Tests for non-correlation of two cointegrated ARMA time series

Dinh Tuan Pham[*] Roch Roy[†‡] Lyne Cédras[§]

CRM-2649
January 2000

[*]Laboratoire de Modélisation et Calcul, Institut d'Informatique et Mathématiques Appliquées, Université de Grenoble I, B.P. 53, 38041 Grenoble cedex 9, France

[†]Département de mathématiques et de statistique, and Centre de recherches mathématiques, Université de Montréal, C.P. 6128, succ. Centre-ville, Montréal, QC H3C 3J7, Canada

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

[§]Direction de la santé publique, 1301 Sherbrooke Est, Montréal, QC H2L 1M3, Canada

Abstract

In multivariate time series modelling, we are often led to investigate the existence of a relationship between two time series. Here we generalize the procedure proposed by Haugh (1976) and extended by El Himdi and Roy (1997) for multivariate stationary ARMA time series to the case of cointegrated (or partially nonstationary) ARMA series. The main contribution consists in showing that in the case of two uncorrelated cointegrated time series, an arbitrary vector of residual cross-correlation matrices asymptotically follows the same distribution as the corresponding vector of cross-correlation matrices between the two innovation series. The estimation method from which the residual are obtained can be the conditional maximum likelihood method as discussed in Yap and Reinsel (1995) or some other which has the same convergence rate. From this result, it follows that the considered test statistics, which are based on residual cross-correlation matrices, asymptotically follow chi-square distributions. The finite sample properties, under the null hypothesis, of the test statistics are studied by simulation. Finally, the proposed procedure is applied to a set of economic data.

Key words and phrases. Independence tests; residual cross-correlation; innovation; co-integration; partially nonstationary.

AMS 1985 subject classification. Primary 62M10; Secondary 62P20.

Résumé

Dans plusieurs situations, nous voulons vérifier l'existence de relations entre deux séries chronologiques. Dans cet article, nous généralisons la procédure proposée par Haugh (1976) et adaptée par El Himdi et Roy (1997) à des séries multivariées ARMA stationnaires au cas de séries ARMA cointégrées (ou partiellement non stationnaires). L'objectif premier est de montrer que dans le cas de deux séries cointégrées ARMA qui sont non corrélées, un vecteur quelconque de corrélations croisées résiduelles suit la même loi asymptotique que le vecteur correspondant des corrélations croisées entre les deux séries d'innovations. La méthode d'estimation utilisée peut être la méthode du maximum de vraisemblance conditionnelle (Yap et Reinsel, 1995) ou une autre méthode ayant la même vitesse de convergence. De ce résultat, il découle que les statistiques de test considérées qui sont basées sur les matrices de corrélations croisées résiduelles suivent asymptotiquement des lois khi-deux. Une expérience de simulation a aussi été réalisée afin d'évaluer la qualité de l'approximation de la loi exacte des statistiques de test, sous l'hypothèse nulle, par la loi khi-deux. Finalement, la méthodologie développée est appliquée à un jeu de données économiques.

1 Introduction

In many situations, we want to verify the existence of a relationship between two time series. It is often the case that investigation of such relations is the first attempt at identifying the models which might govern the global system under study. With multivariate nonstationary series, as mentioned by Engle and Granger (1987), Tsay and Tiao (1990) and many others, it is important to be able to directly model the nonstationarity. Indeed, it is not possible in many situations to transform a nonstationary series in a stationary one by taking differences on each component individually, without introducing unnecessary complications, such as rank deficiency in coefficient matrices and noninvertibility of the moving average operator. The main objective of this paper is to develop a test for non-correlation (or independence in the Gaussian case) of nonstationary series that will be valid in the important case of cointegrated series that are often met with economic data. The concept of cointegration introduced by Engle and Granger (1987) is now discussed in most recent textbooks on time series analysis, namely in Lütkepohl (1991), Reinsel (1993), Hamilton (1994), Fuller (1996).

Among the approaches considered for checking the independence of stationary time series, the one that seems most natural for nonstationary series is the residual cross-covariance approach initiated by Haugh (1976) for univariate ARMA time series. In that procedure, ARMA models are first fitted to the two series and the test statistic is based on the serial cross-correlations $r_{\hat{a}}^{(12)}(k)$ between the two resulting residual series. Under the null hypothesis H_0 of independence, Haugh (1976) showed that for any finite set of lags $k_1,\ldots,\ k_m$, the vector $\sqrt{n}(r_{\hat{a}}^{(12)}(k_1),\ldots,\ r_{\hat{a}}^{(12)}(k_m))'$ asymptotically follows a multinormal distribution with mean vector 0 and covariance matrix I, the identity matrix. This result leads to the definition of the portmanteau statistic

$$Q_M = n \sum_{k=-M}^{M} r_{\hat{a}}^{(12)}(k)^2,$$

where $M \leq n-1$ is a fixed integer and n represents the length of the series. Under H_0, Q_M is asymptotically chi-square distributed and H_0 is rejected for large values of Q_M. A robustified version of Haugh's statistic to outliers is described in Li and Hui (1994). Hong (1996) proposed a modification of Haugh's procedure for stationary infinite order autoregressive series in which a finite-order autoregression is fitted to each series and the test statistic is a properly standardized version of the sum of weighted squared residual cross-correlations, with weights defined by a kernel function. Finally, Hallin et al. (1999) introduced a test for independence of two autoregressive time series which is based on autoregressive rank scores.

For multivariate time series, not many results exist at our knowledge. El Himdi and Roy (1997) extended Haugh'approach to multivariate stationary ARMA time series and here, we will follow that approach for nonstationary series. More precisely, it is assumed that the series are partially nonstationary ARMA that is, the determinant of the autoregressive polynomial $\Phi(\cdot)$ has a unit root of multiplicity $m \in \{0,\ldots,d\}$ where d is the dimension of the process and that all its other roots are outside the unit circle. It is also supposed that $\Phi(1)$ has rank $r = d - m$ and that $\Theta(1)$ is non singular where $\Theta(\cdot)$ denotes the moving average polynomial. Under these assumptions, it can be shown that although the components of the process are nonstationary, there are r linearly independent linear combinations of them which are stationary. For $1 \leq r < d$, this property is often referred to as cointegration of rank r and the process is called partially nonstationary.

The main contribution of this paper is to show that in the case of two uncorrelated partially nonstationary time series, an arbitrary vector of residual cross-correlation matrices asymptotically follows the same distribution that the corresponding vector of cross-correlation matrices between the

two innovation series. The estimation method can be the conditional maximum likelihood method of the error correction form of the model as discussed by Ahn and Reinsel (1990) and Reinsel (1993) in the AR case and by Yap and Reinsel (1995) in the ARMA case, or some other which has the same convergence rate. From this result, it follows that the considered test statistics, which are based on residual cross-correlation matrices, asymptotically follow chi-square distributions.

The method of proof is completely different of the one used in El Himdi and Roy (1997). It relies heavily on the error correction form of a partially nonstationary process and on a decomposition of such a process as the sum of a process with stationary increments and of a stationary process. This decomposition is similar to the one obtained in Yap and Reinsel (1995). However the construction of the matrices P_1 and P_2 is different.

The paper is organized as follows. After presenting some preliminary results in Section 2, the main results are given in Section 3. Two test procedures for the hypothesis of non-correlation are discussed. As in El Himdi and Roy (1997), the first one is based on the cross-correlation matrix at a particular lag while the second one is a portmanteau-type statistic that takes into account the cross-correlation matrices at all lags from -M to M, say. Since the proofs are rather long and technical, they are relegated in an Appendix. The finite sample properties, under the null hypothesis, of the test statistics used are studied by simulation in Section 4. Finally, the proposed procedure is applied to a set of American and Canadian data in Section 5.

2 Preliminaries

Let $\{\mathrm{X}^{(h)}(t) : t \in \mathbb{Z}\}$, $h = 1, 2$, be two multivariate processes, of dimension d_1 and d_2 respectively, for which one would like to test for non-correlation. We shall assume that the processes $\{\mathrm{X}^{(h)}(t)\}$ are of zero mean. Further, we suppose that they can be represented by an autoregressive moving average (ARMA) model of order p_h, q_h:

$$\Phi^{(h)}(B)\mathrm{X}^{(h)}(t) = \Theta^{(h)}(B)\mathrm{a}^{(h)}(t), \quad t \in \mathbb{Z}, \tag{2.1}$$

where $\Phi^{(h)}(B) = \mathrm{I}_{d_h} - \sum_{l=1}^{p_h} \Phi_l^{(h)} B^l$ and $\Theta^{(h)}(B) = \mathrm{I}_{d_h} - \sum_{l=1}^{q_h} \Theta_l^{(h)} B^l$, I_d denoting the identity matrix of dimension d, B denoting the backward shift operator (that is $B\mathrm{X}^{(h)}(t) = \mathrm{X}^{(h)}(t-1)$) and $\{\mathrm{a}^{(h)}(t)\}$ being a weak white noise process in the sense that the $\mathrm{a}^{(h)}(t)$ are uncorrelated random vectors with zero mean and the same covariance matrix Ω_h, assumed to be non singular.

We *do not* assume the stationarity of the process $\{\mathrm{X}^{(h)}(t)\}$: the polynomial $\det \Phi^{(h)}(z)$ may have a unit root, possibly multiple (det is the notation for determinant). But it is assumed that there is no other root inside and on the unit circle $|z| = 1$ and furthermore, the polynomial $\det \Theta^{(h)}(z)$ has no root inside and on the unit circle. Then one can express $\{\mathrm{X}^{(h)}(t)\}$ as an autoregressive process of infinite order:

$$\mathrm{a}^{(h)}(t) = \Theta^{(h)}(B)^{-1}\Phi^{(h)}(B)\mathrm{X}^{(h)}(t), \tag{2.2}$$

where the expression $\Theta^{(h)}(B)^{-1}\Phi^{(h)}(B)$ must be understood as a series, which, in term of the complex variable, is absolutely convergent on the unit disk.

The process $\{\mathrm{a}^{(h)}(t)\}$ is often referred to as the innovation process. As we are concerned about testing the non-correlation between the processes $\{\mathrm{X}^{(1)}(t)\}$ and $\{\mathrm{X}^{(2)}(t)\}$, it is clear from (2.2) that this property implies the non-correlation between the innovation processes $\{\mathrm{a}^{(1)}(t)\}$ and $\{\mathrm{a}^{(2)}(t)\}$. However, in the nonstationary case, the observed processes cannot be defined uniquely in term of their innovation processes, so the converse would require some extra conditions.

Proposition 2.1. *The processes* $\{\mathrm{X}^{(1)}(t)\}$ *and* $\{\mathrm{X}^{(2)}(t)\}$ *are non correlated if and only if for some* $P \in \mathbb{Z}$*:* (i) *the random vectors* $\mathrm{X}^{(1)}(P-1),\ldots,\mathrm{X}^{(1)}(P-p_1)$ *and* $\mathrm{X}^{(2)}(P-1),\ldots,\mathrm{X}^{(2)}(P-p_2)$ *are non-correlated and* (ii) $\mathrm{Cov}\{\mathrm{a}^{(1)}(s),\mathrm{a}^{(2)}(t)\}=0$*, for* $s \geq P-q_1$*,* $t \geq P-q_2$*.*

The proof of the above result is straightforward from (2.2) and the fact that one can write $\mathrm{X}^{(h)}(t)$ as a linear combination of $\mathrm{X}^{(1)}(P-1),\ldots,\mathrm{X}^{(1)}(P-p_h)$ and of $\mathrm{a}^{(h)}(P-q_h),\ldots,\mathrm{a}^{(h)}(t)$. This result means that if one assumes that the remote past of the processes $\{\mathrm{X}^{(1)}(t)\}$ and $\{\mathrm{X}^{(2)}(t)\}$ are uncorrelated, then the non-correlation between them is equivalent to that of their innovation processes. Therefore, our approach is to test for the non-correlation between the innovation processes $\{\mathrm{a}^{(1)}(t)\}$ and $\{\mathrm{a}^{(2)}(t)\}$. It seems impossible to test for non-correlation between the remote past of the observed processes, so we simply assume that they are so. The advantage of this approach is that the sample cross-correlation between these innovation processes have a simple limiting distribution, due to the fact that each of them is serially uncorrelated, while in the nonstationary case, the sample cross correlations between $\{\mathrm{X}^{(1)}(t)\}$ and $\{\mathrm{X}^{(2)}(t)\}$ may not even make sense, because $\mathrm{E}[\mathrm{X}^{(1)}(t)\mathrm{X}^{(2)}(t-k)']$ depends not only on k but also on t.

The above approach requires extracting the innovation processes from the observations according to (2.2). Since the model coefficient matrices $\Phi_l^{(h)}$ and $\Theta_l^{(h)}$ are unknown, one will need to estimate them. To this end, one must first ensure that the two ARMA models defined by (2.1) are identifiable, in the sense that the matrices $\Phi_1^{(h)},\ldots,\Phi_{p_h}^{(h)}$ and $\Theta_1^{(h)},\ldots,\Theta_{q_h}^{(h)}$ can be uniquely determined from the "transfer" function $\Phi^{(h)}(z)^{-1}\Theta^{(h)}(z)$. A set of sufficient condition for identifiability is given in Yap and Reinsel (1995), which is based on Dunsmuir and Hannan's 1976 conditions (see also Hannan 1975). Under such conditions, one can consistently estimate the $\Phi_i^{(h)}$, $\Theta_i^{(h)}$ and Ω_h by the (conditional) Gaussian maximum likelihood method as in Yap and Reinsel (1995). Here the *full rank* estimation procedure will be used, since we do not know *a priori* whether $\Phi^{(h)}(1)$ is singular or not and what is its rank. This procedure consists in maximizing

$$L(\dot\Phi_1^{(h)},\ldots,\dot\Phi_{p_h}^{(h)},\dot\Theta_1^{(h)},\ldots,\dot\Theta_{q_h}^{(h)},\dot\Omega_h) = -\frac{1}{2}\left[(n-p_h)\log\det\dot\Omega_h + \sum_{t=p_h+1}^{n}\dot{\mathrm{a}}^{(h)}(t)\dot\Omega_h^{-1}\dot{\mathrm{a}}^{(h)}(t)'\right]$$

with respect to the generic parameters $\dot\Phi_i^{(h)}$, $\dot\Theta_i^{(h)}$ and $\dot\Omega_h$, where $\dot{\mathrm{a}}^{(h)}(t)$ is computed by the recurrence relation

$$\dot{\mathrm{a}}^{(h)}(t) = \begin{cases} \mathrm{X}^{(h)}(t) - \sum_{l=1}^{p_h}\dot\Phi_l^{(h)}\mathrm{X}^{(h)}(t-l) + \sum_{l=1}^{q_h}\dot\Theta_l^{(h)}\dot{\mathrm{a}}^{(h)}(t-l) & \text{if } p_h < t \leq n, \\ 0 & \text{if } t \leq p_h. \end{cases} \tag{2.3}$$

The minimization will be done *without any restriction* on the rank of $\Phi^{(h)}(1)$, but of course under the above invertibility and identifiability conditions. Call $\hat\Phi_1^{(h)},\ldots,\hat\Phi_{p_h}^{(h)}$, $\hat\Theta_1^{(h)},\ldots,\hat\Theta_{q_h}^{(h)}$ and $\hat\Omega_h$, the obtained estimators, one can compute the estimated innovations (or residuals) by the same recurrence relation (2.3), substituting $\dot\Phi_l^{(h)}$, $\dot\Theta_l^{(h)}$ by $\hat\Phi_l^{(h)}$, $\hat\Theta_l^{(h)}$. One may then form the sample covariance matrices of the combined residual series $\hat{\mathrm{a}}(t) = [\hat{\mathrm{a}}^{(1)}(t)'\,\hat{\mathrm{a}}^{(2)}(t)']'$, $t=1,\ldots,n$, of dimension $d=d_1+d_2$:

$$\mathrm{C}_{\hat{\mathrm{a}}}(k) = (c_{ij}(k))_{d\times d} = \frac{1}{n-p}\sum_{t=p+1+\max(0,k)}^{n+\min(0,k)}\hat{\mathrm{a}}(t)\hat{\mathrm{a}}(t-k)', \quad |k| < n-p \tag{2.4}$$

and $\mathrm{C}_{\hat{\mathrm{a}}}(k)=0$ for $|k|\geq n-p$, where $p=\max(p_1,p_2)$, and the corresponding correlation matrices

$$\mathrm{R}_{\hat{\mathrm{a}}}(k) = (r_{ij}(k))_{d\times d}, \quad r_{ij}(k) = c_{ij}(k)/\sqrt{c_{ii}(0)c_{jj}(0)}. \tag{2.5}$$

The matrix $\mathrm{C}_{\widehat{\mathrm{a}}}(k)$ can be partitioned in a obvious way into blocks $\mathrm{C}_{\widehat{\mathrm{a}}}^{(11)}(k)$, $\mathrm{C}_{\widehat{\mathrm{a}}}^{(22)}(k)$ and $\mathrm{C}_{\widehat{\mathrm{a}}}^{(12)}(k)$, of size $d_1 \times d_1$, $d_2 \times d_2$ and $d_1 \times d_2$, respectively. Similarly for the matrix $\mathrm{R}_{\widehat{\mathrm{a}}}(k)$. The matrices $\mathrm{C}_{\widehat{\mathrm{a}}}^{(hh)}(0)$ is actually the maximum conditional Gaussian likelihood estimate of the innovation covariance matrices Ω_h. The matrix $\mathrm{R}_{\widehat{\mathrm{a}}}^{(12)}(k)$ is the sample cross-correlation matrix at lag k, $|k| < n - p$, between the two residual series $\widehat{\mathrm{a}}^{(1)}(p+1),\ldots,\ \widehat{\mathrm{a}}^{(1)}(n)$ and $\mathrm{a}^{(2)}(p+1),\ldots,\ \mathrm{a}^{(2)}(n)$. These sample cross-correlation matrices will be used to construct the test statistics. To this end, we will need to know their limiting distribution under the null hypothesis. It will be shown below that this limiting distribution is *the same as that of the sample cross-correlation matrices* $\mathrm{R}_{\mathrm{a}}^{(12)}(k)$ *between the true unobservable innovation series* $\mathrm{a}^{(1)}(p+1),\ \ldots,\ \mathrm{a}^{(1)}(n)$ *and* $\mathrm{a}^{(2)}(p+1),\ldots,\ \mathrm{a}^{(2)}(n)$. The later being defined in the same way as (2.4) and (2.5) but with $\mathrm{a}^{(h)}(t)$ in placed of $\widehat{\mathrm{a}}^{(h)}(t)$. Such a result has been shown in El Himdi and Roy (1997), in the *stationary case*. Then, one can construct tests for non-correlation between the innovation processes in the usual way.

3 Asymptotic distribution of test statistics

We first recall the asymptotic distribution of the sample cross-correlation matrices between the true innovation series. It can be obtained from a general result in Roy (1989) (see also Chitturi 1976, Hannan 1976). Let $k_1,\ldots,\ k_m$ be a set of m distinct integers not depending on n such that $|k_i| < n - p$ and consider the random vector

$$\mathrm{r}_{\mathrm{a}}^{(12)} = [\mathrm{vec}\,\mathrm{R}_{\mathrm{a}}^{(12)}(k_1)' \ \cdots\ \mathrm{vec}\,\mathrm{R}_{\mathrm{a}}^{(12)}(k_m)']', \tag{3.1}$$

where the notation vec denotes the vector obtained by stacking the columns of the indicated matrix. Further assume the following assumption.

Assumption A.

$$\mathrm{E}[a_i(t)|\mathcal{F}_{t-1}],\ \mathrm{E}[a_i(t)a_j(t)|\mathcal{F}_{t-1}],\ \mathrm{E}[a_i(t)a_j(t)a_l(t)|\mathcal{F}_{t-1}] \text{ and } \mathrm{E}[a_i(t)a_j(t)a_l(t)a_m(t)|\mathcal{F}_{t-1}]$$

exist and are constant (with respect to t) for all i, j, l, m, where the $a_i(t)$ denote the elements of $\mathrm{a}(t)$ and $\mathcal{F}_{t-1}$ is the σ-field generated by $\mathrm{a}(t-1)$, $\mathrm{a}(t-2)$, ...

Then in the case where the fourth-order cumulants

$$\kappa_{ijlm} = \mathrm{cumulant}\{a_i(t), a_j(t), a_l(t), a_m(t)\}$$

vanish and under the null hypothesis that the processes $\{\mathrm{a}^{(1)}(t)\}$ and $\{\mathrm{a}^{(2)}(t)\}$ are uncorrelated, $\sqrt{n-p}\,\mathrm{r}_{\mathrm{a}}^{(12)}$ converges to a multivariate normal distribution of dimension md_1d_2 with mean 0 and covariance matrix

$$\mathrm{I}_m \otimes [\boldsymbol{\rho}_{\mathrm{a}}^{(22)}(0) \otimes \boldsymbol{\rho}_{\mathrm{a}}^{(11)}(0)],$$

the notation $\otimes$ denoting the tensor (Kronecker) product and $\boldsymbol{\rho}_{\mathrm{a}}^{(hh)}(0)$ denotes the correlation matrix of the process $\{\mathrm{a}^{(h)}(t)\}$, which is no other than the matrix Ω_h pre- and post- divided by the diagonal matrix with diagonal elements being the square root of those of Ω_h.

If the fourth-order cumulants κ_{ijlm} are non zero, the above asymptotic normality still holds but the covariance matrix will be more complex and depends on these cumulants (see Hannan 1976). It also holds if the processes $\{\mathrm{a}^{(1)}(t)\}$ and $\{\mathrm{a}^{(2)}(t)\}$ are correlated, but again with a more complex formula for the covariance matrix. Note further that assumption A holds under the usual condition that the innovation vectors $\mathrm{a}(t)$ are independent and identically distributed with finite fourth-order moments.

We now focus on a particular subprocess $\mathrm{X}^{(h)}(t)$ and thus drop the superscript $^{(h)}$ to simplify the notation. Also the identity matrix will be denoted by I when no confusion on its dimension can occur. We shall be mostly interested in the case where the process is partially nonstationary, that is $\det \Phi(\cdot)$ has a unit root of multiplicity $m \geq 1$. However, the stationary case where there is no such root, which corresponds to $m = 0$, is *not excluded*. More precisely, the following assumption will be made.

Assumption B. $\det \Phi(\cdot)$ has a unit root of multiplicity $m \in \{0, \ldots, d\}$ and $\Phi(1)$ has rank $r = d - m$ (recall that d is the dimension of the process) and $\Theta(1)$ is non-singular.

Proposition 3.1. *Under assumption* B*, one has the following decomposition*

$$\mathrm{X}(t) = \mathrm{P}_1 \mathrm{Y}_1(t) + \mathrm{P}_2 \mathrm{Y}_2(t) \tag{3.2}$$

where P_1 *is a full rank* $d \times m$ *matrix such that* $\Phi(1)\mathrm{P}_1 = 0$, P_2 *is a* $d \times r$ *matrix which completes it to form an invertible matrix,* $\{\mathrm{Y}_1(t)\}$ *is a process of stationary increments and* $\{\mathrm{Y}_2(t)\}$ *is a stationary process. Further,*

$$\mathrm{Y}_1(t) - \mathrm{Y}_1(t-1) = [\mathrm{P}_1^\dagger \Phi^*(1) \mathrm{P}_1]^{-1} \mathrm{P}_1^\dagger \Theta(1) \mathrm{a}(t) + \textit{the increment of a stationary process}$$

where $\mathrm{P}_1^\dagger$ *is any full rank* $m \times d$ *matrix such that* $\mathrm{P}_1^\dagger \Phi(1) = 0$ *and* $\Phi^*(z) = \mathrm{I} - \sum_{j=1}^{p-1} \Phi_j^* z^j$ *with* $\Phi_j^* = -\sum_{k=j+1}^{p} \Phi_k$. *(Clearly the matrix* $[\mathrm{P}_1^\dagger \Phi^*(1) \mathrm{P}_1]^{-1} \mathrm{P}_1^\dagger$ *does not depend on the choice of* $\mathrm{P}_1^\dagger$*.)*

A similar result as above has been obtained in Yap and Reinsel (1995). However, their construction of P_1 and P_2 is based of the Jordan decomposition of the matrix $\sum_{j=1}^{p} \Phi_j$ and is not quite correct. As a counterexample, take

$$\Phi(z) = \mathrm{I} - \begin{bmatrix} 1 & 1 \\ a & 1 \end{bmatrix} z - \begin{bmatrix} 0 & 0 \\ -a & 0 \end{bmatrix} z^2.$$

with $a < \frac{1}{4}$. Then assumption B is satisfied with $m = r = 1$, but

$$\Phi_1 + \Phi_2 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

is already in the Jordan form, and the construction of P_1 and P_2 as in Yap and Reinsel (1995, p. 254) will be wrong. For this reason, we have reworked their result and proof. However, their asymptotic distribution results should remain valid since the decomposition of $\{\mathrm{X}(t)\}$ that they used is similar to (3.2). The only difference lies in the construction of P_1 and P_2.

The above result shows that one can find r linear combinations of the components of the $\{\mathrm{X}(t)\}$ process which are stationary, since $\Phi(1)\mathrm{X}(t) = \Phi(1)\mathrm{P}_2\mathrm{Y}_2(t)$. Further, from the above expression for the increment of the process $\{\mathrm{Y}_1(t)\}$, it is clear that no non trivial linear combination of its components can be stationary. Hence there are exactly r linearly independent linear combinations of the components of the process $\{\mathrm{X}(t)\}$ which are stationary. For $1 < r < d$, this property is often referred to as *cointegration* and the process is called partially nonstationary.

Note that the decomposition (3.2) is not unique since it is constructed from two matrices P_1 and P_2 satisfying the only requirements that $\Phi(1)\mathrm{P}_1 = 0$ and $[\mathrm{P}_1\ \mathrm{P}_2]$ is invertible. More precisely, let $\widetilde{\mathrm{P}}_1$ and $\widetilde{\mathrm{P}}_2$ satisfy the same requirements. Then $\widetilde{\mathrm{P}}_1 = \mathrm{P}_1\mathrm{T}$ and $\widetilde{\mathrm{P}}_2 = \mathrm{P}_2\mathrm{U} + \mathrm{P}_1\mathrm{V}$ for some invertible matrices T, U and some matrix V. Thus $\mathrm{X}(t) = \widetilde{\mathrm{P}}_1\widetilde{\mathrm{Y}}_1(t) + \widetilde{\mathrm{P}}_2\widetilde{\mathrm{Y}}_2(t)$ where $\widetilde{\mathrm{Y}}_1(t) = \mathrm{T}^{-1}[\mathrm{Y}_1(t) - \mathrm{V}\mathrm{U}^{-1}\mathrm{Y}_2(t)]$ and $\widetilde{\mathrm{Y}}_2(t) = \mathrm{U}^{-1}\mathrm{Y}_2(t)$. Again, the process $\{\widetilde{\mathrm{Y}}_1(t)\}$ is the sum of a random

walk with increment $\mathrm{T}^{-1}[\mathrm{P}_1^\dagger\Phi^*(1)\mathrm{P}_1]^{-1}\mathrm{P}_1^\dagger\Theta(1)\mathrm{a}(t) = [\mathrm{P}_1^\dagger\Phi^*(1)\widetilde{\mathrm{P}}_1]^{-1}\mathrm{P}_1^\dagger\Theta(1)\mathrm{a}(t)$ and a stationary process and the process $\{\widetilde{\mathrm{Y}}_2(t)\}$ is stationary.

In the (partially) nonstationary case, it is of interest to consider the error correction form of the model equation:

$$\mathrm{W}(t) = \mathrm{CX}(t-1) + \sum_{j=1}^{p-1} \Phi_j^* \mathrm{W}(t-j) + \Theta(B)\mathrm{a}(t) \tag{3.3}$$

where $\mathrm{W}(t) = \mathrm{X}(t) - \mathrm{X}(t-1)$, $\mathrm{C} = -\Phi(1)$ and Φ_j^* are as in Proposition 2.1. This representation results from the identity

$$\Phi(B) = \Phi^*(B)(\mathrm{I} - B) - \mathrm{C}B \tag{3.4}$$

where $\Phi^*(z)$ is again as in Proposition 2.1.

Since there is a one-to-one map from $\Phi_1, \ldots, \Phi_p$ to $\mathrm{C}, \Phi_1^*, \ldots, \Phi_{p-1}^*$, the maximum conditional Gaussian likelihood estimators $\hat{\Phi}_1, \ldots, \hat{\Phi}_p$ of the former define uniquely those of the later, namely $\widehat{\mathrm{C}}, \hat{\Phi}_1^*, \ldots, \hat{\Phi}_{p-1}^*$, and conversely. It is however much simpler to obtain the asymptotic distribution of $\widehat{\mathrm{C}}, \hat{\Phi}_1^*, \ldots, \hat{\Phi}_{p-1}^*$ than that of $\hat{\Phi}_1, \ldots, \hat{\Phi}_p$. In Yap and Reinsel (1995), it is proved that the random matrices $\sqrt{n-p}\,(\widehat{\mathrm{C}} - \mathrm{C})\mathrm{P}_2$, $\sqrt{n-p}\,(\hat{\Phi}_1^* - \Phi_1^*), \ldots, \sqrt{n-p}\,(\hat{\Phi}_{p-1}^* - \Phi_{p-1}^*)$ and $\sqrt{n-p}(\hat{\Theta}_1 - \Theta_1), \ldots, \sqrt{n-p}\,(\hat{\Theta}_q - \Theta_q)$ converge jointly as $n \to \infty$ to a Gaussian distribution with zero mean and a covariance matrix whose exact value is not of interest to us. Further, $(n-p)(\widehat{\mathrm{C}} - \mathrm{C})\mathrm{P}_1$ converges in distribution to some random matrix, functional of a certain Brownian motion process.

It should be noted that Yap and Reinsel (1995) have proved the above result under the (standard) assumption that the innovation vectors $\mathrm{a}(t)$ are independent identically distributed, while the asymptotic normality of $\mathrm{r_a}$ can be obtained under the weaker assumption A. However, to transfer this asymptotic normality to that of the statistic $\mathrm{r}_{\widehat{\mathrm{a}}}$, defined similarly to $\mathrm{r_a}$ as in (3.1) but with $\mathrm{R}_{\widehat{\mathrm{a}}}^{(12)}(k)$ in place of $\mathrm{R}_{\mathrm{a}}^{(12)}(k)$, we actually only need that the model estimators converge to their true value at an adequate rate. Therefore, it is quite plausible that assumption A is sufficient, but we do not pursuit this matter further. Instead, we will make the following assumption

Assumption C. The estimators $\widehat{\mathrm{C}}, \hat{\Phi}_1^*, \ldots, \hat{\Phi}_{p-1}^*$ and $\hat{\Theta}_1, \ldots, \hat{\Theta}_q$ are such that

$$(n-p)(\widehat{\mathrm{C}} - \mathrm{C})\mathrm{P}_1, \quad \sqrt{n-p}\,(\widehat{\mathrm{C}} - \mathrm{C})\mathrm{P}_2, \quad \sqrt{n-p}\,(\hat{\Phi}_1^* - \Phi_1^*), \quad \ldots, \quad \sqrt{n-p}\,(\hat{\Phi}_{p-1}^* - \Phi_{p-1}^*),$$
$$\sqrt{n-p}\,(\hat{\Theta}_1 - \Theta_1), \quad \ldots, \quad \sqrt{n-p}\,(\hat{\Theta}_q - \Theta_q)$$

are bounded in probability as $n \to \infty$, the matrices P_1 and P_2 being defined as above.

Our main result concerns the joint asymptotic distribution of the sample cross-correlations of the residuals. It relies heavily on the following lemma.

Lemma 3.1. *Let* $\{\mathrm{e}(t)\}$ *be another white noise, uncorrelated with* $\{\mathrm{a}(t)\}$ *and such that*

$$\text{cumulant}\{a_i(t_1), a_j(t_2), e_l(t), e_m(t)\} = 0, \quad (t_1, t_2, t) \in \mathbb{Z}^3$$

($a_i(t)$ *and* $e_i(t)$ *denoting the components of* $\mathrm{a}(t)$ *and* $\mathrm{e}(t)$ *respectively*). *Then under assumptions* B *and* C, *the random variables*

$$\sum_{t=p+1}^{n} [\hat{\mathrm{a}}(t) - \mathrm{a}(t)]\mathrm{e}(t)' \quad \textit{and} \quad \sum_{t=p+1}^{n} \|\hat{\mathrm{a}}(t) - \mathrm{a}(t)\|^2,$$

$\|\ \|$ *denoting the Euclidean norm, are bounded in probability as* $n \to \infty$.

The above lemma will be applied to one of the innovation processes $\{\mathrm{a}^{(h)}(t)\}$ with $\{\mathrm{e}(t)\}$ being the other innovation process. One then get the following result.

Theorem 3.1. *Suppose that the processes* $\{\mathrm{X}^{(1)}(t)\}$ *and* $\{\mathrm{X}^{(2)}(t)\}$ *satisfy the ARMA model* (2.1) *and assumptions* B, C, *that the combined innovation process* $\{\mathrm{a}(t)\}$ *satisfies assumption* A *and that all its fourth-order cumulants* κ_{ijlm} *vanish. Then as* $n \to \infty$*:*

(i) $\mathrm{C}_{\widehat{\mathrm{a}}}^{(hh)}(0)$ *converges in probability to* Ω_h, $h = 1, 2$.

(ii) *In the case where the processes* $\{\mathrm{a}^{(1)}(t)\}$ *and* $\{\mathrm{a}^{(2)}(t)\}$ *are uncorrelated,* $\sqrt{n-p}\,\mathrm{r}_{\widehat{\mathrm{a}}}^{(12)}$ *has the the same limiting distribution as* $\sqrt{n-p}\,\mathrm{r}_{\mathrm{a}}^{(12)}$ *(which is a multivariate normal distribution of dimension* md_1d_2 *with mean* 0 *and covariance matrix* $\mathrm{I}_m \otimes [\boldsymbol{\rho}_{\mathrm{a}}^{(22)}(0) \otimes \boldsymbol{\rho}_{\mathrm{a}}^{(11)}(0)]$*).*

Note that the first result of Theorem 3.1 shows that the matrix $\boldsymbol{\rho}_{\mathrm{a}}^{(hh)}(0)$, appearing in the expression for the asymptotic covariance matrix of $\mathrm{r}_{\mathrm{a}}^{(12)}$, can be estimated consistently by $\mathrm{R}_{\widehat{\mathrm{a}}}^{(hh)}(0)$. One can now construct the statistics in a similar way as in El Himdi and Roy (1997). Two types of test statistics will be considered.

The first type of test is based on the cross-correlations at individual lags. One considers the test statistic

$$\mathrm{QH}(k) = (n-p)\,\mathrm{vec}\,\mathrm{R}_{\widehat{\mathrm{a}}}^{(12)}(k)'[\hat{\boldsymbol{\rho}}_{\mathrm{a}}^{(11)}(0)^{-1} \otimes \hat{\boldsymbol{\rho}}_{\mathrm{a}}^{(22)}(0)^{-1}]\,\mathrm{vec}\,\mathrm{R}_{\widehat{\mathrm{a}}}^{(12)}(k), \tag{3.5}$$

where $\hat{\boldsymbol{\rho}}_{\mathrm{a}}^{(hh)}(0)$, is a consistent estimator of $\boldsymbol{\rho}_{\mathrm{a}}^{(hh)}(0)$, $h = 1, 2$. In particular, one may take it to be $\mathrm{R}_{\widehat{\mathrm{a}}}^{(hh)}(0)$. By Theorem 3.1, under the hypothesis of non-correlation, $\mathrm{QH}(k)$ is asymptotically distributed as a $\chi^2_{d_1d_2}$ variable. The null hypothesis is rejected at a significance level α if $\mathrm{QH}(k) > \chi^2_{d_1d_2,1-\alpha}$ where $\chi^2_{m,p}$ denotes the p-quantile of the χ^2_m distribution.

In practice, one may want to consider simultaneously many lags, for example all lags not greater that M in absolute value. A global test based on the statistics $\mathrm{QH}(k)$, $|k| \le M$, consists in rejecting the null hypothesis if for at least one $k \in \{-M, \ldots, M\}$, $\mathrm{QH}(k) > \chi^2_{d_1d_2,1-\alpha_0}$. Since the test statistics $\mathrm{QH}(k)$ are asymptotically independent, in order to have a global significance level α, the marginal significance level α_0 of each test must be $\alpha_0 = 1 - (1-\alpha)^{1/(2M+1)}$.

The second type of test is a generalization of the global test proposed by Haugh (1976). It is based on the test statistics

$$\mathrm{QH}_M = (n-p)\ (\mathrm{r}_{\widehat{\mathrm{a}}}^{(12)})'[\mathrm{I}_{2M+1} \otimes \hat{\boldsymbol{\rho}}_{\mathrm{a}}^{(22)}(0) \otimes \hat{\boldsymbol{\rho}}_{\mathrm{a}}^{(11)}(0)]^{-1}\mathrm{r}_{\widehat{\mathrm{a}}}^{(12)} = \sum_{k=-M}^{M} \mathrm{QH}(k). \tag{3.6}$$

Under the null hypothesis, the vector $\mathrm{r}_{\widehat{\mathrm{a}}}^{(12)} = [\mathrm{vec}\,\mathrm{R}_{\widehat{\mathrm{a}}}^{(12)}(-M)' \ \cdots \ \mathrm{vec}\,\mathrm{R}_{\widehat{\mathrm{a}}}^{(12)}(M)']'$ obeys asymptotically a multivariate normal distribution of mean zero and covariance matrix $\mathrm{I}_{2M+1} \otimes \boldsymbol{\rho}_{\mathrm{a}}^{(22)}(0) \otimes \boldsymbol{\rho}_{\mathrm{a}}^{(11)}(0)$. Hence QH_M is asymptotically distributed as a $\chi^2_{(2M+1)d_1d_2}$ variable.

The statistic QH_M can also be expressed in terms of the autocovariances $\mathrm{C}_{\widehat{\mathrm{a}}}^{(11)}(0)$, $\mathrm{C}_{\widehat{\mathrm{a}}}^{(22)}(0)$, and the cross-covariances $\mathrm{C}_{\widehat{\mathrm{a}}}^{(12)}(k)$ of the residual series $\widehat{\mathrm{a}}^{(1)}(t)$ and $\widehat{\mathrm{a}}^{(2)}(t)$, as explained in El Himdi and Roy (1997).

As in the stationary case, the modified statistics

$$\mathrm{QH}^*(k) = \frac{n-p}{n-p-|k|}\,\mathrm{QH}(k) \tag{3.7}$$

and

$$\mathrm{QH}_M^* = \sum_{k=-M}^{M} \mathrm{QH}^*(k) \tag{3.8}$$

will be considered, as they seem to be better approximated by a chi-square distribution, for small sample size.

4 Simulation study

4.1 Description of the experiment

Here, we report the results of a small simulation experiment conducted in order to compare the exact distributions of the statistics $\mathrm{QH}(k)$, $\mathrm{QH}^*(k)$, QH_M and QH^*_M with their corresponding asymptotic chi-square distributions, under the null hypothesis of non-correlation. To do that, we examined the empirical frequencies of rejection of the null hypothesis with the proposed tests at three different nominal levels (1, 5 and 10 percent) for each of three series lengths ($n = 50$, 100 and 200) and for two different global models for $\{\mathrm{X}^{(1)}(t)\}$ and $\{\mathrm{X}^{(2)}(t)\}$. These models are described in Table 1.

The dimension of each of the two models is four and for each one, the submodels for $\{\mathrm{X}^{(1)}(t)\}$ and $\{\mathrm{X}^{(2)}(t)\}$ are bivariate. Also, with the considered values for the autoregressive, moving average parameters as well as for the covariance matrix of the innovations, the subprocesses $\{\mathrm{X}^{(1)}(t)\}$ and $\{\mathrm{X}^{(2)}(t)\}$ are uncorrelated and the corresponding submodels are partially nonstationary and invertible.

For each model, the experiment proceeded in the following way.

1. For each model and for each series length n, 2000 independent realizations were generated. First, pseudo-random variables from a $\mathrm{N}(0, 1)$ distribution were obtained with the pseudo-random normal generator of the Splus package and were transformed into independent $\mathrm{N}(0, \Omega_{\mathrm{a}})$ pseudo-random vectors by the Cholesky transformation. Second the $\mathrm{X}(t)$ values were obtained by directly solving the difference equation (2.1). Initial $\mathrm{X}(t)$ values needed to start the recursion were set at 0.

2. In the case of pure AR processes, for both subseries $\mathrm{X}^{(h)}(t)$, $h = 1, 2$, the least squares estimates of the coefficients of the true models were obtained using the procedure described in Reinsel (1993, p. 84). With the ARMA process, each subseries was approximated by a possibly high order AR model, the value of the AR order was obtained by minimizing Hannan-Quinn criterion using conditional least squares estimation. The residual series $\widehat{\mathrm{a}}^{(h)}(t)$, $h = 1, 2$, were cross-correlated by computing $\mathrm{R}^{(12)}_{\widehat{\mathrm{a}}}(k)$ as defined by (2.5)

3. The values of the test statistics $\mathrm{QH}(k)$ and $\mathrm{QH}^*(k)$ were computed for $k = -12, \ldots, 12$ and those for QH_M and QH^*_M for $M = 1, \ldots, 12$. For each test, the value of the statistic was compared with the critical value obtained from the corresponding chi-square distribution.

4. Finally, for each nominal level and for each series length n, we obtained from the 2000 realizations, the empirical frequencies of rejection of the null hypothesis of non-correlation.

4.2 Description of the results

The results, in percentage, for $\mathrm{QH}^*(k)$ and QH^*_M are reported in Figures 1–3. For each nominal level, the corresponding dotted lines represent that level ± 1.96 times the standard error of the empirical level which are 0.22% at 1%, 0.49% at 5%, and 0.67% at 10%. We make the following observations. For the AR(1) model, the chi-square distribution provides a satisfactory approximation even for series of 50 observations for both $\mathrm{QH}^*(k)$ and QH^*_M at the three significance levels since almost all corresponding empirical values lie inside the 5% significant limits. With the AR(2) model, the $\mathrm{QH}^*(k)$ statistics are well approximated by the chi-square even with 50 observations. However, with the global statistic QH^*_M, we notice that the empirical levels are uniformly smaller than the nominal levels and that the discrepancy is significant at 5% and 1%. With 100 observations, the empirical

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

Models	**Equations**
AR(1)	$\begin{bmatrix} \mathrm{X}_t^{(1)} \\ \mathrm{X}_t^{(2)} \end{bmatrix} = \begin{bmatrix} \Phi^{(1)} & 0 \\ 0 & \Phi^{(2)} \end{bmatrix} \begin{bmatrix} \mathrm{X}_{t-1}^{(1)} \\ \mathrm{X}_{t-1}^{(2)} \end{bmatrix} + \begin{bmatrix} \mathrm{a}_t^{(1)} \\ \mathrm{a}_t^{(2)} \end{bmatrix}$
AR(2)	$\begin{bmatrix} \mathrm{X}_t^{(1)} \\ \mathrm{X}_t^{(2)} \end{bmatrix} = \begin{bmatrix} \Phi_1^{(1)} & 0 \\ 0 & \Phi_1^{(2)} \end{bmatrix} \begin{bmatrix} \mathrm{X}_{t-1}^{(1)} \\ \mathrm{X}_{t-1}^{(2)} \end{bmatrix} + \begin{bmatrix} \Phi_2^{(1)} & 0 \\ 0 & \Phi_2^{(2)} \end{bmatrix} \begin{bmatrix} \mathrm{X}_{t-2}^{(1)} \\ \mathrm{X}_{t-2}^{(2)} \end{bmatrix} + \begin{bmatrix} \mathrm{a}_t^{(1)} \\ \mathrm{a}_t^{(2)} \end{bmatrix}$
ARMA(1, 1)	$\begin{bmatrix} \mathrm{X}_t^{(1)} \\ \mathrm{X}_t^{(2)} \end{bmatrix} = \begin{bmatrix} \Phi^{(1)} & 0 \\ 0 & \Phi^{(2)} \end{bmatrix} \begin{bmatrix} \mathrm{X}_{t-1}^{(1)} \\ \mathrm{X}_{t-1}^{(2)} \end{bmatrix} + \begin{bmatrix} \Theta^{(1)} & 0 \\ 0 & \Theta^{(2)} \end{bmatrix} \begin{bmatrix} \mathrm{a}_t^{(1)} \\ \mathrm{a}_t^{(2)} \end{bmatrix}$

Noise covariance matrix

$$\Omega_{\mathrm{a}} = \begin{bmatrix} \Omega_{\mathrm{a}}^{(1)} & 0 \\ 0 & \Omega_{\mathrm{a}}^{(2)} \end{bmatrix}$$

Parameters values

$$\Phi^{(1)} = \begin{bmatrix} 0.4 & 0.0 \\ -1.0 & 1.0 \end{bmatrix} \quad \Phi^{(2)} = \begin{bmatrix} 1.0 & 0.0 \\ -0.8 & 0.5 \end{bmatrix} \quad \Phi_1^{(1)} = \begin{bmatrix} 0.6 & -0.5 \\ 0.3 & 0.4 \end{bmatrix}$$

$$\Phi_1^{(2)} = \begin{bmatrix} -0.5 & -0.8 \\ -0.4 & 0.2 \end{bmatrix} \quad \Phi_2^{(1)} = \begin{bmatrix} 0.4 & 0.5 \\ 0.6 & 0.3 \end{bmatrix} \quad \Phi_2^{(2)} = \begin{bmatrix} -0.5 & 0.3 \\ -0.8 & 0.5 \end{bmatrix}$$

$$\Theta^{(1)} = \begin{bmatrix} 0.7 & 0.6 \\ -0.4 & 0.6 \end{bmatrix} \quad \Theta^{(2)} = \begin{bmatrix} -0.5 & 0.3 \\ -0.7 & 0.6 \end{bmatrix}$$

$$\Omega_{\mathrm{a}}^{(1)} = \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix} \quad \Omega_{\mathrm{a}}^{(2)} = \begin{bmatrix} 1.00 & 0.75 \\ 0.75 & 1.00 \end{bmatrix}$$

Table 2: Roots of the determinant of the autoregressive and moving average polynomials.

Polynomials	**Roots**
$1 - \Phi^{(1)}B$	$\{1.00, 2.50\}$
$1 - \Phi^{(2)}B$	$\{1.00, 2.00\}$
$1 - \Phi_1^{(1)}B - \Phi_2^{(1)}B^2$	$\{1.00, -1.35, 1.53 \pm 1.33i\}$
$1 - \Phi_1^{(2)}B - \Phi_2^{(2)}B^2$	$\{1.00, -0.74 \pm 0.78\,i, -86.51\}$
$1 - \Theta^{(1)}B$	$\{0.10 \pm 1.23i\}$
$1 - \Theta^{(2)}B$	$\{2.82, -3.94\}$

levels are still uniformly smaller than the nominal levels but in general, the discrepancy between them is not significant. With 200 observations, the approximation is very good for the three series lengths considered.

The results illustrated in Figure 3 for the ARMA$(1,1)$ model are very satisfactory even for series of 50 observations. It seems to indicate that our procedure for checking independence of cointegrated series is robust in some sense to specification errors and that our main result is still valid if the residuals of the true models are replaced by those resulting from autoregressive approximations. These empirical results are coherent with the theoretical results of Hong (1996) concerning univariate stationary series. Autoregressive approximations of cointegrated processes for estimation and testing is discussed in Saikkonen (1992).

Finally, the empirical results for $\mathrm{QH}(k)$ and QH_M are not reported since, as in the stationary case, the chi-square approximation is unsatisfactory at least for short series of 50 to 100 observations. For both $\mathrm{QH}(k)$ and QH_M, the empirical levels are systematically smaller than the nominal levels.

5 A numerical example

Here, we consider a set of seven quarterly series of Canadian and American economic indicators used in a study of Canadian monetary policy in order to investigate correlation and causality directions between the two economies, see Racette and Raynauld (1992). These data were also analyzed in El Himdi and Roy (1997). The Canadian economic indicators are gross domestic production (GDP) in constant 1982 dollars, the implicit price index of the gross domestic production (GDPI), the nominal short-term interest rate (TX.CA) and the monetary basis value (M1). The other three variables represent the American gross national product (GNP) in constant 1982 dollars, the implicit price index of the American gross national production (GNPI), and the nominal short-term American interest rate (TX.US). The observation period extends from the first quarter of 1970 through to the last quarter of 1989. The data sources with the corresponding CANSIM series numbers are given in Table 1 of Racette and Raynauld (1992). The natural logarithm of M1 was taken in order to stabilize its variance.

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

$$\mathrm{X}^{(1)}(t) = \begin{bmatrix} \frac{1}{1000}\,\mathrm{GDP}(t) \\ 10\,\mathrm{GDPI}(t) \\ \mathrm{TX.\,CA}(t) \\ 100\ln\Big(\mathrm{M1}(t)\Big) \end{bmatrix}, \quad \mathrm{X}^{(2)}(t) = \begin{bmatrix} \frac{1}{10}\,\mathrm{GNP}(t) \\ 10\,\mathrm{GNPI}(t) \\ \mathrm{TX.\,US}(t) \end{bmatrix}$$

The multiplicative factors appearing in the definition of $\mathrm{X}^{(1)}(t)$ and $\mathrm{X}^{(2)}(t)$ are the same as those used in El Himdi and Roy (1997). With these factors, the orders of magnitude of the components of each of the modified series are similar. The graphs of the various components of the two series are presented in Figure 4. Autoregressive models were fitted to each series using unconstrained least squares estimation (ignoring the cointegration constraint) and minimizing an information criterion for determining the autoregressive order. This procedure led to AR(3) models for both series and they are given in Table 3. They satisfy the diagnostic checks suggested by Tiao and Box (1981) to ensure model adequacy. The roots of the determinant of the autoregressive polynomial were computed for each model. The smallest roots are respectively 1.0002, 1.0504, 1.0532 for Canada and 0.9974, 1.0474 for the United States. Very likely, there is cointegration of order one in both series.

Table 3: Estimated AR models for Canadian and American series.

(a) Canadian series

$$\begin{bmatrix} X_{1t}^{(1)} \\ X_{2t}^{(1)} \\ X_{3t}^{(1)} \\ X_{4t}^{(1)} \end{bmatrix} = \begin{bmatrix} -92.389 \\ -26.385 \\ -7.542 \\ 12.481 \end{bmatrix} + \begin{bmatrix} 1.007 & 0.163 & -0.449 & 0.386 \\ 0.111 & 1.238 & 0.188 & 0.163 \\ 0.123 & 0.094 & 0.833 & 0.179 \\ 0.047 & 0.159 & -0.924 & 0.738 \end{bmatrix} \begin{bmatrix} X_{1,t-1}^{(1)} \\ X_{2,t-1}^{(1)} \\ X_{3,t-1}^{(1)} \\ X_{4,t-1}^{(1)} \end{bmatrix}$$
$$+ \begin{bmatrix} -0.228 & -0.158 & 0.086 & 0.044 \\ -0.027 & -0.096 & 0.258 & 0.568 \\ -0.041 & -0.023 & -0.109 & -0.222 \\ -0.105 & -0.172 & 0.356 & 0.183 \end{bmatrix} \begin{bmatrix} X_{1,t-2}^{(1)} \\ X_{2,t-2}^{(1)} \\ X_{3,t-2}^{(1)} \\ X_{4,t-2}^{(1)} \end{bmatrix}$$
$$+ \begin{bmatrix} 0.117 & 0.013 & -0.641 & -0.297 \\ 0.015 & -0.173 & 0.411 & -0.732 \\ -0.082 & -0.073 & 0.097 & 0.054 \\ 0.058 & 0.012 & 0.092 & 0.070 \end{bmatrix} \begin{bmatrix} X_{1,t-3}^{(1)} \\ X_{2,t-3}^{(1)} \\ X_{3,t-3}^{(1)} \\ X_{4,t-3}^{(1)} \end{bmatrix} + \begin{bmatrix} a_{1t}^{(1)} \\ a_{2t}^{(1)} \\ a_{3t}^{(1)} \\ a_{4t}^{(1)} \end{bmatrix}$$

(b) American series

$$\begin{bmatrix} X_{1t}^{(2)} \\ X_{2t}^{(2)} \\ X_{3t}^{(2)} \end{bmatrix} = \begin{bmatrix} 24.109 \\ -18.248 \\ -3.546 \end{bmatrix} + \begin{bmatrix} 0.866 & 0.167 & -0.036 \\ 0.036 & 1.403 & 0.756 \\ 0.126 & 0.077 & 0.805 \end{bmatrix} \begin{bmatrix} X_{1,t-1}^{(2)} \\ X_{2,t-1}^{(2)} \\ X_{3,t-1}^{(2)} \end{bmatrix}$$
$$+ \begin{bmatrix} 0.198 & -0.072 & -0.968 \\ 0.058 & -0.229 & -0.868 \\ 0.096 & 0.043 & -0.553 \end{bmatrix} \begin{bmatrix} X_{1,t-2}^{(2)} \\ X_{2,t-2}^{(2)} \\ X_{3,t-2}^{(2)} \end{bmatrix}$$
$$+ \begin{bmatrix} -0.172 & -0.073 & 0.038 \\ -0.007 & -0.190 & 0.591 \\ -0.192 & -0.126 & 0.618 \end{bmatrix} \begin{bmatrix} X_{1,t-3}^{(2)} \\ X_{2,t-3}^{(2)} \\ X_{3,t-3}^{(2)} \end{bmatrix} + \begin{bmatrix} a_{1t}^{(2)} \\ a_{2t}^{(2)} \\ a_{3t}^{(2)} \end{bmatrix}$$

Residual covariance matrices

(a)

$$\begin{bmatrix} 8.361 & -4.279 & 0.533 & -0.086 \\ -4.279 & 12.286 & 0.218 & 1.457 \\ 0.533 & 0.218 & 1.193 & -0.072 \\ -0.086 & 1.457 & -0.072 & 1.715 \end{bmatrix}$$

(b)

$$\begin{bmatrix} 5.949 & 0.057 & 1.007 \\ 0.057 & 7.531 & 0.281 \\ 1.007 & 0.281 & 1.507 \end{bmatrix}$$

Table 4: Values of the global statistic QH^*_M defined by (3.8) and its empirical significance level for $M = 1,\ldots, 12$.

M	QH^*_M	α_M	M	QH^*_M	α_M
1	52.338	0.038	7	205.621	0.092
2	80.237	0.042	8	221.767	0.187
3	108.338	0.038	9	250.599	0.145
4	135.703	0.037	10	271.592	0.189
5	159.384	0.052	11	284.563	0.349
6	181.151	0.082	12	306.455	0.386

The values of the statistics $\mathrm{QH}^*(k)$ defined by (3.7) are represented in Figure 5 for $|k| = 1,\ldots 12$. At the significance level $\alpha = 0.05$, the critical value for testing the null hypothesis H_0 of non-correlation between $\mathrm{X}^{(1)}$ and $\mathrm{X}^{(2)}$ against the alternative H_{1k}: $\boldsymbol{\rho}^{(12)}_{\mathrm{a}}(k) \neq 0$ is $\chi^2_{12,0.95} = 21.02$ and only $\boldsymbol{\rho}^{(12)}_{\mathrm{a}}(0)$ is significantly different from zero. The empirical significance levels of the portmanteau test QH^*_M for H_0 are also reported in Table 4, for $M = 1,\ldots, 12$. At the 0.05 significance level H_0 is rejected for $M = 1,\ldots, 4$. With M greater than 4, H_0 is not rejected. The statistic QH^*_M seems to behave in a similar way to the classical Box-Pierce portmanteau statistic in the sense that its power decreases as M increases past some threshold.

By directly modelling the two series, this analysis has led to conclude that the American and Canadian economic indicators considered are correlated, which is not a surprise from an economic point of view. Although this conclusion is coherent with the one obtained in El Himdi and Roy (1997), the present study focused on the original (undifferenced) series while in the latter, all the economic indicators were differenced once except the two interest rate series. >From an economic point of view, it is important to be able to draw conclusions on the original variables rather than on the differenced variables. Furthermore, it is easy to verify that given two scalar processes $\{X(t) : t \in \mathbb{Z}\}$ and $\{Y(t) : t \in \mathbb{Z}\}$, $\mathrm{Cov}\Big(\nabla X(t), \nabla Y(t-k)\Big) = 0$, $k \in \mathbb{Z}$, does not necessarily implies that $\mathrm{Cov}\Big(X(t), Y(t-k)\Big) = 0$, $k \in \mathbb{Z}$, where $\nabla = 1 - B$ is the difference operator.

6 Concluding remarks

In this article, we have extended the methods developed by Haugh (1976) and El Himdi and Roy (1997) for checking the non-correlation of two univariate or multivariate stationary ARMA time series to the case of possibly cointegrated series. It is shown that in the case of two uncorrelated and possibly cointegrated time series, an arbitrary vector of residual cross-correlation matrices asymptotically follows the same distribution that the corresponding vector of cross-correlation matrices between the two innovation series. The estimation method can be the conditional maximum likelihood method or an asymptotically equivalent one. Two types of test statistics are considered. The first one is based on the cross-correlations at individual lags while the second one is a portmanteau-type statistic that simultaneously take into account many lags. Both test statistics asymptotically follow chi-square distributions under the null hypothesis of non-correlation. From a small Monte Carlo experiment, it is seen that a slightly modified version of the test statistics have distributions that are generally better approximated by the chi-square distribution and thus yield more accurate critical values. For the low order models considered (AR(1), AR(2), and ARMA(1,1)), the difference is important and with the modified statistics, the approximation is satisfactory even for short series of 100 observations. Finally, the proposed method avoids the difficulties associated with the

individual differencing of each component in multivariate nonstationary time series modelling and allows us to draw conclusions on the original variables rather than on the corresponding differenced series.

Appendix Proofs

Proof of Proposition 3.1. Let $\mathrm{P} = [\mathrm{P}_1\ \mathrm{P}_2]$ and partition its inverse Q into $[\mathrm{Q}_1'\ \mathrm{Q}_2']'$ with Q_1 and Q_2 having m and r rows respectively. Then, noting that $\mathrm{Q}_2\mathrm{P}_1 = 0$, $\mathrm{P}_1\mathrm{Q}_1 + \mathrm{P}_2\mathrm{Q}_2 = \mathrm{I}$, one gets $\mathrm{I} - B = (\mathrm{I} - \mathrm{P}_2\mathrm{Q}_2 B)(\mathrm{I} - \mathrm{P}_1\mathrm{Q}_1 B)$ and hence by (3.4)

$$\Phi(B) = [\Phi^*(B)(\mathrm{I} - \mathrm{P}_2\mathrm{Q}_2 B) + \Phi(1)B](\mathrm{I} - \mathrm{P}_1\mathrm{Q}_1 B).$$

We denote by $\Phi^{**}(B)$ the expression inside the bracket [ ] in the last right hand side. Thus $\det \Phi(B) = (1-B)^m \det \Phi^{**}(B)$ since $\det(\mathrm{I} - \mathrm{P}_1\mathrm{Q}_1 B) = \det(\mathrm{I} - \mathrm{Q}\mathrm{P}_1\mathrm{Q}_1\mathrm{P}B)$ and the matrix $\mathrm{Q}\mathrm{P}_1\mathrm{Q}_1\mathrm{P}$ has its first m diagonal elements equal to 1 and the others vanish. Our assumption B implies that $\det \Phi^{**}(\cdot)$ is no roots inside and on the unit circle and hence

$$(\mathrm{I} - \mathrm{P}_1\mathrm{Q}_1 B)\mathrm{X}(t) = \Phi^{**}(B)^{-1}\Theta(B)\mathrm{a}(t) \tag{A.1}$$

and the right hand side of (A.1) defines a stationary process. Pre-multiplying both sides of this equality with Q_1 and with Q_2, one gets that the process $\{\mathrm{Y}_1(t) = \mathrm{Q}_1\mathrm{X}(t)\}$ has stationary increments $\mathrm{Q}_1\Phi^{**}(B)^{-1}\Theta(B)\mathrm{a}(t)$ and the process $\{\mathrm{Y}_2(t) = \mathrm{Q}_2\mathrm{X}(t) = \mathrm{Q}_2\Phi^{**}(B)^{-1}\Theta(B)\mathrm{a}(t)\}$ is stationary. Further, from $\mathrm{P}_1\mathrm{Q}_1 + \mathrm{P}_2\mathrm{Q}_2 = \mathrm{I}$, one gets the decomposition (3.2) of the proposition.

Using the identity

$$\Phi^{**}(B)^{-1} = \Phi^{**}(1)^{-1}\{\mathrm{I} - [\Phi^{**}(B) - \Phi^{**}(1)]\Phi^{**}(B)^{-1}\}$$

one can write

$$\mathrm{Y}_1(t) - \mathrm{Y}_1(t-1) = \mathrm{Q}_1\Phi^{**}(1)^{-1}\{\Theta(B)\mathrm{a}(t) - [\Phi^{**}(B) - \Phi^{**}(1)]\Phi^{**}(B)^{-1}\Theta(B)\mathrm{a}(t)\}.$$

Note that by (A.1) and (3.2)

$$\Phi^{**}(B)^{-1}\Theta(B)\mathrm{a}(t) = (\mathrm{I} - \mathrm{P}_1\mathrm{Q}_1 B)[\mathrm{P}_1\mathrm{Y}_1(t) + \mathrm{P}_2\mathrm{Y}_2(t)] = \mathrm{P}_1\mathrm{Q}_1(\mathrm{I} - B)\mathrm{X}(t) + \mathrm{P}_2\mathrm{Y}_2(t),$$

the last equality coming from $\mathrm{Q}_1\mathrm{P}_1 = \mathrm{I}$, $\mathrm{Q}_1\mathrm{P}_2 = 0$ and $\mathrm{Y}_1(t) = \mathrm{Q}_1\mathrm{X}(t)$. Thus, putting $\Pi_1 = \mathrm{Q}_1\Phi^{**}(1)^{-1}$, $\mathrm{W}(t) = \mathrm{X}(t) - \mathrm{X}(t-1)$ and noting that $\Phi^{**}(B)\mathrm{P}_1 = \Phi^*(B)\mathrm{P}_1$, $\Phi^{**}(B)\mathrm{P}_2 = \Phi(1)B\mathrm{P}_2$, we obtain

$$\mathrm{Y}_1(t) - \mathrm{Y}_1(t-1) = \Pi_1\{\Theta(B)\mathrm{a}(t) - [\Phi^*(B) - \Phi^*(1)]\mathrm{P}_1\mathrm{Q}_1\mathrm{W}(t) + (\mathrm{I} - B)\Phi(1)\mathrm{P}_2\mathrm{Y}_2(t)\}.$$

The matrix Π_1 is completely specified by the equation $\Pi_1\Phi^{**}(1)\mathrm{P} = [\mathrm{I}\quad 0]$ or equivalently $\Pi_1\Phi^*(1)\mathrm{P}_1 = \mathrm{I}$, $\Pi_1\mathrm{C}\mathrm{P}_2 = 0$. This shows that it can be written as $[\mathrm{P}_1^\dagger\Phi^*(1)\mathrm{P}_1]^{-1}\mathrm{P}_1^\dagger$ where $\mathrm{P}_1^\dagger$ is any full rank $m \times d$ matrix such that $\mathrm{P}_1^\dagger\Phi(1) = 0$. On the other hand, write $\Phi^*(B) - \Phi^*(1)$ as $(\mathrm{I} - B)\sum_{j=0}^{p-2}(\sum_{k=j+1}^{p-1}\Phi_k^*)B^j$ and $\Theta(B)$ as $\Theta(1) - (\mathrm{I} - B)\sum_{j=0}^{q-1}(\sum_{k=j+1}^{q}\Theta_k)B^j$, the increment $\mathrm{Y}_1(t) - \mathrm{Y}_1(t-1)$ becomes

$$\Pi_1\Bigg\{\Theta(1)\mathrm{a}(t) - \sum_{j=0}^{q-1}\Bigg(\sum_{k=j+1}^{q}\Theta_k\Bigg)[\mathrm{a}(t-j) - \mathrm{a}(t-1-j)] + \sum_{j=0}^{p-2}\Bigg(\sum_{k=j+1}^{p-1}\Phi_k^*\Bigg)\mathrm{P}_1\mathrm{Q}_1[\mathrm{W}(t-j) - \mathrm{W}(t-1-j)] + \Phi(1)\mathrm{P}_2[\mathrm{Y}_2(t) - \mathrm{Y}_2(t-1)]\Bigg\}.$$

This yields the last result of the proposition. □

Proof of Lemma 3.1. We first note from (2.3) that for $t > p$, $\hat{\mathrm{a}}(t) = \hat{\Theta}(B)^{-1}\hat{\mathrm{b}}^+(t)$ where $\hat{\mathrm{b}}^+(t) = \hat{\Phi}(B)\mathrm{X}(t)$ if $t > p$ and 0 otherwise. Thus, we can decompose $\hat{\mathrm{a}}(t) - \mathrm{a}(t)$ into

$$\begin{aligned}\hat{\mathrm{a}}(t) - \mathrm{a}(t) &= \Theta(B)^{-1}[\hat{\mathrm{b}}^+(t) - \mathrm{b}^+(t)] + [\hat{\Theta}(B)^{-1} - \Theta(B)^{-1}]\mathrm{b}^+(t) \\ &\quad + [\hat{\Theta}(B)^{-1} - \Theta(B)^{-1}][\hat{\mathrm{b}}^+(t) - \mathrm{b}^+(t)] + [\Theta(B)^{-1}\mathrm{b}^+(t) - \mathrm{a}(t)] \\ &= \mathrm{d}_1(t) + \mathrm{d}_2(t) + \mathrm{d}_3(t) + \mathrm{d}_4(t),\end{aligned}$$

say, where $\mathrm{b}^+(t) = \Phi(B)\mathrm{X}(t) = \Theta(B)\mathrm{a}(t)$ if $t > p$ and 0 otherwise. Note that $\sum_{t=p+1}^n \|\hat{\mathrm{a}}(t) - \mathrm{a}(t)\|^2$ is the squared norm of the vector $[\hat{\mathrm{a}}(p+1)' - \mathrm{a}(p+1)' \ \cdots \ \hat{\mathrm{a}}(n) - \mathrm{a}(n)']'$. Hence by the triangular inequality, it can be easily seen that one obtains the result of the lemma if one has proved that

$$\sum_{t=p+1}^n \mathrm{d}_k(t)\mathrm{e}(t)' \text{ and } \sum_{t=p+1}^n \|\mathrm{d}_k(t)\|^2, \quad k = 1, \ldots, 4,$$

are bounded in probability as $n \to \infty$.

We will denote by R1,..., R4 each of the above results. They will be proved separately, with an outline first given for ease of reading. We shall use the same notation $\|\ \|$ for the matrix norm which is the Euclidean norm of the vector formed by the elements of the matrix. This norm is sub-multiplicative in the sense that $\|\mathrm{AB}\| \leq \|\mathrm{A}\|\,\|\mathrm{B}\|$ whenever the matrix product AB makes sense.

Proof of R4. We shall prove that $E\|\mathrm{d}_4(t)\|^2 \leq C\rho^{2t}$ for all $t > p$, for some constants $C > 0$ and $0 < \rho < 1$. Therefore, $E\sum_{t=p+1}^n \|\mathrm{d}_4(t)\|^2$ is bounded as $n \to \infty$. Further, by Schwarz's inequality,

$$\left\| \sum_{t=p+1}^n \mathrm{d}_4(t)\mathrm{e}(t)' \right\| \leq \left(\sum_{t=p+1}^n \|\mathrm{d}_4(t)\|^2 / r^{2t} \right)^{1/2} \left(\sum_{t=p+1}^n \|\mathrm{e}(t)\|^2 r^{2t} \right)^{1/2}$$

for any $r \neq 0$. Hence the above left hand side has expectation bounded by

$$\left(\sum_{t=p+1}^n \mathrm{E}\,\|\mathrm{d}_4(t)\|^2 / r^{2t} \right)^{1/2} \left(\sum_{t=p+1}^n \mathrm{E}\,\|\mathrm{e}(t)\|^2 r^{2t} \right)^{1/2}$$

which is bounded as $n \to \infty$ by taking $\rho < r < 1$, since $\mathrm{E}\,\|\mathrm{e}(t)\|^2$ does not depend on t. This yields R4.

We now proceed to prove the announced result. Note that

$$\mathrm{d}_4(t) = \sum_{j=t-p+1}^{\infty} \Lambda_j \Theta(B)\mathrm{a}(t-j), \quad t > p,$$

where Λ_j are the coefficients in the Taylor expansion of $\Theta(z)^{-1}$ in power of z. Since the last function is analytic inside the unit circle, there exist constants $K > 0$, $0 < \rho < 1$ such that $\|\Lambda_j\| \leq K\rho^j$, $\forall j \geq 0$. Hence, by the triangular inequality,

$$\|\mathrm{d}_4(t)\| \leq K \sum_{j=t-p+1}^{\infty} \rho^j \|\Theta(B)\mathrm{a}(t-j)\|$$

and hence

$$[\mathrm{E}\,\|\mathrm{d}_4(t)\|^2]^{1/2} \leq K \left(\sum_{j=t-p+1}^{\infty} \rho^j \right) [\mathrm{E}\,\|\Theta(B)\mathrm{a}(t)\|^2]^{1/2}$$

since the random vector $\Theta(B)\mathrm{a}(t)$ have the same covariance matrix. Thus, $\mathrm{E}\,\|\mathrm{d}_4(t)\|^2 \leq C\rho^{2t}$ where C is another constant, which is the desired result.

Proof of R3. We shall prove that $n\sum_{t=p+1}^{n}\|\mathrm{d}_3(t)\|^2$ is bounded in probability. Then the same is true for $\|\sum_{t=p+1}^{n}\mathrm{d}_3(t)\mathrm{e}(t)'\|^2$, since this random variable, by Schwarz's inequality, is bounded by $(n\sum_{t=p+1}^{T}\|\mathrm{d}_3(t)\|^2)(n^{-1}\sum_{t=p+1}^{n}\|\mathrm{e}(t)\|^2)$ and that the last factor in the above right hand side is bounded in the mean, hence in probability.

To prove the announced result, first note that by (3.4)

$$\widehat{\mathrm{b}}^+(t)-\mathrm{b}^+(t)=[\hat{\Phi}^*(B)-\Phi^*(B)]\mathrm{W}(t)-(\widehat{\mathrm{C}}-\mathrm{C})\mathrm{P}_2\mathrm{Y}_2(t-1)-(\widehat{\mathrm{C}}-\mathrm{C})\mathrm{P}_1\mathrm{Y}_1(t-1),$$

for $t>p$. Then since $\sqrt{n-p}\,(\hat{\Phi}_j^*-\Phi_j^*)$, $j=1,\ldots,p-1$, $\sqrt{n-p}\,(\widehat{\mathrm{C}}-\mathrm{C})\mathrm{P}_2$ and $(n-p)(\widehat{\mathrm{C}}-\mathrm{C})\mathrm{P}_1$ are bounded in probability (by assumption C), there exists for any $\eta>0$ a constant $\delta_1>0$ such that

$$\|\widehat{\mathrm{b}}^+(t)-\mathrm{b}^+(t)\|\le\frac{\delta_1}{\sqrt{n}}\left[\sum_{j=1}^{p-1}\|\mathrm{W}(t-j)\|+\|\mathrm{Y}_2(t)\|+\frac{1}{\sqrt{n}}\|\mathrm{Y}_1(t-1)\|\right],\quad t>p,$$

with probability not less than $1-\eta$, for all n sufficiently large.

On the other hand, write $\hat{\Theta}(B)^{-1}-\Theta(B)^{-1}$ as $\hat{\Theta}(B)^{-1}[\Theta(B)-\hat{\Theta}(B)]\Theta(B)^{-1}$ and expanding $\hat{\Theta}(z)^{-1}$ into $\sum_{j=0}^{\infty}\hat{\Lambda}_jB^j$ and $\Theta(z)^{-1}$ into $\sum_{j=0}^{\infty}\Lambda_jB^j$, one gets

$$\mathrm{d}_3(t)=\sum_{j=0}^{\infty}\hat{\Lambda}_j[\Theta(B)-\hat{\Theta}(B)]\sum_{k=0}^{\infty}\Lambda_k[\widehat{\mathrm{b}}^+(t-j-k)-\mathrm{b}^+(t-j-k)]$$

Since $\sqrt{n}(\hat{\Theta}_j-\Theta_j)$, $j=1,\ldots,q$, are also bounded in probability (by assumption C), there exists for any $\eta>0$ constants $\delta_2>0$ and $0<\rho<1$ such that

$$\begin{aligned}\|\mathrm{d}_3(t)\|&\le\frac{\delta_2}{\sqrt{n}}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\rho^{j+k}\sum_{i=1}^{q}\|\widehat{\mathrm{b}}^+(t-i-j-k)-\mathrm{b}^+(t-i-j-k)\|\\&\le\frac{\delta_2}{\sqrt{n}}\sum_{l=0}^{t-p-2}l\rho^l\sum_{i=1}^{q}\|\widehat{\mathrm{b}}^+(t-i-l)-\mathrm{b}^+(t-i-l)\|\end{aligned}$$

with probability not less than $1-\eta$, for all n sufficiently large. Combining this with the above bound for $\|\widehat{\mathrm{b}}^+(t)-\mathrm{b}^+(t)\|$, $t>p$, one gets that there exists another $\delta>0$ such that

$$\|\mathrm{d}_3(t)\|\le\frac{\delta}{n}\sum_{i=1}^{t-p-2}i\rho^i\left[\sum_{j=2}^{p+q-1}\|\mathrm{W}(t-i-j)\|+\sum_{j=2}^{q+1}\|\mathrm{Y}_2(t-i-j)\|+\frac{1}{\sqrt{n}}\sum_{j=2}^{q+1}\|\mathrm{Y}_1(t-i-j)\|\right]$$

with probability at least $1-2\eta$, for all n sufficiently large. Since $\mathrm{E}\,\|\mathrm{W}(t)\|^2$, $\mathrm{E}\,\|\mathrm{Y}_2(t)\|^2$ and $\mathrm{E}\,\|\mathrm{Y}_1(t))\|^2/t$, $t>p$, are bounded,

$$\sum_{t=p+1}^{n}\mathrm{E}\left\{\sum_{i=1}^{t-p-2}i\rho^i\left[\sum_{j=2}^{p+q-1}\|\mathrm{W}(t-i-j)\|+\sum_{j=2}^{q+1}\|\mathrm{Y}_2(t-i-j)\|+\frac{1}{\sqrt{n}}\sum_{j=2}^{q+1}\|\mathrm{Y}_1(t-i-j)\|\right]\right\}^2\le Cn$$

for some constant C and all n. Therefore, by Markov's inequality, there exists $\Delta>0$ such that

$$\sum_{t=p+1}^{n}\left\{\sum_{i=1}^{t-p-2}i\rho^i\left[\sum_{j=2}^{p+q-1}\|\mathrm{W}(t-i-j)\|+\sum_{j=2}^{q+1}\|\mathrm{Y}_2(t-i-j)\|+\frac{1}{\sqrt{n}}\sum_{j=2}^{q+1}\|\mathrm{Y}_1(t-i-j)\|\right]\right\}^2\le\Delta n$$

with probability at least $1-\eta$. Combining the above results, we get that $\sum_{t=1}^{n}\|\mathrm{d}_3(t)\|^2\le\Delta\delta^2/n$ with probability at least $1-3\eta$. This yields the desired result.

Proof of R2. Write $\hat{\Theta}(B)^{-1} - \Theta(B)^{-1}$ as $\hat{\Theta}(B)^{-1}[\Theta(B) - \hat{\Theta}(B)]\Theta(B)^{-1}$, one gets

$$\begin{aligned}\mathrm{d}_2(t) &= \hat{\Theta}(B)^{-1}[\Theta(B) - \hat{\Theta}(B)]\Theta(B)^{-1}\mathrm{b}^+(t) \\ &= \Theta(B)^{-1}[\Theta(B) - \hat{\Theta}(B)]\Theta(B)^{-1}\mathrm{b}^+(t) \\ &\qquad + \hat{\Theta}(B)^{-1}[\Theta(B) - \hat{\Theta}(B)]\Theta(B)^{-1}[\Theta(B) - \hat{\Theta}(B)]\Theta(B)^{-1}\mathrm{b}^+(t).\end{aligned}$$

Denote by $\widetilde{\mathrm{d}}_2(t)$ and $\widetilde{\widetilde{\mathrm{d}}}_2(t)$ the two terms in the above right hand side. The first term will be further decomposed as

$$\widetilde{\mathrm{d}}_2(t) = \sum_{k=1}^{q}\sum_{l,m=1}^{d}(\hat{\theta}_{lm,k} - \theta_{lm,k})\Theta(B)^{-1}\,\mathrm{E}_{lm}\,\Theta(B)^{-1}\mathrm{b}^+(t-k), \tag{A.2}$$

where $\hat{\theta}_{lm,k}$ and $\theta_{lm,k}$ denote the (l,m) elements of the matrices $\hat{\Theta}_k$ and Θ_k, respectively, and E_{lm} denotes the matrix with 1 at the (l,m) place and zero elsewhere.

We shall prove that

(i) $n\sum_{t=p+1}^{n}\|\widetilde{\widetilde{\mathrm{d}}}_2(t)\|^2$ is bounded in probability. Then, as in the proof of R3, it follows that the same is true for $\sum_{t=p+1}^{n}\widetilde{\widetilde{\mathrm{d}}}_2(t)\mathrm{e}(t)'$.

(ii) For each l, m, k,

$$\frac{1}{n-p}\sum_{t=p+1}^{n}\|\Theta(B)^{-1}\,\mathrm{E}_{lm}\,\Theta(B)^{-1}\mathrm{b}^+(t-k)\|^2$$

and

$$\frac{1}{\sqrt{n-p}}\sum_{t=p+1}^{n}[\Theta(B)^{-1}\,\mathrm{E}_{lm}\,\Theta(B)^{-1}\mathrm{b}^+(t-k)]\mathrm{e}(t)'$$

are bounded in probability as $n\to\infty$. Then by (A.2) and the fact that the $\sqrt{n-p}\times(\hat{\theta}_{lm,k} - \theta_{lm,k})$ converge in distribution, $\sum_{t=p+1}^{n}\|\widetilde{\mathrm{d}}_2(t)\|^2$ and $\sum_{t=p+1}^{n}\widetilde{\mathrm{d}}_2(t)\mathrm{e}(t)'$ are bounded in probability (for the first result, note that $\sum_{t=p+1}^{n}\|\widetilde{\mathrm{d}}_2(t)\|^2$ appears as a squared matrix norm so one can apply the triangular inequality).

We now prove (i). By the same arguments as in the proof of step R3, for any $\eta>0$, there exists constants $\delta>0$, $0<\rho<1$ such that for $t>p$:

$$\|\widetilde{\widetilde{\mathrm{d}}}_2(t)\| \le \frac{\delta}{n}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\rho^{j+k}\sum_{i=1}^{q}\|\Theta(B)^{-1}\mathrm{b}^+(t-i-j-k)\| \le \frac{\delta}{n}\sum_{l=0}^{t-p-2}l\rho^l\sum_{i=1}^{q}\|\Theta(B)^{-1}\mathrm{b}^+(t-i-l)\|$$

with probability not less than $1-\eta$, for all n sufficiently large. On the other hand, $\Theta(B)^{-1}\mathrm{b}^+(t) = \mathrm{a}(t) + \mathrm{d}_4(t)$ for $t>p$ and 0 otherwise, therefore from the result at the beginning of the proof of R4, it can be easily seen that

$$\mathrm{E}\left[\sum_{i=0}^{t-p-2}i\rho^i\sum_{j=1}^{q}\|\Theta(B)^{-1}\mathrm{b}^+(t-i-j)\|\right]^2, \qquad t>p+1,$$

are bounded. Then, using Markov's inequality, one can find a $\Delta>0$ such that

$$\sum_{t=p+1}^{n}\left[\sum_{i=0}^{t-p-2}i\rho^i\sum_{j=1}^{q}\|\Theta(B)^{-1}\mathrm{b}^+(t-i-j)\|\right]^2 \le \Delta n$$

with probability not less than $1-\eta$, for all n. Thus, combining the last two results, one gets that $\sum_{t=p+1}^{n} \|\widetilde{\widetilde{\mathrm{d}}}_4(t)\|^2 \leq \delta^2\Delta/n$, with probability not less than $1-2\eta$. This yields the desired result.

We now prove (ii). Since $\Theta(B)^{-1}\mathrm{b}^+(t) = \mathrm{a}(t) + \mathrm{d}_4(t)$, $t > p$, have bounded expected squared norm, the random variables $n^{-1}\sum_{t=p+1}^{n} \|\Theta(B)^{-1}\,\mathrm{E}_{lm}\,\Theta(B)^{-1}\mathrm{b}^+(t-k)\|^2$ have bounded expectation. Hence they are bounded in probability. On the other hand, since the joint fourth-order cumulant between $a_i(t_1)$, $a_j(t_2)$, $e_l(t)$, $e_m(t)$ vanish, the same is also true for the fourth order joint cumulant between two components of $\Theta(B)^{-1}\,\mathrm{E}_{lm}\,\Theta(B)^{-1}\mathrm{b}^+(t-k)$ and two other components of $\mathrm{e}(t)$. Then by an algebraic calculation, using the formula in Leonov and Shiryaev (1959), the vanishing fourth-order cumulant assumption of the lemma and the fact that the $\mathrm{e}(t)$ have mean zero and the same covariance matrix Σ, say, one gets

$$\begin{aligned}
&\mathrm{E}\left\| \sum_{t=p+1}^{n} [\Theta(B)^{-1}\,\mathrm{E}_{lm}\,\Theta(B)^{-1}\mathrm{b}^+(t-k)]\mathrm{e}(t)' \right\|^2 \\
&\qquad = \mathrm{E}\left[\mathrm{e}(t)' \sum_{t=p+1}^{n} \Theta(B)^{-1}\,\mathrm{E}_{lm}\,\Theta(B)^{-1}\mathrm{b}^+(t-k) \right]^2 \\
&\qquad = \sum_{t=p+1}^{n} \mathrm{E}[\Theta(B)^{-1}\,\mathrm{E}_{lm}\,\Theta(B)^{-1}\mathrm{b}^+(t-k)'\Sigma\Theta(B)^{-1}\,\mathrm{E}_{lm}\,\Theta(B)^{-1}\mathrm{b}^+(t-k)].
\end{aligned}$$

Denote by $|\Sigma|$ the maximum eigenvalue of the non negative matrix Σ, the above right hand side can be bounded by $|\Sigma|\sum_{t=p+1}^{n} \mathrm{E}\,\|\Theta(B)^{-1}\,\mathrm{E}_{lm}\,\Theta(B)^{-1}\mathrm{b}^+(t-k)\|^2$, which can again be bounded by a constant times n, for all n. Therefore by Schwarz's inequality, the random matrices

$$\frac{1}{\sqrt{n-p}} \sum_{t=p+1}^{n} [\Theta(B)^{-1}\,\mathrm{E}_{lm}\,\Theta(B)^{-1}\mathrm{b}^+(t-k)]\mathrm{e}(t)'$$

are bounded in mean squares, hence in probability.

Proof of R1. Recall that $\mathrm{d}_1(t) = \Theta(B)^{-1}[\widehat{\mathrm{b}}^+(t) - \mathrm{b}^+(t)]$ and by the same computation as in the proof of R3 and R2:

$$\begin{aligned}
\widehat{\mathrm{b}}^+(t) - \mathrm{b}^+(t) = &\sum_{k=1}^{p-1}\sum_{i,1=1}^{d} (\hat{\phi}^*_{ij,k} - \phi^*_{ij,k})\,\mathrm{E}_{ij}\,\mathrm{W}(t-k) \\
&- \sum_{i=1}^{d}\sum_{j=1}^{d-r} (\hat{A}_{ij} - A_{ij})\,\mathrm{E}_{ij}\,\mathrm{Y}_2(t-1) - \sum_{i=1}^{d}\sum_{j=1}^{r} (\hat{B}_{ij} - B_{ij})\,\mathrm{E}_{ij}\,\mathrm{Y}_1(t-1)
\end{aligned}$$

for $t > p$, where $\phi^*_{ij,k}$, A_{ij} and B_{ij} are the general terms of the matrices Φ^*_j, CP_2 and CP_1, respectively, and similarly for $\hat{\phi}^*_{ij,k}$, $\hat{A}_{ij}$ and $\hat{B}_{ij}$. Thus expanding $\Theta(z)^{-1}$ into $\sum_{j=0}^{\infty} \Lambda_j B^j$ (as in the proof of R3), one gets

$$\begin{aligned}
\mathrm{d}_1(t) = &\sum_{k=1}^{p-1}\sum_{i,1=1}^{d} (\hat{\phi}^*_{ij,k} - \phi^*_{ij,k}) \sum_{l=0}^{t-p-1} \Lambda_l\,\mathrm{E}_{ij}\,\mathrm{W}(t-k-j) - \sum_{i=1}^{d}\sum_{j=1}^{d-r} (\hat{A}_{ij} - A_{ij}) \sum_{l=0}^{t-p-1} \Lambda_l\,\mathrm{E}_{ij}\,\mathrm{Y}_2(t-1) \\
&- \sum_{i=1}^{d}\sum_{j=1}^{r} (\hat{B}_{ij} - B_{ij}) \sum_{l=0}^{t-p-1} \Lambda_l\,\mathrm{E}_{ij}\,\mathrm{Y}_1(t-1).
\end{aligned}$$

Again, as in the proof of R2, we shall prove that

$$\frac{1}{n-p}\sum_{t=p+1}^{n}\Big\|\sum_{l=0}^{t-p-1}\Lambda_l\,\mathrm{E}_{ij}\,\mathrm{W}(t-k)\Big\|^2, \qquad \frac{1}{\sqrt{n-p}}\sum_{t=p+1}^{n}\Big[\sum_{l=0}^{t-p-1}\Lambda_l\,\mathrm{E}_{ij}\,\mathrm{W}(t-k)\Big]\mathrm{e}(t)'$$

$$\frac{1}{n-p}\sum_{t=p+1}^{n}\Big\|\sum_{l=0}^{t-p-1}\Lambda_l\,\mathrm{E}_{ij}\,\mathrm{Y}_2(t-k)\Big\|^2, \qquad \frac{1}{\sqrt{n-p}}\sum_{t=p+1}^{n}\Big[\sum_{l=0}^{t-p-1}\Lambda_l\,\mathrm{E}_{ij}\,\mathrm{Y}_2(t-k)\Big]\mathrm{e}(t)'$$

$$\frac{1}{n^2}\sum_{t=p+1}^{n}\Big\|\sum_{l=0}^{t-p-1}\Lambda_l\,\mathrm{E}_{ij}\,\mathrm{Y}_1(t-k)\Big\|^2, \qquad \frac{1}{n-p}\sum_{t=p+1}^{n}\Big[\sum_{l=0}^{t-p-1}\Lambda_l\,\mathrm{E}_{ij}\,\mathrm{Y}_1(t-k)\Big]\mathrm{e}(t)'$$

are bounded in probability. Then R1 would follows, since $\sqrt{n-p}\,(\hat{\phi}^*_{ij,k}-\phi_{ij,k})$, $\sqrt{n-p}\,(\hat{\theta}_{ij,k}-\theta_{ij,k})$, $\sqrt{n-p}\,(\hat{A}_{ij}-A_{ij})$ and $(n-p)(\hat{B}_{ij}-B_{ij})$ converge in distribution.

The proof for the above results concerning the random variables $\sum_{l=0}^{t-p-1}\Lambda_l\,\mathrm{E}_{ij}\,\mathrm{W}(t-k)$ and $\sum_{l=0}^{t-p-1}\Lambda_l\,\mathrm{E}_{ij}\,\mathrm{Y}_1(t-k)$ is similar to that of part (ii) in the proof of R3, noting that the above random variables have bounded expected squared norm, for all $t > p$. The random variables $\sum_{l=0}^{t-p-1}\Lambda_l\,\mathrm{E}_{ij}\,\mathrm{Y}_1(t-k)$, $t > p$, however have expected squared norms tending to infinity as $t \to \infty$, since it is the sum of a random walk and a stationary process. But this entails that $(n-p)^{-1} \times \max_{t=p+1,\ldots,n}\mathrm{E}\,\|\mathrm{Y}_1(t)\|^2$ is bounded by a constant independent of n and hence one can still obtain the announced result by a similar calculation. □

Proof of Theorem 3.1. We have

$$\mathrm{C}^{(hh)}_{\widehat{\mathrm{a}}}(0) = \frac{1}{n-p}\sum_{t=p+1}^{n}\widehat{\mathrm{a}}^{(h)}(t)\widehat{\mathrm{a}}^{(h)}(t)',$$

and similarly for $\mathrm{C}^{(hh)}_{\mathrm{a}}(0)$. Hence:

$$\mathrm{C}^{(hh)}_{\widehat{\mathrm{a}}}(0) - \mathrm{C}^{(hh)}_{\mathrm{a}}(0) = \frac{1}{n-p}\sum_{t=p+1}^{n}\{[\widehat{\mathrm{a}}^{(h)}(t)-\mathrm{a}^{(h)}(t)]\mathrm{a}^{(h)}(t)' + \mathrm{a}^{(h)}(t)[\widehat{\mathrm{a}}^{(h)}(t)-\mathrm{a}^{(h)}(t)]'\} + \frac{1}{n-p}\sum_{t=p+1}^{n}[\widehat{\mathrm{a}}^{(h)}(t)-\mathrm{a}^{(h)}(t)][\widehat{\mathrm{a}}^{(h)}(t)-\mathrm{a}^{(h)}(t)]'$$

The first term, by Schwarz's inequality, can be bounded in squared norm by

$$\frac{2}{n-p}\sum_{t=p+1}^{n}\|\widehat{\mathrm{a}}^{(h)}(t)-\mathrm{a}^{(h)}(t)\|^2\frac{1}{n-p}\sum_{t=p+1}^{n}\|\mathrm{a}^{(h)}(t)\|^2.$$

The last term is a positive semi-definite matrix with trace equal to

$$\frac{1}{n-p}\sum_{t=p+1}^{n}\|\widehat{\mathrm{a}}^{(h)}(t)-\mathrm{a}^{(h)}(t)\|^2.$$

Therefore by Lemma 3.1, with $\{\mathrm{a}^{(h)}(t)\}$ in place of $\{\mathrm{a}(t)\}$, both terms tend to zero in probability as $n \to \infty$. Since $\mathrm{C}^{(hh)}_{\mathrm{a}}(0)$ converge in probability to Ω_h, the result (i) follows.

We now prove (ii). For this, it suffices to prove that $\sqrt{n-p}\,(\mathrm{r}^{(12)}_{\widehat{\mathrm{a}}} - \mathrm{r}^{(12)}_{\mathrm{a}}) \to 0$, that is, for each $k \in \{k_1,\ldots,k_m\}$, $\sqrt{n-p}\,[\mathrm{R}^{(12)}_{\widehat{\mathrm{a}}}(k) - \mathrm{R}^{(12)}_{\mathrm{a}}(k)]$ converges in probability to 0 as $n \to \infty$. Denote by D_h and $\widehat{\mathrm{D}}_h$, $h = 1, 2$, the diagonal matrices with diagonal elements those of $\mathrm{C}^{(hh)}_{\mathrm{a}}(0)$ and of $\mathrm{C}^{(hh)}_{\widehat{\mathrm{a}}}(0)$, respectively, then $\mathrm{R}^{(12)}_{\mathrm{a}}(k) = \mathrm{D}_1^{-1}\mathrm{C}^{(12)}_{\mathrm{a}}(k)\mathrm{D}_2^{-1}$ and similarly for $\mathrm{R}^{(12)}_{\widehat{\mathrm{a}}}(k)$. Therefore

$$\mathrm{R}^{(12)}_{\widehat{\mathrm{a}}}(k) - \mathrm{R}^{(12)}_{\mathrm{a}}(k) = \widehat{\mathrm{D}}_1^{-1}[\mathrm{C}^{(12)}_{\widehat{\mathrm{a}}}(k) - \mathrm{C}^{(12)}_{\mathrm{a}}(k)]\widehat{\mathrm{D}}_2^{-1} + \widehat{\mathrm{D}}_1^{-1}\mathrm{D}_1\mathrm{R}^{(12)}_{\mathrm{a}}(k)\mathrm{D}_2\widehat{\mathrm{D}}_2^{-1} - \mathrm{R}^{(12)}_{\mathrm{a}}(k).$$

We already know that $\sqrt{n-p}\,\mathrm{R}_{\mathrm{a}}^{(12)}(k)$ converge in distribution and from the result (i) above, $\widehat{\mathrm{D}}_1^{-1}\mathrm{D}_1$ and $\widehat{\mathrm{D}}_2^{-1}\mathrm{D}_2$ converge in probability to the identity matrix. Hence

$$\sqrt{n-p}\,[\widehat{\mathrm{D}}_1^{-1}\mathrm{D}_1\mathrm{R}_{\mathrm{a}}^{(12)}(k)\mathrm{D}_2\widehat{\mathrm{D}}_2^{-1} - \mathrm{R}_{\mathrm{a}}^{(12)}(k)] \to 0 \quad \text{in probability as } n \to \infty.$$

Thus, since $\widehat{\mathrm{D}}_1^{-1}$ and $\widehat{\mathrm{D}}_2^{-1}$ converge in probability, one would have obtained the desired result if one had shown that $\sqrt{n-p}\,[\mathrm{C}_{\widehat{\mathrm{a}}}^{(12)}(k) - \mathrm{C}_{\mathrm{a}}^{(12)}(k)] \to 0$ in probability as $n \to \infty$. For this end, write

$$\begin{aligned}\sqrt{n-p}\,[\mathrm{C}_{\widehat{\mathrm{a}}}^{(12)}(k) - \mathrm{C}_{\mathrm{a}}^{(12)}(k)] = {} & \frac{1}{\sqrt{n-p}} \sum_{t=p+1+\max(0,k)}^{n+\min(0,k)} [\widehat{\mathrm{a}}^{(1)}(t) - \mathrm{a}^{(1)}(t)]\mathrm{a}^{(2)}(t-k)' \\ & + \frac{1}{\sqrt{n-p}} \sum_{t=p+1+\max(0,k)}^{n+\min(0,k)} \mathrm{a}^{(1)}(t)[\widehat{\mathrm{a}}^{(2)}(t-k) - \mathrm{a}^{(2)}(t-k)]' \\ & + \frac{1}{\sqrt{n-p}} \sum_{t=p+1+\max(0,k)}^{n+\min(0,k)} [\widehat{\mathrm{a}}^{(1)}(t) - \mathrm{a}^{(1)}(t)][\widehat{\mathrm{a}}^{(2)}(t-k) - \mathrm{a}^{(2)}(t-k)]'.\end{aligned}$$

We now apply Lemma 3.1 to the first term of the above right hand side, with $\{\mathrm{a}^{(1)}(t)\}$ and $\{\mathrm{a}^{(2)}(t)\}$ playing the role of $\{\mathrm{a}(t)\}$ and $\{\mathrm{e}(t)\}$. We need to ensure that the vanishing fourth-order cumulant assumption of this lemma is satisfied. But, using the formula in Leonov and Shiryaev (1959), it can be seen that this is a consequence of assumption A and the fact that the cumulants κ_{ijlm} vanish. Thus, the first term of the above right hand side converges to 0 in probability as $n \to \infty$. The reasoning is similar for the second term. Finally, the squared norm of the last term may be bounded, using Schwarz's inequality, by

$$\frac{1}{n-p} \sum_{t=p+1}^{n} \|\widehat{\mathrm{a}}^{(1)}(t) - \mathrm{a}^{(1)}(t)\|^2 \sum_{t=p+1}^{n} \|\widehat{\mathrm{a}}^{(2)}(t) - \mathrm{a}^{(2)}(t)\|^2.$$

Applying again Lemma 3.1 with $\{\mathrm{a}^{(1)}(t)\}$ and then $\{\mathrm{a}^{(2)}(t)\}$ playing the role of $\{\mathrm{a}(t)\}$, one gets that the above expression also converge to 0 in probability as $n \to \infty$. □

References

Ahn, S. K. and Reinsel, G. C. (1990). Estimation for partially nonstationary multivariate autoregressive models. *J. Amer. Statist. Assoc.* 85, 813–823.

Chitturi, R. V. (1976). Distribution of multivariate white noise autocorrelations. *J. Amer. Statist. Assoc.* 71, 224–226.

Dunsmuir, W. and Hannan, E. J. (1976). Vector linear time series models. *Adv. Appl. Probab.* 8, 339–364.

El Himdi, K. and Roy, R. (1997). Tests for non-correlation of two multivariate ARMA time series. *Canad. J. Statist.* 25, 233–256.

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: representation, estimation and testing. *Econometrica* 55, 251–276.

Fuller, W. A. (1996). *Introduction to Statistical Time Series*. Second edition. Wiley, New York.

Hallin, M., Jurečková, J., Picek, J. and Zahaf, T. (1999). Nonparametric tests of independence of two autoregressive time series based on autoregression rank scores. *J. Statist. Plan. Infer.* 75, 319–330.

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

Hannan, E. J. (1975). The estimation of ARMA models. *Ann. Statist.* 3, 975–981.

Hannan, E. J. (1976). The asymptotic distribution of serial covariances. *Ann. Statist.* 4, 396–399.

Haugh, L. D. (1976). Checking the independence of two covariance-stationary time series : a univariate residual cross-correlation approach. *J. Amer. Statist. Assoc.* 71, 378–385.

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

Leonov, V. P. and Shiryaev, A. N. (1959). On a method of calculation of semi-invariants. *Theory Prob. Appl.* 4, 319–329.

Li, W. K. and Hui, Y. V. (1994). Robust residual cross correlation tests for lagged relations in time series. *J. Statist. Comput. Simul.* 49, 103–109.

Li, W. K. and McLeod, A. I. (1981). Distribution of the residual autocorrelations in multivariate ARMA time series models. *J. Roy. Statist. Soc. B* 43, 231–239.

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

McLeod, A. I. (1979). Distribution of the residual cross-correlation in univariate ARMA time series models. *J. Amer. Statist. Assoc.* 74, 849–855.

Racette, D. and Raynauld, J. (1992). Un modèle BVAR de prévision de la dépense nominale et d'analyse de la politique monétaire canadienne. *Séminaire sur les questions monétaires*. Banque du Canada, pp. 347–425

Reinsel, G. C. (1993). *Elements of Multivariate Time Series Analysis*, Springer-Verlag, Berlin.

Roy, R. (1989). Asymptotic covariance structure of serial correlation in multivariate time series. *Biometrika* 76, 824–827.

Saikkonen, P (1992). Estimation and testing of cointegrated systems by an autoregressive approximation. *Econometric Theory* 8, 1–27.

Tiao, G. C. and Box, G. E. P. (1981). Modeling multiple time series with applications. *J. Amer. Statist. Assoc.* 76, 802–816.

Tsay, R. S. and Tiao, G. C. (1990). Asymptotic properties of multivariate nonstationary processes with applications to autoregressions. *Ann. Statist.* 18, 220–250.

Yap, S. F. and Reinsel, G. C. (1995). Estimation and testing for unit roots in a partially nonstationary vector autoregressive moving average model. *J. Amer. Statist. Assoc.* 90, 429, 253–267.

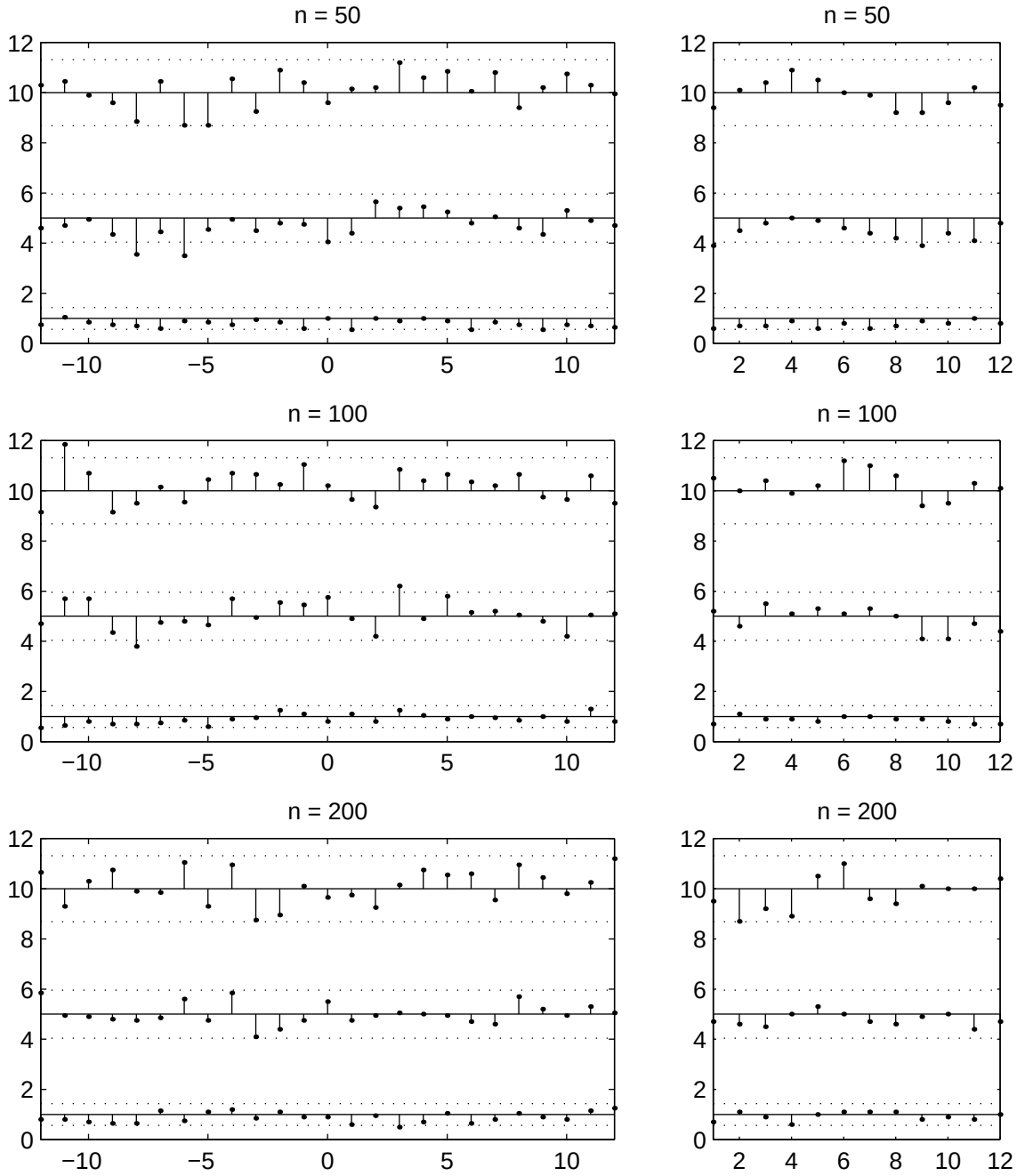

Figure 1: Empirical level of tests at individual lags $\mathrm{QH}^*(k)$ (left) and of the global test QH^*_M (right) respectively defined by (3.7) and (3.8) for the AR(1) model.

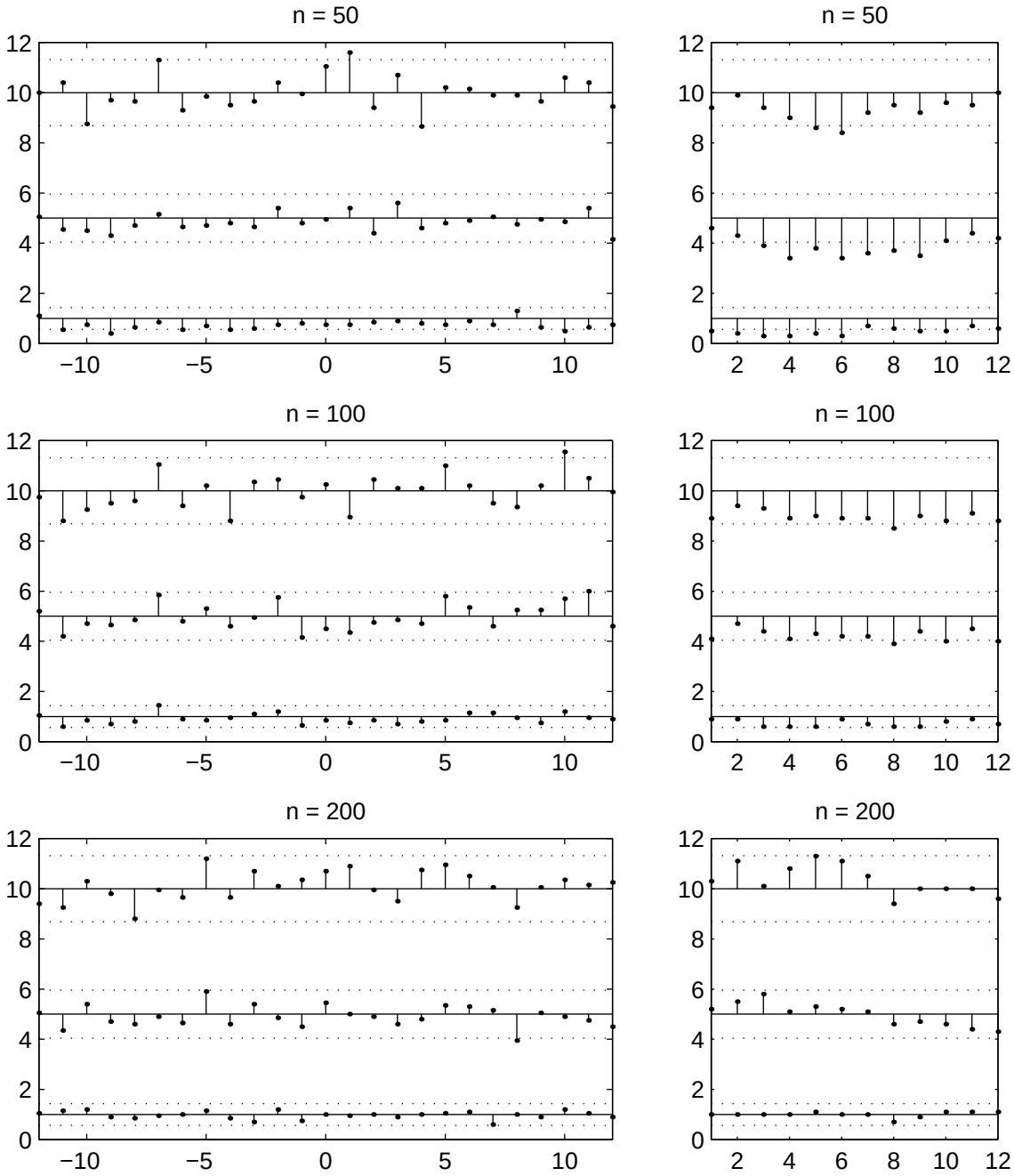

Figure 2: Empirical level of tests at individual lags $\mathrm{QH}^*(k)$ (left) and of the global test QH^*_M (right) respectively defined by (3.7) and (3.8) for the AR(2) model.

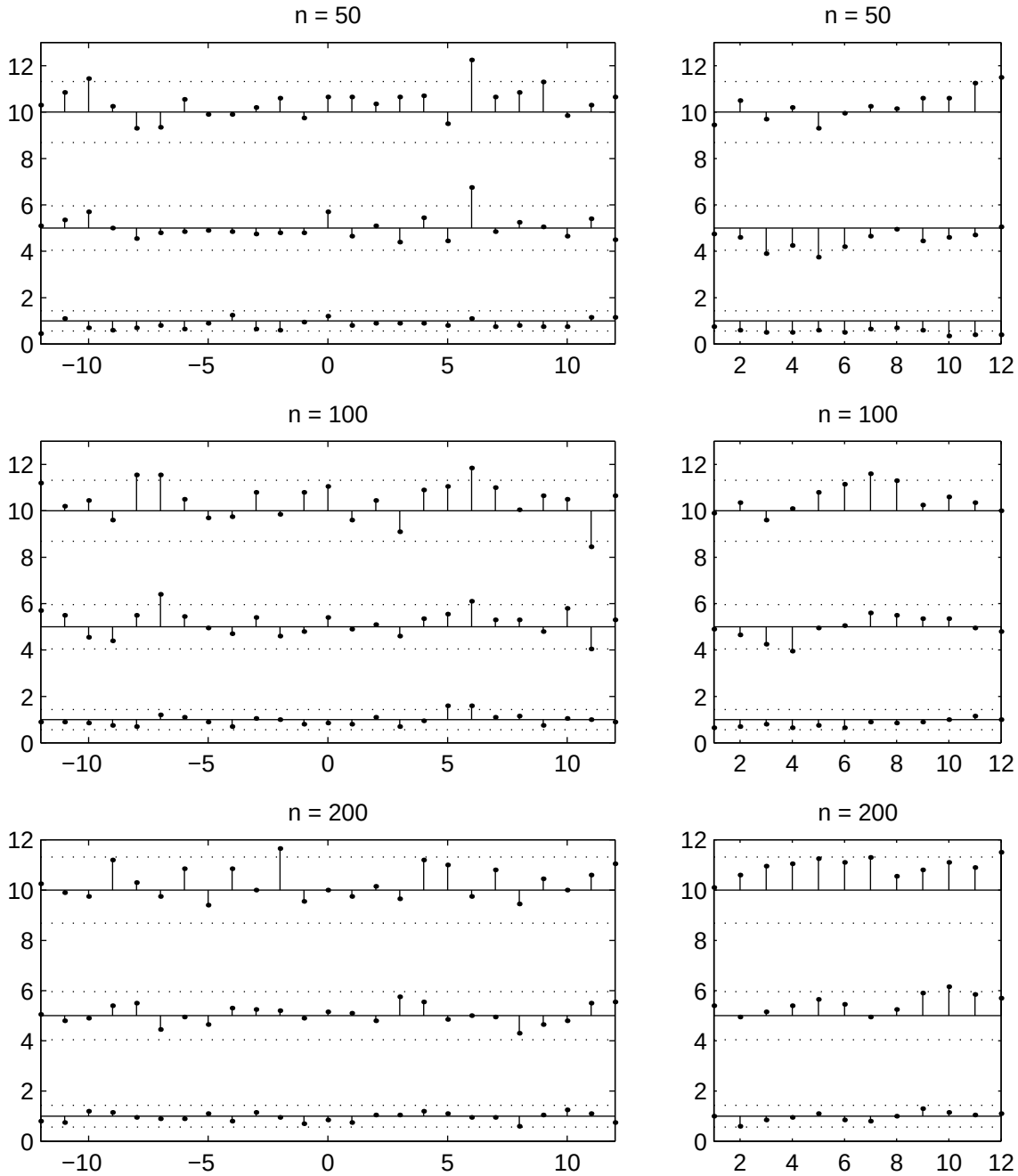

Figure 3: Empirical level of tests at individual lags $\mathrm{QH}^*(k)$ (left) and of the global test QH^*_M (right) respectively defined by (3.7) and (3.8) for the ARMA(1) model.

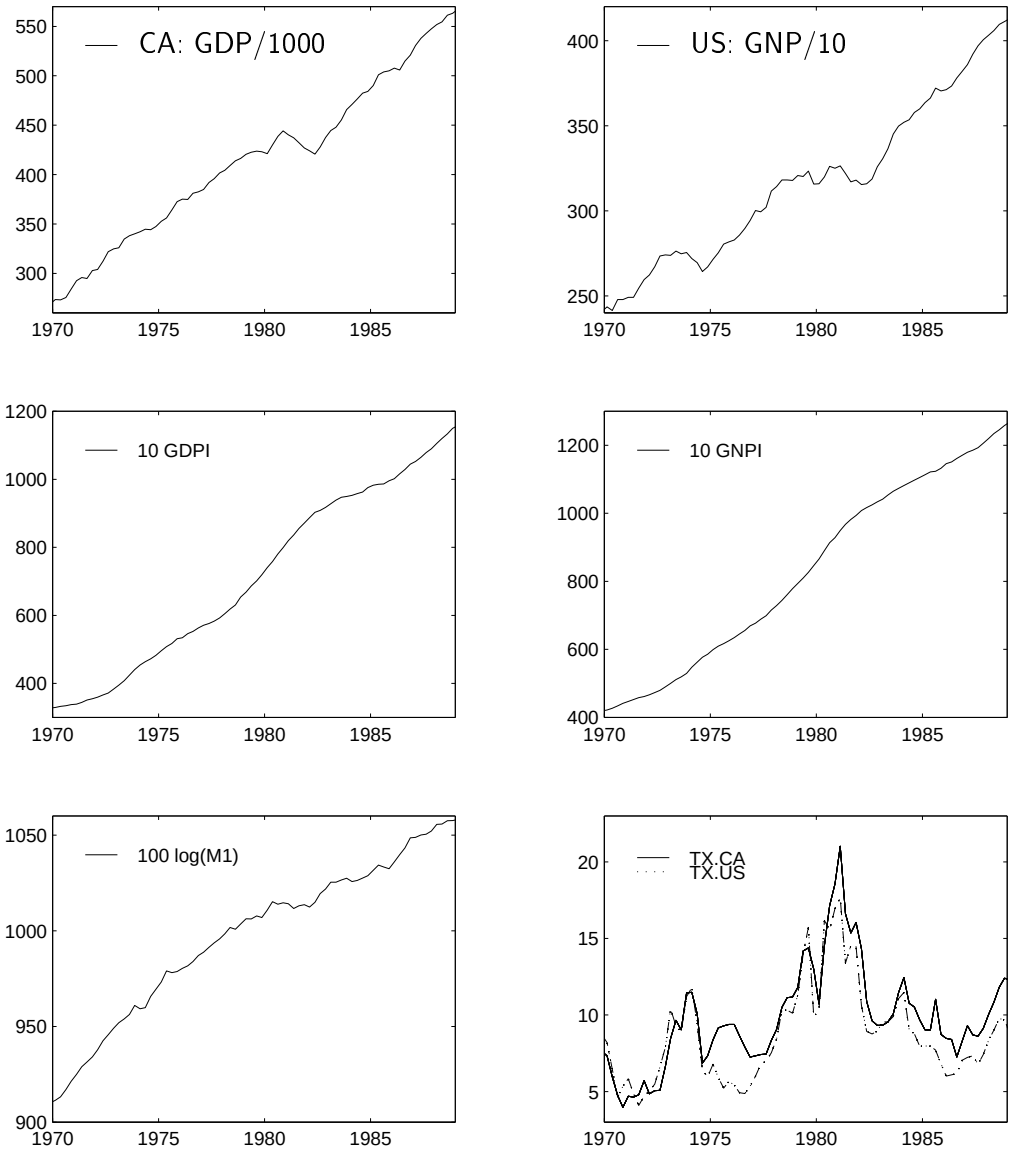

Figure 4: Graphs of the components of the Canadian and United States series.

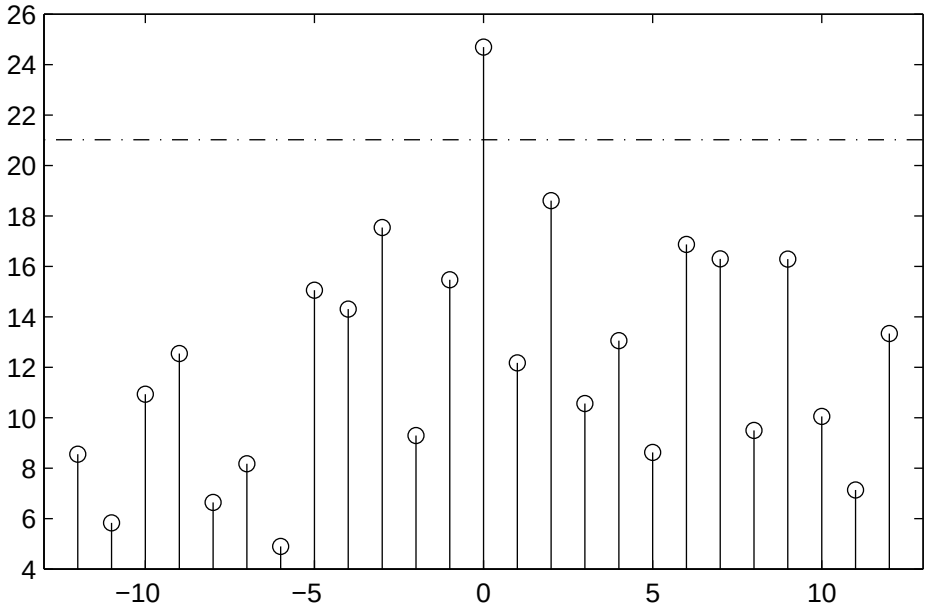

Figure 5: Values of the statistic $\mathrm{QH}^*(k)$ defined by (3.7) for different lags k. The dotted line represent the critical value at $\alpha = 0.05$.