Consistent tests for independence against serial dependence of unknown form in vector time series models

Pierre DUCHESNE[*] Roch ROY[†‡]

CRM-2703

January 2001

[*]Service d'enseignement des Méthodes Quantitatives de Gestion, École des Hautes Études Commerciales, 3000, chemin de la Côte-Sainte-Catherine, Montréal (Québec), H3T 2A7, Canada; pierre.duchesne@hec.ca; tel: (514) 340-6503; fax: (514) 340-5634

[†]Département de mathématiques et statistique, Université de Montréal, C.P. 6128, succ. Centre-Ville, Montréal (Québec) H3C 3J7, Canada; royr@dms.umontreal.ca

[‡]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.

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. Inspired by Hong (1996), we propose a new test statistic for checking this latter hypothesis in this context. 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 (1980). However, some kernels lead to a greater power, as shown in a local and global analysis. The level and power of the resulting tests are also studied by simulation and an application is presented.

Key words and phrases: Vector autoregressive process; exogenous variables; dynamic simultaneous equation model; kernel spectrum estimator; diagnostic test; portmanteau test.

Mathematics subject classification codes (1991): primary 62M10; secondary 62M15.

Résumé

Nous considérons des tests d'absence de corrélation dans le terme d'erreur d'un modèle *VARX*. Il est possible dans cette classe de généraliser au cas multivarié l'approche de Hong (1996). Le test est obtenu en comparant un estimateur de la densité spectrale multivariée calculé par la méthode du noyau avec la densité spectrale sous l'hypothèse nulle d'absence de corrélation dans le terme d'erreur, en utilisant une certaine norme quadratique. Le test généralise la statistique portmanteau de Hosking (1980), qui peut-être vue comme un test basé sur la norme quadratique utilisant le périodogramme tronqué. Cependant, tout comme dans le cas univarié, plusieurs noyaux offrent une meilleure puissance que le test basé sur le noyau tronqué, comme montré dans l'analyse locale et globale. Le niveau et la puissance des tests sont aussi étudiés par simulation et une application est présentée.

1. INTRODUCTION

Vector autoregressive models with explanatory variables, abbreviated by the acronym *VARX* 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 is 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 can 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. (1985, 1988), Hannan and Deistler (1988), Lütkepohl (1993) and Hendry (1995). 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 (1993, Section 10.3) or Hannan and Deistler (1988, Section 4.2).

In the univariate case, Hong (1996) 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 ordinary least squares estimators inconsistent. Three classes of portmanteau type statistics are then proposed. His approach consists in comparing a residual kernel-based spectral density estimator and the spectral density under the null hypothesis. Three distance measures are considered. 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

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

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

The sequence p_n is a sequence of truncation values.

Using a different approach, Paparoditis (2000a) considered goodness-of-fit tests for univariate time series models. His test statistic relies on a distance 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. Power properties of these tests are investigated in Paparoditis (2000b).

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 y that allows us to retrieve Hong's statistic M_{1n} when $d = 1$. 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 (1972) statistic for univariate *ARMA* processes. The multivariate portmanteau statistic was studied by many authors, namely by Chitturi (1974), Hosking (1980, 1981a) and Li and McLeod (1981).

The organization and main results of the paper are as follows. In Section 2, we introduce the notation, the concepts and the results employed thereafter. The new test statistic is introduced in Section 3. Using a central limit theorem for a martingale difference, it is shown that its asymptotic distribution under the null hypothesis 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 proof is long and involved. For that reason, it is presented in an Appendix. The asymptotic power of the tests are studied in Section 4. First, the asymptotic normality of the test statistic is derived under a sequence of local alternatives that converges to the null hypothesis as n tends to ∞. This latter result allows us to compute the asymptotic relative efficiency (ARE) of one kernel with respect to the other and to determine the optimal kernel within a class of sufficiently smooth kernels. Second, it is shown that for an arbitrary fixed alternative, the power tends to one and therefore, our test is consistent. 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. The new test is applied to a real data set in Section 6 and we conclude with some remarks in the last section.

2. PRELIMINARIES

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

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

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

where $\Lambda(B) = \Lambda_0 - \sum_{j=1}^{r} \Lambda_j B^j$, $\mathrm{V}(B) = \sum_{j=0}^{s} \mathrm{V}_j B^j$, B *being the usual backward shift operator and* $\mathrm{u} = \{\mathrm{u}_t : t \in \mathbb{Z}\}$ *is a strong white noise of dimension* d, *that is the* u_t *are independent random vectors with mean* 0, *and regular covariance matrix* Σ_u.

In the sequel, $\det \mathrm{A}$ denotes the determinant of a square matrix A and I_d stands for the identity matrix of dimension d. We suppose that all the roots of $\det \Lambda(z)$ are outside the unit disk, where 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 *VARX* representation of the process y. Since $\det \Lambda(0) = \det \Lambda_0 \neq 0$ by assumption, we can premultiply (3) by Λ_0^{-1} and we obtain an equivalent *VARX* representation in which $\Lambda(0) = \mathrm{I}_d$; it is called the *reduced form* of the model. Hereafter, we will suppose that representation (3) is in reduced form, that is $\Lambda(0) = \Lambda_0 = \mathrm{I}_d$. As mentioned in Lütkepohl (1993, chap. 10), it is more convenient for least squares estimation. Also, predictions of future values of the endogenous variables in a dynamic simultaneous equation model are usually made from the reduced form (Harvey (1990), pp. 352-353).

Let $\mathrm{u} = \{\mathrm{u}_t, t \in \mathbb{Z}\}$ where $\mathrm{u}_t = (u_t(1), \ldots, u_t(d))'$ is an arbitrary second order stationary process whose mean is 0. The autocovariance at lag j will be denoted by

$$\Gamma_u(j) = E(\mathrm{u}_t \mathrm{u}_{t-j}'), \quad j \in \mathbb{Z}.$$

If we write $\Gamma_u(j) = [\Gamma_{u,pq}(j)]_{p,q=1}^{d}$, and if

$$\sum_{j=0}^{\infty} |\Gamma_{u,pq}(j)| < \infty, \quad p, q = 1, \ldots, d,$$

the spectral density $\mathrm{f}(\omega)$ of u is defined by

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

When the existence of the fourth order moments will be required, we will suppose that the process u is fourth order stationary and the fourth order moments and cumulants will be denoted respectively by

$$\mu_4(p,q,r,s) = E(u_t(p)u_t(q)u_t(r)u_t(s))$$

and

$$\kappa_{pqrs}(i,j,k,l) = \mathrm{cum}(u_i(p),u_j(q),u_k(r),u_l(s)),$$

where $p,q,r,s = 1,\ldots,d$ and $i,j,k,l,t \in \mathbb{Z}$. Imposing the existence of the fourth order moments is equivalent to the existence of the fourth order cumulants. If the process u is Gaussian, it is well known that the fourth order cumulants vanish.

Given $\mathrm{u}_1,\ldots,\mathrm{u}_n$ a realization of length n of the process u, the sample autocovariance at lag j, $0 \le |j| \le n-1$, is defined by

$$\mathrm{C}_u(j) = \begin{cases} n^{-1}\sum_{t=j+1}^{n} \mathrm{u}_t\mathrm{u}'_{t-j}, & j = 0,1,\ldots,n-1, \\ \mathrm{C}'_u(-j), & j = -1,\ldots,-n+1. \end{cases} \tag{5}$$

The classical nonparametric kernel-based estimator of the spectral density $\mathrm{f}(\omega)$ of u is given by

$$\mathrm{f}_n(\omega) = \frac{1}{2\pi}\sum_{j=-n+1}^{n-1} k(j/p_n)\mathrm{C}_u(j)e^{-i\omega j}, \tag{6}$$

where $k(\cdot)$ is a kernel or a lag window. The parameter p_n is a truncation point when the kernel is of compact support, or a smoothing parameter when the kernel support is unbounded. We suppose that $p_n \to \infty$ and $p_n/n \to 0$. Examples of p_n satisfying these criteria are $p_n \propto n^{\alpha}$, with $\alpha \in (0,1)$, and $p_n \propto \log(n)$. Using the rectangular or truncated uniform kernel $k_T(z) = I[|z| \le 1]$, where $I(A)$ is the indicator function of the set A, we retrieve the familiar truncated periodogram (Priestley, 1981, Section 6.2.3). Here, we will also use kernel-based estimators of the form (6) with the usual assumptions on the kernel that are summarized as follows.

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

Examples of kernels or lag windows frequently used in time series analysis are given for example in Priestley (1981, Section 6.2.3).

There exists various methods for estimating the parameter of a *VARX* model. The more often use in practice is the generalized least squares (GLS) method as described in Lütkepohl (1993, chap. 10.3).

Often, there are linear contraints on the parameters, for example parameter values that are fixed to zero. Therefore, we suppose that the parameters satisfy the relation $\boldsymbol{\beta} = \mathrm{vec}(\Lambda,\mathrm{V},\mathrm{V}_0) = \mathrm{R}\boldsymbol{\gamma}$, where $\Lambda = (\Lambda_1,\Lambda_2,\ldots,\Lambda_r)$, $\mathrm{V} = (\mathrm{V}_1,\mathrm{V}_2,\ldots,\mathrm{V}_s)$, and R is a known matrix of linear constraints. In GLS, we first estimate $\boldsymbol{\gamma}$, say by $\hat{\boldsymbol{\gamma}}$, and $\hat{\boldsymbol{\beta}} = R\hat{\boldsymbol{\gamma}}$.

An alternative approach is maximum likelihood estimation (MLE) which is described in details in Hannan and Deistler (1988, Chap. 4). Both GLS and MLE lead to $\sqrt{n}$-consistent estimators of $\boldsymbol{\beta}$, that is

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = O_p(n^{-1/2}). \tag{7}$$

The mathematical developments that follow are valid for any estimator of $\boldsymbol{\beta}$ that satisfies relation (7). For that reason, it will be sufficient to make the following assumption on the estimator $\hat{\boldsymbol{\beta}}$.

Assumption B: The estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ in the *VARX* model satisfies relation (7).

Once a *VARX* model is estimated, the residual $\hat{\mathrm{u}}_t$, $t = 1,\ldots,n$, can be computed. The residual autocovariance at lag j, $\mathrm{C}_{\hat{u}}(j)$, is obtained from (5) with u_t replaced by $\hat{\mathrm{u}}_t$. Similarly, the residual spectral density estimator $\hat{\mathrm{f}}_n(\omega)$ is obtained from (6) where $\mathrm{C}_{\hat{u}}(j)$ is substituted for $\mathrm{C}_u(j)$, that is

$$\hat{\mathrm{f}}_n(\omega) = \frac{1}{2\pi}\sum_{j=-n+1}^{n-1} k(j/p_n)\mathrm{C}_{\hat{u}}(j)e^{-i\omega j}. \tag{8}$$

The test statistic introduced in the following section is based on (8).

3. THE TEST STATISTIC AND ITS ASYMPTOTIC NULL DISTRIBUTION

In the sequel, the process u represents the error process in the *VARX* model (3). The hypothesis of interest is that the error process u is a white noise against the alternative of serial correlation of arbitrary form. More formally, it can be written as

$$H_0\colon\ \Gamma_u(j) = 0, \forall j \neq 0, \text{against}$$
$$H_1\colon\ \Gamma_u(j) \neq 0, \text{for at least one } j \neq 0\ .$$

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

Our test statistic will be defined as a global distance measure between f_0 and $\hat{\mathrm{f}}_n$. For two multivariate spectral densities f_1 and f_2, a distance measure between f_1 and f_2 should be such that $D(\mathrm{f}_1;\mathrm{f}_2) \geq 0$ and $D(\mathrm{f}_1;\mathrm{f}_2) = 0$ if and only if $\mathrm{f}_1 = \mathrm{f}_2$. For a given covariance matrix $\Gamma_u(0)$, let us consider the following normalized quadratic distance

$$\begin{aligned} Q^2(\mathrm{f}_1;\mathrm{f}_2) &= 2\pi \int_{-\pi}^{\pi} \mathrm{vec}[\bar{\mathrm{f}}_1(\omega) - \bar{\mathrm{f}}_2(\omega)]'\Gamma_u^{-1}(0) \otimes \Gamma_u^{-1}(0)\, \mathrm{vec}[\mathrm{f}_1(\omega) - \mathrm{f}_2(\omega)]d\omega, \\ &= 2\pi \int_{-\pi}^{\pi} \mathrm{tr}[(\mathrm{f}_1(\omega) - \mathrm{f}_2(\omega))^*\Gamma_u^{-1}(0)(\mathrm{f}_1(\omega) - \mathrm{f}_2(\omega))\Gamma_u^{-1}(0)]d\omega, \\ &= 2\pi \int_{-\pi}^{\pi} \mathrm{tr}[\Gamma_u^{-1}(0)(\mathrm{f}_1(\omega) - \mathrm{f}_2(\omega))^*\Gamma_u^{-1}(0)(\mathrm{f}_1(\omega) - \mathrm{f}_2(\omega))]d\omega. \end{aligned} \tag{9}$$

For a matrix A, A^* denotes the transposed conjugate of A, that is $\mathrm{A}^* = \bar{\mathrm{A}}'$. The second equality is obtained from the following result on matrix calculus (Harville, 1997, Theorem 16.2.2):

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

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

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

is a normalized distance between the two matrices $\mathrm{f}_1(\omega)$ and $\mathrm{f}_2(\omega)$. The global distance $Q^2(\mathrm{f}_1;\mathrm{f}_2)$ is obtained by integrating $Q^2_\omega(\mathrm{f}_1;\mathrm{f}_2)$ over all possible frequencies in $[-\pi,\pi]$. When we compare the true spectral density f of u defined by (4) with f_0, the spectral density of u under H_0, we get the following result.

Proposition 1 *Let $Q^2(\mathrm{f};\mathrm{f}_0)$ be the distance measure given by (9), where* f *is defined by (4) and* $\mathrm{f}_0 = \Gamma_u(0)/(2\pi)$. *Then, we have*

$$Q^2(\mathrm{f};\mathrm{f}_0) = 2\sum_{h=1}^{\infty} \mathrm{tr}[\Gamma_u(h)\Gamma_u^{-1}(0)\Gamma_u(h)'\Gamma_u^{-1}(0)]. \tag{10}$$

Proof: If we reapply the argument followed to obtain (9), we can write $Q^2_\omega(\mathrm{f};\mathrm{f}_0) = \mathrm{tr}[\Gamma_u^{-1}(0)(\bar{\mathrm{f}}(\omega) - \bar{\mathrm{f}}_0(\omega))'\Gamma_u^{-1}(0)(\mathrm{f}(\omega) - \mathrm{f}_0(\omega))]$. Since $\Gamma_u(0)$ is positive definite, by the Cholesky factorization, a lower triangular matrix L exists such that $\Gamma_u^{-1}(0) = \mathrm{L}\mathrm{L}'$ and we have

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

where $\mathrm{f}_L = \mathrm{L}'\mathrm{f}\mathrm{L}$. Integrating $Q^2_\omega(\mathrm{f};\mathrm{f}_0)$ between $-\pi$ et π, we find that

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

We see from (10) that Q^2 is a global measure that takes into account all lags. With similar calculations, we can show that when f is replaced by $\hat{\mathrm{f}}_n$, we have

$$\begin{aligned} Q^2(\hat{\mathrm{f}}_n; \mathrm{f}_0) &= \sum_{j \le |n-1|} k^2(j/p_n)\, \mathrm{tr}[\mathrm{C}'_{\hat{u}}(j)\Gamma_u^{-1}(0)\mathrm{C}_{\hat{u}}(j)\Gamma_u^{-1}(0)] - 2\,\mathrm{tr}[\Gamma_u^{-1}(0)\mathrm{C}'_{\hat{u}}(0)] + d, \\ &= 2\sum_{j=1}^{n-1} k^2(j/p_n)\, \mathrm{tr}[\mathrm{C}'_{\hat{u}}(j)\Gamma_u^{-1}(0)\mathrm{C}_{\hat{u}}(j)\Gamma_u^{-1}(0)] + \\ &\quad \mathrm{tr}[\mathrm{C}'_{\hat{u}}(0)\Gamma_u^{-1}(0)\mathrm{C}_{\hat{u}}(0)\Gamma_u^{-1}(0)] - 2\,\mathrm{tr}[\Gamma_u^{-1}(0)\mathrm{C}'_{\hat{u}}(0)] + d. \end{aligned} \tag{11}$$

Now if we substitute $\mathrm{C}_u(0)$ for $\Gamma_u(0)$ in (11), it follows from (41) and (43) in the Appendix that

$$Q^2(\hat{\mathrm{f}}_n; \mathrm{f}_0) = 2\sum_{j=1}^{n-1} k^2(j/p_n)\, \mathrm{tr}[\mathrm{C}'_{\hat{u}}(j)\mathrm{C}_{\hat{u}}^{-1}(0)\mathrm{C}_{\hat{u}}(j)\mathrm{C}_{\hat{u}}^{-1}(0)] + o_p(\sqrt{p_n}/n). \tag{12}$$

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

$$T_n = \frac{n\sum_{j=1}^{n-1} k^2(j/p_n)\, \mathrm{tr}[\mathrm{C}'_{\hat{u}}(j)\mathrm{C}_{\hat{u}}^{-1}(0)\mathrm{C}_{\hat{u}}(j)\mathrm{C}_{\hat{u}}^{-1}(0)] - d^2 M_n(k)}{[2d^2 V_n(k)]^{1/2}} \tag{13}$$

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 (Robinson, 1994, 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 $O(p_n)$. Using the previous approximations, it is not difficult to see that

$$T_n = \frac{\frac{n}{2}Q^2(\hat{\mathrm{f}}_n; \mathrm{f}_0) - d^2 M_n(k)}{[2d^2 V_n(k)]^{1/2}} + o_p(1). \tag{14}$$

Also, in (13), we can replace $M_n(k)$ and $V_n(k)$ by $p_n M(k)$ and $p_n V(k)$ respectively without modifying the asymptotic distribution. However, these substitutions may lead to better finite sample approximations.

Using the truncated uniform kernel 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

$$T_n = \frac{H_{p_n} - d^2 p_n}{[2d^2 p_n]^{1/2}}, \tag{15}$$

where

$$H_{p_n} = n\sum_{j=1}^{p_n} \mathrm{tr}[\mathrm{C}'_{\hat{u}}(j)\mathrm{C}_{\hat{u}}^{-1}(0)\mathrm{C}_{\hat{u}}(j)\mathrm{C}_{\hat{u}}^{-1}(0)]. \tag{16}$$

When $p_n = M$ is fixed with respect to n, H_M is the multivariate version of the Box-Pierce statistic introduced by Hosking (1980). The latter (Hosking, 1981a) showed that H_M is equivalent to the multivariate portemanteau statistics proposed by Chitturi (1974) and Li and McLeod (1981). Also, Hosking (1981b) described how the statistic H_M may be viewed as a Lagrange multiplier test statistic for zero contraints 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 $\{\Gamma_u(k)\Gamma_u^{-1}(0), k \in \mathbb{Z}\}$ considered in Chitturi (1974), and the unstandardized quadratic norm. However, to avoid complications, we preferred to work with the usual unnormalized spectral density.

Our main result is stated in the following theorem. The symbol $\to_L$ stands for convergence in law.

Theorem 1 *Suppose that* y *is a* VARX(r, s) *process as defined by (1) and that its fourth order moments exist. Under assumptions A and B, the statistic* T_n *defined by (13) has an asymptotic normal distribution under* H_0*, that is* $T_n \to_L N(0,1)$.

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 detailled proof is technical and is presented in the Appendix. As in Hong (1996), it is written in two parts. First, we establish the asymptotic normality of a version $\tilde{T}_n$ of T_n, which is based on the unobservable innovation process u.

Part 1

$$\tilde{T}_n = \frac{n \sum_{j=1}^{n-1} k^2(j/p_n) \operatorname{tr}[\mathrm{C}_u^{-1}(0)\mathrm{C}_u(j)\mathrm{C}_u^{-1}(0)\mathrm{C}_u'(j)] - d^2 M_n(k)}{\sqrt{2d^2 V_n(k)}} \to_L N(0,1). \tag{17}$$

The *VARX* model does not intervene in this part since $\tilde{T}_n$ is completely defined by the innovation series $\mathrm{u}_1, \ldots, \mathrm{u}_n$. The asymptotic normality is derived from a central limit theorem for martingale differences. Here we use the version given by Brown (1971) although similar versions can be found in textbooks, namely in White (1984) and Hamilton (1994). A brief but enlightening discussion of martingale convergence theorems including Brown's central limit theorem is presented in Taniguchi and Kakizawa (2000; Chap. 1). The observed data and the model are taken into account in the second part.

Part 2

$$\sum_{j=1}^{n-1} k^2(j/p_n)\{\operatorname{tr}[\mathrm{C}_u^{-1}(0)\mathrm{C}_u(j)\mathrm{C}_u^{-1}(0)\mathrm{C}_u'(j)] - \operatorname{tr}[\mathrm{C}_{\hat{u}}^{-1}(0)\mathrm{C}_{\hat{u}}(j)\mathrm{C}_{\hat{u}}^{-1}(0)\mathrm{C}_{\hat{u}}'(j)]\} = o_p(\sqrt{p_n}/n). \tag{18}$$

From (18), it is easily seen that $\tilde{T}_n - T_n$ is $o_p(1)$ and Theorem 1 follows.

4. ASYMPTOTIC POWER OF THE GENERALIZED TEST

A test is consistent if its power tends to one as n goes to infinity, for any fixed significance level α. In our context, that means that a consistent test can detect any form of serial dependence. It is particularly useful when we do not have any information on the form of the alternative hypothesis. When the errors are serially correlated, the usual estimators of the coefficients in the *VARX* model (3) are in general inconsistent. It is therefore natural to wonder about the consistency of the proposed test which is based on the residuals from the estimated model. Here, we study its local power for a sequence of alternatives that converges to the null as n goes to infinity and also its global power for a fixed alternative.

4.1 Local asymptotic power

Let us consider a sequence of models for the white noise process u of the *VARX* models whose spectral densities $\mathrm{f}_{0n}(\omega)$ tends to f_0 as n goes to infinity, where f_0 is the spectral density under H_0. To simplify the discussion, we will assume that the white noise process u is as in Definition 1. We will suppose that the alternative hypothesis H_{1n} is of the form

$$H_{1n} : \mathrm{f}_{0n}(\omega) = \mathrm{f}_0 + a_n \mathrm{g}(\omega), \quad n \geq 1, \tag{19}$$

where g is a $d \times d$ matrix-valued function which is hermitian, periodic (of period 2π), continuous and such that $\int_{-\pi}^{\pi} \mathrm{g}(\omega) d\omega = 0$. The a_n's constitute a serie of non-negative numbers that converges to zero. As in the univariate case, if we choose the following separation rate between the null hypothesis and the alternative hypothesis

$$a_n = p_n^{1/4}/n^{1/2}, \tag{20}$$

we can establish the following result which generalizes a part of Theorem 4 in Hong (1996).

Theorem 2 *Let $\hat{\mathrm{f}}$ be the estimator defined by (8). Under assumptions A, B and under the sequence of alternatives H_{1n} defined by (19) with a_n given by (20), we have that*

$$T_n(\mathrm{f}_{0n}) = \frac{\frac{n}{2}Q^2(\hat{\mathrm{f}}_n; \mathrm{f}_{0n}) - d^2 M_n(k)}{\sqrt{2d^2 V_n(k)}} \rightarrow_L N(\mu(k), 1),$$

where $\mu(k) = 2\pi \int_{-\pi}^{\pi} \mathrm{tr}[\Gamma_u^{-1}(0)\mathrm{g}^*(\omega)\Gamma_u^{-1}(0)\mathrm{g}(\omega)]d\omega/[2d^2V(k)]^{1/2}$.

The proof is given in the Appendix. Since the test rejects for the large values of T_n, its asymptotic power under local alternatives is given by

$$\lim_{n\to\infty} Pr(T_n(\mathrm{f}_{0n}) > z_{1-\alpha}) = 1 - \Phi(z_{1-\alpha} - \mu(k)), \tag{21}$$

where $\Phi(\cdot)$ is the $N(0,1)$ cumulative distribution function and $z_{1-\alpha}$ denotes its $(1-\alpha)^{\text{th}}$ quantile. Note that the separation rate given by (20) is slower that the optimal parametric rate $n^{-1/2}$. This is the price to pay in order to obtain some power against a large class of serial correlation under the alternative hypothesis. We should note that price is however not too high, since for example if we let $a_n = n^{\nu/4-1/2}$, then for a relatively small value of ν, say $\nu \leq .2$, we obtain a rate only slightly slower that the parametric rate. Formula (21) being similar to the one derived by Hong (1996) for the ARX model, most of its results and discussion on the local power remain valid for the $VARX$ models. In fact, if we compare with the asymptotic power of the ARX models, only the denominator of $\mu(k)$ depends on the kernel k although its numerator is more complicated.

The Pitman asymptotic relative efficiency (ARE_p) (see for example Pitman (1979)) of a kernel k_2 with respect to the kernel k_1 is given by

$$ARE_P(k_2; k_1) = [V(k_1)/V(k_2)]^{1/(2-\nu)}, \tag{22}$$

if we take $p_n = cn^\nu$ with $0 < \nu < 1$. For example, $ARE_P(k_B; k_T) > 2.23$, where k_B is the Bartlett's kernel. and k_T is the truncated uniform kernel. Many other kernels deliver superior power to the truncated uniform kernel. While that provide interesting comparisons between the different kernels, formula (22) has been calculated assuming that each kernel use the same value of p_n. It is easy to show 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_P of the second kernel relatively to the first one will be zero, meaning that we should always prefer k_1 in such a situation. That remark suggests that we should use a sequence p_n going to infinity at a slower rate, and among the kernels, for a given choice of p_n, we should use a kernel different of the truncated uniform one. In the following class of sufficiently smooth kernels

$$\kappa(\tau) = \{k(\cdot) \,|\, k \text{ satisfies Assumption A with } k^{(2)} = \tau^2/2 > 0\}$$

where $k^{(r)} = \lim_{z\to 0}(1-k(z))/|z|^r$ is the characteristic exponent of k, it follows from Hong's Theorem 5 that the Daniell kernel defined by $k_D(z) = \sin(\sqrt{3}\tau z)/(\sqrt{3}\tau z)$, $z \in \mathbb{R}$, maximized the local power. This result is obtained by minimizing $V(k)$ in the class $\kappa(\tau)$. It is worth mentionning that the class $\kappa(\tau)$ contains Daniell, Parzen and the Bartlett-Priestley kernels among others but rules out the truncated uniform, Bartlett and general Tukey kernels.

4.2 Global asymptotic power

Although local power analyzes brings many useful insights on the asymptotic power of the test, it is far from being exhaustive. For example, it does not say anything on the consistency of the test for a fixed alternative. Here, we consider a fixed alternative H_A of serial dependence of the error u in the $VARX$ model (3) that satisfies the following properties.

Assumption C: Let the dependance structure of the process u be such that $\sum_j ||\Gamma_u(j)||^2 < \infty$ and suppose that the following cumulant condition is satisfied:

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

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

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

$$\mathrm{y}_t = \mathrm{c} + \mathrm{V}(B)\mathrm{x}_t + \mathrm{u}_t, \quad t \in \mathbb{Z}. \tag{23}$$

In this framework, we obtain the following result.

Theorem 3 *Let us consider the model (23), let T_n be the test statistic defined by (13) and suppose that assumptions A,B and C are satisfied. Then, under a fixed serial dependence alternative for the error process* u*, say H_A, we have that*

$$\frac{p_n^{1/2}}{n} T_n \to_P \frac{1}{2} Q^2(\mathrm{f}; \mathrm{f}_0)/[2d^2 D(k)]^{1/2} \tag{24}$$

where f *is the spectral density of* u *under H_A.*

The proof is in the Appendix. This result is a multivariate version of a part of Theorem 6 in Hong (1996). For any fixed alternative H_A, $Q^2(\mathrm{f}, \mathrm{f}_0) > 0$ and it follows from (24) 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 will be more powerful. Also, as illustrated by Hong (1996) for univariate models, Bahadur's (1960) asymptotic slope criterion can used to investigate the relative efficiency of one test with respect to another but we do not pursue in that direction here.

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, we examined the empirical frequencies of rejection of the null hypothesis when the latter is true by tests with three different nominal levels (1, 5 and 10 percent), for each of three series lengths (n=50, 100 and 200) and for five different kernels. Further, for each series length and for each kernel, three truncation values p_n were used. Finally, a data-driven method for the choice of p_n was applied with the non-negative definite kernels. The power properties of the tests were also investigated with the same bivariate VARX model when the error process is moving average of order one.

We also included the Hosking's (1980) multivariate portmanteau statistic defined by (16) and the modified version:

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

In VARMA models, the statistic H_{pn}^* is expected to have better level properties. See Hosking (1980) and Lüthepohl (1993, 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. It is then possible to compare all tests on an equal basis.

5.1 Description of the experiment

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

$$\mathrm{y}_t = \mathrm{c} + \Lambda_1 \mathrm{y}_{t-1} + \mathrm{V}_0 \mathrm{x}_t + \mathrm{V}_1 \mathrm{x}_{t-1} + \mathrm{a}_t, \tag{25}$$

where

$$\mathrm{x}_t = \Phi_x \mathrm{x}_{t-1} + \mathrm{b}_t. \tag{26}$$

The process $\{\mathrm{b}_t\}$ is a Gaussian white noise $N_2(0, \Sigma_\mathrm{b})$. Two cases were considered for the error term a_t: a) $\mathrm{a}_t = \mathrm{e}_t$ and b) $\mathrm{a}_t = \mathrm{e}_t - \Theta_\delta \mathrm{e}_{t-1}$ where $\{\mathrm{e}_t\}$ is another white noise $N_2(0, \Sigma_\mathrm{e})$ independent of $\{\mathrm{b}_t\}$. The

Table 1: Parameter values used in the simulation study for the VARX model (25) and (26).

$$\Lambda_1 = \begin{pmatrix} -0.5 & 0.5 \\ -1.4 & -0.2 \end{pmatrix}, \quad V_0 = \begin{pmatrix} 0.0 & 0.3 \\ 0.1 & 0.6 \end{pmatrix}, \quad V_1 = \begin{pmatrix} 0.7 & 0.0 \\ 0.0 & 0.0 \end{pmatrix},$$
$$\Phi_x = \begin{pmatrix} -1.5 & 1.2 \\ -0.9 & 0.5 \end{pmatrix}, \quad \Theta_\delta = \begin{pmatrix} 0.18\delta & 0.04\delta \\ 0.0 & 0.02\delta \end{pmatrix}, \quad c = \begin{pmatrix} 3.0 \\ 2.0 \end{pmatrix}.$$
$$\Sigma_e = \begin{pmatrix} 1.0 & 0.75 \\ 0.75 & 1.0 \end{pmatrix}, \quad \Sigma_b = \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}.$$

first case allowed us to study the level whilst the second one was chosen in order to study the power. The dependence structure of the a_t's depends on a parameter δ and we considered the following three values δ=0.75, 1.00 and 1.25. The values of the parameters in (25) and (26) used in the experiment are given in the Table 1.

In the level study, 10 000 independent realizations were generated from model (25) and (26) for each value of n anf the computations were made in the following way.

(1) The Gaussian white noise $\{a_t\}$ and $\{b_t\}$ were generated independently using the subroutine G05EZF from the NAG library.

(2) The initial values $\{x_0\}$ and $\{y_0\}$ were generated from the exact distribution of the stationary Gaussian process $\{(y_t', x_t')'\}$ using Ansley's (1980) algorithm. The values $x_t, y_t, t = 1, \ldots, n$, were obtained by solving the difference equations (25) and (26).

(3) For each realization, the true model (25) was estimated by generalized least squares as described in Section 2. The zero-valued parameters in V_0 and V_1 were taken into account by properly defining the constraint matrix R. The residuals $\hat{a}_t, t = 1, \ldots, n$, were obtained.

(4) With each residual series, the test statistic T_n was computed for five different kernels that are described in Table 2. For each kernel, four different values of p_n were considered. We have used 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 (1996). 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 extension by Robinson (1991) of the cross-validation procedure proposed by Beltrao and Bloomfield (1987) for determining the bandwidth of a kernel spectrum estimator of a 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

$$\sum_{j=1}^{n} [\log \det \hat{f}_{(j)}^{m}(\lambda_j) + \mathrm{tr}\{I(\lambda_j)\hat{f}_{(j)}^{m}(\lambda_j)^{-1}\}],$$

where $I(\cdot)$ represents the periodogram, $\hat{f}_{(j)}^{m}(\cdot)$ 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 $m = 2, 3, \ldots, 20$.

(5) 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 percent are reported in Table 3. The standard error of the empirical levels is 0.099% for the nominal level 1%, 0.218% for 5% and 0.300% for 10%.

Table 2: Kernels used in the calculation of the test statistic T_n defined by (13).

Truncated uniform (TR): $k(z) = \begin{cases} 1, & |z| \le 1, \\ 0, & \text{otherwise.} \end{cases}$

Bartlett (BAR): $k(z) = \begin{cases} 1 - |z|, & |z| \le 1, \\ 0, & \text{otherwise.} \end{cases}$

Daniell (DAN): $k(z) = \dfrac{\sin(\pi z)}{\pi z}, \ z \in \mathbb{R}.$

Parzen (PAR): $k(z) = \begin{cases} 1 - 6(\pi z/6)^2 + 6|\pi z/6|^3, & |z| \le 3/\pi, \\ 2(1 - |z|)^3, & 3/\pi \le |z| \le 6/\pi, \\ 0, & \text{otherwise.} \end{cases}$

Bartlett-Priestley (BP): $k(z) = \dfrac{9}{5\pi^2 z^2}\{\dfrac{\sin(\pi\sqrt{5/3}z)}{\pi\sqrt{5/3}z} - \cos(\pi\sqrt{5/3}z)\}, z \in \mathbb{R}.$

The power analysis was conducted in a similar way. The two main differences rely in the number of realizations (2 000 rather than 10 000) and the process $\{a_t\}$ is MA(1) rather than white noise. Three sets of parameters values were considered for the MA(1) model.

5.2 Discussion of the level study

Results from the level study are presented in Table 3. 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 overrejects 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 (1980) test H and its modified version HM clearly over-reject for small p_n, and it seems that an additional adjustement 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 for the values of 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 overrejects slightly more when $n = 50$.

At the 5% level, all kernels (with a fixed value of p_n) lead to rejection rates close to 7% when $n = 50$, between 5.5 and 6.3 when $n = 100$. At $n = 200$, all rejection rates are within two standard errors of 5%. 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 rejection rates that are within two standard errors of 10%. The cross-validation method tends to slightly overreject at $n = 50$ but works reasonably well when $n = 100$ or 200.

5.3 Discussion of the power study

When the error term satisfies $a_t = e_t - \Theta_\delta e_{t-1}$, 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 and the results are presented in Table 4. The powers based on empirical critical values are given in parentheses.

Results in Table 4 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, for $n = 100$ and $n = 200$, 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 phenomena 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

Table 3: Empirical levels (in percentage) of Hosking's test and of the test statistic T_n defined by (13) for different kernels, different truncation values, when the data are generated from model (25) and (26).

	$\alpha = 0.01$							$\alpha = 0.05$							$\alpha = 0.10$						
$n = 50$																					
p_n	BP	BAR	DAN	PAR	TR	H	HM	BP	BAR	DAN	PAR	TR	H	HM	BP	BAR	DAN	PAR	TR	H	HM
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
$n = 100$																					
p_n	BP	BAR	DAN	PAR	TR	H	HM	BP	BAR	DAN	PAR	TR	H	HM	BP	BAR	DAN	PAR	TR	H	HM
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
$n = 200$																					
p_n	BP	BAR	DAN	PAR	TR	H	HM	BP	BAR	DAN	PAR	TR	H	HM	BP	BAR	DAN	PAR	TR	H	HM
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

powerful than the others. It is not in contradiction with the asymptotic power analysis described in Section 4.1, since the assumption made there to conclude that Daniell kernel is optimal rules out Bartlett kernel. This is also in agreement with Hong's (1996) analysis for univariate ARX models.

For the new tests, the results based on empirical and asymptotic quantiles differ considerably at the 1% level when $n = 50$. However, they are closer at the 5% and 10% levels. That difference decreases as n increases, which is not surprising since the level is better controled for large values of n. Since Hosking's test (1980) 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 (1980) test has a lower power than the proposed tests. Indeed, the test H and and the new test based on the truncated uniform kernel 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 that H.

Since the dependence of the errors is of order one, we expect that the tests putting more weight on smaller lags will be more powerful that those putting weight on a large number of lags. This is confirmed by our study since a smaller value of p_n leads to a greater power. The cross-validation procedure of Robinson (1991) seems to work very well here since the resulting power is higher. It leads to a good compromise between errors of types I and II. Finally, 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 other than the truncated uniform one seems appropriate.

6. APPLICATION

We illustrate here the new serial correlation test with a real data set coming from the economic literature. It was first analysed by Gilbert (1993) who aimed at developing a simple econometric monthly model for monitoring the Canadian economy. He estimated a VARX model in which the explanatory variable is the 90-day prime corporate interest rate (R90) and the dependent variables are the following three seasonally adjusted economic indicators: monetary basis value (M1), gross domestic production (GDP) and consumer price index (CPI). The observation period corresponds to a time span of 20 years, from March 1961 through March 1981 (n=240). The data source is the CANSIM database from Statistics Canada.

Table 4: Power based on the asymptotic and empirical (in parentheses) critical values of Hosking's test and of the test T_n define by (13) for different kernels, different truncation values when the data are generated from model (25) and (26) with MA(1) errors.

	$\alpha = 0.01$							$\alpha = 0.05$							$\alpha = 0.10$						
$n = 50$																					
p_n	BP	BAR	DAN	PAR	TR	H	HM	BP	BAR	DAN	PAR	TR	H	HM	BP	BAR	DAN	PAR	TR	H	HM
4	11.7	12.0	11.9	10.9	5.6	7.8	10.9	22.3	22.5	22.2	21.5	14.1	26.2	31.8	29.4	30.0	30.0	28.8	21.4	40.4	46.2
	(6.3)	(6.3)	(6.2)	(5.9)	(3.2)	(3.2)	(3.2)	(18.7)	(19.1)	(18.9)	(17.7)	(10.9)	(10.9)	(10.9)	(29.0)	(30.0)	(29.5)	(28.4)	(19.6)	(19.6)	(19.4)
8	7.7	8.7	7.9	7.7	6.0	3.9	8.1	16.0	18.0	16.1	15.9	13.2	12.6	20.8	25.0	27.0	25.0	24.0	19.6	20.4	33.6
	(4.3)	(5.3)	(4.0)	(3.8)	(2.4)	(2.4)	(2.1)	(13.5)	(14.4)	(13.5)	(12.3)	(9.5)	(9.5)	(9.4)	(23.7)	(25.1)	(23.5)	(22.1)	(17.0)	(17.0)	(16.7)
10	7.4	8.1	7.4	7.0	6.0	2.8	7.8	15.5	16.8	15.7	15.1	13.0	9.1	19.1	23.6	25.2	23.4	22.8	19.0	16.0	30.0
	(3.7)	(4.6)	(3.6)	(3.4)	(2.5)	(2.5)	(2.4)	(12.1)	(13.0)	(12.2)	(11.6)	(9.0)	(9.0)	(8.8)	(20.6)	(23.6)	(20.7)	(20.3)	(16.5)	(16.5)	(15.7)
CV	14.0	15.0	13.5	12.4	NA	NA	NA	25.4	27.2	24.7	24.6	NA	NA	NA	34.1	35.1	33.6	33.5	NA	NA	NA
	(6.7)	(6.3)	(6.7)	(5.4)	NA	NA	NA	(19.4)	(19.4)	(19.4)	(19.1)	NA	NA	NA	(31.6)	(31.2)	(31.0)	(30.3)	NA	NA	NA
$n = 100$																					
p_n	BP	BAR	DAN	PAR	TR	H	HM	BP	BAR	DAN	PAR	TR	H	HM	BP	BAR	DAN	PAR	TR	H	HM
5	27.7	30.3	28.3	26.1	12.5	16.6	18.6	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
	(18.7)	(20.4)	(18.6)	(17.5)	(7.1)	(7.1)	(7.0)	(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	18.9	22.2	18.9	17.9	9.6	8.9	11.6	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
	(11.9)	(14.5)	(11.9)	(10.8)	(4.8)	(4.8)	(4.8)	(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	15.8	19.2	16.1	14.7	8.3	5.9	10.0	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
	(9.0)	(11.7)	(8.9)	(8.0)	(4.5)	(4.5)	(4.3)	(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	38.7	39.7	36.2	31.9	NA	NA	NA	52.2	53.9	51.0	49.0	NA	NA	NA	61.1	62.3	60.1	60.5	NA	NA	NA
	(23.6)	(22.5)	(22.4)	(18.7)	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
$n = 200$																					
p_n	BP	BAR	DAN	PAR	TR	H	HM	BP	BAR	DAN	PAR	TR	H	HM	BP	BAR	DAN	PAR	TR	H	HM
6	63.5	67.6	63.8	61.4	30.0	35.6	37.6	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
	(53.4)	(55.8)	(52.8)	(49.7)	(21.9)	(21.9)	(21.4)	(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	48.2	56.3	48.8	45.3	20.7	22.0	25.9	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
	(38.5)	(46.1)	(37.6)	(36.6)	(15.3)	(15.3)	(14.7)	(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	36.2	45.3	36.8	33.8	16.7	14.4	19.2	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
	(28.9)	(37.0)	(29.0)	(27.1)	(11.9)	(11.9)	(10.8)	(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	76.9	79.1	74.9	72.0	NA	NA	NA	87.5	89.4	84.8	85.5	NA	NA	NA	90.8	92.7	89.8	91.4	NA	NA	NA
	(62.7)	(62.6)	(60.5)	(55.5)	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

Gilbert (1993) considered the following model

$$(I - \Lambda_1 B - \Lambda_2 B^2 - \Lambda_3 B^3)\mathrm{Y}_t = (\mathrm{V}_0 + \mathrm{V}_1 B + \mathrm{V}_2 B^2) X_t + \mathrm{u}_t, \tag{27}$$

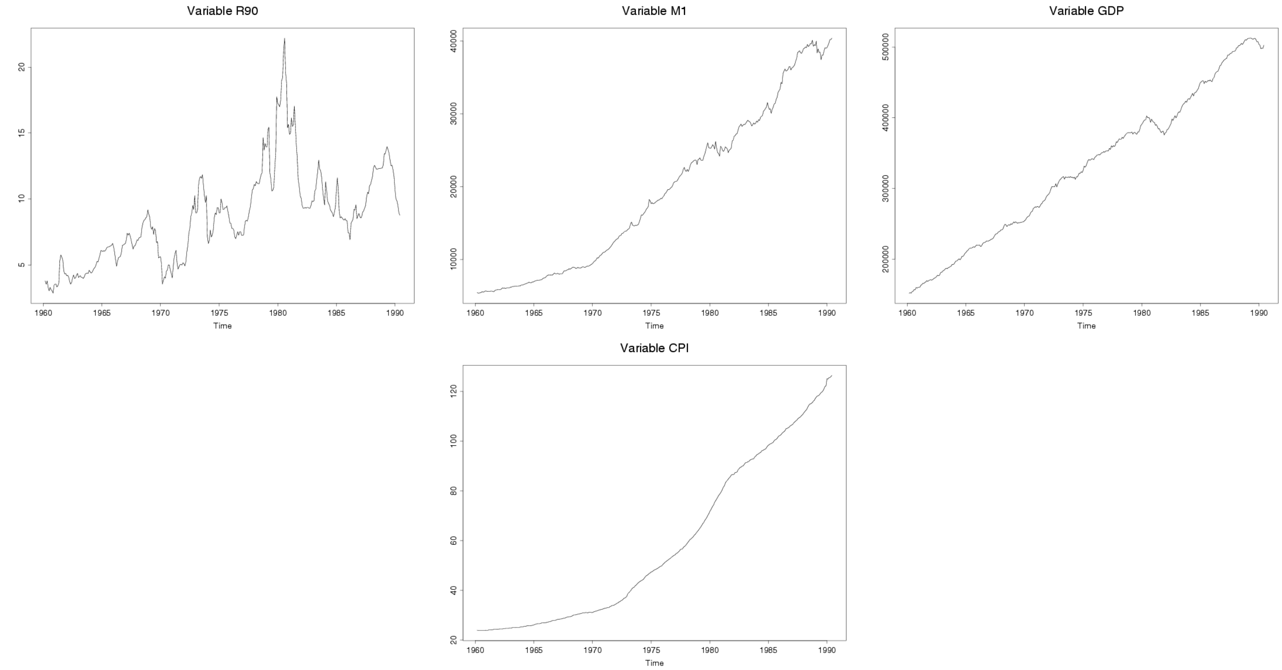
where

$$\mathrm{Y}_t = \begin{bmatrix} 100(1-B)\log(M1_t) \\ 100(1-B)\log(GDP_{t-2}) \\ 100(1-B)\log(CPI_t) \end{bmatrix}, \; X_t = R90_t.$$

A logarithmic transformation was applied to $M1$, GDP and CPI in order to stabilize variances and all three series were differenced to have stationarity. Furthermore, the first difference of $\log(M1)$, $\log(GDP)$ and $\log(CPI)$ were multiplied by 100 to ensure that their sample variances be of the same order of magnitude as the one of $R90$. Gilbert (1993) justified the used of lagged-2 GDP by the delay in the availability of information and chose to used the most recent observation available. However, as he mentioned, the economic interpretation that interest rates can affect GDP two periods in advance of when they are set is questionable unless we presume that economic agents can anticipate interest rate changes.

Model (27) was estimated by the least squares method described in Section 2. In a first step, the full model was estimated and each parameter whose estimate was smaller than one time its standard error was set to zero. In the second step, the reduced model was estimated with a judicious choice of the matrix of linear constraints R. In our situation, R is a 36×24 matrix in which the rows 2, 3, 7, 15, 16, 22, 26, 28, 29, 32, 33 and 36 are null. Starting from the first, the other rows are given by $e_1', e_2', \dots, e_{24}'$, where e_i is the vector with one in position i and zero elsewhere. The final model is given in table 5. Standard errors of the estimators are given in parentheses.

Figure 1: The four economic variables.



The residual covariance matrix is given by

$$\begin{bmatrix} .9960 & .0688 & .01342 \\ .0688 & .3513 & .02090 \\ .01342 & .02090 & .08616 \end{bmatrix}.$$

The values of the portmanteau statistics with Bartlett-Priestley, Bartlett, Daniell, Parzen and truncated uniform kernels are given in Table 6. With each kernel, T_n was computed with $p_n = 5, 9, 15$ and with the estimated value by the cross-validation procedure.

As in the simulation study, the choice of the kernel (except TR) does not affect much the value of T_n. However, the choice of p_n is much more critical and the values obtained with $p_n = 5, 9, 15$ are quite different. It illustrates the importance of having a method for estimating p_n from the data. Here, the cross-validation procedure leads to $p_n = 2$ for each kernel.

At the 5% significance level, the null hypothesis of independent errors in model (27) is rejected if $T_n > 1.645$, the 95% quantile of the $N(0, 1)$ distribution. Therefore, our analysis supports Gilbert's model with the considered data.

Finally, in an other analysis not reported here, we found that a VARX(3,2) model seems also appropriate when the exogenous variable is differenced. However, this last model does not lead exactly to the same interpretation than the one of Gilbert on the relations between the input and output variables.

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 (1980), which can be viewed as a test based on the truncated uniform kernel.

In the power analysis, we first established the asymptotic normality of the test statistic under a sequence of local alternatives that converges to the null hypothesis as n gets large. From this last result, the Pitman asymptotic relative efficiency (ARE) of one kernel with respect to another one is derived. It is also seen that the Daniell kernel is optimal in a given class of sufficiently smooth kernels. Of our analysis, we conclude

Table 5: VARX model (26) estimated with the data of Gilbert (1993), (estimated standard errors are given in parentheses).

$$\begin{bmatrix} 1-.144B+.0626B^2-.113B^3 & -.291B-.260B^2 & -.862B^3 \\ (.0618) \quad (.0636) \quad (.0623) & (.1070) \quad (.106) & (.2160) \\ -.0270B^2-0.07B^3 & 1+.0844B-.170B^2-.214B^3 & .144B+.249B^2 \\ (.0371) \quad (.0369) & (.0631) \quad (.062) \quad (.0632) & (.1250) \quad (.1250) \\ -.0214B^2-.0149B^3 & -.0607B+.097B^3 & 1-.123B-.260B^2-.0955B^3 \\ (.0183) \quad (.0183) & (.0315) \quad (.0314) & (.0641) \quad (.0631) \quad (.0648) \end{bmatrix} \mathrm{Y}_t$$

$$= \begin{bmatrix} -.288+.271B^2 \\ (.0746) \quad (.0724) \\ .0484 \\ (.0123) \\ .0343B \\ (.00647) \end{bmatrix} X_t + \mathrm{u}_t$$

Table 6: Values of the portmanteau statistics with different kernels

	$p_n = 5$	$p_n = 9$	$p_n = 15$	CV
BP	-1.617	-.717	.088	-.883
BAR	-1.629	-.937	-.250	-.912
DAN	-1.580	-.670	.195	-.727
PAR	-1.481	-.517	.216	-.835
TR	.115	.855	1.113	

that it is possible to obtain a greater asymptotic power by using a kernel different of the truncated uniform kernel which leads to Hosking's statistic, in many situations. As illustrated by a simulation study, that latter property also holds in finite samples, at least for the chosen model.

In the simulation experiment, the properties of the new test were investigated for several kernels and several values of the truncation parameter p_n. We also applied the cross-validation method described in Robinson (1991) for choosing p_n. For all kernels considered, the level of the test is reasonably well controled 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. Finally, the cross-validation procedure for choosing p_n works very well here since the resulting power is in general higher. In most practical situations, the new test based on Bartlett or Daniell kernels with p_n chosen by cross-validation should be appropriate. The importance of using a data-driven method for choosing p_n is also illustrated with a real data set.

APPENDIX

Proof of the theorem 1

The following notations are adopted. The scalar product of $\mathrm{x}_t, \mathrm{x}_s \in \mathbb{R}^n$ is denoted by $\langle \mathrm{x}_t, \mathrm{x}_s \rangle = \mathrm{x}_t'\mathrm{x}_s$ and the Euclidian norm of x_t by $||\mathrm{x}_t|| = \sqrt{\langle \mathrm{x}_t, \mathrm{x}_t \rangle}$ The matrix norm used is the Euclidian matrix norm defined by $||\mathrm{A}||_E^2 = \mathrm{tr}(\mathrm{AA}') = \sum_{i=1}^n \sum_{j=1}^n a_{ij}^2$, where $\mathrm{A} = (a_{ij})_{n\times n}$. The notations O_p and o_p are the usual notations for orders in probability and many properties can be found in Fuller (1996). Let $k_{nj} = k(j/p_n)$, $\mathrm{v}_t = \Sigma_u^{-1/2}\mathrm{u}_t$ and $\Sigma_u = \Gamma_u(0)$. The process $\mathrm{v} = \{\mathrm{v}_t : t \in \mathbb{Z}\}$ has mean 0 and variance 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 (chap. 5 and 6). Let A and B and C be any matrices, D and E be symmetric positive definite matrices. Then we have

$$|\,\mathrm{tr}(\mathrm{AB}')| \leq \sqrt{\mathrm{tr}(\mathrm{AA}')}\sqrt{\mathrm{tr}(\mathrm{BB}')}, \tag{28}$$

$$\mathrm{tr}(\mathrm{D}^2) \leq (\mathrm{tr}(\mathrm{D}))^2, \tag{29}$$

$$\mathrm{tr}(\mathrm{DE}) \leq \mathrm{tr}(\mathrm{D})\,\mathrm{tr}(\mathrm{E}), \tag{30}$$

$$\mathrm{tr}(\mathrm{AD}) \leq \sqrt{\mathrm{tr}(\mathrm{AA}')}\,\mathrm{tr}(\mathrm{D}), \tag{31}$$

$$|\,\mathrm{tr}(\mathrm{A}'\mathrm{BAB})| \leq \mathrm{tr}(\mathrm{A}'\mathrm{A})\,\mathrm{tr}(\mathrm{B}'\mathrm{B}), \tag{32}$$

$$\mathrm{tr}[(\mathrm{A}+\mathrm{B}+\mathrm{C})(\mathrm{A}+\mathrm{B}+\mathrm{C})'] \leq 4[\mathrm{tr}(\mathrm{AA}') + \mathrm{tr}(\mathrm{BB}') + \mathrm{tr}(\mathrm{CC}')]. \tag{33}$$

We now prove part 1. First note that $\mathrm{C}_u(0) - \Sigma_u = O_p(n^{-1/2})$ 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 $\mathrm{C}_u^{-1}(0) - \Sigma_u^{-1} = O_p(n^{-1/2})$. We will show that asymptotically, $\mathrm{C}_u(0)$ can be replaced by Σ_u in (17), and in the sequel we will work with the version involving Σ_u.

Result 1

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

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

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

Proof: We have

$$\begin{aligned}\mathrm{C}_v(j)\mathrm{C}_v'(j) &= n^{-2}\sum_{t=j+1}^{n} ||\mathrm{v}_{t-j}||^2\mathrm{v}_t\mathrm{v}_t' + n^{-2}\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1} \mathrm{v}_t\mathrm{v}_{t-j}'\mathrm{v}_{s-j}\mathrm{v}_s' + \\ &\quad n^{-2}\sum_{t=j+1}^{n-1}\sum_{s=t+1}^{n} \mathrm{v}_t\mathrm{v}_{t-j}'\mathrm{v}_{s-j}\mathrm{v}_s'.\end{aligned}$$

Taking expected values on both sides, it is easily seen that $E[\mathrm{C}_v(j)\mathrm{C}_v'(j)] = n^{-2}(n-j)d\mathrm{I}_d$. It follows that $E[\sum_{j=1}^{n-1} k_{jn}^2 \mathrm{C}_v(j)\mathrm{C}_v'(j)] = n^{-1}d\sum_{j=1}^{n-1}(1-j/n)k^2(j/p_n) = O(p_n/n)$.

To show result 1, note that

$$\begin{aligned}\mathrm{C}_u'(j)\mathrm{C}_u^{-1}(0)\mathrm{C}_u(j)\mathrm{C}_u^{-1}(0) &= \mathrm{C}_u'(j)\Sigma_u^{-1}\mathrm{C}_u(j)\Sigma_u^{-1} + \mathrm{C}_u'(j)\Delta_{un}\mathrm{C}_u(j)\Sigma_u^{-1} + \\ &\quad \mathrm{C}_u'(j)\Sigma_u^{-1}\mathrm{C}_u(j)\Delta_{un} + \mathrm{C}_u'(j)\Delta_{un}\mathrm{C}_u(j)\Delta_{un},\end{aligned}$$

where $\mathrm{C}_u^{-1}(0) - \Sigma_u^{-1} = \Delta_{un}$. Then it is sufficient to multiply by $k^2(j/p_n)$, to sum on j, to apply the tr operator, use (31) and (32), $p_n/n \to 0$ and Lemma 1.

Result 1 allows us to work with $\sum_{j=1}^{n-1} k^2(j/p_n)\,\mathrm{tr}[\Sigma_u^{-1}\mathrm{C}_u(j)\Sigma_u^{-1}\mathrm{C}_u'(j)]$ that we now decompose in two parts A_{1n} and A_{2n}:

$$\begin{aligned}\sum_{j=1}^{n-1} k_{nj}^2\,\mathrm{tr}[\mathrm{C}_u'(j)\Sigma_u^{-1}\mathrm{C}_u(j)\Sigma_u^{-1}] &= n^{-2}\sum_{j=1}^{n-1} k_{nj}^2\{\sum_{t=j+1}^{n} ||\mathrm{v}_{t-j}||^2||\mathrm{v}_t||^2\} + \\ &\quad n^{-2}\sum_{j=1}^{n-2} k_{nj}^2\{\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1} 2\langle \mathrm{v}_{s-j},\mathrm{v}_{t-j}\rangle\langle \mathrm{v}_t,\mathrm{v}_s\rangle\}, \\ &= n^{-2}\sum_{j=1}^{n-1} k_{nj}^2\{\sum_{t=j+1}^{n} Z_{jt}^2\} + \\ &\qquad n^{-2}\sum_{j=1}^{n-2} k_{nj}^2\{\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1} w_{jts}\}, \\ &= n^{-1}(A_{1n} + A_{2n}),\end{aligned}$$

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

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

To show result 2, note that $E(A_{1n}) = d^2M_n(k)$ and $\mathrm{var}(A_{1n}) = O(p_n^2/n)$, since

$$\begin{aligned}\mathrm{var}(A_{1n}) &= E\{n^{-1}\sum_{j=1}^{n-1} k_{nj}^2[\sum_{t=j+1}^{n}(Z_{jt}^2 - d^2)]\}^2, \\ &\le n^{-2}\{\sum_{j=1}^{n-1} k_{nj}^2[E(\sum_{t=j+1}^{n}(Z_{jt}^2 - d^2))^2]^{1/2}\}^2 = O(p_n^2/n).\end{aligned}$$

The last relation is obtained with the following lemma.

Lemma 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 2 follows since $E[(Z_{jt}^2 - d^2)^2] = E(||\mathrm{v}_1||^4)^2 - d^4$, and

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

Result 2 is obtained using Proposition 6.2.4 in Brockwell and Davis (1991). To complete the proof of Part 1, we have to show that $(2d^2V_n(k))^{-1/2}A_{2n} \to_L N(0,1)$. 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} in several parts and we eliminate the terms that are $o_p(1)$. Let us write $A_{2n} = B_n + \sum_{i=1}^4 C_{in}$, where

$$\begin{aligned} B_n &= n^{-1}\sum_{j=1}^{l_n} k_{nj}^2\{\sum_{t=2l_n+3}^{n}\sum_{s=l_n+2}^{t-l_n-1} w_{jts}\}, \\ C_{1n} &= n^{-1}\sum_{j=l_n+1}^{n-2} k_{nj}^2\{\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1} w_{jts}\}, \\ C_{2n} &= n^{-1}\sum_{j=1}^{l_n} k_{nj}^2\{\sum_{t=2l_n+3}^{n}\sum_{s=t-l_n}^{t-1} w_{jts}\}, \\ C_{3n} &= n^{-1}\sum_{j=1}^{l_n} k_{nj}^2\{\sum_{t=l_n+3}^{2l_n+2}\sum_{s=l_n+2}^{t-1} w_{jts}\}, \\ C_{4n} &= n^{-1}\sum_{j=1}^{l_n} k_{nj}^2\{\sum_{t=j+1}^{l_n+1}\sum_{s=t+1}^{n} w_{jts}\}. \end{aligned}$$

The following lemma is useful. It generalizes a result in Hong (1996) 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 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*

$$E(w_{j_1t_1s_1}^{l_1l_2l_3l_4} w_{j_2t_2s_2}^{m_1m_2m_3m_4}) = \begin{cases} E(w_{j_1t_1s_1}^{l_1l_2l_3l_4} w_{j_2t_2s_2}^{m_1m_2m_3m_4})\delta_{t_1,t_2}\delta_{s_1,t_1-j_2}\delta_{s_2,t_1-j_1}, & j_1 \neq j_2, \\ E(w_{j_1t_1s_1}^{l_1l_2l_3l_4} w_{j_2t_2s_2}^{m_1m_2m_3m_4})\delta_{t_1,t_2}\delta_{s_1,s_2}, & j_1 = j_2. \end{cases}$$

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

Proof: Let us begin with 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 3, we can show that

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

since $p_n^{-1}\sum_{j=l_n+1}^{n-2} k_{nj}^4 \to 0$ and $p_n/n \to 0$. By similar arguments, we can show that $E(C_{2n}^2) = O(\frac{l_n p_n}{n} + \frac{p_n^2}{n})$, $E(C_{3n}^2) = O(\frac{l_n^2 p_n}{n^2} + \frac{p_n^2 l_n}{n^2})$ and $E(C_{4n}^2) = O(\frac{l_n p_n}{n} + \frac{p_n^2}{n})$, and the announced result is proved.

Result 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 $\sigma^{-2}(n)B_n \to_L N(0,1)$, 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 $B_{nt} = 2\mathrm{v}_t'\{\sum_{j=1}^{l_n} k_{nj}^2 H_{j,t-l_n-1}\mathrm{v}_{t-j}\}$, and $H_{j,t-l_n-1} = \sum_{s=l_n+2}^{t-l_n-1}\mathrm{v}_s\mathrm{v}_{s-j}'$. Note that $\{B_{nt}, \mathcal{F}_{t-1}\}$ is a martingale difference since $E(B_{nt}) = 0$ and $E(B_{nt}|\mathcal{F}_{t-1}) = 0$, where $\mathcal{F}_{t-1}$ is the σ-algebra generated by v_s, $s \leq t$.

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

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

From the previous lemma, we obtain that

$$\begin{aligned}\sigma^2(n) = E(B_n^2) &= n^{-2}\sum_{t=2l_n+3}^{n} E(B_{nt}^2) = 2d^2\frac{(n-2l_n-2)(n-2l_n-1)}{n^2}\sum_{i=1}^{l_n} k_{ni}^4, \\ &= 2d^2 p_n V(k)[1+o(1)].\end{aligned}$$

Result 4 $\sigma^{-1}(n)B_n \rightarrow_L N(0,1)$.

Proof: To apply the central limit theorem of Brown (1971), we have to verify the following two conditions:

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

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

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

We begin with a). It suffices to show that Lyapounov condition is verified. Since $|\langle x, y\rangle| \leq ||x|| \times ||y||$, we have that $|B_{nt}| \leq 2||\mathrm{v}_t|| \times ||\sum_{j=1}^{l_n} k_{nj}^2 H_{j,t-l_n-1}\mathrm{v}_{t-j}||$, and we then obtain

$$E(B_{nt}^4) \leq 16\mu_4(||\mathrm{v}||)E(||\sum_{j=1}^{l_n} k_{nj}^2 H_{j,t-l_n-1}\mathrm{v}_{t-j}||^4),$$

since v_{t-i} is independent of $H_{j,t-l_n-1}$, $1 \leq i,j \leq l_n$. Note that $E(||x||^4) \leq d\sum_{i=1}^{d} E[x^4(i)]$, where $x = (x(1),\ldots,x(d))'$ is a vector of dimension d. Since the l-th component of $\sum_{j=1}^{l_n} k_{nj}^2 H_{j,t-l_n-1}\mathrm{v}_{t-j}$ is given by $\sum_{j=1}^{l_n} k_{nj}^2 \sum_{s=l_n+2}^{t-l_n-1} v_s(l)\langle \mathrm{v}_{s-j}, \mathrm{v}_{t-j}\rangle$, we will make use of the following lemma.

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

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

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

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

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

Since $E(v_s(l)\langle \mathrm{v}_{s-j}, \mathrm{v}_{t-j}\rangle)^4 \leq \mu_4^3(||\mathrm{v}||)$, it follows that $E[\sum_{s=l_n+2}^{t-l_n-1} v_s(l)\langle \mathrm{v}_{s-j}, \mathrm{v}_{t-j}\rangle]^4 = O(t^2)$, independently of l. Regrouping the various results, we obtain that

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

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 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*

$$E[(\sum_{i=1}^{n} X_i)^4] \leq 3\{\sum_{i=1}^{n}[E(X_i^4)]^{1/2}\}^2.$$

To show b), it is sufficient to prove that $\sigma^{-4}(n)E([\ddot{B}_n^2 - \sigma^2(n)]^2) \to 0$, where $\ddot{B}_n^2 = E(B_n^2|\mathcal{F}_{t-1}) = n^{-2}\sum_{t=2l_n+3}^{n} \ddot{B}_{nt}^2$. 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

$$
\begin{aligned}
D_{1nt} &= 2\sum_{j=2}^{l_n}\sum_{i=1}^{j-1} k_{ni}^2 k_{nj}^2 \mathrm{v}_{t-i}' H_{i,t-l_n-1}' H_{j,t-l_n-1}\mathrm{v}_{t-j}, \\
D_{2nt} &= 2\sum_{i=1}^{l_n} k_{ni}^4 \sum_{s_1=l_n+3}^{t-l_n-1}\sum_{s_2=l_n+2}^{s_1-1} \mathrm{v}_{t-i}'\mathrm{v}_{s_1-i}\mathrm{v}_{s_1}'\mathrm{v}_{s_2}\mathrm{v}_{s_2-i}'\mathrm{v}_{t-i}, \\
D_{3nt} &= \sum_{i=1}^{l_n} k_{ni}^4 \mathrm{v}_{t-i}'[\sum_{s=l_n+2}^{t-l_n-1}(\mathrm{v}_{s-i}\mathrm{v}_s'\mathrm{v}_s\mathrm{v}_{s-i}' - dI_d)]\mathrm{v}_{t-i}, \\
D_{4nt} &= d(t-2l_n-2)\sum_{i=1}^{l_n} k_{ni}^4(\mathrm{v}_{t-i}'\mathrm{v}_{t-i} - d).
\end{aligned}
$$

We now prove the two following lemmas.

Lemma 7

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

Proof: First, let us consider D_{1nt}. Let $\mathrm{a}_j = H_{j,t-l_n-1}\mathrm{v}_{t-j}$. We have that

$$
\begin{aligned}
E(D_{1nt}^2) &= 4E(\sum_{j=2}^{l_n}\sum_{i=1}^{j-1} k_{ni}^4 k_{nj}^4 |\langle \mathrm{a}_i, \mathrm{a}_j\rangle|^2), \\
&\le 4\sum_{l_1,l_2=1}^{d}\sum_{j=2}^{l_n}\sum_{i=1}^{j-1} k_{ni}^4 k_{nj}^4 \{E[a_i^4(l_1)]E[a_j^4(l_2)]\}^{1/2},
\end{aligned}
$$

using Cauchy-Schwarz inequality, where $a_i(l) = \sum_{s=l_n+2}^{t-l_n-1} v_s(l)\langle \mathrm{v}_{s-j}, \mathrm{v}_{t-j}\rangle$. We showed in the proof of Lemma 5 that $E[a_i^4(l)] = O(t^2)$. Therefore, $E(D_{1nt}^2) = O(t^2p_n^2)$. For D_{2nt}, first note that

$$D_{2nt} = \sum_{l_1,l_2=1}^{d}\sum_{i=1}^{l_n} k_{ni}^4 b_{t-i}^{l_1l_2} \sum_{s=l_n+3}^{t-l_n-1}\sum_{r=l_n+2}^{s-1} w_{isr}^{l_1l_2}.$$

where $b_t^{l_1l_2} = \mathrm{v}_t(l_1)\mathrm{v}_t(l_2)$. Thus, we have

$$
\begin{aligned}
D_{2nt}^2 &= \sum_{l_1,l_2,m_1,m_2=1}^{d}\sum_{i_1,i_2=1}^{l_n} k_{ni_1}^4 k_{ni_2}^4 b_{t-i_1}^{l_1l_2} b_{t-i_2}^{m_1m_2} \sum_{s_1,s_2=l_n+3}^{t-l_n-1}\sum_{r_1=l_n+2}^{s_1-1}\sum_{r_2=l_n+2}^{s_2-1} w_{i_1s_1r_1}^{l_1l_2} w_{i_2s_2r_2}^{m_1m_2}, \\
&= \sum_{l_1,l_2,m_1,m_2=1}^{d}\sum_{i=1}^{l_n} k_{ni}^8 b_{t-i}^{l_1l_2} b_{t-i}^{m_1m_2} \sum_{s_1,s_2=l_n+3}^{t-l_n-1}\sum_{r_1=l_n+2}^{s_1-1}\sum_{r_2=l_n+2}^{s_2-1} w_{is_1r_1}^{l_1l_2} w_{is_2r_2}^{m_1m_2} + \\
&\quad 2\sum_{l_1,l_2,m_1,m_2=1}^{d}\sum_{i_1=2}^{l_n}\sum_{i_2=1}^{i_1-1} k_{ni_1}^4 k_{ni_2}^4 b_{t-i_1}^{l_1l_2} b_{t-i_2}^{m_1m_2} \sum_{s_1,s_2=l_n+3}^{t-l_n-1}\sum_{r_1=l_n+2}^{s_1-1}\sum_{r_2=l_n+2}^{s_2-1} w_{i_1s_1r_1}^{l_1l_2} w_{i_2s_2r_2}^{m_1m_2}, \\
&= D_{21nt} + D_{22nt}.
\end{aligned}
$$

On taking the expected value of D_{21nt} and using Lemma 3, we show that $E(D_{21nt}) = O(p_nt^2)$. 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

$$
\begin{aligned}
E(D_{3nt}^2) &\le \{\sum_{i=1}^{l_n} k_{ni}^4[E(\mathrm{v}_{t-i}'(\sum_{s=l_n+2}^{t-l_n-1}\mathrm{v}_{s-i}\mathrm{v}_s'\mathrm{v}_s\mathrm{v}_{s-i}' - dI_d)\mathrm{v}_{t-i})^2]^{1/2}\}^2, \\
&\le \mu_4(||\mathrm{v}||)\{\sum_{i=1}^{l_n} k_{ni}^4[E(||\sum_{s=l_n+2}^{t-l_n-1}\mathrm{v}_{s-i}\mathrm{v}_s'\mathrm{v}_s\mathrm{v}_{s-i}' - dI_d||_E^2)]^{1/2}\}^2 = O(tp_n^2)
\end{aligned}
$$

using Lemma 8 that follows.

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

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

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

since

$$\begin{aligned} E(c_{si}^{lm})^2 &= \begin{cases} \mu_4(||\mathrm{v}||)\mu_4(l,l,l,l) - d^2 & \text{if } l = m \\ \mu_4(||\mathrm{v}||)\mu_4(l,l,m,m) & \text{if } l \neq m \end{cases} \\ E(c_{si}^{lm}c_{ri}^{lm}) &= \begin{cases} d[E(||\mathrm{v}_1||^2 v_1(l)^2) - d] & \text{if } l = m,\ r = s-i \\ 0 & \text{elsewhere.} \end{cases} \end{aligned}$$

Finally, we show the result for D_{4nt}. Let us note that

$$E(D_{4nt}^2) = d^2(t - 2l_n - 2)^2 \sum_{i=1}^{l_n} k_{ni}^8 E(\mathrm{v}_{t-i}'\mathrm{v}_{t-i} - d)^2 = O(t^2 p_n),$$

since $E(\mathrm{v}_{t-i}'\mathrm{v}_{t-i} - d)^2 = \mu_4(||\mathrm{v}||) - d^2$.

The following lemma will be useful in the sequel.

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

Proof: Note that in D_{1nt}, $\mathrm{v}_{t-j} \in \{\mathrm{v}_{t-l_n}, \mathrm{v}_{t-l_n+1}, \ldots, \mathrm{v}_{t-1}\}$. Also in D_{1ns}, $\mathrm{v}_{s-j} \in \{\mathrm{v}_{s-l_n}, \mathrm{v}_{s-l_n+1}, \ldots, \mathrm{v}_{s-1}\}$. If $t - l_n > s$, we have the desired result.

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 (1996) for deriving his formulas (A7)-(A10). Using Brown's theorem, it follows that the limiting distribution is normal and the proof of the first part is completed.

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

$$\mathrm{y}_t = \mathrm{c} + \Lambda_1\mathrm{y}_{t-1} + \mathrm{V}_0\mathrm{x}_t + \mathrm{u}_t. \tag{34}$$

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

First, we decompose

$$\sum_{j=1}^{n-1} k_{nj}^2(\mathrm{tr}[\mathrm{C}_{\hat{v}}(j)\mathrm{C}_{\hat{v}}'(j)] - \mathrm{tr}[\mathrm{C}_v(j)\mathrm{C}_v'(j)]). \tag{35}$$

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

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

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

Let

$$\begin{aligned} \hat{\boldsymbol{\lambda}}_{nt} &= (\hat{\mathrm{c}} - \mathrm{c}) + (\hat{\Lambda}_1 - \Lambda_1)\mathrm{y}_{t-1} + (\hat{\mathrm{V}}_0 - \mathrm{V}_0)\mathrm{x}_t, & (36) \\ \hat{\boldsymbol{\gamma}}_{nt} &= \Sigma_u^{-1/2}\hat{\boldsymbol{\lambda}}_{nt}. & (37) \end{aligned}$$

Let also $\hat{\mathrm{u}}_t = \mathrm{u}_t - \hat{\boldsymbol{\lambda}}_{nt}$, and $\hat{\mathrm{v}}_t = \mathrm{v}_t - \hat{\gamma}_{nt}$. First we prove result 5. We can write

$$\mathrm{C}_{\hat{v}}(j) - \mathrm{C}_v(j) = -n^{-1} \sum_{t=j+1}^{n} \hat{\gamma}_{nt}\mathrm{v}'_{t-j} - n^{-1} \sum_{t=j+1}^{n} \mathrm{v}_t\hat{\gamma}'_{n,t-j} + n^{-1} \sum_{t=j+1}^{n} \hat{\gamma}_{nt}\hat{\gamma}'_{n,t-j}. \tag{38}$$

Using (33), we have that

$$\sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}[(\mathrm{C}_{\hat{v}}(j) - \mathrm{C}_v(j))(\mathrm{C}_{\hat{v}}(j) - \mathrm{C}_v(j))'] \le 4(E_{1n} + E_{2n} + E_{3n}), \tag{39}$$

where

$$\begin{aligned} E_{1n} &= n^{-2} \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}[(\sum_{t=j+1}^{n} \hat{\gamma}_{nt}\mathrm{v}'_{t-j})(\sum_{t=j+1}^{n} \hat{\gamma}_{nt}\mathrm{v}'_{t-j})'], \\ E_{2n} &= n^{-2} \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}[(\sum_{t=j+1}^{n} \mathrm{v}_t\hat{\gamma}'_{n,t-j})(\sum_{t=j+1}^{n} \mathrm{v}_t\hat{\gamma}'_{n,t-j})'], \\ E_{3n} &= n^{-2} \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}[(\sum_{t=j+1}^{n} \hat{\gamma}_{nt}\hat{\gamma}'_{n,t-j})(\sum_{t=j+1}^{n} \hat{\gamma}_{nt}\hat{\gamma}'_{n,t-j})']. \end{aligned}$$

Then we show the following result:

Result 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 (30) and (33):

$$\begin{aligned} E_{1n} \le\; & 4 \operatorname{tr}[(\hat{\mathrm{c}} - \mathrm{c})'\Sigma_u^{-1}(\hat{\mathrm{c}} - \mathrm{c})]F_{1n} + 4 \operatorname{tr}[(\hat{\Lambda}_1 - \Lambda_1)'\Sigma_u^{-1}(\hat{\Lambda}_1 - \Lambda_1)]F_{2n} + \\ & 4 \operatorname{tr}[(\hat{\mathrm{c}} - \mathrm{c})'\Sigma_u^{-1}(\hat{\mathrm{c}} - \mathrm{c})]F_{3n}, \end{aligned}$$

where

$$\begin{aligned} F_{1n} &= \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}[(n^{-1} \sum_{t=j+1}^{n} \mathrm{v}'_{t-j})(n^{-1} \sum_{t=j+1}^{n} \mathrm{v}'_{t-j})'], \\ F_{2n} &= \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}[(n^{-1} \sum_{t=j+1}^{n} \mathrm{y}_{t-1}\mathrm{v}'_{t-j})(n^{-1} \sum_{t=j+1}^{n} \mathrm{y}_{t-1}\mathrm{v}'_{t-j})'], \\ F_{3n} &= \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}[(n^{-1} \sum_{t=j+1}^{n} \mathrm{x}_t\mathrm{v}'_{t-j})(n^{-1} \sum_{t=j+1}^{n} \mathrm{x}_t\mathrm{v}'_{t-j})'], \end{aligned}$$

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

Lemma 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 (34) as $\mathrm{y}_t = \mathrm{c}_0 + \Psi(B)\mathrm{V}_0\mathrm{x}_t + \Psi(B)\mathrm{u}_t$, where $\mathrm{c}_0 = (\mathrm{I}_d - \Lambda_1)^{-1}\mathrm{c}$, $\Psi(B) = (\mathrm{I}_d - \Lambda_1 B)^{-1} = \sum_{j\ge 0} \Lambda_1^j B^j$, with $||\Lambda_1|| < 1$. We have that

$$\begin{aligned} n^{-1} \sum_{t=j+1}^{n} \mathrm{y}_{t-1}\mathrm{v}'_{t-j} \;=\; & n^{-1} \sum_{t=j+1}^{n} \mathrm{c}_0\mathrm{v}'_{t-j} + n^{-1} \sum_{t=j+1}^{n} (\Psi(B)\mathrm{V}_0\mathrm{x}_{t-1})\mathrm{v}'_{t-j} + \\ & n^{-1} \sum_{t=j+1}^{n} (\Psi(B)\mathrm{u}_{t-1})\mathrm{v}'_{t-j}, \end{aligned}$$

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

$$\begin{aligned} G_{1n} &= \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}[(n^{-1}\sum_{t=j+1}^{n} \mathrm{c}_0\mathrm{v}_{t-j}')(n^{-1}\sum_{t=j+1}^{n} \mathrm{c}_0\mathrm{v}_{t-j}')'], \\ G_{2n} &= \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}[(n^{-1}\sum_{t=j+1}^{n} (\Psi(B)\mathrm{V}_0\mathrm{x}_{t-1})\mathrm{v}_{t-j}')(n^{-1}\sum_{t=j+1}^{n} (\Psi(B)\mathrm{V}_0\mathrm{x}_{t-1})\mathrm{v}_{t-j}')'], \\ G_{3n} &= \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}[(n^{-1}\sum_{t=j+1}^{n} (\Psi(B)\mathrm{u}_{t-1})\mathrm{v}_{t-j}')(n^{-1}\sum_{t=j+1}^{n} (\Psi(B)\mathrm{u}_{t-1})\mathrm{v}_{t-j}')']. \end{aligned}$$

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

$$\begin{aligned} G_{2n} &= n^{-2}\sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}\{\sum_{t=j+1}^{n} ||\mathrm{v}_{s-j}||^2((\Psi(B)\mathrm{V}_0\mathrm{x}_{t-1})\mathrm{v}_{t-j}')((\Psi(B)\mathrm{V}_0\mathrm{x}_{t-1})\mathrm{v}_{t-j}')' \\ &\quad +2\sum_{s=j+1}^{n}\sum_{j+2}^{t-1}\mathrm{v}_{t-j}'\mathrm{v}_{s-j}((\Psi(B)\mathrm{V}_0\mathrm{x}_{t-1})\mathrm{v}_{t-j}')((\Psi(B)\mathrm{V}_0\mathrm{x}_{s-1})\mathrm{v}_{s-j}')'\}. \end{aligned}$$

It follows that

$$\begin{aligned} E(|G_{2n}|) &= dn^{-2}\sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}\{\sum_{t=j+1}^{n} E[((\Psi(B)\mathrm{V}_0\mathrm{x}_{t-1})\mathrm{v}_{t-j}')((\Psi(B)\mathrm{V}_0\mathrm{x}_{t-1})\mathrm{v}_{t-j}')']\}, \\ &= O(p_n/n), \end{aligned}$$

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 (1996).

Lemma 11

$$E\{\operatorname{tr}[(\sum_{t=j+1}^{n}(\Psi(B)\mathrm{u}_{t-1})\mathrm{v}_{t-j}')(\sum_{t=j+1}^{n}(\Psi(B)\mathrm{u}_{t-1})\mathrm{v}_{t-j}')']\} \leq \Delta_1 n + \Delta_2 n^2 |||\Lambda_1||_E^{2(j-1)}.$$

Proof: We note that

$$\begin{aligned} &(\sum_{t=j+1}^{n}(\Psi(B)\mathrm{u}_{t-1})\mathrm{v}_{t-j}')(\sum_{t=j+1}^{n}(\Psi(B)\mathrm{u}_{t-1})\mathrm{v}_{t-j}')' = \\ &\quad \sum_{s=j+1}^{n}||\mathrm{v}_{s-j}||^2||\Psi(B)\mathrm{u}_{s-1}||^2 + 2\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1}\langle \mathrm{v}_{t-j},\mathrm{v}_{s-j}\rangle\langle\Psi(B)\mathrm{u}_{t-1},\Psi(B)\mathrm{u}_{s-1}\rangle. \end{aligned}$$

Let us consider the first term. We note that

$$||\Psi(B)\mathrm{u}_{t-1}||^2 = \sum_{j_1\geq 0}\mathrm{u}_{t-j_1-1}'(\Lambda_1^{j_1})'(\Lambda_1^{j_1})\mathrm{u}_{t-j_1-1} + \sum_{j_1\neq j_2}\mathrm{u}_{t-j_1-1}'(\Lambda_1^{j_1})'(\Lambda_1^{j_2})\mathrm{u}_{t-j_2-1}$$

and thus

$$\begin{aligned} E(||\mathrm{v}_{t-j}||^2||\Psi(B)\mathrm{u}_{t-1}||^2) &= \sum_{j_1\geq 0}E(||\mathrm{v}_{t-j}||^2\mathrm{u}_{t-j_1-1}'(\Lambda_1^{j_1})'(\Lambda_1^{j_1})\mathrm{u}_{t-j_1-1}) \\ &\leq \sum_{j_1\geq 0}E(||\mathrm{v}_{t-j}||^2||\mathrm{u}_{t-j_1-1}||^2)||(\Lambda_1^{j_1})'(\Lambda_1^{j_1})||_E \\ &\leq \mu_4(||\mathrm{v}||)||\Sigma_u||\sum_{j\geq 0}||\Lambda_1||_E^{2j} \leq \Delta_1. \end{aligned}$$

For the second term, we get

$$\langle \Psi(B)\mathrm{u}_{t-1}, \Psi(B)\mathrm{u}_{s-1}\rangle = \sum_{j_1\geq 0} \mathrm{u}'_{t-j_1-1}(\Lambda_1^{j_1})'(\Lambda_1^{j_1})\mathrm{u}_{s-j_1-1} + \sum_{j_1\neq j_2} \mathrm{u}'_{t-j_1-1}(\Lambda_1^{j_1})'(\Lambda_1^{j_2})\mathrm{u}_{s-j_2-1}.$$

Finally, we have that

$$\begin{aligned} &E(\mathrm{v}'_{t-j}\mathrm{v}_{s-j}\langle \Psi(B)\mathrm{u}_{t-1}, \Psi(B)\mathrm{u}_{s-1}\rangle) = \\ &\qquad \sum_{j_1\geq 0} E[\mathrm{v}'_{t-j}\mathrm{v}_{s-j}\mathrm{u}'_{t-j_1-1}(\Lambda_1^{j_1})'(\Lambda_1^{j_1})\mathrm{u}_{s-j_1-1}] + \\ &\qquad \sum_{j_1\neq j_2} E[\mathrm{v}'_{t-j}\mathrm{v}_{s-j}\mathrm{u}'_{t-j_1-1}(\Lambda_1^{j_1})'(\Lambda_1^{j_2})\mathrm{u}_{s-j_2-1}] \\ &\qquad \leq 2||\Lambda_1||_E^{2(j-1)}\,\mathrm{tr}(\Sigma_u). \end{aligned}$$

Regrouping the results, we obtain

$$\begin{aligned} &E\{\mathrm{tr}[(n^{-1}\sum_{t=j+1}^{n}(\Psi(B)\mathrm{u}_{t-1})\mathrm{v}'_{t-j})(n^{-1}\sum_{t=j+1}^{n}(\Psi(B)\mathrm{u}_{t-1})\mathrm{v}'_{t-j})']\} \leq \\ &\qquad n^{-2}\sum_{t=j+1}^{n}|E(||\Psi(B)\mathrm{u}_{t-1}||^2||\mathrm{v}_{t-j}||^2)| + \\ &\qquad 2n^{-2}\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1}|E(\langle \Psi(B)\mathrm{u}_{s-1}, \Psi(B)\mathrm{u}_{t-1}\rangle\langle \mathrm{v}_{t-j}, \mathrm{v}_{s-j}\rangle| \\ &\qquad \leq n^{-1}\Delta_1 + \Delta_2||\Lambda_1||_E^{2(j-1)}. \end{aligned}$$

and the proof of Lemma 11 is completed.

Remark 1 *It follows from Lemma 11 that*

$$E\{\mathrm{tr}[(\sum_{t=j+1}^{n}\mathrm{y}_{t-1}\mathrm{v}'_{t-j})(\sum_{t=j+1}^{n}\mathrm{y}_{t-1}\mathrm{v}'_{t-j})']\} \leq \Delta_1 n + \Delta_2 n^2|||\Lambda_1||_E^{2(j-1)}. \tag{40}$$

With Lemma 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 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 $|E_{3n}| \leq \sum_{j=1}^{n-1}k_{nj}^2(n^{-1}\sum_{t=1}^{n}||\hat{\gamma}_{nt}||^2)^2$. The following lemma is also needed.

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

Proof: Noting that $||\hat{\gamma}_{nt}||^2 = \mathrm{tr}[\hat{\gamma}_{nt}\hat{\gamma}'_{nt}]$ and using equations (36) and (37), we have that

$$\begin{aligned} \mathrm{tr}(\hat{\gamma}_{nt}\hat{\gamma}'_{nt}) \leq\ & 4(\hat{\mathrm{c}}-\mathrm{c})'\Sigma_u^{-1}(\hat{\mathrm{c}}-\mathrm{c}) + 4\,\mathrm{tr}[(\hat{\Lambda}_1-\Lambda_1)'\Sigma_u^{-1}(\hat{\Lambda}_1-\Lambda_1)]||\mathrm{y}_{t-1}||^2 + \\ & 4\,\mathrm{tr}[(\hat{\mathrm{V}}_0-\mathrm{V}_0)'\Sigma_u^{-1}(\hat{\mathrm{V}}_0-\mathrm{V}_0)]||\mathrm{x}_t||^2. \end{aligned}$$

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

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

$$\begin{aligned} E_{4n} &= \sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}[(n^{-1} \sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}')(n^{-1} \sum_{t=j+1}^{n} \hat{\gamma}_{nt} \mathrm{v}_{t-j}')'], \\ E_{5n} &= \sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}[(n^{-1} \sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}')(n^{-1} \sum_{t=j+1}^{n} \mathrm{v}_t \hat{\gamma}_{n,t-j}')'], \\ E_{6n} &= \sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}[(n^{-1} \sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}')(n^{-1} \sum_{t=j+1}^{n} \hat{\gamma}_{nt} \hat{\gamma}_{n,t-j}')']. \end{aligned}$$

We complete the proof by showing that

Result 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

$$\begin{aligned} F_{4n} &= \sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}[(n^{-1} \sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}')(\Sigma_u^{-1/2}(\hat{\mathrm{c}} - \mathrm{c}) n^{-1} \sum_{t=j+1}^{n} \mathrm{v}_{t-j}')'], \\ F_{5n} &= \sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}[(n^{-1} \sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}')(\Sigma_u^{-1/2}(\hat{\Lambda}_1 - \Lambda_1) n^{-1} \sum_{t=j+1}^{n} \mathrm{y}_{t-1} \mathrm{v}_{t-j}')'], \\ F_{6n} &= \sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}[(n^{-1} \sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}')(\Sigma_u^{-1/2}(\hat{\mathrm{V}}_0 - \mathrm{V}_0) n^{-1} \sum_{t=j+1}^{n} \mathrm{x}_t \mathrm{v}_{t-j}')']. \end{aligned}$$

Note that

$$\begin{aligned} |F_{4n}| \;\le\; & n^{-2}\{\mathrm{tr}[(\hat{c} - c)' \Sigma_u^{-1} (\hat{c} - c)]\}^{1/2} \times \\ & \sum_{j=1}^{n-1} k_{nj}^2 \{\mathrm{tr}[(\sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}')(\sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}')']\}^{1/2} \{\mathrm{tr}[(\sum_{t=j+1}^{n} \mathrm{v}_{t-j}')(\sum_{t=j+1}^{n} \mathrm{v}_{t-j}')'\}^{1/2}. \end{aligned}$$

Since $E[(|X|\,|Y|)^{1/2}] \le \{E|X|\,E|Y|\}^{1/2}$ by Cauchy-Schwarz inequality, that

$$E(\mathrm{tr}[(\sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}')(\sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}')']) = O(n)$$

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

$$\begin{aligned} |F_{5n}| \;\le\; & \{\mathrm{tr}[(\hat{\Lambda}_1 - \Lambda_1)' \Sigma_u^{-1} (\hat{\Lambda}_1 - \Lambda_1)]\}^{1/2} n^{-2} \sum_{j=1}^{n-1} k_{nj}^2 \{\mathrm{tr}[(\sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}') \times \\ & (\sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}')']\}^{1/2} \{\mathrm{tr}[(\sum_{t=j+1}^{n} \mathrm{y}_{t-1} \mathrm{v}_{t-j}')(\sum_{t=j+1}^{n} \mathrm{y}_{t-1} \mathrm{v}_{t-j}')'\}^{1/2}. \end{aligned}$$

Since

$$\begin{aligned} & E[\sum_{j=1}^{n-1} k_{nj}^2 \{\mathrm{tr}[(\sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}')(\sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}')']\}^{1/2} \{\mathrm{tr}[(\sum_{t=j+1}^{n} \mathrm{y}_{t-1} \mathrm{v}_{t-j}') \times \\ & (\sum_{t=j+1}^{n} \mathrm{y}_{t-1} \mathrm{v}_{t-j}')'\}^{1/2}] \le \Delta_1^{1/2} n \sum_{j=1}^{n-1} k_{nj}^2 + n^{3/2} \Delta_2^{1/2} \sum_{j=1}^{n-1} k_{nj}^2 ||\Lambda_1||_E^{j-1}, \end{aligned}$$

and using (40), we have that $F_{5n} = O_p(p_n/n^{3/2} + 1/n)$. Similarly, note that

$$\begin{aligned} |F_{6n}| &\leq \{\mathrm{tr}[(\hat{V}_0 - V_0)'\Sigma_u^{-1}(\hat{V}_0 - V_0)]\}^{1/2} n^{-2} \sum_{j=1}^{n-1} k_{nj}^2 \{\mathrm{tr}[(\sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}') \times \\ &\qquad (\sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}')']\}^{1/2} \{\mathrm{tr}[(\sum_{t=j+1}^{n} \mathrm{x}_t \mathrm{v}_{t-j}')(\sum_{t=j+1}^{n} \mathrm{x}_t \mathrm{v}_{t-j}')'\}^{1/2}. \end{aligned}$$

Since $E(\mathrm{tr}[(\sum_{t=j+1}^{n} \mathrm{x}_t \mathrm{v}_{t-j}')(\sum_{t=j+1}^{n} \mathrm{x}_t \mathrm{v}_{t-j}')']) = O(n)$, by the strict exogeneity of 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

$$\begin{aligned} |E_{5n}| &\leq n^{-2} \sum_{t=j+1}^{n} k_{nj}^2 \{\mathrm{tr}[(\sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}')(\sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}')']\}^{1/2} \times \\ &\qquad \{\mathrm{tr}[(\sum_{t=j+1}^{n} \mathrm{v}_t \hat{\gamma}_{n,t-j}')(\sum_{t=j+1}^{n} \mathrm{v}_t \hat{\gamma}_{n,t-j}')']\}^{1/2}, \\ &\leq n^{-2} \{\sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}[(\sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}')(\sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}')']\}^{1/2} \times \\ &\qquad \{\sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}[(\sum_{t=j+1}^{n} \mathrm{v}_t \hat{\gamma}_{n,t-j}')(\sum_{t=j+1}^{n} \mathrm{v}_t \hat{\gamma}_{n,t-j}')']\}^{1/2}. \end{aligned}$$

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

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

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

$$\begin{aligned} |E_{6n}| &\leq n^{-2} \sum_{t=j+1}^{n} k_{nj}^2 \{\mathrm{tr}[(\sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}')(\sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}')']\}^{1/2} \times \\ &\qquad \{\mathrm{tr}[(\sum_{t=j+1}^{n} \hat{\gamma}_{nt} \hat{\gamma}_{n,t-j}')(\sum_{t=j+1}^{n} \hat{\gamma}_{nt} \hat{\gamma}_{n,t-j}')']\}^{1/2}, \\ &\leq n^{-2} \{\sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}[(\sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}')(\sum_{t=j+1}^{n} \mathrm{v}_t \mathrm{v}_{t-j}')']\}^{1/2} \times \\ &\qquad \{\sum_{j=1}^{n-1} k_{nj}^2 \,\mathrm{tr}[(\sum_{t=j+1}^{n} \hat{\gamma}_{nt} \hat{\gamma}_{n,t-j}')(\sum_{t=j+1}^{n} \hat{\gamma}_{nt} \hat{\gamma}_{n,t-j}')']\}^{1/2}. \end{aligned}$$

We can then conclude that $E_{6n} = o_p(p_n^{1/2}/n)$. By adding and substracting $\mathrm{tr}[\Sigma_u^{-1} \mathrm{C}_u'(j) \Sigma_u^{-1} \mathrm{C}_u(j)]$ in the left hand side of (18) and by using result 1, the proof of part 2 will be completed if we can show that

$$\sum_{j=1}^{n-1} k_{nj}^2 (\mathrm{tr}[\mathrm{C}_{\hat{u}}^{-1}(0) \mathrm{C}_{\hat{u}}'(j) \mathrm{C}_{\hat{u}}^{-1}(0) \mathrm{C}_{\hat{u}}(j)] - \mathrm{tr}[\Sigma_u^{-1} \mathrm{C}_u'(j) \Sigma_u^{-1} \mathrm{C}_u(j)]) = o_p(\sqrt{p_n}/n). \quad (41)$$

From (35), and results 5 and 6, it is sufficient to show that

$$\sum_{j=1}^{n-1} k_{nj}^2 (\mathrm{tr}[\mathrm{C}_{\hat{u}}^{-1}(0) \mathrm{C}_{\hat{u}}'(j) \mathrm{C}_{\hat{u}}^{-1}(0) \mathrm{C}_{\hat{u}}(j)] - \mathrm{tr}[\Sigma_u^{-1} \mathrm{C}_{\hat{u}}'(j) \Sigma_u^{-1} \mathrm{C}_{\hat{u}}(j)]) = o_p(\sqrt{p_n}/n), \quad (42)$$

We already know that $C_{\hat{u}}(0) - \Sigma_u = O_p(n^{-1/2})$, which implies that

$$C_{\hat{u}}^{-1}(0) - \Sigma_u^{-1} = O_p(n^{-1/2}), \quad (43)$$

and (42) follows using inequality (28).

Proof of Theorem 2

First, remark that $\hat{f}_n - f_{0n} = (\hat{f}_n - f_0) - a_n g$ and by direct calculation, we have that

$$\begin{aligned} Q^2(\hat{f}_n; f_{0n}) &= Q^2(\hat{f}_n; f_0) + 2\pi \frac{p_n^{1/2}}{n} \int_{-\pi}^{\pi} \mathrm{tr}[\Gamma_u^{-1}(0) g(\omega) \Gamma_u^{-1}(0) g(\omega)] d\omega - \\ & \quad 4\pi \frac{p_n^{1/4}}{n^{1/2}} \int_{-\pi}^{\pi} \mathrm{tr}[\Gamma_u^{-1}(0) g(\omega) \Gamma_u^{-1}(0)(\hat{f}_n - f_0)] d\omega. \end{aligned}$$

Let $\tilde{f}_n(\omega) = \frac{1}{2\pi} \sum_{j=-n+1}^{n-1} k_{nj} C_u(j) e^{-i\omega j}$. Writing $\hat{f}_n - f_0 = (\hat{f}_n - \tilde{f}_n) + (\tilde{f}_n - f_0)$, it follows that

$$\begin{aligned} &\int_{-\pi}^{\pi} \mathrm{tr}[\Gamma_u^{-1}(0) g(\omega) \Gamma_u^{-1}(0)(\hat{f}_n - f_0)] d\omega = \\ &\quad \int_{-\pi}^{\pi} \{\mathrm{tr}[\Gamma_u^{-1}(0) g(\omega) \Gamma_u^{-1}(0)(\hat{f}_n - \tilde{f}_n)] + \mathrm{tr}[\Gamma_u^{-1}(0) g(\omega) \Gamma_u^{-1}(0)(\tilde{f}_n - f_0)]\} d\omega. \end{aligned}$$

Lemma 13

$$\int_{-\pi}^{\pi} \mathrm{tr}[\Gamma_u^{-1}(0) g(\omega) \Gamma_u^{-1}(0)(\hat{f}_n - \tilde{f}_n)] d\omega = O_p(n^{-1/2}).$$

Proof: Since

$$\hat{f}_n - \tilde{f}_n = \frac{1}{2\pi} \sum_{j=-n+1}^{n-1} k_{nj} [C_{\hat{u}}(j) - C_u(j)] e^{-i\omega j}, \quad (44)$$

we can write

$$\begin{aligned} &\int_{-\pi}^{\pi} \mathrm{tr}[\Gamma_u^{-1}(0) g(\omega) \Gamma_u^{-1}(0)(\hat{f}_n - \tilde{f}_n)] d\omega = \\ &\qquad \frac{1}{2\pi} \sum_{j=-n+1}^{n-1} k_{nj}\, \mathrm{tr}[\Gamma_u^{-1}(0) g_j \Gamma_u^{-1}(0)(C_{\hat{u}}(j) - C_u(j))], \end{aligned}$$

where $g_j = \int_{-\pi}^{\pi} g(\omega) e^{i\omega j} d\omega$. Also, we have

$$\begin{aligned} &| \sum_{j=-n+1}^{n-1} k_{nj}\, \mathrm{tr}[\Gamma_u^{-1}(0) g_j \Gamma_u^{-1}(0)(C_{\hat{u}}(j) - C_u(j))]| \leq \\ &\quad \sum_{j=-n+1}^{n-1} k_{nj}\, \mathrm{tr}[\Gamma_u^{-1}(0) g_j \Gamma_u^{-1}(0) g_j']^{1/2}\, \mathrm{tr}[(C_{\hat{v}}(j) - C_v(j))(C_{\hat{v}}(j) - C_v(j))']^{1/2} \leq \\ &\quad \{\sum_{j=-n+1}^{n-1} \mathrm{tr}[\Gamma_u^{-1}(0) g_j \Gamma_u^{-1}(0) g_j']\}^{1/2} \{\sum_{j=-n+1}^{n-1} k_{nj}^2\, \mathrm{tr}[(C_{\hat{v}}(j) - C_v(j))(C_{\hat{v}}(j) - C_v(j))']\}^{1/2}, \end{aligned}$$

and Lemma 13 follows since $\sum_{j=-\infty}^{\infty} \mathrm{tr}[\Gamma_u^{-1}(0) g_j \Gamma_u^{-1}(0) g_j'] < \infty$ and using result 5 of the second part of the proof of Theorem 1.

Lemma 14

$$\int_{-\pi}^{\pi} \mathrm{tr}[\Gamma_u^{-1}(0) g(\omega) \Gamma_u^{-1}(0)(\tilde{f}_n - f_0)] d\omega = O_p(n^{-1/2}).$$

Proof: Since $\tilde{\mathrm{f}}_n - \mathrm{f}_0 = \frac{1}{2\pi}\sum_{j=-n+1}^{n-1} k_{nj}[\mathrm{C}_u(j) - \Gamma_u(j)]e^{-i\omega j}$, we have that

$$\int_{-\pi}^{\pi} \mathrm{tr}[\Gamma_u^{-1}(0)\mathrm{g}(\omega)\Gamma_u^{-1}(0)(\tilde{\mathrm{f}}_n - \mathrm{f}_0)]d\omega = \frac{1}{2\pi}\sum_{j=-n+1}^{n-1} k_{nj}\,\mathrm{tr}[\Gamma_u^{-1}(0)\mathrm{g}_j\Gamma_u^{-1}(0)(\mathrm{C}_u(j) - \Gamma_u(j))'],$$

where g_j is as before. If $\mathrm{g}_{jS} = \Gamma_u^{-1}(0)\mathrm{g}_j\Gamma_u^{-1}(0)$, then we have that $\mathrm{tr}[\mathrm{g}_{jS}(\mathrm{C}_u(j)-\Gamma_u(j))'] = \sum_{l,m=1}^{d} g_{jS,lm}[C_{u,ml}(j)-\Gamma_{u,ml}(j)]$, where $g_{jS,lm}$ and $C_{u,lm}(j) - \Gamma_{u,lm}(j)$ are the (l,m)-components of g_{jS} and $\mathrm{C}_u(j) - \Gamma_u(j)$, respectively. We note that $\sum g_{jS,lm}^2$ is a convergent series. Therefore, since $E[(\sum_{j=1}^{n-1} g_{jS,lm}C_{u,lm}(j))^2] = O(n^{-1})$, the announced result is proved.

Proof of Theorem 3

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

$$\begin{aligned} Q^2(\hat{\mathrm{f}}_n;\mathrm{f}_0) &= Q^2(\mathrm{f};\mathrm{f}_0) + 4\pi\int_{-\pi}^{\pi} \mathrm{tr}[\Gamma_u^{-1}(0)(\mathrm{f} - \mathrm{f}_0)^*\Gamma_u^{-1}(0)(\hat{\mathrm{f}}_n - \mathrm{f})]d\omega + \\ &\quad 2\pi\int_{-\pi}^{\pi} \mathrm{tr}[\Gamma_u^{-1}(0)(\hat{\mathrm{f}}_n - \mathrm{f})^*\Gamma_u^{-1}(0)(\hat{\mathrm{f}}_n - \mathrm{f})]d\omega. \end{aligned}$$

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

Result 9

$$\int_{-\pi}^{\pi} \mathrm{tr}[\Gamma_u^{-1}(0)(\hat{\mathrm{f}}_n - \mathrm{f})^*\Gamma_u^{-1}(0)(\hat{\mathrm{f}}_n - \mathrm{f})]d\omega = o_p(1).$$

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

$$\int_{-\pi}^{\pi} \mathrm{tr}[\Gamma_u^{-1}(0)(\hat{\mathrm{f}}_n - \mathrm{f})^*\Gamma_u^{-1}(0)(\hat{\mathrm{f}}_n - \mathrm{f})]d\omega \le 2\int_{-\pi}^{\pi}\{\mathrm{tr}[\Gamma_u^{-1}(0)(\hat{\mathrm{f}}_n - \tilde{\mathrm{f}}_n)^*\Gamma_u^{-1}(0)(\hat{\mathrm{f}}_n - \tilde{\mathrm{f}}_n)] + \mathrm{tr}[\Gamma_u^{-1}(0)(\tilde{\mathrm{f}}_n - \mathrm{f})^*\Gamma_u^{-1}(0)(\tilde{\mathrm{f}}_n - \mathrm{f})]\}d\omega.$$

Now, result 9 follows from the next two lemmas.

Lemma 15

$$\int_{-\pi}^{\pi} \mathrm{tr}[\Gamma_u^{-1}(0)(\hat{\mathrm{f}}_n - \tilde{\mathrm{f}}_n)^*\Gamma_u^{-1}(0)(\hat{\mathrm{f}}_n - \tilde{\mathrm{f}}_n)]d\omega = o_p(1).$$

Proof: Using (44), we have that

$$\int_{-\pi}^{\pi} \mathrm{tr}[\Gamma_u^{-1}(0)(\hat{\mathrm{f}}_n - \tilde{\mathrm{f}}_n)^*\Gamma_u^{-1}(0)(\hat{\mathrm{f}}_n - \tilde{\mathrm{f}}_n)]d\omega = \frac{1}{2\pi}\sum_{j=-n+1}^{n-1} k_{nj}^2\,\mathrm{tr}[(\mathrm{C}_{\hat{v}}(j) - \mathrm{C}_v(j))(\mathrm{C}_{\hat{v}}(j) - \mathrm{C}_v(j))'].$$

Inequality (39) provides an upper bound for $\sum_{j=1}^{n-1} k_{nj}^2\,\mathrm{tr}[(\mathrm{C}_{\hat{v}}(j) - \mathrm{C}_v(j))(\mathrm{C}_{\hat{v}}(j) - \mathrm{C}_v(j))']$. 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 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 dependence structure given in Assumption C. However, by Cauchy-Schwarz inequality, we obtain that

$$|E_{1n}| = E_{1n} \le (\sum_{j=1}^{n} k_{nj}^2)(n^{-1}\sum_{j=1}^{n} ||\mathrm{v}_t||^2)(n^{-1}\sum_{j=1}^{n} ||\hat{\gamma}_{nt}||^2).$$

But we have $n^{-1}\sum_{j=1}^{n}||\hat{\gamma}_{nt}||^2 = O_p(n^{-1})$, since

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

and in the static model (23) the least square 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 complete the proof of Lemma 15.

Lemma 16

$$\int_{-\pi}^{\pi}\mathrm{tr}[\Gamma_u^{-1}(0)(\tilde{\mathrm{f}}_n-\mathrm{f})^*\Gamma_u^{-1}(0)(\tilde{\mathrm{f}}_n-\mathrm{f})]d\omega = o_p(1).$$

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

$$\begin{aligned}&\int_{-\pi}^{\pi}\mathrm{tr}[\Gamma_u^{-1}(0)(\tilde{\mathrm{f}}_n-\mathrm{f})^*\Gamma_u^{-1}(0)(\tilde{\mathrm{f}}_n-\mathrm{f})]d\omega = \\ &\quad\frac{1}{2\pi}\sum_{|j|\leq n-1}\mathrm{tr}[\Gamma_u^{-1}(0)(k_{nj}\mathrm{C}_u(j)-\Gamma_u(j))'\Gamma_u^{-1}(0)(k_{nj}\mathrm{C}_u(j)-\Gamma_u(j))] \\ &\quad+\frac{1}{2\pi}\sum_{|j|\geq n}\mathrm{tr}[\Gamma_u^{-1}(0)\Gamma_u(j)'\Gamma_u^{-1}(0)\Gamma_u(j)].\end{aligned}$$

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

$$\begin{aligned}&\sum_{|j|\leq n-1}\mathrm{tr}[\Gamma_u^{-1}(0)(k_{nj}\mathrm{C}_u(j)-\Gamma_u(j))'\Gamma_u^{-1}(0)(k_{nj}\mathrm{C}_u(j)-\Gamma_u(j))] \leq \\ &\quad 2\sum_{|j|\leq n-1}(k_{nj}-1)^2\,\mathrm{tr}[\Gamma_u^{-1}(0)\Gamma_u(j)'\Gamma_u^{-1}(0)\Gamma_u(j)] + \\ &\quad 2\sum_{|j|\leq n-1}k_{nj}^2\,\mathrm{tr}[\Gamma_u^{-1}(0)(\mathrm{C}_u(j)-\Gamma_u(j))'\Gamma_u^{-1}(0)(\mathrm{C}_u(j)-\Gamma_u(j))].\end{aligned}$$

By an argument similar to the one used by Hong (1996, 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 $\mathrm{tr}[\Gamma_u^{-1}(0)(\mathrm{C}_u(j)-\Gamma_u(j))'\Gamma_u^{-1}(0)(\mathrm{C}_u(j)-\Gamma_u(j)) = \sum_{t=1}^{d}\sum_{s=1}^{d}(C_{v,st}(j)-\Gamma_{v,st}(j))^2$, where $C_{v,st}(j)$ and $\Gamma_{v,st}(j)$ are the (s,t)-components of $\mathrm{C}_v(j)$ and $\Gamma_v(j)$, respectively. From a general result for the variance of cross-covariances, see for example Hannan (1970, pp. 208-211) or Hannan (1976). The variance of $C_{v,st}(j)$ is given by $\mathrm{var}(C_{v,st}(j)) = n^{-1}\sum_{|i|\leq n-1}(1-|i|/n)[\Gamma_{v,st}(i+j)\Gamma_{v,st}(i-j) + \kappa_{stst}(0,j,i,i+j)]$. From Assumption C, we have that $\sup_{j\geq 1}\mathrm{var}[C_{v,st}(j)] = O(n^{-1})$. Therefore

$$\sum_{|j|\leq n-1}k_{nj}^2\sum_{t=1}^{d}\sum_{s=1}^{d}[C_{v,st}(j)-\Gamma_{v,st}(j)]^2 = O_p(p_n/n).$$

and the proof of Lemma 16 is completed. Consequently, result 9 hold and Theorem 3 is proved.

REFERENCES

Ansley, C. F. (1980), 'Computation of the theoretical autocovariance function for a vector *ARMA* process', *Journal of Statistical Computation and Simulation* **12**, 15–24.

Bahadur, R. R. (1960), 'Stochastic comparison of tests', *Annals of Mathematical Statistics* **31**, 276–295.

Beltrao, K. and Bloomfield, P. (1987), 'Determining the bandwidth of a kernel spectrum estimate', *Journal of Time Series Analysis* **8**, 21–38.

Box,G. E. P. and Pierce, D. A. (1970), ‘Distribution of residual autocorrelations in autoregressive-integrated moving average time series models’, *Journal of the American Statistical Association* **65**, 1509–1526.

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

Brown, B. M. (1971), ‘Martingale central limit theorems’, *Annals of Mathematical Statistics* **42**, 59–66.

Chitturi, R. V. (1974), ‘Distribution of residual autocorrelations in multiple autoregressive schemes’, *Journal of the American Statistical Association* **69**, 928–934.

Chitturi, R. V. (1976), ‘Distribution of multivariate white noise autocorrelations’, *Journal of the American Statistical Association* **71**, 223–226.

Fuller, W. A. (1996), *Introduction to Statistical Time Series*, second edn, Wiley:New-York.

Gilbert, P. D. (1993), ‘State space and *ARMA* models: An overview of the equivalence’. Working Paper 93-4, Bank of Canada.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press:Princeton.

Hannan, E. (1970), *Multiple time series*, Wiley:New-York.

Hannan, E. J. (1976), ‘The asymptotic distribution of serial covariances’, *Annals of Statistics* **4**, 396–399.

Hannan, E. J. and Deistler, M. (1988), ‘The statistical theory of linear systems’, Wiley:New-York.

Harvey, A. C. (1990), *The econometric analysis of time series*, 2nd ed, The MIT Press: Cambridge, Massachusetts.

Harville, D. A. (1997), *Matrix Algebra from a Statistician’s perspective*, Springer-Verlag:Berlin.

Hendry, D. F. (1995), *Dynamic Econometrics*, Oxford University Press:Oxford.

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

Hosking, J. (1980), ‘The multivariate portmanteau statistic’, *Journal of the American Statistical Association* **75**, 602–608.

Hosking, J. (1981a), ‘Equivalent forms of the multivariate portmanteau statistic’, *Journal of the Royal Statistical Society, Ser. B* **43**, 261–262, (Corr: V51/1989, p.303).

Hosking, J. (1981b), ‘Lagrange-multiplier tests of multivariate time series models’, *Journal of the Royal Statistical Society, Ser. B* **43**, 219–230.

Judge, G. G., Hill, R. C., Grittiths, W. E., Lütkepohl, H. and Lee, T.-C. (1985), *The theory and practice of econometrics*, second edn, Wiley:New-York.

Judge, G. G., Hill, R. C., Grittiths, W. E., Lütkepohl, H. and Lee, T.-C. (1988), *Introduction to the theory and practice of econometrics*, second edn, Wiley:New-York.

Li, W. K. and McLeod, A. I. (1981), ’Distribution of the residual autocorrelations in multivariate *ARMA* time series models, *Journal of the Royal Statistical Society, Ser. B* **43**, 231–239.

Lütkepohl, H. (1993), *Introduction to Multiple Time Series Analysis*, second edn, Springer-Verlag:Berlin.

Paparoditis, E. (2000a), ’Spectral density based on goodness-of-fit tests for time series analysis’, *Scandinavian Journal of Statistics*, **27**, 143–176.

Paparoditis, E. (2000b), ’On some power properties of goodness-of-fit tests for time series models’, To appear in a *Festschrift* by Chapman and Hall.

Pitman, E. J. G. (1979), *Some Basic Theory for Statistical Inference*, Halsted:New-York.

Priestley, M. B. (1981), *Univariate Series*, Vol. 1 of *Spectral Analysis and Time Series*, Academic Press:New-York.

Robinson, P. M. (1991), ‘Automatic frequency domain inference on semiparametric and non-parametric models’, *Econometrica* **59**, 1329–1363.

Robinson, P. M. (1994), ‘Time series with strong dependence’, in *Advances in Econometrics, Sixth World Congress, Vol.1*, ed. by C. Sims. Cambridge University Press, 47–95.

Shiryaev, A. N. (1995), *Probability*, second edn, Springer-Verlag:Berlin.

Taniguchi, M. and Kakizawa, Y. (2000), *Asymptotic Theory of Statistical Inference for Time Series*, Springer:Berlin.

White, H. (1984), *Asymptotic Theory for Econometricians*, Academic Press:New-York.