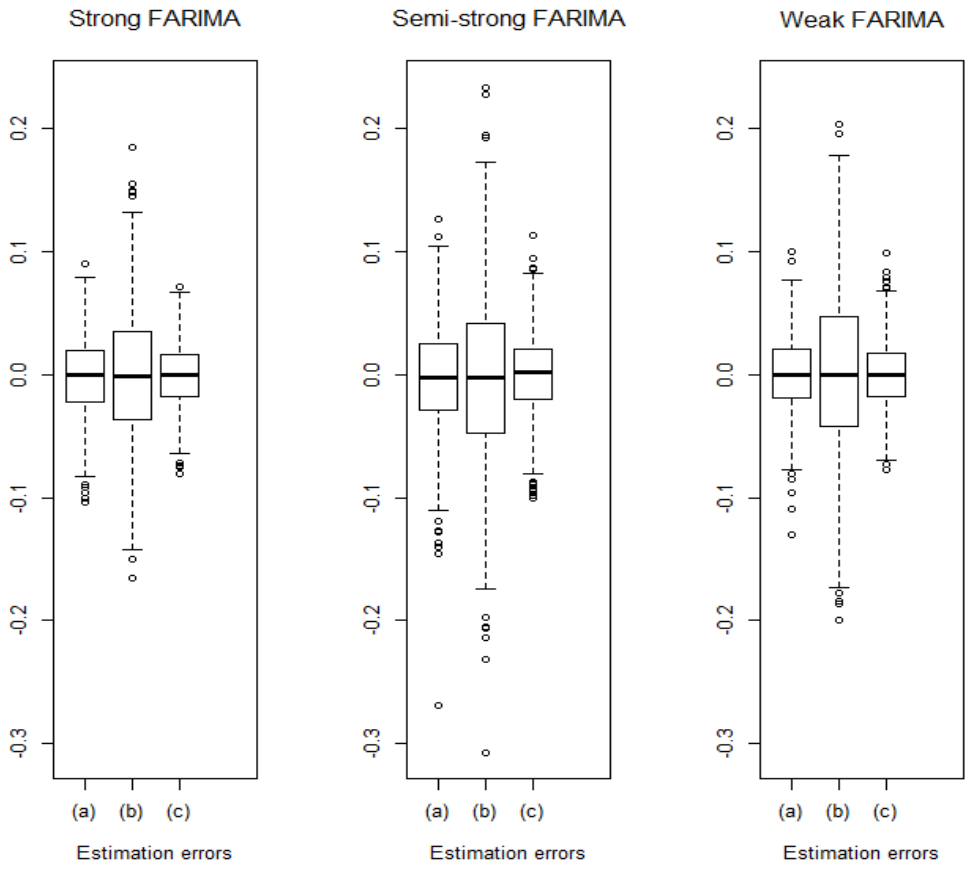


Figure 2.1 – LSE of $N = 1,000$ independent simulations of the FARIMA$(1, d, 1)$ model (2.19) with size $n = 2,000$ and unknown parameter $\theta_0 = (a, b, d) = (-0.7, -0.2, 0.4)$, when the noise is strong (left panel), when the noise is semi-strong (2.20) (middle panel) and when the noise is weak of the form (2.21) (right panel). Points (a)-(c), in the box-plots, display the distribution of the estimation error $\hat{\theta}_n(i) - \theta_0(i)$ for $i = 1, 2, 3$.

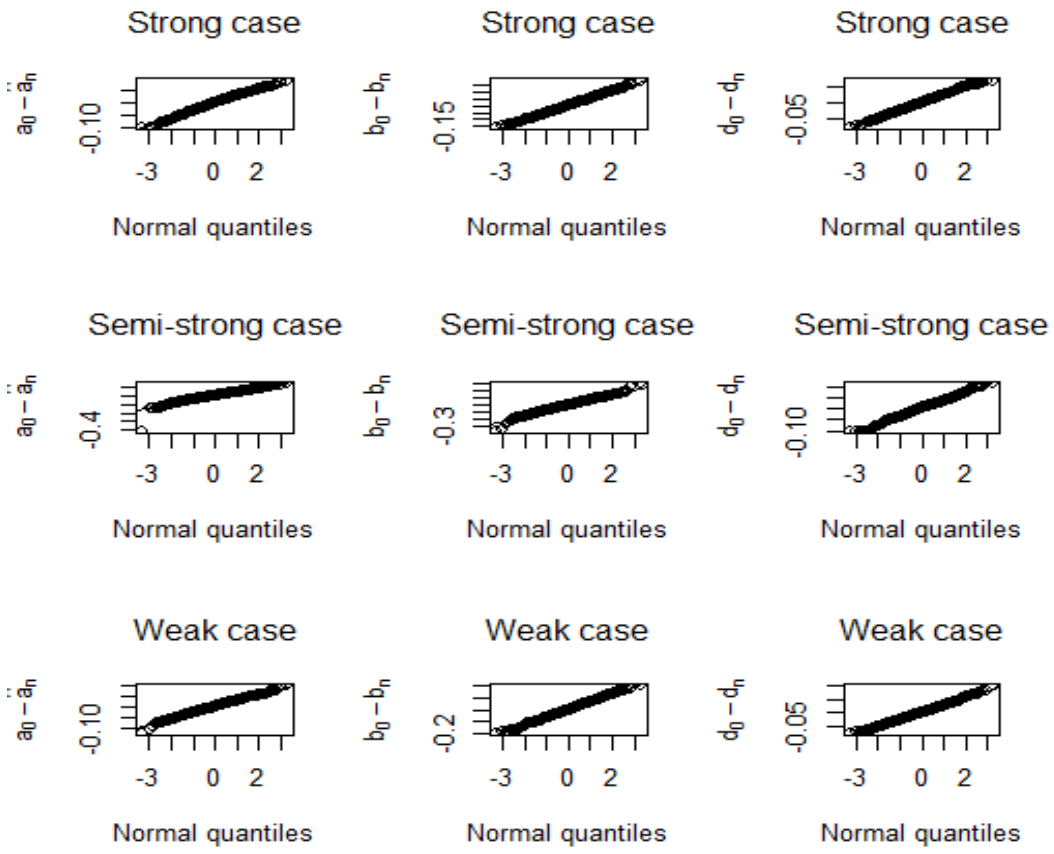


Figure 2.2 – LSE of $N = 1,000$ independent simulations of the FARIMA$(1, d, 1)$ model (2.19) with size $n = 2,000$ and unknown parameter $\theta_0 = (a, b, d) = (-0.7, -0.2, 0.4)$. The top panels present respectively, from left to right, the Q-Q plot of the estimates $\hat{a}_n$, $\hat{b}_n$ and $\hat{d}_n$ of a, b and d in the strong case. Similarly the middle and the bottom panels present respectively, from left to right, the Q-Q plot of the estimates $\hat{a}_n$, $\hat{b}_n$ and $\hat{d}_n$ of a, b and d in the semi-strong and weak cases.

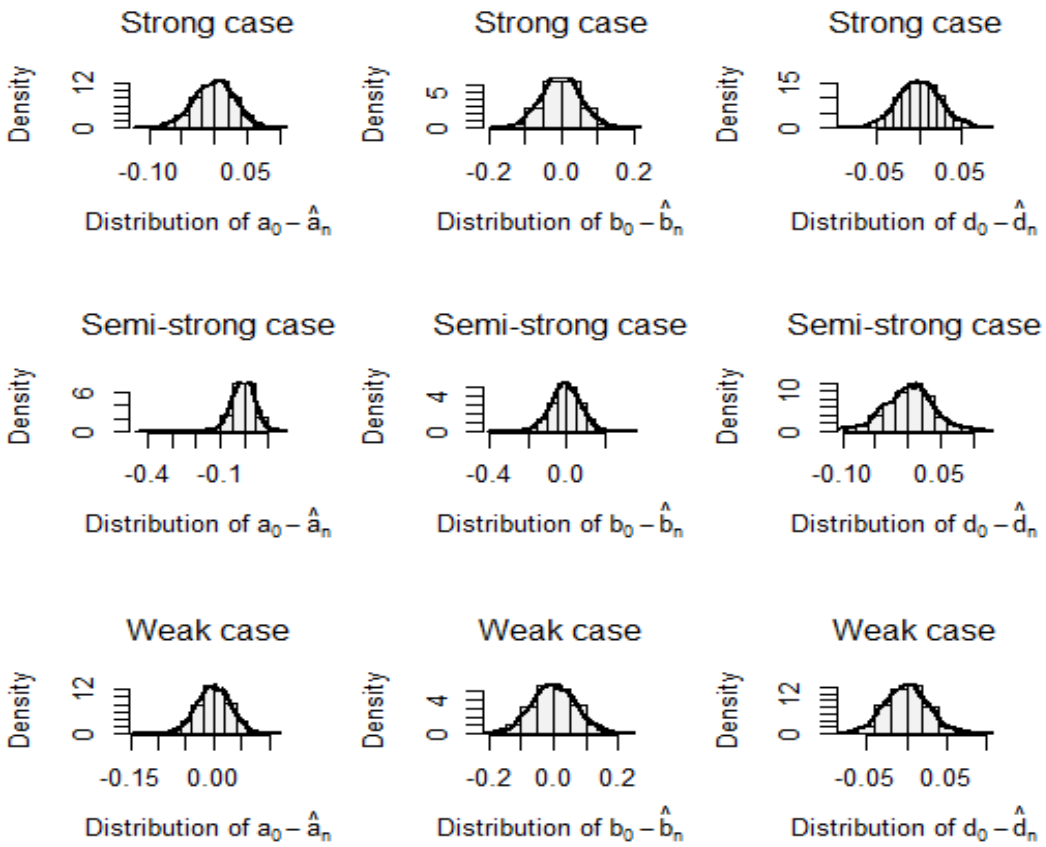


Figure 2.3 – LSE of $N = 1,000$ independent simulations of the FARIMA$(1, d, 1)$ model (2.19) with size $n = 2,000$ and unknown parameter $\theta_0 = (a, b, d) = (-0.7, -0.2, 0.4)$. The top panels present respectively, from left to right, the distribution of the estimates $\hat{a}_n$, $\hat{b}_n$ and $\hat{d}_n$ of a, b and d in the strong case. Similarly the middle and the bottom panels present respectively, from left to right, the distribution of the estimates $\hat{a}_n$, $\hat{b}_n$ and $\hat{d}_n$ of a, b and d in the semi-strong and weak cases. The kernel density estimate is displayed in full line, and the centered Gaussian density with the same variance is plotted in dotted line.

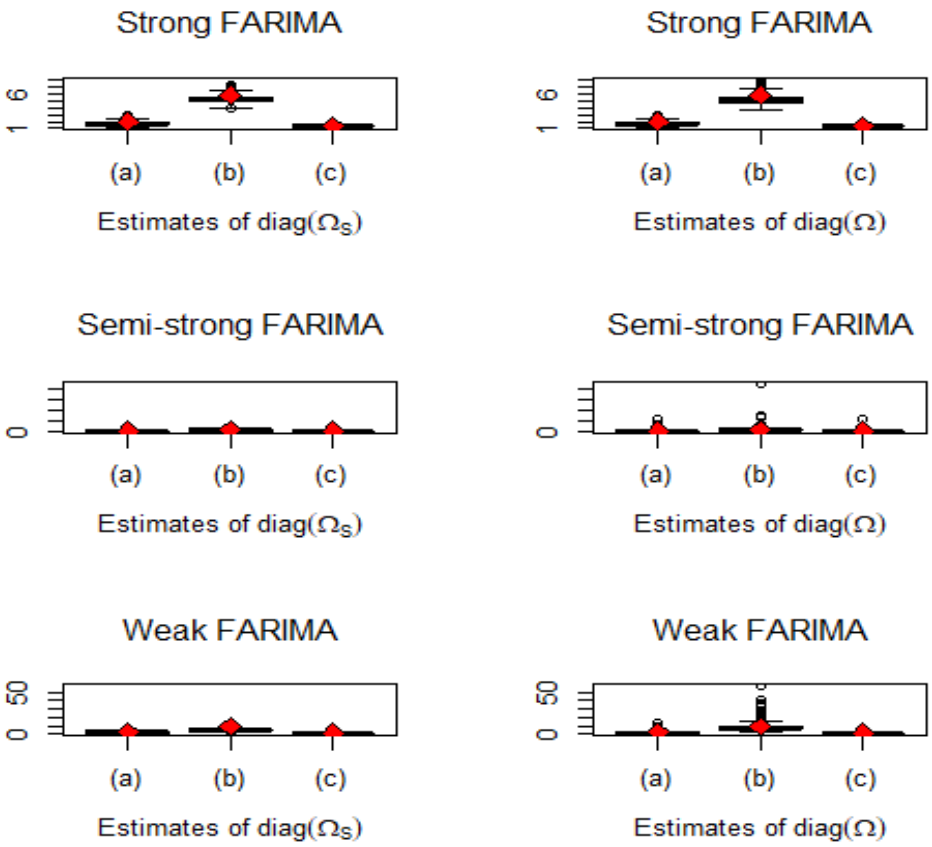


Figure 2.4 – Comparison of standard and modified estimates of the asymptotic variance Ω of the LSE, on the simulated models presented in Figure 2.1. The diamond symbols represent the mean, over $N = 1,000$ replications, of the standardized errors $n(\hat{a}_n + 0.7)^2$ for (a) (1.90 in the strong case and 4.32 (resp. 1.80) in the semi-strong case (resp. in the weak case)), $n(\hat{b}_n + 0.2)^2$ for (b) (5.81 in the strong case and 11.33 (resp. 8.88) in the semi-strong case (resp. in the weak case)) and $n(\hat{d}_n - 0.4)^2$ for (c) (1.28 in the strong case and 2.65 (resp. 1.40) in the semi-strong case (resp. in the weak case)).

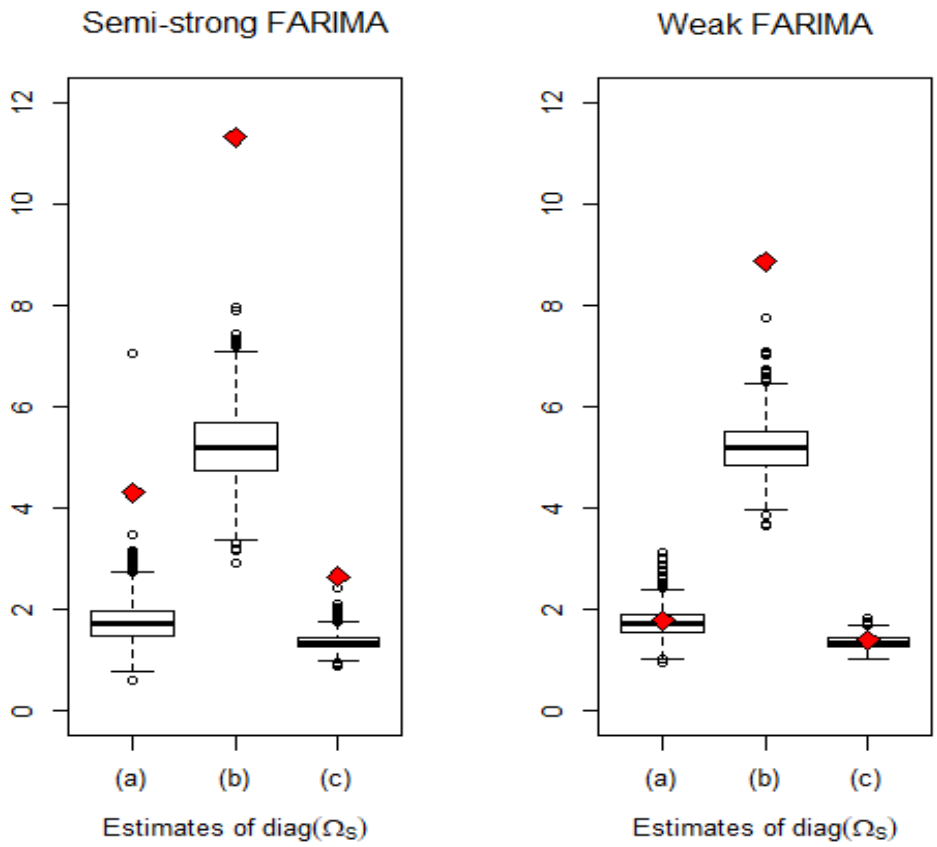


Figure 2.5 – A zoom of the left-middle and left-bottom panels of Figure 2.4

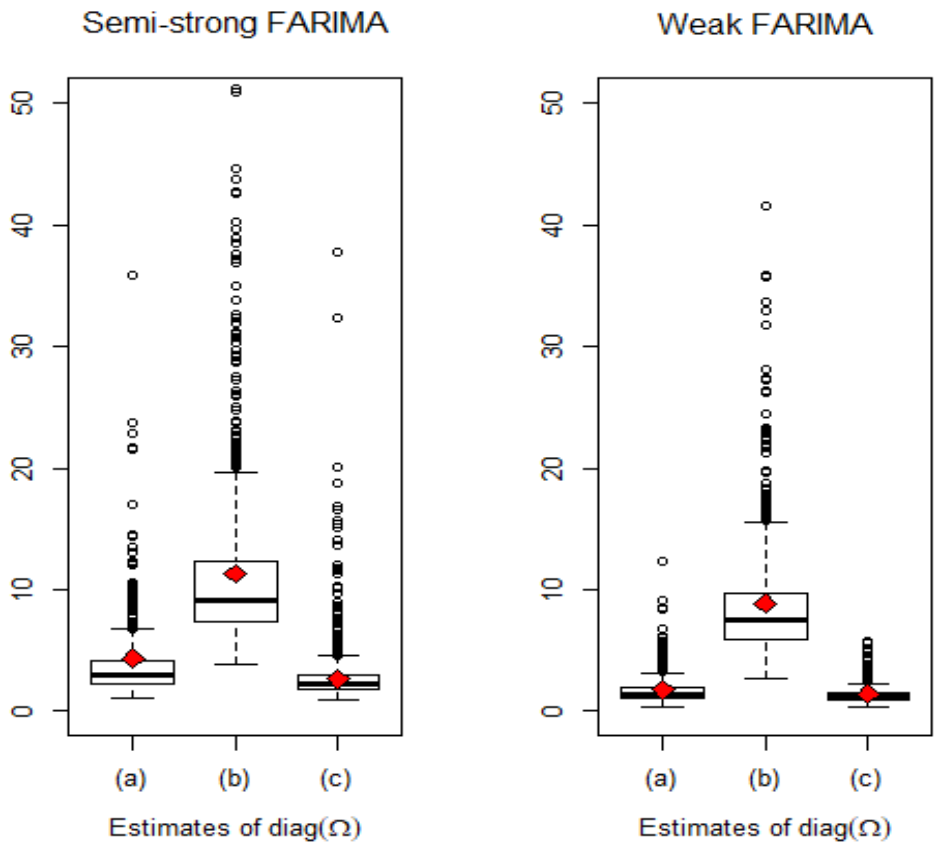


Figure 2.6 – A zoom of the right-middle and right-bottom panels of Figure 2.4

Table 2.1 – Empirical size of standard and modified confidence interval: relative frequencies (in %) of rejection. Modified SN stands for the self-normalized approach. In Modified we use the sandwich estimator of the asymptotic variance Ω of the LSE while in Standard we use $\hat{\Omega}_S$. The number of replications is $N = 1000$.

Model	Length n	Level	Standard			Modified			Modified SN		
			$\hat{a}_n$	$\hat{b}_n$	$\hat{d}_n$	$\hat{a}_n$	$\hat{b}_n$	$\hat{d}_n$	$\hat{a}_n$	$\hat{b}_n$	$\hat{d}_n$
		$\alpha = 1\%$	**2.8**	**2.7**	**2.1**	**3.7**	**3.1**	**2.5**	**2.5**	**2.5**	**2.0**
Strong FARIMA	$n = 200$	$\alpha = 5\%$	**7.1**	**7.3**	5.2	**8.2**	**8.0**	5.4	**8.4**	**6.9**	5.6
		$\alpha = 10\%$	11.8	11.2	8.3	**12.8**	**12.4**	9.5	**14.5**	11.5	10.6
		$\alpha = 1\%$	1.1	1.6	0.7	1.3	1.6	1.0	1.6	1.0	0.8
Strong FARIMA	$n = 2,000$	$\alpha = 5\%$	5.8	**6.9**	5.1	6.1	**6.8**	5.3	5.6	6.4	3.8
		$\alpha = 10\%$	10.9	**13.0**	9.5	11.4	**12.8**	9.5	10.3	**12.4**	9.0
		$\alpha = 1\%$	1.3	1.2	0.7	1.2	1.2	0.8	0.8	1.3	1.2
Strong FARIMA	$n = 5,000$	$\alpha = 5\%$	5.3	4.9	5.2	5.7	5.1	5.4	5.1	4.9	5.3
		$\alpha = 10\%$	10.6	10.3	11.4	10.7	10.2	11.7	11.8	10.8	11.4
		$\alpha = 1\%$	**5.3**	**3.7**	**2.3**	**4.8**	**3.4**	**3.2**	**4.0**	**2.8**	1.4
Semi-strong FARIMA	$n = 200$	$\alpha = 5\%$	**11.2**	**9.5**	6.1	**10.7**	**9.1**	5.9	**11.1**	**8.5**	5.7
		$\alpha = 10\%$	**16.8**	**14.5**	8.6	**16.7**	**14.7**	10.0	**17.1**	**13.9**	10.9
		$\alpha = 1\%$	**6.8**	**7.7**	**7.8**	1.7	0.9	1.4	**2.3**	**2.0**	0.8
Semi-strong FARIMA	$n = 2,000$	$\alpha = 5\%$	**19.5**	**17.7**	**14.9**	**6.5**	5.8	5.5	**8.1**	**6.8**	**6.5**
		$\alpha = 10\%$	**26.5**	**26.7**	**21.5**	**13.5**	11.0	9.9	**14.6**	**12.8**	**12.5**
		$\alpha = 1\%$	**11.2**	**9.8**	**9.4**	1.6	1.5	1.1	1.6	1.3	1.2
Semi-strong FARIMA	$n = 5,000$	$\alpha = 5\%$	**20.8**	**20.2**	**20.9**	6.4	5.7	5.3	5.7	6.2	**7.2**
		$\alpha = 10\%$	**28.2**	**28.4**	**28.4**	**12.2**	9.8	11.3	**12.0**	**13.0**	**13.9**
		$\alpha = 1\%$	**2.6**	**4.4**	1.2	**6.2**	**6.9**	**4.3**	**4.1**	**4.2**	**2.6**
Weak FARIMA	$n = 200$	$\alpha = 5\%$	**6.6**	**11.3**	4.3	**13.8**	**14.6**	**10.2**	**12.0**	**10.5**	**8.9**
		$\alpha = 10\%$	10.9	**18.8**	**7.1**	**20.3**	**21.9**	**16.3**	**17.7**	17.4	**15.4**
		$\alpha = 1\%$	1.1	**5.3**	1.3	1.5	1.2	1.6	1.2	1.1	0.9
Weak FARIMA	$n = 2,000$	$\alpha = 5\%$	5.4	**13.4**	5.8	**7.0**	**6.8**	5.5	5.7	**6.5**	6.4
		$\alpha = 10\%$	11.4	**21.2**	9.6	**12.8**	**12.0**	11.2	11.3	**11.9**	**12.2**
		$\alpha = 1\%$	1.3	**4.6**	1.7	1.2	1.3	1.2	1.3	1.4	0.9
Weak FARIMA	$n = 5,000$	$\alpha = 5\%$	6.3	**14.4**	6.0	**6.7**	6.3	5.9	6.2	6.2	5.0
		$\alpha = 10\%$	11.5	**22.3**	11.6	**12.1**	**12.3**	10.8	10.6	10.9	10.0

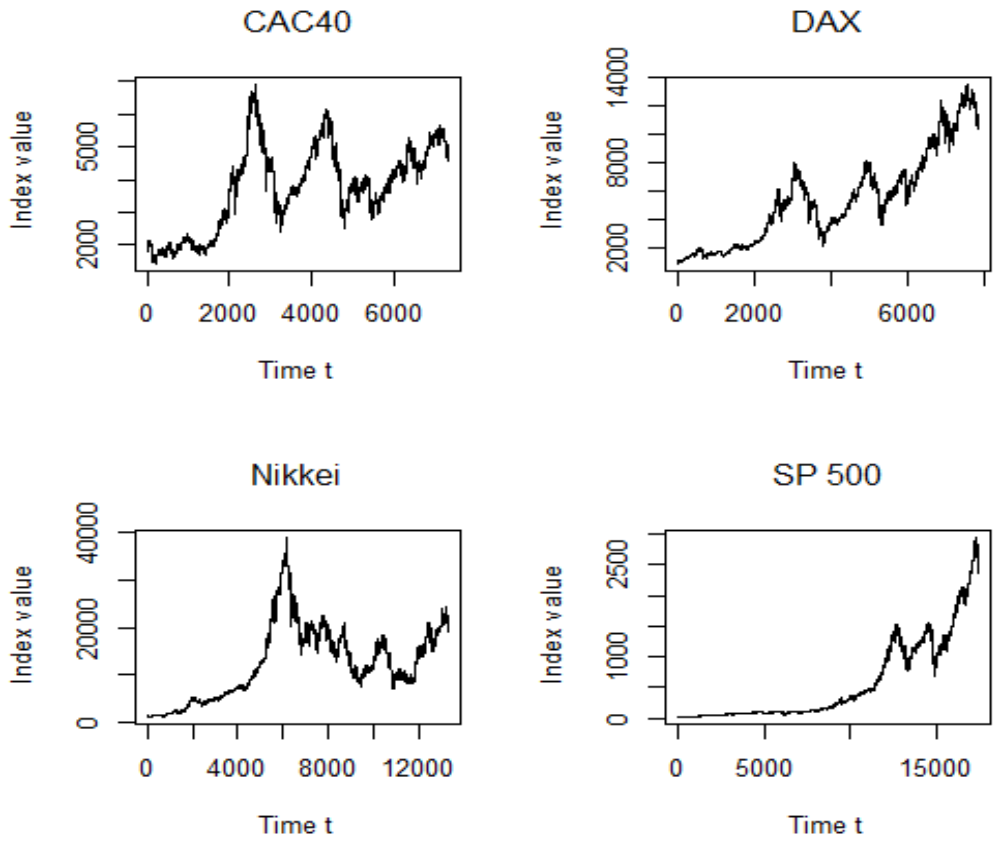


Figure 2.7 – Closing prices of the four stock market indices from the starting date of each index to February 14, 2019.

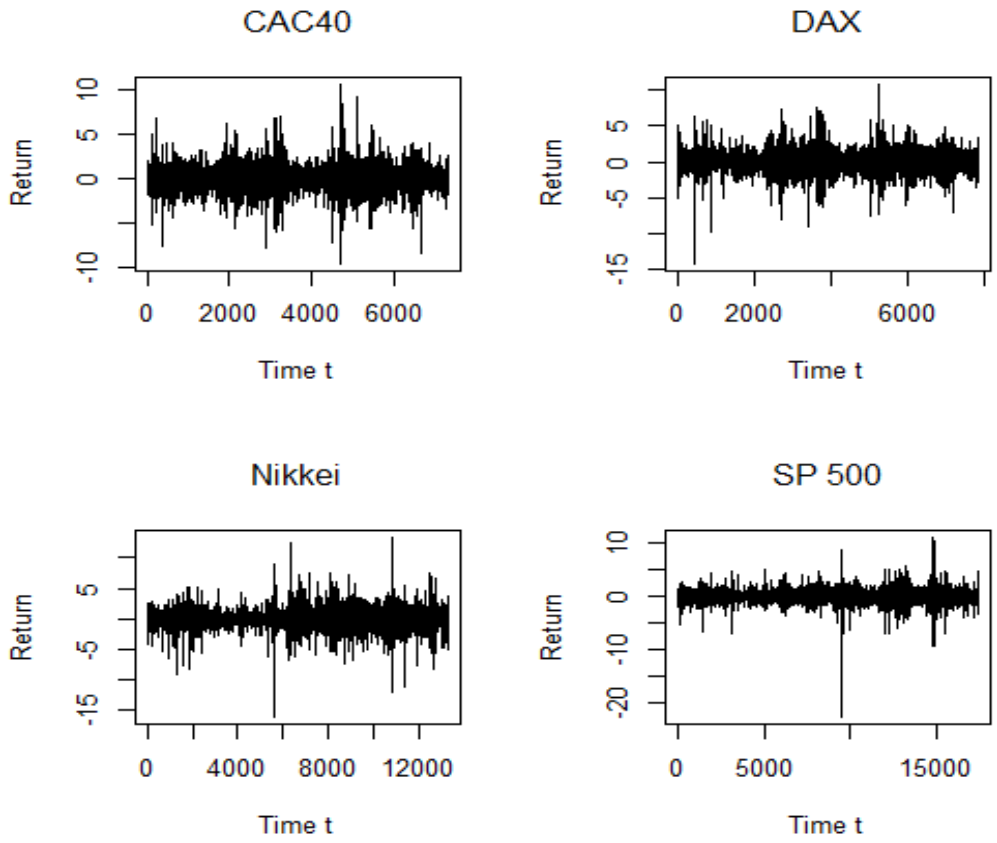


Figure 2.8 – Returns of the four stock market indices from the starting date of each index to February 14, 2019.

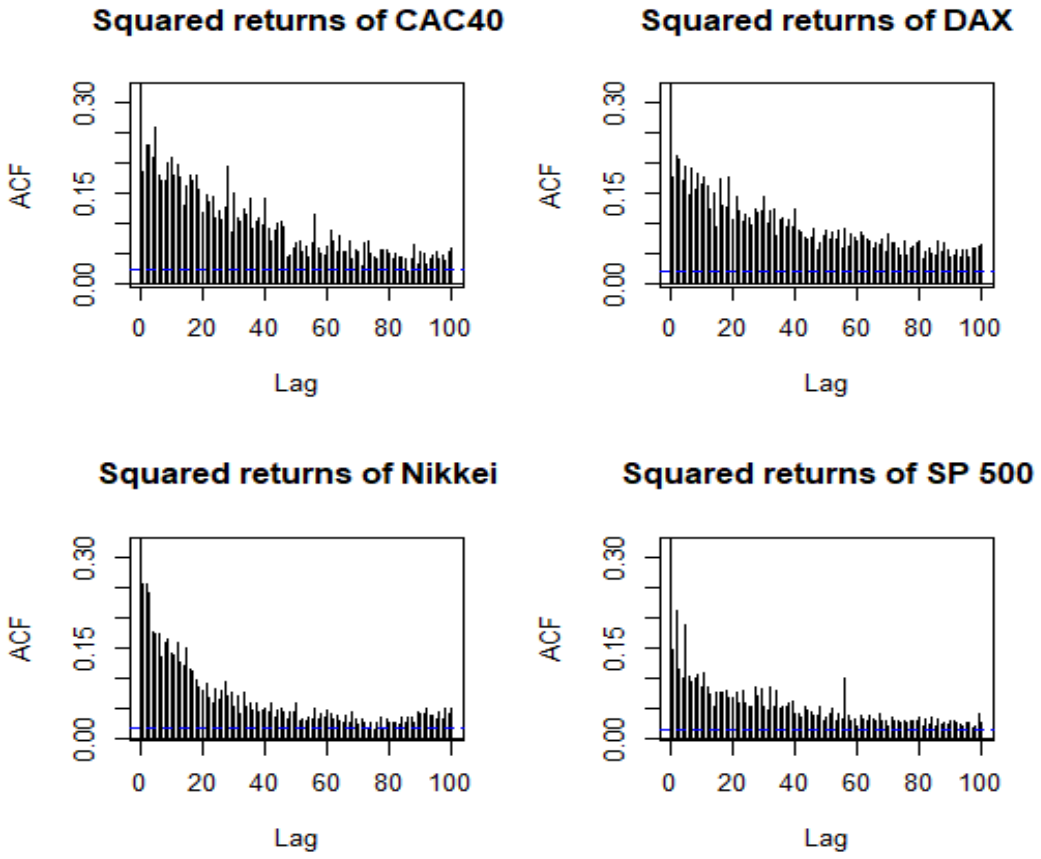


Figure 2.9 – Sample autocorrelations of squared returns of the four stock market indices.

Table 2.2 – Fitting a FARIMA$(1, d, 1)$ model to the squares of the 4 daily returns considered. The corresponding $p-$values are reported in parentheses. The last column presents the estimated residual variance.

Series	Length n	$\hat{a}_n$	$\hat{\theta}_n$	$\hat{b}_n$	$\hat{d}_n$	$\mathrm{Var}(\epsilon_t)$ $\hat{\sigma}^2_\epsilon$
CAC	$n = 7,341$	0.1199 (0.1524)	0.5296 (0.0000)	0.4506 (0.0000)		19.6244×10^{-8}
DAX	$n = 7,860$	0.1598 (0.1819)	0.4926 (0.0000)	0.3894 (0.0000)		25.9383×10^{-8}
Nikkei	$n = 13,318$	-0.0217 (0.9528)	0.1579 (0.6050)	0.3217 (0.0000)		25.6844×10^{-8}
S&P 500	$n = 17,390$	-0.3371 (0.0023)	-0.1795 (0.0227)	0.2338 (0.0000)		22.9076×10^{-8}

Table 2.3 – Modified confidence interval at the asymptotic level $\alpha = 5\%$ for the parameters estimated in Table 2.2. Modified SN stands for the self-normalized approach while Modified corresponds to the confidence interval obtained by using the sandwich estimator of the asymptotic variance Ω of the LSE.

Series	Modified			Modified SN		
	$\hat{a}_n$	$\hat{b}_n$	$\hat{d}_n$	$\hat{a}_n$	$\hat{b}_n$	$\hat{d}_n$
CAC	$[-0.044, 0.284]$	$[0.319, 0.740]$	$[0.251, 0.651]$	$[0.049, 0.193]$	$[0.432, 0.627]$	$[0.358, 0.543]$
DAX	$[-0.075, 0.394]$	$[0.272, 0.714]$	$[0.263, 0.516]$	$[-0.219, 0.537]$	$[0.422, 0.563]$	$[0.158, 0.621]$
Nikkei	$[-0.741, 0.698]$	$[-0.440, 0.756]$	$[0.206, 0.437]$	$[-0.823, 0.779]$	$[-0.407, 0.717]$	$[0.156, 0.488]$
S&P 500	$[-0.554, -0.121]$	$[-0.334, -0.025]$	$[0.162, 0.306]$	$[-0.430, -0.244]$	$[-0.297, -0.062]$	$[0.101, 0.366]$

2.8 Details on the proof of Theorem 2.1

Note that for all $\delta > 0$, Θ_δ is covered by $V(\theta_0)$ and the union of all $V(\overline{\theta})$ such that $\overline{\theta} \in \Theta_\delta \setminus V(\theta_0)$, where $V(\theta_0)$ is a neighbourhood of θ_0 and similarly, for all $\overline{\theta} \in \Theta_\delta \setminus V(\theta_0)$, $V(\overline{\theta})$ is a neighbourhood of $\overline{\theta}$.

Recall that Θ_δ is a compact space, thus, there exists $\theta_1, \theta_2, \ldots, \theta_\ell$, $\ell \in \mathbb{N}^*$ such that Θ_δ is covered by $V(\theta_0), V(\theta_1), \ldots, V(\theta_\ell)$. Then, for large n, we have

$$\inf_{\theta\in\Theta_\delta} Q_n(\theta) = \min_{j\in\{0,1,\ldots,\ell\}} \inf_{\theta\in V(\theta_i)\cap\Theta_\delta} Q_n(\theta) \ .$$

To prove Theorem 2.1, it is sufficient to show that for large n

$$\min_{j\in\{0,1,\ldots,\ell\}} \inf_{\theta\in V(\theta_i)\cap\Theta_\delta} Q_n(\theta) = \inf_{\theta\in V(\theta_0)\cap\Theta_\delta} Q_n(\theta) \ . \tag{2.86}$$

The equation (2.86) will be proved by the absurd.

Let (H) be the assumption that "for all $N \in \mathbb{N}, \exists n \geq N$ such that $\hat{\theta}_n \notin V(\theta_0) \cap \Theta_\delta$".

Under (H), there exists $i \in \{1, 2, \ldots, \ell\}$ such that $\hat{\theta}_n \in V(\theta_i) \cap \Theta_\delta$.

The two following assertions are direct consequences of the definition of the limit inferior and the limit superior of a sequence. For all $\epsilon > 0$, we have

a) $\exists N_1 \in \mathbb{N}$ such that $\forall n \geq N_1$,

$$\inf_{\theta\in V(\theta_0)\cap\Theta_\delta} Q_n(\theta) < \left\{ \limsup \left(\inf_{\theta\in V(\theta_0)\cap\Theta_\delta} Q_n(\theta) \right) \right\} + \epsilon.$$

b) $\exists N_2 \in \mathbb{N}$ such that $\forall n \geq N_2$,

$$\inf_{\theta\in V(\theta_i)\cap\Theta_\delta} Q_n(\theta) > \left\{ \liminf \left(\inf_{\theta\in V(\theta_i)\cap\Theta_\delta} Q_n(\theta) \right) \right\} - \epsilon.$$

The following lemmas allow us to contradict the assumption (H).

Lemma 2.18. *For all $\theta \in \Theta$ and all $t \in \mathbb{Z}$, we have $\epsilon_t(\theta) = \epsilon_t$ a.s. $\Rightarrow \theta = \theta_0$.*

Proof. The fact that the variance σ_ϵ^2 of the linear innovations is supposed to be strictly positive implies that, for any sequence $(r_i)_{i\in\mathbb{N}}$ of real numbers, we have

$$\sum_{i=0}^{\infty} r_i X_{t-i} = 0 \ a.s. \Rightarrow r_i = 0, \forall i \in \mathbb{N}.$$

To see that, it is enough to calculate the variance of the random variable $\sum_{i=0}^{\infty} r_i X_{t-i}$. In fact, using (2.26), we get

$$\begin{aligned}\mathrm{Var}_{\theta_0}\left(\sum_{i=0}^{\infty} r_i X_{t-i}\right) &= \mathrm{Var}_{\theta_0}\left(\sum_{i=0}^{\infty} r_i \sum_{j=0}^{\infty} \eta_j(\theta_0)\epsilon_{t-i-j}\right) = \mathrm{Var}_{\theta_0}\left\{\sum_{k=0}^{\infty}\left(\sum_{i=0}^{k} r_i \eta_{k-i}(\theta_0)\right)\epsilon_{t-k}\right\} \\ &= \sum_{k=0}^{\infty}\left(\sum_{i=0}^{k} r_i \eta_{k-i}(\theta_0)\right)^2 \sigma_\epsilon^2.\end{aligned}$$

Then

$$\sum_{i=0}^{\infty} r_i X_{t-i} = 0 \ a.s. \Rightarrow \mathrm{Var}_{\theta_0}\left(\sum_{i=0}^{\infty} r_i X_{t-i}\right) = 0$$
$$\Rightarrow \sum_{i=0}^{k} r_i \eta_{k-i}(\theta_0) = 0, \ \forall k \geq 0.$$

Observe that, for $k = 0$, we have

$$r_0 \eta_0(\theta_0) = 0 \Leftrightarrow r_0 = 0 \quad (\eta_0(\theta_0) = 1),$$

and for $k = 1$,

$$r_1 \eta_0(\theta_0) = 0 \Leftrightarrow r_1 = 0.$$

By an immediate recurrence, we obtain that

$$\forall i \geq 0 \ , r_i = 0.$$

Thus, in view of (2.25) and for all $\theta \in \Theta$, we have

$$\epsilon_t(\theta) = \epsilon_t \ a.s. \Rightarrow \sum_{j=0}^{\infty} (\gamma_j(\theta) - \gamma_j(\theta_0)) X_{t-j} = 0 \ a.s.$$
$$\Rightarrow (\gamma_j(\theta))_{j\in\mathbb{N}} = (\gamma_j(\theta_0))_{j\in\mathbb{N}}$$
$$\Rightarrow a_\theta(z) b_{\theta_0}(z) = (1-z)^{d_0-d} a_{\theta_0}(z) b_\theta(z), \ \mid z \mid \leq 1.$$

By identifying the coefficients of the power series $a_\theta(z)b_{\theta_0}(z)$ and $a_{\theta_0}(z)b_\theta(z)$, we obtain $d = d_0$.

The proof is completed since it is excluded that both polynomials a_θ and b_θ have greater degrees than a_{θ_0} and b_{θ_0} (respectively), and since a_{θ_0} and b_{θ_0} do not share any root (see Assumption **(A1)**). □

Lemma 2.19. *For any $\theta \in \Theta$ let*

$$O_\infty(\theta) = \mathbb{E}_{\theta_0}\left[\epsilon_t^2(\theta)\right].$$

Then for any $\theta \in \Theta \setminus \{\theta_0\}$, we have

$$O_\infty(\theta_0) = \sigma_\epsilon^2 < O_\infty(\theta).$$

Proof. Observe that for all $t \in \mathbb{Z}$ and for all $\theta \in \Theta \setminus \{\theta_0\}$, the random variable $\epsilon_t(\theta) - \epsilon_t$ belongs to the Hilbert space $H_X(t-1)$, the space generated by $(X_s, s < t)$. This can be seen clearly from (2.25). Therefore the linear innovation ϵ_t is not correlated with $\epsilon_t(\theta) - \epsilon_t$. Thus

$$\mathbb{E}_{\theta_0}\left[\epsilon_t^2(\theta)\right] = \mathbb{E}_{\theta_0}\left[(\epsilon_t(\theta) - \epsilon_t + \epsilon_t)^2\right]$$
$$= \mathbb{E}_{\theta_0}\left[(\epsilon_t(\theta) - \epsilon_t)^2\right] + \mathbb{E}_{\theta_0}\left[\epsilon_t^2\right] + 2Cov_{\theta_0}\left(\epsilon_t, \epsilon_t(\theta) - \epsilon_t\right)$$
$$= \sigma_\epsilon^2 + \mathbb{E}_{\theta_0}\left[(\epsilon_t(\theta) - \epsilon_t)^2\right].$$

Then, we obtain that

$$O_\infty(\theta) \geq \sigma_\epsilon^2,$$

with equality if and only if $\epsilon_t(\theta) = \epsilon_t$ with probability 1. The proof is ended according to the result of Lemma 2.18. □

Lemma 2.20. *For all $\overline{\theta} \in \Theta \backslash \{\theta_0\}$, there exists a neighbourhood $V(\overline{\theta})$ of $\overline{\theta}$ such that $V(\overline{\theta}) \subset \Theta$ and*

$$\liminf_{n\to\infty} \inf_{\theta\in V(\overline{\theta})} O_n(\theta) > \sigma_\epsilon^2, \quad a.s.$$

Proof. Let $V_r(\overline{\theta})$ be the open ball with center $\overline{\theta}$ and radius $1/r$. For all $t \in \mathbb{Z}$, let

$$S_r(t) = \inf_{\theta\in V_r(\overline{\theta})\cap\Theta} \epsilon_t^2(\theta).$$

It is clear that $S_r(t)$ is measurable because it can be written as the infimum over a dense countable subset. Note also that $S_r(t)$ clearly belongs to $\mathbb{L}^1$. This is due to the belonging of $\epsilon_t(\theta)$, for all $\theta \in \Theta$, to $\mathbb{L}^2$. In fact, we have

$$\mathbb{E}\left[\mid S_r(t) \mid\right] = \mathbb{E}\left[\inf_{\theta\in V_r(\overline{\theta})\cap\Theta} \epsilon_t^2(\theta)\right] \leq \inf_{\theta\in V_r(\overline{\theta})\cap\Theta} \mathbb{E}\left[\epsilon_t^2(\theta)\right] < \infty.$$

As n tends to infinity, by the ergodic theorem, we have almost surely

$$\inf_{\theta\in V_r(\overline{\theta})\cap\Theta} O_n(\theta) \geq \frac{1}{n}\sum_{t=1}^{n} S_r(t) \to \mathbb{E}_{\theta_0}\left[S_r(t)\right].$$

Since $\epsilon_t^2(\theta)$ is a smooth function of θ, $S_r(t)$ increases to $\epsilon_t^2(\overline{\theta})$ as r tends to infinity. By Beppo-Levi's theorem we obtain

$$\lim_{r\to\infty} \mathbb{E}_{\theta_0}\left[S_r(t)\right] = \mathbb{E}_{\theta_0}\left[\lim_{r\to\infty} S_r(t)\right] = \mathbb{E}_{\theta_0}\left[\epsilon_t^2(\overline{\theta})\right] = O_\infty(\overline{\theta}).$$

From Lemma 2.19 we have

$$\liminf_{r\to\infty} \liminf_{n\to\infty} \inf_{\theta\in V_r(\overline{\theta})} O_n(\theta) \geq O_\infty(\overline{\theta}) > \sigma_\epsilon^2,$$

and thus the lemma is proved. □

Lemma 2.21. *We have*

$$\mathbb{E}_{\theta_0}\left[\sup_{\theta\in\Theta_\delta} \epsilon_t^2(\theta)\right] < \infty.$$

Proof. For all $t \in \mathbb{Z}$, in view of (2.25), we have

$$\sup_{\theta\in\Theta_\delta} |\epsilon_t(\theta)| = \sup_{\theta\in\Theta_\delta} \left| \sum_{j\geq 0} \gamma_j(\theta) X_{t-j} \right| \leq \sup_{\theta\in\Theta_\delta} \sum_{j\geq 0} |\gamma_j(\theta)|\, |X_{t-j}| \leq \sum_{j\geq 0} \sup_{\theta\in\Theta_\delta} |\gamma_j(\theta)|\, |X_{t-j}|\,.$$

From (2.27), there exists a constant $C > 0$ such that, for large j and for any $\theta \in \Theta_\delta$, we have $\sup_{\theta\in\Theta_\delta} |\gamma_j(\theta)| \leq C j^{-d_1-1}$. It is clear that $\sum_{j\geq 0} \sup_{\theta\in\Theta_\delta} |\gamma_j(\theta)| < \infty$ and $\sup_t \mathbb{E}\left[X_t^2\right] < \infty$. Then, by stationarity and Cauchy-Schwarz inequality, we have

$$\mathbb{E}_{\theta_0}\left[\left| \sum_{j\geq 0} \sup_{\theta\in\Theta_\delta} |\gamma_j(\theta)|\, |X_{t-j}| \right|^2 \right] \leq \left(\sum_{j\geq 0} \sup_{\theta\in\Theta_\delta} |\gamma_j(\theta)| \right)^2 \sup_{t\in\mathbb{Z}} \mathbb{E}\left[X_t^2\right] < \infty.$$

Since $\forall t \in \mathbb{Z}$, $\sup_{\theta\in\Theta_\delta} \epsilon_t^2(\theta) \leq \left(\sup_{\theta\in\Theta_\delta} |\epsilon_t(\theta)|\right)^2$, we then deduce that

$$\mathbb{E}_{\theta_0}\left[\sup_{\theta\in\Theta_\delta} \left\{\epsilon_t^2(\theta)\right\} \right] \leq \mathbb{E}_{\theta_0}\left[\left(\sup_{\theta\in\Theta_\delta} |\epsilon_t(\theta)| \right)^2 \right] < \infty.$$

The conclusion follows. □

Lemma 2.22. *We have almost surely*

$$\lim_{n\to\infty} \sup_{\theta\in\Theta_\delta} |O_n(\theta) - Q_n(\theta)| = 0.$$

Proof. By the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
\sup_{\theta\in\Theta_\delta} |Q_n(\theta) - O_n(\theta)| &= \sup_{\theta\in\Theta_\delta} \left| \frac{1}{n}\sum_{t=1}^n \left\{\tilde{\epsilon}_t(\theta) - \epsilon_t(\theta) + \epsilon_t(\theta)\right\}^2 - \frac{1}{n}\sum_{t=1}^n \epsilon_t^2(\theta) \right| \\
&= \sup_{\theta\in\Theta_\delta} \left| \frac{1}{n}\sum_{t=1}^n \left\{\tilde{\epsilon}_t(\theta) - \epsilon_t(\theta)\right\}^2 + \frac{2}{n}\sum_{t=1}^n \epsilon_t(\theta)\left[\tilde{\epsilon}_t(\theta) - \epsilon_t(\theta)\right] \right| \\
&\leq \sup_{\theta\in\Theta_\delta} \left\{ \frac{1}{n}\sum_{t=1}^n |\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)|^2 + \frac{2}{n}\sum_{t=1}^n |\epsilon_t(\theta)|\, |\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)| \right\} \\
&\leq \frac{1}{n}\sum_{t=1}^n \sup_{\theta\in\Theta_\delta} \left\{ |\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)|^2 \right\} + \frac{2}{n}\sum_{t=1}^n \sup_{\theta\in\Theta_\delta} |\epsilon_t(\theta)| \sup_{\theta\in\Theta_\delta} |\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)| \\
&\leq \frac{1}{n}\sum_{t=1}^n \sup_{\theta\in\Theta_\delta} \left\{ |\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)|^2 \right\} \\
&\quad + 2\left\{ \left(\frac{1}{n}\sum_{t=1}^n \left[\sup_{\theta\in\Theta_\delta} |\epsilon_t(\theta)| \right]^2 \right) \left(\frac{1}{n}\sum_{t=1}^n \left[\sup_{\theta\in\Theta_\delta} |\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)| \right]^2 \right) \right\}^{1/2} \\
&\leq \frac{1}{n}\sum_{t=1}^n \left\{ \sup_{\theta\in\Theta_\delta} |\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)| \right\}^2
\end{aligned}$$

$$+2\left\{\left(\frac{1}{n}\sum_{t=1}^{n}\left[\sup_{\theta\in\Theta_\delta}|\epsilon_t(\theta)|\right]^2\right)\left(\frac{1}{n}\sum_{t=1}^{n}\left[\sup_{\theta\in\Theta_\delta}|\epsilon_t(\theta)-\tilde{\epsilon}_t(\theta)|\right]^2\right)\right\}^{1/2}.$$

The conclusion follows from Cesàro's Lemma, Lemma 2.4, Lemma 2.21 and the ergodic theorem. □

Lemma 2.23. *For any $\overline{\theta}\in\Theta_\delta\setminus\{\theta_0\}$, there exists a neighbourhood $V\left(\overline{\theta}\right)$ of $\overline{\theta}$ such that $V\left(\overline{\theta}\right)\subset\Theta$ and*

$$\liminf_{n\to\infty}\inf_{\theta\in V\left(\overline{\theta}\right)}Q_n(\theta)>\sigma_\epsilon^2,\ a.s.$$

Proof. Observe that

$$\inf_{\theta\in V\left(\overline{\theta}\right)}Q_n(\theta)=\inf_{\theta\in V\left(\overline{\theta}\right)}\{Q_n(\theta)-O_n(\theta)+O_n(\theta)\}\geq\inf_{\theta\in V\left(\overline{\theta}\right)}O_n(\theta)-\sup_{\theta\in V\left(\overline{\theta}\right)}|O_n(\theta)-Q_n(\theta)|,$$

since from Lemma 2.20 $V\left(\overline{\theta}\right)$ can be chosen as being included in $\Theta_{\delta/2}$. Clearly, we have

$$\inf_{\theta\in V\left(\overline{\theta}\right)}Q_n(\theta)\geq\inf_{\theta\in V\left(\overline{\theta}\right)}O_n(\theta)-\sup_{\theta\in\Theta_{\delta/2}}|O_n(\theta)-Q_n(\theta)|.$$

The conclusion follows from Lemma 2.20 and Lemma 2.22.

□

Let us now return to the proof by the absurd of the Theorem. Using the two assertions *a*) and *b*) given before, under the assumption (H) we have

$$\liminf\left(\inf_{\theta\in V(\theta_i)\cap\Theta_\delta}Q_n(\theta)\right)-\epsilon<\inf_{\theta\in V(\theta_i)\cap\Theta_\delta}Q_n(\theta,$$

for all $n\geq N$ such that $N=\sup(N_1,N_2)$. Moreover, we have

$$\inf_{\theta\in V(\theta_i)\cap\Theta_\delta}Q_n(\theta)=\inf_{\theta\in\Theta_\delta}Q_n(\theta)\leq\inf_{\theta\in V(\theta_0)\cap\Theta_\delta}Q_n(\theta)<\limsup\left(\inf_{\theta\in V(\theta_0)\cap\Theta_\delta}Q_n(\theta)\right)+\epsilon,$$

and thus

$$\liminf\left(\inf_{\theta\in V(\theta_i)\cap\Theta_\delta}Q_n(\theta)\right)<\limsup\left(\inf_{\theta\in V(\theta_0)\cap\Theta_\delta}Q_n(\theta)\right)+2\epsilon.$$

Note also that, for all $k\in\mathbb{N}^*$, we have

$$\inf_{\theta\in V(\theta_0)\cap\Theta_\delta}Q_k(\theta)\leq Q_k(\theta_0).$$

Then, for all $n\in\mathbb{N}^*$, we obtain that

$$\sup_{k\geq n}\left\{\inf_{\theta\in V(\theta_0)\cap\Theta_\delta}Q_k(\theta)\right\}\leq\sup_{k\geq n}Q_k(\theta_0).$$

Hence

$$\lim_{n\to\infty}\sup_{k\geq n}\left\{\inf_{\theta\in V(\theta_0)\cap\Theta_\delta} Q_k(\theta)\right\}\leq \lim_{n\to\infty}\sup_{k\geq n} Q_k(\theta_0).$$

According to the definition of the limit superior, from Lemma 2.22 and the ergodic theorem, we have

$$\lim_{n\to\infty}\sup_{k\geq n} Q_k(\theta_0)=\limsup_{n\to\infty} Q_n(\theta_0)=\lim_{n\to\infty} Q_n(\theta_0)=\lim_{n\to\infty} O_n(\theta_0)=\sigma_\epsilon^2.$$

Thus

$$\limsup_{n\to\infty}\left(\inf_{\theta\in V(\theta_0)\cap\Theta_\delta} Q_n(\theta)\right)\leq\sigma_\epsilon^2,$$

and we therefore have

$$\liminf_{n\to\infty}\left(\inf_{\theta\in V(\theta_i)\cap\Theta_\delta} Q_n(\theta)\right)<\sigma_\epsilon^2+2\epsilon.$$

This contradicts the result of Lemma 2.23. The proof of Theorem 2.1 is completed.

Bibliography

Anderson, T. W. (1971). *The statistical analysis of time series*. John Wiley & Sons, Inc., New York-London-Sydney.

Baillie, R. T., Chung, C.-F., and Tieslau, M. A. (1996). Analysing inflation by the fractionally integrated ARFIMA-GARCH model. *Journal of applied econometrics*, 11(1):23–40.

Beran, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *J. Roy. Statist. Soc. Ser. B*, 57(4):659–672.

Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2013). *Long-memory processes*. Springer, Heidelberg. Probabilistic properties and statistical methods.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *Ann. Statist.*, 2:489–502. Collection of articles dedicated to Jerzy Neyman on his 80th birthday.

Boubacar Maïnassara, Y. (2009). *Estimation, validation et identification des modèles ARMA faibles multivariés*. PHD thesis of Université Lille 3.

Boubacar Maïnassara, Y. (2012). Selection of weak VARMA models by modified Akaike's information criteria. *J. Time Series Anal.*, 33(1):121–130.

Boubacar Mainassara, Y., Carbon, M., and Francq, C. (2012). Computing and estimating information matrices of weak ARMA models. *Comput. Statist. Data Anal.*, 56(2):345–361.

Boubacar Mainassara, Y. and Francq, C. (2011). Estimating structural VARMA models with uncorrelated but non-independent error terms. *J. Multivariate Anal.*, 102(3):496–505.

Boubacar Maïnassara, Y. and Kokonendji, C. C. (2016). Modified Schwarz and Hannan-Quinn information criteria for weak VARMA models. *Stat. Inference Stoch. Process.*, 19(2):199–217.

Boubacar Maïnassara, Y. and Saussereau, B. (2018). Diagnostic checking in multivariate arma models with dependent errors using normalized residual autocorrelations. *J. Amer. Statist. Assoc.*, 113(524):1813–1827.

Brillinger, D. R. (1981). *Time series: data analysis and theory*, volume 36. Siam.

Chiu, S.-T. (1988). Weighted least squares estimators on the frequency domain for the parameters of a time series. *Ann. Statist.*, 16(3):1315–1326.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.*, 17(4):1749–1766.

Davidson, J. (1994). *Stochastic limit theory*. Advanced Texts in Econometrics. The Clarendon Press, Oxford University Press, New York. An introduction for econometricians.

Ding, Z., Granger, C. W., and Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1:83–106.

Doukhan, P. and León, J. (1989). Cumulants for stationary mixing random sequences and applications to empirical spectral density. *Probab. Math. Stat*, 10:11–26.

Fan, J. and Yao, Q. (2008). *Nonlinear time series: nonparametric and parametric methods*. Springer Science & Business Media.

Fox, R. and Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.*, 14(2):517–532.

Francq, C., Roy, R., and Zakoïan, J.-M. (2005). Diagnostic checking in ARMA models with uncorrelated errors. *J. Amer. Statist. Assoc.*, 100(470):532–544.

Francq, C. and Zakoïan, J.-M. (1998). Estimating linear representations of nonlinear processes. *J. Statist. Plann. Inference*, 68(1):145–165.

Francq, C. and Zakoïan, J.-M. (2005). Recent results for linear time series models with non independent innovations. In *Statistical modeling and analysis for complex data problems*, volume 1 of *GERAD 25th Anniv. Ser.*, pages 241–265. Springer, New York.

Francq, C. and Zakoïan, J.-M. (2007). HAC estimation and strong linearity testing in weak ARMA models. *J. Multivariate Anal.*, 98(1):114–144.

Francq, C. and Zakoïan, J.-M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley.

Giraitis, L. and Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probab. Theory Related Fields*, 86(1):87–104.

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal.*, 1(1):15–29.

Hallin, M., Taniguchi, M., Serroukh, A., and Choy, K. (1999). Local asymptotic normality for regression models with long-memory disturbance. *Ann. Statist.*, 27(6):2054–2080.

Hauser, M. and Kunst, R. (1998). Fractionally integrated models with arch errors: with an application to the swiss 1-month euromarket interest rate. *Review of Quantitative Finance and Accounting*, 10(1):95–113.

Herrndorf, N. (1984). A functional central limit theorem for weakly dependent sequences of random variables. *Ann. Probab.*, 12(1):141–153.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68(1):165–176.

Keenan, D. M. (1987). Limiting behavior of functionals of higher-order sample cumulant spectra. *Ann. Statist.*, 15(1):134–151.

Kuan, C.-M. and Lee, W.-M. (2006). Robust M tests without consistent estimation of the asymptotic covariance matrix. *J. Amer. Statist. Assoc.*, 101(475):1264–1275.

Ling, S. and Li, W. K. (1997). On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *J. Amer. Statist. Assoc.*, 92(439):1184–1194.

Lobato, I. N. (2001). Testing that a dependent process is uncorrelated. *J. Amer. Statist. Assoc.*, 96(455):1066–1076.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer-Verlag, Berlin.

Magnus, J. R. and Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester. Revised reprint of the 1988 original.

Palma, W. (2007). *Long-memory time series*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ. Theory and methods.

Shao, X. (2010a). Corrigendum: A self-normalized approach to confidence interval construction in time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(5):695–696.

Shao, X. (2010b). Nonstationarity-extended Whittle estimation. *Econometric Theory*, 26(4):1060–1087.

Shao, X. (2010c). A self-normalized approach to confidence interval construction in time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(3):343–366.

Shao, X. (2011). Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models. *Econometric Theory*, 27(2):312–343.

Shao, X. (2012). Parametric inference in stationary time series models with dependent errors. *Scand. J. Stat.*, 39(4):772–783.

Shao, X. (2015). Self-normalization for time series: a review of recent developments. *J. Amer. Statist. Assoc.*, 110(512):1797–1817.

Taniguchi, M. (1982). On estimation of the integrals of the fourth order cumulant spectral density. *Biometrika*, 69(1):117–122.

Taqqu, M. S. and Teverovsky, V. (1997). Robustness of Whittle-type estimators for time series with long-range dependence. *Comm. Statist. Stochastic Models*, 13(4):723–757. Heavy tails and highly volatile phenomena.

Tong, H. (1990). *Non-linear time series: a dynamical system approach*. Oxford University Press.

Whittle, P. (1953). Estimation and information in stationary time series. *Ark. Mat.*, 2:423–434.

Zhu, K. and Li, W. K. (2015). A bootstrapped spectral test for adequacy in weak ARMA models. *J. Econometrics*, 187(1):113–130.

Chapter 3

Diagnostic checking in FARIMA models with uncorrelated but non-independent error terms

Contents

3.1 Introduction **131**
3.2 Assumptions and estimation procedure **133**
3.3 Diagnostic checking in weak FARIMA models **136**
3.3.1 Asymptotic distribution of the residual autocorrelations 136
3.3.2 Modified version of the portmanteau test 139
3.3.3 Self-normalized asymptotic distribution of the residual autocorrelations 141
3.3.4 Example of explicit calculation of $\Sigma_{\hat{\rho}_m}$ and C_m 143
3.4 Numerical illustrations **148**
3.4.1 Simulation studies and empirical sizes 148
3.4.2 Empirical power 149
3.4.3 Illustrative example 150
3.5 Proofs **152**
3.5.1 Preliminary results 152
3.5.2 Proof of Proposition 3.1 154
3.5.3 Proof of Theorem 3.1 156
3.5.4 Proof of Proposition 3.2 165
3.5.5 Proof of Theorem 3.4 166
3.5.6 Proof of Theorem 3.5 170
3.6 Figures and tables **171**
3.7 Supplementary material: additional Monte Carlo experiments **194**
3.8 Another one **201**
Bibliographie **205**

Abstract: This work considers the problem of modified portmanteau tests for testing the adequacy of FARIMA models under the assumption that the errors are uncorrelated but not necessarily independent (*i.e.* weak FARIMA). We first study the joint distribution of the least squares estimator and the noise empirical autocovariances. We then derive the asymptotic distribution of residual empirical autocovariances and autocorrelations. We deduce the asymptotic distribution of the Ljung-Box (or Box-Pierce) modified portmanteau statistics for weak FARIMA models. We also propose another method based on a self-normalization approach to test the adequacy of FARIMA models. Finally some simulation studies are presented to corroborate our theoretical work. An application to the Standard & Poor's 500 and Nikkei returns also illustrate the practical relevance of our theoretical results.
AMS 2000 subject classifications: Primary 62M10, 62F03, 62F05; secondary 91B84, 62P05.
Keywords and phrases: Nonlinear processes; Long-memory processes; Weak FARIMA models; Least squares estimator; Box-Pierce and Ljung-Box portmanteau tests; Residual autocorrelations; Self-normalization.

3.1 Introduction

To model the long memory phenomenon, a widely used model is the fractional autoregressive integrated moving average (FARIMA, for short) model (see for instance Granger and Joyeux [1980], Fox and Taqqu [1986], Dahlhaus [1989], Hosking [1981], Beran et al. [2013], Palma [2007], among others). This model plays an important role in many scientific disciplines and applied fields such as hydrology, climatology, economics, finance, to name a few.

We consider a centered stationary process $X := (X_t)_{t\in\mathbb{Z}}$ satisfying a FARIMA(p, d_0, q) representation of the form

$$a(L)(1-L)^{d_0}X_t = b(L)\epsilon_t, \tag{3.1}$$

where d_0 is the long memory parameter, L stands for the back-shift operator and $a(L) = 1-\sum_{i=1}^{p} a_i L^i$, respectively $b(L) = 1-\sum_{i=1}^{q} b_i L^i$, is the autoregressive, respectively the moving average, operator. These operators represent the short memory part of the model (by convention $a_0 = b_0 = 1$). In the standard situation $\epsilon := (\epsilon_t)_{t\in\mathbb{Z}}$ is assumed to be a sequence of independent and identically distributed (iid for short) random variables with zero mean and with a common variance. In this standard framework, ϵ is said to be a *strong white noise* and the representation (3.1) is called a strong FARIMA(p, d_0, q) process. In contrast with this previous definition, the representation (3.1) is said to be a weak FARIMA(p, d_0, q) if the noise process ϵ is a *weak white noise*, that is, if it satisfies

(A0): $\mathbb{E}(\epsilon_t) = 0$, $\mathrm{Var}\,(\epsilon_t) = \sigma_0^2$ and $\mathrm{Cov}\,(\epsilon_t, \epsilon_{t-h}) = 0$ for all $t \in \mathbb{Z}$ and all $h \neq 0$.

A strong white noise is obviously a weak white noise because independence entails uncorrelatedness. Of course the converse is not true. The strong FARIMA model was introduced by Hosking [1981]. The particular strong FARIMA$(0, d_0, 0)$ process was discussed by Granger and

Joyeux [1980]. To ensure the stationarity and the invertibility of the model defined by (3.1), we assume that $0 < d_0 < 1/2$ and all roots of $a(z)b(z) = 0$ are outside the unit disk (see Granger and Joyeux [1980] and Hosking [1981] for details). It is also assumed that $a(z)$ and $b(z)$ have no common factors in order to insure unique identifiability of the parameters.

The validity of the different steps of the traditional methodology of Box and Jenkins (identification, estimation and validation) depends on the noise properties. After estimating the FARIMA process, the next important step in the modeling consists in checking if the estimated model fits satisfactorily the data. Thus, under the null hypothesis that the model has been correctly identified, the residuals $(\hat{\epsilon}_t)$ are approximately a white noise. This adequacy checking step allows to validate or invalidate the choice of the orders p and q. The choice of p and q is particularly important because the number of parameters $(p + q + 1)$ quickly increases with p and q, which entails statistical difficulties. In particular, the selection of too large orders p and q may introduce terms that are not necessarily relevant in the model. In other terms, overidentification generally leads to a loss of precision in parameter estimation. Conversely, the selection of too small orders p and q causes loss of some information, that can be detected by the correlation of the residuals.

Thus it is important to check the validity of a FARIMA(p, d_0, q) model, for given orders p and q. Based on the residual empirical autocorrelation, Box and Pierce [1970] have proposed a goodness-of-fit test, the so-called portmanteau test, for strong ARMA models. The intuition behind these portmanteau tests is that if a given time series model with iid innovation is appropriate for the data at hand, the autocorrelations of the residuals $\hat{\epsilon}_t$ should be close to zero, which is the theoretical value of the autocorrelations of ϵ_t (see Assumption **(A0)** below). A modification of the test of Box and Pierce [1970] has been proposed by Ljung and Box [1978] which is nowadays one of the most popular diagnostic checking tools in strong ARMA modeling of time series. A modified portmanteau test statistic was proposed by Li and McLeod [1986] for checking the overall significance of the residual autocorrelations of a strong FARIMA(p, d_0, q) model. All these above test statistics have been obtained under the iid assumption on the noise and they may be invalid when the series is uncorrelated but dependent (see Romano and Thombs [1996], Francq et al. [2005], Boubacar Maïnassara and Saussereau [2018], Zhu and Li [2015], Lobato et al. [2001], Lobato et al. [2002], to name a few).

As mentioned above, the works on the portmanteau statistic are generally performed under the assumption that the errors ϵ_t are independent (see for instance Li and McLeod [1986]). This independence assumption is often considered too restrictive by practitioners. It precludes conditional heteroscedasticity and/or other forms of nonlinearity (see Francq and Zakoïan [2005] for a review on weak univariate ARMA models) which can not be generated by FARIMA models with iid noises. [1] Relaxing this independence assumption allows to cover linear representations of general nonlinear processes and to extend the range of application of the FARIMA models.

This paper is devoted to the problem of the validation step of weak FARIMA processes.

1. To cite few examples of nonlinear processes, let us mention: the generalized autoregressive conditional heteroscedastic (GARCH) model (see Francq and Zakoïan [2010]), the self-exciting threshold autoregressive (SETAR), the smooth transition autoregressive (STAR), the exponential autoregressive (EXPAR), the bilinear, the random coefficient autoregressive (RCA), the functional autoregressive (FAR) (see Tong [1990] and Fan and Yao [2008], for references on these nonlinear time series models).

For the asymptotic theory of weak FARIMA model validation, recently Shao [2011] studied the diagnostic checking for long memory time series models with nonparametric conditionally heteroscedastic martingale difference errors. This author also generalized the test statistic based on the kernel-based spectral proposed by Hong [1996] under weak assumptions on the innovation process. As mentioned by the author, his work is no longer valid for $d_0 \in [1/4, 1/2[$ (see Remark 3.3 in Shao [2011]). Note also that Ling and Li [1997] have studied the Box and Pierce [1970] type test for FARIMA-GARCH models by assuming a parametric form for the GARCH model.

To our knowledge, it does not exist any diagnostic checking methodology for FARIMA models when the (possibly dependent) error is subject to unknown conditional heteroscedasticity. We think that this is due to the difficulty that arises when one has to estimate the asymptotic covariance matrix. In our paper, thanks to the asymptotic results obtained by Boubacar Maïnassara et al. [2019], we are able to extend for weak FARIMA models the diagnostic checking methodology proposed by Francq et al. [2005] as well as the self-normalized approach proposed by Boubacar Maïnassara and Saussereau [2018]. We strength the fact that, contrarily to Shao [2011], our results are valid for $d_0 \in]0, 1/2[$.

The paper is organized as follows. In Section 3.2, we recall the results on the least-squares estimator asymptotic distribution of weak FARIMA models obtained by Boubacar Maïnassara et al. [2019]. In Section 3.3, a modified version of the portmanteau test is proposed thanks to the investigation of the asymptotic distribution of the residual autocorrelations. Our first main result is stated in Theorem 3.1. The second main result of this section is obtained in Theorem 3.5 by means of a self-normalized approach. Two examples are also proposed in order to illustrate our results. Some numerical illustrations are gathered in Section 3.4. They corroborate our theoretical work. An application to the Standard & Poor's 500 and Nikkei returns also illustrate the practical relevance of our theoretical results. All our proofs are given in Section 3.5 and figures and tables are brought together in Section 3.6.

3.2 Assumptions and estimation procedure

In this section, we recall the results on the least-squares estimator asymptotic distribution of weak FARIMA models obtained by Boubacar Maïnassara et al. [2019] in order to have a self-containing paper.

Let Θ^* be the parameter space

$$\Theta^* := \Big\{ (\theta_1, \theta_2, \ldots, \theta_{p+q}) \in \mathbb{R}^{p+q}, \text{ where } a_\theta(z) = 1 - \sum_{i=1}^{p} \theta_i z^i, \text{ and } b_\theta(z) = 1 - \sum_{j=1}^{q} \theta_{p+j} z^j \text{ have all their zeros outside the unit disk and have no zero in common} \Big\}.$$

Denote by Θ the cartesian product $\Theta^* \times [d_1, d_2]$, where $[d_1, d_2] \subset]0, 1/2[$ with $d_1 \leq d_0 \leq d_2$. The unknown parameter of interest $\theta_0 = (a_1, a_2, \ldots, a_p, b_1, b_2, \ldots, b_q, d_0)'$ is supposed to belong to the parameter space Θ.

The fractional difference operator $(1-L)^{d_0}$ is defined, using the generalized binomial series, by

$$(1-L)^{d_0} = \sum_{j\geq 0} \alpha_j(d_0) L^j,$$

where for all $j \geq 0$, $\alpha_j(d_0) = \Gamma(j-d_0)/\{\Gamma(j+1)\Gamma(-d_0)\}$ and $\Gamma(\cdot)$ is the Gamma function. Using the Stirling formula we obtain that for large j, $\alpha_j(d_0) \sim j^{-d_0-1}/\Gamma(-d_0)$ (one refers to Beran et al. [2013] for further details).

For all $\theta \in \Theta$ we define $(\epsilon_t(\theta))_{t\in\mathbb{Z}}$ as the second order stationary process which is the solution of

$$\epsilon_t(\theta) = \sum_{j\geq 0} \alpha_j(d) X_{t-j} - \sum_{i=1}^{p} \theta_i \sum_{j\geq 0} \alpha_j(d) X_{t-i-j} + \sum_{j=1}^{q} \theta_{p+j} \epsilon_{t-j}(\theta). \tag{3.2}$$

Observe that, for all $t \in \mathbb{Z}$, $\epsilon_t(\theta_0) = \epsilon_t$ a.s. Given a realization $X_1, \ldots, X_n$ of length n, $\epsilon_t(\theta)$ can be approximated, for $0 < t \leq n$, by $\tilde{\epsilon}_t(\theta)$ defined recursively by

$$\tilde{\epsilon}_t(\theta) = \sum_{j=0}^{t-1} \alpha_j(d) X_{t-j} - \sum_{i=1}^{p} \theta_i \sum_{j=0}^{t-i-1} \alpha_j(d) X_{t-i-j} + \sum_{j=1}^{q} \theta_{p+j} \tilde{\epsilon}_{t-j}(\theta), \tag{3.3}$$

with $\tilde{\epsilon}_t(\theta) = X_t = 0$ if $t \leq 0$.

As shown in Lemma 4 of Boubacar Maïnassara et al. [2019], these initial values are asymptotically negligible uniformly in θ and in particular it holds that $\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta) \to 0$ almost surely as $t \to \infty$. Thus the choice of the initial values has no influence on the asymptotic properties of the model parameters estimator. Let Θ_δ^* denotes the compact set

$$\Theta_\delta^* = \left\{\theta \in \mathbb{R}^{p+q};\ \text{the roots of the polynomials } a_\theta(z) \text{ and } b_\theta(z) \text{ have modulus } \geq 1+\delta\right\}.$$

We define the set Θ_δ as the cartesian product of Θ_δ^* by $[d_1, d_2]$, *i.e.* $\Theta_\delta = \Theta_\delta^* \times [d_1, d_2]$, where δ is a strictly positive constant chosen such that θ_0 belongs to Θ_δ.

The least square estimator is defined, almost-surely, by

$$\hat{\theta}_n = \underset{\theta\in\Theta_\delta}{\operatorname{argmin}}\ Q_n(\theta), \text{ where } Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \tilde{\epsilon}_t^2(\theta). \tag{3.4}$$

To ensure the strong consistency of the least square estimator, we assume as in Boubacar Maïnassara et al. [2019] that the parametrization satisfies the following condition.

(A1): The process $(\epsilon_t)_{t\in\mathbb{Z}}$ is strictly stationary and ergodic.

The strong consistency of the estimator is obtained under the assumptions **(A0)** and **(A1)**. Additional assumptions are required in order to establish the asymptotic normality of the least square estimator. We assume that θ_0 is not on the boundary of the parameter space Θ.

(A2): We have $\theta_0 \in \overset{\circ}{\Theta}$, where $\overset{\circ}{\Theta}$ denotes the interior of Θ.

The stationary process ϵ is not supposed to be an independent sequence. So one needs to control its dependency by means of its strong mixing coefficients $\{\alpha_\epsilon(h)\}_{h\in\mathbb{N}}$ defined by

$$\alpha_\epsilon(h) = \sup_{A\in\mathcal{F}_{-\infty}^{t}, B\in\mathcal{F}_{t+h}^{\infty}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|,$$

where $\mathcal{F}_{-\infty}^{t} = \sigma(\epsilon_u, u \leq t)$ and $\mathcal{F}_{t+h}^{\infty} = \sigma(\epsilon_u, u \geq t+h)$.

We shall need an integrability assumption on the moment of the noise ϵ and a summability condition on the strong mixing coefficients $(\alpha_\epsilon(k))_{k\geq 0}$.

(A3): There exists an integer τ such that for some $\nu \in]0,1]$, we have $\mathbb{E}|\epsilon_t|^{\tau+\nu} < \infty$ and $\sum_{h=0}^{\infty}(h+1)^{k-2}\{\alpha_\epsilon(h)\}^{\frac{\nu}{k+\nu}} < \infty$ for $k = 1,\ldots,\tau$.

Note that **(A3)** implies the following weak assumption on the joint cumulants of the innovation process ϵ (see Doukhan and León [1989], for more details).

(A3'): There exists an integer $\tau \geq 2$ such that $C_\tau := \sum_{i_1,\ldots,i_{\tau-1}\in\mathbb{Z}} |\mathrm{cum}(\epsilon_0, \epsilon_{i_1}, \ldots, \epsilon_{i_{\tau-1}})| < \infty$.

In the above expression, $\mathrm{cum}(\epsilon_0, \epsilon_{i_1}, \ldots, \epsilon_{i_{\tau-1}})$ denotes the $\tau-$th order joint cumulant of the stationary process. Due to the fact that the ϵ_t's are centered, we notice that for fixed (i,j,k)

$$\mathrm{cum}(\epsilon_0, \epsilon_i, \epsilon_j, \epsilon_k) = \mathbb{E}[\epsilon_0\epsilon_i\epsilon_j\epsilon_k] - \mathbb{E}[\epsilon_0\epsilon_i]\mathbb{E}[\epsilon_j\epsilon_k] - \mathbb{E}[\epsilon_0\epsilon_j]\mathbb{E}[\epsilon_i\epsilon_k] - \mathbb{E}[\epsilon_0\epsilon_k]\mathbb{E}[\epsilon_i\epsilon_j].$$

Assumption **(A3)** is a usual technical hypothesis which is useful when one proves the asymptotic normality (see Francq and Zakoïan [1998] for example). Let us notice however that we impose a stronger convergence speed for the mixing coefficients than in the works on weak ARMA processes. This is due to the fact that the coefficients in the infinite AR or MA representation of $\epsilon_t(\theta)$ have no more exponential decay because of the fractional operator (see Subsection 6.1 in Boubacar Maïnassara et al. [2019] for details and comments).

As mentioned before, Hypothesis **(A3)** implies **(A3')** which is also a technical assumption usually used in the fractional ARIMA processes framework (see for instance Shao [2010b, 2011]) or even in an ARMA context (see Francq and Zakoïan [2007], Zhu and Li [2015]).

Under the assumptions **(A0)**, **(A1)**, **(A2)** and **(A3)** with $\tau = 4$, Boubacar Maïnassara et al. [2019] showed that $\hat{\theta}_n \to \theta_0$ *a.s.* as $n \to \infty$ and $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotically normal with mean 0 and covariance matrix $\Sigma_{\hat{\theta}} := J^{-1}IJ^{-1}$, where $J = J(\theta_0)$ and $I = I(\theta_0)$, with

$$I(\theta) = \lim_{n\to\infty} Var\left\{\sqrt{n}\frac{\partial}{\partial\theta}O_n(\theta)\right\} \text{ and } J(\theta) = \lim_{n\to\infty}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}O_n(\theta)\right] \text{ a.s.},$$

where $O_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2(\theta)$.

Remind that the sequence $(\epsilon_t(\theta))_{t\in\mathbb{Z}}$ is given by (3.2).

3.3 Diagnostic checking in weak FARIMA models

After the estimation phase, the next important step consists in checking if the estimated model fits satisfactorily the data. In this section we derive the limiting distribution of the residual autocorrelations and that of the portmanteau statistics (based on the standard and the self-normalized approaches) in the framework of weak FARIMA models.

For $t \geq 1$, let $\hat{e}_t = \tilde{\epsilon}_t(\hat{\theta}_n)$ be the least-square residuals. By (3.3) we notice that $\hat{e}_t = 0$ for $t \leq 0$ and $t > n$. By (3.1) it holds that

$$\hat{e}_t = \sum_{j=0}^{t-1} \alpha_j(\hat{d})\hat{X}_{t-j} - \sum_{i=1}^{p} \hat{\theta}_i \sum_{j=0}^{t-i-1} \alpha_j(\hat{d})\hat{X}_{t-i-j} + \sum_{j=1}^{q} \hat{\theta}_{p+j}\hat{e}_{t-j},$$

for $t = 1, \ldots, n$, with $\hat{X}_t = 0$ for $t \leq 0$ and $\hat{X}_t = X_t$ for $t \geq 1$.

For a fixed integer $m \geq 1$ consider the vector of residual autocovariances

$$\hat{\gamma}_m = (\hat{\gamma}(1), \ldots, \hat{\gamma}(m))' \text{ where } \hat{\gamma}(h) = \frac{1}{n}\sum_{t=h+1}^{n} \hat{e}_t\, \hat{e}_{t-h} \text{ for } 0 \leq h < n.$$

In the sequel we will also need the vector of the first m sample autocorrelations

$$\hat{\rho}_m = (\hat{\rho}(1), \ldots, \hat{\rho}(m))' \text{ where } \hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0).$$

Since the papers by Box and Pierce [1970] and Ljung and Box [1978], portmanteau tests have been popular diagnostic checking tools in the ARMA modeling of time series. Based on the residual empirical autocorrelations, their test statistics are defined respectively by

$$Q_m^{\text{BP}} = n\sum_{h=1}^{m} \hat{\rho}^2(h) \text{ and } Q_m^{\text{LB}} = n(n+2)\sum_{h=1}^{m} \frac{\hat{\rho}^2(h)}{n-h}. \tag{3.5}$$

These statistics are usually used to test the null hypothesis

(H0): $(X_t)_{t\in\mathbb{Z}}$ satisfies a FARIMA(p, d_0, q) representation;

against the alternative

(H1): $(X_t)_{t\in\mathbb{Z}}$ does not admit a FARIMA representation or admits a FARIMA(p', d_0, q') representation with $p' > p$ or $q' > q$.

These tests are very useful tools to check the global significance of the residual autocorrelations.

3.3.1 Asymptotic distribution of the residual autocorrelations

First of all, the mixing assumption **(A3)** will entail the asymptotic normality of the "empirical" autocovariances

$$\gamma_m = (\gamma(1), \ldots, \gamma(m))' \text{ where } \gamma(h) = \frac{1}{n}\sum_{t=h+1}^{n} \epsilon_t\, \epsilon_{t-h} \text{ for } 0 \leq h < n. \tag{3.6}$$

It should be noted that $\gamma(h)$ is not a computable statistic because it depends on the unobserved innovations $\epsilon_t = \epsilon_t(\theta_0)$. They are introduced as a device to facilitate future derivations. Let Ψ_m be the $m \times (p+q+1)$ matrix defined by

$$\Psi_m = \mathbb{E}\left\{ \begin{pmatrix} \epsilon_{t-1} \\ \vdots \\ \epsilon_{t-m} \end{pmatrix} \frac{\partial \epsilon_t}{\partial \theta'} \right\}. \tag{3.7}$$

By a Taylor expansion of $\sqrt{n}\hat{\gamma}_m$, one should prove that (see Section 3.5.3)

$$\sqrt{n}\hat{\gamma}_m = \sqrt{n}\gamma_m + \Psi_m \sqrt{n}\left(\hat{\theta}_n - \theta_0\right) + o_{\mathbb{P}}(1), \tag{3.8}$$

where Ψ_m is given in (3.7). We shall also prove (see Section 3.5.3 again) that

$$\sqrt{n}\hat{\rho}_m = \sqrt{n}\frac{\hat{\gamma}_m}{\sigma_\epsilon^2} + o_{\mathbb{P}}(1). \tag{3.9}$$

Thus from (3.9) the asymptotic distribution of the residual autocorrelations $\sqrt{n}\hat{\rho}_m$ depends on the distribution of $\hat{\gamma}_m$. In view of (3.8) the asymptotic distribution of the residual autocovariances $\sqrt{n}\hat{\gamma}_m$ will be obtained from the joint asymptotic behavior of $\sqrt{n}(\hat{\theta}_n' - \theta_0', \gamma_m')'$.

In view of Theorem 1 in Boubacar Maïnassara et al. [2019] and **(A2)**, we have almost surely $\hat{\theta}_n \to \theta_0 \in \overset{\circ}{\Theta}$. Thus $\partial Q_n(\hat{\theta}_n)/\partial\theta = 0$ for sufficiently large n and a Taylor expansion gives

$$\sqrt{n}\frac{\partial}{\partial\theta} O_n(\theta_0) + J(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0) = o_{\mathbb{P}}(1). \tag{3.10}$$

The equation (3.10) is proved in Boubacar Maïnassara et al. [2019] (see the proof of Theorem 2). Consequently from (3.10) we have

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -\frac{2}{\sqrt{n}} \sum_{t=1}^{n} J^{-1}(\theta_0)\epsilon_t(\theta_0)\frac{\partial \epsilon_t(\theta_0)}{\partial\theta} + o_{\mathbb{P}}(1). \tag{3.11}$$

For integers $m, m' \geq 1$, one needs the matrix $\Gamma_{m,m'} = [\Gamma(\ell, \ell')]_{1\leq\ell\leq m, 1\leq\ell'\leq m'}$ where

$$\Gamma(\ell, \ell') = \sum_{h=-\infty}^{\infty} \mathbb{E}\left[\epsilon_t \epsilon_{t-\ell}\epsilon_{t-h}\epsilon_{t-h-\ell'}\right].$$

The existence of $\Gamma(\ell, \ell')$ will be justified in Lemma 3.3 of the appendix.

Proposition 3.1. *Under the assumptions* **(A0)**, **(A1)**, **(A2)** *and* **(A3)** *with* $\tau = 4$, *the random vector*

$$\sqrt{n}\left(\left(\hat{\theta}_n - \theta_0\right)', \gamma_m'\right)'$$

has a limiting centered normal distribution with covariance matrix

$$\Xi = \begin{pmatrix} \Sigma_{\hat{\theta}} & \Sigma_{\hat{\theta},\gamma_m} \\ \Sigma'_{\hat{\theta},\gamma_m} & \Gamma_{m,m} \end{pmatrix} = \sum_{h=-\infty}^{\infty} \mathbb{E}\left[U_t U'_{t-h}\right],$$

where from (3.6) *and* (3.11) *we have*

$$U_t = \begin{pmatrix} U_{1t} \\ U_{2t} \end{pmatrix} = \begin{pmatrix} -2J^{-1}(\theta_0)\epsilon_t(\theta_0)\frac{\partial}{\partial\theta}\epsilon_t(\theta_0) \\ (\epsilon_{t-1},\ldots,\epsilon_{t-m})'\epsilon_t \end{pmatrix}. \tag{3.12}$$

The proof of the proposition is given in Subsection 3.5.2 of the appendix.

The following theorem which is an extension of the result given in Francq et al. [2005] provides the limit distribution of the residual autocovariances and autocorrelations of weak FARIMA models.

Theorem 3.1. *Under the assumptions of Proposition 3.1, we have*

$$\sqrt{n}\hat{\gamma}_m \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0,\Sigma_{\hat{\gamma}_m}) \quad \textit{where} \quad \Sigma_{\hat{\gamma}_m} = \Gamma_{m,m} + \Psi_m\Sigma_{\hat{\theta}}\Psi'_m + \Psi_m\Sigma_{\hat{\theta},\gamma_m} + \Sigma'_{\hat{\theta},\gamma_m}\Psi'_m \tag{3.13}$$

and

$$\sqrt{n}\hat{\rho}_m \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0,\Sigma_{\hat{\rho}_m}) \quad \textit{where} \quad \Sigma_{\hat{\rho}_m} = \frac{1}{\sigma_\epsilon^4}\Sigma_{\hat{\gamma}_m}. \tag{3.14}$$

The detailed proof of this result is postponed to the Subsection 3.5.3 of Appendix.

Remark 3.1. *It is clear from Theorem 3.1 that for a given FARIMA(p,d_0,q) model, the asymptotic distribution of the residual autocorrelations depends only on the noise distribution through the quantities $\Gamma(\ell,\ell')$ (which depends on the fourth-order structure of the noise). It is also worth noting that this asymptotic distribution depends on the asymptotic normality of the least square estimator of the FARIMA(p,d_0,q) only through the matrix $\Sigma_{\hat{\theta}}$.*

Remark 3.2. *In the standard strong FARIMA case,* i.e. *when* **(A1)** *is replaced by the assumption that (ϵ_t) is iid, Boubacar Maïnassara et al. [2019] have showed in Remark 2 that $I(\theta_0) = 2\sigma_\epsilon^2 J(\theta_0)$. Thus the asymptotic covariance matrix is then reduced as $\Sigma_{\hat{\theta}} = 2\sigma_\epsilon^2 J^{-1}(\theta_0)$. In the strong case, we also have: $\Gamma(\ell,\ell') = 0$ when $\ell \neq \ell'$ and $\Gamma(\ell,\ell) = \sigma_\epsilon^4$ for $h = 0$. Thus $\Gamma_{m,m}$ is reduced as $\Gamma_{m,m} = \sigma_\epsilon^4 I_m$, where I_m denotes the $m \times m$ identity matrix. Because $\Sigma_{\hat{\theta}} = 2\sigma_\epsilon^2 J^{-1}(\theta_0)$ we obtain that*

$$\begin{aligned}
\Sigma_{\hat{\theta},\gamma_m} &= -2\sum_{h=-\infty}^{\infty} \mathbb{E}\left\{\epsilon_t J^{-1}(\theta_0)\frac{\partial\epsilon_t(\theta_0)}{\partial\theta}\right\}\left\{\begin{pmatrix}\epsilon_{t-1-h}\\ \vdots \\ \epsilon_{t-m-h}\end{pmatrix}\epsilon_{t-h}\right\}' \\
&= -\left(2\sigma_\epsilon^2 J^{-1}(\theta_0)\right)\left\{\mathbb{E}\left[\begin{pmatrix}\epsilon_{t-1}\\ \vdots \\ \epsilon_{t-m}\end{pmatrix}\frac{\partial\epsilon_t(\theta_0)}{\partial\theta'}\right]\right\}' = -\Sigma_{\hat{\theta}}\Psi'_m.
\end{aligned}$$

We denote by $\Sigma^{\mathrm{S}}_{\hat{\gamma}_m}$ *and* $\Sigma^{\mathrm{S}}_{\hat{\rho}_m}$ *the asymptotic variances obtained respectively in* (3.13) *and* (3.14) *for the strong FARIMA case. Thus we obtain, in the strong case, the following simpler expressions*

$$\Sigma^{\mathrm{S}}_{\hat{\gamma}_m} = \sigma_\epsilon^4 I_m - 2\sigma_\epsilon^2 \Psi_m J^{-1}(\theta_0)\Psi_m' \quad \text{and} \quad \Sigma^{\mathrm{S}}_{\hat{\rho}_m} = I_m - \frac{2}{\sigma_\epsilon^2}\Psi_m J^{-1}(\theta_0)\Psi_m',$$

which are the matrices obtained by Li and McLeod [1986].

To validate a FARIMA(p, d_0, q) model, the most basic technique is to examine the autocorrelation function of the residuals. Theorem 3.1 can be used to obtain asymptotic significance limits for the residual autocorrelations. However, the asymptotic variance matrices $\Sigma_{\hat{\gamma}_m}$ and $\Sigma_{\hat{\rho}_m}$ depend on the unknown matrices Ξ, Ψ_m and the strictly positive scalar σ_ϵ^2 which need to be estimated. This is the purpose of the following discussion.

3.3.2 Modified version of the portmanteau test

From Theorem 3.1 we can deduce the following result, which gives the limiting distribution of the standard portmanteau statistics (3.5) under general assumptions on the innovation process of the fitted FARIMA(p, d_0, q) model.

Theorem 3.2. *Under the assumptions of Theorem 3.1 and* **(H0)**, *the statistics* Q_m^{BP} *and* Q_m^{LB} *defined by* (3.5) *converge in distribution, as* $n \to \infty$, *to*

$$Z_m(\xi_m) = \sum_{k=1}^{m} \xi_{k,m} Z_k^2,$$

where $\xi_m = (\xi_{1,m}, \ldots, \xi_{m,m})'$ *is the vector of the eigenvalues of the matrix* $\Sigma_{\hat{\rho}_m} = \sigma_\epsilon^{-4}\Sigma_{\hat{\gamma}_m}$ *and* $Z_1, \ldots, Z_m$ *are independent* $\mathcal{N}(0,1)$ *variables.*

It is possible to evaluate the distribution of a quadratic form of a Gaussian vector by means of the Imhof algorithm (see Imhof [1961]).

Remark 3.3. *In view of remark 3.2 when m is large,* $\Sigma^{\mathrm{S}}_{\hat{\rho}_m} \simeq I_m - 2\sigma_\epsilon^{-2}\Psi_m J^{-1}(\theta_0)\Psi_m'$ *is close to a projection matrix. Its eigenvalues are therefore equal to 0 and 1. The number of eigenvalues equal to 1 is* $\mathrm{Tr}(I_m - 2\sigma_\epsilon^{-2}\Psi_m J^{-1}(\theta_0)\Psi_m') = \mathrm{Tr}(I_{m-(p+q+1)}) = m - (p+q+1)$ *and* $p+q+1$ *eigenvalues equal to 0,* $\mathrm{Tr}(\cdot)$ *denotes the trace of a matrix. Therefore we retrieve the well-known result obtained by Li and McLeod [1986]. More precisely, under* **(H0)** *and in the strong FARIMA case, the asymptotic distributions of the statistics* Q_m^{BP} *and* Q_m^{LB} *are approximated by a* $\mathcal{X}^2_{m-(p+q+1)}$, *where* $m > p+q+1$ *and* $\mathcal{X}^2_k$ *denotes the chi-squared distribution with k degrees of freedom. Theorem 3.2 shows that this approximation is no longer valid in the framework of weak FARIMA*(p, d, q) *models and that the asymptotic null distributions of the statistics* Q_m^{BP} *and* Q_m^{LB} *are more complicated.*

The limit distribution $Z_m(\xi_m)$ depends on the nuisance parameter σ_ϵ^2, the matrix Ψ_m and the elements of Ξ. Therefore, the asymptotic distribution of the portmanteau statistics (3.5),

under weak assumptions on the noise, requires a computation of a consistent estimator of the asymptotic covariance matrix $\Sigma_{\hat{\rho}_m}$. The $m \times (p+q+1)$ matrix Ψ_m and the noise variance σ_ϵ^2 can be estimated by its empirical counterpart. Thus we may use

$$\hat{\Psi}_m = \frac{1}{n}\sum_{t=1}^{n}\left\{(\hat{e}_{t-1},\ldots,\hat{e}_{t-m})' \frac{\partial \hat{e}_t}{\partial \theta'}\right\} \quad \text{and} \quad \hat{\sigma}_\epsilon^2 = \hat{\gamma}(0) = \frac{1}{n}\sum_{t=1}^{n}\hat{e}_t^2.$$

A consistent estimator of Ξ is obtained by means of an autoregressive spectral estimator, as in Boubacar Maïnassara et al. [2019] (see also Berk [1974], Boubacar Mainassara et al. [2012] and den Haan and Levin [1997], to name a few for a more comprehensive exposition of this method). In view of (3.12), the matrix Ξ can be interpreted as 2π times the spectral density of the stationary process $(U_t)_{t\in\mathbb{Z}} = ((U_{1t}', U_{2t}')')_{t\in\mathbb{Z}}$ evaluated at frequency 0 (see p. 459 of Brockwell and Davis [1991]). So this estimation is based on the following expression

$$\Xi = A^{-1}(1)\Sigma_v A'^{-1}(1)$$

when $(U_t)_{t\in\mathbb{Z}}$ satisfies an AR(∞) representation of the form

$$A(L)U_t := U_t - \sum_{k=1}^{\infty} A_k U_{t-k} = v_t, \tag{3.15}$$

such that $\sum_{k=1}^{\infty}\|A_k\| < \infty$ and $\det\{A(z)\} \neq 0$ for all $|z| \leq 1$ and where $(v_t)_{t\in\mathbb{Z}}$ is a $(p+q+1+m)$-variate weak white noise with variance matrix Σ_v. It is proved in Boubacar Maïnassara [2009], Lütkepohl [2005] that one may find a constant K and $0 < \rho < 1$ such that

$$\|A_k\| \leq K\,\rho^k. \tag{3.16}$$

Since U_t is unobservable, we introduce $\hat{U}_t \in \mathbb{R}^{p+q+1+m}$ obtained by replacing $\epsilon_t(\theta_0)$ by $\tilde{\epsilon}_t(\hat{\theta}_n)$ and $J(\theta_0)$ by its empirical or observable counterpart $\hat{J}_n$ in (3.12). Let $\hat{A}_r(z) = I_{p+q+1+m} - \sum_{k=1}^{r}\hat{A}_{r,k}z^k$, where $\hat{A}_{r,1},\ldots,\hat{A}_{r,r}$ denote the coefficients of the least squares regression of $\hat{U}_t$ on $\hat{U}_{t-1},\ldots,\hat{U}_{t-r}$. Let $\hat{v}_{r,t}$ be the residuals of this regression, and let $\hat{\Sigma}_{\hat{v}_r}$ be the empirical variance of $\hat{v}_{r,1},\ldots,\hat{v}_{r,n}$. We are now able to state Theorem 3.3 which is an extension of a result given in Boubacar Mainassara et al. [2012].

Theorem 3.3. *We assume* ***(A0)***, ***(A1)***, ***(A2)*** *and Assumption* **(A3')** *with* $\tau = 8$. *In addition, we assume that the process* $(U_t)_{t\in\mathbb{Z}}$ *defined in* (3.12) *admits a multivariate AR*(∞) *representation* (3.15) *such that the coefficients satisfy* (3.16) *and* $\Sigma_v = \mathrm{Var}(v_t)$ *is non-singular. Then the spectral estimator of* Ξ *satisfies*

$$\hat{\Xi}_n^{\mathrm{SP}} := \hat{A}_r^{-1}(1)\hat{\Sigma}_{\hat{v}_r}\hat{A}_r'^{-1}(1) \xrightarrow[n\to\infty]{\mathbb{P}} \Xi = A^{-1}(1)\Sigma_v A^{-1}(1)$$

where r *depends on* n *and satisfies* $\lim_{n\to\infty} r^5(n)/n^{1-2(d_2-d_1)} = 0$ *(remind that* $d_0 \in [d_1,d_2] \subset]0,1/2[$*).*

The proof of this theorem is similar to the proof of Theorem 3 in Boubacar Maïnassara et al. [2019] and it is omitted.

We are now in a position to define the modified versions of the Box-Pierce (BP) and Ljung-Box (LB) goodness-of-fit portmanteau tests. The standard versions of the portmanteau tests are useful tools to detect if the orders p and q of a FARIMA(p, d_0, q) model are well chosen, provided the error terms $(\epsilon_t)_{t\in\mathbb{Z}}$ of the FARIMA(p, d_0, q) equation be a strong white noise and provided the number m of residual autocorrelations is not too small (see Remark 3.3). Now we define the modified versions which are aimed to detect if the orders p and q of a weak FARIMA(p, d_0, q) model are well chosen. These tests are also asymptotically valid for strong FARIMA(p, d_0, q) even for small m. The modified versions of the portmanteau tests will be denoted by BP_{W} and LB_{W}, the subscript W referring to the term weak.

Let $\hat{\Sigma}_{\hat{\rho}_m}$ be the matrix obtained by replacing Ξ by $\hat{\Xi}$ and σ_ϵ^2 by $\hat{\sigma}_\epsilon^2$ in $\Sigma_{\hat{\rho}_m}$. Denote by $\hat{\xi}_m = (\hat{\xi}_{1,m}, \ldots, \hat{\xi}_{m,m})'$ the vector of the eigenvalues of $\hat{\Sigma}_{\hat{\rho}_m}$. At the asymptotic level α, the BP_{W} test (resp. the LB_{W} test) consists in rejecting the null hypothesis of the weak FARIMA(p, d_0, q) model (the adequacy of the weak FARIMA(p, d_0, q) model) when

$$Q_m^{\mathrm{BP}} > S_m(1-\alpha) \quad (\text{resp.} \quad Q_m^{\mathrm{LB}} > S_m(1-\alpha)),$$

where $S_m(1-\alpha)$ is such that $\mathbb{P}(Z_m(\hat{\xi}_m) > S_m(1-\alpha)) = \alpha$. We emphasize the fact that the proposed modified versions of the Box-Pierce and Ljung-Box statistics are more difficult to implement because their critical values have to be computed from the data while the critical values of the standard method are simply deduced from a χ^2-table. We shall evaluate the p-values

$$\mathbb{P}\left\{Z_m(\hat{\xi}_m) > Q_m^{\mathrm{BP}}\right\} \quad \text{and} \quad \mathbb{P}\left\{Z_m(\hat{\xi}_m) > Q_m^{\mathrm{LB}}\right\}, \text{with } Z_m(\hat{\xi}_m) = \sum_{i=1}^{m} \hat{\xi}_{i,m} Z_i^2,$$

by means of the Imhof algorithm (see Imhof [1961]).

A second method avoiding the estimation of the asymptotic matrix is proposed in the next Subsection.

3.3.3 Self-normalized asymptotic distribution of the residual autocorrelations

In view of Theorem 3.2, the asymptotic distributions of the statistics defined in (3.5) are a mixture of chi-squared distributions, weighted by eigenvalues of the asymptotic covariance matrix $\Sigma_{\hat{\rho}_m}$ of the vector of autocorrelations obtained in Theorem 3.1. However, this asymptotic variance matrix depends on the unknown matrices Ξ, Ψ_m and the noise variance σ_ϵ^2. Consequently, in order to obtain a consistent estimator of the asymptotic covariance matrix $\Sigma_{\hat{\rho}_m}$ of the residual autocorrelations vector we have used an autoregressive spectral estimator of the spectral density of the stationary process $(U_t)_{t\in\mathbb{Z}}$ to get a consistency estimator of the matrix Ξ (see Theorem 3.3). However, this approach presents the problem of choosing the truncation parameter. Indeed this method is based on an infinite autoregressive representation

of the stationary process $(U_t)_{t\in\mathbb{Z}}$ (see (3.15)). So the choice of the order of truncation is crucial and difficult.

In this section, we propose an alternative method where we do not estimate an asymptotic covariance matrix which is an extension to the results obtained by Boubacar Maïnassara and Saussereau [2018]. It is based on a self-normalization approach to construct a test-statistic which is asymptotically distribution-free under the null hypothesis. This approach has been studied by Boubacar Maïnassara and Saussereau [2018] in the weak ARMA case, by proposing new portmanteau statistics. In this case the critical values are not computed from the data since they are tabulated by Lobato [2001]. In some sense this method is finally closer to the standard method in which the critical values are simply deduced from a $\mathcal{X}^2$-table. The idea comes from Lobato [2001] and has been already extended by Boubacar Maïnassara and Saussereau [2018], Kuan and Lee [2006], Shao [2010b], Shao [2010a] and Shao [2012] to name a few in more general frameworks. See also Shao [2015] for a review on some recent developments on the inference of time series data using the self-normalized approach.

We denote by Λ the block matrix of $\mathbb{R}^{m\times(p+q+1+m)}$ defined by $\Lambda = (\Psi_m|I_m)$. In view of (3.8) and (3.11) we deduce that

$$\sqrt{n}\hat{\gamma}_m = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Lambda U_t + o_{\mathbb{P}}(1).$$

At this stage, we do not rely on the classical method that would consist in estimating the asymptotic covariance matrix Ξ. We rather try to apply Lemma 1 in Lobato [2001]. So we need to check that a functional central limit theorem holds for the process $U := (U_t)_{t\geq 1}$. For that sake, we define the normalization matrix C_m of $\mathbb{R}^{m\times m}$ by

$$C_m = \frac{1}{n^2}\sum_{t=1}^{n} S_t S_t' \text{ where } S_t = \sum_{j=1}^{t}(\Lambda U_j - \gamma_m).$$

To ensure the invertibility of the normalization matrix C_m (it is the result stated in the next proposition), we need the following technical assumption on the distribution of ϵ_t.

(A4): The process $(\epsilon_t)_{t\in\mathbb{Z}}$ has a positive density on some neighbourhood of zero.

Proposition 3.2. *Under the assumptions of Theorem 3.1 and* **(A4)**, *the matrix C_m is almost surely non singular.*

The proof of this proposition is given in Subsection 3.5.4 of the appendix.

Let $(B_K(r))_{r\geq 0}$ be a K-dimensional Brownian motion starting from 0. For $K \geq 1$, we denote by $\mathcal{U}_K$ the random variable defined by:

$$\mathcal{U}_K = B_K'(1)V_K^{-1}B_K(1), \tag{3.17}$$

where

$$V_K = \int_0^1 (B_K(r) - rB_K(1))(B_K(r) - rB_K(1))' \, dr. \tag{3.18}$$

The critical values of $\mathcal{U}_K$ have been tabulated by Lobato [2001].

The following theorem states the asymptotic distributions of the sample autocovariances and autocorrelations.

Theorem 3.4. *Under the assumptions of Theorem 3.1,* **(A4)** *and under the null hypothesis* **(H0)** *we have*

$$n\hat{\gamma}_m' C_m^{-1}\hat{\gamma}_m \xrightarrow[n\to\infty]{\text{in law}} \mathcal{U}_m \text{ and } n\sigma_\epsilon^4 \hat{\rho}_m' C_m^{-1}\hat{\rho}_m \xrightarrow[n\to\infty]{\text{in law}} \mathcal{U}_m.$$

The proof of this theorem is given in Subsection 3.5.5 of Appendix.

Of course, the above theorem is useless for practical purpose because the normalization matrix C_m and the nuisance parameter σ_ϵ^2 are not observable. This gap will be fixed below (see Theorem 3.5) when one replaces the matrix C_m and the scalar σ_ϵ^2 by their empirical or observable counterparts. Then we denote

$$\hat{C}_m = \frac{1}{n^2}\sum_{t=1}^{n} \hat{S}_t \hat{S}_t' \text{ where } \hat{S}_t = \sum_{j=1}^{t} \left(\hat{\Lambda}\hat{U}_j - \hat{\gamma}_m\right),$$

with $\hat{\Lambda} = (\hat{\Psi}_m | I_m)$ and where $\hat{U}_t$ and $\hat{\sigma}_\epsilon^2$ are defined in Subsection 3.3.2.

The above quantities are all observable and the following result is the applicable counterpart of Theorem 3.4.

Theorem 3.5. *Under the assumptions of Theorem 3.4, we have*

$$n\hat{\gamma}_m' \hat{C}_m^{-1}\hat{\gamma}_m \xrightarrow[n\to\infty]{\text{in law}} \mathcal{U}_m \text{ and } Q_m^{\text{SN}} = n\hat{\sigma}_\epsilon^4 \hat{\rho}_m' \hat{C}_m^{-1}\hat{\rho}_m \xrightarrow[n\to\infty]{\text{in law}} \mathcal{U}_m.$$

The proof of this result is postponed in Subsection 3.5.6 of Appendix.

Based on the above result, we propose a modified version of the Ljung-Box statistic when one uses the statistic

$$\tilde{Q}_m^{\text{SN}} = n\hat{\sigma}_\epsilon^4 \hat{\rho}_m' D_{n,m}^{1/2} \hat{C}_m^{-1} D_{n,m}^{1/2}\hat{\rho}_m,$$

where $D_{n,m} \in \mathbb{R}^{m\times m}$ is diagonal with $(n+2)/(n-1), \ldots, (n+2)/(n-m)$ as diagonal terms. These modified versions of the portmanteau tests will be denoted by BP_{SN} and LB_{SN}, the subscript SN referring to the term self-normalized.

3.3.4 Example of explicit calculation of $\Sigma_{\hat{\rho}_m}$ and C_m

The results of the previous subsections 3.3.2 and 3.3.3 are particularized in the FARIMA$(1,d,0)$ and FARIMA$(0,d,1)$ cases. First we consider the case of a FARIMA$(1,d,0)$ model of the form

$$(1-L)^d (X_t - aX_{t-1}) = \epsilon_t, \tag{3.19}$$

where the unknown parameter is $\theta_0 = (a,d)$. We assume that in (3.19) the innovation process $(\epsilon_t)_{t\in\mathbb{Z}}$ is a GARCH$(1,1)$ process given by the model

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \end{cases} \tag{3.20}$$

with $\omega > 0$, $\alpha_1 \geq 0$ and where $(\eta_t)_{t\in\mathbb{Z}}$ is a sequence of iid centered Gaussian random variables with variance 1. We also assume that $\alpha_1^2\kappa + \beta_1^2 + 2\alpha_1\beta_1 < 1$,[2] where $\kappa := \mathbb{E}\eta_1^4$ and we assume that $\kappa > 1$.

For the sake of simplicity we assume that the variables (η_t) have a symmetric distribution. More precisely, we have the following symmetry assumption

$$\mathbb{E}[\epsilon_{t_1}\epsilon_{t_2}\epsilon_{t_3}\epsilon_{t_4}] = 0 \qquad \text{when} \qquad t_1 \neq t_2,\ t_1 \neq t_3 \text{ and } t_1 \neq t_4, \tag{3.21}$$

made in Boubacar Mainassara et al. [2012], Francq and Zakoïan [2009]. For this particular GARCH(1, 1) model with fourth-order moments and symmetric innovations satisfying (3.21), it can be shown that

$$\mathbb{E}\left[\epsilon_t\epsilon_{t-\ell}\epsilon_{t-h}\epsilon_{t-h-\ell'}\right] = \begin{cases} \mathbb{E}\left[\epsilon_t^2\epsilon_{t-\ell}^2\right] & \text{if } h = 0 \text{ and } \ell = \ell' \\ 0 & \text{otherwise.} \end{cases} \tag{3.22}$$

Now we need to compute the autocovariance structure of (ϵ_t^2). We will use the fact that the GARCH process (ϵ_t) is fourth-order stationary, then (ϵ_t^2) is a solution of the following ARMA(1, 1) model

$$\epsilon_t^2 = \omega + (\alpha_1 + \beta_1)\epsilon_{t-1}^2 + \nu_t - \beta_1\nu_{t-1}, \quad t \in \mathbb{Z} \tag{3.23}$$

where $\nu_t = \epsilon_t^2 - \sigma_t^2$ is the innovation of (ϵ_t^2). From (3.23) the autocovariances of (ϵ_t^2) take the form

$$\gamma_{\epsilon^2}(\ell) := \mathrm{Cov}(\epsilon_t^2, \epsilon_{t-\ell}^2) = \gamma_{\epsilon^2}(1)(\alpha_1 + \beta_1)^{\ell-1}, \qquad \ell \geq 1, \tag{3.24}$$

where

$$\begin{aligned} \gamma_{\epsilon^2}(1) &= \frac{(\kappa-1)(\alpha_1 - \alpha_1\beta_1^2 - \alpha_1^2\beta_1)}{1 - \beta_1^2 - 2\alpha_1\beta_1 - \alpha_1^2\kappa}\sigma_\epsilon^4, \\ \gamma_{\epsilon^2}(0) &:= \mathrm{Var}(\epsilon_t^2) = \frac{(\kappa-1)(1 - \beta_1^2 - 2\alpha_1\beta_1)}{1 - \beta_1^2 - 2\alpha_1\beta_1 - \alpha_1^2\kappa}\sigma_\epsilon^4, \\ \text{and } \sigma_\epsilon^2 &:= \frac{\omega}{1 - \alpha_1 - \beta_1}. \end{aligned}$$

From (3.22) and (3.24) we deduce that for any $\ell \geq 1$

$$\begin{aligned} \Gamma(\ell, \ell) = \mathbb{E}\left[\epsilon_t^2\epsilon_{t-\ell}^2\right] &= \mathrm{Cov}(\epsilon_t^2, \epsilon_{t-\ell}^2) + \mathbb{E}\left[\epsilon_t^2\right]\mathbb{E}\left[\epsilon_{t-\ell}^2\right] \\ &= \left\{1 + \frac{1}{\sigma_\epsilon^4}\gamma_{\epsilon^2}(1)(\alpha_1 + \beta_1)^{\ell-1}\right\}\sigma_\epsilon^4. \end{aligned} \tag{3.25}$$

2. This is a necessary and sufficient condition for the existence of a nonanticipative stationary solution process $(\epsilon_t)_{t\in\mathbb{Z}}$ with fourth-order moments (see [Francq and Zakoïan, 2010, Example 2.3]).

Examples of analytic and numerical computations of $\Sigma_{\hat{\rho}_m}$

As mentioned before, the subject of this subsection is to give an explicit expression of the asymptotic variance of residual autocorrelations $\Sigma_{\hat{\rho}_m}$ defined in (3.14) in the particular case of model (3.19). For that sake, we need the following additional expressions. It is classical that the noise derivatives $(\partial\epsilon_t(\theta_0)/\partial a, \partial\epsilon_t(\theta_0)/\partial d)'$ in (3.19) can be represented as

$$\begin{pmatrix} \frac{\partial\epsilon_t(\theta_0)}{\partial a} \\ \frac{\partial\epsilon_t(\theta_0)}{\partial d} \end{pmatrix} = -\sum_{j\geq 1} \begin{pmatrix} a^{j-1} \\ \frac{1}{j} \end{pmatrix} \epsilon_{t-j}. \tag{3.26}$$

We compute the information matrices $J(\theta_0)$ and $I(\theta_0)$ by using (3.26). Then we have

$$J(\theta_0) = 2\sigma_\epsilon^2 \begin{pmatrix} \frac{1}{1-a^2} & -\frac{\ln(1-a)}{a} \\ -\frac{\ln(1-a)}{a} & \frac{\pi^2}{6} \end{pmatrix}. \tag{3.27}$$

A simple calculation implies that

$$J^{-1}(\theta_0) = \frac{1}{2\sigma_\epsilon^2 c(a)} \begin{pmatrix} \frac{\pi^2}{6} & \frac{\ln(1-a)}{a} \\ \frac{\ln(1-a)}{a} & \frac{1}{1-a^2} \end{pmatrix}, \tag{3.28}$$

where

$$c(a) = \frac{\pi^2}{6(1-a^2)} - \left(\frac{\ln(1-a)}{a}\right)^2. \tag{3.29}$$

We now investigate a similar tractable expression for $I(\theta_0)$. Using (3.26) and (3.21) we have

$$I(\theta_0) = 2\sigma_\epsilon^2 J(\theta_0) + 4\sigma_\epsilon^4 \frac{(\kappa-1)(\alpha_1 - \alpha_1\beta_1^2 - \alpha_1^2\beta_1)}{1-\beta_1^2-2\alpha_1\beta_1-\alpha_1^2\kappa} \begin{pmatrix} \frac{1}{1-a^2(\alpha_1+\beta_1)} & -\frac{\ln[1-a(\alpha_1+\beta_1)]}{a(\alpha_1+\beta_1)} \\ -\frac{\ln[1-a(\alpha_1+\beta_1)]}{a(\alpha_1+\beta_1)} & \frac{\mathrm{Li}_2(\alpha_1+\beta_1)}{\alpha_1+\beta_1} \end{pmatrix}, \tag{3.30}$$

where Li_2 is the Spence function defined by $\mathrm{Li}_2(z) = \sum_{k=1}^{\infty} z^k k^{-2}$. Note that we retrieve the well know result: $I(\theta_0) = 2\sigma_\epsilon^2 J(\theta_0)$ in the strong FARIMA case (*i.e.* when $\alpha_1 = \beta_1 = 0$ in (3.30)).

The matrix defined in (3.7) can be rewritten as

$$\Psi_m = -\sigma_\epsilon^2 \begin{pmatrix} 1 & a & \dots & a^{m-1} \\ 1 & \frac{1}{2} & \dots & \frac{1}{m} \end{pmatrix}'. \tag{3.31}$$

Using (3.25) and under the symmetry assumption (3.21), the matrix $\Gamma_{m,m}$ takes the simple following diagonal form

$$\Gamma_{m,m} = \sigma_\epsilon^4 I_m + \sigma_\epsilon^4 \frac{(\kappa-1)(\alpha_1 - \alpha_1\beta_1^2 - \alpha_1^2\beta_1)}{1-\beta_1^2-2\alpha_1\beta_1-\alpha_1^2\kappa} \mathrm{diag}(1, (\alpha_1+\beta_1), \dots, (\alpha_1+\beta_1)^{m-1}). \tag{3.32}$$

Using (3.21), (3.26) and (3.28), the matrix $\Sigma'_{\hat{\theta},\gamma_m}$ is given by

$$\Sigma'_{\hat{\theta},\gamma_m} = \frac{1}{\sigma_\epsilon^2 c(a)} \begin{pmatrix} \left\{\frac{\pi^2}{6} + \frac{\ln(1-a)}{a}\right\} \Gamma_{m,m}(1,1) & \left\{\frac{1}{1-a^2} + \frac{\ln(1-a)}{a}\right\} \Gamma_{m,m}(1,1) \\ \left\{a\frac{\pi^2}{6} + \frac{\ln(1-a)}{2a}\right\} \Gamma_{m,m}(2,2) & \left\{\frac{1}{2(1-a^2)} + \ln(1-a)\right\} \Gamma_{m,m}(2,2) \\ \vdots & \vdots \\ \left\{a^{m-1}\frac{\pi^2}{6} + \frac{\ln(1-a)}{ma}\right\} \Gamma_{m,m}(m,m) & \left\{\frac{1}{m(1-a^2)} + a^{m-2}\ln(1-a)\right\} \Gamma_{m,m}(m,m) \end{pmatrix}, \tag{3.33}$$

where for any $1 \leq i,j \leq m$, $\Gamma_{m,m}(i,j)$ is given by (3.32).

From Remark 3.2, in the strong FARIMA case the asymptotic variance of residual autocorrelations takes a simpler form

$$\Sigma^{\mathrm{S}}_{\hat{\rho}_m} = I_m - \frac{1}{c(a)}\left[\frac{\pi^2}{6}\left(a^{i+j-2}\right) + \frac{1}{1-a^2}\left(\frac{1}{ij}\right) + \frac{\ln(1-a)}{a}\left(\frac{a^{j-1}}{i} + \frac{a^{i-1}}{j}\right)\right]_{1\leq i,j\leq m}$$

where $c(a)$ is the constant given in (3.29).

From the above explicit expressions we deduce that the asymptotic variance of residual autocorrelations for this model is in the form

$$\begin{aligned}\Sigma_{\hat{\rho}_m} = \Sigma^{\mathrm{S}}_{\hat{\rho}_m} + \frac{(\kappa-1)(\alpha_1 - \alpha_1\beta_1^2 - \alpha_1^2\beta_1)}{1-\beta_1^2-2\alpha_1\beta_1-\alpha_1^2\kappa}\Bigg[&(\alpha_1+\beta_1)^{i-1}\mathbb{1}_{\{i=j\}} + \frac{1}{c(a)}M(i,j) - \Big\{(\alpha_1+\beta_1)^{i-1} \\ &+(\alpha_1+\beta_1)^{j-1}\Big\}\frac{1}{c(a)}\left\{\frac{\pi^2}{6}\left(a^{i+j-2}\right) + \frac{1}{1-a^2}\left(\frac{1}{ij}\right) + \frac{\ln(1-a)}{a}\left(\frac{a^{j-1}}{i} + \frac{a^{i-1}}{j}\right)\right\}\Bigg]_{1\leq i,j\leq m},\end{aligned}$$

where

$$\begin{aligned}M(i,j) = &\left[\frac{\ln(1-a)}{a}\frac{1}{1-a^2(\alpha_1+\beta_1)} - \frac{1}{1-a^2}\frac{\ln(1-a(\alpha_1+\beta_1))}{a(\alpha_1+\beta_1)}\right]\left[\frac{\pi^2}{6}\frac{a^{j-1}}{i} + \frac{1}{ij}\frac{\ln(1-a)}{a}\right] \\ &+\left[\frac{\mathrm{Li}_2(\alpha_1+\beta_1)}{\alpha_1+\beta_1}\frac{1}{1-a^2} - \frac{\ln(1-a)}{a}\frac{\ln(1-a(\alpha_1+\beta_1))}{a(\alpha_1+\beta_1)}\right]\left[\frac{\ln(1-a)}{a}\frac{a^{j-1}}{i} + \frac{1}{ij}\frac{1}{1-a^2}\right] \\ &+\left[\frac{\pi^2}{6}\frac{1}{1-a^2(\alpha_1+\beta_1)} - \frac{\ln(1-a)}{a}\frac{\ln(1-a(\alpha_1+\beta_1))}{a(\alpha_1+\beta_1)}\right]\left[\frac{\pi^2}{6}a^{i+j-2} + \frac{a^{i-1}}{j}\frac{\ln(1-a)}{a}\right] \\ &+\left[\frac{\mathrm{Li}_2(\alpha_1+\beta_1)}{\alpha_1+\beta_1}\frac{\ln(1-a)}{a} - \frac{\pi^2}{6}\frac{\ln(1-a(\alpha_1+\beta_1))}{a(\alpha_1+\beta_1)}\right]\left[\frac{\ln(1-a)}{a}a^{i+j-2} + \frac{a^{i-1}}{j}\frac{1}{1-a^2}\right].\end{aligned}$$

For simplicity, we take in the sequel $\beta_1 = 0$ to consider the case of an ARCH(1) model. For instance when $m = 3$, $\kappa = 3$, $\omega = 1$ and $a = -0.55$ we have

	$\Sigma_{\hat{\rho}_3}$	Eigenvalues $\xi_3 = (\xi_{1,3}, \xi_{2,3}, \xi_{3,3})$	$Z_3(\xi_3)$
$\alpha_1 = 0$	$\begin{pmatrix} 0.1383 & 0.0859 & -0.2720 \\ 0.0859 & 0.2490 & 0.0053 \\ -0.2720 & 0.0053 & 0.9135 \end{pmatrix}$	$(1.0000, 0.2791, 0.0217)$	$\chi_1^2 + 0.2791\chi_1^2 + 0.0217\chi_1^2$
$\alpha_1 = 0.55$	$\begin{pmatrix} 0.6989 & 0.3825 & -1.6041 \\ 0.3825 & 0.9351 & -0.2342 \\ -1.6041 & -0.2342 & 4.7979 \end{pmatrix}$	$(5.3780, 1.0025, 0.0513)$	$5.3780\chi_1^2 + 1.0025\chi_1^2 + 0.0513\chi_1^2$

It is clear that for $\alpha_1 = 0.55$, the Li and McLeod [1986] approximation by a χ^2_1 distribution will be disastrous. The eigenvalues ξ_m can be very different from those of strong FARIMA models which are close to 1 or 0 when the lag m is large enough (see Remark 3.3). More precisely, for instance for $\alpha_1 = 0$ and $m = 12$ we obtain

$$\xi_{12} = (1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 0.0665, 0.0000)',$$

In this weak FARIMA$(1, d, 0)$ with $\alpha_1 = 0.55$ and $m = 12$ we also obtain

$$\xi_{12} = (5.4628, 3.7524, 2.3222, 1.7930, 1.4152, 1.2405, 1.1295, 1.0723, 1.0387, 1.0207, 0.0827, 0.0000)'.$$

The same result holds for FARIMA$(0, d, 1)$ model with a replaced by b in θ_0.

Explicit form of the matrix C_m

The following example gives an explicit form of the normalization matrix C_m for the model given in (3.19). For reading convenience, we restrict ourselves to the case $m = 3$. Using the expression of $J^{-1}(\theta_0)$ given in (3.28) and Equation (3.26), we obtain that for all $1 \leq j \leq n$

$$-2J^{-1}(\theta_0)\epsilon_j \begin{pmatrix} \frac{\partial \epsilon_j(\theta_0)}{\partial a} \\ \frac{\partial \epsilon_j(\theta_0)}{\partial d} \end{pmatrix} = \begin{pmatrix} v_j^{(1)}(a) \\ v_j^{(2)}(a) \end{pmatrix},$$

where

$$v_j^{(1)}(a) = \frac{1}{\sigma_\epsilon^2 c(a)} \sum_{k \geq 1} \left\{ \frac{\pi^2}{6} a^{k-1} + \frac{\ln(1-a)}{a} \frac{1}{k} \right\} \epsilon_j \epsilon_{j-k}$$

and

$$v_j^{(2)}(a) = \frac{1}{\sigma_\epsilon^2 c(a)} \sum_{k \geq 1} \left\{ \frac{\ln(1-a)}{a} a^{k-1} + \frac{1}{1-a^2} \frac{1}{k} \right\} \epsilon_j \epsilon_{j-k}.$$

Thus, the vector ΛU_j is given by

$$\Lambda U_j = \begin{pmatrix} -\sigma_\epsilon^2 v_j^{(1)}(a) - \sigma_\epsilon^2 v_j^{(2)}(a) + \epsilon_j \epsilon_{j-1} \\ -\sigma_\epsilon^2 a v_j^{(1)}(a) - \sigma_\epsilon^2 v_j^{(2)}(a)/2 + \epsilon_j \epsilon_{j-2} \\ -\sigma_\epsilon^2 a^2 v_j^{(1)}(a) - \sigma_\epsilon^2 v_j^{(2)}(a)/3 + \epsilon_j \epsilon_{j-3} \end{pmatrix}.$$

A simple calculation shows that, for any $1 \leq j_1, j_2 \leq n$,

$$(\Lambda U_{j_1})(\Lambda U_{j_2})' = \begin{pmatrix} K_{j_1}^{(1)}(a) K_{j_2}^{(1)}(a) & K_{j_1}^{(1)}(a) K_{j_2}^{(2)}(a) & K_{j_1}^{(1)}(a) K_{j_2}^{(3)}(a) \\ K_{j_1}^{(2)}(a) K_{j_2}^{(1)}(a) & K_{j_1}^{(2)}(a) K_{j_2}^{(2)}(a) & K_{j_1}^{(2)}(a) K_{j_2}^{(3)}(a) \\ K_{j_1}^{(3)}(a) K_{j_2}^{(1)}(a) & K_{j_1}^{(3)}(a) K_{j_2}^{(2)}(a) & K_{j_1}^{(3)}(a) K_{j_2}^{(3)}(a) \end{pmatrix},$$

where

$$\begin{aligned} K_j^{(1)}(a) &= -\sigma_\epsilon^2 v_j^{(1)}(a) - \sigma_\epsilon^2 v_j^{(2)}(a) + \epsilon_j\epsilon_{j-1}, \\ K_j^{(2)}(a) &= -\sigma_\epsilon^2 a v_j^{(1)}(a) - \sigma_\epsilon^2 v_j^{(2)}(a)/2 + \epsilon_j\epsilon_{j-2} \\ \text{and} \quad K_j^{(3)}(a) &= -\sigma_\epsilon^2 a^2 v_j^{(1)}(a) - \sigma_\epsilon^2 v_j^{(2)}(a)/3 + \epsilon_j\epsilon_{j-3}. \end{aligned}$$

Therefore we deduce that for all positive integer t

$$S_t = \sum_{j=1}^{t} (\Lambda U_j - \gamma_3) = \sum_{j=1}^{t} \begin{pmatrix} -\sigma_\epsilon^2 v_j^{(1)}(a) - \sigma_\epsilon^2 v_j^{(2)}(a) + \epsilon_j\epsilon_{j-1} \\ -\sigma_\epsilon^2 a v_j^{(1)}(a) - \sigma_\epsilon^2 v_j^{(2)}(a)/2 + \epsilon_j\epsilon_{j-2} \\ -\sigma_\epsilon^2 a^2 v_j^{(1)}(a) - \sigma_\epsilon^2 v_j^{(2)}(a)/3 + \epsilon_j\epsilon_{j-3} \end{pmatrix} - \frac{t}{n} \begin{pmatrix} \sum_{j=2}^{n} \epsilon_j\epsilon_{j-1} \\ \sum_{j=3}^{n} \epsilon_j\epsilon_{j-2} \\ \sum_{j=4}^{n} \epsilon_j\epsilon_{j-3} \end{pmatrix}.$$

The same result holds for FARIMA$(0, d, 1)$ model with a replaced by b in θ_0.

3.4 Numerical illustrations

In this section, by means of Monte Carlo experiments, we investigate the finite sample properties of the asymptotic results that we introduced in this work. The numerical illustrations of this section are made with the open source statistical software R (see http://cran.r-project.org/).

3.4.1 Simulation studies and empirical sizes

We study numerically the behavior of the least square estimator for FARIMA models of the form

$$(1 - L)^d (X_t - aX_{t-1}) = \epsilon_t - b\epsilon_{t-1}, \tag{3.34}$$

where the unknown parameter is $\theta = (a, b, d)$. First we assume that in (3.34) the innovation process $(\epsilon_t)_{t\in\mathbb{Z}}$ is an iid centered Gaussian process with common variance 1 which corresponds to the strong FARIMA case. We consider that in (3.34) the innovation process $(\epsilon_t)_{t\in\mathbb{Z}}$ follows firstly a GARCH$(1, 1)$ given by (3.20) and secondly a noise defined by

$$\epsilon_t = \eta_t^2 \eta_{t-1} \tag{3.35}$$

where $(\eta_t)_{t\in\mathbb{Z}}$ is a sequence of iid centered Gaussian random variables with variance 1.

We simulate $N = 1,000$ independent trajectories of size $n = 10,000$ of models (3.34). The same series is partitioned as three series of sizes $n = 1,000$, $n = 5,000$ and $n = 10,000$. For each of these N replications, we use the least square estimation method to estimate the coefficient θ and we apply portmanteau tests to the residuals for different values of $m \in \{1, 2, 3, 6, 12, 15\}$, where m is the number of autocorrelations used in the portmanteau test statistic. For the nominal level $\alpha = 5\%$, the empirical size over the N independent replications

should vary between the significant limits 3.6% and 6.4% with probability 95%. When the relative rejection frequencies are outside the 95% significant limits, they are displayed in bold type in Tables 3.1, 3.2, 3.3, 3.4, 3.5 and 3.6.

For the standard Box-Pierce test, the model is therefore rejected when the statistic Q_m^{BP} or Q_m^{LB} is larger than $\chi^2_{(m-p-q-1)}(0.95)$ in a FARIMA(p,d,q) case (see Li and McLeod [1986]). Consequently the empirical size is not available (n.a.) for the statistic Q_m^{BP} or Q_m^{LB} because they are not applicable for $m \leq p+q+1$. For the proposed self-normalized test $\mathrm{BP}_{\mathrm{SN}}$ or $\mathrm{LB}_{\mathrm{SN}}$, the model is rejected when the statistic Q_m^{SN} or $\tilde{Q}_m^{\mathrm{SN}}$ is larger than $\mathcal{U}_m(0.95)$, where the critical values $\mathcal{U}_K(0.95)$ (for $K = 1,\ldots,20$) are tabulated in Lobato (see Table 1 in Lobato [2001]).

Tables 3.1 and 3.4 display the relative rejection frequencies of the null hypothesis **(H0)** that the data generating process (DGP for short) follows a strong FARIMA model (3.34), over the N independent replications. When $p = q = 1$ (resp. $p = q = 0$) for all tests, the percentages of rejection belong to the confident interval with probabilities 95%, except for LB_{s} and BP_{s} (see Table 3.1). Consequently all these tests well control the error of first kind.

We draw the conclusion that in these strong FARIMA cases the proposed modified version may be clearly preferable to the standard ones.

Now, we repeat the same experiments on two weak FARIMA models. As expected Tables 3.2, 3.3, 3.5 and 3.6 show that the standard LB_{s} or BP_{s} test poorly performs in assessing the adequacy of these particular weak FARIMA models. Indeed, we observe that

— the observed relative rejection frequencies of LB_{s} and BP_{s} are definitely outside the significant limits

— the errors of the first kind are only globally well controlled by the proposed tests when n is large.

We also investigate the case where the GARCH model (3.20) have infinite fourth moments. As showing in Figures 3.5,...,3.10 the results are qualitatively similar to what we observe here in Tables 3.2, 3.3, 3.5 and 3.6.

Figures 3.5,...,3.10 display the residual autocorrelations of a realization of size $n = 2,000$ for weak FARIMA models (3.34)–(3.20) with $\omega = 0.04$, $\alpha_1 = 0.13$, $\beta_1 = 0.88$ and three values of d, and their 5% significance limits under the strong FARIMA and weak FARIMA assumptions. These figures confirm clearly the conclusions drawn in Subsection 3.4.1. The horizontal dotted lines (blue color) correspond to the 5% significant limits obtained under the strong FARIMA assumption. The solid lines (red color) and dashed lines (green color) correspond also to the 5% significant limits under the weak FARIMA assumption. The full lines correspond to the asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.1. The dashed lines (green color) correspond to the self-normalized asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.5.

3.4.2 Empirical power

In this section, we repeat the same experiments as in Section 3.4.1 to examine the power of the tests for the null hypothesis of Model (3.34) against the following FARIMA alternative defined by

$$(1-L)^d (X_t - aX_{t-1}) = \epsilon_t - b_1\epsilon_{t-1} - b_2\epsilon_{t-2}, \tag{3.36}$$

with $\theta_0 = (a, b_1, b_2, d_0)$ and where the innovation process $(\epsilon_t)_{t\in\mathbb{Z}}$ follows a strong or weak white noise introduced in Section 3.4.1.

For each of these N replications we fit a FARIMA$(1, d, 1)$ model (3.34) and perform standard and modified tests based on $m = 1, 2, 3, 6, 12$ and 15 residual autocorrelations.

Tables 3.7, 3.8 and 3.9 (resp. Tables 3.10 and 3.11) compare the empirical powers of Model (3.36) with $\theta_0 = (0.9, 1, -0.2, d_0)$ (resp. with $\theta_0 = (0, 0.2, 0, d_0)$) over the N independent replications. For these particular strong and weak FARIMA models, we notice that the standard BP_{S} and LB_{S} and our proposed tests have very similar powers except for $\mathrm{BP}_{\mathrm{SN}}$ and $\mathrm{LB}_{\mathrm{SN}}$ when $n = 1,000$.

In these Monte Carlo experiments, we illustrate that the proposed test statistics have reasonable finite sample performance. Under nonindependent errors, it appears that the standard test statistics are generally non reliable, overrejecting severely, while the proposed tests statistics offer satisfactory levels. Even for independent errors, they seem preferable to the standard ones when the number m of autocorrelations is small. Moreover, the error of first kind is well controlled. Contrarily to the standard tests based on BP_{S} or LB_{S}, the proposed tests can be used safely for m small (see for instance Figure 3.5). For all these above reasons, we think that the modified versions that we propose in this paper are preferable to the standard ones for diagnosing FARIMA models under nonindependent errors.

3.4.3 Illustrative example

We now consider an application to the daily log returns (also simply called the returns) of the Nikkei and Standard & Poor's 500 indices (S&P 500, for short). The returns are defined by $r_t = \log(p_t/p_{t-1})$ where p_t denotes the price index of the S&P 500 index at time t. The observations of the S&P 500 (resp. the Nikkei) index cover the period from January 3, 1950 to to February 14, 2019 (resp. from January 5, 1965 to February 14, 2019). The length of the series is $n = 17,391$ (resp. $n = 13,319$) for the S&P 500 (resp. the Nikkei) index. The data can be downloaded from the website Yahoo Finance: http://fr.finance.yahoo.com/. Figure 3.1 (resp. Figure 3.3) plots the returns and the sample autocorrelations of squared returns of the S&P 500 (resp. of the Nikkei).

In Financial Econometrics the returns are often assumed to be a white noise. In view of the so-called volatility clustering, it is well known that the strong white noise model is not adequate for these series (see for instance Boubacar Mainassara et al. [2012], Boubacar Maïnassara and Saussereau [2018], Francq and Zakoïan [2010], Lobato et al. [2001]). A long-range memory property of the stock market returns series was also largely investigated by Ding et al. [1993] (see also Beran et al. [2013], Palma [2007], Baillie et al. [1996] and Ling and Li [1997]). The squared returns $(r_t^2)_{t\geq 1}$ have significant positive autocorrelations at least up to lag 80 (see Figure 3.1 and Figure 3.3) which confirm the claim that stock market returns have long-term memory (see for instance Ding et al. [1993], for more details). In particular the returns $(r_t)_{t\geq 1}$ process is characterized by substantially more correlation between absolute or squared returns than between the returns themselves.

Therefore we focus on the dynamics of the squared returns and we first fit a FARIMA$(1, d, 1)$ model to the squares of the S&P 500 and Nikkei returns. Denoting by $(X_t)_{t\geq 1}$ the mean

corrected series of the squared returns, we adjust the following model

$$(1-L)^d(X_t - aX_{t-1}) = \epsilon_t - b\epsilon_{t-1}. \tag{3.37}$$

Let $\hat{\theta}_n^{\text{SP500}}$ and $\hat{\theta}_n^{\text{Nikkei}}$ be respectively the least squares estimators of the parameter $\theta = (a, b, d)$ for the model (3.37) in the case of the S&P 500 and Nikkei. The least square estimators were obtained as

$$\hat{\theta}_n^{\text{SP500}} = \begin{pmatrix} -0.3371 & [0.1105] & (0.0023) \\ -0.1795 & [0.0788] & (0.0227) \\ 0.2338 & [0.0367] & (0.0000) \end{pmatrix} \text{ and } \hat{\sigma}_\epsilon^2 = 22.9076 \times 10^{-8}$$

and

$$\hat{\theta}_n^{\text{Nikkei}} = \begin{pmatrix} -0.0217 & [0.1105] & (0.9528) \\ 0.1579 & [0.0788] & (0.6050) \\ 0.3217 & [0.0367] & (0.0000) \end{pmatrix} \text{ and } \hat{\sigma}_\epsilon^2 = 25.6844 \times 10^{-8}, \tag{3.38}$$

where the estimated asymptotic standard errors obtained from $\Sigma_{\hat{\theta}} := J^{-1}IJ^{-1}$ (respectively the p-values), of the estimated parameters (first column), are given into brackets (respectively in parentheses). Note that for these series, the estimated coefficients $|\hat{a}_n|$ and $|\hat{b}_n|$ are smaller than one. This is in accordance with the assumptions that the power series a_θ^{-1} and b_θ^{-1} are well defined (remind that the moving average polynomial is denoted b_θ and the autoregressive polynomials a_θ). We also observe that the estimated long-range dependence coefficients $\hat{d}_n$ is significant for any reasonable asymptotic level and is inside $]0, 0.5[$. So we think that the assumption **(A2)** is satisfied and thus our asymptotic normality theorem on the residual autocorrelations can be applied.

Concerning the S&P 500, the estimators of the parameters a and b are significant whereas it is not the case for the Nikkei (see (3.38)). In the Nikkei case, the coefficients could reasonably be set to zero. So we adjust a FARIMA$(0, d, 0)$ for the squares of Nikkei returns and (3.38) is reduced as

$$\hat{\theta}_n^{\text{Nikkei}} = \begin{pmatrix} 0.2132 & [0.0259] & (0.0000) \end{pmatrix} \text{ and } \hat{\sigma}_\epsilon^2 = 25.9793 \times 10^{-8}.$$

We thus apply portmanteau tests to the residuals of FARIMA$(1, d, 1)$ model for the squares of S&P 500 and FARIMA$(0, d, 0)$ model for the squares of Nikkei. Table 3.12 (resp. Table 3.13) displays the statistics and the p-values of the standard and modified versions of BP and LB tests of model (3.37) (resp. of FARIMA$(0, d, 0)$). From Tables 3.12 and 3.13, we draw the conclusion that the strong FARIMA$(1, 0.2338, 1)$ and FARIMA$(0, 0.2132, 0)$ models are rejected but the weak FARIMA$(1, 0.2338, 1)$ and FARIMA$(0, 0.2132, 0)$ models are not rejected.

Figure 3.2 (resp. Figure 3.4) displays the residual autocorrelations and their 5% significance limits under the strong FARIMA and weak FARIMA assumptions. In view of Figures 3.2 and 3.4, the diagnostic checking of residuals does not indicate any inadequacy for the proposed tests. All of the sample autocorrelations should lie between the bands (at 95%) shown as dashed lines (green color) and solid lines (red color) for the modified tests, while the horizontal dotted (blue color) for standard test indicate that strong FARIMA is not adequate. Figure 3.2 (resp. Figure 3.4) confirms the conclusions drawn from Table 3.12 (resp. Table 3.13).

3.5 Proofs

The following proofs are quite technical and are adaptations of the arguments used in Francq and Zakoïan [1998], Francq et al. [2005] and Boubacar Maïnassara and Saussereau [2018].

The results of Boubacar Maïnassara et al. [2019] which will be needed for all the proofs are collected in the following Subsection 3.5.1 in order to have a self-containing paper.

In all our proofs, K is a strictly positive constant that may vary from line to line.

3.5.1 Preliminary results

In this subsection, we shall give some results on estimations of the coefficient of formal power series that will arise in our study.

We begin by recalling the following properties on power series. If for $|z| \leq R$, the power series $f(z) = \sum_{i\geq 0} a_i z^i$ and $g(z) = \sum_{i\geq 0} b_i z^i$ are well defined, then one has $(fg)(z) = \sum_{i\geq 0} c_i z^i$ is also well defined for $|z| \leq R$ with the sequence $(c_i)_{i\geq 0}$ which is given by $c = a * b$ where $*$ denotes the convolution product between a and b defined by $c_i = \sum_{k=0}^{i} a_k b_{i-k} = \sum_{k=0}^{i} a_{i-k} b_k$. We will make use of the Young inequality that states that if the sequence $a \in \ell^p$ and $b \in \ell^q$ and such that $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$ with $1 \leq p, q, r \leq \infty$, then

$$\|a * b\|_{\ell^r} \leq \|a\|_{\ell^p} \times \|b\|_{\ell^q} .$$

Now we come back to the power series that arise in our context. Remind that for the true value of the parameter,

$$a_{\theta_0}(L)(1-L)^{d_0} X_t = b_{\theta_0}(L)\epsilon_t. \tag{3.39}$$

Thanks to the assumptions on the moving average polynomials b_θ and the autoregressive polynomials a_θ, the power series a_θ^{-1} and b_θ^{-1} are well defined.

Thus the functions $\epsilon_t(\theta)$ defined in (3.2) can be written as

$$\epsilon_t(\theta) = b_\theta^{-1}(L)a_\theta(L)(1-L)^d X_t \tag{3.40}$$

$$= b_\theta^{-1}(L)a_\theta(L)(1-L)^{d-d_0} a_{\theta_0}^{-1}(L)b_{\theta_0}(L)\epsilon_t \tag{3.41}$$

and if we denote $\gamma(\theta) = (\gamma_i(\theta))_{i\geq 0}$ the sequence of coefficients of the power series $b_\theta^{-1}(z)a_\theta(z)(1-z)^d$ (which is absolutely convergent for at least for $|z| \leq 1$), we may write for all $t \in \mathbb{Z}$:

$$\epsilon_t(\theta) = \sum_{i\geq 0} \gamma_i(\theta) X_{t-i}. \tag{3.42}$$

In the same way, by (3.40) one has

$$X_t = (1-L)^{-d} a_\theta^{-1}(L)b_\theta(L)\epsilon_t(\theta)$$

and if we denote $\eta(\theta) = (\eta_i(\theta))_{i\geq 0}$ the coefficients of the power series $(1-z)^{-d}a_\theta^{-1}(z)b_\theta(z)$ one has

$$X_t = \sum_{i\geq 0} \eta_i(\theta)\epsilon_{t-i}(\theta) \ . \tag{3.43}$$

We strength the fact that $\gamma_0(\theta) = \eta_0(\theta) = 1$ for all θ.

For large j, Hallin et al. [1999] have shown that uniformly in θ the sequences $\gamma(\theta)$ and $\eta(\theta)$ satisfy

$$\frac{\partial^k \gamma_j(\theta)}{\partial\theta_{i_1}\cdots\partial\theta_{i_k}} = \mathrm{O}\left(j^{-1-d}\left\{\log(j)\right\}^k\right), \text{ for } k = 0,1,2,3, \tag{3.44}$$

and

$$\frac{\partial^k \eta_j(\theta)}{\partial\theta_{i_1}\cdots\partial\theta_{i_k}} = \mathrm{O}\left(j^{-1+d}\left\{\log(j)\right\}^k\right), \text{ for } k = 0,1,2,3. \tag{3.45}$$

Note that, in view of (3.42), (3.43) and (3.44), for all $\theta \in \Theta_\delta$, $\epsilon_t(\theta)$ belongs to $\mathbb{L}^2$, that $(\epsilon_t(\theta))_{t\in\mathbb{Z}}$ is an ergodic sequence and that, for all $t \in \mathbb{Z}$, the function $\epsilon_t(\cdot)$ is a continuous function.

One difficulty that has to be addressed is that (3.42) includes the infinite past $(X_{t-i})_{i\geq 0}$ whereas only a finite number of observations $(X_t)_{1\leq t\leq n}$ are available to compute the estimators defined in (3.4). The simplest solution is truncation which amounts to setting all unobserved values equal to zero. Thus, for all $\theta \in \Theta$ and $1 \leq t \leq n$ one defines

$$\tilde{\epsilon}_t(\theta) = \sum_{i=0}^{t-1} \gamma_i(\theta) X_{t-i} = \sum_{i\geq 0} \gamma_i^t(\theta) X_{t-i} \tag{3.46}$$

where the truncated sequence $\gamma^t(\theta) = (\gamma_i^t(\theta))_{i\geq 0}$ is defined by

$$\gamma_i^t(\theta) = \begin{cases} \gamma_i(\theta) & \text{if } 0 \leq j \leq t-1\ , \\ 0 & \text{otherwise.} \end{cases}$$

Since our assumptions are made on the noise in (3.1), it will be useful to express the random variables $\epsilon_t(\theta)$ and its partial derivatives with respect to θ, as a function of $(\epsilon_{t-i})_{i\geq 0}$.

From (3.41), there exists a sequence $\lambda(\theta) = (\lambda_i(\theta))_{i\geq 0}$ such that

$$\epsilon_t(\theta) = \sum_{i=0}^{\infty} \lambda_i(\theta)\, \epsilon_{t-i}, \tag{3.47}$$

where the sequence $\lambda(\theta)$ is given by the sequence of the coefficients of the power series $b_\theta^{-1}(z)a_\theta(z)(1-z)^{d-d_0}a_{\theta_0}^{-1}(z)b_{\theta_0}(z)$. Consequently $\lambda(\theta) = \gamma(\theta) * \eta(\theta_0)$ or, equivalently,

$$\lambda_i(\theta) = \sum_{j=0}^{i} \gamma_j(\theta)\eta_{i-j}(\theta_0). \tag{3.48}$$

We proceed in the same way as regard to the derivatives of $\epsilon_t(\theta)$. More precisely, for any $\theta \in \Theta$, $t \in \mathbb{Z}$ and $1 \leq k, l \leq p+q+1$ there exists sequences $\dot{\lambda}_k(\theta) = (\dot{\lambda}_{i,k}(\theta))_{i\geq 1}$ and $\ddot{\lambda}_{k,l}(\theta) = (\ddot{\lambda}_{i,k,l}(\theta))_{i\geq 1}$ such that

$$\frac{\partial \epsilon_t(\theta)}{\partial\theta_k} = \sum_{i=1}^{\infty} \dot{\lambda}_{i,k}(\theta)\, \epsilon_{t-i} \tag{3.49}$$

$$\frac{\partial^2 \epsilon_t(\theta)}{\partial\theta_k\partial\theta_l} = \sum_{i=1}^{\infty} \ddot{\lambda}_{i,k,l}(\theta)\, \epsilon_{t-i}. \tag{3.50}$$

Of course it holds that $\dot{\lambda}_k(\theta) = \frac{\partial \gamma(\theta)}{\partial \theta_k} * \eta(\theta_0)$ and $\ddot{\lambda}_{k,l}(\theta) = \frac{\partial^2 \gamma(\theta)}{\partial \theta_k \partial \theta_l} * \eta(\theta_0)$.

Similarly we have

$$\tilde{\epsilon}_t(\theta) = \sum_{i=0}^{\infty} \lambda_i^t(\theta)\, \epsilon_{t-i}, \tag{3.51}$$

$$\frac{\partial \tilde{\epsilon}_t(\theta)}{\partial \theta_k} = \sum_{i=1}^{\infty} \dot{\lambda}_{i,k}^{\,t}(\theta)\, \epsilon_{t-i} \qquad \text{and} \tag{3.52}$$

$$\frac{\partial^2 \tilde{\epsilon}_t(\theta)}{\partial \theta_k \partial \theta_l} = \sum_{i=1}^{\infty} \ddot{\lambda}_{i,k,l}^{\,t}(\theta)\, \epsilon_{t-i}, \tag{3.53}$$

where $\lambda^t(\theta) = \gamma^t(\theta) * \eta(\theta_0)$, $\dot{\lambda}_k^{\,t}(\theta) = \frac{\partial \gamma^t(\theta)}{\partial \theta_k} * \eta(\theta_0)$ and $\ddot{\lambda}_{k,l}^{\,t}(\theta) = \frac{\partial^2 \gamma^t(\theta)}{\partial \theta_k \partial \theta_l} * \eta(\theta_0)$.

In order to handle the truncation error $\epsilon_t(\theta) - \tilde{\epsilon}_t(\theta)$, one needs some information on the sequence $\lambda(\theta) - \lambda^t(\theta)$. In Boubacar Maïnassara et al. [2019] the following lemmas are proved.

Lemma 3.1. *For $2 \leq r \leq \infty$, $1 \leq k \leq p+q+1$ and $\theta \in \Theta$, we have*

$$\| \lambda(\theta) - \lambda^t(\theta) \|_{\ell^r} = O\left(t^{-1+\frac{1}{r}-(d-d_0)}\right)$$

and

$$\| \dot{\lambda}_k(\theta) - \dot{\lambda}_k^{\,t}(\theta) \|_{\ell^r} = O\left(t^{-1+\frac{1}{r}-(d-d_0)}\right).$$

Lemma 3.2. *For any $2 \leq r \leq \infty$, $1 \leq k \leq p+q+1$ and $\theta \in \Theta$, there exists a constant K such that we have*

$$\| \lambda_k^t(\theta) \|_{\ell^r} \leq K \qquad \text{and}$$
$$\| \dot{\lambda}_k^{\,t}(\theta) \|_{\ell^r} \leq K.$$

3.5.2 Proof of Proposition 3.1

First we remark that the asymptotic normality of the joint distribution of $\sqrt{n}(\hat{\theta}_n' - \theta_0', \gamma_m')'$ can be established along the same lines as the proof of Theorem 2 in Boubacar Maïnassara et al. [2019]. The detailed proof is omitted. From (3.6) and (3.11) we have

$$\sqrt{n}\begin{pmatrix} \hat{\theta}_n - \theta_0 \\ \gamma_m \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \begin{pmatrix} -2J^{-1}(\theta_0)\epsilon_t \frac{\partial}{\partial \theta}\epsilon_t(\theta_0) \\ (\epsilon_{t-1}, \ldots, \epsilon_{t-m})'\epsilon_t \end{pmatrix} + \begin{pmatrix} o_{\mathbb{P}}(1) \\ 0_m \end{pmatrix}$$
$$= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t + o_{\mathbb{P}}(1),$$

where 0_m is the vector of $\mathbb{R}^{m\times 1}$ with zero components. It is clear that U_t is a measurable function of $\epsilon_t, \epsilon_{t-1}, \ldots$ Thus by using the same arguments as in Boubacar Maïnassara et al. [2019] (see proof of Theorem 2), the central limit theorem (CLT) for strongly mixing processes

$(U_t)_{t\in\mathbb{Z}}$ of Herrndorf [1984] implies that $(1/\sqrt{n})\sum_{t=1}^{n} U_t$ has a limiting normal distribution with mean 0 and covariance matrix Ξ.

For $i \geq 1$, we denote $\Lambda_i(\theta_0) = (\dot{\lambda}_{i,1}(\theta_0), \ldots, \dot{\lambda}_{i,p+q+1}(\theta_0))'$. From (3.49) we deduce that

$$\frac{\partial \epsilon_t(\theta_0)}{\partial \theta} = \sum_{i=1}^{\infty} \Lambda_i(\theta_0)\epsilon_{t-i}. \tag{3.54}$$

In view of (3.11) and (3.54), by applying the CLT for mixing processes we directly obtain

$$\begin{aligned}
\Sigma_{\hat{\theta}} &= \lim_{n\to\infty} \mathrm{Var}\left(2J^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t\frac{\partial}{\partial\theta}\epsilon_t(\theta_0)\right) := J^{-1}IJ^{-1} \\
&= 4J^{-1}\sum_{\ell,\ell'=1}^{\infty}\Lambda_\ell(\theta_0)\Lambda'_{\ell'}(\theta_0)\sum_{h=-\infty}^{\infty}\mathbb{E}\left(\epsilon_t\epsilon_{t-\ell}\epsilon_{t-h}\epsilon_{t-\ell'-h}\right)J^{-1} \\
&= 4J^{-1}\sum_{\ell,\ell'=1}^{\infty}\Lambda_\ell(\theta_0)\Lambda'_{\ell'}(\theta_0)\Gamma(\ell,\ell')J^{-1},
\end{aligned}$$

which gives the first block of the asymptotic covariance matrix of Proposition 3.1.

By the stationarity of $(\epsilon_t)_{t\in\mathbb{Z}}$ and Lebesgue's dominated convergence theorem, we obtain the (ℓ,ℓ')-th entry of the matrix $\Gamma_{m,m}$:

$$\begin{aligned}
\lim_{n\to\infty}\mathrm{Cov}(\sqrt{n}\gamma(\ell),\sqrt{n}\gamma(\ell')) &= \lim_{n\to\infty}\frac{1}{n}\sum_{t=\ell+1}^{n}\sum_{s=\ell'+1}^{n}\mathbb{E}\left[\epsilon_t\epsilon_{t-\ell}\epsilon_s\epsilon_{s-\ell'}\right] \\
&= \sum_{h=-\infty}^{\infty}\mathbb{E}\left[\epsilon_t\epsilon_{t-\ell}\epsilon_{t-h}\epsilon_{t-h-\ell'}\right] := \Gamma(\ell,\ell').
\end{aligned}$$

We thus have $\Gamma_{m,m} = [\Gamma(\ell,\ell')]_{1\leq\ell,\ell'\leq m}$.

Finally, by the stationarity of $(\epsilon_t)_{t\in\mathbb{Z}}$ and $(\epsilon_t\partial\epsilon_t(\theta_0)/\partial\theta)_{t\in\mathbb{Z}}$ we have

$$\begin{aligned}
\mathrm{Cov}\left(-2J^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t\frac{\partial}{\partial\theta}\epsilon_t(\theta_0),\sqrt{n}\gamma(\ell')\right) &= -2J^{-1}\frac{1}{n}\sum_{t=1}^{n}\sum_{t'=\ell'+1}^{n}\mathrm{Cov}\left(\epsilon_t\frac{\partial}{\partial\theta}\epsilon_t(\theta_0),\epsilon_{t'}\epsilon_{t'-\ell'}\right) \\
&= -2J^{-1}\frac{1}{n}\sum_{h=-n+1}^{n-1}(n-|h|)\mathrm{Cov}\left(\epsilon_t\frac{\partial\epsilon_t(\theta_0)}{\partial\theta},\epsilon_{t-h}\epsilon_{t-\ell'-h}\right).
\end{aligned}$$

By the dominated convergence theorem and from (3.54), it follows that

$$\begin{aligned}
\lim_{n\to\infty}\mathrm{Cov}\left(-2J^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t\frac{\partial}{\partial\theta}\epsilon_t(\theta_0),\sqrt{n}\gamma(\ell')\right) &= -2J^{-1}\sum_{h=-\infty}^{\infty}\mathrm{Cov}\left(\epsilon_t\frac{\partial}{\partial\theta}\epsilon_t(\theta_0),\epsilon_{t-h}\epsilon_{t-\ell'-h}\right) \\
&= -2J^{-1}\sum_{j\geq1}\Lambda_j(\theta_0)\sum_{h=-\infty}^{\infty}\mathbb{E}\left(\epsilon_t\epsilon_{t-j}\epsilon_{t-h}\epsilon_{t-\ell'-h}\right) \\
&= -2J^{-1}\sum_{j\geq1}\Lambda_j(\theta_0)\Gamma(j,\ell') := \Sigma_{\hat{\theta},\gamma_m}(\cdot,\ell').
\end{aligned}$$

It is clear that the existence of the above matrices is ensured by the existence of $\Gamma(\ell,\ell')$ and $\sum_{\ell,\ell'=1}^{\infty}\|\Lambda_\ell(\theta_0)\Lambda'_{\ell'}(\theta_0)\Gamma(\ell,\ell')\|$. The proof will thus follow from Lemma 3.3 below. □

We now justify the existence of the $\Gamma(\ell,\ell')$ and $\sum_{\ell,\ell'=1}^{\infty}\|\Lambda_\ell(\theta_0)\Lambda'_{\ell'}(\theta_0)\Gamma(\ell,\ell')\|$ in the following result.

Lemma 3.3. *Under the assumptions* **(A0)** *and* **(A3')** *with* $\tau=4$*, we have for* $(\ell,\ell')\neq(0,0)$

$$\Gamma(\ell,\ell')=\sum_{h=-\infty}^{\infty}\mathbb{E}\left(\epsilon_t\epsilon_{t-\ell}\epsilon_{t-h}\epsilon_{t-h-\ell'}\right)<\infty \quad \textit{and} \tag{3.55}$$

$$\sum_{\ell,\ell'=1}^{\infty}\left\|\Lambda_\ell\left(\theta_0\right)\Lambda'_{\ell'}\left(\theta_0\right)\Gamma\left(\ell,\ell'\right)\right\|<\infty. \tag{3.56}$$

Proof. Note that, for all $h\in\mathbb{Z}$ and all $(\ell,\ell')\neq(0,0)$ we have

$$\begin{aligned}\left|\mathbb{E}\left[\epsilon_t\epsilon_{t-\ell}\epsilon_{t-h}\epsilon_{t-h-\ell'}\right]\right| \leq{}& \left|\mathrm{cum}\left(\epsilon_t,\epsilon_{t-\ell},\epsilon_{t-h},\epsilon_{t-h-\ell'}\right)\right|+\left|\mathbb{E}\left[\epsilon_t\epsilon_{t-\ell}\right]\right|\left|\mathbb{E}\left[\epsilon_{t-h}\epsilon_{t-h-\ell'}\right]\right|\\ &+\left|\mathbb{E}\left[\epsilon_t\epsilon_{t-h}\right]\right|\left|\mathbb{E}\left[\epsilon_{t-\ell}\epsilon_{t-h-\ell'}\right]\right|+\left|\mathbb{E}\left[\epsilon_t\epsilon_{t-h-\ell'}\right]\right|\left|\mathbb{E}\left[\epsilon_{t-\ell}\epsilon_{t-h}\right]\right|.\end{aligned}$$

Then, using the stationarity of $(\epsilon_t)_{t\in\mathbb{Z}}$, and under the assumptions **(A0)** and **(A3')** with $\tau=4$ it follows that

$$\Gamma(\ell,\ell')\leq\left[\mathbb{E}\left(\epsilon_t^2\right)\right]^2+\sum_{h=-\infty}^{\infty}\left|\mathrm{cum}\left(\epsilon_0,\epsilon_{-\ell},\epsilon_{-h},\epsilon_{-h-\ell'}\right)\right|\leq K$$

which proves (3.55). Similarly, we obtain

$$\begin{aligned}\sum_{\ell,\ell'=1}^{\infty}\left\|\Lambda_\ell\left(\theta_0\right)\Lambda'_{\ell'}\left(\theta_0\right)\Gamma\left(\ell,\ell'\right)\right\| &\leq \sum_{h=-\infty}^{\infty}\sum_{\ell,\ell'=1}^{\infty}\left|\mathrm{cum}\left(\epsilon_0,\epsilon_{-\ell},\epsilon_{-h},\epsilon_{-h-\ell'}\right)\right|\\ &\quad+\left[\mathbb{E}\left(\epsilon_t^2\right)\right]^2\sum_{\ell=1}^{\infty}\left\|\Lambda_\ell\left(\theta_0\right)\right\|^2\\ &\leq K\end{aligned}$$

where we have used Lemma 3.2. The conclusion follows. □

3.5.3 Proof of Theorem 3.1

The proof is divided in two steps.

Step 1: Taylor's expansion of $\sqrt{n}\hat{\gamma}_m$ and $\sqrt{n}\hat{\rho}_m$

The aim of this step is to prove (3.8) and (3.9). First we prove that for $h=1,\ldots,m$

$$\sqrt{n}\hat{\gamma}(h)=\sqrt{n}\gamma(h)+\left(\mathbb{E}\left[\epsilon_{t-h}\frac{\partial}{\partial\theta'}\epsilon_t(\theta_0)\right]\right)\sqrt{n}\left(\hat{\theta}_n-\theta_0\right)+\mathrm{o}_{\mathbb{P}}(1). \tag{3.57}$$

A Taylor expansion of $(1/\sqrt{n})\sum_{t=1+h}^{n}\tilde{\epsilon}_t(\cdot)\tilde{\epsilon}_{t-h}(\cdot)$ around θ_0 gives

$$\begin{aligned}\sqrt{n}\hat{\gamma}(h) &= \frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\tilde{\epsilon}_t(\theta_0)\tilde{\epsilon}_{t-h}(\theta_0) + \left(\frac{1}{n}\sum_{t=1+h}^{n}\tilde{D}_t(\theta_n^*)\right)\sqrt{n}\left(\hat{\theta}_n-\theta_0\right)\\ &= \sqrt{n}\gamma(h) + \left(\mathbb{E}\left[D_t(\theta_0)\right]\right)\sqrt{n}\left(\hat{\theta}_n-\theta_0\right) + R_{n,h,1} + R_{n,h,2} + R_{n,h,3},\end{aligned}$$

where

$$\begin{aligned}\tilde{D}_t(\theta) &= \frac{\partial\tilde{\epsilon}_t(\theta)}{\partial\theta'}\tilde{\epsilon}_{t-h}(\theta) + \tilde{\epsilon}_t(\theta)\frac{\partial\tilde{\epsilon}_{t-h}(\theta)}{\partial\theta'},\\ D_t(\theta_0) &= \frac{\partial\epsilon_t(\theta_0)}{\partial\theta'}\epsilon_{t-h} + \epsilon_t\frac{\partial\epsilon_{t-h}(\theta_0)}{\partial\theta'},\\ R_{n,h,1} &= \frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\left\{\tilde{\epsilon}_t(\theta_0)\tilde{\epsilon}_{t-h}(\theta_0) - \epsilon_t(\theta_0)\epsilon_{t-h}(\theta_0)\right\},\\ R_{n,h,2} &= \left(\frac{1}{n}\sum_{t=1+h}^{n}\left(\tilde{D}_t(\theta_n^*) - D_t(\theta_0)\right)\right)\sqrt{n}\left(\hat{\theta}_n-\theta_0\right),\\ R_{n,h,3} &= \left(\frac{1}{n}\sum_{t=1+h}^{n}D_t(\theta_0) - \mathbb{E}\left[D_t(\theta_0)\right]\right)\sqrt{n}\left(\hat{\theta}_n-\theta_0\right),\end{aligned}$$

and where θ_n^* is between $\hat{\theta}_n$ and θ_0. Using the orthogonality between ϵ_t and any linear combination of the past values of ϵ_t (in particular $\partial\epsilon_{t-h}/\partial\theta$), we have

$$\sqrt{n}\hat{\gamma}(h) = \sqrt{n}\gamma(h) + \left(\mathbb{E}\left[\epsilon_{t-h}\frac{\partial}{\partial\theta'}\epsilon_t(\theta_0)\right]\right)\sqrt{n}\left(\hat{\theta}_n-\theta_0\right) + R_{n,h,1} + R_{n,h,2} + R_{n,h,3}. \quad (3.58)$$

Thus, to obtain (3.57), we just need to prove that in (3.58) the sequences of random variables $(R_{n,h,1})_n$, $(R_{n,h,2})_n$ and $(R_{n,h,3})_n$ converge in probability to 0.

One of the three above term is easy to handle. Indeed, by the ergodic theorem, we have $n^{-1}\sum_{t=1+h}^{n}D_t(\theta_0) - \mathbb{E}\left[D_t(\theta_0)\right] \to 0$ almost-surely as $n\to\infty$. Thus using the tightness of the sequence $(\sqrt{n}(\hat{\theta}_n-\theta_0))_n$, we deduce that $R_{n,h,3} = o_{\mathbb{P}}(1)$.

The proof of (3.57) will thus follow from Lemmas 3.4 and 3.5 in which the two others terms $R_{n,h,1}$ and $R_{n,h,2}$ are discussed. These lemmas are stated and proved hereafter (see subsections 3.5.3 and 3.5.3).

We now remark that in Equation (3.57), $\mathbb{E}[\epsilon_{t-h}(\partial\epsilon_t(\theta_0)/\partial\theta')]$ is the line h of the matrix $\Psi_m\in\mathbb{R}^{m\times(p+q+1)}$ defined by (3.7). So for $h=1,\ldots,m$, Equation (3.57) becomes

$$\sqrt{n}\hat{\gamma}_m = \left(\sqrt{n}\hat{\gamma}(1),\ldots,\sqrt{n}\hat{\gamma}(m)\right)' = \sqrt{n}\gamma_m + \Psi_m\sqrt{n}\left(\hat{\theta}_n-\theta_0\right) + o_{\mathbb{P}}(1).$$

Therefore the Taylor expansion (3.8) of $\hat{\gamma}_m$ is proved.

Now, it is clear that the asymptotic distribution of the residual autocovariances $\sqrt{n}\hat{\gamma}_m$ is related to the asymptotic behavior of $\sqrt{n}(\hat{\theta}_n'-\theta_0',\gamma_m')'$ obtained in Subsection 3.5.2. We come

back to the vector $\hat{\rho}_m = (\hat{\rho}(1), \ldots, \hat{\rho}(m))'$. Note that from (3.57), we have $\sqrt{n}(\hat{\gamma}(0) - \gamma(0)) = o_{\mathbb{P}}(1)$. Applying the CLT for mixing processes (see Herrndorf [1984]) to the process $(\epsilon_t^2)_{t\in\mathbb{Z}}$, we obtain

$$\sqrt{n}\left(\hat{\sigma}_\epsilon^2 - \sigma_\epsilon^2\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\epsilon_t^2 - \mathbb{E}[\epsilon_t^2]\right) + o_{\mathbb{P}}(1) \xrightarrow[n\to\infty]{\text{in law}} \mathcal{N}\left(0, \sum_{h=-\infty}^{\infty} \text{Cov}\left(\epsilon_t^2, \epsilon_{t-h}^2\right)\right).$$

So we have $\sqrt{n}(\hat{\sigma}_\epsilon^2 - \sigma_\epsilon^2) = O_{\mathbb{P}}(1)$ and $\sqrt{n}(\gamma(0) - \sigma_\epsilon^2) = O_{\mathbb{P}}(1)$. Now, using (3.13) and the ergodic theorem, we have

$$n\left(\frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} - \frac{\hat{\gamma}(h)}{\sigma_\epsilon^2}\right) = \sqrt{n}\hat{\gamma}(h)\frac{\sqrt{n}\left(\sigma_\epsilon^2 - \hat{\gamma}(0)\right)}{\sigma_\epsilon^2\hat{\gamma}(0)} = O_{\mathbb{P}}(1),$$

which means $\sqrt{n}\hat{\rho}(h) = \sqrt{n}\hat{\gamma}(h)/\sigma_\epsilon^2 + O_{\mathbb{P}}(n^{-1/2})$. For $h = 1, \ldots, m$, it follows that

$$\sqrt{n}\hat{\rho}_m = \frac{\sqrt{n}\hat{\gamma}_m}{\sigma_\epsilon^2} + o_{\mathbb{P}}(1),$$

and the Taylor expansion (3.9) of $\hat{\rho}_m$ is proved. This ends our first step.

The next step deals with the asymptotic distributions of $\sqrt{n}\hat{\gamma}_m$ and $\sqrt{n}\hat{\rho}_m$.

Step 2: Asymptotic distributions of $\sqrt{n}\hat{\gamma}_m$ and $\sqrt{n}\hat{\rho}_m$

The joint asymptotic distribution of $\sqrt{n}\gamma_m$ and $\sqrt{n}(\hat{\theta}_n - \theta_0)$ shows that $\sqrt{n}\hat{\gamma}_m$ has a limiting normal distribution with mean zero and covariance matrix

$$\begin{aligned}\lim_{n\to\infty}\text{Var}\left(\sqrt{n}\hat{\gamma}_m\right) &= \lim_{n\to\infty}\text{Var}\left(\sqrt{n}\gamma_m\right) + \Psi_m \lim_{n\to\infty}\text{Var}\left(\sqrt{n}(\hat{\theta}_n - \theta_0)\right)\Psi_m' \\ &+ \Psi_m \lim_{n\to\infty}\text{Cov}\left(\sqrt{n}(\hat{\theta}_n - \theta_0), \sqrt{n}\gamma_m\right) + \lim_{n\to\infty}\text{Cov}\left(\sqrt{n}\gamma_m, \sqrt{n}(\hat{\theta}_n - \theta_0)\right)\Psi_m' \\ &= \Gamma_{m,m} + \Psi_m\Sigma_{\hat{\theta}}\Psi_m' + \Psi_m\Sigma_{\hat{\theta},\gamma_m} + \Sigma_{\hat{\theta},\gamma_m}'\Psi_m'.\end{aligned}$$

Consequently, we have

$$\lim_{n\to\infty}\text{Var}\left(\sqrt{n}\hat{\rho}_m\right) = \lim_{n\to\infty}\text{Var}\left(\sqrt{n}\frac{\hat{\gamma}_m}{\sigma_\epsilon^2}\right) = \frac{1}{\sigma_\epsilon^4}\Sigma_{\hat{\gamma}_m}.$$

This ends our second step and the proof is completed. □

In the following, we justify the convergence of $R_{n,h,1}$, $R_{n,h,2}$.

Step 3: convergence of $R_{n,h,1}$

Lemma 3.4. *Under the assumptions of Theorem 3.1, the sequence of random variables*

$$R_{n,h,1} = \frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\left\{\tilde{\epsilon}_t(\theta_0)\tilde{\epsilon}_{t-h}(\theta_0) - \epsilon_t(\theta_0)\epsilon_{t-h}(\theta_0)\right\}. \tag{3.59}$$

tends to zero in probability as $n \to \infty$.

Proof. Throughout this proof, $\theta = (\theta_1, \ldots, \theta_{p+q}, d)' \in \Theta_\delta$ is such that $d_0 < d \leq d_2$ where d_2 is the upper bound of the support of the long-range parameter d_0. Let

$$R^1_{n,h,1} = \frac{1}{\sqrt{n}} \sum_{t=1+h}^{n} \{\tilde{\epsilon}_t(\theta_0) - \epsilon_t(\theta_0)\} \tilde{\epsilon}_{t-h}(\theta_0) \qquad \text{and} \tag{3.60}$$

$$R^2_{n,h,1} = \frac{1}{\sqrt{n}} \sum_{t=1+h}^{n} \epsilon_t(\theta_0) \{\tilde{\epsilon}_{t-h}(\theta_0) - \epsilon_{t-h}(\theta_0)\}. \tag{3.61}$$

The lemma will be proved as soon as we show that $R^1_{n,h,1}$ and $R^2_{n,h,1}$ tend to zero in probability when $n \to \infty$.

Proof of the convergence in probability of $R^1_{n,h,1}$
The arguments follow the one of Lemma 5 in Boubacar Maïnassara et al. [2019] in a simpler context. The proof is quite long so we divide it in four steps.

◇ **Step 1: preliminaries.** We have

$$\begin{aligned} R^1_{n,h,1} &= \frac{1}{\sqrt{n}} \sum_{t=1+h}^{n} \{\tilde{\epsilon}_t(\theta_0) - \tilde{\epsilon}_t(\theta)\} \tilde{\epsilon}_{t-h}(\theta_0) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{t=1+h}^{n} \{\tilde{\epsilon}_t(\theta) - \epsilon_t(\theta)\} \tilde{\epsilon}_{t-h}(\theta_0) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{t=1+h}^{n} \{\epsilon_t(\theta) - \epsilon_t(\theta_0)\} \tilde{\epsilon}_{t-h}(\theta_0) \\ &= \omega_{n,h,1}(\theta) + \omega_{n,h,2}(\theta) + \omega_{n,h,3}(\theta), \end{aligned}$$

where

$$\omega_{n,h,1}(\theta) = \frac{1}{\sqrt{n}} \sum_{t=1+h}^{n} \{\tilde{\epsilon}_t(\theta_0) - \tilde{\epsilon}_t(\theta)\} \tilde{\epsilon}_{t-h}(\theta_0),$$

$$\omega_{n,h,2}(\theta) = \frac{1}{\sqrt{n}} \sum_{t=1+h}^{n} \{\tilde{\epsilon}_t(\theta) - \epsilon_t(\theta)\} \tilde{\epsilon}_{t-h}(\theta_0) \qquad \text{and}$$

$$\omega_{n,h,3}(\theta) = \frac{1}{\sqrt{n}} \sum_{t=1+h}^{n} \{\epsilon_t(\theta) - \epsilon_t(\theta_0)\} \tilde{\epsilon}_{t-h}(\theta_0).$$

Therefore, if we prove that the two sequences of random variables $(\omega_{n,h,2}(\theta))_n$ and $(\omega_{n,h,1}(\theta) + \omega_{n,h,3}(\theta))_n$ converge in probability towards 0, then the convergence in probability of $R^1_{n,h,1}$ to zero will be true.

◇ **Step 2: convergence in probability of $(\omega_{n,h,2}(\theta))_n$ to 0**

For all $\beta > 0$, we have

$$\begin{aligned}\mathbb{P}\left(|\omega_{n,h,2}| \geq \beta\right) &\leq \frac{1}{\sqrt{n}\beta} \sum_{t=1+h}^{n} \mathbb{E}\left[|\tilde{\epsilon}_t(\theta) - \epsilon_t(\theta)|\, |\tilde{\epsilon}_{t-h}(\theta_0)|\right] \\ &\leq \frac{1}{\sqrt{n}\beta} \sum_{t=1+h}^{n} \|\tilde{\epsilon}_t(\theta) - \epsilon_t(\theta)\|_{\mathbb{L}^2} \|\tilde{\epsilon}_{t-h}(\theta_0)\|_{\mathbb{L}^2}.\end{aligned}$$

First, from (3.51) and using Lemma 3.2, we have

$$\begin{aligned}\|\tilde{\epsilon}_{t-h}(\theta_0)\|^2_{\mathbb{L}^2} &= \mathbb{E}\left[\left(\sum_{i=0}^{\infty} \lambda_i^t(\theta_0)\, \epsilon_{t-i-h}\right)^2\right] \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \lambda_i^t(\theta_0)\, \lambda_j^t(\theta_0)\, \mathbb{E}\left[\epsilon_{t-i-h}\epsilon_{t-j-h}\right] + \sigma_\epsilon^2 \left\{\lambda_0^t(\theta_0)\right\}^2 \\ &= \sigma_\epsilon^2 \sum_{i=1}^{\infty} \left\{\lambda_i^t(\theta_0)\right\}^2 + \sigma_\epsilon^2 \\ &\leq K. \end{aligned} \tag{3.62}$$

In view of (3.47), (3.51) and (3.62), we may write

$$\begin{aligned}\mathbb{P}\left(|\omega_{n,h,2}(\theta)| \geq \beta\right) &\leq \frac{K}{\beta\sqrt{n}} \sum_{t=1+h}^{n} \left(\mathbb{E}\left[\left(\tilde{\epsilon}_{t-h}(\theta) - \epsilon_{t-h}(\theta)\right)^2\right]\right)^{1/2} \\ &\leq \frac{K}{\beta\sqrt{n}} \sum_{t=1+h}^{n} \left(\sum_{i\geq 0}\sum_{j\geq 0} \left(\lambda_i^t(\theta) - \lambda_i(\theta)\right)\left(\lambda_j^t(\theta) - \lambda_j(\theta)\right) \mathbb{E}\left[\epsilon_{t-i-h}\epsilon_{t-j-h}\right]\right)^{1/2} \\ &\leq \frac{\sigma_\epsilon K}{\beta\sqrt{n}} \sum_{t=1}^{n} \left(\sum_{i\geq 0} \left(\lambda_i^t(\theta) - \lambda_i(\theta)\right)^2\right)^{1/2} \\ &\leq \frac{\sigma_\epsilon K}{\beta\sqrt{n}} \sum_{t=1}^{n} \|\lambda(\theta) - \lambda^t(\theta)\|_{\ell^2}.\end{aligned}$$

We use Lemma 3.1, the fact that $d > d_0$ and the fractional version of Cesàro's Lemma [3] to obtain

$$\mathbb{P}\left(|\omega_{n,h,2}(\theta)| \geq \beta\right) \leq \frac{\sigma_\epsilon K}{\beta} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{1}{t^{1/2+(d-d_0)}} \xrightarrow[n\to\infty]{} 0.$$

3. Recall that the fractional version of Cesàro's Lemma states that for $(h_t)_t$ a sequence of positive real numbers, $\kappa > 0$ and $c \geq 0$ we have

$$\lim_{t\to\infty} h_t t^{1-\kappa} = |\kappa|\, c \Rightarrow \lim_{n\to\infty} \frac{1}{n^\kappa} \sum_{t=0}^{n} h_t = c.$$

This proves the expected convergence in probability.

◇ **Step 3: Convergence of** $(\omega_{n,h,1}(\theta)+\omega_{n,h,3}(\theta))_n$

Note now that, for all $n \geq 1$, we have

$$\omega_{n,h,1}(\theta)+\omega_{n,h,3}(\theta) = \frac{1}{\sqrt{n}}\sum_{t=1+h}^{n}\Big\{ (\epsilon_t(\theta)-\tilde{\epsilon}_t(\theta)) - (\epsilon_t(\theta_0)-\tilde{\epsilon}_t(\theta_0)) \Big\}\tilde{\epsilon}_{t-h}(\theta_0).$$

By the mean value theorem, there exists $0 < c_\omega < 1$ such that

$$\Big|(\epsilon_t(\theta)-\tilde{\epsilon}_t(\theta)) - (\epsilon_t(\theta_0)-\tilde{\epsilon}_t(\theta_0))\Big| \leq \left\|\frac{\partial(\epsilon_t-\tilde{\epsilon}_t)}{\partial\theta}((1-c_\omega)\theta+c_\omega\theta_0)\right\|_{\mathbb{R}^{p+q+1}} \|\theta-\theta_0\|_{\mathbb{R}^{p+q+1}} \ . \tag{3.63}$$

Following the same method than in Step 2 we obtain

$$\begin{aligned}
&\mathbb{E}\Big((\epsilon_t(\theta)-\tilde{\epsilon}_t(\theta)) - (\epsilon_t(\theta_0)-\tilde{\epsilon}_t(\theta_0)) \Big)^2 \\
&\qquad \leq \|\theta-\theta_0\|^2_{\mathbb{R}^{p+q+1}} \sum_{k=1}^{p+q+1} \mathbb{E}\left[\left|\frac{\partial(\epsilon_t-\tilde{\epsilon}_t)}{\partial\theta_k}((1-c_\omega)\theta+c_\omega\theta_0)\right|^2\right] \\
&\qquad \leq \|\theta-\theta_0\|^2_{\mathbb{R}^{p+q+1}} \sum_{k=1}^{p+q+1} \sup_\theta \mathbb{E}\left[\left|\frac{\partial(\epsilon_t-\tilde{\epsilon}_t)}{\partial\theta_k}(\theta)\right|^2\right] \\
&\qquad \leq \|\theta-\theta_0\|^2_{\mathbb{R}^{p+q+1}} \sum_{k=1}^{p+q+1} \sigma_\epsilon^2 \sup_\theta \left\|(\dot{\lambda}_k-\dot{\lambda}_k^t)(\theta)\right\|^2_{\ell^2} \\
&\qquad \leq K\,\|\theta-\theta_0\|^2_{\mathbb{R}^{p+q+1}} \sup_{d;d_0\leq d\leq d_2}\left(\frac{1}{t^{1/2+(d-d_0)}}\right)^2 \\
&\qquad \leq K\,\|\theta-\theta_0\|^2_{\mathbb{R}^{p+q+1}} \frac{1}{t},
\end{aligned} \tag{3.64}$$

where we have used the fact that the function

$$\theta \mapsto \mathbb{E}\left[\left|\frac{\partial(\epsilon_t-\tilde{\epsilon}_t)}{\partial\theta_k}(\theta)\right|^2\right]$$

is bounded and continuous. By (3.62) and (3.64), it follows that

$$\mathbb{P}\left(|\omega_{n,h,1}(\theta)+\omega_{n,h,3}(\theta)| \geq \beta\right) \leq \frac{K}{\beta}\|\theta-\theta_0\|_{\mathbb{R}^{p+q+1}} \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{1}{t^{1/2}}$$

and the fractional version of Cesàro's Lemma implies

$$\lim_{n\to\infty}\mathbb{P}\left(|\omega_{n,h,1}(\theta)+\omega_{n,h,3}(\theta)| \geq \beta\right) \leq \frac{K}{\beta}\|\theta-\theta_0\|_{\mathbb{R}^{p+q+1}}. \tag{3.65}$$

◇ **Step 4: end of the proof of the convergence in probability of $R^1_{n,h,1}$ to 0.**
For any $\varepsilon > 0$, we choose θ such that $(K/\beta)\,\|\theta - \theta_0\|_{\mathbb{R}^{p+q+1}} \leq \varepsilon$. Then, from (3.65), there exists n_0 such that for all $n \geq n_0$,

$$\mathbb{P}\left(|\omega_{n,h,1}(\theta) + \omega_{n,h,3}(\theta)| \geq \beta\right) \leq \varepsilon.$$

By Step 2, one also has for $n \geq n_0$

$$\mathbb{P}\left(|\omega_{n,h,2}(\theta)| \geq \beta\right) \leq \varepsilon.$$

Therefore, for all $n \geq n_0$,

$$\mathbb{P}\left(\left|R^1_{n,h,1}\right| \geq 2\beta\right) \leq \mathbb{P}\left(|\omega_{n,h,1}(\theta) + \omega_{n,h,3}(\theta)| \geq \beta\right) + \mathbb{P}\left(|\omega_{n,h,2}(\theta)| \geq \beta\right) \leq \varepsilon$$

and the expected convergence is proved.

Proof of the convergence in probability of $R^2_{n,h,1}$
Under Assumption **(A3)** with $\tau = 2$ it follows that $\epsilon_t(\theta_0)$ belongs to $\mathbb{L}^2$. Thus the proof of the convergence in probability of $R^2_{n,h,1}$ to zero is shown in the same way as the proof of the convergence in probability of $R^1_{n,h,1}$ to 0.

Conclusion : convergence in probability of $R_{n,h,1}$
The conclusion is a consequence of the above convergences. □

Step 4: convergence of $R_{n,h,2}$

Lemma 3.5. *Under the assumptions of Theorem 3.1, the sequences of random variables*

$$R_{n,h,2} = \left(\frac{1}{n}\sum_{t=1+h}^{n}\left(\tilde{D}_t(\theta_n^*) - D_t(\theta_0)\right)\right)\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \tag{3.66}$$

tend to zero in probability as $n \to \infty$ and where θ_n^* is between $\hat{\theta}_n$ and θ_0.

Proof. Since $(\sqrt{n}(\hat{\theta}_n - \theta_0))_n$ is a tight sequence, we have $\sqrt{n}(\hat{\theta}_n - \theta_0) = \mathrm{O}_{\mathbb{P}}(1)$. Hence, to prove the convergence in probability of $(R_{n,h,2})_n$ to 0, it suffices to show that

$$\frac{1}{n}\sum_{t=1+h}^{n}\left(\tilde{D}_t(\theta_n^*) - D_t(\theta_0)\right) = \mathrm{o}_{\mathbb{P}}(1). \tag{3.67}$$

This will be proved using Lemma 3.1 and Cesàro's Lemma. Nevertheless, the proof is quite long so we divide it in four steps.

◇ **Step 1: preliminaries.** We have

$$\frac{1}{n}\sum_{t=1+h}^{n}\left(\tilde{D}_t(\theta_n^*) - D_t(\theta_0)\right) = T_{n,h,1}(\theta_n^*) + T_{n,h,2}(\theta_n^*) + T_{n,h,3}(\theta_n^*) + T_{n,h,4}(\theta_n^*) + T_{n,h,5}(\theta_n^*),$$

where

$$\begin{aligned}
T_{n,h,1}(\theta) &= \frac{1}{n}\sum_{t=1+h}^{n}\frac{\partial\tilde{\epsilon}_t(\theta)}{\partial\theta'}\left(\tilde{\epsilon}_{t-h}(\theta) - \epsilon_{t-h}(\theta)\right),\\
T_{n,h,2}(\theta) &= \frac{1}{n}\sum_{t=1+h}^{n}\left(\tilde{\epsilon}_t(\theta) - \epsilon_t(\theta)\right)\frac{\partial\tilde{\epsilon}_{t-h}(\theta)}{\partial\theta'},\\
T_{n,h,3}(\theta) &= \frac{1}{n}\sum_{t=1+h}^{n}\left(\frac{\partial\tilde{\epsilon}_t(\theta)}{\partial\theta'} - \frac{\partial\epsilon_t(\theta)}{\partial\theta'}\right)\epsilon_{t-h}(\theta),\\
T_{n,h,4}(\theta) &= \frac{1}{n}\sum_{t=1+h}^{n}\epsilon_t(\theta)\left(\frac{\partial\tilde{\epsilon}_{t-h}(\theta)}{\partial\theta'} - \frac{\partial\epsilon_{t-h}(\theta)}{\partial\theta'}\right) \quad \textit{and}\\
T_{n,h,5}(\theta) &= \frac{1}{n}\sum_{t=1+h}^{n}\left(D_t(\theta) - D_t(\theta_0)\right).
\end{aligned}$$

Therefore, if we prove that the five sequences of random variables $(T_{n,h,i}(\theta))_n$ (for $i = 1,\ldots,5$) converge in probability towards 0, then (3.67) will be true.

◇ **Step 2: convergence in probability of $(T_{n,h,1}(\theta))_n$ to 0**
For all $\beta > 0$, we have

$$\begin{aligned}
\mathbb{P}\left(\|T_{n,h,1}(\theta)\| \geq \beta\right) &\leq \frac{1}{n\beta}\sum_{t=1+h}^{n}\mathbb{E}\left[\left\|\frac{\partial\tilde{\epsilon}_t(\theta)}{\partial\theta'}\right\|\left|\tilde{\epsilon}_{t-h}(\theta) - \epsilon_{t-h}(\theta)\right|\right]\\
&\leq \frac{1}{n\beta}\sum_{t=1+h}^{n}\left\|\tilde{\epsilon}_{t-h}(\theta) - \epsilon_{t-h}(\theta)\right\|_{\mathbb{L}^2}\left\|\frac{\partial\tilde{\epsilon}_t(\theta)}{\partial\theta'}\right\|_{\mathbb{L}^2}.
\end{aligned}$$

First, from (3.51) and using Lemma 3.2 we have for $1 \leq k \leq p+q+1$

$$\begin{aligned}
\left\|\frac{\partial}{\partial\theta_k}\tilde{\epsilon}_t(\theta)\right\|_{\mathbb{L}^2}^2 &= \mathbb{E}\left[\left(\sum_{i=1}^{\infty}\dot{\lambda}_{i,k}^{t}(\theta)\epsilon_{t-i}\right)^2\right]\\
&= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\dot{\lambda}_{i,k}^{t}(\theta)\dot{\lambda}_{j,k}^{t}(\theta)\mathbb{E}\left[\epsilon_{t-i}\epsilon_{t-j}\right]\\
&= \sigma_\epsilon^2\sum_{i=1}^{\infty}\left\{\dot{\lambda}_{i,k}^{t}(\theta)\right\}^2\\
&\leq K. \qquad (3.68)
\end{aligned}$$

In view of (3.47), (3.51), (3.68) and following the same way as the step 2 of Lemma 3.4 we have

$$\begin{aligned}
\mathbb{P}\left(|T_{n,h,1}(\theta)| \geq \beta\right) &\leq \frac{K}{\beta n}\sum_{t=1+h}^{n}\left(\mathbb{E}\left[\left(\tilde{\epsilon}_{t-h}(\theta)-\epsilon_{t-h}(\theta)\right)^2\right]\right)^{1/2} \\
&\leq \frac{K}{\beta n}\sum_{t=1+h}^{n}\left(\sum_{i\geq 0}\sum_{j\geq 0}\left(\lambda_i^t(\theta)-\lambda_i(\theta)\right)\left(\lambda_j^t(\theta)-\lambda_j(\theta)\right)\mathbb{E}\left[\epsilon_{t-i-h}\epsilon_{t-j-h}\right]\right)^{1/2} \\
&\leq \frac{\sigma_\epsilon K}{\beta n}\sum_{t=1}^{n}\left(\sum_{i\geq 0}\left(\lambda_i^t(\theta)-\lambda_i(\theta)\right)^2\right)^{1/2} \\
&\leq \frac{\sigma_\epsilon K}{\beta n}\sum_{t=1}^{n}\left\|\lambda(\theta)-\lambda^t(\theta)\right\|_{\ell^2}.
\end{aligned}$$

We use Lemma 3.1, the fact that $|d-d_0|<1/2$ and the Cesàro Lemma to obtain

$$\mathbb{P}\left(|T_{n,h,1}(\theta)| \geq \beta\right) \leq \frac{\sigma_\epsilon K}{\beta}\frac{1}{n}\sum_{t=1}^{n}\frac{1}{t^{1/2+(d-d_0)}} \xrightarrow[n\to\infty]{} 0.$$

This proves the expected convergence in probability of $T_{n,h,1}(\theta)$.

The same calculations holds for the sequences of random variables $(T_{n,h,2}(\theta))_n$, $(T_{n,h,3}(\theta))_n$ and $(T_{n,h,4}(\theta))_n$.

◇ **Step 3: convergence in probability of $(T_{n,h,5}(\theta))_n$ to 0**

For $1\leq i,j\leq p+q+1$ and in view of (3.42), (3.44), we have

$$\begin{aligned}
\sup_{\theta\in\Theta_\delta}\left|\frac{\partial}{\partial\theta_i}\epsilon_t(\theta)\frac{\partial}{\partial\theta_j}\epsilon_t(\theta)\right| &= \sup_{\theta\in\Theta_\delta}\left|\sum_{k_1,k_2\geq 1}\frac{\partial}{\partial\theta_i}\gamma_{k_1}(\theta)\frac{\partial}{\partial\theta_j}\gamma_{k_2}(\theta)X_{t-k_1}X_{t-k_2}\right| \\
&\leq \sum_{k_1,k_2\geq 1}\sup_{\theta\in\Theta_\delta}\left|\frac{\partial}{\partial\theta_i}\gamma_{k_1}(\theta)\right|\sup_{\theta\in\Theta_\delta}\left|\frac{\partial}{\partial\theta_j}\gamma_{k_2}(\theta)\right||X_{t-k_1}|\,|X_{t-k_2}| \\
&\leq K\sum_{k_1,k_2\geq 1}\log(k_1)k_1^{-1-d_1}\log(k_2)k_2^{-1-d_1}|X_{t-k_1}|\,|X_{t-k_2}|.
\end{aligned}$$

Consequently we obtain

$$\begin{aligned}
\mathbb{E}_{\theta_0}\left[\sup_{\theta\in\Theta_\delta}\left|\frac{\partial}{\partial\theta_i}\epsilon_t(\theta)\frac{\partial}{\partial\theta_j}\epsilon_t(\theta)\right|\right] &\leq K\sum_{k_1,k_2\geq 1}\log(k_1)k_1^{-1-d_1}\log(k_2)k_2^{-1-d_1}\sup_{t\in\mathbb{Z}}\mathbb{E}_{\theta_0}|X_t|^2 \\
&\leq K. \qquad (3.69)
\end{aligned}$$

Following the same approach used to obtain (3.69), we have

$$\mathbb{E}_{\theta_0}\left[\sup_{\theta\in\Theta_\delta}\left|\epsilon_t(\theta)\frac{\partial^2}{\partial\theta_i\partial\theta_j}\epsilon_t(\theta)\right|\right] < \infty. \qquad (3.70)$$

A Taylor expansion of $D_t(\cdot)$ around θ_0 implies that

$$\|T_{n,h,5}(\theta)\| \leq \frac{1}{n}\sum_{t=1}^{n} \sup_{\theta\in\Theta_\delta} \left\| \frac{\partial}{\partial\theta} D_t(\theta) \right\| \|\theta_n^* - \theta_0\| .$$

From (3.69) and (3.70), it follows that

$$\begin{aligned}\mathbb{E}\left[\sup_{\theta\in\Theta_\delta}\left\|\frac{\partial}{\partial\theta}D_t(\theta)\right\|\right] &= \mathbb{E}\left[\sup_{\theta\in\Theta_\delta}\left\|\epsilon_{t-h}(\theta)\frac{\partial^2}{\partial\theta\partial\theta'}\epsilon_t(\theta) + \frac{\partial}{\partial\theta}\epsilon_{t-h}(\theta)\frac{\partial}{\partial\theta'}\epsilon_t(\theta)\right.\right. \\ &\qquad \left.\left. + \frac{\partial}{\partial\theta}\epsilon_t(\theta)\frac{\partial}{\partial\theta'}\epsilon_{t-h}(\theta) + \epsilon_t(\theta)\frac{\partial^2}{\partial\theta\partial\theta'}\epsilon_{t-h}(\theta)\right\|\right] \\ &\leq K. \end{aligned} \tag{3.71}$$

Using Equation (3.71), the ergodic theorem and the almost-sure convergence of $(\hat{\theta}_n - \theta_0)_n$ to 0 imply that $T_{n,h,5}(\theta)$ tends to 0 almost-surely.

◇ **Step 4: end of the proof of the convergence in probability of $R_{n,h,2}$ to zero.**
By Step 2 and 3 we deduce that

$$R_{n,h,2} = \mathrm{o}_{\mathbb{P}}(1)$$

and the convergence in probability is proved.

The proof of the lemma is completed. □

3.5.4 Proof of Proposition 3.2

The following proofs are quite technical and are adaptations of the arguments used in Boubacar Maïnassara and Saussereau [2018].

To prove the invertibility of the normalized matrix C_m, we need to introduce the following notation.

Let $S_t(i)$ be the i-th component of the vector $S_t = \sum_{j=1}^{t}(\Lambda U_j - \gamma_m) \in \mathbb{R}^m$. We remark that

$$S_{t-1}(i) = S_t(i) - \sum_{k=1}^{p+q+1} \delta_{i,k}\epsilon_t \frac{\partial}{\partial\theta_k}\epsilon_t(\theta_0) - \epsilon_t\epsilon_{t-i} + \gamma(i), \tag{3.72}$$

where $\delta_{i,k}$ is the $(i,k)-$th entry of the $m\times(p+q+1)$ matrix $\Delta := -2\Psi_m J^{-1}$.

If the matrix C_m is not invertible, there exists some real constants $c_1,\ldots,c_m$ not all equal to zero, such that we have

$$\sum_{i=1}^{m}\sum_{j=1}^{m} c_j C_m(j,i) c_i = \frac{1}{n^2}\sum_{t=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{m} c_j S_t(j) S_t(i) c_i = \frac{1}{n^2}\sum_{t=1}^{n}\left(\sum_{i=1}^{m} c_i S_t(i)\right)^2 = 0,$$

which implies that $\sum_{i=1}^{m} c_i S_t(i) = 0$ for all $t \geq 1$.

Then by (3.72), it would imply that

$$\sum_{i=1}^{m}\sum_{k=1}^{p+q+1} c_i\delta_{i,k}\epsilon_t\frac{\partial}{\partial\theta_k}\epsilon_t(\theta_0)+\sum_{i=1}^{m} c_i\epsilon_t\epsilon_{t-i}=\sum_{i=1}^{m} c_i\gamma(i). \tag{3.73}$$

By the ergodic Theorem, we also have $\sum_{i=1}^{m} c_i\gamma(i)\to 0$ almost-surely as n goes to infinity.

Consequently replacing this convergence in (3.73) implies that for all $t\geq 1$

$$\sum_{i=1}^{m}\sum_{k=1}^{p+q+1} c_i\delta_{i,k}\epsilon_t\frac{\partial}{\partial\theta_k}\epsilon_t(\theta_0)+\sum_{i=1}^{m} c_i\epsilon_t\epsilon_{t-i}=0, \quad \text{a.s.}$$

Using (3.47), it yields that

$$\epsilon_t\left\{\sum_{\ell\geq 1}\left(\sum_{i=1}^{m}\sum_{k=1}^{p+q+1} c_i\delta_{i,k}\dot{\lambda}_{\ell,k}(\theta_0)\right)\epsilon_{t-\ell}+\sum_{\ell=1}^{m} c_\ell\epsilon_{t-\ell}\right\}=0, \quad \text{a.s.}$$

Or equivalently,

$$\epsilon_t\left\{\sum_{\ell=1}^{m}\left(\sum_{i=1}^{m} c_i\sum_{k=1}^{p+q+1}\delta_{i,k}\dot{\lambda}_{\ell,k}(\theta_0)+c_\ell\right)\epsilon_{t-\ell}+\sum_{\ell\geq m+1}\left(\sum_{i=1}^{m} c_i\sum_{k=1}^{p+q+1}\delta_{i,k}\dot{\lambda}_{\ell,k}(\theta_0)\right)\epsilon_{t-\ell}\right\}=0, \quad \text{a.s.}$$

Thanks to Assumption **(A4)**, ϵ_t has a positive density in some neighborhood of zero and then $\epsilon_t\neq 0$ almost-surely. Hence we obtain

$$\sum_{\ell=1}^{m}\left(\sum_{i=1}^{m} c_i\sum_{k=1}^{p+q+1}\delta_{i,k}\dot{\lambda}_{\ell,k}(\theta_0)+c_\ell\right)\epsilon_{t-\ell}+\sum_{\ell\geq m+1}\left(\sum_{i=1}^{m} c_i\sum_{k=1}^{p+q+1}\delta_{i,k}\dot{\lambda}_{\ell,k}(\theta_0)\right)\epsilon_{t-\ell}=0, \quad \text{a.s.}$$

Since the variance of the linear innovation process in not equal to zero, we deduce that

$$\begin{cases} \sum_{i=1}^{m} c_i\sum_{k=1}^{p+q+1}\delta_{i,k}\dot{\lambda}_{\ell,k}(\theta_0)+c_\ell=0 & \text{for all } \ell\in\{1,\ldots,m\} \\ \sum_{i=1}^{m} c_i\sum_{k=1}^{p+q+1}\delta_{i,k}\dot{\lambda}_{\ell,k}(\theta_0)=0 & \text{for all } \ell\in\{m+1,\ldots\}. \end{cases}$$

Then we would have $c_1=\cdots=c_m=0$ which is impossible. Thus we have a contradiction and the matrix $C_m\in\mathbb{R}^{m\times m}$ is non singular. □

3.5.5 Proof of Theorem 3.4

We recall that the Skorohod space $\mathbb{D}^\ell[0,1]$ is the set of $\mathbb{R}^\ell-$valued functions on [0,1] which are right-continuous and have left limits everywhere. It is endowed with the Skorohod topology and the weak convergence on $\mathbb{D}^\ell[0,1]$ is mentioned by $\xrightarrow{\mathbb{D}^\ell}$. The integer part of x will be denoted by $\lfloor x\rfloor$.

The proof is divided in two steps.

Functional central limit theorem for $(\Lambda U_t)_{t\geq 1}$

In view of (3.8) and (3.12), we deduce that

$$\begin{aligned}\sqrt{n}\hat{\gamma}_m &= \sqrt{n}\gamma_m + \sqrt{n}\Psi_m\left(\hat{\theta}_n - \theta_0\right) + o_{\mathbb{P}}(1)\\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_{2t} + \Psi_m\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_{1t} + o_{\mathbb{P}}(1)\right) + o_{\mathbb{P}}(1)\\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Lambda U_t + o_{\mathbb{P}}(1). \end{aligned} \tag{3.74}$$

Now, it is clear that the asymptotic behaviour of $\hat{\gamma}_m$ is related to the limit distribution of $U_t = (U_{1t}', U_{2t}')'$. Our first goal is to show that there exists a lower triangular matrix Π with nonnegative diagonal entries such that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr\rfloor}\Lambda U_t \xrightarrow[n\to\infty]{\mathbb{D}^m} \left(\Pi\Pi'\right)^{1/2} B_m(r), \tag{3.75}$$

where $(B_m(r))_{r\geq 0}$ is a $m-$dimensional standard Brownian motion. Using (3.47), U_t can be rewritten as

$$U_t = \left(-2\left\{\sum_{i=1}^{\infty}\dot{\lambda}_{i,1}(\theta_0)\,\epsilon_t\epsilon_{t-i}, \ldots, \sum_{i=1}^{\infty}\dot{\lambda}_{i,p+q+1}(\theta_0)\,\epsilon_t\epsilon_{t-i}\right\}J^{-1'}, \epsilon_t\epsilon_{t-1}, \ldots, \epsilon_t\epsilon_{t-m}\right)'.$$

The non-correlation between ϵ_t's implies that the process $(U_t)_{t\in\mathbb{Z}}$ of $\mathbb{R}^{p+q+1+m}$ is centered. In order to apply the functional central limit theorem for strongly mixing process (see Herrndorf [1984]), we need to identify the asymptotic covariance matrix in the classical central limit theorem for the sequence $(U_t)_{t\in\mathbb{Z}}$. It is proved in Proposition 3.1 that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t \xrightarrow[n\to\infty]{\text{in law}} \mathcal{N}\left(0, \Xi := 2\pi f_U(0)\right),$$

where $f_U(0)$ is the spectral density of the stationary process $(U_t)_{t\in\mathbb{Z}}$ evaluated at frequency 0. The existence of the matrix Ξ has already been discussed in Lemma 3.3.

Since the matrix Ξ is positive definite, it can be factorized as $\Xi = \Upsilon\Upsilon'$, where the $(p+q+1+m)\times(p+q+1+m)$ lower triangular matrix Υ has nonnegative diagonal entries. Therefore, we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Lambda U_t \xrightarrow[n\to\infty]{\text{in law}} \mathcal{N}\left(0, \Lambda\Xi\Lambda'\right),$$

and the new variance matrix can also been factorized as $\Lambda\Xi\Lambda' = (\Lambda\Upsilon)(\Lambda\Upsilon)' := \Pi\Pi'$, where $\Pi \in \mathbb{R}^{m\times(p+q+1)}$. Thus

$$n^{-1/2}\sum_{t=1}^{n}(\Pi\Pi')^{-1/2}\Lambda U_t \xrightarrow[n\to\infty]{\text{in law}} \mathcal{N}(0, I_m),$$

where $(\Pi\Pi')^{-1/2}$ is the Moore-Penrose inverse (see Magnus and Neudecker [1999], p. 36) of $(\Pi\Pi')^{1/2}$.

Using the same arguments as in the proof of Theorem 2 in Boubacar Maïnassara et al. [2019], the asymptotic distribution of $n^{-1/2}\sum_{t=1}^{n} U_t$ when n tends to infinity is obtained by introducing the random vector U_t^k defined for any strictly positive integer k by

$$U_t^k = \left(-2 \left\{ \sum_{i=1}^{k} \dot{\lambda}_{i,1}(\theta_0)\, \epsilon_t \epsilon_{t-i}, \ldots, \sum_{i=1}^{k} \dot{\lambda}_{i,p+q+1}(\theta_0)\, \epsilon_t \epsilon_{t-i} \right\} J^{-1'}, \epsilon_t\epsilon_{t-1}, \ldots, \epsilon_t\epsilon_{t-m} \right)'.$$

Since U_t^k depends on a finite number of values of the noise-process $(\epsilon_t)_{t\in\mathbb{Z}}$, it also satisfies a mixing property (see Theorem 14.1 in Davidson [1994], p. 210). Then applying the central limit theorem for strongly mixing process of Herrndorf [1984] shows that its asymptotic distribution is normal with zero mean and variance matrix Ξ_k that converges when k tends to infinity to Ξ. More precisely we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t^k \xrightarrow[n\to\infty]{\text{in law}} \mathcal{N}(0, \Xi_k).$$

The above arguments also apply to matrix Ξ_k with some matrix Π_k which is defined analogously as Π. Consequently we obtain

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \Lambda U_t^k \xrightarrow[n\to\infty]{\text{in law}} \mathcal{N}\left(0, \Lambda\Xi_k\Lambda'\right)$$

and we also have $n^{-1/2}\sum_{t=1}^{n}(\Pi_k\Pi_k')^{-1/2}\Lambda U_t^k \xrightarrow[n\to\infty]{\text{in law}} \mathcal{N}(0, I_m)$.

Now we are able to apply the functional central limit theorem (see Herrndorf [1984]) and we obtain that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor}(\Pi_k\Pi_k')^{-1/2}\Lambda U_t^k \xrightarrow[n\to\infty]{\mathbb{D}^m} B_m(r).$$

Since for all $t \in \{1, \ldots, \lfloor nr \rfloor\}$ we write

$$(\Pi\Pi')^{-1/2}\Lambda U_t^k = \left((\Pi\Pi')^{-1/2} - (\Pi_k\Pi_k')^{-1/2}\right)\Lambda U_t^k + (\Pi_k\Pi_k')^{-1/2}\Lambda U_t^k,$$

we obtain the following weak convergence on $\mathbb{D}^m[0,1]$:

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor}(\Pi\Pi')^{-1/2}\Lambda U_t^k \xrightarrow[n\to\infty]{\mathbb{D}^m} B_m(r).$$

In order to conclude that (3.75) is true, it remains to observe that uniformly with respect to n

$$Y_n^k(r) := \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor}(\Pi\Pi')^{-1/2}\Lambda Z_t^k \xrightarrow[k\to\infty]{\mathbb{D}^m} 0, \tag{3.76}$$

where

$$Z_t^k = \left(-2 \left\{ \sum_{i=k+1}^{\infty} \dot{\lambda}_{i,1}(\theta_0)\, \epsilon_t \epsilon_{t-i}, \ldots, \sum_{i=k+1}^{\infty} \dot{\lambda}_{i,p+q+1}(\theta_0)\, \epsilon_t \epsilon_{t-i} \right\} J^{-1'}, \epsilon_t \epsilon_{t-1}, \ldots, \epsilon_t \epsilon_{t-m} \right)'.$$

Using the same arguments as those used in the proof of Theorem 2 in Boubacar Maïnassara et al. [2019], we have

$$\sup_n \mathrm{Var}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} Z_t^k \right) \underset{k\to\infty}{\longrightarrow} 0$$

and since $\lfloor nr \rfloor \leq n$,

$$\sup_{0\leq r\leq 1} \sup_n \left\{ \left\| Y_n^k(r) \right\| \right\} \underset{k\to\infty}{\longrightarrow} 0.$$

Thus (3.76) is true and the proof of (3.75) is achieved.

Limit theorem

To conclude the prove of Theorem 3.4, we follow the arguments developed in Boubacar Maïnassara and Saussereau [2018]. Note that the previous step ensures us that Assumption 1 in Lobato [2001] is satisfied for the sequence $(\Lambda U_t)_{t\geq 1}$. Firstly from (3.75) we deduce that

$$\begin{aligned} \frac{1}{\sqrt{n}} S_{\lfloor nr \rfloor} &= \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \Lambda U_t - \frac{\lfloor nr \rfloor}{n} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \Lambda U_t \right) \\ &\underset{n\to\infty}{\overset{\mathbb{D}^m}{\longrightarrow}} (\Pi\Pi')^{1/2} B_m(r) - r(\Pi\Pi')^{1/2} B_m(1). \end{aligned} \tag{3.77}$$

Observe now that the continuous mapping theorem implies

$$\begin{aligned} C_m &= \frac{1}{n} \sum_{t=1}^{n} \left(\frac{1}{\sqrt{n}} S_t \right) \left(\frac{1}{\sqrt{n}} S_t \right)' \\ &\underset{n\to\infty}{\overset{\mathbb{D}^m}{\longrightarrow}} (\Pi\Pi')^{1/2} \left[\int_0^1 \{B_m(r) - rB_m(1)\} \{B_m(r) - rB_m(1)\}' \, \mathrm{d}r \right] (\Pi\Pi')^{1/2} = (\Pi\Pi')^{1/2} V_m (\Pi\Pi')^{1/2}. \end{aligned}$$

Using (3.74), (3.77) and again the continuous mapping theorem on the Skorohod space, one finally obtains

$$\begin{aligned} n\hat{\gamma}_m' C_m^{-1} \hat{\gamma}_m &\underset{n\to\infty}{\overset{\mathbb{D}^m}{\longrightarrow}} \left\{ (\Pi\Pi')^{1/2} B_m(1) \right\}' \left\{ (\Pi\Pi')^{1/2} V_m (\Pi\Pi')^{1/2} \right\}^{-1} \left\{ (\Pi\Pi')^{1/2} B_m(1) \right\} \\ &= B_m'(1) V_m^{-1} B_m(1) := \mathcal{U}_m. \end{aligned}$$

Consequently, from (3.9) it follows that

$$n\sigma_\epsilon^4 \hat{\rho}_m' C_m^{-1} \hat{\rho}_m \underset{n\to\infty}{\overset{\mathbb{D}^m}{\longrightarrow}} \mathcal{U}_m,$$

which completes the proof of Theorem 3.4. □

3.5.6 Proof of Theorem 3.5

The proof follows the same line than in the proof of Theorem 2 in Boubacar Maïnassara and Saussereau [2018] (see also the proof of in Boubacar Maïnassara et al. [2019]).

3.6 Figures and tables

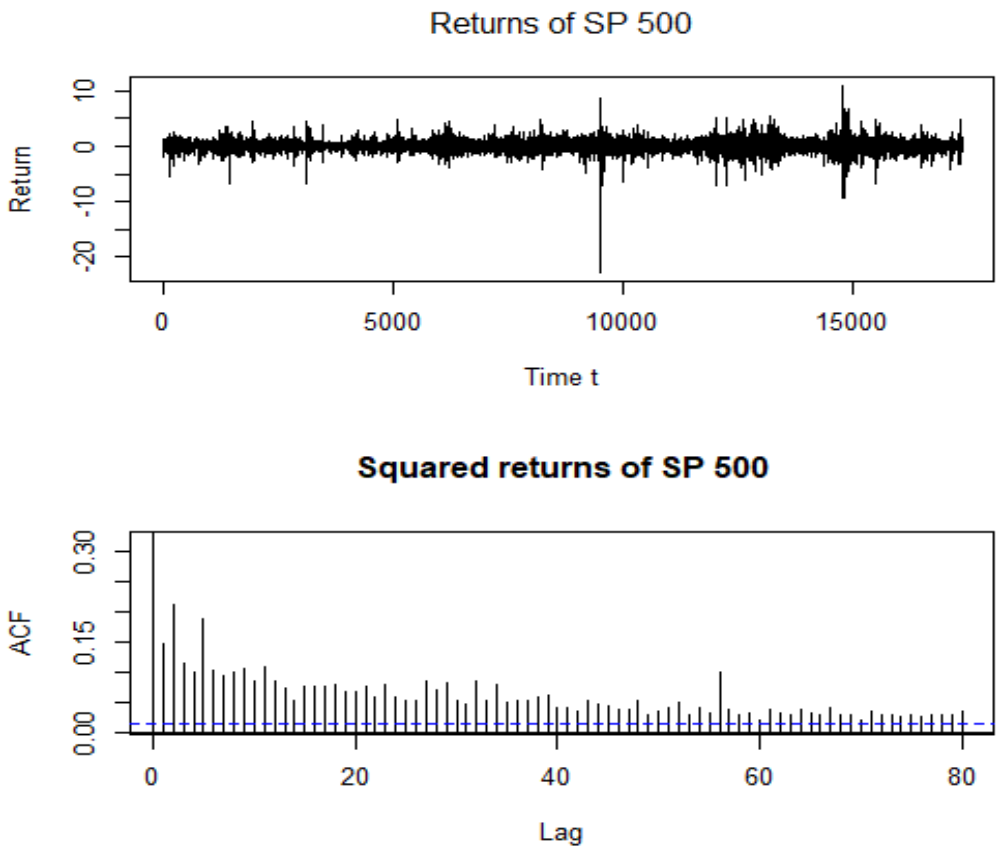


Figure 3.1 – Returns and the sample autocorrelations of squared returns of the S&P 500.

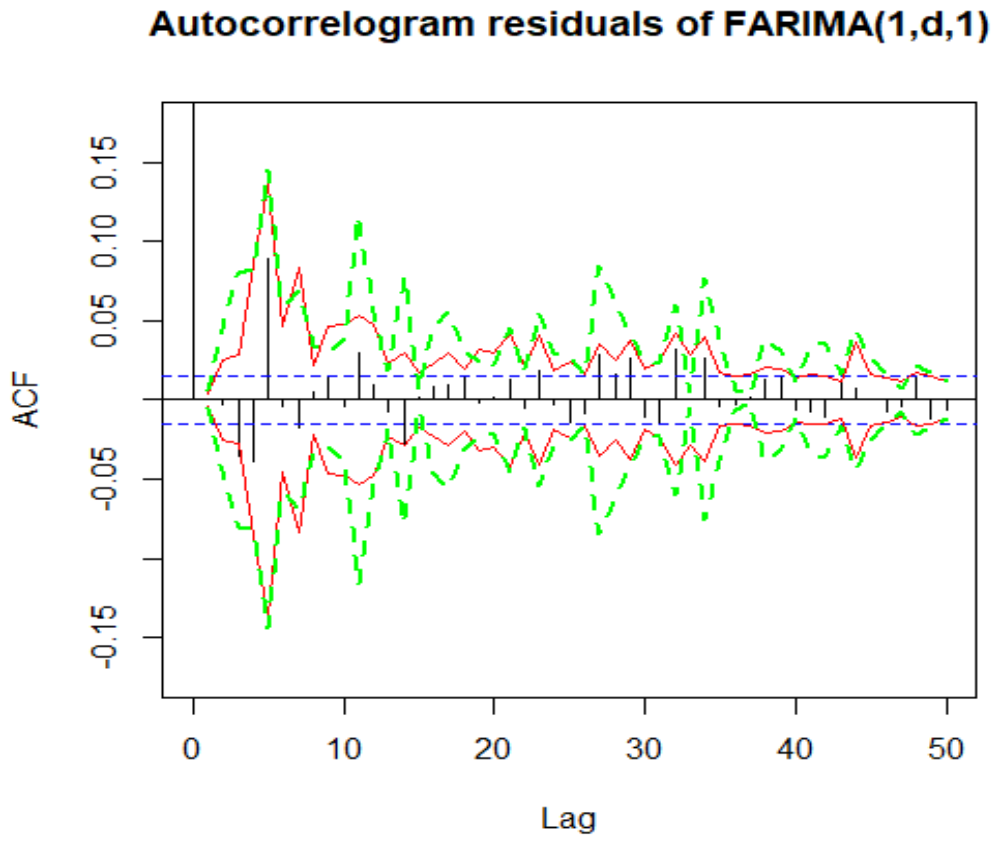


Figure 3.2 – Autocorrelation of the FARIMA$(1, 0.2338, 1)$ residuals for the squares of the S&P 500 returns. The horizontal dotted lines (blue color) correspond to the 5% significant limits obtained under the strong FARIMA assumption. The solid lines (red color) and dashed lines (green color) correspond also to the 5% significant limits under the weak FARIMA assumption. The full lines correspond to the asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.1. The dashed lines (green color) correspond to the self-normalized asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.5.

Table 3.1 – Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a strong FARIMA$(1, d_0, 1)$ defined by (3.34) with $\theta_0 = (0.9, 0.2, d_0)$. The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$.

d_0	Length n	Lag m	LB_{SN}	BP_{SN}	LB_W	BP_W	LB_S	BP_S
0.05	$n = 1,000$	1	5.8	5.7	**7.4**	**7.3**	n.a.	n.a.
		2	5.0	5.0	**7.4**	**7.3**	n.a.	n.a.
		3	4.3	4.3	5.8	5.8	n.a.	n.a.
		6	4.1	4.1	5.6	5.5	**10.9**	**10.9**
		12	5.1	4.6	4.7	4.5	**6.9**	**6.6**
		15	5.0	4.7	5.0	4.8	**6.9**	5.9
0.05	$n = 5,000$	1	6.0	6.0	**7.4**	**7.4**	n.a.	n.a.
		2	**6.5**	**6.5**	**7.9**	**7.9**	n.a.	n.a.
		3	4.7	4.7	**6.7**	**6.7**	n.a.	n.a.
		6	**3.5**	**3.5**	5.2	5.1	**11.0**	**10.9**
		12	5.3	5.3	5.8	5.8	**7.9**	**7.6**
		15	4.5	4.5	5.8	5.5	**7.0**	**6.9**
0.05	$n = 10,000$	1	4.2	4.2	6.1	6.1	n.a.	n.a.
		2	4.2	4.2	6.3	6.4	n.a.	n.a.
		3	3.8	3.8	5.9	5.9	n.a.	n.a.
		6	**3.5**	**3.5**	4.7	4.7	**10.4**	**10.4**
		12	4.2	4.2	6.1	6.1	**7.6**	**7.6**
		15	4.0	3.8	5.7	5.7	**7.4**	**7.4**
0.20	$n = 1,000$	1	5.8	5.8	**9.2**	**9.1**	n.a.	n.a.
		2	4.9	4.9	**7.5**	**7.5**	n.a.	n.a.
		3	4.6	4.5	5.9	5.9	n.a.	n.a.
		6	4.2	4.1	5.6	5.4	**10.3**	**10.2**
		12	5.4	4.9	4.7	4.4	6.4	5.9
		15	5.5	4.9	5.1	4.4	6.8	6.2
0.20	$n = 5,000$	1	6.4	6.4	6.1	6.2	n.a.	n.a.
		2	**6.8**	**6.8**	**6.9**	**6.9**	n.a.	n.a.
		3	4.3	4.3	5.9	5.8	n.a.	n.a.
		6	3.8	3.8	4.6	4.6	**10.0**	**10.0**
		12	5.2	5.2	5.7	5.6	**7.6**	**7.5**
		15	4.5	4.5	5.6	5.3	**6.8**	**6.7**
0.20	$n = 10,000$	1	4.5	4.5	5.5	5.5	n.a.	n.a.
		2	4.1	4.1	5.8	5.8	n.a.	n.a.
		3	**3.1**	**3.1**	5.3	5.3	n.a.	n.a.
		6	3.7	3.6	4.3	4.3	**10.1**	**10.1**
		12	3.8	3.8	6.1	6.1	**7.5**	**7.5**
		15	3.7	3.7	5.8	5.7	**7.0**	**6.9**
0.45	$n = 1,000$	1	4.3	4.3	**8.7**	**8.7**	n.a.	n.a.
		2	**3.0**	**3.0**	5.9	5.9	n.a.	n.a.
		3	3.7	3.7	4.4	4.4	n.a.	n.a.
		6	3.8	3.8	4.7	4.5	**8.1**	**7.8**
		12	5.1	4.6	4.3	4.2	5.1	4.9
		15	4.6	4.5	4.7	4.3	5.0	4.7
0.45	$n = 5,000$	1	5.6	5.5	6.0	6.0	n.a.	n.a.
		2	5.2	5.2	6.4	6.4	n.a.	n.a.
		3	4.0	4.0	5.9	5.9	n.a.	n.a.
		6	3.8	3.8	4.6	4.6	**10.1**	**9.9**
		12	5.2	5.2	5.4	5.4	**7.2**	**7.1**
		15	4.6	4.6	5.0	4.9	**6.7**	**6.6**
0.45	$n = 10,000$	1	4.3	4.3	5.3	5.3	n.a.	n.a.
		2	3.2	3.2	5.7	5.7	n.a.	n.a.
		3	**3.1**	**3.0**	5.4	5.4	n.a.	n.a.
		6	3.7	3.7	4.3	4.3	**9.8**	**9.8**
		12	4.3	4.3	5.8	5.8	**7.2**	**7.0**
		15	3.6	3.3	5.7	5.7	**6.8**	**6.8**

Table 3.2 – Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(1, d_0, 1)$ defined by (3.34) with $\theta_0 = (0.9, 0.2, d_0)$ and where $\omega = 0.4$, $\alpha_1 = 0.3$ and $\beta_1 = 0.3$ in (3.20). The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$.

d_0	Length n	Lag m	LB_{SN}	BP_{SN}	LB_W	BP_W	LB_S	BP_S
		1	4.9	4.9	**6.7**	**6.7**	n.a.	n.a.
		2	3.8	3.8	6.3	6.3	n.a.	n.a.
0.05	$n = 1,000$	3	**3.2**	**3.2**	5.2	5.2	n.a.	n.a.
		6	3.9	3.8	4.9	4.8	**18.5**	**18.3**
		12	**2.3**	**2.3**	4.1	4.0	**10.2**	**9.7**
		15	**2.7**	**2.3**	4.4	4.2	**9.7**	**9.3**
		1	5.1	5.1	5.6	5.6	n.a.	n.a.
		2	4.9	4.9	5.4	5.4	n.a.	n.a.
0.05	$n = 5,000$	3	**2.6**	**2.6**	5.0	5.0	n.a.	n.a.
		6	**3.5**	**3.5**	4.4	4.4	**19.6**	**19.6**
		12	**2.7**	**2.7**	**3.3**	**3.2**	**11.4**	**11.4**
		15	**3.4**	**3.4**	4.2	4.1	**10.8**	**10.7**
		1	4.8	4.8	**6.9**	**6.9**	n.a.	n.a.
		2	4.8	4.8	**6.7**	**6.7**	n.a.	n.a.
0.05	$n = 10,000$	3	4.7	4.7	5.5	5.5	n.a.	n.a.
		6	**3.3**	**3.3**	6.4	6.4	**20.2**	**20.2**
		12	4.2	4.2	6.3	6.3	**12.4**	**12.3**
		15	3.6	3.6	5.5	5.5	**11.6**	**11.6**
		1	5.3	5.3	**7.8**	**7.7**	n.a.	n.a.
		2	3.6	**3.4**	5.7	5.7	n.a.	n.a.
0.20	$n = 1,000$	3	**3.1**	**3.1**	4.9	4.8	n.a.	n.a.
		6	**3.3**	**3.2**	4.5	4.5	**17.6**	**17.4**
		12	**2.3**	**2.0**	4.1	4.1	**9.4**	**8.9**
		15	**2.4**	**2.1**	4.4	4.2	**9.0**	**8.1**
		1	4.6	4.6	4.3	4.3	n.a.	n.a.
		2	4.3	4.3	4.4	4.4	n.a.	n.a.
0.20	$n = 5,000$	3	**3.1**	**3.1**	4.4	4.3	n.a.	n.a.
		6	4.1	4.1	3.9	3.9	**19.0**	**19.0**
		12	**2.6**	**2.6**	**2.9**	**2.9**	**10.9**	**10.6**
		15	**3.4**	**3.3**	4.0	4.0	**10.0**	**9.9**
		1	4.8	4.8	5.1	5.1	n.a.	n.a.
		2	4.7	4.7	5.0	5.0	n.a.	n.a.
0.20	$n = 10,000$	3	4.5	4.5	4.8	4.8	n.a.	n.a.
		6	**3.5**	**3.5**	5.6	5.6	**19.1**	**19.1**
		12	4.1	4.1	5.9	5.9	**12.1**	**12.1**
		15	3.7	3.7	5.3	5.3	**11.3**	**11.3**
		1	4.4	4.4	**11.1**	**11.0**	n.a.	n.a.
		2	**3.4**	**3.4**	5.4	5.3	n.a.	n.a.
0.45	$n = 1,000$	3	**3.1**	**3.1**	4.9	4.9	n.a.	n.a.
		6	**3.1**	**2.9**	4.5	4.4	**15.3**	**15.1**
		12	**2.2**	**2.1**	4.0	4.0	**7.9**	**7.5**
		15	**2.1**	**2.0**	4.4	4.3	**7.0**	**6.5**
		1	3.9	3.9	4.2	4.2	n.a.	n.a.
		2	**3.4**	**3.4**	4.2	4.2	n.a.	n.a.
0.45	$n = 5,000$	3	**2.9**	**2.9**	4.4	4.4	n.a.	n.a.
		6	**3.5**	**3.5**	3.9	3.9	**18.4**	**18.4**
		12	**2.4**	**2.4**	**2.8**	**2.7**	**9.9**	**9.8**
		15	**3.2**	**3.2**	3.9	3.8	**9.2**	**9.2**
		1	4.6	4.6	5.3	5.3	n.a.	n.a.
		2	4.3	4.3	5.1	5.0	n.a.	n.a.
0.45	$n = 10,000$	3	**3.5**	**3.5**	5.0	5.0	n.a.	n.a.
		6	**2.8**	**2.8**	5.3	5.3	**19.3**	**19.3**
		12	4.2	4.2	5.5	5.5	**12.2**	**12.2**
		15	3.6	**3.5**	5.5	5.5	**11.4**	**11.4**

Table 3.3 – Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(1, d_0, 1)$ defined by (3.34)–(3.35) with $\theta_0 = (0.9, 0.2, d_0)$. The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$.

d_0	Length n	Lag m	LB_{SN}	BP_{SN}	LB_W	BP_W	LB_S	BP_S
0.05	$n = 1,000$	1	5.1	5.1	**7.3**	**7.3**	n.a.	n.a.
		2	3.6	3.6	**6.9**	**6.9**	n.a.	n.a.
		3	**2.9**	**2.9**	4.3	4.1	n.a.	n.a.
		6	**2.6**	**2.5**	**3.1**	**3.0**	**10.3**	**10.3**
		12	**0.9**	**0.9**	**1.2**	**1.1**	**8.7**	**8.3**
		15	**0.4**	**0.4**	**1.0**	**0.8**	**8.0**	**7.3**
0.05	$n = 5,000$	1	3.9	3.9	5.4	5.4	n.a.	n.a.
		2	3.9	3.9	5.9	5.9	n.a.	n.a.
		3	3.9	3.9	5.5	5.5	n.a.	n.a.
		6	**3.2**	**3.1**	3.8	3.8	**10.6**	**10.6**
		12	**2.4**	**2.4**	**3.5**	**3.4**	**8.3**	**8.2**
		15	**2.7**	**2.7**	**3.3**	**3.3**	**8.4**	**8.3**
0.05	$n = 10,000$	1	5.0	5.0	5.2	5.2	n.a.	n.a.
		2	4.9	4.9	4.5	4.5	n.a.	n.a.
		3	3.8	3.8	5.6	5.6	n.a.	n.a.
		6	3.6	3.6	4.5	4.5	**10.4**	**10.4**
		12	**3.3**	**3.3**	4.3	4.3	**8.5**	**8.4**
		15	4.7	4.7	3.8	3.8	**7.7**	**7.4**
0.20	$n = 1,000$	1	5.7	5.6	**10.1**	**10.0**	n.a.	n.a.
		2	**3.4**	**3.4**	5.5	5.5	n.a.	n.a.
		3	3.7	3.7	4.0	4.0	n.a.	n.a.
		6	**2.9**	**2.8**	**2.5**	**2.4**	**10.2**	**9.7**
		12	**0.9**	**0.9**	**1.1**	**1.1**	**7.9**	**7.2**
		15	**0.5**	**0.5**	**0.8**	**0.8**	**7.5**	**6.9**
0.20	$n = 5,000$	1	**3.5**	**3.5**	4.0	3.9	n.a.	n.a.
		2	3.7	3.7	4.3	4.3	n.a.	n.a.
		3	4.1	4.1	5.0	5.0	n.a.	n.a.
		6	**3.1**	**3.1**	**3.5**	**3.5**	**10.0**	**10.0**
		12	**2.8**	**2.8**	**3.3**	**3.3**	**8.2**	**8.2**
		15	**2.4**	**2.4**	**3.1**	**3.1**	**7.9**	**7.8**
0.20	$n = 10,000$	1	5.1	5.1	4.8	4.8	n.a.	n.a.
		2	4.7	4.7	4.2	4.2	n.a.	n.a.
		3	3.8	3.8	4.7	4.7	n.a.	n.a.
		6	3.8	3.8	4.1	4.1	**10.1**	**10.1**
		12	**3.4**	**3.4**	4.0	4.0	**8.0**	**8.0**
		15	4.8	4.8	3.6	3.6	**7.5**	**7.4**
0.45	$n = 1,000$	1	3.8	3.8	**12.1**	**12.0**	n.a.	n.a.
		2	**2.4**	**2.4**	4.4	4.4	n.a.	n.a.
		3	**2.7**	**2.6**	3.8	3.7	n.a.	n.a.
		6	**3.2**	**3.0**	**2.3**	**2.3**	**8.3**	**7.9**
		12	**1.1**	**0.9**	**1.0**	**0.9**	6.4	6.3
		15	**0.3**	**0.3**	**1.4**	**1.1**	**6.8**	6.4
0.45	$n = 5,000$	1	**3.1**	**3.1**	4.4	4.4	n.a.	n.a.
		2	**2.7**	**2.7**	4.5	4.5	n.a.	n.a.
		3	**3.2**	**3.2**	4.9	4.9	n.a.	n.a.
		6	**3.2**	**3.1**	**3.4**	**3.4**	**9.7**	**9.7**
		12	**3.3**	**3.3**	**3.3**	**3.3**	**7.3**	**7.3**
		15	**2.4**	**2.4**	**3.2**	**3.1**	**7.2**	**7.0**
0.45	$n = 10,000$	1	5.1	5.1	4.8	4.8	n.a.	n.a.
		2	4.9	4.9	4.3	4.3	n.a.	n.a.
		3	3.6	3.6	4.9	4.9	n.a.	n.a.
		6	**3.5**	**3.5**	4.3	4.2	**10.2**	**10.2**
		12	3.7	3.7	3.7	3.7	**7.7**	**7.6**
		15	4.8	4.8	3.9	3.9	**7.2**	**7.1**

Table 3.4 – Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a strong FARIMA$(0, d_0, 0)$ defined by (3.34) with $\theta_0 = (0, 0, d_0)$. The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$.

d_0	Length n	Lag m	LB_{SN}	BP_{SN}	LB_W	BP_W	LB_S	BP_S
0.05	$n = 1,000$	1	**3.3**	**3.3**	4.6	4.5	n.a.	n.a.
		2	4.5	4.5	4.9	4.9	5.8	5.8
		3	5.2	5.1	4.7	4.4	4.9	4.8
		6	5.8	5.8	4.6	4.5	5.1	5.0
		12	6.0	5.6	5.3	4.6	5.2	5.0
		15	5.6	5.2	4.7	4.3	5.3	4.7
0.05	$n = 5,000$	1	**6.8**	**6.8**	**6.6**	**6.6**	n.a.	n.a.
		2	**6.8**	**6.8**	6.4	6.4	**7.9**	**7.9**
		3	**6.6**	**6.6**	5.7	5.7	5.8	5.8
		6	**6.5**	6.4	5.6	5.6	5.7	5.6
		12	6.4	6.4	5.3	5.3	6.0	5.9
		15	6.1	6.0	4.7	4.6	5.3	5.2
0.05	$n = 10,000$	1	4.9	4.9	5.3	5.3	n.a.	n.a.
		2	5.4	5.4	6.6	6.6	7.8	7.8
		3	5.7	5.7	5.9	5.9	6.2	6.2
		6	5.9	5.8	4.5	4.5	4.6	4.6
		12	5.3	5.3	5.4	5.4	5.6	5.6
		15	4.4	4.3	4.8	4.8	4.9	4.9
0.20	$n = 1,000$	1	3.6	**3.5**	4.3	4.3	n.a.	n.a.
		2	4.7	4.7	4.7	4.7	5.8	5.7
		3	5.2	5.0	4.3	4.3	4.9	4.7
		6	6.0	5.9	4.7	4.5	5.0	4.9
		12	5.7	5.4	5.3	4.7	5.2	4.9
		15	5.9	5.6	4.8	4.2	5.2	4.8
0.20	$n = 5,000$	1	**6.6**	**6.6**	**6.5**	**6.5**	n.a.	n.a.
		2	**6.6**	**6.6**	6.4	6.4	**7.9**	**7.9**
		3	**6.7**	**6.7**	5.7	5.7	5.8	5.8
		6	6.3	6.3	5.6	5.6	5.7	5.5
		12	6.3	6.2	5.5	5.3	6.0	5.9
		15	6.1	5.9	4.7	4.6	5.3	5.2
0.20	$n = 10,000$	1	4.8	4.8	5.3	5.3	n.a.	n.a.
		2	5.4	5.4	**6.6**	**6.6**	**7.8**	**7.8**
		3	5.5	5.5	5.9	5.9	6.3	6.3
		6	5.8	5.8	4.5	4.5	4.6	4.6
		12	5.4	5.3	5.5	5.5	5.6	5.6
		15	4.4	4.3	4.7	4.7	4.9	4.9
0.45	$n = 1,000$	1	3.9	3.8	4.9	4.9	n.a.	n.a.
		2	5.1	5.0	4.8	4.6	5.9	5.9
		3	5.2	5.2	4.3	4.3	4.8	4.8
		6	6.2	6.0	4.7	4.3	4.9	4.9
		12	5.8	5.4	4.8	4.7	4.9	4.8
		15	5.6	5.5	4.5	4.2	5.0	4.8
0.45	$n = 5,000$	1	**6.6**	**6.6**	**6.6**	**6.6**	n.a.	n.a.
		2	**6.7**	**6.7**	**6.5**	**6.5**	**8.0**	**8.0**
		3	**6.6**	**6.6**	5.7	5.7	5.8	5.8
		6	6.3	6.3	5.4	5.4	5.6	5.5
		12	6.2	6.2	5.5	5.5	6.0	5.9
		15	6.2	5.9	4.6	4.6	5.5	5.3
0.45	$n = 10,000$	1	5.0	5.0	5.3	5.3	n.a.	n.a.
		2	5.4	5.4	**6.6**	**6.6**	**7.9**	**7.9**
		3	5.3	5.3	5.9	5.9	6.3	6.3
		6	5.8	5.8	4.7	4.6	4.7	4.7
		12	5.4	5.4	5.5	5.5	5.7	5.7
		15	4.6	4.5	4.9	4.8	4.9	4.9

Table 3.5 – Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(0, d_0, 0)$ defined by (3.34) with $\theta_0 = (0, 0, d_0)$ and where $\omega = 0.4$, $\alpha_1 = 0.3$ and $\beta_1 = 0.3$ in (3.20). The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$.

d_0	Length n	Lag m	LB_{SN}	BP_{SN}	LB_W	BP_W	LB_S	BP_S
0.05	$n = 1,000$	1	4.4	4.4	5.4	5.4	n.a.	n.a.
		2	4.3	4.2	5.7	5.7	**15.6**	**15.5**
		3	5.9	5.9	5.3	5.0	**14.2**	**14.0**
		6	5.2	5.1	6.0	6.0	**14.6**	**14.4**
		12	4.5	4.1	4.2	4.0	**11.0**	**10.7**
		15	4.0	3.9	4.2	3.9	**11.1**	**10.6**
0.05	$n = 5,000$	1	4.3	4.3	5.1	5.1	n.a.	n.a.
		2	4.4	4.4	5.8	5.8	**16.9**	**16.8**
		3	5.0	5.0	5.5	5.5	**16.5**	**16.5**
		6	5.6	5.6	4.5	4.5	**14.8**	**14.6**
		12	5.1	5.1	5.0	4.9	**12.6**	**12.5**
		15	5.2	5.1	4.9	4.7	**11.8**	**11.6**
0.05	$n = 10,000$	1	5.7	5.7	5.3	5.1	n.a.	n.a.
		2	5.0	5.0	4.5	4.5	**17.4**	**17.4**
		3	5.5	5.5	4.7	4.6	**17.2**	**17.2**
		6	5.3	5.3	5.0	5.0	**14.2**	**14.1**
		12	4.9	4.9	4.7	4.7	**11.0**	**11.0**
		15	4.9	4.8	4.7	4.6	**10.2**	**10.2**
0.20	$n = 1,000$	1	4.9	4.9	4.3	4.3	n.a.	n.a.
		2	4.0	4.0	5.7	5.6	**15.5**	**15.4**
		3	6.0	6.0	5.0	4.8	**14.0**	**13.8**
		6	5.2	5.1	5.7	5.6	**14.3**	**14.2**
		12	4.4	4.0	4.3	4.0	**10.8**	**10.5**
		15	3.9	3.8	4.2	3.9	**10.8**	**10.1**
0.20	$n = 5,000$	1	4.3	4.3	5.0	5.0	n.a.	n.a.
		2	4.3	4.3	5.9	5.8	**16.9**	**16.9**
		3	5.2	5.2	5.4	5.4	**16.7**	**16.7**
		6	5.6	5.5	4.6	4.5	**14.8**	**14.7**
		12	5.2	5.2	5.0	4.9	**12.5**	**12.4**
		15	5.2	5.2	4.8	4.6	11.7	11.7
0.20	$n = 10,000$	1	5.7	5.7	5.2	5.2	n.a.	n.a.
		2	5.1	5.1	4.5	4.5	**17.3**	**17.3**
		3	5.7	5.6	4.7	4.7	**17.2**	**17.2**
		6	5.1	5.1	4.9	4.9	**14.2**	**14.2**
		12	4.8	4.8	4.7	4.7	**11.0**	**11.0**
		15	4.9	4.7	4.6	4.6	**10.2**	**10.2**
0.45	$n = 1,000$	1	4.5	4.5	5.4	5.4	n.a.	n.a.
		2	4.1	4.1	6.0	6.0	**16.2**	**16.1**
		3	5.9	5.7	5.3	5.3	**14.6**	**14.5**
		6	5.2	4.8	5.5	5.4	**14.4**	**14.1**
		12	4.0	3.7	4.2	4.2	**11.2**	**10.8**
		15	3.8	3.7	4.3	3.9	**10.6**	**10.4**
0.45	$n = 5,000$	1	4.6	4.6	5.0	5.0	n.a.	n.a.
		2	4.3	4.3	5.9	5.9	**16.7**	**16.7**
		3	4.9	4.9	5.4	5.4	**16.8**	**16.7**
		6	5.7	5.6	4.6	4.6	**15.1**	**14.9**
		12	5.3	5.3	5.1	5.1	**12.7**	**12.4**
		15	5.1	5.0	4.8	4.8	**11.7**	**11.7**
0.45	$n = 10,000$	1	5.7	5.7	5.2	5.2	n.a.	n.a.
		2	5.0	5.0	4.7	4.7	**17.2**	**17.2**
		3	5.8	5.7	4.7	4.7	**17.5**	**17.4**
		6	5.1	5.1	5.0	4.9	**14.3**	**14.3**
		12	4.8	4.8	4.7	4.7	**10.9**	**10.9**
		15	4.9	4.7	4.6	4.6	**10.2**	**10.2**

Table 3.6 – Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of weak FARIMA$(0, d_0, 0)$ defined by (3.34)–(3.35) with $\theta_0 = (0, 0, d_0)$. The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$.

d_0	Length n	Lag m	LB_{SN}	BP_{SN}	LB_W	BP_W	LB_S	BP_S
0.05	$n = 1,000$	1	**3.3**	**3.3**	**8.7**	**8.6**	n.a.	n.a.
		2	3.8	3.7	6.1	6.1	**16.9**	**16.9**
		3	**3.5**	**3.5**	4.8	4.7	**14.8**	**14.8**
		6	**3.3**	**3.2**	4.0	4.0	**14.1**	**14.0**
		12	**1.0**	**0.9**	**2.5**	**2.4**	**13.0**	**12.8**
		15	**1.0**	**0.9**	**2.3**	**2.1**	**12.8**	**12.2**
0.05	$n = 5,000$	1	3.9	3.9	5.3	5.3	n.a.	n.a.
		2	4.8	4.8	5.2	5.2	**18.7**	**18.7**
		3	5.6	5.6	5.3	5.3	**15.1**	**15.0**
		6	4.8	4.8	4.3	4.3	**12.4**	**12.4**
		12	3.9	3.9	**3.3**	**3.3**	**11.2**	**11.1**
		15	**3.5**	**3.5**	**2.7**	**2.7**	**10.2**	**10.1**
0.05	$n = 10,000$	1	5.4	5.4	5.2	5.2	n.a.	n.a.
		2	5.6	5.6	5.3	5.3	**18.6**	**18.6**
		3	4.9	4.9	5.3	5.2	**16.6**	**16.5**
		6	4.8	4.8	5.5	5.4	**13.3**	**13.3**
		12	4.1	4.0	4.0	4.0	**12.2**	**12.2**
		15	5.0	5.0	**3.5**	**3.5**	**11.2**	**11.2**
0.20	$n = 1,000$	1	**3.3**	**3.3**	4.9	4.9	n.a.	n.a.
		2	4.2	4.1	4.4	4.3	**14.7**	**14.7**
		3	3.7	3.7	**3.4**	**3.2**	**12.8**	**12.8**
		6	3.6	**3.4**	**2.7**	**2.7**	**12.9**	**12.8**
		12	**1.1**	**1.0**	**1.9**	**1.7**	**11.8**	**11.3**
		15	**0.9**	**0.6**	**1.8**	**1.7**	**12.0**	**11.5**
0.20	$n = 5,000$	1	3.8	3.8	5.5	5.5	n.a.	n.a.
		2	4.7	4.7	5.1	5.1	**18.8**	**18.8**
		3	5.8	5.8	5.2	5.2	**15.0**	**15.0**
		6	4.9	4.9	4.3	4.3	**12.5**	**12.4**
		12	3.9	3.9	**3.4**	**3.4**	**11.1**	**11.1**
		15	**3.5**	**3.3**	**2.7**	**2.7**	**10.2**	**10.1**
0.20	$n = 10,000$	1	5.4	5.4	5.1	5.1	n.a.	n.a.
		2	5.6	5.6	5.3	5.3	**18.8**	**18.8**
		3	5.0	5.0	5.2	5.2	**16.6**	**16.6**
		6	4.8	4.8	5.4	5.4	**13.3**	**13.3**
		12	4.0	4.0	4.0	4.0	**12.1**	**12.1**
		15	5.3	5.3	**3.4**	**3.4**	**11.2**	**11.2**
0.45	$n = 1,000$	1	**3.5**	**3.5**	**9.0**	**9.0**	n.a.	n.a.
		2	4.1	4.1	5.9	5.9	17.5	17.5
		3	3.9	3.7	5.0	4.8	**15.0**	**14.6**
		6	**3.4**	**3.4**	3.7	3.7	**14.1**	**13.9**
		12	**0.9**	**0.9**	**2.0**	**2.0**	**12.9**	**12.2**
		15	**1.0**	**0.5**	**1.9**	**1.7**	**13.1**	**12.8**
0.45	$n = 5,000$	1	4.1	4.1	5.4	5.4	n.a.	n.a.
		2	4.6	4.6	5.2	5.2	**18.8**	**18.7**
		3	5.6	5.6	5.2	5.2	**15.2**	**15.2**
		6	5.1	5.0	4.4	4.4	**12.5**	**12.4**
		12	4.0	3.8	**3.5**	**3.5**	**11.1**	**11.1**
		15	**3.5**	**3.5**	**2.6**	**2.6**	**10.0**	**9.9**
0.45	$n = 10,000$	1	5.5	5.5	5.1	5.1	n.a.	n.a.
		2	5.6	5.6	5.3	5.3	**18.7**	**18.6**
		3	4.7	4.7	5.2	5.2	**16.6**	**16.6**
		6	4.8	4.8	5.3	5.3	**13.3**	**13.3**
		12	4.0	4.0	4.0	4.0	**12.1**	**12.1**
		15	5.2	5.2	**3.5**	**3.5**	**11.1**	**11.1**

Table 3.7 – Empirical power (in %) of the modified and standard versions of the LB and BP tests in the case of a strong FARIMA$(1, d_0, 2)$ defined by (3.36) with $\theta_0 = (0.9, 1, -0.2, d_0)$. The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$.

d_0	Length n	Lag m	LB_{SN}	BP_{SN}	LB_W	BP_W	LB_S	BP_S
0.05	$n = 1,000$	1	7.1	7.1	15.0	15.0	n.a.	n.a.
		2	7.4	7.4	15.4	15.3	n.a.	n.a.
		3	9.7	9.5	8.8	8.8	n.a.	n.a.
		6	10.9	10.6	9.7	8.8	13.0	12.7
		12	11.2	10.8	6.9	6.4	8.1	7.9
		15	11.8	10.6	4.4	3.9	7.3	7.1
0.05	$n = 5,000$	1	24.5	24.5	37.9	37.9	n.a.	n.a.
		2	28.8	28.8	46.1	46.1	n.a.	n.a.
		3	36.7	36.7	22.2	22.1	n.a.	n.a.
		6	55.7	55.7	40.6	40.3	47.6	47.6
		12	54.9	54.7	27.2	27.2	28.3	28.0
		15	54.0	53.6	18.0	17.8	27.9	27.7
0.05	$n = 10,000$	1	44.9	44.9	62.8	62.7	n.a.	n.a.
		2	51.4	51.3	76.1	76.0	n.a.	n.a.
		3	62.8	62.8	39.9	39.9	n.a.	n.a.
		6	86.5	86.5	80.9	80.8	84.7	84.7
		12	85.8	85.8	64.9	64.8	66.4	66.2
		15	82.0	82.0	43.2	43.2	60.8	60.8
0.20	$n = 1,000$	1	4.7	4.7	26.2	26.2	n.a.	n.a.
		2	7.4	7.3	30.3	30.2	n.a.	n.a.
		3	9.2	9.1	15.7	15.7	n.a.	n.a.
		6	9.3	9.2	23.2	22.4	27.2	26.8
		12	11.5	11.0	13.0	12.4	18.2	17.9
		15	10.8	10.5	3.5	3.3	16.0	15.7
0.20	$n = 5,000$	1	14.0	14.0	58.0	57.9	n.a.	n.a.
		2	22.2	22.2	71.1	71.1	n.a.	n.a.
		3	24.1	23.8	40.7	40.7	n.a.	n.a.
		6	32.1	32.0	74.4	74.4	78.5	78.5
		12	52.3	52.2	62.4	62.2	67.7	67.6
		15	51.6	51.3	14.1	14.0	62.1	61.7
0.20	$n = 10,000$	1	21.4	21.4	84.9	85.0	n.a.	n.a.
		2	30.6	30.6	93.1	93.1	n.a.	n.a.
		3	35.6	35.6	65.9	65.7	n.a.	n.a.
		6	44.1	44.1	96.9	96.9	97.8	97.8
		12	76.3	76.2	93.2	93.2	94.3	94.3
		15	73.7	73.7	43.9	43.9	91.6	91.6
0.45	$n = 1,000$	1	0.2	0.2	67.0	66.8	n.a.	n.a.
		2	9.3	9.0	94.9	94.9	n.a.	n.a.
		3	18.4	18.3	99.7	99.7	n.a.	n.a.
		6	21.6	21.3	99.6	99.6	99.7	99.7
		12	22.2	21.2	99.4	99.4	99.4	99.4
		15	23.4	21.8	99.2	99.2	99.3	99.3
0.45	$n = 5,000$	1	0.0	0.	100.0	100.0	n.a.	n.a.
		2	49.1	49.1	100.0	100.0	n.a.	n.a.
		3	69.0	69.0	100.0	100.0	n.a.	n.a.
		6	76.7	76.6	100.0	100.0	100.0	100.0
		12	86.8	86.7	100.0	100.0	100.0	100.0
		15	90.9	90.7	100.0	100.0	100.0	100.0
0.45	$n = 10,000$	1	0.0	0.0	100.0	100.0	n.a.	n.a.
		2	77.9	77.9	100.0	100.0	n.a.	n.a.
		3	90.3	90.2	100.0	100.0	n.a.	n.a.
		6	94.2	94.2	100.0	100.0	100.0	100.0
		12	98.9	98.9	100.0	100.0	100.0	100.0
		15	99.5	99.4	100.0	100.0	100.0	100.0

Table 3.8 – Empirical power (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(1, d_0, 2)$ defined by (3.36) with $\theta_0 = (0.9, 1, -0.2, d_0)$ and where $\omega = 0.4$, $\alpha_1 = 0.3$ and $\beta_1 = 0.3$ in (3.20). The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$.

d_0	Length n	Lag m	LB_{SN}	BP_{SN}	LB_W	BP_W	LB_S	BP_S
		1	6.2	6.2	13.1	13.3	n.a.	n.a.
		2	6.9	6.9	14.3	14.0	n.a.	n.a.
0.05	$n = 1,000$	3	9.0	9.0	7.7	7.6	n.a.	n.a.
		6	10.1	9.9	7.4	7.3	18.3	18.2
		12	8.3	8.2	5.4	5.1	9.1	8.6
		15	6.9	6.0	4.6	4.2	9.8	9.1
		1	22.5	22.5	32.8	32.7	n.a.	n.a.
		2	27.3	27.3	41.7	41.8	n.a.	n.a.
0.05	$n = 5,000$	3	32.4	32.3	20.1	20.0	n.a.	n.a.
		6	52.1	52.0	34.0	34.0	55.8	55.7
		12	54.1	54.1	23.5	23.5	34.2	34.1
		15	53.9	53.4	17.1	16.9	31.9	31.8
		1	36.1	36.1	53.2	53.2	n.a.	n.a.
		2	44.9	44.9	64.5	64.5	n.a.	n.a.
0.05	$n = 10,000$	3	56.5	56.5	33.1	33.1	n.a.	n.a.
		6	83.1	83.1	71.2	71.2	86.4	86.2
		12	84.0	83.9	59.0	59.0	70.4	70.2
		15	80.6	80.5	40.1	40.1	67.4	67.2
		1	4.8	4.8	25.1	24.6	n.a.	n.a.
		2	8.1	7.8	25.9	25.8	n.a.	n.a.
0.20	$n = 1,000$	3	8.1	8.1	14.9	14.5	n.a.	n.a.
		6	8.9	8.7	19.6	19.3	32.8	32.3
		12	8.5	7.9	11.9	11.7	20.4	19.8
		15	6.8	5.7	4.4	4.2	17.9	17.8
		1	14.6	14.5	51.0	50.9	n.a.	n.a.
		2	21.8	21.8	67.1	67.1	n.a.	n.a.
0.20	$n = 5,000$	3	22.4	22.3	37.7	37.7	n.a.	n.a.
		6	32.3	32.3	68.3	68.3	81.9	81.9
		12	51.6	51.5	55.9	55.8	68.7	68.5
		15	51.7	51.6	14.2	14.1	64.8	64.6
		1	22.8	22.8	74.1	74.0	n.a.	n.a.
		2	29.6	29.6	86.2	86.2	n.a.	n.a.
0.20	$n = 10,000$	3	32.9	32.9	56.6	56.5	n.a.	n.a.
		6	43.1	43.1	92.3	92.3	97.1	97.1
		12	72.9	72.8	88.3	88.3	93.8	93.8
		15	71.2	71.1	39.1	38.9	92.0	92.0
		1	0.5	0.5	66.1	66.0	n.a.	n.a.
		2	9.0	9.0	92.4	92.4	n.a.	n.a.
0.45	$n = 1,000$	3	13.6	13.5	98.1	98.1	n.a.	n.a.
		6	18.3	18.3	99.2	99.2	99.6	99.6
		12	16.8	16.2	97.8	97.8	99.3	99.3
		15	15.6	14.5	97.6	97.4	99.2	99.2
		1	0.0	0.0	99.8	99.8	n.a.	n.a.
		2	40.1	40.1	100.0	100.0	n.a.	n.a.
0.45	$n = 5,000$	3	57.9	57.9	100.0	100.0	n.a.	n.a.
		6	65.7	65.7	100.0	100.0	100.0	100.0
		12	78.8	78.5	100.0	100.0	100.0	100.0
		15	84.7	84.6	100.0	100.0	100.0	100.0
		1	0.0	0.0	99.9	99.9	n.a.	n.a.
		2	72.2	72.2	100.0	99.9	n.a.	n.a.
0.45	$n = 10,000$	3	84.8	84.8	100.0	100.0	n.a.	n.a.
		6	89.8	89.7	100.0	100.0	100.0	100.0
		12	97.7	97.7	100.0	100.0	100.0	100.0
		15	99.0	99.0	100.0	100.0	100.0	100.0

Table 3.9 – Empirical power (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(1, d_0, 2)$ defined by (3.36)–(3.35) with $\theta_0 = (0.9, 1, -0.2, d_0)$. The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$.

d_0	Length n	Lag m	LB_{SN}	BP_{SN}	LB_W	BP_W	LB_S	BP_S
0.05	$n = 1,000$	1	10.1	10.1	18.2	18.2	n.a.	n.a.
		2	8.1	8.1	16.8	16.8	n.a.	n.a.
		3	8.6	8.6	10.1	9.9	n.a.	n.a.
		6	9.6	9.2	8.0	7.8	11.4	11.3
		12	4.7	4.4	3.9	3.7	9.1	8.8
		15	2.6	2.4	1.7	1.6	8.8	8.5
0.05	$n = 5,000$	1	27.6	27.6	42.6	42.7	n.a.	n.a.
		2	32.7	32.6	51.4	51.3	n.a.	n.a.
		3	36.9	36.9	23.7	23.7	n.a.	n.a.
		6	53.3	53.0	39.7	39.7	46.0	45.9
		12	49.6	49.3	23.7	23.7	29.3	29.2
		15	44.4	44.2	17.5	17.4	28.5	28.1
0.05	$n = 10,000$	1	48.5	48.5	68.3	68.3	n.a.	n.a.
		2	58.7	58.6	76.6	76.5	n.a.	n.a.
		3	66.8	66.8	42.5	42.5	n.a.	n.a.
		6	84.2	84.0	77.0	76.9	83.2	83.2
		12	79.9	79.9	62.7	62.6	66.0	66.0
		15	75.8	75.8	40.5	40.5	61.4	61.3
0.20	$n = 1,000$	1	5.1	5.1	30.1	30.3	n.a.	n.a.
		2	8.0	8.0	33.8	33.7	n.a.	n.a.
		3	7.9	7.9	18.1	18.1	n.a.	n.a.
		6	7.4	7.2	23.4	22.9	25.4	25.3
		12	4.7	4.4	9.5	9.0	17.8	17.3
		15	2.9	2.5	2.5	2.3	16.3	15.6
0.20	$n = 5,000$	1	15.3	15.3	62.4	62.5	n.a.	n.a.
		2	23.5	23.4	74.6	74.6	n.a.	n.a.
		3	25.9	25.9	45.3	45.2	n.a.	n.a.
		6	34.0	34.0	73.1	72.9	78.5	78.4
		12	51.3	50.8	56.8	56.6	64.5	64.4
		15	46.3	45.8	15.0	14.9	60.1	60.1
0.20	$n = 10,000$	1	23.0	23.0	85.2	85.2	n.a.	n.a.
		2	33.8	33.8	93.6	93.6	n.a.	n.a.
		3	36.5	36.5	68.3	68.3	n.a.	n.a.
		6	46.8	46.7	95.4	95.4	97.1	97.1
		12	81.7	81.7	90.8	90.8	93.7	93.6
		15	79.0	78.7	44.2	44.0	91.7	91.7
0.45	$n = 1,000$	1	0.3	0.3	65.2	65.3	n.a.	n.a.
		2	9.4	9.2	90.2	90.2	n.a.	n.a.
		3	15.6	15.6	95.1	95.1	n.a.	n.a.
		6	16.5	16.0	94.8	94.8	96.4	96.4
		12	9.7	9.2	94.7	94.7	96.4	96.4
		15	12.5	12.0	93.0	93.0	96.0	96.0
0.45	$n = 5,000$	1	0.0	0.0	99.9	99.9	n.a.	n.a.
		2	51.9	51.9	100.0	100.0	n.a.	n.a.
		3	66.7	66.7	100.0	100.0	n.a.	n.a.
		6	73.6	73.6	100.0	100.0	100.0	100.0
		12	83.1	83.0	100.0	100.0	100.0	100.0
		15	85.5	85.4	100.0	100.0	100.0	100.0
0.45	$n = 10,000$	1	0.0	0.0	100.0	99.9	n.a.	n.a.
		2	79.2	79.2	100.0	100.0	n.a.	n.a.
		3	90.8	90.8	100.0	100.0	n.a.	n.a.
		6	93.6	93.6	100.0	100.0	100.0	100.0
		12	97.8	97.8	100.0	100.0	100.0	100.0
		15	99.1	99.1	100.0	100.0	100.0	100.0

Table 3.10 – Empirical power (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(1, d_0, 2)$ defined by (3.36) with $\theta_0 = (0., 0.2, 0., d_0)$ and where $\omega = 0.4$, $\alpha_1 = 0.3$ and $\beta_1 = 0.3$ in (3.20). The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$.

d_0	Length n	Lag m	LB_{SN}	BP_{SN}	LB_W	BP_W	LB_S	BP_S
0.05	$n = 1,000$	1	2.9	2.9	98.8	98.8	n.a.	n.a.
		2	13.3	13.2	95.0	95.0	98.1	98.1
		3	18.9	18.7	92.5	92.4	97.6	97.6
		6	25.6	25.1	85.9	85.9	95.7	95.6
		12	20.8	19.9	78.8	78.4	90.8	90.6
		15	19.6	19.1	74.5	74.2	87.7	87.1
0.05	$n = 5,000$	1	0.1	0.1	100.0	100.0	n.a.	n.a.
		2	55.7	55.7	100.0	100.0	100.0	100.0
		3	75.7	75.7	100.0	100.0	100.0	100.0
		6	87.1	87.1	100.0	100.0	100.0	100.0
		12	87.0	86.8	100.0	100.0	100.0	100.0
		15	87.3	87.2	100.0	100.0	100.0	100.0
0.05	$n = 10,000$	1	0.0	0.0	100.0	100.0	n.a.	n.a.
		2	79.5	79.4	100.0	100.0	100.0	100.0
		3	95.2	95.2	100.0	100.0	100.0	100.0
		6	98.0	98.0	100.0	100.0	100.0	100.0
		12	98.6	98.6	100.0	100.0	100.0	100.0
		15	99.0	99.0	100.0	100.0	100.0	100.0
0.20	$n = 1,000$	1	58.4	58.3	82.7	82.5	n.a.	n.a.
		2	47.7	47.6	56.2	56.1	78.7	78.6
		3	40.1	39.9	53.9	53.7	75.6	75.5
		6	32.2	31.2	49.0	48.7	70.3	69.8
		12	23.0	22.3	43.8	43.0	63.0	62.1
		15	21.0	20.0	40.4	39.9	58.0	57.6
0.20	$n = 5,000$	1	98.2	98.2	99.9	99.9	n.a.	n.a.
		2	94.6	94.6	99.5	99.5	100.0	100.0
		3	92.3	92.3	99.6	99.6	100.0	100.0
		6	91.0	91.0	99.6	99.6	100.0	100.0
		12	88.8	88.7	99.8	99.8	100.0	100.0
		15	88.6	88.6	99.8	99.8	100.0	100.0
0.20	$n = 10,000$	1	99.7	99.7	100.0	100.0	n.a.	n.a.
		2	99.2	99.2	100.0	100.0	100.0	100.0
		3	99.3	99.2	100.0	100.0	100.0	100.0
		6	98.8	98.8	100.0	100.0	100.0	100.0
		12	99.3	99.3	100.0	100.0	100.0	100.0
		15	99.3	99.3	100.0	100.0	100.0	100.0
0.45	$n = 1,000$	1	59.0	59.0	82.4	82.4	n.a.	n.a.
		2	47.7	47.7	56.7	56.4	78.7	78.6
		3	41.0	41.0	54.6	54.2	76.4	76.4
		6	33.1	32.9	49.9	49.7	70.5	70.4
		12	24.1	23.1	44.3	43.9	63.5	63.3
		15	21.9	20.6	41.0	40.7	58.2	58.0
0.45	$n = 5,000$	1	98.2	98.2	99.8	99.8	n.a.	n.a.
		2	94.4	94.3	99.5	99.5	100.0	100.0
		3	92.4	92.4	99.6	99.6	100.0	100.0
		6	90.9	90.8	99.6	99.6	100.0	100.0
		12	88.9	88.9	99.8	99.8	100.0	100.0
		15	88.8	88.5	99.8	99.8	100.0	100.0
0.45	$n = 10,000$	1	99.7	99.7	100.0	100.0	n.a.	n.a.
		2	99.0	99.0	100.0	100.0	100.0	100.0
		3	99.2	99.2	100.0	100.0	100.0	100.0
		6	98.9	98.9	100.0	100.0	100.0	100.0
		12	99.3	99.3	100.0	100.0	100.0	100.0
		15	99.3	99.3	100.0	100.0	100.0	100.0

Table 3.11 – Empirical power (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(1, d_0, 2)$ defined by (3.36)–(3.35) with $\theta_0 = (0., 0.2, 0., d_0)$. The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$.

d_0	Length n	Lag m	LB_{SN}	BP_{SN}	LB_W	BP_W	LB_S	BP_S
0.05	$n = 1,000$	1	2.7	2.7	91.7	91.7	n.a.	n.a.
		2	11.1	11.0	85.5	85.4	93.4	93.3
		3	15.7	15.6	82.7	82.7	91.2	91.2
		6	15.8	15.7	77.1	77.0	87.1	87.1
		12	6.5	5.8	65.8	65.7	80.5	80.4
		15	3.9	3.5	62.3	61.8	78.6	78.4
0.05	$n = 5,000$	1	0.3	0.3	99.9	99.9	n.a.	n.a.
		2	56.7	56.6	99.9	99.9	99.9	99.9
		3	69.1	69.1	99.9	99.9	99.9	99.9
		6	75.9	75.9	99.9	99.9	99.9	99.9
		12	71.9	71.4	99.9	99.9	99.9	99.9
		15	68.5	68.0	99.9	99.9	99.9	99.9
0.05	$n = 10,000$	1	0.0	0.0	100.0	100.0	n.a.	n.a.
		2	81.8	81.8	100.0	100.0	100.0	100.0
		3	90.3	90.3	100.0	100.0	100.0	100.0
		6	93.9	93.9	100.0	100.0	100.0	100.0
		12	93.8	93.8	100.0	100.0	100.0	100.0
		15	93.7	93.7	100.0	100.0	100.0	100.0
0.20	$n = 1,000$	1	43.3	43.2	64.5	64.5	n.a.	n.a.
		2	31.3	31.1	53.2	53.1	75.1	75.1
		3	26.6	26.5	50.1	49.8	71.2	71.2
		6	18.6	18.4	43.1	42.8	66.5	66.4
		12	7.9	7.7	32.9	32.3	55.1	54.4
		15	5.2	4.6	29.0	28.4	51.8	51.3
0.20	$n = 5,000$	1	92.3	92.3	99.9	99.9	n.a.	n.a.
		2	86.1	86.0	98.6	98.6	99.8	99.8
		3	82.3	82.3	99.2	99.1	99.8	99.8
		6	80.0	80.0	98.9	98.9	99.9	99.9
		12	73.1	72.8	98.7	98.7	99.6	99.6
		15	68.3	68.0	98.4	98.4	99.5	99.5
0.20	$n = 10,000$	1	99.2	99.2	100.0	100.0	n.a.	n.a.
		2	96.4	96.4	100.0	100.0	100.0	100.0
		3	94.6	94.6	100.0	100.0	100.0	100.0
		6	95.1	95.1	100.0	100.0	100.0	100.0
		12	95.2	95.2	100.0	100.0	100.0	100.0
		15	94.0	94.0	100.0	100.0	100.0	100.0
0.45	$n = 1,000$	1	43.6	43.2	64.6	64.6	n.a.	n.a.
		2	31.9	31.8	53.4	53.1	75.7	75.6
		3	27.1	27.0	50.2	50.0	71.8	71.6
		6	18.8	18.2	43.4	43.3	67.1	67.0
		12	8.7	8.1	33.7	33.2	55.6	54.8
		15	5.5	4.7	29.1	28.5	52.8	52.0
0.45	$n = 5,000$	1	92.4	92.4	99.9	99.9	n.a.	n.a.
		2	85.6	85.6	98.6	98.6	99.8	99.8
		3	82.1	82.0	99.3	99.3	99.8	99.8
		6	80.3	80.3	98.9	98.9	99.9	99.9
		12	73.0	72.7	98.7	98.7	99.6	99.6
		15	68.2	68.1	98.4	98.4	99.5	99.5
0.45	$n = 10,000$	1	99.2	99.2	100.0	100.0	n.a.	n.a.
		2	96.4	96.4	100.0	100.0	100.0	100.0
		3	94.8	94.8	100.0	100.0	100.0	100.0
		6	95.2	95.2	100.0	100.0	100.0	100.0
		12	95.0	95.0	100.0	100.0	100.0	100.0
		15	94.0	94.0	100.0	100.0	100.0	100.0

Table 3.12 – Modified and standard versions of portmanteau tests to check the null hypothesis that the S&P 500 squared returns follow a FARIMA$(1, 0.2338, 1)$ model (3.37).

Lag m	1	2	3	4	5	6	7
$\hat{\rho}(m)$	0.0002	-0.0033	-0.0350	-0.0393	0.0893	-0.0040	-0.0179
$\mathrm{LB}_{\mathrm{SN}}$	0.0653	18.150	41.924	58.057	186.72	313.78	341.38
$\mathrm{BP}_{\mathrm{SN}}$	0.0653	18.146	41.912	58.037	186.64	313.64	341.20
LB_{W}	0.0008	0.1885	21.445	48.248	186.95	187.23	192.77
BP_{W}	0.0008	0.1884	21.439	48.232	186.88	187.15	192.67
$\mathrm{p}_{\mathrm{W}}^{\mathrm{LB}}$	0.8525	0.6985	0.0916	0.3137	0.0678	0.0717	0.0752
$\mathrm{p}_{\mathrm{W}}^{\mathrm{BP}}$	0.8525	0.6986	0.0917	0.3138	0.0679	0.0718	0.0753
$\mathrm{p}_{\mathrm{S}}^{\mathrm{LB}}$	n.a.	n.a.	n.a.	0.0000	0.0000	0.0000	0.0000
$\mathrm{p}_{\mathrm{S}}^{\mathrm{BP}}$	n.a.	n.a.	n.a.	0.0000	0.0000	0.0000	0.0000
Lag m	8	9	10	11	12	13	14
$\hat{\rho}(m)$	0.0047	0.0137	-0.0040	0.0295	0.0093	-0.0077	-0.0286
$\mathrm{LB}_{\mathrm{SN}}$	397.27	397.38	415.22	465.52	468.76	567.87	573.02
$\mathrm{BP}_{\mathrm{SN}}$	397.04	397.13	414.93	465.17	468.33	567.38	572.49
LB_{W}	193.16	196.42	196.69	211.82	213.31	214.34	228.55
BP_{W}	193.09	196.34	196.61	211.74	213.22	214.25	228.45
$\mathrm{p}_{\mathrm{W}}^{\mathrm{LB}}$	0.0758	0.0786	0.0986	0.1053	0.1148	0.1226	0.1047
$\mathrm{p}_{\mathrm{W}}^{\mathrm{BP}}$	0.0758	0.0787	0.0987	0.1054	0.1150	0.1228	0.1048
$\mathrm{p}_{\mathrm{S}}^{\mathrm{LB}}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\mathrm{p}_{\mathrm{S}}^{\mathrm{BP}}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Lag m	15	16	17	18	19	20	21
$\hat{\rho}(m)$	0.0021	0.0086	0.0097	0.0137	-0.0023	0.0016	0.0132
$\mathrm{LB}_{\mathrm{SN}}$	588.61	701.16	738.23	738.58	749.24	778.88	788.01
$\mathrm{BP}_{\mathrm{SN}}$	588.04	700.44	737.42	737.73	748.33	777.90	786.97
LB_{W}	228.63	229.91	231.54	234.83	234.92	234.97	238.00
BP_{W}	228.52	229.80	231.44	234.72	234.81	234.86	237.89
$\mathrm{p}_{\mathrm{W}}^{\mathrm{LB}}$	0.1079	0.1113	0.2212	0.2138	0.2127	0.2169	0.2324
$\mathrm{p}_{\mathrm{W}}^{\mathrm{BP}}$	0.1080	0.1114	0.2214	0.2140	0.2130	0.2171	0.2327
$\mathrm{p}_{\mathrm{S}}^{\mathrm{LB}}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\mathrm{p}_{\mathrm{S}}^{\mathrm{BP}}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 3.13 – Modified and standard versions of portmanteau tests to check the null hypothesis that the Nikkei squared returns follow a FARIMA$(0, 0.2132, 0)$ model as in (3.37) with $a = b = 0$.

Lag m	1	2	3	4	5	6	7
$\hat{\rho}(m)$	-0.0678	0.0400	0.0634	-0.0022	0.0165	0.0320	-0.0158
$\mathrm{LB}_{\mathrm{SN}}$	5.7332	29.005	34.758	34.779	66.692	288.57	324.46
$\mathrm{BP}_{\mathrm{SN}}$	5.7319	28.997	34.745	34.764	66.657	288.40	324.24
LB_{W}	61.211	82.507	136.13	136.20	139.84	153.46	156.78
BP_{W}	61.198	82.487	136.09	136.16	139.76	153.41	156.73
$\mathrm{p}_{\mathrm{W}}^{\mathrm{LB}}$	0.1086	0.2186	0.1830	0.2551	0.3002	0.3519	0.3609
$\mathrm{p}_{\mathrm{W}}^{\mathrm{BP}}$	0.1086	0.2187	0.1831	0.2552	0.3003	0.3521	0.3611
$\mathrm{p}_{\mathrm{S}}^{\mathrm{LB}}$	n.a.	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\mathrm{p}_{\mathrm{S}}^{\mathrm{BP}}$	n.a.	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Lag m	8	9	10	11	12	13	14
$\hat{\rho}(m)$	0.0295	0.0384	0.0121	0.0133	0.0503	0.0076	0.0068
$\mathrm{LB}_{\mathrm{SN}}$	387.88	512.70	575.09	600.81	791.67	808.20	808.27
$\mathrm{BP}_{\mathrm{SN}}$	387.59	512.28	574.57	600.22	790.83	807.29	807.30
LB_{W}	168.41	188.08	190.01	192.36	226.12	226.89	227.50
BP_{W}	168.35	187.10	189.93	192.29	225.10	226.76	227.39
$\mathrm{p}_{\mathrm{W}}^{\mathrm{LB}}$	0.3627	0.3757	0.3802	0.3825	0.3320	0.3447	0.3526
$\mathrm{p}_{\mathrm{W}}^{\mathrm{BP}}$	0.3629	0.3759	0.3804	0.3827	0.3323	0.3450	0.3529
$\mathrm{p}_{\mathrm{S}}^{\mathrm{LB}}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\mathrm{p}_{\mathrm{S}}^{\mathrm{BP}}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Lag m	15	16	17	18	19	20	21
$\hat{\rho}(m)$	0.0538	0.0073	0.0173	0.0067	-0.0027	-0.0057	0.0153
$\mathrm{LB}_{\mathrm{SN}}$	839.87	842.24	842.31	845.36	885.74	935.70	946.03
$\mathrm{BP}_{\mathrm{SN}}$	838.80	841.10	841.11	844.10	884.35	934.15	944.40
LB_{W}	266.16	266.88	270.85	271.45	271.56	271.99	275.13
BP_{W}	265.99	266.71	270.68	271.28	271.38	271.82	274.94
$\mathrm{p}_{\mathrm{W}}^{\mathrm{LB}}$	0.3105	0.3163	0.3161	0.3264	0.3289	0.3329	0.3366
$\mathrm{p}_{\mathrm{W}}^{\mathrm{BP}}$	0.3108	0.3166	0.3165	0.3268	0.3293	0.3333	0.3369
$\mathrm{p}_{\mathrm{S}}^{\mathrm{LB}}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\mathrm{p}_{\mathrm{S}}^{\mathrm{BP}}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

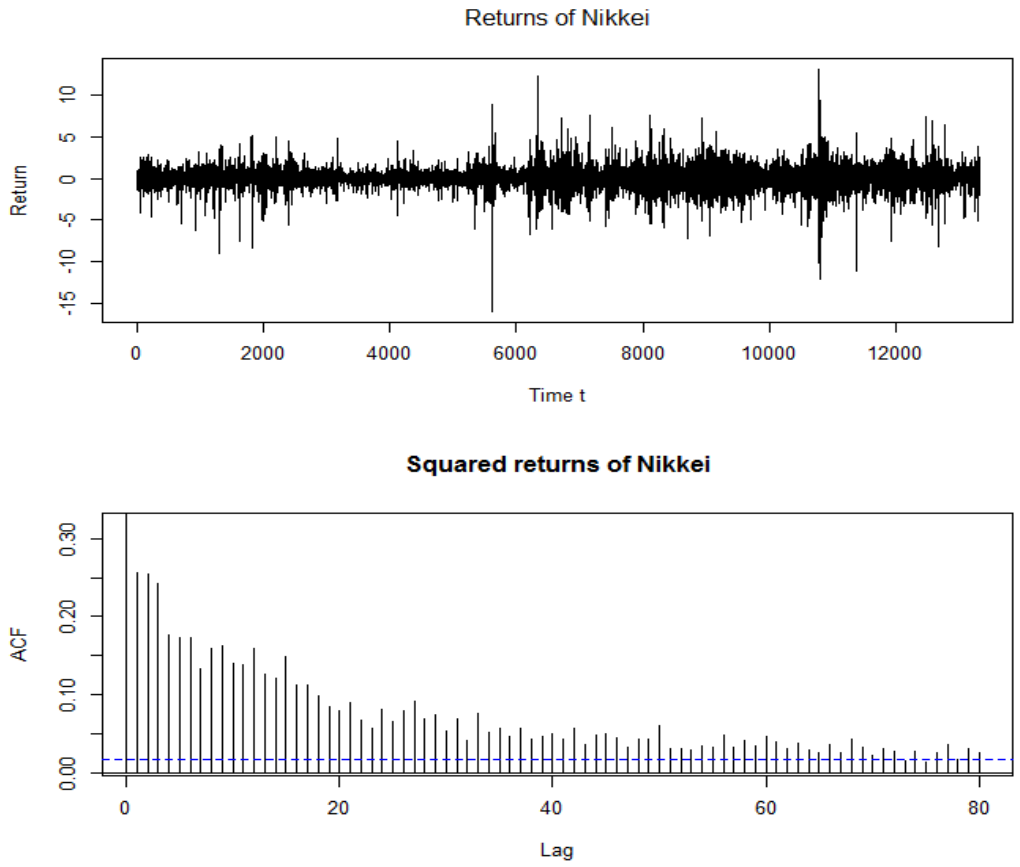


Figure 3.3 – Returns and the sample autocorrelations of squared returns of the Nikkei.

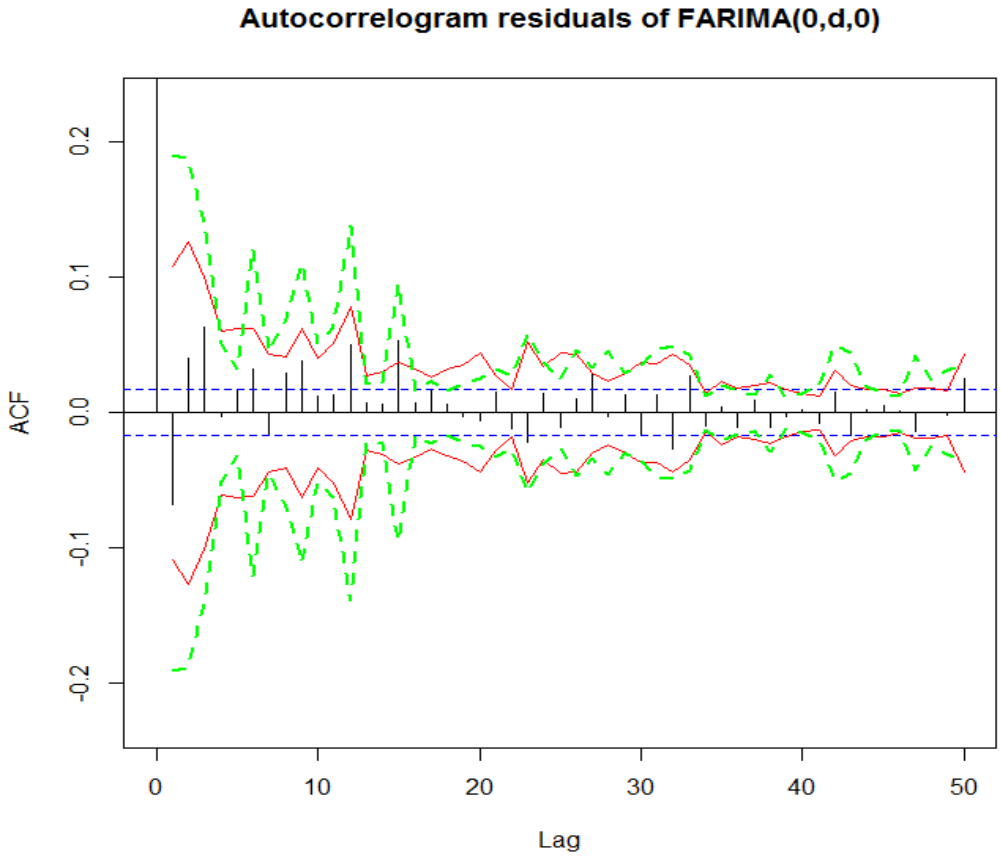


Figure 3.4 – Autocorrelation of the FARIMA$(0, 0.2132, 0)$ residuals for the squares of the Nikkei returns. The horizontal dotted lines (blue color) correspond to the 5% significant limits obtained under the strong FARIMA assumption. The solid lines (red color) and dashed lines (green color) correspond also to the 5% significant limits under the weak FARIMA assumption. The full lines correspond to the asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.1. The dashed lines (green color) correspond to the self-normalized asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.5.

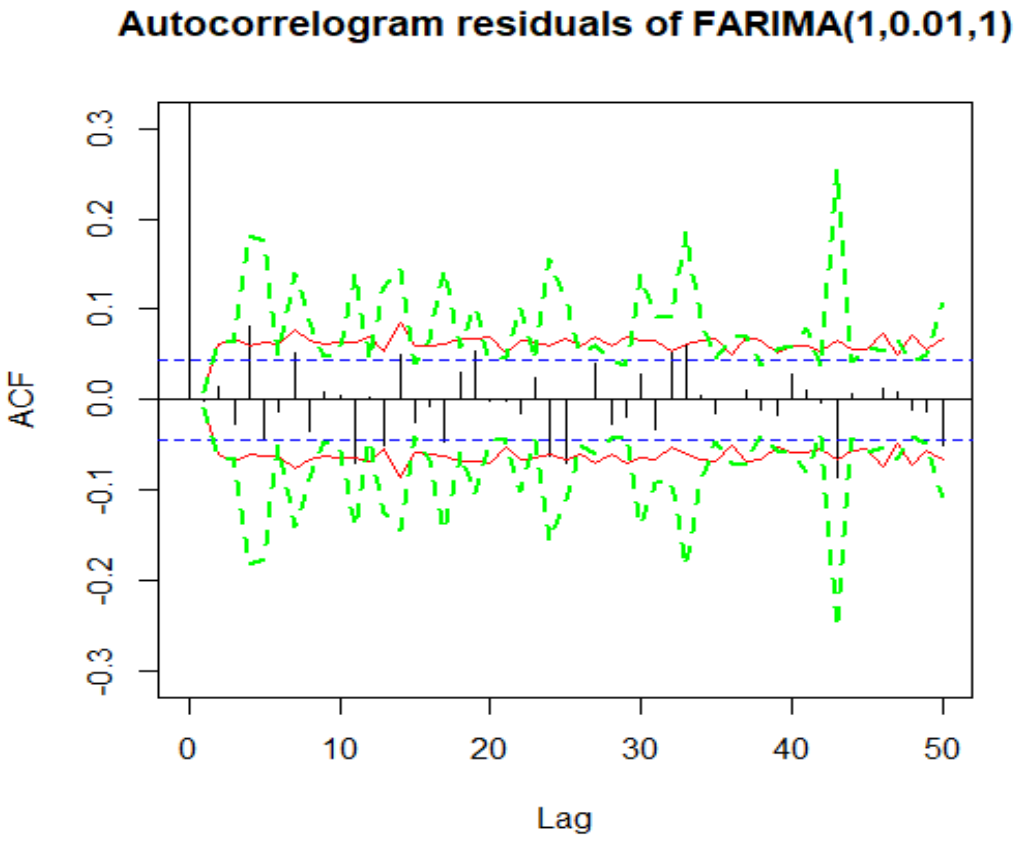


Figure 3.5 – Autocorrelation of a realization of size $n = 2,000$ for a weak FARIMA$(1, 0.01, 1)$ model (3.34)–(3.20) with $\theta_0 = (0.9, 0.2, 0.01)$ and where $\omega = 0.04$, $\alpha_1 = 0.13$ and $\beta_1 = 0.88$. The horizontal dotted lines (blue color) correspond to the 5% significant limits obtained under the strong FARIMA assumption. The solid lines (red color) and dashed lines (green color) correspond also to the 5% significant limits under the weak FARIMA assumption. The full lines correspond to the asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.1. The dashed lines (green color) correspond to the self-normalized asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.5.

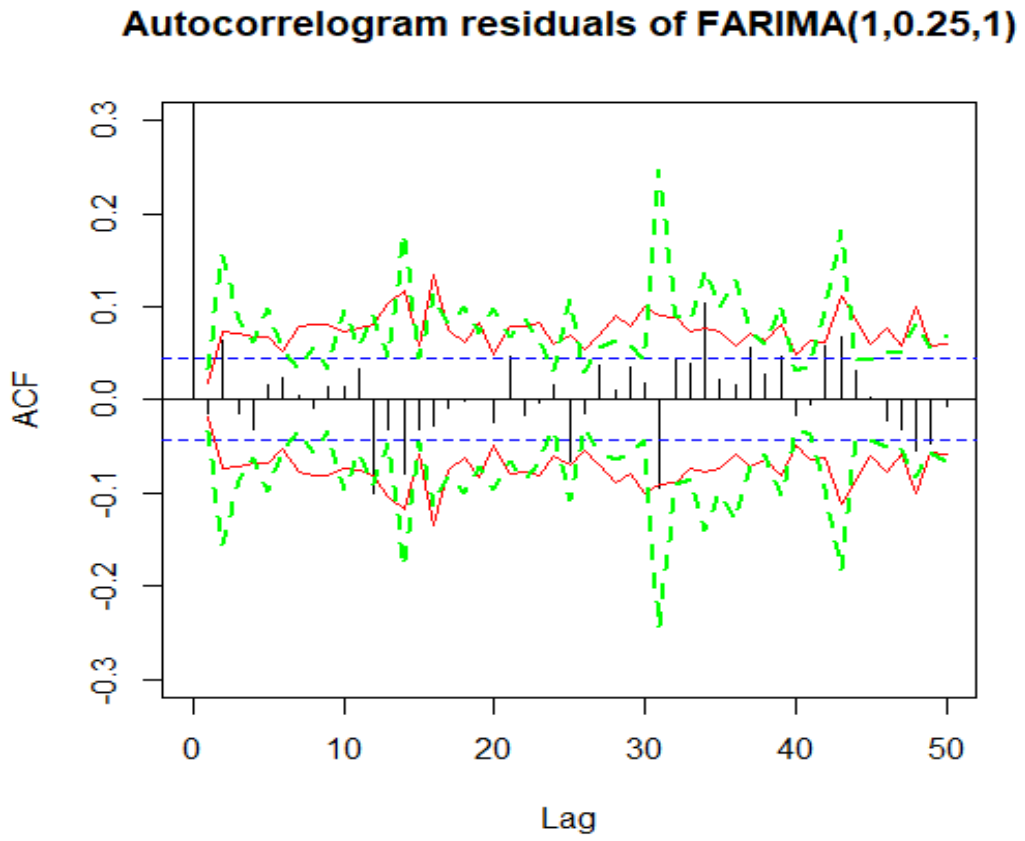


Figure 3.6 – Autocorrelation of a realization of size $n = 2,000$ for a weak FARIMA$(1, 0.25, 1)$ model (3.34)–(3.20) with $\theta_0 = (0.9, 0.2, 0.25)$ and where $\omega = 0.04$, $\alpha_1 = 0.13$ and $\beta_1 = 0.88$. The horizontal dotted lines (blue color) correspond to the 5% significant limits obtained under the strong FARIMA assumption. The solid lines (red color) and dashed lines (green color) correspond also to the 5% significant limits under the weak FARIMA assumption. The full lines correspond to the asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.1. The dashed lines (green color) correspond to the self-normalized asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.5.

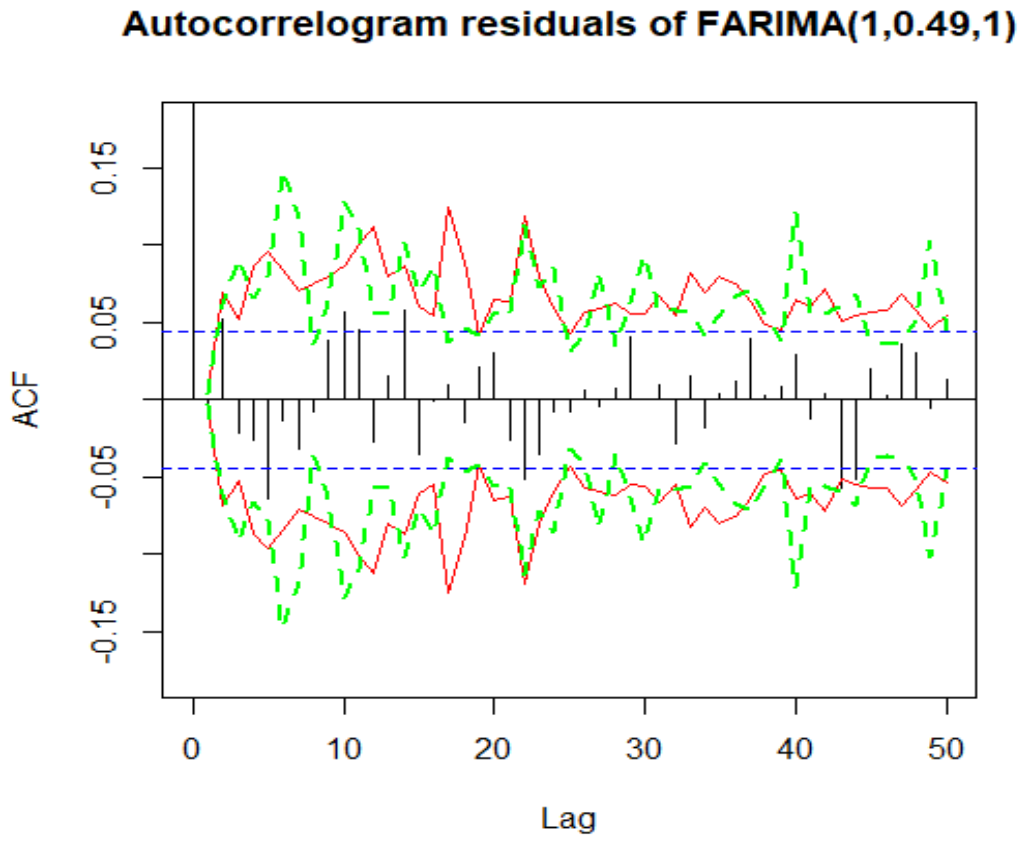


Figure 3.7 – Autocorrelation of a realization of size $n = 2,000$ for a weak FARIMA$(1, 0.49, 1)$ model (3.34)–(3.20) with $\theta_0 = (0.9, 0.2, 0.49)$ and where $\omega = 0.04$, $\alpha_1 = 0.13$ and $\beta_1 = 0.88$. The horizontal dotted lines (blue color) correspond to the 5% significant limits obtained under the strong FARIMA assumption. The solid lines (red color) and dashed lines (green color) correspond also to the 5% significant limits under the weak FARIMA assumption. The full lines correspond to the asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.1. The dashed lines (green color) correspond to the self-normalized asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.5.

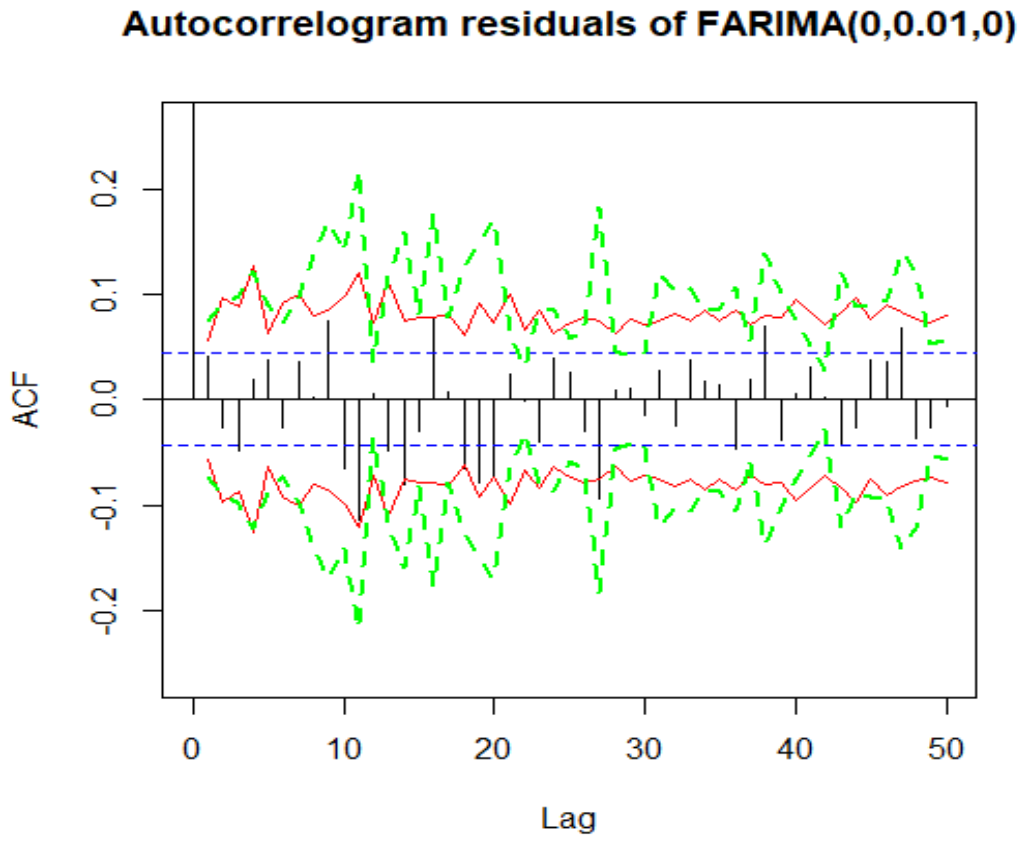


Figure 3.8 – Autocorrelation of a realization of size $n = 2,000$ for a weak FARIMA$(0, 0.01, 0)$ model (3.34)–(3.20) with $\theta_0 = (0, 0, 0.01)$ and where $\omega = 0.04$, $\alpha_1 = 0.13$ and $\beta_1 = 0.88$. The horizontal dotted lines (blue color) correspond to the 5% significant limits obtained under the strong FARIMA assumption. The solid lines (red color) and dashed lines (green color) correspond also to the 5% significant limits under the weak FARIMA assumption. The full lines correspond to the asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.1. The dashed lines (green color) correspond to the self-normalized asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.5.

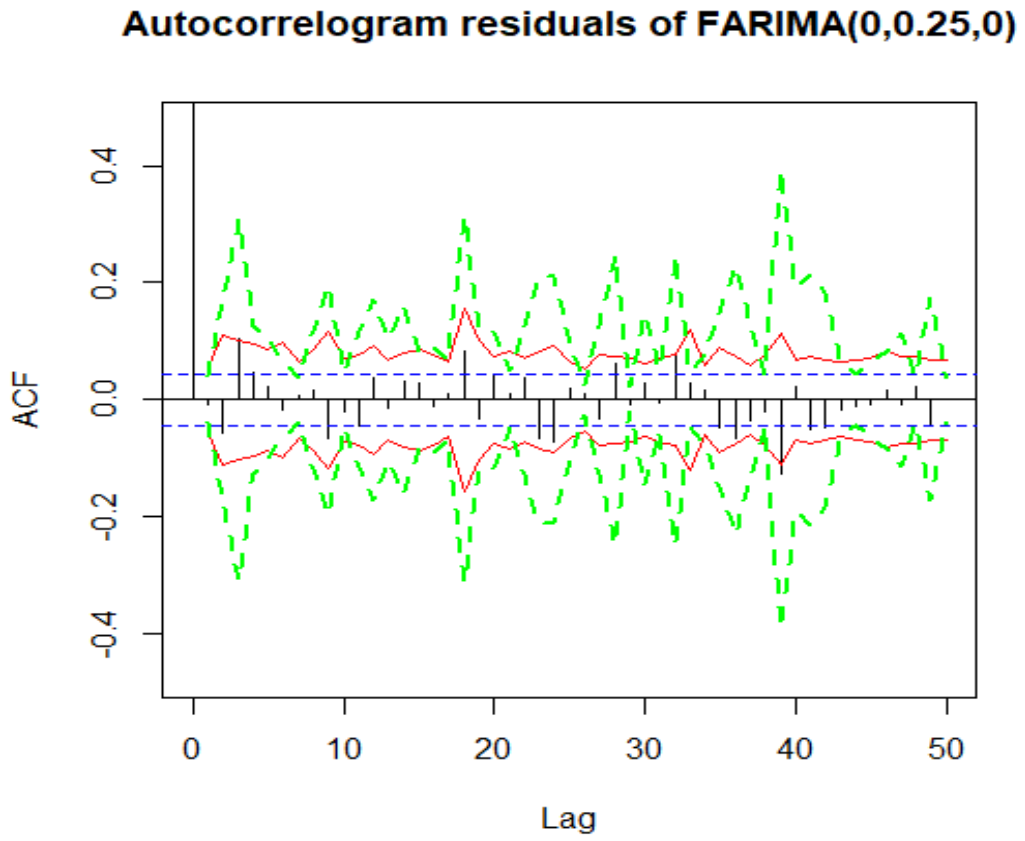


Figure 3.9 – Autocorrelation of a realization of size $n = 2,000$ for a weak FARIMA$(0, 0.25, 0)$ model (3.34)–(3.20) with $\theta_0 = (0, 0, 0.25)$ and where $\omega = 0.04$, $\alpha_1 = 0.13$ and $\beta_1 = 0.88$. The horizontal dotted lines (blue color) correspond to the 5% significant limits obtained under the strong FARIMA assumption. The solid lines (red color) and dashed lines (green color) correspond also to the 5% significant limits under the weak FARIMA assumption. The full lines correspond to the asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.1. The dashed lines (green color) correspond to the self-normalized asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.5.

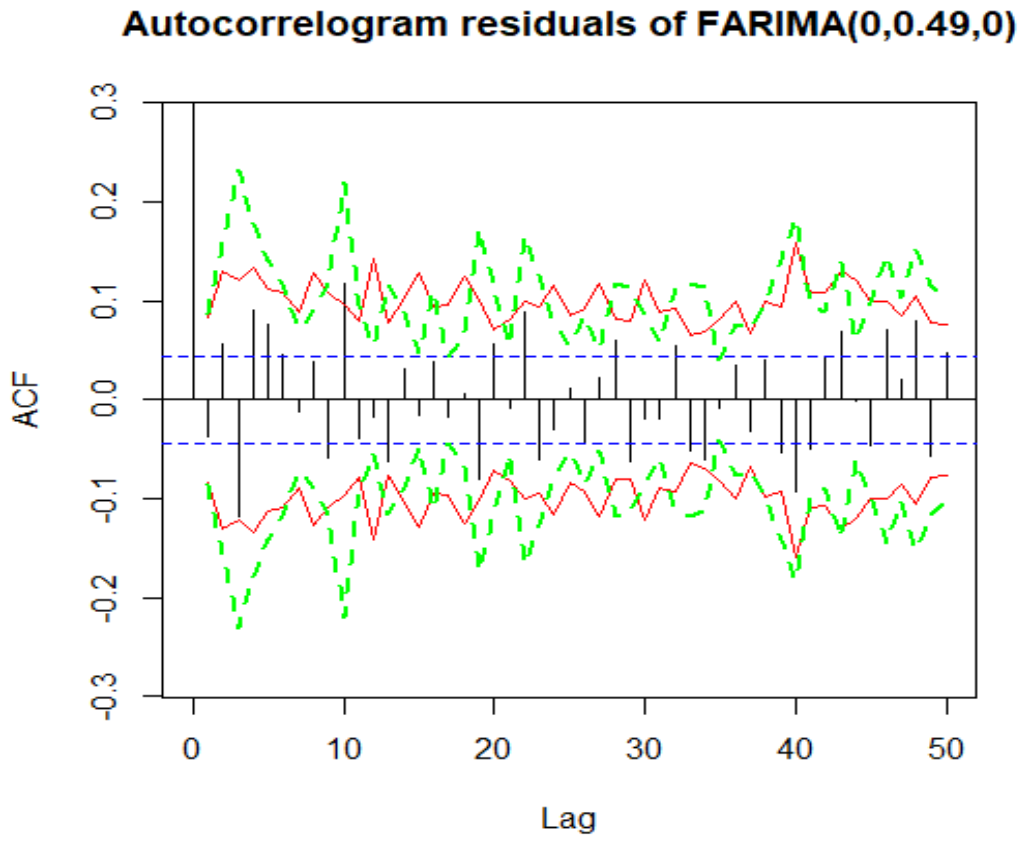


Figure 3.10 – Autocorrelation of a realization of size $n = 2,000$ for a weak FARIMA$(0, 0.49, 0)$ model (3.34)–(3.20) with $\theta_0 = (0, 0, 0.49)$ and where $\omega = 0.04$, $\alpha_1 = 0.13$ and $\beta_1 = 0.88$. The horizontal dotted lines (blue color) correspond to the 5% significant limits obtained under the strong FARIMA assumption. The solid lines (red color) and dashed lines (green color) correspond also to the 5% significant limits under the weak FARIMA assumption. The full lines correspond to the asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.1. The dashed lines (green color) correspond to the self-normalized asymptotic significance limits for the residual autocorrelations obtained in Theorem 3.5.

3.7 Supplementary material: additional Monte Carlo experiments

The following tables deal with the same numerical experiments that in Section 3.4 when the sample sizes are less than 500.

Table 3.14 – Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a strong FARIMA$(1, d_0, 1)$ defined by (3.34) with $\theta_0 = (0.9, 0.2, d_0)$. The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$.

d_0	Length n	Lag m	LB_{SN}	BP_{SN}	LB_W	BP_W	LB_S	BP_S
0.05	$n = 100$	1	4.7	4.4	**23.2**	**22.9**	n.a.	n.a.
		2	3.9	3.6	**8.1**	**7.5**	n.a.	n.a.
		3	4.3	4.0	**6.9**	6.1	n.a.	n.a.
		6	4.7	3.6	5.4	3.7	**8.4**	5.9
		12	5.1	**2.9**	4.2	**2.3**	5.0	**2.6**
		15	6.2	3.5	4.9	**2.5**	5.8	**2.5**
0.05	$n = 250$	1	5.3	5.3	**10.8**	**10.7**	n.a.	n.a.
		2	3.6	**3.3**	**6.8**	**6.8**	n.a.	n.a.
		3	4.0	3.7	5.7	5.4	n.a.	n.a.
		6	4.2	3.7	5.4	5.1	**10.6**	**9.6**
		12	**3.1**	**2.2**	5.3	4.2	**6.5**	5.7
		15	**3.3**	**2.5**	5.6	4.3	6.4	5.2
0.05	$n = 500$	1	4.6	4.6	**6.9**	**6.8**	n.a.	n.a.
		2	4.3	4.2	5.8	5.6	n.a.	n.a.
		3	4.3	4.2	5.7	5.5	n.a.	n.a.
		6	5.0	4.8	**6.7**	**6.5**	**11.0**	**10.7**
		12	4.9	4.2	5.5	4.6	**7.1**	6.2
		15	5.6	4.3	5.7	4.5	**7.1**	6.2
0.20	$n = 100$	1	5.1	4.8	**27.1**	**25.9**	n.a.	n.a.
		2	4.0	3.8	**8.7**	**8.2**	n.a.	n.a.
		3	4.1	4.0	**7.5**	**6.9**	n.a.	n.a.
		6	5.5	3.9	5.3	3.9	**7.6**	6.2
		12	4.9	**3.0**	4.3	**2.6**	4.3	**2.9**
		15	**6.9**	**2.4**	5.1	**2.9**	5.2	**2.7**
0.20	$n = 250$	1	5.1	5.0	**14.0**	**13.9**	n.a.	n.a.
		2	3.4	**3.1**	**7.3**	**7.2**	n.a.	n.a.
		3	4.3	4.1	6.2	5.9	n.a.	n.a.
		6	4.7	4.3	6.0	5.5	**10.3**	**9.8**
		12	3.8	**2.6**	5.1	4.3	5.7	5.1
		15	3.9	**2.8**	5.9	4.4	5.7	5.0
0.20	$n = 500$	1	5.6	5.6	**12.1**	**12.1**	n.a.	n.a.
		2	4.9	4.9	**7.0**	**6.9**	n.a.	n.a.
		3	5.0	4.9	**6.7**	6.4	n.a.	n.a.
		6	5.5	5.2	6.2	5.7	**10.1**	**9.6**
		12	5.6	4.8	5.3	4.6	6.3	5.3
		15	5.7	4.4	5.4	4.5	5.9	5.1
0.45	$n = 100$	1	**3.2**	**3.1**	**32.0**	**31.6**	n.a.	n.a.
		2	**3.5**	**3.4**	**8.3**	**7.3**	n.a.	n.a.
		3	**2.9**	**2.5**	**6.9**	6.4	n.a.	n.a.
		6	3.8	**2.9**	3.6	**2.8**	4.6	**3.5**
		12	3.6	**1.3**	**2.7**	**1.8**	**2.1**	**1.2**
		15	4.1	**1.9**	3.7	**1.5**	**2.2**	**0.9**
0.45	$n = 250$	1	**3.4**	**3.3**	**18.3**	**18.0**	n.a.	n.a.
		2	**3.2**	**3.2**	6.4	6.1	n.a.	n.a.
		3	3.6	**3.4**	5.2	5.1	n.a.	n.a.
		6	3.8	**3.3**	4.8	4.4	**7.9**	**7.3**
		12	**3.1**	**2.3**	4.0	**3.2**	4.4	3.7
		15	**3.2**	**2.3**	4.7	**3.3**	4.0	**3.1**
0.45	$n = 500$	1	3.6	3.6	**14.5**	**14.4**	n.a.	n.a.
		2	**3.4**	**3.4**	5.3	5.3	n.a.	n.a.
		3	**3.4**	**3.4**	5.5	5.5	n.a.	n.a.
		6	5.0	4.7	4.9	4.6	**7.2**	**7.0**
		12	5.2	4.7	4.4	3.9	4.2	4.0
		15	5.0	4.3	4.4	3.6	4.2	3.7

Table 3.15 – Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(1, d_0, 1)$ defined by (3.34) with $\theta_0 = (0.9, 0.2, d_0)$ and where $\omega = 0.4$, $\alpha_1 = 0.3$ and $\beta_1 = 0.3$ in (3.20). The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$.

d_0	Length n	Lag m	LB_{SN}	BP_{SN}	LB_W	BP_W	LB_S	BP_S
0.05	$n = 100$	1	**3.1**	**3.1**	**19.7**	**18.7**	n.a.	n.a.
		2	**2.0**	**1.7**	**7.8**	**7.3**	n.a.	n.a.
		3	**1.7**	**1.6**	**6.8**	6.2	n.a.	n.a.
		6	**1.4**	**0.9**	6.1	4.7	**15.6**	**12.5**
		12	**1.5**	**0.9**	5.1	3.7	**13.5**	**8.9**
		15	**2.0**	**1.2**	5.1	**2.6**	**13.1**	**8.9**
0.05	$n = 250$	1	**2.5**	**2.4**	**10.6**	**10.0**	n.a.	n.a.
		2	**2.1**	**1.7**	**6.6**	6.4	n.a.	n.a.
		3	**1.2**	**1.1**	5.7	5.2	n.a.	n.a.
		6	**0.8**	**0.8**	5.3	4.7	**25.0**	**24.2**
		12	**0.8**	**0.7**	3.7	**3.3**	**23.5**	**21.5**
		15	**1.1**	**1.1**	3.8	**3.0**	**24.7**	**21.8**
0.05	$n = 500$	1	**2.4**	**2.4**	**8.1**	**8.1**	n.a.	n.a.
		2	**1.7**	**1.7**	**7.1**	**7.0**	n.a.	n.a.
		3	**0.8**	**0.7**	6.1	6.0	n.a.	n.a.
		6	**0.7**	**0.6**	4.6	4.2	**31.5**	**31.0**
		12	**1.1**	**1.1**	3.9	3.8	**33.5**	**32.3**
		15	**1.0**	**0.9**	4.6	4.0	**35.0**	**33.4**
0.20	$n = 100$	1	**2.6**	**2.6**	**24.0**	**23.4**	n.a.	n.a.
		2	**1.7**	**1.6**	**9.0**	**8.4**	n.a.	n.a.
		3	**2.3**	**1.7**	**6.7**	6.2	n.a.	n.a.
		6	**1.5**	**0.8**	5.5	4.2	**15.2**	**12.3**
		12	**1.4**	**0.6**	4.5	**3.1**	**12.0**	**7.7**
		15	**2.0**	**0.8**	4.7	**2.8**	**11.2**	**7.5**
0.20	$n = 250$	1	**3.5**	**3.5**	**17.1**	**16.8**	n.a.	n.a.
		2	**1.9**	**1.9**	**8.5**	**8.0**	n.a.	n.a.
		3	**1.1**	**1.0**	5.5	5.0	n.a.	n.a.
		6	**0.7**	**0.7**	4.3	4.1	**24.2**	**23.4**
		12	**0.6**	**0.6**	**3.3**	**2.9**	**22.1**	**19.7**
		15	**0.6**	**0.5**	3.8	**3.1**	**22.9**	**20.1**
0.20	$n = 500$	1	**2.5**	**2.4**	**12.0**	**11.8**	n.a.	n.a.
		2	**2.0**	**2.0**	**7.7**	**7.7**	n.a.	n.a.
		3	**1.4**	**1.4**	6.1	5.6	n.a.	n.a.
		6	**0.8**	**0.8**	4.3	4.0	**30.2**	**29.6**
		12	**0.8**	**0.7**	**3.4**	**3.2**	**33.2**	**31.7**
		15	**0.7**	**0.6**	4.3	3.8	**34.3**	**32.7**
0.45	$n = 100$	1	**2.4**	**2.3**	**33.2**	**32.9**	n.a.	n.a.
		2	**1.4**	**1.3**	**8.5**	**7.8**	n.a.	n.a.
		3	**1.5**	**1.2**	6.3	5.4	n.a.	n.a.
		6	**1.4**	**0.8**	4.5	**3.5**	**10.5**	**8.3**
		12	**0.8**	**0.3**	4.3	**2.7**	**7.0**	5.0
		15	**1.5**	**0.4**	4.1	**2.4**	**7.5**	4.3
0.45	$n = 250$	1	**2.1**	**2.1**	**20.1**	**20.1**	n.a.	n.a.
		2	**1.7**	**1.7**	5.9	5.8	n.a.	n.a.
		3	**1.1**	**0.8**	5.2	4.9	n.a.	n.a.
		6	**0.9**	**0.9**	4.1	3.7	**18.8**	**18.0**
		12	**0.4**	**0.4**	**2.6**	**2.1**	**17.4**	**15.4**
		15	**0.2**	**0.2**	4.2	**3.0**	**18.4**	**15.7**
0.45	$n = 500$	1	**2.1**	**2.1**	**13.3**	**13.2**	n.a.	n.a.
		2	**1.2**	**1.2**	5.8	5.7	n.a.	n.a.
		3	**1.1**	**1.0**	4.9	4.9	n.a.	n.a.
		6	**0.6**	**0.6**	4.0	3.8	**27.3**	**26.4**
		12	**0.2**	**0.2**	**3.1**	**2.8**	**28.3**	**27.0**
		15	**0.2**	**0.1**	4.3	3.8	**28.4**	**26.8**

Table 3.16 – Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(1, d_0, 1)$ defined by (3.34) with $\theta_0 = (0.9, 0.2, d_0)$ and where $\omega = 0.04$, $\alpha_1 = 0.12$ and $\beta_1 = 0.85$ in (3.20). The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$.

d_0	Length n	Lag m	LB_{SN}	BP_{SN}	LB_W	BP_W	LB_S	BP_S
0.05	$n = 100$	1	**3.1**	**3.1**	**19.7**	**18.7**	n.a.	n.a.
		2	**2.0**	**1.7**	**7.8**	**7.3**	n.a.	n.a.
		3	**1.7**	**1.6**	**6.8**	6.2	n.a.	n.a.
		6	**1.4**	**0.9**	6.1	4.7	**15.6**	**12.5**
		12	**1.5**	**0.9**	5.1	3.7	**13.5**	**8.9**
		15	**2.0**	**1.2**	5.1	**2.6**	**13.1**	**8.9**
0.05	$n = 250$	1	**2.5**	**2.4**	**10.6**	**10.0**	n.a.	n.a.
		2	**2.1**	**1.7**	**6.6**	6.4	n.a.	n.a.
		3	**1.2**	**1.1**	5.7	5.2	n.a.	n.a.
		6	**0.8**	**0.8**	5.3	4.7	**25.0**	**24.2**
		12	**0.8**	**0.7**	3.7	**3.3**	**23.5**	**21.5**
		15	**1.1**	**1.1**	3.8	**3.0**	**24.7**	**21.8**
0.05	$n = 500$	1	**2.4**	**2.4**	**8.1**	**8.1**	n.a.	n.a.
		2	**1.7**	**1.7**	**7.1**	**7.0**	n.a.	n.a.
		3	**0.8**	**0.7**	6.1	6.0	n.a.	n.a.
		6	**0.7**	**0.6**	4.6	4.2	**31.5**	**31.0**
		12	**1.1**	**1.1**	3.9	3.8	**33.5**	**32.3**
		15	**1.0**	**0.9**	4.6	4.0	**35.0**	**33.4**
0.20	$n = 100$	1	**2.6**	**2.6**	**24.0**	**23.4**	n.a.	n.a.
		2	**1.7**	**1.6**	**9.0**	**8.4**	n.a.	n.a.
		3	**2.3**	**1.7**	**6.7**	6.2	n.a.	n.a.
		6	**1.5**	**0.8**	5.5	4.2	**15.2**	**12.3**
		12	**1.4**	**0.6**	4.5	**3.1**	**12.0**	**7.7**
		15	**2.0**	**0.8**	4.7	**2.8**	**11.2**	**7.5**
0.20	$n = 250$	1	**3.5**	**3.5**	**17.1**	**16.8**	n.a.	n.a.
		2	**1.9**	**1.9**	**8.5**	**8.0**	n.a.	n.a.
		3	**1.1**	**1.0**	5.5	5.0	n.a.	n.a.
		6	**0.7**	**0.7**	4.3	4.1	**24.2**	**23.4**
		12	**0.6**	**0.6**	**3.3**	**2.9**	**22.1**	**19.7**
		15	**0.6**	**0.5**	3.8	**3.1**	**22.9**	**20.1**
0.20	$n = 500$	1	**2.5**	**2.4**	**12.0**	**11.8**	n.a.	n.a.
		2	**2.0**	**2.0**	**7.7**	**7.7**	n.a.	n.a.
		3	**1.4**	**1.4**	6.1	5.6	n.a.	n.a.
		6	**0.8**	**0.8**	4.3	4.0	**30.2**	**29.6**
		12	**0.8**	**0.7**	**3.4**	**3.2**	**33.2**	**31.7**
		15	**0.7**	**0.6**	4.3	3.8	**34.3**	**32.7**
0.45	$n = 100$	1	**2.4**	**2.3**	**33.2**	**32.9**	n.a.	n.a.
		2	**1.4**	**1.3**	**8.5**	**7.8**	n.a.	n.a.
		3	**1.5**	**1.2**	6.3	5.4	n.a.	n.a.
		6	**1.4**	**0.8**	4.5	**3.5**	**10.5**	**8.3**
		12	**0.8**	**0.3**	4.3	**2.7**	**7.0**	5.0
		15	**1.5**	**0.4**	4.1	**2.4**	**7.5**	4.3
0.45	$n = 250$	1	**2.1**	**2.1**	**20.1**	**20.1**	n.a.	n.a.
		2	**1.7**	**1.7**	5.9	5.8	n.a.	n.a.
		3	**1.1**	**0.8**	5.2	4.9	n.a.	n.a.
		6	**0.9**	**0.9**	4.1	3.7	**18.8**	**18.0**
		12	**0.4**	**0.4**	**2.6**	**2.1**	**17.4**	**15.4**
		15	**0.2**	**0.2**	4.2	**3.0**	**18.4**	**15.7**
0.45	$n = 500$	1	**2.1**	**2.1**	**13.3**	**13.2**	n.a.	n.a.
		2	**1.2**	**1.2**	5.8	5.7	n.a.	n.a.
		3	**1.1**	**1.0**	4.9	4.9	n.a.	n.a.
		6	**0.6**	**0.6**	4.0	3.8	**27.3**	**26.4**
		12	**0.2**	**0.2**	**3.1**	**2.8**	**28.3**	**27.0**
		15	**0.2**	**0.1**	4.3	3.8	**28.4**	**26.8**

Table 3.17 – Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a strong FARIMA$(0, d_0, 0)$ defined by (3.34) with $\theta_0 = (0, 0, d_0)$. The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$.

d_0	Length n	Lag m	LB_{SN}	BP_{SN}	LB_W	BP_W	LB_S	BP_S
0.05	$n = 100$	1	3.9	3.6	**10.1**	**9.6**	n.a.	n.a.
		2	**3.3**	**3.2**	**8.1**	**7.4**	**7.6**	**7.1**
		3	3.8	**3.1**	5.9	5.2	**8.1**	**6.8**
		6	**3.1**	**2.7**	5.0	3.9	**6.9**	5.9
		12	**2.4**	**1.3**	3.9	**2.1**	5.8	3.8
		15	**2.8**	**1.0**	4.5	**2.3**	**6.9**	4.3
0.05	$n = 250$	1	5.3	5.2	**7.6**	**7.3**	n.a.	n.a.
		2	5.0	4.7	5.4	5.3	6.1	6.0
		3	4.7	4.5	5.6	5.5	5.8	5.6
		6	5.2	4.8	6.4	6.1	6.7	6.3
		12	5.0	3.8	4.4	3.7	6.2	5.3
		15	4.6	**3.2**	4.4	**3.5**	6.0	4.9
0.05	$n = 500$	1	5.0	5.0	5.6	5.6	n.a.	n.a.
		2	5.5	5.5	5.7	5.6	6.0	5.8
		3	5.9	5.7	5.9	5.7	**6.6**	**6.5**
		6	5.3	5.1	5.6	5.2	6.0	5.9
		12	5.1	4.3	5.0	4.7	5.9	5.0
		15	5.4	4.5	4.6	4.2	6.0	5.2
0.20	$n = 100$	1	4.5	4.0	5.9	5.3	n.a.	n.a.
		2	4.1	3.7	6.5	6.0	**6.5**	5.8
		3	4.1	**3.5**	5.3	4.9	6.4	6.1
		6	3.3	**2.9**	4.6	3.7	6.1	4.9
		12	3.6	**1.5**	4.1	**2.0**	5.5	**3.4**
		15	**2.9**	**0.9**	4.4	**2.0**	6.5	**3.5**
0.20	$n = 250$	1	5.8	5.7	5.8	5.7	n.a.	n.a.
		2	5.2	5.1	5.2	4.8	5.8	5.6
		3	5.1	5.0	5.5	5.4	5.4	5.1
		6	5.7	5.4	5.9	5.3	6.3	5.7
		12	5.6	4.0	4.2	3.8	5.8	5.1
		15	4.8	3.6	4.5	3.6	6.2	4.7
0.20	$n = 500$	1	5.7	5.5	5.0	5.0	n.a.	n.a.
		2	5.4	5.4	5.4	5.3	5.5	5.3
		3	6.2	6.1	5.7	5.6	6.3	6.2
		6	5.4	5.0	5.5	5.0	5.6	5.6
		12	5.1	4.4	5.0	4.7	6.0	5.0
		15	5.2	4.3	4.4	4.2	5.9	5.1
0.45	$n = 100$	1	4.3	4.1	**9.4**	**8.9**	n.a.	n.a.
		2	3.9	3.4	**8.3**	**7.5**	**7.7**	**7.3**
		3	4.0	**3.3**	**6.5**	5.7	**7.0**	**6.5**
		6	**3.3**	**2.4**	4.7	**3.5**	**6.5**	5.3
		12	**3.5**	**1.7**	**3.9**	**2.3**	5.5	**3.2**
		15	3.9	**1.4**	4.2	**2.2**	6.1	3.7
0.45	$n = 250$	1	5.4	5.4	**8.2**	**7.9**	n.a.	n.a.
		2	5.0	4.9	5.3	5.1	5.5	5.3
		3	5.1	5.0	5.8	5.3	5.3	5.0
		6	5.6	5.2	6.0	5.2	6.2	5.4
		12	5.4	3.9	4.6	3.9	5.8	5.2
		15	5.1	4.0	4.7	3.7	6.2	5.0
0.45	$n = 500$	1	5.4	5.2	5.6	5.6	n.a.	n.a.
		2	5.2	5.2	5.4	5.3	5.9	5.8
		3	5.9	5.8	6.3	6.1	6.4	6.4
		6	6.0	5.6	5.6	5.0	5.6	5.5
		12	4.9	3.9	5.6	4.8	5.7	5.1
		15	5.2	4.3	4.6	4.2	6.1	4.9

Table 3.18 – Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(0, d_0, 0)$ defined by (3.34) with $\theta_0 = (0, 0, d_0)$ with $\omega = 0.4$, $\alpha_1 = 0.3$ and $\beta_1 = 0.3$ in (3.35). The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$.

d_0	Length n	Lag m	LB_{SN}	BP_{SN}	LB_W	BP_W	LB_S	BP_S
0.05	$n = 100$	1	**2.3**	**2.3**	**10.1**	**9.6**	n.a.	n.a.
		2	**2.6**	**2.6**	5.9	5.3	**13.1**	**12.4**
		3	**1.9**	**1.6**	4.0	**3.1**	**11.1**	**9.9**
		6	**1.4**	**1.1**	**3.0**	**2.5**	**12.8**	**11.2**
		12	**1.0**	**0.3**	**3.5**	**2.0**	**14.5**	**10.8**
		15	**0.8**	**0.1**	**2.6**	**0.8**	**16.1**	**11.0**
0.05	$n = 250$	1	**3.0**	**3.0**	**8.1**	**8.1**	n.a.	n.a.
		2	**2.6**	**2.4**	5.3	5.2	**16.4**	**16.4**
		3	**1.9**	**1.8**	4.3	3.9	**16.2**	**15.6**
		6	**0.7**	**0.4**	4.3	4.1	**20.1**	**18.8**
		12	**0.6**	**0.5**	3.6	**2.6**	**24.6**	**22.4**
		15	**0.2**	**0.2**	4.0	**2.9**	**25.7**	**22.4**
0.05	$n = 500$	1	**3.4**	**3.4**	**7.2**	**7.0**	n.a.	n.a.
		2	**2.0**	**2.0**	6.3	6.3	**20.4**	**20.3**
		3	**1.5**	**1.5**	5.1	5.0	**21.1**	**20.7**
		6	**0.9**	**0.9**	4.6	4.6	**28.0**	**27.6**
		12	**0.4**	**0.4**	4.0	**3.2**	**34.2**	**32.8**
		15	**0.1**	**0.0**	**3.3**	**3.0**	**36.2**	**34.7**
0.20	$n = 100$	1	**2.8**	**2.7**	5.3	5.0	n.a.	n.a.
		2	**3.1**	**3.1**	4.9	4.2	**10.9**	**10.1**
		3	**1.8**	**1.6**	3.8	**2.9**	**9.9**	**8.3**
		6	**1.9**	**1.1**	**2.9**	**2.0**	**10.8**	**9.0**
		12	**0.8**	**0.3**	**3.1**	**1.8**	**13.1**	**9.7**
		15	**0.7**	**0.1**	**2.3**	**0.7**	**14.7**	**9.6**
0.20	$n = 250$	1	**3.2**	**3.2**	5.5	5.4	n.a.	n.a.
		2	**3.0**	**3.0**	4.3	4.2	**14.4**	**14.3**
		3	**2.4**	**2.3**	3.6	**3.4**	**14.9**	**14.2**
		6	**0.7**	**0.7**	4.3	3.8	**18.3**	**17.3**
		12	**0.6**	**0.4**	**3.5**	**2.6**	**23.6**	**21.2**
		15	**0.4**	**0.1**	3.8	**2.5**	**23.9**	**21.0**
0.20	$n = 500$	1	3.8	3.8	5.3	5.3	n.a.	n.a.
		2	**2.4**	**2.3**	6.1	6.1	**18.9**	**18.9**
		3	**1.8**	**1.7**	4.9	4.6	**19.9**	**19.6**
		6	**0.9**	**0.9**	4.4	4.3	**26.5**	**26.2**
		12	**0.4**	**0.4**	3.7	**3.2**	**33.5**	**31.5**
		15	**0.1**	**0.1**	**3.3**	**3.0**	**35.4**	**33.8**
0.45	$n = 100$	1	**2.8**	**2.6**	**8.9**	**8.3**	n.a.	n.a.
		2	**2.5**	**2.2**	**6.9**	**6.5**	**12.1**	**11.4**
		3	**1.6**	**1.5**	5.0	4.1	**11.4**	**10.0**
		6	**1.6**	**1.2**	3.4	**2.2**	**10.9**	**8.4**
		12	**0.9**	**0.5**	**3.2**	**1.9**	**13.5**	**10.0**
		15	**0.9**	**0.3**	**2.2**	**0.8**	**14.3**	**9.0**
0.45	$n = 250$	1	**3.3**	**3.1**	**8.7**	**8.6**	n.a.	n.a.
		2	**3.3**	**3.1**	6.1	6.1	**16.8**	**16.2**
		3	**2.6**	**2.5**	4.3	4.2	**15.5**	**15.1**
		6	**1.0**	**0.9**	4.5	4.3	**19.0**	**18.0**
		12	**0.6**	**0.4**	3.9	**2.8**	**23.7**	**21.8**
		15	**0.4**	**0.3**	3.6	**2.5**	**24.5**	**21.6**
0.45	$n = 500$	1	3.6	**3.5**	**6.7**	**6.6**	n.a.	n.a.
		2	**2.4**	**2.3**	**6.9**	**6.8**	**20.0**	**20.0**
		3	**1.7**	**1.7**	5.4	5.2	**21.3**	**21.2**
		6	**1.0**	**0.9**	4.8	4.5	**26.9**	**26.4**
		12	**0.5**	**0.4**	3.7	**3.5**	**33.2**	**32.0**
		15	**0.1**	**0.1**	**3.5**	**3.1**	**36.3**	**34.8**

Table 3.19 – Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(0, d_0, 0)$ defined by (3.34)–(3.35) with $\theta_0 = (0, 0, d_0)$. The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$.

d_0	Length n	Lag m	LB_{SN}	BP_{SN}	LB_W	BP_W	LB_S	BP_S
0.05	$n = 100$	1	**2.2**	**2.1**	**20.0**	**19.5**	n.a.	n.a.
		2	**1.5**	**1.5**	**15.2**	**14.7**	**18.3**	**17.3**
		3	**1.1**	**0.9**	**10.7**	**10.1**	**15.3**	**14.4**
		6	**0.4**	**0.2**	6.0	5.2	**10.4**	**9.7**
		12	**0.0**	**0.0**	**3.2**	**2.5**	**8.2**	5.9
		15	**0.2**	**0.0**	**2.4**	**1.7**	**7.7**	5.0
0.05	$n = 250$	1	**3.2**	**2.9**	**14.4**	**14.2**	n.a.	n.a.
		2	**3.1**	**2.9**	**10.7**	**10.6**	**18.7**	**18.3**
		3	**1.9**	**1.8**	**7.8**	**7.6**	**16.3**	**16.0**
		6	**0.9**	**0.6**	4.5	4.2	**12.6**	**12.0**
		12	**0.4**	**0.3**	**2.0**	**1.5**	**10.6**	**8.8**
		15	**0.2**	**0.2**	**1.3**	**1.3**	**10.0**	**8.2**
0.05	$n = 500$	1	4.3	4.3	**11.7**	**11.6**	n.a.	n.a.
		2	3.7	3.7	**8.7**	**8.6**	**18.7**	**18.6**
		3	**2.9**	**2.7**	**6.5**	6.4	**16.7**	**16.6**
		6	**1.8**	**1.6**	**3.4**	**3.2**	**14.4**	**14.1**
		12	**0.3**	**0.2**	**2.2**	**1.7**	**10.9**	**10.4**
		15	**0.2**	**0.2**	**1.1**	**1.0**	**10.2**	**9.7**
0.20	$n = 100$	1	3.9	3.7	**11.9**	**11.3**	n.a.	n.a.
		2	**1.5**	**1.5**	**7.4**	**6.8**	**12.3**	**11.4**
		3	**1.4**	**1.4**	5.2	4.5	**10.7**	**9.6**
		6	**0.3**	**0.2**	**2.3**	**1.8**	**8.4**	**7.6**
		12	**0.1**	**0.0**	**1.1**	**0.8**	**6.5**	4.2
		15	**0.2**	**0.0**	**0.9**	**0.4**	5.8	**3.4**
0.20	$n = 250$	1	3.9	3.8	**7.1**	**6.9**	n.a.	n.a.
		2	3.6	**3.4**	6.1	5.7	**13.2**	**13.1**
		3	**1.9**	**1.8**	3.8	**3.4**	**11.7**	**11.3**
		6	**0.9**	**0.6**	**2.6**	**2.3**	**9.8**	**9.3**
		12	**0.3**	**0.3**	**1.0**	**0.6**	**8.8**	**7.6**
		15	**0.2**	**0.2**	**0.5**	**0.5**	**8.9**	**7.2**
0.20	$n = 500$	1	5.3	5.3	6.3	6.1	n.a.	n.a.
		2	4.0	3.9	5.4	5.3	**15.8**	**15.6**
		3	**3.3**	**3.3**	3.7	3.6	**12.9**	**12.9**
		6	**1.9**	**1.5**	**1.4**	**1.4**	**11.9**	**11.5**
		12	**0.2**	**0.1**	**1.2**	**0.9**	**9.8**	**9.2**
		15	**0.3**	**0.2**	**0.5**	**0.5**	**9.2**	**8.9**
0.45	$n = 100$	1	3.9	3.8	**21.5**	**20.2**	n.a.	n.a.
		2	**1.6**	**1.5**	**13.1**	**11.9**	**16.5**	**16.4**
		3	**1.2**	**0.9**	**7.5**	**7.2**	**13.7**	**12.7**
		6	**0.7**	**0.7**	**3.1**	**2.4**	**10.6**	**9.2**
		12	**0.1**	**0.0**	**1.3**	**0.8**	**6.9**	5.2
		15	**0.2**	**0.0**	**1.3**	**0.3**	6.2	3.8
0.45	$n = 250$	1	5.0	5.0	**15.7**	**15.5**	n.a.	n.a.
		2	**3.0**	**3.0**	**10.4**	**10.0**	**18.6**	**18.2**
		3	**2.3**	**2.3**	**7.5**	**7.3**	**16.1**	**15.9**
		6	**0.6**	**0.4**	3.6	3.6	**12.1**	**11.4**
		12	**0.4**	**0.3**	**1.5**	**1.1**	**9.7**	**8.6**
		15	**0.2**	**0.2**	**1.1**	**0.8**	**10.1**	**8.8**
0.45	$n = 500$	1	4.8	4.8	**12.5**	**12.5**	n.a.	n.a.
		2	4.2	4.0	**8.9**	**8.7**	**19.6**	**19.5**
		3	3.2	3.2	5.7	5.6	16.6	16.6
		6	**2.0**	**1.8**	**2.6**	**2.5**	**13.7**	**13.4**
		12	**0.1**	**0.1**	**1.5**	**1.1**	**10.8**	**10.3**
		15	0.3	0.2	0.6	0.6	10.4	10.1

3.8 Another one

In order the see if the test procedures remain reliable for GARCH process with infinite moment (for $\alpha_1 + \beta_1 \geq 1$), we replicate the numerical experiments made on Model (3.34)–(3.20) with $\omega = 0.04$, $\alpha_1 = 0.13$ and $\beta_1 = 0.88$.

Table 3.20 – Empirical size (in %) of the modified and standard versions of the LB and BP tests in the case of a weak FARIMA$(1, d_0, 1)$ defined by (3.34) with $\theta_0 = (0.9, 0.2, d_0)$ and where $\omega = 0.04$, $\alpha_1 = 0.12$ and $\beta_1 = 0.85$ in (3.20). The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1,000$.

d_0	Length n	Lag m	LB_{SN}	BP_{SN}	LB_W	BP_W	LB_S	BP_S
0.05	$n = 1,000$	1	**3.3**	**3.3**	**8.9**	**8.9**	n.a.	n.a.
		2	**2.5**	**2.5**	**7.5**	**7.5**	n.a.	n.a.
		3	**2.1**	**2.1**	5.4	5.3	n.a.	n.a.
		6	**1.1**	**1.0**	4.3	4.1	**38.4**	**38.1**
		12	**0.6**	**0.6**	3.7	**3.2**	**43.3**	**42.7**
		15	**0.2**	**0.2**	**3.5**	**3.5**	**45.7**	**44.6**
0.05	$n = 10,000$	1	4.4	4.4	5.4	5.4	n.a.	n.a.
		2	**3.5**	**3.5**	5.0	5.0	n.a.	n.a.
		3	**3.3**	**3.3**	6.3	6.3	n.a.	n.a.
		6	**2.3**	**2.2**	4.5	4.4	**57.7**	**57.6**
		12	**1.6**	**1.6**	4.0	3.9	**68.9**	**68.8**
		15	**1.3**	**1.3**	4.6	4.6	**73.8**	**73.8**
0.05	$n = 20,000$	1	4.6	4.6	5.2	5.2	n.a.	n.a.
		2	4.3	4.3	5.0	5.0	n.a.	n.a.
		3	3.9	3.9	4.8	4.8	n.a.	n.a.
		6	**2.5**	**2.5**	5.0	4.8	**41.7**	**41.7**
		12	3.8	3.8	4.2	4.2	**51.1**	**51.1**
		15	**3.3**	**3.2**	3.7	3.7	**52.8**	**52.7**
0.20	$n = 1,000$	1	**3.5**	**3.4**	**9.9**	**9.7**	n.a.	n.a.
		2	**2.4**	**2.4**	6.4	6.4	n.a.	n.a.
		3	**2.2**	**2.1**	4.9	4.9	n.a.	n.a.
		6	**1.1**	**0.8**	**3.5**	**3.4**	**37.4**	**37.2**
		12	**0.3**	**0.3**	**3.5**	**3.3**	**42.9**	**42.4**
		15	**0.0**	**0.0**	3.6	**3.5**	**44.4**	**43.2**
0.20	$n = 10,000$	1	4.2	4.2	4.0	4.0	n.a.	n.a.
		2	**3.4**	**3.4**	4.1	4.1	n.a.	n.a.
		3	**3.3**	**3.3**	5.3	5.3	n.a.	n.a.
		6	**2.2**	**2.2**	4.3	4.3	**55.8**	**55.8**
		12	**1.6**	**1.6**	3.9	3.9	**67.7**	**67.7**
		15	**1.3**	**1.3**	4.1	4.1	**72.9**	**72.9**
0.20	$n = 20,000$	1	5.0	5.0	4.3	4.3	n.a.	n.a.
		2	4.6	4.6	4.4	4.4	n.a.	n.a.
		3	3.9	3.9	4.7	4.7	n.a.	n.a.
		6	**2.7**	**2.7**	4.7	4.7	**41.0**	**41.0**
		12	3.7	3.7	4.0	4.0	**50.3**	**50.3**
		15	**3.4**	**3.4**	3.6	**3.5**	**51.9**	**51.8**
0.45	$n = 1,000$	1	**3.0**	**3.0**	**12.1**	**12.2**	n.a.	n.a.
		2	**1.8**	**1.8**	5.5	5.4	n.a.	n.a.
		3	**1.7**	**1.6**	4.4	4.4	n.a.	n.a.
		6	**0.6**	**0.6**	**3.4**	**3.3**	**34.7**	**34.4**
		12	**0.4**	**0.4**	**3.3**	**3.0**	**38.6**	**38.0**
		15	**0.2**	**0.2**	3.6	**3.5**	**40.0**	**38.9**
0.45	$n = 10,000$	1	3.7	3.6	3.7	3.7	n.a.	n.a.
		2	**3.0**	**3.0**	3.7	3.8	n.a.	n.a.
		3	**3.0**	**3.0**	5.1	5.1	n.a.	n.a.
		6	**2.0**	**2.0**	4.7	4.7	**55.3**	**55.3**
		12	**1.7**	**1.7**	3.8	3.8	**67.6**	**67.4**
		15	**1.3**	**1.3**	4.1	4.1	**72.0**	**71.8**
0.45	$n = 20,000$	1	5.0	5.0	4.1	4.1	n.a.	n.a.
		2	4.5	4.5	4.1	4.1	n.a.	n.a.
		3	3.7	3.7	4.8	4.8	n.a.	n.a.
		6	**2.9**	**2.9**	4.5	4.4	**40.5**	**40.5**
		12	3.7	3.7	3.8	3.8	**49.8**	**49.7**
		15	**3.5**	**3.5**	3.6	3.6	**51.3**	**51.3**

Bibliography

Baillie, R. T., Chung, C.-F., and Tieslau, M. A. (1996). Analysing inflation by the fractionally integrated ARFIMA-GARCH model. *Journal of applied econometrics*, 11(1):23–40.

Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2013). *Long-memory processes*. Springer, Heidelberg. Probabilistic properties and statistical methods.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *Ann. Statist.*, 2:489–502. Collection of articles dedicated to Jerzy Neyman on his 80th birthday.

Boubacar Maïnassara, Y. (2009). *Estimation, validation et identification des modèles ARMA faibles multivariés*. PHD thesis of Université Lille 3.

Boubacar Mainassara, Y., Carbon, M., and Francq, C. (2012). Computing and estimating information matrices of weak ARMA models. *Comput. Statist. Data Anal.*, 56(2):345–361.

Boubacar Maïnassara, Y., Esstafa, Y., and Saussereau, B. (2019). Estimating FARIMA models with uncorrelated but non-independent error terms. *Working document*.

Boubacar Maïnassara, Y. and Saussereau, B. (2018). Diagnostic Checking in Multivariate ARMA Models With Dependent Errors Using Normalized Residual Autocorrelations. *J. Amer. Statist. Assoc.*, 113(524):1813–1827.

Box, G. E. P. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Amer. Statist. Assoc.*, 65:1509–1526.

Brockwell, P. J. and Davis, R. A. (1991). *Time series: theory and methods*. Springer Series in Statistics. Springer-Verlag, New York, second edition.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.*, 17(4):1749–1766.

Davidson, J. (1994). *Stochastic limit theory*. Advanced Texts in Econometrics. The Clarendon Press, Oxford University Press, New York. An introduction for econometricians.

den Haan, W. J. and Levin, A. T. (1997). A practitioner's guide to robust covariance matrix estimation. In *Robust inference*, volume 15 of *Handbook of Statist.*, pages 299–342. North-Holland, Amsterdam.

Ding, Z., Granger, C. W., and Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of empirical finance*, 1:83–106.

Doukhan, P. and León, J. (1989). Cumulants for stationary mixing random sequences and applications to empirical spectral density. *Probab. Math. Stat*, 10:11–26.

Fan, J. and Yao, Q. (2008). *Nonlinear time series: nonparametric and parametric methods*. Springer Science & Business Media.

Fox, R. and Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.*, 14(2):517–532.

Francq, C., Roy, R., and Zakoïan, J.-M. (2005). Diagnostic Checking in ARMA Models With Uncorrelated Errors. *J. Amer. Statist. Assoc.*, 100(470):532–544.

Francq, C. and Zakoïan, J.-M. (1998). Estimating linear representations of nonlinear processes. *J. Statist. Plann. Inference*, 68(1):145–165.

Francq, C. and Zakoïan, J.-M. (2005). Recent results for linear time series models with non independent innovations. In *Statistical modeling and analysis for complex data problems*, volume 1 of *GERAD 25th Anniv. Ser.*, pages 241–265. Springer, New York.

Francq, C. and Zakoïan, J.-M. (2007). HAC estimation and strong linearity testing in weak ARMA models. *J. Multivariate Anal.*, 98(1):114–144.

Francq, C. and Zakoïan, J.-M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley.

Francq, C. and Zakoïan, J.-M. (2009). Bartlett's formula for a general class of nonlinear processes. *Journal of Time Series Analysis*, 30(4):449–465.

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal.*, 1(1):15–29.

Hallin, M., Taniguchi, M., Serroukh, A., and Choy, K. (1999). Local asymptotic normality for regression models with long-memory disturbance. *Ann. Statist.*, 27(6):2054–2080.

Herrndorf, N. (1984). A functional central limit theorem for weakly dependent sequences of random variables. *Ann. Probab.*, 12(1):141–153.

Hong, Y. (1996). Consistent testing for serial correlation of unknown form. *Econometrica*, 64(4):837–864.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68(1):165–176.

Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48:419–426.

Kuan, C.-M. and Lee, W.-M. (2006). Robust M tests without consistent estimation of the asymptotic covariance matrix. *J. Amer. Statist. Assoc.*, 101(475):1264–1275.

Li, W. K. and McLeod, A. I. (1986). Fractional time series modelling. *Biometrika*, 73(1):217–221.

Ling, S. and Li, W. K. (1997). On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *J. Amer. Statist. Assoc.*, 92(439):1184–1194.

Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2):pp. 297–303.

Lobato, I. N. (2001). Testing that a dependent process is uncorrelated. *J. Amer. Statist. Assoc.*, 96(455):1066–1076.

Lobato, I. N., Nankervis, J. C., and Savin, N. E. (2001). Testing for autocorrelation using a modified box- pierce q test. *Inter. Econ. Review*, 42(1):187–205.

Lobato, I. N., Nankervis, J. C., and Savin, N. E. (2002). Testing for zero autocorrelation in the presence of statistical dependence. *Econ. Theory*, 18(3):730–743.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer-Verlag, Berlin.

Magnus, J. R. and Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester. Revised reprint of the 1988 original.

Palma, W. (2007). *Long-memory time series*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ. Theory and methods.

Romano, J. P. and Thombs, L. A. (1996). Inference for autocorrelations under weak assumptions. *J. Amer. Statist. Assoc.*, 91(434):590–600.

Shao, X. (2010a). Corrigendum: A self-normalized approach to confidence interval construction in time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(5):695–696.

Shao, X. (2010b). A self-normalized approach to confidence interval construction in time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(3):343–366.

Shao, X. (2011). Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models. *Econometric Theory*, 27(2):312–343.

Shao, X. (2012). Parametric inference in stationary time series models with dependent errors. *Scand. J. Stat.*, 39(4):772–783.

Shao, X. (2015). Self-normalization for time series: a review of recent developments. *J. Amer. Statist. Assoc.*, 110(512):1797–1817.

Tong, H. (1990). *Non-linear time series: a dynamical system approach*. Oxford University Press.

Zhu, K. and Li, W. K. (2015). A bootstrapped spectral test for adequacy in weak ARMA models. *J. Econometrics*, 187(1):113–130.

Chapter 4

Generalized least squares estimation of fractional autoregressive models

Contents

4.1 Introduction **207**
4.2 Generalized least squares estimation of FrAR models of order 1 when the Hurst exponent is known **209**
4.2.1 Statement of the problem and notations 209
4.2.2 Asymptotic properties 211
4.3 Numerical illustrations **211**
4.3.1 Simulation studies 212
4.4 Proofs **212**
4.4.1 Preliminary results 212
4.4.2 Proof of Theorem 4.1 216
4.4.3 Proof of Theorem 4.2 228
4.5 Figures **234**
Bibliographie **241**

Abstract: In this paper we derive the asymptotic properties of the generalized least squares estimator (GLSE) of autoregressive models of order 1 endowed with fractional Gaussian noise (the so-called fractional autoregressive models). We establish the consistency and the asymptotic normality of the GLSE. Some simulation studies are presented to corroborate our theoretical work.

AMS 2000 subject classifications: Primary 62M10; secondary 91B84.
Keywords and phrases: Fractional Brownian motion; Fractional Gaussian noise; FrAR models; Generalized least squares estimator; Consistency; Asymptotic normality.

4.1 Introduction

The error term in the classical autoregressive models is usually assumed to be a strong white noise (namely an independent and identically distributed (iid for short) sequence of random variables with mean 0 and common variance), a martingale difference or a stationary process with a bounded spectral density (see Anderson and Taylor [1976], Anderson and Taylor [1979], Lai and Robbins [1977], Lai et al. [1978], Lai et al. [1979] and Solo [1981]). In the late 1990s, C. Francq and J.-M. Zakoïan have released the assumption of the independence on the noise and have considered the problem of estimating ARMA models with uncorrelated but non-independent error term, the so-called weak ARMA models (see Francq and Zakoïan [1998]). This extension allowed to ARMA models to be less restrictive and to cover a large class of general nonlinear models. Nevertheless, the uncorrelatedness of the noise assumed in these classes of models considerably restricts their application to short-memory stochastic processes.

Yet in many scientific disciplines and applied fields such as hydrology, climatology, economics, finance, to name a few, time series exhibit a long-range dependence (see for instance Granger and Joyeux [1980], Fox and Taqqu [1986], Dahlhaus [1989], Hosking [1981], Beran et al. [2013], Palma [2007], among others), that is dependence between distant observations is so strong that classical or weak ARMA models are unable to identify and then to take into consideration.

To model the long memory phenomenon, a widely used models is the fractional Gaussian noise (see Mandelbrot and Van Ness [1968]) and the fractional autoregressive integrated moving average (FARIMA, for short) models (see Granger and Joyeux [1980] and Hosking [1981]). We shall consider in this work the problem of generalized least squares estimation of autoregressive models endowed with fractional Gaussian noise. More explicitly, we shall study the asymptotic properties of the generalized least squares estimator of the parameters of the centered stationary fractional autoregressive process of order 1 (FrAR(1)) $(X_t)_{t\in\mathbb{Z}}$ defined, for all $t \in \mathbb{Z}$, by

$$X_t = a_0 X_{t-1} + \epsilon_t^{\mathsf{H}}, \tag{4.1}$$

where $(\epsilon_t^{\mathsf{H}})_{t\in\mathbb{Z}}$ is a fractional Gaussian noise and where the autoregressive parameter a_0 and the Hurst index or Hurst parameter H associated with the fractional Gaussian noise $(\epsilon_t^{\mathsf{H}})_{t\in\mathbb{Z}}$ are assumed to satisfy the following standard assumption:

(A0): $a_0 \in]-1,1[$ and $\mathsf{H} \in]0,1[$.

The condition $a_0 \in]-1,1[$ ensures the causality and the second-order stationarity of the process $(X_t)_{t\in\mathbb{Z}}$. In fact, for AR(p) models a causal weak-sense stationary solution is obtained once the roots of the corresponding autoregressive polynomial lie outside the closed unit disk.

The process $(\epsilon_t^{\mathsf{H}})_{t\in\mathbb{Z}}$ is the increment process of the fractional Brownian motion $(B_{\mathsf{H}}(t))_{t\in\mathbb{Z}}$ of Hurst index H, namely, for any $t \in \mathbb{Z}$, it holds

$$\epsilon_t^{\mathsf{H}} = B_{\mathsf{H}}(t+1) - B_{\mathsf{H}}(t).$$

Based on the definition and the properties of the fractional Brownian motion $(B_{\mathsf{H}}(t))_{t\in\mathbb{Z}}$, the process $(\epsilon_t^{\mathsf{H}})_{t\in\mathbb{Z}}$ is the unique continuous stationary Gaussian centered process with autocovari-

ance function defined for $k \in \mathbb{Z}$ by:

$$\gamma_{\epsilon^H}(k) = \frac{1}{2}\left(|k-1|^{2H} - 2|k|^{2H} + |k+1|^{2H}\right). \tag{4.2}$$

A Taylor expansion of the function $\ell : x \to (1-x)^{2H} - 2 + (1+x)^{2H}$ around 0 implies that for large k,

$$\gamma_{\epsilon^H}(k) = \frac{1}{2}k^{2H}\ell(1/k) = H(2H-1)k^{2H-2} + o(k^{2H-2}).$$

The last result implies that $\gamma_{\epsilon^H}(k) \to 0$ as $k \to \infty$. So the process $(\epsilon_t^H)_{t\in\mathbb{Z}}$ is mixing (see Ito [1944]). Mixing is a stronger property than ergodicity: this result is known as the Khinchin's theorem (see Khinchin [1949]). Hence $\lim_{k\to\infty} \gamma_{\epsilon^H}(k) = 0$ implies that the process $(\epsilon_t^H)_{t\in\mathbb{Z}}$ is ergodic.

When $1/2 < H < 1$, the series of the autocovariances of $(\epsilon_t^H)_{t\in\mathbb{Z}}$ is absolutely divergent and thus the process $(\epsilon_t^H)_{t\in\mathbb{Z}}$ exhibits long-range dependence.

In the literature, many approaches have been proposed to estimate the parameters of strongly dependent Gaussian processes. We cite for example the R/S method, periodogram estimation and maximum likelihood estimation. Mandelbrot [1975] and Mandelbrot and Taqqu [1979] have established the theoretical properties of the R/S estimates. Mohr [1981], Graf [1980], Geweke and Porter-Hudak [1983], Fox and Taqqu [1986] and Dahlhaus [1989] have considered the problem of periodogram estimation. McLeod and Hipel [1978] have discussed computational considerations involved in the application of maximum likelihood estimation. Sweeting [1980] has proposed assumptions to obtain the consistency and the asymptotic normality of maximum likelihood estimator. Dahlhaus [1989] has established the asymptotic properties of the exact maximum likelihood estimation of the parameters of strongly dependent Gaussien processes.

In the particular case of autoregressive models endowed with a strongly dependent noise, Brouste et al. [2014] have proposed a new approach based on a transformation of the initial model to establish the asymptotic properties of the maximum likelihood estimator. See also Yajima [1988] for the case of a regression problem with nonstochastic regressors.

We study in this paper the asymptotic properties of the generalized least squares estimator of autoregressive models endowed with fractional Gaussian noise. We propose a simple and natural transformation of the initial model and we establish the convergence and asymptotic normality of the generalized least squares estimator based on this transformation.

In all this work, x^T designates the transpose of the vector x. We shall also use the matrix norm defined by $\|A\| = \sup_{\|x\|_{\mathbb{R}^{k_2}}\leq 1} \|Ax\|_{\mathbb{R}^{k_1}} = \rho^{1/2}(A^TA)$, when A is a $\mathbb{R}^{k_1\times k_2}$ matrix, $\|x\|^2_{\mathbb{R}^{k_2}} = x^Tx$ is the Euclidean norm of the vector $x \in \mathbb{R}^{k_2}$, and $\rho(\cdot)$ denotes the spectral radius.

4.2 Generalized least squares estimation of FrAR models of order 1 when the Hurst exponent is known

In this section we present the parametrization that is used in the sequel. Then we state the asymptotic properties of the GLSE of FrAR(1) when the Hurst exponent H is known.

4.2.1 Statement of the problem and notations

For all $a \in]-1,1[$, let $(\epsilon_t^{\mathsf{H}}(a))_{t\in\mathbb{Z}}$ be the second order stationary fractional Gaussian noise defined as the solution of

$$\epsilon_t^{\mathsf{H}}(a) = X_t - aX_{t-1}, \quad \forall t \in \mathbb{Z}. \tag{4.3}$$

Observe that the distribution of $\epsilon_t^{\mathsf{H}}(a)$ does not depend on a. Note also that, for all $t \in \mathbb{Z}$, $\epsilon_t^{\mathsf{H}}(a_0) = \epsilon_t^{\mathsf{H}}$ a.s. Given a realization $X_1, \ldots, X_{n+1}$ of length $n+1$, $\epsilon_t^{\mathsf{H}}(a)$ can be approximated, for $0 < t \leq n+1$, by $e_t^{\mathsf{H}}(a)$ defined recursively by

$$e_t^{\mathsf{H}}(a) = X_t - aX_{t-1}, \tag{4.4}$$

with $e_t^{\mathsf{H}}(a) = X_t = 0$ if $t \leq 0$. It is clear that these initial values are asymptotically negligible uniformly in a. In fact, $\epsilon_t^{\mathsf{H}}(a) - e_t^{\mathsf{H}}(a) = 0$ for all $t \geq 2$. Thus the choice of the initial values has no influence on the asymptotic properties of the model parameter estimator.

From (4.4), one can write

$$\begin{pmatrix} X_2 \\ \vdots \\ X_{n+1} \end{pmatrix} = a \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} + \begin{pmatrix} e_2^{\mathsf{H}}(a) \\ \vdots \\ e_{n+1}^{\mathsf{H}}(a) \end{pmatrix}. \tag{4.5}$$

By construction, the random vector $(e_2^{\mathsf{H}}(a), \ldots, e_{n+1}^{\mathsf{H}}(a))^{\mathsf{T}}$ in equation (4.5) is equal to the vector $(\epsilon_2^{\mathsf{H}}(a), \ldots \epsilon_{n+1}^{\mathsf{H}}(a))^{\mathsf{T}}$, where the random process $(\epsilon_t^{\mathsf{H}}(a))_{t\in\mathbb{Z}}$ represents the theoretical noise introduced in (4.3). By exploiting the properties of the fractional Gaussian noise $(\epsilon_t^{\mathsf{H}}(a))_{t\in\mathbb{Z}}$, one can emphasize that the random vector $(e_2^{\mathsf{H}}(a), \ldots, e_{n+1}^{\mathsf{H}}(a))^{\mathsf{T}}$ is a centered normal random vector with covariance matrix $\Omega_{n,\mathsf{H}} := [\gamma_{\epsilon^{\mathsf{H}}}(j-i)]_{1\leq i,j\leq n} = [\mathrm{Cov}(\epsilon_i^{\mathsf{H}}, \epsilon_j^{\mathsf{H}})]_{1\leq i,j\leq n}$.

Remark 4.1. *The covariance matrix $\Omega_{n,\mathsf{H}}$ depends only on the known Hurst parameter* H. *As a symmetric real matrix, $\Omega_{n,\mathsf{H}}$ can be diagonalized by an orthogonal matrix. More explicitly, there exists a real orthogonal matrix $P_{n,\mathsf{H}}$ such that $D_{n,\mathsf{H}} = P_{n,\mathsf{H}}^{\mathsf{T}}\Omega_{n,\mathsf{H}}P_{n,\mathsf{H}}$ is a diagonal matrix. By taking into consideration the positive definiteness of the matrix $\Omega_{n,\mathsf{H}}$ (see Proposition 4.1 hereafter), we have $\Omega_{n,\mathsf{H}}^{1/2} := P_{n,\mathsf{H}}D_{n,\mathsf{H}}^{1/2}P_{n,\mathsf{H}}^{\mathsf{T}}$ where $D_{n,\mathsf{H}}^{1/2} = \mathrm{diag}((\lambda_{1,\mathsf{H}}^{(n)})^{1/2}, \ldots, (\lambda_{n,\mathsf{H}}^{(n)})^{1/2})$ and the $\lambda_{i,\mathsf{H}}^{(n)}$ are the eigenvalues of $\Omega_{n,\mathsf{H}}$. The matrix $\Omega_{n,\mathsf{H}}^{1/2}$ is a square root of $\Omega_{n,\mathsf{H}}$ since it satisfies $\Omega_{n,\mathsf{H}}^{1/2}\Omega_{n,\mathsf{H}}^{1/2} = \Omega_{n,\mathsf{H}}$.*

Remark 4.2. *The ordinary least squares estimator of the parameter a_0 is biased. To see that, let us denote by $\hat{a}_n^{(\mathrm{OLS})}$ the standard least squares estimator of a_0. It is well known that*

$$\hat{a}_n^{(\mathrm{OLS})} = \frac{\sum_{t=1}^n X_tX_{t-1}}{\sum_{t=1}^n X_{t-1}^2}.$$

Thanks to (4.1)*, one has*

$$\hat{a}_n^{(\mathrm{OLS})} - a_0 = \frac{\sum_{t=1}^n \epsilon_t^{\mathsf{H}} X_{t-1}}{\sum_{t=1}^n X_{t-1}^2}.$$

The ergodicity of $(\epsilon_t^{\mathsf{H}})_{t\in\mathbb{Z}}$ *and the stationarity of* $(X_t)_{t\in\mathbb{Z}}$ *imply that*

$$\hat{a}_n^{(\mathrm{OLS})} - a_0 \xrightarrow[n\to\infty]{\text{a.s.}} \frac{\mathbb{E}\left[\epsilon_2^{\mathsf{H}} X_1\right]}{\mathrm{Var}\,(X_1)}.$$

From (4.2) *and* (4.13)*, one can easily obtain that*

$$\mathbb{E}\left[\epsilon_2^{\mathsf{H}} X_1\right] = \frac{a_0 - 2}{2} + \frac{(a_0-1)^2}{2} \sum_{j\geq 2} a_0^{j-2} j^{2\mathsf{H}}.$$

Since the function $\mathsf{H} \in]0,1[\to \sum_{j\geq 2} a_0^{j-2} j^{2\mathsf{H}}$ *is bijective, we deduce that* $\mathbb{E}[\epsilon_2^{\mathsf{H}} X_1] = 0$ *if and only if* $\mathsf{H} = 1/2$*. This leads us to the standard case of an autoregressive model where the noise is an iid sequence with standard normal distribution.*

Due to the bias problem of the ordinary least squares estimator of the parameter a_0 in model (4.1) (problem caused by the correlation of the noise), we consider here the generalized least squares estimation procedure. This technique consists in transforming model (4.4) to another autoregressive model, with the same unknown autoregressive parameter a_0, endowed with a standard white noise (*i.e.* an independent and standard normal distributed sequence of random variables). Then we consider the ordinary least squares estimation on the new model that we obtained.

In view of Remark 4.1, the vector $\Omega_{n,\mathsf{H}}^{-1/2}(e_2^{\mathsf{H}}(a), \ldots, e_{n+1}^{\mathsf{H}}(a))^{\mathsf{T}}$ is a standard normal random vector. Thus, a rather natural and simple transformation that can be proposed to obtain this standard intermediate autoregressive model is to consider a vector representation of the model (4.4) and to multiply all the terms of the obtained equation by the appropriate matrix $\Omega_{n,\mathsf{H}}^{-1/2}$. More precisely, let

$$Y_{n,\mathsf{H}} = \Omega_{n,\mathsf{H}}^{-1/2} \begin{pmatrix} X_2 \\ \vdots \\ X_{n+1} \end{pmatrix}, \quad Z_{n,\mathsf{H}} = \Omega_{n,\mathsf{H}}^{-1/2} \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \text{ and } U_{n,\mathsf{H}}(a) = \Omega_{n,\mathsf{H}}^{-1/2} \begin{pmatrix} e_2^{\mathsf{H}}(a) \\ \vdots \\ e_{n+1}^{\mathsf{H}}(a) \end{pmatrix}. \tag{4.6}$$

Then, from (4.4) we have

$$Y_{n,\mathsf{H}} = a Z_{n,\mathsf{H}} + U_{n,\mathsf{H}}(a). \tag{4.7}$$

The model (4.7) presents a theoritical standard autoregressive model associated to the starting model (4.1). The advantage of this model is the structure of its noise, the error term of the model (4.7) is a standard normal vector. Hence, to estimate the unknown parameter a_0, one uses this model and we apply the ordinary least squares estimation procedure to obtain a consistent and asymptotically normally distributed estimator.

The random variable $\hat{a}_n$ is called generalized least squares estimator if it satisfies, almost surely,

$$\hat{a}_n = \underset{a\in]-1,1[}{\operatorname{argmin}} \frac{1}{n} \|U_{n,\mathrm{H}}(a)\|^2_{\mathbb{R}^n}. \tag{4.8}$$

The resolution of the optimization problem presented in (4.8) allows to obtain the explicit expression of $\hat{a}_n$. One can easily prove that

$$\hat{a}_n = \frac{Z^{\mathsf{T}}_{n,\mathrm{H}} Y_{n,\mathrm{H}}}{\|Z_{n,\mathrm{H}}\|^2_{\mathbb{R}^n}}. \tag{4.9}$$

Moreover, in view of this last equality and Equation (4.7), one can deduce that

$$\hat{a}_n - a_0 = \frac{Z^{\mathsf{T}}_{n,\mathrm{H}} U_{n,\mathrm{H}}(a_0)}{\|Z_{n,\mathrm{H}}\|^2_{\mathbb{R}^n}}. \tag{4.10}$$

4.2.2 Asymptotic properties

The asymptotic properties of the generalized least squares estimator of the fractional autoregressive model of order 1 when the Hurst parameter H is known are stated in the following two theorems.

Theorem 4.1. *(Consistency). We assume that $(X_t)_{t\in\mathbb{Z}}$ satisfies (4.1). Under **(A0)**, we have*

$$\hat{a}_n \xrightarrow[n\to\infty]{\mathbb{P}} a_0,$$

where $(\hat{a}_n)_{n\geq 1}$ is a sequence of generalized least squares estimators as defined in (4.9).

The proof of this theorem is given in Subsection 4.4.2.

Theorem 4.2. *(Asymptotic normality). We assume that $(X_t)_{t\in\mathbb{Z}}$ satisfies (4.1). Under **(A0)**, the sequence $(\sqrt{n}(\hat{a}_n - a_0))_{n\geq 1}$ has a limiting centered normal distribution with variance $1 - a_0^2$.*

The proof of this theorem is given in Subsection 4.4.3.

4.3 Numerical illustrations

In this section, we investigate the finite sample properties of the asymptotic results that we introduced in this work. For that sake we use Monte Carlo experiments. The numerical illustrations of this section are made with the open source statistical software R (see R Development Core Team, 2017) or (see http://cran.r-project.org/).

4.3.1 Simulation studies

We study numerically the behavior of the generalized least squares estimator for fractional autoregressive models of the form

$$X_t = aX_{t-1} + \epsilon_t^{\mathsf{H}}, \tag{4.11}$$

where the unknown autoregressive parameter is taken as $a = 0.3$, and where $(\epsilon_t^{\mathsf{H}})_{t\in\mathbb{Z}}$ is a fractional Gaussian noise with Hurst index $\mathsf{H} = 0.7$.

We simulated $N = 2,000$ independent trajectories of sizes $n = 30$, $n = 500$ and $n = 2,000$ of Model (4.11).

Figures 4.1, 4.2 and 4.3 illustrate the distribution of the generalized least squares estimator $\hat{a}_n$ of the autoregressive parameter a. We can notice that these numerical results are consistent with the convergence established in Theorem 4.2.

In Figures 4.4, 4.5 and 4.6, we corroborate the convergence of the estimator $1 - \hat{a}_n^2$ of the asymptotic variance of the generalized least squares estimator $\hat{a}_n$ given in Theorem 4.2.

4.4 Proofs

In all our proofs, K is a positive constant that may vary from line to line.

4.4.1 Preliminary results

In this subsection, we state some results that will be useful for the proofs of the asymptotic properties of the generalized least squares estimator introduced in Section 4.2. We exploit **(A0)** to obtain the expression of $Z_{n,\mathsf{H}}$ as a function of $U_{n,\mathsf{H}}(a_0)$. This expression will provide very intuitive demonstrations of the asymptotic properties of the GLSE.
We also give the asymptotic behavior of the components of the inverse variance matrix of the fractional Gaussian noise $(\epsilon_t^{\mathsf{H}})_{t\in\mathbb{Z}}$ based on the behavior in a neighbourhood of zero of the spectral density of this process.

The expression of $Z_{n,\mathsf{H}}$ as a function of $U_{n,\mathsf{H}}(a_0)$

We recall that, when $|a| < 1$, the polynomial $a(z) = 1 - az$ is invertible and we have

$$a^{-1}(z) = \sum_{j\geq 0} a^j z^j. \tag{4.12}$$

This implies that under Assumption **(A0)**, one may write the process $(X_t)_{t\in\mathbb{Z}}$ as a linear combination of the infinite past $(\epsilon_{t-j}^{\mathsf{H}})_{j\geq 0}$. More explicitly, we have from (4.1) that for any

$t \in \mathbb{Z}$,

$$X_t = (1 - a_0 L)^{-1} \epsilon_t^{\mathsf{H}} = \sum_{j \geq 0} a_0^j \epsilon_{t-j}^{\mathsf{H}}. \tag{4.13}$$

In view of the definition of the process $(e_t^{\mathsf{H}}(a))_{t \in \mathbb{Z}}$ introduced in (4.4) and (4.12), one can express X_t when $1 \leq t \leq n+1$ as a function of the finite past $(e_{t-j}^{\mathsf{H}}(a_0))_{0 \leq j \leq t-1}$. Indeed, Equation (4.4) and Assumption **(A0)** imply that

$$X_t = (1 - a_0 L)^{-1} e_t^{\mathsf{H}}(a_0) = \sum_{j=0}^{t-1} a_0^j e_{t-j}^{\mathsf{H}}(a_0), \quad \forall t \in \{1, \ldots, n+1\}. \tag{4.14}$$

From (4.14), we deduce the vector $(X_1, \ldots, X_n)^{\mathsf{T}}$ can be rewritten in the form

$$\begin{aligned}
\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} &= \begin{pmatrix} \sum_{j=0}^{1-1} a_0^j e_{1-j}^{\mathsf{H}}(a_0) \\ \sum_{j=0}^{2-1} a_0^j e_{2-j}^{\mathsf{H}}(a_0) \\ \vdots \\ \sum_{j=0}^{n-1} a_0^j e_{n-j}^{\mathsf{H}}(a_0) \end{pmatrix} = \begin{pmatrix} e_1^{\mathsf{H}}(a_0) \\ e_2^{\mathsf{H}}(a_0) + a_0 e_1^{\mathsf{H}}(a_0) \\ \vdots \\ e_n^{\mathsf{H}}(a_0) + \cdots + a_0^{n-1} e_1^{\mathsf{H}}(a_0) \end{pmatrix} \\
&= \begin{pmatrix} e_1^{\mathsf{H}}(a_0) \\ e_2^{\mathsf{H}}(a_0) \\ \vdots \\ e_n^{\mathsf{H}}(a_0) \end{pmatrix} + a_0 \begin{pmatrix} e_0^{\mathsf{H}}(a_0) \\ e_1^{\mathsf{H}}(a_0) \\ \vdots \\ e_{n-1}^{\mathsf{H}}(a_0) \end{pmatrix} + \cdots + a_0^{n-1} \begin{pmatrix} e_{2-n}^{\mathsf{H}}(a_0) \\ e_{3-n}^{\mathsf{H}}(a_0) \\ \vdots \\ e_1^{\mathsf{H}}(a_0) \end{pmatrix} \\
&= \sum_{j=0}^{n-1} a_0^j \begin{pmatrix} e_{1-j}^{\mathsf{H}}(a_0) \\ e_{2-j}^{\mathsf{H}}(a_0) \\ \vdots \\ e_{n-j}^{\mathsf{H}}(a_0) \end{pmatrix} = \sum_{j=0}^{n-1} a_0^j \left[L^{j+1} \begin{pmatrix} e_2^{\mathsf{H}}(a_0) \\ e_3^{\mathsf{H}}(a_0) \\ \vdots \\ e_{n+1}^{\mathsf{H}}(a_0) \end{pmatrix} \right],
\end{aligned} \tag{4.15}$$

where the shift operator L^{j+1} acts on all the components of the vector $(e_2^{\mathsf{H}}(a_0), \ldots, e_{n+1}^{\mathsf{H}}(a_0))^{\mathsf{T}}$, *i.e.*

$$L^{j+1} \begin{pmatrix} e_2^{\mathsf{H}}(a_0) \\ \vdots \\ e_{n+1}^{\mathsf{H}}(a_0) \end{pmatrix} = \begin{pmatrix} e_{1-j}^{\mathsf{H}}(a_0) \\ \vdots \\ e_{n-j}^{\mathsf{H}}(a_0) \end{pmatrix}.$$

In view of (4.15), it follows that the vector $Z_{n,\mathsf{H}}$ defined in (4.6) can be expressed as a function of $U_{n,\mathsf{H}}(a_0)$:

$$Z_{n,\mathsf{H}} = \Omega_{n,\mathsf{H}}^{-1/2} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = \sum_{j=0}^{n-1} a_0^j \left[L^{j+1} \Omega_{n,\mathsf{H}}^{-1/2} \begin{pmatrix} e_2^{\mathsf{H}}(a_0) \\ e_3^{\mathsf{H}}(a_0) \\ \vdots \\ e_{n+1}^{\mathsf{H}}(a_0) \end{pmatrix} \right] = \sum_{j=1}^{n} a_0^{j-1} \left(L^j U_{n,\mathsf{H}}(a_0) \right). \tag{4.16}$$

Asymptotic properties of the components of the inverse covariance matrix of $(\epsilon_t^{\mathsf{H}})_{1\le t\le n}$

By Fox and Taqqu [1986], Sinai [1976], the spectral density $f_{\epsilon^{\mathsf{H}}}(\cdot)$ of the fractional Gaussian noise $(\epsilon_t^{\mathsf{H}})_{t\in\mathbb{Z}}$ is defined for all $\omega\in[-\pi,\pi]$ by:

$$\begin{aligned} f_{\epsilon^{\mathsf{H}}}(\omega) &= \frac{1}{2\pi}\sum_{h\in\mathbb{Z}}\gamma_{\epsilon^{\mathsf{H}}}(h)\mathrm{e}^{-i\omega h} \\ &= C_{\mathsf{H}}(1-\cos(\omega))\sum_{j\in\mathbb{Z}}\frac{1}{|\omega+2j\pi|^{1+2\mathsf{H}}} \\ &= C_{\mathsf{H}}(1-\cos(\omega))\,|\omega|^{-1-2\mathsf{H}} + C_{\mathsf{H}}(1-\cos(\omega))\sum_{\substack{j\in\mathbb{Z}\\ j\neq 0}}\frac{1}{|\omega+2j\pi|^{1+2\mathsf{H}}}, \end{aligned}$$

where $C_{\mathsf{H}}=2\Gamma(2\mathsf{H}+1)\sin(\pi\mathsf{H})$ and $\Gamma(\cdot)$ denotes the Gamma function.
As $\omega\to 0$, we have

$$f_{\epsilon^{\mathsf{H}}}(\omega)\sim\frac{C_{\mathsf{H}}}{2}\,|\omega|^{1-2\mathsf{H}}. \tag{4.17}$$

Proposition 4.1. *Under Assumption **(A0)**, the following three assertions hold true.*

1. *The covariance matrix $\Omega_{n,\mathsf{H}}$ of the centered Gaussian random vector $(\epsilon_1^{\mathsf{H}},\ldots,\epsilon_n^{\mathsf{H}})^{\mathsf{T}}$ is a symmetric Toeplitz positive definite matrix.*
2. *For large k, the behavior of the autocovariance $\gamma_{\epsilon^{\mathsf{H}}}(k)$ is given by:*

$$(\Omega_{n,\mathsf{H}})_{1,k}=\gamma_{\epsilon^{\mathsf{H}}}(k)=\mathsf{H}(2\mathsf{H}-1)k^{2\mathsf{H}-2}+\mathrm{o}(k^{2\mathsf{H}-2}). \tag{4.18}$$

3. *The inverse autocovariances of $(\epsilon_t^{\mathsf{H}})_{1\le t\le n}$ satisfy*

$$\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{j,k}\sim\left(\frac{1}{k-j}\right)^{2\mathsf{H}}\frac{\Gamma(2\mathsf{H})\mathrm{e}^{i\pi\mathsf{H}}}{C_{\mathsf{H}}\pi^2}\ \text{ when } |k-j|\to\infty. \tag{4.19}$$

Proof. To prove the first point, let us observe at first that $\Omega_{n,\mathsf{H}}$ can be considered as a self-adjoint endomorphism of the vector space $\mathbb{C}^n$. Note now that for any $C\in\mathbb{C}^n\setminus\{0_{\mathbb{C}^n}\}$, the spectral representation of $\Omega_{n,\mathsf{H}}$ implies that

$$C^{\mathsf{T}}\Omega_{n,\mathsf{H}}C=\sum_{j,k=1}^{n}C_k\left(\Omega_{n,\mathsf{H}}\right)_{k,j}C_j=\sum_{j,k=1}^{n}C_k\left(\int_{-\pi}^{\pi}f_{\epsilon^{\mathsf{H}}}(\omega)\mathrm{e}^{i(j-k)\omega}\mathrm{d}\omega\right)C_j=\int_{-\pi}^{\pi}\left|\sum_{j=1}^{n}C_j\mathrm{e}^{ij\omega}\right|^2 f_{\epsilon^{\mathsf{H}}}(\omega)\mathrm{d}\omega.$$

The function $\omega\to\sum_{j=1}^{n}C_j\mathrm{e}^{ij\omega}$ is an analytic function on the disc and is not identically zero. Since $f_{\epsilon^{\mathsf{H}}}(\omega)>0$ for $\omega\neq 0$ the last integral is positive. Thus we conclude that the symmetric Toeplitz matrix $\Omega_{n,\mathsf{H}}$ is positive definite.

In view of (4.2), a Taylor expansion of the function $\ell: x\to(1-x)^{2\mathsf{H}}-2+(1+x)^{2\mathsf{H}}$ around 0 implies that for large k,

$$(\Omega_{n,\mathsf{H}})_{1,k}=\gamma_{\epsilon^{\mathsf{H}}}(k)=\frac{1}{2}k^{2\mathsf{H}}\ell(1/k)=\mathsf{H}(2\mathsf{H}-1)k^{2\mathsf{H}-2}+\mathrm{o}(k^{2\mathsf{H}-2}).$$

This gives the second point of the proposition.

The components of the inverse matrix of $\Omega_{n,\mathsf{H}}$ can also be expressed as a function of the spectral density of $(\epsilon_t^{\mathsf{H}})_{t\in\mathbb{Z}}$ (see Fox and Taqqu [1986]). More precisely we have that for all $j,k=1,\ldots,n$,

$$\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{j,k}=\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\frac{1}{f_{\epsilon^{\mathsf{H}}}(\omega)}\mathrm{e}^{i(k-j)\omega}\mathrm{d}\omega. \tag{4.20}$$

Using the parity of the function $f_{\epsilon^{\mathsf{H}}}(\cdot)$, observe that when $j=k$ we have

$$\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{j,j}=\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\frac{1}{f_{\epsilon^{\mathsf{H}}}(\omega)}\mathrm{d}\omega=\frac{1}{2\pi^2}\int_{0}^{\pi}\frac{1}{f_{\epsilon^{\mathsf{H}}}(\omega)}\mathrm{d}\omega. \tag{4.21}$$

From equivalence in (4.17), one has when $\omega\to 0$ that

$$\frac{1}{f_{\epsilon^{\mathsf{H}}}(\omega)}=\frac{2}{C_{\mathsf{H}}}\left|\omega\right|^{2\mathsf{H}-1}+\mathrm{o}\left(\frac{2}{C_{\mathsf{H}}}\left|\omega\right|^{2\mathsf{H}-1}\right).$$

This implies that for $\kappa>0$ there exits $\delta_\kappa>0$ such that for any $\omega\in]-\delta_\kappa,\delta_\kappa[$ we have

$$(1-\kappa)\frac{2}{C_{\mathsf{H}}}\left|\omega\right|^{2\mathsf{H}-1}\leq\frac{1}{f_{\epsilon^{\mathsf{H}}}(\omega)}\leq(1+\kappa)\frac{2}{C_{\mathsf{H}}}\left|\omega\right|^{2\mathsf{H}-1}. \tag{4.22}$$

Thanks to (4.21), (4.22) and the boundedness of the function $\omega\to 1/f_{\epsilon^{\mathsf{H}}}(\omega)$ on $]\delta_\kappa,\pi]$, one can show that there exists a positive constant K such that for any $j=1,\ldots,n$,

$$\begin{aligned}\left|\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{j,j}\right|&\leq\frac{1+\kappa}{C_{\mathsf{H}}\pi^2}\int_0^{\delta_\kappa}\omega^{2\mathsf{H}-1}\mathrm{d}\omega+\frac{1}{2\pi^2}\int_{\delta_\kappa}^{\pi}\frac{1}{f_{\epsilon^{\mathsf{H}}}(\omega)}\mathrm{d}\omega\\&\leq\frac{\delta_\kappa^{2\mathsf{H}}(1+\kappa)}{2\mathsf{H}C_{\mathsf{H}}\pi^2}+\frac{\pi-\delta_\kappa}{2\pi^2}\sup_{\omega\in]\delta_\kappa,\pi]}\frac{1}{f_{\epsilon^{\mathsf{H}}}(\omega)}\\&\leq K.\end{aligned} \tag{4.23}$$

When $j\neq k$ (suppose without loss of generality that $k>j$), one can use Lemma 6.3.2 (page 442) of Ablowitz and Fokas [2003] which is an analog version of Watson's Lemma (see for instance Watson [1918] for the original proof or Miller [2006] for a more recent development) to obtain the asymptotic equivalent of the (j,k)-th component of $\Omega_{n,\mathsf{H}}^{-1}$ when $|k-j|\to\infty$ stated in the third point of the proposition.

□

Remark 4.3. *From (4.18), for any $\rho>0$ there exists $N_\rho\in\mathbb{N}$ such that for any $k\geq N_\rho$ we have*

$$\mathsf{H}\left\{(2\mathsf{H}-1)-\rho\left|2\mathsf{H}-1\right|\right\}k^{2\mathsf{H}-2}\leq\left(\Omega_{n,\mathsf{H}}\right)_{1,k}\leq\mathsf{H}\left\{\rho\left|2\mathsf{H}-1\right|+(2\mathsf{H}-1)\right\}k^{2\mathsf{H}-2}.$$

Since $-1<2\mathsf{H}-1<1$ we obtain that

$$\left|\left(\Omega_{n,\mathsf{H}}\right)_{1,k}\right|\leq\mathsf{H}\left\{\rho\left|2\mathsf{H}-1\right|+1\right\}k^{2\mathsf{H}-2}. \tag{4.24}$$

Observe that for $k = 1, \ldots, N_\rho$, *we have in view of* (4.2) *that the autocovariances* $(\Omega_{n,\mathsf{H}})_{1,k}$ *are bounded, i.e. there exists* $K_{\mathsf{H},\rho} > 0$ *such that*

$$\left|(\Omega_{n,\mathsf{H}})_{1,k}\right| \leq K_{\mathsf{H},\rho}.$$

Since $k^{2\mathsf{H}-2} \leq 1$ *for* $k = 1, \ldots, N_\rho$, *it follows that*

$$\left|(\Omega_{n,\mathsf{H}})_{1,k}\right| \leq K_{\mathsf{H},\rho} k^{2\mathsf{H}-2}, \qquad k = 1, \ldots, N_\rho. \tag{4.25}$$

Consequently, we obtain from (4.24) *and* (4.25) *that for any* $k \geq 1$,

$$\left|(\Omega_{n,\mathsf{H}})_{1,k}\right| \leq K k^{2\mathsf{H}-2}, \tag{4.26}$$

where $K = \max(\mathsf{H}\{\rho\,|2\mathsf{H}-1|+1\}, K_{\mathsf{H},\rho})$.

A similar calculation can be done to show that there exists a positive constant K *such that for any* $j, k = 1, 2, \ldots$, *we have*

$$\left|\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{j,k}\right| \leq K\left(\frac{1}{k-j}\right)^{2\mathsf{H}}. \tag{4.27}$$

4.4.2 Proof of Theorem 4.1

The proof of convergence in probability of $(\hat{a}_n)_{n\geq 1}$ to a_0 will be done in two steps. We first show that $(\|Z_{n,\mathsf{H}}\|^2_{\mathbb{R}^n}/n)_{n\geq 1}$ converges in mean-square to $1/(1-a_0^2)$ and we prove in a second time that $(Z_{n,\mathsf{H}}^{\mathsf{T}} U_{n,\mathsf{H}}(a_0)/n)_{n\geq 1}$ converges in mean-square to 0.

Proof of the convergence in mean-square of $(\|Z_{n,\mathsf{H}}\|^2_{\mathbb{R}^n}/n)_{n\geq 1}$ to $1/(1-a_0^2)$

To establish the mean-square convergence of $(\|Z_{n,\mathsf{H}}\|^2_{\mathbb{R}^n}/n)_{n\geq 1}$ to $1/(1-a_0^2)$, we consider the following lemmas and remark.

Lemma 4.1. *Under Assumption* ***(A0)****, we have*

$$\mathbb{E}\left[\frac{1}{n}\|Z_{n,\mathsf{H}}\|^2_{\mathbb{R}^n}\right] \xrightarrow[n\to\infty]{} \frac{1}{1-a_0^2}.$$

Proof. In view of (4.16), we have

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{n}\|Z_{n,\mathsf{H}}\|^2_{\mathbb{R}^n}\right] &= \frac{1}{n}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n} a_0^{j_1+j_2-2}\mathbb{E}\left[\left\{L^{j_1}U_{n,\mathsf{H}}^{\mathsf{T}}(a_0)\right\}\left\{L^{j_2}U_{n,\mathsf{H}}(a_0)\right\}\right] \\
&= \frac{1}{n}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n} a_0^{j_1+j_2-2}\sum_{k=1}^{n}\mathbb{E}\left[\left\{L^{j_1}(U_{n,\mathsf{H}}(a_0))_k\right\}\left\{L^{j_2}(U_{n,\mathsf{H}}(a_0))_k\right\}\right] \\
&= \frac{1}{n}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n} a_0^{j_1+j_2-2}\sum_{k=1}^{n}\sum_{r_1=1}^{n}\sum_{r_2=1}^{n}\left(\Omega_{n,\mathsf{H}}^{-1/2}\right)_{k,r_1}\left(\Omega_{n,\mathsf{H}}^{-1/2}\right)_{k,r_2}\mathbb{E}\left[e_{r_1+1-j_1}^{\mathsf{H}}(a_0)e_{r_2+1-j_2}^{\mathsf{H}}(a_0)\right] \\
&= \frac{1}{n}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_2=j_2}^{n}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_1,r_2}(\Omega_{n,\mathsf{H}})_{r_1+1-j_1,r_2+1-j_2}.
\end{aligned}$$

The stationarity of $(\epsilon_t^{\mathsf{H}})_{t\in\mathbb{Z}}$ implies that:

— for $j_2 < j_1$:

$$(\Omega_{n,\mathsf{H}})_{r_1+1-j_1,r_2+1-j_2} = (\Omega_{n,\mathsf{H}})_{r_2,r_1+j_2-j_1}\,,$$

— for $j_2 \geq j_1$:

$$(\Omega_{n,\mathsf{H}})_{r_1+1-j_1,r_2+1-j_2} = (\Omega_{n,\mathsf{H}})_{r_1,r_2+j_1-j_2}\,.$$

Thus, we obtain that

$$\begin{aligned}\mathbb{E}\left[\frac{1}{n}\left\|Z_{n,\mathsf{H}}\right\|_{\mathbb{R}^n}^2\right] &= \frac{1}{n}\sum_{j_1=1}^{n}\sum_{j_2=1}^{j_1-1} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_2=j_2}^{n}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_1,r_2}(\Omega_{n,\mathsf{H}})_{r_2,r_1+j_2-j_1}\\ &\quad+\frac{1}{n}\sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_2=j_2}^{n}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_1,r_2}(\Omega_{n,\mathsf{H}})_{r_1,r_2+j_1-j_2}\,.\end{aligned}$$

This implies that

$$\begin{aligned}\mathbb{E}\left[\frac{1}{n}\left\|Z_{n,\mathsf{H}}\right\|_{\mathbb{R}^n}^2\right] &= \frac{1}{n}\sum_{j_1=1}^{n}\sum_{j_2=1}^{j_1-1} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}(I_n)_{r_1,r_1+j_2-j_1}\\ &\quad-\frac{1}{n}\sum_{j_1=1}^{n}\sum_{j_2=1}^{j_1-1} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_2=1}^{j_2-1}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_1,r_2}(\Omega_{n,\mathsf{H}})_{r_2,r_1+j_2-j_1}\\ &\quad+\frac{1}{n}\sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n} a_0^{j_1+j_2-2}\sum_{r_2=j_2}^{n}(I_n)_{r_2,r_2+j_1-j_2}\\ &\quad-\frac{1}{n}\sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=1}^{j_1-1}\sum_{r_2=j_2}^{n}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_1,r_2}(\Omega_{n,\mathsf{H}})_{r_1,r_2+j_1-j_2}\,.\end{aligned}$$

We have

$$\frac{1}{n}\sum_{j_1=1}^{n}\sum_{j_2=1}^{j_1-1} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}(I_n)_{r_1,r_1+j_2-j_1} = 0$$

and

$$\frac{1}{n}\sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n} a_0^{j_1+j_2-2}\sum_{r_2=j_2}^{n}(I_n)_{r_2,r_2+j_1-j_2} = \frac{1}{n}\sum_{j=1}^{n} a_0^{2j-2}(n-j+1),$$

thus we may write

$$\mathbb{E}\left[\frac{1}{n}\left\|Z_{n,\mathsf{H}}\right\|_{\mathbb{R}^n}^2\right] = \frac{1}{n}\sum_{j=1}^{n} a_0^{2j-2}(n-j+1) - \mathsf{W}_{n,\mathsf{H}}, \tag{4.28}$$

where

$$\mathrm{W}_{n,\mathrm{H}} = \frac{1}{n}\sum_{j_1=1}^{n}\sum_{j_2=1}^{j_1-1} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_2=1}^{j_2-1}\left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_1,r_2}\left(\Omega_{n,\mathrm{H}}\right)_{r_2,r_1+j_2-j_1} + \frac{1}{n}\sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=1}^{j_1-1}\sum_{r_2=j_2}^{n}\left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_1,r_2}\left(\Omega_{n,\mathrm{H}}\right)_{r_1,r_2+j_1-j_2}. \tag{4.29}$$

We shall prove that $\mathrm{W}_{n,\mathrm{H}} \xrightarrow[n\to\infty]{} 0$. Consider the first term on the right side of Equation (4.29) (the second term can be treated in a similar way) and note that $r_2 - r_1 \leq j_1 - r_1 - 1$ and $r_2 - r_1 + j_1 - j_2 \leq j_1 - r_1 - 1$. By the properties of the components of the matrix $\Omega_{n,\mathrm{H}}$ and its inverse given in (4.26) and (4.27), there exists a positive constant K such that

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{j_1=1}^{n}\sum_{j_2=1}^{j_1-1} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_2=1}^{j_2-1}\left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_1,r_2}\left(\Omega_{n,\mathrm{H}}\right)_{r_2,r_1+j_2-j_1}\right| \\
&\quad\leq \frac{K}{n}\sum_{j_1=1}^{n}\sum_{j_2=1}^{j_1-1}|a_0|^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_2=1}^{j_2-1}\frac{1}{(r_2-r_1)^{2\mathrm{H}}}\frac{1}{(r_2-r_1+j_1-j_2)^{2-2\mathrm{H}}} \\
&\quad\leq \frac{K}{n}\sum_{j_1=1}^{n}\sum_{j_2=1}^{j_1-1}|a_0|^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_2=1}^{j_2-1}\frac{1}{(j_1-r_1-1)^{2\mathrm{H}}}\frac{1}{(j_1-r_1-1)^{2-2\mathrm{H}}} \\
&\quad= \frac{K}{n}\sum_{j_1=1}^{n}\sum_{r_1=j_1}^{n}\frac{|a_0|^{j_1-1}}{(j_1-r_1-1)^2}\sum_{j_2=1}^{j_1-1}(j_2-1)\,|a_0|^{j_2-1}.
\end{aligned}$$

Since

$$\sum_{j_2=1}^{j_1-1}(j_2-1)\,|a_0|^{j_2-1} = \frac{1-|a_0|^{j_1-1}\{(j_1-1)(1-|a_0|)+1\}}{(1-|a_0|)^2} \leq \frac{1}{(1-|a_0|)^2},$$

we deduce that

$$\left|\frac{1}{n}\sum_{j_1=1}^{n}\sum_{j_2=1}^{j_1-1} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_2=1}^{j_2-1}\left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_1,r_2}\left(\Omega_{n,\mathrm{H}}\right)_{r_2,r_1+j_2-j_1}\right| \leq \frac{K}{n(1-|a_0|)^2}\sum_{j_1=1}^{n}\sum_{r_1=j_1}^{n}\frac{|a_0|^{j_1-1}}{(j_1-r_1-1)^2}.$$

Now we consider the change of indices $r_1 = j_1 + k - 1$ to obtain that

$$\left|\frac{1}{n}\sum_{j_1=1}^{n}\sum_{j_2=1}^{j_1-1} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_2=1}^{j_2-1}\left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_1,r_2}\left(\Omega_{n,\mathrm{H}}\right)_{r_2,r_1+j_2-j_1}\right| \leq \frac{K}{n(1-|a_0|)^2}\sum_{j_1=1}^{n}|a_0|^{j_1-1}\sum_{k\geq 1}\frac{1}{k^2} \xrightarrow[n\to\infty]{} 0.$$

The second term on the right side of Equation (4.29) can be treated in the same way. The proof of the convergence of the sequence $(\mathrm{W}_{n,\mathrm{H}})_{n\geq 1}$ towards 0 is then completed.

Finally, from (4.28) we conclude that

$$\mathbb{E}\left[\frac{1}{n}\left\|Z_{n,\mathrm{H}}\right\|^2_{\mathbb{R}^n}\right] = \left(\frac{n+1}{n}\right)\left(\frac{1-a_0^{2n}}{1-a_0^2}\right) - \frac{1}{n}\left(\frac{1-a_0^{2n}}{1-a_0^2}\right) - \frac{1}{n}\left(\frac{1-a_0^{2n-2}\left\{n(1-a_0^2)+a_0^2\right\}}{(1-a_0^2)^2}\right)$$
$$- \mathrm{W}_{n,\mathrm{H}} \xrightarrow[n\to\infty]{} \frac{1}{1-a_0^2},$$

and the lemma is proved. □

Remark 4.4. *Denote by $\Sigma_{n,X,\mathrm{H}}$ the variance matrix of the random vector $(X_1,\ldots,X_n)^\mathsf{T}$, i.e. for any $1 \le i,j \le n$, $(\Sigma_{n,X,\mathrm{H}})_{i,j} = \mathrm{Cov}(X_i,X_j)$, and let $Z^{(1)}_{n,\mathrm{H}}$ be the random vector defined by $Z^{(1)}_{n,\mathrm{H}} = \Sigma^{-1/2}_{n,X,\mathrm{H}}(X_1,\ldots,X_n)^\mathsf{T}$. From (4.6), one has*

$$\frac{1}{n}\left\|Z_{n,\mathrm{H}}\right\|^2_{\mathbb{R}^n} = \frac{1}{n}Z^\mathsf{T}_{n,\mathrm{H}}Z_{n,\mathrm{H}} = \frac{1}{n}\begin{pmatrix}X_1\\ \vdots\\ X_n\end{pmatrix}^\mathsf{T}\Omega^{-1}_{n,\mathrm{H}}\begin{pmatrix}X_1\\ \vdots\\ X_n\end{pmatrix} = \frac{1}{n}(Z^{(1)}_{n,\mathrm{H}})^\mathsf{T}\Sigma^{1/2}_{n,X,\mathrm{H}}\Omega^{-1}_{n,\mathrm{H}}\Sigma^{1/2}_{n,X,\mathrm{H}}Z^{(1)}_{n,\mathrm{H}}.$$

The covariance matrix $\Sigma^{1/2}_{n,X,\mathrm{H}}\Omega^{-1}_{n,\mathrm{H}}\Sigma^{1/2}_{n,X,\mathrm{H}}$ is a symmetric real matrix. So there exists an orthogonal matrix $Q_{n,\mathrm{H}}$ such that $\Lambda_{n,\mathrm{H}} = Q^\mathsf{T}_{n,\mathrm{H}}\Sigma^{1/2}_{n,X,\mathrm{H}}\Omega^{-1}_{n,\mathrm{H}}\Sigma^{1/2}_{n,X,\mathrm{H}}Q_{n,\mathrm{H}}$ is a diagonal matrix. Hence, the random variable $\left\|Z_{n,\mathrm{H}}\right\|^2_{\mathbb{R}^n}/n$ can be rewritten as

$$\begin{aligned}\frac{1}{n}\left\|Z_{n,\mathrm{H}}\right\|^2_{\mathbb{R}^n} &= \frac{1}{n}(Q^\mathsf{T}_{n,\mathrm{H}}Z^{(1)}_{n,\mathrm{H}})^\mathsf{T}\Lambda_{n,\mathrm{H}}\left(Q^\mathsf{T}_{n,\mathrm{H}}Z^{(1)}_{n,\mathrm{H}}\right)\\ &= \frac{1}{n}\sum_{j=1}^{n}\eta^{(n)}_{j,\mathrm{H}}\left[\left(Q^\mathsf{T}_{n,\mathrm{H}}Z^{(1)}_{n,\mathrm{H}}\right)_j\right]^2, \qquad (4.30)\end{aligned}$$

where $\eta^{(n)}_{1,\mathrm{H}},\ldots,\eta^{(n)}_{n,\mathrm{H}}$ are the eigenvalues of the symmetric real matrix $\Sigma^{1/2}_{n,X,\mathrm{H}}\Omega^{-1}_{n,\mathrm{H}}\Sigma^{1/2}_{n,X,\mathrm{H}}$. The expression of $\left\|Z_{n,\mathrm{H}}\right\|^2_{\mathbb{R}^n}/n$ given in (4.30) has the advantage of involving the standard normal random vector $Q^\mathsf{T}_{n,\mathrm{H}}Z^{(1)}_{n,\mathrm{H}}$. This will considerably simplify the calculations in the following.

We show in the lemma below the boundedness of the eigenvalues $\eta^{(n)}_{1,\mathrm{H}},\ldots,\eta^{(n)}_{n,\mathrm{H}}$ of the matrix $\Sigma^{1/2}_{n,X,\mathrm{H}}\Omega^{-1}_{n,\mathrm{H}}\Sigma^{1/2}_{n,X,\mathrm{H}}$ uniformly in n.

Lemma 4.2. *Under Assumption **(A0)** and for any $1 \le j \le n$, we have*

$$\frac{1}{(1+|a_0|)^2} \le \eta^{(n)}_{j,\mathrm{H}} \le \frac{1}{(1-|a_0|)^2}.$$

Proof. Let $f_X(\cdot)$ be the spectral density of $(X_t)_{t\in\mathbb{Z}}$ and denote by $f_{\epsilon^\mathrm{H}}(\cdot)$ the spectral density of the stationary process $(\epsilon^\mathrm{H}_t)_{t\in\mathbb{Z}}$. One can express $f_X(\cdot)$ as a function of $f_{\epsilon^\mathrm{H}}(\cdot)$. Indeed, using

(4.13) and the stationarity of $(X_t)_{t\in\mathbb{Z}}$ and $(\epsilon_t^{\mathsf{H}})_{t\in\mathbb{Z}}$ we have for any $\omega\in\mathbb{R}$,

$$\begin{aligned} f_X(\omega) &= \frac{1}{2\pi}\sum_{h\in\mathbb{Z}}\gamma_X(h)\mathrm{e}^{-i\omega h} = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}}\mathrm{Cov}(X_t,X_{t-h})\mathrm{e}^{-i\omega h} \\ &= \frac{1}{2\pi}\sum_{h\in\mathbb{Z}}\mathrm{Cov}\left(\sum_{j\in\mathbb{N}}a_0^j\epsilon_{t-j}^{\mathsf{H}},\sum_{k\in\mathbb{N}}a_0^k\epsilon_{t-h-k}^{\mathsf{H}}\right)\mathrm{e}^{-i\omega h} \\ &= \frac{1}{2\pi}\sum_{h\in\mathbb{Z}}\sum_{j,k\in\mathbb{N}}a_0^{j+k}\gamma_{\epsilon^{\mathsf{H}}}(h+k-j)\mathrm{e}^{-i\omega(h+k-j)}\mathrm{e}^{-i\omega(j-k)} \\ &= f_{\epsilon^{\mathsf{H}}}(\omega)\left(\sum_{j\in\mathbb{N}}a_0^j\mathrm{e}^{-i\omega j}\right)\left(\overline{\sum_{j\in\mathbb{N}}a_0^j\mathrm{e}^{-i\omega j}}\right) = f_{\epsilon^{\mathsf{H}}}(\omega)\left|\sum_{j\in\mathbb{N}}a_0^j\mathrm{e}^{-i\omega j}\right|^2, \end{aligned}$$

where $\gamma_X(\cdot)$ denotes the autocovariance function of the process $(X_t)_{t\in\mathbb{Z}}$ and $\overline{z}$ designates the complex conjugate of the complex number z. Furthermore, Assumption **(A0)** implies that the modulus of the complex number $a_0\mathrm{e}^{-i\omega}$ is strictly less than 1. Hence, we deduce that the spectral density of the stationary process $(X_t)_{t\in\mathbb{Z}}$ is related to $f_{\epsilon^{\mathsf{H}}}(\cdot)$ as follows:

$$f_X(\omega) = f_{\epsilon^{\mathsf{H}}}(\omega)\left|\frac{1}{1-a_0\mathrm{e}^{-i\omega}}\right|^2 = \frac{f_{\epsilon^{\mathsf{H}}}(\omega)}{1-2a_0\cos(\omega)+a_0^2},\quad \forall\omega\in\mathbb{R}. \tag{4.31}$$

From this last equation and taking into consideration the fact that $(1-|a_0|)^2\leq 1-2a_0\cos(\omega)+a_0^2\leq(1+|a_0|)^2$, one obtains that for any $\omega\in\mathbb{R}$,

$$(1-|a_0|)^2 f_X(\omega)\leq f_{\epsilon^{\mathsf{H}}}(\omega)\leq(1+|a_0|)^2 f_X(\omega). \tag{4.32}$$

Note now that for any vector V of $\mathbb{R}^{n\times 1}$, we have

$$\begin{aligned} V^{\mathsf{T}}\left(\Omega_{n,\mathsf{H}}-(1+|a_0|)^2\Sigma_{n,X,\mathsf{H}}\right)V &= \sum_{k=1}^{n}\sum_{j=1}^{n}V_j\left(\Omega_{n,\mathsf{H}}-(1+|a_0|)^2\Sigma_{n,X,\mathsf{H}}\right)_{j,k}V_k \\ &= \sum_{k,j=1}^{n}V_j\int_{-\pi}^{\pi}\left\{f_{\epsilon^{\mathsf{H}}}(\omega)-(1+|a_0|)^2 f_X(\omega)\right\}\mathrm{e}^{i\omega(k-j)}\mathrm{d}\omega V_k \\ &= \int_{-\pi}^{\pi}\left\{f_{\epsilon^{\mathsf{H}}}(\omega)-(1+|a_0|)^2 f_X(\omega)\right\}V^{\mathsf{T}}\Upsilon_n(\omega)V\mathrm{d}\omega, \end{aligned} \tag{4.33}$$

where $\Upsilon_n(\omega)=V_n(\omega)V_n^*(\omega)$ with $V_n(\omega)=(\mathrm{e}^{i\omega},\mathrm{e}^{2i\omega},\ldots,\mathrm{e}^{ni\omega})^{\mathsf{T}}$. The notation V^* designates the conjugate transpose of the vector V.

Furthermore, the real $V^{\mathsf{T}}\Upsilon_n(\omega)V$ is a non-negative number. In fact,

$$V^{\mathsf{T}}\Upsilon_n(\omega)V = V^{\mathsf{T}}V_n(\omega)V_n^*(\omega)V = \left(V_n^*(\omega)V\right)^*\left(V_n^*(\omega)V\right) = \left|V_n^*(\omega)V\right|^2.$$

Then, from (4.32) and (4.33), one has

$$V^{\mathsf{T}}(\Omega_{n,\mathsf{H}}-(1+|a_0|)^2\Sigma_{n,X,\mathsf{H}})V\leq 0. \tag{4.34}$$

Let now ζ be an element of the spectrum of $\Omega_{n,\mathrm{H}}^{-1/2}\Sigma_{n,X,\mathrm{H}}\Omega_{n,\mathrm{H}}^{-1/2}$, there exists then $C_1 \in \mathbb{R}^{n\times 1} \setminus \{0_{\mathbb{R}^{n\times 1}}\}$ such that $\Omega_{n,\mathrm{H}}^{-1/2}\Sigma_{n,X,\mathrm{H}}\Omega_{n,\mathrm{H}}^{-1/2}C_1 = \zeta C_1$. Hence, it follows that

$$C_1^{\mathsf{T}}\Omega_{n,\mathrm{H}}^{-1/2}\Sigma_{n,X,\mathrm{H}}\Omega_{n,\mathrm{H}}^{-1/2}C_1 = \zeta \left\| C_1 \right\|_{\mathbb{R}^n}^2 .$$

Taking $C_2 = \Omega_{n,\mathrm{H}}^{-1/2}C_1$, we obtain from this last equation that

$$C_2^{\mathsf{T}}\Sigma_{n,X,\mathrm{H}}C_2 = \zeta \left\| \Omega_{n,\mathrm{H}}^{1/2}C_2 \right\|_{\mathbb{R}^n}^2 .$$

Using (4.34), we deduce that

$$C_2^{\mathsf{T}}\Omega_{n,\mathrm{H}}C_2 \leq (1+|a_0|)^2\zeta \left\| \Omega_{n,\mathrm{H}}^{1/2}C_2 \right\|_{\mathbb{R}^n}^2 .$$

Since $C_1 = \Omega_{n,\mathrm{H}}^{1/2}C_2 \neq 0_{\mathbb{R}^{n\times 1}}$ as an eigenvector, we conclude that

$$\zeta \geq \frac{C_2^{\mathsf{T}}\Omega_{n,\mathrm{H}}C_2}{\left\| \Omega_{n,\mathrm{H}}^{1/2}C_2 \right\|_{\mathbb{R}^n}^2}\frac{1}{(1+|a_0|)^2} = \frac{1}{(1+|a_0|)^2}. \tag{4.35}$$

A similar calculation can be done to prove that, for any vector V of $\mathbb{R}^{n\times 1}$,

$$V^{\mathsf{T}}(\Omega_{n,\mathrm{H}} - (1-|a_0|)^2\Sigma_{n,X,\mathrm{H}})V \geq 0.$$

Arguing as before, one obtains that

$$\zeta \leq \frac{C_2^{\mathsf{T}}\Omega_{n,\mathrm{H}}C_2}{\left\| \Omega_{n,\mathrm{H}}^{1/2}C_2 \right\|_{\mathbb{R}^n}^2}\frac{1}{(1-|a_0|)^2} = \frac{1}{(1-|a_0|)^2}.$$

Since for any $\alpha \in \mathbb{R}$,

$$\begin{aligned}
\det\left(\Sigma_{n,X,\mathrm{H}}^{1/2}\Omega_{n,\mathrm{H}}^{-1}\Sigma_{n,X,\mathrm{H}}^{1/2} - \alpha I_n\right) &= \det\left(\Sigma_{n,X,\mathrm{H}}^{1/2}\left\{\Omega_{n,\mathrm{H}}^{-1}\Sigma_{n,X,\mathrm{H}} - \alpha I_n\right\}\Sigma_{n,X,\mathrm{H}}^{-1/2}\right) \\
&= \det\left(\Omega_{n,\mathrm{H}}^{-1}\Sigma_{n,X,\mathrm{H}} - \alpha I_n\right) \\
&= \det\left(\Omega_{n,\mathrm{H}}^{-1/2}\left\{\Omega_{n,\mathrm{H}}^{-1/2}\Sigma_{n,X,\mathrm{H}}\Omega_{n,\mathrm{H}}^{-1/2} - \alpha I_n\right\}\Omega_{n,\mathrm{H}}^{1/2}\right) \\
&= \det\left(\Omega_{n,\mathrm{H}}^{-1/2}\Sigma_{n,X,\mathrm{H}}\Omega_{n,\mathrm{H}}^{-1/2} - \alpha I_n\right),
\end{aligned}$$

the spectrum of the matrix $\Sigma_{n,X,\mathrm{H}}^{1/2}\Omega_{n,\mathrm{H}}^{-1}\Sigma_{n,X,\mathrm{H}}^{1/2}$ is equal to the set of the eigenvalues of $\Omega_{n,\mathrm{H}}^{-1/2}\Sigma_{n,X,\mathrm{H}}\Omega_{n,\mathrm{H}}^{-1/2}$, the lemma is then proved. □

Lemma 4.3. *Under Assumption **(A0)**, we have*

$$\frac{1}{n}\left\| Z_{n,\mathrm{H}} \right\|_{\mathbb{R}^n}^2 \xrightarrow[n\to\infty]{\mathbb{L}^2} \frac{1}{1-a_0^2}.$$

Proof. In view of Lemma 4.2, Remark 4.4 and (4.30), we have

$$\begin{aligned}\mathrm{Var}\left(\frac{1}{n}\left\|Z_{n,\mathrm{H}}\right\|_{\mathbb{R}^n}^2\right) &= \frac{1}{n^2}\sum_{j=1}^{n}\left(\eta_{j,\mathrm{H}}^{(n)}\right)^2 \mathrm{Var}\left(\left[\left(Q_{n,\mathrm{H}}^{\mathsf{T}} Z_{n,\mathrm{H}}^{(1)}\right)_j\right]^2\right) = \frac{2}{n^2}\sum_{j=1}^{n}\left(\eta_{j,\mathrm{H}}^{(n)}\right)^2 \\ &\leq \frac{2}{n(1-|a_0|)^4} \xrightarrow[n\to\infty]{} 0.\end{aligned}$$

Observe now that

$$\mathbb{E}\left[\left|\frac{1}{n}\left\|Z_{n,\mathrm{H}}\right\|_{\mathbb{R}^n}^2 - \frac{1}{1-a_0^2}\right|^2\right] = \mathrm{Var}\left(\frac{1}{n}\left\|Z_{n,\mathrm{H}}\right\|_{\mathbb{R}^n}^2\right) + \left(\mathbb{E}\left[\frac{1}{n}\left\|Z_{n,\mathrm{H}}\right\|_{\mathbb{R}^n}^2\right] - \frac{1}{1-a_0^2}\right)^2.$$

Thanks to Lemma 4.1 and the convergence stated just before, the expected result is obtained. □

Proof of the convergence in mean-square of $(Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)/n)_{n\geq 1}$ to 0

Lemma 4.4. *Under Assumption **(A0)**, we have*

$$\frac{1}{n}\left[Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)\right] \xrightarrow[n\to\infty]{\mathbb{L}^2} 0.$$

Proof. Due to the fact that $e_t^{\mathrm{H}}(a_0)$ is centered for any t, we notice that for fixed (i,j,k) the joint cumulant of $e_0^{\mathrm{H}}(a_0), e_i^{\mathrm{H}}(a_0), e_j^{\mathrm{H}}(a_0)$ and $e_k^{\mathrm{H}}(a_0)$ is given by:

$$\begin{aligned}\mathrm{cum}(e_0^{\mathrm{H}}(a_0), e_i^{\mathrm{H}}(a_0), e_j^{\mathrm{H}}(a_0), e_k^{\mathrm{H}}(a_0)) &= \mathbb{E}\left[e_0^{\mathrm{H}}(a_0)e_i^{\mathrm{H}}(a_0)e_j^{\mathrm{H}}(a_0)e_k^{\mathrm{H}}(a_0)\right] \\ &\quad - \mathbb{E}\left[e_0^{\mathrm{H}}(a_0)e_i^{\mathrm{H}}(a_0)\right]\mathbb{E}\left[e_j^{\mathrm{H}}(a_0)e_k^{\mathrm{H}}(a_0)\right] - \mathbb{E}\left[e_0^{\mathrm{H}}(a_0)e_j^{\mathrm{H}}(a_0)\right]\mathbb{E}\left[e_i^{\mathrm{H}}(a_0)e_k^{\mathrm{H}}(a_0)\right] \\ &\quad - \mathbb{E}\left[e_0^{\mathrm{H}}(a_0)e_k^{\mathrm{H}}(a_0)\right]\mathbb{E}\left[e_i^{\mathrm{H}}(a_0)e_j^{\mathrm{H}}(a_0)\right].\end{aligned}$$

From (4.6) and (4.16), the positive definiteness of $\Omega_{n,\mathrm{H}}$ and the fact that the process $(e_t^{\mathrm{H}}(a_0))_{t\in\mathbb{Z}}$ defined in (4.4) is centered, we obtain that

$$\begin{aligned}\mathbb{E}\left[\left(\frac{1}{n}\left\{Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)\right\}\right)^2\right] &= \frac{1}{n^2}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n} a_0^{j_1+j_2-2}\mathbb{E}\left[\left(L^{j_1}U_{n,\mathrm{H}}(a_0)\right)^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)\left(L^{j_2}U_{n,\mathrm{H}}(a_0)\right)^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)\right] \\ &= \frac{1}{n^2}\sum_{j_1,j_2=1}^{n} a_0^{j_1+j_2-2}\sum_{k_1,k_2=1}^{n}\sum_{r_1,\ldots,r_4=1}^{n}\left(\Omega_{n,\mathrm{H}}^{-1/2}\right)_{k_1,r_1}\left(\Omega_{n,\mathrm{H}}^{-1/2}\right)_{k_1,r_2}\left(\Omega_{n,\mathrm{H}}^{-1/2}\right)_{k_2,r_3}\left(\Omega_{n,\mathrm{H}}^{-1/2}\right)_{k_2,r_4} \\ &\qquad \times \mathbb{E}\left[e_{r_1+1-j_1}^{\mathrm{H}}(a_0)e_{r_2+1}^{\mathrm{H}}(a_0)e_{r_3+1-j_2}^{\mathrm{H}}(a_0)e_{r_4+1}^{\mathrm{H}}(a_0)\right] \\ &= \Delta_{1,n,\mathrm{H}} + \Delta_{2,n,\mathrm{H}} + \Delta_{3,n,\mathrm{H}} + \Delta_{4,n,\mathrm{H}}, \qquad (4.36)\end{aligned}$$

where

$$\begin{aligned}
\Delta_{1,n,\mathsf{H}} &= \frac{1}{n^2}\sum_{j_1,j_2=1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_2=1}^{n}\sum_{r_3=j_2}^{n}\sum_{r_4=1}^{n}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_1,r_2}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_3,r_4}\\
&\qquad\times \operatorname{cum}\left(e^{\mathsf{H}}_{r_1+1-j_1}(a_0), e^{\mathsf{H}}_{r_2+1}(a_0), e^{\mathsf{H}}_{r_3+1-j_2}(a_0), e^{\mathsf{H}}_{r_4+1}(a_0)\right),\\
\Delta_{2,n,\mathsf{H}} &= \frac{1}{n^2}\sum_{j_1,j_2=1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_2=1}^{n}\sum_{r_3=j_2}^{n}\sum_{r_4=1}^{n}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_1,r_2}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_3,r_4}\\
&\qquad\times \mathbb{E}\left[e^{\mathsf{H}}_{r_1+1-j_1}(a_0)e^{\mathsf{H}}_{r_2+1}(a_0)\right]\mathbb{E}\left[e^{\mathsf{H}}_{r_3+1-j_2}(a_0)e^{\mathsf{H}}_{r_4+1}(a_0)\right],\\
\Delta_{3,n,\mathsf{H}} &= \frac{1}{n^2}\sum_{j_1,j_2=1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_2=1}^{n}\sum_{r_3=j_2}^{n}\sum_{r_4=1}^{n}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_1,r_2}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_3,r_4}\\
&\qquad\times \mathbb{E}\left[e^{\mathsf{H}}_{r_1+1-j_1}(a_0)e^{\mathsf{H}}_{r_3+1-j_2}(a_0)\right]\mathbb{E}\left[e^{\mathsf{H}}_{r_2+1}(a_0)e^{\mathsf{H}}_{r_4+1}(a_0)\right] \text{ and}\\
\Delta_{4,n,\mathsf{H}} &= \frac{1}{n^2}\sum_{j_1,j_2=1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_2=1}^{n}\sum_{r_3=j_2}^{n}\sum_{r_4=1}^{n}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_1,r_2}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_3,r_4}\\
&\qquad\times \mathbb{E}\left[e^{\mathsf{H}}_{r_1+1-j_1}(a_0)e^{\mathsf{H}}_{r_4+1}(a_0)\right]\mathbb{E}\left[e^{\mathsf{H}}_{r_2+1}(a_0)e^{\mathsf{H}}_{r_3+1-j_2}(a_0)\right].
\end{aligned}$$

Since the vector $(\epsilon_1^{\mathsf{H}},\ldots,\epsilon_{n+1}^{\mathsf{H}})^{\mathsf{T}}$ is Gaussian random vector, any joint cumulant involving three (or more) elements of $(\epsilon_1^{\mathsf{H}},\ldots,\epsilon_{n+1}^{\mathsf{H}})^{\mathsf{T}}$ is zero. The random variables $e^{\mathsf{H}}_{r_1+1-j_1}(a_0)$, $e^{\mathsf{H}}_{r_2+1}(a_0)$, $e^{\mathsf{H}}_{r_3+1-j_2}(a_0)$ and $e^{\mathsf{H}}_{r_4+1}(a_0)$ in (4.36) are elements of the Gaussian vector $(\epsilon_1^{\mathsf{H}},\ldots,\epsilon_{n+1}^{\mathsf{H}})^{\mathsf{T}}$ so

$$\operatorname{cum}\left(e^{\mathsf{H}}_{r_1+1-j_1}(a_0), e^{\mathsf{H}}_{r_2+1}(a_0), e^{\mathsf{H}}_{r_3+1-j_2}(a_0), e^{\mathsf{H}}_{r_4+1}(a_0)\right) = 0.$$

Thus, we have

$$\Delta_{1,n,\mathsf{H}} = 0.$$

Let $\Omega_{n+1,\mathsf{H}}$ be the covariance matrix of the random vector $(\epsilon_1^{\mathsf{H}},\ldots,\epsilon_{n+1}^{\mathsf{H}})^{\mathsf{T}}$. Thanks to the stationarity of the process $(\epsilon_t^{\mathsf{H}})_{t\in\mathbb{Z}}$ one has $(\Omega_{n+1,\mathsf{H}})_{r_3+1-j_2,r_4+1} = (\Omega_{n,\mathsf{H}})_{r_3-j_2,r_4}$ when $j_2, r_4 = 1,\ldots,n$ and $j_2+1 \le r_3 \le n$. Thus $\Delta_{2,n,\mathsf{H}}$ can be rewritten as:

$$\begin{aligned}
\Delta_{2,n,\mathsf{H}} &= \frac{1}{n^2}\sum_{j_1,j_2=1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_2=1}^{n}\sum_{r_3=j_2}^{n}\sum_{r_4=1}^{n}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_1,r_2}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_3,r_4}\left(\Omega_{n+1,\mathsf{H}}\right)_{r_1+1-j_1,r_2+1}\left(\Omega_{n+1,\mathsf{H}}\right)_{r_3+1-j_2,r_4+1}\\
&= \frac{1}{n^2}\sum_{j_1,j_2=1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_2=1}^{n}\sum_{r_3=j_2+1}^{n}\sum_{r_4=1}^{n}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_1,r_2}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_3,r_4}\left(\Omega_{n+1,\mathsf{H}}\right)_{r_1+1-j_1,r_2+1}\left(\Omega_{n,\mathsf{H}}\right)_{r_3-j_2,r_4}\\
&\quad+\frac{1}{n^2}\sum_{j_1,j_2=1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_2=1}^{n}\sum_{r_4=1}^{n}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_1,r_2}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{j_2,r_4}\left(\Omega_{n+1,\mathsf{H}}\right)_{r_1+1-j_1,r_2+1}\left(\Omega_{n+1,\mathsf{H}}\right)_{1,r_4+1}
\end{aligned}$$

This implies that

$$\begin{aligned}
\Delta_{2,n,\mathsf{H}} &= \frac{1}{n^2}\sum_{j_1,j_2=1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_2=1}^{n}\sum_{r_3=j_2+1}^{n}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_1,r_2}\left(\Omega_{n+1,\mathsf{H}}\right)_{r_1+1-j_1,r_2+1}\left(I_n\right)_{r_3,r_3-j_2} \\
&\quad + \frac{1}{n^2}\sum_{j_1,j_2=1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=j_1+1}^{n}\sum_{r_2=1}^{n}\sum_{r_4=1}^{n}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_1,r_2}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{j_2,r_4}\left(\Omega_{n,\mathsf{H}}\right)_{r_1-j_1,r_2}\left(\Omega_{n+1,\mathsf{H}}\right)_{1,r_4+1} \\
&\quad + \frac{1}{n^2}\sum_{j_1,j_2=1}^{n} a_0^{j_1+j_2-2}\sum_{r_2=1}^{n}\sum_{r_4=1}^{n}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{j_1,r_2}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{j_2,r_4}\left(\Omega_{n+1,\mathsf{H}}\right)_{1,r_2+1}\left(\Omega_{n+1,\mathsf{H}}\right)_{1,r_4+1} \\
&= \frac{1}{n^2}\sum_{j_1,j_2=1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_2=1}^{n}\sum_{r_3=j_2+1}^{n}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_1,r_2}\left(\Omega_{n+1,\mathsf{H}}\right)_{r_1+1-j_1,r_2+1}\left(I_n\right)_{r_3,r_3-j_2} \\
&\quad + \frac{1}{n^2}\sum_{j_1,j_2=1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=j_1+1}^{n}\sum_{r_4=1}^{n}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{j_2,r_4}\left(\Omega_{n+1,\mathsf{H}}\right)_{1,r_4+1}\left(I_n\right)_{r_1,r_1-j_1} \\
&\quad + \frac{1}{n^2}\sum_{j_1,j_2=1}^{n} a_0^{j_1+j_2-2}\sum_{r_2=1}^{n}\sum_{r_4=1}^{n}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{j_1,r_2}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{j_2,r_4}\left(\Omega_{n+1,\mathsf{H}}\right)_{1,r_2+1}\left(\Omega_{n+1,\mathsf{H}}\right)_{1,r_4+1}
\end{aligned}$$

Since $j_1, j_2 = 1, \ldots, n$, $\left(I_n\right)_{r_1,r_1-j_1} = \left(I_n\right)_{r_3,r_3-j_2} = 0$, we obtain that

$$\begin{aligned}
\Delta_{2,n,\mathsf{H}} &= \frac{1}{n^2}\sum_{j_1,j_2=1}^{n} a_0^{j_1+j_2-2}\sum_{r_2=1}^{n}\sum_{r_4=1}^{n}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{j_1,r_2}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{j_2,r_4}\left(\Omega_{n+1,\mathsf{H}}\right)_{1,r_2+1}\left(\Omega_{n+1,\mathsf{H}}\right)_{1,r_4+1} \\
&= \left(\frac{1}{n}\sum_{j=1}^{n} a_0^{j-1}\sum_{r=1}^{n}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{j,r}\left(\Omega_{n+1,\mathsf{H}}\right)_{1,r+1}\right)^2 . \qquad (4.37)
\end{aligned}$$

Using a similar approach as that given in the proof of Lemma 4.1, one can confirm that the term in the right side of Equation (4.37) is equal to the square of the expected value of the random variable $Z_{n,\mathsf{H}}^{\mathsf{T}} U_{n,\mathsf{H}}(a_0)/n$. Similar calculation as we did in the proof of Lemma 4.1 implies that

$$\mathbb{E}\left[\frac{1}{n}\left\{Z_{n,\mathsf{H}}^{\mathsf{T}} U_{n,\mathsf{H}}(a_0)\right\}\right] \xrightarrow[n\to\infty]{} 0.$$

Let us be more precise.

From (4.16), we have

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{n}\left\{Z_{n,\mathsf{H}}^{\mathsf{T}} U_{n,\mathsf{H}}(a_0)\right\}\right] &= \frac{1}{n}\sum_{j=1}^{n} a_0^{j-1}\sum_{k=1}^{n}\mathbb{E}\left[\left\{L^j\left(U_{n,\mathsf{H}}(a_0)\right)_k\right\}\left(U_{n,\mathsf{H}}(a_0)\right)_k\right] \\
&= \frac{1}{n}\sum_{j=1}^{n} a_0^{j-1}\sum_{k=1}^{n}\sum_{r_1,r_2=1}^{n}\left(\Omega_{n,\mathsf{H}}^{-1/2}\right)_{k,r_1}\left(\Omega_{n,\mathsf{H}}^{-1/2}\right)_{k,r_2}\mathbb{E}\left[e_{r_1+1-j}^{\mathsf{H}}(a_0)\, e_{r_2+1}^{\mathsf{H}}(a_0)\right] \qquad (4.38)
\end{aligned}$$

$$= \frac{1}{n}\sum_{j=1}^{n} a_0^{j-1} \sum_{r_1=j}^{n}\sum_{r_2=1}^{n} \left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_1,r_2} \left(\Omega_{n+1,\mathrm{H}}\right)_{r_1+1-j,r_2+1}. \tag{4.39}$$

As before, we proceed by cases to deduce that

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{n}\left\{Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)\right\}\right] &= \frac{1}{n}\sum_{j=1}^{n} a_0^{j-1} \sum_{r_1=j+1}^{n}\sum_{r_2=1}^{n} \left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_1,r_2} \left(\Omega_{n,\mathrm{H}}\right)_{r_1-j,r_2} \\
&\qquad + \frac{1}{n}\sum_{j=1}^{n} a_0^{j-1} \sum_{r_2=1}^{n} \left(\Omega_{n,\mathrm{H}}^{-1}\right)_{j,r_2} \left(\Omega_{n+1,\mathrm{H}}\right)_{1,r_2+1} \\
&= \frac{1}{n}\sum_{j=1}^{n} a_0^{j-1} \sum_{r_1=j+1}^{n} \left(I_n\right)_{r_1,r_1-j} + \frac{1}{n}\sum_{j=1}^{n} a_0^{j-1} \sum_{r_2=1}^{n} \left(\Omega_{n,\mathrm{H}}^{-1}\right)_{j,r_2} \left(\Omega_{n+1,\mathrm{H}}\right)_{1,r_2+1} \\
&= \frac{1}{n}\sum_{j=1}^{n} a_0^{j-1} \sum_{r_2=1}^{n} \left(\Omega_{n,\mathrm{H}}^{-1}\right)_{j,r_2} \left(\Omega_{n+1,\mathrm{H}}\right)_{1,r_2+1} \\
&= \frac{1}{n}\sum_{j=1}^{n} a_0^{j-1} \left(\Omega_{n,\mathrm{H}}^{-1}\right)_{j,j} \left(\Omega_{n+1,\mathrm{H}}\right)_{1,j+1} + \frac{1}{n}\sum_{j=1}^{n} a_0^{j-1} \sum_{\substack{1\le r_2\le n \\ r_2\neq j}} \left(\Omega_{n,\mathrm{H}}^{-1}\right)_{j,r_2} \left(\Omega_{n+1,\mathrm{H}}\right)_{1,r_2+1}.
\end{aligned}$$

Finally, in view of (4.26) and (4.27) we conclude that

$$\left|\mathbb{E}\left[\frac{1}{n}\left\{Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)\right\}\right]\right| \le K\left(\frac{1}{n}\sum_{j=1}^{n}|a_0|^{j-1} + \frac{1}{n}\sum_{j=1}^{n}|a_0|^{j-1}\sum_{\substack{1\le r\le n \\ r\neq j}} \frac{1}{(r-j)^{2\mathrm{H}}}\frac{1}{r^{2-2\mathrm{H}}}\right). \tag{4.40}$$

Using the fact that the autoregressive parameter a_0 is assumed to belong to $]-1,1[$, Cesàro's Lemma implies that the first term in the right hand side of (4.40) tends to 0.

Proceeding by a simple calculation and using Cesàro's Lemma, one can also prove that the second term in the right hand side of (4.40) converges to 0. More explicitly,

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{n}|a_0|^{j-1}\sum_{\substack{1\le r\le n \\ r\neq j}} \frac{1}{(r-j)^{2\mathrm{H}}}\frac{1}{r^{2-2\mathrm{H}}} \\
&\quad = \frac{1}{n}\sum_{r=1}^{n}\frac{1}{r^{2-2\mathrm{H}}}\sum_{\substack{1\le j\le n \\ j\neq r}}|a_0|^{j-1}\frac{1}{(r-j)^{2\mathrm{H}}} \\
&\quad = \frac{1}{n}\sum_{r=1}^{n}\frac{1}{r^{2-2\mathrm{H}}}|a_0|^{r-1}\sum_{\substack{1-r\le j\le n-r \\ j\neq 0}}|a_0|^{j}\frac{1}{j^{2\mathrm{H}}} \\
&\quad = \frac{1}{n}\sum_{r=2}^{n}\frac{1}{r^{2-2\mathrm{H}}}|a_0|^{r-1}\sum_{j=1-r}^{-1}|a_0|^{j}\frac{1}{j^{2\mathrm{H}}} + \frac{1}{n}\sum_{r=1}^{n}\frac{1}{r^{2-2\mathrm{H}}}|a_0|^{r-1}\sum_{j=1}^{n-r}|a_0|^{j}\frac{1}{j^{2\mathrm{H}}}
\end{aligned}$$

$$\leq \frac{1}{n}\sum_{r=1}^{n}\frac{1}{r^{1-2\mathrm{H}}}\left|a_0\right|^{r-1}\max\left(\frac{1}{|a_0|},\frac{1}{|a_0|^r\,r^{2\mathrm{H}}}\right)+\frac{1}{n}\sum_{r=1}^{n}\frac{1}{r^{2-2\mathrm{H}}}\left|a_0\right|^{r}\left(\frac{1-|a_0|^{n-r}}{1-|a_0|}\right)$$
$$\xrightarrow[n\to\infty]{} 0.$$

Therefore, we deduce that

$$\Delta_{2,n,\mathrm{H}} \xrightarrow[n\to\infty]{} 0.$$

It is obvious that by exchanging the roles of j_1 and j_2 and of r_2 and r_4 in the expression of $\Delta_{4,n,\mathrm{H}}$, we obtain that $\Delta_{4,n,\mathrm{H}} = \Delta_{2,n,\mathrm{H}}$. Then, the sequence $(\Delta_{4,n,\mathrm{H}})_{n\geq 1}$ also converges to 0.

We therefore just have to show that $(\Delta_{3,n,\mathrm{H}})_{n\geq 1}$ converges to 0 to obtain the convergence in mean-square of $(Z_{n,\mathrm{H}}^{\mathsf{T}}U_{n,\mathrm{H}}(a_0)/n)_{n\geq 1}$ to 0 and thus the convergence in probability of the generalized least squares estimator $\hat{a}_n$ defined in (4.9) to the true parameter a_0.

We use the stationarity of the process $(e_t^{\mathrm{H}}(a_0))_{t\in\mathbb{Z}}$ and we follow the same calculation procedures developed to obtain the explicit expression of $\Delta_{2,n,\mathrm{H}}$ given in (4.37). This leads to:

$$\begin{aligned}
\Delta_{3,n,\mathrm{H}} &= \frac{1}{n^2}\sum_{j_1,j_2=1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_2=1}^{n}\sum_{r_3=j_2}^{n}\sum_{r_4=1}^{n}\left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_1,r_2}\left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_3,r_4}\left(\Omega_{n,\mathrm{H}}\right)_{r_1+1-j_1,r_3+1-j_2}\left(\Omega_{n+1,\mathrm{H}}\right)_{r_2+1,r_4+1}\\
&= \frac{1}{n^2}\sum_{j_1,j_2=1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_2=1}^{n}\sum_{r_3=j_2}^{n}\sum_{r_4=1}^{n}\left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_1,r_2}\left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_3,r_4}\left(\Omega_{n,\mathrm{H}}\right)_{r_1+1-j_1,r_3+1-j_2}\left(\Omega_{n,\mathrm{H}}\right)_{r_2,r_4}\\
&= \frac{1}{n^2}\sum_{j_1,j_2=1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_3=j_2}^{n}\sum_{r_4=1}^{n}\left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_3,r_4}\left(\Omega_{n,\mathrm{H}}\right)_{r_1+1-j_1,r_3+1-j_2}\left(I_n\right)_{r_1,r_4}\\
&= \frac{1}{n^2}\sum_{j_1,j_2=1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_3=j_2}^{n}\left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_3,r_1}\left(\Omega_{n,\mathrm{H}}\right)_{r_1+1-j_1,r_3+1-j_2}\\
&= \frac{1}{n^2}\sum_{j_1=1}^{n}\sum_{j_2=1}^{j_1} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_3=j_2}^{n}\left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_3,r_1}\left(\Omega_{n,\mathrm{H}}\right)_{r_1+1-j_1,r_3+1-j_2}\\
&\qquad + \frac{1}{n^2}\sum_{j_1=1}^{n}\sum_{j_2=j_1+1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_3=j_2}^{n}\left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_3,r_1}\left(\Omega_{n,\mathrm{H}}\right)_{r_1+1-j_1,r_3+1-j_2}.
\end{aligned}$$

The stationarity of $(\epsilon_t^{\mathrm{H}})_{t\in\mathbb{Z}}$ implies that:

— for $j_2 \leq j_1$:

$$\left(\Omega_{n,\mathrm{H}}\right)_{r_1+1-j_1,r_3+1-j_2} = \left(\Omega_{n,\mathrm{H}}\right)_{r_1+j_2-j_1,r_3},$$

— for $j_2 > j_1$:

$$\left(\Omega_{n,\mathrm{H}}\right)_{r_1+1-j_1,r_3+1-j_2} = \left(\Omega_{n,\mathrm{H}}\right)_{r_1,r_3+j_1-j_2}.$$

It follows that

$$\begin{aligned}\Delta_{3,n,\mathrm{H}} &= \frac{1}{n^2}\sum_{j_1=1}^{n}\sum_{j_2=1}^{j_1} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_3=j_2}^{n}\left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_3,r_1}\left(\Omega_{n,\mathrm{H}}\right)_{r_1+j_2-j_1,r_3}\\ &\quad+\frac{1}{n^2}\sum_{j_1=1}^{n}\sum_{j_2=j_1+1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_3=j_2}^{n}\left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_3,r_1}\left(\Omega_{n,\mathrm{H}}\right)_{r_1,r_3+j_1-j_2}\\ &= \frac{1}{n^2}\sum_{j_1=1}^{n}\sum_{j_2=1}^{j_1} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_3=1}^{n}\left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_3,r_1}\left(\Omega_{n,\mathrm{H}}\right)_{r_1+j_2-j_1,r_3}\\ &\quad-\frac{1}{n^2}\sum_{j_1=1}^{n}\sum_{j_2=1}^{j_1} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_3=1}^{j_2-1}\left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_3,r_1}\left(\Omega_{n,\mathrm{H}}\right)_{r_1+j_2-j_1,r_3}\\ &\quad+\frac{1}{n^2}\sum_{j_1=1}^{n}\sum_{j_2=j_1+1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=1}^{n}\sum_{r_3=j_2}^{n}\left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_3,r_1}\left(\Omega_{n,\mathrm{H}}\right)_{r_1,r_3+j_1-j_2}\\ &\quad-\frac{1}{n^2}\sum_{j_1=1}^{n}\sum_{j_2=j_1+1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=1}^{j_1-1}\sum_{r_3=j_2}^{n}\left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_3,r_1}\left(\Omega_{n,\mathrm{H}}\right)_{r_1,r_3+j_1-j_2}.\end{aligned} \tag{4.41}$$

Thus we have

$$\begin{aligned}\Delta_{3,n,\mathrm{H}} &= \frac{1}{n^2}\sum_{j_1=1}^{n}(n-j_1+1)a_0^{2j_1-2}-\frac{1}{n^2}\sum_{j_1=1}^{n}\sum_{j_2=1}^{j_1} a_0^{j_1+j_2-2}\sum_{r_1=j_1}^{n}\sum_{r_3=1}^{j_2-1}\left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_3,r_1}\left(\Omega_{n,\mathrm{H}}\right)_{r_1+j_2-j_1,r_3}\\ &\quad-\frac{1}{n^2}\sum_{j_1=1}^{n}\sum_{j_2=j_1+1}^{n} a_0^{j_1+j_2-2}\sum_{r_1=1}^{j_1-1}\sum_{r_3=j_2}^{n}\left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_3,r_1}\left(\Omega_{n,\mathrm{H}}\right)_{r_1,r_3+j_1-j_2}\\ &= \frac{1}{n}\mathbb{E}\left[\frac{1}{n}\left\|Z_{n,\mathrm{H}}\right\|_{\mathbb{R}^n}^2\right].\end{aligned} \tag{4.42}$$

Now, using Lemma 4.1 we deduce that

$$\Delta_{3,n,\mathrm{H}} \xrightarrow[n\to\infty]{} 0.$$

The lemma is then proved. □

The convergences established in Lemmas 4.3 and 4.4 imply that the joint vector

$$\left(Z_{n,\mathrm{H}}^{\mathsf{T}}U_{n,\mathrm{H}}(a_0), \left\|Z_{n,\mathrm{H}}\right\|_{\mathbb{R}^n}^2\right)/n$$

converges in probability to $(0, 1/(1-a_0^2))$. Let now g be the continuous function defined on $\mathbb{R}\times\mathbb{R}^*$ by $g(x,y) = xy^{-1}$, it is obvious that the vector $(0, 1/(1-a_0^2))$ belongs to the set of continuity points of the function g. The continuous mapping theorem yields

$$\hat{a}_n - a_0 = g\left(Z_{n,\mathrm{H}}^{\mathsf{T}}U_{n,\mathrm{H}}(a_0)/n, \left\|Z_{n,\mathrm{H}}\right\|_{\mathbb{R}^n}^2/n\right) \xrightarrow[n\to\infty]{\mathbb{P}} g\left(0, 1/(1-a_0^2)\right) = 0.$$

The proof of Theorem 4.1 is then complete.

Remark 4.5. *The following equivalent formulation of Equality* (4.38)

$$\mathbb{E}\left[\frac{1}{\sqrt{n}}\left\{Z_{n,\mathrm{H}}^{\mathsf{T}}U_{n,\mathrm{H}}(a_0)\right\}\right]=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}a_0^{j-1}\sum_{r_1=j}^{n}\sum_{r_2=1}^{n}\left(\Omega_{n,\mathrm{H}}^{-1}\right)_{r_1,r_2}\mathbb{E}\left[e_{r_1+1-j}^{\mathrm{H}}(a_0)e_{r_2+1}^{\mathrm{H}}(a_0)\right] \tag{4.43}$$

will be helpful in the proof of Theorem 4.2.

4.4.3 Proof of Theorem 4.2

We establish in this section the proof of the asymptotic normality of the generalized least squares estimator $\hat{a}_n$ given in (4.9) under the hypothesis of the stationarity of $(X_t)_{t\in\mathbb{Z}}$ (*i.e.* when the true parameter a_0 is assumed to belong to $]-1,1[$) and in the case where H is known.

Recall that the normal distribution admits moments of all orders and is completely determined by its moments. The expected convergence in distribution stated in Theorem 4.2 can be obtained using the method of moments (see Theorem 30.2. of Billingsley [1986]).

In our case, Theorem 4.2 will be proved as soon as for any $r\geq 1$

$$\lim_{n\to\infty}\mathbb{E}\left[\left\{\sqrt{n}(\hat{a}_n-a_0)\right\}^r\right]=\mathbb{E}\left[G_1^r\right], \tag{4.44}$$

where G_1 is normally distributed with mean 0 and variance $1-a_0^2$.

In view of the explicit expression of $\hat{a}_n-a_0$ (see (4.10)), the random variable $\sqrt{n}(\hat{a}_n-a_0)$ can be rewritten in the form:

$$\sqrt{n}(\hat{a}_n-a_0)=\frac{\frac{1}{\sqrt{n}}Z_{n,\mathrm{H}}^{\mathsf{T}}U_{n,\mathrm{H}}(a_0)}{\frac{1}{n}\left\|Z_{n,\mathrm{H}}\right\|_{\mathbb{R}^n}^2}. \tag{4.45}$$

From Lemma 4.3, the random variable $\left\|Z_{n,\mathrm{H}}\right\|_{\mathbb{R}^n}^2/n$ converges in mean-square to $1/(1-a_0^2)$. This implies that the sequence $\left(\left\|Z_{n,\mathrm{H}}\right\|_{\mathbb{R}^n}^2/n\right)_{n\geq 1}$ converges in probability to $1/(1-a_0^2)$.

The idea of the proof is to show at first using the method of moments that $Z_{n,\mathrm{H}}^{\mathsf{T}}U_{n,\mathrm{H}}(a_0)/\sqrt{n}$ has a limiting centered normal distribution with variance $1/(1-a_0^2)$. The proof will be concluded by the application of Slutsky's theorem on variables $\left\|Z_{n,\mathrm{H}}\right\|_{\mathbb{R}^n}^2/n$ and $Z_{n,\mathrm{H}}^{\mathsf{T}}U_{n,\mathrm{H}}(a_0)/\sqrt{n}$.

The asymptotic distribution of the sequence $\left(Z_{n,\mathrm{H}}^{\mathsf{T}}U_{n,\mathrm{H}}(a_0)/\sqrt{n}\right)_{n\geq 1}$ is given in the following lemma.

Lemma 4.5. *Under Assumption* ***(A0)****, the sequence of random variables* $\left(Z_{n,\mathrm{H}}^{\mathsf{T}}U_{n,\mathrm{H}}(a_0)/\sqrt{n}\right)_{n\geq 1}$ *has a limiting centered normal distribution with variance* $1/(1-a_0^2)$.

Proof. Using the existence of moments of all orders of $Z_{n,\mathrm{H}}^{\mathsf{T}}U_{n,\mathrm{H}}(a_0)/\sqrt{n}$ and in view of Theorem 30.2. of Billingsley [1986], the proof of Lemma 4.5 is obtained once

$$\lim_{n\to\infty}\mathbb{E}\left[\left\{\frac{1}{\sqrt{n}}Z_{n,\mathrm{H}}^{\mathsf{T}}U_{n,\mathrm{H}}(a_0)\right\}^r\right]=\mathbb{E}\left[G_2^r\right],\ \text{for all } r\geq 1, \tag{4.46}$$

where G_2 is is normally distributed with mean 0 and variance $1/(1-a_0^2)$.

To prove (4.46), in a first step we consider an odd order r and in a second step r will be assumed to be even.

Recall that for all $r \geq 1$, we have

$$\mathbb{E}\left[G_2^r\right] = \begin{cases} 0 & \text{if } r \text{ is odd} \\ \left\{1/(1-a_0^2)\right\}^{r/2}(r-1)!! & \text{if } r \text{ is even,} \end{cases} \tag{4.47}$$

where $r!! = r(r-2)(r-4)\cdots 1$.

We start from (4.16), we use the fact that the random vector $(e_1^{\mathsf{H}}(a_0),\ldots,e_{n+1}^{\mathsf{H}}(a_0))^{\mathsf{T}}$ is a centered Gaussian random vector and we exploit the expression of moments in terms of cumulants to obtain that

$$\begin{aligned}
&\mathbb{E}\left[\left\{\frac{1}{\sqrt{n}}Z_{n,\mathsf{H}}^{\mathsf{T}}U_{n,\mathsf{H}}(a_0)\right\}^r\right] \\
&= \frac{1}{n^{r/2}}\sum_{j_1,j_2,\ldots,j_r=1}^{n} a_0^{\sum_{i=1}^r (j_i-1)} \sum_{k_1=j_1}^{n}\sum_{k_2=1}^{n}\sum_{k_3=j_2}^{n}\sum_{k_4=1}^{n}\cdots\sum_{k_{2r-1}=j_r}^{n}\sum_{k_{2r}=1}^{n}\left\{\prod_{i=1}^{r}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{k_{2i-1},k_{2i}}\right. \\
&\qquad \left.\times\mathbb{E}\left[\prod_{i=1}^{r} e_{k_{2i-1}+1-j_i}^{\mathsf{H}}(a_0)e_{k_{2i}+1}^{\mathsf{H}}(a_0)\right]\right\} \\
&= \frac{1}{n^{r/2}}\sum_{j_1,j_2,\ldots,j_r=1}^{n} a_0^{\sum_{i=1}^r (j_i-1)} \sum_{k_1=j_1}^{n}\sum_{k_2=1}^{n}\sum_{k_3=j_2}^{n}\sum_{k_4=1}^{n}\cdots\sum_{k_{2r-1}=j_r}^{n}\sum_{k_{2r}=1}^{n}\left\{\prod_{i=1}^{r}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{k_{2i-1},k_{2i}}\right. \\
&\qquad \left.\times\sum_{\pi}\prod_{B\in\pi}\operatorname{cum}\left(e_i^{\mathsf{H}}(a_0) : i\in B\right)\right\},
\end{aligned} \tag{4.48}$$

where π runs through the list of all unordered pair partitions of $\{k_1+1-j_1, k_2+1, k_3+1-j_2, k_4+1,\ldots,k_{2r-1}+1-j_r, k_{2r}+1\}$ and B runs through the list of all blocks of the partition π (see Brillinger [1981]).

Remark 4.6. *In the above expression, we only kept unordered pair partitions of* $\{k_1+1-j_1, k_2+1, k_3+1-j_2, k_4+1,\ldots,k_{2r-1}+1-j_r, k_{2r}+1\}$. *This is due to the fact that the random vector* $(e_1^{\mathsf{H}}(a_0),\ldots,e_{n+1}^{\mathsf{H}}(a_0))^{\mathsf{T}}$ *is a centered Gaussian vector.*

Any joint cumulant involving three (or more) elements of $(e_1^{\mathsf{H}}(a_0),\ldots,e_{n+1}^{\mathsf{H}}(a_0))^{\mathsf{T}}$ *is zero. Moreover, since the joint cumulant of just one random variable is its expectation and the process* $(e_t^{\mathsf{H}}(a_0))_{t\in\mathbb{Z}}$ *is centered, the first cumulant of any element of* $(e_1^{\mathsf{H}}(a_0),\ldots,e_{n+1}^{\mathsf{H}}(a_0))^{\mathsf{T}}$ *vanishes.*

Thanks to the above remark, we may write (4.48) as

$$\begin{aligned}
&\mathbb{E}\left[\left\{\frac{1}{\sqrt{n}}Z_{n,\mathsf{H}}^{\mathsf{T}}U_{n,\mathsf{H}}(a_0)\right\}^r\right] \\
&= \frac{1}{n^{r/2}}\sum_{j_1,j_2,\ldots,j_r=1}^{n} a_0^{\sum_{i=1}^{r}(j_i-1)}\sum_{k_1=j_1}^{n}\sum_{k_2=1}^{n}\sum_{k_3=j_2}^{n}\sum_{k_4=1}^{n}\cdots\sum_{k_{2r-1}=j_r}^{n}\sum_{k_{2r}=1}^{n}\left\{\prod_{i=1}^{r}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{k_{2i-1},k_{2i}}\right. \\
&\qquad\left.\times\sum_{\pi}\prod_{B\in\pi}\mathbb{E}\left[e_i^{\mathsf{H}}(a_0) : i\in B\right]\right\}. \qquad (4.49)
\end{aligned}$$

We strength the fact that when $r=1$, (4.49) is (4.43) and for $r=2$ we retrieve an analogous formula to (4.36).

The following remark will help us to develop more explicitly (4.48).

Remark 4.7. — *In Equation (4.48), the exchange of the roles of indices with even sub-indices does not modify the term in the right hand side. It is the same when we exchange the roles of indices with odd sub-indices. For example, for the indices k_1 and k_3 it holds that*

$$\begin{aligned}
&\frac{1}{n^{r/2}}\sum_{j_1,j_2,\ldots,j_r=1}^{n} a_0^{\sum_{i=1}^{r}(j_i-1)}\sum_{\mathbf{k_1}=\mathbf{j_1}}^{n}\sum_{k_2=1}^{n}\sum_{\mathbf{k_3}=\mathbf{j_2}}^{n}\sum_{k_4=1}^{n}\cdots\sum_{k_{2r-1}=j_r}^{n}\sum_{k_{2r}=1}^{n}\left\{\prod_{i=1}^{r}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{k_{2i-1},k_{2i}}\right. \\
&\qquad\left.\times\sum_{\pi}\prod_{B\in\pi}\operatorname{cum}\left(e_i^{\mathsf{H}}(a_0) : i\in B\right)\right\} \\
&= \frac{1}{n^{r/2}}\sum_{j_1,j_2,\ldots,j_r=1}^{n} a_0^{\sum_{i=1}^{r}(j_i-1)}\sum_{\mathbf{k_1}=\mathbf{j_2}}^{n}\sum_{k_2=1}^{n}\sum_{\mathbf{k_3}=\mathbf{j_1}}^{n}\sum_{k_4=1}^{n}\cdots\sum_{k_{2r-1}=j_r}^{n}\sum_{k_{2r}=1}^{n}\left\{\prod_{i=1}^{r}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{k_{2i-1},k_{2i}}\right. \\
&\qquad\left.\times\sum_{\pi}\prod_{B\in\pi}\operatorname{cum}\left(e_i^{\mathsf{H}}(a_0) : i\in B\right)\right\},
\end{aligned}$$

— *Again in Equation (4.48) when r is odd, every pair partition π of the set of indices $\{k_1+1-j_1, k_2+1, k_3+1-j_2, k_4+1, \ldots, k_{2r-1}+1-j_r, k_{2r}+1\}$ contains at least one block of the type $\{k_{2i-1}+1-j_i, k_{i'}+1\}$. To see that, let us consider the case when $r=3$. The third moment of $Z_{n,\mathsf{H}}^{\mathsf{T}}U_{n,\mathsf{H}}(a_0)/\sqrt{n}$ is given by:*

$$\begin{aligned}
&\mathbb{E}\left[\left\{\frac{1}{\sqrt{n}}Z_{n,\mathsf{H}}^{\mathsf{T}}U_{n,\mathsf{H}}(a_0)\right\}^3\right] \\
&= \frac{1}{n^{3/2}}\sum_{j_1,j_2,j_3=1}^{n} a_0^{j_1+j_2+j_3-3}\sum_{r_1=j_1}^{n}\sum_{r_2=1}^{n}\sum_{r_3=j_2}^{n}\sum_{r_4=1}^{n}\sum_{r_5=j_3}^{n}\sum_{r_6=1}^{n}\left\{\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_1,r_2}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_3,r_4}\left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_5,r_6}\right. \\
&\qquad\left.\times\,\mathbb{E}\left[e_{r_1+1-j_1}^{\mathsf{H}}(a_0)e_{r_2+1}^{\mathsf{H}}(a_0)e_{r_3+1-j_2}^{\mathsf{H}}(a_0)e_{r_4+1}^{\mathsf{H}}(a_0)e_{r_5+1-j_3}^{\mathsf{H}}(a_0)e_{r_6+1}^{\mathsf{H}}(a_0)\right]\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n^{3/2}} \sum_{j_1,j_2,j_3=1}^{n} a_0^{j_1+j_2+j_3-3} \sum_{r_1=j_1}^{n} \sum_{r_2=1}^{n} \sum_{r_3=j_2}^{n} \sum_{r_4=1}^{n} \sum_{r_5=j_3}^{n} \sum_{r_6=1}^{n} \Bigg\{ \left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_1,r_2} \left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_3,r_4} \left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_5,r_6} \\
&\quad \times \Bigg[\mathbb{E}\left[e^{\mathsf{H}}_{r_1+1-j_1}(a_0) e^{\mathsf{H}}_{r_2+1}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_3+1-j_2}(a_0) e^{\mathsf{H}}_{r_4+1}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_5+1-j_3}(a_0) e^{\mathsf{H}}_{r_6+1}(a_0)\right] \\
&\quad + \mathbb{E}\left[e^{\mathsf{H}}_{r_1+1-j_1}(a_0) e^{\mathsf{H}}_{r_2+1}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_3+1-j_2}(a_0) e^{\mathsf{H}}_{r_5+1-j_3}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_4+1}(a_0) e^{\mathsf{H}}_{r_6+1}(a_0)\right] \\
&\quad + \mathbb{E}\left[e^{\mathsf{H}}_{r_1+1-j_1}(a_0) e^{\mathsf{H}}_{r_2+1}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_3+1-j_2}(a_0) e^{\mathsf{H}}_{r_6+1}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_4+1}(a_0) e^{\mathsf{H}}_{r_5+1-j_3}(a_0)\right] \\
&\quad + \mathbb{E}\left[e^{\mathsf{H}}_{r_1+1-j_1}(a_0) e^{\mathsf{H}}_{r_3+1-j_2}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_2+1}(a_0) e^{\mathsf{H}}_{r_4+1}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_5+1-j_3}(a_0) e^{\mathsf{H}}_{r_6+1}(a_0)\right] \\
&\quad + \mathbb{E}\left[e^{\mathsf{H}}_{r_1+1-j_1}(a_0) e^{\mathsf{H}}_{r_3+1-j_2}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_2+1}(a_0) e^{\mathsf{H}}_{r_5+1-j_3}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_4+1}(a_0) e^{\mathsf{H}}_{r_6+1}(a_0)\right] \\
&\quad + \mathbb{E}\left[e^{\mathsf{H}}_{r_1+1-j_1}(a_0) e^{\mathsf{H}}_{r_3+1-j_2}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_2+1}(a_0) e^{\mathsf{H}}_{r_6+1}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_4+1}(a_0) e^{\mathsf{H}}_{r_5+1-j_3}(a_0)\right] \\
&\quad + \mathbb{E}\left[e^{\mathsf{H}}_{r_1+1-j_1}(a_0) e^{\mathsf{H}}_{r_4+1}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_2+1}(a_0) e^{\mathsf{H}}_{r_3+1-j_2}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_5+1-j_3}(a_0) e^{\mathsf{H}}_{r_6+1}(a_0)\right] \\
&\quad + \mathbb{E}\left[e^{\mathsf{H}}_{r_1+1-j_1}(a_0) e^{\mathsf{H}}_{r_4+1}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_2+1}(a_0) e^{\mathsf{H}}_{r_5+1-j_3}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_3+1-j_2}(a_0) e^{\mathsf{H}}_{r_6+1}(a_0)\right] \\
&\quad + \mathbb{E}\left[e^{\mathsf{H}}_{r_1+1-j_1}(a_0) e^{\mathsf{H}}_{r_4+1}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_2+1}(a_0) e^{\mathsf{H}}_{r_6+1}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_3+1-j_2}(a_0) e^{\mathsf{H}}_{r_5+1-j_3}(a_0)\right] \\
&\quad + \mathbb{E}\left[e^{\mathsf{H}}_{r_1+1-j_1}(a_0) e^{\mathsf{H}}_{r_5+1-j_3}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_2+1}(a_0) e^{\mathsf{H}}_{r_3+1-j_2}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_4+1}(a_0) e^{\mathsf{H}}_{r_6+1}(a_0)\right] \\
&\quad + \mathbb{E}\left[e^{\mathsf{H}}_{r_1+1-j_1}(a_0) e^{\mathsf{H}}_{r_5+1-j_3}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_2+1}(a_0) e^{\mathsf{H}}_{r_4+1}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_3+1-j_2}(a_0) e^{\mathsf{H}}_{r_6+1}(a_0)\right] \\
&\quad + \mathbb{E}\left[e^{\mathsf{H}}_{r_1+1-j_1}(a_0) e^{\mathsf{H}}_{r_5+1-j_3}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_2+1}(a_0) e^{\mathsf{H}}_{r_6+1}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_3+1-j_2}(a_0) e^{\mathsf{H}}_{r_4+1}(a_0)\right] \\
&\quad + \mathbb{E}\left[e^{\mathsf{H}}_{r_1+1-j_1}(a_0) e^{\mathsf{H}}_{r_6+1}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_2+1}(a_0) e^{\mathsf{H}}_{r_3+1-j_2}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_4+1}(a_0) e^{\mathsf{H}}_{r_5+1-j_3}(a_0)\right] \\
&\quad + \mathbb{E}\left[e^{\mathsf{H}}_{r_1+1-j_1}(a_0) e^{\mathsf{H}}_{r_6+1}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_2+1}(a_0) e^{\mathsf{H}}_{r_4+1}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_3+1-j_2}(a_0) e^{\mathsf{H}}_{r_5+1-j_3}(a_0)\right] \\
&\quad + \mathbb{E}\left[e^{\mathsf{H}}_{r_1+1-j_1}(a_0) e^{\mathsf{H}}_{r_6+1}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_2+1}(a_0) e^{\mathsf{H}}_{r_5+1-j_3}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_3+1-j_2}(a_0) e^{\mathsf{H}}_{r_4+1}(a_0)\right] \Bigg] \Bigg\}.
\end{aligned}
$$

We can use this last remark and Equations (4.43) and (4.36) to show that

$$
\begin{aligned}
\mathbb{E}\left[\left\{\frac{1}{\sqrt{n}} Z_{n,\mathsf{H}}^{\mathsf{T}} U_{n,\mathsf{H}}(a_0)\right\}^3\right] &= \left(\frac{1}{\sqrt{n}} \sum_{j_1=1}^{n} a_0^{j_1-1} \sum_{r_1=j_1}^{n} \sum_{r_2=1}^{n} \left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_1,r_2} \mathbb{E}\left[e^{\mathsf{H}}_{r_1+1-j_1}(a_0) e^{\mathsf{H}}_{r_2+1}(a_0)\right]\right) \\
&\quad \times \Bigg(\frac{1}{n} \sum_{j_2,j_3=1}^{n} a_0^{j_2+j_3-2} \sum_{r_3=j_2}^{n} \sum_{r_4=1}^{n} \sum_{r_5=j_3}^{n} \sum_{r_6=1}^{n} \Bigg\{ \left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_3,r_4} \left(\Omega_{n,\mathsf{H}}^{-1}\right)_{r_5,r_6} \\
&\quad \times \Bigg[\mathbb{E}\left[e^{\mathsf{H}}_{r_3+1-j_2}(a_0) e^{\mathsf{H}}_{r_4+1}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_5+1-j_3}(a_0) e^{\mathsf{H}}_{r_6+1}(a_0)\right] \\
&\quad + \mathbb{E}\left[e^{\mathsf{H}}_{r_3+1-j_2}(a_0) e^{\mathsf{H}}_{r_5+1-j_3}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_4+1}(a_0) e^{\mathsf{H}}_{r_6+1}(a_0)\right] \\
&\quad + \mathbb{E}\left[e^{\mathsf{H}}_{r_3+1-j_2}(a_0) e^{\mathsf{H}}_{r_6+1}(a_0)\right] \mathbb{E}\left[e^{\mathsf{H}}_{r_4+1}(a_0) e^{\mathsf{H}}_{r_5+1-j_3}(a_0)\right] \Bigg] \Bigg\}\Bigg) \\
&= \mathbb{E}\left[\frac{1}{\sqrt{n}} Z_{n,\mathsf{H}}^{\mathsf{T}} U_{n,\mathsf{H}}(a_0)\right] \mathbb{E}\left[\left\{\frac{1}{\sqrt{n}} Z_{n,\mathsf{H}}^{\mathsf{T}} U_{n,\mathsf{H}}(a_0)\right\}^2\right].
\end{aligned}
$$

In view of (4.36), (4.39) and Remarks 4.6 and 4.7, one can rewrite the r-th moment of the random variable $Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)/\sqrt{n}$ when r is odd in the form:

$$\mathbb{E}\left[\left\{\frac{1}{\sqrt{n}} Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)\right\}^r\right] = \mathbb{E}\left[\frac{1}{\sqrt{n}} Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)\right] \mathbb{E}\left[\left\{\frac{1}{\sqrt{n}} Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)\right\}^{r-1}\right]. \tag{4.50}$$

The calculation given in the proof of Lemma 4.4 to show the convergence to 0 of the sequence $(\mathbb{E}[Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)/n])_{n\geq 1}$ remains valid when we multiply the general term of the last sequence by $\sqrt{n}$ and when we use the fractional version of Cesàro's Lemma [1] to conclude.

In fact, from (4.40) we have

$$\left|\mathbb{E}\left[\frac{1}{\sqrt{n}}\left\{Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)\right\}\right]\right| \leq K\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}|a_0|^{j-1} + \frac{1}{\sqrt{n}}\sum_{j=1}^{n}|a_0|^{j-1}\sum_{\substack{1\leq r\leq n\\ r\neq j}}\frac{1}{(r-j)^{2\mathrm{H}}}\frac{1}{r^{2-2\mathrm{H}}}\right). \tag{4.51}$$

Since $\sqrt{j}\,|a_0|^{j-1} \to 0$ when $j \to \infty$, fractional version of Cesàro's Lemma implies that the first term in the right hand side of (4.51) tends to 0 when n goes to infinity. Based on the calculation given in the proof of Lemma 4.4 and using the fractional version of Cesàro's Lemma, the second term in the right hand side of (4.51) verifies

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{j=1}^{n}|a_0|^{j-1}\sum_{\substack{1\leq r\leq n\\ r\neq j}}\frac{1}{(r-j)^{2\mathrm{H}}}\frac{1}{r^{2-2\mathrm{H}}}\\
&\qquad\leq \frac{1}{\sqrt{n}}\sum_{r=1}^{n}\frac{1}{r^{1-2\mathrm{H}}}|a_0|^{r-1}\max\left(\frac{1}{|a_0|},\frac{1}{|a_0|^r r^{2\mathrm{H}}}\right) + \frac{1}{\sqrt{n}}\sum_{r=1}^{n}\frac{1}{r^{2-2\mathrm{H}}}|a_0|^r\left(\frac{1-|a_0|^{n-r}}{1-|a_0|}\right)\\
&\qquad\xrightarrow[n\to\infty]{} 0.
\end{aligned}$$

Then the sequence $(\mathbb{E}[Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)/\sqrt{n}])_{n\geq 1}$ converges to 0.

Furthermore, Equation (4.42) and Lemma 4.1 imply that

$$\lim_{n\to\infty}\mathbb{E}\left[\left\{\frac{1}{\sqrt{n}} Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)\right\}^2\right] = \lim_{n\to\infty}\mathbb{E}\left[\frac{1}{n}\left\|Z_{n,\mathrm{H}}\right\|_{\mathbb{R}^n}^2\right] = \frac{1}{1-a_0^2}.$$

Using (4.50) we deduce that

$$\lim_{n\to\infty}\mathbb{E}\left[\left\{\frac{1}{\sqrt{n}} Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)\right\}^3\right] = 0. \tag{4.52}$$

1. Recall that the fractional version of Cesàro's Lemma states that for $(h_t)_{t\geq 0}$ a sequence of positive real numbers, $\kappa > 0$ and $c \geq 0$ we have

$$\lim_{t\to\infty} h_t t^{1-\kappa} = |\kappa|\, c \Rightarrow \lim_{n\to\infty}\frac{1}{n^\kappa}\sum_{t=0}^{n} h_t = c.$$

When $r = 4$, the set $\{k_1 + 1 - j_1, k_2 + 1, k_3 + 1 - j_2, k_4 + 1, \ldots, k_{2r-1} + 1 - j_r, k_{2r} + 1\}$ can be partitioned by pairs in two different ways. The first is to consider partitions that contain at least one block of type $\{k_{2i-1} + 1 - j_i, k_{i'} + 1\}$ and the second is to take partitions that do not contain any such block. In the first case, we use the same remarks introduced to obtain (4.50) and we factorize by the expectation of $Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)/\sqrt{n}$. In the second case, we use the result stated in Lemma 4.1. More precisely, we have

$$\mathbb{E}\left[\left\{\frac{1}{\sqrt{n}} Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)\right\}^4\right] = 3\left(\mathbb{E}\left[\frac{1}{n}\left\|Z_{n,\mathrm{H}}\right\|_{\mathbb{R}^n}^2\right]\right)^2 + \mathbb{E}\left[\frac{1}{\sqrt{n}} Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)\right] \mathbb{E}\left[\left\{\frac{1}{\sqrt{n}} Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)\right\}^3\right].$$

Therefore, in view of Lemma 4.1 and (4.52) we have

$$\lim_{n\to\infty} \mathbb{E}\left[\left\{\frac{1}{\sqrt{n}} Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)\right\}^4\right] = \frac{3}{(1 - a_0^2)^2}.$$

We can generalize this result by an immediate recurrence on r to obtain

$$\mathbb{E}\left[\left\{\frac{1}{\sqrt{n}} Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)\right\}^r\right] = \begin{cases} \mathbb{E}\left[\frac{1}{\sqrt{n}} Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)\right] \mathbb{E}\left[\left\{\frac{1}{\sqrt{n}} Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)\right\}^{r-1}\right] & \text{if } r \text{ is odd} \\ (r-1)!! \left(\mathbb{E}\left[\frac{1}{n}\left\|Z_{n,\mathrm{H}}\right\|_{\mathbb{R}^n}^2\right]\right)^{r/2} + \mathbb{E}\left[\frac{1}{\sqrt{n}} Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)\right] \mathbb{E}\left[\left\{\frac{1}{\sqrt{n}} Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)\right\}^{r-1}\right] & \text{if } r \text{ is even.} \end{cases}$$

This implies that the r-th moment of the random variable $Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)/\sqrt{n}$ converges to

$$\begin{cases} 0 & \text{if } r \text{ is odd} \\ (r-1)!! \left(\frac{1}{1-a_0^2}\right)^{r/2} & \text{if } r \text{ is even,} \end{cases} \tag{4.53}$$

and this is exactly the result recalled in (4.47). The lemma is then proved.

□

The convergences established in Lemmas 4.3 and 4.5 imply (by Slutsky's theorem) that the vector $(Z_{n,\mathrm{H}}^{\mathsf{T}} U_{n,\mathrm{H}}(a_0)/\sqrt{n}, \|Z_{n,\mathrm{H}}\|_{\mathbb{R}^n}^2/n)$ converges in distribution to $(G_2, 1/(1 - a_0^2))$, where G_2 is the random variable defined in the proof of Lemma 4.5. Continuous mapping theorem allows to complete the proof of Theorem 4.2.

4.5 Figures

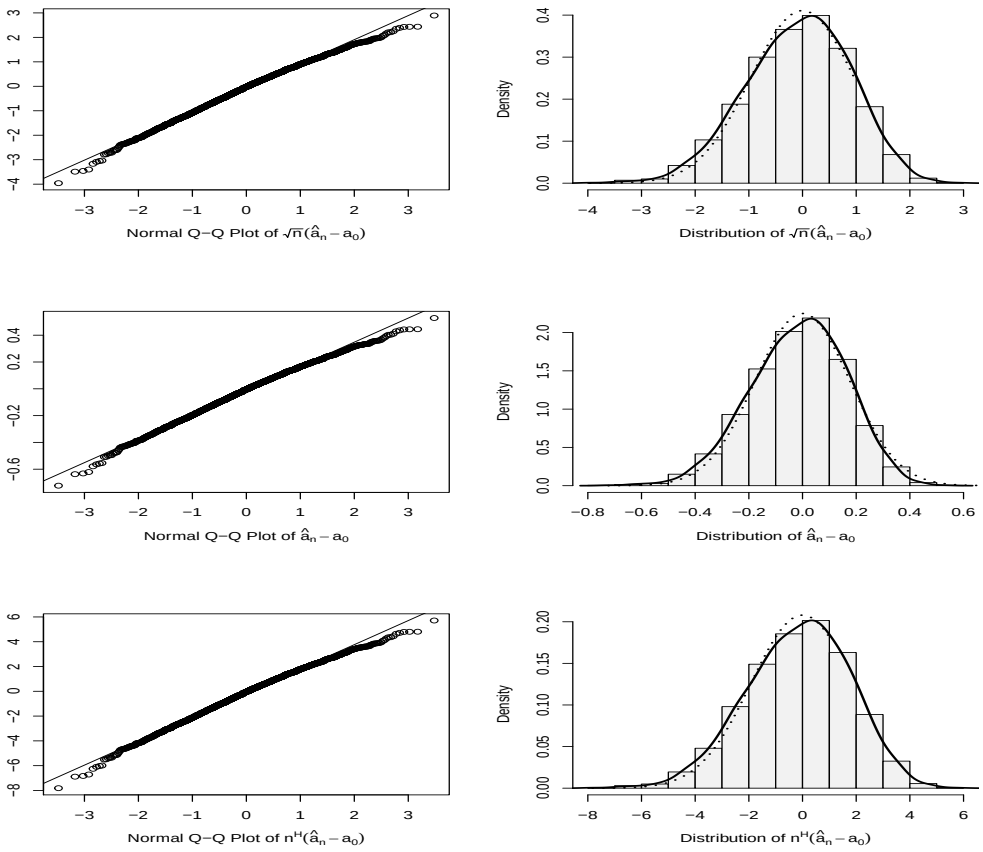


Figure 4.1 – Distribution of the generalized least squares estimator of $N = 2,000$ independent simulations of the fractional autoregressive model (4.11) with size $n = 30$ and unknown parameter $a = 0.3$. The Hurst index is assumed to be equal to 0.7.

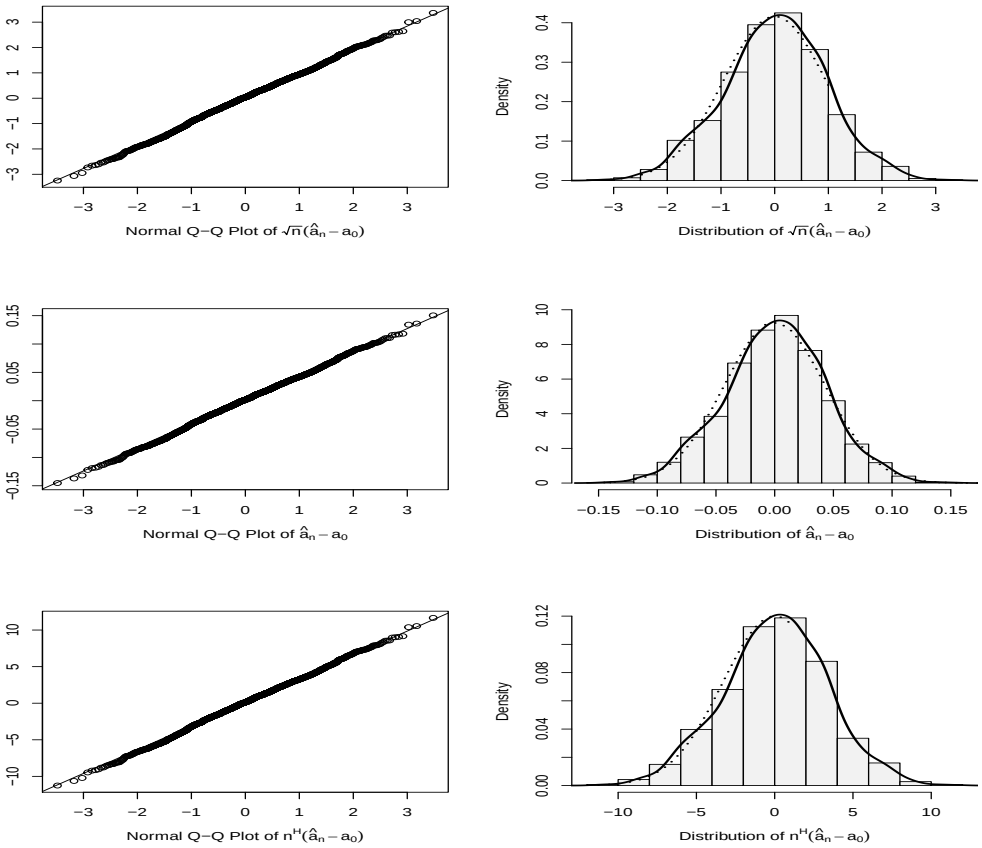


Figure 4.2 – Distribution of the generalized least squares estimator of $N = 2,000$ independent simulations of the fractional autoregressive model (4.11) with size $n = 500$ and unknown parameter $a = 0.3$. The Hurst index is assumed to be equal to 0.7.

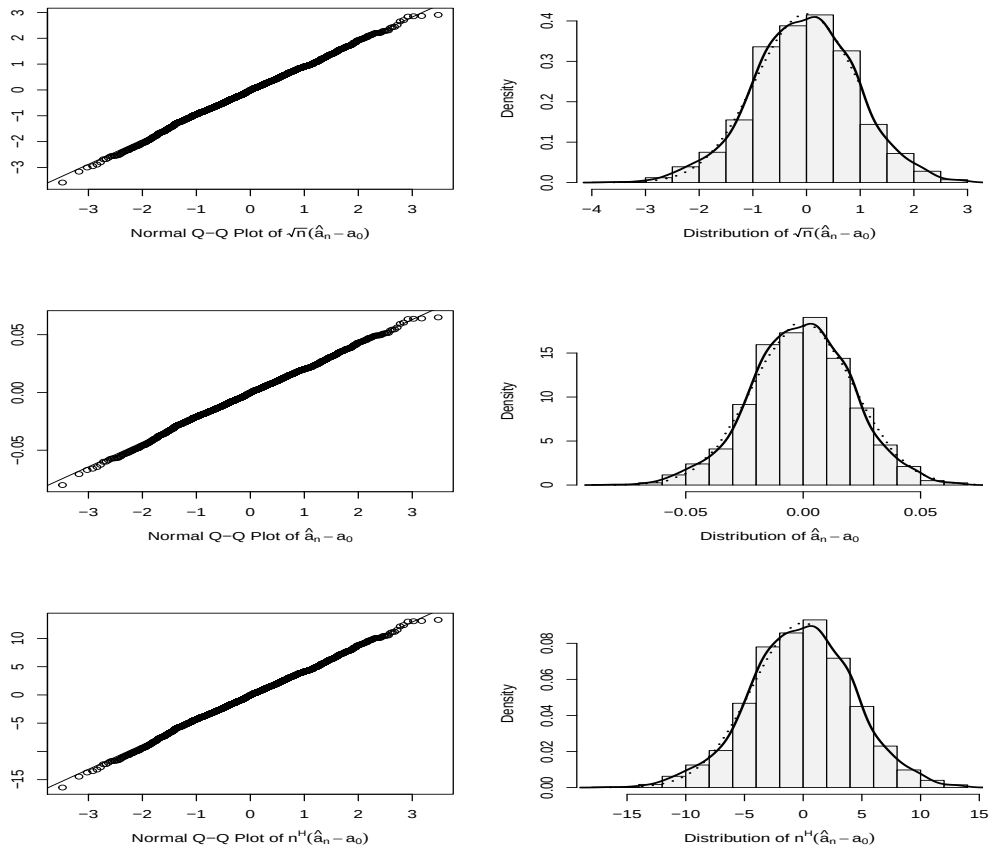


Figure 4.3 – Distribution of the generalized least squares estimator of $N = 2,000$ independent simulations of the fractional autoregressive model (4.11) with size $n = 2,000$ and unknown parameter $a = 0.3$. The Hurst index is assumed to be equal to 0.7.

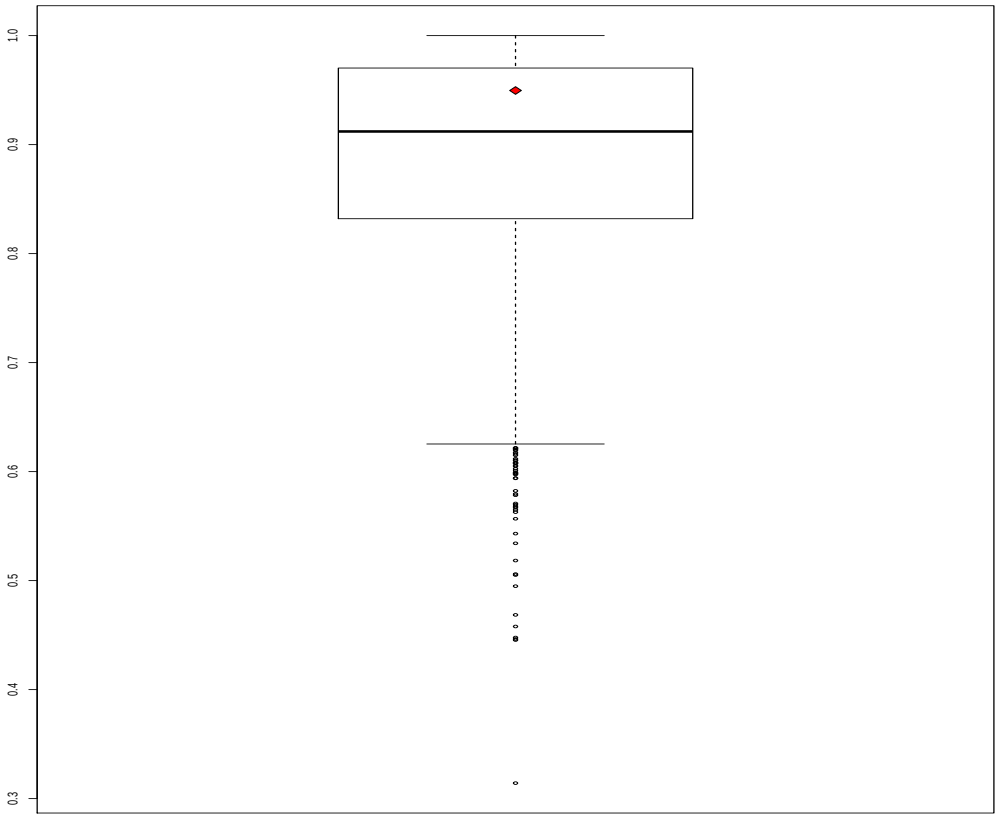


Figure 4.4 – Box plot describing the distribution of the variance estimates of the generalized least squares estimator of $N = 2,000$ independent simulations of the fractional autoregressive model (4.11) with size $n = 30$ and unknown parameter $a = 0.3$. The Hurst index is assumed to be equal to 0.7. The diamond symbol represent the mean of the standardized errors $n(\hat{a}_n - 0.3)^2$, on the simulated models presented in Figure 4.1.

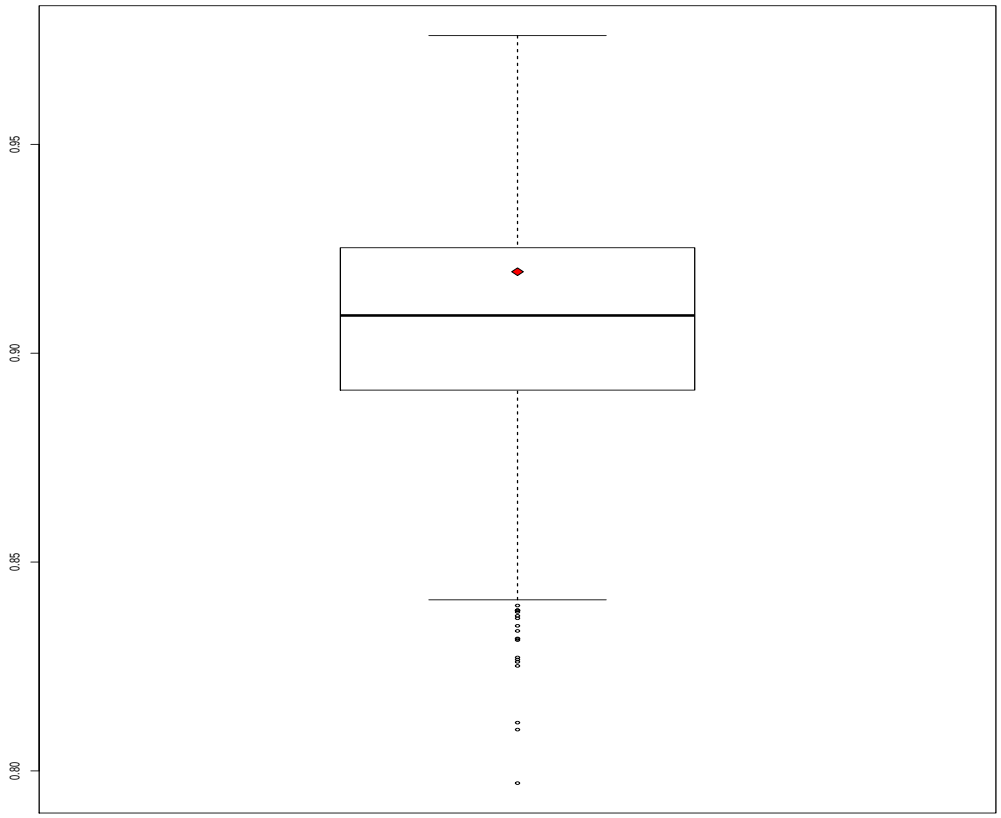


Figure 4.5 – Box plot describing the distribution of the variance estimates of the generalized least squares estimator of $N = 2,000$ independent simulations of the fractional autoregressive model (4.11) with size $n = 500$ and unknown parameter $a = 0.3$. The Hurst index is assumed to be equal to 0.7. The diamond symbol represent the mean of the standardized errors $n(\hat{a}_n - 0.3)^2$, on the simulated models presented in Figure 4.2.

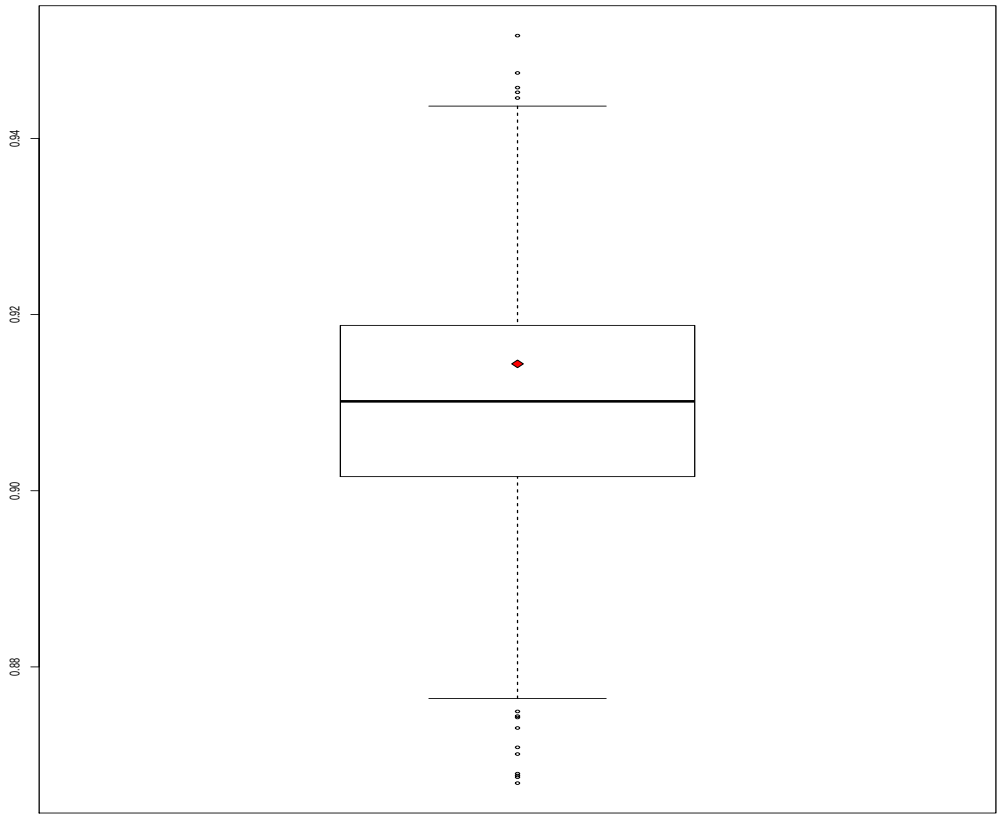


Figure 4.6 – Box plot describing the distribution of the variance estimates of the generalized least squares estimator of $N = 2,000$ independent simulations of the fractional autoregressive model (4.11) with size $n = 2,000$ and unknown parameter $a = 0.3$. The Hurst index is assumed to be equal to 0.7. The diamond symbol represent the mean of the standardized errors $n(\hat{a}_n - 0.3)^2$, on the simulated models presented in Figure 4.3.

Bibliography

Ablowitz, M. J. and Fokas, A. S. (2003). *Complex Variables: Introduction and Applications*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2 edition.

Anderson, T. W. and Taylor, J. B. (1976). Strong consistency of least squares estimates in normal linear regression. *Ann. Statist.*, 4(4):788–790.

Anderson, T. W. and Taylor, J. B. (1979). Strong consistency of least squares estimates in dynamic models. *Ann. Statist.*, 7(3):484–489.

Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2013). *Long-memory processes*. Springer, Heidelberg. Probabilistic properties and statistical methods.

Billingsley, P. (1986). *Probability and measure*. Wiley series in probability and mathematical statistics. Wiley.

Brillinger, D. R. (1981). *Time series: data analysis and theory*, volume 36. Siam.

Brouste, A., Cai, C., and Kleptsyna, M. (2014). Asymptotic properties of the MLE for the autoregressive process coefficients under stationary gaussian noise. *Mathematical Methods of Statistics*, 23(2):103–115.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.*, 17(4):1749–1766.

Fox, R. and Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series. *Ann. Statist.*, 14(2):517–532.

Francq, C. and Zakoïan, J.-M. (1998). Estimating linear representations of nonlinear processes. *J. Statist. Plann. Inference*, 68(1):145–165.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4(4):221–238.

Graf, H.-P. (1980). *Long-range correlations and estimation of the self-similarity parameter*. Ph.D. dissertation, Swiss Federal Institute of Technoloy.

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal.*, 1(1):15–29.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68(1):165–176.

Ito, K. (1944). On the ergodicity of a certain stationary process. *Proc. Imp. Acad.*, 20(2):54–55.

Khinchin, A. I. (1949). Mathematical foundations of statistical mechanics dover publications. *Inc. New York*.

Lai, T., Robbins, H., and Wei, C. (1979). Strong consistency of least squares estimates in multiple regression ii. *Journal of Multivariate Analysis*, 9(3):343 – 361.

Lai, T. L. and Robbins, H. (1977). Strong consistency of least-squares estimates in regression models. *Proceedings of the National Academy of Sciences*, 74(7):2667–2669.

Lai, T. L., Robbins, H., and Wei, C. Z. (1978). Strong consistency of least squares estimates in multiple regression. *Proceedings of the National Academy of Sciences*, 75(7):3034–3036.

Mandelbrot, B. and Taqqu, M. (1979). *Robust R/S Analysis of Long Run Serial Correlation*. IBM Thomas J. Watson Research Division.

Mandelbrot, B. B. (1975). Limit theorems on the self-normalized range for weakly and strongly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 31(4):271–285.

Mandelbrot, B. B. and Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, 10:422–437.

McLeod, A. I. and Hipel, K. W. (1978). Preservation of the rescaled adjusted range: 1. a reassessment of the hurst phenomenon. *Water Resources Research*, 14(3):491–508.

Miller, P. D. (2006). *Applied asymptotic analysis*, volume 75. American Mathematical Soc.

Mohr, D. (1981). *Modeling data as a fractional Gaussian noise*. Ph.D. dissertation, Princeton University.

Palma, W. (2007). *Long-memory time series*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ. Theory and methods.

Sinai, Y. (1976). Self-similar probability distributions. *Theory of Probability & Its Applications*, 21(1):64–80.

Solo, V. (1981). Strong consistency of least squares estimators in regression with correlated disturbances. *Ann. Statist.*, 9(3):689–693.

Sweeting, T. J. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *The Annals of Statistics*, 8(6):1375–1381.

Watson, G. N. (1918). The harmonic functions associated with the parabolic cylinder. *Proceedings of the London Mathematical Society*, s2-17(1):116–148.

Yajima, Y. (1988). On estimation of a regression model with long-memory stationary errors. *Ann. Statist.*, 16(2):791–807.

Conclusion générale et perspectives

Dans cette thèse, nous avons étudié, dans un premier temps, les modèles FARIMA faibles. Nous avons montré que les hypothèses exigées dans le cadre faible sont moins restrictives que celles imposées dans la cadre standard habituellement utilisé dans la littérature. Relâcher l'hypothèse d'indépendance sur le processus d'innovation a permis aux modèles FARIMA faibles d'élargir leurs champs d'application en couvrant une large classe de processus non linéaires à mémoire longue. Sous des hypothèses faibles de régularités sur le bruit, nous avons établi la convergence forte et la normalité asymptotique de l'estimateur des moindres carrés des paramètres des modèles FARIMA(p, d, q) faibles. Nous avons ensuite proposé, par deux approches différentes, un estimateur convergent de la matrice de variance asymptotique de cet estimateur. Cela nous a permis par la suite de construire les intervalles de confiance des paramètres des modèles FARIMA(p, d, q) faibles. Une deuxième méthode basée sur une approche d'auto-normalisation a été également développée pour la construction de ces intervalles de confiance.

Le problème de la validation des modèles FARIMA faibles a aussi été abordé en proposant deux techniques différentes permettant de tester l'adéquation de ces modèles. Nous avons montré que les autocorrélations résiduelles ont une distribution asymptotique normale de matrice de covariance différente de celle obtenue dans le cadre des FARIMA standards avec un bruit indépendant et identiquement distribué (iid). Cela nous a permis de déduire la loi asymptotique exacte des statistiques portmanteau et de proposer ainsi des versions modifiées des tests portmanteau standards de Box-Pierce et Ljung-Box. Dans le cas où le bruit est supposé iid, il est connu que la distribution asymptotique des tests portmanteau est correctement approximée par un khi-deux. Dans le cas général, nous avons montré que cette distribution asymptotique est celle d'une somme pondérée de khi-deux. Elle peut être très différente de l'approximation khi-deux usuelle du cas standard. Afin de contourner le problème de l'estimation de la matrice de covariance asymptotique du vecteur joint de l'estimateur des moindres carrés et des autocovariances empiriques du bruit, une deuxième méthode alternative basée sur une approche d'auto-normalisation a été proposée. Nous avons introduit une nouvelle statistique de test dont la distribution asymptotique dépend uniquement des ordres autorégressifs et moyennes-mobiles. Les valeurs critiques de la loi limite de cette statistique ont été déjà tabulées.

Les logiciels actuellement disponibles n'évaluent pas correctement la précision des estimations des paramètres des représentations FARIMA faibles. Cela peut entraîner de graves erreurs de spécifications, notamment la sélection des modèles avec des ordres très grands. Nous avons justifié une pratique très courante, qui consiste à ajuster un modèle FARIMA, même si les données dont on dispose présentent des non linéarités évidentes.

Dans un second temps, nous avons traité dans cette thèse le problème de l'estimation des modèles autorégressifs d'ordre 1 induits par un bruit gaussien fractionnaire d'indice de Hurst H supposé connu. Nous avons étudié, les propriétés asymptotiques de l'estimateur des moindres carrés généralisés du paramètre autorégressif de ces modèles. Plus précisément, nous avons établi la convergence et la normalité asymptotique de cet estimateur.

Le travail que nous avons réalisé comporte de nombreuses extensions possibles. Parmi ces développements, nous citons :

Estimation non-paramétrique de la matrice de variance asymptotique de l'estimateur des moindres carrés des modèles FARIMA faibles

Dans le chapitre 2, afin d'estimer la matrice $I(\theta_0)$ intervenant dans l'expression de la variance asymptotique $\Omega = J^{-1}(\theta_0)I(\theta_0)J^{-1}(\theta_0)$ de l'estimateur des moindres carrés des paramètres des modèles FARIMA faibles, nous avons proposé une méthode paramétrique basée sur l'estimation de la densité spectrale du processus $(H_t(\theta_0))_{t\in\mathbb{Z}}$. Une autre voie de recherche concernant l'estimation de cette matrice de variance serait l'étude de l'extension de l'approche non-paramétrique de Francq and Zakoïan [2000] aux cas des modèles FARIMA faibles. Plus précisément, si $(X_t)_{t\in\mathbb{Z}}$ satisfait (2.1) et si $(h_n)_n$ est une suite de nombres réels et $\omega : \mathbb{R} \to \mathbb{R}$ est une fonction bornée, à support compact $[-c, c]$ et continue à l'origine avec $\omega(0) = 1$, sous quelles conditions sur $(h_n)_n$ l'estimateur

$$\hat{I}_n(\hat{\theta}_n) = \sum_{k=-[c/h_n]}^{[c/h_n]} \omega(kh_n)\hat{\Delta}_k(\hat{\theta}_n),$$

avec

$$\hat{\Delta}_k(\theta) = \frac{4}{n}\sum_{t=1}^{n-k} \tilde{\epsilon}_t(\theta)\frac{\partial\tilde{\epsilon}_t(\theta)}{\partial\theta}\left(\tilde{\epsilon}_{t+k}(\theta)\frac{\partial\tilde{\epsilon}_{t+k}(\theta)}{\partial\theta}\right)'$$

et

$$\hat{\Delta}_{-k}(\theta) = \hat{\Delta}_k(\theta)',$$

converge en probabilité vers la matrice $I(\theta_0)$?

Il serait intéressant de comparer cette méthode avec l'approche paramétrique de Berk [1974] que nous avons établi dans le chapitre 2.

Sélection des modèles FARIMA faibles par un critère d'information d'Akaike modifié

Un des problèmes les plus délicats est celui de la sélection d'un petit nombre de valeurs plausibles pour les ordres p et q du modèle FARIMA. Cette étape d'identification de la traditionnelle méthodologie de Box et Jenkins est d'importance cruciale dans la modélisation des modèles de séries temporelles.

Parmi les méthodes d'identification, les plus populaires sont celles basées sur l'optimisation d'un critère d'information. Le critère d'information d'Akaike (noté AIC pour Akaike's information criterion) est le plus utilisé dans ce contexte. L'idée est d'étudier le problème de la sélection des ordres p et q des modèles FARIMA induits par un bruit non corrélés mais qui peut contenir des dépendances non linéaires très générales. Le critère AIC est souvent utilisé, dans la littérature des séries temporelles, pour sélectionner les ordres de modèles avec un bruit iid. Boubacar Maïnassara [2012] avait proposé un critère d'information d'Akaike modifié pour sélectionner les ordres des modèles VARMA avec des termes d'erreurs non corrélés mais non nécessairement indépendants. Il sera donc intéressant d'étudier l'extension de son résultat aux modèles FARIMA faibles à mémoire longue.

Estimation des modèles FARIMA faibles à coefficients dépendants du temps

Une voie de recherche future pour le chapitre 2 est d'étendre l'étude des propriétés asymptotiques de l'estimateur des moindres carrés des modèles FARIMA faibles au cas des modèles FARIMA à coefficients dépendants du temps induits par un bruit blanc faible. En raison de la perte de la stationnarité et de l'ergodicité, il sera donc nécessaire de développer d'autres démarches, différentes de celles utilisées dans cette thèse, pour établir les propriétés asymptotiques des modèles FARIMA faibles à coefficients dépendants du temps. Azrak and Mélard [2006] ont étudié les propriétés asymptotiques de l'estimateur du quasi-maximum de vraisemblance (QMLE) pour les modèles ARMA à coefficients dépendants du temps. Ils ont considéré le cas où la variance du bruit est hétéroscédastique. Ils ont établi la convergence forte et la normalité asymptotique du QMLE sous certaines hypothèses sur les coefficients et sur le bruit permettant d'atténuer la non stationnarité du processus. L'idée est donc d'étudier, dans un premier temps, l'extension de leurs résultats aux cas des modèles FARIMA faibles à coefficients dépendants du temps et d'aborder, dans un second temps, le problème de la validation de ces modèles.

Estimation des modèles AR fractionnaires lorsque l'indice de Hurst H est supposé inconnu

Nous avons établi dans le chapitre 4 les propriétés asymptotiques de l'estimateur des moindres carrés généralisés du paramètre autorégressif des modèles autorégressifs d'ordre 1 induits par un bruit gaussien fractionnaire d'indice de Hurst H supposé connu. Ces propriétés ont été établies sous l'unique hypothèse standard de l'appartenance du paramètre autorégressif à $]-1,1[$ (pour la stationnarité et la causalité du processus) et de l'appartenance de H à $]0,1[$.

Dans la pratique, le paramètre de Hurst est généralement inconnu et nécessite d'être estimé. Il semble alors judicieux d'étendre les résultats obtenus dans chapitre 4 aux cas des modèles autorégressifs induits par un bruit gaussien fractionnaire d'indice de Hurst inconnu.

Bibliographie

Ablowitz, M. J. and Fokas, A. S. (2003). *Complex Variables : Introduction and Applications*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2 edition.

Anderson, T. W. (1971). *The statistical analysis of time series*. John Wiley & Sons, Inc., New York-London-Sydney.

Anderson, T. W. and Taylor, J. B. (1976). Strong consistency of least squares estimates in normal linear regression. *Ann. Statist.*, 4(4) :788–790.

Anderson, T. W. and Taylor, J. B. (1979). Strong consistency of least squares estimates in dynamic models. *Ann. Statist.*, 7(3) :484–489.

Azrak, R. and Mélard, G. (2006). Asymptotic properties of quasi-maximum likelihood estimators for ARMA models with time-dependent coefficients. *Statistical Inference for Stochastic Processes*, 9(3) :279–330.

Baillie, R. T., Chung, C.-F., and Tieslau, M. A. (1996). Analysing inflation by the fractionally integrated ARFIMA-GARCH model. *Journal of applied econometrics*, 11(1) :23–40.

Bauwens, L., Laurent, S., and Rombouts, J. V. K. (2006). Multivariate GARCH models : a survey. *J. Appl. Econometrics*, 21(1) :79–109.

Beran, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *J. Roy. Statist. Soc. Ser. B*, 57(4) :659–672.

Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2013). *Long-memory processes*. Springer, Heidelberg. Probabilistic properties and statistical methods.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *Ann. Statist.*, 2 :489–502. Collection of articles dedicated to Jerzy Neyman on his 80th birthday.

Billingsley, P. (1986). *Probability and measure*. Wiley series in probability and mathematical statistics. Wiley.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3) :307 – 327.

Boubacar Maïnassara, Y. (2009). *Estimation, validation et identification des modèles ARMA faibles multivariés*. PHD thesis of Université Lille 3.

Boubacar Mainassara, Y. (2011). Multivariate portmanteau test for structural VARMA models with uncorrelated but non-independent error terms. *J. Statist. Plann. Inference*, 141(8) :2961–2975.

Boubacar Maïnassara, Y. (2012). Selection of weak VARMA models by modified Akaike's information criteria. *J. Time Series Anal.*, 33(1) :121–130.

Boubacar Maïnassara, Y. (2014). Estimation of the variance of the quasi-maximum likelihood estimator of weak VARMA models. *Electron. J. Stat.*, 8(2) :2701–2740.

Boubacar Mainassara, Y., Carbon, M., and Francq, C. (2012a). Computing and estimating information matrices of weak ARMA models. *Comput. Statist. Data Anal.*, 56(2) :345–361.

Boubacar Mainassara, Y., Carbon, M., and Francq, C. (2012b). Computing and estimating information matrices of weak ARMA models. *Comput. Statist. Data Anal.*, 56(2) :345–361.

Boubacar Maïnassara, Y., Esstafa, Y., and Saussereau, B. (2019). Estimating FARIMA models with uncorrelated but non-independent error terms. *Working document*.

Boubacar Mainassara, Y. and Francq, C. (2011). Estimating structural VARMA models with uncorrelated but non-independent error terms. *J. Multivariate Anal.*, 102(3) :496–505.

Boubacar Maïnassara, Y. and Kokonendji, C. C. (2016). Modified Schwarz and Hannan-Quinn information criteria for weak VARMA models. *Stat. Inference Stoch. Process.*, 19(2) :199–217.

Boubacar Maïnassara, Y. and Saussereau, B. (2018). Diagnostic checking in multivariate arma models with dependent errors using normalized residual autocorrelations. *J. Amer. Statist. Assoc.*, 113(524) :1813–1827.

Boubacar Maïnassara, Y. (2012). Selection of weak varma models by modified akaike's information criteria. *Journal of Time Series Analysis*, 33(1) :121–130.

Box, G. E. P. and Jenkins, G. M. (1970). *Times series analysis. Forecasting and control*. Holden-Day, San Francisco, Calif.-London-Amsterdam.

Box, G. E. P. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Amer. Statist. Assoc.*, 65 :1509–1526.

Brillinger, D. R. (1981). *Time series : data analysis and theory*, volume 36. Siam.

Brockwell, P. J. and Davis, R. A. (1991). *Time series : theory and methods*. Springer Series in Statistics. Springer-Verlag, New York, second edition.

Chiu, S.-T. (1988). Weighted least squares estimators on the frequency domain for the parameters of a time series. *Ann. Statist.*, 16(3) :1315–1326.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.*, 17(4) :1749–1766.

Davidson, J. (1994). *Stochastic limit theory*. Advanced Texts in Econometrics. The Clarendon Press, Oxford University Press, New York. An introduction for econometricians.

den Haan, W. J. and Levin, A. T. (1997). A practitioner's guide to robust covariance matrix estimation. In *Robust inference*, volume 15 of *Handbook of Statist.*, pages 299–342. North-Holland, Amsterdam.

Ding, Z., Granger, C. W., and Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1 :83–106.

Doukhan, P. and León, J. (1989). Cumulants for stationary mixing random sequences and applications to empirical spectral density. *Probab. Math. Stat*, 10 :11–26.

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingrom inflation. *Econometrica*, 50 :391–407.

Fan, J. and Yao, Q. (2008). *Nonlinear time series : nonparametric and parametric methods*. Springer Science & Business Media.

Fox, R. and Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.*, 14(2) :517–532.

Francq, C., Roy, R., and Zakoïan, J.-M. (2005). Diagnostic checking in ARMA models with uncorrelated errors. *J. Amer. Statist. Assoc.*, 100(470) :532–544.

Francq, C. and Zakoïan, J.-M. (1998). Estimating linear representations of nonlinear processes. *J. Statist. Plann. Inference*, 68(1) :145–165.

Francq, C. and Zakoïan, J.-M. (2000). Covariance matrix estimation for estimators of mixing weak ARMA models. *J. Statist. Plann. Inference*, 83(2) :369–394.

Francq, C. and Zakoïan, J.-M. (2000). Covariance matrix estimation for estimators of mixing weak arma models. *Journal of Statistical Planning and Inference*, 83(2) :369 – 394.

Francq, C. and Zakoïan, J.-M. (2005). Recent results for linear time series models with non independent innovations. In *Statistical modeling and analysis for complex data problems*, volume 1 of *GERAD 25th Anniv. Ser.*, pages 241–265. Springer, New York.

Francq, C. and Zakoïan, J.-M. (2007). HAC estimation and strong linearity testing in weak ARMA models. *J. Multivariate Anal.*, 98(1) :114–144.

Francq, C. and Zakoïan, J.-M. (2010). *GARCH Models : Structure, Statistical Inference and Financial Applications*. Wiley.

Giraitis, L. and Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probab. Theory Related Fields*, 86(1) :87–104.

Gradshteyn, I. S. and Ryzhik, I. M. (2007). *Table of integrals, series, and products*. Elsevier/Academic Press, Amsterdam, seventh edition. Translated from the Russian, Translation edited and with a preface by Alan Jeffrey and Daniel Zwillinger, With one CD-ROM (Windows, Macintosh and UNIX).

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal.*, 1(1) :15–29.

Gromykov, G., Haye, M. O., and Philippe, A. (2018). A frequency-domain test for long range dependence. *Statistical Inference for Stochastic Processes*, 21(3) :513–526.

Hallin, M., Taniguchi, M., Serroukh, A., and Choy, K. (1999). Local asymptotic normality for regression models with long-memory disturbance. *Ann. Statist.*, 27(6) :2054–2080.

Hauser, M. and Kunst, R. (1998). Fractionally integrated models with arch errors : with an application to the swiss 1-month euromarket interest rate. *Review of Quantitative Finance and Accounting*, 10(1) :95–113.

Herrndorf, N. (1984). A functional central limit theorem for weakly dependent sequences of random variables. *Ann. Probab.*, 12(1) :141–153.

Hong, Y. (1996). Consistent testing for serial correlation of unknown form. *Econometrica*, 64(4) :837–864.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68(1) :165–176.

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng.*, 116 :770–799.

Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48 :419–426.

Ito, K. (1944). On the ergodicity of a certain stationary process. *Proc. Imp. Acad.*, 20(2) :54–55.

Keenan, D. M. (1987). Limiting behavior of functionals of higher-order sample cumulant spectra. *Ann. Statist.*, 15(1) :134–151.

Khinchin, A. I. (1949). Mathematical foundations of statistical mechanics dover publications. *Inc. New York*.

Kuan, C.-M. and Lee, W.-M. (2006). Robust M tests without consistent estimation of the asymptotic covariance matrix. *J. Amer. Statist. Assoc.*, 101(475) :1264–1275.

Lai, T., Robbins, H., and Wei, C. (1979). Strong consistency of least squares estimates in multiple regression ii. *Journal of Multivariate Analysis*, 9(3) :343 – 361.

Lai, T. L. and Robbins, H. (1977). Strong consistency of least-squares estimates in regression models. *Proceedings of the National Academy of Sciences*, 74(7) :2667–2669.

Lai, T. L., Robbins, H., and Wei, C. Z. (1978). Strong consistency of least squares estimates in multiple regression. *Proceedings of the National Academy of Sciences*, 75(7) :3034–3036.

Li, W. K. and McLeod, A. I. (1986). Fractional time series modelling. *Biometrika*, 73(1) :217–221.

Ling, S. and Li, W. K. (1997). On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *J. Amer. Statist. Assoc.*, 92(439) :1184–1194.

Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2) :pp. 297–303.

Lobato, I. N. (2001). Testing that a dependent process is uncorrelated. *J. Amer. Statist. Assoc.*, 96(455) :1066–1076.

Lobato, I. N., Nankervis, J. C., and Savin, N. E. (2001). Testing for autocorrelation using a modified box- pierce q test. *Inter. Econ. Review*, 42(1) :187–205.

Lobato, I. N., Nankervis, J. C., and Savin, N. E. (2002). Testing for zero autocorrelation in the presence of statistical dependence. *Econ. Theory*, 18(3) :730–743.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer-Verlag, Berlin.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer-Verlag, Berlin.

Magnus, J. R. and Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester. Revised reprint of the 1988 original.

Mandelbrot, B. B. (1965). Une classe de processus stochastiques homothétiques à soi ; application à la loi climatologique de H. E. Hurst. 260 :3274–3277.

Mandelbrot, B. B. and Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, 10 :422–437.

Mandelbrot, B. B. and Wallis, J. R. (1969a). Computer experiments with fractional gaussian noises : Part 1, averages and variances. *Water Resources Research*, 5(1) :228–241.

Mandelbrot, B. B. and Wallis, J. R. (1969b). Computer experiments with fractional gaussian noises : Part 2, rescaled ranges and spectra. *Water Resources Research*, 5(1) :242–259.

Mandelbrot, B. B. and Wallis, J. R. (1969c). Computer experiments with fractional gaussian noises : Part 3, mathematical appendix. *Water Resources Research*, 5(1) :260–267.

Miller, P. D. (2006). *Applied asymptotic analysis*, volume 75. American Mathematical Soc.

Palma, W. (2007). *Long-memory time series*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ. Theory and methods.

Priestley, M. B. (1981). *Spectral analysis and time series*. Academic Press London ; New York.

Romano, J. P. and Thombs, L. A. (1996). Inference for autocorrelations under weak assumptions. *J. Amer. Statist. Assoc.*, 91(434) :590–600.

Shao, X. (2010a). Corrigendum : A self-normalized approach to confidence interval construction in time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(5) :695–696.

Shao, X. (2010b). Nonstationarity-extended Whittle estimation. *Econometric Theory*, 26(4) :1060–1087.

Shao, X. (2010c). A self-normalized approach to confidence interval construction in time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(3) :343–366.

Shao, X. (2011a). Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models. *Econometric Theory*, 27(2) :312–343.

Shao, X. (2011b). Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models. *Econometric Theory*, 27(2) :312–343.

Shao, X. (2012). Parametric inference in stationary time series models with dependent errors. *Scand. J. Stat.*, 39(4) :772–783.

Shao, X. (2015). Self-normalization for time series : a review of recent developments. *J. Amer. Statist. Assoc.*, 110(512) :1797–1817.

Sinai, Y. (1976). Self-similar probability distributions. *Theory of Probability & Its Applications*, 21(1) :64–80.

Solo, V. (1981). Strong consistency of least squares estimators in regression with correlated disturbances. *Ann. Statist.*, 9(3) :689–693.

Taniguchi, M. (1982). On estimation of the integrals of the fourth order cumulant spectral density. *Biometrika*, 69(1) :117–122.

Taqqu, M. S. and Teverovsky, V. (1997). Robustness of Whittle-type estimators for time series with long-range dependence. *Comm. Statist. Stochastic Models*, 13(4) :723–757. Heavy tails and highly volatile phenomena.

Tong, H. (1990). *Non-linear time series : a dynamical system approach*. Oxford University Press.

Watson, G. N. (1918). The harmonic functions associated with the parabolic cylinder. *Proceedings of the London Mathematical Society*, s2-17(1) :116–148.

Whittle, P. (1953). Estimation and information in stationary time series. *Ark. Mat.*, 2 :423–434.

Wold, H. (1938). *A study in the analysis of stationary time series*. PhD thesis, Almqvist & Wiksell.

Zhu, K. and Li, W. K. (2015a). A bootstrapped spectral test for adequacy in weak ARMA models. *J. Econometrics*, 187(1) :113–130.

Zhu, K. and Li, W. K. (2015b). A bootstrapped spectral test for adequacy in weak ARMA models. *J. Econometrics*, 187(1) :113–130.