



Departure from normality of increasing-dimension martingales

Ignacio Arbués

S.G. de Estadísticas Industriales y Agrarias, Instituto Nacional de Estadística, Castellana, 183, 28071, Madrid, Spain

ARTICLE INFO

Article history:

Received 14 August 2008

Available online 27 November 2008

AMS 1991 subject classifications:

60F05

60B12

62M10

Keywords:

Central limit theorem

Banach spaces

Residual autocorrelation

Confidence regions

Approximate models

ABSTRACT

In this paper, we consider sequences of vector martingale differences of increasing dimension. We show that the Kantorovich distance from the distribution of the $k(n)$ -dimensional average of n martingale differences to the corresponding Gaussian distribution satisfies certain inequalities. As a consequence, if the growth of $k(n)$ is not too fast, then the Kantorovich distance converges to zero. Two applications of this result are presented. The first is a precise proof of the asymptotic distribution of the multivariate portmanteau statistic applied to the residuals of an autoregressive model and the second is a proof of the asymptotic normality of the estimates of a finite autoregressive model when the process is an $AR(\infty)$ and the order of the model grows with the length of the series.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Many versions of the Central Limit Theorem and convergence rate estimates have been proved under different conditions. The case considered here is that of a triangular array X_{ni} of vector martingale differences of dimension k depending on n . We are interested, in particular, in the case that $k(n) \rightarrow \infty$ for otherwise, all X_{ni} can be considered as vectors of dimension $\max\{k(n) : n \in \mathbb{N}\}$. Even in the case that k diverges, the convergence can in principle be analyzed by considering X_{ni} as an infinite sequence completed with zeros and using Banach Space techniques. Nevertheless, for some applications, to establish the convergence to an infinite random sequence is not so useful as to measure how much the $k(n)$ -variate distribution of the average differs from a $k(n)$ -variate Gaussian. Therefore, we focus on calculating bounds for the distance from the distribution of $Z_n = n^{-1/2} \sum_i X_{ni}$ to the $k(n)$ -variate normal distribution, measured with a certain metric. The metric that is most commonly used for this kind of analysis, since the classical Berry–Esséen theorem, is probably the uniform metric (for example, [1,2]). Unfortunately, this one is not the most convenient for our purposes. As we will show, bounds expressed in terms of the Prokhorov metric allow us to prove the results presented in Section 4. On the other hand, the Kantorovich metric provides an upper bound on the Prokhorov metric and behaves well under Lipschitz transformations of the variables. For these reasons, we take as the starting point for our work the result by Rachev and Rüschendorf [3] for martingales in Banach spaces, which is stated in terms of the Kantorovich metric. However, we cannot directly use their result. A generalization is required and we prove it in Section 2.

Our results are not just a theoretical exercise. In fact, they have been worked out to fill a theoretical gap in the diagnostic for time series models. Since the work of Box, Pierce and Ljung [4–6], a number of tests for residual autocorrelation have been proposed. The multivariate case was analyzed by Hosking in [7]. Ahn generalized the multivariate test for the constrained autoregressive case in [8]. More recently, it has been proved in [9] that the test can be applied to Vector Error Correction models. A variation of the test is proposed in [10]. A common feature of all these papers is the vagueness with which the

E-mail address: iarbues@ine.es.

asymptotic distribution property is stated (with the exception of the \hat{D}_m statistic of [10], that has a different form and distribution). Generally, it is claimed that the distribution of the statistic, say Q_k , where k is the greatest autocorrelation order, approximates a chi-square with $d(k)$ degrees of freedom for large k and T , where T is the number of observations. This is argued by expressing Q_k as a quadratic form evaluated at a martingale difference average plus terms that vanish as $T \rightarrow \infty$. Then, the CLT applies to that average, but the matrix of the form is only approximately idempotent for large k . Hence, any convergence analysis should consider the limit when both T and k go to infinity.

The first result of this kind that provides a precise convergence result seems to be [11], but it pays the price of taking sequential limits ($\lim_k \lim_T \dots$). This kind of asymptotic property is not the most adequate for applications because it is not realistic. In real life, T is usually given and k chosen by the analyst, so the desired result should provide convergence when k and T tend to infinity satisfying some joint condition. In which sense should this convergence be established? Clearly, it should be ensured that the error due to the use of the theoretical rejection region instead of the true one, converges to zero. If we choose $k = k(T)$ satisfying the joint convergence condition, $F_T(x)$ is the distribution function of the true statistic Q_k and $G_T(x)$ the one of a chi-square with the theoretical degrees of freedom corresponding to $k(T)$, then we want that for any $u \in (0, 1)$,

$$\lim_{T \rightarrow \infty} F_T(G_T^{-1}(u)) = u.$$

In Section 4.1 we prove that the relation above holds under some assumptions.

The second application is presented in Section 4.2 and it is related to the inference of autoregressive models when the true model is an $AR(\infty)$. If we fit an autoregressive model to a time series of length T that is generated by an $AR(\infty)$ process, then the order p of the model and T have to satisfy some joint conditions for the estimates to have good properties. These conditions were analyzed in [12] for the univariate case and in [13] for the multivariate case. In these articles, the asymptotic normality of the estimates was proved when $p^3/T \rightarrow 0$ and $T^{1/2} \sum_{j=p}^{\infty} \|\phi_j\| \rightarrow 0$, where ϕ_j is the j th autoregressive coefficient matrix of the true model. Unfortunately, the asymptotic normality is not established for the vector of estimates but for a linear combination of its components. Specifically, if $\hat{\phi}(p)$ is a vector containing the estimate coefficients and $l(p)$ is a sequence of constant vectors satisfying certain conditions, then $l(p)' \hat{\phi}(p)$ is asymptotically normal. Instead, we will establish the asymptotic normality of $\hat{\phi}(p)$ by proving that the distance from the distribution of $\hat{\phi}(p)$ to a certain pr^2 -variate Gaussian converges to zero. This allows, for example, to build confidence regions for $\phi(p)$.

2. CLT rates for martingales in Banach spaces

First we need a generalization of Theorem 3.6 in [3]. This theorem applies not only to the Gaussian case, but to the more general α -stable case. The generalization consists of relaxing an assumption that can usually be checked only when the conditional first and second order moments of the i th martingale difference with respect to the $(i-1)$ th field are almost surely constant. This condition is too strong for our applications. Hence, we will prove our result with the relaxed condition that the moments are constant with respect to the $(i-\nu)$ th field, for a certain $\nu \geq 0$. Besides this generalization, we need to state the main proposition in such a form that all the constants appearing in the inequality are absolute. This will allow us to apply in Section 3 the result to different spaces for each n . No substantial modifications of the proof in [3] are needed for this.

In order to make our results easier to relate to the corresponding ones in [3], we adhere as much as possible to their notation, both in this section and in the next one. In these two sections, we will use the first upper case roman letters with or without numeral subscripts (A, B, C, A_1, \dots) to denote absolute constants. We use also capital roman letters with subscripts that indicate dependence with respect to variables or parameters, such as C_θ or L_r . The last capital roman letters U, V, \dots, Z are reserved for random variables.

For two random variables X and Y defined on a probability space (Ω, \mathcal{F}, P) taking values in a separable Banach Space $(\mathcal{X}, \|\cdot\|)$, let us denote by $\ell_1(X, Y)$ the Kantorovich metric,

$$\ell_1(X, Y) = \sup\{|E(f(X) - f(Y))| : f \in \mathcal{L}_1^B\},$$

where \mathcal{L}_1^B is the set of the bounded real functions f defined on \mathcal{X} such that $|f(x) - f(y)| \leq \|x - y\|$.

We also need the total variation metric,

$$\sigma(X, Y) = \sup\{|E(f(X) - f(Y))| : f \in C^0(\mathcal{X}; [0, 1])\},$$

where $C^0(\mathcal{X}; [0, 1])$ denotes the set of the continuous functions defined from \mathcal{X} into $[0, 1]$.

Let θ be a symmetric α -stable random variable independent from X and Y . We define the smoothing distances,

$$\ell_r(X, Y) = \sup_{h>0} h^{r-1} \ell_1(X + h\theta, Y + h\theta), \quad r > 1,$$

$$\sigma_r(X, Y) = \sup_{h>0} h^r \sigma(X + h\theta, Y + h\theta), \quad r > 0.$$

Let X_i be a sequence of martingale differences and θ_i a sequence of independent variables distributed as θ . Let us define for $\nu \geq 0$,

$$\begin{aligned}
X_{i,v} &= \sum_{j=i-v}^i X_j, & X_{i,-v} &= \sum_{j=i}^{i+v} X_j, \\
\tilde{X}_{i,v} &= \sum_{j=i-v}^{i-1} X_j + W_i, & \tilde{X}_{i,-v} &= W_i + \sum_{j=i+1}^{i+v} X_j, \\
\hat{X}_{i,v} &= \sum_{j=i-v}^{i-1} X_j + \theta_i, & \hat{X}_{i,-v} &= \theta_i + \sum_{j=i+1}^{i+v} X_j,
\end{aligned}$$

where W_i is a random variable with the same distribution as X_i and independent from $\{X_j : j \neq i\}$. In order to establish our results, we will use the constants,

$$\begin{aligned}
\ell_r &= \sup_i \ell_r(X_i, \theta_i), \\
\tau_{r,v} &= \sup_i E \ell_r(P_{X_{i,v}|\mathcal{F}_{i-v}}, P_{\tilde{X}_{i,v}|\mathcal{F}_{i-v}}), \\
\tilde{\tau}_{r,v} &= \sup_i E \ell_r(P_{X_{i,v}|\mathcal{F}_{i-v-1}}, P_{\tilde{X}_{i,v}|\mathcal{F}_{i-v-1}}), \\
\hat{\tau}_{r,v} &= \sup_i E \ell_r(P_{X_{i,-v}|\hat{\mathcal{G}}_{i+v+1}}, P_{\hat{X}_{i,-v}|\hat{\mathcal{G}}_{i+v+1}}), \\
\sigma_r &= \sup_i \sigma_r(X_i, \theta_i), \\
t_{r,v} &= \max\{\ell_1, \sigma_1, \sigma_r^{(1/(r-2))}, \hat{\tau}_{r,v}^{(1/(r-2))}, \tilde{\tau}_{1,v}\} \quad \text{and} \\
\tilde{\ell}_{r,v} &= \max\{\ell_r, \tau_{r,v}\},
\end{aligned}$$

where $P_{X|\mathcal{F}}$ is the conditional distribution of X with respect to the σ -field \mathcal{F} and $\mathcal{F}_i = \sigma(X_j : j \leq i)$, $\hat{\mathcal{G}}_i = \sigma(X_j : j \geq i)$. Let $Z_n = n^{-1/\alpha} \sum_{i=1}^n X_i$.

Proposition 2.1. *If $E\|\theta\| < +\infty$, then there exists a constant C_θ such that*

$$\ell_1(Z_n, \theta) \leq C_\theta(n^{1-r/\alpha} \tilde{\ell}_{r,v} + n^{-1/\alpha} t_{r,v}). \quad (2.1)$$

Moreover, there exist M, N such that C_θ can be chosen satisfying $C_\theta \leq M + NE\|\theta\|$.

If $v = 0$, we obtain Theorem 3.6 from [3] as a particular case. Before going on to the proof of Proposition 2.1, we present a modified version of Lemma 3.3 in [3],

Lemma 2.2. *Let (X_i, \mathcal{F}_i) be a stochastic sequence and (\mathcal{G}_i) a decreasing sequence of sub σ -fields such that Y_j are \mathcal{G}_i measurable for $j \leq i$. Then, for $v \geq 0$,*

$$\ell_r \left(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i \right) \leq \sum_{i=1}^n E \ell_r(P_{U|\mathcal{F}_{i-v-1} \vee \mathcal{G}_{i+v+1}}, P_{V|\mathcal{F}_{i-v-1} \vee \mathcal{G}_{i+v+1}}),$$

where $U = \sum_{j=i-v}^i X_j + \sum_{j=i+1}^{i+v} Y_j$ and $V = \sum_{j=i-v}^{i-1} X_j + \sum_{j=i}^{i+v} Y_j$.

The proof of our Lemma 2.2 is essentially the same as that of lemma 3.3 in [3], whereas the proof of Theorem 3.6 in [3] has to be modified at three points. We summarize the modifications in Lemma 2.3. Let us define,

$$\gamma_1 = \ell_1 \left(n^{-1/\alpha} \sum_{i=1}^n X_i + \varepsilon \theta, n^{-1/\alpha} \left[\sum_{i=1}^{n-1} X_i + W_n \right] + \varepsilon \theta \right), \quad (2.2)$$

$$\gamma_2 = \sum_{j=1}^m \ell_1 \left(n^{-1/\alpha} \left[\sum_{l=1}^j \theta_l + \sum_{l=j+1}^n X_l \right] + \varepsilon \theta, n^{-1/\alpha} \left[\sum_{l=1}^j \theta_l + W_{j+1} + \sum_{l=j+2}^n X_l \right] + \varepsilon \theta \right) \quad \text{and} \quad (2.3)$$

$$\gamma_3 = \ell_1 \left(n^{-1/\alpha} \left[\sum_{i=1}^{m+1} \theta_i + \sum_{i=m+2}^n X_i \right] + \varepsilon \theta, n^{-1/\alpha} \sum_{i=1}^n \theta_i + \varepsilon \theta \right) \quad (2.4)$$

Lemma 2.3. *There exists a constant A such that*

$$\gamma_1 \leq n^{-1/\alpha} \tilde{\tau}_{1,v}, \quad (2.5)$$

$$\gamma_2 \leq A a^{-r+1+\alpha} n^{-1/\alpha} \hat{\tau}_{r,v}^{1/(r-\alpha)}, \quad (2.6)$$

$$\gamma_3 \leq A n^{1-r/\alpha} \tau_{r,v}, \quad (2.7)$$

where $a = \varepsilon n^{1/\alpha} / \max\{\sigma_1, \sigma_r^{1/(r-\alpha)}, \ell_r^{1/(r-\alpha)}, \hat{\tau}_{r,v}^{1/(r-\alpha)}\}$.

Proof. Using the dependence metric, defined for a metric μ as,

$$\mu(X, Y|\mathcal{F}) = \sup_{V \in \mathcal{F}} \mu(X + V, Y + V),$$

where $V \in \mathcal{F}$ means that V is \mathcal{F} -measurable, we have,

$$\begin{aligned} \gamma_1 &\leq \ell_1(n^{-1/\alpha}X_{n,v} + \varepsilon\theta, n^{-1/\alpha}\tilde{X}_{n,v} + \varepsilon\theta|\mathcal{F}_{n-v-1}) \\ &\leq \ell_1(n^{-1/\alpha}X_{n,v}, n^{-1/\alpha}\tilde{X}_{n,v}|\mathcal{F}_{n-v-1}) \\ &\leq n^{-1/\alpha}E\ell_1(P_{X_{n,v}|\mathcal{F}_{n-v-1}}, P_{\tilde{X}_{n,v}|\mathcal{F}_{n-v-1}}) \leq n^{-1/\alpha}\tilde{\tau}_{1,v}, \end{aligned}$$

where the first inequality holds by the definition of the dependence metric, the second is a consequence of the regularity of ℓ_1 and the third follows from homogeneity and the property $\ell_r(X, Y|\mathcal{F}) \leq E\ell_r(P_{X|\mathcal{F}}, P_{Y|\mathcal{F}})$ (Lemma 3.2 in [3]).

Using the α -stability of θ and the inequality $\ell_r(X, Y) \geq h^{r-1}\ell_1(X + h\theta, Y + h\theta)$,

$$\begin{aligned} \gamma_2 &\leq \sum_{j=1}^m \left(\frac{j}{n} + \varepsilon^\alpha\right)^{(1-r)/\alpha} \ell_r\left(n^{-1/\alpha} \sum_{l=j+1}^n X_l, n^{-1/\alpha} \left[\tilde{X}_{j+1} + \sum_{l=j+2}^n X_l\right]\right) \\ &\leq \sum_{j=1}^m \frac{(j + n\varepsilon^\alpha)^{(1-r)/\alpha}}{n^{(1-r)/\alpha}} n^{-r/\alpha} \ell_r(X_{j+1, -v}, \tilde{X}_{j+1, -v}|\hat{\mathcal{G}}_{j+2+v}) \\ &\leq \sum_{j=1}^m (j + na^\alpha \hat{\tau}_{r,v}^{\alpha/(r-\alpha)} n^{-1})^{(1-r)/\alpha} n^{-1/\alpha} \hat{\tau}_{r,v} \\ &\leq An^{-1/\alpha} \frac{1}{a^{r-1-\alpha}} \tilde{\tau}_{r,v}^{1/(r-\alpha)}, \end{aligned}$$

where the last inequality follows from the approximate identity (see [14], p. 379),

$$\sum_{k=m+1}^{\infty} \frac{1}{k^s} \approx \frac{1}{(s-1)m^{s-1}}.$$

Finally,

$$\begin{aligned} \gamma_3 &\leq \ell_1\left(\left(\frac{m+1}{n}\right)^{1/\alpha} \theta + n^{-1/\alpha} \sum_{j=m+1}^n X_j, \left(\frac{m+1}{n}\right)^{1/\alpha} \theta + n^{-1/\alpha} \sum_{j=m+1}^n \theta_j\right) \\ &\leq \left(\frac{m+1}{n}\right)^{(1-r)/\alpha} n^{-r/\alpha} \sum_{j=m+1}^n E\ell_r(P_{X_{i,v}|\mathcal{F}_{i-v-1} \vee \mathcal{G}_{i+v+1}}, P_{\tilde{X}_{i,v}|\mathcal{F}_{i-v-1} \vee \mathcal{G}_{i+v+1}}), \end{aligned}$$

where we have used Lemma 2.2 with $\mathcal{G}_i = \sigma(\theta_j : j \geq i)$. Since $\{\theta_i\}$ are independent among them and from $\{X_i\}$, $P_{X_{i,v}|\mathcal{F}_{i-v-1} \vee \mathcal{G}_{i+v+1}} = P_{X_{i,v}|\mathcal{F}_{i-v-1}}$ and $P_{\tilde{X}_{i,v}|\mathcal{F}_{i-v-1} \vee \mathcal{G}_{i+v+1}} = P_{\tilde{X}_{i,v}|\mathcal{F}_{i-v-1}}$. Hence,

$$\gamma_3 \leq An^{1-r/\alpha} \tau_{r,v}. \quad \square$$

The proof of Proposition 2.1 is the same as that of Theorem 3.6 in [3] except that we use the inequalities in Lemma 2.3 at the appropriate points, with $\gamma_1 = \Delta_3$, $\gamma_2 = \Delta_6$ and $\gamma_3 = \Delta_7$. The fact that the constant C_θ satisfies $C_\theta \leq M + NE\|\theta\|$ is a consequence of how the constant C is chosen in [3].

3. Increasing-dimension martingales

We will use the result of the previous section to prove that the Kantorovich distance from $n^{-1/2}$ times a sum of n vector martingale differences to a Gaussian distribution converges to zero when the growth of the dimension $k(n)$ satisfies some conditions. This means that we will focus on the Gaussian case $\alpha = 2$, $r = 3$. The fact that the constants N and M in Proposition 2.1 are absolute is critical because we need the inequality (2.1) to hold for spaces that have a different dimension for each n . This is indicated by an additional subscript n in some places.

Let $\{Y_i^j\}_{i,j \in \mathbb{N}}$ be real random variables defined in a probability space (Ω, \mathcal{F}, P) and $n \mapsto k(n) \in \mathbb{N}$ a nondecreasing function. We define $X_{ni} = (Y_i^1, \dots, Y_i^{k(n)})$ and $Z_n = n^{-1/2} \sum_{i=1}^n X_{ni}$. We denote by θ_n and θ_{ni} , random variables defined in the same probability space and distributed as a $k(n)$ -dimensional Gaussian with zero mean and unit covariance matrix. In order to measure the difference between the distributions of Z_n and θ_n , we consider them as random variables taking values in the Banach space $\mathbb{R}^{k(n)}$ endowed with the norm $\|(x_1, \dots, x_{k(n)})\|_p = (\sum_{i=1}^{k(n)} |x_i|^p)^{1/p}$.

Assumption 3.1. For any n , the sequence $\{X_{ni}\}_i \in \mathbb{N}$ is a martingale difference sequence w.r.t the sequence of σ -fields \mathcal{F}_i .

Assumption 3.2. (i) $EX_{ni}X'_{ni} = I_{k(n)}$, the identity matrix of dimension $k(n)$. (ii) There exists a function $\nu(n)$ such that

- (a) $\text{Cov}[X_{ni} | \mathcal{F}_{i-\nu(n)-1}] = I_{k(n)}$, a.s.
 (b) $\text{Cov}[X_{ni} | X_{nj} : j \geq i + \nu(n) + 1] = I_{k(n)}$, a.s.

We need the following lemma before stating the main result of this section,

Lemma 3.3. *There exist constants B_j , for $j = 1, 3$ such that if $\varphi(x)$ is the density function of a m -variate Gaussian distribution with zero mean and unit covariance matrix, then,*

$$\sup_{x \in \mathbb{R}^m} \sup_{\|z\| \leq 1} |\varphi^{(j)}(x)(z)| \leq B_j (2\pi)^{-m/2}, \quad (3.1)$$

where $\varphi^{(j)}(x)(z)$ is the j th differential of φ as a j -linear form, evaluated at (z, \dots, z) . Besides that, the following inequality holds,

$$\int_{\mathbb{R}^m} \sup_{\|z\| \leq 1} |\varphi^{(3)}(x)(z)| dx \leq m^{3/2} + 3m^{1/2}. \quad (3.2)$$

Proof. The partial derivatives of φ are,

$$\begin{aligned} \frac{\partial \varphi}{\partial x_i} &= (2\pi)^{-m/2} \exp\left(-\frac{\|x\|^2}{2}\right) x_i \\ \frac{\partial^3 \varphi}{\partial x_i \partial x_j \partial x_k} &= (2\pi)^{-m/2} \exp\left(-\frac{\|x\|^2}{2}\right) \{-x_i x_j x_k + x_i \delta_{jk} + x_j \delta_{ki} + x_k \delta_{ij}\}, \end{aligned}$$

where δ_{uv} is the Kronecker delta.

Then, if $\|z\| \leq 1$,

$$\begin{aligned} |\varphi'(x)(z)| &= (2\pi)^{-m/2} \exp\left(-\frac{\|x\|^2}{2}\right) |(x, z)| \leq (2\pi)^{-m/2} \exp\left(-\frac{\|x\|^2}{2}\right) \|x\| \\ |\varphi^{(3)}(x)(z)| &= (2\pi)^{-m/2} \exp\left(-\frac{\|x\|^2}{2}\right) |-(x, z)^3 + 3(x, z)| \\ &\leq (2\pi)^{-m/2} \exp\left(-\frac{\|x\|^2}{2}\right) \{\|x\|^3 + 3\|x\|\}. \end{aligned} \quad (3.3)$$

Consequently,

$$\begin{aligned} \sup_{x \in \mathbb{R}^m} \sup_{\|z\| \leq 1} |\varphi'(x)(z)| &\leq (2\pi)^{-m/2} S_1 \\ \sup_{x \in \mathbb{R}^m} \sup_{\|z\| \leq 1} |\varphi^{(3)}(x)(z)| &\leq (2\pi)^{-m/2} (S_3 + 3S_1), \end{aligned}$$

where $S_j = \sup_{r>0} r^j \exp\{-r^2/2\}$.

On the other hand, (3.3) implies,

$$\int_{\mathbb{R}^m} \sup_{\|z\| \leq 1} |\varphi^{(3)}(x)(z)| dx \leq E\|\theta\|^3 + 3E\|\theta\|.$$

We get (3.2) using that by Jensen's inequality, $E\|\theta\|^3 \leq \{E(\sum_i \theta_i^2)^3\}^{1/2}$ and $E\|\theta\| \leq \{E \sum_i \theta_i^2\}^{1/2}$, and this completes the proof of the lemma. \square

Let us write $\mu_q = \sup_{i,j} E|Y_i^j|^q$.

Proposition 3.4. *If Assumptions 3.1 and 3.2 hold and $\mu_{3p} < +\infty$, then there exist constants $D_{1,Y}$ and $D_{2,Y}$ such that*

$$\ell_1(Z_n, \theta_n) \leq \frac{D_{1,Y} k(n)^{3/2+4/p} + D_{2,Y} k(n)^{3/2+2/p} \nu(n)}{n^{1/2}}. \quad (3.4)$$

Proof. Using that the constants M and N in Proposition 2.1 are absolute, we can use inequality (2.1) for the case that the Banach space \mathcal{X} varies with n . Therefore, we assume that $\mathcal{X}_n = (\mathbb{R}^{k(n)}, \|\cdot\|_p)$.

We will need bounds for some moments of X_{ni} ,

$$E\|X_{ni}\|_p \leq (k(n)\mu_p)^{1/p} \quad E\|X_{ni}\|_p^3 \leq (k(n)^3 \mu_{3p})^{1/p}.$$

Provided the conditions of [Proposition 2.1](#) hold with $\alpha = 2$ and $r = 3$,

$$\ell_1(Z_n, \theta_n) \leq (M + NE\|\theta_n\|_p) \frac{\tilde{\ell}_{3,v} + t_{3,v}}{n^{1/2}}.$$

In order to estimate these constants, we will use that for any integer r , the metrics ℓ_r and σ_r satisfy the inequalities,

$$\ell_r(X, Y) \leq G_{r,\theta} \zeta_r(X, Y) \quad (3.5)$$

$$\sigma_r(X, Y) \leq L_{r,\theta} \zeta_r(X, Y), \quad (3.6)$$

where ζ_r is the Zolotarev metric (defined in [15]), the constant $G_{r,\theta}$ is defined as $\int \sup_{\|z\|_p \leq 1} |\varphi^{(r)}(x)(z)| dx$, $L_{r,\theta}$ is a Lipschitz constant of $\varphi^{(r-1)}$ and φ is the density function of θ_n . For the inequality on σ_r , see [Proposition 4.4](#) in [16]; for ℓ_r , see [17].

From (3.2), the constant $G_{3,\theta}$ is bounded by $k(n)^{3/2} + 3k(n)^{1/2}$ and from (3.1), the constants $L_{1,\theta}$ and $L_{3,\theta}$ are bounded by $B_1(2\pi)^{-k(n)/2}$ and $B_3(2\pi)^{-k(n)/2}$, respectively.

In turn, $\zeta_r(X, Y)$ is bounded by $2s + r$ times the pseudomoment metric $\kappa_r(X, Y)$ when $s = \max\{j \in \mathbb{N} : j < r\}$ and the j th order moments of X and Y are equal for any integer $j \leq s$ (see [15] or [16]). The pseudomoment metric is defined as,

$$\kappa_r(X, Y) = \sup\{|E(f(X) - f(Y))| : f \in \mathcal{M}_r\},$$

where $\mathcal{M}_r = \{f \in \mathbb{R}^X : |f(x) - f(y)| \leq \|x\|^{r-1}x - \|y\|^{r-1}y\}$.

We will use relations (3.5) and (3.6) together with the obvious $\kappa_r(X, Y) \leq E\|X\|^r + E\|Y\|^r$ to find bounds for all constants involved in [Proposition 2.1](#).

$$(a) \ell_1 \quad (b) \ell_3 \quad (c) \sigma_1 \quad (d) \sigma_3 \quad (e) \hat{\tau}_{3,v} \quad (f) \tilde{\tau}_{1,v} \quad (g) \tau_{3,v}$$

(a) Since $\ell_1(X, Y) \leq E\|X - Y\|$,

$$\ell_1(X_{ni}, \theta_{ni}) \leq E\|X_{ni} - \theta_{ni}\|_p \leq E\|X_{ni}\|_p + E\|\theta_{ni}\|_p \leq (k(n)\mu_p)^{1/p} + (k(n)\mu_p(\theta))^{1/p},$$

where $\mu_p(\theta) = E\|\theta\|_p$.

(b) We can use the inequalities involving ℓ_r , ζ_r and κ_r to get,

$$\begin{aligned} \ell_3(X_{ni}, \theta_{ni}) &\leq G_{3,\theta} \zeta_3(X_{ni}, \theta_{ni}) \leq G_{3,\theta} 7 \kappa_3(X_{ni}, \theta_{ni}) \leq G_{3,\theta} 7 (E\|X_{ni}\|_p^3 + E\|\theta_{ni}\|_p^3) \\ &\leq G_{3,\theta} 7 \left((k(n)^3 \mu_{3p})^{1/p} + (k(n)^3 \mu_{3p}(\theta))^{1/p} \right) \leq 7 (k(n)^{3/2} + 3k(n)^{1/2}) k(n)^{3/p} (\mu_{3p}^{1/p} + \mu_{3p}(\theta)^{1/p}). \end{aligned}$$

The identity of the first and second-order moments of X_{ni} and θ_{ni} is a consequence of [Assumptions 3.1](#) and [3.2\(i\)](#).

(c) The bound for σ_1 is obtained in a similar way,

$$\begin{aligned} \sigma_1(X_{ni}, \theta_{ni}) &\leq L_{1,\theta} \zeta_1(X_{ni}, \theta_{ni}) \leq B_1(2\pi)^{-k(n)/2} \zeta_1(X_{ni}, \theta_{ni}) \\ &\leq B_1(2\pi)^{-k(n)/2} \kappa_1(X_{ni}, \theta_{ni}) \\ &\leq B_1(2\pi)^{-k(n)/2} \{ (k(n)\mu_p)^{1/p} + (k(n)\mu_p(\theta))^{1/p} \}. \end{aligned}$$

(d) For σ_3 , as for ℓ_3 ,

$$\sigma_3(X_{ni}, \theta_{ni}) \leq L_{3,\theta} 7 (k(n)^3 \mu_{3p})^{1/p} \leq 7 B_3(2\pi)^{-k(n)/2} (k(n)^3 \mu_{3p})^{1/p}.$$

(e) For $\hat{\tau}_{3,v}$, we can use again the bound on ℓ_3 in terms of ζ_3 and, given [Assumption 3.2\(ii\)](#), also the bound on ζ_3 in terms of κ_3 . Thus,

$$\ell_3(P_{X_{i,-v}|\hat{\theta}_{i+v+1}}, P_{\hat{X}_{i,-v}|\hat{\theta}_{i+v+1}}) \leq G_{3,\theta} 7 \left(2E \left[\left\| \sum_{j=i+1}^{i+v} X_j \right\|_p^3 \right] \mathcal{G}_{i+v(n)+1} + E[\|X_i\|_p^3] \mathcal{G}_{i+v(n)+1} + E[\|\theta_{ni}\|_p^3] \mathcal{G}_{i+v(n)+1} \right).$$

Taking expectation in both sides we have,

$$E\ell_3(P_{X_{i,-v}|\hat{\theta}_{i+v+1}}, P_{\hat{X}_{i,-v}|\hat{\theta}_{i+v+1}}) \leq 14 (k(n)^{3/2} + 3k(n)^{1/2}) v(n) k(n)^{1/p} (\mu_{3p}^{1/p} + \mu_{3p}(\theta)^{1/p}).$$

(f) Let us now estimate $\tilde{\tau}_{1,v}$.

$$\begin{aligned} \tilde{\tau}_{1,v} &= \sup_i E\ell_1(P_{X_{i,v}|\mathcal{F}_{i-v}}, P_{\tilde{X}_{i,v}|\mathcal{F}_{i-v}}) \leq E\|X_{i,v}\|_p + E\|\tilde{X}_{i,v}\|_p \\ &\leq \sum_{j=i-v}^{i-1} E\|X_j\|_p + E\|X_i\|_p + E\|W_i\|_p \leq 2v(n) (k(n)\mu_p)^{1/p}. \end{aligned}$$

(g) $\tau_{3,v}$ is bounded as $\hat{\tau}_{3,v}$.

Gathering all these bounds, we get,

$$\ell_1(Z_n, \theta_n) \leq (M + N(k(n)\mu_p)^{1/p}) \frac{H_{1,\gamma} k(n)^{3/2+3/p} + H_{2,\gamma} v(n) k(n)^{3/2+1/p}}{n^{1/2}},$$

so (3.4) is established. \square

4. Applications

In this section, we present two applications of Proposition 3.4. The first is a proof of the asymptotic distribution of the multivariate portmanteau statistic Q_k [7] when it is applied to the residuals of an autoregressive models and in the second we state sufficient conditions for obtaining asymptotic confidence regions for the coefficients of an autoregressive model when the process is an $AR(\infty)$.

4.1. Residual autocorrelation tests

The residual autocorrelation tests have been analyzed in many cases. To avoid inessential complications, we will assume that the r -variate process x_t satisfies, rather than a general ARMA, an autoregressive model,

$$x_t = \sum_{j=1}^p \phi_j x_{t-j} + \varepsilon_t,$$

where $\varepsilon_t = (\varepsilon_{t1}, \dots, \varepsilon_{tr})'$, $E\varepsilon_t = 0$ and $E\varepsilon_t \varepsilon_t' = \Sigma$. We denote by $\hat{\varepsilon}_t$ the residuals of a model with coefficients $\hat{\phi}_1, \dots, \hat{\phi}_p$ estimated by Gaussian maximum likelihood using a series of length T . The residual autocovariances are $\hat{C}_j = T^{-1} \sum_t \hat{\varepsilon}_t \hat{\varepsilon}_{t-j}'$ (we also use the notation $\hat{\Sigma}$ for \hat{C}_0) and the statistic of the test is defined as,

$$Q_k = T \sum_{j=1}^k \text{tr}(\hat{C}_0^{-1} \hat{C}_j \hat{C}_0^{-1} \hat{C}_j').$$

Assumption 4.1. (i) The polynomial $|I_r - \phi_1 z - \dots - \phi_p z^p|$ has its roots $\{z_i\}$ outside the unit circle. (ii) ε_t is i.i.d. and $E|\varepsilon_{ti}|^{6\beta} < +\infty$, $\beta \geq 2$.

Let $\rho_* = (\min_i |z_i|)^{-1}$.

Proposition 4.2. If Assumption 4.1 holds and $T\rho^k, k^\alpha T^{-1} \rightarrow 0$ then, for any $\rho > \rho_*$ and $u \in (0, 1)$,

$$F_T(G_T^{-1}(u)) = u + O\left(\left[T\rho^k + \frac{k^{\alpha/2}}{T^{1/2}}\right]^{1/2}\right), \quad (4.1)$$

where $\alpha = (\beta - 2)/\beta + \max\{5 + 4/\beta, 3 + 8/\beta\}$, F_T is the distribution function of Q_k and G_T is the distribution function of a chi-square with $r^2(k - p)$ degrees of freedom.

Remark 4.3. The fastest convergence rate is achieved in (4.1) when $k = O(\log T)$. Hence, regarding the convergence under the null, the growth of k is not effectively limited by the condition $k^\alpha T^{-1} \rightarrow 0$. However, a faster growth of k could be convenient for increasing the power of the test under an alternative hypothesis.

Before going on to the proof of Proposition 4.2, we need an auxiliary lemma.

Lemma 4.4. Let $\{F_T\}_{T \in \mathbb{N}}$ and $\{G_T\}_{T \in \mathbb{N}}$ be two classes of distribution functions in \mathbb{R} and assume that the elements of $\{G_T\}_{T \in \mathbb{N}}$ satisfy a Lipschitz condition with the common constant L . Then for any $u \in (0, 1)$,

$$|F_T(G_T^{-1}(u)) - u| \leq (1 + L)\pi(F_T, G_T),$$

where $\pi(\cdot, \cdot)$ denotes the Prokhorov metric.

Proof of Lemma 4.4. Let δ be such that $\pi(F_T, G_T) < \delta$. Then,

$$\mathbb{P}_T(A) \leq \mathbb{Q}_T(A^\delta) + \delta, \quad \mathbb{Q}_T(A) \leq \mathbb{P}_T(A^\delta) + \delta, \quad (4.2)$$

where \mathbb{P}_T and \mathbb{Q}_T are the probability measures of F_T and G_T respectively, A is any Borel set of \mathbb{R} and $A^\gamma := \{x : d(x, A) \leq \gamma\}$. If we put for $u \in (0, 1)$, $A = (-\infty, G_T^{-1}(u)]$ in the first inequality in (4.2) and $A = (-\infty, G_T^{-1}(u) - \delta]$ in the second, we get,

$$F_T(G_T^{-1}(u)) \leq u + L\delta + \delta \quad u - L\delta \leq F_T(G_T^{-1}(u)) + \delta.$$

Then, $|F_T(G_T^{-1}(u)) - u| \leq \delta + L\delta$, so we conclude. \square

Proof of Proposition 4.2. In what follows, we denote by $O_m(\cdot)$ order in mean, that is, for any random variable $X_{T,k}$, the identity $X_{T,k} = O_m(\omega(T, k))$ means that $\omega(T, k)^{-1}E\|X_{T,k}\|$ is bounded independently from k and T . In this section, $\|\cdot\|$ means the euclidean norm, while the β -norm is denoted by $\|\cdot\|_\beta$. When the ℓ_1 metric refers to a β -norm with $\beta \neq 2$, we write ℓ_1^β .

Assumption 4.1 implies that the process x_t has a Wold representation,

$$x_t = \sum_{l=0}^{\infty} \psi_l \varepsilon_{t-l}$$

such that the coefficients decay exponentially. Moreover, for any $\rho > \rho_*$, there exists a constant M_ρ such that $\|\psi_l\| \leq M_\rho \rho^l$.

Another consequence of Assumption 4.1 is that the maximum likelihood estimates are consistent (see, for example, [18]) and satisfy a CLT. Thus, if we stack all the true coefficients in $\phi = (\text{vec}(\Phi_1)', \dots, \text{vec}(\Phi_p)')'$ and the estimates in $\hat{\phi} = (\text{vec}(\hat{\Phi}_1)', \dots, \text{vec}(\hat{\Phi}_p)')'$, then $\hat{\phi} - \phi = O_m(T^{-1/2})$.

For $c_j = \text{vec}(T^{-1} \sum_t \varepsilon_t \varepsilon_{t-j}')$ and $\hat{c}_j = \text{vec}(\hat{C}_j)$, by a Taylor expansion, we get,

$$\hat{c}_j = c_j + \frac{\partial c_j}{\partial \phi'} (\hat{\phi} - \phi) + \frac{1}{2} D^2 c_j (\hat{\phi} - \phi) \otimes (\hat{\phi} - \phi), \quad (4.3)$$

where $D^2 c_j$ is a matrix containing the second derivatives of the elements of \hat{c}_j with respect to ϕ arranged in the appropriate way. It can be proved that,

$$\frac{\partial c_j}{\partial \phi'} = -\hat{W}_j$$

where,

$$\begin{aligned} \hat{W}_j &= (\hat{W}_{j1}, \dots, \hat{W}_{jp}) \\ \hat{W}_{ji} &= T^{-1} \sum_t \{(\varepsilon_{t-j} \otimes I_r)(x'_{t-i} \otimes I_r) + (I_r \otimes \varepsilon_t)(x'_{t-j-i} \otimes I_r)\}. \end{aligned}$$

The quadratic part of the Taylor expansion (4.3) satisfies,

$$\sum_{j=1}^k \|D^2 c_j (\hat{\phi} - \phi) \otimes (\hat{\phi} - \phi)\|^2 = O\left(\frac{k}{T^2}\right).$$

On the other hand, if we put $c = (c_1, \dots, c_k)'$, $\hat{W} = [\hat{W}'_1, \dots, \hat{W}'_k]'$ and $W = [W'_1, \dots, W'_k]'$, with $W_j = [\Sigma \Psi'_{j-1} \otimes I_r, \dots, \Sigma \Psi'_{j-p} \otimes I_r]$, then it can be proved that $\|W\|$ is bounded uniformly in k (because of the exponential decay of ψ_l) and $\hat{W} = W + O([k/T]^{1/2})$. Consequently, \hat{c} can be expressed as,

$$\hat{c} = c - W(\hat{\phi} - \phi) + O_m(k^{1/2} T^{-1}).$$

The statistic Q_k can be written as,

$$Q_k = T \hat{c}' (I_k \otimes \hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1}) \hat{c}.$$

If we put $\Omega = (I_k \otimes \Sigma^{-1} \otimes \Sigma^{-1})$ and $\hat{\Omega} = (I_k \otimes \hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1})$, using that $\hat{\Sigma} = \Sigma + O_m(T^{-1/2})$ we get,

$$Q_k = T(c - W(\hat{\phi} - \phi))' \Omega^{-1} (c - W(\hat{\phi} - \phi)) + O_m(k^{1/2} T^{-1/2}). \quad (4.4)$$

A Taylor expansion of the log-likelihood l yields,

$$\hat{\phi} - \phi = \frac{1}{T} I(\phi)^{-1} \frac{\partial l}{\partial \phi} + O_m(T^{-1}), \quad (4.5)$$

where $I(\phi) = (\gamma'_{u-v} \otimes \Sigma)_{u,v}$ is the information matrix, with $\gamma_l = E x_t x'_{t+l}$. The derivative of the log-likelihood is given by,

$$\begin{aligned} \frac{\partial l}{\partial \phi_i} &= - \sum_t (x_{t-i} \otimes I_r) \Sigma^{-1} \varepsilon_t = - \sum_t \sum_{u=0}^{\infty} (\Psi_u \varepsilon_{t-i-u} \otimes I_r) \Sigma^{-1} \varepsilon_t \\ &= - \sum_{u=0}^{k-i} (\Psi_u \Sigma \otimes I_r) (\Sigma^{-1} \otimes \Sigma^{-1}) \sum_t \text{vec}(\varepsilon_t \varepsilon'_{t-i-u}) + O(\rho^{k+1-i}) O_m(T^{1/2}). \end{aligned}$$

Hence,

$$\frac{\partial l}{\partial \phi} = -TW' \Omega^{-1} c + O(\rho^k) O_m(T^{1/2}). \quad (4.6)$$

Now put $E_T := T^{-1/2} \sum_{t=1}^T e_t$, where e_t is the increasing-dimension martingale difference $e_t = \Omega^{-1/2}(\text{vec}(\varepsilon_t \varepsilon'_{t-1}), \dots, \text{vec}(\varepsilon_t \varepsilon'_{t-k}))'$. If we replace $\partial l / \partial \phi$ by the right hand side of (4.6) in (4.5) we get,

$$\hat{\phi} - \phi = -T^{-1/2} I(\phi)^{-1} W \Omega^{-1/2} E_T + O_m(T^{-1}) + O(\rho^k) O_m(T^{1/2}). \quad (4.7)$$

Now, using (4.4) and (4.7) we obtain,

$$Q_k = E'_T (I_{kr^2} - \Omega^{-1/2} W I(\phi)^{-1} W' \Omega^{-1/2})' (I_{kr^2} - \Omega^{-1/2} W I(\phi)^{-1} W' \Omega^{-1/2}) E_T + O_m(T^{-1/2} + k^{1/2} T^{-1} + \rho^k T). \quad (4.8)$$

We can write more succinctly,

$$Q_k = \xi_T^2 + O_m(T^{-1/2} + k^{1/2} T^{-1} + \rho^k T), \quad (4.9)$$

where $\xi_T^2 = E'_T A'_T A_T E_T$ and $A_T = I_{kr^2} - \Omega^{-1/2} W I(\phi)^{-1} W' \Omega^{-1/2}$. The matrix,

$$U_T = I_{kr^2} - \Omega^{-1/2} W (W' \Omega^{-1} W)^{-1} W' \Omega^{-1/2}$$

is idempotent and,

$$\begin{aligned} \|A_T - U_T\| &\leq \|\Omega^{-1}\| \cdot \|W\|^2 \cdot \|I(\phi)^{-1} - (W' \Omega^{-1} W)^{-1}\| \\ &\leq \|\Omega^{-1}\| \cdot \|W\|^2 \cdot \|I(\phi)^{-1}\| \cdot \|(W' \Omega^{-1} W)^{-1}\| \cdot \|I(\phi) - W' \Omega^{-1} W\| = O(\rho^k), \end{aligned} \quad (4.10)$$

where the last identity is a consequence of the fact that $W' \Omega^{-1} W = (\hat{\gamma}'_{u-v} \otimes \Sigma)_{u,v}$ where $\hat{\gamma}_{u-v} = \gamma_{u-v} + O(\rho^k)$.

Since ε_t are i.i.d., Assumption 3.2(ii) holds for E_T with $n = T$ and $v = k$ and thus, by Proposition 3.4, $\ell_1(E_T, \theta_T) = O([k^\varpi / T]^{1/2})$, with $\varpi = \max\{5 + 4/\beta, 3 + 8/\beta\}$. Here we consider E_T, θ_T as elements of the Banach space \mathbb{R}^k with the norm $\|\cdot\|_\beta$. From now on, k is a function of T , but in order to simplify the notation, we omit this dependency. By the triangular inequality,

$$\ell_1(A_T E_T, U_T \theta_T) \leq \ell_1(A_T E_T, U_T E_T) + \ell_1(U_T E_T, U_T \theta_T).$$

We can estimate both terms as,

$$\begin{aligned} \ell_1(A_T E_T, U_T E_T) &\leq \|A_T - U_T\| \cdot E\|E_T\| = O(\rho^k) \\ \ell_1(U_T E_T, U_T \theta_T) &\leq \|U_T\| \ell_1(E_T, \theta_T) \\ &\leq \|U_T\| k^{(\beta-2)/(2\beta)} \ell_1^\beta(E_T, \theta_T) = k^{(\beta-2)/(2\beta)} O\left(\left[\frac{k^\varpi}{T}\right]^{1/2}\right), \end{aligned}$$

where we have used that $\|z\|_2 \leq k^{(\beta-2)/(2\beta)} \|z\|_\beta$. Then, for $\eta_T = (\theta'_T U_T \theta_T)^{1/2}$,

$$\ell_1(\xi_T, \eta_T) = O\left(\rho^k + \left[\frac{k^\alpha}{T}\right]^{1/2}\right)$$

with $\alpha = (\beta - 2)/\beta + \varpi$. The variable η_T is the square root of a chi-square with $k - p$ degrees of freedom, that is, a chi distribution. The inequality $\pi(X, Y)^2 \leq \ell_1(X, Y)$ (see [3]) implies

$$\pi(\xi_T, \eta_T) = O\left(\left[\rho^k + \frac{k^{\alpha/2}}{T^{1/2}}\right]^{1/2}\right). \quad (4.11)$$

Now put $\omega_1 = [\rho^k + T^{-1/2} k^{\alpha/2}]^{1/2}$ and $\omega_2 = [T^{-1/2} + k^{1/2} T^{-1} + \rho^k T]^{1/2}$. From (4.9), we have $\ell_1(Q_k, \xi_T^2) = O(\omega_2^2)$ and then, again from $\pi(X, Y)^2 \leq \ell_1(X, Y)$, we get,

$$\pi(Q_k, \xi_T^2) = O(\omega_2). \quad (4.12)$$

On the other hand, since $\ell_1(\xi_T, \eta_T) = O(\omega_1^2)$, for f defined as

$$f(x) = \begin{cases} 1 & x \in (-\infty, 1] \\ 2 - x & x \in (1, 2] \\ 0 & x \in (2, +\infty), \end{cases}$$

we have that $|Ef(\xi_T) - Ef(\eta_T)| = O(\omega_1^2)$. The properties of the distribution of η_T imply that $Ef(\eta_T) = O(k^{-k/\tau})$, and then $Ef(\xi_T) = O(\omega_1^2 + k^{-k/\tau})$, but $k^{-k/\tau}$ converges so fast as to be neglected. Since $\xi_T \geq 0$, $P[\xi_T \leq 1] \leq Ef(\xi_T)$. From, (4.12), we know that there exists a sequence $\epsilon_{T,k}$ such that $\lim_{k,T} \epsilon_{T,k} \omega_2^{-1/2} = 0$ and for any Borel set $A \subset \mathbb{R}_+$,

$$P[Q_k \in A] \leq P[\xi_T^2 \in A^{\epsilon_{T,k}}] + \epsilon_{T,k},$$

which implies,

$$\begin{aligned} P[Q_k^{1/2} \in A] &= P[Q_k \in A^2] \leq P[\xi_T^2 \in (A^2)^{\epsilon_{T,k}}] + \epsilon_{T,k} \\ &\leq P[\xi_T \in A^{\epsilon_{T,k}}] + \epsilon_{T,k} + P[\xi_T \leq 1] \\ &\leq P[\xi_T \in A^{\epsilon_{T,k} + P[\xi_T \leq 1]}] + \epsilon_{T,k} + P[\xi_T \leq 1], \end{aligned} \quad (4.13)$$

where the second inequality follows from the relation $|\xi_T - \zeta| \leq |\xi_T^2 - \zeta^2|$, that holds when $\xi_T \geq 1$ and $\zeta > 0$. Consequently,

$$\pi(Q_k^{1/2}, \xi_T) = O(\omega_2 + \omega_1^2). \quad (4.14)$$

From (4.11) and (4.14),

$$\pi(Q_k^{1/2}, \eta_T) \leq \pi(Q_k^{1/2}, \xi_T) + \pi(\xi_T, \eta_T) = O(\omega_1 + \omega_2).$$

A chi distribution with d degrees of freedom has as density function,

$$g_d(x) = \frac{2^{1-d/2}}{\Gamma(d/2)} x^{d-1} e^{-x^2/2}.$$

It is easy to see that the class $\{g_d : d > 1\}$ is uniformly bounded by, say, M . This implies that the distribution functions satisfy a common Lipschitz condition with constant M . If F_T^0 is the distribution function of $Q_k^{1/2}$ and G_T^0 is the distribution function of a chi with $kr^2 - p$ degrees of freedom, that is, the distribution function of η_T , then by Lemma 4.4,

$$|F_T^0(G_T^{0-1}(u)) - u| = O(\omega_1 + \omega_2).$$

Since $F_T^0(G_T^{0-1}(u)) = F_T(G_T^{-1}(u))$, (4.1) is established. \square

4.2. Confidence regions for approximate autoregressive models

We now consider stationary processes x_t such that,

Assumption 4.5. x_t satisfies

$$x_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad (4.15)$$

where ε_t is a sequence of i.i.d. random variables with zero mean and covariance matrix Σ . We also assume that $\lim_n n \sum_{j=n+1}^{\infty} \|\psi_j\|^2 < +\infty$ and $|\sum_{j=0}^{\infty} \psi_j z^j|$ has its roots outside the unit circle $\{z : |z| \leq 1\}$.

The identity (4.15) can be inverted as,

$$x_t = \sum_{j=1}^{\infty} \Phi_j x_{t-j} + \varepsilon_t$$

where $\sum_{j=0}^{\infty} \|\Phi_j\| < +\infty$ and $|\sum_{j=0}^{\infty} \Phi_j z^j|$ also has no roots on or inside the unit circle. For a certain p , we compute the estimates $\hat{\Phi}_1, \dots, \hat{\Phi}_p$ using the Yule–Walker equations. Let us stack them into $\hat{\phi}(p) = (\text{vec}(\hat{\Phi}_1)', \dots, \text{vec}(\hat{\Phi}_p)')'$ and the true values into $\phi(p) = (\text{vec}(\Phi_1)', \dots, \text{vec}(\Phi_p)')'$. Let us also define the matrix $\Gamma_p = (\gamma_{u-v})_{u,v=0}^{p-1}$, with $\gamma_j = E x_t x_{t+j}'$.

Proposition 4.6. If Assumption 4.5 holds, ε_t satisfies Assumption 4.1(ii) and p is such that $T^{1/2} \sum_{j=p}^{\infty} \|\Phi_j\| \rightarrow 0$ and $p^\alpha/T \rightarrow 0$ where α is defined as in Proposition 4.2, then,

$$\ell_1((T-p)^{1/2}(\Gamma_p^{1/2} \otimes \Sigma^{-1/2})(\hat{\phi}(p) - \phi(p)), Z_p) \rightarrow 0, \quad (4.16)$$

where Z_p is a pr^2 -dimensional normally distributed r. v. with zero mean and unit covariance matrix.

Proof. In this proof we use the notation $\xi_T = o_m(u_T)$ when $\lim_T u_T^{-1} E \|\xi_T\| = 0$. It is easy to modify the proof of Theorem 2 in [13] to establish,

$$(T-p)^{1/2}(\hat{\phi}(p) - \phi(p)) = \frac{1}{(T-p)^{1/2}} \text{vec} \left[\sum_{t=p}^{T-1} \varepsilon_{t+1} X_{t,p}' \tilde{\Gamma}_p^{-1} \right] + o_m(1),$$

where $X_{t,p} = (x'_t, x'_{t-1}, \dots, x'_{t-p+1})'$. Let us denote by s_T the first term in the right hand side of (4.2). We can use the identity,

$$X_{t,p} = \sum_{u=0}^{\infty} (I_p \otimes \Psi_u) \epsilon_{t-u,p},$$

with $\epsilon_{t,p} = (\epsilon'_t, \epsilon'_{t-1}, \dots, \epsilon'_{t-p+1})'$ to write,

$$\begin{aligned} s_T &= \frac{1}{(T-p)^{1/2}} \text{vec} \left[\sum_{t=p}^{T-1} \sum_{u=0}^{\infty} \epsilon_{t+1} \epsilon'_{t-u,p} (I_p \otimes \Psi'_u) \Gamma_p^{-1} \right] \\ &= \frac{1}{(T-p)^{1/2}} \text{vec} \left[\sum_{t=p}^{T-1} \sum_{u=0}^j \epsilon_{t+1} \epsilon'_{t-u,p} (I_p \otimes \Psi'_u) \Gamma_p^{-1} \right] + \frac{1}{(T-p)^{1/2}} \text{vec} \left[\sum_{t=p}^{T-1} \sum_{u=j+1}^{\infty} \epsilon_{t+1} \epsilon'_{t-u,p} (I_p \otimes \Psi'_u) \Gamma_p^{-1} \right] \\ &= \frac{1}{(T-p)^{1/2}} (\Gamma_p^{-1} \otimes I_r) \sum_{t=p}^{T-1} G e_t + \eta_T, \end{aligned} \quad (4.17)$$

where $e_t = (\text{vec}(\epsilon_{t+1} \epsilon'_t), \text{vec}(\epsilon_{t+1} \epsilon'_{t-1}), \dots, \text{vec}(\epsilon_{t+1} \epsilon'_{t-p+1}))'$ and G is defined as the block matrix $(G_{uv})_{u=0, v=0}^{p-1, p+j-1}$ with $G_{uv} = \Psi_{v-u} \otimes I_r$ if $0 \leq v-u \leq j$ and $G_{uv} = 0$ otherwise. Let us see that $\eta_T = o_m(1)$,

$$\begin{aligned} E \left\| \text{vec} \left[\sum_{t=p}^{T-1} \sum_{u=j+1}^{\infty} \epsilon_{t+1} \epsilon'_{t-u,p} (I_p \otimes \Psi'_u) \right] \right\|^2 &\leq E \sum_{l=0}^p \sum_{t=p}^{T-1} \sum_{u=j+1}^{\infty} \left\| \text{vec} [\epsilon_{t+1} \epsilon'_{t-l-u} \Psi'_u] \right\|^2 \\ &\leq \sum_{l=0}^p \sum_{t, s=p}^{T-1} \sum_{u, v=j+1}^{\infty} \text{vec}(\Psi'_u)' E [(I_r \otimes \epsilon_{t+1} \epsilon'_{t-l-u})' (I_r \otimes \epsilon_{s+1} \epsilon'_{s-l-v})] \text{vec}(\Psi'_v) \\ &= O \left(T p \sum_{u=j+1}^{\infty} \|\Psi_u\|^2 \right). \end{aligned}$$

If we set $j \propto p$, given that the norm of Γ_p^{-1} is uniformly bounded ([12], p. 491), we find that $E \|\eta_T\| \rightarrow 0$ and thus,

$$(T-p)^{1/2} (\hat{\phi}(p) - \phi(p)) = \frac{1}{(T-p)^{1/2}} (\Gamma_p^{-1} \otimes I_r) \sum_{t=p}^{T-1} G e_t + o_m(1). \quad (4.18)$$

Consequently, we can write,

$$(T-p)^{1/2} (\Gamma_p^{1/2} \otimes \Sigma^{-1/2}) (\hat{\phi}(p) - \phi(p)) = \frac{1}{(T-p)^{1/2}} (\Gamma_p^{-1/2} \otimes \Sigma^{-1/2}) \sum_{t=p}^{T-1} G e_t + o_m(1). \quad (4.19)$$

The process $(G \Omega G')^{-1/2} G e_t$, where Ω is defined as in the proof of Proposition 4.2 has unit covariance matrix and then Proposition 3.4 can be applied to it. Therefore, we have to prove that $\|(\Gamma_p \otimes \Sigma)^{-1/2} - (G \Omega G')^{-1/2}\| \rightarrow 0$ and then, since $\|G\|$ and $E \|(T-p)^{-1/2} \sum_t e_t\|$ are bounded, we can replace $(\Gamma_p^{-1/2} \otimes \Sigma^{-1/2}) G e_t$ by $u_t = (G \Omega G')^{-1/2} G e_t$ in (4.19) obtaining,

$$(T-p)^{1/2} (\Gamma_p^{1/2} \otimes \Sigma^{-1/2}) (\hat{\phi}(p) - \phi(p)) = \frac{1}{(T-p)^{1/2}} \sum_{t=p}^{T-1} u_t + o_m(1). \quad (4.20)$$

Let us see that $\|(\Gamma_p \otimes \Sigma) - (G \Omega G')\| \rightarrow 0$. It can be proved that $\|(\Gamma_p \otimes \Sigma) - (G \Omega G')\| \leq \|\Sigma\| \cdot \|\Gamma_p - \hat{\Gamma}_p\|$ where $\hat{\Gamma}_p = (\hat{\gamma}_{u-v})_{u,v}$ and,

$$\hat{\gamma}_{u-v} = \sum_{w=\max(u,v)}^{\min(u,v)+j} \Psi_{w-u} \Sigma \Psi'_{w-v}.$$

On the other hand, γ_{u-v} can be written as the sum above but with ∞ as the upper limit. If we put $j = 2k$, then at least the first p terms of the sum are common to γ_{u-v} and $\hat{\gamma}_{u-v}$. For a symmetric matrix A , $\|A\| \leq \|A\|_{\infty} = \max_u \sum_v |a_{uv}|$. Then,

$$\begin{aligned} \|\Gamma_p - \hat{\Gamma}_p\| &\leq \max_{u=1, \dots, p} \sum_{v=1}^p \|\gamma_{u-v} - \hat{\gamma}_{u-v}\| \leq 2 \sum_{v=1}^p \|\gamma_v - \hat{\gamma}_v\| \\ &\leq 2p \left\| \sum_{\mu=p+1}^{\infty} \Psi_{\mu} \Sigma \Psi'_{\mu+v} \right\| \leq 2k \|\Sigma\| \sum_{\mu=p+1}^{\infty} \|\Psi_{\mu}\|^2, \end{aligned}$$

which by the assumptions converges to zero. Now, let us consider all matrices Γ_p and $\hat{\Gamma}_p$ as $\infty \times \infty$ matrices completed with zeros or, equivalently, as elements of the Banach space of the self-adjoint bounded linear operators from ℓ^2 (the space of the sequences of real numbers $x = (x_n)_n$ with the norm $\|x\| = \{\sum_n |x_n|^2\}^{1/2}$) into itself. Both sequences $\{\Gamma_p\}_p$ and $\{\hat{\Gamma}_p\}_p$ converge to the limit Γ_∞ . Let us now consider the mapping $s : A \mapsto s(A) = (AA)^{-1}$. The differential of s at $\Gamma_\infty^{1/2}$ is the linear operator $ds(\Gamma_\infty^{1/2})(A) = \Gamma_\infty^{-1}(\Gamma_\infty^{1/2}A + A\Gamma_\infty^{1/2})\Gamma_\infty^{-1}$.

Let us see that $ds(\Gamma_\infty^{1/2})$ is injective. Under the assumptions, Γ_∞ is nonsingular (see [19]). Thus, if $ds(\Gamma_\infty^{1/2})(A) = 0$, then $\Gamma_\infty^{1/2}A + A\Gamma_\infty^{1/2} = 0$. Consequently, for any $x \in \ell^2$, $x'(\Gamma_\infty^{1/2}A + A\Gamma_\infty^{1/2})x = 2x'A\Gamma_\infty^{1/2}x = 0$, and thus, $A\Gamma_\infty^{1/2} = 0$, but then $A = 0$. Therefore, we can apply the Inverse Function Theorem to s , so there is a differentiable inverse s^{-1} in a neighborhood of Γ_∞ . As a consequence of the definition of differential, we have that $\|s^{-1}(\Gamma_p) - s^{-1}(\hat{\Gamma}_p)\| \leq \|ds^{-1}(\Gamma_\infty)\| \cdot \|\Gamma_p - \hat{\Gamma}_p\| + o(\|\hat{\Gamma}_p - \Gamma_\infty\|) + o(\|\Gamma_p - \Gamma_\infty\|) \rightarrow 0$.

On the other hand,

$$\begin{aligned} \left\| T^{-1/2} \sum_{t=1}^T u_t - (T-p)^{-1/2} \sum_{t=p}^{T-1} u_t \right\| &\leq T^{-1/2} \left\{ E\|u_T\| + \sum_{t=1}^{p-1} E\|u_t\| \right\} + ((T-p)^{-1/2} - T^{-1/2}) \sum_{t=p}^{T-1} E\|u_t\| \\ &\leq C \frac{p(p+j)^{1/2}}{T^{1/2}} + D \frac{p(T-p-1)(p+j)^{1/2}}{T^{1/2}(T-p)^{1/2}(T^{1/2} + (T-p)^{1/2})}. \end{aligned} \quad (4.21)$$

Again, with $j = 2p$, the expression above converges to zero.

Then, from (4.20) and (4.21) we have,

$$(T-p)^{1/2}(\Gamma_p^{1/2} \otimes \Sigma^{-1/2})(\hat{\phi}(p) - \phi(p)) = \frac{1}{T^{1/2}} \sum_{t=1}^T u_t + o_m(1). \quad (4.22)$$

We can now apply Proposition 3.4 to u_t with $k = p + j$ and $v = k$ and we get, as in the proof of Proposition 4.2,

$$\ell_1 \left(T^{-1/2} \sum_{t=1}^T u_t, Z_p \right) \rightarrow 0. \quad (4.23)$$

Finally (4.16) is a consequence of (4.22) and (4.23). \square

We can apply Proposition 4.6 to compute confidence regions for the parameter vectors $\Phi(p)$. Let $\varphi_{p,u}$ be such that if $\chi_{kr,2}^2$ is a chi-square with pr^2 degrees of freedom, then $P\{\chi_{kr,2}^2 \leq \varphi_{p,u}\} = u$. We can build a confidence region $C_{p,u} = \{\Phi : (T-p)(\Phi - \hat{\Phi})'(\Gamma_p \otimes \Sigma^{-1})(\Phi - \hat{\Phi})' \leq \varphi_{p,u}\}$. If we proceed as in the proof of Proposition 4.2, we can check that $\lim_T P\{\Phi \in C_{p,u}\} = u$.

References

- [1] C.C. Heyde, G.C. Brown, On the departure from normality of a certain class of martingales, *Annals of Mathematical Statistics* 41 (1970) 2161–2165.
- [2] V.M. Bentkus, F. Götze, Uniform rates of convergence in the CLT for quadratic forms in multidimensional spaces, *Probability Theory and Related Fields* 109 (1997) 367–416.
- [3] S.T. Rachev, L. Rüschendorf, in: J. Hoffmann-Jorgensen, J. Kuelbs, B. Marcus (Eds.), On the Rate of Convergence in the CLT with respect to the Kantorovich Metric, in: *Probability in Banach spaces*, vol. 9, Birkhäuser, London, 1994, pp. 193–207.
- [4] G.E.P. Box, D.A. Pierce, Distribution of autocorrelations in autoregressive moving average time series models, *Journal of the American Statistical Association* 65 (1970) 1509–1536.
- [5] A.I. McLeod, On the distribution of residual autocorrelations in Box–Jenkins models, *Journal of the Royal Statistical Society, Series B* 40 (1978) 296–302.
- [6] G.M. Ljung, G.E.P. Box, On a measure of lack of fit in time series models, *Biometrika* 65 (1978) 297–303.
- [7] J.R.M. Hosking, The multivariate portmanteau statistic, *Journal of the American Statistical Association* 75 (1980) 602–608.
- [8] S.K. Ahn, Distribution for residual autocorrelations in multivariate autoregressive models with structured parameterization, *Biometrika* 75 (1988) 590–593.
- [9] R. Brüggemann, H. Lütkepohl, P. Saikkonen, Residual autocorrelation testing for vector error correction models, *Journal of Econometrics* 134 (2006) 579–604.
- [10] D. Peña, J. Rodríguez, A powerful portmanteau test of lack of fit for time series, *Journal of the American Statistical Association* 97 (2002) 601–619.
- [11] I. Arbués, An extended portmanteau test for VARMA models with mixing nonlinear constraints, *Journal of Time Series Analysis* 29 (2008) 741–761.
- [12] K.N. Berk, Consistent autoregressive spectral estimates, *Annals of Statistics* 2 (1974) 489–502.
- [13] R. Lewis, G.C. Reinsel, Prediction of multivariate time series by autoregressive model fitting, *Journal of Multivariate Analysis* 16 (1985) 393–411.
- [14] C.M. Bender, S.A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers*, Springer, New York, 1999.
- [15] S.T. Rachev, *Probability Metrics and the Stability of Stochastic Models*, Wiley, New York, 1991.
- [16] R. Neininger, L. Rüschendorf, A general limit for recursive algorithms and combinatorial structures, *Annals of Applied Probability* 14 (2004) 378–418.
- [17] V.M. Zolotarev, *Probability metrics, Theory of Probability and its Applications* 28 (1983) 278–302.
- [18] E.J. Hannan, M. Deistler, *The Statistical Theory of Linear Systems*, Wiley, New York, 1988.
- [19] E.J. Hannan, D.S. Poskitt, Unit canonical correlation between future and past, *Annals of Statistics* 16 (1987) 784–790.