ACADEMIC PRESS

Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE DIRECT

Journal of Multivariate Analysis 89 (2004) 148–180

Journal of Multivariate Analysis

http://www.elsevier.com/locate/jmva

On consistent testing for serial correlation of unknown form in vector time series models

Pierre Duchesne*,[1] and Roch Roy[2]

Département de mathématiques et de statistique, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Qué., Canada H3C 3J7

Received 21 December 2000

Abstract

Multivariate autoregressive models with exogenous variables (VARX) are often used in econometric applications. Many properties of the basic statistics for this class of models rely on the assumption of independent errors. Using results of Hong (Econometrica 64 (1996) 837), we propose a new test statistic for checking the hypothesis of non-correlation or independence in the Gaussian case. The test statistic is obtained by comparing the spectral density of the errors under the null hypothesis of independence with a kernel-based spectral density estimator. The asymptotic distribution of the statistic is derived under the null hypothesis. This test generalizes the portmanteau test of Hosking (J. Amer. Statist. Assoc. 75 (1980) 602). The consistency of the test is established for a general class of static regression models with autocorrelated errors. Its asymptotic slope is derived and the asymptotic relative efficiency within the class of possible kernels is also investigated. Finally, the level and power of the resulting tests are also studied by simulation.

AMS 1991 subject classifications: primary 62M10; secondary 62M15

Keywords: Vector autoregressive process; Exogenous variables; Dynamic simultaneous equation model; Kernel spectrum estimator; Diagnostic test; Portmanteau test

*Corresponding author.

E-mail addresses: duchesne@dms.umontreal.ca (P. Duchesne), royr@dms.umontreal.ca (R. Roy).

[1] This work was supported by a research grant from the Natural Science and Engineering Research Council of Canada and a grant from the Institut de Finance Mathématique de Montréal.

[2] This work was supported by grants from the Natural Science and Engineering Research Council of Canada, the Fonds FCAR (Government of Québec) and the Network of Centres of Excellence on the Mathematics of Information Technology and Complex Systems.

doi:10.1016/S0047-259X(03)00126-X

1. Introduction

Vector autoregressive with explanatory variables (VARX) models are used in many fields of study. In the econometric literature, they are also called dynamic simultaneous equation models and then, the dependent variables are said to be endogenous while the explanatory variables are called exogenous. These models generalize multivariate linear regression models in the sense that the explanatory variables may include lagged values of the endogenous variables. When there are no explanatory variables, we retrieve the popular class of vector autoregressive (VAR) models. Dictated by theoretical or empirical considerations, these models allow us to describe situations where causal relationships between stochastic economic variables may exist, that is, the present values of the dependent variables can be influenced by present and past states of the variables in the system. These models were studied by many authors and are discussed for example in Judge et al. [16], Hannan and Deistler [9], Lütkepohl [18]. A key assumption for obtaining consistent estimators of the coefficients in VARX models and for deriving their asymptotic covariance structure is the independence or at least the non-correlation of the errors, see for example Lütkepohl [18, Section 10.3] or Hannan and Deistler [9, Section 4.2].

In the univariate case, Hong [12] proposed several classes of consistent tests for checking the null hypothesis that the errors in an ARX model constitute a white noise against serial correlation of unknown form. His work is motivated by the fact that any form of serial correlation in the errors term will render the least-squares (LS) estimators inconsistent. His approach consists in comparing a residual kernel-based spectral density estimator and the spectral density of the noise under the null hypothesis, using different norms. With the quadratic norm, Hong's statistic for series of length n can be written as

$$M_{1n} = \frac{n \sum_{j=1}^{n-1} k^2(j/p_n)\hat{\rho}^2(j) - M_n(k)}{\sqrt{2V_n(k)}},$$

where $\hat{\rho}(j) = C_{\hat{u}}(j)/C_{\hat{u}}(0)$ is the residual autocorrelation at lag j and $C_{\hat{u}}(j) = n^{-1} \sum_{t=|j|+1}^{n} \hat{u}_t \hat{u}_{t-|j|}$ is the residual autocovariance at lag j. The function k is a kernel or a lag window in the spectral analysis terminology and

$$M_n(k) = \sum_{j=1}^{n-1} (1 - j/n) k^2(j/p_n), \tag{1}$$

$$V_n(k) = \sum_{j=1}^{n-2} (1 - j/n)(1 - (j+1)/n) k^4(j/p_n). \tag{2}$$

The sequence p_n is a sequence of truncation values.

Using a different approach, Paparoditis [19] considered goodness-of-fit tests for univariate time series models. The power properties of these tests are investigated in Paparoditis [20]. His test statistic relies on a distance measure between a kernel estimator of the ratio between the true and hypothesized spectral density and the expected value of the estimator under the null hypothesis. A multivariate version of this test for vector autoregressive moving average models is studied in Paparoditis [21].

The main objective of this paper is to extend Hong's approach to VARX models. Using a normalized version of the quadratic distance between two multivariate spectral densities, we introduce a kernel-based statistic for a d-dimensional process $\mathbf{y}$ that allows us to retrieve Hong's statistic M_{1n} when $d = 1$. In a static regression model, the corresponding tests are also consistent for the null hypothesis of multivariate white noise against any alternative of serial correlation of arbitrary form. With the truncated uniform or rectangular kernel, we obtain a normalized version of the multivariate portmanteau statistic for VARMA processes that generalizes the well-known Box and Pierce [3] statistic for univariate ARMA processes. The multivariate portmanteau statistic was studied by many authors, namely by Chitturi [5], Hosking [13,14] and Li and McLeod [17]. The flexible weighting of our test procedure allows us to assign different weights to the various lags. Often in practice, only the low-order autocorrelations are of interest. With an appropriate kernel, our test procedure will assign more weight to low lags and should, therefore, lead to a greater power.

The organization of the paper is as follows. In Section 2, we give some preliminaries. The new test statistic is introduced in Section 3. It is shown that its asymptotic distribution under a correctly specified VARX model is $N(0, 1)$ when the estimators of the model parameters are $\sqrt{n}$-consistent. This result contrasts strongly with the multivariate portmanteau statistic whose chi-squared asymptotic distribution depends on the estimated VARMA model. The power properties of the test are discussed in Section 4. The consistency and the asymptotic slope are studied for an arbitrary fixed alternative in a static regression model. Furthermore, the asymptotic relative efficiency in the Bahadur sense ($\mathrm{ARE_B}$) of one kernel with respect to another is also presented. Many of the currently used kernels in spectral density estimation lead to an $\mathrm{ARE_B}$ greater than one with respect to the truncated uniform kernel. In Section 5, we present the results of a small Monte Carlo experiment conducted in order to study the exact level and power of the test for finite samples and to analyse the impact of the kernel on the power. In particular, it is observed that with the considered model, Hosking's test and its normalized version defined from our statistic with the truncated uniform kernel are in general less powerful than the new statistic computed with other kernels than the uniform one. We conclude with some remarks and the appendix contains the proof of our main results.

2. Preliminaries

Let $\mathbf{y} = \{\mathbf{y}_t\colon\ t \in \mathbb{Z}\}$ and $\mathbf{x} = \{\mathbf{x}_t\colon\ t \in \mathbb{Z}\}$ be two multivariate second-order stationary processes of dimension d and m, respectively. Without loss of generality, we assume that $\mathbf{x}$ is of mean $\mathbf{0}$.

Definition 1. The process $\mathbf{y}$ is a multivariate autoregressive process with explanatory variables, noted $VARX(r,s)$, if there exists matrices $\mathbf{\Lambda}_j$ of dimension $d \times d$, $j = 0, \ldots, r$, and matrices $\mathbf{V}_j$, of dimension $d \times m$, $j = 0, \ldots, s$, such that $\mathbf{\Lambda}_r \neq \mathbf{0}$, $\mathbf{V}_s \neq \mathbf{0}$, and

$$\mathbf{\Lambda}(B)\mathbf{y}_t = \mathbf{c} + \mathbf{V}(B)\mathbf{x}_t + \mathbf{u}_t, \quad t \in \mathbb{Z}, \tag{3}$$

where $\mathbf{c}$ is the constant term, $\mathbf{\Lambda}(B) = \mathbf{\Lambda}_0 - \sum_{j=1}^{r} \mathbf{\Lambda}_j B^j$, $\mathbf{V}(B) = \sum_{j=0}^{s} \mathbf{V}_j B^j$, B being the usual backward shift operator and $\mathbf{u} = \{\mathbf{u}_t\colon t \in \mathbb{Z}\}$ a strong white noise of dimension d, where $\mathbf{u}_t = (u_t(1), \ldots, u_t(d))'$, that is the $\mathbf{u}_t$ are independent random vectors with mean $\mathbf{0}$ and regular covariance matrix $\mathbf{\Sigma}_u$. We suppose that all the roots of $\det \mathbf{\Lambda}(z)$ are outside the unit disk, where det denotes the determinant of a square matrix and z is a complex variable.

In economics, representation (3) is often called the *structural form* of the model when it represents the instantaneous and lagged effects of the endogenous variables as suggested by the economic theory. However, from a statistical point of view, representation (3) is unidentifiable without a priori information since the premultiplication of the two members of (3) by any $d \times d$ regular matrix leads to an equivalent (identical covariance structure) VARX representation of the process $\mathbf{y}$. Since $\det \mathbf{\Lambda}(0) = \det \mathbf{\Lambda}_0 \neq 0$ by assumption, we can premultiply (3) by $\mathbf{\Lambda}_0^{-1}$ and we obtain an equivalent VARX representation in which $\mathbf{\Lambda}(0) = \mathbf{\Lambda}_0 = \mathbf{I}_d$, the $d \times d$ identity matrix; it is called the *reduced form* of the model. Hereafter, we will suppose that representation (3) is in reduced form, which is more convenient for LS estimation [18, Chapter 10]. Also, predictions of future values of the endogenous variables are usually made from the reduced form [10, pp. 352–353].

The spectral density $\mathbf{f}(\omega)$ of an arbitrary second-order stationary process $\mathbf{a} = \{\mathbf{a}_t, t \in \mathbb{Z}\}$ with mean $\mathbf{0}$ is defined by

$$\mathbf{f}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \mathbf{\Gamma}_a(h) e^{-i\omega h}, \quad \omega \in [-\pi, \pi], \tag{4}$$

where $\mathbf{\Gamma}_a(j) = [\Gamma_{a,pq}(j)]_{p,q=1}^{d} = E(\mathbf{a}_t \mathbf{a}'_{t-j})$, $j \in \mathbb{Z}$, denoted the autocovariance at lag j and we assume that $\sum_{j=0}^{\infty} |\Gamma_{a,pq}(j)| < \infty$, $p, q = 1, \ldots, d$. The fourth-order moments of $\mathbf{a}$ will be denoted by $\mu_4(p,q,r,s) = E(a_t(p)a_t(q)a_t(r)a_t(s))$ and the fourth-order cumulants by $\kappa_{pqrs}(i,j,k,l) = \text{cum}(a_i(p), a_j(q), a_k(r), a_l(s))$, where $p,q,r,s = 1, \ldots, d$ and $i,j,k,l,t \in \mathbb{Z}$.

The generalized least-squares (GLS) method is popular [18, Chapter 10.3] for estimating the parameters of a VARX model. Often, there are linear constraints on the parameters, for example parameter values that are fixed to zero. Therefore, we suppose that the parameters satisfy the relation $\boldsymbol{\beta} = \text{vec}(\mathbf{\Lambda}, \mathbf{V}, \mathbf{V}_0) = \mathbf{R}\boldsymbol{\gamma}$, where $\mathbf{\Lambda} = (\mathbf{\Lambda}_1, \mathbf{\Lambda}_2, \ldots, \mathbf{\Lambda}_r)$, $\mathbf{V} = (\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_s)$, and $\mathbf{R}$ is a known matrix of linear constraints. In GLS, we first estimate $\boldsymbol{\gamma}$, say by $\hat{\boldsymbol{\gamma}}$, and $\hat{\boldsymbol{\beta}} = \mathbf{R}\hat{\boldsymbol{\gamma}}$. We make the following assumption on the estimator $\hat{\boldsymbol{\beta}}$.

Assumption A. The estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ in the VARX model satisfies $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = O_p(n^{-1/2})$.

Another method is maximum likelihood estimation (MLE), which is described in detail in Hannan and Deistler [9, Chapter 4]. Both MLE and GLS lead to $\sqrt{n}$-consistent estimators of $\boldsymbol{\beta}$.

Once a VARX model is estimated, the residual $\hat{\mathbf{u}}_t$, $t = 1, \ldots, n$, can be computed. A residual non-parametric spectral density estimator $\hat{\mathbf{f}}_n(\omega)$ is given by

$$\hat{\mathbf{f}}_n(\omega) = \frac{1}{2\pi} \sum_{j=-n+1}^{n-1} k(j/p_n)\mathbf{C}_{\hat{u}}(j)e^{-i\omega j}, \tag{5}$$

where the residual autocovariance at lag j is $\mathbf{C}_{\hat{u}}(j) = n^{-1}\sum_{t=j+1}^{n} \hat{\mathbf{u}}_t\hat{\mathbf{u}}'_{t-j}$, $j = 0, 1, \ldots, n-1$, and $\mathbf{C}_{\hat{u}}(j) = \mathbf{C}'_{\hat{u}}(-j)$ if $j = -1, \ldots, -n+1$. The function $k(\cdot)$ is a kernel or a lag window. The parameter p_n is a truncation point (smoothing parameter) when the kernel is of compact support (unbounded support). We suppose that $p_n \to \infty$ and $p_n/n \to 0$. The usual assumptions on the kernel are summarized as follows.

Assumption B. The kernel $k : \mathbb{R} \to [-1, 1]$ is a symmetric function, continuous at 0, having at most a finite number of discontinuity points, such that $k(0) = 1$ and $\int_{-\infty}^{\infty} k^2(z)\, dz < \infty$.

Using the rectangular or truncated uniform kernel $k_{\mathrm{T}}(z) = I[|z| \leqslant 1]$, where $I(A)$ is the indicator function of the set A, we retrieve the familiar truncated periodogram. Other kernels frequently used in time series analysis are given in Priestley [22, Section 6.2.3].

3. The test statistic and its asymptotic null distribution

The hypothesis of interest is that the error process $\mathbf{u}$ is, as in Definition 1, a white-noise process against the alternative of serial correlation of arbitrary form. More formally, it can be written as

$$H_0\colon \boldsymbol{\Gamma}_u(j) = \mathbf{0}, \quad \forall j \neq 0, \text{ against}$$

$$H_1\colon \boldsymbol{\Gamma}_u(j) \neq \mathbf{0}, \quad \text{for at least one } j \neq 0.$$

In terms of the spectral density $\mathbf{f}(\omega)$ of $\mathbf{u}$, H_0 can be written as $\mathbf{f}(\omega) = \mathbf{f}_0(\omega)$, $\omega \in [-\pi, \pi]$, where $\mathbf{f}_0(\omega) = \boldsymbol{\Gamma}_u(0)/(2\pi)$, $\omega \in [-\pi, \pi]$.

Our test statistic will be defined as a global distance measure between $\mathbf{f}_0$ and $\hat{\mathbf{f}}_n$. For two multivariate spectral densities $\mathbf{f}_1$ and $\mathbf{f}_2$, a distance measure between $\mathbf{f}_1$ and $\mathbf{f}_2$ such that $D(\mathbf{f}_1; \mathbf{f}_2) \geqslant 0$ and $D(\mathbf{f}_1; \mathbf{f}_2) = 0$ if and only if $\mathbf{f}_1 = \mathbf{f}_2$ is the following. In this work, the triangular inequality is not needed. For a given covariance matrix

$\mathbf{\Gamma}_u(0)$, let us consider the following normalized quadratic distance:

$$\begin{aligned} Q^2(\mathbf{f}_1;\mathbf{f}_2) &= 2\pi\int_{-\pi}^{\pi} \mathrm{vec}[\bar{\mathbf{f}}_1(\omega)-\bar{\mathbf{f}}_2(\omega)]'\mathbf{\Gamma}_u^{-1}(0)\otimes\mathbf{\Gamma}_u^{-1}(0)\,\mathrm{vec}[\mathbf{f}_1(\omega)-\mathbf{f}_2(\omega)]\,d\omega \\ &= 2\pi\int_{-\pi}^{\pi} \mathrm{tr}[(\mathbf{f}_1(\omega)-\mathbf{f}_2(\omega))^*\mathbf{\Gamma}_u^{-1}(0)(\mathbf{f}_1(\omega)-\mathbf{f}_2(\omega))\mathbf{\Gamma}_u^{-1}(0)]\,d\omega \\ &= 2\pi\int_{-\pi}^{\pi} \mathrm{tr}[\mathbf{\Gamma}_u^{-1}(0)(\mathbf{f}_1(\omega)-\mathbf{f}_2(\omega))^*\mathbf{\Gamma}_u^{-1}(0)(\mathbf{f}_1(\omega)-\mathbf{f}_2(\omega))]\,d\omega. \qquad (6)\end{aligned}$$

For a matrix $\mathbf{A}$, $\mathbf{A}^*$ denotes the transposed conjugate of $\mathbf{A}$, that is $\mathbf{A}^* = \bar{\mathbf{A}}'$. The second equality is obtained from the following result on matrix calculus [11, Theorem 16.2.2]:

$$\mathrm{tr}(\mathbf{A}'\mathbf{BCD}') = (\mathrm{vec}(\mathbf{A}))'(\mathbf{D}\otimes\mathbf{B})(\mathrm{vec}(\mathbf{C}))$$

for any matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ for which the above product is defined. For a given frequency ω,

$$Q^2_\omega(\mathbf{f}_1;\mathbf{f}_2) = \mathrm{vec}(\mathbf{f}_1(\omega)-\mathbf{f}_2(\omega))^*(\mathbf{\Gamma}_u^{-1}(0)\otimes\mathbf{\Gamma}_u^{-1}(0))\,\mathrm{vec}(\mathbf{f}_1(\omega)-\mathbf{f}_2(\omega))$$

is a normalized distance between the two matrices $\mathbf{f}_1(\omega)$ and $\mathbf{f}_2(\omega)$. The global distance $Q^2(\mathbf{f}_1;\mathbf{f}_2)$ is obtained by integrating $Q^2_\omega(\mathbf{f}_1;\mathbf{f}_2)$ over all possible frequencies in $[-\pi,\pi]$. When we compare the true spectral density $\mathbf{f}$ of $\mathbf{u}$ with the spectral density $\mathbf{f}_0$ of $\mathbf{u}$ under H_0, we get the following result.

Proposition 1. *Let $Q^2(\mathbf{f};\mathbf{f}_0)$ be the distance measure given by* (6), *where $\mathbf{f}$ is defined by* (4) *and $\mathbf{f}_0 = \mathbf{\Gamma}_u(0)/(2\pi)$. Then, we have*

$$Q^2(\mathbf{f};\mathbf{f}_0) = 2\sum_{h=1}^{\infty}\mathrm{tr}[\mathbf{\Gamma}_u(h)\mathbf{\Gamma}_u^{-1}(0)\mathbf{\Gamma}_u(h)'\mathbf{\Gamma}_u^{-1}(0)]. \qquad (7)$$

Proof. If we reapply the argument followed to obtain (6), we can write $Q^2_\omega(\mathbf{f};\mathbf{f}_0) = \mathrm{tr}[\mathbf{\Gamma}_u^{-1}(0)(\bar{\mathbf{f}}(\omega)-\bar{\mathbf{f}}_0(\omega))'\mathbf{\Gamma}_u^{-1}(0)(\mathbf{f}(\omega)-\mathbf{f}_0(\omega))]$. Since $\mathbf{\Gamma}_u(0)$ is positive definite, by the Cholesky factorization, a lower triangular matrix $\mathbf{L}$ exists such that $\mathbf{\Gamma}_u^{-1}(0) = \mathbf{LL}'$ and we have

$$\begin{aligned} Q^2_\omega(\mathbf{f};\mathbf{f}_0) &= \mathrm{tr}[(\mathbf{L}'(\mathbf{f}(\omega)-\mathbf{f}_0(\omega))\mathbf{L})(\mathbf{L}'(\mathbf{f}(\omega)-\mathbf{f}_0(\omega))\mathbf{L})^*] \\ &= \mathrm{tr}[(\mathbf{f}_\mathrm{L}-\mathbf{I}_d/(2\pi))(\mathbf{f}_\mathrm{L}-\mathbf{I}_d/(2\pi))^*],\end{aligned}$$

where $\mathbf{f}_\mathrm{L} = \mathbf{L}'\mathbf{fL}$. Integrating $Q^2_\omega(\mathbf{f};\mathbf{f}_0)$, we find that

$$\begin{aligned} Q^2(\mathbf{f};\mathbf{f}_0) &= 2\pi\int_{-\pi}^{\pi} Q^2_\omega(\mathbf{f};\mathbf{f}_0)\,d\omega = \sum_{h=-\infty}^{\infty}\mathrm{tr}[\mathbf{\Gamma}_u(h)\mathbf{\Gamma}_u^{-1}(0)\mathbf{\Gamma}_u(h)'\mathbf{\Gamma}_u^{-1}(0)] - d \\ &= 2\sum_{h=1}^{\infty}\mathrm{tr}[\mathbf{\Gamma}_u(h)\mathbf{\Gamma}_u^{-1}(0)\mathbf{\Gamma}_u(h)'\mathbf{\Gamma}_u^{-1}(0)].\end{aligned}$$

We see from (7) that Q^2 is a global measure that takes into account all lags. With similar calculations, we can show that when $\mathbf{f}$ is replaced by $\hat{\mathbf{f}}_n$, we have

$$\begin{aligned} Q^2(\hat{\mathbf{f}}_n; \mathbf{f}_0) &= \sum_{j \leqslant |n-1|} k^2(j/p_n)\, \mathrm{tr}[\mathbf{C}'_{\hat{u}}(j)\boldsymbol{\Gamma}_u^{-1}(0)\mathbf{C}_{\hat{u}}(j)\boldsymbol{\Gamma}_u^{-1}(0)] \\ &\quad - 2\,\mathrm{tr}[\boldsymbol{\Gamma}_u^{-1}(0)\mathbf{C}'_{\hat{u}}(0)] + d \\ &= 2\sum_{j=1}^{n-1} k^2(j/p_n)\, \mathrm{tr}[\mathbf{C}'_{\hat{u}}(j)\boldsymbol{\Gamma}_u^{-1}(0)\mathbf{C}_{\hat{u}}(j)\boldsymbol{\Gamma}_u^{-1}(0)] \\ &\quad + \mathrm{tr}[\mathbf{C}'_{\hat{u}}(0)\boldsymbol{\Gamma}_u^{-1}(0)\mathbf{C}_{\hat{u}}(0)\boldsymbol{\Gamma}_u^{-1}(0)] - 2\,\mathrm{tr}[\boldsymbol{\Gamma}_u^{-1}(0)\mathbf{C}'_{\hat{u}}(0)] + d. \end{aligned} \tag{8}$$

Now if we substitute $\mathbf{C}_u(0)$ for $\boldsymbol{\Gamma}_u(0)$ in (8), it follows from (A.12) and (A.14) in the appendix that

$$Q^2(\hat{\mathbf{f}}_n; \mathbf{f}_0) = 2\sum_{j=1}^{n-1} k^2(j/p_n)\, \mathrm{tr}[\mathbf{C}'_{\hat{u}}(j)\mathbf{C}_{\hat{u}}^{-1}(0)\mathbf{C}_{\hat{u}}(j)\mathbf{C}_{\hat{u}}^{-1}(0)] + o_p(\sqrt{p_n}/n). \tag{9}$$

The proposed test statistic is essentially a standardized version of $Q^2(\hat{\mathbf{f}}_n; \mathbf{f}_0)$ and is defined by

$$T_n = \frac{n\sum_{j=1}^{n-1} k^2(j/p_n)\, \mathrm{tr}[\mathbf{C}'_{\hat{u}}(j)\mathbf{C}_{\hat{u}}^{-1}(0)\mathbf{C}_{\hat{u}}(j)\mathbf{C}_{\hat{u}}^{-1}(0)] - d^2 M_n(k)}{[2d^2 V_n(k)]^{1/2}}, \tag{10}$$

where $M_n(k)$ and $V_n(k)$ are given by (1) and (2). If $p_n \to \infty$ and $p_n/n \to 0$, we can show that $p_n^{-1} M_n(k) \to M(k) = \int_0^\infty k^2(z)\,dz$ and $p_n^{-1} V_n(k) \to V(k) = \int_0^\infty k^4(z)\,dz$. Under some additional assumptions on k and/or p_n [24, p. 73], $p_n^{-1} M_n(k) = M(k) + o(p_n^{-1/2})$. In particular, $M_n(k)$ and $V_n(k)$ are both of the order $O(p_n)$. It is not difficult to see that

$$T_n = \frac{\frac{n}{2} Q^2(\hat{\mathbf{f}}_n; \mathbf{f}_0) - d^2 M_n(k)}{[2d^2 V_n(k)]^{1/2}} + o_p(1). \tag{11}$$

Also, in (10), we can replace $M_n(k)$ and $V_n(k)$ by $p_n M(k)$ and $p_n V(k)$ without modifying the asymptotic distribution. However, they may lead to better finite sample approximations.

Using k_{T} and replacing $M_n(k)$ and $V_n(k)$ by $p_n M(k) = p_n$ and $p_n V(k) = p_n$, we obtain

$$T_n = \frac{H_{p_n} - d^2 p_n}{[2d^2 p_n]^{1/2}}, \tag{12}$$

where

$$H_{p_n} = n\sum_{j=1}^{p_n} \mathrm{tr}[\mathbf{C}'_{\hat{u}}(j)\mathbf{C}_{\hat{u}}^{-1}(0)\mathbf{C}_{\hat{u}}(j)\mathbf{C}_{\hat{u}}^{-1}(0)]. \tag{13}$$

When $p_n = M$ is fixed with respect to n, H_M is Hosking's [13] multivariate version of the Box–Pierce statistic. In Hosking [14], it is shown that H_M is equivalent to the multivariate portmanteau statistics proposed by Chitturi [5] and by Li and McLeod [17]. Also, Hosking [15] described how the statistic H_M may be viewed as a Lagrange multiplier test statistic for zero constraints on VAR coefficients. Thus, T_n leads to a standardized version of H_M whose asymptotic $N(0,1)$ distribution is independent of the estimated model whilst the asymptotic chi-square distribution of H_M depends on the orders of the estimated VARMA model.

Our approach differs slightly from Hong's approach since he compared a *standardized* spectral density based on the autocorrelation function using for example a quadratic norm. In the multivariate case, we decided to work with the usual (unnormalized) multivariate spectral density (based on the matrix autocovariance function), and we compare the spectral densities using a standardized norm. It is possible to extend the univariate approach in different ways, and to define a normalized spectral density using for example the pseudo-autocorrelation functions $\{\mathbf{\Gamma}_u(k)\mathbf{\Gamma}_u^{-1}(0),\ k \in \mathbb{Z}\}$ considered in Chitturi [5], and the unstandardized quadratic norm. However, to avoid complications, we preferred to work with the usual unnormalized spectral density.

We now state our main result. The symbol $\rightarrow_{\mathrm{L}}$ stands for convergence in law.

Theorem 1. *Suppose that* $\mathbf{y}$ *is a* $\mathrm{VARX}(r,s)$ *process as in Definition* 1 *and that the fourth-order moments of* $\mathbf{u}$ *exist. Under assumptions* A, B, *and if* $p_n \rightarrow \infty$ *with* $p_n/n \rightarrow 0$, *the statistic* T_n *defined by* (10) *has an asymptotic normal distribution, that is* $T_n \rightarrow_{\mathrm{L}} N(0,1)$.

The test statistic T_n can be used to test for independent errors when $\mathbf{u}$ is Gaussian or to check for the hypothesis H_0 of no serial correlation. Note that in Theorem 1, we do not assume that the innovations are Gaussian. Also, for a multivariate white noise, the asymptotic covariance structure of the sample autocovariances involve fourth-order cumulants and several authors suppose that they are zero in order to avoid complications. Here, we do not need to assume that the fourth-order cumulants vanish, the main reason being that our proof does not make use of the asymptotic distributions of the sample autocovariances. The proof of Theorem 1 is written in two parts:

Part 1.

$$\tilde{T}_n = \frac{n\sum_{j=1}^{n-1} k^2(j/p_n)\,\mathrm{tr}[\mathbf{C}_u^{-1}(0)\mathbf{C}_u(j)\mathbf{C}_u^{-1}(0)\mathbf{C}_u'(j)] - d^2 M_n(k)}{\sqrt{2d^2 V_n(k)}} \rightarrow_{\mathrm{L}} N(0,1). \tag{14}$$

Part 2.

$$\sum_{j=1}^{n-1} k^2(j/p_n)\{\mathrm{tr}[\mathbf{C}_u^{-1}(0)\mathbf{C}_u(j)\mathbf{C}_u^{-1}(0)\mathbf{C}_u'(j)] - \mathrm{tr}[\mathbf{C}_{\hat{u}}^{-1}(0)\mathbf{C}_{\hat{u}}(j)\mathbf{C}_{\hat{u}}^{-1}(0)\mathbf{C}_{\hat{u}}'(j)]\} = o_p(\sqrt{p_n}/n). \tag{15}$$

In the first part, we establish the asymptotic normality of a version $\tilde{T}_n$ of T_n, which is based on the unobservable process $\mathbf{u}$. The VARX model does not intervene in this part since $\tilde{T}_n$ is completely defined by the innovation series $\mathbf{u}_1, \dots, \mathbf{u}_n$. The asymptotic normality is derived from the central limit theorem (CLT) for martingale differences of Brown [4]. The observed data and the model are taken into account in the second part. Since $\tilde{T}_n - T_n$ is $o_p(1)$ from Part 2, Theorem 1 follows. The detailed proof is presented in the appendix.

4. Consistency of the generalized test

We now investigate the power properties of the new test statistic T_n under fixed alternatives. We consider a fixed alternative H_{A} of serial correlation of the error $\mathbf{u}$ in the *VARX* model (3) that satisfies the following properties.

Assumption C. Let the correlation structure of the process $\mathbf{u}$ be such that $\mathbf{\Gamma}_u(j) \neq 0$ for at least one value of $j \neq 0$, $\sum_j ||\mathbf{\Gamma}_u(j)||^2 < \infty$, and that the following cumulant condition is satisfied:

$$\sum_i \sum_j \sum_l |\kappa_{pqrs}(t, t+i, t+j, t+l)| < \infty,$$

where $p, q, r, s \in \{1, \dots, d\}$.

When the process $\mathbf{u}$ is Gaussian the fourth-order cumulants are zero and the cumulant condition is trivially satisfied. Linear processes which are fourth-order stationary with absolutely summable coefficients and innovations whose fourth-order moments exist satisfy also the cumulant condition. See Hannan [8, p. 211].

For simplicity we restrict ourselves to the subclass of (3) in which there are no lagged values of the dependent variables, that is, the following static regression model:

$$\mathbf{y}_t = \mathbf{c} + \mathbf{V}(B)\mathbf{x}_t + \mathbf{u}_t, \quad t \in \mathbb{Z}. \tag{16}$$

Note that when the errors are serially correlated, the usual estimators of the coefficients in the VARX model (3) are in general inconsistent, but not in the static regression model where $\hat{\beta}$ is still consistent for β, but potentially inefficient. In this framework, we obtain the following result.

Theorem 2. *Let us consider model* (16), *let T_n be the test statistic defined by* (10) *and suppose that Assumptions* A *and* B *are satisfied. Then, under a fixed serial correlation alternative for the error process* $\mathbf{u}$ *satisfying Assumption* C, *say H_{A}, if $p_n \to \infty$ with*

$p_n/n \to 0$, we have that

$$\frac{p_n^{1/2}}{n} T_n \to_P \frac{1}{2} Q^2(\mathbf{f}; \mathbf{f}_0)/[2d^2 V(k)]^{1/2}, \tag{17}$$

where $\mathbf{f}$ *is the spectral density of* $\mathbf{u}$ *under* H_{A}.

The proof is in the appendix. This result is a multivariate version of a part of Theorem 6 in Hong [12]. For any fixed alternative H_{A} satisfying Assumption C, $Q^2(\mathbf{f}, \mathbf{f}_0) > 0$ and it follows from (17) that the statistics T_n is consistent. The rate of convergence of T_n toward infinity is given by $n/p_n^{1/2}$. Therefore, the slower p_n tends to infinity, the faster T_n will approach infinity and the resulting test should be more powerful.

For a given kernel k, let $T_n(k)$ be the corresponding T_n statistic. Since $T_n(k)$ is asymptotically normal under H_0, the concept of asymptotic slope introduced by Bahadur [2] can be used to compare two kernels k_1 and k_2 for a given alternative H_{A}. Bahadur's criterion is based on the rate at which the asymptotic p-value converges to zero as the sample size $n \to \infty$. For the test $T_n(k_i)$, the asymptotic p-value is given by $\alpha_{T_n(k_i)} = 1 - \Phi(T_n(k_i))$, where $\Phi(\cdot)$ is the cumulative distribution function of the $N(0,1)$ probability law. Now define $S_n(k_i) = -2\log[\alpha_{T_n(k_i)}]$. Using the relation $\log(1 - \Phi(a)) = -a^2/2[1 + o(1)]$ for large a as shown in Bahadur [2], we have from (17) that

$$\frac{p_n}{n^2} S_n(k_i) \to_P \frac{1}{4} Q^4(\mathbf{f}; \mathbf{f}_0)/(2d^2 V(k_i)), \tag{18}$$

under a fixed alternative as $n \to \infty$, $p_n/n \to 0$. The Bahadur's asymptotic relative efficiency ($\mathrm{ARE_B}$) of $T_n(k_1)$ with respect to $T_n(k_2)$ is by definition the limit ratio of the two sample sizes n_1 and n_2 required by the two test statistics $T_n(k_1)$ and $T_n(k_2)$, respectively, to attain the same asymptotic significance level under the alternative H_{A}. Therefore, $S_n(k_1)/S_n(k_2) \to_P 1$ as $n_1, n_2 \to \infty$ and if we take $p_{n_i} = c n_i^{\nu}$, $\nu \in (0,1)$, it is easily shown by standard arguments that

$$\mathrm{ARE_B}(T_n(k_2), T_n(k_1)) = \lim_{n_1, n_2 \to \infty} \frac{n_1}{n_2} = [V(k_1)/V(k_2)]^{1/(2-\nu)}. \tag{19}$$

For example, $\mathrm{ARE_B}(k_{\mathrm{B}}; k_{\mathrm{T}}) > 2.23$, where k_{B} and k_{T} are, respectively, the Bartlett' and the truncated uniform kernels (the different kernels used in that work are defined in Table 1 of Section 5).

Many of the most popular kernels used in spectral density estimation deliver an $\mathrm{ARE_B}$ greater than one with respect to the truncated uniform kernel. A test with a greater asymptotic slope may be expected to have a greater power for a fixed alternative than one with a smaller asymptotic slope. However, as pointed out by Geweke [7], there is no clear analytical relationship between the asymptotic slope of a test and its power function. Therefore, for a given alternative, we cannot conclude that a test with a greater asymptotic slope should be automatically preferred to one with a smaller asymptotic slope without investigating further the finite sample properties of the two statistics for the alternatives of interest. For example, with an

Table 1
Kernels used in the calculation of the test statistic T_n defined by (10)

Truncated uniform (TR):	$k(z) = \begin{cases} 1, & \lvert z\rvert \leqslant 1, \\ 0, & \text{otherwise} \end{cases}$
Bartlett (BAR):	$k(z) = \begin{cases} 1-\lvert z\rvert, & \lvert z\rvert \leqslant 1, \\ 0, & \text{otherwise} \end{cases}$
Daniell (DAN):	$k(z) = \dfrac{\sin(\pi z)}{\pi z}, \quad z \in \mathbb{R}$
Parzen (PAR):	$k(z) = \begin{cases} 1-6(\pi z/6)^2 + 6\lvert \pi z/6\rvert^3, & \lvert z\rvert \leqslant 3/\pi, \\ 2(1-\lvert \pi z/6\rvert)^3, & 3/\pi \leqslant \lvert z\rvert \leqslant 6/\pi, \\ 0, & \text{otherwise} \end{cases}$
Bartlett–Priestley (BP):	$k(z) = \dfrac{9}{5\pi^2 z^2}\left\{ \dfrac{\sin(\pi\sqrt{5/3}z)}{\pi\sqrt{5/3}z} - \cos(\pi\sqrt{5/3}z) \right\}, \quad z \in \mathbb{R}.$

alternative of the form $\mathbf{\Gamma}_u(j_0) \neq \mathbf{0}$ and $\mathbf{\Gamma}_u(j) \equiv \mathbf{0}$, $j > 0$ and $j \neq j_0$, it is likely that $T_n(k_T)$ will be more powerful than $T_n(k_B)$ with very small values of p_n since the kernel k_B might assign a too small weight to the lag j_0. However, with low-order autoregressive models, $\mathbf{\Gamma}_u(j)$ decreases rapidly to zero as $j \to \infty$, and another kernel than k_T should be preferable in such situations. This point is illustrated in Section 5.

Result (19) which was derived under the assumption that the same value of p_n is employed for the two kernels provides interesting comparisons of different kernels. However, it is easily shown that if for the first kernel we use $p_{n,1}$, while for the second one we choose $p_{n,2}$, and if these two sequences satisfy the relation $p_{n,1} = o(p_{n,2})$, then the ARE_B of the second kernel relatively to the first one will be zero, meaning that we should always prefer k_1 in such a situation. This is an additional argument suggesting that we should use a sequence p_n going to infinity at a slower rate.

5. Simulation results

In the previous sections, we have studied a new class of test statistics which have interesting asymptotic properties. However, from a practitioner point of view, it is natural to inquire for their finite sample properties, in particular their exact level and power. To partially answer that question, we have conducted a small Monte Carlo experiment. For a given bivariate VARX model described below, the new test statistics are studied empirically and compared with Hosking's [13] multivariate portmanteau statistic defined by (13) and the modified version:

$$H^*_{p_n} = n^2 \sum_{j=1}^{p_n} (n-j)^{-1} \operatorname{tr}[\mathbf{C}'_{\hat{u}}(j)\mathbf{C}_{\hat{u}}^{-1}(0)\mathbf{C}_{\hat{u}}(j)\mathbf{C}_{\hat{u}}^{-1}(0)].$$

In VARMA models, the statistic $H^*_{p_n}$ is expected to have better level properties. See Hosking [13] and Lüthepohl [18, p. 152]. Although Hosking's test was developed for VARMA models, it is tempting to use it for VARX models even if its validity is not yet established. For that reason, we included it in our simulation study. The power values obtained with the asymptotical critical points are not necessarily valid but those computed with the exact (empirical) points are correct and allows sound comparison with the new tests.

5.1. Description of the experiment

The following VARX(1,1) model was used:

$$\mathbf{y}_t = \mathbf{c} + \mathbf{\Lambda}_1 \mathbf{y}_{t-1} + \mathbf{V}_0 \mathbf{x}_t + \mathbf{V}_1 \mathbf{x}_{t-1} + \mathbf{a}_t, \tag{20}$$

$$\mathbf{x}_t = \mathbf{\Phi}_x \mathbf{x}_{t-1} + \mathbf{b}_t. \tag{21}$$

The process $\{\mathbf{b}_t\}$ is a Gaussian white noise $N_2(\mathbf{0}, \mathbf{\Sigma}_b)$. Two cases were considered for the error term $\mathbf{a}_t$: (a) $\mathbf{a}_t = \mathbf{e}_t$ and (b) $\mathbf{a}_t = \mathbf{e}_t - \mathbf{\Theta}_\delta \mathbf{e}_{t-1}$, where $\{\mathbf{e}_t\}$ is another white noise $N_2(\mathbf{0}, \mathbf{\Sigma}_e)$ independent of $\{\mathbf{b}_t\}$. The first case allowed us to study the level whilst the second one was chosen in order to study the power. The correlation structure of the $\mathbf{a}_t$'s depends on a parameter δ. The values of the parameters in (20) and (21) used in the experiment are:

$$\mathbf{c} = \begin{pmatrix} 3.0 \\ 2.0 \end{pmatrix}, \quad \mathbf{\Lambda}_1 = \begin{pmatrix} -0.5 & 0.5 \\ -1.4 & -0.2 \end{pmatrix},$$
$$\mathbf{V}_0 = \begin{pmatrix} 0.0 & 0.3 \\ 0.1 & 0.6 \end{pmatrix}, \quad \mathbf{V}_1 = \begin{pmatrix} 0.7 & 0.0 \\ 0.0 & 0.0 \end{pmatrix},$$
$$\mathbf{\Phi}_x = \begin{pmatrix} -1.5 & 1.2 \\ -0.9 & 0.5 \end{pmatrix}, \quad \mathbf{\Theta}_\delta = \begin{pmatrix} 0.18\delta & 0.04\delta \\ 0.0 & 0.02\delta \end{pmatrix},$$
$$\mathbf{\Sigma}_e = \begin{pmatrix} 1.0 & 0.75 \\ 0.75 & 1.0 \end{pmatrix}, \quad \mathbf{\Sigma}_b = \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}.$$

In the level study, 10 000 independent realizations were generated from model (20) and (21) for three series lengths ($n = 50$, 100 and 200) and the computations were made in the following way. First, the Gaussian white noise $\{\mathbf{a}_t\}$ and $\{\mathbf{b}_t\}$ were generated independently using the subroutine G05EZF from the NAG library. The initial values $\{\mathbf{x}_0\}$ and $\{\mathbf{y}_0\}$ were generated from the exact distribution of the stationary Gaussian process $\{(\mathbf{y}_t', \mathbf{x}_t')'\}$ using Ansley's [1] algorithm. Then, the values $\mathbf{x}_t, \mathbf{y}_t, t = 1, \ldots, n$, were obtained by solving the difference equations (20) and (21). For each realization, the true model (20) was estimated by generalized least squares as described in Section 2. The zero-valued parameters in $\mathbf{V}_0$ and $\mathbf{V}_1$ were taken into account by properly defining the constraint matrix $\mathbf{R}$. The residuals $\hat{\mathbf{a}}_t$, $t = 1, \ldots, n$, were obtained. With each residual series, the test statistic T_n was computed for five different kernels that are described in Table 1 at three nominal levels (1%, 5% and 10%). For each kernel, six different values of p_n were considered. We have used

$p_n = 2, 3$ and the three rates $p_n = [\log(n)]$, $p_n = [3.5n^{0.2}]$ and $p_n = [3n^{0.3}]$. Similar rates are discussed in Hong [12]. They lead, respectively, to the values $p_n = 4, 8, 10$ for $n = 50$, $p_n = 5, 9, 12$ for $n = 100$ and finally to $p_n = 6, 10, 15$ for $n = 200$. The multivariate version of Robinson's [23] cross-validation procedure for determining the bandwidth of a kernel spectrum estimator in univariate time series was also employed. Besides establishing the consistency of the procedure for non-Gaussian time series, Robinson also discusses a multivariate generalization and gives practical solutions. In our simulation, we retained for p_n the value of m that minimizes the pseudo-log-likelihood defined by $\sum_{j=1}^{n} [\log\det \hat{\mathbf{f}}^{m}_{(j)}(\lambda_j) + \mathrm{tr}\{\mathbf{I}(\lambda_j)\hat{\mathbf{f}}^{m}_{(j)}(\lambda_j)^{-1}\}]$, where $\mathbf{I}(\cdot)$ represents the periodogram, $\hat{\mathbf{f}}^{m}_{(j)}(\cdot)$ a leave-two-out-type smooth periodogram and $\lambda_j = 2\pi j/n$, $j = 1, \ldots, n$ are the Fourier frequencies. The optimization was performed for the values $m = 2, 3, \ldots, 20$. Note that we cannot use k_{T} with the cross-validation procedure, since Robinson's procedure necessitates positive definite kernels. Finally, for each series of length n, for each kernel, for each value of p_n and for each nominal level, we obtained from the 10 000 realizations the empirical frequencies of rejection of the null hypothesis of independence. The results in per cent are reported in Table 2. The standard error of the empirical levels is 0.099% for the nominal level 1%, 0.218% for 5% and 0.300% for 10%.

The power analysis was conducted in a similar way. The two main differences rely in the number of realizations (2000 rather than 10 000) and the process $\{\mathbf{a}_t\}$ is MA(1) rather than white noise. Three sets of parameter values were considered for the MA(1) model.

5.2. Discussion of the level study

Results from the level study are presented in Table 2. As expected, the approximation of the exact distribution by the asymptotic one improves in general as the series length n increases. The approximation is reasonably good at the 5% and 10% levels but the proposed test considerably over-rejects at the 1% level. That situation occurs since the finite sample distribution of the test statistic seems to be skew with a long right tail. Hosking's [13] test H and its modified version HM clearly over-reject for small p_n, and it seems that an additional adjustment is needed with models containing exogenous variables. The H test gives better size results than HM for large values of p_n. Since the new tests have good level properties at 5% and 10% levels, the rest of the discussion focuses on these nominal levels. Globally, the various kernels and truncation values lead to similar results except for TR which over-rejects slightly more when $n = 50$.

At the 5% level, all kernels (with p_n fixed) lead to rejection rates close to 7% when $n = 50$, between 5.5 and 6.3 when $n = 100$. For $n = 200$, all rejection rates are within two standard errors of 5% for $p_n = 6, 10, 15$. The cross-validation leads in general to rejection rates that are slightly higher than those obtained with the fixed values of p_n and the over-rejection tendency does not seem to decrease as n increases. At the 10% level, the rejection rates are much closer to the nominal level when $n = 50$ or 100 but the test under-rejects at $n = 200$. When $n = 100$ with fixed p_n, all kernels lead to

Table 2
Empirical levels (in percentage) of Hosking's test and of the test statistic T_n defined by (10) for different kernels, different truncation values, when the data are generated from model (20) and (21)

	$\alpha = 0.01$							$\alpha = 0.05$							$\alpha = 0.10$						
p_n	BP	BAR	DAN	PAR	TR	H	HM	BP	BAR	DAN	PAR	TR	H	HM	BP	BAR	DAN	PAR	TR	H	HM
$n = 50$																					
2	2.7	2.7	2.7	2.7	2.3	8.5	9.8	6.4	6.4	6.5	6.4	6.5	28.0	30.3	10.3	10.3	10.4	10.2	10.4	44.6	47.2
3	2.5	2.6	2.6	2.5	2.3	4.9	6.2	6.5	6.4	6.5	6.4	6.4	18.7	21.7	10.2	10.2	10.3	10.1	10.7	30.8	35.1
4	2.4	2.5	2.4	2.4	2.4	3.4	5.0	6.5	6.5	6.6	6.6	6.5	14.0	18.3	10.2	10.1	10.2	10.3	11.0	25.3	30.7
8	2.3	2.5	2.4	2.4	2.8	1.8	4.2	6.8	6.7	6.8	6.9	7.4	7.0	14.0	10.9	10.9	11.1	11.1	12.3	13.1	23.2
10	2.5	2.4	2.6	2.6	3.0	1.3	4.1	7.0	6.9	6.8	7.2	8.0	5.1	13.1	11.3	11.1	11.4	11.6	12.4	9.8	21.9
CV	2.9	3.2	2.9	2.8	NA	NA	NA	7.3	7.7	7.2	7.2	NA	NA	NA	11.5	12.2	11.6	11.5	NA	NA	NA
$n = 100$																					
2	2.3	2.3	2.3	2.4	2.2	8.4	8.9	6.0	6.0	6.0	5.9	5.8	27.1	28.2	9.7	9.7	9.7	9.7	9.5	42.6	43.8
3	2.3	2.3	2.3	2.2	2.1	5.1	5.8	5.9	5.9	5.9	5.8	5.9	18.2	19.9	9.4	9.5	9.5	9.4	9.8	30.8	32.8
5	2.2	2.2	2.2	2.1	2.1	3.2	4.0	5.7	5.7	5.7	5.7	5.7	12.1	14.4	9.6	9.4	9.6	9.6	9.8	21.3	24.3
9	2.2	2.2	2.2	2.3	2.2	2.1	3.4	5.6	5.6	5.5	5.7	6.1	7.5	11.0	9.7	9.7	9.6	9.5	10.1	14.6	19.3
12	2.2	2.2	2.3	2.2	2.2	1.4	3.2	5.8	5.7	5.9	5.9	6.3	5.7	10.6	9.9	9.7	9.9	10.1	10.5	10.9	18.0
CV	2.7	3.0	2.7	2.7	NA	NA	NA	6.7	7.5	6.6	7.1	NA	NA	NA	10.8	11.9	10.7	11.6	NA	NA	NA
$n = 200$																					
2	2.6	2.6	2.6	2.6	2.2	8.0	8.2	6.1	6.1	6.0	6.1	5.4	26.5	27.1	9.4	9.4	9.4	9.3	8.7	41.2	41.8
3	2.5	2.6	2.6	2.4	1.9	4.8	5.0	5.9	5.9	5.9	5.8	5.1	17.8	18.4	9.1	9.1	9.0	9.0	8.5	29.3	30.2
6	2.1	2.3	2.1	2.1	1.8	2.8	3.2	5.3	5.4	5.2	5.2	5.1	10.7	12.0	8.6	8.7	8.5	8.5	9.0	19.3	21.1
10	1.8	1.8	1.9	1.9	1.7	1.9	2.7	5.1	5.4	5.2	5.1	5.3	7.8	9.9	8.8	8.7	8.8	8.7	9.2	14.7	17.7
15	1.8	1.8	1.8	1.7	1.9	1.5	2.5	5.2	5.0	5.3	5.3	5.8	6.1	9.2	8.8	8.6	8.8	8.9	9.8	11.2	15.9
CV	2.9	3.4	2.8	3.0	NA	NA	NA	6.7	7.7	6.5	7.4	NA	NA	NA	10.6	11.6	10.3	11.8	NA	NA	NA

rejection rates that are within two standard errors of 10%. The cross-validation method tends to slightly over-reject at $n = 50$ but works reasonably well when $n = 100$ or 200.

5.3. Discussion of the power study

When the error term satisfies $\mathbf{a}_t = \mathbf{e}_t - \mathbf{\Theta}_\delta \mathbf{e}_{t-1}$, for $\delta \neq 0$, the errors are serially correlated. For large values of δ, the correlation is stronger and the test is more powerful. We made simulations with several values of $\delta > 0$, but we only reproduce the results for $\delta = 1.0$. We computed the power using the asymptotic critical values and the exact (empirical) critical values obtained from the level study. The results are presented in Table 3 for $n = 100, 200$ and the two nominal levels 5% and 10%. The powers based on empirical critical values are given in parentheses.

Table 3
Power based on the asymptotic and empirical (in parentheses) critical values of Hosking's test and of the test T_n for different kernels and truncation values when the data are generated from model (20) and (21) with MA(1) errors

	$\alpha = 0.05$							$\alpha = 0.10$						
p_n	BP	BAR	DAN	PAR	TR	H	HM	BP	BAR	DAN	PAR	TR	H	HM
$n = 100$														
2	51.8	52.0	51.7	51.6	37.7	74.2	74.8	60.3	60.4	60.1	60.6	47.6	84.9	85.5
	(49.5)	(49.5)	(50.0)	(49.4)	(35.1)	(35.1)	(35.0)	(60.9)	(61.1)	(60.5)	(61.1)	(48.9)	(48.9)	(48.6)
3	50.3	50.6	50.6	49.0	30.0	56.0	57.9	59.5	59.9	59.8	58.3	40.3	70.7	72.1
	(48.3)	(48.4)	(48.0)	(47.1)	(28.1)	(28.1)	(27.5)	(60.4)	(60.8)	(60.8)	(59.4)	(40.6)	(40.6)	(40.3)
5	43.4	45.0	43.6	41.6	24.2	37.1	40.1	52.6	54.7	52.6	50.5	32.8	52.8	56.2
	(41.2)	(43.3)	(41.5)	(40.0)	(22.2)	(22.2)	(21.8)	(53.6)	(55.6)	(53.8)	(51.4)	(33.1)	(33.1)	(32.6)
9	33.0	37.7	33.1	31.2	20.1	22.9	29.0	43.1	47.3	43.0	41.6	27.7	34.4	41.8
	(30.2)	(34.9)	(30.7)	(29.1)	(16.9)	(16.9)	(16.6)	(43.8)	(47.8)	(44.2)	(42.4)	(27.5)	(27.5)	(27.0)
12	28.5	32.7	28.8	27.1	17.9	16.6	24.3	37.8	43.5	37.6	36.2	25.2	25.8	38.3
	(25.9)	(31.1)	(25.7)	(24.3)	(15.2)	(15.2)	(14.7)	(37.9)	(43.8)	(37.8)	(36.1)	(24.7)	(24.7)	(23.8)
CV	52.2	53.9	51.0	49.0	NA	NA	NA	61.1	62.3	60.1	60.5	NA	NA	NA
	(48.5)	(47.5)	(47.3)	(43.4)	NA	NA	NA	(59.6)	(59.7)	(58.7)	(57.0)	NA	NA	NA
$n = 200$														
2	87.8	88.0	87.4	87.8	75.7	94.9	95.1	91.2	91.3	91.1	91.3	82.8	97.8	97.8
	(85.4)	(85.3)	(85.2)	(85.4)	(74.5)	(74.4)	(74.3)	(91.6)	(91.6)	(91.7)	(91.7)	(84.5)	(84.5)	(84.5)
3	86.6	86.8	86.6	85.7	66.0	85.8	86.4	90.9	91.2	90.9	90.2	76.0	92.7	93.1
	(85.4)	(85.5)	(85.3)	(83.9)	(65.7)	(65.7)	(65.7)	(91.7)	(91.6)	(91.8)	(91.2)	(78.2)	(78.2)	(78.1)
6	77.2	80.1	77.4	75.4	47.7	63.5	64.9	83.9	86.4	84.1	82.4	58.8	75.5	77.0
	(76.4)	(79.1)	(76.4)	(74.5)	(47.4)	(47.4)	(46.9)	(86.1)	(88.0)	(86.5)	(84.5)	(61.2)	(61.2)	(60.6)
10	66.2	71.7	66.5	63.4	37.7	43.8	47.6	74.3	79.6	74.5	72.2	46.8	58.1	61.6
	(65.7)	(70.3)	(65.6)	(63.4)	(36.6)	(36.6)	(35.9)	(76.8)	(81.7)	(76.8)	(74.5)	(49.0)	(49.0)	(47.8)
15	54.5	63.7	55.0	52.2	31.2	32.2	40.0	65.1	72.7	65.6	63.1	42.1	45.7	53.4
	(53.9)	(63.6)	(53.4)	(51.4)	(28.2)	(28.2)	(27.2)	(68.1)	(74.9)	(67.9)	(65.6)	(42.6)	(42.6)	(41.2)
CV	87.5	89.4	84.8	85.5	NA	NA	NA	90.8	92.7	89.8	91.4	NA	NA	NA
	(83.8)	(84.1)	(81.4)	(79.6)	NA	NA	NA	(90.7)	(91.4)	(89.6)	(89.7)	NA	NA	NA

Results in Table 3 show that the power seems to behave in the same manner, for all kernels, except the truncated one when p_n is fixed. Results for DAN, PAR and BP are similar. When p_n is fixed, it seems that the DAN kernel is slightly superior. More particularly, DAN and BP seem to behave in the same manner, and seem to be more powerful than PAR. This phenomenon is perceptible with both the asymptotic and empirical quantiles. We reach a different conclusion with the cross-validation procedure, where BP seems to be more powerful than DAN and PAR. The BAR kernel seems to be more powerful than the others. This is in agreement with Hong's [12] analysis for univariate ARX models.

For the new tests, the results based on empirical and asymptotic quantiles do not differ considerably at the 5% and 10% levels. That difference decreases as n increases, which is not surprising since the level is better controlled for large values of n. Since Hosking's test [13] over-rejects under the null hypothesis using the asymptotic quantiles, we have the false impression that its power is higher for low values of p_n. The results based on the empirical quantiles show that in fact Hosking's [13] test has a lower power than the proposed tests. Indeed, the test H and the new test based on k_{T} lead to the same power, based on the empirical quantiles, since they are related by a linear transformation. In our study, HM seems to be slightly less powerful than H.

Since the autocorrelation of the errors is of order one, we expect that the tests assigning more weight to small lags will be more powerful that those assigning weights to a large number of lags. This is confirmed by our study since a small value of p_n leads to a greater power. With the considered VARX model, the truncated uniform kernel is inferior in our simulation for a fixed p_n, but the difference among the kernels is rather small when $p_n = 2$. The cross-validation procedure of Robinson [23] seems to work very well here since the resulting power is higher than for fixed values of p_n that are moderately large. A very small value of p_n gives a slightly better power than the cross-validation procedure. However, without any knowledge on the alternative hypothesis, the cross-validation seems to reveal some valuable information on the shape of the spectral density, and the resulting power of T_n is quite close to the one obtained with $p_n = 2$. In practice, the analyst could not want to systematically use a very low value of p_n, since that choice might ignore important high-order autocorrelations. The cross-validation represents an objective choice and it leads to a good compromise between errors of types I and II. Finally, in our experiment, the truncated uniform kernel and Hosking's tests H and HM are the less powerful and the use of the new test based on another kernel than the truncated uniform one seems appropriate, at least for the chosen model.

6. Conclusion

In this paper, new consistent tests of serial correlation are proposed in the VARX model, when there is no information on the true alternative hypothesis. Our approach relies on a comparison between a multivariate spectral density estimator calculated with the kernel method, and the true spectral density under the null hypothesis of absence of correlation in the error term. The test generalizes the multivariate portmanteau statistic of Hosking [13], which can be viewed as a test based on the truncated uniform kernel.

In the simulation experiment, the properties of the new test were investigated for several kernels and several values of the truncation parameter p_n. Since our test procedure relies on a multivariate kernel-based spectral density estimator, we also applied the cross-validation method described in Robinson [23] for choosing p_n when the employed kernel is positive definite. For all kernels considered, the level of the test is reasonably well controlled at the nominal levels 5% and 10% with series of 100 and 200 observations. The data-driven method for choosing p_n works quite well when $n = 100$ or 200 even if it tends to over-reject slightly at the 5% nominal level. Bartlett, Daniell, Parzen and Bartlett–Priestley kernels lead to similar powers which are systematically higher that the one obtained with the truncated uniform kernel, in our experiment. Finally, the cross-validation procedure for choosing p_n works well here since the resulting power is high. That procedure provides an objective choice of the smoothing parameter which takes into account the form of the spectral density specified by the alternative hypothesis. In practical situations, the new test based on Bartlett or Daniell kernels with p_n chosen by cross-validation should be appropriate.

Acknowledgments

The authors thank the Associate Editor and two anonymous referees for their helpful comments. In particular, many constructive remarks by the referees on earlier drafts of the paper greatly enhanced the presentation.

Appendix

Proof of Theorem 1. The following notations are adopted. The scalar product of $\mathbf{x}_t, \mathbf{x}_s \in \mathbb{R}^n$ is denoted by $\langle \mathbf{x}_t, \mathbf{x}_s \rangle = \mathbf{x}_t' \mathbf{x}_s$ and the Euclidian norm of $\mathbf{x}_t$ by $||\mathbf{x}_t|| = \sqrt{\langle \mathbf{x}_t, \mathbf{x}_t \rangle}$. The Euclidian matrix norm defined by $||\mathbf{A}||_{\mathrm{E}}^2 = \mathrm{tr}(\mathbf{A}\mathbf{A}') = \sum_{i=1}^n \sum_{j=1}^n a_{ij}^2$, where $\mathbf{A} = (a_{ij})_{n \times n}$ is also used. The notations O_p and o_p are the usual notations for orders in probability. Let $k_{nj} = k(j/p_n)$, $\mathbf{v}_t = \mathbf{\Sigma}_u^{-1/2} \mathbf{u}_t$ and $\mathbf{\Sigma}_u = \mathbf{\Gamma}_u(0)$. The process $\mathbf{v} = \{\mathbf{v}_t \colon t \in \mathbb{Z}\}$ has mean $\mathbf{0}$ and variance $\mathbf{I}_d$.

We will intensively use Cauchy–Schwarz type inequalities involving the trace (tr) operator. The most useful are presented here. More details are given in Harville (Chapters 5 and 6). Let $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ be arbitrary matrices, $\mathbf{D}$ and $\mathbf{E}$ be symmetric positive definite matrices. Then we have

$$|\mathrm{tr}(\mathbf{A}\mathbf{B}')| \leqslant \sqrt{\mathrm{tr}(\mathbf{A}\mathbf{A}')} \sqrt{\mathrm{tr}(\mathbf{B}\mathbf{B}')}, \tag{A.1}$$

$$\mathrm{tr}(\mathbf{D}^2) \leqslant (\mathrm{tr}(\mathbf{D}))^2, \tag{A.2}$$

$$\mathrm{tr}(\mathbf{DE}) \leqslant \mathrm{tr}(\mathbf{D})\,\mathrm{tr}(\mathbf{E}), \tag{A.3}$$

$$\mathrm{tr}(\mathbf{AD}) \leqslant \sqrt{\mathrm{tr}(\mathbf{AA}')}\,\mathrm{tr}(\mathbf{D}), \tag{A.4}$$

$$|\mathrm{tr}(\mathbf{A}'\mathbf{BAB})| \leqslant \mathrm{tr}(\mathbf{A}'\mathbf{A})\,\mathrm{tr}(\mathbf{B}'\mathbf{B}), \tag{A.5}$$

$$\mathrm{tr}[(\mathbf{A}+\mathbf{B}+\mathbf{C})(\mathbf{A}+\mathbf{B}+\mathbf{C})'] \leqslant 4[\mathrm{tr}(\mathbf{AA}')+\mathrm{tr}(\mathbf{BB}')+\mathrm{tr}(\mathbf{CC}')]. \tag{A.6}$$

We now prove part 1. First note that $\mathbf{C}_u(0) - \boldsymbol{\Sigma}_u = O_p(n^{-1/2})$ since $E(C_{u,ij}(0) - \sigma_{ij}) = 0$ and $\mathrm{var}(C_{u,ij}(0)) = n^{-1}(\mu(i,i,j,j) - \sigma_{ij}^2)$. Then it follows that $\mathbf{C}_u^{-1}(0) - \boldsymbol{\Sigma}_u^{-1} = O_p(n^{-1/2})$. We will show that asymptotically, $\mathbf{C}_u(0)$ can be replaced by $\boldsymbol{\Sigma}_u$ in (14).

Result A.1.

$$\sum_{j=1}^{n-1} k_{nj}^2\{\mathrm{tr}[\mathbf{C}_u^{-1}(0)\mathbf{C}_u(j)\mathbf{C}_u^{-1}(0)\mathbf{C}_u'(j)] - \mathrm{tr}[\boldsymbol{\Sigma}_u^{-1}\mathbf{C}_u(j)\boldsymbol{\Sigma}_u^{-1}\mathbf{C}_u'(j)]\}$$
$$= o_p(\sqrt{p_n}/n).$$

To prove this latter result, the following lemma is needed.

Lemma A.1. $\sum_{j=1}^{n-1} k^2(j/p_n)\mathbf{C}_v(j)\mathbf{C}_v'(j) = O_p(p_n/n)$.

Proof. We have

$$\mathbf{C}_v(j)\mathbf{C}_v'(j) = n^{-2}\sum_{t=j+1}^{n} ||\mathbf{v}_{t-j}||^2\mathbf{v}_t\mathbf{v}_t' + n^{-2}\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1}\mathbf{v}_t\mathbf{v}_{t-j}'\mathbf{v}_{s-j}\mathbf{v}_s'$$
$$+ n^{-2}\sum_{t=j+1}^{n-1}\sum_{s=t+1}^{n}\mathbf{v}_t\mathbf{v}_{t-j}'\mathbf{v}_{s-j}\mathbf{v}_s'.$$

Taking expected values on both sides, it is easily seen that $E[\mathbf{C}_v(j)\mathbf{C}_v'(j)] = n^{-2}(n-j)d\mathbf{I}_d$. We have that $E[\sum_{j=1}^{n-1} k_{jn}^2\mathbf{C}_v(j)\mathbf{C}_v'(j)] = n^{-1}d\sum_{j=1}^{n-1}(1-j/n)k^2(j/p_n)\mathbf{I}_d = \mathbf{O}(p_n/n)$. Lemma A.1 follows with a judicious choice of $\mathbf{X}_n$ since for an arbitrary random matrix $\mathbf{X}_n$, $E(\mathbf{X}_n\mathbf{X}_n') = \mathbf{O}(a_n)$ implies that $\mathbf{X}_n\mathbf{X}_n' = \mathbf{O}_p(a_n)$. $\square$

To show Result A.1, note that

$$\mathbf{C}_u'(j)\mathbf{C}_u^{-1}(0)\mathbf{C}_u(j)\mathbf{C}_u^{-1}(0) = \mathbf{C}_u'(j)\boldsymbol{\Sigma}_u^{-1}\mathbf{C}_u(j)\boldsymbol{\Sigma}_u^{-1} + \mathbf{C}_u'(j)\Delta_{un}\mathbf{C}_u(j)\boldsymbol{\Sigma}_u^{-1}$$
$$+ \mathbf{C}_u'(j)\boldsymbol{\Sigma}_u^{-1}\mathbf{C}_u(j)\Delta_{un} + \mathbf{C}_u'(j)\Delta_{un}\mathbf{C}_u(j)\Delta_{un},$$

where $\mathbf{C}_u^{-1}(0) - \boldsymbol{\Sigma}_u^{-1} = \Delta_{un}$. Then it is sufficient to multiply by $k^2(j/p_n)$, to sum on j, to apply the tr operator, use (A.4) and (A.5), $p_n/n \to 0$ and Lemma A.1. $\square$

We now decompose $\sum_{j=1}^{n-1} k^2(j/p_n)\,\mathrm{tr}[\mathbf{\Sigma}_u^{-1}\mathbf{C}_u(j)\mathbf{\Sigma}_u^{-1}\mathbf{C}_u'(j)]$ into two parts A_{1n} and A_{2n}:

$$\sum_{j=1}^{n-1} k_{nj}^2\,\mathrm{tr}[\mathbf{C}_u'(j)\mathbf{\Sigma}_u^{-1}\mathbf{C}_u(j)\mathbf{\Sigma}_u^{-1}] = n^{-2}\sum_{j=1}^{n-1} k_{nj}^2\left\{\sum_{t=j+1}^{n} Z_{jt}^2\right\} + n^{-2}\sum_{j=1}^{n-2} k_{nj}^2\left\{\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1} w_{jts}\right\} = n^{-1}(A_{1n}+A_{2n}),$$

where $Z_{jt} = ||\mathbf{v}_{t-j}|| \times ||\mathbf{v}_t||$ and $w_{jts} = 2\langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j}\rangle\langle \mathbf{v}_t, \mathbf{v}_s\rangle$.

Result A.2. $p_n^{-1/2}(A_{1n} - d^2 M_n(k)) \to_P 0$.

To show Result A.2, note that $E(A_{1n}) = d^2 M_n(k)$, $\mathrm{var}(A_{1n}) = O(p_n^2/n)$, and using Lemma A.2, $\mathrm{var}(A_{1n}) \leqslant n^{-2}\{\sum_{j=1}^{n-1} k_{nj}^2 [E(\sum_{t=j+1}^{n} (Z_{jt}^2 - d^2))^2]^{1/2}\}^2 = O(p_n^2/n)$.

Lemma A.2. $E[\sum_{t=j+1}^{n} (Z_{jt}^2 - d^2)]^2 = O(n)$.

Proof. First note that $(\sum_{t=j+1}^{n} (Z_{jt}^2 - d^2))^2 = \sum_{t=j+1}^{n} (Z_{jt}^2 - d^2)^2 + 2\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1} (Z_{jt}^2 - d^2)(Z_{js}^2 - d^2)$. Then Lemma A.2 follows since $E[(Z_{jt}^2 - d^2)^2] = E(||\mathbf{v}_1||^4)^2 - d^4$, and

$$E(Z_{jt}^2 - d^2)(Z_{js}^2 - d^2) = \begin{cases} (E(||\mathbf{v}_1||^4) - d^2)d^2 & \text{if } s = t-j, \\ 0 & \text{elsewhere.} \end{cases} \quad \square$$

This shows Result A.2. $\square$

To complete the proof of Part 1, we have to show that $(2d^2 V_n(k))^{-1/2} A_{2n} \to_L N(0,1)$. To prove that result, let l_n be such that $l_n/p_n \to \infty$ and $l_n/n \to 0$. We decompose A_{2n} as $A_{2n} = B_n + \sum_{i=1}^{4} C_{in}$, where

$$B_n = n^{-1}\sum_{j=1}^{l_n} k_{nj}^2\left\{\sum_{t=2l_n+3}^{n}\sum_{s=l_n+2}^{t-l_n-1} w_{jts}\right\},$$

$$C_{1n} = n^{-1}\sum_{j=l_n+1}^{n-2} k_{nj}^2\left\{\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1} w_{jts}\right\},$$

$$C_{2n} = n^{-1}\sum_{j=1}^{l_n} k_{nj}^2\left\{\sum_{t=2l_n+3}^{n}\sum_{s=t-l_n}^{t-1} w_{jts}\right\},$$

$$C_{3n} = n^{-1} \sum_{j=1}^{l_n} k_{nj}^2 \left\{ \sum_{t=l_n+3}^{2l_n+2} \sum_{s=l_n+2}^{t-1} w_{jts} \right\},$$

$$C_{4n} = n^{-1} \sum_{j=1}^{l_n} k_{nj}^2 \left\{ \sum_{t=j+1}^{l_n+1} \sum_{s=t+1}^{n} w_{jts} \right\}.$$

The following lemma is useful. It generalizes a result in Hong [12] that we corrected very slightly since in his paper, he did not distinguish the two cases $j_1 \neq j_2$ and $j_1 = j_2$.

Lemma A.3. *Let* $w_{j_1 t_1 s_1}^{l_1 l_2 l_3 l_4} = 2v_{t_1}(l_1)v_{s_1}(l_2)v_{t_1-j_1}(l_3)v_{s_1-j_1}(l_4)$. *Then we have that*

$$E(w_{j_1 t_1 s_1}^{l_1 l_2 l_3 l_4} w_{j_2 t_2 s_2}^{m_1 m_2 m_3 m_4}) = \begin{cases} E(w_{j_1 t_1 s_1}^{l_1 l_2 l_3 l_4} w_{j_2 t_2 s_2}^{m_1 m_2 m_3 m_4})\delta_{t_1,t_2}\delta_{s_1,t_1-j_2}\delta_{s_2,t_1-j_1}, & j_1 \neq j_2, \\ E(w_{j_1 t_1 s_1}^{l_1 l_2 l_3 l_4} w_{j_2 t_2 s_2}^{m_1 m_2 m_3 m_4})\delta_{t_1,t_2}\delta_{s_1,s_2}, & j_1 = j_2. \end{cases}$$

Proof. The proof can be done case by case and is tedious but straightforward. We do not reproduce it here. □

We then show the following result.

Result A.3. $p_n^{-1/2} C_{in} = o_p(1)$, $i = 1, 2, 3, 4$.

Proof. For C_{1n}, it is sufficient to show that $E(C_{1n}^2) = o(p_n)$. Squaring C_{1n}, breaking the sum according to $j_1 = j_2$ and $j_1 \neq j_2$, taking the expected value and using Lemma A.3, we can show that

$$E(C_{1n}^2) \leqslant 4d^2 \mu_4(\|\mathbf{v}\|) \left(\sum_{j=l_n+1}^{n-2} k_{nj}^4 \right) + \frac{8d}{n} \left(\sum_{l_n+2}^{n-2} k_{nj}^2 \right)^2 = o(p_n),$$

since $p_n^{-1} \sum_{j=l_n+1}^{n-2} k_{nj}^4 \to 0$ and $p_n/n \to 0$. Similarly, $E(C_{2n}^2) = O(\frac{l_n p_n}{n} + \frac{p_n^2}{n})$, $E(C_{3n}^2) = O(\frac{l_n^2 p_n}{n^2} + \frac{p_n^2 l_n}{n^2})$ and $E(C_{4n}^2) = O(\frac{l_n p_n}{n} + \frac{p_n^2}{n})$. □

Result A.3 shows that the only important term in the asymptotic distribution of A_{2n} is B_n. The proof of the first step will be completed if we can show that $\sigma^{-2}(n) B_n \to_{\mathrm{L}} N(0,1)$, where $\sigma^2(n) = E(B_n^2)$. We will show later that $E(B_n^2) = 2d^2 p_n V(k)[1 + o(1)]$. The term B_n can be written as the following average: $B_n = n^{-1} \sum_{t=2l_n+3}^{n} B_{nt}$, where $B_{nt} = 2\mathbf{v}_t' \{ \sum_{j=1}^{l_n} k_{nj}^2 H_{j,t-l_n-1} \mathbf{v}_{t-j} \}$, and $H_{j,t-l_n-1} = \sum_{s=l_n+2}^{t-l_n-1} \mathbf{v}_s \mathbf{v}_{s-j}'$. Note that $\{B_{nt}, \mathscr{F}_{t-1}\}$ is a martingale difference since $E(B_{nt}) = 0$ and $E(B_{nt} | \mathscr{F}_{t-1}) = 0$, where $\mathscr{F}_{t-1}$ is the σ-algebra generated by $\mathbf{v}_s$, $s \leqslant t$.

Lemma A.4. $E(B_{nt}^2) = 4d^2(t - 2l_n - 2) \sum_{i=1}^{l_n} k_{ni}^4$.

Proof. $E(B_{nt}^2) = 4\,\mathrm{tr}[\sum_{i,j=1}^{l_n} k_{ni}^2 k_{nj}^2 E(\mathbf{v}_{t-j}\mathbf{v}'_{t-i})E(H'_{i,t-l_n-1}H_{j,t-l_n-1})]$ since $\mathbf{v}_{t-i}$ is independent of $H_{j,t-l_n-1}$, $1 \leqslant i,j \leqslant l_n$. The result follows by evaluating the latter sum. □

From the previous lemma, we obtain that $E(B_n^2) = n^{-2}\sum_{t=2l_n+3}^{n} E(B_{nt}^2) = 2d^2 p_n V(k)[1+o(1)]$.

Result A.4. $\sigma^{-1}(n)B_n \to_{\mathrm{L}} N(0,1)$.

Proof. To apply the CLT of Brown [4], we have to verify the following two conditions:

(a) $\sigma^{-2}(n)n^{-2}\sum_{t=2l_n+3}^{n} E(B_{nt}^2 I[|B_{nt}| > \varepsilon n\sigma(n)]) \to 0,\ \forall \varepsilon > 0$,

(b) $\sigma^{-2}(n)n^{-2}\sum_{t=2l_n+3}^{n} \ddot{B}_{nt}^2 \to_P 1$,

where $\ddot{B}_{nt}^2 = E(B_{nt}^2 | \mathscr{F}_{t-1})$.

We begin with (a). It suffices to show that Lyapounov condition is verified. We have that $|B_{nt}| \leqslant 2||\mathbf{v}_t|| \times ||\sum_{j=1}^{l_n} k_{nj}^2 H_{j,t-l_n-1}\mathbf{v}_{t-j}||$, and we obtain $E(B_{nt}^4) \leqslant 16\mu_4(||\mathbf{v}||)E(||\sum_{j=1}^{l_n} k_{nj}^2 H_{j,t-l_n-1}\mathbf{v}_{t-j}||^4)$ since $\mathbf{v}_{t-i}$ is independent of $H_{j,t-l_n-1}$, $1 \leqslant i,j \leqslant l_n$. Note that $E(||x||^4) \leqslant d\sum_{i=1}^{d} E[x^4(i)]$, where $x = (x(1), \ldots, x(d))'$ is a vector of dimension d. Since the lth component of $\sum_{j=1}^{l_n} k_{nj}^2 H_{j,t-l_n-1}\mathbf{v}_{t-j}$ is given by $\sum_{j=1}^{l_n} k_{nj}^2 \sum_{s=l_n+2}^{t-l_n-1} v_s(l)\langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j}\rangle$, we will make use of the following lemma.

Lemma A.5. $E([\sum_{j=1}^{l_n} k_{nj}^2 \sum_{s=l_n+2}^{t-l_n-1} v_s(l)\langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j}\rangle]^4) = O(p_n^2 t^2)$, *independently of* l.

Proof. First, by applying Lemma A.6 that follows to the variables $\{k_{nj}^2 \sum_{s=l_n+2}^{t-l_n-1} v_s(l)\langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j}\rangle,\ j = 1, \ldots, l_n\}$, we get

$$E\left[\sum_{j=1}^{l_n}\left(k_{nj}^2 \sum_{s=l_n+2}^{t-l_n-1} v_s(l)\langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j}\rangle\right)\right]^4 \leqslant 3\left\{\sum_{j=1}^{l_n} k_{nj}^4 \left[E\left(\sum_{s=l_n+2}^{t-l_n-1} v_s(l)\langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j}\rangle\right)^4\right]^{1/2}\right\}^2.$$

We apply a second time Lemma A.6 to the variables $\{v_s(l)\langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j}\rangle,\ s = l_n+2, \ldots, t-l_n-1\}$, and we obtain

$$E\left[\sum_{s=l_n+2}^{t-l_n-1} v_s(l)\langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j}\rangle\right]^4 \leqslant 3\left\{\sum_{s=l_n+2}^{t-l_n-1} [E(v_s(l)\langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j}\rangle)^4]^{1/2}\right\}^2.$$

Since $E(v_s(l)\langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j}\rangle)^4 \leqslant \mu_4^3(||\mathbf{v}||)$, it follows that $E[\sum_{s=l_n+2}^{t-l_n-1} v_s(l)\langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j}\rangle]^4 = O(t^2)$, independently of l. □

Regrouping the various results, we obtain that

$$E(B_{nt}^4) \leqslant 144 d\mu_4(\|\mathbf{v}\|) \sum_{l=1}^{d} \left\{ \sum_{j=1}^{l_n} k_{nj}^4 \sum_{s=l_n+2}^{t-l_n-1} (E[v_s(l)\langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j}\rangle]^4)^{1/2} \right\}^2$$
$$= O(p_n^2 t^2).$$

Then, $\sigma^{-4}(n)n^{-4}\sum_{t=2l_n+3}^{n} E(B_{nt}^4) = O(n^{-1})$, since $\sigma^2(n) = O(p_n)$, and condition (a) holds.

Lemma A.6. *Let $X_1, \ldots, X_n$ be random variables such that $E(X_i) = 0, i = 1, \ldots, n$. If $E[X_i g(X_j, X_k, X_l)] = 0$, $i \neq j, k, l$ for any function g, then $E[(\sum_{i=1}^{n} X_i)^4] \leqslant 3\{\sum_{i=1}^{n} [E(X_i^4)]^{1/2}\}^2$.*

To show (b), it is sufficient to prove that $\sigma^{-4}(n)E([\ddot{B}_n^2 - \sigma^2(n)]^2) \to 0$, where $\ddot{B}_n^2 = E(B_n^2 | \mathscr{F}_{t-1}) = n^{-2}\sum_{t=2l_n+3}^{n} \ddot{B}_{nt}^2$. We begin be writing $\ddot{B}_{nt}^2$ as $\ddot{B}_{nt}^2 = E(B_{nt}^2) + 4\sum_{i=1}^{4} D_{int}$, where

$$D_{1nt} = 2\sum_{j=2}^{l_n}\sum_{i=1}^{j-1} k_{ni}^2 k_{nj}^2 \mathbf{v}_{t-i}' H_{i,t-l_n-1}' H_{j,t-l_n-1} \mathbf{v}_{t-j},$$

$$D_{2nt} = 2\sum_{i=1}^{l_n} k_{ni}^4 \sum_{s_1=l_n+3}^{t-l_n-1} \sum_{s_2=l_n+2}^{s_1-1} \mathbf{v}_{t-i}'\mathbf{v}_{s_1-i}\mathbf{v}_{s_1}'\mathbf{v}_{s_2}\mathbf{v}_{s_2-i}'\mathbf{v}_{t-i},$$

$$D_{3nt} = \sum_{i=1}^{l_n} k_{ni}^4 \mathbf{v}_{t-i}' \left[\sum_{s=l_n+2}^{t-l_n-1} (\mathbf{v}_{s-i}\mathbf{v}_s'\mathbf{v}_s\mathbf{v}_{s-i}' - dI_d) \right] \mathbf{v}_{t-i},$$

$$D_{4nt} = d(t - 2l_n - 2)\sum_{i=1}^{l_n} k_{ni}^4 (\mathbf{v}_{t-i}'\mathbf{v}_{t-i} - d).$$

We now prove the two following lemmas.

Lemma A.7. $E(D_{1nt}^2) = O(t^2 p_n^2)$, $E(D_{2nt}^2) = O(t^2 p_n + t p_n^2)$, $E(D_{3nt}^2) = O(t p_n^2)$, $E(D_{4nt}^2) = O(t^2 p_n)$.

Proof. For D_{1nt}, we have that $E(D_{1nt}^2) \leqslant 4\sum_{l_1,l_2=1}^{d}\sum_{j=2}^{l_n}\sum_{i=1}^{j-1} k_{ni}^4 k_{nj}^4 \{E[a_i^4(l_1)] E[a_j^4(l_2)]\}^{1/2}$, using Cauchy–Schwarz inequality, where $\mathbf{a}_j = H_{j,t-l_n-1}\mathbf{v}_{t-j}$ and $a_i(l) = \sum_{s=l_n+2}^{t-l_n-1} v_s(l)\langle \mathbf{v}_{s-j}, \mathbf{v}_{t-j}\rangle$. Since $E[a_i^4(l)] = O(t^2)$, then $E(D_{1nt}^2) = O(t^2 p_n^2)$. For D_{2nt}, first note that

$$D_{2nt} = \sum_{l_1,l_2=1}^{d}\sum_{i=1}^{l_n} k_{ni}^4 b_{t-i}^{l_1 l_2} \sum_{s=l_n+3}^{t-l_n-1}\sum_{r=l_n+2}^{s-1} w_{isr}^{l_1 l_2},$$

where $b_t^{l_1 l_2} = \mathbf{v}_t(l_1)\mathbf{v}_t(l_2)$. Thus, we have

$$\begin{aligned}
D_{2nt}^2 &= \sum_{l_1,l_2,m_1,m_2=1}^{d} \sum_{i=1}^{l_n} k_{ni}^8 b_{t-i}^{l_1 l_2} b_{t-i}^{m_1 m_2} \sum_{s_1,s_2=l_n+3}^{t-l_n-1} \sum_{r_1=l_n+2}^{s_1-1} \sum_{r_2=l_n+2}^{s_2-1} w_{is_1r_1}^{l_1 l_2} w_{is_2r_2}^{m_1 m_2} \\
&\quad + 2 \sum_{l_1,l_2,m_1,m_2=1}^{d} \sum_{i_1=2}^{l_n} \sum_{i_2=1}^{i_1-1} k_{ni_1}^4 k_{ni_2}^4 b_{t-i_1}^{l_1 l_2} b_{t-i_2}^{m_1 m_2} \\
&\quad \times \sum_{s_1,s_2=l_n+3}^{t-l_n-1} \sum_{r_1=l_n+2}^{s_1-1} \sum_{r_2=l_n+2}^{s_2-1} w_{i_1 s_1 r_1}^{l_1 l_2} w_{i_2 s_2 r_2}^{m_1 m_2} \\
&= D_{21nt} + D_{22nt}.
\end{aligned}$$

On taking the expected value of D_{21nt} and using Lemma A.3, we show that $E(D_{21nt}) = O(p_n t^2)$. Similarly, we can show that $E(D_{22nt}) = O(tp_n^2)$, and the result for D_{2nt} follows. For D_{3nt}, let us note that

$$\begin{aligned}
E(D_{3nt}^2) &\leqslant \left\{ \sum_{i=1}^{l_n} k_{ni}^4 \left[E\left(\mathbf{v}_{t-i}' \left(\sum_{s=l_n+2}^{t-l_n-1} \mathbf{v}_{s-i}\mathbf{v}_s'\mathbf{v}_s\mathbf{v}_{s-i}' - d\mathbf{I}_d \right) \mathbf{v}_{t-i} \right)^2 \right]^{1/2} \right\}^2 \\
&\leqslant \mu_4(||\mathbf{v}||) \left\{ \sum_{i=1}^{l_n} k_{ni}^4 \left[E\left(\left|\left| \sum_{s=l_n+2}^{t-l_n-1} \mathbf{v}_{s-i}\mathbf{v}_s'\mathbf{v}_s\mathbf{v}_{s-i}' - d\mathbf{I}_d \right|\right|_{\mathrm{E}}^2 \right) \right]^{1/2} \right\}^2 \\
&= O(tp_n^2)
\end{aligned}$$

using Lemma A.8 that follows.

Lemma A.8. $E(||\sum_{s=l_n+2}^{t-l_n-1} (\mathbf{v}_{s-i}\mathbf{v}_s'\mathbf{v}_s\mathbf{v}_{s-i}' - d\mathbf{I}_d)||_{\mathrm{E}}^2) = O(t)$.

Proof. Let $c_{si}^{lm} = \mathbf{v}_s'\mathbf{v}_s\mathbf{v}_{s-i}(l)\mathbf{v}_{s-i}(m) - d\delta_{lm}$. We have that

$$\begin{aligned}
&E\left(\left|\left| \sum_{s=l_n+2}^{t-l_n-1} (\mathbf{v}_{s-i}\mathbf{v}_s'\mathbf{v}_s\mathbf{v}_{s-i}' - d\mathbf{I}_d) \right|\right|_{\mathrm{E}}^2 \right) \\
&\quad = E\left[\sum_{l,m=1}^{d} \left\{ \sum_{s=l_n+2}^{t-l_n-1} (c_{si}^{lm})^2 + 2 \sum_{s=l_n+3}^{t-l_n-1} \sum_{r=l_n+2}^{s-1} (c_{si}^{lm} c_{ri}^{lm}) \right\} \right] \\
&\quad = O(t),
\end{aligned}$$

since

$$E(c_{si}^{lm})^2 = \begin{cases} \mu_4(||\mathbf{v}||)\mu_4(l,l,l,l) - d^2 & \text{if } l = m, \\ \mu_4(||\mathbf{v}||)\mu_4(l,l,m,m) & \text{if } l \neq m, \end{cases}$$

$$E(c_{si}^{lm} c_{ri}^{lm}) = \begin{cases} d[E(||\mathbf{v}_1||^2 v_1(l)^2) - d] & \text{if } l = m,\ r = s - i, \\ 0 & \text{elsewhere.} \quad \square \end{cases}$$

Finally, we show the result for D_{4nt}. Let us note that $E(D_{4nt}^2) = d^2(t-2l_n-2)^2\sum_{i=1}^{l_n} k_{ni}^8 E(\mathbf{v}_{t-i}'\mathbf{v}_{t-i}-d)^2 = O(t^2 p_n)$, since $E(\mathbf{v}_{t-i}'\mathbf{v}_{t-i}-d)^2 = \mu_4(||\mathbf{v}||)-d^2$. $\square$

The following lemma that is easy to prove will be useful in the sequel.

Lemma A.9. $E(D_{1nt}D_{1ns}) = 0,\ t-s > l_n$.

Then we have that $\ddot{B}_n^2 = \sigma^2(n) + 4n^{-2}\sum_{j=1}^4\sum_{t=2l_n+3}^n D_{jnt}$ and the validity of condition (b) will be established once it is shown that $E[p_n^{-2}(n^{-2}\sum_{t=2l_n+3}^n D_{jnt})^2] \to 0$, $j=1,2,3,4$. This latter result can be obtained with a reasoning similar to the one made by Hong [12] for deriving his formulas (A.7)–(A.10). Using Brown's theorem, the proof of the first part is completed. $\square$

We now show the second part. To reduce the length of the proof, we restrict ourselves to the following model:

$$\mathbf{y}_t = \mathbf{c} + \mathbf{\Lambda}_1\mathbf{y}_{t-1} + \mathbf{V}_0\mathbf{x}_t + \mathbf{u}_t. \tag{A.7}$$

The proof for the general model (3) is in all points similar, except that the algebraic developments are heavier.

First, we decompose

$$\sum_{j=1}^{n-1} k_{nj}^2(\mathrm{tr}[\mathbf{C}_{\hat{v}}(j)\mathbf{C}_{\hat{v}}'(j)] - \mathrm{tr}[\mathbf{C}_v(j)\mathbf{C}_v'(j)]). \tag{A.8}$$

Since $\mathrm{tr}(\mathbf{AA}') - \mathrm{tr}(\mathbf{BB}') = \mathrm{tr}[(\mathbf{A}-\mathbf{B})(\mathbf{A}-\mathbf{B})'] + 2\,\mathrm{tr}[\mathbf{B}(\mathbf{A}-\mathbf{B})']$, it suffices to show the two following results.

Result A.5. $\sum_{j=1}^{n-1} k_{nj}^2\,\mathrm{tr}[(\mathbf{C}_{\hat{v}}(j)-\mathbf{C}_v(j))(\mathbf{C}_{\hat{v}}(j)-\mathbf{C}_v(j))'] = O_p(n^{-1})$.

Result A.6. $\sum_{j=1}^{n-1} k_{nj}^2\,\mathrm{tr}[\mathbf{C}_v(j)(\mathbf{C}_{\hat{v}}(j)-\mathbf{C}_v(j))'] = o_p(\sqrt{p_n}/n)$.

Let $\hat{\boldsymbol{\lambda}}_{nt} = (\hat{\mathbf{c}}-\mathbf{c}) + (\hat{\mathbf{\Lambda}}_1-\mathbf{\Lambda}_1)\mathbf{y}_{t-1} + (\hat{\mathbf{V}}_0-\mathbf{V}_0)\mathbf{x}_t$ and $\hat{\boldsymbol{\gamma}}_{nt} = \mathbf{\Sigma}_u^{-1/2}\hat{\boldsymbol{\lambda}}_{nt}$. Let also $\hat{\mathbf{u}}_t = \mathbf{u}_t - \hat{\boldsymbol{\lambda}}_{nt}$, and $\hat{\mathbf{v}}_t = \mathbf{v}_t - \hat{\boldsymbol{\gamma}}_{nt}$. First we prove result Result A.5. We can write

$$\mathbf{C}_{\hat{v}}(j) - \mathbf{C}_v(j) = -n^{-1}\sum_{t=j+1}^n \hat{\boldsymbol{\gamma}}_{nt}\mathbf{v}_{t-j}' - n^{-1}\sum_{t=j+1}^n \mathbf{v}_t\hat{\boldsymbol{\gamma}}_{n,t-j}' + n^{-1}\sum_{t=j+1}^n \hat{\boldsymbol{\gamma}}_{nt}\hat{\boldsymbol{\gamma}}_{n,t-j}'. \tag{A.9}$$

Using (A.6), we have that

$$\sum_{j=1}^{n-1} k_{nj}^2\,\mathrm{tr}[(\mathbf{C}_{\hat{v}}(j)-\mathbf{C}_v(j))(\mathbf{C}_{\hat{v}}(j)-\mathbf{C}_v(j))'] \leqslant 4(E_{1n}+E_{2n}+E_{3n}), \tag{A.10}$$

where

$$E_{1n} = n^{-2} \sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}\left[\left(\sum_{t=j+1}^{n} \hat{\gamma}_{nt} \mathbf{v}'_{t-j} \right) \left(\sum_{t=j+1}^{n} \hat{\gamma}_{nt} \mathbf{v}'_{t-j} \right)' \right],$$

$$E_{2n} = n^{-2} \sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}\left[\left(\sum_{t=j+1}^{n} \mathbf{v}_t \hat{\gamma}'_{n,t-j} \right) \left(\sum_{t=j+1}^{n} \mathbf{v}_t \hat{\gamma}'_{n,t-j} \right)' \right],$$

$$E_{3n} = n^{-2} \sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}\left[\left(\sum_{t=j+1}^{n} \hat{\gamma}_{nt} \hat{\gamma}'_{n,t-j} \right) \left(\sum_{t=j+1}^{n} \hat{\gamma}_{nt} \hat{\gamma}'_{n,t-j} \right)' \right].$$

Then we show the following result:

Result A.7. $E_{jn} = O_p(n^{-1}),\ j = 1, 2, 3.$

Proof. Let us begin with E_{1n} that we bound in the following manner using (A.3) and (A.6):

$$\begin{aligned} E_{1n} \leqslant{} & 4\,\mathrm{tr}[(\hat{\mathbf{c}} - \mathbf{c})' \boldsymbol{\Sigma}_u^{-1} (\hat{\mathbf{c}} - \mathbf{c})] F_{1n} + 4\,\mathrm{tr}[(\hat{\boldsymbol{\Lambda}}_1 - \boldsymbol{\Lambda}_1)' \boldsymbol{\Sigma}_u^{-1} (\hat{\boldsymbol{\Lambda}}_1 - \boldsymbol{\Lambda}_1)] F_{2n} \\ & + 4\,\mathrm{tr}[(\hat{\mathbf{c}} - \mathbf{c})' \boldsymbol{\Sigma}_u^{-1} (\hat{\mathbf{c}} - \mathbf{c})] F_{3n}, \end{aligned}$$

where

$$F_{1n} = \sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}\left[\left(n^{-1} \sum_{t=j+1}^{n} \mathbf{v}'_{t-j} \right) \left(n^{-1} \sum_{t=j+1}^{n} \mathbf{v}'_{t-j} \right)' \right],$$

$$F_{2n} = \sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}\left[\left(n^{-1} \sum_{t=j+1}^{n} \mathbf{y}_{t-1} \mathbf{v}'_{t-j} \right) \left(n^{-1} \sum_{t=j+1}^{n} \mathbf{y}_{t-1} \mathbf{v}'_{t-j} \right)' \right],$$

$$F_{3n} = \sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}\left[\left(n^{-1} \sum_{t=j+1}^{n} \mathbf{x}_t \mathbf{v}'_{t-j} \right) \left(n^{-1} \sum_{t=j+1}^{n} \mathbf{x}_t \mathbf{v}'_{t-j} \right)' \right].$$

The result for E_{1n} is based on the following lemma.

Lemma A.10. $F_{1n} = O_p(p_n/n)$, $F_{2n} = O_p(1)$, $F_{3n} = O_p(p_n/n)$.

Proof. The result for F_{1n} is immediate noting that $E(|F_{1n}|) = dn^{-2} \sum_{j=1}^{n-1} (n-j) k_{nj}^2 = O(p_n/n)$. To show the result for F_{2n}, we write the model (A.7) as $\mathbf{y}_t = \mathbf{c}_0 + \boldsymbol{\Psi}(B) \mathbf{V}_0 \mathbf{x}_t + \boldsymbol{\Psi}(B) \mathbf{u}_t$, where $\mathbf{c}_0 = (\mathbf{I}_d - \boldsymbol{\Lambda}_1)^{-1} \mathbf{c}$, $\boldsymbol{\Psi}(B) = (\mathbf{I}_d - \boldsymbol{\Lambda}_1 B)^{-1} = \sum_{j \geqslant 0} \boldsymbol{\Lambda}_1^j B^j$, with $||\mathbf{\Lambda}_1|| < 1$. We have that

$$n^{-1} \sum_{t=j+1}^{n} \mathbf{y}_{t-1}\mathbf{v}'_{t-j} = n^{-1} \sum_{t=j+1}^{n} \mathbf{c}_0\mathbf{v}'_{t-j} + n^{-1} \sum_{t=j+1}^{n} (\mathbf{\Psi}(B)\mathbf{V}_0\mathbf{x}_{t-1})\mathbf{v}'_{t-j} + n^{-1} \sum_{t=j+1}^{n} (\mathbf{\Psi}(B)\mathbf{u}_{t-1})\mathbf{v}'_{t-j},$$

and, therefore, $F_{2n} \leqslant 4(G_{1n} + G_{2n} + G_{3n})$, where

$$G_{1n} = \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}\left[\left(n^{-1} \sum_{t=j+1}^{n} \mathbf{c}_0\mathbf{v}'_{t-j}\right)\left(n^{-1} \sum_{t=j+1}^{n} \mathbf{c}_0\mathbf{v}'_{t-j}\right)'\right],$$

$$G_{2n} = \sum_{j=1}^{n-1} k_{nj}^2 \times \operatorname{tr}\left[\left(n^{-1} \sum_{t=j+1}^{n} (\mathbf{\Psi}(B)\mathbf{V}_0\mathbf{x}_{t-1})\mathbf{v}'_{t-j}\right)\left(n^{-1} \sum_{t=j+1}^{n} (\mathbf{\Psi}(B)\mathbf{V}_0\mathbf{x}_{t-1})\mathbf{v}'_{t-j}\right)'\right],$$

$$G_{3n} = \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}\left[\left(n^{-1} \sum_{t=j+1}^{n} (\mathbf{\Psi}(B)\mathbf{u}_{t-1})\mathbf{v}'_{t-j}\right)\left(n^{-1} \sum_{t=j+1}^{n} (\mathbf{\Psi}(B)\mathbf{u}_{t-1})\mathbf{v}'_{t-j}\right)'\right].$$

We note that $G_{1n} = O_p(p_n/n)$, since $\mathbf{c}_0 \sum_{s,t=j+1}^{n} \mathbf{v}'_{t-j}\mathbf{v}_{s-j}\mathbf{c}'_0 = O_p(n)$. With G_{2n}, we have that

$$G_{2n} = n^{-2} \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}\left\{ \sum_{t=j+1}^{n} ||\mathbf{v}_{s-j}||^2((\mathbf{\Psi}(B)\mathbf{V}_0\mathbf{x}_{t-1})\mathbf{v}'_{t-j})((\mathbf{\Psi}(B)\mathbf{V}_0\mathbf{x}_{t-1})\mathbf{v}'_{t-j})' + 2 \sum_{s=j+1}^{n} \sum_{j+2}^{t-1} \mathbf{v}'_{t-j}\mathbf{v}_{s-j}((\mathbf{\Psi}(B)\mathbf{V}_0\mathbf{x}_{t-1})\mathbf{v}'_{t-j})((\mathbf{\Psi}(B)\mathbf{V}_0\mathbf{x}_{s-1})\mathbf{v}'_{s-j})' \right\}.$$

It follows that $E(|G_{2n}|) = O(p_n/n)$ and $G_{2n} = O_p(p_n/n)$. The proof for G_{3n} is based on the following lemma, that generalizes Lemma A.1 of Hong [12].

Lemma A.11.

$$E\left\{\operatorname{tr}\left[\left(\sum_{t=j+1}^{n} (\mathbf{\Psi}(B)\mathbf{u}_{t-1})\mathbf{v}'_{t-j}\right)\left(\sum_{t=j+1}^{n} (\mathbf{\Psi}(B)\mathbf{u}_{t-1})\mathbf{v}'_{t-j}\right)'\right]\right\} \leqslant \Delta_1 n + \Delta_2 n^2 ||\mathbf{\Lambda}_1||_{\mathrm{E}}^{2(j-1)}.$$

Proof. We note that

$$\left(\sum_{t=j+1}^{n} (\mathbf{\Psi}(B)\mathbf{u}_{t-1})\mathbf{v}'_{t-j}\right)\left(\sum_{t=j+1}^{n} (\mathbf{\Psi}(B)\mathbf{u}_{t-1})\mathbf{v}'_{t-j}\right)'$$
$$= \sum_{s=j+1}^{n} ||\mathbf{v}_{s-j}||^2||\mathbf{\Psi}(B)\mathbf{u}_{s-1}||^2$$
$$+\ 2\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1} \langle \mathbf{v}_{t-j}, \mathbf{v}_{s-j}\rangle\langle \mathbf{\Psi}(B)\mathbf{u}_{t-1}, \mathbf{\Psi}(B)\mathbf{u}_{s-1}\rangle.$$

We note that

$$\langle \mathbf{\Psi}(B)\mathbf{u}_{t-1}, \mathbf{\Psi}(B)\mathbf{u}_{s-1}\rangle = \sum_{j_1\geqslant 0} \mathbf{u}'_{t-j_1-1}(\mathbf{\Lambda}_1^{j_1})'(\mathbf{\Lambda}_1^{j_1})\mathbf{u}_{s-j_1-1}$$
$$+ \sum_{j_1\neq j_2} \mathbf{u}'_{t-j_1-1}(\mathbf{\Lambda}_1^{j_1})'(\mathbf{\Lambda}_1^{j_2})\mathbf{u}_{s-j_2-1}.$$

Let us consider the first term. We have that

$$E(||\mathbf{v}_{t-j}||^2||\mathbf{\Psi}(B)\mathbf{u}_{t-1}||^2) \leqslant \sum_{j_1\geqslant 0} E(||\mathbf{v}_{t-j}||^2||\mathbf{u}_{t-j_1-1}||^2)||(\mathbf{\Lambda}_1^{j_1})'(\mathbf{\Lambda}_1^{j_1})||_{\mathrm{E}}$$
$$\leqslant \mu_4(||\mathbf{v}||)||\mathbf{\Sigma}_u|| \sum_{j\geqslant 0} ||\mathbf{\Lambda}_1||_{\mathrm{E}}^{2j} \leqslant \Delta_1.$$

Finally, we have that

$$E(\mathbf{v}'_{t-j}\mathbf{v}_{s-j}\langle \mathbf{\Psi}(B)\mathbf{u}_{t-1}, \mathbf{\Psi}(B)\mathbf{u}_{s-1}\rangle)$$
$$= \sum_{j_1\geqslant 0} E[\mathbf{v}'_{t-j}\mathbf{v}_{s-j}\mathbf{u}'_{t-j_1-1}(\mathbf{\Lambda}_1^{j_1})'(\mathbf{\Lambda}_1^{j_1})\mathbf{u}_{s-j_1-1}]$$
$$+ \sum_{j_1\neq j_2} E[\mathbf{v}'_{t-j}\mathbf{v}_{s-j}\mathbf{u}'_{t-j_1-1}(\mathbf{\Lambda}_1^{j_1})'(\mathbf{\Lambda}_1^{j_2})\mathbf{u}_{s-j_2-1}]$$
$$\leqslant 2||\mathbf{\Lambda}_1||_{\mathrm{E}}^{2(j-1)}\,\mathrm{tr}(\mathbf{\Sigma}_u).$$

Regrouping the results, we obtain

$$E\left\{\mathrm{tr}\left[\left(n^{-1}\sum_{t=j+1}^{n} (\mathbf{\Psi}(B)\mathbf{u}_{t-1})\mathbf{v}'_{t-j}\right)\left(n^{-1}\sum_{t=j+1}^{n} (\mathbf{\Psi}(B)\mathbf{u}_{t-1})\mathbf{v}'_{t-j}\right)'\right]\right\}$$
$$\leqslant n^{-1}\Delta_1 + \Delta_2||\mathbf{\Lambda}_1||_{\mathrm{E}}^{2(j-1)}$$

and the proof of Lemma A.11 is completed. □

Note that it follows from Lemma A.11 that

$$E\left\{\mathrm{tr}\left[\left(\sum_{t=j+1}^{n} \mathbf{y}_{t-1}\mathbf{v}'_{t-j}\right)\left(\sum_{t=j+1}^{n} \mathbf{y}_{t-1}\mathbf{v}'_{t-j}\right)'\right]\right\} \leqslant \Delta_1 n + \Delta_2 n^2||\mathbf{\Lambda}_1||_{\mathrm{E}}^{2(j-1)}. \qquad \text{(A.11)}$$

With Lemma A.11, we can conclude that G_{3n} is bounded in probability. Regrouping the results for G_{1n}, G_{2n} and G_{3n}, we conclude that $F_{2n} = O_p(1)$.

It is easy to verify that $E(|F_{3n}|) = dn^{-1}E(||x_1||^2)\sum_{j=1}^{n-1}(1-j/n)k_{nj}^2 = O(p_n/n)$, by the strict exogeneity of the $\mathbf{x}_t$ process and the result for F_{3n} follows. □

The proof for E_{1n} is therefore completed. The proof for E_{2n} is similar. It remains to study E_{3n}. We remark first that $|E_{3n}| \leqslant \sum_{j=1}^{n-1} k_{nj}^2 (n^{-1}\sum_{t=1}^{n} ||\hat{\gamma}_{nt}||^2)^2$. The following lemma whose proof is straightforward is needed.

Lemma A.12. $n^{-1}\sum_{t=1}^{n} ||\hat{\gamma}_{nt}||^2 = O_p(n^{-1})$.

This shows that $E_{3n} = O_p(p_n/n^2) = o_p(n^{-1})$ and the proof of Result A.5 is completed. □

To prove Result A.6, we write $\sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}[\mathbf{C}_v(j)(\mathbf{C}_{\hat{v}}(j) - \mathbf{C}_v(j))'] = -E_{4n} - E_{5n} + E_{6n}$, where

$$E_{4n} = \sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}\left[\left(n^{-1}\sum_{t=j+1}^{n} \mathbf{v}_t\mathbf{v}_{t-j}'\right)\left(n^{-1}\sum_{t=j+1}^{n} \hat{\gamma}_{nt}\mathbf{v}_{t-j}'\right)'\right],$$

$$E_{5n} = \sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}\left[\left(n^{-1}\sum_{t=j+1}^{n} \mathbf{v}_t\mathbf{v}_{t-j}'\right)\left(n^{-1}\sum_{t=j+1}^{n} \mathbf{v}_t\hat{\gamma}_{n,t-j}'\right)'\right],$$

$$E_{6n} = \sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}\left[\left(n^{-1}\sum_{t=j+1}^{n} \mathbf{v}_t\mathbf{v}_{t-j}'\right)\left(n^{-1}\sum_{t=j+1}^{n} \hat{\gamma}_{nt}\hat{\gamma}_{n,t-j}'\right)'\right].$$

We complete the proof by showing that

Result A.8. $E_{jn} = o_p(\sqrt{p_n}/n),\ j = 4, 5, 6.$

Proof. Let us first consider E_{4n} that we decompose it in the following manner: $E_{4n} = F_{4n} + F_{5n} + F_{6n}$, where

$$F_{4n} = \sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}\left[\left(n^{-1}\sum_{t=j+1}^{n} \mathbf{v}_t\mathbf{v}_{t-j}'\right)\left(\boldsymbol{\Sigma}_u^{-1/2}(\hat{\mathbf{c}} - \mathbf{c})n^{-1}\sum_{t=j+1}^{n} \mathbf{v}_{t-j}'\right)'\right],$$

$$F_{5n} = \sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}\left[\left(n^{-1}\sum_{t=j+1}^{n} \mathbf{v}_t\mathbf{v}_{t-j}'\right)\left(\boldsymbol{\Sigma}_u^{-1/2}(\hat{\boldsymbol{\Lambda}}_1 - \boldsymbol{\Lambda}_1)n^{-1}\sum_{t=j+1}^{n} \mathbf{y}_{t-1}\mathbf{v}_{t-j}'\right)'\right],$$

$$F_{6n} = \sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}\left[\left(n^{-1}\sum_{t=j+1}^{n} \mathbf{v}_t\mathbf{v}_{t-j}'\right)\left(\boldsymbol{\Sigma}_u^{-1/2}(\hat{\mathbf{V}}_0 - \mathbf{V}_0)n^{-1}\sum_{t=j+1}^{n} \mathbf{x}_t\mathbf{v}_{t-j}'\right)'\right].$$

Note that

$$|F_{4n}| \leqslant n^{-2}\{\mathrm{tr}[(\hat{\mathbf{c}}-\mathbf{c})'\boldsymbol{\Sigma}_u^{-1}(\hat{\mathbf{c}}-\mathbf{c})]\}^{1/2} \times \sum_{j=1}^{n-1} k_{nj}^2 \left\{\mathrm{tr}\left[\left(\sum_{t=j+1}^{n} \mathbf{v}_t\mathbf{v}'_{t-j}\right)\left(\sum_{t=j+1}^{n} \mathbf{v}_t\mathbf{v}'_{t-j}\right)'\right]\right\}^{1/2} \times \left\{\mathrm{tr}\left[\left(\sum_{t=j+1}^{n} \mathbf{v}'_{t-j}\right)\left(\sum_{t=j+1}^{n} \mathbf{v}'_{t-j}\right)'\right]\right\}^{1/2}.$$

By Cauchy–Schwarz inequality, since $E(\mathrm{tr}[(\sum_{t=j+1}^{n}\mathbf{v}_t\mathbf{v}'_{t-j})(\sum_{t=j+1}^{n}\mathbf{v}_t\mathbf{v}'_{t-j})'])=O(n)$ and since we also have that $E(\mathrm{tr}[(\sum_{t=j+1}^{n}\mathbf{v}'_{t-j})(\sum_{t=j+1}^{n}\mathbf{v}'_{t-j})'])=O(n)$, we can conclude that $F_{4n}=O_p(p_n/n^{3/2})$. Similarly, we have that

$$|F_{5n}| \leqslant \{\mathrm{tr}[(\hat{\boldsymbol{\Lambda}}_1-\boldsymbol{\Lambda}_1)'\boldsymbol{\Sigma}_u^{-1}(\hat{\boldsymbol{\Lambda}}_1-\boldsymbol{\Lambda}_1)]\}^{1/2}n^{-2} \times \sum_{j=1}^{n-1} k_{nj}^2 \left\{\mathrm{tr}\left[\left(\sum_{t=j+1}^{n} \mathbf{v}_t\mathbf{v}'_{t-j}\right)\left(\sum_{t=j+1}^{n} \mathbf{v}_t\mathbf{v}'_{t-j}\right)'\right]\right\}^{1/2} \times \left\{\mathrm{tr}\left[\left(\sum_{t=j+1}^{n} \mathbf{y}_{t-1}\mathbf{v}'_{t-j}\right)\left(\sum_{t=j+1}^{n} \mathbf{y}_{t-1}\mathbf{v}'_{t-j}\right)'\right]\right\}^{1/2}.$$

Since

$$E\left[\sum_{j=1}^{n-1} k_{nj}^2 \left\{\mathrm{tr}\left[\left(\sum_{t=j+1}^{n} \mathbf{v}_t\mathbf{v}'_{t-j}\right)\left(\sum_{t=j+1}^{n} \mathbf{v}_t\mathbf{v}'_{t-j}\right)'\right]\right\}^{1/2} \times \left\{\mathrm{tr}\left[\left(\sum_{t=j+1}^{n} \mathbf{y}_{t-1}\mathbf{v}'_{t-j}\right)\left(\sum_{t=j+1}^{n} \mathbf{y}_{t-1}\mathbf{v}'_{t-j}\right)'\right]\right\}^{1/2}\right] \leqslant \Delta_1^{1/2} n \sum_{j=1}^{n-1} k_{nj}^2 + n^{3/2}\Delta_2^{1/2} \sum_{j=1}^{n-1} k_{nj}^2 ||\boldsymbol{\Lambda}_1||_{\mathrm{E}}^{j-1},$$

and using (A.11), we have that $F_{5n}=O_p(p_n/n^{3/2}+1/n)$. Similarly, using the strict exogeneity of $\mathbf{x}_t$, we can conclude that $F_{6n}=O_p(p_n/n^{3/2})$. It follows therefore that $E_{4n}=o_p(p_n^{1/2}/n)$ since $p_n/n\to 0$. For E_{5n}, note that

$$|E_{5n}| \leqslant n^{-2}\left\{\sum_{j=1}^{n-1} k_{nj}^2\, \mathrm{tr}\left[\left(\sum_{t=j+1}^{n} \mathbf{v}_t\mathbf{v}'_{t-j}\right)\left(\sum_{t=j+1}^{n} \mathbf{v}_t\mathbf{v}'_{t-j}\right)'\right]\right\}^{1/2} \times \left\{\sum_{j=1}^{n-1} k_{nj}^2\, \mathrm{tr}\left[\left(\sum_{t=j+1}^{n} \mathbf{v}_t\hat{\boldsymbol{\gamma}}'_{n,t-j}\right)\left(\sum_{t=j+1}^{n} \mathbf{v}_t\hat{\boldsymbol{\gamma}}'_{n,t-j}\right)'\right]\right\}^{1/2}.$$

We can conclude that $E_{5n} = o_p(p_n^{1/2}/n)$, since it is easily shown that

$$\sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}\left[\left(n^{-1}\sum_{t=j+1}^{n} \mathbf{v}_t\hat{\gamma}'_{n,t-j}\right)\left(n^{-1}\sum_{t=j+1}^{n} \mathbf{v}_t\hat{\gamma}'_{n,t-j}\right)'\right] = o_p(n^{-1}).$$

For E_{6n}, we proceed in a similar way showing that

$$|E_{6n}| \leqslant n^{-2}\left\{\sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}\left[\left(\sum_{t=j+1}^{n} \mathbf{v}_t\mathbf{v}'_{t-j}\right)\left(\sum_{t=j+1}^{n} \mathbf{v}_t\mathbf{v}'_{t-j}\right)'\right]\right\}^{1/2} \times \left\{\sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}\left[\left(\sum_{t=j+1}^{n} \hat{\gamma}_{nt}\hat{\gamma}'_{n,t-j}\right)\left(\sum_{t=j+1}^{n} \hat{\gamma}_{nt}\hat{\gamma}'_{n,t-j}\right)'\right]\right\}^{1/2}.$$

We can then conclude that $E_{6n} = o_p(p_n^{1/2}/n)$. □

By adding and subtracting $\operatorname{tr}[\mathbf{\Sigma}_u^{-1}\mathbf{C}'_u(j)\mathbf{\Sigma}_u^{-1}\mathbf{C}_u(j)]$ in the left-hand side of (15) and by using Result A.1, the proof of part 2 will be completed if we can show that

$$\sum_{j=1}^{n-1} k_{nj}^2(\operatorname{tr}[\mathbf{C}_{\hat{u}}^{-1}(0)\mathbf{C}'_{\hat{u}}(j)\mathbf{C}_{\hat{u}}^{-1}(0)\mathbf{C}_{\hat{u}}(j)] - \operatorname{tr}[\mathbf{\Sigma}_u^{-1}\mathbf{C}'_u(j)\mathbf{\Sigma}_u^{-1}\mathbf{C}_u(j)]) = o_p(\sqrt{p_n}/n). \tag{A.12}$$

From (A.8), and Results A.5 and A.6, it is sufficient to show that

$$\sum_{j=1}^{n-1} k_{nj}^2(\operatorname{tr}[\mathbf{C}_{\hat{u}}^{-1}(0)\mathbf{C}'_{\hat{u}}(j)\mathbf{C}_{\hat{u}}^{-1}(0)\mathbf{C}_{\hat{u}}(j)] - \operatorname{tr}[\mathbf{\Sigma}_u^{-1}\mathbf{C}'_{\hat{u}}(j)\mathbf{\Sigma}_u^{-1}\mathbf{C}_{\hat{u}}(j)]) = o_p(\sqrt{p_n}/n). \tag{A.13}$$

We already know that $\mathbf{C}_{\hat{u}}(0) - \mathbf{\Sigma}_u = O_p(n^{-1/2})$, which implies that

$$\mathbf{C}_{\hat{u}}^{-1}(0) - \mathbf{\Sigma}_u^{-1} = O_p(n^{-1/2}), \tag{A.14}$$

and (A.13) follows using inequality (A.1). □

Proof of Theorem 2. We have that $Q^2(\hat{\mathbf{f}}_n;\mathbf{f}_0) = 2\pi\int_{-\pi}^{\pi}\operatorname{tr}[\mathbf{\Gamma}_u^{-1}(0)(\hat{\mathbf{f}}_n - \mathbf{f}_0)^*\mathbf{\Gamma}_u^{-1}(0)(\hat{\mathbf{f}}_n - \mathbf{f}_0)]\,d\omega$. Since $\hat{\mathbf{f}}_n - \mathbf{f}_0 = (\hat{\mathbf{f}}_n - \mathbf{f}) + (\mathbf{f} - \mathbf{f}_0)$, a direct calculation leads to

$$Q^2(\hat{\mathbf{f}}_n;\mathbf{f}_0) = Q^2(\mathbf{f};\mathbf{f}_0) + 4\pi\int_{-\pi}^{\pi}\operatorname{tr}[\mathbf{\Gamma}_u^{-1}(0)(\mathbf{f} - \mathbf{f}_0)^*\mathbf{\Gamma}_u^{-1}(0)(\hat{\mathbf{f}}_n - \mathbf{f})]\,d\omega + 2\pi\int_{-\pi}^{\pi}\operatorname{tr}[\mathbf{\Gamma}_u^{-1}(0)(\hat{\mathbf{f}}_n - \mathbf{f})^*\mathbf{\Gamma}_u^{-1}(0)(\hat{\mathbf{f}}_n - \mathbf{f})]\,d\omega.$$

By showing that $\int_{-\pi}^{\pi}\operatorname{tr}[\mathbf{\Gamma}_u^{-1}(0)(\hat{\mathbf{f}}_n - \mathbf{f})^*\mathbf{\Gamma}_u^{-1}(0)(\hat{\mathbf{f}}_n - \mathbf{f})]\,d\omega = o_p(1)$, we obtain from Cauchy–Schwarz inequality that $4\pi\int_{-\pi}^{\pi}\operatorname{tr}[\mathbf{\Gamma}_u^{-1}(0)(\mathbf{f} - \mathbf{f}_0)^*\mathbf{\Gamma}_u^{-1}(0)(\hat{\mathbf{f}}_n - \mathbf{f})]\,d\omega = o_p(1)$.

Result A.9. $\int_{-\pi}^{\pi} \mathrm{tr}[\boldsymbol{\Gamma}_u^{-1}(0)(\hat{\mathbf{f}}_n - \mathbf{f})^* \boldsymbol{\Gamma}_u^{-1}(0)(\hat{\mathbf{f}}_n - \mathbf{f})]\, d\omega = o_p(1)$.

Proof. Since $\mathrm{tr}[\mathbf{A}(\mathbf{B}+\mathbf{C})^*\mathbf{A}(\mathbf{B}+\mathbf{C})] \leqslant 2\,\mathrm{tr}[\mathbf{A}\mathbf{B}^*\mathbf{A}\mathbf{B}] + 2\,\mathrm{tr}[\mathbf{A}\mathbf{C}^*\mathbf{A}\mathbf{C}]$, where $\mathbf{A}$ is symmetric and non-singular, using the decomposition $\hat{\mathbf{f}}_n - \mathbf{f} = (\hat{\mathbf{f}}_n - \tilde{\mathbf{f}}_n) + (\tilde{\mathbf{f}}_n - \mathbf{f})$, we can write

$$\begin{aligned}
&\int_{-\pi}^{\pi} \mathrm{tr}[\boldsymbol{\Gamma}_u^{-1}(0)(\hat{\mathbf{f}}_n - \mathbf{f})^* \boldsymbol{\Gamma}_u^{-1}(0)(\hat{\mathbf{f}}_n - \mathbf{f})]\, d\omega \\
&\quad \leqslant 2\int_{-\pi}^{\pi} \{\mathrm{tr}[\boldsymbol{\Gamma}_u^{-1}(0)(\hat{\mathbf{f}}_n - \tilde{\mathbf{f}}_n)^* \boldsymbol{\Gamma}_u^{-1}(0)(\hat{\mathbf{f}}_n - \tilde{\mathbf{f}}_n)] \\
&\qquad + \mathrm{tr}[\boldsymbol{\Gamma}_u^{-1}(0)(\tilde{\mathbf{f}}_n - \mathbf{f})^* \boldsymbol{\Gamma}_u^{-1}(0)(\tilde{\mathbf{f}}_n - \mathbf{f})]\}\, d\omega.
\end{aligned}$$

Now, Result A.9 follows from the next two lemmas. □

Lemma A.13. $\int_{-\pi}^{\pi} \mathrm{tr}[\boldsymbol{\Gamma}_u^{-1}(0)(\hat{\mathbf{f}}_n - \tilde{\mathbf{f}}_n)^* \boldsymbol{\Gamma}_u^{-1}(0)(\hat{\mathbf{f}}_n - \tilde{\mathbf{f}}_n)]\, d\omega = o_p(1)$.

Proof. Using $\hat{\mathbf{f}}_n - \tilde{\mathbf{f}}_n = \frac{1}{2\pi}\sum_{j=-n+1}^{n-1} k_{nj}[\mathbf{C}_{\hat{u}}(j) - \mathbf{C}_u(j)]e^{-i\omega j}$, we have that

$$\begin{aligned}
&\int_{-\pi}^{\pi} \mathrm{tr}[\boldsymbol{\Gamma}_u^{-1}(0)(\hat{\mathbf{f}}_n - \tilde{\mathbf{f}}_n)^* \boldsymbol{\Gamma}_u^{-1}(0)(\hat{\mathbf{f}}_n - \tilde{\mathbf{f}}_n)]\, d\omega \\
&\quad = \frac{1}{2\pi}\sum_{j=-n+1}^{n-1} k_{nj}^2\, \mathrm{tr}[(\mathbf{C}_{\hat{v}}(j) - \mathbf{C}_v(j))(\mathbf{C}_{\hat{v}}(j) - \mathbf{C}_v(j))'].
\end{aligned}$$

Inequality (A.10) provides an upper bound for $\sum_{j=1}^{n-1} k_{nj}^2\, \mathrm{tr}[(\mathbf{C}_{\hat{v}}(j) - \mathbf{C}_v(j))(\mathbf{C}_{\hat{v}}(j) - \mathbf{C}_v(j))']$. The sum for negative j can be bounded in a similar way and it is easy to deal with the term corresponding to $j=0$. Here, Result A.7 in the second part of the proof of Theorem 1 does not necessarily hold since we now are under the alternative hypothesis. More precisely, we have to treat differently the E_{jn}'s, $j = 1, 2, 3$, under the correlation structure given in Assumption C. However, by Cauchy–Schwarz inequality, we obtain that $|E_{1n}| = E_{1n} \leqslant (\sum_{j=1}^n k_{nj}^2)(n^{-1}\sum_{j=1}^n ||\mathbf{v}_t||^2)(n^{-1}\sum_{j=1}^n ||\hat{\gamma}_{nt}||^2)$. But we have $n^{-1}\sum_{j=1}^n ||\hat{\gamma}_{nt}||^2 = O_p(n^{-1})$, since

$$||\hat{\gamma}_{nt}||^2 \leqslant 4(\hat{\mathbf{c}} - \mathbf{c})'\boldsymbol{\Sigma}_u^{-1}(\hat{\mathbf{c}} - \mathbf{c}) + 4\,\mathrm{tr}[(\hat{\mathbf{V}}_0 - \mathbf{V}_0)'\boldsymbol{\Sigma}_u^{-1}(\hat{\mathbf{V}}_0 - \mathbf{V}_0)]||\mathbf{x}_t||^2,$$

and in the static model (16) the LS estimators are $\sqrt{n}$-consistent. Thus, we have $E_{1n} = O_p(p_n/n)$ and the terms E_{2n} and E_{3n} can be dealt with in a similar way. This completes the proof of Lemma A.13. □

Lemma A.14. $\int_{-\pi}^{\pi} \mathrm{tr}[\boldsymbol{\Gamma}_u^{-1}(0)(\tilde{\mathbf{f}}_n - \mathbf{f})^* \boldsymbol{\Gamma}_u^{-1}(0)(\tilde{\mathbf{f}}_n - \mathbf{f})]\, d\omega = o_p(1)$.

Proof. We can write $\tilde{\mathbf{f}}_n - \mathbf{f} = \frac{1}{2\pi}\sum_{|j|\leqslant n-1}[k_{nj}\mathbf{C}_u(j) - \mathbf{\Gamma}_u(j)]e^{-ij\omega} - \frac{1}{2\pi}\sum_{|j|>n-1}\mathbf{\Gamma}_u(j)e^{-ij\omega}$ and after integrating, we find that

$$\int_{-\pi}^{\pi} \mathrm{tr}[\mathbf{\Gamma}_u^{-1}(0)(\tilde{\mathbf{f}}_n - \mathbf{f})^* \mathbf{\Gamma}_u^{-1}(0)(\tilde{\mathbf{f}}_n - \mathbf{f})]\, d\omega$$
$$= \frac{1}{2\pi}\sum_{|j|\leqslant n-1} \mathrm{tr}[\mathbf{\Gamma}_u^{-1}(0)(k_{nj}\mathbf{C}_u(j) - \mathbf{\Gamma}_u(j))' \mathbf{\Gamma}_u^{-1}(0)(k_{nj}\mathbf{C}_u(j) - \mathbf{\Gamma}_u(j))]$$
$$+ \frac{1}{2\pi}\sum_{|j|\geqslant n} \mathrm{tr}[\mathbf{\Gamma}_u^{-1}(0)\mathbf{\Gamma}_u(j)'\mathbf{\Gamma}_u^{-1}(0)\mathbf{\Gamma}_u(j)].$$

Under Assumption C, we have that $\sum_{|j|\geqslant n} \mathrm{tr}[\mathbf{\Gamma}_u^{-1}(0)\mathbf{\Gamma}_u(j)'\mathbf{\Gamma}_u^{-1}(0)\mathbf{\Gamma}_u(j)] = o_p(1)$. It remains to verify that the first term in the right-hand-side member is also $o_p(1)$. However, using $k_{nj}\mathbf{C}_u(j) - \mathbf{\Gamma}_u(j) = (k_{nj}-1)\mathbf{\Gamma}_u(j) + k_{nj}(\mathbf{C}_u(j) - \mathbf{\Gamma}_u(j))$, we can show that

$$\sum_{|j|\leqslant n-1} \mathrm{tr}[\mathbf{\Gamma}_u^{-1}(0)(k_{nj}\mathbf{C}_u(j) - \mathbf{\Gamma}_u(j))' \mathbf{\Gamma}_u^{-1}(0)(k_{nj}\mathbf{C}_u(j) - \mathbf{\Gamma}_u(j))]$$
$$\leqslant 2\sum_{|j|\leqslant n-1} (k_{nj}-1)^2\, \mathrm{tr}[\mathbf{\Gamma}_u^{-1}(0)\mathbf{\Gamma}_u(j)'\mathbf{\Gamma}_u^{-1}(0)\mathbf{\Gamma}_u(j)]$$
$$+\ 2\sum_{|j|\leqslant n-1} k_{nj}^2\, \mathrm{tr}[\mathbf{\Gamma}_u^{-1}(0)(\mathbf{C}_u(j) - \mathbf{\Gamma}_u(j))'\mathbf{\Gamma}_u^{-1}(0)(\mathbf{C}_u(j) - \mathbf{\Gamma}_u(j))].$$

By an argument similar to the one used by Hong [12, p. 861], the first term in the right-hand side is $o(1)$ by Lebesgue dominated convergence theorem and Assumption A on the kernel k. For the other term, note that $\mathrm{tr}[\mathbf{\Gamma}_u^{-1}(0)(\mathbf{C}_u(j) - \mathbf{\Gamma}_u(j))'\mathbf{\Gamma}_u^{-1}(0)(\mathbf{C}_u(j) - \mathbf{\Gamma}_u(j)) = \sum_{t=1}^{d}\sum_{s=1}^{d}(C_{v,st}(j) - \Gamma_{v,st}(j))^2$, where $C_{v,st}(j)$ and $\Gamma_{v,st}(j)$ are the (s,t)-components of $\mathbf{C}_v(j)$ and $\mathbf{\Gamma}_v(j)$, respectively. From a general result for the variance of cross-covariances, see for example Hannan [8, pp. 208–211] or Chitturi [6]. The variance of $C_{v,st}(j)$ is given by $\mathrm{var}(C_{v,st}(j)) = n^{-1}\sum_{|i|\leqslant n-1}(1-|i|/n)[\Gamma_{v,st}(i+j)\Gamma_{v,st}(i-j) + \kappa_{stst}(0,j,i,i+j)]$. From Assumption C, we have that $\sup_{j\geqslant 1}\mathrm{var}[C_{v,st}(j)] = O(n^{-1})$. Therefore, $\sum_{|j|\leqslant n-1} k_{nj}^2 \sum_{t=1}^{d}\sum_{s=1}^{d}[C_{v,st}(j) - \Gamma_{v,st}(j)]^2 = O_p(p_n/n)$ and the proof of Lemma A.14 is completed. Consequently, Result A.29 holds and Theorem 2 is proved. □

References

[1] C.F. Ansley, Computation of the theoretical autocovariance function for a vector *ARMA* process, J. Statist. Comput. Simulation 12 (1980) 15–24.

[2] R.R. Bahadur, Stochastic comparison of tests, Ann. Math. Statist. 31 (1960) 276–295.

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

[4] B.M. Brown, Martingale central limit theorems, Ann. Math. Statist. 42 (1971) 59–66.

[5] R.V. Chitturi, Distribution of residual autocorrelations in multiple autoregressive schemes, J. Amer. Statist. Assoc. 69 (1974) 928–934.
[6] R.V. Chitturi, Distribution of multivariate white noise autocorrelations, J. Amer. Statist. Assoc. 71 (1976) 223–226.
[7] J. Geweke, The approximate slopes of econometric tests, Econometrica 49 (1981) 1427–1442.
[8] E. Hannan, Multiple Time Series, Wiley, New York, 1970.
[9] E.J. Hannan, M. Deistler, The Statistical Theory of Linear Systems, Wiley, New York, 1988.
[10] A.C. Harvey, The Econometric Analysis of Time Series, 2nd Edition, MIT Press, Cambridge, MA, 1990.
[11] D.A. Harville, Matrix Algebra from a Statistician's Perspective, Springer, Berlin, 1997.
[12] Y. Hong, Consistent testing for serial correlation of unknown form, Econometrica 64 (1996) 837–864.
[13] J. Hosking, The multivariate portmanteau statistic, J. Amer. Statist. Assoc. 75 (1980) 602–608.
[14] J. Hosking, Equivalent forms of the multivariate portmanteau statistic, J. Roy. Statist. Soc. Ser. B 43 (1981) 261–262 (Corr: V51/1989, p. 303).
[15] J. Hosking, Lagrange-multiplier tests of multivariate time series models, J. Roy. Statist. Soc. Ser. B 43 (1981) 219–230.
[16] G.G. Judge, R.C. Hill, W.E. Griffiths, H. Lütkepohl, T.-C. Lee, The Theory and Practice of Econometrics, 2nd Edition, Wiley, New York, 1985.
[17] W.K. Li, A.I. McLeod, Distribution of the residual autocorrelations in multivariate *ARMA* time series models, J. Roy. Statist. Soc. Ser. B 43 (1981) 231–239.
[18] H. Lütkepohl, Introduction to Multiple Time Series Analysis, 2nd Edition, Springer, Berlin, 1993.
[19] E. Paparoditis, Spectral density based goodness-of-fit tests for time series analysis, Scand. J. Statist. 27 (2000) 143–176.
[20] E. Paparoditis, On some power properties of goodness-of-fit tests for time series models, in: Ch.A. Charalambides, M.V., Koutras, N. Balakrishnan (Eds.), Probability and Statistical Models with Applications, Chapman & Hall, London, 2000.
[21] E. Paparoditis, Testing the fit of a vector autoregressive moving average model, Preprint, University of Cyprus, 2001.
[22] M.B. Priestley, Spectral Analysis and Time Series, Vol. 1: Univariate Series, Academic Press, New York, 1981.
[23] P.M. Robinson, Automatic frequency domain inference on semiparametric and non-parametric models, Econometrica 59 (1991) 1329–1363.
[24] P.M. Robinson, Time series with strong dependence, in: C. Sims (Ed.), Advances in Econometrics, Sixth World Congress, Vol. 1, Cambridge University Press, Cambridge, 1994, pp. 47–95.