HAC Estimation and Strong Linearity Testing in Weak ARMA Models

Christian Francq

Université Lille III, GREMARS

E-mail: francq@univ-lille3.fr

Jean-Michel Zakoïan

Université Lille III, GREMARS

and CREST

E-mail: zakoian@ensae.fr

Running head: HAC Estimation in Weak ARMA Models.

Corresponding Author: Jean-Michel Zakoïan

CREST, MK1, 3 Avenue Pierre Larousse, 92245 Malakoff Cedex France

Email :zakoian@ensae.fr

Phone number: 33(0)141176840

Fax: 33(0)141176840.

Abstract

The paper develops a procedure for testing the null hypothesis that the errors of an ARMA model are independent and identically distributed against the alternative that they are uncorrelated but not independent. The test statistic is based on the difference between a conventional estimator of the asymptotic covariance matrix of the least-squares estimator of the ARMA coefficients and its robust HAC-type version. The asymptotic distribution of the HAC estimator is established under the null hypothesis of independence, and under a large class of alternatives. The asymptotic distribution of the proposed statistic is shown to be a standard chi-square under the null, and a noncentral chi-square under the alternatives. The choice of the HAC estimator is discussed through a local power analysis. An automatic bandwidth selection method is also considered. The finite sample properties of the test are analyzed via Monte Carlo simulation.

1 Introduction

The standard statistical inference of time series models relies on methods for fitting ARMA models by model selection and estimation followed by model criticism through significance tests and diagnostic checks on the adequacy of the fitted model. The large sample distributions of the statistics involved in the so-called Box-Jenkins methodology have been established under the assumption of independent white noise. Departure from this assumption can severely alter the large sample distributions of standard statistics, such as the empirical autocovariances (Romano and Thombs (1996) and Berlinet and Francq (1999)), the estimators of the ARMA parameters (Francq and Zakoïan (1998)) or the portmanteau statistics (Lobato (2002), Francq, Roy and Zakoïan (2004)).

Unfortunately, ARMA models endowed with an independent white noise sequence, referred to as *strong* ARMA models, are often found to be unrealistic in economic applications. On the other hand, *weak* ARMA models (*i.e.* in which the noise is not required to be independent nor to be a martingale difference) have found increasing interest in the recent statistical and econometric literatures. As we will see, relaxing the strong assumptions on the noise allows for a great generality. Indeed, a large variety of well-known nonlinear processes admit (weak) ARMA representations. Other examples of weak ARMA processes are obtained from usual transformations (such as aggregation) of strong ARMA processes.

It is therefore crucial, at the model criticism stage, to be able to detect departure from the key assumption that the white noise disturbances are independent. Apart from the technical reasons already mentioned, one important motivation for testing this assumption is to find out whether or not the estimated model captures the whole dynamics of the series. Only optimal *linear* predictions can be deduced from non-strong ARMA models. Thus, when the strong ARMA assumption is rejected, improvements can be expected from using nonlinear models.

In this paper we propose a test for white noise independence in ARMA models. An informal presentation of the test is as follows, precise notations and assumptions being introduced in the next section. Let θ denote the vector of the ARMA coefficients, and let $\hat{\theta}_n$ denote the standard Least-Squares Estimator (LSE) of θ for a sample of size n. Then, under appropriate conditions, the $\sqrt{n}$-difference between $\hat{\theta}_n$ and the true parameter value converges in distribution to a $\mathcal{N}(0, \Sigma^{(1)})$ when the ARMA is strong. When the noise is only uncorrelated, the asymptotic distribution turns out to be of the form $\mathcal{N}(0, \Sigma^{(2)})$. The aim of the paper is to test

$$\mathbf{H}_0: \quad \Sigma^{(1)} = \Sigma^{(2)}, \quad \text{against} \quad \mathbf{H}_1: \quad \Sigma^{(1)} \neq \Sigma^{(2)}.$$

The Asymptotic Covariance Matrices (ACM) $\Sigma^{(1)}$ and $\Sigma^{(2)}$ coincide in case of a strong ARMA process, but they will be different, in general, in the weak situation. It should be noted that $\mathbf{H}_0$ is not equivalent to the noise independence. Hypothesis $\mathbf{H}_0$ is the consequence of independence that matters as far as the asymptotic precision of the LS estimator is concerned.

The test proposed in this paper is built on the difference between consistent estimators of the two ACM. Scaled in an appropriate way, this difference will have two different behaviors: it will converge to a non degenerate distribution if the ARMA process is strong and to infinity otherwise.

The matrix $\Sigma^{(2)}$ is function of a matrix of the form $\lim_{n\to\infty} \frac{1}{n} \sum_{s,t=1}^{n} \mathrm{Cov}(V_t, V_s)$, which is also 2π times the spectral density matrix of the multivariate process (V_t) evaluated at frequency zero. Many papers in the statistical and econometric literatures deal with estimating such 'long-run variance' matrices. Examples include the estimation of the optimal weighting matrix for General Method of Moments estimators (Hansen, 1982), the estimation of the covariance matrix of the error term in unit root tests (Phillips, 1987), the estimation of the asymptotic variance of sample autocovariances of nonlinear processes (Berlinet and Francq, 1999); other important econometric contributions include Newey and West (1987), Gallant and White (1988), Andrews (1991), Hansen (1992), de Jong and Davidson (2000).

However, the asymptotic distribution of the ACM estimators is seldom considered in the literature dealing with HAC estimation. See Phillips, Sun and Jin (2003, Theorem 2) for the asymptotic distribution for a HAC estimator in the framework of robust regression. The main difficulty of the present paper is to derive the asymptotic distributions of the HAC estimator of $\Sigma^{(2)}$, under both assumptions of weak and strong ARMA. Those distributions are needed to construct the asymptotic critical region of our test and to derive its local asymptotic power.

The order of the paper is as follows. Section 2 introduces the notion of weak ARMA representations. Examples treated explicitly are the first component of a strong bivariate MA(1) model and a chaotic process. Section 3 presents notations and briefly overviews results concerning the asymptotic behaviour of the LSE in the weak ARMA framework. Section 4 establishes the asymptotic distribution of the above-mentioned HAC estimator. Section 5 introduces the test statistic and derives its asymptotic properties under the null of strong ARMA, and under the alternative of weak ARMA model. In Section 6, the choice of the HAC estimator is discussed through a local power analysis and an automatic bandwidth selection method. The finite sample performance of the tests is studied. Proofs and additional notations are in Section 7. Section 8 concludes.

2 Weak ARMA models

It is not very surprising that number of non linear processes admit an ARMA representation. Indeed, the Wold Theorem states that any purely non deterministic, second-order stationary process (X_t) can be represented by an infinite MA representation. When the infinite MA polynomial is obtained as the ratio of two finite-order polynomials, the nonlinear process also admits a finite-order ARMA representation. This representation is weak because the noise is only the linear innovation of (X_t) (otherwise (X_t) would be a linear process). This distinction has important consequences in terms of prediction. The predictors obtained from a weak ARMA model are only optimal in the linear sense. They are particularly useful if the nonlinear dynamics of (X_t) is difficult to identify, which is often the case given the variety of non linear models. It is also crucial to take into account the differences between weak and strong ARMA models in the different steps of the methodology of Box and Jenkins. When the noise is only uncorrelated, the tools provided by the standard Box-Jenkins methodology can be quite misleading. Recent papers cited in the introduction have been devoted to the correction of such tools and have attempted to develop tests and methods allowing to work with a broad class of ARMA models. It is, however, of importance to know, in practical situations, when these new methods are required and when the classical ones are reliable. Needless to say that the latter procedures are simpler to use, in particular because they are widely implemented in all standard statistical softwares. It is precisely the purpose of the present paper to develop a test of the reliability of such procedures.

Examples of ARMA representations of bilinear processes, Markov switching processes, threshold processes, processes obtained by temporal aggregation of a strong ARMA, or by linear combination of independent ARMA processes can be found in Romano and Thombs (1996), Francq and Zakoïan (1998, 2000). We will give two new illustrations which have their own interest. To construct a simple example of weak ARMA model obtained by transformation of a strong ARMA process, let us consider the following bivariate MA(1) model

$$\begin{pmatrix} X_{1\,t} \\ X_{2\,t} \end{pmatrix} = \begin{pmatrix} \eta_{1\,t} \\ \eta_{2\,t} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} \eta_{1\,t-1} \\ \eta_{2\,t-1} \end{pmatrix}$$

where $\{\eta_{1\,t}, \eta_{2\,t}\}_t'$ is a centered iid sequence with covariance matrix (ξ_{ij}). It is easy to see that the first component of the strong bivariate MA(1) model satisfies a univariate MA(1) model of the form $X_{1\,t} = \epsilon_t + \theta\epsilon_{t-1}$, where $\theta \in (-1, 1)$ is such that

$$\frac{\theta}{1+\theta^2} = \frac{EX_{1\,t}X_{1\,t-1}}{EX_{1\,t}^2} = \frac{\xi_{11}b_{11} + \xi_{12}b_{12}}{\xi_{11}(1+b_{11}^2) + \xi_{22}b_{12}^2 + 2\xi_{12}b_{11}b_{12}}.$$

From $\epsilon_t + \theta\epsilon_{t-1} = \eta_{1\,t} + b_{11}\eta_{1\,t-1} + b_{12}\eta_{2\,t-1}$, we find

$$\epsilon_t = (b_{11} - \theta)(\eta_{1\,t-1} - \theta\eta_{1\,t-2}) + b_{12}(\eta_{2\,t-1} - \theta\eta_{2\,t-2}) + R_t,$$

where R_t is centered and independent of $X_{1\,t-1}$. Therefore

$$\begin{aligned} EX_{1\,t-1}^2\epsilon_t &= (b_{11} - \theta)\left\{(1 - \theta b_{11}^2)E\eta_{1\,t}^3 - \theta b_{12}^2 E\eta_{1\,t}\eta_{2\,t}^2 - 2\theta b_{11}b_{12}E\eta_{1\,t}^2\eta_{2\,t}\right\} \\ &\quad + b_{12}\left\{(1 - \theta b_{11}^2)E\eta_{1\,t}^2\eta_{2\,t} - \theta b_{12}^2 E\eta_{2\,t}^3 - 2\theta b_{11}b_{12}E\eta_{1\,t}\eta_{2\,t}^2\right\}. \end{aligned}$$

The random variable $X_{1\,t-1}^2$ belongs to $\sigma\{\epsilon_s, s < t\}$. It can be seen that, in general, $EX_{1\,t-1}^2\epsilon_t = E[X_{1\,t-1}^2 E(\epsilon_t|\epsilon_{t-1}, \ldots)] \neq 0$. Thus, (ϵ_t) is not a martingale difference. So the MA(1) for $X_{1\,t}$ is only weak.

Now we will give an example based on a chaotic process (see May (1976)). Let

$$\epsilon_t = u_t + \eta_t, \quad u_t = 4u_{t-1}(1 - u_{t-1}), \quad t \geq 1 \tag{1}$$

where u_0 has the arc-sinus density $f(x) = \pi^{-1}\{x(1-x)\}^{-1/2}$ on the interval $[0, 1]$, $(\eta_t)_{t\geq 1}$ is an iid sequence, independent from u_0, with mean -1/2 and finite variance. Since f is the invariant density of (u_t), this process is stationary. We have $E\epsilon_t = 0$ and, since u_t and $1 - u_t$ have the same law,

$$\begin{aligned} \mathrm{Cov}(\epsilon_t, \epsilon_{t-1}) &= \mathrm{Cov}(u_t, u_{t-1}) \\ &= \mathrm{Cov}\{4u_{t-1}(1 - u_{t-1}), u_{t-1}\} = \mathrm{Cov}(u_t, 1 - u_{t-1}) = 0. \end{aligned}$$

The same symmetry argument shows that $\mathrm{Cov}(\epsilon_t, \epsilon_{t-h}) = 0$ for all $h \neq 0$. Therefore (ϵ_t) is a white noise. Consequently, given $\{\epsilon_u, u \leq t\}$, the best linear predictor of ϵ_{t+h} is equal to zero, for any horizon h. However, in general, the best (nonlinear) predictor is quite different. For illustrative purpose, Figure 1 displays the scatter plot of the pairs $(\epsilon_t, \epsilon_{t-2})$, for $t = 1, \ldots, 1,000$, obtained by simulation, and the nonlinear regression obtained by tedious computation, in the case where η_t is uniformly distributed over $[-0.6, -0.4]$.

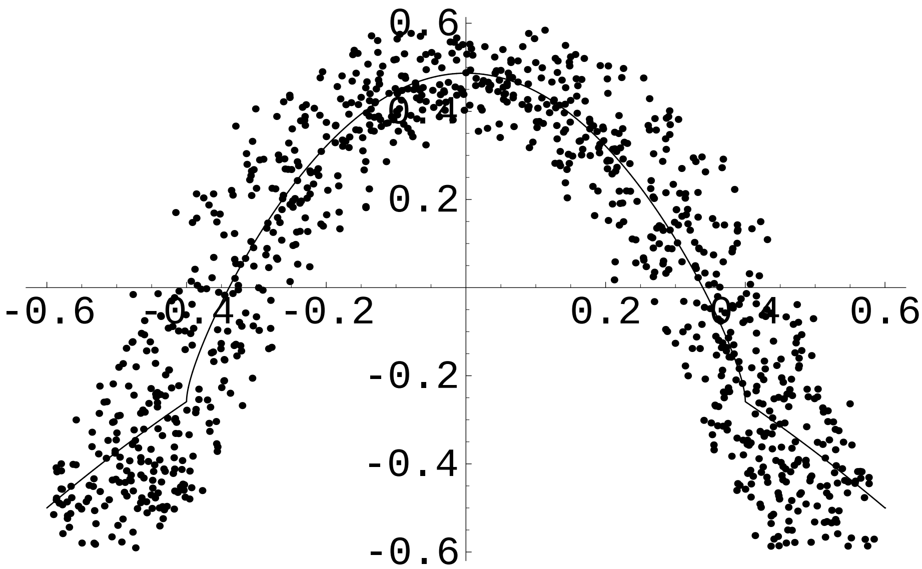

Figure 1: Scatter plot of 1,000 pairs $(\epsilon_t, \epsilon_{t-2})$, simulated from (1) with η_t uniformly distributed over $[-0.6, -0.4]$. The full line is the theoretical (nonlinear) regression of ϵ_t on ϵ_{t-2}.

This example illustrates the, possibly dramatic, differences between linear and nonlinear predictions of a given weak ARMA process.

3 Notations and preliminary asymptotic results

In this section we introduce the main notations and recall some results established in Francq and Zakoïan (1998, 2000). Let $X = (X_t)_{t\in\mathbb{Z}}$ be a real second-order stationary ARMA(p,q) process such that, for all $t \in \mathbb{Z}$,

$$X_t + \sum_{i=1}^{p} \phi_i X_{t-i} = \epsilon_t + \sum_{i=1}^{q} \psi_i \epsilon_{t-i}. \tag{2}$$

We consider the estimation of parameter $\theta = (\theta_1, \ldots, \theta_{p+q})' \in \mathbb{R}^{p+q}$ with true value $\theta_0 = (\phi_1, \ldots, \phi_p, \psi_1, \ldots, \psi_q)'$. Let $\Phi_\theta(z) = 1 + \theta_1 z + \cdots + \theta_p z^p$ and $\Psi_\theta(z) = 1 + \theta_{p+1} z + \cdots + \theta_{p+q} z^q$ be the AR and MA polynomials.[1] For any $\delta > 0$, let the compact set

$$\Theta_\delta = \{\theta \in \mathbb{R}^{p+q};\ \text{ the zeros of polynomials } \Phi_\theta(z) \text{ and } \Psi_\theta(z) \text{ have moduli} \geq 1+\delta\}.$$

We make the following assumptions.

A1. $\epsilon = (\epsilon_t)$ is a strictly stationary sequence of uncorrelated random variables with zero mean and variance $\sigma^2 > 0$, defined on some probability space (Ω, $\mathcal{A}$, P).

A2. θ_0 belongs to the interior of Θ_δ, and the polynomials $\Phi_{\theta_0}(z)$ and $\Psi_{\theta_0}(z)$ have no zero in common.

A3. ϕ_p and ψ_q are not both equal to zero (by convention $\phi_0 = \psi_0 = 1$).

For all $\theta \in \Theta$, let $\epsilon_t(\theta) = \Psi_\theta^{-1}(B)\Phi_\theta(B)X_t$, where B denotes the backshift operator. Given a realization $X_1, X_2, \ldots, X_n$ of X, the $\epsilon_t(\theta)$ can be approximated, for $0 < t \leq n$, by $e_t(\theta) = \Psi_\theta^{-1}(B)\Phi_\theta(B)(X_t 1_{1\leq t\leq n})$.

The random variable $\hat{\theta}_n$ is called Least Squares Estimator (LSE) if it satisfies, almost surely,

$$Q_n(\hat{\theta}_n) = \min_{\theta\in\Theta_\delta} Q_n(\theta) \quad \text{where} \quad Q_n(\theta) = \frac{1}{2n}\sum_{t=1}^{n} e_t^2(\theta). \tag{3}$$

Define the strong ARMA assumption:

A4. The process X is solution of Model (2) where the random variables ϵ_t are independent and identically distributed (iid).

In the ARMA models statistical literature, most results are obtained under **A4**. Less restrictive hypotheses rely on the $\alpha-$mixing (strong mixing) coefficients $\{\alpha_\epsilon(h)\}_{h\geq 0}$ for (ϵ_t), or $\{\alpha_X(h)\}_{h\geq 0}$ for (X_t). Let us consider the assumptions:

A5. The process X is solution of Model (2) where $\sum_{h=0}^{\infty}\{\alpha_\epsilon(h)\}^{\frac{\nu}{2+\nu}} < \infty$, for some $\nu > 0$.

A5′. The process X is solution of Model (2) where $\sum_{h=0}^{\infty}\{\alpha_X(h)\}^{\frac{\nu}{2+\nu}} < \infty$, for some $\nu > 0$.

It is well known that **A5** and **A5′** are not equivalent. Pham (1986), Carrasco and Chen (2002) have shown that, for a wide class of processes, the mixing conditions **A5** and/or **A5′** are satisfied. Let

$$J(\theta) = \lim_{n\to\infty} \frac{\partial^2}{\partial\theta\partial\theta'} Q_n(\theta) \quad a.s., \qquad I(\theta) = \lim_{n\to\infty} \mathrm{Var}(\sqrt{n}\frac{\partial}{\partial\theta} Q_n(\theta)).$$

The following theorem gives already-established results concerning the asymptotic behaviour of the LSE. The symbol $\overset{d}{\leadsto}$ denotes convergence in distribution as the sample size n goes to infinity.

Theorem 1 *Assume that* **A1** $-$ **A3** *hold.*

If $E\epsilon_t^2 < \infty$ *and* **A4** *holds, then* $\hat{\theta}_n$ *is strongly consistent and*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{d}{\leadsto} \mathcal{N}(0, \Sigma^{(1)}), \qquad \textit{where} \quad \Sigma^{(1)} = \sigma^2 J^{-1}(\theta_0). \tag{4}$$

If $E\epsilon_t^{4+2\nu} < \infty$ *and either* **A4**, **A5** *or* **A5′** *holds, then* $\hat{\theta}_n$ *is strongly consistent and*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{d}{\leadsto} \mathcal{N}(0, \Sigma^{(2)}), \qquad \textit{where} \quad \Sigma^{(2)} = J^{-1}(\theta_0) I(\theta_0) J^{-1}(\theta_0). \tag{5}$$

Obviously the moment assumptions on ϵ could be replaced by the same assumption on X. The convergence under **A4** is standard (see e.g. Brockwell and Davis (1991)). Francq and Zakoïan (1998) established (5) under **A5′**. It can be shown that the result remains valid under **A5**. Note that the finiteness of $E\epsilon_t^4$ is required for the existence of $\Sigma^{(2)}$.[2]

Remarks

(a) Straightforward computations show that

$$I(\theta) = \sum_{i=-\infty}^{+\infty} \Delta_i(\theta), \quad \text{where} \quad \Delta_i(\theta) = E\mathbf{u}_t(\theta)\mathbf{u}'_{t+i}(\theta), \quad \mathbf{u}_t(\theta) = \epsilon_t(\theta)\frac{\partial \epsilon_t}{\partial \theta}(\theta).$$

(b) Under **A4** the asymptotic variances $\Sigma^{(1)}$ and $\Sigma^{(2)}$ are clearly equal. However it may also be the case that $\Sigma^{(1)} = \Sigma^{(2)}$ under **A5** or **A5′**. More precisely we have

$$\begin{aligned} \Sigma^{(1)} = \Sigma^{(2)} \quad &\Leftrightarrow \quad I(\theta_0) = \sigma^2 J(\theta_0) \\ &\Leftrightarrow \quad \sum_{i=-\infty}^{+\infty} E\mathbf{u}_t(\theta_0)\mathbf{u}'_{t+i}(\theta_0) = \sigma^2 E\left(\frac{\partial \epsilon_t}{\partial \theta}(\theta_0)\right)\left(\frac{\partial \epsilon_t}{\partial \theta}(\theta_0)\right)' \end{aligned}$$

where all quantities are taken at θ_0. In particular, the asymptotic variances are equal when the process $\left(\epsilon_t \frac{\partial \epsilon_t}{\partial \theta}\right)$ is a noise and ϵ_t^2 is uncorrelated with $\left(\frac{\partial \epsilon_t}{\partial \theta}\right)\left(\frac{\partial \epsilon_t}{\partial \theta}\right)'$. In the case of a martingale difference the former condition holds but, in general, the latter condition does not. More precisely we have, when (ϵ_t) is a martingale difference

$$\Sigma^{(1)} - \Sigma^{(2)} = E\left(\epsilon_t^2 \frac{\partial \epsilon_t}{\partial \theta}\frac{\partial \epsilon_t}{\partial \theta'}\right) - E\left(\epsilon_t^2\right) E\left(\frac{\partial \epsilon_t}{\partial \theta}\right)\left(\frac{\partial \epsilon_t}{\partial \theta}\right)'.$$

For example if the model is an AR(1) with true value $\theta_0 = 0$, then $\frac{\partial \epsilon_t}{\partial \theta} = X_{t-1} = \epsilon_{t-1}$. Thus we have $\Sigma^{(1)} - \Sigma^{(2)} = \mathrm{Cov}(\epsilon_t^2, \epsilon_{t-1}^2)$, which is not equal to zero in presence of ARCH-type conditional heteroskedasticity. When (ϵ_t) is not a martingale difference, the sequence $\left(\epsilon_t \frac{\partial \epsilon_t}{\partial \theta}\right)$ is not uncorrelated in general and the difference between the matrices $\Sigma^{(1)}$ and $\Sigma^{(2)}$ can be substantial, as illustrated in Figure 2. Interestingly, it is also seen on this example that $\Sigma^{(1)} - \Sigma^{(2)}$ is not always positive definite. Therefore, for some linear combinations of the ARMA parameters, a better asymptotic accuracy may be obtained when the noise is weak than when it is strong. The same remark was made by Romano and Thombs (1996) on another example.

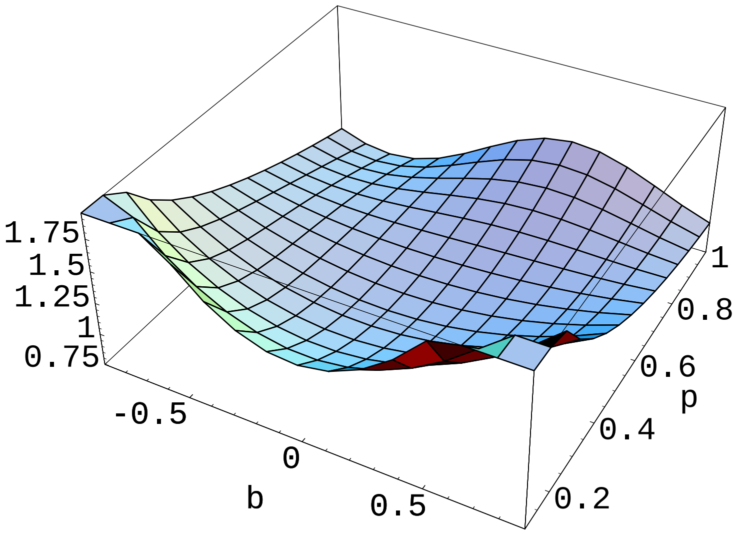

Figure 2: $\Sigma^{(2)}(1,1)/\Sigma^{(1)}(1,1)$ as function of $b \in [-0.9, 0.9]$ and p $\in [0.1, 1]$ for the model

$$\forall t \in Z, \quad \begin{cases} X_t + aX_{t-1} = \epsilon_t + b\epsilon_{t-1} \\ \epsilon_t = \eta_t + (c - 2c\Delta_t)\eta_{t-1}, \end{cases}$$

where $a = -0.5$ and $c = 1$, (η_t) is an i.i.d. $\mathcal{N}(0,1)$ sequence, (Δ_t) is a stationary Markov chain, independent of (η_t), with state space $\{0,1\}$ and transition probabilities p $= P(\Delta_t = 1|\Delta_{t-1} = 0) = P(\Delta_t = 0|\Delta_{t-1} = 1) \in (0,1)$. It can be shown that (ϵ_t) is a white noise.

Consistent estimation of the matrix $J = J(\theta_0) = E\frac{\partial}{\partial\theta}\epsilon_t(\theta_0)\frac{\partial}{\partial\theta'}\epsilon_t(\theta_0)$ involved in (4) and (5) is straightforward, for example by taking

$$\hat{J} = \hat{J}_n(\hat{\theta}_n), \qquad \hat{J}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}e_t(\theta)\frac{\partial}{\partial\theta'}e_t(\theta). \tag{6}$$

Estimation of the matrix $I = I(\theta_0)$ is a much more intricate problem and is the object of the next section.

4 Asymptotic distribution of the HAC estimator of the covariance matrix I

The formula displayed in Remark **(a)** of Theorem 1 motivates the introduction of a HAC estimator of I of the general form

$$\hat{I} \;=\; \hat{I}_n(\hat{\theta}_n) = \sum_{i=-\infty}^{+\infty}\omega(ib_n)\hat{\Delta}_i(\hat{\theta}_n), \tag{7}$$

involving the ARMA residuals $e_t(\hat{\theta}_n)$ through the functions

$$\hat{\Delta}_i(\theta) = \hat{\Delta}_{-i}(\theta)' = \begin{cases} \frac{1}{n}\sum_{t=1}^{n-i} e_t(\theta)\frac{\partial}{\partial\theta}e_t(\theta)\frac{\partial}{\partial\theta'}e_{t+i}(\theta)e_{t+i}(\theta), & 0 \le i < n, \\ 0, & i \ge n. \end{cases}$$

In (7), $\omega(\cdot)$ is a kernel function belonging to the set $\mathcal{K}$ defined below, and b_n is a size-dependent bandwidth parameter. When i is large relative to n, $\Delta_i(\theta_0)$ is in general poorly estimated by $\hat{\Delta}_i(\hat{\theta}_n)$ which is based on too few observations. Consistency of $\hat{I}$ therefore requires that the weights $\omega(ib_n)$ be close to one for small i and close to zero for large i. In particular, the naive estimator $\sum_{i=-\infty}^{\infty}\hat{\Delta}_i(\hat{\theta}_n)$ is inconsistent.[3] The set $\mathcal{K}$ is defined by

$$\begin{aligned} \mathcal{K} \;=\; & \{\omega(\cdot) : \mathbb{R} \to \mathbb{R} \;\;|\;\; \omega(0) = 1,\;\; \omega(\cdot) \text{ is bounded, even,} \\ & \quad \text{has a compact support } [-a, a] \text{ and is continuous on } [-a, a]\}\,. \end{aligned} \tag{8}$$

Various kernels $\omega(\cdot)$ belonging to $\mathcal{K}$ are available for use in (7). Standard examples are the rectangular kernel $\omega(x) = 1_{[-1,1]}(x)$, the Bartlett kernel $\omega(x) = (1-|x|)1_{[-1,1]}(x)$, the Parzen kernel $\omega(x) = (1-6x^2+6|x|^3)1_{[0,1/2]}(|x|) + 2(1-|x|)^3 1_{(1/2,1]}(|x|)$ or the Tukey-Hanning kernel $\omega(x) = \{1/2+\cos(\pi x)/2\}\, 1_{[-1,1]}(x)$. The properties of kernel functions have been extensively studied in the time series literature (see e.g. Priestley (1981)). For a given kernel $\omega(\cdot)$ in $\mathcal{K}$, let $\varpi^2 = \int \omega^2(x)dx$. Our asymptotic normality result on $\hat{I}$ requires[4]

$$\lim_{n\to\infty} b_n = 0, \quad \lim_{n\to\infty} nb_n^4 = +\infty. \tag{9}$$

We denote by $A \otimes B$ the Kronecker product of two matrices A and B, vec$A$ denotes the vector obtained by stacking the columns of A, and vech$A$ denotes the vector obtained by stacking the diagonal and subdiagonal elements of A (see *e.g.* Harville (1997) for more details about these matrix operators). Let D_m denote the $m^2 \times m(m+1)/2$ duplication matrix,[5] and let $D_m^+ = (D_m' D_m)^{-1} D_m'$. Thus $D_m^+ \text{vec}\,(A) = \text{vech}\,(A)$ when $A = A'$.

4.1 Strong ARMA

Our first result is stated in the following theorem.

Theorem 2 *Let* **A1** − **A3** *hold and let (9) be satisfied.*

If $E\epsilon_t^{4+\nu} < \infty$ *and* **A4** *holds, then* $\hat{I}$ *defined by (7), with* $\omega(\cdot) \in \mathcal{K}$, *is consistent and*

$$\sqrt{nb_n}\ vech\left(\hat{I} - I\right) \overset{d}{\rightsquigarrow} \mathcal{N}\left(0, 2\varpi^2 D_{p+q}^+ (I \otimes I) D_{p+q}^{+'}\right)$$

Next, we give a corresponding result for the estimator of the difference between the ACM under the two set of assumptions, $\Sigma^{(1)} - \Sigma^{(2)}$, upon which the test of the next section will be based.

Let

$$\hat{\Sigma}^{(1)} = \hat{\sigma}^2 \hat{J}^{-1}, \quad \hat{\Sigma}^{(2)} = \hat{J}^{-1}\hat{I}\hat{J}^{-1}, \quad \hat{\sigma}^2 = Q_n(\hat{\theta}_n).$$

Theorem 3 *Let the assumptions of Theorem 2 hold. Then*

$$\sqrt{nb_n}\ vech\left(\hat{\Sigma}^{(2)} - \hat{\Sigma}^{(1)}\right) \overset{d}{\rightsquigarrow} \mathcal{N}\left(0, 2\varpi^2 D^+_{p+q}(\Sigma^{(1)} \otimes \Sigma^{(1)})D^{+'}_{p+q}\right) := \mathcal{N}(0, \Lambda)$$

The asymptotic distribution of $\sqrt{nb_n}$ vech $\left(\hat{\Sigma}^{(2)} - \hat{\Sigma}^{(1)}\right)$ is non degenerate. The non-singularity of $\Sigma^{(1)}$ results from Theorem 1 and the determinant of Λ is, up to a constant, equal to the determinant of $\Sigma^{(1)}$ at the power $(p+q)^2+1$ (see Magnus and Neudecker, Theorem 3.14, 1988). Moreover, we have

$$\Lambda^{-1} = (2\sigma^4\varpi^2)^{-1}D'_{p+q}(J \otimes J)D_{p+q} = (2\varpi^2)^{-1}D'_{p+q}(\Sigma^{(1)} \otimes \Sigma^{(1)})^{-1}D_{p+q}. \quad (10)$$

Explicit expressions for the asymptotic covariance matrices appearing in Theorems 1 and 3 can be obtained for the MA(1) and AR(1) models.

Corollary 1 *Let $X_t = \epsilon_t + \psi_1\epsilon_{t-1}$, where ϵ_t iid $(0, \sigma^2)$, $\sigma^2 > 0$ and $|\psi_1| < 1$. Then Theorem 1 holds with $\Sigma^{(1)} = \Sigma^{(2)} = 1 - \psi_1^2$ and*

$$\sqrt{nb_n}\left(\hat{\Sigma}^{(2)} - \hat{\Sigma}^{(1)}\right) \overset{d}{\rightsquigarrow} \mathcal{N}(0, 2\varpi^2(1-\psi_1^2)^2).$$

The same results hold for the stationary solution of the AR(1) model $X_t + \phi_1 X_{t-1} = \epsilon_t$ with $\phi_1 = \psi_1$ and under the same noise assumptions.

4.2 Weak ARMA

Next we state additional assumptions on the kernel at zero and on $\mathbf{u}_t(\theta_0)$ to obtain asymptotic results for the HAC estimator when the ARMA is not strong. Following Parzen (1957), define

$$\omega_{(r)} = \lim_{x \to 0} \frac{1-\omega(x)}{|x|^r} \qquad \text{for} \quad r \in [0, +\infty).$$

The largest exponent r such that $\omega_{(r)}$ exists and is finite characterizes the smoothness of $\omega(\cdot)$ at zero.[6] Let the matrix

$$I^{(r)} = \sum_{i=-\infty}^{\infty} |i|^r \Delta_i(\theta_0) \qquad \text{for} \quad r \in [0, +\infty).$$

Let

$$\mathbf{u}_t = \mathbf{u}_t(\theta_0) = \epsilon_t \frac{\partial \epsilon_t}{\partial \theta}(\theta_0) := (u_t(1), \dots, u_t(p+q))'.$$

Denote by $\kappa_{\ell_1,\dots,\ell_8}(0, j_1, \dots, j_7)$ the eighth order cumulant of $(u_t(\ell_1), u_{t+j_1}(\ell_2), \dots, u_{t+j_7}(\ell_8))$ (see Brillinger, 1981 p.19), where $\ell_1, \dots, \ell_8$ are positive integers less than $p+q$ and $j_1, \dots, j_7$ are integers. We consider the following assumptions.

A6. $E(\epsilon_t^{16}) < \infty$ and for all $\ell_1, \dots, \ell_8$

$$\sum_{j_1=-\infty}^{+\infty} \cdots \sum_{j_7=-\infty}^{+\infty} |\kappa_{\ell_1,\dots,\ell_8}(0, j_1, \dots, j_7)| < \infty$$

A7. For some $r_0 \in (0, +\infty)$,

$$\lim_{n\to\infty} n b_n^{2r_0+1} = \gamma \in [0, +\infty), \quad \omega_{(r_0)} < \infty \quad \text{and} \quad \|I^{(r_0)}\| < \infty.$$

The following result is a consequence of Andrews (Theorem 1, 1991) (and partly of Francq and Zakoïan (1998, 2000)):

Theorem 4 *Let* **A1** − **A3** *and* **A6** − **A7** *hold with* $\gamma > 0$. *Then for* $\omega(\cdot) \in \mathcal{K}$,

$$\begin{aligned} & \lim_{n\to\infty} n b_n E \left\{ vec\left(\hat{I} - I\right) \right\}' vec\left(\hat{I} - I\right) \\ = \;& \omega_{(r_0)}^2 \left\{ vec I^{(r_0)} \right\}' \left\{ vec I^{(r_0)} \right\} \gamma + 2\varpi^2 \mathrm{tr}\left\{ D_{p+q} D_{p+q}^{+} (I \otimes I) \right\}. \qquad (11) \end{aligned}$$

Note that this result is not in contradiction with Theorem 2. Indeed, under **A4**, $I^{(r)} = 0$ for $r > 0$, so the first term in the right-hand side of (11) vanishes.

As noted by Andrews (1991), it seems likely that Assumption **A6** could be replaced by mixing plus moment conditions, such as Assumption **A5**, or **A5′**, and the moment condition $E\epsilon_t^{16+\nu} < \infty$. We will do so in the next theorem.

We introduce the following mixing and moment assumptions on the process $\mathbf{u} = (\mathbf{u}_t)$.

A5″. **A7** holds with $\gamma = 0$, $\|\mathbf{u}_t\|_{8+\nu} < \infty$ and $\sum_{h=0}^{\infty} h^r \{\alpha_{\mathbf{u}}(h)\}^{\frac{\nu}{8+\nu}} < \infty$, for some $\nu > 0$ and some r such that $r \geq 2$, $r > (3\nu - 8)/(8 + \nu)$ and $r \geq r_0$.

The extensions of Theorems 2 and 3 to weak ARMA can be formulated as follows.

Theorem 5 *Let* **A1** − **A3** *and* **A5″** *hold. Assume there exists* $\kappa < 1/6$ *such that* $\liminf_{n\to\infty} nb_n^{1/\kappa} > 0$. *Then the convergence in distribution of Theorem 2 holds, and that of Theorem 3 becomes*

$$\sqrt{nb_n}\ vech\left(\hat{\Sigma}^{(2)} - \hat{\Sigma}^{(1)} - \Sigma^{(2)} + \Sigma^{(1)}\right) \overset{d}{\leadsto} \mathcal{N}\left(0, 2\varpi^2 D_{p+q}^{+}(\Sigma^{(2)} \otimes \Sigma^{(2)}) D_{p+q}^{+'}\right).$$

To prove this theorem we searched for a CLT for a triangular array $(x_{n,t})$ (see Equation (41) below) such that $x_{n,t}$ is a measurable function of $\mathbf{u}_{t-T_n}$, $\mathbf{u}_{t-T_n+1}$, $\dots$,$\mathbf{u}_{t+T_n}$, for some mixing process $\mathbf{u}_t$, with $T_n = [a/b_n] \to \infty$. Denote by $\alpha_n(\cdot)$ the mixing coefficients of $(x_{n,t})_t$. It seems that the existing CLT's (see *e.g.* Withers (1981)) require conditions on $\sup_n \alpha_n(\cdot)$ or other conditions which are difficult to check in our framework. We therefore establish the following Lemma, which can be viewed as a direct extension of the CLT given by Herrndorf (1984) in the case of a non stationary sequence (x_t), but may have its own interest.

Lemma 1 *Let* $(x_{n,t})_{n\geq 1, 1\leq t\leq n}$ *be a triangular array of centered real-valued random variables. For each* $n \geq 2$, *let* $\alpha_n(h)$, $h = 1, \dots, n-1$, *be the strong mixing coefficients of* $x_{n,1}, \dots, x_{n,n}$, *defined by:*

$$\alpha_n(h) = \sup_{1\leq t\leq n-h} \sup_{A\in\mathcal{A}_{n,t},\ B\in\mathcal{B}_{n,t+h}} |P(A \cap B) - P(A)P(B)|.$$

where $\mathcal{A}_{n,t} = \sigma(x_{n,u} : 1 \le u \le t)$ *and* $\mathcal{B}_{n,t} = \sigma(x_{n,u} : t \le u \le n)$. *By convention, we set* $\alpha_n(h) = 1/4$ *for* $h \le 0$ *and* $\alpha_n(h) = 0$ *for* $h \ge n$. *Put* $S_n = \sum_{t=1}^n x_{n,t}$ *and assume that*

$$\sup_{n\ge 1}\sup_{1\le t\le n} \|x_{n,t}\|_{2+\nu^*} < \infty \quad \textit{for some } \nu^*(0,\infty], \tag{12}$$

$$\lim_{n\to\infty} n^{-1}\, VarS_n = \sigma^2 > 0, \tag{13}$$

there exist a sequence of integers (T_n) *such that*

$$T_n = O(n^\kappa) \quad \textit{for some } \kappa \in [0, \nu^*/\{4(1+\nu^*)\}), \tag{14}$$

and a sequence $\{\alpha(h)\}_{h\ge 1}$ *such that*

$$\alpha_n(h) \le \alpha(h - T_n), \quad \textit{for all } h > T_n, \tag{15}$$

$$\sum_{h=1}^{\infty} h^{r^*}\alpha(h)^{\frac{\nu^*}{2+\nu^*}} < \infty \quad \textit{for some } r^* > \frac{2\kappa(1+\nu^*)}{\nu^* - 2\kappa(1+\nu^*)}. \tag{16}$$

Then

$$n^{-1/2}S_n \overset{d}{\leadsto} \mathcal{N}(0,\sigma^2).$$

Theorem 5 applies for kernels which are very smooth at zero. Indeed, the conditions $\lim_{n\to\infty} nb_n^{2r_0+1} = 0$ and $\liminf_{n\to\infty} nb_n^6 \ne 0$ imply $r_0 > 5/2$. The following theorem shows that this smoothness condition can be weaken when moment assumptions are added. The proof is similar to that of Theorem 5 and is therefore skipped.

Theorem 6 *Let* **A1**$-$**A3** *and* **A5**$''$ *hold with* $\nu = \infty$. *Assume there exists* $\kappa < 1/4$ *such that* $\liminf_{n\to\infty} nb_n^{1/\kappa} > 0$. *Then the convergences in distribution of Theorem 5 hold.*

The results of this section will now be used to derive the asymptotic level of our test statistic.

5 Testing adequacy of the standard asymptotic distribution

The results of Section 3 show that the asymptotic variances of the LSE under strong and weak assumptions can be dramatically different. Standard statistical routines estimate the asymptotic variance corresponding to strong ARMA models and it is of importance to know if the resulting tests (or confidence intervals) are reliable. The aim of the present section is therefore to test the assumptions presented in the introduction, which we recall for convenience:

$$\mathbf{H}_0: \quad \Sigma^{(1)} = \Sigma^{(2)}, \qquad \text{against} \qquad \mathbf{H}_1: \quad \Sigma^{(1)} \neq \Sigma^{(2)}.$$

It should be clear that under both assumptions the ARMA model is well-specified. In particular, the case of serial correlation of (ϵ_t) is not considered. A statistic derived from Theorem 3 is

$$\Upsilon_n = nb_n \left\{ \text{vech} \left(\hat{\Sigma}^{(2)} - \hat{\Sigma}^{(1)} \right) \right\}' \hat{\Lambda}^{-1} \left\{ \text{vech} \left(\hat{\Sigma}^{(2)} - \hat{\Sigma}^{(1)} \right) \right\} \tag{17}$$

where $\hat{\Lambda}^{-1}$ is any consistent estimator of Λ^{-1}. In view of (10) we can take

$$\hat{\Lambda}^{-1} = (2\hat{\sigma}^4 \varpi^2)^{-1} D'_{p+q} (\hat{J} \otimes \hat{J}) D_{p+q},$$

which does not require any matrix inversion. Then we have

$$\Upsilon_n = nb_n (2\hat{\sigma}^4 \varpi^2)^{-1} \left\{ \text{vec} \left(\hat{\Sigma}^{(2)} - \hat{\Sigma}^{(1)} \right) \right\}' (\hat{J} \otimes \hat{J}) \left\{ \text{vec} \left(\hat{\Sigma}^{(2)} - \hat{\Sigma}^{(1)} \right) \right\}.$$

But since

$$\begin{aligned}
\left\{ \text{vec}\hat{\Sigma}^{(1)} \right\}' (\hat{J} \otimes \hat{J}) \text{vec}\hat{\Sigma}^{(1)} &= \hat{\sigma}^4 \left\{ \text{vec}\hat{J}^{-1} \right\}' \text{vec}\hat{J} = \hat{\sigma}^4 \text{tr}(\hat{J}^{-1}\hat{J}) = \hat{\sigma}^4 (p+q), \\
\left\{ \text{vec}\hat{\Sigma}^{(1)} \right\}' (\hat{J} \otimes \hat{J}) \text{vec}\hat{\Sigma}^{(2)} &= \hat{\sigma}^2 \left\{ \text{vec}\hat{J}^{-1} \right\}' \text{vec}\hat{I} = \hat{\sigma}^2 \text{tr}(\hat{J}^{-1}\hat{I}), \\
\left\{ \text{vec}\hat{\Sigma}^{(2)} \right\}' (\hat{J} \otimes \hat{J}) \text{vec}\hat{\Sigma}^{(2)} &= \left\{ \text{vec}\hat{J}^{-1}\hat{I}\hat{J}^{-1} \right\}' \text{vec}\hat{I} = \text{tr}\{(\hat{J}^{-1}\hat{I})^2\},
\end{aligned}$$

we get

$$\Upsilon_n = nb_n(2\hat{\sigma}^4\varpi^2)^{-1}\left\{\hat{\sigma}^4(p+q) - \hat{\sigma}^2\mathrm{tr}(\hat{J}^{-1}\hat{I}) + \mathrm{tr}\{(\hat{J}^{-1}\hat{I})^2\}\right\}$$

and therefore, denoting by $\mathbf{I}_{p+q}$ the $(p+q)\times(p+q)$ identity matrix, we obtain

$$\Upsilon_n = \frac{nb_n}{2\varpi^2}\ \mathrm{tr}\left(\mathbf{I}_{p+q} - \frac{1}{\hat{\sigma}^2}\hat{J}^{-1}\hat{I}\right)^2. \tag{18}$$

Note that when the ARMA is strong, $\frac{1}{\hat{\sigma}^2}\hat{J}^{-1}\hat{I}$ converges to the identity matrix.

The next result, which is a straightforward consequence of Theorem 3 and (17), provides a critical region of asymptotic level $\alpha \in (0,1)$.

Theorem 7 *Let the assumptions of Theorem 2, in particular* $\mathbf{H_0}$*, hold. Then*

$$\lim_{n\to\infty} P\left\{\Upsilon_n > \chi^2_{\frac{1}{2}(p+q)(p+q+1),1-\alpha}\right\} = \alpha,$$

where $\chi^2_{r,\alpha}$ *denotes the* α*-quantile of the* χ^2_r *distribution.*

The following theorem gives conditions for the consistency of our test.

Theorem 8 *Assume that* $\mathbf{A1}-\mathbf{A3}$ *and* $\mathbf{A5}$ *(or* $\mathbf{A5'}$*) hold and that* $E|X_t|^{4+2\nu} < \infty$*. Let* $\omega(\cdot)\in\mathcal{K}$*, and* (b_n) *satisfying* $\lim_{n\to\infty} b_n = 0$ *and* $\lim_{n\to\infty} nb_n^{4+10/\nu} = +\infty$*.*

Then, under $\mathbf{H_1}$ *we have*

$$\lim_{n\to\infty} P\left\{\Upsilon_n > \chi^2_{\frac{1}{2}(p+q)(p+q+1),1-\alpha}\right\} = 1,$$

for any $\alpha \in (0,1)$*.*

The test procedure is quite simple. For a given time series it consists in : (i) fitting an ARMA(p,q) model (after an identification step of the orders p and q, which is not the subject of this paper); (ii) estimating the matrices I and J by (7) and (6); (iii) computing the statistic Υ_n by (18) and rejecting $\mathbf{H_0}$ when $\Upsilon_n > \chi^2_{\frac{1}{2}(p+q)(p+q+1),1-\alpha}$. Choice of the bandwidth and kernel used to define the estimator of matrix I will be discussed in the next section.

Remarks

(a) To our knowledge, this is the first test designed for the purpose of distinguishing between weak and strong ARMA models. In other words, this statistic allows to test whether the error term of an ARMA model is independent or simply non correlated.

(b) Our test is related to other tests recently introduced in the time series literature. Some of them are goodness-of-fit tests (which is not the case of ours, since under both $\mathbf{H_0}$ and $\mathbf{H_1}$ the ARMA model (2) is well-specified). Let us mention Hong (1996), who proposes a test, based on a kernel-based spectral density estimator, for uncorrelatedness of the residuals of AR models with exogeneous variables. Its asymptotic distribution is established under the null assumption of independent errors and his consistent for serial correlation of unknown form. In the framework of ARMA models, Francq, Roy and Zakoian (2004) propose a modification of the standard portmanteau test for serial correlation, when the errors are only assumed to be uncorrelated (not independent) under the null hypothesis. The test of the present paper can be viewed as complementary to those goodness-of-fit tests.

Another approach very closely related to ours is taken by Hong (1999) who proposes a nonparametric test for serial independence, based on a generalization of the spectral density. Hong's test has power against various types of pairwise dependencies, including cases of absence of correlation, and is aimed to detect any departure from independence. Our test has a more limited scope since it is devoted to ARMA models. Moreover it is aimed to test one consequence of independence, the one that matters for the reliability of the standard routines. Another difference between the two approaches, is that Hong's test is aimed to detect departure from independence of the observed process. It cannot be straightforwardly applied to our framework because, even under the null hypothesis of independence, the residuals are dependent.

6 Choice of the bandwidth and kernel, and finite sample performances of the test

To make the test procedure fully operational, it is necessary to specify how to chose the kernel and bandwidth parameters. To this aim we will first consider two standard asymptotic local efficiency criteria, respectively derived from the Bahadur and Pitman approaches. The reader is referred to Van der Vaart (1998) for details concerning the asymptotic local efficiency of tests.

6.1 Bahadur's approach

In view of (18), under the assumptions of Theorem 8,

$$\frac{1}{nb_n}\Upsilon_n \to \frac{1}{2\varpi^2}\ \mathrm{tr}\left(\mathbf{I}_{p+q} - \frac{1}{\sigma^2}J^{-1}I\right)^2 \tag{19}$$

in probability as $n \to \infty$. Let $\Upsilon_n^{(1)}$ and $\Upsilon_n^{(2)}$ be two test statistics of the form Υ_n, with respective kernels ω_1 and ω_2. The p-value of the tests can be approximated by $1-F_{\chi^2_{\frac{1}{2}(p+q)(p+q+1)}}(\Upsilon_n^{(i)})$, where $F_{\chi^2_k}$ denotes the cumulative distribution function of the χ^2_k distribution. Assume we are under the alternative. Let $\{n_1(n)\}_n$ and $\{n_2(n)\}_n$ be two sequences of integers tending to infinity such that

$$\lim_{n\to\infty}\frac{\log\left\{1 - F_{\chi^2_{\frac{1}{2}(p+q)(p+q+1)}}(\Upsilon^{(1)}_{n_1(n)})\right\}}{\log\left\{1 - F_{\chi^2_{\frac{1}{2}(p+q)(p+q+1)}}(\Upsilon^{(2)}_{n_2(n)})\right\}} = 1 \quad a.s.$$

One can say that, for large n, the two tests require respectively $n_1(n)$ and $n_2(n)$ observations to reach the same (log) p-value. The Bahadur asymptotic relative efficiency (ARE) of $\Upsilon_n^{(1)}$ with respect to $\Upsilon_n^{(2)}$ is defined by $ARE(\Upsilon_n^{(1)}, \Upsilon_n^{(2)}) =$

$\lim_{n\to\infty} n_2(n)/n_1(n)$. To make the comparison meaningful we use the same bandwidth for the two tests. Assume that $b_n = cn^{-\nu}$, $c > 0$, $\nu \in (0,1)$. Using $\log\left\{1 - F_{\chi^2_k}(x)\right\} \sim -x/2$ as $x \to \infty$, we obtain $ARE(\Upsilon_n^{(1)}, \Upsilon_n^{(2)}) = \left(\varpi_2^2/\varpi_1^2\right)^{1/(1-\nu)}$, where $\varpi_i^2 = \int \omega_i^2(x)dx$. Thus, one can consider that the asymptotic superiority of the first test over the second test hold if $\varpi_1^2 < \varpi_2^2$. In this sense, it is easy to see that the tests based on the truncated, Bartlett, Tukey-Hanning and Parzen kernels are ranked in an increasing order of asymptotic efficiency. A similar argument shows that, when the kernel is fixed and ν varies, the Bahadur efficiency decreases when ν increases. Thus, in the Bahadur sense, there is no optimal choice of b_n: the slower b_n tends to zero, the asymptotically more efficient the tests are. Unfortunately, this result does not give indication on how to choose the bandwidth parameter for finite samples. If b_n tends to zero too slowly and/or if ϖ_i^2 is too small, the finite sample bias of $\hat{I}$ is likely to be important, and the rate of convergence in (19) is likely to be very slow.

6.2 Pitman's approach

Another popular approach used to compare the asymptotic local powers of tests is that of Pitman. Consider local alternatives of the form $H_{1n} : I = \sigma^2 J + \Delta/\sqrt{nb_n}$, where $\Delta \neq 0$ is a symmetric positive definite matrix. Alternatively, one could formulate these alternatives as $H_{1n} : \Sigma^{(2)} = \Sigma^{(1)} + J^{-1}\Delta J^{-1}/\sqrt{nb_n}$. Under standard assumptions, $\hat{\Sigma}^{(2)} - \hat{\Sigma}^{(1)}$ is a regular estimator of $\Sigma^{(2)} - \Sigma^{(1)}$ (see Van der Vaart (1998), Section 8.5). Therefore, in view of Theorem 5, under H_{1n},

$$\sqrt{nb_n}\ \mathrm{vech}\left(\hat{\Sigma}^{(2)} - \hat{\Sigma}^{(1)}\right) \overset{d}{\rightsquigarrow} \mathcal{N}\left(\mathrm{vech} J^{-1}\Delta J^{-1}, \Lambda\right).$$

It follows that, under H_{1n},

$$\Upsilon_n \overset{d}{\rightsquigarrow} \chi^2_{\frac{1}{2}(p+q)(p+q+1)}\left\{\left(\mathrm{vech} J^{-1}\Delta J^{-1}\right)' \Lambda^{-1}\left(\mathrm{vech} J^{-1}\Delta J^{-1}\right)\right\},$$

where $\chi^2_k(\delta)$ denotes the noncentral χ^2_k distribution with noncentrality parameter δ. It follows that the Pitman asymptotic local power increases with the noncentrality parameter. In view of (10), we draw the same conclusion as for the Bahadur approach: tests with small ϖ_i^2 are preferred. Amazingly, the asymptotic distribution of Υ_n does not depend on the asymptotic behaviour of b_n. However the slower b_n tends to zero, the faster H_{1n} tends to H_0. It is therefore preferable to chose b_n as large as possible. This was also our conclusion with the previous approach.

6.3 Automatic bandwidth estimators

Since the approaches of the previous sections do not allow to chose b_n in practice, we now turn to an automatic bandwidth method. Andrews (1991) obtained asymptotically optimal data-dependent automatic bandwidth parameters for HAC estimators. We will apply his results to our framework. It should be noted, however, that optimal HAC estimators do not necessarily provide asymptotically optimal test statistics. This issue is left for further investigation.

Andrews showed that, under the assumption $\|I^{(r)}\| > 0$ and other regularity assumptions, the asymptotically optimal bandwidth parameter (leading to an estimator with the best bias-variance trade-off) is given by

$$b_n^* = c^{-1}\left\{\alpha(r)n\right\}^{-1/(2r+1)}, \quad \alpha(r) = \frac{\sum_{i=1}^{p+q}\left(e_i' I^{(r)} e_i\right)^2}{\sum_{i=1}^{p+q}\left(e_i' I e_i\right)^2},$$

where (c, r) is equal to $(1.1447, 1)$ for the Bartlett kernel, to $(2.6614, 2)$ for the Parzen kernel, and to $(1.7462, 2)$ for the Tukey-Hanning kernel, and where e_i denotes the i-th vector of the canonical basis of $\mathbb{R}^{p+q}$. Approximating, for $j = 1, \ldots, p+q$, the dynamics of $\epsilon_t \partial\epsilon_t(\theta_0)/\partial\theta_j$ by a simple AR(1) model with autoregressive parameter $\hat{a}_j$ and variance parameter $\hat{\sigma}_j^2$, Andrews obtained the data-dependent estimate $\hat{b}_n^* = c^{-1}\left\{\hat{\alpha}(r)n\right\}^{-1/(2r+1)}$ of b_n^*, by setting

$$\hat{\alpha}(1) = \frac{\sum_{i=1}^{p+q} 4\hat{a}_j^2 \hat{\sigma}_j^4 (1-\hat{a}_j)^{-6}(1+\hat{a}_j)^{-2}}{\sum_{i=1}^{p+q} \hat{\sigma}_j^4 (1-\hat{a}_j)^{-4}}, \quad \hat{\alpha}(2) = \frac{\sum_{i=1}^{p+q} 4\hat{a}_j^2 \hat{\sigma}_j^4 (1-\hat{a}_j)^{-8}}{\sum_{i=1}^{p+q} \hat{\sigma}_j^4 (1-\hat{a}_j)^{-4}}.$$

6.4 Finite sample performance

To assess the finite sample performances of the tests proposed in this paper, we first simulate 1,000 replications of several strong ARMA models of size $n = 200, n = 400$ and $n = 800$. We consider tests with nominal level $\alpha = 5\%$. We use the estimated optimal bandwidth given by Andrews, as described previously, and 3 different kernels. The relative rejection frequencies are given in Table 1. All the empirical sizes are less than the nominal 5% level. It seems that the tests are slightly conservative, and that, in terms of control of type I error, the performance of the three kernels is very similar.

Now we turn to experiments under the alternative of non independent errors. Five models were considered: i) an AR(1) with GARCH(1,1) errors; ii) the square of a GARCH(1,1); iii) an ARMA(1,1) with a Markov-Switching white noise; iv) the first component of a strong bivariate MA(1); v) an AR(1) with a chaotic noise. Precise specifications are displayed in Table 2. All these examples have been shown to provide ARMA models with non independent errors: an ARMA(1,1) for models ii) and iii), an AR(1) for models i) and v), and a MA(1) for model iv). It should be emphasized that we only need to estimate the ARMA representation, not the DGP.

Andrews (1991) showed that, for HAC estimation, the Bartlett kernel is less efficient than the two other kernels. In these examples, the power of the test does not appear to be very sensitive to the kernel choice. There is no particular user-chosen kernel that is the most satisfactory for all cases. For each sample size, the best performance is obtained for the Markov-switching model. For this model, the tests almost always take the right decision, at least when $n \geq 500$. Slower convergences to 1 are obtained for the powers in models i), iv) and v). The very slow power convergence in model ii) can be explained as follows. The weak ARMA$(1,1)$ representation of ϵ_t^2 is $\epsilon_t^2 - 0.97\epsilon_{t-1}^2 = 1 + u_t - 0.85u_{t-1}$ for some white noise u_t. It is seen that the AR and MA parts have close roots, making statistical inference difficult. As continuous functions of the ARMA estimators, the estimators of I and J inherit their poor accuracy. Another explanation, which also holds for the relatively poor performance of model i), is that, in models based on GARCH errors, the noise is a martingale difference. For this reason, departure from the strong assumption can be more difficult to detect than in cases when the noise is only uncorrelated (as in models iii)-v)).

Table 1: Size (in % of relative rejection frequencies) of Υ_n-tests with estimated optimal bandwidth. The nominal significance level is $\alpha = 5\%$. The number of replications is $N = 1,000$.

		Kernel		
Model	n	Bartlett	Parzen	Tukey-Hanning
	200	2.3	1.5	2.2
Strong AR(1)[1]	500	2.0	3.8	2.3
	800	2.9	2.1	2.6
	200	1.3	3.1	1.3
Strong MA(1)[2]	500	1.9	2.7	2.1
	800	2.2	2.5	3.0
	200	1.9	3.0	3.3
Strong ARMA(1,1)[3]	500	3.2	4.8	3.8
	800	3.6	4.4	4.0

1: $X_t = 0.5X_{t-1} + \epsilon_t$, ϵ_t iid Student with $\nu = 5$ degrees of freedom
2: $X_t = \epsilon_t + 0.7\epsilon_{t-1}$, ϵ_t iid with centered exponential density
3: $X_t - 0.5X_{t-1} = \epsilon_t + 0.7\epsilon_{t-1}$, ϵ_t iid $\mathcal{N}(0,1)$

Table 2: Power (in % of relative rejection frequencies) of Υ_n-tests with estimated optimal bandwidth. The nominal significance level is $\alpha = 5\%$. The number of replications is $N = 1,000$.

		Kernel		
Model	n	Bartlett	Parzen	Tukey-Hanning
AR(1)-GARCH(1,1)[4]	200	14.1	18.9	19.7
	500	53.5	50.0	51.0
	800	74.6	72.2	70.4
Square of a GARCH(1,1)[5]	200	9.4	9.4	9.7
	500	23.6	24.9	26.8
	800	38.8	39.0	36.1
MS-ARMA(1,1)[6]	200	75.8	81.1	79.8
	500	98.4	98.4	98.3
	800	99.9	99.9	99.9
MA(1) marginal [7]	200	32.6	30.5	36.2
	500	70.4	73.1	78.8
	800	86.0	89.4	94.1
AR(1) [8]	200	22.5	25.8	21.0
	500	42.1	46.2	48.9
	800	53.9	63.3	62.2

4: $X_t = 0.5X_{t-1} + \epsilon_t, \quad \epsilon_t = \sqrt{h_t}\eta_t,\ h_t = 1 + 0.12\epsilon_{t-1}^2 + 0.85h_{t-1},\ \eta_t$ iid $\mathcal{N}(0,1)$

5: $X_t = \epsilon_t^2$, where ϵ_t is as in 4

6: $X_t - 0.5X_{t-1} = \epsilon_t + 0.7\epsilon_{t-1}, \epsilon_t = \eta_t + (1-2\Delta_t)\eta_{t-1}$, (Δ_t) is a Markov Chain with state-space $\{0,1\}$ and transition probabilities $P(\Delta_t = 1|\Delta_{t-1} = 0) = P(\Delta_t = 0|\Delta_{t-1} = 1) = 0.01$, η_t iid $\mathcal{N}(0,1)$

7: $X_{1\,t} = \epsilon_{1\,t} + 0.8\epsilon_{1\,t-1} - 0.9\epsilon_{2\,t-1}$, where $\epsilon_{1\,t} = \eta_{1\,t}^2 - 1$, $\epsilon_{2\,t} = \eta_{2\,t}^2 - 1$, and $\begin{pmatrix} \eta_{1\,t} \\ \eta_{2\,t} \end{pmatrix} \sim \mathcal{N}\left\{0, \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}\right\}$

8: $X_t = 0.5X_{t-1} + \epsilon_t$, where (ϵ_t) is the noise defined by (1) with $\eta_t \sim \mathcal{N}(-0.5, 0.05^2)$

7 Proofs

7.1 Additional notations and scheme of proof for Theorem 2

Throughout this section, the letter K (resp. ρ) will be used to denote positive constants (resp. constants in $(0,1)$) whose values are unimportant and may vary. We will use the norm $\|A\| = \sum |a_{ij}|$ for any matrix $A = (a_{ij})$. Let

$$\mathbf{u}_t(\theta) = \epsilon_t(\theta)\frac{\partial}{\partial\theta}\epsilon_t(\theta), \qquad \mathbf{u}_t = \mathbf{u}_t(\theta_0),$$

$$\mathbf{v}_t(\theta) = e_t(\theta)\frac{\partial}{\partial\theta}e_t(\theta), \qquad \mathbf{v}_t = \mathbf{v}_t(\theta_0).$$

Hence, for $0 \leq i < n$, and for $i \geq 0$, respectively,

$$\hat{\Delta}_i(\theta) = \hat{\Delta}'_{-i}(\theta) = \frac{1}{n}\sum_{t=1}^{n-i}\mathbf{v}_t(\theta)\mathbf{v}'_{t+i}(\theta), \qquad \Delta_i(\theta) = \Delta'_{-i}(\theta) = E\{\mathbf{u}_t(\theta)\mathbf{u}'_{t+i}(\theta)\}.$$

We set, for $0 \leq i < n$,

$$\hat{\Delta}_i^{\mathbf{u}}(\theta) = \left\{\hat{\Delta}_{-i}^{\mathbf{u}}(\theta)\right\}' = \frac{1}{n}\sum_{t=1}^{n-i}\mathbf{u}_t(\theta)\mathbf{u}'_{t+i}(\theta), \quad \hat{I}_n^{\mathbf{u}}(\theta) = \sum_{i=-[a/b_n]}^{[a/b_n]}\omega(ib_n)\hat{\Delta}_i^{\mathbf{u}}(\theta),$$

assuming without loss of generality that $a/b_n \leq n$. Similarly, we can write $\hat{\Delta}_i(\theta) = \hat{\Delta}_i^{\mathbf{v}}(\theta)$ and $\hat{I}_n(\theta) = \hat{I}_n^{\mathbf{v}}(\theta)$. In the notation of all those quantities the parameter θ will be removed when equal to θ_0.

It will also be convenient to modify the number of terms taken into account in the definition of $\hat{\Delta}_i^{\mathbf{u}}(\theta_0) = \hat{\Delta}_i^{\mathbf{u}}$. Define, for $i \in \mathbb{N} = \{0, 1, \dots\}$,

$$\Delta_i^{\mathbf{u}} = (\Delta_{-i}^{\mathbf{u}})' = \frac{1}{n}\sum_{t=1}^{n}\mathbf{u}_t\mathbf{u}'_{t+i}, \qquad I_n^{\mathbf{u}} = \sum_{i=-[a/b_n]}^{[a/b_n]}\omega(ib_n)\Delta_i^{\mathbf{u}},$$

and the centered matrix

$$\overline{I}_n^{\mathbf{u}} = I_n^{\mathbf{u}} - \Delta_0^{\mathbf{u}} = \sum_{0<|i|\leq[a/b_n]} \omega(ib_n)\Delta_i^{\mathbf{u}} = \sum_{0<i\leq[a/b_n]} \omega(ib_n)\left\{\Delta_i^{\mathbf{u}} + (\Delta_i^{\mathbf{u}})'\right\}.$$

It will be shown that $\sqrt{nb_n}\left(\hat{I} - I\right) = \sqrt{nb_n}\left\{\hat{I}_n^{\mathbf{v}}(\hat{\theta}_n) - I\right\}$ and $\sqrt{nb_n}\quad \overline{I}_n^{\mathbf{u}}$ have the same asymptotic distribution (see Lemma 9 below).

Define the sequences $(c_{k,\ell})_k$ by $\frac{\partial}{\partial\theta_\ell}\epsilon_t = \sum_{k=1}^{\infty} c_{k,\ell}\epsilon_{t-k}$. Since a central limit theorem for r-dependent sequences will be used, it will be useful to consider the following truncated variables. For any positive integer r, let

$$_r\!\left(\frac{\partial}{\partial\theta_\ell}\epsilon_t\right) = \sum_{k=1}^{r} c_{k,\ell}\epsilon_{t-k}, \quad {}_r\mathbf{u}_t = \epsilon_t\, {}_r\!\left(\frac{\partial}{\partial\theta}\epsilon_t\right), \quad {}^r\mathbf{u}_t = \mathbf{u}_t - {}_r\mathbf{u}_t, \qquad (20)$$

and for $m < k$

$${}_r\Delta_{i,(m,k)}^{\mathbf{u}} = \left\{ {}_r\Delta_{-i,(m,k)}^{\mathbf{u}}\right\}' = \frac{1}{k-m+1}\sum_{t=m}^{k} {}_r\mathbf{u}_t\; {}_r\mathbf{u}_{t+i}' \quad (i \in \mathbb{N}),$$

$${}_r\overline{I}_{n,(m,k)}^{\mathbf{u}} = \sum_{0<|i|\leq[a/b_n]} \omega(ib_n)\,{}_r\Delta_{i,(m,k)}^{\mathbf{u}}.$$

When $m = 1$ and $k = n$ we will write ${}_r\Delta_{i,(m,k)}^{\mathbf{u}} = {}_r\Delta_i^{\mathbf{u}}$ and ${}_r\overline{I}_{n,(m,k)}^{\mathbf{u}} = {}_r\overline{I}_n^{\mathbf{u}}$. By a standard argument, we will obtain the limit distribution of $\sqrt{nb_n}\quad \overline{I}_n^{\mathbf{u}}$ by taking the limit as $r \to \infty$ of the asymptotic distribution, as $n \to \infty$, of $\sqrt{nb_n}\quad {}_r\overline{I}_n^{\mathbf{u}}$. We also denote

$${}_rJ = E\, {}_r\!\left(\frac{\partial}{\partial\theta}\epsilon_t\right) {}_r\!\left(\frac{\partial}{\partial\theta'}\epsilon_t\right).$$

Now note that ${}_r\overline{I}_n^{\mathbf{u}} = \sum_{t=1}^{n} Z_{n,t}$, where

$$Z_{n,t} = \frac{1}{n}\sum_{0<i\leq[a/b_n]} \omega(ib_n)\left({}_r\mathbf{u}_t\; {}_r\mathbf{u}_{t+i}' + {}_r\mathbf{u}_{t+i}\; {}_r\mathbf{u}_t'\right).$$

Process $(Z_{n,t})_t$ is m_n-dependent, with $m_n = [ab_n^{-1}]+r$. Next we split the sum $\sum_{t=1}^n Z_{n,t}$ into alternate blocks of length $k_n - m_n$ and m_n (with a remaining block of size $n - p_n k_n + m_n$):

$$\begin{aligned}
{}_r\overline{I}_n^{\mathbf{u}} &= {}_rS_n + \left(\sum_{\ell=0}^{p_n-2} Z_{n,(\ell+1)k_n-m_n+1} + \cdots + Z_{n,(\ell+1)k_n} \right) \\
&\quad + Z_{n,p_nk_n-m_n+1} + \cdots + Z_{n,n}, \qquad (21) \\
{}_rS_n &= \sum_{\ell=0}^{p_n-1} Z_{n,\ell k_n+1} + \cdots + Z_{n,(\ell+1)k_n-m_n} \\
&= \frac{k_n - m_n}{n} \sum_{\ell=0}^{p_n-1} {}_r\overline{I}^{\mathbf{u}}_{n,(\ell k_n+1,(\ell+1)k_n-m_n)},
\end{aligned}$$

where k_n is an integer in (m_n, n) to be specified later and $p_n = [n/k_n]$, the integer part of n/k_n (assuming that n is sufficiently large, so that $m_n \leq n$). It will be shown that, when $m_n = o(k_n - m_n)$ and $p_n \to \infty$, the asymptotic distributions of $\sqrt{nb_n}\ {}_r\overline{I}_n^{\mathbf{u}}$ and $\sqrt{nb_n}\ {}_rS_n$ are identical (see part (b) of Lemma 13).

To avoid moment assumptions of excessive order, we then introduce variables that are truncated in level. For any positive constant κ and for $m < k$, let

$$\epsilon_t^\kappa = \epsilon_t 1_{\{|\epsilon_t| \leq \kappa\}} - E\epsilon_t 1_{\{|\epsilon_t| \leq \kappa\}}, \qquad {}_r\left(\frac{\partial}{\partial \theta_\ell} \epsilon_t \right)^\kappa = \sum_{k=1}^r c_{k,\ell} \epsilon_{t-k}^\kappa,$$

$${}_r\mathbf{u}_t^\kappa = \epsilon_t^\kappa\ {}_r\left(\frac{\partial}{\partial \theta} \epsilon_t \right)^\kappa, \qquad {}^\kappa_r\mathbf{u}_t = {}_r\mathbf{u}_t - {}_r\mathbf{u}_t^\kappa,$$

let ${}_r\Delta^{\mathbf{u},\kappa}_{i,(m,k)}$ (resp. ${}_r\overline{I}^{\mathbf{u},\kappa}_{n,(m,k)}$) be the matrix obtained by replacing the variables ${}_r\mathbf{u}_t\ {}_r\mathbf{u}'_{t+i}$ by ${}_r\mathbf{u}_t^\kappa\ {}_r\mathbf{u}_{t+i}^{\kappa\prime}$ in ${}_r\Delta^{\mathbf{u}}_{i,(m,k)}$ (resp. ${}_r\overline{I}^{\mathbf{u}}_{n,(m,k)}$) and let

$${}_rS_n{}^\kappa = \frac{k_n - m_n}{n} \sum_{\ell=0}^{p_n-1} {}_r\overline{I}^{\mathbf{u},\kappa}_{n,(\ell k_n+1,\ell k_n+k_n-m_n)}.$$

We will show that, when $\kappa_n \to \infty$, the asymptotic distributions of $\sqrt{nb_n}\ {}_rS_n$ and $\sqrt{nb_n}\ {}_rS_n^{\kappa_n}$ are identical (see part (c) of Lemma 13). The Lindeberg central limit theorem will be used to show the asymptotic normality of $\sqrt{nb_n}\ {}_rS_n^{\kappa_n}$.

7.2 Lemmas and proofs for Theorem 2

In all subsequent lemmas, the assumptions of Theorem 2 are supposed to be satisfied. The first four lemmas are concerned with some fourth-order properties of the process $(\mathbf{u}_t)$.

Lemma 2 *Let $k \in \mathbb{N}^* = \{1, 2, \dots\}$, $t_1, t_2, \dots, t_k \in \mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ and $i_1, i_2, \dots, i_k \in \mathbb{N}^*$. If the indexes $t_1, t_2, \dots, t_k$ are all distinct and the indices $i_1, i_2, \dots, i_k$ are all less than or equal to 4, then*

$$E \prod_{j=1}^{k} \left|\epsilon_{t_j}\right|^{i_j} \leq M_k < \infty.$$

Proof. Arguing by induction, it suffices to note that $E|X\epsilon_t^i| = E|X|E|\epsilon_t^i| \leq E|X| \max_{j\in\{1,\dots,4\}} E|\epsilon_t|^j < \infty$, when X and ϵ_t are independent, $E|X| < \infty$, $i \in \{1, \dots, 4\}$ and $E\epsilon_t^4 < \infty$. □

Lemma 3 *For all $i, j, h \in \mathbb{Z}$, the matrix $Cov\{\mathbf{u}_1 \otimes \mathbf{u}_{1+i}, \mathbf{u}_{1+h} \otimes \mathbf{u}_{1+h+j}\}$ is well defined in $\mathbb{R}^{(p+q)^2} \times \mathbb{R}^{(p+q)^2}$ and is bounded in norm by a constant independent of i, j and h.*

Proof. For $\ell \in \{1, \dots, p+q\}$, recall that $u_t(\ell) = \epsilon_t \frac{\partial}{\partial\theta_\ell}\epsilon_t(\theta_0)$ is the ℓ-th element of $\mathbf{u}_t$. Assumption **A2** entails that $\frac{\partial}{\partial\theta_\ell}\epsilon_t(\theta_0) = \sum_{k\geq1} c_{k,\ell}\epsilon_{t-k}$, where

$|c_{k,\ell}| < K\rho^k$. Therefore we have, for all $\ell_1, \ell_2, \ell_3, \ell_4 \in \{1, \dots, p+q\}$,

$$E\left|u_1(\ell_1)u_{1+i}(\ell_2)u_{1+h}(\ell_3)u_{1+h+j}(\ell_4)\right| \leq K \sum_{i_1,i_2,i_3,i_4 \geq 1} \rho^{i_1+i_2+i_3+i_4} E \prod_{j=1}^{8} |\epsilon_{t_j}| \quad (22)$$

with $t_1 := 1 \neq t_2 := 1 - i_1$, $t_3 := 1 + i \neq t_4 := 1 + i - i_2$, $t_5 := 1 + h \neq t_6 := 1 + h - i_3$ and $t_7 := 1 + h + j \neq t_8 := 1 + h + j - i_4$. Since, at most four of the indices $t_1, \dots, t_8$ are equal, Lemma 2 shows that the right-hand side of (22) is bounded. The proof follows. □

Lemma 4 *For* $i, h \in \mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$,

$$Cov\{\mathbf{u}_1 \otimes \mathbf{u}_{1+i}, \mathbf{u}_{1+h} \otimes \mathbf{u}_{1+h+i}\} = \begin{cases} 0 & \text{when} \quad ih \neq 0, \\ O(\rho^{|h|}) & \text{when} \quad i = 0, \\ \sigma^4 J \otimes J + O(\rho^{|i|}) & \text{when} \quad h = 0. \end{cases}$$

Proof. We keep the notations of the proof of Lemma 3. First note that, for $\ell_1, \ell_2, \ell_3, \ell_4 \in \{1, \dots, p+q\}$, the $(p+q)(\ell_1-1)+\ell_2$-th row and $(p+q)(\ell_3-1)+\ell_4$-th column element of the previous covariance matrix (resp. of $J \otimes J$) is

$$\text{Cov}\left\{u_1(\ell_1)u_{1+i}(\ell_2), u_{1+h}(\ell_3)u_{1+h+i}(\ell_4)\right\} \quad (23)$$

(resp. $J(\ell_1, \ell_3)J(\ell_2, \ell_4)$).

The case $ih \neq 0$ is obvious by noting that, in (23), one of the u's has an index that is strictly greater than the other indexes.

To deal with the case $i = 0$, without loss of generality suppose that $h > 0$. The covariance in (23) is given by

$$\sigma^2 \sum_{i_1,i_2,i_3,i_4 \geq 1} c_{i_1,\ell_1} c_{i_2,\ell_2} c_{i_3,\ell_3} c_{i_4,\ell_4} \text{Cov}\left\{\epsilon_1^2 \epsilon_{1-i_1} \epsilon_{1-i_2}, \epsilon_{1+h-i_3} \epsilon_{1+h-i_4}\right\}.$$

When $i_4 < h$, $\text{Cov}\{\epsilon_1^2\epsilon_{1-i_1}\epsilon_{1-i_2}, \epsilon_{1+h-i_3}\epsilon_{1+h-i_4}\} = 0$ because when $i_3 < h$, $\epsilon_1^2\epsilon_{1-i_1}\epsilon_{1-i_2}$ and $\epsilon_{1+h-i_3}\epsilon_{1+h-i_4}$ are independent, and when $i_3 \geq h$ this covariance is given by $E\epsilon_1^2\epsilon_{1-i_1}\epsilon_{1-i_2}\epsilon_{1+h-i_3}E\epsilon_{1+h-i_4} = 0$. Therefore

$$|\text{Cov}\{u_1(\ell_1)u_1(\ell_2), u_{1+h}(\ell_3)u_{1+h}(\ell_4)\}| \leq K \sum_{i_4 \geq h} c_{i_4,\ell_4} \leq K\rho^h.$$

Now we consider the case $h = 0$. For $i \neq 0$, (23) is given by

$$\begin{aligned} Eu_1(\ell_1)u_1(\ell_3)u_{1+i}(\ell_2)u_{1+i}(\ell_4) &= Eu_1(\ell_1)u_1(\ell_3)Eu_{1+i}(\ell_2)u_{1+i}(\ell_4) + \nabla \\ &= \sigma^4 J(\ell_1,\ell_3)J(\ell_2,\ell_4) + \nabla, \end{aligned}$$

where $\nabla = \text{Cov}\{u_1(\ell_1)u_1(\ell_3), u_{1+i}(\ell_2)u_{1+i}(\ell_4)\} = O(\rho^{|i|})$, as already proven. $\square$

Lemma 5 *For* $|i| \neq |j|$,

$$\sum_{h=-\infty}^{+\infty} \| Cov\{\mathbf{u}_1 \otimes \mathbf{u}_{1+i}, \mathbf{u}_{1+h} \otimes \mathbf{u}_{1+h+j}\}\| \leq K\rho^{|i|}\rho^{|j|}. \qquad (24)$$

Proof. Without loss of generality, assume that $i \geq 0$ and $j \geq 0$ $(i \neq j)$. For $i > 0$ and $j > 0$, all norms in (24) vanish, except perhaps one which is the sum (over the ℓ_i's) of

$$\begin{aligned} &|\text{Cov}\{u_1(\ell_1)u_{1+i}(\ell_2), u_{1-j+i}(\ell_3)u_{1+i}(\ell_4)\}| \\ \leq & \sum_{i_1,i_2,i_3,i_4 \geq 1} |c_{i_1,\ell_1}c_{i_2,\ell_2}c_{i_3,\ell_3}c_{i_4,\ell_4}| \left| E\epsilon_{1-i_1}\epsilon_1\epsilon_{1+i-i_2}\epsilon_{1+i}^2\epsilon_{1-j+i-i_3}\epsilon_{1-j+i}\epsilon_{1+i-i_4} \right| \\ \leq & \; K \sum_{i_1,i_2,i_3,i_4 \geq 1} \rho^{i_1+i_2+i_3+i_4} \left| E\epsilon_{1-i_1}\epsilon_1\epsilon_{1+i-i_2}\epsilon_{1-j+i-i_3}\epsilon_{1-j+i}\epsilon_{1+i-i_4} \right|. \end{aligned}$$

Let $\mathcal{S}_i = \{(i_1,i_2,i_3,i_4) | \max(i_1,i_2,i_3,i_4) \geq i\}$. Clearly,

$$\sum_{\mathcal{S}_i} \rho^{i_1+i_2+i_3+i_4} \left| E\epsilon_{1-i_1}\epsilon_1\epsilon_{1+i-i_2}\epsilon_{1-j+i-i_3}\epsilon_{1-j+i}\epsilon_{1+i-i_4} \right| \leq K\rho^i.$$

Now for (i_1, i_2, i_3, i_4) not belonging to $\mathcal{S}_i$, the indices of the ϵ's in the expectation can be ranked as follows:

$$1 - i_1 < 1 < \min(1 + i - i_2, 1 + i - i_4) \quad \text{and} \quad 1 - j + i - i_3 < 1 - j + i. \tag{25}$$

It is therefore clear that, at least one of the ϵ's in the previous expectations has an index different from the others, making these expectations equal to 0. We conclude that the left-hand side of (24) is bounded by $K\rho^i$ uniformly in $j > 0$, and, by symmetry, by $K\rho^j$ uniformly in $i > 0$.

For $i = 0$ and $j > 0$, the left-hand side of (24) reduces to a sum (over the ℓ_i's) of

$$\begin{aligned} &\sum_{h=-\infty}^{-j} \left| \mathrm{Cov}\left\{u_1(\ell_1)u_1(\ell_2), u_{1+h}(\ell_3)u_{1+h+j}(\ell_4)\right\}\right| \\ \leq\ & K \sum_{h=-\infty}^{-j} \sum_{i_1,i_2,i_3,i_4 \geq 1} \rho^{i_1+i_2+i_3+i_4} \left| E\epsilon_{1-i_1}\epsilon_1^2\epsilon_{1-i_2}\epsilon_{1+h-i_3}\epsilon_{1+h}\epsilon_{1+h+j-i_4}\epsilon_{1+h+j}\right|. \end{aligned}$$

In this sum all terms vanish, except when $i_1 = -h$ or $i_2 = -h$ or $i_4 = j$ (in which case at least two indices are equal to $1+h$) and when $i_1 = -h-j$ or $h = -j$ or $i_2 = -h - j$ (in which case at least two indices are equal to $1 + h + j$). Therefore, it can be seen that the left-hand side of (24) is also bounded by $K\rho^j$ when $i = 0$. The case $j = 0$ is handled in the same way, by symmetry. The conclusion follows. □

Lemma 6

$$\lim_{n\to\infty} nb_n\, Var\{vech\ I_n^{\mathbf{u}}\} = 2\sigma^4\varpi^2 D_{p+q}^{+}(J \otimes J)D_{p+q}^{+'}.$$

Proof. By stationarity of the processes $(\mathbf{u}_t)$, and by the elementary relations $\mathrm{vech}(A) = D_{p+q}^{+}\mathrm{vec}(A)$, for any $p + q \times p + q$ symmetric matrix A, and

$\text{vec}(\mathbf{u}_i\mathbf{u}_j') = \mathbf{u}_j \otimes \mathbf{u}_i$, we have

$$nb_n \text{Var}\left\{\text{vech } I_n^{\mathbf{u}}\right\} = \sum_{k=1}^{5} \sum_{(i,j,h)\in\mathcal{I}_k} \mathbf{A}_n(i,j,h)$$

where

$$\mathbf{A}_n(i,j,h) = \frac{b_n}{n}\omega(ib_n)\omega(jb_n)(n-|h|)D_{p+q}^{+}\text{Cov}\left\{\mathbf{u}_1 \otimes \mathbf{u}_{1+i}, \mathbf{u}_{1+h} \otimes \mathbf{u}_{1+h+j}\right\} D_{p+q}^{+'}$$

and the $\mathcal{I}_k$'s are subsets of $\mathbb{Z}^2 \times \{-n+1,\ldots,n-1\}$ defined by

$$\begin{aligned}
\mathcal{I}_1 &= \{i=j, h=0\}, \quad \mathcal{I}_2 = \{i=-j=h\neq 0\}, \\
\mathcal{I}_3 &= \{|i|\neq|j|\}, \quad \mathcal{I}_4 = \{i=j, h\neq 0\}, \quad \mathcal{I}_5 = \{i=-j, h\neq i\neq 0\}.
\end{aligned}$$

In view of Lemma 4,

$$\begin{aligned}
\sum_{(i,j,h)\in\mathcal{I}_1} \mathbf{A}_n(i,j,h) &= \sigma^4 b_n \sum_{i=-\infty}^{+\infty} \left\{\omega^2(ib_n)D_{p+q}^{+}(J\otimes J)D_{p+q}^{+'} + O(\rho^{|i|})\right\} \\
&\overset{n\to\infty}{\longrightarrow} \sigma^4\varpi^2 D_{p+q}^{+}(J\otimes J)D_{p+q}^{+'}.
\end{aligned}$$

We have $\text{Cov}\left\{\mathbf{u}_1 \otimes \mathbf{u}_{1+i}, \mathbf{u}_{1+i} \otimes \mathbf{u}_1\right\} = \text{Cov}\left\{\mathbf{u}_1 \otimes \mathbf{u}_{1+i}, \mathbf{u}_1 \otimes \mathbf{u}_{1+i}\right\} K_{p+q}$ where K_{p+q} is the (symmetric) commutation matrix such that $K_{p+q}\text{vec}A = \text{vec}A'$ for any $(p+q)\times(p+q)$ matrix A. Thus,

$$\begin{aligned}
\sum_{(i,j,h)\in\mathcal{I}_2} \mathbf{A}_n(i,j,h) &= \sigma^4 b_n \sum_{i=-n+1}^{n-1} \frac{n-|i|}{n} \left\{\omega^2(ib_n)D_{p+q}^{+}(J\otimes J)K_{p+q}D_{p+q}^{+'} + O(\rho^{|i|})\right\} \\
&\overset{n\to\infty}{\longrightarrow} \sigma^4\varpi^2 D_{p+q}^{+}(J\otimes J)K_{p+q}D_{p+q}^{+'}
\end{aligned}$$

From Lemma 5,

$$\sum_{(i,j,h)\in\mathcal{I}_3} \mathbf{A}_n(i,j,h) = O(b_n) \overset{n\to\infty}{\longrightarrow} 0.$$

From Lemma 4,

$$\sum_{(i,j,h)\in\mathcal{I}_4} \mathbf{A}_n(i,j,h) = b_n \sum_{0<|h|<n} \rho^{|h|} \stackrel{n\to\infty}{\longrightarrow} 0, \qquad \sum_{(i,j,h)\in\mathcal{I}_5} \mathbf{A}_n(i,j,h) = 0.$$

Therefore

$$\lim_{n\to\infty} nb_n \mathrm{Var}\left\{\mathrm{vech}\; I_n^{\mathbf{u}}\right\} = \sigma^4 \varpi^2 D_{p+q}^{+}(J\otimes J)(\mathbf{I}_{(p+q)^2} + K_{p+q})D_{p+q}^{+'}.$$

The conclusion follows from the relation

$$(\mathbf{I}_{(p+q)^2} + K_{p+q})D_{p+q}^{+'} = 2D_{p+q}D_{p+q}^{+}D_{p+q}^{+'} = 2D_{p+q}^{+'}.$$

(see Magnus and Neudecker, 1988, Theorem 3.12). □

Lemma 7

$$E\left\|\frac{\partial}{\partial\theta'}(vec\;\hat{\Delta}_i^{\mathbf{u}})\right\| < K\rho^{|i|} + \frac{K}{\sqrt{n}}.$$

Proof. Because $\hat{\Delta}_i^{\mathbf{u}} = \hat{\Delta}_{-i}^{\mathbf{u}'}$, we will only consider the case $0 \le i < n$. We have

$$\begin{aligned}
\frac{\partial}{\partial\theta'}(\mathrm{vec}\;\hat{\Delta}_i^{\mathbf{u}}) &= \frac{4}{n}\sum_{t=1}^{n-i} \frac{\partial \mathbf{u}_{t+i}}{\partial\theta'} \otimes \mathbf{u}_t + \mathbf{u}_{t+i} \otimes \frac{\partial \mathbf{u}_t}{\partial\theta'} \\
&= \frac{4}{n}\sum_{t=1}^{n-i} \epsilon_t \left(\frac{\partial\epsilon_{t+i}}{\partial\theta}\frac{\partial\epsilon_{t+i}}{\partial\theta'}\right) \otimes \frac{\partial\epsilon_t}{\partial\theta} + \frac{4}{n}\sum_{t=1}^{n-i} \epsilon_t\epsilon_{t+i}\frac{\partial^2\epsilon_{t+i}}{\partial\theta\partial\theta'} \otimes \frac{\partial\epsilon_t}{\partial\theta} \\
&\quad + \frac{4}{n}\sum_{t=1}^{n-i} \epsilon_{t+i}\frac{\partial\epsilon_{t+i}}{\partial\theta} \otimes \left(\frac{\partial\epsilon_t}{\partial\theta}\frac{\partial\epsilon_t}{\partial\theta'}\right) + \frac{4}{n}\sum_{t=1}^{n-i} \epsilon_t\epsilon_{t+i}\frac{\partial\epsilon_{t+i}}{\partial\theta} \otimes \frac{\partial^2\epsilon_t}{\partial\theta\partial\theta'}.
\end{aligned}$$

Considering the first sum in the right-hand side, we will prove that, for any $\ell_1, \ell_2, \ell_3 \in \{1,\ldots,p+q\}$

$$E\left\|\frac{1}{n}\sum_{t=1}^{n-i} \epsilon_t \frac{\partial\epsilon_{t+i}}{\partial\theta_{\ell_1}}\frac{\partial\epsilon_{t+i}}{\partial\theta_{\ell_2}}\frac{\partial\epsilon_t}{\partial\theta_{\ell_3}}\right\| < K\rho^i + \frac{K}{\sqrt{n}}. \qquad (26)$$

The other sums can be handled in a similar fashion. For $i = 0$ or $i = 1$, (26) holds straightforwardly because the L^1 norm of the term inside the sum exists. Now, for $i > 1$ write

$$\frac{\partial \epsilon_{t+i}}{\partial \theta_\ell} = {}_{i-1}\left(\frac{\partial \epsilon_{t+i}}{\partial \theta_\ell}\right) + {}^{i-1}\left(\frac{\partial \epsilon_{t+i}}{\partial \theta_\ell}\right),$$

with the notation in (20). Note that the truncated derivative, *i.e.* the first term in the right-hand side of the previous equality, is independent of ϵ_{t-j}, $j \geq 0$.

We will first prove that (26) holds when $\partial \epsilon_{t+i}/\partial \theta_{\ell_1}$ and $\partial \epsilon_{t+i}/\partial \theta_{\ell_2}$ are replaced by the truncated derivatives. It will be sufficient to show that

$$E\left\{\frac{1}{n}\sum_{t=1}^{n-i} \epsilon_t \; {}_{i-1}\left(\frac{\partial \epsilon_{t+i}}{\partial \theta_{\ell_1}}\right) {}_{i-1}\left(\frac{\partial \epsilon_{t+i}}{\partial \theta_{\ell_2}}\right) \frac{\partial \epsilon_t}{\partial \theta_{\ell_3}}\right\}^2 < \frac{K}{n}. \tag{27}$$

By stationarity the left-hand side in (27) is bounded by

$$\begin{aligned}
& \frac{1}{n}\sum_{h=-\infty}^{+\infty}\left| E\left(\epsilon_1 \; {}_{i-1}\left(\frac{\partial \epsilon_{1+i}}{\partial \theta_{\ell_1}}\right) {}_{i-1}\left(\frac{\partial \epsilon_{1+i}}{\partial \theta_{\ell_2}}\right) \frac{\partial \epsilon_1}{\partial \theta_{\ell_3}}\right.\right. \\
& \qquad \left.\left. \times \; \epsilon_{1+|h|} \; {}_{i-1}\left(\frac{\partial \epsilon_{1+|h|+i}}{\partial \theta_{\ell_1}}\right) {}_{i-1}\left(\frac{\partial \epsilon_{1+|h|+i}}{\partial \theta_{\ell_2}}\right) \frac{\partial \epsilon_{1+|h|}}{\partial \theta_{\ell_3}}\right)\right| \\
\leq \; & \frac{1}{n}\sum_{h=-\infty}^{+\infty} \sum_{k_1,k_2,k_4,k_5=1}^{i-1} \sum_{k_3,k_6=1}^{\infty} \left| c_{k_1,\ell_1} c_{k_2,\ell_2} c_{k_3,\ell_3} c_{k_4,\ell_1} c_{k_5,\ell_2} c_{k_6,\ell_3} \right| \\
& \times \left| E\left(\epsilon_1 \epsilon_{1+i-k_1} \epsilon_{1+i-k_2} \epsilon_{1-k_3} \epsilon_{1+|h|} \epsilon_{1+|h|+i-k_4} \epsilon_{1+|h|+i-k_5} \epsilon_{1+|h|-k_6}\right)\right|.
\end{aligned}$$

It is easily seen that in the last expectation at most four indexes can be equal which, by Lemma 2 ensures its existence. Moreover, when $h \neq 0$ the expectation vanishes. Therefore (27) holds.

It remains to show that (26) holds when $\partial \epsilon_{t+i}/\partial \theta_{\ell_1}$ and/or $\partial \epsilon_{t+i}/\partial \theta_{\ell_2}$ are replaced by the complements of the truncated derivatives. For instance we have

$$\begin{aligned} E\left\| \frac{1}{n}\sum_{t=1}^{n-i} \epsilon_t \;{}^{i-1}\!\left(\frac{\partial \epsilon_{t+i}}{\partial \theta_{\ell_1}}\right) \frac{\partial \epsilon_{t+i}}{\partial \theta_{\ell_2}} \frac{\partial \epsilon_{t}}{\partial \theta_{\ell_3}} \right\| &\leq \|\epsilon_t\|_4 \left\| \;{}^{i-1}\!\left(\frac{\partial \epsilon_{t+i}}{\partial \theta_{\ell_1}}\right) \right\|_4 \left\| \frac{\partial \epsilon_{t+i}}{\partial \theta_{\ell_2}} \right\|_4 \left\| \frac{\partial \epsilon_{t}}{\partial \theta_{\ell_3}} \right\|_4 \\ &< K \times K\rho^i \times K \times K. \end{aligned}$$

The proof of (26) is completed. Hence the Lemma is proved. $\square$

Lemma 8

$$\sqrt{nb_n} E\left\{ \sup_{\theta\in\Theta_\delta} \left\| \hat{I}_n(\theta) - \hat{I}_n^{\mathbf{u}}(\theta) \right\| \right\} \overset{n\to\infty}{\longrightarrow} 0.$$

Proof. The matrix norm being multiplicative, the supremum inside the brackets is bounded by

$$\frac{2}{n}\sum_{i=0}^{[a/b_n]} \omega(ib_n) \sum_{t=1}^{n-i} \lambda_{i,t}$$

where

$$\lambda_{i,t} = \sup_{\theta\in\Theta_\delta} \{ \|\mathbf{u}_t(\theta) - \mathbf{v}_t(\theta)\| \|\mathbf{u}_{t+i}(\theta)\| + \|\mathbf{u}_{t+i}(\theta) - \mathbf{v}_{t+i}(\theta)\| \|\mathbf{v}_t(\theta)\| \}.$$

Note that

$$\max\left\{ \sup_{\theta\in\Theta_\delta} |\epsilon_t(\theta) - e_t(\theta)|, \sup_{\theta\in\Theta_\delta} \left\| \frac{\partial}{\partial\theta}\epsilon_t(\theta) - \frac{\partial}{\partial\theta} e_t(\theta) \right\| \right\} \leq K \sum_{m\geq 1} \rho^{t+m} |\epsilon_{1-m}|.$$

Hence

$$\begin{aligned} \sup_{\theta\in\Theta_\delta} \|\mathbf{u}_t(\theta) - \mathbf{v}_t(\theta)\| &\leq \sup_{\theta\in\Theta_\delta} \left\{ |\epsilon_t(\theta) - e_t(\theta)| \left\| \frac{\partial}{\partial\theta}\epsilon_t(\theta) \right\| + |e_t(\theta)| \left\| \frac{\partial}{\partial\theta}\epsilon_t(\theta) - \frac{\partial}{\partial\theta} e_t(\theta) \right\| \right\} \\ &\leq K \sum_{m_1,m_2\geq 1} \rho^{t+m_1+m_2} |\epsilon_{1-m_1}| |\epsilon_{t-m_2}|. \end{aligned}$$

It follows that, for $i \geq 0$

$$\lambda_{i,t} \leq K \sum_{m_1,m_2,m_3,m_4\geq 1} \rho^{t+m_1+m_2+m_3+m_4} |\epsilon_{1-m_1}| |\epsilon_{t-m_2}| |\epsilon_{t+i-m_3}| \left(|\epsilon_{t+i-m_4}| + |\epsilon_{t-m_4}|\right).$$

Therefore, in view of $E\epsilon_t^4 < \infty$, we get $E(\lambda_{i,t}) \leq K\rho^t$, from which we deduce

$$\begin{aligned}\sqrt{nb_n}E\left\{\sup_{\theta\in\Theta_\delta}\left\|\hat{I}_n(\theta)-\hat{I}_n^{\mathbf{u}}(\theta)\right\|\right\} &\leq K\sqrt{\frac{b_n}{n}}\sum_{i=0}^{[a/b_n]}\omega(ib_n)\sum_{t=1}^{\infty}\rho^t \\ &\leq \frac{K}{\sqrt{nb_n}} = o(1). \quad \square\end{aligned}$$

Lemma 9

$$\sqrt{nb_n}\left(\hat{I} - I - \overline{I}_n^{\mathbf{u}}\right) = o_P(1).$$

Proof. We prove this lemma by showing that:

i) $\sqrt{nb_n}\left(\hat{I} - \hat{I}_n^{\mathbf{u}}(\hat{\theta}_n)\right) = o_P(1)$;

ii) $\sqrt{nb_n}\left(\hat{I}_n^{\mathbf{u}}(\hat{\theta}_n) - \hat{I}_n^{\mathbf{u}}\right) = o_P(1)$;

iii) $\sqrt{nb_n}\left(\hat{I}_n^{\mathbf{u}} - I_n^{\mathbf{u}}\right) = o_P(1)$;

iv) $\sqrt{nb_n}\left(I_n^{\mathbf{u}} - I - \overline{I}_n^{\mathbf{u}}\right) = o_P(1)$.

Result *i)* is a straightforward consequence of Lemma 8. To prove *ii)* we proceed by applying the mean-value theorem to the (ℓ_1,ℓ_2)-th component of $\hat{I}_n^{\mathbf{u}}$. For some θ between $\hat{\theta}_n$ and θ_0 we have

$$\hat{I}_n^{\mathbf{u}}(\hat{\theta}_n)(\ell_1,\ell_2) - \hat{I}_n^{\mathbf{u}}(\ell_1,\ell_2) = (\hat{\theta}_n - \theta_0)'\frac{\partial}{\partial\theta}\hat{I}_n^{\mathbf{u}}(\theta)(\ell_1,\ell_2).$$

Since $\|\hat{\theta}_n - \theta_0\| = O_P(n^{-1/2})$, it is sufficient to prove that

$$\sup_{\theta\in\Theta_\delta}\left\|\frac{\partial}{\partial\theta}\hat{I}_n^{\mathbf{u}}(\theta)(\ell_1,\ell_2)\right\| = O_P(1). \tag{28}$$

Straightforward algebra shows that

$$\begin{aligned}&\frac{\partial}{\partial\theta'}\left[\text{vec}\left\{\frac{\partial}{\partial\theta'}\text{vec }\hat{\Delta}_i^{\mathbf{u}}(\theta)\right\}\right] \\ =\ &\frac{1}{n}\sum_{t=1}^{n-i}\left(\frac{\partial}{\partial\theta'}\text{vec }\frac{\partial\mathbf{u}_{t+i}}{\partial\theta'}\right)\otimes\mathbf{u}_t(\theta) + \frac{1}{n}\sum_{t=1}^{n-i}\text{vec}\left(\frac{\partial\mathbf{u}_{t+i}}{\partial\theta'}\right)\otimes\frac{\partial\mathbf{u}_t}{\partial\theta'}(\theta) \\ &+\frac{1}{n}\sum_{t=1}^{n-i}\frac{\partial\mathbf{u}_{t+i}}{\partial\theta'}\otimes\text{vec}\left(\frac{\partial\mathbf{u}_t}{\partial\theta'}\right)(\theta) + \frac{1}{n}\sum_{t=1}^{n-i}\mathbf{u}_{t+i}\otimes\left(\frac{\partial}{\partial\theta'}\text{vec }\frac{\partial\mathbf{u}_t}{\partial\theta'}\right)(\theta). \end{aligned} \tag{29}$$

Using the Cauchy-Schwarz inequality and the ergodic theorem, we have, almost surely,

$$\begin{aligned}
&\lim_{n\to\infty}\sup_i \sup_{\theta\in\Theta_\delta}\left\|\frac{1}{n}\sum_{t=1}^{n-i}\left(\frac{\partial}{\partial\theta'}\text{vec}\,\frac{\partial \mathbf{u}_{t+i}}{\partial\theta'}\right)\otimes \mathbf{u}_t(\theta)\right\| \\
\leq\ & \lim_{n\to\infty}\sup_i \frac{1}{n}\sum_{t=1}^{n-i}\sup_{\theta\in\Theta_\delta}\left\|\left(\frac{\partial}{\partial\theta'}\text{vec}\,\frac{\partial \mathbf{u}_{t+i}}{\partial\theta'}(\theta)\right)\right\|\sup_{\theta\in\Theta_\delta}\|\mathbf{u}_t(\theta)\| \\
\leq\ & \lim_{n\to\infty}\left\{\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta_\delta}\left\|\left(\frac{\partial}{\partial\theta'}\text{vec}\,\frac{\partial \mathbf{u}_{t}}{\partial\theta'}(\theta)\right)\right\|^2\right\}^{1/2}\left\{\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta_\delta}\|\mathbf{u}_t(\theta)\|^2\right\}^{1/2} < \infty.
\end{aligned}$$

Treating in the same way the three other sums of (29), we deduce

$$\lim_{n\to\infty}\sup_i \sup_{\theta\in\Theta_\delta}\left\|\frac{\partial}{\partial\theta'}\left[\text{vec}\,\left\{\frac{\partial}{\partial\theta'}\text{vec}\,\hat{\Delta}_i^{\mathbf{u}}(\theta)\right\}\right]\right\| < \infty \quad \text{a.s.} \tag{30}$$

Now a Taylor expansion gives, for any ℓ_1, ℓ_2, ℓ_3

$$\frac{\partial}{\partial\theta_{\ell_3}}\hat{\Delta}_i^{\mathbf{u}}(\theta)(\ell_1,\ell_2) = \frac{\partial}{\partial\theta_{\ell_3}}\hat{\Delta}_i^{\mathbf{u}}(\ell_1,\ell_2) + (\theta-\theta_0)'\frac{\partial}{\partial\theta}\left\{\frac{\partial}{\partial\theta_{\ell_3}}\hat{\Delta}_i^{\mathbf{u}}(\theta^*)(\ell_1,\ell_2)\right\}, \tag{31}$$

where θ^* is between θ and θ_0. From (31), (30), Lemma 7 and the Cesaro Lemma we obtain

$$\begin{aligned}
\left\|\frac{\partial}{\partial\theta}\hat{I}_n^{\mathbf{u}}(\theta)(\ell_1,\ell_2)\right\| &\leq K\sum_{i=-[ab_n^{-1}]}^{[ab_n^{-1}]}\left\|\frac{\partial}{\partial\theta}\hat{\Delta}_i^{\mathbf{u}}(\ell_1,\ell_2)\right\| \\
&\quad + K\|\theta-\theta_0\|\sum_{i=-[ab_n^{-1}]}^{[ab_n^{-1}]}\sup_i\sup_{\theta^*\in\Theta_\delta}\left\|\frac{\partial^2}{\partial\theta\partial\theta'}\hat{\Delta}_i^{\mathbf{u}}(\theta^*)(\ell_1,\ell_2)\right\| \\
&= O_P(1) + O_P(n^{-1/2}b_n^{-1}) + O_P(b_n^{-1}\|\theta-\theta_0\|).
\end{aligned} \tag{32}$$

Hence (28), and thus *ii)* is proved.

Next we prove *iii)*. We have

$$\text{vec}\ (\hat{I}_n^{\mathbf{u}} - I_n^{\mathbf{u}}) = \sum_{i=1}^{[a/b_n]}\omega(ib_n)\frac{1}{n}\sum_{t=n-i+1}^{n}\mathbf{u}_{t+i}\otimes\mathbf{u}_t + \mathbf{u}_t\otimes\mathbf{u}_{t+i}$$

Hence, by Lemma 3

$$\begin{aligned}
& nb_n \text{Var}\left\{\text{vec }(\hat{I}_n^{\mathbf{u}} - I_n^{\mathbf{u}})\right\} \\
= \; & \frac{b_n}{n} \sum_{i,j=1}^{[a/b_n]} \omega(ib_n)\omega(jb_n) \\
& \times \sum_{t=n-i+1}^{n} \sum_{s=n-j+1}^{n} \text{Cov}\left\{\mathbf{u}_{t+i} \otimes \mathbf{u}_t + \mathbf{u}_t \otimes \mathbf{u}_{t+i}, \mathbf{u}_{s+j} \otimes \mathbf{u}_s + \mathbf{u}_s \otimes \mathbf{u}_{s+j}\right\} \\
\leq \; & \frac{Kb_n}{n} \sum_{i,j=1}^{[a/b_n]} ij = O\left(\frac{1}{nb_n^3}\right) = o(1),
\end{aligned}$$

which establishes *iii)*.

Note that, under **A4**, we have $I = E(I_n^{\mathbf{u}} - \overline{I}_n^{\mathbf{u}}) = E\Delta_0^{\mathbf{u}}$. To prove *iv)* it suffices therefore to show that

$$\sqrt{n} \text{ vec}\left\{I_n^{\mathbf{u}} - I - \overline{I}_n^{\mathbf{u}}\right\} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{ \mathbf{u}_t \otimes \mathbf{u}_t - E(\mathbf{u}_t \otimes \mathbf{u}_t)\}$$

is bounded in probability. This is straightforward from Lemma 4. □

Lemma 10

$$\lim_{n\to\infty} Var\left[\sqrt{nb_n}\ vec\ (\overline{I}_n^{\mathbf{u}} - \ {}_r\overline{I}_n^{\mathbf{u}})\right] = O(\rho^r) \quad as\ r \to \infty.$$

Proof. We start by writing

$$\text{vec } (\overline{I}_n^{\mathbf{u}} - {}_r\overline{I}_n^{\mathbf{u}}) = \sum_{0<|i|\leq[a/b_n]} \omega(ib_n)\frac{1}{n}\sum_{t=1}^{n} {}^r\mathbf{u}_{t+i} \otimes \mathbf{u}_t + \mathbf{u}_{t+i} \otimes {}^r\mathbf{u}_t - {}^r\mathbf{u}_{t+i} \otimes {}^r\mathbf{u}_t.$$

This double sum can obviously be split into six parts (distinguishing $i > 0$ and $i < 0$), and it will be sufficient to show for instance that

$$\lim_{n\to\infty} \text{Var}\left[\sqrt{nb_n} \sum_{i=1}^{[a/b_n]} \omega(ib_n)\frac{1}{n}\sum_{t=1}^{n} {}^r\mathbf{u}_{t+i} \otimes \mathbf{u}_t\right] = O(\rho^r) \qquad (33)$$

since the other terms can be treated in precisely the same way. The variance in (33) can be written as

$$\frac{b_n}{n}\sum_{i,j=1}^{[a/b_n]}\omega(ib_n)\omega(jb_n)\sum_{|h|<n}(n-|h|)\mathrm{Cov}\left\{\,^r\mathbf{u}_{1+i}\otimes\mathbf{u}_1,{}^r\mathbf{u}_{1+h+j}\otimes\mathbf{u}_{1+h}\right\}. \quad (34)$$

The existence of these covariances is obtained by a straightforward extension of Lemma 3. Proceeding as in the proofs of Lemmas 4 and 5 we find, for $i,j>0$

$$\mathrm{Cov}\left\{\,^r\mathbf{u}_{1+i}\otimes\mathbf{u}_1,{}^r\mathbf{u}_{1+h+i}\otimes\mathbf{u}_{1+h}\right\} = \begin{cases} 0 & \text{when} \quad h\neq 0, \\ O(\rho^r) & \text{when} \quad h=0, \end{cases}$$

uniformly in i, and

$$\sum_{h=-\infty}^{+\infty}\|\mathrm{Cov}\left\{\,^r\mathbf{u}_{1+i}\otimes\mathbf{u}_1,{}^r\mathbf{u}_{1+h+j}\otimes\mathbf{u}_{1+h}\right\}\| \;\leq\; K\rho^i\rho^j\rho^r \quad \text{when} \quad i\neq j.$$

It follows from (34) that (33) holds, which concludes the proof of this lemma. □

Lemma 11 *For $m<k$,*

$$Var\left[\sqrt{(k-m+1)b_n}\; vec\left(\,_r\overline{I}^{\mathbf{u}}_{n,(m,k)} - \,_r\overline{I}^{\mathbf{u},\kappa}_{n,(m,k)}\right)\right] \leq K\kappa^{-\nu/2},$$

where K is independent of m,k,n,κ.

Proof. We have

$$\begin{aligned}&\mathrm{vec}\left(\,_r\overline{I}^{\mathbf{u}}_{n,(m,k)} - \,_r\overline{I}^{\mathbf{u},\kappa}_{n,(m,k)}\right)\\ =&\sum_{0<|i|\leq[a/b_n]}\omega(ib_n)\frac{1}{k-m+1}\sum_{t=m}^{k}\,^\kappa_r\mathbf{u}_{t+i}\otimes\,_r\mathbf{u}_t+\,_r\mathbf{u}_{t+i}\otimes\,^\kappa_r\mathbf{u}_t-\,^\kappa_r\mathbf{u}_{t+i}\otimes\,^\kappa_r\mathbf{u}_t.\end{aligned}$$

Again, this double sum can be split into six parts and, as in the above lemma, it will be sufficient to show for instance that

$$\operatorname{Var}\left[\sqrt{\frac{b_n}{k-m+1}}\sum_{i=1}^{[a/b_n]}\omega(ib_n)\sum_{t=m}^{k}{}_r^{\kappa}\mathbf{u}_{t+i}\otimes{}_r\mathbf{u}_t\right]\leq K\kappa^{-\nu/4}. \qquad (35)$$

Let ${}^{\kappa}\epsilon_t=\epsilon_t-\epsilon_t^{\kappa}$. Recall that $E^{\kappa}\epsilon_t=E\epsilon_t=E\epsilon_t^{\kappa}=0$. Note that the ℓ-th components of ${}_r^{\kappa}\mathbf{u}_{t+i}$ and ${}_r\mathbf{u}_t$ are

$${}_r^{\kappa}\mathbf{u}_{t+i}(\ell)=\sum_{k_1=1}^{r}c_{k_1,\ell}\left({}^{\kappa}\epsilon_{t+i}\epsilon_{t+i-k_1}+\epsilon_{t+i}^{\kappa}\,{}^{\kappa}\epsilon_{t+i-k_1}\right)\quad\text{and}\quad{}_r\mathbf{u}_t(\ell)=\sum_{k_2=1}^{r}c_{k_2,\ell}\epsilon_t\epsilon_{t-k_2}.$$

It is clear that in Lemma 3, some or all of the ϵ_t's can be replaced by the truncated variables ${}^{\kappa}\epsilon_t$ or ϵ_t^{κ}. Thus the variance in (35) is well defined. In addition, by

$$\|{}^{\kappa}\epsilon_t\|_4\leq\left\|\epsilon_t1_{\{|\epsilon_t|>\kappa\}}\right\|_4+\left|E\epsilon_t1_{\{|\epsilon_t|\leq\kappa\}}\right|\leq\left\{\frac{1}{\kappa^{\nu}}E\left|\epsilon_t\right|^{4+\nu}\right\}^{1/4}+\left|E\epsilon_t1_{\{|\epsilon_t|>\kappa\}}\right|\leq\frac{K}{\kappa^{\nu/4}}$$

and by the Hölder inequality we get

$$\left|\operatorname{Cov}\left({}^{\kappa}\epsilon_{1+i}\epsilon_{1+i-k_1}\epsilon_1\epsilon_{1-k_2},{}^{\kappa}\epsilon_{1+i}\epsilon_{1+i-k_3}\epsilon_1\epsilon_{1-k_4}\right)\right|\leq\|{}^{\kappa}\epsilon_t\|_4^2\|\epsilon_t\|_4^6\leq\frac{K}{\kappa^{\nu/2}}.$$

The same inequality holds when the indexes are permuted and/or when the ϵ_t's are replaced by the ϵ_t^{κ}'s. Therefore we have, for $i,j>0$

$$\operatorname{Cov}\left\{{}_r^{\kappa}\mathbf{u}_{1+i}\otimes{}_r\mathbf{u}_1,\ {}_r^{\kappa}\mathbf{u}_{1+i+h}\otimes{}_r\mathbf{u}_{1+h}\right\}=\begin{cases}0 & \text{when}\quad h\neq0,\\ O(\kappa^{-\nu/2}) & \text{when}\quad h=0,\end{cases}$$

$$\sum_{h=-\infty}^{+\infty}\left\|\operatorname{Cov}\left\{{}_r^{\kappa}\mathbf{u}_{1+i}\otimes{}_r\mathbf{u}_1,\ {}_r^{\kappa}\mathbf{u}_{1+j+h}\otimes{}_r\mathbf{u}_{1+h}\right\}\right\|\leq K\rho^i\rho^j\kappa^{-\nu/2}\quad\text{when}\quad i\neq j$$

and the conclusion follows as in Lemma 10. □

Lemma 12 *For any* $\lambda \in \mathbb{R}^{(p+q)^2}$, $\lambda \neq 0$

$$\left\| \sqrt{b_n}\, \lambda' \; vec\left({}_r\overline{I}^{\mathbf{u},\kappa}_{n,(m,k)} \right) \right\|_4 = O(\kappa^2 b_n^{-1/4}) \quad as\ \kappa \to \infty$$

uniformly in m *and* k.

Proof. The variable in the left-hand side can be written as $\frac{1}{k-m+1}\sum_{t=m}^{k} U_t$ where

$$U_t = \sqrt{b_n} \sum_{0<i\leq[a/b_n]} \omega(ib_n)\lambda' \left({}_r\mathbf{u}^{\kappa}_{t+i} \otimes {}_r\mathbf{u}^{\kappa}_{t} + {}_r\mathbf{u}^{\kappa}_{t} \otimes {}_r\mathbf{u}^{\kappa}_{t+i} \right).$$

It is clear that $\|{}_r\mathbf{u}^{\kappa}_t\| \leq K\kappa^2$. Hence $|U_t| \leq K\kappa^4 b_n^{-1/2}$ and $EU_t^4 \leq K\kappa^8 EU_t^2 b_n^{-1}$.

By arguments used in the proof of Lemma 4,

$$\mathrm{Cov}\left({}_r\mathbf{u}^{\kappa}_{t} \otimes {}_r\mathbf{u}^{\kappa}_{t+i},\ {}_r\mathbf{u}^{\kappa}_{t} \otimes {}_r\mathbf{u}^{\kappa}_{t+j} \right) = 0$$

for $i, j > 0$ and $i \neq j$. Therefore

$$\mathrm{Var}U_t = b_n \sum_{0<i\leq[a/b_n]} \omega^2(ib_n)\lambda' \mathrm{Var}\left({}_r\mathbf{u}^{\kappa}_{t+i} \otimes {}_r\mathbf{u}^{\kappa}_{t} + {}_r\mathbf{u}^{\kappa}_{t} \otimes {}_r\mathbf{u}^{\kappa}_{t+i} \right) = O(1)$$

uniformly in κ and r. The conclusion follows. $\square$

Lemma 13 *The following hold*

$$\begin{array}{llcl}
(a) & \lim_{r\to\infty} \overline{\lim}_{n\to\infty} Var\left\{ \sqrt{nb_n}\ vec\left(\overline{I}^{\mathbf{u}}_n - {}_r\overline{I}^{\mathbf{u}}_n \right) \right\} & = & 0; \\
(b)\ \ if\ \ k_n b_n \to \infty\ \ and\ \ k_n n^{-1} \to 0, & \sqrt{nb_n}\ vec\left({}_r\overline{I}^{\mathbf{u}}_n - {}_rS_n \right) & = & o_P(1); \\
(c)\ \ if, moreover,\ \ \kappa_n \to \infty, & \sqrt{nb_n}\ vec\left({}_rS_n - {}_rS^{\kappa_n}_n \right) & = & o_P(1);
\end{array}$$

Proof. Part (a) is a direct consequence of Lemma 10.

Next we turn to (b). Observe that, in view of (21),

$$\begin{aligned}
&\sqrt{nb_n}\ \mathrm{vec}\left({}_r\overline{I}^{\mathbf{u}}_n - {}_rS_n\right)\\
=\ &\sum_{\ell=0}^{p_n-2}\frac{\sqrt{b_n}}{\sqrt{n}}\sum_{0<i\le[a/b_n]}\omega(ib_n)\sum_{t=(\ell+1)k_n-m_n+1}^{(\ell+1)k_n}{}_r\mathbf{u}_{t+i}\otimes{}_r\mathbf{u}_t + {}_r\mathbf{u}_t\otimes{}_r\mathbf{u}_{t+i}\\
&+\frac{\sqrt{b_n}}{\sqrt{n}}\sum_{0<i\le[a/b_n]}\omega(ib_n)\sum_{t=p_nk_n-m_n+1}^{n}{}_r\mathbf{u}_{t+i}\otimes{}_r\mathbf{u}_t + {}_r\mathbf{u}_t\otimes{}_r\mathbf{u}_{t+i}
\end{aligned}$$

is a sum of $p_n - 1$ independent random matrices (for n large enough so that $k_n - m_n > m_n$). Now, by the arguments of the proofs of Lemmas 4 and 5, we have for $i, j > 0$

$$\mathrm{Cov}\left({}_r\mathbf{u}_t\otimes{}_r\mathbf{u}_{t+i},\ {}_r\mathbf{u}_s\otimes{}_r\mathbf{u}_{s+j}\right) = \begin{cases} 0 & \text{when}\quad t+i\neq s+j,\\ O(\rho^i\rho^j) & \text{when}\quad t+i = s+j.\end{cases}$$

Therefore

$$\begin{aligned}
&\mathrm{Var}\left\{\frac{\sqrt{b_n}}{\sqrt{n}}\sum_{0<i\le[a/b_n]}\omega(ib_n)\sum_{t=k}^{m}{}_r\mathbf{u}_{t+i}\otimes{}_r\mathbf{u}_t + {}_r\mathbf{u}_t\otimes{}_r\mathbf{u}_{t+i}\right\}\\
=\ &O\left(\frac{b_n(m-k+1)}{n}\right) = O\left(\frac{m-k+1}{n}\right).
\end{aligned}$$

Then

$$nb_n\ \mathrm{Var}\left\{\ \mathrm{vec}\left({}_r\overline{I}^{\mathbf{u}}_n - {}_rS_n\right)\right\} = O\left(\frac{(p_n-1)m_n}{n}\right) + O\left(\frac{n-p_nk_n+m_n}{n}\right) = o(1).$$

Hence, since $E\left\{\ \mathrm{vec}\left({}_r\overline{I}^{\mathbf{u}}_n - {}_rS_n\right)\right\} = 0$, (b) is proved.

For part (c), we note that $\sqrt{nb_n}\ \mathrm{vec}\left({}_rS_n - {}_rS_n^{\kappa_n}\right)$ is a sum of p_n i.i.d. variables whose common variance is

$$\frac{k_n-m_n}{n}\ \mathrm{Var}\left\{\sqrt{(k_n-m_n)b_n}\ \mathrm{vec}\left({}_r\overline{I}^{\mathbf{u}}_{n,(1,k_n-m_n)} - {}_r\overline{I}^{\mathbf{u},\kappa_n}_{n,(1,k_n-m_n)}\right)\right\}.$$

Thus, in view of Lemma 11

$$\text{Var}\left\{\sqrt{nb_n}\ \text{vec}\left({}_rS_n - {}_rS_n^{\kappa_n}\right)\right\} = O\left(\frac{p_n(k_n - m_n)}{n\kappa_n^{\nu/2}}\right) = O\left(\frac{1}{\kappa_n^{\nu/2}}\right) = o(1) \tag{36}$$

when $n \to \infty$. This establishes (c) and completes the proof of Lemma 13. □

Lemma 14

$$\sqrt{nb_n}\ \textit{vech}\left({}_r\overline{I}_n^{\mathbf{u}}\right) \overset{d}{\rightsquigarrow} \mathcal{N}\left(0, 2\sigma^4\varpi^2 D_{p+q}^+({}_rJ \otimes {}_rJ)D_{p+q}^{+'}\right).$$

Proof. The random matrices $\sqrt{nb_n}$ vech ${}_r\overline{I}_n^{\mathbf{u}}$ are centered. By a trivial extension of part *iv)* of the proof of Lemma 9, $\sqrt{nb_n}$ vech ${}_r\overline{I}_n^{\mathbf{u}}$ and $\sqrt{nb_n}$ vech $({}_rI_n^{\mathbf{u}} - E_rI_n^{\mathbf{u}})$ have the same asymptotic distribution. Is is easy to see that Lemmas 4 and 5 still hold when $(\mathbf{u}_t)$ and J are replaced by $({}_r\mathbf{u}_t)$ and ${}_rJ$. Therefore, by the proof of Lemma 6,

$$\lim_{n\to\infty} \text{Var}\left\{\sqrt{nb_n}\ \text{vech}\ {}_r\overline{I}_n^{\mathbf{u}}\right\} = 2\sigma^4\varpi^2 D_{p+q}^+({}_rJ \otimes {}_rJ)D_{p+q}^{+'}.$$

By virtue of Lemma 13 (b) and (c), vech ${}_r\overline{I}_n^{\mathbf{u}}$, vech ${}_rS_n$ and vech ${}_rS_n^{\kappa_n}$ have the same asymptotic distribution for appropriately chosen sequences (k_n) and (κ_n).

To establish the asymptotic normality of $\sqrt{nb_n}$ vech ${}_rS_n^{\kappa_n}$, we will use the Cramér-Wold device. Therefore we will show the asymptotic normality of

$$\begin{aligned} X_n &:= \sqrt{nb_n}\ \lambda'\ \text{vec}\ {}_rS_n^{\kappa_n} \\ &= \sum_{\ell=0}^{p_n-1} \frac{k_n - m_n}{\sqrt{n}} \sqrt{b_n}\ \lambda'\ \text{vec}\ {}_r\overline{I}^{\mathbf{u},\kappa_n}_{n,(\ell k_n+1,\ell k_n+k_n-m_n)} := \sum_{\ell=0}^{p_n-1} X_{n\ell}, \end{aligned}$$

for any non trivial $\lambda \in \mathbb{R}^{(p+q)^2}$.

Observe that X_n is a sum of p_n i.i.d. centered variables with common variance

$$v_n^{\kappa_n} = \text{Var}\left\{\frac{\sqrt{b_n}}{\sqrt{n}}\sum_{0<i\leq[a/b_n]}\omega(ib_n)\sum_{t=1}^{k_n-m_n}\lambda'\left({}_r\mathbf{u}_{t+i}^{\kappa_n}\otimes {}_r\mathbf{u}_t^{\kappa_n} + {}_r\mathbf{u}_t^{\kappa_n}\otimes {}_r\mathbf{u}_{t+i}^{\kappa_n}\right)\right\}.$$

By (36), $v_n^{\kappa_n}$ is asymptotically equivalent, when $\kappa_n\to\infty$, to

$$v_n = \text{Var}\left\{\frac{\sqrt{b_n}}{\sqrt{n}}\sum_{0<i\leq[a/b_n]}\omega(ib_n)\sum_{t=1}^{k_n-m_n}\lambda'\left({}_r\mathbf{u}_{t+i}\otimes {}_r\mathbf{u}_t + {}_r\mathbf{u}_t\otimes {}_r\mathbf{u}_{t+i}\right)\right\}.$$

The arguments used to prove Lemmas 4 and 5 show that $v_n = O\left(\frac{k_n-m_n}{n}\right)$. Next, we will verify the Lindeberg condition. For any $\varepsilon>0$,

$$\begin{aligned}\sum_{\ell=0}^{p_n-1}\frac{1}{p_n v_n^{\kappa_n}}\int_{\left\{|X_{n\ell}|\geq\varepsilon\sqrt{p_n v_n^{\kappa_n}}\right\}}X_{n\ell}^2 dP &= \frac{1}{v_n^{\kappa_n}}\int_{\left\{|X_{n1}|\geq\varepsilon\sqrt{p_n v_n^{\kappa_n}}\right\}}X_{n1}^2 dP\\ &\leq \frac{EX_{n1}^4}{\varepsilon^2 p_n (v_n^{\kappa_n})^2}.\end{aligned}$$

Now Lemma 12 implies that $EX_{n1}^4 = O\left(\frac{k_n^4\kappa_n^8}{n^2 b_n}\right)$. Therefore

$$\sum_{\ell=0}^{p_n-1}\frac{1}{p_n v_n^{\kappa_n}}\int_{\left\{|X_{n\ell}|\geq\varepsilon\sqrt{p_n v_n^{\kappa_n}}\right\}}X_{n\ell}^2 dP = O\left(\frac{k_n^3\kappa_n^8}{nb_n}\right).$$

To fulfill the Lindeberg condition, it therefore suffices that $\frac{k_n^3\kappa_n^8}{nb_n}\to 0$. From Lemma 13 (c), it is also required that $\kappa_n\to\infty$. Therefore, in view Lemma 13 (b), it remains to show that we can find k_n such that $\frac{k_n^3}{nb_n}\to 0$ and $k_n b_n\to\infty$. This is obvious because $\frac{k_n^3}{nb_n} = \frac{(k_n b_n)^3}{nb_n^4}$ and it is supposed that $nb_n^4\to\infty$. □

Proof of Theorem 2. It comes from Lemma 9, Lemma 14, Lemma 13 (a), and a standard argument (see *e.g.* Billingsley (1995), Theorem 25.5). □

7.3 Lemmas and proofs for Theorem 3

In all subsequent lemmas, the assumptions of Theorem 2 are supposed to be satisfied.

Lemma 15 *Under the assumptions of Theorem 3,*

$$\sqrt{nb_n} vec\left(\hat{J}-J\right)=o_P(1).$$

Proof. By arguments already employed, we have

$$\begin{aligned}\text{Var}\left\{\sqrt{n}\text{ vec}\left(\hat{J}-J\right)\right\} &= \frac{1}{n}\sum_{|h|<n}(n-|h|)\text{Cov}\left(\frac{\partial\epsilon_1}{\partial\theta}\otimes\frac{\partial\epsilon_1}{\partial\theta},\frac{\partial\epsilon_{1+h}}{\partial\theta}\otimes\frac{\partial\epsilon_{1+h}}{\partial\theta}\right)+o(1)\\ &= O(1).\end{aligned}$$

Since $b_n\to 0$, the conclusion follows. □

Proof of Theorem 3. From Lemma 15 and straightforward algebra, we have

$$\begin{aligned}\sqrt{nb_n}\text{vech}\left(\hat{\Sigma}^{(2)}-\hat{\Sigma}^{(1)}\right) &= \sqrt{nb_n}\text{vech}\left\{J^{-1}\left(\hat{I}-I\right)J^{-1}\right\}+o_P(1)\\ &= D^+_{p+q}\left(J^{-1}\otimes J^{-1}\right)D_{p+q}\text{vech}\sqrt{nb_n}\left(\hat{I}-I\right)+o_P(1).\end{aligned}$$

The first convergence follows from Theorem 2, by application of the relation

$$D_{p+q}D^+_{p+q}(J\otimes J)D_{p+q}=(J\otimes J)D_{p+q}.$$

(see Magnus and Neudecker, 1988, Theorem 3.13). The second convergence is deduced, as in Theorem 3. □

Proof of Theorem 8. Under the assumptions of Theorem 8, Francq and Zakoïan (2000, Theorems 2 and 3) have shown that $\hat{I}$ and $\hat{J}$ are weakly consistent estimators of I and J. We deduce that $\hat{\Sigma}^{(1)}$, $\hat{\Sigma}^{(2)}$ and $\hat{\Lambda}^{-1}$ are weakly consistent estimators of $\Sigma^{(1)}$, $\Sigma^{(2)}$ and Λ^{-1}. Therefore $\Upsilon_n = nb_n\left(c + o_P(1)\right)$, with $c = \text{vech}\left(\Sigma^{(2)} - \Sigma^{(1)}\right)' \Lambda^{-1} \text{vech}\left(\Sigma^{(2)} - \Sigma^{(1)}\right)$. Since $\Sigma^{(1)} \neq \Sigma^{(2)}$ and Λ^{-1} is positive definite we have $c > 0$. Because nb_n tends to infinity, we have

$$\begin{aligned}
&\lim_{n\to\infty} P\left\{\Upsilon_n > \chi^2_{(p+q)(p+q+1)/2}(1-\alpha)\right\} \\
= &\lim_{n\to\infty} P\left\{nb_n\left(c + o_P(1)\right) > \chi^2_{(p+q)(p+q+1)/2}(1-\alpha)\right\} = 1,
\end{aligned}$$

for any $\alpha \in (0,1)$. $\square$

Proof of Lemma 1. Let κ_1 and κ_2 be constants such that

$$\begin{aligned}
&0 < \kappa_1 < \kappa_2 < 1, \quad \kappa < \kappa_1, \quad \kappa_1 + \kappa < \kappa_2 \\
&\kappa_2 > 1 - \frac{(1+r^*)(2+\nu^*)}{\nu^*}\kappa_1, \quad \kappa_2 < \frac{\nu^*}{2(1+\nu^*)}.
\end{aligned} \tag{37}$$

Figure 3 shows that these inequalities are compatible. Define sequences of integers (k_n), (m_n), (q_n) and (p_n) by

$$k_n = [n^{\kappa_2}], \quad m_n = [n^{\kappa_1}], \quad q_n = k_n - m_n, \quad p_n = \left[\frac{n}{k_n}\right].$$

Note that

$$q_n \sim k_n \sim n^{\kappa_2}, \quad m_n \sim n^{\kappa_1}, \quad p_n \sim n^{1-\kappa_2} \tag{38}$$

as $n \to \infty$. Employing a standard technique, we split the sum $S_n = \sum_{t=1}^n x_{n,t}$ into p_n alternate "big" blocks of length q_n and "small" blocks of length m_n. More precisely, write $S_n = S_n' + S_n''$, where

$$\begin{aligned}
S_n' &= \sum_{\ell=0}^{p_n-1} \xi_{n,\ell} \qquad \xi_{n,\ell} = x_{n,\ell k_n+1} + \cdots + x_{n,\ell k_n+q_n} \\
S_n'' &= \sum_{\ell=0}^{p_n} \zeta_{n,\ell} \qquad \zeta_{n,\ell} = x_{n,\ell k_n+q_n+1} + \cdots + x_{n,(\ell+1)k_n} \text{ for } \ell \leq p_n - 1 \\
&\qquad\qquad\qquad \zeta_{n,p_n} = x_{n,p_n k_n+1} + \cdots + x_{n,n}.
\end{aligned}$$

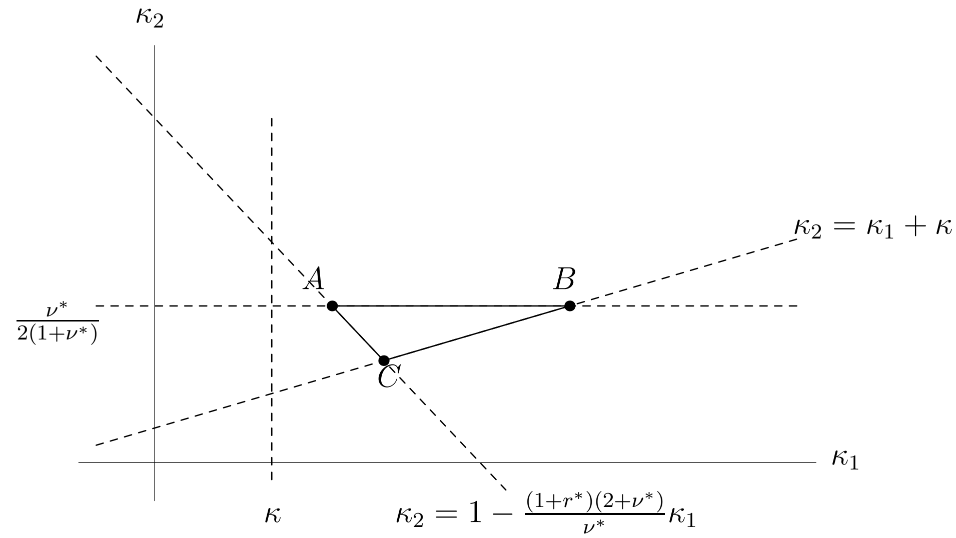

Figure 3: Inequalities (37). Let $A = \left(\frac{\nu^*}{2(1+r^*)(1+\nu^*)}, \frac{\nu^*}{2(1+\nu^*)}\right)$, $B = \left(\frac{\nu^*}{2(1+\nu^*)} - \kappa, \frac{\nu^*}{2(1+\nu^*)}\right)$, $C = \left(\frac{\nu^*(1-\kappa)}{2+2r^*+2\nu^*+r^*\nu^*}, \kappa + \frac{\nu^*(1-\kappa)}{2+2r^*+2\nu^*+r^*\nu^*}\right)$. We have $x_B := \frac{\nu^*}{2(1+\nu^*)} - \kappa > \kappa$ iff $\kappa < \frac{\nu^*}{4(1+\nu^*)}$, and $y_C := \kappa + \frac{\nu^*(1-\kappa)}{2+2r^*+2\nu^*+r^*\nu^*} < \frac{\nu^*}{2(1+\nu^*)}$ iff $r^* > \frac{2\kappa(1+\nu^*)}{\nu^*-2\kappa(1+\nu^*)}$.

To prove the CLT it suffices to show that for $n \to \infty$

i) $n^{-1}ES_n''^2 \to 0$,

ii) $E\exp(itn^{-1/2}S_n') \sim \prod_{\ell=0}^{p_n-1} E\exp(itn^{-1/2}\xi_{n,\ell})$ with $i^2 = -1$,

iii) $p_n \to \infty$ and $n^{-1}\sum_{\ell=0}^{p_n-1} E\xi_{n,\ell}^2 I_{\{|\xi_{n,\ell}| \geq n^{1/2}\epsilon\}} \to 0$ for every $\epsilon > 0$.

Indeed, i) implies that $n^{-1/2}S_n$ and $n^{-1/2}S_n'$ have the same asymptotic distribution, ii) implies that the characteristic function of $n^{-1/2}S_n'$ is asymptotically equivalent to that of $n^{-1/2}\sum_{\ell=0}^{p_n-1}\xi_{n,\ell}'$ where the $\xi_{n,\ell}'$ are independent, and distributed like $\xi_{n,\ell}$, and iii) is simply the Lindeberg condition ensuring the central limit theorem for independent, but not necessarily identically distributed, random variables.

Using the Davydov (1968) inequality, there exists an universal constant K_0 (from Rio (1993), one can take $K_0 = 4$), such that for $\nu^* < \infty$,

$$\begin{aligned}
\sum_{t=1}^{n} \sum_{s=t+q+1}^{n} |Ex_{n,t}x_{n,s}| &\leq K_0 \sum_{t=1}^{n} \sum_{s=t+q+1}^{n} \|x_{n,t}\|_{2+\nu^*} \|x_{n,s}\|_{2+\nu^*} \{\alpha_n(s-t)\}^{\frac{\nu^*}{2+\nu^*}} \\
&\leq K_0 n \sup_t \|x_{n,t}\|_{2+\nu^*}^2 \sum_{k>q} \{\alpha_n(k)\}^{\frac{\nu^*}{2+\nu^*}}
\end{aligned}$$

and

$$E\left(\sum_{t=t_0}^{t_0+m-1} x_{n,t}\right)^2 \leq K_0 \sup_t \|x_{n,t}\|_{2+\nu^*}^2 \sum_{k=0}^{m-1} (m-|k|) \{\alpha_n(k)\}^{\frac{\nu^*}{2+\nu^*}}.$$

Using inequality (1.4) in Ibragimov (1962), the same inequalities hold when $\nu^* = \infty$ (with $2+\nu^* = \infty$ and $\nu^*/(2+\nu^*) = 1$). Thus

$$\begin{aligned}
ES_n^{''2} &\leq \sum_{\ell=0}^{p_n-1} E\zeta_{n,\ell}^2 + E\zeta_{n,p_n}^2 + 2\sum_{t=1}^{n} \sum_{s=t+q_n+1}^{n} |Ex_{n,t}x_{n,s}| \\
&\leq Kp_n m_n \sum_{k=0}^{m_n-1} \{\alpha_n(k)\}^{\frac{\nu^*}{2+\nu^*}} + Kk_n \sum_{k=0}^{k_n-1} \{\alpha_n(k)\}^{\frac{\nu^*}{2+\nu^*}} + Kn \sum_{k>q_n} \{\alpha_n(k)\}^{\frac{\nu^*}{2+\nu^*}} \\
&\leq Kp_n m_n (T_n+1) + Kk_n^2 + Kn \sum_{k>q_n-T_n} \{\alpha(k)\}^{\frac{\nu^*}{2+\nu^*}}
\end{aligned}$$

for some positive constant K. Hence i) holds since, in view of (37), $p_n m_n(T_n + 1)/n \sim n^{-\kappa_2+\kappa_1+\kappa} \to 0$, $k_n^2/n \sim n^{2\kappa_2-1} \to 0$ and $q_n - T_n \sim n^{\kappa_2} \to \infty$.

Using the Ibragimov inequality (1962) and the fact that $u_n = O\left(n^{-r^*-1}\right)$ when $u_n \downarrow 0$ and $\sum_n n^{r^*} u_n < \infty$, we obtain

$$\begin{aligned}
&\left| E \prod_{\ell=0}^{p_n-1} \exp(itn^{-1/2}\xi_{n,\ell}) - \prod_{\ell=0}^{p_n-1} E\exp(itn^{-1/2}\xi_{n,\ell}) \right| \\
&\leq 4p_n\alpha_n(m_n) \leq 4p_n \frac{1}{(m_n-T_n)^{\frac{(1+r^*)(2+\nu^*)}{\nu^*}}} \sim n^{1-\kappa_2-\frac{(1+r^*)(2+\nu^*)}{\nu^*}\kappa_1} \to 0
\end{aligned}$$

which establishes ii).

When $\nu^* < \infty$, by the Hölder and Markov inequalities, we have

$$\begin{aligned} & n^{-1}\sum_{\ell=0}^{p_n-1} E\xi_{n,\ell}^2 I_{\{|\xi_{n,\ell}|\geq n^{1/2}\epsilon\}} \leq n^{-1}p_n \left\{n^{1/2}\epsilon\right\}^{-\nu^*} E\xi_{n,\ell}^{2+\nu^*} \\ \leq \;& n^{-1}p_n \left\{n^{1/2}\epsilon\right\}^{-\nu^*} (q_n)^{2+\nu^*} \sup_t \|x_{n,t}\|_{2+\nu^*}^{2+\nu^*} \sim n^{(1+\nu^*)\kappa_2-\nu^*/2} \to 0, \end{aligned}$$

which establishes iii) in the case $\nu^* < \infty$. Finally, when $M := \sup_{n\geq 1}\sup_{1\leq t\leq n}\|x_{n,t}\|_\infty < \infty$, we have

$$\|\xi_{n,\ell}\|_\infty \leq Mq_n < \sqrt{n}\epsilon,$$

and the sum in iii) equals 0, for sufficiently large n. The proof is complete.
□

Lemma 16 *Let* $(\mathbf{u}_t)$ *be any stationary centered process verifying* $\mathbf{A5}''$. *Then*

$$\left\| \sum_{i_1,i_2,i_3,i_4=1}^{n} E\left\{\mathbf{u}_t^{\otimes 4} \otimes \mathbf{u}_{t+i_1} \otimes \mathbf{u}_{t+i_2} \otimes \mathbf{u}_{t+i_3} \otimes \mathbf{u}_{t+i_4}\right\}\right\| = O(n^2).$$

Proof. We will only consider the first component of the vector inside the norm. Let u_t the first component of $\mathbf{u}_t$. The first component of the sum is bounded by $4!\sum_{k=1}^4\sum_{(k)}|Eu_t^4u_{t+i_1}u_{t+i_2}u_{t+i_3}u_{t+i_4}|$, where $\sum_{(k)}$ denotes the sum over the indices such that $i_1\leq i_2\leq i_3\leq i_4$ and $i_k - i_{k-1} = \max_{1\leq r\leq 4} i_r - i_{r-1}$ (with $i_0 = 0$).

Using the Davydov (1968) inequality, for $i_1\leq i_2\leq i_3\leq i_4$

$$\begin{aligned} \left|\mathrm{Cov}\left(u_t^4u_{t+i_1}u_{t+i_2}u_{t+i_3}, u_{t+i_4}\right)\right| &\leq K_0\|u_t^4u_{t+i_1}u_{t+i_2}u_{t+i_3}\|_{\frac{8+\nu}{7}}\|u_{t+i_4}\|_{8+\nu}\alpha_{\mathbf{u}}(|i_4-i_3|)^{\frac{\nu}{8+\nu}} \\ &\leq K_0\|u_t\|_{8+\nu}^8\alpha_{\mathbf{u}}(|i_4-i_3|)^{\frac{\nu}{8+\nu}}. \end{aligned}$$

Therefore, we have

$$\sum_{(4)} \left| E u_t^4 u_{t+i_1} u_{t+i_2} u_{t+i_3} u_{t+i_4} \right| = \sum_{(4)} \left| \mathrm{Cov}\left(u_t^4 u_{t+i_1} u_{t+i_2} u_{t+i_3}, u_{t+i_4} \right) \right|$$

$$\leq \quad n \sum_{h=0}^{n} (h+1)^2 K_0 \|u_t\|_{8+\nu}^8 \alpha_{\mathbf{u}}(h)^{\frac{\nu}{8+\nu}} = O(n). \qquad (39)$$

The last inequality has been obtained by setting $h = i_4 - i_3$, by noting that there exist at most n values for the subscript i_4, and that, when h and i_4 are fixed, it remains at most $h+1$ possibilities for i_1 and for i_2.

The term involving $\sum_{(3)}$ is bounded by

$$\sum_{(3)} \left| \mathrm{Cov}\left(u_t^4 u_{t+i_1} u_{t+i_2}, u_{t+i_3} u_{t+i_4} \right) \right| \quad + \quad \sum_{(3)} \left| E u_t^4 u_{t+i_1} u_{t+i_2} E u_{t+i_3} u_{t+i_4} \right|.$$

By the arguments used to show (39), it can be shown that the first sum is $O(n)$. The second sum is bounded by

$$n \sum_{h=0}^{n} \left\{ \sum_{\ell=0}^{h} \alpha_{\mathbf{u}}(\ell)^{\frac{\nu}{8+\nu}} \right\}^2 = O(n^2),$$

setting $h = i_3 - i_2$, arguing that there exist at most n possibilities for i_2, and that we have $|E u_t^4 u_{t+i_1} u_{t+i_2}| \leq K_0 \|u_t\|_{8+\nu}^6 \alpha_{\mathbf{u}}(\ell)^{\frac{2+\nu}{8+\nu}}$ with $\ell = i_2 - i_1 \leq h$, and $|E u_{t+i_3} u_{t+i_4}| \leq K_0 \|u_t\|_{8+\nu}^2 \alpha_{\mathbf{u}}(\ell)^{\frac{6+\nu}{8+\nu}}$ with $\ell = i_4 - i_3 \leq h$. Similarly, we have $\sum_{(2)} |E u_t^4 u_{t+i_1} u_{t+i_2} u_{t+i_3} u_{t+i_4}| = O(n^2)$.

The last term, $\sum_{(1)}$, is bounded by

$$\sum_{(1)} \left| \mathrm{Cov}\left(u_t^4, u_{t+i_1} u_{t+i_2} u_{t+i_3} u_{t+i_4} \right) \right| + \sum_{(1)} \left| E u_t^4 E u_{t+i_1} u_{t+i_2} u_{t+i_3} u_{t+i_4} \right|$$

$$\leq \quad \sum_{h=0}^{n} K_0 \|u_t\|_{8+\nu}^8 (h+1)^3 \alpha_{\mathbf{u}}(h)^{\frac{\nu}{8+\nu}} + E u_t^4 \sum_{h=0}^{n} \sum_{0 \leq i_1 \leq i_2 \leq i_3 \leq i_4 \leq h} |E u_{t+i_1} u_{t+i_2} u_{t+i_3} u_{t+i_4}|$$

$$\leq \quad O(n) + E u_t^4 \sum_{h=0}^{n} \sum_{k=1}^{3} \sum^{(k)} \left| E u_0 u_{i_2 - i_1} u_{i_3 - i_1} u_{i_4 - i_1} \right|, \qquad (40)$$

where $\sum^{(k)}$ denotes the sum over the indices such that $0 \le i_1 \le i_2 \le i_3 \le i_4 \le h$ and $i_{k+1} - i_k = \max_{1\le r\le 3} i_{r+1} - i_r$. We have

$$\begin{aligned}\sum^{(3)} |Eu_0 u_{i_2-i_1} u_{i_3-i_1} u_{i_4-i_1}| &= \sum^{(3)} |\mathrm{Cov}\,(u_0 u_{i_2-i_1} u_{i_3-i_1}, u_{i_4-i_1})| \\ &\le \sum_{\ell=0}^{h} (\ell+1)^2 K_0 \|u_t\|_{8+\nu}^4 \alpha_{\mathbf{u}}(\ell)^{\frac{4+\nu}{8+\nu}} = O(1).\end{aligned}$$

The same bound holds when $\sum^{(3)}$ is replaced by $\sum^{(1)}$. The term $\sum^{(2)}$ is equal to

$$\begin{aligned}&\sum^{(2)} |\mathrm{Cov}\,(u_0 u_{i_2-i_1}, u_{i_3-i_1} u_{i_4-i_1})| + \sum^{(2)} |Eu_0 u_{i_2-i_1} E u_{i_3-i_1} u_{i_4-i_1}| \\ \le\ & O(1) + \sum_{\ell=0}^{h} \left\{ \sum_{\ell'=0}^{\ell} \alpha_{\mathbf{u}}(\ell)^{\frac{6+\nu}{8+\nu}} \right\}^2 = O(h).\end{aligned}$$

It follows that the right-hand side of the inequality (40) is $O(n^2)$, which completes the proof. $\square$

Proof of Theorem 5. It can be shown by adapting the proof of *i)-iii)* in Lemma 9, that

$$\sqrt{nb_n}\ \mathrm{vech}\left(\hat{I} - I\right) = \sqrt{nb_n}\ \mathrm{vech}\,(I_n^{\mathbf{u}} - I) + o_P(1).$$

Write

$$\sqrt{nb_n}\ \mathrm{vech}\,(I_n^{\mathbf{u}} - I) := a_{1,n} + a_{2,n}$$

where

$$a_{1,n} = \sqrt{nb_n}\ \mathrm{vech}\,(I_n^{\mathbf{u}} - EI_n^{\mathbf{u}})\,, \qquad a_{2,n} = \sqrt{nb_n}\mathrm{vech}\,(EI_n^{\mathbf{u}} - I)\,.$$

We will show that

(i) $a_{1,n}$ converges to the normal distribution of the theorem,

(ii) $a_{2,n} = o_P(1)$.

For any $\lambda \in \mathbb{R}^{(p+q)(p+q+1)/2}$, $\lambda \neq 0$, we have $\lambda' a_{1,n} = n^{-1/2} \sum_{t=1}^n x_{n,t}$ where

$$x_{n,t} = y_{n,t} - Ey_{n,t}, \qquad y_{n,t} = \sqrt{b_n} \sum_{|i| \leq [a/b_n]} \omega(ib_n) \lambda' D_{p+q}^{+} \mathbf{u}_{t+i} \otimes \mathbf{u}_t. \tag{41}$$

To show (i) we will use Lemma 1.

We begin to show that $\sup_n \|x_{t,n}\|_4 < \infty$. We have, by the Davydov (1968) inequality

$$\|Ey_{n,t}\| \leq K\sqrt{b_n} \sum_{|i| \leq [a/b_n]} \omega(ib_n) \|\mathbf{u}_t\|_{8+\nu}^2 \left\{\alpha_{\mathbf{u}}(i)\right\}^{\frac{6+\nu}{8+\nu}} = o(1).$$

Moreover, since ω is bounded,

$$\begin{aligned} \|y_{n,t}\|_4^4 &\leq Kb_n^2 \sum_{j_1,\dots,j_8} \sum_{|i_1|,\dots,|i_4| \leq [a/b_n]} Eu_t(j_1)\dots u_t(j_4) u_{t+i_1}(j_5)\dots u_{t+i_4}(j_8) \\ &= O(1) \end{aligned}$$

where the last equality follows from Lemma 16. Hence

$$\|x_{t,n}\|_4 \leq \|y_{n,t}\|_4 + \|Ey_{n,t}\| = O(1),$$

and (12) holds with $\nu^* = 2$.

From Lemma 1 in Andrews (1991), Assumption **A5"** implies that, for all $\ell_1, \dots, \ell_4$

$$\sum_{j_1,j_2,j_3=-\infty}^{+\infty} \left|\kappa_{\ell_1,\dots,\ell_4}^{\mathbf{u}}(0, j_1, j_2, j_3)\right| < \infty \tag{42}$$

where the $\kappa_{\ell_1,\dots,\ell_4}^{\mathbf{u}}(0, j_1, j_2, j_3)$'s denote the fourth order cumulants of $(u_t(\ell_1), u_{t+j_1}(\ell_2), u_{t+j_2}(\ell_3), u_{t+j_3}(\ell_4))$. Hence, by Proposition 1 (a) of Andrews (1991), $n^{-1}\mathrm{Var}\left(\sum_{t=1}^n x_{t,n}\right)$ converges to $2\varpi^2 \lambda' D_{p+q}^{+}(I \otimes I) D_{p+q}^{+'} \lambda > 0$. Thus (13) holds.

Note that $x_{n,t}$ is a measurable function of the $\mathbf{u}_{t-[a/b_n]}, \ldots, \mathbf{u}_{t+[a/b_n]}$. Note that, by an argument already used in the proof of Lemma 1, **A5"** implies $\{\alpha_{\mathbf{u}}(k)\}^{\frac{\nu}{8+\nu}} = O(k^{-(r+1)})$, from which we deduce

$$\sum_{k=1}^{\infty} k \{\alpha_{\mathbf{u}}(k)\}^{1/2} = O\left(\sum_{k=1}^{\infty} k^{1-\frac{(r+1)(8+\nu)}{2\nu}}\right) = O(1).$$

Thus (15) holds with $T_n = 2[a/b_n]$, $\alpha(\cdot) = \alpha_{\mathbf{u}}(\cdot)$, $r^* = 1$ and $\nu^* = 2$. Clearly, $\liminf nb_n^{1/\kappa} > 0$ implies $T_n = O(n^\kappa)$. Therefore (14) holds.

Finally, Lemma 1 entails (i).

To show (ii), write

$$a_{2,n} = \sqrt{nb_n} \sum_{|i| \leq [a/b_n]} \{\omega(ib_n) - 1\} \operatorname{vech}\Delta_i - \sqrt{nb_n} \sum_{|i| > [a/b_n]} \operatorname{vech}\Delta_i.$$

Since

$$\|\Delta_i\| \leq K \|\mathbf{u}_t\|_{8+\nu}^2 \{\alpha_{\mathbf{u}}(i)\}^{\frac{6+\nu}{8+\nu}}$$

and $\{\alpha_{\mathbf{u}}(i)\}$ is a decreasing sequence, Assumption **A5"** implies that $\|\Delta_i\| = O(i^{-(r+1)(1+6/\nu)})$. Therefore

$$\sqrt{nb_n} \sum_{|i| > [a/b_n]} \|\operatorname{vech}\Delta_i\| = O(\sqrt{nb_n^{2r+1}}) = O(\sqrt{nb_n^{2r_0+1}}) = o(1).$$

Now

$$\begin{aligned} &\left\| \sqrt{nb_n} \sum_{|i| \leq [a/b_n]} \{\omega(ib_n) - 1\} \operatorname{vech}\Delta_i \right\| \\ \leq &\sum_{|i| \leq [a/b_n]} \sqrt{nb_n^{2r_0+1}} \frac{\omega(ib_n) - 1}{(ib_n)^{r_0}} i^{r_0} \|\operatorname{vech}\Delta_i\| = o(1) \end{aligned}$$

in view of the Lebesgue theorem, $\|I^{(r_0)}\| < \infty$, $\lim nb_n^{2r_0+1} = 0$ and since the function $(\omega(x) - 1)x^{-r_0}$ is bounded. Hence (ii) is shown and the proof is complete.

8 Conclusion

In this paper we have proposed a test of strong linearity in the framework of weak ARMA models. We have derived the asymptotic distribution of the test statistic under the null hypothesis and we have shown the consistency of the test. The usefulness of this test is the following. When the null assumption is not rejected, there is no evidence against standard strong ARMA models. In this case, there is no reason to think that the ARMA predictions are not optimal in the least squares sense. When the null assumption is rejected, two different strategies can be considered. A weak ARMA model can be fitted, following the lines of Francq and Zakoïan (1998, 2000), and used for optimal linear prediction. Another approach is to fit a nonlinear model to provide the (nonlinear) optimal prediction, though it can be a difficult task to determine the most appropriate nonlinear model.

Finally, we believe that the asymptotic distribution of the HAC estimator established in this paper has its own interest, apart from the proposed test. Other assumptions on the ACM could be tested, such as noninvertibility which may indicate some misspecification of the model.

FOOTNOTES

1. For ease of presentation we have not included a constant in the ARMA model. This can be done without altering the asymptotic behaviours of the estimators and test statistics introduced in the paper. The subsequent analysis apply to data that have been adjusted by subtraction of the mean.

2. When the strong mixing coefficients decrease at an exponential rate (which is the case for a large class of processes) ν can be chosen arbitrarily small in **A5** or **A5′**. Thus $E\epsilon_t^{4+2\nu} < \infty$ is a mild assumption.

3. Indeed, we have

$$\sum_i \hat{\Delta}_i(\hat{\theta}_n) = n^{-1}\left\{\sum_t e_t(\hat{\theta}_n)\frac{\partial}{\partial\theta}e_t(\hat{\theta}_n)\right\}^2 = 0.$$

4. Recent papers by Kiefer and Vogelsang (2002a, 2002b) have suggested the use of kernels with bandwidth equal to the sample size, *i.e.* $b_n = 1/n$ in our notations. It is well known that this choice of bandwidth results in inconsistent long-run variance estimators. However, Kiefer and Vogelsang have shown that, because the long-run variance matrix plays the role of a nuisance parameter, asymptotically valid tests statistics (nuisance parameter free) can be constructed based on such inconsistent estimators. In our framework it is of course crucial to have a consistent estimator of the matrix I.

5. D_m transforms vech(A) into vec(A), for any symmetric $m \times m$ matrix A.

6. As noted by Andrews (1991), for the rectangular kernel, $\omega_{(r)} = 0$ for all $r \geq 0$. For the Bartlett kernel, $\omega_{(1)} = 1$ and $\omega_{(r)} = \infty$ for all $r > 1$. For the Parzen and Tukey-Hanning kernels, $\omega_{(2)} = 6$ and $\pi^2/4$, respectively, $\omega_{(r)} = 0$ for $r < 2$, $\omega_{(r)} = \infty$ for $r > 2$.

References

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Berlinet, A. & C. Francq (1999) Estimation des covariances entre autocovariances empiriques de processus multivariés non linéaires. *La revue Canadienne de Statistique* 27, 1–22.

Billingsley, P. (1995) *Probability and Measure.* Wiley, New-York.

Brockwell, P.J. & R.A. Davis (1991) *Time Series: Theory and Methods.* Springer-Verlag, New-York.

de Jong, R.M. & J. Davidson (2000) Consistency of Kernel Estimators of Heteroskedastic and Autocorrelated Covariance Matrices. *Econometrica* 68, 407–424.

Carrasco, M. & X. Chen (2002) Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models. *Econometric Theory* 18, 17–39.

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

Francq, C. & J.M. Zakoïan (1998) Estimating linear representations of nonlinear processes. *Journal of Statistical Planning and Inference* 68, 145–165.

Francq, C. & J.M. Zakoïan (2000) Covariance matrix estimation for estimators of mixing weak ARMA models. *Journal of Statistical Planning and Inference* 83, 369–394.

Francq, C., Roy, R. & J.M. Zakoïan (2004) Goodness-of-fit Tests for ARMA Models with Uncorrelated Errors. Technical report CRM-2925, Centre de recherches mathématiques, université de Montréal.

Gallant, A.R. & H. White (1988) *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models.* New York: Basil Blackwell.

Hansen, B.E. (1992) Consistent covariance matrix estimation for dependent heterogeneous processes. *Econometrica* 60, 967–972.

Hansen, L.P. (1982) Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* 50, 1029–1054.

Harville, D.A. (1997) *Matrix Algebra From a Statistician's Perspective.* Springer-Verlag, New-York.

Herrndorf, N. (1984) A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables *The Annals of Probability* 12, 141–153.

Hong, Y. (1996) Consistent Testing for Serial Correlation of Unknown Form. *Econometrica* 64, 837–864.

Hong, Y. (1999) Testing for Serial Independence via the Empirical Characteristic Function, Working Paper, Department of Economics, Cornell University.

Ibragimov, I.A. (1962) Some Limit Theorems for Stationary Processes. *Theory Probability and its Applications* 7, 349–382.

Kiefer, N.M. & T.J. Vogelsang (2002a) Heteroskedasticity-autocorrelation Robust Testing Using Bandwidth equal to Sample Size. *Econometric Theory* 18, 1350–1366.

Kiefer, N.M. & T.J. Vogelsang (2002b) Heteroskedasticity-autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation. *Econometrica* 70, 2093–2095.

Lobato, I.N. (2002) Testing for Zero Autocorrelation in the Presence of Statistical Dependence. *Econometric Theory* 18, 730–743.

Magnus, J.R. & H. Neudecker (1988) *Matrix Differential Calculus with Application in Statistics and Econometrics.* New-York, Wiley.

May, R.M. (1976) Simple mathematical models with very complicated dynamics. *Nature* 261, 459–467.

Newey, W.K. & K.D. West (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

Parzen, E. (1957) On Consistent Estimates of the Spectrum of a Stationary Time Series. *Annals of Mathematical Statistics* 28, 329–348.

Pham, D.T. (1986) The Mixing Property of Bilinear and Generalized Random Coefficient Autoregressive Models. *Stochastic Processes and their Applications* 23, 291–300.

Phillips, P.C.B (1987) Time series regression with a unit root. *Econometrica* 55, 277-301.

Phillips, P.C.B., Sun, Y. & S. Jin (2003) Consistent HAC Estimation and Robust Regression testing Using Sharp Origin Kernels with No Truncation. *Discussion Paper*, Yale University.

Priestley, M.B. (1981) *Spectral Analysis and Time Series.* Vols. 1 and 2, Academic press, New York.

Romano, J.P. & L.A. Thombs (1996) Inference for autocorrelations under weak assumptions. Journal of the American Statistical Association 91, 590–600.

Van Der Vaart, A.W. (1998) *Asymptotic Statistics.* Cambridge University Press, Cambridge.

Withers, C.S. (1981) Central Limit Theorems for Dependent Variables. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandete Gebiete* 57, 509–534