2004s-06

A Generalized Portmanteau Test for Independence of Two Infinite Order Vector Autoregressive Series

Chafik Bouhaddioui, Roch Roy

Série Scientifique
Scientific Series

Montréal
Février 2004

CIRANO
Centre interuniversitaire de recherche
en analyse des organisations

CIRANO

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du ministère de la Recherche, de la Science et de la Technologie, de même que des subventions et mandats obtenus par ses équipes de recherche.

CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère de la Recherche, de la Science et de la Technologie, and grants and research mandates obtained by its research teams.

Les organisations-partenaires / The Partner Organizations

A Generalized Portmanteau Test for Independence of Two Infinite Order Vector Autoregressive Series[*]

Chafik Bouhaddioui[†], Roch Roy[‡]

Résumé / *Abstract*

Dans ce travail, nous proposons une généralisation au cas multivarié de l'approche de Hong (1996) afin de tester l'indépendance de deux séries multivariées stationnaires et autorégressives d'ordre infini. Il s'agit d'une approche semiparamétrique où chaque série est d'abord filtrée par une autorégression d'ordre fini et où la statistique de test est une version normalisée d'une somme pondérée de formes quadratiques dans les matrices de corrélations croisées résiduelles résultantes à tous les délais. Les poids sont définis par une fonction de noyau et un point de troncature. En utilisant un résultat de Lewis et Reinsel (1985), la loi asymptotique de la statistique de test est obtenue sous l'hypothèse nulle et la convergence du test est établie pour une contre-hypothèse fixée de corrélation sérielle de forme quelconque. À des facteurs de normalisation près, la statistique portmanteau étudiée dans Bouhaddioui et Roy (2003), qui est basée sur un nombre fixé de délais, peut être vue comme un cas particulier en utilisant le noyau uniforme tronqué. Cependant, plusieurs noyaux produisent une plus grande puissance comme le montrent une analyse asymptotique de la puissance ainsi que des simulations de Monte Carlo en échantillons finis. Un exemple avec des données réelles est aussi présenté.

> **Mots clés** : vecteurs autorégressifs d'ordre infini, indépendance, corrélations croisées résiduelles, noyaux, statistique portmanteau, puissance. asymptotique.

[†] CIRANO and Département de mathématiques et de statistique, Université de Montréal, Montréal (Québec), Canada, email : chafik.bouhaddioui@cirano.qc.ca.
[‡] Département de mathématiques et de statistique and Centre de Recherches mathématiques, Université de Montréal. Adresse for correspondence : Département de mathématiques et de statistique, Université de Montréal, Case postale 6128; succursale Centre-ville, Montréal, Québec, Canada, H3C 3J7, email : royr@dms.umontreal.ca.

In many situations, we want to verify the existence of a relationship between multivariate time series. Here, we propose a semiparametric approach for testing the independence between two infinite order vector autoregressive (VAR(∞)) series which is an extension of Hong's (1996a) univariate results. We first filter each series by a finite-order autoregression and the test statistic is a standardized version of a weighted sum of quadratic form in residual cross-correlation at all possible lags. The weights depend on a kernel function and on a truncation parameter. Using a result of Lewis and Reinsel (1985), the asymptotic distribution of the statistic test is derived under the null hypothesis and its consistency is also established for a fixed alternative of serial cross-correlation of unknown form. Apart from standardization factors, the multivariate portmanteau statistic proposed by Bouhaddioui and Roy (2003) that takes into account a fixed number of lags can be viewed as a special case by using the truncated uniform kernel. However, many kernels lead to a greater power, as shown in an asymptotic power analysis and by a small simulation study in finite samples. A numerical example with real data is also presented.

1   Introduction

In multivariate time series, many recent papers address the problem of checking the hypothesis independence or non-correlation between two vector series. Most of these studies concentrated on tests for independence of two multivariate finite order vector autoregressive (VAR) or vector autoregressive moving average (VARMA) time series. El Himdi and Roy (1997) generalized the procedure developed by Haugh (1976) for univariate time series in order to test the null hypothesis of noncorrelation between two multivariate stationary and invertible VARMA series. They proposed test statistics based on the residual cross-correlation matrices $\mathbf{R}_{\hat{a}}^{(12)}(j)$, $|j| \leq M$ for a given $M < N$, (N being the sample size), between the two residual series $\{\hat{a}_t^{(1)}\}$ and $\{\hat{a}_t^{(2)}\}$ resulting from fitting the *true* VARMA models to each of the original series $\{\mathbf{X}_t^{(1)}\}$ and $\{\mathbf{X}_t^{(2)}\}$. Under the hypothesis of non-correlation between the two series, they showed in particular that an arbitrary vector of residual cross-correlations asymptotically follows a multivariate normal distribution. Hallin and Saidi (2002) used that result to develop a test statistic that takes into account a possible pattern in the signs of cross-correlations at different lags. This latter test is a generalization to the multivariate case of the procedure introduced by Koch and Yang (1986). Also, El Himdi and Roy's procedure was extended to partially nonstationary (cointegrated) VARMA series by Pham, Roy and Cédras (2003).

In the univariate case, another important extension of Haugh's test is the class of spectral test statistics introduced by Hong (1996a). His approach is semiparametric and is valid for two infinite order autoregressive series AR(∞), say $\{X_t^{(1)}\}$ and $\{X_t^{(2)}\}$. Its consists to fitting an autoregressive model of order p to a series of N observations from each infinite order autoregressive process $X^{(h)}$, $h = 1, 2$. The order p_h of the fitted autoregression is a function of the sample size, see Berk (1974). The portmanteau type statistic is based on the sum of the weighted squared cross-correlations $r_{\hat{a}}^{(12)}(j)$ at all possible lags between the residual series and is defined by

$$\mathcal{Q}_N = \frac{N \sum_{j=1-N}^{N-1} k^2(j/M) r_{\hat{a}}^{(12)}(j)^2 - S_N(k)}{\{2D_N(k)\}^{1/2}}, \tag{1.1}$$

where $k(.)$ is an arbitrary kernel function, M is either a truncation or smoothing parameter. The quantities $S_N(k)$ and $D_N(k)$ that depend only on the kernel $k(.)$ are defined by equation (3.4). They essentially correspond to the asymptotic mean and variance of the weighted sum. Using the truncated uniform kernel, $\mathcal{Q}_N$ can be viewed as a normalized version of Haugh's statistic. Under the null hypothesis of independence, the statistic $\mathcal{Q}_N$ is asymptotically $\mathcal{N}(0,1)$. The test is unilateral and rejects for large values of $\mathcal{Q}_N$. Hong's approach presents two important practical advantages. First, it protects us against mis-specifications of the true underlying ARMA models,

as in Haugh approach that may lead to wrong conclusions because they invalidate the asymptotic theory. Second, in contrast to Haugh's and Koch and Yang's tests that are based on the residual cross-correlations at lag j such that $|j| \leq M$, where M is fixed with respect to N, the portmanteau test $\mathcal{Q}_N$ takes into account all lags and is therefore consistent for a large class of alternatives of serial cross-correlations of arbitrary form between the two series. A robust version of Hong's test for univariate ARMA series is described in Duchesne and Roy (2003). Bouhaddioui and Roy (2003) derived the asymptotic distribution of an arbitrary vector of residual cross-correlation under the hypothesis of non-correlation of two series. That result allows them to propose test statistics for the alternative of serial cross-correlation at a particular lag or at a fixed number of lags j, for example $|j| \leq M$.

The main objective of this work is to extend Hong's approach to VAR(∞) models. These models were studied by many authors and are discussed for example in Lütkepohl (1991) and Reinsel (1993). Using the cross-correlation matrices $\mathbf{R}_{\hat{\boldsymbol{a}}}^{(12)}(j)$, $|j| \leq N - 1$, between the two residuals series $\{\hat{\boldsymbol{a}}_t^{(1)}\}$ and $\{\hat{\boldsymbol{a}}_t^{(2)}\}$, obtained by approximating the two multivariate VAR(∞) series by finite order autoregressions, we introduce a multivariate version of the weighted portmanteau statistic $\mathcal{Q}_N$. It can be viewed as a normalized version of the $\mathcal{L}_2$-norm of a kernel-based estimator of the cross-coherency function between the two innovation series. With univariate series, we retrieve Hong's statistic. The test statistic continues to have an asymptotic $\mathcal{N}(0, 1)$ distribution under the hypothesis of independence of the two series and is also consistent for any alternative of serial cross-correlation of arbitrary form.

The organization of the paper is as follows. Section 2 contains preliminary results. 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 $\mathcal{N}(0, 1)$. The consistency of the test is established in Section 4. The Bahadur's asymptotic relative efficiency of one kernel with respect to another is also discussed. 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 analyze the impact of the kernel and the truncation parameter on the level and power. Finally, the new test is applied to a set of American and Canadian economic data. Most of the proofs are relegated to the Appendix.

2 Preliminaries

Let $\mathbf{X} = \{\mathbf{X}_t \, , \, t \in \mathbb{Z}\}$ be a multivariate second-order stationary process of dimension m. Without loss of generality, we can assume that $\mathbb{E}(\mathbf{X}_t) = \mathbf{0}$. The autocovariance matrix at lag j, $j \in \mathbb{Z}$, is given by

$$\boldsymbol{\Gamma}_\mathbf{X}(j) \;\; = \;\; \mathbb{E}(\mathbf{X}_t \mathbf{X}_{t-j})^T = (\gamma_{uv}(j))_{m \times m},$$

with $\boldsymbol{\Gamma}_\mathbf{X}(j) = \boldsymbol{\Gamma}_\mathbf{X}(-j)^T$. We suppose that

$$\sum_{j=-\infty}^{\infty} |\gamma_{uv}(j)| \;\; < \;\; \infty \, , \; u, v = 1, ..., m, \; j \in \mathbb{Z}. \tag{2.1}$$

The autocorrelation matrix at lag j, $j \in \mathbb{Z}$, is denoted by

$$\boldsymbol{\rho}_\mathbf{X}(j) = (\rho_{uv}(j))_{m \times m}, \qquad \rho_{uv}(j) = \gamma_{uv}(j)\{\gamma_{uu}(0)\gamma_{vv}(0)\}^{-1/2},$$

with $\boldsymbol{\rho}_\mathbf{X}(j) = \boldsymbol{\rho}_\mathbf{X}(-j)^T$. If we denote by $\mathbf{D}\{b_i\}$ a diagonal matrix whose elements are $b_1, ..., b_m$, the matrix form of $\boldsymbol{\rho}_\mathbf{X}(j)$ is given by

$$\boldsymbol{\rho}_\mathbf{X}(j) \;\; = \;\; \mathbf{D}\{\gamma_{ii}(0)^{-1/2}\}\boldsymbol{\Gamma}_\mathbf{X}(j)\mathbf{D}\{\gamma_{ii}(0)^{-1/2}\} \, , \; j \in \mathbb{Z}. \tag{2.2}$$

Under condition (2.1), the standardized spectral density function $\boldsymbol{f}_\mathbf{X}(\lambda) = (f_{uv}(\lambda))_{m \times m}$ is defined by

$$\boldsymbol{f}_\mathbf{X}(\lambda) \;\; = \;\; \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \boldsymbol{\rho}_\mathbf{X}(j) e^{-i\lambda j} \, , \; -\pi \leq \lambda \leq \pi,$$

The coherency function $\boldsymbol{S}_\mathbf{X}(\lambda) = (\mathcal{S}_{uv}(\lambda))_{m \times m}$ at the frequency λ is given by

$$\boldsymbol{S}_\mathbf{X}(\lambda) \;\; = \;\; \mathbf{D}\{f_{uu}(\lambda)\}^{-1/2} \boldsymbol{f}_\mathbf{X}(\lambda)\mathbf{D}\{f_{vv}(\lambda)\}^{-1/2} \, , \; -\pi \leq \lambda \leq \pi.$$

The coherence function at the frequency λ is the modulus of the coherency function.

Given a realization $\mathbf{X}_1, ..., \mathbf{X}_N$ of length N of the process $\mathbf{X}$, the $m \times m$ sample autocovariance and autocorrelation matrices at lag j, $0 \leq j \leq N-1$, are denoted by $\boldsymbol{C}_\mathbf{X}(j) = (c_{uv}(j))_{m \times m}$ and $\mathbf{R}_\mathbf{X}(j) = (r_{uv}(j))_{m \times m}$ and are defined by

$$\boldsymbol{C}_\mathbf{X}(j) \;\; = \;\; \frac{1}{N} \sum_{t=j+1}^{N} \mathbf{X}_t \mathbf{X}_{t-j}^T, \qquad \mathbf{R}_\mathbf{X}(j) = \mathbf{D}\{c_{ii}(0)^{-1/2}\}\boldsymbol{C}_\mathbf{X}(j)\mathbf{D}\{c_{ii}(0)^{-1/2}\}.$$

For $-N+1 \leq j \leq 0$, we have $\boldsymbol{C}_\mathbf{X}(j) = \boldsymbol{C}_\mathbf{X}(-j)^T$ and $\mathbf{R}_\mathbf{X}(j) = \mathbf{R}_\mathbf{X}(-j)^T$.

In the sequel, we suppose that the process $\mathbf{X}$ admits a stationary infinite-order autoregressive representation, VAR(∞), i.e., there exists matrices $\boldsymbol{\Phi}_l$, $l \in \mathbb{N}$, such that

$$\mathbf{X}_t - \sum_{l=1}^{\infty} \boldsymbol{\Phi}_l \mathbf{X}_{t-l} \;=\; \boldsymbol{\Phi}(B)\mathbf{X}_t = \boldsymbol{a}_t, \; t \in \mathbb{Z}, \tag{2.3}$$

where $\sum_{l=1}^{\infty} \|\boldsymbol{\Phi}_l\| < \infty$, $\boldsymbol{\Phi}(z) = \boldsymbol{I}_m - \sum_{l=1}^{\infty} \boldsymbol{\Phi}_l z^l$ and $\det\{\boldsymbol{\Phi}(z)\} \neq 0$, $|z| \leq 1$, where $\boldsymbol{I}_m$ is the identity matrix of dimension m, B is the backward shift operator defined by $B^l \mathbf{X}_t = \mathbf{X}_{t-l}$ and $\|.\|$ is the Euclidean matrix norm defined by $\|\mathbf{A}\|^2 = tr(\mathbf{A}^T \mathbf{A})$ for a given matrix $\mathbf{A}$. The process $\boldsymbol{a}$ is a strong white noise that is, a sequence of independent identically distributed random vectors with mean $\mathbf{0}$ and regular covariance matrix $\boldsymbol{\Sigma}$. The stationarity assumption ensures that the process $\mathbf{X}$ also admits a causal linear representation.

Based on a realization $\mathbf{X}_1, ..., \mathbf{X}_N$ of length N, we fit an autoregressive model of order p, VAR(p), whose coefficients are denoted by $\boldsymbol{\Phi}_{1,p}, ..., \boldsymbol{\Phi}_{p,p}$ and we write $\boldsymbol{\Phi}(p) = (\boldsymbol{\Phi}_{1,p}, ..., \boldsymbol{\Phi}_{p,p})$. The corresponding Yule-Walker estimator $\hat{\boldsymbol{\Phi}}(p) = (\hat{\boldsymbol{\Phi}}_{1,p}, ..., \hat{\boldsymbol{\Phi}}_{p,p})$ is given by

$$\hat{\boldsymbol{\Phi}}(p) \;=\; \hat{\mathbf{A}}_{1,p}^T \hat{\mathbf{A}}_p^{-1}, \tag{2.4}$$

where $\hat{\mathbf{A}}_{1,p} = (N-p)^{-1} \sum_{t=p+1}^{N} \mathbf{X}_t(p)\mathbf{X}_t^T$, $\hat{\mathbf{A}}_p = (N-p)^{-1} \sum_{t=p+1}^{N} \mathbf{X}_t(p)\mathbf{X}_t^T(p)$ and $\mathbf{X}_t(p) = (\mathbf{X}_{t-1}^T, \mathbf{X}_{t-2}^T, ..., \mathbf{X}_{t-p}^T)^T$. To obtain a consistent estimator $\hat{\boldsymbol{\Phi}}(p)$, we must let p tends to infinity as N increases but not too fast. The following assumption on the noise process is also needed.

Assumption A *The m-dimensional strong white noise $\boldsymbol{a} = \{\boldsymbol{a}_t = (a_{1t}, .., a_{mt})^T\}$ is such that $\mathbb{E}(\boldsymbol{a}) = \mathbf{0}$, its covariance matrix $\boldsymbol{\Sigma}$ is regular and*

$$\mathbb{E}|a_{i,t} a_{j,t} a_{k,t} a_{l,t}| \;<\; \gamma_4 < \infty \, , \; i,j,k,l \in \{1, ..., m\} \text{ and } t \in \mathbb{Z}.$$

The following proposition that gives the consistency rate of $\hat{\boldsymbol{\Phi}}(p)$ is a multivariate generalization of a univariate result given by Berk (1974). It follows from Eq. (2.8) of Lewis and Reinsel (1985, p. 397), see also Theorem 2.1 in Paparoditis (1996).

Proposition 2.1 *Let $\{\mathbf{X}_t\}$ be a VAR(∞) process given by (2.3) and satisfying Assumption A. Also, suppose that the following two conditions are verified:*
(i) p is chosen as a function of N such that $p \to \infty$ and $p^2/N \to 0$ as $N \to \infty$;
(ii) $\sqrt{p} \sum_{j=p+1}^{\infty} \|\boldsymbol{\Phi}_j\| \to 0$ as $N \to \infty$.
Then, the estimator $\hat{\boldsymbol{\Phi}}(p)$ defined by (2.4) is such that

$$\|\hat{\boldsymbol{\Phi}}(p) - \boldsymbol{\Phi}(p)\| \;=\; O_p\left(\frac{p^{1/2}}{N^{1/2}}\right). \tag{2.5}$$

4

In this result, the condition $p = o(N^{1/2})$ for the rate of increase of p ensures that asymptotically, enough sample information is available for the estimators to have standard limiting distributions. The condition $\sqrt{p} \sum_{j=p+1}^{\infty} \|\mathbf{\Phi}_j\| \to 0$ imposes a lower bound on the growth rate of p, which ensures that the approximation error of the true underlying model by a finite order autoregression gets small when the sample size increases. A more detailed discussion of these conditions is available in Lütkepohl (1991), see also Hong (1996a).

Now, suppose that the process $\mathbf{X}$ is partitioned into two subprocesses $\mathbf{X}^{(h)} = \{\mathbf{X}_t^{(h)}, \; t \in \mathbb{Z}\}$, $h = 1, 2$, with m_1 and m_2 components respectively $(m_1 + m_2 = m)$, that is $\mathbf{X}_t = (\mathbf{X}_t^{(1)^T}, \mathbf{X}_t^{(2)^T})^T$. The autocovariance matrix $\mathbf{\Gamma}_{\mathbf{X}}(j)$ is partitioned as

$$\mathbf{\Gamma}_{\mathbf{X}}(j) \;\; = \;\; \begin{pmatrix} \mathbf{\Gamma}_{\mathbf{X}}^{(11)}(j) & \mathbf{\Gamma}_{\mathbf{X}}^{(12)}(j) \\ \mathbf{\Gamma}_{\mathbf{X}}^{(21)}(j) & \mathbf{\Gamma}_{\mathbf{X}}^{(22)}(j) \end{pmatrix} \;, \; j \in \mathbb{Z},$$

where $\mathbf{\Gamma}_{\mathbf{X}}^{(hh)}(j)$ is the autocovariance matrix at lag j of the process $\mathbf{X}^{(h)}$, $h = 1, 2$, and $\mathbf{\Gamma}_{\mathbf{X}}^{(12)}(j)$ is the cross-covariance matrix at lag j between $\{\mathbf{X}_t^{(1)}\}$ and $\{\mathbf{X}_t^{(2)}\}$ and $\mathbf{\Gamma}_{\mathbf{X}}^{(21)}(j) = \mathbf{\Gamma}_{\mathbf{X}}^{(12)}(-j)^T$. The autocorrelation matrix $\boldsymbol{\rho}_{\mathbf{X}}(j)$, the spectral density matrix $\boldsymbol{f}_{\mathbf{X}}(\lambda)$ and the coherency function $\boldsymbol{S}_{\mathbf{X}}(\lambda)$ can also be partitioned in a similar way. Given a realization of length N of the process $\mathbf{X}$, the sample cross-covariance matrix at lag j is defined by

$$\boldsymbol{C}_{\mathbf{X}}^{(12)}(j) \;\; = \;\; N^{-1} \sum_{t=j+1}^{N} \mathbf{X}_t^{(1)} \mathbf{X}_{t-j}^{(2)^T}, \; 0 \leq j \leq N-1. \tag{2.6}$$

Also, for $-N+1 \leq j \leq 0$, $\boldsymbol{C}_{\mathbf{X}}^{(12)}(-j) = \boldsymbol{C}_{\mathbf{X}}^{(21)}(j)^T$ and $\boldsymbol{C}_{\mathbf{X}}^{(12)}(j) = 0$ for $|j| \geq N$. The sample cross-correlation matrix at lag j is given by

$$\mathbf{R}_{\mathbf{X}}^{(12)}(j) \;\; = \;\; \mathbf{D}\{c_{\mathbf{X},ii}^{(11)}(0)^{-1}\} \boldsymbol{C}_{\mathbf{X}}^{(12)}(j) \mathbf{D}\{c_{\mathbf{X},ii}^{(22)}(0)^{-1}\}. \tag{2.7}$$

We denote by $\boldsymbol{r}_{\mathbf{X}}^{(12)}(j) = vec(\mathbf{R}_{\mathbf{X}}^{(12)}(j))$ where the symbol vec stands for the usual operator that transforms a matrix into a vector by stacking its columns.

In the sequel, we suppose that for $h = 1, 2$, $\mathbf{X}^{(h)}$ satisfy (2.3) and Assumption A and we want to test the null hypothesis that they are uncorrelated (or independent in the Gaussian case), that is $\boldsymbol{\rho}_{\mathbf{X}}^{(12)}(j) = \mathbf{0}$, $j \in \mathbb{Z}$. As in El Himdi and Roy (1997), this hypothesis is equivalent to

$$\boldsymbol{\rho}_a^{(12)}(j) \;\; = \;\; \mathbf{0} \;, \qquad j \in \mathbb{Z}. \tag{2.8}$$

Each series $\mathbf{X}_1^{(h)}, ..., \mathbf{X}_N^{(h)}$ is described by a finite-order autoregressive model VAR(p_h). The order p_h depends on N. The resulting residuals are given by

$$\hat{\boldsymbol{a}}_t^{(h)} = \begin{cases} \mathbf{X}_t^{(h)} - \sum_{l=1}^{p_h} \hat{\boldsymbol{\Phi}}_{l,p_h} \mathbf{X}_{t-l}^{(h)} & \text{if } t = p_h + 1, ..., N \ , \\ \mathbf{0} & \text{if } t \le p_h, \end{cases} \tag{2.9}$$

where the $\hat{\boldsymbol{\Phi}}_{l,p_h}$ are the Yule-Walker estimators defined by (2.4). The residual cross-covariance matrix $\boldsymbol{C}_{\hat{\boldsymbol{a}}}^{(12)}(j)$ is defined by

$$\boldsymbol{C}_{\hat{\boldsymbol{a}}}^{(12)}(j) = \begin{cases} N^{-1} \sum_{t=j+1}^{N} \hat{\boldsymbol{a}}_t^{(1)} \hat{\boldsymbol{a}}_{t-j}^{(2)^T} & \text{if } 0 \le j \le N - 1 \ , \\ N^{-1} \sum_{t=-j+1}^{N} \hat{\boldsymbol{a}}_{t+j}^{(1)} \hat{\boldsymbol{a}}_t^{(2)^T} & \text{if } -N + 1 \le j \le 0 \ , \end{cases} \tag{2.10}$$

the corresponding residual cross-correlation matrix is given by

$$\mathbf{R}_{\hat{\boldsymbol{a}}}^{(12)}(j) \quad = \quad \mathbf{D}\{c_{\hat{\boldsymbol{a}},ii}^{(11)}(0)^{-1}\} \boldsymbol{C}_{\hat{\boldsymbol{a}}}^{(12)}(j) \mathbf{D}\{c_{\hat{\boldsymbol{a}},ii}^{(11)}(0)^{-1}\}.$$

A commonly used nonparametric kernel-based estimator of the standardized cross-spectral density function $\mathbf{f}_{\boldsymbol{a}}^{(12)}(\lambda)$ is given by

$$\hat{\boldsymbol{f}}_{\boldsymbol{a}}^{(12)}(\lambda) \quad = \quad \frac{1}{2\pi} \sum_{j=-N+1}^{N-1} k(j/M) \mathbf{R}_{\hat{\boldsymbol{a}}}^{(12)}(j) e^{-i\lambda j}, \tag{2.11}$$

where $k(.)$ is a suitable kernel function. The parameter M is a truncation point when the kernel is of compact support, or a smoothing parameter when the kernel support is unbounded. We suppose that M is function of N such that $M \to \infty$ and $M/N \to 0$ when $N \to \infty$. The most commonly used kernels typically give more weight to lower lags and less weight to higher ones. An exception is the truncated uniform kernel $k_T(z) = \mathbb{I}[|z| \le 1]$, where $\mathbb{I}(A)$ represents the indicator function of the set A, which gives the same weight to all lags. In the sequel, we suppose that the kernel function k satisfies the following assumption.

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

The property $k(0) = 1$ implies that the weights assigned to the lower lags are close to unity. The square integrability of the kernel k implies that $k(z) \to 0$ as $|z| \to \infty$. Thus, eventually, less weight is given to $\mathbf{R}_{\hat{\boldsymbol{a}}}^{(12)}(j)$ as j increases. It is worth noting that all the kernels used in spectral analysis satisfy Assumption B, see Priestley (1981, Section 6.2.3).

3 The test statistic and its asymptotic null distribution

With multivariate time series, the squared cross-correlation $r_{\hat{a}}^{(12)}(j)^2$ in (1.1) is replaced by a quadratic form in the vector $\boldsymbol{r}_{\hat{a}}^{(12)}(j) = vec(\mathbf{R}_{\hat{a}}^{(12)}(j))$. The test is based on the following sum of weighted quadratic forms at all possible lags

$$\mathcal{T}(\hat{\boldsymbol{a}}, \hat{\boldsymbol{\Sigma}}) = \sum_{j=1-N}^{N-1} k^2(j/M) Q_{\hat{a}}^{(12)}(j) \tag{3.1}$$

where

$$Q_{\hat{a}}(j) = N \boldsymbol{r}_{\hat{a}}^{(12)}(j)^T \left(\mathbf{R}_{\hat{a}}^{(22)}(0)^{-1} \otimes \mathbf{R}_{\hat{a}}^{(11)}(0)^{-1} \right) \boldsymbol{r}_{\hat{a}}^{(12)}(j), \tag{3.2}$$

and $k(.)$ is a kernel function satisfying Assumption B. The test statistic is a standardized version of $\mathcal{T}(\hat{\boldsymbol{a}}, \hat{\boldsymbol{\Sigma}})$ given by

$$\mathcal{Q}_N = \frac{\mathcal{T}(\hat{\boldsymbol{a}}, \hat{\boldsymbol{\Sigma}}) - m_1 m_2 S_N(k)}{\sqrt{2 m_1 m_2 D_N(k)}}, \tag{3.3}$$

where the smoothing parameter $M = M(N) \to \infty$ and $M/N \to 0$ when $N \to \infty$, and

$$S_N(k) = \sum_{j=1-N}^{N-1} (1 - |j|/N) k^2(j/M), \; D_N(k) = \sum_{j=2-N}^{N-2} (1 - |j|/N)(1 - (|j|+1)/N) k^4(j/M) \tag{3.4}$$

Note that $S_N(k)$ and $D_N(k)$ are essentially the asymptotic mean and variance of $\mathcal{T}(\hat{\boldsymbol{a}}, \hat{\boldsymbol{\Sigma}})$ under $\mathcal{H}_0$. If k is the truncated uniform kernel, apart from the standardization factors $S_N(k)$ and $D_N(k)$, $\mathcal{Q}_N$ corresponds to the multivariate version of Haugh's statistic introduced in El Himdi and Roy (1997) that is defined by

$$P_M = \sum_{j=-M}^{M} Q_{\hat{a}}(j). \tag{3.5}$$

In that case, M is a fixed integer that does not depend on the sample size N. The properties of P_M in the VAR(∞) context are studied in Bouhaddioui and Roy (2003). As it will be seen below, many kernels k yeald to a greater power than P_M.

Under some conditions on the smoothing parameter M and if the kernel k verifies the Assumption B, it is easily seen that

$$M^{-1} S_N(k) \to S(k) , \; M^{-1} D_N(k) \to D(k), \tag{3.6}$$

where

$$S(k) = \int_{-\infty}^{+\infty} k^2(z) dz , \; D(k) = \int_{-\infty}^{+\infty} k^4(z) dz. \tag{3.7}$$

An alternative statistic is obtained by replacing $S_N(k)$ and $D_N(k)$ by their asymptotic approximations $MS(k)$ an $MD(k)$ respectively and is defined by

$$\mathcal{Q}_N^* \;=\; \frac{\mathcal{T}(\hat{\boldsymbol{a}}, \hat{\boldsymbol{\Sigma}}) - M m_1 m_2 S(k)}{\sqrt{2 M m_1 m_2 D(k)}}. \tag{3.8}$$

Both $\mathcal{Q}_N$ and $\mathcal{Q}_N^*$ have the same asymptotic null distribution and power properties.

The statistic $\mathcal{Q}_N$ can also be expressed in term of the autocovariances $\mathbf{C}_{\hat{\boldsymbol{a}}}^{(hh)}(0)$ and the cross-covariances $\mathbf{C}_{\hat{\boldsymbol{a}}}^{(12)}(j)$ of the same residual series. Invoking Lemma 4.1 of El Himdi and Roy (1997), the quadratic form $\mathcal{T}(\hat{\boldsymbol{a}}, \hat{\boldsymbol{\Sigma}})$ can be written as follows in terms of the residual covariances:

$$\mathcal{T}(\hat{\boldsymbol{a}}, \hat{\boldsymbol{\Sigma}}) \;=\; N \sum_{j=1-N}^{N-1} k^2(j/M) \mathbf{c}_{\hat{\boldsymbol{a}}}^{(12)}(j)^T \left(\mathbf{C}_{\hat{\boldsymbol{a}}}^{(22)}(0)^{-1} \otimes \mathbf{C}_{\hat{\boldsymbol{a}}}^{(11)}(0)^{-1} \right) \mathbf{c}_{\hat{\boldsymbol{a}}}^{(12)}(j)$$

with $\mathbf{c}_{\hat{\boldsymbol{a}}}^{(12)}(j) = vec\left(\mathbf{C}_{\hat{\boldsymbol{a}}}^{(12)}(j) \right)$. Our main result is stated in the following theorem. In the sequel, $\xrightarrow{L}$ stands for convergence in law and $\xrightarrow{p}$ for convergence in probability.

Theorem 3.1 *Let $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ be two multivariate stationary processes that satisfy the VAR(∞) model (2.3) and suppose that Assumptions A and B hold. Let $M = M(N) \to \infty$, $M/N \to 0$ when $N \to \infty$ and let p_h, $h = 1, 2$, satisfy the following conditions*
(i) $p_h = o\left(\frac{N^{1/2}}{M^{1/4}} \right)$.
(ii) $N \sum_{j=p_h+1}^{\infty} ||\mathbf{\Phi}_j^{(h)}||^2 = o\left(\frac{N^{1/2}}{M^{1/4}} \right)$.
If the processes $\boldsymbol{a}^{(1)}$ and $\boldsymbol{a}^{(2)}$ are independent, the statistic $\mathcal{Q}_N$ defined by (3.3) has an asymptotic normal distribution, that is $\mathcal{Q}_N \xrightarrow{L} \mathcal{N}(0, 1)$.

Remark: For the practical implications of the conditions *(i)* and *(ii)*, let us consider the case of a stationary and invertible VARMA(p,q) process. In that situation, we know that for large j, $||\mathbf{\Phi}_j|| \leq C \rho^j$ where C and ρ are constants (independent of j) such that $C > 0$ and $\rho \in [0, 1]$. It follows that $N \sum_{j=p_h+1}^{\infty} \leq C_1 \rho^{2 p_h}$ where $C_1 > 0$ is another constant and Condition *(ii)* holds provided that $p_h(N) \to \infty$ at any rate faster that $ln(N)$.

PROOF OF THEOREM 3.1.

We start by defining the pseudo-statistic

$$\mathcal{T}(\boldsymbol{a}, \boldsymbol{\Sigma}) \;=\; N \sum_{j=1-N}^{N-1} k^2(j/M) \mathbf{c}_{\boldsymbol{a}}^{(12)}(j)^T \left(\boldsymbol{\Sigma}_2^{-1} \otimes \boldsymbol{\Sigma}_1^{-1} \right) \mathbf{c}_{\boldsymbol{a}}^{(12)}(j),$$

where $\mathbf{c}_{\mathbf{a}}^{(12)}(j)$ is defined as $\mathbf{c}_{\hat{\mathbf{a}}}^{(12)}(j)$ with the residual series $\left(\hat{\mathbf{a}}_t^{(1)}, \hat{\mathbf{a}}_t^{(2)}\right)_{t=1}^N$ replaced by the unobservable innovation series $\left(\mathbf{a}_t^{(1)}, \mathbf{a}_t^{(2)}\right)_{t=1}^N$ and $\mathbf{\Sigma}_h = \mathbf{\Gamma}_{\mathbf{a}}^{(hh)}(0)$, $h = 1, 2$, is the covariance matrix of $\mathbf{a}^{(h)}$. Also, we define $\mathcal{T}(\hat{\mathbf{a}}, \mathbf{\Sigma})$ by

$$\mathcal{T}(\hat{\mathbf{a}}, \mathbf{\Sigma}) \;=\; N \sum_{j=1-N}^{N-1} k^2(j/M) \mathbf{c}_{\hat{\mathbf{a}}}^{(12)}(j)^T (\mathbf{\Sigma}_2^{-1} \otimes \mathbf{\Sigma}_1^{-1}) \mathbf{c}_{\hat{\mathbf{a}}}^{(12)}(j).$$

Thus, with $\hat{\mathbf{\Sigma}}_h = \mathbf{C}_{\hat{\mathbf{a}}}^{(hh)}(0)$, $h = 1, 2$, we can write the statistic $\mathcal{Q}_N$ as

$$\begin{aligned}
\mathcal{Q}_N \;&=\; \frac{\mathcal{T}(\hat{\mathbf{a}}, \hat{\mathbf{\Sigma}}) - m_1 m_2 S_N(k)}{\sqrt{2 m_1 m_2 D_N(k)}} \\
&=\; \frac{\mathcal{T}(\hat{\mathbf{a}}, \mathbf{\Sigma}) - m_1 m_2 S_N(k)}{\sqrt{2 m_1 m_2 D_N(k)}} + \frac{\mathcal{T}(\hat{\mathbf{a}}, \hat{\mathbf{\Sigma}}) - \mathcal{T}(\hat{\mathbf{a}}, \mathbf{\Sigma})}{\sqrt{2 m_1 m_2 D_N(k)}}.
\end{aligned} \tag{3.9}$$

The asymptotic distribution of $\mathcal{Q}_N$ follows from the next two propositions.

Proposition 3.1 *Under the assumptions of Theorem 3.1, we have that*

$$\frac{\mathcal{T}(\hat{\mathbf{a}}, \mathbf{\Sigma}) - m_1 m_2 S_N(k)}{\sqrt{2 m_1 m_2 D_N(k)}} \quad \xrightarrow{L} \quad \mathcal{N}(0, 1).$$

PROOF.

With a decomposition similar to the one for $\mathcal{Q}_N$ in (3.9), we can write

$$\frac{\mathcal{T}(\hat{\mathbf{a}}, \mathbf{\Sigma}) - m_1 m_2 S_N(k)}{\sqrt{2 m_1 m_2 D_N(k)}} \;=\; \frac{\mathcal{T}(\mathbf{a}, \mathbf{\Sigma}) - m_1 m_2 S_N(k)}{\sqrt{2 m_1 m_2 D_N(k)}} + \frac{\mathcal{T}(\hat{\mathbf{a}}, \mathbf{\Sigma}) - \mathcal{T}(\mathbf{a}, \mathbf{\Sigma})}{\sqrt{2 m_1 m_2 D_N(k)}}$$

and Proposition 3.1 can be established using the following two lemmas.

Lemma 3.1 *Under the assumptions of Theorem 3.1, we have that*

$$\frac{\mathcal{T}(\mathbf{a}, \mathbf{\Sigma}) - m_1 m_2 S_N(k)}{\sqrt{2 m_1 m_2 D_N(k)}} \quad \xrightarrow{L} \quad \mathcal{N}(0, 1).$$

Lemma 3.2 *Under the assumptions of Theorem 3.1, it follows that*

$$\mathcal{T}(\hat{\mathbf{a}}, \mathbf{\Sigma}) - \mathcal{T}(\mathbf{a}, \mathbf{\Sigma}) \;=\; o_p(M^{1/2}).$$

The proofs of the these two lemmas are long and technical and are deferred to the Appendix.

Proposition 3.2 *Under the assumptions of Theorem 3.1, it follows that*

$$\frac{\mathcal{T}(\hat{\mathbf{a}}, \hat{\mathbf{\Sigma}}) - \mathcal{T}(\hat{\mathbf{a}}, \mathbf{\Sigma})}{\sqrt{2 m_1 m_2 D_N(k)}} \quad \xrightarrow{p} \quad 0.$$

PROOF.

Since $D_N(k) = MD(k)\{1 + o(1)\}$, it is sufficient to show that

$$\mathcal{T}(\hat{\mathbf{a}}, \hat{\boldsymbol{\Sigma}}) - \mathcal{T}(\hat{\mathbf{a}}, \boldsymbol{\Sigma}) = O_p(M/N^{1/2}).$$

Using the fact that $\mathbf{C}_{\hat{\mathbf{a}}}^{(hh)}(0) - \boldsymbol{\Sigma}_h = O_p(N^{-1/2})$, (see Lütkepohl (1991, p.309)), for $h = 1, 2$, it follows that

$$\mathbf{C}_{\hat{\mathbf{a}}}^{(22)}(0)^{-1} \otimes \mathbf{C}_{\hat{\mathbf{a}}}^{(11)}(0)^{-1} - \boldsymbol{\Sigma}_2^{-1} \otimes \boldsymbol{\Sigma}_1^{-1} = O_p(N^{-1/2}).$$

Thus,

$$\mathcal{T}(\hat{\mathbf{a}}, \hat{\boldsymbol{\Sigma}}) - \mathcal{T}(\hat{\mathbf{a}}, \boldsymbol{\Sigma}) = N \sum_{j=1-N}^{N-1} k^2(j/M) \mathbf{c}_{\hat{\mathbf{a}}}^{(12)}(j)^T O_p(N^{-1/2}) \mathbf{c}_{\hat{\mathbf{a}}}^{(12)}(j)$$

$$= O_p(N^{1/2}) \sum_{j=1-N}^{N-1} k^2(j/M) \mathbf{c}_{\hat{\mathbf{a}}}^{(12)}(j)^T \mathbf{c}_{\hat{\mathbf{a}}}^{(12)}(j).$$

To complete the proof, it remains to prove that

$$\mathcal{B}(N) = \sum_{j=1-N}^{N-1} k^2(j/M) \mathbf{c}_{\hat{\mathbf{a}}}^{(12)}(j)^T \mathbf{c}_{\hat{\mathbf{a}}}^{(12)}(j) = O_p(M/N).$$

First, let us decompose $\mathcal{B}(N)$ in two parts

$$\mathcal{B}(N) = \sum_{j=1-N}^{N-1} k^2(j/M)\{\mathbf{c}_{\hat{\mathbf{a}}}^{(12)}(j)^T \mathbf{c}_{\hat{\mathbf{a}}}^{(12)}(j) - \mathbf{c}_{\mathbf{a}}^{(12)}(j)^T \mathbf{c}_{\mathbf{a}}^{(12)}(j)\} + \sum_{j=1-N}^{N-1} k^2(j/M) \mathbf{c}_{\mathbf{a}}^{(12)}(j)^T \mathbf{c}_{\mathbf{a}}^{(12)}(j)$$

$$= \mathcal{B}_1 + \mathcal{B}_2.$$

By an argument similar to the one used in the second part of the proof of Lemma 3.2 in the Appendix, we have that

$$\mathcal{B}_1(N) = \sum_{j=1-N}^{N-1} k^2(j/M)\{\mathbf{c}_{\hat{\mathbf{a}}}^{(12)}(j)^T \mathbf{c}_{\hat{\mathbf{a}}}^{(12)}(j) - \mathbf{c}_{\mathbf{a}}^{(12)}(j)^T \mathbf{c}_{\mathbf{a}}^{(12)}(j)\} = o_p(M^{1/2}/N),$$

and by Markov inequality, it follows that

$$\mathcal{B}_2(N) = \sum_{j=1-N}^{N-1} k^2(j/M) \mathbf{c}_{\mathbf{a}}^{(12)}(j)^T \mathbf{c}_{\mathbf{a}}^{(12)}(j) = O_p(M/N).$$

Combining the results for $\mathcal{B}_1(N)$ and $\mathcal{B}_2(N)$, we obtain that

$$\mathcal{T}(\hat{\mathbf{a}}, \hat{\boldsymbol{\Sigma}}) - \mathcal{T}(\hat{\mathbf{a}}, \boldsymbol{\Sigma})) = O_p(N^{1/2}) O_p(M/N) = O_p(M/N^{1/2}),$$

and the proof of Proposition 3.2 is completed.

10

4 Consistency of the generalized test

We now investigate the asymptotic power of the test $\mathcal{Q}_N$ under fixed alternatives. We consider an alternative $\mathcal{H}_1$ of serial cross-correlation between the two innovation processes $\boldsymbol{a}^{(1)}$ and $\boldsymbol{a}^{(2)}$ that satisfies the following general assumptions.

Assumption C *The two innovation processes $\boldsymbol{a}^{(1)}$ et $\boldsymbol{a}^{(2)}$ are jointly fourth-order stationary and their cross-correlation structure is such that $\boldsymbol{\Gamma}_{\boldsymbol{a}}^{(12)}(j) \neq \boldsymbol{0}$ for at least one value of j and*

$$\sum_{j=-\infty}^{+\infty} \|\boldsymbol{\Gamma}_{\boldsymbol{a}}^{(12)}(j)\|^2 < \infty \,, \quad \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} |\kappa_{uvuv}(0,i,j,l)| < \infty,$$

where $\kappa_{uvuv}(0,i,j,l)$ is the fourth-order cumulant of the joint distribution of $a_{u,t}^{(1)}$, $a_{v,t+i}^{(2)}$, $a_{u,t+j}^{(1)}$, $a_{v,t+l}^{(2)}$.

If the joint process $\{(\boldsymbol{a}_t^{(1)^T}, \boldsymbol{a}_t^{(1)^T})^T\}$ is Gaussian, the fourth-order cumulants are zero and the cumulant condition is trivially satisfied. Fourth-order stationary linear processes with absolutely summable coefficients and with innovations whose fourth-order moments exist, also satisfy the cumulant condition, see Hannan (1970, p. 211).

The statistic $\mathcal{Q}_N$ is a normalized version of $\mathcal{T}(\hat{\boldsymbol{a}}, \hat{\boldsymbol{\Sigma}})$ which can be viewed as the $\mathcal{L}_2$-norm of a kernel-based estimator of the cross-coherency function between the two innovations processes. Indeed, the cross-coherency function $\boldsymbol{S}_{\boldsymbol{a}}^{(12)}(w)$ between the two innovation processes is given by

$$\boldsymbol{S}_{\boldsymbol{a}}^{(12)}(w) \quad = \quad \sum_{j=-\infty}^{\infty} \boldsymbol{\Sigma}_1^{-1/2} \boldsymbol{\Gamma}_{\boldsymbol{a}}^{(12)}(j) \boldsymbol{\Sigma}_2^{-1/2} e^{-iwj},$$

where $\boldsymbol{\Sigma}_h$ is the covariance matrix of $\boldsymbol{a}^{(h)}$, $h = 1, 2$. Using the relation $vec\,(ABC) = (C \otimes A)\,vec\,(B)$, we have

$$\boldsymbol{s}_{\boldsymbol{a}}^{(12)}(w) \quad = \quad \sum_{j=-\infty}^{\infty} (\boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_1)^{-1/2} \boldsymbol{\gamma}_{\boldsymbol{a}}^{(12)}(j) e^{-iwj}$$

where $\boldsymbol{s}_{\boldsymbol{a}}^{(12)}(w) = vec\left(\boldsymbol{S}_{\boldsymbol{a}}^{(12)}(w)\right)$ and $\boldsymbol{\gamma}_{\boldsymbol{a}}^{(12)}(j) = vec(\boldsymbol{\Gamma}_{\boldsymbol{a}}^{(12)}(j))$. Employing the following quadratic norm $\|.\|_2^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |.|^2 dw$, where $|.|$ represents the modulus of a complex number, we have

$$\|\boldsymbol{s}_{\boldsymbol{a}}^{(12)}\|_2^2 \quad = \quad \sum_{j=-\infty}^{\infty} \boldsymbol{\gamma}_{\boldsymbol{a}}^{(12)}(j)^T (\boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_1)^{-1} \boldsymbol{\gamma}_{\boldsymbol{a}}^{(12)}(j). \tag{4.1}$$

The following theorem gives sufficient conditions for the consistency of $\mathcal{Q}_N$ under a fixed alternative hypothesis.

Theorem 4.1 *Let $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ be two multivariate stationary processes that satisfy the VAR(∞) model (2.3) and suppose that their innovation processes $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(1)}$ follow Assumptions A and C. If the kernel $k(.)$ satisfies B and if p_h, $h = 1, 2$, are such that*

$$p_h^2 \;=\; o\left(\frac{N}{M}\right) \qquad , \qquad \sum_{j=p_h+1}^{\infty} \|\mathbf{\Phi}_j^{(h)}\|^2 = o\left(M^{-1}\right).$$

Then, we have

$$\left(\frac{M^{1/2}}{N}\right)\mathcal{Q}_N \;\overset{p}{\to}\; \frac{\|\mathbf{s_x}^{(12)}\|_2^2}{\{2m_1 m_2 D(k)\}^{1/2}} \tag{4.2}$$

where $\mathbf{s_x}^{(12)} = vec(\mathbf{S_x}^{(12)})$, $\mathbf{S_x}^{(12)}$ being the cross-coherency function between $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$.

The proof is given in the appendix. This result is a multivariate version of Theorem 4 in Hong (1996a). Under a fixed alternative, this theorem implies that $\mathcal{Q}_N$ goes to infinity at the rate $N/M^{1/2}$. Thus, the slower M grows, the faster $\mathcal{Q}_N$ will approach to infinity and the test will be more powerful.

Since $\mathcal{Q}_N$ depends on a kernel function and that under the null hypothesis $\mathcal{H}_0$, the statistic $\mathcal{Q}_N$ is asymptotically normal, we can use the concept of asymptotic slope introduced by Bahadur (1960) to compare two kernels k_1 and k_2 for a given alternative $\mathcal{H}_1$. For a given kernel k, let $\mathcal{Q}_N(k)$ be the corresponding statistic. The Bahadur's slope criterion is useful for large sample tests under fixed alternatives and is defined as the rate at which the asymptotic p-value goes to zero as $N \to \infty$. Thus, for the test $\mathcal{Q}_N(k)$, the asymptotic p-value is given by $1 - \Phi(\mathcal{Q}_N)$ where $\Phi(.)$ denotes the $N(0,1)$ cumulative distribution function. Now define

$$\aleph_N(k) \;=\; -2ln\{1 - \Phi(\mathcal{Q}_N)\}. \tag{4.3}$$

As shown in Bahadur (1960, p. 283), we can use the relation $ln\{1 - \Phi(\xi)\} = -\frac{1}{2}\xi^2\{1 + o(1)\}$ for a large ξ, and by (4.2), we have that

$$\left(\frac{M}{N^2}\right)\aleph_N(k) \;\overset{p}{\to}\; \frac{\|\mathbf{s_x}^{(12)}\|^4}{2m_1 m_2 D(k)}, \tag{4.4}$$

under a fixed alternative as $M \to \infty$ and $M/N \to 0$. Following Bahadur (1960), the quantity $\frac{\|\mathbf{s_x}^{(12)}\|^4}{2m_1 m_2 D(k)}$ is the asymptotic slope of $\mathcal{Q}_N$. The Bahadur's asymptotic relative efficiency $ARE_B(k_2, k_1)$ of k_2 with respect of k_1 is by definition the limit ratio of the two sample sizes N_1 and N_2 required by the two test statistics to obtain the same asymptotic significance level under the alternative $\mathcal{H}_1$. If we take $M = N^{\nu}$, it is easily shown by standard arguments that

$$ARE_B(k_2, k_1) \;=\; \lim_{N_1, N_2 \to \infty} \frac{N_1}{N_2} = \left\{\frac{D(k_1)}{D(k_2)}\right\}^{\frac{1}{2-\nu}}. \tag{4.5}$$

12

For example, $ARE_B(k_{BAR}, k_{TR}) > 2.23$, where k_{BAR} and k_{TR} denote respectively the Bartlett and the truncated uniform kernels which are given in Table 2.

Many of the commonly used kernels in spectral estimation lead to an ARE_B greater than one with respect to the truncated uniform kernel. A test with a greater asymptotic slope may be expected to have a greater power for a fixed alternative than one with a smaller asymptotic slope. However, Geweke (1981) noticed that there is no clear analytical relationship between the slope of a test and its power function. Hence, for a specific alternative, we cannot conclude that a test with a greater asymptotic slope should be automatically preferred to one with a smaller asymptotic slope without analyzing further the finite sample properties of the two test statistics.

5   Simulation study

It is natural to inquire after the finite sample properties of the proposed test statistics , in particular their exact level and power. At this aim, a small Monte Carlo simulation was conducted. In addition to the test statistics discussed in the preceding sections, the multivariate version of Haugh's statistic P_M^* previously studied by El Himdi and Roy (1997) and Bouhaddioui and Roy (2003) was also included:

$$P_M^* \quad = \quad \sum_{j=-M}^{M} \frac{N}{N-|j|} Q_{\hat{a}}(j), \tag{5.1}$$

where $Q_{\hat{a}}^{(12)}(j)$ is given by (3.2). The statistic P_M^* is a slightly modified version of P_M defined by (3.5).

5.1   Description of the experiment

In the simulation experiment, we considered bivariate series $\{\mathbf{X}_t^{(1)}\}$ and $\{\mathbf{X}_t^{(2)}\}$ generated from the global 4-dimensional, stationary and invertible $\mathbf{AR}(1)$, $\mathbf{MA}(1)$ and $\mathbf{AR}_\delta(1)$ models described in Table 1. In the first two models, the two subprocesses $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are independent bivariate $\mathbf{AR}(1)$ or $\mathbf{MA}(1)$ and served for the level study. The third one, in which there is instantaneous correlation between the two innovation series, was used for the power study. The correlation depends on a parameter δ and the values $\delta = 1.0$, 1.5 and 2 were chosen. For each model, two series lengths ($N = 100$, 200), were considered. With the statistics $\mathcal{Q}_N$ and $\mathcal{Q}_N^*$ defined by (3.3) and (3.8) respectively, we used the four kernels described in Table 2. For each kernel, the following three truncation values M were employed: $M = [ln(N)]$, $[3N^{0.2}]$ et $[3N^{0.3}]$ ([a] denotes the integer part

of a). These rates are discussed in Hong (1996c, p. 849). They lead respectively to $M = 5, 8, 12$ for the length series $N = 100$, and to $M = 5, 9, 15$ for $N = 200$. The same truncation values were used for P_M^*.

In the level study, 5.000 independent realizations were generated from both $\mathbf{AR}(1)$ and $\mathbf{MA}(1)$ models for each series length N. Computations were made in the following way.

1. First, pseudo-random variables from the $\mathcal{N}(0,1)$ distribution were obtained with the pseudo-random normal generator of the S-plus package and were transformed into independent $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma_a})$ pseudo-random vectors using the Cholesky decomposition. Second, the $\mathbf{X}_t$ values were obtained by directly solving the difference equation defining a VARMA model. For the $\mathbf{AR}(1)$ model, $\mathbf{X}_1$ was generated from the exact $\mathcal{N}(\mathbf{0}, \mathbf{\Gamma_X(0)})$ distribution of the $\mathbf{X}_t$'s. The covariance $\mathbf{\Gamma_X}(0)$ was obtained by an algorithm of Ansley (1980).

2. For both series $\{\mathbf{X}_t^{(h)}, t = 1, \ldots, N\}$, $h = 1, 2$, autoregressions were fitted by conditional least squares. The autoregressive order was obtained by minimizing the AIC criterion for $p \le P$, where P was fixed to 12. The residual series $\{\hat{\boldsymbol{a}}_t^{(h)}\}$, $h = 1, 2$, were cross-correlated by computing the $\mathbf{R}_{\hat{\boldsymbol{a}}}^{(12)}(j)$'s as defined by (2.7).

3. For each realization, the test statistics $\mathcal{Q}_N$ and $\mathcal{Q}_N^*$ were compared for each of the four kernels and the three values of M. The same values of M were used for the statistic P_M^*. The values of the statistics $\mathcal{Q}_N$ and $\mathcal{Q}_N^*$ were compared with the $\mathcal{N}(0,1)$ critical values and those of P_M^* to the $\chi^2_{4(2M+1)}$ critical values.

4. Finally, for each model, each series length and each nominal level, the empirical frequencies of rejection of the null hypothesis of non-correlation were obtained from the 5.000 realizations. The results in percentage are reported in Table 3. The standard error of the empirical level is 0.14% for the nominal level 1%, 0.31% for 5% and 0.42% for 10%.

Computations for the power analysis were made in a similar way using the $\mathbf{AR}_\delta(1)$ model with different values of δ.

5.2 Discussion of the level study

Results from the level study are presented in Table 3. For both AR(1) and MA(1) models, we make the following observations. The asymptotic $\mathcal{N}(0,1)$ distribution provides a good approximation of the exact distributions of $\mathcal{Q}_N$ and $\mathcal{Q}_N^*$ at the three nominal levels, for the five considered kernels and for the three truncation values chosen. Almost all empirical levels are within three standard

errors of the corresponding nominal levels and the majority are within two standard errors. The statistic $\mathcal{Q}_N^*$ is slightly better approximated than $\mathcal{Q}_N$ since most of its empirical levels are within two standard errors of the nominal level. At the 1% and 10% nominal levels, both statistics have a small tendancy to under or over-reject. There is no significant difference between the kernels. The best approximation is obtained with the Bartlett-Priestley kernel and the less good one corresponds to the Parzen kernel. With the Bartlett-Priestley kernel, the empirical size is always within two standard errors of the nominal size. The size of P_M^* and of $\mathcal{Q}_N$ and $\mathcal{Q}_N^*$ with the truncated uniform kernel are very close. It is not surprising since $\mathcal{Q}_N$ and $\mathcal{Q}_N^*$ are linear transformations of P_M and P_M^* is nothing else than a finite sample-size modification of P_M. For the models considered, the values of the truncation parameter M has no significant effect on the size of the tests. Finally, when the series length N goes from 100 to 200, the approximation improves very slightly.

5.3 Discussion of the power study

The results are given in Table 4. With the $\mathbf{AR}_\delta(1)$, the cross-correlation at lag 0 between the two innovation series increases with δ and as expected, the power of the three tests considered also increases with δ. Since the relative behaviors pf the various tests are similar for the three values of δ (1, 1.5, 2), only the results for $\delta = 2$ are presented. Furthermore, we only present the result for $\mathcal{Q}_N^*$ since $\mathcal{Q}_N$ and $\mathcal{Q}_N^*$ have a similar behavior with respect to the kernels and the truncation values.

The following observations are made from Table 4. First, the power of all tests increases when the sample size varies from 100 to 200. Also, the power decreases as M increases. It is not surprising since the model considered is characterized by the lag 0 serial correlation and in such a situation, we expect that the tests assigning more weight to small lags will be more powerful than those assigning weights to a large number of lags. For the three significance levels and the three truncation values, the Daniel, Parzen, Bartlett and Bartlett-Priestley kernels lead to similar powers for the test $\mathcal{Q}_N^*$. However, the power of $\mathcal{Q}_N^*$ with the truncated uniform kernels is much smaller and is comparable to the power of P_M^*. At least for the chosen model, the new tests $\mathcal{Q}_N$ or $\mathcal{Q}_N^*$ with another kernel than the truncated uniform one should be prefer to the multivariate version of Haugh's test P_M^*.

6 Application

Here, we consider a set of seven quarterly series of Canadian and American economic indicators used in a study of Canadian monetary policy in order to investigate the relationships between the

Table 1: Time series models used in the simulation study.

MODELS	EQUATIONS	Σ_a
$\mathbf{AR}(1)$	$\begin{bmatrix} \mathbf{X}_t^{(1)} \\ \mathbf{X}_t^{(2)} \end{bmatrix} = \begin{bmatrix} \Phi^{(1)} & 0 \\ 0 & \Phi^{(2)} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{t-1}^{(1)} \\ \mathbf{X}_{t-1}^{(2)} \end{bmatrix} + \begin{bmatrix} \mathbf{a}_t^{(1)} \\ \mathbf{a}_t^{(2)} \end{bmatrix}$	$\begin{bmatrix} \Sigma_a^{(1)} & 0 \\ 0 & \Sigma_a^{(2)} \end{bmatrix}$
$\mathbf{MA}(1)$	$\begin{bmatrix} \mathbf{X}_t^{(1)} \\ \mathbf{X}_t^{(2)} \end{bmatrix} = \begin{bmatrix} \Psi^{(1)} & 0 \\ 0 & \Psi^{(2)} \end{bmatrix} \begin{bmatrix} \mathbf{a}_{t-1}^{(1)} \\ \mathbf{a}_{t-1}^{(2)} \end{bmatrix} + \begin{bmatrix} \mathbf{a}_t^{(1)} \\ \mathbf{a}_t^{(2)} \end{bmatrix}$	$\begin{bmatrix} \Sigma_a^{(1)} & 0 \\ 0 & \Sigma_a^{(2)} \end{bmatrix}$
$\mathbf{AR}_\delta(1)$	$\begin{bmatrix} \mathbf{X}_t^{(1)} \\ \mathbf{X}_t^{(2)} \end{bmatrix} = \begin{bmatrix} \Phi^{(1)} & 0 \\ 0 & \Phi^{(2)} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{t-1}^{(1)} \\ \mathbf{X}_{t-1}^{(2)} \end{bmatrix} + \begin{bmatrix} \mathbf{a}_t^{(1)} \\ \mathbf{a}_t^{(2)} \end{bmatrix}$	$\begin{bmatrix} \Sigma_a^{(1)} & \Sigma_{a,\delta}^{(12)} \\ \Sigma_{a,\delta}^{(12)} & \Sigma_a^{(2)} \end{bmatrix}$

PARAMETERS VALUES

$$\Phi^{(1)} = \begin{bmatrix} 1.2 & -0.5 \\ 0.6 & 0.3 \end{bmatrix} \qquad \Phi^{(2)} = \begin{bmatrix} -0.6 & 0.3 \\ 0.3 & 0.6 \end{bmatrix} \qquad \Psi^{(1)} = \begin{bmatrix} -0.2 & 0.3 \\ -0.6 & 1.1 \end{bmatrix}$$

$$\Psi^{(2)} = \begin{bmatrix} 0.8 & 0.3 \\ 0.1 & 0.6 \end{bmatrix} \qquad \Sigma_a^{(1)} = \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix} \qquad \Sigma_a^{(2)} = \begin{bmatrix} 1.0 & 0.75 \\ 0.75 & 1.0 \end{bmatrix}$$

$$\Sigma_{a,\delta}^{(12)} = \begin{bmatrix} 0.1\delta & 0 \\ 0 & 0.05\delta \end{bmatrix}$$

Table 2: Kernels used with the test statistics $\mathcal{Q}_N$ and $\mathcal{Q}_N^*$.

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

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

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

Parzen (PAR): $\quad k(z) = \begin{cases} 1 - 6z^2 + 6|z|^3, & \text{if } |z| \leq 0.5, \\ 2\left(1 - |z|\right)^3, & \text{if } 0.5 \leq |z| \leq 1, \\ 0, & \text{otherwise.} \end{cases}$

Bartlett-Priestley (BP): $\quad k(z) = \frac{3}{(\pi z)^2}\{\frac{\sin(\pi z)}{\pi z} - \cos(\pi z)\}, \ z \in \mathbb{R}.$

Table 3: Empirical level (in percentage) of the test $\mathcal{Q}_N$, $\mathcal{Q}_N^*$ and P_M^* based on 5000 realizations for different kernels, different truncation values, for the **AR**(1) and **MA**(1) models.

	N	M	$\alpha\%$	$\mathcal{Q}_N$					$\mathcal{Q}_N^*$					P_M^*
				DAN	PAR	BAR	BP	TR	DAN	PAR	BAR	BP	TR	
AR(1)	100	5	1	0.8	0.5	0.7	0.7	0.6	0.9	0.6	0.8	0.8	0.7	0.7
			5	6.1	3.4	5.7	4.8	4.6	4.8	4.2	4.8	5.1	4.7	4.4
			10	9.4	7.8	9.5	10.2	8.4	9.7	8.2	9.2	10.3	9.3	9.1
		8	1	0.9	0.6	0.8	1.3	0.6	1.5	0.8	0.9	1.2	0.9	0.7
			5	5.4	3.9	5.9	5.2	3.6	5.2	4.2	5.8	4.8	4.1	4.7
			10	10.5	8.9	11.2	10.7	7.1	9.4	9.3	10.8	10.2	8.9	8.9
		12	1	0.7	0.6	0.8	0.9	0.5	0.8	0.7	1.2	1.1	0.6	0.7
			5	5.1	4.6	5.2	4.8	4.1	4.2	4.3	4.9	4.7	4.4	4.2
			10	10.6	8.6	11.0	11.1	7.5	10.3	9.2	10.6	10.4	7.9	8.2
	200	5	1	0.8	0.6	0.7	0.8	0.7	0.8	0.8	0.8	1.1	0.8	0.8
			5	5.9	4.8	5.5	5.7	4.2	6.1	4.5	4.9	5.2	4.8	4.1
			10	9.1	8.3	9.2	9.4	7.9	8.5	8.4	9.4	9.6	8.6	8.7
		9	1	0.8	0.6	0.7	0.8	0.7	0.8	0.7	0.7	1.2	0.8	0.7
			5	6.3	4.2	5.7	5.7	4.2	5.8	4.4	5.4	5.6	4.7	4.4
			10	9.7	8.9	10.4	10.2	7.3	8.9	9.0	9.5	9.7	8.9	9.0
		15	1	1.2	0.7	1.3	0.8	0.7	0.9	0.8	1.2	0.8	0.7	0.8
			5	6.4	4.1	6.0	5.8	3.9	5.1	4.3	5.8	5.2	4.3	4.6
			10	10.2	8.9	11.1	10.4	6.9	9.7	9.1	10.8	10.6	8.5	8.9
MA(1)	100	5	1	0.8	0.6	0.7	0.9	0.6	0.9	0.7	1.3	1.1	0.7	0.8
			5	5.9	4.2	4.4	5.2	4.6	4.8	4.3	5.7	5.1	4.7	4.8
			10	10.3	7.8	8.5	10.2	8.2	9.6	9.5	8.9	9.8	8.7	8.7
		8	1	0.8	0.5	0.6	1.3	0.6	1.1	0.6	0.7	1.1	0.7	1.1
			5	5.4	4.2	5.7	5.6	3.6	5.2	4.3	4.6	4.8	4.1	5.2
			10	10.3	8.2	9.1	10.1	7.1	9.4	8.6	9.3	9.2	7.9	9.4
		12	1	0.8	0.6	1.4	0.9	0.7	0.8	0.7	1.2	1.1	0.8	0.8
			5	5.4	4.7	5.6	5.8	4.6	5.2	4.9	5.6	5.2	4.8	4.3
			10	9.3	8.6	9.1	9.2	7.5	9.5	8.4	8.9	9.4	7.9	7.9
	200	5	1	0.7	0.6	0.7	0.8	0.7	0.8	0.7	0.9	1.1	0.8	0.7
			5	5.9	4.5	5.3	5.7	4.2	6.1	4.4	5.4	5.2	4.8	6.1
			10	9.2	8.3	8.5	9.0	7.9	9.5	8.4	10.2	9.6	8.6	8.5
		9	1	0.8	0.9	1.4	0.8	0.7	0.8	0.9	9.5	1.0	0.8	0.8
			5	6.3	4.1	4.6	5.7	4.2	5.8	4.5	4.8	5.6	4.7	5.8
			10	9.7	8.9	9.3	10.5	7.3	9.2	9.1	9.5	9.7	8.9	8.9
		15	1	1.1	0.8	0.9	0.9	0.7	0.8	0.8	0.9	1.1	0.9	0.8
			5	6.4	4.4	5.5	5.8	4.5	5.1	4.5	4.6	5.4	4.6	5.1
			10	10.2	9.1	10.3	9.2	6.9	9.7	9.5	10.4	10.1	9.5	8.7

Table 4: Power of the tests $\mathcal{Q}_N$, $\mathcal{Q}_N^*$ and P_M^* based on their asymptotic critical values for different kernels and different truncation values for the $AR_\delta(1)$ data generation process with $\delta = 2$.

N	M	$\alpha\%$	$\mathcal{Q}_N^*$					P_M^*
			DAN	PAR	BAR	BP	TR	
100	5	1	54.2	52.7	52.1	53.1	31.2	27.8
		5	59.1	61.9	56.8	58.7	34.2	28.4
		10	64.8	66.2	65.3	67.3	36.7	29.3
	8	1	51.2	48.9	50.2	49.0	25.7	19.4
		5	56.2	54.0	56.0	55.0	30.1	20.2
		10	62.3	60.6	61.4	63.5	31.9	22.6
	12	1	46.2	44.7	46.1	45.3	23.4	19.8
		5	51.2	48.3	50.2	52.0	26.4	20.5
		10	54.9	52.6	54.6	53.5	26.8	22.9
200	5	1	76.8	72.6	73.6	74.0	50.8	54.8
		5	84.2	82.6	80.4	84.8	52.8	57.1
		10	92.5	90.4	88.8	90.6	60.6	58.9
	9	1	66.4	62.9	64.0	63.8	40.2	45.7
		5	71.2	72.4	70.6	74.6	42.5	47.9
		10	77.8	75.4	79.6	78.8	47.4	48.6
	15	1	54.8	53.4	50.8	55.1	29.9	39.8
		5	62.1	56.4	51.0	61.1	31.3	42.5
		10	62.6	54.6	55.2	60.6	29.5	44.6

two economies; see Racette and Raynauld (1992). The Canadian economic indicators are the gross domestic production (GDP) in constant 1982 dollars, the implicit price index of the gross domestic production (GDPI), the nominal short-term interest rate (TX.CA) and the monetary basis value (M1). The other three variables represent the American gross national product (GNP) in constant 1982 dollars, the implicit price index of the American gross national production (GNPI), and the nominal short-term American interest rate (TX.US). In this study, the observation period extends from the first quarter of 1970 through to the last quarter of 1989. The data sources with the corresponding CANSIM series numbers are given in Table 1 of Racette and Raynauld (1992). The natural logarithm of M1 was taken in order to stabilize its variance, and all series except interest rates were differenced to have stationarity.

In the sequel, the two vector series of Canadian and American data, denoted by $\{\mathbf{X}_t^{(1)}\}$ and $\{\mathbf{X}_t^{(2)}\}$, are defined by

$$
\mathbf{X}_t^{(1)} = \begin{bmatrix} \frac{1}{1000}(1-B)GDP_t \\ 10(1-B)GDPI_t \\ TX.CA_t \\ 100(1-B)ln(M1_t) \end{bmatrix}, \qquad \mathbf{X}_t^{(2)} = \begin{bmatrix} \frac{1}{10}(1-B)GNP_t, \\ 10(1-B)GNPI_t, \\ TX.US_t. \end{bmatrix}.
$$

The multiplicative factors appearing in the definition of these series are the same as those used in El Himdi and Roy (1997). With these factors, the sample variances of the variables within each of the two vector series are of the same order of magnitude. Autoregressive AR(p) models were fitted to each series using the STEPAR procedure of the SCA statistical package. The autoregressive order p was obtained by minimizing the AIC criterion. For the Canadian series, this procedure led to $p = 11$. After deleting the non-significant matrix coefficients at 5% significance level, we finally obtain the following model:

$$
\mathbf{X}_t^{(1)} = \boldsymbol{\mu}_0^{(1)} + \boldsymbol{\Phi}_1^{(1)}\mathbf{X}_{t-1}^{(1)} + \boldsymbol{\Phi}_2^{(1)}\mathbf{X}_{t-2}^{(1)} + \boldsymbol{\Phi}_{11}^{(1)}\mathbf{X}_{t-11}^{(1)} + \mathbf{a}_t^{(1)} , \; t \in \mathbb{Z}. \tag{6.1}
$$

At the estimation stage, the full model was reestimated by the Gaussian maximum likelihood method (the "exact" method available in SCA) and then, each parameter estimate smaller than one standard error, in absolute value, was set at zero. The reduced model was reestimated until all the parameter estimates were greater than one standard error, in absolute value. The final parameter estimates of model (6.1) are given in Table 5.

In similar way, an AR(10) model was identified for the American series

$$
\mathbf{X}_t^{(2)} = \boldsymbol{\mu}_0^{(2)} + \boldsymbol{\Phi}_1^{(2)}\mathbf{X}_{t-1}^{(2)} + \boldsymbol{\Phi}_2^{(2)}\mathbf{X}_{t-2}^{(2)} + \boldsymbol{\Phi}_3^{(2)}\mathbf{X}_{t-3}^{(2)} + \boldsymbol{\Phi}_{10}^{(2)}\mathbf{X}_{t-10}^{(2)} + \mathbf{a}_t^{(2)} , \; t \in \mathbb{Z}. \tag{6.2}
$$

19

and the maximum likelihood method led to the final estimates given in Table 5. The two final models are stationary and satisfy the diagnostic checking procedure suggested by Tiao and Box (1981) for model adequacy.

Values of the statistics

$$Q_{\hat{a}}(j)^* = \frac{N}{N - |j|} Q_{\hat{a}}(j) \tag{6.3}$$

where $Q_{\hat{a}}(j)$ defined by (3.2), are displayed in Figure 1. At the significance level $\alpha = 0.05$, the asymptotic critical value for testing the null hypothesis $\mathcal{H}_0$ of noncorrelation between $\boldsymbol{a}^{(1)}$ and $\boldsymbol{a}^{(2)}$ against the alternative $\mathcal{H}_{1j} : \boldsymbol{\rho}_{\boldsymbol{a}}^{(12)}(j) \neq \boldsymbol{0}$ is 21.02 and only the three cross-correlation $\boldsymbol{\rho}_{\boldsymbol{a}}^{(12)}(j)$, $j = -1, 0, 2$, are significantly different from zero. The p-value of the portmanteau test P_M^* for $\mathcal{H}_0$ are also reported in Table 6, for $M = 1, ..12$. At the 0.05 significance level, $\mathcal{H}_0$ is rejected for all values of M such that $M \leq 9$.

The values of the global tests $\mathcal{Q}_N$ and $\mathcal{Q}_N^*$ and the corresponding p-values are reported in Table 8 for the truncated uniform , Daniell and Bartlett-Priestley kernels. As in the simulation study, the truncation values are $[ln(N)]$, $[3N^{0.2}]$ and $[3N^{0.3}]$ which correspond to 4, 7 and 11 respectively. The Daniell kernel is more powerful in the Bahadur's sense whilst the Bartlett-Priestley one has had the best size in the simulation study. At the 5% significance level, the tests based on $\mathcal{Q}_N$ and $\mathcal{Q}_N^*$ reject the hypothesis of non-correlation between the two series with DAN and BP kernels for the three values of M. With the truncated uniform kernel, the conclusion is the same with $\mathcal{Q}_N$ and $\mathcal{Q}_N^*$ does not reject when $M = 11$. The difference between the values of $\mathcal{Q}_N$ and $\mathcal{Q}_N^*$ is due to the fact that the length of the series is pretty small ($N = 68$) and that $MS(k)$ and $MD(k)$ do not provide good approximations of $S_N(k)$ and $D_N(k)$ as illustrated in Table 7.

Figure 1 indicates that there is a rather strong instantaneous correlation between the two series and the null hypothesis of non-correlation between them is rejected with the test $Q_{\hat{a}}(0)$ which is solely based on the cross-correlation matrix at lag 0. The portmanteau test P_M^* does not reject when $M > 9$ whilst the new tests $\mathcal{Q}_N$ or $\mathcal{Q}_N^*$, with DAN or BP kernels, reject for the three values of M considered. This conclusion is coherent with the simulation study which shows that $\mathcal{Q}_N$ and $\mathcal{Q}_N^*$ with one of the four kernels DAN, PAR, BAR and PB, are considerably more powerful than P_M^*.

Table 5: Estimated AR models for the Canadian and American series

(a) Canadian series

$$
\begin{bmatrix}
1 - .109\,B^{11} & 1.07\,B & -.177\,B^{11} & -.516\,B^2 - .049\,B^{11} \\
(0.088) & (0.117) & (0.112) & (0.089)\ (0.104) \\
0 & 1 - .576\,B + .189\,B^{11} & 0 & 0 \\
& (0.094)\ (0.075) & & \\
-.179\,B & -.177\,B - .39\,B^2 & 1 - 0.888\,B & -.106\,B \\
(0.044) & (0.060)\ (0.061) & (0.074) & (0.049) \\
1.129\,B^2 & -.445\,B & 1.718\,B - .779\,B^2 & 1 + .253\,B \\
(0.093) & (0.110) & (0.212)\ (0.219) & (0.101)
\end{bmatrix}
\begin{bmatrix}
X_{1t}^{(1)} \\
X_{2t}^{(1)} \\
X_{3t}^{(1)} \\
X_{4t}^{(1)}
\end{bmatrix}
$$

$$
=
\begin{bmatrix}
0.256 \\
(0.554) \\
1.275 \\
(0.272) \\
-0.357 \\
(0.221) \\
3.339 \\
(0.476)
\end{bmatrix}
+
\begin{bmatrix}
a_{1t}^{(1)} \\
a_{2t}^{(1)} \\
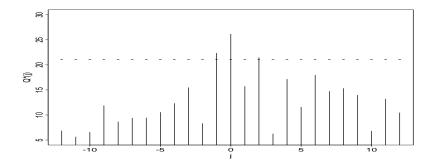a_{3t}^{(1)} \\
a_{4t}^{(1)}
\end{bmatrix}
$$

(b) American series

$$
\begin{bmatrix}
1 & 0 & .618\,B^2 - .371\,B^{10} \\
& & (0.093)\ (0.084) \\
0 & 1 - .554\,B & -.511\,B + .335\,B^2 \\
& (0.105) & (0.163)\ (0.154) \\
-.203\,B^2 - .125\,B^{10} & -.230\,B^2 & 1 - .910\,B - .482\,B^2 + .359\,B^3 \\
(0.054)\ (0.044) & (0.062) & (0.092)\ (0.131)\ (0.103)
\end{bmatrix}
\begin{bmatrix}
X_{1t}^{(1)} \\
X_{2t}^{(1)} \\
X_{3t}^{(1)}
\end{bmatrix}
$$

$$
=
\begin{bmatrix}
0.144 \\
(0.342) \\
0.627 \\
(0.254) \\
-.156 \\
(0.192)
\end{bmatrix}
+
\begin{bmatrix}
a_{1t}^{(1)} \\
a_{2t}^{(1)} \\
a_{3t}^{(1)}
\end{bmatrix}
$$

(c) Residual covariance matrices

$$
\hat{\Sigma}_{\mathbf{a}}^{(1)} =
\begin{bmatrix}
0.746 & -.327 & 0.033 & -.007 \\
-.327 & 0.507 & 0.021 & 0.097 \\
0.033 & 0.021 & 0.108 & -.006 \\
-.007 & 0.097 & -.006 & 0.447
\end{bmatrix}
, \quad
\hat{\Sigma}_{\mathbf{a}}^{(2)} =
\begin{bmatrix}
0.589 & -.025 & 0.097 \\
-.025 & 0.406 & 0.017 \\
0.097 & 0.017 & 0.149
\end{bmatrix}.
$$

Figure 1: Values of the statistic $Q_{\hat{a}}(j)^*$ defined by (6.3) at different lags j. The horizontal dotted line represent the marginal critical value at the significance level $\alpha = 0.05$.

7   Conclusion

Following the idea of Hong (1996a), we have introduced a new semi-parametric approach to test the non-correlation (or independence in the Gaussian case) between two multivariate stationary VAR(∞) series. The approach is semi-parametric in the sens that if the two series are VARMA, we do not need to separately estimate the true model for each of the series. We rather fit a vector autoregression to each series and the test statistic is based on residual cross-correlations at all possible lags. The weights assigned to the lags are determined by a kernel function and a truncation parameter. With univariate series, we retrieve Hong's (1996a) test. Under the hypothesis of independence of the two series, the asymptotic normality of the test statistic is established. For a general class of fixed alternatives of cross-correlation between the two series, the consistency of the test is also derived. The finite sample properties of the test were investigated by a Monte Carlo experiment. It is seen that the level is reasonably well controlled with short series of 100 observations. Furthermore, with the model considered, the four kernels DAN, PAR, BAR, BP lead to similar powers and are more powerful than the truncated uniform kernel which corresponds to the multivariate version of Haugh's portmanteau test.

Table 6: Values of the global statistic P_M^* defined by (5.1) and its empirical significance level for $M = 1, ..12$.

M	P_M^*	α_M	M	P_M^*	α_M
1	64.254	0.003	7	218.771	0.026
2	94.028	0.003	8	242.721	0.033
3	115.723	0.012	9	268.552	0.034
4	145.184	0.010	10	281.917	0.095
5	167.284	0.021	11	300.734	0.147
6	194.675	0.019	12	318.024	0.227

Table 7: Value of $S_N(k)$, $MS(k)$, $D_N(k)$ and $MD(k)$ with the truncated uniform kernel for variaous values of M when $N = 68$.

M	$S_N(k)$	$MS(k)$	$D_N(k)$	$MD(k)$
4	8.7	8	8.29	8
7	14.2	14	13.2	14
11	21.1	22	19.0	22

Table 8: Values of the statistics $\mathcal{Q}_N$ and $\mathcal{Q}_N^*$ and their p-values for three kernels and three values of M.

Kernels	$M = 4$			$M = 7$			$M = 11$		
	TR	DAN	BP	TR	DAN	BP	TR	DAN	BP
$\mathcal{Q}_N$	2.922	2.191	2.672	2.640	2.779	2.964	2.189	2.969	2.838
α	0.002	0.014	0.004	0.004	0.003	0.002	0.014	0.002	0.002
$\mathcal{Q}_{N}*$	3.586	1.931	2.474	2.680	2.560	2.537	1.544	2.564	2.054
α	0.0001	0.027	0.007	0.004	0.005	0.006	0.061	0.005	0.002

Appendix

The following notations are adopted. The Euclidian scalar product of $\mathbf{x}_t$ and $\mathbf{x}_s$ is defined by $\langle \mathbf{x}_t, \mathbf{x}_s \rangle = \mathbf{x}_t^T \mathbf{x}_s$ and the Euclidean norm of $\mathbf{x}_t$ by $\|\mathbf{x}_t\| = \sqrt{\langle \mathbf{x}_t, \mathbf{x}_t \rangle}$. The scalar Δ denotes a generic positive bounded constant that may differ from place to place. Important parts of the proofs presented in this appendix are adaptations of those presented in Hong (1996b) for univariate case. However, the multivariate context involves a lot of matrix calculus.

PROOF OF LEMMA 3.1.

Consider the following linear transformation $\boldsymbol{b}_t = \boldsymbol{\Sigma}^{-1/2} \boldsymbol{a}_t$. The process $\boldsymbol{b} = \{\boldsymbol{b}_t, \ t \in \mathbb{Z}\}$ is a white noise process with mean $\mathbf{0}$ and variance $\mathbb{I}_m$. Since $\boldsymbol{C}_{\boldsymbol{b}}^{(12)}(j) = \boldsymbol{\Sigma}_1^{-1/2} \boldsymbol{C}_{\boldsymbol{a}}^{(12)}(j) \boldsymbol{\Sigma}_2^{-1/2}$, and using the property $vec(ABC) = (C^T \otimes A)vec(B)$, we have that

$$
\begin{aligned}
\mathcal{T}(\boldsymbol{a}, \boldsymbol{\Sigma}) &= N \sum_{j=1-N}^{N-1} k^2(j/M) \mathbf{c}_{\boldsymbol{a}}^{(12)}(j)^T (\boldsymbol{\Sigma}_2^{-1} \otimes \boldsymbol{\Sigma}_1^{-1}) \mathbf{c}_{\boldsymbol{a}}^{(12)}(j) \\
&= N \sum_{j=1-N}^{N-1} k^2(j/M) \mathbf{c}_{\boldsymbol{b}}^{(12)}(j)^T \mathbf{c}_{\boldsymbol{b}}^{(12)}(j) \\
&= \mathcal{T}_{\boldsymbol{b}}^{(12)},
\end{aligned}
$$

(say). We can decompose the last quantity in two parts. First, using definition (2.6), and since $\boldsymbol{C}_{\boldsymbol{b}}^{(12)}(-j) = \boldsymbol{C}_{\boldsymbol{b}}^{(21)}(j)^T$, for $j > 0$, we have

$$
\text{tr}[\boldsymbol{C}_{\boldsymbol{b}}^{(12)}(-j)^T \boldsymbol{C}_{\boldsymbol{b}}^{(12)}(-j)] = N^{-2} \Big[\sum_{t=j+1}^{N} \|\boldsymbol{b}_{t-j}^{(1)}\|^2 \|\boldsymbol{b}_t^{(2)}\|^2 + 2 \sum_{t=j+2}^{N} \sum_{s=j+1}^{t-1} \boldsymbol{\pi}_{jts}^{(21)} \Big],
$$

and for $j \geq 0$, we also have

$$
\text{tr}[\boldsymbol{C}_{\boldsymbol{b}}^{(12)}(j)^T \boldsymbol{C}_{\boldsymbol{b}}^{(12)}(j)] = N^{-2} \Big[\sum_{t=j+1}^{N} \|\boldsymbol{b}_t^{(1)}\|^2 \|\boldsymbol{b}_{t-j}^{(2)}\|^2 + 2 \sum_{t=j+2}^{N} \sum_{s=j+1}^{t-1} \boldsymbol{\pi}_{jts}^{(12)} \Big],
$$

where $\boldsymbol{\pi}_{jts}^{(12)} = \langle \boldsymbol{b}_t^{(1)}, \boldsymbol{b}_s^{(1)} \rangle \langle \boldsymbol{b}_{t-j}^{(2)}, \boldsymbol{b}_{s-j}^{(2)} \rangle$ and $\boldsymbol{\pi}_{jts}^{(21)} = \langle \boldsymbol{b}_t^{(2)}, \boldsymbol{b}_s^{(2)} \rangle \langle \boldsymbol{b}_{t-j}^{(1)}, \boldsymbol{b}_{s-j}^{(1)} \rangle$. If

$$
H_{1N} = N^{-1} \sum_{j=0}^{N-1} k^2(j/M) \sum_{t=j+1}^{N} \|\boldsymbol{b}_t^{(1)}\|^2 \|\boldsymbol{b}_{t-j}^{(2)}\|^2, \quad H_{2N} = N^{-1} \sum_{j=1}^{N-1} k^2(j/M) \sum_{t=j+1}^{N} \|\boldsymbol{b}_{t-j}^{(1)}\|^2 \|\boldsymbol{b}_t^{(2)}\|^2
$$

$$
W_{1N}^* = 2N^{-1} \sum_{j=0}^{N-2} k^2(j/M) \sum_{t=j+2}^{N} \sum_{s=j+1}^{t-1} \boldsymbol{\pi}_{jts}^{(12)}, \quad W_{2N}^* = 2N^{-1} \sum_{j=1}^{N-2} k^2(j/M) \sum_{t=j+2}^{N} \sum_{s=j+1}^{t-1} \boldsymbol{\pi}_{jts}^{(21)}.
$$

24

Thus, we can write $\mathcal{T}_{\boldsymbol{b}}^{(12)}$ as

$$
\begin{aligned}
\mathcal{T}_{\boldsymbol{b}}^{(12)} &= N \sum_{j=1}^{N-1} k^2(j/M)\mathrm{tr}[\boldsymbol{C}_{\boldsymbol{b}}^{(12)}(-j)^T \boldsymbol{C}_{\boldsymbol{b}}^{(12)}(-j)] + N \sum_{j=0}^{N-1} k^2(j/M)\mathrm{tr}[\boldsymbol{C}_{\boldsymbol{b}}^{(12)}(j)^T \boldsymbol{C}_{\boldsymbol{b}}^{(12)}(j)] \\
&= H_N + W_N^*,
\end{aligned}
$$

with $H_N = H_{1N} + H_{2N}$ and $W_N^* = W_{1N}^* + W_{2N}^*$. The proof of Lemma 3.1 can be completed by proving the following two lemmas.

Lemma A.1 $\sigma^{-1}(N)\{H_N - m_1 m_2 S_N(k)\} \overset{p}{\to} 0$ where $\sigma^2(N) = 2m_1 m_2 D_N(k)$.

Lemma A.2 $\sigma^{-1}(N)W_N^* \overset{L}{\to} N(0, 1)$.

PROOF OF LEMMA A.1.

First, we note that $\mathbb{E}(H_N) = m_1 m_2 S_N(k)$ since

$$
\mathbb{E}(H_N) = N^{-1}\mathbb{E}\{\sum_{j=0}^{N-1} k^2(j/M) \sum_{t=j+1}^{N} \|\boldsymbol{b}_t^{(1)}\|^2 \|\boldsymbol{b}_{t-j}^{(2)}\|^2 + \sum_{j=1}^{N-1} k^2(j/M) \sum_{t=j+1}^{N} \|\boldsymbol{b}_{t-j}^{(1)}\|^2 \|\boldsymbol{b}_t^{(2)}\|^2\}.
$$

By symmetry of the kernel k, the independence of the two processes and since $\mathbb{E}(\|\boldsymbol{b}_t^{(h)}\|^2) = m_h$, for $h = 1, 2$, we have

$$
\begin{aligned}
\mathbb{E}(H_N) &= \frac{m_1 m_2}{N}\left(\sum_{j=0}^{N-1} \sum_{t=j+1}^{N} k^2(j/M) + \sum_{j=1}^{N-1} \sum_{t=j+1}^{N} k^2(j/M)\right) \\
&= m_1 m_2 \sum_{j=1-N}^{N-1} (1 - \frac{|j|}{N})k^2(j/M) = m_1 m_2 S_N(k).
\end{aligned}
$$

Also, using Minkowski inequality, we obtain

$$
\begin{aligned}
\mathbb{E}(H_{1N} - \mathbb{E}H_{1N})^2 &= \mathbb{E}[N^{-1} \sum_{j=0}^{N-1} k^2(j/M) \sum_{t=j+1}^{N} (\|\boldsymbol{b}_t^{(1)}\|^2 \|\boldsymbol{b}_{t-j}^{(2)}\|^2 - m_1 m_2)]^2 \\
&\leq \frac{\Delta M^2}{N}\{M^{-1} \sum_{j=0}^{N-1} k^2(j/M)\}^2.
\end{aligned}
$$

Given assumption B and since
$M \to \infty$ as $N \to \infty$, we have $M^{-1} \sum_{j=0}^{N-1} k^2(j/M) \to \int_0^\infty k^2(z)dz < \infty$ and thus $\mathbb{E}(H_{1N} - \mathbb{E}H_{1N})^2 = O(M^2/N)$. By symmetry, we have $\mathbb{E}(H_{2N} - \mathbb{E}H_{2N})^2 = O(M^2/N)$. Using Minkowski inequality once more, it follows that $\mathbb{E}(H_N - \mathbb{E}H_N)^2 = O(M^2/N)$. Since $M^{-1}D_N(k) \to D(k)$ as $N \to \infty$ and $M/N \to 0$, we have $\sigma^2(N) = 2m_1 m_2 MD(k)\{1 + o(1)\} = O(M)$ and it follows that $\sigma^{-1}(N)\{H_N - m_1 m_2 S_N(k)\} \overset{p}{\to} 0$.

PROOF OF LEMMA A.2.

First, we denote by $W_{1Nt} = 2\sum_{s=2}^{t-1}\sum_{j=0}^{s-1} k^2(j/M)\boldsymbol{\pi}_{jts}^{(12)}$, $W_{2Nt} = 2\sum_{s=2}^{t-1}\sum_{j=1}^{s-1} k^2(j/M)\boldsymbol{\pi}_{jts}^{(21)}$ and $\omega_N = 2N^{-1}\sum_{t=2}^{N}\boldsymbol{\pi}_{0t1}^{(12)}$. Using the properties $\sum_{j=0}^{N-2}\sum_{t=j+2}^{N}\sum_{s=j+1}^{t-1} = \sum_{t=2}^{N}\sum_{s=1}^{t-1}\sum_{j=0}^{s-1}$ and $\sum_{j=1}^{N-2}\sum_{t=j+2}^{N}\sum_{s=j+1}^{t-1} = \sum_{t=3}^{N}\sum_{s=1}^{t-1}\sum_{j=1}^{s-1}$, it can be shown that

$$W_N^* = N^{-1}\sum_{t=3}^{N}(W_{1Nt} + W_{2Nt}) + \omega_N = N^{-1}\sum_{t=3}^{N}W_{Nt} + \omega_N.$$

By the hypothesis of independence of the two processes, note that $\omega_N = o_p(1)$. We also have that $\sigma(N)^{-1} = O(M^{-1/2})$ and we obtain that $\sigma^{-1}(N)\omega_N \xrightarrow{p} 0$. Therefore, the asymptotic behaviour of W_N^* is determined by the one of $W_N = N^{-1}\sum_{t=3}^{N}W_{Nt}$. Its exact variance is the following.

Lemma A.3

$$Var(W_N) = \sigma^2(N) \text{ where } \sigma^2(N) = 2m_1 m_2 D_N(k).$$

PROOF.

From the independence of the two innovation processes, it follows that $\mathbb{E}(W_{1Nt}) = 2\sum_{s=2}^{t-1}\sum_{j=0}^{s-1}$ $k^2(j/M)\mathbb{E}(\langle\boldsymbol{b}_t^{(1)},\boldsymbol{b}_s^{(1)}\rangle)\mathbb{E}(\langle\boldsymbol{b}_{t-j}^{(2)},\boldsymbol{b}_{s-j}^{(2)}\rangle) = 0$ since for $t \neq s$, $\mathbb{E}(\langle\boldsymbol{b}_t^{(1)},\boldsymbol{b}_s^{(1)}\rangle) = 0$. Also, we have $\mathbb{E}(W_{1Nt}W_{2Nt}) = 0$ since $\mathbb{E}(\pi_{jts}^{(12)}\pi_{j_1ts_1}^{(21)}) = 0$ and

$$Var(W_{Nt}) = \mathbb{E}(W_{Nt}^2) = \mathbb{E}(W_{1Nt}^2 + W_{2Nt}^2).$$

By developing W_{1Nt}^2, using the independence of the two innovation processes and the properties $\mathbb{E}(\langle\boldsymbol{b}_t^{(1)},\boldsymbol{b}_s^{(1)}\rangle)^2 = m_1$ and $\mathbb{E}(\langle\boldsymbol{b}_{t-j}^{(2)},\boldsymbol{b}_{s-j}^{(2)}\rangle)^2 = m_2$, we have

$$\mathbb{E}(W_{1Nt}^2) = 4\sum_{s=2}^{t-1}\sum_{j=0}^{s-1} k^4(j/M)\mathbb{E}\langle\boldsymbol{b}_t^{(1)},\boldsymbol{b}_s^{(1)}\rangle^2\mathbb{E}\langle\boldsymbol{b}_{t-j}^{(2)},\boldsymbol{b}_{s-j}^{(2)}\rangle^2 = 4m_1 m_2\sum_{s=2}^{t-1}\sum_{j=0}^{s-1} k^4(j/M).$$

Similarly, $\mathbb{E}(W_{2Nt}^2) = 4m_1 m_2\sum_{s=2}^{t-1}\sum_{j=1}^{s-1} k^4(j/M)$, and we obtain $Var(W_{Nt}) = 4m_1 m_2\sum_{s=2}^{t-1}\sum_{|j|=0}^{s-1}$ $k^4(j/M)$. Finally, using the fact that $\mathbb{E}(W_{Nt}) = 0$ and for $t \neq t_1$, $\mathbb{E}(W_{Nt}W_{Nt_1}) = 0$, it follows that

$$Var(W_N) = \frac{1}{N^2}\sum_{t=3}^{N}\mathbb{E}(W_{Nt}^2) = \frac{4m_1 m_2}{N^2}\sum_{|j|=0}^{N-2}\sum_{t=|j|+2}^{N}\sum_{s=2}^{t} k^4(j/M)$$

$$= 2m_1 m_2\sum_{|j|=0}^{N-2}\left(1 - \frac{|j|}{N}\right)\left(1 - \frac{|j|+1}{N}\right)k^4(j/M) = \sigma^2(N).$$

To continue the proof of Lemma A.2, we note that $\{(W_{Nt},\mathfrak{F}_{t-1}); t \in \mathbb{Z}\}$, where $\mathfrak{F}_{t-1}$ is the σ-algebra generated by $\{(\boldsymbol{b}_s^{(1)},\boldsymbol{b}_s^{(2)})^T, s \leq t\}$, is a martingale difference since $\mathbb{E}(W_{Nt}|\mathfrak{F}_{t-1}) = 0$.

26

As in Hong(1996a, 1996b), the asymptotic normality of W_N^* follows from the central limit theorem for a martingale difference derived in Brown (1971). To apply this later theorem, it is sufficient to verify the following two conditions stated in the next two lemmas.

Lemma A.4 $\sigma^{-2}(N)\frac{1}{N^2}\sum_{t=3}^N \mathbb{E}[W_{N,t}^2 \mathbb{I}\{|W_{N,t}| > \epsilon\sigma(N)\}] \to 0$, $\forall \epsilon > 0$.

Lemma A.5 $\sigma^{-2}(N)\frac{1}{N^2}\sum_{t=3}^N \ddot{W}_{Nt}^2 \xrightarrow{p} 1$, where $\ddot{W}_{Nt}^2 = \mathbb{E}\left(W_{Nt}^2|\mathfrak{F}_{t-1}\right)$.

PROOF OF LEMMA A.4.

To prove this lemma, it is sufficient to verify the Lyapounov condition $\sigma^{-4}(N)N^{-4}\sum_{t=3}^N \mathbb{E}\left(W_{Nt}^4\right) \to 0$. To do that, denote $G_{ts}^{(1)} = \sum_{j=1}^{s-1} k^2(j/M)\langle \boldsymbol{b}_{t-j}^{(1)}, \boldsymbol{b}_{s-j}^{(1)}\rangle$ and $G_{ts}^{(2)} = \sum_{j=0}^{s-1} k^2(j/M)\langle \boldsymbol{b}_{t-j}^{(2)}, \boldsymbol{b}_{s-j}^{(2)}\rangle$. Then, we can write $W_{1Nt} = 2\sum_{s=2}^{t-1}\langle \boldsymbol{b}_t^{(1)}, \boldsymbol{b}_s^{(1)}\rangle G_{ts}^{(2)}$ and $W_{2Nt} = 2\sum_{s=2}^{t-1}\langle \boldsymbol{b}_t^{(2)}, \boldsymbol{b}_s^{(2)}\rangle G_{ts}^{(1)}$. Given the assumption of independence of the two innovation processes, we have

$$
\begin{aligned}
\mathbb{E}(W_{1Nt}^4) &\leq 16\mathbb{E}[\sum_{s=2}^{t-1}\|\boldsymbol{b}_t^{(1)}\|\|\boldsymbol{b}_s^{(1)}\|G_{ts}^{(2)}]^4 \\
&\leq 48\mathbb{E}\|\boldsymbol{b}_t^{(1)}\|^4\{\sum_{s=2}^{t-1}[\mathbb{E}\|\boldsymbol{b}_s^{(1)}\|^4\mathbb{E}(G_{ts}^{(2)})^4]^{1/2}\}^2 \leq \Delta\{\sum_{s=2}^{t-1}[\mathbb{E}(G_{ts}^{(2)})^4]^{1/2}\}^2.
\end{aligned}
$$

The second inequality follows by applying the inequality $\mathbb{E}(\sum_{i=1}^n Y_i)^4 \leq 3\{\sum_{i=1}^n [\mathbb{E}(Y_i^4)]^{1/2}\}^{1/2}$ where the sequence of random variables $\{Y_i\}$ verify $\mathbb{E}(Y_i) = 0$ and $\mathbb{E}(Y_i f(Y_j, Y_k, Y_l)) = 0$ for $i \neq j, k, l$ and for any function f. Also, using the same inequality, and for $t > s$, we have

$$
\mathbb{E}(G_{ts}^{(2)})^4 \leq 3\{\sum_{j=1}^{s-1} k^4(j/M)[\mathbb{E}\|\boldsymbol{b}_{t-j}^{(2)}\|^4\mathbb{E}\|\boldsymbol{b}_{s-j}^{(2)}\|^4]^{1/2}\}^2 \leq \Delta M^2\{\frac{1}{M}\sum_{j=1}^{s-1} k^4(j/M)\}^2 = \mathcal{O}(M^2).
$$

Thus, we obtain that $\mathbb{E}(W_{1Nt}^4) \leq \Delta t^2 M^2 = \mathcal{O}(t^2 M^2)$. By symmetry, we also have $\mathbb{E}(W_{2Nt}^4) = \mathcal{O}(t^2 M^2)$. Since $(a+b)^4 \leq 8(a^4 + b^4)$, it follows that

$$
\sigma^{-4}(N)N^{-4}\sum_{t=3}^N \mathbb{E}\left(W_{Nt}^4\right) \leq \sigma^{-4}(N)\frac{8}{N^4}\sum_{t=3}^N \mathbb{E}\left(W_{1Nt}^4 + W_{2Nt}^4\right).
$$

Since $\sigma^{-4}(N) = O(M^{-2})$ and $\sum_{t=3}^N \mathbb{E}(W_{1Nt}^4 + W_{2Nt}^4) \leq \Delta\sum_{t=3}^N t^2 M^2 \leq \Delta N^3 M^2$, then

$$
\sigma^{-4}(N)N^{-4}\sum_{t=3}^N \mathbb{E}\left(W_{Nt}^4\right) = O(N^{-1})
$$

and the proof of Lemma A.4 is completed.

PROOF OF LEMMA A.5.

To prove the second condition, it is sufficient to show that $\sigma^{-4}(N)var(N^{-2}\sum_{t=3}^{N}\ddot{W}_{Nt}^2) \to 0$. By definition of $\ddot{W}_{Nt}^2$, we can write

$$
\begin{aligned}
\ddot{W}_{Nt}^2 &= \mathbb{E}[W_{1Nt}^2|\mathcal{F}_{t-1}] + \mathbb{E}[W_{2Nt}^2|\mathcal{F}_{t-1}] \\
&= 4tr\{\mathbb{E}[\sum_{s=2}^{t-1} G_{ts}^{(2)}\boldsymbol{b}_s^{(1)^T}\boldsymbol{b}_t^{(1)}\boldsymbol{b}_t^{(1)^T}\sum_{s_1=2}^{t-1}\boldsymbol{b}_{s_1}^{(1)}G_{ts_1}^{(2)}|\mathcal{F}_{t-1}]\} + 4tr\{\mathbb{E}[\sum_{s=2}^{t-1} G_{ts}^{(1)}\boldsymbol{b}_s^{(2)^T}\boldsymbol{b}_t^{(2)}\boldsymbol{b}_t^{(2)^T}\sum_{s_1=2}^{t-1}\boldsymbol{b}_{s_1}^{(2)}G_{ts_1}^{(1)}|\mathcal{F}_{t-1}]\}
\end{aligned}
$$

We denote $\lambda_{1Nt} = \sum_{s=2}^{t-1}G_{ts}^{(1)}\boldsymbol{b}_s^{(2)^T}$ and $\lambda_{2Nt} = \sum_{s=2}^{t-1}G_{ts}^{(2)}\boldsymbol{b}_s^{(1)^T}$. Since $tr(AB) \le tr(A)tr(B)$ and that the processes $\{\boldsymbol{b}_t^{(h)}\}$, $h=1,2$, are independent, we have

$$
\begin{aligned}
\ddot{W}_{Nt}^2 &\le 4tr\{\mathbb{E}[\lambda_{1Nt}\lambda_{1Nt}^T|\mathcal{F}_{t-1}]\}tr\{\mathbb{E}[\boldsymbol{b}_t^{(1)}\boldsymbol{b}_t^{(1)^T}|\mathcal{F}_{t-1}]\} + 4tr\{\mathbb{E}[\lambda_{2Nt}\lambda_{2Nt}^T|\mathcal{F}_{t-1}]\}tr\{\mathbb{E}[\boldsymbol{b}_t^{(2)}\boldsymbol{b}_t^{(2)^T}|\mathcal{F}_{t-1}]\} \\
&\le 4(\ddot{W}_{1Nt}^2 + \ddot{W}_{2Nt}^2),
\end{aligned}
$$

where $\ddot{W}_{1Nt}^2 = m_1\|\lambda_{1Nt}\|^2$ and $\ddot{W}_{2Nt}^2 = m_2\|\lambda_{2Nt}\|^2$. The second inequality follows since by conditioning on $\mathcal{F}_{t-1}$, the terms λ_{1Nt} and λ_{2Nt} become constant. Thus, to prove the lemma, it is sufficient to show that $M^{-2}var(N^{-2}\sum_{t=3}^{N}\ddot{W}_{lNt}^2)\to0$, for $l=1,2$. We consider the case $l=1$, and we write

$$
\begin{aligned}
\ddot{W}_{1Nt}^2 &= m_1 tr\{(\sum_{s=2}^{t-1}G_{ts}^{(2)}\boldsymbol{b}_s^{(1)})(\sum_{s=2}^{t-1}G_{ts}^{(2)}\boldsymbol{b}_s^{(1)^T})\} \\
&= m_1\sum_{s=2}^{t-1}\|\boldsymbol{b}_s^{(1)}\|^2(G_{ts}^{(2)})^2 + 2m_1\sum_{s_2=3}^{t-1}\sum_{s_1=2}^{s_2-1}G_{ts_1}^{(2)}G_{ts_2}^{(2)}\langle\boldsymbol{b}_{s_1}^{(1)},\boldsymbol{b}_{s_2}^{(1)}\rangle \\
&= \tilde{B}_{1Nt} + \tilde{A}_{1Nt}.
\end{aligned}
$$

Note that $\tilde{A}_{1Nt} = \sum_{s_2=3}^{t-1}\tilde{A}_{s_2t}^{(12)}$ where $\tilde{A}_{s_2t}^{(12)} = \sum_{s_1=2}^{s_2-1}G_{ts_1}^{(2)}G_{ts_2}^{(2)}\langle\boldsymbol{b}_{s_1}^{(1)},\boldsymbol{b}_{s_2}^{(1)}\rangle$ is a sum of a martingale difference sequence over s_2 since $\mathbb{E}[\tilde{A}_{s_2t}^{(12)}|\mathcal{F}_{\tau-1}^{(1)}] = 0$ where $\mathcal{F}_{\tau-1}^{(1)}$ is the σ-algebra generated by $\{\boldsymbol{b}_{s_2}^{(1)}, s_2 \le \tau\}$. Given the assumption of independence between $\{\boldsymbol{b}_{s_1}^{(1)}\}$ and $\{\boldsymbol{b}_{s_2}^{(2)}\}$, for $t_2 \ge t_1$, by developping the product $\tilde{A}_{1Nt_2}\tilde{A}_{1Nt_1}$, see Bouhaddioui (2002), and using $\mathbb{E}(\langle\boldsymbol{b}_{s_1}^{(1)},\boldsymbol{b}_{s_2}^{(1)}\rangle) = m_1$, we obtain that

$$
\mathbb{E}(\tilde{A}_{1Nt_2}\tilde{A}_{1Nt_1}) = 4m_1^3\sum_{s_2=3}^{t_1-1}\sum_{s_1=2}^{s_2-1}\mathbb{E}(G_{t_2s_2}^{(2)}G_{t_2s_1}^{(2)}G_{t_1s_2}^{(2)}G_{t_1s_1}^{(2)}).
$$

Using the definition of $G_{ts}^{(h)}$, $h=1,2$, and by straightforward but tedious calculus, we get for $t_2 > s_2$, $t_1 > s_1$,

$$
\left|\mathbb{E}(G_{t_2s_2}^{(2)}G_{t_2s_1}^{(2)}G_{t_1s_2}^{(2)}G_{t_1s_1}^{(2)})\right| \le \begin{cases} \Delta M^2\left\{\frac{1}{M}\sum_{j=0}^{N-1}k^2(j/M)\right\}^2 & \text{if } t_2 = t_1, \\ \Delta M\left\{\frac{1}{M}\sum_{j=0}^{N-1}k^2(j/M)\right\} & \text{if } t_2 > t_1, \end{cases}
$$

28

and it follows that $\mathbb{E}(\tilde{A}_{1Nt_2}\tilde{A}_{1Nt_1}) \le \Delta M t_1^2$.

Also, we have $\frac{1}{M^2}\mathbb{E}(\frac{1}{N^2}\sum_{t=3}^{N}\tilde{A}_{1Nt})^2 = \frac{1}{M^2N^4}\sum_{t=3}^{N}\mathbb{E}(\tilde{A}_{1Nt}^2) + \frac{2}{M^2N^4}\sum_{t_2=4}^{N}\sum_{t_1=3}^{t_2-1}\mathbb{E}(\tilde{A}_{1Nt_2}\tilde{A}_{1Nt_1})$.
Since $\mathbb{E}(\tilde{A}_{1Nt}^2) \le M^2\sum_{s_2=3}^{t-1}\sum_{s_1=2}^{s_2-1}\Delta \le \Delta M^2 t^2$, we obtain that

$$\frac{1}{M^2}\mathbb{E}(\frac{1}{N^2}\sum_{t=3}^{N}\tilde{A}_{1Nt})^2 = \mathcal{O}(N^{-1} + M^{-1}).$$

Now, the first term $\tilde{B}_{1Nt}$ can be decomposed in two parts.

$$\tilde{B}_{1Nt} = m_1^2\sum_{s=2}^{t-1}(G_{ts}^{(2)})^2 + m_1\sum_{s=2}^{t-1}(\|\boldsymbol{b}_s^{(1)}\|^2 - m_1)(G_{ts}^{(2)})^2 = \tilde{B}_{2Nt} + \tilde{A}_{2Nt}.$$

By conditioning on $(\boldsymbol{b}_t^{(2)})_{t=1}^{N}$, $\tilde{A}_{2Nt}$ is a weighted sum of independent and identically distributed random variables. Using the fact that $\mathbb{E}(G_{ts}^{(2)})^4 \le \Delta M^2\{\frac{1}{M}\sum_{j=0}^{N-1}k^2(j/M)\}^2$, it follows that

$$\mathbb{E}\tilde{A}_{2Nt}^2 = \sum_{s=2}^{t-1}\mathbb{E}(\|\boldsymbol{b}_s^{(1)}\|^2 - m_1)^2\mathbb{E}(G_{ts}^{(2)})^4 \le \Delta t M^2\{\frac{1}{M}\sum_{j=0}^{N-1}k^2(j/M)\}^2,$$

Thus, using Minkowski inequality, we have

$$\frac{1}{M^2}\mathbb{E}(\frac{1}{N^2}\sum_{t=3}^{N}\tilde{A}_{2Nt})^2 \le \frac{1}{M^2N^4}\{\sum_{t=3}^{N}(\mathbb{E}\tilde{A}_{2Nt}^2)^{1/2}\}^2 = \mathcal{O}(N^{-1}).$$

The term $\tilde{B}_{2Nt}$ can be decomposed it in two parts $\tilde{B}_{2Nt} = \tilde{B}_{3Nt} + \tilde{A}_{3Nt}$ where

$$\tilde{B}_{3Nt} = m_1^2\sum_{s=2}^{t-1}\sum_{j=0}^{s-1}k^4(j/M)\boldsymbol{b}_{t-j}^{(2)T}\boldsymbol{b}_{s-j}^{(2)}\boldsymbol{b}_{s-j}^{(2)T}\boldsymbol{b}_{t-j}^{(2)},$$

$$\tilde{A}_{3Nt} = \sum_{j_1=0}^{t-3}\sum_{j_2=j_1+1}^{t-2}\sum_{s=0}^{t-1}k^2(j_2/M)k^2(j_1/M)\boldsymbol{b}_{t-j_2}^{(2)T}\boldsymbol{b}_{s-j_2}^{(2)}\boldsymbol{b}_{s-j_1}^{(2)T}\boldsymbol{b}_{t-j_1}^{(2)}.$$

Using the fact that $tr(\boldsymbol{b}_{t-j_2}^{(2)T}\boldsymbol{b}_{s-j_2}^{(2)}\boldsymbol{b}_{s-j_1}^{(2)T}\boldsymbol{b}_{t-j_1}^{(2)}) = tr(\boldsymbol{b}_{t-j_1}^{(2)}\boldsymbol{b}_{t-j_2}^{(2)T}\boldsymbol{b}_{s-j_2}^{(2)}\boldsymbol{b}_{s-j_1}^{(2)T})$, we can write

$$\tilde{A}_{3Nt} = tr(\sum_{j_1=0}^{t-3}k^2(j_1/M)\boldsymbol{b}_{t-j_1}^{(2)}\sum_{j_2=j_1+1}^{t-2}k^2(j_2/M)\boldsymbol{b}_{t-j_2}^{(2)T}\sum_{s=0}^{t-1}\boldsymbol{b}_{s-j_2}^{(2)}\boldsymbol{b}_{s-j_1}^{(2)T}).$$

Similarly, $\tilde{A}_{3Nt}$ is a sum over j_1 of a martingale difference. By using the inequality $tr(AB) \le tr(A)tr(B)$, Minkowski inequality and the fact that $\mathbb{E}(\sum_{s=j_2+1}^{t-1}\|\boldsymbol{b}_{s-j_1}^{(2)}\|\|\boldsymbol{b}_{s-j_2}^{(2)}\|)^2 \le \Delta t$ for $t > s > j_2 > j_1$, it follows that

$$\mathbb{E}\tilde{A}_{3Nt}^2 \le 4m_1^4m_2^2\sum_{j_1=1}^{t-3}k^4(j_1/M)[\sum_{j_2=j_1+1}^{t-2}k^2(j_2/M)\{\mathbb{E}(\sum_{s=j_2+1}^{t-1}\|\boldsymbol{b}_{s-j_1}^{(2)}\|\|\boldsymbol{b}_{s-j_2}^{(2)}\|)^2\}^{1/2}]^2$$

$$\le \Delta t M^3\{\frac{1}{M}\sum_{j_1=1}^{N-1}k^4(j_1/M)\}\{\frac{1}{M}\sum_{j_2=1}^{N-1}k^2(j_2/M)\}^2 = \mathcal{O}(tM^3).$$

Thus, we can write

$$\frac{1}{M^2}\mathbb{E}(\frac{1}{N^2}\sum_{t=3}^{N}\tilde{A}_{3Nt})^2 \leq \frac{1}{M^2N^4}\{\sum_{t=3}^{N}\mathbb{E}^{1/2}(\tilde{A}_{3Nt}^2)\}^2 = \mathcal{O}(M/N).$$

Finally, for the term $\tilde{B}_{3Nt}$, we have

$$\begin{aligned}
\tilde{B}_{3Nt} &= m_1^2 m_2^2 \sum_{s=2}^{t-1}\sum_{j=0}^{s-1}k^4(j/M) + m_1^2\sum_{s=2}^{t-1}\sum_{j=0}^{s-1}k^4(j/M)(\boldsymbol{b}_{t-j}^{(2)^T}\boldsymbol{b}_{s-j}^{(2)}\boldsymbol{b}_{s-j}^{(2)^T}\boldsymbol{b}_{t-j}^{(2)} - m_2^2) \\
&= \frac{m_1 m_2}{4}\mathbb{E}(W_{1Nt}^2) + \tilde{A}_{4Nt}.
\end{aligned}$$

By an argument similar to the one employed for $\tilde{A}_{3Nt}$, we have that $\mathbb{E}(\tilde{A}_{4Nt})^2 \leq \Delta t^2 M$, and it follows that $M^{-2}N^{-4}\mathbb{E}(\sum_{t=3}^{N}\tilde{A}_{4Nt})^2 \leq M^{-2}N^{-4}\{\sum_{t=3}^{N}\mathbb{E}^{1/2}(\tilde{A}_{4Nt}^2)\}^2 = \mathcal{O}(M^{-1})$. Thus, $\ddot{W}_{1Nt}^2$ can be written as

$$\ddot{W}_{1Nt}^2 = \frac{m_1 m_2}{4}\mathbb{E}(W_{1Nt}^2) + \sum_{l=1}^{4}\tilde{A}_{lNt},$$

and by combining all above results, we obtain

$$\frac{1}{M^2}var(\frac{1}{N^2}\sum_{t=3}^{N}\ddot{W}_{1Nt}^2) \leq \frac{\Delta}{M^2}\sum_{l=1}^{4}\mathbb{E}(\frac{1}{N^2}\sum_{t=3}^{N}\tilde{A}_{lNt})^2 = \mathcal{O}(M/N + M^{-1}).$$

A similar result for $\ddot{W}_{2Nt}^2$ can be proven by symmetry. The result of Lemma A.5 follows if $M/N \to 0$ and $M \to \infty$, as $N \to \infty$.

PROOF OF LEMMA 3.2.

Considering the same linear transformation as the one in the proof of Lemma 3.1, that is $\hat{\boldsymbol{b}}_t = \boldsymbol{\Sigma}^{-1/2}\hat{\boldsymbol{a}}_t$ and noting that $\boldsymbol{C}_{\hat{\boldsymbol{b}}}^{(12)}(j) = \boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{C}_{\hat{\boldsymbol{a}}}^{(12)}(j)\boldsymbol{\Sigma}_2^{-1/2}$, we can write

$$\begin{aligned}
\mathcal{T}(\hat{\boldsymbol{a}}, \boldsymbol{\Sigma}) &= N\sum_{j=1-N}^{N-1}k^2(j/M)\mathbf{c}_{\hat{\boldsymbol{a}}}^{(12)}(j)^T(\boldsymbol{\Sigma}_2^{-1}\otimes\boldsymbol{\Sigma}_1^{-1})\mathbf{c}_{\hat{\boldsymbol{a}}}^{(12)}(j) = N\sum_{j=1-N}^{N-1}k^2(j/M)\mathbf{c}_{\hat{\boldsymbol{b}}}^{(12)}(j)^T\mathbf{c}_{\hat{\boldsymbol{b}}}^{(12)}(j) \\
&= \mathcal{T}_{\hat{\boldsymbol{b}}}^{(12)}.
\end{aligned}$$

Thus, to prove the result, it is sufficient to show that $\mathcal{T}_{\boldsymbol{b}}^{(12)} - \mathcal{T}_{\hat{\boldsymbol{b}}}^{(12)} = o_p(M^{1/2})$. The result follows by decomposing the latter difference in two parts, that is,

$$\mathcal{T}_{\boldsymbol{b}}^{(12)} - \mathcal{T}_{\hat{\boldsymbol{b}}}^{(12)} = N\sum_{j=1-N}^{N-1}k^2(j/M)(\|\mathbf{c}_{\hat{\boldsymbol{b}}}^{(12)}(j) - \mathbf{c}_{\boldsymbol{b}}^{(12)}(j)\|^2 + 2\langle\mathbf{c}_{\boldsymbol{b}}^{(12)}(j), \mathbf{c}_{\hat{\boldsymbol{b}}}^{(12)}(j) - \mathbf{c}_{\boldsymbol{b}}^{(12)}(j)\rangle) = T_N^{(1)} + T_N^{(2)},$$

and by showing that each part is $o_p(M^{1/2})$. To prove that, we only consider the positive lags $j \geq 0$, since for negative lags, the proof is similar by symmetry.

Define $\hat{\boldsymbol{\delta}}_t = \boldsymbol{b}_t^{(1)} - \hat{\boldsymbol{b}}_t^{(1)}$ and $\hat{\boldsymbol{\eta}}_t = \boldsymbol{b}_t^{(2)} - \hat{\boldsymbol{b}}_t^{(2)}$. From (2.10), we have

$$T_N^{(1)} = N \sum_{j=0}^{N-1} k^2(j/M) \| \mathbf{c}_{\hat{\boldsymbol{b}}}^{(12)}(j) - \mathbf{c}_{\boldsymbol{b}}^{(12)}(j) \|^2 = N \sum_{j=0}^{N-1} k^2(j/M) \| \frac{1}{N} \sum_{t=j+1}^{N} (\boldsymbol{b}_t^{(1)} \boldsymbol{b}_{t-j}^{(2)T} - \hat{\boldsymbol{b}}_t^{(1)} \hat{\boldsymbol{b}}_{t-j}^{(2)T}) \|^2,$$

and using Cauchy-Schwarz inequality, we obtain

$$T_N^{(1)} = N \sum_{j=0}^{N-1} k^2(j/M) \| \frac{1}{N} \sum_{t=j+1}^{N} (\boldsymbol{b}_t^{(1)} \hat{\boldsymbol{\eta}}_{t-j}^{T} + \hat{\boldsymbol{\delta}}_t \boldsymbol{b}_{t-j}^{(2)T} - \hat{\boldsymbol{\delta}}_t \hat{\boldsymbol{\eta}}_{t-j}^{T}) \|^2 \leq 4N(T_{1N} + T_{2N} + T_{3N}),$$

with $T_{1N} = \sum_{j=0}^{N-1} k^2(j/M) \| \frac{1}{N} \sum_{t=j+1}^{N} \boldsymbol{b}_t^{(1)} \hat{\boldsymbol{\eta}}_{t-j}^{T} \|^2$, $T_{2N} = \sum_{j=0}^{N-1} k^2(j/M) \| \frac{1}{N} \sum_{t=j+1}^{N} \hat{\boldsymbol{\delta}}_t \boldsymbol{b}_{t-j}^{(2)T} \|^2$ and $T_{3N} = \sum_{j=0}^{N-1} k^2(j/M) \| \frac{1}{N} \sum_{t=j+1}^{N} \hat{\boldsymbol{\delta}}_t \hat{\boldsymbol{\eta}}_{t-j}^{T} \|^2$. Now, it suffices to show that the terms T_{jN}, $j = 1, 2, 3$, are $o_p(M^{1/2}/N)$. The techniques used in this part of the proof are similar to those using in the proof of Theorem 3.1 in Bouhaddioui and Roy (2003). From (2.9), we can write that

$$\hat{\boldsymbol{\eta}}_t = \{\boldsymbol{\Phi}(p_2) - \hat{\boldsymbol{\Phi}}(p_2)\} \mathbf{X}_t^{(2)}(p_2) + \boldsymbol{\xi}_t(p_2),$$

where $\boldsymbol{\xi}_t(p_2) = \sum_{l=p_2+1}^{\infty} \boldsymbol{\Phi}_l^{(2)} \mathbf{X}_{t-l}^{(2)}$ represents the bias of the VAR(p_2) approximation of $\{\mathbf{X}_t^{(2)}\}$. By equation (3.15) in Bouhaddioui and Roy (2003), we have $\mathbb{E}\left(\|\boldsymbol{\xi}_t(p_h)\|^2\right) = O\left(\sum_{l=p_h+1}^{\infty} \|\boldsymbol{\Phi}_l^{(h)}\|\right)^2$, $h = 1, 2$. Also, by result (3.17) in the same paper and equation (2.5), we obtain that

$$T_{1N} = \sum_{j=0}^{N-1} k^2(j/M) \| \frac{1}{N} \sum_{t=j+1}^{N} \hat{\boldsymbol{\eta}}_t \boldsymbol{b}_{t-j}^{(2)T} \|^2 = O_p(\frac{p_1^2 M}{N^2}) \{ \frac{1}{M} \sum_{j=0}^{N-1} k^2(j/M) \}$$

Since $p_1^2 = o(\frac{N}{M^{1/2}})$, we have $T_{1N} = o_p(\frac{M^{1/2}}{N})$. By symmetry, we can prove that $T_{2N} = o_p(\frac{M^{1/2}}{N})$.

For the third term T_{3N}, using the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
T_{3N} &= \sum_{j=0}^{N-1} k^2(j/M) \| N^{-1} \sum_{t=j+1}^{N} \hat{\boldsymbol{\eta}}_t \hat{\boldsymbol{\delta}}_{t-j}^{T} \|^2 \\
&\leq \| \boldsymbol{\Phi}(p_1) - \hat{\boldsymbol{\Phi}}(p_1) \|^2 \| \boldsymbol{\Phi}(p_2) - \hat{\boldsymbol{\Phi}}(p_2) \|^2 \sum_{j=0}^{N-1} k^2(j/M) \| N^{-1} \sum_{t=j+1}^{N} \mathbf{X}_t^{(1)}(p_1) \mathbf{X}_{t-j}^{(2)}(p_2)^{T} \|^2 \\
&+ \| \boldsymbol{\Phi}(p_1) - \hat{\boldsymbol{\Phi}}(p_1) \|^2 \sum_{j=0}^{N-1} k^2(j/M) \| N^{-1} \sum_{t=j+1}^{N} \mathbf{X}_t^{(1)}(p_1) \boldsymbol{\xi}_{t-j}(p_2)^{T} \|^2 \\
&+ \| \boldsymbol{\Phi}(p_2) - \hat{\boldsymbol{\Phi}}(p_2) \|^2 \sum_{j=0}^{N-1} k^2(j/M) \| N^{-1} \sum_{t=j+1}^{N} \boldsymbol{\xi}_t(p_1) \mathbf{X}_{t-j}^{(2)}(p_2)^{T} \|^2 \\
&+ \sum_{j=0}^{N-1} k^2(j/M) \| N^{-1} \sum_{t=j+1}^{N} \boldsymbol{\xi}_t(p_1) \boldsymbol{\xi}_{t-j}(p_2)^{T} \|^2
\end{aligned}$$

Using the equations (3.19) - (3.22) in Bouhaddioui and Roy (2003), assumptions (i) and (ii) of Theorem 3.1 and result (2.5), we can conclude that $T_{3N} = o_p(M^{1/2}/N)$. Therefore, we obtain

$$T_N^{(1)} = N \sum_{j=0}^{N-1} k^2(j/M) \| \mathbf{c}_{\hat{\boldsymbol{b}}}^{(12)}(j) - \mathbf{c}_{\boldsymbol{b}}^{(12)}(j) \|^2 = o_p\left(M^{1/2}\right). \tag{A.1}$$

Finally, for $T_N^{(2)}$, using the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
T_N^{(2)} &= N \sum_{j=1-N}^{N-1} k^2(j/M) \langle \mathbf{c}_{\boldsymbol{b}}^{(12)}(j), \mathbf{c}_{\hat{\boldsymbol{b}}}^{(12)}(j) - \mathbf{c}_{\boldsymbol{b}}^{(12)}(j) \rangle \\
&\leq N \{ \sum_{j=1-N}^{N-1} k^2(j/M) \| \mathbf{c}_{\boldsymbol{b}}^{(12)}(j) \|^2 \}^{1/2} \{ \sum_{j=1-N}^{N-1} k^2(j/M) \| \mathbf{c}_{\boldsymbol{b}}^{(12)}(j) - \mathbf{c}_{\hat{\boldsymbol{b}}}^{(12)}(j) \|^2 \}^{1/2}.
\end{aligned}$$

From Markov inequality, we have $\sum_{j=1}^{N-1} k^2(j/M) \| \mathbf{c}_{\boldsymbol{b}}^{(12)}(j) \|^2 = O_p(M/N)$. From (A.1) and the fact that $M/N \to 0$, we obtain that

$$T_N^{(2)} = o_p(M^{1/2}).$$

Thus, the second result is verified. This completes the proof of Lemma 3.2.

PROOF OF THEOREM 4.1.

By definition of $\mathcal{Q}_N$, we can write

$$\begin{aligned}
\left(\frac{M^{1/2}}{N} \right) \mathcal{Q}_N &= \frac{M^{1/2} \| \boldsymbol{s}_{\hat{\boldsymbol{a}}}^{(12)} \|_2^2 - \left(\frac{M^{1/2}}{N} \right) m_1 m_2 S_N(k)}{\{ 2 m_1 m_2 D(k) \}^{1/2}} \\
&= \frac{\| \boldsymbol{s}_{\hat{\boldsymbol{a}}}^{(12)} \|_2^2}{\{ 2 m_1 m_2 M^{-1} D_N(k) \}^{1/2}} - \frac{N^{-1} S_N(k)}{\{ 2 M^{-1} D_N(k) \}^{1/2}} (m_1 m_2)^{1/2}.
\end{aligned}$$

From (3.6), the last term of the previous equation goes to zero when $M/N \to 0$ as $N \to \infty$. By the invariance property of the coherency function under linear transformations, see Priestley (1981, p. 661), we have $\| \boldsymbol{s}_{\mathbf{x}}^{(12)} \| = \| \boldsymbol{s}_{\boldsymbol{a}}^{(12)} \|$. Using, as in Lemma 3.1, the linear transformation $\boldsymbol{b}_t = \boldsymbol{\Sigma}^{-1/2} \boldsymbol{a}_t$, we also get $\| \boldsymbol{s}_{\boldsymbol{a}}^{(12)} \| = \| \boldsymbol{s}_{\boldsymbol{b}}^{(12)} \|$. Thus, to prove the consistency result (4.2), it is sufficient to verify that $\| \boldsymbol{s}_{\hat{\boldsymbol{b}}}^{(12)} \|^2 - \| \boldsymbol{s}_{\boldsymbol{b}}^{(12)} \|^2 \xrightarrow{p} 0$, which follows from the two following lemmas. We first note that $\tilde{\boldsymbol{s}}_{\boldsymbol{b}}^{(12)}(w)$ is defined as $\boldsymbol{s}_{\hat{\boldsymbol{b}}}^{(12)}(w)$, the residual series $(\hat{\boldsymbol{b}}_t^{(1)}, \hat{\boldsymbol{b}}_t^{(2)})_{t=1}^N$ being replaced by the innovation series $(\boldsymbol{b}_t^{(1)}, \boldsymbol{b}_t^{(2)})_{t=1}^N$.

Lemma A.6 *Under the assumptions of Theorem 4.1, we have*

$$\| \boldsymbol{s}_{\hat{\boldsymbol{b}}}^{(12)} \|_2^2 - \| \tilde{\boldsymbol{s}}_{\boldsymbol{b}}^{(12)} \|_2^2 \xrightarrow{p} 0$$

Lemma A.7 *Under the assumptions of Theorem 4.1, we have*

$$\|\tilde{s}_b^{(12)}\|^2 - \|s_b^{(12)}\|^2 \xrightarrow{p} 0$$

PROOF OF LEMMA A.6.

By definition of $s_{\hat{b}}^{(12)}$ and $\tilde{s}_b^{(12)}$, and by similar calculations to those for the proof in Lemma 3.2, we obtain

$$
\begin{aligned}
\|s_{\hat{b}}^{(12)}\|_2^2 - \|\tilde{s}_b^{(12)}\|_2^2 &= \sum_{j=1-N}^{N-1} k^2(j/M)(\|c_{\hat{b}}^{(12)}(j)\|^2 - \|c_b^{(12)}(j)\|^2) \\
&= \sum_{j=1-N}^{N-1} k^2(j/M)\|c_{\hat{b}}^{(12)}(j) - c_b^{(12)}(j)\|^2 \\
&\quad + 2\sum_{j=1-N}^{N-1} k^2(j/M)\langle c_b^{(12)}(j), c_{\hat{b}}^{(12)}(j) - c_b^{(12)}(j)\rangle.
\end{aligned}
$$

It is sufficient to prove that the first term goes to zero in probability, because the second term can be bounded by a product of the first term and a finite quantity, using the Cauchy-Schwarz inequality. With the notations of Lemma 3.2, we can write

$$\sum_{j=1-N}^{N-1} k^2(j/M)\|c_{\hat{b}}^{(12)}(j) - c_b^{(12)}(j)\|^2 \leq 4\sum_{l=1}^{3} T_{lN},$$

where T_{lN}, $l = 1, 2, 3$, are defined as Lemma 3.2. We first prove that $T_{1N} \to 0$ in probability. By the Cauchy-Schwarz inequality, we obtain

$$T_{1N} \leq M\{\frac{1}{M}\sum_{j=0}^{N-1} k^2(j/M)\}\{\frac{1}{N}\sum_{t=1}^{N}\|b_t^{(1)}\|^2\}\{\frac{1}{N}\sum_{t=1}^{N}\|\hat{\eta}_t\|^2\}.$$

By definition of $\hat{\eta}_t$, it follows that

$$\frac{1}{N}\sum_{t=j}^{N}\|\hat{\eta}_t\|^2 \leq \frac{1}{N}\sum_{t=1}^{N}\{\|\Phi(p_2) - \hat{\Phi}(p_2)X_t^{(2)}(p_2)\|^2 + \|\sum_{l=p_2+1}^{\infty}\Phi_l^{(2)}X_{t-l}^{(2)}\|^2\}.$$

Under the assumptions on the process b, on p_2 and on the parameters $(\Phi_l^{(2)})$, we have

$$T_{1N} = O_p(\frac{Mp_2^2}{N}) + O_p(M\sum_{l=p_2+1}^{\infty}\|\Phi_l^{(2)}\|^2) = o_p(1).$$

By symmetry, we can verify that $T_{2N} = o_p(1)$. For T_{3N}, we can write

$$
\begin{aligned}
T_{3N} &= \sum_{j=0}^{N-1} k^2(j/M)\|\frac{1}{N}\sum_{t=j+1}^{N}\hat{\delta}_t\hat{\eta}_{t-j}^T\|^2 \\
&\leq M\{\frac{1}{M}\sum_{j=0}^{N-1} k^2(j/M)\}\{\frac{1}{N}\sum_{t=1}^{N}\|\hat{\delta}_t\|^2\}\{\frac{1}{N}\sum_{t=1}^{N}\|\hat{\eta}_t\|^2\}.
\end{aligned}
$$

33

By symmetry, we can prove that $\frac{1}{N}\sum_{t=1}^{N}\|\hat{\boldsymbol{\delta}}_t\|^2 = O_p(\frac{p_1^2}{N}) + O_p(1)\sum_{l=p_1+1}^{\infty}\|\boldsymbol{\Phi}_l^{(1)}\|^2$, and using the same assumptions as those for T_{1N}, we obtain that $T_{3N} = o_p(1)$. Finally, we conclude that

$$\|\boldsymbol{s}_{\hat{\boldsymbol{b}}}^{(12)}\|^2 - \|\tilde{\boldsymbol{s}}_{\boldsymbol{b}}^{(12)}\|^2 \;\; = \;\; o_p(1).$$

This complete the proof of Lemma A.6.

PROOF OF LEMMA A.7.

The development here is similar to the one in the previous proof. By definition of $\|\tilde{\boldsymbol{s}}_{\boldsymbol{b}}^{(12)}\|^2$ and $\|\boldsymbol{s}_{\boldsymbol{b}}^{(12)}\|^2$, we can write

$$
\begin{aligned}
\|\tilde{\boldsymbol{s}}_{\boldsymbol{b}}^{(12)}\|^2 - \|\boldsymbol{s}_{\boldsymbol{b}}^{(12)}\|^2 \;\; &= \;\; \sum_{j=1-N}^{N-1} k^2(j/M)\|\boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j)\|^2 - \sum_{j=-\infty}^{\infty}\|\boldsymbol{\gamma}_{\boldsymbol{b}}^{(12)}(j)\|^2 \\
&= \;\; \sum_{j=1-N}^{N-1} k^2(j/M)[\|\boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j)\|^2 - \|\boldsymbol{\gamma}_{\boldsymbol{b}}^{(12)}(j)\|^2] \\
&+ \;\; \sum_{j=1-N}^{N-1} (k^2(j/M)-1)\|\boldsymbol{\gamma}_{\boldsymbol{b}}^{(12)}(j)\|^2 + \sum_{|j|>N}\|\boldsymbol{\gamma}_{\boldsymbol{b}}^{(12)}(j)\|^2.
\end{aligned}
$$

Given Assumptions B and C, the second term of this equality goes to zero as $M \to \infty$ by the dominate convergence theorem. Also the last term goes to zero since $\sum_{j=-\infty}^{+\infty}\|\boldsymbol{\Gamma}_{\boldsymbol{X}}^{(12)}(j)\|^2 < \infty$. It remains to show that the first term tends to 0 in probability. Since

$$\sum_{j=1-N}^{N-1} k^2(j/M)[\|\boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j)\|^2 - \|\boldsymbol{\gamma}_{\boldsymbol{b}}^{(12)}(j)\|^2] = \sum_{j=1-N}^{N-1} k^2(j/M)\|\boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j) - \boldsymbol{\gamma}_{\boldsymbol{b}}^{(12)}(j)\|^2 + 2\sum_{j=1-N}^{N-1} k^2(j/M)\lambda(j),$$

where $\lambda(j) = \langle \boldsymbol{\gamma}_{\boldsymbol{b}}^{(12)}(j), \boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j) - \boldsymbol{\gamma}_{\boldsymbol{b}}^{(12)}(j)\rangle$. As in the proof of the Lemma A.6, it suffices to show that $\sum_{j=1-N}^{N-1} k^2(j/M)\|\boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j) - \boldsymbol{\gamma}_{\boldsymbol{b}}^{(12)}(j)\|^2 = \sum_{u,v=1}^{m_1,m_2}\sum_{j=-N+1}^{N-1} k^2(j/M)\{c_{uv}^{(12)}(j) - \gamma_{uv}^{(12)}(j)\}^2 = o_p(1)$ where $c_{uv}^{(12)}(j)$ and $\gamma_{uv}^{(12)}(j)$ denote the (u,v) element of the matrices $\boldsymbol{C}_{\boldsymbol{b}}^{(12)}(j)$ and $\boldsymbol{\Gamma}_{\boldsymbol{b}}^{(12)}(j)$ respectively. Under Assumptions A, C and by Priestley (1981, p. 325-26), we have

$$Var\{c_{uv}^{(12)}(j)\} \;\; = \;\; \frac{1}{N}\sum_{i=-N+1}^{N-1}(1 - \frac{|i|}{N})\{\gamma_{uv}^{(12)}(i+j)\gamma_{uv}^{(12)}(i-j) + \kappa_{uvuv}(0,j,i,i+j)\}.$$

It follows, from Assumptions B and C, that

$$
\begin{aligned}
\sum_{j=-N+1}^{N-1} k^2(j/M)Var\{c_{uv}^{(12)}(j)\} \;\; &= \;\; \frac{1}{N}\sum_{j=-N+1}^{N-1} k^2(j/M)\sum_{i=-N+1}^{N-1}(1-\frac{|i|}{N})\gamma_{uv}^{(12)}(i+j)\gamma_{uv}^{(12)}(i-j) \\
&+ \;\; \frac{1}{N}\sum_{j=-N+1}^{N-1} k^2(j/M)\sum_{i=-N+1}^{N-1}(1-\frac{|i|}{N})\kappa_{uvuv}(0,j,i,i+j) \\
&= \;\; O(\frac{M}{N} + \frac{1}{N}).
\end{aligned}
$$

34

Thus, we conclude that $\sum_{j=1-N}^{N-1} k^2(j/M)\|\boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j) - \boldsymbol{\gamma}_{\boldsymbol{b}}^{(12)}(j)\|^2 = o_p(1)$. This completes the proof of Lemma A.7 and Theorem 4.1 follows from Lemmas A.6 and A.7.

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.

Berk, K. N. (1974), 'Consistent autoregressive spectral estimates', *Annals of Statistics* **2**, 489–502.

Bouhaddioui, C. (2002), 'Tests d'indépendance de deux séries multivariées autorégressives d'ordre infini', *Unpublished Ph.D. Thesis, Département de mathématiques et de statistique*, Université de Montréal.

Bouhaddioui, C. and Roy, R. (2003), 'On the asymptotic distribution of residual cross-correlation between two independent infinite order vector autoregressive series', *Technical Report CRM-2924, Centre de recherches mathématiques*, Université de Montréal.

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

Duchesne, P. and Roy, R. (2003), 'Robust tests for independence of two time series', *Statistica Sinica* **13**, 827–852.

El Himdi, K. and Roy, R. (1997), 'Tests for noncorrelation of two multivariate ARMA time series', *Canadian Journal of Statistics* **25**, 233–256.

Geweke, J. (1981), 'The approximate slopes of econometric tests', *Econometrica* **49**, 1427–1442.

Hallin, M. and Saidi, A. (2002), 'Testing independance and causality between multivariate ARMA times series', *Journal of Time Series Analysis, forthcoming.*

Hannan, E. (1970), *Multiple Time Series*, John Wiley and Sons: New York.

Haugh, L. D. (1976), 'Checking the independence of two covariance-stationary time series: a univariate residual cross correlation approch', *Journal of the American Statistical Association* **71**, 376–385.

Hong, Y. (1996a), 'Testing for independence between two covariance stationary time series', *Biometrika* **83, 3**, 615–625.

Hong, Y. (1996b), 'A separate mathematical appendix for testing for independence between two covariance stationary time series', *Mimeo, Department of Economics and Department of Statistical Sciences,Cornell University.*

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

Koch, P. and Yang, S.-S. (1986), 'A method for testing the independance of two time series that accounts for a potential pattern in the cross-correlation function', *Journal of the American Statistical Analysis* **81**, 533–544.

Lewis, R. and Reinsel, G. (1985), 'Prediction of multivariate time series by autoregressive model fitting', *Journal of Multivariate Analysis* **16**, 393–411.

Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Springer-Verlag: Berlin.

Paparoditis, E. (1996), 'Bootstrapping autoregressive and moving average parameter estimates of infinite order vector autoregressive processes', *Journal of Multivariate Analysis* **57**, 277–296.

Pham, D., Roy, R. and Cédras, L. (2003), 'Tests for non-correlation of two cointegrated ARMA time series', *Journal of Time Series Analysis* **24**, 553–577.

Priestley, M. B. (1981), *Spectral Analysis and Time Series*, Vol. 1, Academic Press: London.

Racette, D. and Raynauld, J. (1992), 'Un modèle BVAR de prévision de la dépense nominale et d'analyse de la politique monétaire canadienne', *Séminaire sur les questions monétaires, Banque du Canada* pp. 317–325.

Reinsel, G. (1993), *Elements of Multivariate Time Series Analysis*, Springer-Verlag: New York.

Tiao, G. C. and Box, G. (1981), 'Modeling multiple time series with applications', *Journal of the American Statistical Association* **76**, 802–816.