

2011s-23

**Semiparametric Innovation-Based Tests of Orthogonality and
Causality Between Two Infinite-Order Cointegrated Series
with Application to Canada/US Monetary Interactions**

Chafik Bouhaddioui, Jean-Marie Dufour

Série Scientifique
Scientific Series

Montréal
Février 2011

© 2011 *Chafik Bouhaddioui, Jean-Marie Dufour*. Tous droits réservés. *All rights reserved*. Reproduction partielle permise avec citation du document source, incluant la notice ©.
Short sections may be quoted without explicit permission, if full credit, including © notice, is given to the source.



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 du Développement économique et régional et de la Recherche, 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 du Développement économique et régional et de la Recherche, and grants and research mandates obtained by its research teams.

Les partenaires du CIRANO

Partenaire majeur

Ministère du Développement économique, de l'Innovation et de l'Exportation

Partenaires corporatifs

Banque de développement du Canada
Banque du Canada
Banque Laurentienne du Canada
Banque Nationale du Canada
Banque Royale du Canada
Banque Scotia
Bell Canada
BMO Groupe financier
Caisse de dépôt et placement du Québec
Fédération des caisses Desjardins du Québec
Financière Sun Life, Québec
Gaz Métro
Hydro-Québec
Industrie Canada
Investissements PSP
Ministère des Finances du Québec
Power Corporation du Canada
Raymond Chabot Grant Thornton
Rio Tinto
State Street Global Advisors
Transat A.T.
Ville de Montréal

Partenaires universitaires

École Polytechnique de Montréal
HEC Montréal
McGill University
Université Concordia
Université de Montréal
Université de Sherbrooke
Université du Québec
Université du Québec à Montréal
Université Laval

Le CIRANO collabore avec de nombreux centres et chaires de recherche universitaires dont on peut consulter la liste sur son site web.

Les cahiers de la série scientifique (CS) visent à rendre accessibles des résultats de recherche effectuée au CIRANO afin de susciter échanges et commentaires. Ces cahiers sont écrits dans le style des publications scientifiques. Les idées et les opinions émises sont sous l'unique responsabilité des auteurs et ne représentent pas nécessairement les positions du CIRANO ou de ses partenaires.

This paper presents research carried out at CIRANO and aims at encouraging discussion and comment. The observations and viewpoints expressed are the sole responsibility of the authors. They do not necessarily represent positions of CIRANO or its partners.

ISSN 1198-8177

Partenaire financier

Développement
économique, Innovation
et Exportation
Québec 

Semiparametric Innovation-Based Tests of Orthogonality and Causality Between Two Infinite-Order Cointegrated Series with Application to Canada/US Monetary Interactions

Chafik Bouhaddioui^{*}, Jean-Marie Dufour[†]

Abstract

We propose a semiparametric approach for testing orthogonality and causality between two infinite-order cointegrated vector autoregressive IVAR(1) series. The procedures considered can be viewed as extensions of classical methods proposed by Haugh (1976, *JASA*) and Hong (1996, *Biometrika*) for testing independence between stationary univariate time series. The tests are based on the residuals of long autoregressions, hence allowing for computational simplicity, weak assumptions on the form of the underlying process, and a direct interpretation of the results in terms of innovations (or reduced-form shocks). The test statistics are standardized versions of the sum of weighted squares of residual cross-correlation matrices. The weights depend on a kernel function and a truncation parameter. The asymptotic distributions of the test statistics under the null hypothesis are derived, and consistency is established against fixed alternatives of serial cross-correlation of unknown form. Apart from standardization factors, the multivariate portmanteau statistic which takes into account a fixed number of lags, can be viewed as a special case of our procedure based on the truncated uniform kernel. A simulation study is presented which indicates that the proposed tests have good size and power properties in finite samples. The proposed procedures are applied to study interactions between Canadian and American monetary quarterly variables associated with monetary policy (money, interest rates, prices, aggregate output). The empirical results clearly allow to reject the absence of correlation between the shocks in both countries, and indicate a unidirectional Granger causality running from the U.S. variables to the Canadian ones.

Key words Infinite-order cointegrated vector autoregressive process; independence; causality; residual cross-correlation; consistency; asymptotic power.

^{*} Department of Statistics, United Arab Emirates University and Centre interuniversitaire de recherche en analyse des organisations (CIRANO). Mailing address: College of Business and Economics, Department of Statistics, United Arab Emirates University, P.O. Box 17555 Al Ain, United Arab Emirates. TEL: 971 3 713 3345; FAX 971 3 762 4385; e-mail: chafikb@uaeu.ac.ae.

[†] William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 8879; FAX: (1) 514 398 4938; e-mail: jean-marie.dufour@mcgill.ca. Webpage: <http://www.jeanmariedufour.com>

1. Introduction

Studying the dynamic relationship between two multivariate series is a fundamental objective of time series analysis in statistics and econometrics. For example, in econometrics, this can help one to understand the associated economic mechanisms. In this context, a basic problem consists in testing independence (or the absence of serial cross-correlation) between two vector processes. The seminal paper on this problem is due to Haugh (1976), who proposed a general procedure for testing independence between two covariance-stationary ARMA time series. His method is based on considering cross-correlations between residuals obtained after fitting univariate ARMA models on each series. Since the innovations of an ARMA model follow a white noise by assumption, this considerably simplifies the underlying distributional theory, and the corresponding tests are relatively simple to apply. Further, the corresponding statistics have a direct interpretation in terms of process innovations (or reduced-form shocks), a feature of interest in econometrics since innovations can often be interpreted as “shocks” to economic systems. Consequently, the possibility of focusing on “shock cross-correlations” should be useful in econometric research. Furthermore, Wiener-Granger causality properties can also be represented in terms of the innovation cross-correlations considered by Haugh (1976), so that causality properties can also be assessed by checking whether relevant subsets of innovation cross-correlations are close to zero (or not); see Pierce and Haugh (1977, 1979). The required distributional theory is however more complex under usual noncausality hypotheses (due to the form of the required standard errors), and no solution was proposed at that time.

The work of Haugh (1976) has been extended by several authors; see Hong (1996a), El Himdi and Roy (1997), Pham, Roy and Cédras (2003), Hallin and Saidi (2005), Bouhaddioui and Roy (2006a, 2006b), Hallin and Saidi (2007), Saidi (2007), and Bouhaddioui and Dufour (2008). Most of these studies focus on independence between two multivariate finite-order vector autoregressive (VAR) or vector autoregressive moving-average (VARMA) models. El Himdi and Roy (1997) extended the procedure developed by Haugh (1976) in order to test non-correlation and non-causality between two time series in the context of multivariate stationary and invertible VARMA models. This result was used by Hallin and Saidi (2005) to

develop a test that takes into account a possible pattern in the signs of cross-correlations at different lags. In a nonparametric setup, Hallin, Jurečková, Picek and Zahaf (1999) proposed a test for independence between two autoregressive time series which is based on autoregressive rank score, while Hong (1998) proposed a test based on the empirical distribution functions.

In many cases, however, the stationarity condition is unrealistic and constitutes a heavy constraint. Further, in nonstationary cases, it is more important to work directly with the original series without transformations (by taking differences on each component) which causes distributional complications and can lead to misleading results. Engle and Granger (1987) introduced the concept of cointegration which is used in many recent studies across several fields. In the case of a finite-order autoregressive cointegrated vector, Ahn and Reinsel (1990) developed an efficient method estimation in Gaussian processes. Yap and Reinsel (1995) proposed a full-rank and reduced rank Gaussian estimation procedures for cointegrated VARMA processes. For a good discussion of the related models, see Lütkepohl (2001). By exploiting the estimation methods proposed by Yap and Reinsel (1995), Pham et al. (2003) generalized the main result of El Himdi and Roy (1997) to the case of two cointegrated (or partially nonstationary) VARMA series. They proposed test statistics based on residual cross-correlation matrices $\mathbf{R}_{\hat{\mathbf{a}}}^{(12)}(j)$, $|j| \leq M$, where M does not depend on the sample size N , between the two residual series $\{\hat{\mathbf{a}}_t^{(1)}\}$ and $\{\hat{\mathbf{a}}_t^{(2)}\}$ 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 show that an arbitrary vector of residual cross-correlations asymptotically follows a multivariate normal distribution.

In practice, a finite-order VAR model is a rough approximation to the true data generating process of a given multivariate time series. It is not common for the true model to be a function of a small number of unknown parameters. From this perspective, a more flexible alternative approach assumes that the data are generated by an infinite-order autoregressive process. Such models lead one to consider a truncated (potentially long) autoregression as an approximation of the underlying process. Very often in statistics, the properties of estimators and test statistics under the assumption of a true model, even if model assumptions are clearly not fulfilled. For example, in VARMA estimation, it is well known that misspecification of the

AR or MA orders can lead to inconsistent estimators. Further, the estimation of VARMA models is highly nonlinear and raises difficult identification issues (in the sense of model non-unicity). Correspondingly, non-causality hypotheses become nonlinear and lead to non-standard distributional problems; see Boudjellaba, Dufour and Roy (1992, 1994).

The autoregressive model fitting approach has been successfully applied by several authors: Akaike (1969), Berk (1974) and Parzen (1974) for spectral density estimation, Parzen (1974), Lütkepohl (1985), Lewis and Reinsel (1985) and Bhansali (1996) for prediction, Saikkonen (1992) for inference in cointegrated systems; see also Lütkepohl (1991) and Reinsel (1997). In previous work [Bouhaddioui and Roy (2006*b*)], we have generalized the work of El Himdi and Roy (1997) to the case of two stationary multivariate infinite-order autoregressive series $\text{VAR}(\infty)$. This result allows one to develop tests against serial cross-correlation at a particular lag or at a fixed number of lags j such as $|j| \leq M$, where M does not depend on the sample size N .

In the univariate stationary case, Hong (1996*c*) introduced an important extension of Haugh's procedure by proposing a class of spectral test statistics. His approach is semiparametric and valid for two infinite-order autoregressive series $\text{AR}(\infty)$. It is based on fitting an autoregressive model of order p to a series of N observations from each infinite-order autoregressive process. Following Berk (1974), the order p of the fitted autoregression is a function of the sample size. This approach was also used by Hong (1999), Duchesne and Roy (2003), Duchesne (2005), and recently by Shao (2008) for the case of two univariate long memory processes. In Bouhaddioui and Roy (2006*a*), this approach is extended to $\text{VAR}(\infty)$ models, hence protecting against misspecifications of the true underlying VARMA models. In contrast to Haugh's test, which is based on the residual cross-correlations at lag j such that $|j| \leq M$, the portmanteau test \mathcal{Q}_N is consistent for a large class of serial cross-correlations alternatives of an arbitrary form between the two series.

The main objective of this paper is to propose a semiparametric approach to test orthogonality (non-correlation) and causality between two infinite-order cointegrated autoregressive [IVAR(∞)] models. These models were introduced by Saikkonen (1992) and involve much weaker conditions than those considered by

Yap and Reinsel (1995), Pham et al. (2003), Hallin and Saidi (2005) and Saidi (2007); for further, discussion of this setup, see Saikkonen and Lütkepohl (1996), Saikkonen and Luukkonen (1997) and Lütkepohl and Saikkonen (1997). The problem of testing the absence of correlation between two $\text{IVAR}(\infty)$ was first considered in Bouhaddioui and Dufour (2008), where the asymptotic distribution of an arbitrary vector of residual cross-correlations and partial cross-correlations under the hypothesis of non-correlation of the two series is derived under the assumption process innovations are a strong white noise. However, the test statistics considered in the latter paper only consider one lag at a time or a fixed number of lags j (for example $|j| \leq M$).

In this article, we propose a multivariate version of the weighted portmanteau statistic \mathcal{Q}_N , which is based on the sample cross-correlation matrices $\mathbf{R}_{\hat{\mathbf{a}}}^{(12)}(j)$, $|j| \leq N - 1$, between residual series $\{\hat{\mathbf{a}}_t^{(1)}\}$ and $\{\hat{\mathbf{a}}_t^{(2)}\}$. The residuals are obtained by approximating the two multivariate $\text{IVAR}(\infty)$ series by finite-order autoregressions whose order increases with the sample size at an appropriate rate. The test statistics continue to have an asymptotic $\mathcal{N}(0, 1)$ distribution under the hypothesis of independence of the two series and are also consistent for any alternative of serial cross-correlation of arbitrary form. Further, by restricting the test statistics to positive lags or negative lags, we propose modified statistics \mathcal{Q}_N^+ and \mathcal{Q}_N^- for testing Granger noncausality (as opposed to the absence of any correlation) between the two vector processes considered. This is both a technically more difficult problem and one of potentially much greater econometric interest, since the hypothesis of no correlation typically does not hold in economic data.

The proposed tests can be applied in different contexts and may help one to understand the dynamic properties of economic and financial time series. In this paper, we illustrate their use by studying the link [in the sense of orthogonality and causality] between a set of Canadian macroeconomic and monetary variables [real income, prices, interest rates, and money] and a set of corresponding U.S. variables. For the sake of comparability, we consider quarterly data previously studied by El Himdi and Roy (1997) and Pham et al. (2003). The results appear to be less dependent on the choice of truncation parameters in portmanteau statistics. The independence assumption is clearly rejected, the strongest link being contemporaneous. Further, our results indicate unidirectional Granger causality running from the U.S. variables to the Canadian ones.

The organization of the paper is as follows. Section 2 describes the statistical framework as well as some preliminary results. The new test statistics are introduced in Section 3. We show that their asymptotic distributions under the null hypothesis are $\mathcal{N}(0, 1)$. In section 4, we establish the consistency of the tests. In Section 5, we present the results of a small Monte Carlo experiment studying the level and power of the tests in finite samples, including the effect of the kernel. Finally, the new tests are applied in Section 6 to a set of American and Canadian macroeconomic indicators to study monetary interactions. We conclude in Section 7. The proofs of all results are relegated to the Appendix.

2. Framework and preliminary results

Following the notations of Saikkonen (1992) and Saikkonen and Lütkepohl (1996), we consider a d -dimensional process $\mathbf{X} = \{\mathbf{X}_t, t \in \mathbb{Z}\}$ partitioned into two subprocesses $\mathbf{X}_i = \{\mathbf{X}_{it}, t \in \mathbb{Z}\}$, $i = 1, 2$, with d_1 and d_2 components respectively ($d_1 + d_2 = d$). The data generating process has the form:

$$\mathbf{X}_{1t} = \mathbf{C}_1 \mathbf{X}_{2t} + \boldsymbol{\varepsilon}_{1t}, \quad (2.1)$$

$$\Delta \mathbf{X}_{2t} = \boldsymbol{\varepsilon}_{2t}, \quad (2.2)$$

where \mathbf{C}_1 is a given $d_1 \times d_2$ matrix, Δ is the usual difference operator, and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}'_{1t}, \boldsymbol{\varepsilon}'_{2t})'$ is a stationary process with zero mean and continuous spectral density matrix which is positive definite at zero frequency. \mathbf{X}_{2t} is an integrated vector process of order one (with no cointegrating relationship), while \mathbf{X}_{1t} and \mathbf{X}_{2t} are cointegrated.

By taking first differences in (2.1), the above system can be written in the form

$$\Delta \mathbf{X}_t = \begin{bmatrix} -\mathbb{I}_{d_1} & \mathbf{C}_1 \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{X}_{t-1} + \mathbf{b}_t = \mathbf{J} \boldsymbol{\Theta}' \mathbf{X}_{t-1} + \mathbf{b}_t \quad (2.3)$$

where \mathbb{I}_d represents the $d \times d$ identity matrix, $\mathbf{J}' = [-\mathbb{I}_{d_1} : \mathbf{0}]$, $\boldsymbol{\Theta}' = [\mathbb{I}_{d_1} : -\mathbf{C}_1]$, $\mathbf{b}_t = [\mathbf{b}'_{1t} : \mathbf{b}'_{2t}]'$ is nonsingular transformation of $\boldsymbol{\varepsilon}_t$ defined by

$$\mathbf{b}_{1t} = \boldsymbol{\varepsilon}_{1t} + \mathbf{C}_1 \boldsymbol{\varepsilon}_{2t}, \quad \mathbf{b}_{2t} = \boldsymbol{\varepsilon}_{2t}. \quad (2.4)$$

The notation $\mathbf{A} = [\mathbf{A}_1 : \mathbf{A}_2]$ means that the matrix \mathbf{A} is partitioned into a matrix \mathbf{A}_1 consisting of the first

d_1 columns and a matrix \mathbf{A}_2 with d_2 columns.

We suppose also that the process \mathbf{b}_t (and hence ε_t) has an infinite-order autoregressive representation

$$\sum_{l=0}^{\infty} \mathbf{G}_l \mathbf{b}_{t-l} = \mathbf{a}_t, \quad \mathbf{G}_0 = \mathbb{I}_d, \quad (2.5)$$

where \mathbf{a}_t is independent and identically distributed white noise process with $\mathbb{E}(\mathbf{a}_t) = \mathbf{0}$ and $\mathbb{E}(\mathbf{a}_t \mathbf{a}_t') = \boldsymbol{\Sigma}_a$ is a definite positive matrix. Setting $\mathbf{G}(z) = \mathbf{I}_d - \sum_{l=1}^{\infty} \mathbf{G}_l z^l$, the stationarity hypothesis of the process \mathbf{b}_t implies that the zeros of the equation $\det\{\mathbf{G}(z)\} = 0$ all lie outside the unit circle $|z| = 1$, where $\det\{\mathbf{A}\}$ denotes the determinant of the square matrix \mathbf{A} . A further assumption is that the coefficient matrices \mathbf{G}_l satisfy the summability condition

$$\sum_{l=1}^{\infty} l^n \|\mathbf{G}_l\| < \infty \quad (2.6)$$

for some $n \geq 1$ and $\|\cdot\|$ is the Euclidean matrix norm defined by $\|\mathbf{A}\|^2 = \text{tr}(\mathbf{A}'\mathbf{A})$. This is a standard condition for weakly stationary processes, which ensures that the process be well defined. Depending on n , it imposes weak restrictions on the autocorrelation structure of the process \mathbf{b}_t . Also, it implies that the process \mathbf{b}_t and, consequently, \mathbf{X}_t can be approximate by a finite-order autoregression. The order p_N of the fitted autoregression is a function of the sample size; *i.e.*, $p_N = p(N)$. In order to reduce approximation errors, we allow the maximal order p_N to increase to infinity, at some rate, simultaneously with realization length N , see Burnham and Anderson (2002). In the sequel, we assume the following assumption on the finite autoregressive order.

Assumption 2.1 $N^{-1/3} p_N \rightarrow 0$ and $\sqrt{N} \sum_{l=p_N+1}^{\infty} \|\mathbf{G}_l\| \rightarrow 0$ as $N \rightarrow \infty$.

The condition $p_N = o(N^{1/3})$ for the rate of increase of p_N ensures that enough sample information is asymptotically available for estimators to have standard limiting distributions. The condition $\sqrt{p_N} \sum_{j=p_N+1}^{\infty} \|\mathbf{G}_j\| \rightarrow 0$ imposes a lower bound on the growth rate of p_N , 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) and Burnham and Anderson (2002).

Using the equations (2.3) - (2.5) and rearranging terms, we obtain the autoregressive *error correction*

model (ECM) representation

$$\Delta \mathbf{X}_t = \boldsymbol{\Psi} \boldsymbol{\Theta}' \mathbf{X}_{t-1} + \sum_{l=1}^{p_N} \boldsymbol{\Pi}_l \Delta \mathbf{X}_{t-l} + \mathbf{e}_t, \quad t = p+1, p+2, \dots \quad (2.7)$$

where $\mathbf{e}_t = \mathbf{a}_t - \sum_{l=p_N+1}^{\infty} \mathbf{G}_l \mathbf{b}_{t-l}$, $\boldsymbol{\Psi} = -\sum_{l=0}^{p_N} \mathbf{G}_l \mathbf{J}$, and the $d \times d_1$ matrix $\boldsymbol{\Psi}$ is of full column rank (at least for p_N large enough). Details for this derivation can be found in Saikkonen and Lütkepohl (1994). Note that the coefficient matrices $\boldsymbol{\Pi}_l (l = 1, \dots, p_N)$ are functions of $\boldsymbol{\Theta}$ and $\mathbf{G}_l (l = 1, 2, \dots)$, and they depend on p_N . Furthermore, the sequence $\boldsymbol{\Pi}_l (l = 1, \dots, p_N)$ is absolutely summable as $p_N \rightarrow \infty$.

The autoregressive ECM in (2.7) can also be rewritten in a pure vector autoregressive (VAR) form

$$\mathbf{X}_t = \sum_{l=1}^{p_N+1} \boldsymbol{\Phi}_l \mathbf{X}_{t-l} + \mathbf{e}_t \quad (2.8)$$

where $\boldsymbol{\Phi}_1 = \mathbb{I}_d + \boldsymbol{\Psi} \boldsymbol{\Theta}' + \boldsymbol{\Pi}_1$, $\boldsymbol{\Phi}_l = \boldsymbol{\Pi}_l - \boldsymbol{\Pi}_{l-1}$, $l = 2, \dots, p_N$ and $\boldsymbol{\Phi}_{p_N+1} = -\boldsymbol{\Pi}_{p_N}$. Although the $\boldsymbol{\Pi}_l$ depend on p_N , the same is not true for the $\boldsymbol{\Phi}_l$ except for $\boldsymbol{\Phi}_{p_N+1}$.

Saikkonen and Lütkepohl (1996) derived the asymptotic properties of the multivariate least square (LS) estimators of the VAR coefficients under a standard assumption. Let $\boldsymbol{\Phi}(p_N) = (\boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_{p_N})$ be the matrix of the first p_N autoregressive parameter matrices in the representation (2.7) and denote by $\hat{\boldsymbol{\Phi}}(p_N) = (\hat{\boldsymbol{\Phi}}_1, \dots, \hat{\boldsymbol{\Phi}}_{p_N})$ the corresponding LS estimator. The following proposition gives a direct result on the asymptotic properties of the estimator $\hat{\boldsymbol{\Phi}}(p_N)$. It can be proved using the same straightforward techniques that in part (i) of Theorem 3.2 in Saikkonen (1992); see also Theorem 2 in Saikkonen and Lütkepohl (1996).

Proposition 2.2 ASYMPTOTIC PROPERTIES OF THE AUTOREGRESSIVE PARAMETER ESTIMATORS.

Let $\{\mathbf{X}_t\}$ a process given by (2.8) and assume that $\mathbb{E}|a_{i,t}a_{j,t}a_{k,t}a_{l,t}| < \gamma_4 < \infty$, $1 \leq i, j, k, l \leq d$. Then, under Assumption 2.1,

$$\|\hat{\boldsymbol{\Phi}}(p_N) - \boldsymbol{\Phi}(p_N)\| = O_p\left(\frac{p_N^{1/2}}{N^{1/2}}\right). \quad (2.9)$$

Note that this proposition is formulated for the first p_N coefficient matrices, whereas the underlying process fitted to the data is a VAR($p_N + 1$), where p_N goes to infinity with the sample size N . The details of the estimates of the $\boldsymbol{\Phi}_l$ are given in Saikkonen and Lütkepohl (1994). This result can be considered as a generalization of Theorem 1 in Lewis and Reinsel (1985) in the infinite-order stationary vector autoregressive case. Also, in the stationary case, Paparoditis (1996) established this result under the same assumption when the estimators of the parameters are based on a bootstrap procedure.

We consider now two processes $\mathbf{X}^{(h)} = \{\mathbf{X}_t^{(h)}, t \in \mathbb{Z}\}$, $h = 1, 2$, with m_1 and m_2 components respectively, which satisfy the $IVAR(\infty)$ model (2.8). We are interested in whether $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are uncorrelated (or independent in the Gaussian case) at different lags. Also, based on the sign of the lags, this question can be generalized to study the causality in different directions between the two processes. Further, we suppose that for $h = 1, 2$, $\mathbf{X}^{(h)}$ follows an infinite-order cointegrated vector autoregressive model $IVAR(\infty)$ given by (2.3) and are uncorrelated. The non-correlation between $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ is equivalent to the non-correlation between the corresponding innovation processes $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$, see Proposition 2.1 in Pham et al. (2003). Thus, this hypothesis is equivalent to

$$\mathcal{H}_0 : \boldsymbol{\rho}_{\mathbf{a}}^{(12)}(j) = \mathbf{0}, \quad j \in \mathbb{Z},$$

where

$$\boldsymbol{\rho}_{\mathbf{a}}^{(12)}(j) = \boldsymbol{\Sigma}_1^{-1/2} \mathbf{E}[\mathbf{a}_t^{(1)} \mathbf{a}_{t-j}^{(2)T}] \boldsymbol{\Sigma}_2^{-1/2}$$

represents the cross-correlation matrix at lag j between the two innovation processes. $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ denote respectively the covariance matrices of the innovation processes $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$. We can also consider the two following hypotheses to study the causality between the two processes. For $n_0 = 0$ or 1, let

$$\mathcal{H}_0^+ : \boldsymbol{\rho}_{\mathbf{a}}^{(12)}(j) = \mathbf{0}, \quad j \geq n_0,$$

and

$$\mathcal{H}_0^- : \boldsymbol{\rho}_{\mathbf{a}}^{(12)}(j) = \mathbf{0}, \quad j \leq -n_0,$$

be, respectively, the hypotheses for testing non-correlation (or non-causality) in positive and negative lags. For $n_0 = 1$, the hypothesis \mathcal{H}_0^+ means that $\mathbf{X}^{(2)}$ does not cause $\mathbf{X}^{(1)}$ ($\mathbf{X}^{(2)} \not\rightarrow \mathbf{X}^{(1)}$) and under \mathcal{H}_0^- , an instantaneous causality exists between $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$. For $n_0 = 0$, the hypothesis \mathcal{H}_0^- is equivalent to $\mathbf{X}^{(1)}$ does not cause $\mathbf{X}^{(2)}$ ($\mathbf{X}^{(1)} \not\rightarrow \mathbf{X}^{(2)}$) and under \mathcal{H}_0^+ , an instantaneous causality exists between $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$. In the sequel, \xrightarrow{L} stands for convergence in law and \xrightarrow{p} for convergence in probability.

3. Test statistics and asymptotic null distributions

Based on a realization $\mathbf{X}_1^{(h)}, \dots, \mathbf{X}_N^{(h)}$ of length N , each process is fitted by a finite-order autoregressive model $\text{VAR}(p_N^{(h)})$. The order $p_N^{(h)}$ depends on the sample size N . The resulting residuals are given by

$$\hat{\mathbf{a}}_t^{(h)} = \begin{cases} \mathbf{X}_t^{(h)} - \sum_{l=1}^{p_N^{(h)}+1} \hat{\Phi}_{l,p_N^{(h)}} \mathbf{X}_{t-l}^{(h)} & \text{if } t = p_N^{(h)} + 2, \dots, N, \\ \mathbf{0} & \text{if } t \leq p_N^{(h)} + 1, \end{cases} \quad (3.1)$$

where the $\hat{\Phi}_{l,p_N^{(h)}}$ are the OLS estimators of $\Phi_{l,p_N^{(h)}}$. We can also use the conditional maximum likelihood estimator of the error correction form of the model as discussed by Ahn and Reinsel (1990) and Reinsel (1993) or some other estimator which has the same rate of convergence. The residual cross-covariance matrix $\mathbf{C}_{\hat{\mathbf{a}}}^{(12)}(j)$ is defined by

$$\mathbf{C}_{\hat{\mathbf{a}}}^{(12)}(j) = \begin{cases} N^{-1} \sum_{t=j+1}^N \hat{\mathbf{a}}_t^{(1)} \hat{\mathbf{a}}_{t-j}^{(2)T} & \text{if } 0 \leq j \leq N-1, \\ N^{-1} \sum_{t=-j+1}^N \hat{\mathbf{a}}_{t+j}^{(1)} \hat{\mathbf{a}}_t^{(2)T} & \text{if } -N+1 \leq j \leq 0, \end{cases} \quad (3.2)$$

while the corresponding residual cross-correlation matrix is

$$\mathbf{R}_{\hat{\mathbf{a}}}^{(12)}(j) = \mathbf{D}\{c_{\hat{\mathbf{a}},ii}^{(11)}(0)^{-1/2}\} \mathbf{C}_{\hat{\mathbf{a}}}^{(12)}(j) \mathbf{D}\{c_{\hat{\mathbf{a}},ii}^{(11)}(0)^{-1/2}\} \quad (3.3)$$

where $\mathbf{D}\{b_i\}$ a diagonal matrix whose elements are b_1, \dots, b_m . In the sequel, we suppose that for $h = 1, 2$, $\mathbf{X}^{(h)}$ satisfies (2.3). We wish to test the null hypothesis that they are uncorrelated (or independent in the Gaussian case), *i.e.* $\rho_{\mathbf{X}}^{(12)}(j) = \mathbf{0}, j \in \mathbb{Z}$. In the nonstationary case, we also need to work with the sample covariance and the sample correlation of the innovation process instead of the sample covariance of $\{\mathbf{X}_t\}$, because $\mathbb{E}[\mathbf{X}_t \mathbf{X}_{t-j}']$ depends not only on the lag j but also on t .

3.1. Orthogonality tests

In the univariate case, Hong (1996c) proposed a portmanteau-type statistic which is based on the sum of the weighted squared cross-correlations $r_{\hat{\mathbf{a}}}^{(12)}(j)$ at all possible lags between the residual series defined by

$$\mathcal{Q}_N = \frac{N \sum_{j=1-N}^{N-1} k^2(j/M) r_{\hat{\mathbf{a}}}^{(12)}(j)^2 - S_N(k)}{\{2D_N(k)\}^{1/2}} \quad (3.4)$$

where $k(\cdot)$ is an arbitrary kernel function, and M is a smoothing parameter, while $S_N(k)$ and $D_N(k)$ are normalization coefficients which depend on the kernel $k(\cdot)$:

$$S_N(k) = \sum_{j=1-N}^{N-1} \left(1 - \frac{|j|}{N}\right) k^2(j/M), \quad D_N(k) = \sum_{j=2-N}^{N-2} \left(1 - \frac{|j|}{N}\right) \left(1 - \frac{|j|+1}{N}\right) k^4(j/M). \quad (3.5)$$

They correspond to the asymptotic mean and variance of the weighted sum. In multivariate time series, the squared cross-correlation $r_{\hat{\mathbf{a}}}^{(12)}(j)^2$ in (3.4) is replaced by a quadratic form in the vector $\mathbf{r}_{\hat{\mathbf{a}}}^{(12)}(j) = \text{vec} \left[\mathbf{R}_{\hat{\mathbf{a}}}^{(12)}(j) \right]$. For \mathcal{H}_0 , the test statistic is based on the following sum of weighted quadratic forms at all possible lags:

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

where

$$Q_{\hat{\mathbf{a}}}(j) = N \mathbf{r}_{\hat{\mathbf{a}}}^{(12)}(j)' \left[\mathbf{R}_{\hat{\mathbf{a}}}^{(22)}(0)^{-1} \otimes \mathbf{R}_{\hat{\mathbf{a}}}^{(11)}(0)^{-1} \right] \mathbf{r}_{\hat{\mathbf{a}}}^{(12)}(j), \quad (3.6)$$

and $k(\cdot)$ is a suitable kernel function. The parameter M is a truncation point when the kernel has compact support, or a smoothing parameter when the kernel support is unbounded. We suppose that M is function of N such that $M \rightarrow \infty$ and $M/N \rightarrow 0$ when $N \rightarrow \infty$. 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 $k_T(z) = \mathbb{I}[|z| \leq 1]$, where $\mathbb{I}(A)$ represents the indicator function of the set A , which gives the same weight to all lags. The asymptotic distribution of $Q_{\hat{\mathbf{a}}}(j)$ is given in Bouhaddioui and Dufour (2008). In the sequel, we suppose that the kernel function k satisfies the following assumption.

Assumption 3.1 *The kernel $k : \mathbb{R} \rightarrow [-1, 1]$ is a symmetric function, continuous at zero, with at most a finite number of discontinuity points, such that $k(0) = 1$ and $\int_{-\infty}^{+\infty} k^2(z) dz < \infty$.*

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 $k(z) \rightarrow 0$ as $|z| \rightarrow \infty$. Thus, eventually, less weight is given to $\mathbf{R}_{\hat{\mathbf{a}}}^{(12)}(j)$ as j increases. It is worth noting that all the kernels used in spectral analysis satisfy Assumption 3.1; see Priestley (1981, Section 6.2.3). For the hypothesis \mathcal{H}_0 , the test statistic is a standardized version of $\mathcal{T}(\hat{\mathbf{a}}, \hat{\Sigma})$:

$$Q_N = \frac{\mathcal{T}(\hat{\mathbf{a}}, \hat{\Sigma}) - m_1 m_2 S_N(k)}{\sqrt{2 m_1 m_2 D_N(k)}}, \quad (3.7)$$

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

This test statistic 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. Note that $\{S_N(k), D_N(k)\}$ are essentially the asymptotic mean and variance of $\mathcal{T}(\hat{\mathbf{a}}, \hat{\mathbf{\Sigma}})$ 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 used in Pham et al. (2003) for finite-order cointegrated case and in Bouhaddioui and Dufour (2008) for the infinite-order case, namely

$$P_M = \sum_{j=-M}^M Q_{\hat{\mathbf{a}}}(j). \quad (3.8)$$

In that case, M is a fixed integer that does not depend on the sample size N . The properties of P_M in the stationary $\text{VAR}(\infty)$ context and cointegrated $\text{IVAR}(\infty)$ are studied respectively in Bouhaddioui and Roy (2006b) and Bouhaddioui and Dufour (2008). As it will be seen below, many kernels k yield tests that are more powerful than P_M .

In the case of testing independence, under some conditions on the smoothing parameter M and if the kernel k verifies the Assumption 3.1, one sees easily that

$$M^{-1}S_N(k) \rightarrow S(k), \quad M^{-1}D_N(k) \rightarrow D(k),$$

where

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

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

$$\mathcal{Q}_N^* = \frac{\mathcal{T}(\hat{\mathbf{a}}, \hat{\mathbf{\Sigma}}) - Mm_1m_2S(k)}{\sqrt{2Mm_1m_2D(k)}}. \quad (3.9)$$

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

$$\mathcal{T}(\hat{\mathbf{a}}, \hat{\mathbf{\Sigma}}) = N \sum_{j=1-N}^{N-1} k^2(j/M) \mathbf{c}_{\hat{\mathbf{a}}}^{(12)}(j)' \left[\mathbf{C}_{\hat{\mathbf{a}}}^{(22)}(0)^{-1} \otimes \mathbf{C}_{\hat{\mathbf{a}}}^{(11)}(0)^{-1} \right] \mathbf{c}_{\hat{\mathbf{a}}}^{(12)}(j)$$

with $\mathbf{c}_{\hat{\mathbf{a}}}^{(12)}(j) = \text{vec} \left[\mathbf{C}_{\hat{\mathbf{a}}}^{(12)}(j) \right]$.

We start by defining the pseudo-statistic

$$\mathcal{T}(\mathbf{a}, \boldsymbol{\Sigma}) = N \sum_{j=1-N}^{N-1} k^2(j/M) \mathbf{c}_{\mathbf{a}}^{(12)}(j)' (\boldsymbol{\Sigma}_2^{-1} \otimes \boldsymbol{\Sigma}_1^{-1}) \mathbf{c}_{\mathbf{a}}^{(12)}(j),$$

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

$$\mathcal{T}(\hat{\mathbf{a}}, \boldsymbol{\Sigma}) = N \sum_{j=1-N}^{N-1} k^2(j/M) \mathbf{c}_{\hat{\mathbf{a}}}^{(12)}(j)' (\boldsymbol{\Sigma}_2^{-1} \otimes \boldsymbol{\Sigma}_1^{-1}) \mathbf{c}_{\hat{\mathbf{a}}}^{(12)}(j).$$

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

$$\begin{aligned} \mathcal{Q}_N &= \frac{\mathcal{T}(\hat{\mathbf{a}}, \hat{\boldsymbol{\Sigma}}) - m_1 m_2 S_N(k)}{\sqrt{2m_1 m_2 D_N(k)}} \\ &= \frac{\mathcal{T}(\mathbf{a}, \boldsymbol{\Sigma}) - m_1 m_2 S_N(k)}{\sqrt{2m_1 m_2 D_N(k)}} + \frac{\mathcal{T}(\hat{\mathbf{a}}, \boldsymbol{\Sigma}) - \mathcal{T}(\mathbf{a}, \boldsymbol{\Sigma})}{\sqrt{2m_1 m_2 D_N(k)}} + \frac{\mathcal{T}(\hat{\mathbf{a}}, \hat{\boldsymbol{\Sigma}}) - \mathcal{T}(\hat{\mathbf{a}}, \boldsymbol{\Sigma})}{\sqrt{2m_1 m_2 D_N(k)}}. \end{aligned} \quad (3.10)$$

Since the quantity $\mathcal{T}(\mathbf{a}, \boldsymbol{\Sigma})$ depends only on the stationary process \mathbf{a} , the result of Lemma 3.1 in Bouhadoui and Roy (2006a) is still valid. Thus, we conclude that

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

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

Proposition 3.2 APPROXIMATION OF THE PSEUDO-STATISTIC. *Let $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ be two multivariate processes that satisfy the IVAR(∞) model (2.8) with $\mathbb{E}|a_{i,t}^{(h)} a_{j,t}^{(h)} a_{k,t}^{(h)} a_{l,t}^{(h)}| < \gamma_4 < \infty$, $1 \leq i, j, k, l \leq m_h$, and suppose that Assumption 3.1 hold. Let $M = M(N) \rightarrow \infty$, $M/N \rightarrow 0$ when $N \rightarrow \infty$ and let $p_N^{(h)}$, $h = 1, 2$, satisfy the following conditions:*

$$(i) \ p_N^{(h)} = o\left(\frac{N^{1/2}}{M^{1/4}}\right),$$

$$(ii) N \sum_{j=p_N^{(h)}+1}^{\infty} \|\Phi_j^{(h)}\|^2 = o\left(\frac{N^{1/2}}{M^{1/4}}\right).$$

If the processes $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$ are independent, we have

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

Note that the two conditions (i) and (ii) imply that the order $p_N^{(h)}$ satisfies Assumption 2.1.

Proposition 3.3 ASYMPTOTIC EQUIVALENCE OF THE TEST STATISTIC. *Under the assumptions of Proposition 3.2, we have*

$$\frac{\mathcal{T}(\hat{\mathbf{a}}, \hat{\Sigma}) - \mathcal{T}(\hat{\mathbf{a}}, \Sigma)}{\sqrt{2m_1m_2D_N(k)}} \xrightarrow{p} 0.$$

Our main result is stated in the following theorem. The proof is based on the results of the two propositions above.

Theorem 3.4 NULL ASYMPTOTIC DISTRIBUTION. *Under the assumptions of Proposition 3.2, the statistic \mathcal{Q}_N defined by (3.7) has an asymptotic normal distribution, i.e. $\mathcal{Q}_N \xrightarrow{L} \mathcal{N}(0, 1)$.*

3.2. Causality tests

For the hypotheses \mathcal{H}_0^+ and \mathcal{H}_0^- , we consider the test statistics $\mathcal{T}^+(\hat{\mathbf{a}}, \hat{\Sigma})$ and $\mathcal{T}^-(\hat{\mathbf{a}}, \hat{\Sigma})$, which are obtained by considering $Q_{\hat{\mathbf{a}}}^{(12)}(j)$ associated with either positive or negative lags:

$$\mathcal{T}^+(\hat{\mathbf{a}}, \hat{\Sigma}) = \sum_{j=1}^{N-1} k^2(j/M) Q_{\hat{\mathbf{a}}}^{(12)}(j) \quad \text{and} \quad \mathcal{T}^-(\hat{\mathbf{a}}, \hat{\Sigma}) = \sum_{j=1-N}^{-1} k^2(j/M) Q_{\hat{\mathbf{a}}}^{(12)}(j)$$

For the hypothesis \mathcal{H}_0^+ and \mathcal{H}_0^- , we have:

$$\mathcal{Q}_N^+ = \frac{\mathcal{T}^+(\hat{\mathbf{a}}, \hat{\Sigma}) - m_1m_2S_N^+(k)}{\sqrt{2m_1m_2D_N^+(k)}}, \quad (3.11)$$

$$\mathcal{Q}_N^- = \frac{\mathcal{T}^-(\hat{\mathbf{a}}, \hat{\Sigma}) - m_1m_2S_N^-(k)}{\sqrt{2m_1m_2D_N^-(k)}}, \quad (3.12)$$

where, by symmetry,

$$S_N^+(k) = S_N^-(k) = \sum_{j=1}^{N-1} \left(1 - \frac{j}{N}\right) k^2(j/M), \quad (3.13)$$

$$D_N^+(k) = D_N^-(k) = \sum_{j=1}^{N-2} \left(1 - \frac{j}{N}\right) \left(1 - \frac{j+1}{N}\right) k^4(j/M). \quad (3.14)$$

From theorem 3.4, we can derive the two following results on testing noncausality between the two processes.

Proposition 3.5 ASYMPTOTIC DISTRIBUTION OF THE POSITIVE CAUSALITY TEST STATISTICS. *Under the assumptions of Theorem 3.4, if the processes $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$ satisfy \mathcal{H}_0^+ , the test statistic \mathcal{Q}_N^+ has an asymptotic standard normal distribution, i.e. $\mathcal{Q}_N^+ \xrightarrow{L} \mathcal{N}(0, 1)$.*

Proposition 3.6 ASYMPTOTIC DISTRIBUTION OF THE NEGATIVE CAUSALITY TEST STATISTICS. *Under the assumptions of Theorem 3.4, if the processes $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$ satisfy \mathcal{H}_0^- , the test statistic \mathcal{Q}_N^- defined respectively by (3.12) has an asymptotic standard normal distribution, i.e. $\mathcal{Q}_N^- \xrightarrow{L} \mathcal{N}(0, 1)$.*

The proof of these two propositions 3.5 and 3.6 can be deduced easily from the proof of theorem 3.4. We thus see that the statistics for testing noncausality follow exactly the same $\mathcal{N}(0, 1)$ asymptotic distribution under the null hypothesis as the statistics for testing the absence of correlation between the two series. The only adjustments required involve the centering and scale parameters used for building the test statistics.

4. Consistency of the generalized tests

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

Assumption 4.1 *The two innovation processes $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$ are jointly fourth-order stationary and their cross-correlation structure is such that $\mathbf{\Gamma}_{\mathbf{a}}^{(12)}(j) \neq \mathbf{0}$ for at least one value of j and*

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

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

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

Theorem 4.2 GLOBAL POWER. *Let $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ be two multivariate processes that satisfy the IVAR(∞) model (2.8) and suppose that their innovation processes $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$ follow Assumption 4.1. If the kernel $k(\cdot)$ satisfies 3.1 and if $p_N^{(h)}$, $h = 1, 2$, satisfy*

$$p_N^{(h)2} = o\left(\frac{N}{M}\right) \quad , \quad \sum_{j=p_N^{(h)}+1}^{\infty} \|\Phi_j^{(h)}\|^2 = o(M^{-1}) \quad ,$$

then

$$P[\mathcal{Q}_N > C(N, M)] \rightarrow 1 \tag{4.1}$$

for any sequence of constants $\{C(N, M) = o(N/M^{1/2})\}$.

This theorem entails that the test based on \mathcal{Q}_N is consistent against every alternative for which the sub-processes are dependent: $\mathcal{Q}_N \rightarrow \infty$ with probability approaching 1 under a fixed alternative of dependency. Thus, the slower M grows, the faster \mathcal{Q}_N will approach infinity and the test will be more powerful. To investigate the relative efficiency of \mathcal{Q}_N , one can use the Bahadur's asymptotic slope criterion defined in Bahadur (1960); see also Hong (1996a, 1996c) and Bouhaddioui and Roy (2006a). Similarly to Bouhaddioui and Roy (2006a), we can show that the relative efficiency of kernels k_2 with respect to k_1 when $M = N^\nu$ is given by

$$ARE_B(k_2, k_1) = \left\{ \frac{D(k_1)}{D(k_2)} \right\}^{\frac{1}{2-\nu}} .$$

We can then proceed as in Bouhaddioui (2002) and Hong (1996a, 1996c) to derive the kernel which maximizes the asymptotic slope over some appropriate classes of kernel functions. For example, consider the following class of kernels:

$$\kappa(\tau) = \{k(\cdot) \text{ satisfies Assumption 4.1, } k^{(2)} = \tau^2/2, K(\lambda) \geq 0 \text{ for } \lambda \in (-\infty, +\infty)\}$$

where $k^{(2)} = \lim_{z \rightarrow 0} (1 - k(z))/z^2$ and $K(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} k(z) e^{-iz\lambda} dz$. This class contains the Daniel, Parzen and quadratic-spectral kernels among others. Using Theorem 1 of Ghosh and Huang (1991) along with similar to the one in Bouhaddioui (2002), we can see that the Daniell kernel [see Table 2] maximizes the asymptotic slope of \mathcal{Q}_N over $\kappa(\tau)$. As mentioned in Bouhaddioui and Roy (2006a), 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 further analysis of the finite-sample properties of the two test statistics.

5. Simulation study

In the previous section, we studied the asymptotic distribution of the test statistics under the null hypothesis. Here we investigate the finite-sample properties of the proposed test statistics, in particular their exact level and power. To do this, we performed a small Monte Carlo study. In addition to the test statistics discussed in the preceding sections, the nonstationary multivariate version of Haugh's statistic P_M^* previously studied by Pham et al. (2003) was also included:

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

where $Q_{\hat{a}}^{(12)}(j)$ is given by (3.6). The statistic P_M^* is a slightly modified version of P_M defined by (3.8).

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 **VAR**(2), **VARMA**(1,1) and **VAR** $_{\delta}$ (1) models described in Table 1. In the first two models, the two subprocesses $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are independent bivariate **VAR**(2) or **VARMA**(1,1) and served for the level study and the corresponding submodels are partially nonstationary and invertible. 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 (3.7) and (3.9), we used the four kernels described in Table 2. For each kernel, the following three truncation values M were employed: $M = [\ln(N)]$, $[3N^{0.2}]$ and $[3N^{0.3}]$ ($[a]$ denotes the integer part of a). These rates are discussed in Hong (1996a, p. 849). They lead respectively to $M = 5, 8, 12$ for the series length $N = 100$, and to $M = 5, 9, 15$ for $N = 200$. The same truncation values were used for P_M^* .

In the level study, 5000 independent realizations were generated from both models **VAR**(2) and **VARMA**(1,1) 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 $\mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{a}})$ pseudo-random vectors using the Cholesky decomposition. Second, the \mathbf{X}_t values were obtained by directly solving the model difference equation.
2. For the **VAR**(2) model, the least squares estimates of the coefficients of the true models were obtained using the procedure described in Reinsel (1993). The autoregressive order was obtained by minimizing the AIC criterion for $p \leq P$, where P is set to $N^{1/3}$. With the **VARMA**(1,1), each subseries was approximated by a possible high order **VAR** model. The value of the **VAR** order was obtained by minimizing Hannan-Quinn criterion using conditional least square estimation. The residual series $\{\hat{\mathbf{a}}_t^{(h)}\}$, $h = 1, 2$, were cross-correlated by computing the $\mathbf{R}_{\hat{\mathbf{a}}}^{(12)}(j)$'s as defined by (3.3).
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 $\chi_{4(2M+1)}^2$ critical values.
4. Finally, for each model, each series length and 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 **VAR** $_{\delta}$ (1) model with different values of δ .

5.2. Level

5.2.1. Gaussian innovations

Results from the level study are presented in Table 3. For both **VAR**(2) and **VARMA**(1, 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. These results are similar to the stationary case. At the 1% and 10% nominal levels, both statistics have a small tendency to under or over-reject. There is no significant difference between the kernels. The best approximation is obtained with the Bartlett and Bartlett-Priestley kernels and the less good one corresponds

to the Parzen kernel. With the Bartlett kernel, the empirical size is always within two standard errors of the nominal size. For the truncated uniform kernel, the size of \mathcal{Q}_N and \mathcal{Q}_N^* are very close to the size of P_M^* , which is normal since \mathcal{Q}_N and \mathcal{Q}_N^* are linear transformations of P_M and P_M^* is a finite-sample version 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.

5.2.2. Non-Gaussian innovations

Here, we propose another simulation experiment with innovations having a multivariate contaminated normal distribution. Let us consider the following model

$$p\mathcal{N}_m(\mathbf{0}, \mathbf{\Gamma}) + (1 - p)\mathcal{N}_m(\mathbf{0}, \mathbf{A})$$

to denote the m -dimensional contaminated normal distribution in which the $\mathcal{N}_m(\mathbf{0}, \mathbf{\Gamma})$ distribution is contaminated with probability $1 - p$, by the $\mathcal{N}_m(\mathbf{0}, \mathbf{A})$ distribution. We can verify that the fourth-order cumulants of this distribution depends on p , $\mathbf{\Gamma}$, \mathbf{A} and is different from zero. Thus, we consider in this part of the simulation two innovations series $\{\mathbf{a}_t^{(1)}\}$ and $\{\mathbf{a}_t^{(2)}\}$ generated independently according to the following two distributions:

$$p_1\mathcal{N}_{m_1}(\mathbf{0}, \mathbb{I}_{m_1}) + (1 - p_1)\mathcal{N}_{m_1}(\mathbf{0}, \mathbf{\Omega}_a^{(1)}) , p_2\mathcal{N}_{m_2}(\mathbf{0}, \mathbb{I}_{m_2}) + (1 - p_2)\mathcal{N}_{m_2}(\mathbf{0}, \mathbf{\Omega}_a^{(2)})$$

with

$$\mathbf{\Omega}_a^{(1)} = \begin{bmatrix} 25 & 5 \\ 5 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{\Omega}_a^{(2)} = \begin{bmatrix} 25 & 7.5 \\ 7.5 & 4 \end{bmatrix}. \quad (5.2)$$

Simulations were made for different pair of values for (p_1, p_2) and for the two models of Table 1 where $\Sigma_a^{(1)}$ and $\Sigma_a^{(2)}$ are now the covariance matrices of the two contaminated normal distributions given in (5.2). The results in the Table 4 are obtained by using $(p_1, p_2) = (0.7, 0.9)$, since the results for the other values of (p_1, p_2) gave similar results. Finally, we see from the Table 4 that the non-normality of the innovations does not significantly affect the behavior of the test statistic \mathcal{Q}_N with the associate kernel function and truncation parameter for the two sizes $N = 100$ and $N = 200$.

5.3. Power

The results are given in Table 5. With the $\text{VAR}_\delta(1)$, 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 of 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 Q_N^* since Q_N and Q_N^* have a similar behavior with respect to the kernels and the truncation values.

The following observations are made from Table 5. First, power decreases as M increases. Indeed, the model considered here is characterized by the lag 0 serial correlation. 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 kernel provided the powerful test, while the Parzen, Bartlett and Bartlett-Priestley kernels led to similar powers for the test Q_N^* . However, the power of Q_N^* with the truncated uniform kernel is much smaller and is comparable to the power of P_M^* . At least for the chosen model, the new tests Q_N or Q_N^* with kernels other than the truncated uniform preferred to the nonstationary multivariate version of Haugh's test P_M^* . Finally, the power of all tests increases when the sample size varies from 100 to 200.

6. Canada/US monetary interactions

We will now study 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 two economies. The data sources with the corresponding CANSIM series numbers are given in Table 1 of Racette and Raynauld (1992). The Canadian economic indicators are gross domestic production (GDP) in constant 1982 dollars, the implicit price index of gross domestic production (GDPI), the nominal short-term interest rate (TX.CA), and a monetary aggregate (M1). The other three variables represent American real 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 natural logarithm of M1 was taken in order to stabilize its variance. These data were first analyzed in El Himdi and Roy (1997), who considered first differences of the series to achieve stationarity, while Pham et al. (2003) analyzed the same data (undifferenced) and applied the tests P_M and P_M^* directly to the series. Since we also work directly with the original series, we will reproduce the results of Pham et al. (2003) to compare them to our test

statistics Q_N and Q_N^* .

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

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

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 have 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 selected by minimizing the AIC criterion. For both series, using the diagnostic checks of Tiao and Box (1981), this procedure led to $p = 3$. The roots of the determinant of the autoregressive polynomial were computed for each model. The smallest roots were respectively 1.002, 1.0504, 1.0532 for Canada and 0.9974, 1.0474 for the United States.

The statistics

$$Q_{\hat{\mathbf{a}}}(j)^* = \frac{N}{N - |j|} Q_{\hat{\mathbf{a}}}(j) \quad (6.1)$$

where $Q_{\hat{\mathbf{a}}}(j)$ is defined by (3.6), are displayed in Figure 1. At level $\alpha = 0.05$, the asymptotic critical value for testing the null hypothesis \mathcal{H}_0 of non-correlation between $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$ against the alternative $\mathcal{H}_{1j} : \boldsymbol{\rho}_{\hat{\mathbf{a}}}^{(12)}(j) \neq \mathbf{0}$ is 21.02 and only one cross-correlation vector $\boldsymbol{\rho}_{\hat{\mathbf{a}}}^{(12)}(j)$ for $j = 0$ significantly differs from zero. Figure 1 suggests 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 $Q_{\hat{\mathbf{a}}}(0)$ based on the cross-correlation matrix at lag 0. The p-values of the portmanteau test P_M^* for \mathcal{H}_0 are also reported in Table 6 for $M = 1, \dots, 12$. At the 0.05 significance level, \mathcal{H}_0 is rejected only for values of M such that $M \leq 4$. These results are similar to those in El Himdi and Roy (1997) and Pham et al. (2003).

With the new tests statistics, the values of the global test statistics Q_N and Q_N^* with the corresponding p-values are reported in Table 9 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. At the 5% significance level, the tests based on Q_N and 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 (TR), the conclusion is the same with Q_N^* but Q_N does not reject when $M = 7$

and $M = 11$. Thus, contrary to the conclusions of El Himdi and Roy (1997) and Pham et al. (2003) which reject the correlation between the two series after lag 4, our results confirm the economic point of view which stipulate that the two economies are known to be more correlated even for a long run. This conclusion is also coherent with the simulation study which shows that Q_N and Q_N^* with any of the four kernels DAN, PAR, BAR and PB are considerably more powerful than P_M^* .

Finally, to determine the direction of causality between the two series, we computed the tests statistics Q_N^+ to test \mathcal{H}_0^+ which is $\mathbf{X}^{(2)}$ does not cause $\mathbf{X}^{(1)}$. The empirical significance levels of Q_N^+ are reported in Table 7. We conclude that we reject \mathcal{H}_0 for all values of $M = 1, \dots, 12$ with the three kernels DAN, PZ, and QS. For the truncated kernel, we reject the non-causality for $M \leq 9$. From table 8, all empirical significance levels Q_N^- are greater than 5%, which means that \mathcal{H}_0^- is not rejected for all values of M and for all used kernels. We conclude that $\mathbf{X}^{(2)}$ causes $\mathbf{X}^{(1)}$ unidirectionally, *i.e.* there is a unidirectional Granger causality running from the U.S. variables to the Canadian ones.

7. Conclusion

In this paper, we have proposed a semiparametric approach to test the non-correlation (or independence in the Gaussian case) and non-causality between infinite-order cointegrated series $\text{IVAR}(\infty)$. The approach is semiparametric in the sense that if the two series are VARMA, we do not need to separately estimate the *true* model for each of the series. Instead, we fit a vector autoregression to each series and the tests statistics are based on residual cross-correlations at all possible lags. The weights assigned to the lags are defined by a kernel function and a smoothing parameter. Under the hypothesis of independence or non-causality of the two series, the asymptotic normality of the tests statistics are established. The finite-sample properties of the test were investigated by a Monte Carlo experiment which shows that the level is reasonably well controlled for both series lengths 100 and 200. 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 the portmanteau test proposed by Bouhaddioui and Dufour (2008). Further, we applied these tests to a set of Canadian and American macroeconomic and monetary variables used in El Himdi and Roy (1997). We have shown that the choice of truncation parameter has no effect in portmanteau statistics. Thus, the independence hypothesis between the two series is rejected for all possible truncation parameters. Also, an unidirectional Granger causality running from the U.S. variables to the Canadian ones is clearly shown.

Acknowledgements

This work was supported by the William Dow Chair in Political Economy (McGill University), the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Bank of Canada (Research Fellowship), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Institut de finance mathématique de Montréal (IFM2), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (Québec).

A. Appendix: proofs

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.

PROOF OF PROPOSITION 3.2 First, let

$$\Xi \stackrel{def}{=} [\Xi_1 : \dots : \Xi_p : \Xi_{p+1,1}] = [\Psi : \Pi_1 : \dots : \Pi_p] D_p \stackrel{def}{=} \Pi D_p$$

where D_p is a suitable nonsingular transformation matrix containing the unknown matrix C_1 . The ECM representation (2.7) can be written as

$$\Delta \mathbf{X}_t = \Psi_0 \mathbf{X}_{2,t-1} + \sum_{l=1}^p \Xi_l \varepsilon_{t-j} + \Xi_{p+1,1} \varepsilon_{1,t-p-1} + \mathbf{e}_t. \quad (\text{A.1})$$

The matrices Ξ and Ψ_0 are defined in equation (A.2) in Saikkonen (1992). Also, let $\Lambda = [\Xi : \Psi_0]$ and $\mathbf{W}_t = [\mathbf{Y}'_t, \mathbf{X}'_{2,t-1}]$ where $\mathbf{Y}'_t = [\varepsilon'_{t-1}, \dots, \varepsilon'_{t-p}, \varepsilon'_{1,t-p-1}]$.

Consider the following linear transformation $\hat{\mathbf{b}}_t = \Sigma^{-1/2} \hat{\mathbf{a}}_t$. Since $\mathbf{C}_{\hat{\mathbf{b}}}^{(12)}(j) = \Sigma_1^{-1/2} \mathbf{C}_{\hat{\mathbf{a}}}^{(12)}(j) \Sigma_2^{-1/2}$, and using the property $vec(ABC) = (C' \otimes A)vec(B)$, we have that

$$\begin{aligned} \mathcal{T}(\hat{\mathbf{a}}, \Sigma) &= N \sum_{j=1-N}^{N-1} k^2(j/M) \mathbf{c}_{\hat{\mathbf{a}}}^{(12)}(j)' (\Sigma_2^{-1} \otimes \Sigma_1^{-1}) \mathbf{c}_{\hat{\mathbf{a}}}^{(12)}(j) \\ &= N \sum_{j=1-N}^{N-1} k^2(j/M) \mathbf{c}_{\hat{\mathbf{b}}}^{(12)}(j)' \mathbf{c}_{\hat{\mathbf{b}}}^{(12)}(j) = \mathcal{T}_{\hat{\mathbf{b}}}^{(12)}. \end{aligned}$$

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

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

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

Define $\hat{\delta}_t = \mathbf{b}_t^{(1)} - \hat{\mathbf{b}}_t^{(1)}$ and $\hat{\eta}_t = \mathbf{b}_t^{(2)} - \hat{\mathbf{b}}_t^{(2)}$. From (3.2), we have

$$T_N^{(1)} = N \sum_{j=0}^{N-1} k^2(j/M) \|\mathbf{c}_{\hat{\mathbf{b}}}^{(12)}(j) - \mathbf{c}_{\mathbf{b}}^{(12)}(j)\|^2 = N \sum_{j=0}^{N-1} k^2(j/M) \left\| \frac{1}{N} \sum_{t=j+1}^N (\mathbf{b}_t^{(1)} \mathbf{b}_{t-j}^{(2)'} - \hat{\mathbf{b}}_t^{(1)} \hat{\mathbf{b}}_{t-j}^{(2)'}) \right\|^2,$$

and using Cauchy-Schwarz inequality, we obtain

$$T_N^{(1)} = N \sum_{j=0}^{N-1} k^2(j/M) \left\| \frac{1}{N} \sum_{t=j+1}^N (\mathbf{b}_t^{(1)} \hat{\eta}'_{t-j} + \hat{\delta}_t \mathbf{b}_{t-j}^{(2)'} - \hat{\delta}_t \hat{\eta}'_{t-j}) \right\|^2 \leq 4N(T_{1N} + T_{2N} + T_{3N}),$$

with $T_{1N} = \sum_{j=0}^{N-1} k^2(j/M) \left\| \frac{1}{N} \sum_{t=j+1}^N \mathbf{b}_t^{(1)} \hat{\eta}'_{t-j} \right\|^2$, $T_{2N} = \sum_{j=0}^{N-1} k^2(j/M) \left\| \frac{1}{N} \sum_{t=j+1}^N \hat{\delta}_t \mathbf{b}_{t-j}^{(2)'} \right\|^2$ and $T_{3N} = \sum_{j=0}^{N-1} k^2(j/M) \left\| \frac{1}{N} \sum_{t=j+1}^N \hat{\delta}_t \hat{\eta}'_{t-j} \right\|^2$. It suffices to show that the terms T_{jN} , $j = 1, 2, 3$, are $o_p(M^{1/2}/N)$. Now, we can write

$$\begin{aligned} \hat{\delta}_t &= (\hat{\mathbf{b}}_t^{(1)} - \Sigma_1^{-1/2} \mathbf{e}_t^{(1)}) + (\Sigma_1^{-1/2} \mathbf{e}_t^{(1)} - \mathbf{b}_t^{(1)}) \\ &= \Sigma_1^{-1/2} \{(\hat{\mathbf{a}}_t^{(1)} - \mathbf{e}_t^{(1)}) + (\mathbf{a}_t^{(1)})\} \\ &= \Sigma_1^{-1/2} \{(\hat{\Lambda}^{(1)} - \Lambda^{(1)}) \mathbf{W}_t^{(1)} + \boldsymbol{\xi}_t(p_1)\} \end{aligned}$$

Where $\Lambda^{(h)}$ and $\mathbf{W}_t^{(h)}$, $h = 1, 2$ are defined as in (A.1) for each process. $\hat{\Lambda}$ is the LS estimator of Λ and $\boldsymbol{\xi}_t(p_1)$ represents the bias of the VAR(p_1) approximation of $\{\mathbf{X}_t^{(1)}\}$.

The second equality is from Saikkonen and Lütkepohl (1996, page 832). Also, using the result of Proposition 2.2, we deduce that

$$\|\hat{\Lambda}^{(1)} - \Lambda^{(1)}\|^2 = O_p\left(\frac{p_1}{N}\right).$$

By equation 3.15 in Bouhaddioui and Roy (2006a), we have $\mathbb{E}\left(\|\boldsymbol{\xi}_t(p_N^{(h)})\|^2\right) = O\left(\sum_{l=p_N^{(h)}+1}^{\infty} \|\Phi_l^{(h)}\|\right)^2$, $h = 1, 2$. Also, based on the result 3.17 in Bouhaddioui and Roy (2006a) and equation (2.9), we obtain that

$$T_{1N} = \sum_{j=0}^{N-1} k^2(j/M) \left\| \frac{1}{N} \sum_{t=j+1}^N \mathbf{b}_t^{(1)} \hat{\eta}'_{t-j} \right\|^2 = O_p\left(\frac{p_2^2 M}{N^2}\right) \left\{ \frac{1}{M} \sum_{j=0}^{N-1} k^2(j/M) \right\}$$

Since $p_2^2 = o(\frac{N}{M^{1/2}})$, 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

$$\begin{aligned}
T_{3N} &= \sum_{j=0}^{N-1} k^2(j/M) \|N^{-1} \sum_{t=j+1}^N \hat{\boldsymbol{\eta}}_t \hat{\boldsymbol{\delta}}'_{t-j}\|^2 \\
&\leq \| \mathbf{A}^{(1)} - \hat{\mathbf{A}}^{(1)} \|^2 \| \mathbf{A}^{(2)} - \hat{\mathbf{A}}^{(2)} \|^2 \sum_{j=0}^{N-1} k^2(j/M) \|N^{-1} \sum_{t=j+1}^N \mathbf{W}_t^{(1)}(p_1) \mathbf{W}_{t-j}^{(2)}(p_2)'\|^2 \\
&+ \| \mathbf{A}^{(1)} - \hat{\mathbf{A}}^{(1)} \|^2 \sum_{j=0}^{N-1} k^2(j/M) \|N^{-1} \sum_{t=j+1}^N \mathbf{W}_t^{(1)}(p_1) \boldsymbol{\xi}_{t-j}(p_2)'\|^2 \\
&+ \| \mathbf{A}^{(2)} - \hat{\mathbf{A}} \|^2 \sum_{j=0}^{N-1} k^2(j/M) \|N^{-1} \sum_{t=j+1}^N \boldsymbol{\xi}_t(p_1) \mathbf{W}_{t-j}^{(2)}(p_2)'\|^2 \\
&+ \sum_{j=0}^{N-1} k^2(j/M) \|N^{-1} \sum_{t=j+1}^N \boldsymbol{\xi}_t(p_1) \boldsymbol{\xi}_{t-j}(p_2)'\|^2.
\end{aligned}$$

Using the equations 3.19-3.22 in Bouhaddioui and Roy (2006a), the assumptions $p_N^{(h)} = o(N^{1/2}/M^{1/4})$, $N \sum_{l=p_N^{(h)}+1}^{\infty} \|\boldsymbol{\Phi}_l^{(h)}\|^2 = o(N^{1/2}/M^{1/4})$ and the result (2.9), we conclude that $T_{3N} = o_p(M^{1/2}/N)$. Therefore, we obtain

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

Finally, using Cauchy-Schwarz inequality once more, we have

$$|T_N^{(2)}| \leq N \sum_{j=1-N}^{N-1} k^2(j/M) |\langle \mathbf{c}_{\mathbf{b}}^{(12)}(j), \mathbf{c}_{\mathbf{b}}^{(12)}(j) - \mathbf{c}_{\mathbf{b}}^{(12)}(j) \rangle| \leq N \sum_{l=4}^6 T_{lN},$$

with

$$\begin{aligned}
T_{4N} &= \sum_{j=0}^{N-1} k^2(j/M) \|\mathbf{c}_{\mathbf{b}}^{(12)}(j)\| \left\| \frac{1}{N} \sum_{t=j+1}^N \hat{\boldsymbol{\delta}}_t \mathbf{b}_{t-j}^{(2)T} \right\|, \\
T_{5N} &= \sum_{j=0}^{N-1} k^2(j/M) \|\mathbf{c}_{\mathbf{b}}^{(12)}(j)\| \left\| \frac{1}{N} \sum_{t=j+1}^N \mathbf{b}_t^{(1)} \hat{\boldsymbol{\eta}}_{t-j}^T \right\|, \\
T_{6N} &= \sum_{j=0}^{N-1} k^2(j/M) \|\mathbf{c}_{\mathbf{b}}^{(12)}(j)\| \left\| \frac{1}{N} \sum_{t=j+1}^N \hat{\boldsymbol{\delta}}_t \hat{\boldsymbol{\eta}}_{t-j}^T \right\|.
\end{aligned}$$

Thus, it is sufficient to show that the terms T_{jN} , $j = 4, 5, 6$, are $o_p(M^{1/2}/N)$. By conditioning on

$(\mathbf{b}_s^{(2)})_{s=-\infty}^N$ and using Jensen's inequality, we have

$$\begin{aligned}
\mathbb{E} \left(T_{4N} \mid (\mathbf{b}_s^{(2)})_{s=-\infty}^N \right) &\leq \sum_{j=1-N}^{N-1} k^2(j/M) \\
&\times [\mathbb{E}(\{(\frac{1}{N} \sum_{\tau=1}^N \|\mathbf{b}_\tau^{(1)} \mathbf{b}_{\tau-j}^{(2)T}\|)(\frac{1}{N} \sum_{t=j+1}^N \|\hat{\boldsymbol{\delta}}_t \mathbf{b}_{t-j}^{(2)T}\|)\}^2 \mid (\mathbf{b}_s^{(2)})_{s=-\infty}^N)]^{1/2} \\
&\leq \frac{M\Delta}{N^2} \{ \frac{1}{M} \sum_{1-N}^{N-1} k^2(j/M) \} (\frac{1}{N} \sum_{\tau=1}^N \|\mathbf{b}_\tau^{(2)}\|^2)^{1/2} (\frac{1}{N} \sum_{t=1}^N \|\hat{\boldsymbol{\delta}}_t\|^2)^{1/2} \\
&= O_p \{ \frac{Mp_2^{1/2}}{N^{5/2}} \} = o_p(M^{1/2}/N).
\end{aligned}$$

The first equality is obtained by using the conditions on p_2 , $\boldsymbol{\Phi}^{(2)}$, and the assumption of independence of the two innovation series. Then, $T_{4N} = o_p(M^{1/2}/N)$. By symmetry, we have also $T_{5N} = o_p(M^{1/2}/N)$. Finally, from Markov inequality, we have $\sum_{j=1}^{N-1} k^2(j/M) \|\mathbf{c}_b^{(12)}(j)\|^2 = O_p(M/N)$, and using Cauchy-Schwarz inequality and the result for T_{3N} , we obtain that $T_{6N} = o_p(M/N)$. Thus, $T_N^{(2)} = o_p(M^{1/2})$ and the proof of Proposition 3.2 is completed. \square

PROOF OF PROPOSITION 3.5 Let

$$\begin{aligned}
T_b^{(12)+} &= N \sum_{j=1}^{N-1} k^2(j/M) \mathbf{c}_b^{(12)}(j)' \mathbf{c}_b^{(12)}(j) \\
&= N \sum_{j=1}^{N-1} k^2(j/M) \text{tr}[\mathbf{C}_b^{(12)}(j)^T \mathbf{C}_b^{(12)}(j)]
\end{aligned}$$

Using definition (3.2), we have

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

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

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

$$W_N^* = 2N^{-1} \sum_{j=1}^{N-2} k^2(j/M) \sum_{t=j+2}^N \sum_{s=j+1}^{t-1} \pi_{jts}^{(12)},$$

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

$$\mathcal{T}_{\mathbf{b}}^{(12)} = H_N + W_N^*$$

The proof of Proposition 3.5 can be completed by proving the following two lemmas.

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

PROOF OF LEMMA A.1 First, we note that $\mathbb{E}(H_N) = m_1 m_2 S_N(k)$ since

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

Under the assumption of independence and since $\mathbb{E}(\|\mathbf{b}_t^{(h)}\|^2) = m_N^h$, for $h = 1, 2$, we have

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

Also, using Minkowski inequality, we obtain

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

Given assumption 3.1 and since $M \rightarrow \infty$ as $N \rightarrow \infty$, we have $M^{-1} \sum_{j=0}^{N-1} k^2(j/M) \rightarrow \int_0^\infty k^2(z) dz < \infty$ and thus, $\mathbb{E}(H_N - \mathbb{E}H_N)^2 = O(M^2/N)$. Since $M^{-1} D_N^+(k) \rightarrow D^+(k) = \int_0^\infty k^4(z) dz$ as $N \rightarrow \infty$ and $M/N \rightarrow 0$, we have $\sigma^2(N) = 2m_1 m_2 M D^+(k) \{1 + o(1)\} = O(M)$ and $\sigma^{-1}(N)\{H_N - m_1 m_2 S_N^+(k)\} \xrightarrow{p} 0$. \square

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

PROOF OF LEMMA A.2 Let $W_{Nt} = 2 \sum_{s=2}^{t-1} \sum_{j=1}^{s-1} k^2(j/M) \boldsymbol{\pi}_{jts}^{(12)}$. Using the property $\sum_{j=1}^{N-2} \sum_{t=j+2}^N \sum_{s=j+1}^{t-1} = \sum_{t=3}^N \sum_{s=1}^{t-1} \sum_{j=1}^{s-1}$, it can be shown that

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

By the hypothesis of non-causality between 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 behavior of W_N^* is determined by the one of $W_N = N^{-1} \sum_{t=3}^N W_{Nt}$. Also, using the same techniques used in Lemma A.3 of Bouhaddioui and Roy (2006a), under the assumption of independence and the hypothesis \mathcal{H}_0^+ , the exact variance of W_N is given by:

$$\text{Var}(W_N) = \sigma^2(N) \text{ where } \sigma^2(N) = 2m_1m_2D_N^+(k).$$

To continue the proof of Lemma A.2, we note that $\{(W_{Nt}, \mathfrak{F}_t); t \in \mathbb{Z}\}$, where \mathfrak{F}_t is the σ -algebra generated by $\{(\mathbf{b}_s^{(1)}, \mathbf{b}_s^{(2)})^T, s \leq t\}$, is a martingale difference since $\mathbb{E}(W_{Nt} | \mathfrak{F}_{t-1}) = 0$. As in Hong(1996b, 1996c), the asymptotic normality of W_N^* follows from the martingale central limit theorem 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.3 $\sigma^{-2}(N) \frac{1}{N^2} \sum_{t=3}^N \mathbb{E}[W_{N,t}^2 \mathbb{I}\{|W_{N,t}| > \epsilon \sigma(N)\}] \rightarrow 0, \forall \epsilon > 0$.

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

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

The second inequality follows by applying the inequality $\mathbb{E}(\sum_{i=1}^n Y_i)^4 \leq 3 \{ \sum_{i=1}^n [\mathbb{E}(Y_i^4)]^{1/2} \}^2$ where the sequence of random variables $\{Y_i\}$ verifies $\mathbb{E}(Y_i) = 0$ 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

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

Thus, we obtain that $\mathbb{E}(W_{Nt}^4) \leq \Delta t^2 M^2 = \mathcal{O}(t^2 M^2)$. Since $\sigma^{-4}(N) = O(M^{-2})$ and $\sum_{t=3}^N \mathbb{E}(W_{Nt}^4) \leq \Delta \sum_{t=3}^N t^2 M^2 \leq \Delta N^3 M^2$, then

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

and the proof of Lemma A.3 is completed. \square

Lemma A.4 $\sigma^{-2}(N) \frac{1}{N^2} \sum_{t=3}^N \ddot{W}_{Nt}^2 \xrightarrow{p} 1$, where $\ddot{W}_{Nt}^2 = \mathbb{E}(W_{Nt}^2 | \mathfrak{F}_{t-1})$.

PROOF OF LEMMA A.4 To prove this lemma, it is sufficient to show that $\sigma^{-4}(N) \text{var}(N^{-2} \sum_{t=3}^N \ddot{W}_{Nt}^2) \rightarrow 0$. By definition of \ddot{W}_{Nt}^2 , we can write

$$\ddot{W}_{Nt}^2 = 4 \text{tr} \left\{ \mathbb{E} \left[\sum_{s=2}^{t-1} G_{ts}^{(2)} \mathbf{b}_s^{(1)T} \mathbf{b}_t^{(1)} \mathbf{b}_t^{(1)T} \sum_{s_1=2}^{t-1} \mathbf{b}_{s_1}^{(1)} G_{ts_1}^{(2)} | \mathcal{F}_{t-1} \right] \right\}$$

let $\lambda_{Nt} = \sum_{s=2}^{t-1} G_{ts}^{(2)} \mathbf{b}_s^{(1)T}$. Since $\text{tr}(AB) \leq \text{tr}(A)\text{tr}(B)$ and that the processes $\{\mathbf{b}_t^{(h)}\}$, $h = 1, 2$, satisfy the assumption of independence, we have

$$\ddot{W}_{Nt}^2 \leq 4 \text{tr} \{ \mathbb{E}[\lambda_{1Nt} \lambda_{1Nt}^T | \mathcal{F}_{t-1}] \} \text{tr} \{ \mathbb{E}[\mathbf{b}_t^{(1)} \mathbf{b}_t^{(1)T} | \mathcal{F}_{t-1}] \} \leq 4 \ddot{W}_{1Nt}^2,$$

where $\ddot{W}_{1Nt}^2 = m_1 \|\lambda_{Nt}\|^2$. The second inequality follows since by conditioning on \mathcal{F}_{t-1} , the terms λ_{Nt} becomes constant. Thus, to prove the lemma, it is sufficient to show that $M^{-2} \text{var}(N^{-2} \sum_{t=3}^N \ddot{W}_{1Nt}^2) \rightarrow 0$ which was done in Lemma A.5 in Bouhaddioui and Roy (2006a).

PROOF OF PROPOSITION 3.3 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 $\mathbf{C}_{\hat{\mathbf{a}}}^{(hh)}(0) - \boldsymbol{\Sigma}_N^h = O_p(N^{-1/2})$, (see Lütkepohl and Saikkonen (1997, p.133)), 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,

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

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

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

By an argument similar to the one used in the second part of the proof of Proposition 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

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

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 3.3 is completed. \square

PROOF OF THEOREM 4.2 First, we note that the statistic \mathcal{Q}_N is a normalized version of $\mathcal{T}(\hat{\mathbf{a}}, \hat{\boldsymbol{\Sigma}})$ 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. Thus, the statistic \mathcal{Q}_N can be expressed as

$$\mathcal{Q}_N = \frac{N \|\mathbf{s}_{\hat{\mathbf{a}}}^{(12)}\|_2^2 - m_1 m_2 S_N(k)}{\sqrt{2m_1 m_2 D_N(k)}},$$

where $\mathbf{s}_{\hat{\mathbf{a}}}^{(12)}$ is the estimator of the cross-coherency function between the two innovations processes given by

$$\|\mathbf{s}_{\hat{\mathbf{a}}}^{(12)}\|_2^2 = \sum_{j=-\infty}^{\infty} \boldsymbol{\gamma}_{\hat{\mathbf{a}}}^{(12)}(j)' (\boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_1)^{-1} \boldsymbol{\gamma}_{\hat{\mathbf{a}}}^{(12)}(j).$$

For details, see Section 4 in Bouhaddioui and Roy (2006a). By definition of \mathcal{Q}_N , we can write

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

$$= \frac{\|\mathbf{s}_{\hat{\mathbf{a}}}^{(12)}\|_2^2}{\{2m_1m_2M^{-1}D_N(k)\}^{1/2}} - \frac{N^{-1}S_N(k)}{\{2M^{-1}D_N(k)\}^{1/2}} (m_1m_2)^{1/2}.$$

From (3.13), the last term of the previous equation goes to zero when $M/N \rightarrow 0$ as $N \rightarrow \infty$. Using the linear transformation $\mathbf{b}_t = \boldsymbol{\Sigma}^{-1/2}\mathbf{a}_t$, as in Proposition 3.2, we have $\|\mathbf{s}_{\hat{\mathbf{a}}}^{(12)}\| = \|\mathbf{s}_{\hat{\mathbf{b}}}^{(12)}\|$. Also, since the processes $\mathbf{b}^{(1)}$ and $\mathbf{b}^{(2)}$ are stationary and by Lemma A.7 in Bouhaddioui and Roy (2006a), we have that

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

where $\|\tilde{\mathbf{s}}_{\hat{\mathbf{b}}}^{(12)}\|$ is defined as $\|\mathbf{s}_{\hat{\mathbf{b}}}^{(12)}\|$, the residual series $(\hat{\mathbf{b}}_t^{(1)}, \hat{\mathbf{b}}_t^{(2)})_{t=1}^N$ being replaced by the innovation series $(\mathbf{b}_t^{(1)}, \mathbf{b}_t^{(2)})_{t=1}^N$. Thus, to prove the consistency result (4.1), it is sufficient to verify that $\|\mathbf{s}_{\hat{\mathbf{b}}}^{(12)}\|_2^2 - \|\tilde{\mathbf{s}}_{\hat{\mathbf{b}}}^{(12)}\|_2^2 \xrightarrow{p} 0$, which follows from the following lemma.

Lemma A.5 *Under the assumptions of Theorem 4.2, we have*

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

PROOF OF LEMMA A.5. By definition of $\mathbf{s}_{\hat{\mathbf{b}}}^{(12)}$ and $\tilde{\mathbf{s}}_{\hat{\mathbf{b}}}^{(12)}$, and by similar calculations to those for the proof in Proposition 3.2, we obtain

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

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 Proposition 3.2, we can write

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

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

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

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

$$\frac{1}{N} \sum_{t=j}^N \|\hat{\boldsymbol{\eta}}_t\|^2 \leq \frac{1}{N} \sum_{t=1}^N \{ \|(\mathbf{A}^{(2)} - \hat{\mathbf{A}}) \mathbf{W}_t^{(2)}\|^2 + \|\boldsymbol{\xi}_t(p_2)\|^2 \}.$$

Since $\|\boldsymbol{\Gamma}_\alpha^{(11)}(l)\|$ is uniformly bounded by a positive constant Δ , and the parameters $\{\boldsymbol{\Phi}_l\}$ are a linear function of the original parameters $\{\mathbf{G}_l\}$, then the bias approximation can be bounded by

$$\mathbb{E} \|\boldsymbol{\xi}_t(p_2)\|^2 \leq \Delta \left(\sum_{l=p_2+1}^{\infty} \|\boldsymbol{\Phi}_l^{(2)}\|^2 \right) = o(N^{-1}).$$

See also the result (A.12) in Saikkonen (1992). Under the assumptions on the process \mathbf{b} , on p_2 and on the parameters $(\boldsymbol{\Phi}_l^{(2)})$, we have

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

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

$$\begin{aligned} T_{3N} &= \sum_{j=0}^{N-1} k^2(j/M) \left\| \frac{1}{N} \sum_{t=j+1}^N \hat{\boldsymbol{\delta}}_t \hat{\boldsymbol{\eta}}'_{t-j} \right\|^2 \\ &\leq M \left\{ \frac{1}{M} \sum_{j=0}^{N-1} k^2(j/M) \right\} \left\{ \frac{1}{N} \sum_{t=1}^N \|\hat{\boldsymbol{\delta}}_t\|^2 \right\} \left\{ \frac{1}{N} \sum_{t=1}^N \|\hat{\boldsymbol{\eta}}_t\|^2 \right\}. \end{aligned}$$

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

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

This completes the proof of Lemma A.5 and then Theorem 4.2. □

References

Ahn, S. k. and Reinsel, G. C. (1990), ‘Estimation of partially nonstationary multivariate autoregressive models’, *Journal of the American Statistical Association* **85**, 813–823.

- Akaike, H. (1969), 'Power spectrum estimation through autoregressive model fitting', *Annals of the Institute of Statistical Mathematics* **21**, 407–419.
- 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.
- Bhansali, R. J. (1996), 'Asymptotically efficient autoregressive model selection for multistep prediction', *Annals of the Institute of Statistical Mathematics* **48**, 577–602.
- Boudjellaba, H., Dufour, J.-M. and Roy, R. (1992), 'Testing causality between two vectors in multivariate arma models', *Journal of the American Statistical Association* **87**, 1082–1090.
- Boudjellaba, H., Dufour, J.-M. and Roy, R. (1994), 'Simplified conditions for non-causality between two vectors in multivariate arma models', *Journal of Econometrics* **63**(2), 271–287.
- Bouhaddioui, C. (2002), Tests d'indépendance de deux séries multivariées autorégressives d'ordre infini, PhD thesis, Département de mathématiques et de statistique, Université de Montréal.
- Bouhaddioui, C. and Dufour, J.-M. (2008), 'Tests for non-correlation of two infinite-order cointegrated vector autoregressive series', *Journal of Applied Probability and Statistics* **3**(1), 78–94.
- Bouhaddioui, C. and Roy, R. (2006a), 'A generalized portmanteau test for independence of two infinite order vector autoregressive series', *Journal of Time Series Analysis* **27**(4), 505–544.
- Bouhaddioui, C. and Roy, R. (2006b), 'On the distribution of residual cross-correlations between two uncorrelated infinite order vector autoregressive series', *Statistics and Probability Letters* **76**(1), 58–68.
- Brown, B. (1971), 'Martingale central limit theorems', *Annals of Mathematical Statistics* **42**, 59–66.
- Burnham, K. P. and Anderson, D. R. (2002), *Model selection and multimodel inference : a practical information-theoretic approach*, Springer-Verlag, New York.
- Duchesne, P. (2005), 'Testing for serial correlation of unknown form in cointegrated time series models?', *Annals of the Institute of Statistical Mathematics* **57**, 575–595.
- 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.
- Engle, R. and Granger, C. (1987), 'Co-integration and error correction: Representation, estimation and testing', *Econometrica* **55**, 251–276.
- Geweke, J. (1981), 'The approximate slopes of econometric tests', *Econometrica* **49**, 1427–1442.
- Ghosh, B. K. and Huang, W. M. (1991), 'The power and optimal kernel of the bickel-rosenblatt test for goodness of fit', *Annals of Statistics* **17**, 999–1009.
- Hallin, M., Jurečková, J., Picek, J. and Zahaf, T. (1999), 'Nonparametric tests of independence of two autoregressive time series based on autoregression rank scores', *Journal of Statistical Planning and Inference* **75**, 319–330.
- Hallin, M. and Saidi, A. (2005), 'Testing independence and causality between multivariate ARMA times series', *Journal of Time Series Analysis* **26**, 83–106.
- Hallin, M. and Saidi, A. (2007), 'Optimal tests for non-correlation between multivariate time series', *Journal of The American Statistical Association* **102**(479), 938–951.
- Haugh, L. D. (1976), 'Checking the independence of two covariance-stationary time series: A univariate residual cross-correlation approach', *Journal of the American Statistical Association* **71**, 378–385.
- Hong, Y. (1996a), 'Consistent testing for serial correlation of unknown form', *Econometrica* **64**, 837–864.
- Hong, Y. (1996b), A separate mathematical appendix for testing for independence between two covariance stationary time series, Technical report, Department of Economics and Department of Statistical Sciences, Cornell University, Ithaca, New York.
- Hong, Y. (1996c), 'Testing for independence between two covariance stationary time series', *Biometrika* **83**(3), 615–625.
- Hong, Y. (1998), 'Testing for pairwise serial independence via the empirical distribution function', *Journal of the Royal Statistical Society, Series B* **60**(2), 271–496.
- Hong, Y. (1999), 'Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach', *Journal of the American Statistical Association* **94**(448), 1201–1220.

- Lewis, R. and Reinsel, G. C. (1985), 'Prediction of multivariate time series by autoregressive model fitting', *Journal of Multivariate Analysis* **16**, 393–411.
- Lütkepohl, H. (1985), 'The joint asymptotic distribution of multistep prediction errors of estimated vector autoregressions', *Economics Letters* **17**, 103–106.
- Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.
- Lütkepohl, H. (2001), Vector autoregressions, in B. Baltagi, ed., 'Companion to Theoretical Econometrics', Blackwell Companions to Contemporary Economics, Basil Blackwell, Oxford, U.K., chapter 32, pp. 678–699.
- Lütkepohl, H. and Saikkonen, P. (1997), 'Impulse response analysis in infinite order cointegrated vector autoregressive processes', *Journal of Econometrics* **81**, 127–157.
- Paparoditis, E. (1996), 'Bootstrapping autoregressive and moving average parameter estimates of infinite order vector autoregressive processes', *Journal of Multivariate Analysis* **57**, 277–296.
- Parzen, E. (1974), 'Some recent advances in time series modeling', *IEEE Transactions on Automatic Control* pp. 723–730.
- Pham, D. T., Roy, R. and Cédras, L. (2003), 'Tests for non-correlation of two cointegrated ARMA time series', *Journal of Time Series Analysis* **24**, 553–577.
- Pierce, D. A. and Haugh, L. D. (1977), 'Causality in temporal systems: Characterizations and survey', *Journal of Econometrics* **5**(3), 265–293.
- Pierce, D. A. and Haugh, L. D. (1979), 'The characterization of instantaneous causality, a comment', *Journal of Econometrics* **10**(2), 257–259.
- Priestley, M. B. (1981), *Spectral Analysis and Time Series*, Vol. 1, Academic Press.
- Racette, D. and Raynauld, J. (1992), Un modèle BVAR de prévision de la dépense nominale et d'analyse de la politique monétaire canadienne, in D. Longworth, ed., 'Monetary Seminar: A Seminar Sponsored by the Bank of Canada, May 7-9, 1990', Bank of Canada, Ottawa, pp. 317–325.
- Reinsel, G. C. (1993), *Elements of Multivariate Time Series Analysis*, Springer-Verlag, New York.

- Reinsel, G. C. (1997), *Elements of Multivariate Time Series Analysis*, second edn, Springer-Verlag, New York.
- Saidi, A. (2007), 'Consistent testing for non-correlation of two cointegrated ARMA time series', *Canadian Journal of Statistics* **35**(1), 169–188.
- Saikkonen, P. (1992), 'Estimation and testing of cointegrated systems by an autoregressive approximation', *Econometric Theory* **8**, 1–27.
- Saikkonen, P. and Lütkepohl, H. (1994), Infinite order cointegrated autoregressive processes: Estimation and inference, Technical Report 5, Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin.
- Saikkonen, P. and Lütkepohl, H. (1996), 'Infinite-order cointegrated vector autoregressive processes: Estimation and inference', *Econometric Theory* **12**, 814–844.
- Saikkonen, P. and Luukkonen, R. (1997), 'Testing cointegration in infinite order vector autoregressive processes', *Journal of Econometrics* **81**, 93–126.
- Shao, X. (2008), 'A generalized portmanteau test for independence between two stationary time series', *Econometric Theory* **forthcoming**.
- Tiao, G. and Box, G. (1981), 'Modeling multiple time series with applications', *Journal of the American Statistical Association* **76**, 802–816.
- Yap, S. F. and Reinsel, G. C. (1995), 'Estimation and testing for unit roots in a partially nonstationary vector autoregressive moving average model', *Journal of the American Statistical Analysis* **90**, 253–267.

Table 1. Time series models used in the simulation study

MODELS	EQUATIONS		
VAR(2)	$\begin{bmatrix} \mathbf{X}_t^{(1)} \\ \mathbf{X}_t^{(2)} \end{bmatrix} = \begin{bmatrix} \phi_1^{(1)} & 0 \\ 0 & \phi_1^{(2)} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{t-1}^{(1)} \\ \mathbf{X}_{t-1}^{(2)} \end{bmatrix} + \begin{bmatrix} \phi_2^{(1)} & 0 \\ 0 & \phi_2^{(2)} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{t-2}^{(1)} \\ \mathbf{X}_{t-2}^{(2)} \end{bmatrix} + \begin{bmatrix} \mathbf{a}_t^{(1)} \\ \mathbf{a}_t^{(2)} \end{bmatrix}$		
VARMA(1, 1)	$\begin{bmatrix} \mathbf{X}_t^{(1)} \\ \mathbf{X}_t^{(2)} \end{bmatrix} = \begin{bmatrix} \phi_1^{(1)} & 0 \\ 0 & \phi_1^{(2)} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{t-1}^{(1)} \\ \mathbf{X}_{t-1}^{(2)} \end{bmatrix} + \begin{bmatrix} \psi^{(1)} & 0 \\ 0 & \psi^{(2)} \end{bmatrix} \begin{bmatrix} \mathbf{a}_{t-1}^{(1)} \\ \mathbf{a}_{t-1}^{(2)} \end{bmatrix} + \begin{bmatrix} \mathbf{a}_t^{(1)} \\ \mathbf{a}_t^{(2)} \end{bmatrix}$		
VAR$_{\delta}$(1)	$\begin{bmatrix} \mathbf{X}_t^{(1)} \\ \mathbf{X}_t^{(2)} \end{bmatrix} = \begin{bmatrix} \phi_1^{(1)} & 0 \\ 0 & \phi_1^{(2)} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{t-1}^{(1)} \\ \mathbf{X}_{t-1}^{(2)} \end{bmatrix} + \begin{bmatrix} \mathbf{a}_t^{(1)} \\ \mathbf{a}_t^{(2)} \end{bmatrix}$		
Noise covariance matrices			
$\Sigma_{\mathbf{a}} = \begin{bmatrix} \Sigma_{\mathbf{a}}^{(1)} & 0 \\ 0 & \Sigma_{\mathbf{a}}^{(2)} \end{bmatrix}$	$\Sigma_{\mathbf{a},\delta} = \begin{bmatrix} \Sigma_{\mathbf{a}}^{(1)} & \Sigma_{\mathbf{a},\delta}^{(12)} \\ \Sigma_{\mathbf{a},\delta}^{(21)} & \Sigma_{\mathbf{a}}^{(2)} \end{bmatrix}$		
Parameters values			
$\Phi_1^{(1)} = \begin{bmatrix} 0.4 & 0.0 \\ -1.0 & 1.0 \end{bmatrix}$	$\Phi_1^{(2)} = \begin{bmatrix} 1.0 & 0.0 \\ -0.8 & 0.5 \end{bmatrix}$	$\Phi_2^{(1)} = \begin{bmatrix} 0.6 & -0.5 \\ 0.3 & 0.4 \end{bmatrix}$	$\Phi_2^{(2)} = \begin{bmatrix} -0.5 & -0.8 \\ -0.4 & 0.2 \end{bmatrix}$
$\Psi^{(1)} = \begin{bmatrix} -0.2 & 0.3 \\ -0.6 & 1.1 \end{bmatrix}$	$\Psi^{(2)} = \begin{bmatrix} 0.8 & 0.3 \\ 0.1 & 0.6 \end{bmatrix}$	$\Sigma_{\mathbf{a}}^{(1)} = \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix}$	$\Sigma_{\mathbf{a}}^{(2)} = \begin{bmatrix} 1.0 & 0.75 \\ 0.75 & 1.0 \end{bmatrix}$
$\Sigma_{\mathbf{a},\delta}^{(12)} = \begin{bmatrix} 0.1\delta & 0 \\ 0 & 0.05\delta \end{bmatrix}$			

Table 2. Kernels used with the test statistics Q_N and Q_N^*

Truncated Uniform (TR):	$k(z) = \begin{cases} 1, & z \leq 1, \\ 0, & \text{otherwise.} \end{cases}$
Bartlett (BAR):	$k(z) = \begin{cases} 1 - z , & z \leq 1, \\ 0, & \text{otherwise.} \end{cases}$
Daniell (DAN):	$k(z) = \frac{\sin(\pi z)}{\pi z}, \quad z \in \mathbb{R}.$
Parzen (PAR):	$k(z) = \begin{cases} 1 - 6z^2 + 6 z ^3, & \text{if } z \leq 0.5, \\ 2(1 - z)^3, & \text{if } 0.5 \leq z \leq 1, \\ 0, & \text{otherwise.} \end{cases}$
Bartlett-Priestley (BP):	$k(z) = \frac{3}{(\pi z)^2} \left\{ \frac{\sin(\pi z)}{\pi z} - \cos(\pi z) \right\}, \quad z \in \mathbb{R}.$

Table 3. Empirical level (in percentage) of the test Q_N , Q_N^* and P_M^* based on 5000 realizations for different kernels, different truncation values, for the VAR(2) and VARMA(1,1) models

	N	M	$\alpha\%$	Q_N					Q_N^*					P_M^*
				DAN	PAR	BAR	BP	TR	DAN	PAR	BAR	BP	TR	
VAR(2)	100	5	1	0.7	0.6	0.8	0.7	0.6	1.2	0.9	0.7	0.7	1.3	0.7
			5	5.8	3.9	5.7	5.2	4.4	5.9	4.3	5.8	6.1	3.7	4.2
			10	9.6	8.0	9.5	10.6	8.3	10.3	8.8	9.4	10.7	9.0	8.8
		8	1	1.3	0.6	0.9	1.2	0.7	1.4	1.2	1.0	1.5	0.6	0.8
			5	5.6	4.1	5.9	5.6	4.0	5.4	4.0	5.2	4.8	4.0	4.3
			10	10.7	9.2	10.8	10.7	7.4	10.6	9.6	11.0	10.4	8.2	8.4
		12	1	0.8	0.7	0.8	1.2	0.6	1.3	0.8	1.4	1.5	0.7	0.8
			5	5.4	4.8	5.3	5.4	4.2	5.6	4.5	4.9	5.7	4.2	4.5
			10	10.4	8.7	11.2	10.8	7.8	10.8	10.4	11.2	10.5	8.1	8.4
	200	5	1	0.8	1.2	0.8	1.2	0.8	0.7	0.8	1.2	1.3	0.7	0.9
			5	5.7	5.2	5.8	5.5	4.1	5.5	4.2	5.9	5.7	4.4	4.2
			10	9.1	9.2	10.4	10.6	8.3	8.4	10.2	10.6	10.2	8.7	8.9
		9	1	1.2	1.1	0.9	0.8	0.7	1.4	0.9	0.8	1.2	0.7	0.7
			5	6.1	4.3	5.5	5.7	4.4	6.3	4.6	5.5	5.9	4.5	4.1
			10	10.9	9.5	10.5	11.0	7.6	11.2	9.3	10.6	10.7	8.6	9.2
		15	1	1.4	0.8	1.2	1.4	1.2	0.9	1.2	1.4	0.8	0.6	0.6
			5	6.0	4.5	6.2	5.4	4.1	5.8	4.7	5.8	5.6	4.3	4.5
			10	10.6	10.3	11.2	10.6	7.9	11.0	10.5	10.8	10.4	8.2	8.9
VARMA(1,1)	100	5	1	1.3	1.1	0.7	0.8	0.7	1.2	0.7	1.4	1.2	0.6	0.8
			5	5.7	4.7	6.2	4.5	4.3	5.8	4.4	5.8	4.6	3.9	4.3
			10	9.6	8.6	9.3	10.4	8.3	9.6	9.0	9.5	10.8	8.2	8.4
		8	1	1.4	0.7	0.8	1.2	0.7	1.3	0.8	1.2	0.9	0.8	1.3
			5	5.6	4.4	5.9	5.6	3.9	5.4	4.1	5.5	5.5	4.3	5.6
			10	10.6	8.5	11.3	10.6	7.3	9.4	9.0	11.0	10.7	8.0	9.4
		12	1	0.9	1.2	0.7	0.8	0.6	1.1	0.9	0.9	1.3	0.7	1.4
			5	5.4	5.1	6.0	5.6	4.2	5.6	5.4	5.8	5.6	4.1	4.5
			10	9.4	8.8	10.4	10.2	7.9	9.1	8.2	9.1	10.6	7.5	8.3
	200	5	1	0.8	1.3	0.7	0.9	0.7	1.2	0.8	1.2	1.2	0.7	1.3
			5	5.6	4.7	5.4	5.9	4.0	6.2	4.8	5.7	6.3	4.6	5.9
			10	9.0	9.3	10.6	11.0	8.9	10.5	9.2	10.5	9.6	8.2	8.9
		9	1	1.3	0.7	1.2	1.1	0.8	0.9	0.8	1.3	0.8	0.8	0.9
			5	6.1	5.2	4.2	6.1	4.3	5.7	5.1	5.5	6.3	4.3	5.6
			10	9.4	10.5	11.0	10.7	8.4	10.7	9.5	10.8	10.3	8.7	8.9
		15	1	1.4	1.1	0.8	0.9	0.7	1.3	0.9	0.9	0.8	0.7	0.8
			5	6.2	4.6	5.2	6.0	4.3	5.3	5.1	5.3	6.0	4.6	5.5
			10	10.3	10.5	10.8	10.6	7.9	10.7	10.2	11.2	10.7	8.4	9.1

Table 4. Empirical level (in percentage) of the test Q_N , Q_N^* and P_M^* based on 5000 realizations for different kernels, different truncation values, for the **VAR(2)** and **VARMA(1,1)** models with non-Gaussian innovations

	N	M	$\alpha\%$	Q_N					Q_N^*					P_M^*
				DAN	PAR	BAR	BP	TR	DAN	PAR	BAR	BP	TR	
VAR(2)	100	5	1	1.3	0.7	1.2	1.3	0.6	0.8	1.3	0.9	0.8	1.4	1.3
			5	5.4	4.6	5.8	5.3	4.1	5.5	4.4	5.9	5.8	4.0	4.2
			10	9.8	8.4	10.5	10.7	8.2	10.5	9.0	9.1	9.3	8.5	8.9
		8	1	0.7	1.2	0.8	1.3	0.7	1.2	0.8	0.8	1.3	0.7	0.8
			5	6.0	5.4	4.6	5.8	3.8	5.7	4.2	5.6	4.4	4.0	4.2
			10	11.0	9.4	10.6	9.5	8.2	10.8	9.4	10.8	10.6	8.4	8.8
	12	1	1.2	0.9	0.7	1.3	0.7	1.4	1.2	0.8	1.3	0.6	0.8	
		5	5.8	5.6	5.2	5.6	4.0	4.6	4.8	5.3	5.4	3.8	4.2	
		10	11.3	10.9	11.0	10.6	8.4	10.6	9.8	10.8	9.5	8.3	8.8	
	200	5	1	1.2	0.9	0.8	1.3	0.7	0.8	1.3	1.1	0.8	0.8	1.2
			5	6.0	5.8	5.4	5.6	3.9	6.1	5.9	5.5	5.3	4.0	4.4
			10	10.6	9.0	10.2	10.4	8.4	9.4	10.8	11.0	10.6	8.4	9.2
		9	1	0.7	0.9	0.7	0.8	0.8	1.3	0.7	0.7	1.1	0.8	0.8
			5	5.8	5.6	5.2	4.7	4.2	6.0	4.8	5.8	5.8	4.2	4.6
			10	11.2	9.3	9.6	10.6	8.8	11.4	9.7	10.3	10.9	8.6	9.4
	15	1	1.3	1.1	0.8	0.7	0.7	1.1	1.3	0.9	0.8	0.6	0.7	
		5	5.6	5.8	6.0	5.6	4.2	5.6	4.4	6.0	6.2	4.1	4.6	
		10	11.2	10.6	10.2	10.8	8.6	11.0	10.8	10.3	10.2	8.6	9.0	
VARMA(1,1)	100	5	1	0.8	1.2	1.3	0.7	0.6	1.1	0.8	1.2	1.2	0.6	0.7
			5	5.9	6.1	5.6	4.4	4.0	5.7	5.9	4.8	4.8	4.0	4.4
			10	10.6	9.2	9.6	11.0	8.5	10.9	10.4	9.2	11.0	8.0	9.0
		8	1	1.4	1.2	1.2	0.8	0.7	1.2	1.4	1.3	0.8	0.7	1.4
			5	6.0	4.2	5.6	5.8	3.8	6.2	4.0	6.1	6.3	4.2	6.0
			10	11.6	9.6	10.4	10.8	8.0	11.2	9.4	11.2	10.6	8.0	9.6
	12	1	0.8	1.3	0.8	0.9	0.7	1.2	1.1	0.9	1.1	0.8	1.3	
		5	5.8	5.3	5.8	6.0	4.4	6.0	5.2	5.4	5.8	4.0	5.8	
		10	10.8	9.2	11.4	10.6	8.1	11.2	9.4	9.3	11.0	8.4	8.8	
	200	5	1	1.1	1.2	0.9	1.3	0.7	1.2	1.3	1.1	0.8	0.8	1.2
			5	6.1	5.4	4.8	6.1	4.2	5.9	4.7	5.4	6.0	4.4	5.8
			10	10.6	10.3	11.3	11.5	8.4	11.3	10.4	11.0	10.8	8.4	9.2
		9	1	1.3	1.2	0.9	1.2	0.8	1.2	1.3	1.1	0.9	0.7	1.3
			5	5.9	5.9	4.6	5.4	4.1	5.7	6.1	5.2	5.8	4.4	5.8
			10	11.4	10.8	10.6	10.6	8.8	11.2	10.8	10.4	9.8	8.6	9.3
	15	1	0.9	1.3	0.8	1.2	0.8	1.3	1.2	1.3	1.1	0.7	1.3	
		5	5.4	5.8	6.2	5.6	4.0	5.5	5.6	5.8	5.4	4.2	5.8	
		10	11.0	10.8	9.8	10.2	8.2	10.6	10.6	10.2	10.4	8.6	9.3	

Table 5. 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 $\text{VAR}_\delta(1)$ data generation process with $\delta = 2$

N	M	$\alpha\%$	\mathcal{Q}_N^*					P_M^*
			DAN	PAR	BAR	BP	TR	
100	5	1	57.3	53.5	54.6	52.6	35.3	24.6
		5	63.2	60.1	56.4	58.6	36.8	26.8
		10	72.6	70.8	62.5	64.3	38.2	27.5
	8	1	49.6	46.1	51.4	48.0	27.5	22.6
		5	58.4	53.2	55.8	51.6	31.2	23.8
		10	63.7	60.8	62.6	61.7	34.6	25.8
	12	1	43.6	38.5	41.8	42.6	23.3	18.9
		5	50.2	44.7	40.3	43.0	26.4	21.2
		10	56.8	50.6	48.8	46.5	28.8	23.7
200	5	1	78.4	74.5	74.8	76.2	54.8	50.6
		5	85.6	82.6	81.6	85.8	56.4	54.1
		10	93.4	89.5	87.5	90.2	60.4	56.8
	9	1	69.5	65.2	63.0	66.8	42.4	40.7
		5	75.6	76.6	72.4	78.2	46.2	44.6
		10	80.8	78.5	77.6	82.8	50.4	46.4
	15	1	56.8	52.4	54.8	56.1	36.8	32.8
		5	60.1	57.4	53.2	60.1	40.2	35.0
		10	64.8	54.4	54.2	62.6	44.8	40.4

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

M	P_M^*	α_M	M	P_M^*	α_M
1	52.34	0.038	7	205.62	0.092
2	80.24	0.042	8	221.77	0.187
3	108.34	0.038	9	250.60	0.145
4	135.70	0.037	10	271.59	0.189
5	159.38	0.052	11	284.56	0.348
6	181.15	0.082	12	306.46	0.386

Table 7. Empirical significance level of the global statistics \mathcal{Q}_N^+ defined by (3.11) for $M = 1, \dots, 12$ using TR, DAN, PZ and QS kernels

M	TR	DAN	PZ	QS	M	TR	DAN	PZ	QS
1	0.0301	0.0342	0.0401	0.0403	7	0.0421	0.0267	0.0422	0.02955
2	0.0406	0.0447	0.0420	0.0465	8	0.0332	0.0288	0.0337	0.02972
3	0.0191	0.0452	0.0218	0.0448	9	0.0538	0.0273	0.0317	0.03028
4	0.0401	0.0429	0.0202	0.0310	10	0.0555	0.0278	0.0341	0.03098
5	0.0438	0.0305	0.0219	0.0279	11	0.0602	0.0279	0.0287	0.03176
6	0.0486	0.0247	0.0210	0.0288	12	0.0899	0.0276	0.0312	0.03263

Table 8. Empirical significance level of the global statistics \mathcal{Q}_N^- defined by (3.12) for $M = 1, \dots, 12$ using TR, DAN, PZ and QS kernels

M	TR	DAN	PZ	QS	M	TR	DAN	PZ	QS
1	0.6711	0.4601	0.5950	0.4935	7	0.5971	0.6941	0.5848	0.7778
2	0.2035	0.4939	0.3875	0.4895	8	0.5984	0.7367	0.6500	0.7884
3	0.4605	0.5722	0.4826	0.6736	9	0.6392	0.7782	0.6731	0.8945
4	0.5028	0.5851	0.6866	0.6733	10	0.6456	0.7990	0.7546	0.7976
5	0.3925	0.6955	0.5916	0.7821	11	0.7598	0.8135	0.6970	0.6450
6	0.4680	0.6985	0.6950	0.7896	12	0.6996	0.7998	0.6790	0.7234

Table 9. Values of the statistics \mathcal{Q}_N and \mathcal{Q}_N^* and their p-values for three kernels and three values of $M = 4, 7$ and 11

Kernels	$M = 4$			$M = 7$			$M = 11$		
	TR	DAN	BP	TR	DAN	BP	TR	DAN	BP
\mathcal{Q}_N	2.072	1.771	2.022	0.915	2.011	1.815	0.534	1.684	1.347
α	0.019	0.038	0.021	0.179	0.022	0.034	0.297	0.046	0.08
\mathcal{Q}_N^*	3.656	2.118	2.474	2.072	2.405	3.732	1.929	1.899	3.660
α	0.0001	0.017	0.007	0.019	0.007	0.0001	0.026	0.028	0.0001

Figure 1. Values of the statistic $Q_{\hat{\alpha}}(j)^*$ defined by (6.1) at different lags j . The horizontal dotted line represents the marginal critical value at the significance level $\alpha = 0.05$

