Ann. Inst. Statist. Math.
Vol. 57, No. 3, 575–595 (2005)

TESTING FOR SERIAL CORRELATION OF UNKNOWN FORM IN COINTEGRATED TIME SERIES MODELS

PIERRE DUCHESNE

Département de mathématiques et de statistique, Université de Montréal, CP 6128, Succursale Centre-Ville, Montréal, Québec, Canada H3C 3J7, e-mail: duchesne@dms.umontreal.ca

(Received November 25, 2003; revised July 12, 2004)

Abstract. Portmanteau test statistics are useful for checking the adequacy of many time series models. Here we generalize the omnibus procedure proposed by Duchesne and Roy (2004, *Journal of Multivariate Analysis*, **89**, 148–180) for multivariate stationary autoregressive models with exogenous variables (VARX) to the case of cointegrated (or partially nonstationary) VARX models. We show that for cointegrated VARX time series, the test statistic obtained by comparing the spectral density of the errors under the null hypothesis of non-correlation with a kernel-based spectral density estimator, is asymptotically standard normal. The parameters of the model can be estimated by conditional maximum likelihood or by asymptotically equivalent estimation procedures. The procedure relies on a truncation point or a smoothing parameter. We state conditions under which the asymptotic distribution of the test statistic is unaffected by a data-dependent method. The finite sample properties of the test statistics are studied via a small simulation study.

Key words and phrases: Vector autoregressive process, cointegration, exogenous variables, kernel spectrum estimator, diagnostic test, portmanteau test.

1. Introduction

Many real life situations can be described by vector autoregressive models with exogenous variables (VARX). These models represent a generalization of the popular vector autoregressive models (VAR) in the sense that explanatory variables (or exogenous variables) can be included, which is a desirable property in many practical situations; see for econometric applications Judge *et al.* (1985) and Lütkepohl (1993), among others. Statistical properties of VARX models are described in Hannan and Deistler (1988), among others. Many economic and financial time series exhibit characteristics that are believed to be nonstationary. An important nonstationary model of considerable practical importance occurs when the determinant of the autoregressive operator admits unit roots, while all other roots are outside the unit circle. With multivariate VARX models with unit roots, a classical approach consists to differentiate each component in order to obtain stationarity. However, differencing the series tends to introduce complications, as for example rank deficiency in coefficient matrices and noninvertibility problems (see, e.g., Lütkepohl (1982) and Ahn and Reinsel (1990)). Therefore, it appears important to be able to directly describe and model the nonstationarity, since this approach gives an increased understanding of the nature of the nonstationarities (see, e.g., Granger (1981), Engle and Granger (1987)). Furthermore, direct modelling of the unit roots provides more accurate long-term forecasts (Yap and Reinsel (1995)).

Diagnostic checking of VARX models (stationary or partially nonstationary) appears to be a critical step. Test procedures based on the residual autocovariances and/or autocorrelations have been found useful for diagnosing the adequacy of time series models. In stationary VARMA models, Hosking (1980) proposed a test procedure based on the residual autocovariances. In that approach, VARMA models are fitted and the test statistic relies on the residual autocovariance matrices of the residual time series. Duchesne and Roy (2004) proposed test statistics, using a spectral approach, for checking the adequacy of VARX models using kernel-based spectral density estimators of the residual time series. Their test statistic is based on a distance measure between a nonparametric kernel-based spectral density estimator and the spectral density under the null hypothesis of non-correlation. It can be expressed as a weighted sum of a function of the residual autocovariance matrices at all lags j, $1 \leq j \leq n-1$, n being the sample size. The kernel-based test statistic depends on a kernel function $k(\cdot)$ and a truncation point or a smoothing parameter, noted p_n. When the truncated uniform kernel is adopted, the resulting test statistic corresponds to a generalized Hosking's (1980) test. However, many kernels $k(\cdot)$ exhibit better power than the uniform weighting offered by the truncated uniform kernel.

The main contribution of this paper is to establish the asymptotic distribution of the kernel-based test statistic in the case of a cointegrated VARX model. Using the error correction representation, the parameters of the VARX model can be estimated by the conditional maximum likelihood method, or with other methods which are asymptotically equivalent. Such estimation procedures have been studied in Ahn and Reinsel (1990) for VAR models, and we discuss how to adapt them when exogenous variables are included in the model. We will show that the kernel-based test statistic introduced in Duchesne and Roy (2004) admits a standard normal distribution under the null hypothesis of adequacy.

The method of proof differs considerably from the one developed in Duchesne and Roy (2004). In partially nonstationary models, we exploit the error correction form of the cointegrated VARX model; such a process can be written as the sum of a nonstationary process with stationary increments and of a stationary process. The decomposition has been studied in Ahn and Reinsel (1990) in VAR models. See also Yap and Reinsel (1995) and Pham *et al.* (2003) in VARMA models. To handle properly the nonstationary component represents one of the technical achievements of the paper.

In practice, the choice of the smoothing parameter in the kernel-based test statistic needs to be determined. In the stationary VARX model, Duchesne and Roy (2004) considered in their empirical study the cross-validation procedure. However, their asymptotic analysis establishing the asymptotic distribution of their test statistic assumed a non-stochastic p_n. Here, we would like to study the asymptotic distribution of the test statistic when p_n is determined by the observed data. Consequently, a second objective of the present paper is to state precise conditions under which the asymptotic distribution of the kernel-based test statistic is not affected by a data-dependent method. The method of proof is valid for the cointegrated VARX model and it can be easily adapted for the stationary case. Under the stated conditions, the test statistic admits a standard normal distribution under the null hypothesis of adequacy, when the smoothing parameter has been determined with the data available.

The paper is organized as follows. In Section 2, some preliminaries are introduced, where the cointegrated VARX model is presented and the sample residual autocovariance matrices are defined. In Section 3, we give the hypotheses of interest for model adequacy. The estimation of the cointegrated VARX model is considered and we present the test procedure for diagnostic checking. The asymptotic distribution of our portmanteau procedure is derived under the null hypothesis of non correlation in the error term. In Section 4, we discuss data-driven methods for the bandwidth. We give precise conditions under which the asymptotic distribution of our kernel-based test statistic, based on a data-driven smoothing parameter, is still standard normal under the null hypothesis. Some simulation results are reported in Section 5. The proofs of our results are contained in the Appendix.

2. Preliminaries

Let $\boldsymbol{Y} = \{\boldsymbol{Y}_t : t \in \mathbb{Z}\}$ and $\boldsymbol{X} = \{\boldsymbol{X}_t : t \in \mathbb{Z}\}$ be two multivariate processes, where $\boldsymbol{Y}_t = (Y_t(1), \ldots, Y_t(d))'$ and $\boldsymbol{X}_t = (X_t(1), \ldots, X_t(m))'$. We assume that $\boldsymbol{X}$ is of mean $\boldsymbol{0}$. We suppose that these processes can be represented by a multivariate linear dynamic model with second-order stationary exogenous variables VARX(p,s), defined by:

$$\boldsymbol{\Lambda}(B)\boldsymbol{Y}_t = \boldsymbol{V}(B)\boldsymbol{X}_t + \boldsymbol{a}_t, \tag{2.1}$$

where $\boldsymbol{\Lambda}(B) = \boldsymbol{I}_d - \sum_{i=1}^{p} \boldsymbol{\Lambda}_i B^i$, $\boldsymbol{\Lambda}_p \neq \boldsymbol{0}$, $\boldsymbol{I}_d$ being the $d \times d$ identity matrix, $\boldsymbol{V}(B) = \sum_{i=0}^{s} \boldsymbol{V}_i B^i$, $\boldsymbol{V}_s \neq \boldsymbol{0}$, B denoting the usual backward shift operator, $\boldsymbol{a} = \{\boldsymbol{a}_t, t \in \mathbb{Z}\}$ is a white noise, that is $\boldsymbol{a}_t = (a_t(1), \ldots, a_t(d))'$, $t \in \mathbb{Z}$, are identically and independent distributed (iid) random vectors with mean zero and regular covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{a}} = (\sigma_{\boldsymbol{a},ij})_{i,j=1,\ldots,d}$. The process $\boldsymbol{X}$ is supposed strictly exogenous, that is $\boldsymbol{X}$ and $\boldsymbol{a}$ are assumed independent.

Partial nonstationarity of the process $\boldsymbol{Y}$ is assumed, as characterized by the existence of $d_0 < d$ unit roots in the VAR operator $\boldsymbol{\Lambda}(B)$. More precisely, $\det\{\boldsymbol{\Lambda}(B)\} = 0$ admits $d_0 < d$ unit roots, where $\det\{\boldsymbol{A}\}$ stands for the determinant of the square matrix $\boldsymbol{A}$. However, it is supposed that the remaining roots are outside the unit circle. The rank of $\boldsymbol{\Lambda}(1)$ is r, that is $\operatorname{rank}\{\boldsymbol{\Lambda}(1)\} = r$, $r = d - d_0 > 0$. Now let $\boldsymbol{P} = (\boldsymbol{P}_1, \boldsymbol{P}_2)$, with $\boldsymbol{P}_1$ a full rank $d \times d_0$ matrix such that $\boldsymbol{\Lambda}(1)\boldsymbol{P}_1 = \boldsymbol{0}$, $\boldsymbol{P}_2$ being a $d \times r$ matrix such that $\boldsymbol{P}$ is nonsingular. Partitioning $\boldsymbol{P}^{-1} = \boldsymbol{Q}$ into $\boldsymbol{Q} = (\boldsymbol{Q}_1', \boldsymbol{Q}_2')'$, and noting that $\boldsymbol{P}_1\boldsymbol{Q}_1 + \boldsymbol{P}_2\boldsymbol{Q}_2 = \boldsymbol{I}_d$, it is possible to show, using the arguments of Pham *et al.* (2003) that $\boldsymbol{Y}_t = \boldsymbol{P}_1\boldsymbol{Z}_{1t} + \boldsymbol{P}_2\boldsymbol{Z}_{2t}$, where $\{\boldsymbol{Z}_{1t} = \boldsymbol{Q}_1\boldsymbol{Y}_t\}$ is nonstationary with $\{\boldsymbol{Z}_{1t} - \boldsymbol{Z}_{1,t-1}\}$ stationary, while $\{\boldsymbol{Z}_{2t} = \boldsymbol{Q}_2\boldsymbol{Y}_t\}$ is stationary.

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

$$\boldsymbol{\Gamma}_{\boldsymbol{u}}(j) = E(\boldsymbol{u}_t\boldsymbol{u}_{t-j}'), \qquad j \in \mathbb{Z}.$$

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

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

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

$$\boldsymbol{f}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \boldsymbol{\Gamma}_{\boldsymbol{u}}(h) e^{-i\omega h}, \qquad \omega \in [-\pi, \pi]. \tag{2.2}$$

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

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

and

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

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

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

$$\boldsymbol{C}_{\boldsymbol{u}}(j) = \begin{cases} n^{-1}\sum_{t=j+1}^{n} \boldsymbol{u}_t\boldsymbol{u}'_{t-j}, & j = 0,1,\dots,n-1, \\ \boldsymbol{C}'_{\boldsymbol{u}}(-j), & j = -1,\dots,-n+1. \end{cases} \tag{2.3}$$

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

$$\boldsymbol{f}_n(\omega) = \frac{1}{2\pi}\sum_{j=-n+1}^{n-1} k(j/p_n)\boldsymbol{C}_{\boldsymbol{u}}(j)e^{-i\omega j}, \tag{2.4}$$

where $k(\cdot)$ is a kernel or a lag window. The parameter p_n is a truncation point when the kernel is of compact support, or a smoothing parameter when the kernel support is unbounded. We suppose that $p_n \to \infty$ and $p_n/n \to 0$. Examples of p_n satisfying these criteria are $p_n \propto n^\alpha$, with $\alpha \in (0,1)$, and $p_n \propto \log(n)$. Using the rectangular or truncated uniform kernel $k_{TR}(z) = I[|z| \le 1]$, where $I(A)$ is the indicator function of the set A, we retrieve the familiar truncated periodogram (Priestley (1981), Section 6.2.3). Here, we will also use kernel-based estimators of the form (2.4) with the usual assumptions on the kernel that are summarized as follows.

ASSUMPTION 2.1. The kernel $k : \mathbb{R} \to [-1,1]$ is a symmetric function, continuous at 0, having at most a finite number of discontinuity points and such that $k(0) = 1$, $\int_{-\infty}^{\infty} k^2(z)dz < \infty$.

Examples of kernels or lag windows frequently used in time series analysis are studied in Priestley ((1981), Section 6.2.3), among others.

Once a cointegrated VARX model is estimated, the residual time series $\hat{\boldsymbol{a}}_t$, $t = 1,\dots,n$, can be computed. The residual autocovariance at lag j is obtained from (2.3) with $\boldsymbol{u}_t$ replaced by $\hat{\boldsymbol{a}}_t$ and it is noted $\boldsymbol{C}_{\hat{\boldsymbol{a}}}(j)$, $j = -n+1,\dots,n-1$. Similarly, the residual spectral density estimator $\hat{\boldsymbol{f}}_n(\omega)$ is obtained from (2.4) where $\boldsymbol{C}_{\hat{\boldsymbol{a}}}(j)$ is substituted for $\boldsymbol{C}_{\boldsymbol{u}}(j)$, that is

$$\hat{\boldsymbol{f}}_n(\omega) = \frac{1}{2\pi}\sum_{j=-n+1}^{n-1} k(j/p_n)\boldsymbol{C}_{\hat{\boldsymbol{a}}}(j)e^{-i\omega j}. \tag{2.5}$$

The estimation of the cointegrated VARX model, the hypotheses of interest for diagnosing the model, the kernel-based test statistic and its asymptotic distribution are discussed in the following section.

3. Estimation of the model and the test statistic for adequacy

The hypothesis of interest states that the error process $\boldsymbol{a}$ is a white noise against the alternative of serial correlation of arbitrary form. More formally, it can be written as

$$H_0 : \boldsymbol{\Gamma}_{\boldsymbol{a}}(j) = \boldsymbol{0}, \ \forall j \neq 0, \text{ against}$$
$$H_1 : \boldsymbol{\Gamma}_{\boldsymbol{a}}(j) \neq \boldsymbol{0}, \text{ for at least one } j \neq 0.$$

The hypotheses can be written in terms of the spectral density $\boldsymbol{f}(\omega)$ of $\boldsymbol{a}$. For example, H_0 can be expressed as $\boldsymbol{f}(\omega) = \boldsymbol{f}_0(\omega)$, $\omega \in [-\pi, \pi]$, where $\boldsymbol{f}_0(\omega) = \boldsymbol{\Gamma}_{\boldsymbol{a}}(0)/(2\pi) = \boldsymbol{\Sigma}_{\boldsymbol{a}}/(2\pi)$, $\omega \in [-\pi, \pi]$.

We now discuss estimation of the cointegrated VARX model. The model (2.1) can be written as:

$$\boldsymbol{\Lambda}^*(B)\,\boldsymbol{W}_t = \boldsymbol{C}\,\boldsymbol{Y}_{t-1} + \boldsymbol{V}(B)\boldsymbol{X}_t + \boldsymbol{a}_t, \tag{3.1}$$

where $\boldsymbol{W}_t = (1-B)\boldsymbol{Y}_t$, $\boldsymbol{C} = -\boldsymbol{\Lambda}(1)$, $\boldsymbol{\Lambda}^*(B) = \boldsymbol{I}_d - \sum_{j=1}^{p-1} \boldsymbol{\Lambda}_j^* B^j$, $\boldsymbol{\Lambda}_j^* = -\sum_{k=j+1}^{p} \boldsymbol{\Lambda}_k$ and $\boldsymbol{\Lambda}(1) = \boldsymbol{I}_d - \sum_{j=1}^{p} \boldsymbol{\Lambda}_j$. Equation (3.1) represents the error correction form of the model (2.1) (see, e.g., Granger and Weiss (1983), Engle and Granger (1987), Ahn and Reinsel (1990)). As noted by Yap and Reinsel (1995), a model form as the one given by (3.1) is particularly convenient, since the nonstationarity of the VAR operator $\boldsymbol{\Lambda}(B)$ is concentrated in the behavior of the matrix $\boldsymbol{C}$.

Let $\boldsymbol{\Theta} = (\boldsymbol{C}, \boldsymbol{\Lambda}_1^*, \ldots, \boldsymbol{\Lambda}_{p-1}^*, \boldsymbol{V}_0, \boldsymbol{V}_1, \ldots, \boldsymbol{V}_s)$ be a $d \times (dp + m(s+1))$ matrix of parameters and $\boldsymbol{H}_t = (\boldsymbol{Y}_{t-1}', \boldsymbol{W}_{t-1}', \ldots, \boldsymbol{W}_{t-p+1}', \boldsymbol{X}_t', \ldots, \boldsymbol{X}_{t-s}')'$ be a $(dp + m(s+1)) \times 1$ vector. The model (3.1) can be written in the compact form $\boldsymbol{W}_t = \boldsymbol{\Theta}\boldsymbol{H}_t + \boldsymbol{a}_t$ and the least squares (LS) estimator $\hat{\boldsymbol{\Theta}}$ of $\boldsymbol{\Theta}$ is given by

$$\hat{\boldsymbol{\Theta}} = \left(\sum \boldsymbol{W}_t \boldsymbol{H}_t'\right)\left(\sum \boldsymbol{H}_t \boldsymbol{H}_t'\right)^{-1}.$$

The LS estimator of the error covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{a}}$ is given by $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{a}} = \boldsymbol{C}_{\hat{\boldsymbol{a}}}(0)$, where $\hat{\boldsymbol{a}}$, $t = 1, \ldots, n$ denote the LS residuals. It can be shown, using an approach similar to Ahn and Reinsel (1990), and using results of Phillips and Durlauf (1986) and Sims *et al.* (1990), that the convergence rates of $\hat{\boldsymbol{C}}$, $\hat{\boldsymbol{\Lambda}}_i^*$, $i = 1, \ldots, p-1$, $\hat{\boldsymbol{V}}_i$, $i = 0, \ldots, s$ are such that

$$\begin{aligned}
&(\hat{\boldsymbol{C}} - \boldsymbol{C})\boldsymbol{P}_1 = \boldsymbol{O}_p(n^{-1}), \qquad (\hat{\boldsymbol{C}} - \boldsymbol{C})\boldsymbol{P}_2 = \boldsymbol{O}_p(n^{-1/2}),\\
&\hat{\boldsymbol{\Lambda}}_i^* - \boldsymbol{\Lambda}_i^* = \boldsymbol{O}_p(n^{-1/2}), \qquad i = 1, \ldots, p-1,\\
&\hat{\boldsymbol{V}}_i - \boldsymbol{V}_i = \boldsymbol{O}_p(n^{-1/2}), \qquad i = 0, \ldots, s.
\end{aligned} \tag{3.2}$$

The matrices $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are defined in Section 2. The mathematical developments that follow are valid for any estimator of $\boldsymbol{\Theta}$ that verifies (3.2). For this reason, it is hypothesized that our estimation method satisfies the following assumption for the estimator $\hat{\boldsymbol{\Theta}}$.

ASSUMPTION 3.1. The estimator $\hat{\boldsymbol{\Theta}}$ of $\boldsymbol{\Theta}$ in the VARX model satisfies (3.2).

Remark 1. Often, there are linear constraints on the parameters, for example parameter values that are fixed to zero. Let $\boldsymbol{\theta} = \mathrm{vec}(\boldsymbol{\Theta}) = \boldsymbol{R}\boldsymbol{\gamma}$, for a known matrix $\boldsymbol{R}$.

Then, the model (3.1) can be written as $\boldsymbol{W}_t = (\boldsymbol{H}_t' \otimes \boldsymbol{I})\boldsymbol{R}\gamma + \boldsymbol{a}_t$ and the LS estimator $\hat{\gamma}$ can be derived easily. In this situation, the LS estimator of $\boldsymbol{\Theta}$ is given by $\hat{\boldsymbol{\Theta}} = \boldsymbol{R}\hat{\gamma}$.

Explicit imposition of unit roots may lead to considerable improvement on prediction performance. Consequently, to incorporate the d_0 unit roots in the estimation of the partially nonstationary VARX model may be highly desirable. Estimation of the error correction model (3.1) with the reduced rank structure $\boldsymbol{C} = \boldsymbol{AB}$ imposed, where $\boldsymbol{A}$ and $\boldsymbol{B}$ are full-rank matrices of dimension $d \times r$ and $r \times d$, has been investigated in Ahn and Reinsel (1990) and Yap and Reinsel (1995) in VAR and VARMA models. For a unique parameterization, the matrix $\boldsymbol{B}$ can be normalized such that $\boldsymbol{B} = (\boldsymbol{I}_r, \boldsymbol{B}_0)$, where $\boldsymbol{B}_0$ is an $r \times d_0$ matrix of parameters. This leads to the reduced rank model

$$\boldsymbol{\Lambda}^*(B)\boldsymbol{W}_t = \boldsymbol{AB}\boldsymbol{Y}_{t-1} + \boldsymbol{V}(B)\boldsymbol{X}_t + \boldsymbol{a}_t. \tag{3.3}$$

Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_0', \boldsymbol{\alpha}')'$, where $\boldsymbol{\beta}_0 = \mathrm{vec}(\boldsymbol{B}_0)$ and $\boldsymbol{\alpha} = \mathrm{vec}(\boldsymbol{A}, \boldsymbol{\Lambda}_1^*, \ldots, \boldsymbol{\Lambda}_{p-1}^*, \boldsymbol{V}_0, \ldots, \boldsymbol{V}_s)$. Based on the n observations $\{\boldsymbol{Y}_t, t = 1, \ldots, n\}$, the Gaussian log-likelihood becomes

$$l(\boldsymbol{\beta}, \boldsymbol{\Sigma}_a) = -\frac{n}{2}\log\det\boldsymbol{\Sigma}_a - \frac{1}{2}\sum_{t=1}^{n}\boldsymbol{a}_t'\boldsymbol{\Sigma}_a^{-1}\boldsymbol{a}_t.$$

Taking the same approach that Ahn and Reinsel (1988, 1990), Gaussian estimators are obtained using the Newton-Raphson algorithm, which is based on an initial consistent estimator satisfying Assumption 3.1. Using a first-order Taylor expansion similar to Ahn and Reinsel (1990), an asymptotic representation is obtained and we can establish that the Gaussian estimator $\tilde{\boldsymbol{\beta}}$ satisfies

$$\begin{aligned} &\tilde{\boldsymbol{A}} - \boldsymbol{A} = \boldsymbol{O}_p(n^{-1/2}), \qquad \tilde{\boldsymbol{B}}_0 - \boldsymbol{B}_0 = \boldsymbol{O}_p(n^{-1}), \\ &\tilde{\boldsymbol{\Lambda}}_i^* - \boldsymbol{\Lambda}_i^* = \boldsymbol{O}_p(n^{-1/2}), \qquad i = 1, \ldots, p-1, \\ &\tilde{\boldsymbol{V}}_i - \boldsymbol{V}_i = \boldsymbol{O}_p(n^{-1/2}), \qquad i = 0, \ldots, s. \end{aligned} \tag{3.4}$$

The assumptions on the reduced rank estimator are summarized as follows.

Assumption 3.2. The estimator $\tilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ in the VARX model satisfies (3.4).

Other estimation methods verify Assumption 3.2, as discussed in Ahn and Reinsel ((1990), Section 8) in the context of VAR estimation. Let the LS estimators of $\boldsymbol{C}$ be written as $\hat{\boldsymbol{C}} = (\hat{\boldsymbol{C}}_1, \hat{\boldsymbol{C}}_2)$, where $\tilde{\boldsymbol{A}}_{alt} = \hat{\boldsymbol{C}}_1$ is $d \times r$, and let $\tilde{\boldsymbol{B}}_{0,alt} = (\tilde{\boldsymbol{A}}_{alt}'\hat{\boldsymbol{\Sigma}}_a^{-1}\tilde{\boldsymbol{A}}_{alt})^{-1} \cdot (\tilde{\boldsymbol{A}}_{alt}'\hat{\boldsymbol{\Sigma}}_a^{-1}\hat{\boldsymbol{C}}_2)$. It follows that $\tilde{\boldsymbol{B}}_{0,alt} - \boldsymbol{B}_0 = \boldsymbol{O}_p(n^{-1})$ and this estimator can be shown asymptotically efficient. Let $\tilde{\boldsymbol{B}}_{alt} = (\boldsymbol{I}_r, \tilde{\boldsymbol{B}}_{0,alt})$. A possible two-step estimator, which delivers asymptotic efficiency, of $\boldsymbol{A}$, $\boldsymbol{\Lambda}_j^*$, $j = 1, \ldots, p-1$, $\boldsymbol{V}_i$, $i = 0, \ldots, s$ based on this initial $\tilde{\boldsymbol{B}}_{0,alt}$, is obtained as the usual least squares regression of $\boldsymbol{W}_t$ on $\tilde{\boldsymbol{B}}_{alt}\boldsymbol{Y}_{t-1}$, $\boldsymbol{W}_{t-1}, \ldots, \boldsymbol{W}_{t-p+1}$, $\boldsymbol{X}_t, \ldots, \boldsymbol{X}_{t-s}$. Using a cointegrated VAR(1) model, Ahn and Reinsel (1990) compared the two-step estimator with the Gaussian estimator and they found, from their empirical study, that these methods perform rather comparably. The two-step procedure represents a particularly convenient method, since efficient estimators are obtained using a computationally attractive procedure. Obviously, this is highly desirable in practical applications.

Let $\tilde{C} = \tilde{A}\tilde{B}$ be an estimator of $C = AB$, such that $\tilde{A}$ and $\tilde{B}$ satisfy Assumption 3.2. Note that using the straightforward identity $\tilde{A}\tilde{B} - AB = (\tilde{A} - A)(\tilde{B} - B) + A(\tilde{B} - B) + (\tilde{A} - A)B$, and since $BP_1 = 0$ (Ahn and Reinsel (1990), p. 817), it follows from Assumption 3.2 that $(\tilde{C} - C)P_1 = O_p(n^{-1})$ and $(\tilde{C} - C)P_2 = O_p(n^{-1/2})$.

Once a particular VARX model has been adjusted, a sound practice consists to test the null hypothesis of non-correlation in the error term $\{a_t, t \in \mathbb{Z}\}$. To this end, we first consistently estimate $f_0(\omega)$:

$$\tilde{f}_0(\omega) \equiv \frac{C_{\tilde{a}}(0)}{2\pi}, \qquad \omega \in [-\pi, \pi].$$

Secondly, we construct a non-parametric spectral density estimator of the residual time series, $\tilde{a}_t$, $t = 1, \dots, n$, based on the reduced rank estimator:

$$\tilde{f}_n(\omega) = \frac{1}{2\pi} \sum_{j=-n+1}^{n-1} k(j/p_n) C_{\tilde{a}}(j) e^{-i\omega j}. \tag{3.5}$$

Note that $\tilde{a}_t$ and $C_{\tilde{a}}(j)$ are defined similarly as $\hat{a}_t$ and $C_{\hat{a}}(j)$, where we recall that the $\hat{a}_t$'s denote the LS residuals and the $\tilde{a}_t$'s correspond to the reduced rank residuals. Consider the normalized quadratic distance

$$Q^2(f_1, f_2) = 2\pi \int_{-\pi}^{\pi} \operatorname{tr}[\Gamma_a^{-1}(0)\{f_1(\omega) - f_2(\omega)\}^\star \Gamma_a^{-1}(0)\{f_1(\omega) - f_2(\omega)\}] d\omega,$$

where for a matrix A, $A^\star$ denotes the transposed conjugate of A, that is $A^\star = \bar{A}'$. The proposed test statistic is essentially a standardized version of $Q^2(\tilde{f}_n, \tilde{f}_0)$, defined by:

$$T_n = \frac{n \sum_{j=1}^{n-1} k^2(j/p_n) \operatorname{tr}\{C'_{\tilde{a}}(j) C_{\tilde{a}}^{-1}(0) C_{\tilde{a}}(j) C_{\tilde{a}}^{-1}(0)\} - d^2 M_n(k)}{\{2d^2 V_n(k)\}^{1/2}} \tag{3.6}$$

where $M_n(k)$ and $V_n(k)$ are given by:

$$M_n(k) = \sum_{j=1}^{n-1} (1 - j/n) k^2(j/p_n), \tag{3.7}$$

$$V_n(k) = \sum_{j=1}^{n-2} (1 - j/n)(1 - (j+1)/n) k^4(j/p_n). \tag{3.8}$$

If $p_n \to \infty$ and $p_n/n \to 0$, we can show that $p_n^{-1} M_n(k) \to M(k) = \int_0^\infty k^2(z) dz$ and $p_n^{-1} V_n(k) \to V(k) = \int_0^\infty k^4(z) dz$. Under some additional assumptions on k and/or p_n (Robinson (1994), p. 73), $p_n^{-1} M_n(k) = M(k) + o(p_n^{-1/2})$. Consequently, asymptotically equivalent test statistics are obtained from (3.6), where $M_n(k)$ ($V_n(k)$) is substituted for $p_n M(k)$ ($p_n V(k)$); see Duchesne and Roy (2004).

For a stationary VARX model, Duchesne and Roy (2004) have proposed the test T_n for testing serial correlation of arbitrary form. In a multivariate regression model, they showed that if the dependence structure of $\{a_t\}$ is such that $\sum_j \|\Gamma_a(j)\|^2 < \infty$ and $\sum_i \sum_j \sum_l |\kappa_{pqrs}(t, t+i, t+j, t+l)| < \infty$, then $T_n = O_p(n/p_n^{1/2})$, meaning that the test procedure T_n is consistent, since $p_n \to \infty$ and p_n/n is assumed to converge to zero.

Computing the asymptotic relative efficiency in the Bahadur sense (ARE_B) of one kernel with respect to another, they found that many currently used kernels in spectral density estimation lead to an ARE_B greater than one with respect to the truncated uniform kernel.

However, there is a growing consensus among statisticians and econometricians that many vector time series may contain unit roots. Recent studies (see, e.g., Engle and Granger (1987), Ahn and Reinsel (1990), Reinsel and Ahn (1992), Yap and Reinsel (1995)) find the presence of unit roots in economic and financial time series. In light of this, we justify the asymptotic distribution of the test statistic given by (3.6) for cointegrated time series. The asymptotic analysis for the generalization from stationary VARX to partially non stationary is nontrivial, because one can no longer exploit the stationarity of $\{\boldsymbol{Y}_t\}$ under the null hypothesis. The asymptotic analysis relies strongly on the error correction form of the model, where we decompose $\boldsymbol{Y}_t$ as the sum of a nonstationary process with stationary increments and of a stationary process. Different arguments are needed to manipulate adequately the nonstationary component; see the proofs of Theorems A.1 and A.2 in the Appendix. Furthermore, the different convergence rates of the estimators of the parameters in Assumption 3.2 need to be properly taken into account in the proof of the result.

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

THEOREM 3.1. *Suppose that $\boldsymbol{Y}$ is a cointegrated VARX(p,s) process as defined by* (2.1) *and that the fourth order moments of $\{\boldsymbol{a}_t\}$ exist. Under Assumptions* 2.1 *and* 3.2, *$p_n \rightarrow \infty$ and $p_n/n \rightarrow 0$, the statistic T_n defined by* (3.6) *admits an asymptotic normal distribution, that is $T_n \rightarrow_L N(0,1)$.*

When $\{\boldsymbol{a}_t\}$ is a Gaussian process, the test statistic T_n can be used to test for the hypothesis H_0 of independent errors. In general, (3.6) can be used to check for the hypothesis of no serial correlation in the error term. As in the stationary case, we do not assume that the innovations are Gaussian and we do not need to assume that the fourth order cumulants vanish. The detailed proof is technical and is presented in the Appendix. Note that since $\{\boldsymbol{a}_t\}$ is stationary, it follows according Theorem 1 of Duchesne and Roy (2004) that

$$(3.9)\quad \tilde{T}_n = \frac{n\sum_{j=1}^{n-1} k^2(j/p_n)\,\mathrm{tr}\{\boldsymbol{C}_{\boldsymbol{a}}^{-1}(0)\boldsymbol{C}_{\boldsymbol{a}}(j)\boldsymbol{C}_{\boldsymbol{a}}^{-1}(0)\boldsymbol{C}'_{\boldsymbol{a}}(j)\} - d^2M_n(k)}{\{2d^2V_n(k)\}^{1/2}} \rightarrow_L N(0,1).$$

The cointegrated VARX model does not intervene in this part since $\tilde{T}_n$ is completely defined by the innovation series $\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n$. The proof of our main theorem is completed if we establish that

$$(3.10)\quad \sum_{j=1}^{n-1} k^2(j/p_n)[\mathrm{tr}\{\boldsymbol{C}_{\boldsymbol{a}}^{-1}(0)\boldsymbol{C}_{\boldsymbol{a}}(j)\boldsymbol{C}_{\boldsymbol{a}}^{-1}(0)\boldsymbol{C}'_{\boldsymbol{a}}(j)\} - \mathrm{tr}\{\boldsymbol{C}_{\tilde{\boldsymbol{a}}}^{-1}(0)\boldsymbol{C}_{\tilde{\boldsymbol{a}}}(j)\boldsymbol{C}_{\tilde{\boldsymbol{a}}}^{-1}(0)\boldsymbol{C}'_{\tilde{\boldsymbol{a}}}(j)\}] = o_p(\sqrt{p_n}/n).$$

From (3.10), it is easily seen that $\tilde{T}_n - T_n$ is $o_p(1)$ and Theorem 3.1 follows. The proof of Theorem 3.1 assumes a non-stochastic p_n. Let the dependence of the test statistic

(3.6) on p_n be more explicit: $T_n \equiv T_n(p_n)$. Conditions under which the asymptotic distribution of the test statistic $T_n(\hat{p}_n)$ is the same as the one of $T_n(p_n)$, where $\hat{p}_n$ is a data-driven bandwidth, are given in the next section.

4. Choice of the smoothing parameter in the new test statistic

The kernel-based spectral density estimator given by (3.5) and the test statistic (3.6) rely on a truncation point or a smoothing parameter, depending of the nature of the kernel $k(\cdot)$. The choice of p_n may affect the power of the test statistic T_n. However, this is a difficult issue to choose an optimal p_n to maximize the power. When $k(\cdot)$ corresponds to the truncated uniform kernel, the weighting is uniform. In this situation p_n possesses an easy interpretation: it corresponds to a lag order. However, when the kernel is of unbounded support, p_n is a smoothing parameter, and it may be more difficult to interpret. Therefore, it seems highly desirable to choose p_n via appropriate data-dependent methods. These methods will reveal some important information on the shape of the true spectral density. Theorem 4.1 gives conditions on the data-driven $\hat{p}_n$, under which $T_n(\hat{p}_n) - T_n(p_n) = o_p(1)$ under the null hypothesis. In particular, it follows by the Slutsky's Theorem that $T_n(\hat{p}_n) \to_L N(0,1)$.

ASSUMPTION 4.1. The kernel $k : \mathbb{R} \to [-1,1]$ satisfies a Lipschitz condition: $|k(z_1) - k(z_2)| \leq \Delta_1 |z_1 - z_2|$, $z_1, z_2 \in \mathbb{R}$, $\Delta_1 \in (0,\infty)$ and $|k(z)| \leq \Delta_2 |z|^{-b}$, $\forall z \in \mathbb{R}$ and for some $b > 1/2$.

Assumption 4.1 allows for most commonly used kernels. It rules out, however, the truncated uniform kernel. We now state Theorem 4.1, which is an extension of Theorem 3.1, allowing for a data-dependent bandwidth $\hat{p}_n$.

THEOREM 4.1. *Suppose that $\boldsymbol{Y}$ is a cointegrated VARX(p,s) process as defined by* (2.1) *and that the fourth order moments of $\{\boldsymbol{a}_t\}$ exist. Suppose that the data-dependent $\hat{p}_n$ satisfies the relation*

$$\frac{\hat{p}_n}{p_n} - 1 = o_p(p_n^{-3/2a+1}), \tag{4.1}$$

where $a > (2b - 1/2)/(2b-1)$. Suppose that p_n is such that $p_n \to \infty$ and $p_n^a/n \to 0$. Under Assumptions 2.1, 3.2 *and* 4.1, *the statistic $T_n(p_n)$ defined by* (3.6) *satisfies*

$$T_n(\hat{p}_n) - T_n(p_n) = o_p(1).$$

In Theorem 4.1, the range of possible $\hat{p}_n$ is function of a, and a depends on b and the kernel $k(\cdot)$. A small a is associated with a large range of possible $\hat{p}_n$. When the kernel admits a compact support, such as the Bartlett and Parzen kernels, $k(z) = 0$ if $z > c$, for a certain c. Consequently, any $a > 1$ is allowed. With the Daniell kernel, we have that $b = 1$ and any $a > 3/2$ is permitted. In all these situations, to satisfy the relation (4.1) seems rather easy. Note that in a different context, Hong and Shehadeh ((1999), Section 2) proved a result similar to Theorem 4.1, when testing for conditional heteroskedasticity in univariate stationary time series. Our proof represents an adaptation of their result in a multivariate framework, when testing for serial correlation in vector cointegrated time series.

The cross-validation procedure of Robinson (1991) for determining the bandwidth of a kernel spectrum estimator of a time series is an appropriate data-driven procedure. In a multiple regression model, Robinson (1991) justified under certain conditions the consistency of a particular cross-validation method of automatically determining a desirable degree of smoothing. Robinson (1991) showed that, asymptotically, such chosen cross-validated $\hat{p}_n$ minimizes a weighted integrated mean squared error of the spectral density estimator with suitable weights depending on the true spectral density. Robinson's procedure represents a generalization of the method of Beltrao and Bloomfield (1987), which is valid for Gaussian time series. Besides establishing the consistency of the procedure for non-Gaussian time series, Robinson (1991) also discusses various multivariate generalizations and he presents practical implementations. A possible procedure is to retain for p_n the value of M that minimizes the pseudo-log-likelihood defined by

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

where $\boldsymbol{I}(\cdot)$ represents the periodogram, $\hat{\boldsymbol{f}}^{M}_{(j)}(\cdot)$ is a leave-two-out type smooth periodogram and $\lambda_j = 2\pi j/n$, $j = 1,\ldots,n$ denote the Fourier frequencies. In practice, the optimization can be performed using a grid search, for example for the values $M = 2,3,\ldots,20$. Note that M is real-valued. However, the impact of integer-clipping of M on spectral density estimators is likely to be negligible.

The cross-validation procedure delivers a $\hat{p}_n$ which is asymptotically optimal for the estimation of the spectral density, when the criterion is a certain integrated weighted mean squared error. Consequently, in an hypothesis testing framework, this choice of p_n does not give necessarily the T_n with the best power. This is a theoretically interesting issue, but the asymptotic analysis involved seems complicated. Nevertheless, the simulations of the next section suggest that the cross-validated $\hat{p}_n$ delivers very reasonable power in many situations.

5. Simulation results

In the previous sections, we have studied kernel-based test statistics, for diagnosing cointegrated VARX models. However, from the point of view of the applied statistician, it is natural to inquire for their finite sample properties. In particular, it is relevant to study the level and power of the kernel-based tests for reasonable time series length, to investigate if the properties in the partially nonstationary case are similar to what have been observed for stationary series. For a given bivariate data generating process (DGP) described below, we examined frequencies of rejection of the null hypothesis, a) when it is in fact true, b) for a fixed alternative, using the kernel-based tests with nominal levels 1, 5 and 10%. We considered three series length, $n = 50$, 100 and 200. Five kernels have been adopted: truncated uniform (TR), Bartlett (BAR), Daniell (DAN), Parzen (PAR) and Bartlett-Priestley (BP). The precise definitions of the kernels are given in Table 1 of Duchesne and Roy (2004).

The following DGP was used:

$$\boldsymbol{Y}_t = \boldsymbol{\Lambda}_1\boldsymbol{Y}_{t-1} + \boldsymbol{V}_0 X_t + \boldsymbol{a}_t, \tag{5.1}$$

where $\boldsymbol{Y}_t = (Y_t(1), Y_t(2))'$ and $\{X_t\}$ is a scalar exogenous variable satisfying $X_t =$

$0.8X_{t-1} + b_t$, $\{b_t\}$ iid $N(0,4)$ (giving $(d,m) = (2,1)$), and

$$\boldsymbol{\Lambda}_1 = \begin{pmatrix} 0.60 & 1.00 \\ 0.12 & 0.70 \end{pmatrix}, \qquad \boldsymbol{V}_0 = \begin{pmatrix} 2.0 \\ 4.0 \end{pmatrix}.$$

The process $\{\boldsymbol{Y}_t\}$ can be written in the following error correction form:

$$\boldsymbol{W}_t = \boldsymbol{C}\boldsymbol{Y}_{t-1} + \boldsymbol{V}_0 X_t + \boldsymbol{a}_t,$$

with $\boldsymbol{C} = (-0.40, 0.12)'(1.00, -2.50) = \boldsymbol{AB}$. Two cases were considered for the error term $\{\boldsymbol{a}_t\}$:

$$\boldsymbol{a}_t = \boldsymbol{e}_t, \tag{5.2}$$

$$\boldsymbol{a}_t = \boldsymbol{e}_t - \boldsymbol{\Xi}\boldsymbol{e}_{t-1}, \tag{5.3}$$

where $\{\boldsymbol{e}_t\}$ is iid $N_2(\mathbf{0}, \boldsymbol{\Sigma}_e)$, independent of $\{b_t\}$, with

$$\boldsymbol{\Xi} = \begin{pmatrix} 0.24 & 0.08 \\ 0.00 & 0.04 \end{pmatrix}, \qquad \boldsymbol{\Sigma}_e = \begin{pmatrix} 25.0 & 5.4 \\ 5.4 & 9.0 \end{pmatrix}.$$

The form of this DGP is inspired of the structure of the model studied in the simulation example of Ahn and Reinsel ((1990), Section 8). The first case (5.2) allowed us to study the level whilst the second one (5.3) was chosen in order to study the power.

In the level study, 10000 independent realizations were generated from DGP (5.1), for each value of n, and the computations were made in the following way.

(1) The Gaussian white noise $\{\boldsymbol{a}_t\}$, which satisfies (5.2), and $\{b_t\}$ were generated independently using the subroutines G05EZF and G05FDF from the NAG library.

(2) Using the initial values $\boldsymbol{Y}_0 = \mathbf{0}$, $X_0 = 0$, $N = 2n+1$ values X_t and $\boldsymbol{Y}_t$, $t = 1, \ldots, N$ were obtained. The first $n+1$ values were discarded, giving a time series of length n, in order to minimize the effect of the initial values.

(3) For each realization, the true DGP was estimated by the two-step approach, as described in Section 3. In a first step, least squares are calculated and in a second step, reduced rank estimators are determined. The residuals $\tilde{\boldsymbol{a}}_t$, $t = 1, \ldots, n$ were obtained.

(4) With each residual time series, the test statistic T_n was computed for the five different kernels TR, BAR, DAN, PAR and BP. For each kernel, the three rates $p_n = [\log(n)]$, $p_n = [3.5n^{0.2}]$ and $p_n = [3n^{0.3}]$ have been used (see Duchesne and Roy (2004) for more details on the different rates). The data-driven procedure of Robinson (1991) was also employed. Note that Robinson's procedure necessitates positive definite kernels. It rules out the truncated uniform kernel.

(5) Finally, for each series of length n, for each kernel k, for each value of p_n and for each nominal level, we obtained from the 10000 realizations the empirical frequencies of rejection of the null hypothesis of non-correlation. The results in percent are reported in Table 1. The standard errors of the empirical levels is 0.099% for the nominal 1%, 0.218% for 5% and 0.300% for 10%.

The power analysis was conducted in a similar way, except that $\{\boldsymbol{a}_t\}$ satisfies (5.3) and that the number of realizations was set to 1000.

We now discuss the results from the level study. They are presented in Table 1. In general, the normal approximation improves with the time series length, as expected.

Table 1. Empirical levels (in percentage) of the test statistic T_n defined by (3.6) for different kernels, different truncation values, when the data are generated from (5.1) with an error term satisfying (5.2).

	$\alpha = 0.01$					$\alpha = 0.05$					$\alpha = 0.10$				
p_n	BP	BAR	DAN	PAR	TR	BP	BAR	DAN	PAR	TR	BP	BAR	DAN	PAR	TR
$n = 50$															
4	2.52	2.52	2.59	2.63	2.42	6.66	6.48	6.71	6.71	6.72	10.33	10.35	10.47	10.37	10.95
8	2.57	2.64	2.65	2.59	2.83	6.89	6.85	6.85	6.87	7.00	10.82	10.61	10.90	10.77	11.19
10	2.63	2.66	2.69	2.71	3.01	6.98	6.88	7.07	6.95	7.22	10.89	10.73	11.09	11.15	11.43
CV	3.08	3.42	3.04	2.95	NA	7.55	8.20	7.60	7.35	NA	11.60	12.52	11.85	11.54	NA
$n = 100$															
5	2.22	2.21	2.15	2.14	2.10	5.97	5.65	5.85	5.87	5.92	9.68	9.63	9.72	9.68	10.24
9	2.14	2.16	2.18	2.16	2.12	5.87	5.84	5.91	5.95	6.50	10.22	10.05	10.18	10.11	10.61
12	2.20	2.13	2.19	2.21	2.28	6.02	6.03	6.13	6.16	6.58	10.44	10.19	10.46	10.69	11.38
CV	2.75	3.01	2.65	2.88	NA	6.95	7.37	6.93	7.20	NA	11.11	11.68	11.26	11.75	NA
$n = 200$															
6	2.15	2.22	2.19	2.13	1.77	6.23	6.22	6.30	6.19	5.70	10.00	10.00	10.07	10.16	10.10
10	2.02	2.02	2.04	1.91	1.86	6.17	6.25	6.26	5.98	6.11	10.00	10.01	10.06	10.08	10.40
15	1.86	1.88	1.86	1.78	1.97	5.87	6.06	5.94	5.85	6.41	10.14	10.02	10.13	10.24	10.67
CV	2.88	3.29	2.87	3.07	NA	7.43	8.02	7.38	8.35	NA	11.51	12.41	11.51	12.76	NA

Some overrejection is observed at the 1% level, but in general, the approximation appeared satisfactory, particularly at the 10% level. Except for the truncated uniform kernel, the choice of the kernel had little impact in our experiments, since the various kernels and p_n values gave comparable results. The truncated uniform kernel seemed slightly inferior, specially when p_n was larger.

At the 5% level, all the test statistics based on non-stochastic p_n lead to rejection rate around 6.5–7% when $n = 50$, between 5.7%–6.5% when $n = 100$ and $n = 200$. The cross-validation lead to rejection rates slightly higher than those obtained with non-stochastic p_n. This additional noise in the rejection rates represents the price to pay to estimate p_n with the available data. However, this price does not seem too high, at least in our experiments. As reported in the power analysis, the power gain that emanates from the cross-validated $\hat{p}_n$ compensates largely. At the 10% level, the rejection rates of the test statistics based on non-stochastic p_n are much closer to the nominal level for all sample sizes, particularly for $n = 200$. For this nominal level, the observed rejection rates are within two standard errors in most cases, or very close of this. The cross-validation exhibited slight overrejection, but for all sample sizes the empirical levels were reasonable.

We now turn to the analysis of the power results, which are presented in Table 2. We computed the power using the asymptotic critical values and using the empirical critical values (given in parentheses), obtained from the 10000 realizations of the level study.

It is interesting to note that the power results indicate that, as in the stationary case, the kernel DAN, PAR and BP behave similarly, BAR seemed slightly superior, and the truncated uniform kernel was inferior. As an illustration, for $n = 200$, $\alpha = 5\%$, using

Table 2. Empirical powers (in percentage) based on the asymptotic and empirical (in parentheses) critical values of the test T_n defined by (3.6) for different kernels, different truncation values when the data are generated from (5.1) with an error term satisfying (5.3).

	$\alpha = 0.01$					$\alpha = 0.05$					$\alpha = 0.10$				
p_n	BP	BAR	DAN	PAR	TR	BP	BAR	DAN	PAR	TR	BP	BAR	DAN	PAR	TR
$n = 50$															
4	18.2	18.6	18.4	17.2	11.1	29.1	29.8	30.2	28.3	20.6	37.5	37.8	37.5	37.1	27.0
	(10.2)	(10.6)	(10.1)	(10.2)	(5.3)	(25.8)	(26.1)	(26.1)	(24.9)	(17.5)	(36.5)	(37.5)	(36.8)	(36.1)	(25.5)
8	14.2	15.5	14.5	13.3	9.0	23.9	26.6	23.7	23.7	18.3	32.6	34.4	32.4	31.7	26.0
	(8.2)	(9.4)	(8.3)	(7.7)	(3.7)	(20.0)	(22.0)	(19.9)	(20.3)	(14.8)	(30.9)	(33.2)	(30.6)	(29.8)	(24.6)
10	12.9	14.5	12.9	12.4	8.1	22.9	24.6	23.3	23.4	17.5	29.7	33.0	29.9	29.8	25.1
	(6.8)	(8.9)	(7.3)	(6.9)	(3.0)	(20.0)	(20.9)	(20.1)	(18.8)	(13.0)	(28.8)	(31.3)	(29.0)	(28.5)	(22.6)
CV	21.4	22.0	20.9	17.6	NA	33.1	34.5	33.1	30.0	NA	43.1	44.5	42.0	40.7	NA
	(11.0)	(11.1)	(11.7)	(9.8)	NA	(27.9)	(26.7)	(26.4)	(25.9)	NA	(38.7)	(39.0)	(37.6)	(37.4)	NA
$n = 100$															
5	31.9	33.7	32.1	30.7	14.3	47.7	49.6	48.3	45.2	29.1	56.8	57.7	56.6	55.0	38.3
	(23.0)	(24.6)	(23.1)	(22.4)	(8.8)	(44.5)	(46.7)	(45.6)	(42.7)	(25.6)	(57.4)	(58.3)	(57.3)	(55.3)	(37.4)
9	24.4	27.5	23.7	22.6	12.8	37.2	41.5	37.4	35.7	23.0	46.9	50.7	47.3	45.5	31.6
	(14.1)	(18.5)	(14.4)	(13.4)	(7.1)	(34.7)	(38.7)	(34.8)	(33.5)	(20.4)	(46.6)	(50.7)	(47.1)	(45.4)	(30.3)
12	19.6	24.2	19.5	18.7	11.2	33.4	37.1	33.1	32.2	21.4	42.4	47.7	42.6	41.2	30.3
	(11.6)	(14.8)	(11.5)	(11.0)	(6.8)	(30.1)	(35.1)	(30.5)	(30.0)	(17.7)	(41.3)	(47.1)	(41.0)	(39.9)	(28.1)
CV	41.5	42.6	41.1	35.2	NA	56.4	57.2	55.1	51.6	NA	65.3	66.7	65.2	61.9	NA
	(29.0)	(28.5)	(28.9)	(22.0)	NA	(51.5)	(51.8)	(50.1)	(44.8)	NA	(63.2)	(63.3)	(61.3)	(58.6)	NA
$n = 200$															
6	66.9	70.8	67.6	64.2	35.5	80.2	84.1	80.6	78.4	53.5	87.7	90.0	87.6	86.5	64.8
	(58.0)	(61.8)	(58.2)	(55.5)	(27.8)	(77.9)	(81.0)	(78.0)	(76.4)	(51.1)	(87.7)	(90.0)	(87.6)	(86.3)	(64.3)
10	53.1	60.1	53.8	50.0	25.9	69.7	76.3	69.5	67.7	43.3	78.6	83.4	79.0	76.5	53.7
	(44.6)	(51.3)	(44.4)	(41.4)	(17.8)	(66.0)	(72.6)	(66.6)	(63.9)	(40.4)	(78.6)	(83.4)	(78.9)	(76.5)	(52.8)
15	41.6	51.4	42.0	38.8	20.1	60.6	68.4	61.1	58.7	37.3	70.6	76.3	71.3	68.3	49.5
	(33.5)	(43.0)	(33.2)	(31.3)	(13.7)	(57.8)	(65.2)	(58.2)	(55.8)	(31.5)	(70.6)	(76.2)	(71.0)	(68.0)	(47.9)
CV	81.1	83.0	79.6	74.3	NA	89.8	91.0	88.5	88.1	NA	92.9	93.3	92.3	91.9	NA
	(70.6)	(69.1)	(68.8)	(62.0)	NA	(87.5)	(87.8)	(86.0)	(80.3)	NA	(91.7)	(92.2)	(91.1)	(90.1)	NA

the empirical critical values, $T_n(p_n; k)$, $k \neq k_{TR}$, was at least 50% more powerful than $T_n(p_n; k_{TR})$, for all the p_n's under investigation.

In general, results based on the empirical and asymptotic quantiles are seen reasonably comparable, except maybe at the nominal level 1%. The rejection rates were similar at the 5% level, and even closer at the 10% level.

The data-driven procedure for p_n worked very well. Without any information on the true alternative, the cross-validation procedure revealed valuable information on the true alternative.

Note that in unreported results, we included in our experiments Hosking's (1980) test. However, as in the stationary case, it seems that an adjustment is needed for this test statistic, to take into account the presence of exogenous variables. See Duchesne and Roy (2004), who observed serious overrejection of the null hypothesis in their level study, when the critical values are taken from the chi-square distribution. However, based on the empirical critical values, the power results of Hosking's (1980) test based on $P \equiv p_n$ lags were identical to those of $T_n(p_n; k_{TR})$, as expected.

Overall, from this very limited empirical study, it appeared that the tests statistics behave as satisfactory in the partially non-stationary situation than in the stationary case, since with time series length as low as $n = 50$, reasonable levels have been observed. The flexible weighting, offering the possibility to give more weight to low order of lags and less weight to high order of lags, displayed powerful procedures, since kernels different from the truncated uniform had interesting power properties, at least in our experiments. Data-driven p_n revealed some valuable information on the true alternative and were powerful for hypothesis testing. Test statistics based on such adaptive procedure are fully operational in practice, which is another serious advantage in real applications.

6. Conclusion

In this paper, we generalized the omnibus procedure proposed by Duchesne and Roy (2004) for multivariate stationary autoregressive models with exogenous variables (VARX) to the case of cointegrated (or partially nonstationary) VARX models. We showed that in the case of a cointegrated VARX time series, the test statistic obtained by comparing the spectral density of the errors under the null hypothesis of non-correlation with a kernel-based spectral density estimator, is still asymptotically standard normal under the null hypothesis. The parameters of the model can be estimated by the conditional maximum likelihood method, or by asymptotically equivalent estimation procedures. We discussed how to adapt the estimation methods of Ahn and Reinsel (1990) to VARX models. Since the proposed methodology relies on a kernel function, a truncation point or a smoothing parameter has to be determined. We stated conditions under which the asymptotic distribution of the test statistics is unaffected by a data-dependent method. The finite sample properties of the test statistics were studied in a small simulation study for non-stochastic smoothing parameter and when it is chosen via the cross-validation method. Our main conclusions are that the new test statistics perform well in partially nonstationary VARX models, since they displayed reasonable levels. Under the considered alternative, using a flexible non-uniform weighting gave better power properties than uniform weighting. The cross-validation procedure offered some overrejetion in the level study but exhibited high power. Overall, for diagnosing cointegrated VARX models, the kernel-based test statistics based on Bartlett or Daniell kernels with p_n chosen by cross-validation should be appropriate in practice.

Acknowledgements

The author would like to thank an anonymous referee for a careful reading of the manuscript and for many constructive suggestions. Discussions with Yongmiao Hong are gratefully acknowledged. This work was supported by grants from FQRNT (Québec) and NSERC (Canada).

Appendix

PROOF OF THEOREM 3.1. The following notations are adopted. The scalar product of $\boldsymbol{x}_t, \boldsymbol{x}_s \in \mathbb{R}^n$ is denoted by $\langle \boldsymbol{x}_t, \boldsymbol{x}_s \rangle = \boldsymbol{x}_t' \boldsymbol{x}_s$ and the norm of $\boldsymbol{x}_t$ by $\|\boldsymbol{x}_t\| = \sqrt{\langle \boldsymbol{x}_t, \boldsymbol{x}_t \rangle}$. The matrix norm of a matrix $\boldsymbol{A} = (a_{ij})_{n\times m}$ is defined by $\|\boldsymbol{A}\|_E^2 = \mathrm{tr}(\boldsymbol{A}\boldsymbol{A}') = \sum_{i=1}^n \sum_{j=1}^m a_{ij}^2$. Let $k_{nj} = k(j/p_n)$, $\boldsymbol{b}_t = \boldsymbol{\Sigma}_a^{-1/2} \boldsymbol{a}_t$. The process $\boldsymbol{b} = \{\boldsymbol{b}_t, t \in \mathbb{Z}\}$ has mean $\boldsymbol{0}$ and variance $\boldsymbol{I}_d$. Let $\boldsymbol{\delta}_{nt} = \tilde{\boldsymbol{a}}_t - \boldsymbol{a}_t$ and $\boldsymbol{\xi}_{nt} = \boldsymbol{\Sigma}_a^{-1/2} \boldsymbol{\delta}_{nt}$, where $\{\tilde{\boldsymbol{a}}_t, t = 1, \ldots, n\}$ denote the residuals of the adjusted reduced rank model. Using the error correction form, we can write

$$\boldsymbol{\delta}_{nt} = \{\tilde{\boldsymbol{\Lambda}}^*(B) - \boldsymbol{\Lambda}^*(B)\} \boldsymbol{W}_t - (\tilde{\boldsymbol{C}} - \boldsymbol{C}) \boldsymbol{P}_2 \boldsymbol{Z}_{2,t-1} - (\tilde{\boldsymbol{C}} - \boldsymbol{C}) \boldsymbol{P}_1 \boldsymbol{Z}_{1,t-1}$$

$$-\{\tilde{\boldsymbol{V}}(B)-\boldsymbol{V}(B)\}\boldsymbol{X}_t,$$

where $\boldsymbol{Y}_t = \boldsymbol{P}_1\boldsymbol{Q}_1\boldsymbol{Y}_t + \boldsymbol{P}_2\boldsymbol{Q}_2\boldsymbol{Y}_t = \boldsymbol{P}_1\boldsymbol{Z}_{1t} + \boldsymbol{P}_2\boldsymbol{Z}_{2t}$, that is $\boldsymbol{Q}\boldsymbol{Y}_t = (\boldsymbol{Z}'_{1t}, \boldsymbol{Z}'_{2t})'$. Recall that $\{\boldsymbol{Z}_{1t}\}$ is nonstationary, with stationary increments, while $\{\boldsymbol{Z}_{2t}\}$ denotes a stationary process. In the whole proof, special attention has to be paid to $\{\boldsymbol{Z}_{1t}\}$, since the arguments in the stationary case do not extend to the nonstationary situation.

The following lemma will be useful:

LEMMA A.1. $n^{-1}\sum_{t=1}^n \|\boldsymbol{\xi}_{nt}\|^2 = O_p(n^{-1})$.

PROOF OF LEMMA A.1. Note that $\|\boldsymbol{\xi}_{nt}\|^2 = \mathrm{tr}(\boldsymbol{\xi}_{nt}\boldsymbol{\xi}'_{nt}) = \mathrm{tr}(\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1/2}\boldsymbol{\delta}_{nt}\boldsymbol{\delta}'_{nt}\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1/2})$. However, using the inequality $\mathrm{tr}\{(\sum_{i=1}^p \boldsymbol{A}_i)(\sum_{i=1}^p \boldsymbol{A}_i)'\} \le 2^{p-1}\sum_{i=1}^p \mathrm{tr}\{\boldsymbol{A}_i\boldsymbol{A}'_i\}$ and Cauchy-Schwarz inequality, it follows that

$$\begin{aligned}\mathrm{tr}(\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1/2}\boldsymbol{\delta}_{nt}\boldsymbol{\delta}'_{nt}\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1/2}) \le \Delta\Bigg[&\sum_{i=1}^{p-1}\mathrm{tr}\{(\tilde{\boldsymbol{\Lambda}}_i^* - \boldsymbol{\Lambda}_i^*)'\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1}(\tilde{\boldsymbol{\Lambda}}_i^* - \boldsymbol{\Lambda}_i^*)\}\|\boldsymbol{W}_{t-i}\|^2\\ &+\mathrm{tr}\{\boldsymbol{P}'_2(\tilde{\boldsymbol{C}}-\boldsymbol{C})'\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1}(\tilde{\boldsymbol{C}}-\boldsymbol{C})\boldsymbol{P}_2\}\|\boldsymbol{Z}_{2,t-1}\|^2\\ &+\mathrm{tr}\{\boldsymbol{P}'_1(\tilde{\boldsymbol{C}}-\boldsymbol{C})'\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1}(\tilde{\boldsymbol{C}}-\boldsymbol{C})\boldsymbol{P}_1\}\|\boldsymbol{Z}_{1,t-1}\|^2\\ &+\sum_{i=0}^{s}\mathrm{tr}\{(\tilde{\boldsymbol{V}}_i-\boldsymbol{V}_i)'\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1}(\tilde{\boldsymbol{V}}_i-\boldsymbol{V}_i)\}\|\boldsymbol{X}_{t-i}\|^2\Bigg],\end{aligned}$$

where Δ is a generic constant. We have that $n^{-1}\sum_{t=1}^n \|\boldsymbol{W}_{t-i}\|^2 = O_p(1)$, $i = 1,\ldots,p-1$, $n^{-1}\sum_{t=1}^n \|\boldsymbol{Z}_{2,t-1}\|^2 = O_p(1)$, $n^{-1}\sum_{t=1}^n \|\boldsymbol{X}_{t-i}\|^2 = O_p(1)$, $i = 0,\ldots,s$, since $\{\boldsymbol{W}_t\}$, $\{\boldsymbol{Z}_{2t}\}$ and $\{\boldsymbol{X}_t\}$ are stationary processes. Because $\{\boldsymbol{Z}_{1t}\}$ is nonstationary with stationary increments, using Lemma 3.1 of Phillips and Durlauf (1986), it follows that $n^{-2}\sum_{t=1}^n \|\boldsymbol{Z}_{1,t-1}\|^2 = O_p(1)$ and by Assumption 3.2 it follows that $n^{-1}\sum_{t=1}^n \|\boldsymbol{\xi}_{nt}\|^2 = O_p(n^{-1})$, as announced.

Let $\tilde{\boldsymbol{b}}_t = \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1/2}\tilde{\boldsymbol{a}}_t$. As in Duchesne and Roy (2004), it can be shown that to replace in (3.10) expressions $\boldsymbol{C}_{\boldsymbol{a}}(0)$ and $\boldsymbol{C}_{\tilde{\boldsymbol{a}}}(0)$ by $\boldsymbol{\Sigma}_{\boldsymbol{a}}$ has no impact asymptotically. Consequently, we decompose

$$\sum_{j=1}^{n-1} k_{nj}^2[\mathrm{tr}\{\boldsymbol{C}_{\tilde{\boldsymbol{b}}}(j)\boldsymbol{C}'_{\tilde{\boldsymbol{b}}}(j)\} - \mathrm{tr}\{\boldsymbol{C}_{\boldsymbol{b}}(j)\boldsymbol{C}'_{\boldsymbol{b}}(j)\}] = S_{n1} + 2S_{n2},$$

where

$$S_{n1} = \sum_{j=1}^{n-1} k_{nj}^2\,\mathrm{tr}[\{\boldsymbol{C}_{\tilde{\boldsymbol{b}}}(j) - \boldsymbol{C}_{\boldsymbol{b}}(j)\}\{\boldsymbol{C}_{\tilde{\boldsymbol{b}}}(j) - \boldsymbol{C}_{\boldsymbol{b}}(j)\}'],$$

$$S_{n2} = \sum_{j=1}^{n-1} k_{nj}^2\,\mathrm{tr}[\boldsymbol{C}_{\boldsymbol{b}}(j)\{\boldsymbol{C}_{\tilde{\boldsymbol{b}}}(j) - \boldsymbol{C}_{\boldsymbol{b}}(j)\}'].$$

To show Theorem 3.1, it suffices to establish Theorems A.1 and A.2.

THEOREM A.1. $S_{n1} = O_p(n^{-1})$.

THEOREM A.2. $S_{n2} = o_p(p_n^{1/2}/n)$.

PROOF OF THEOREM A.1. We write $\boldsymbol{C}_{\tilde{b}}(j) - \boldsymbol{C}_b(j) = \Delta\boldsymbol{C}_1(j) + \Delta\boldsymbol{C}_2(j) + \Delta\boldsymbol{C}_3(j)$, where $\Delta\boldsymbol{C}_1(j) = n^{-1}\sum_{t=j+1}^{n} \boldsymbol{\xi}_{nt}\boldsymbol{b}'_{t-j}$, $\Delta\boldsymbol{C}_2(j) = n^{-1}\sum_{t=j+1}^{n} \boldsymbol{b}_t\boldsymbol{\xi}'_{n,t-j}$, $\Delta\boldsymbol{C}_3(j) = n^{-1}\sum_{t=j+1}^{n} \boldsymbol{\xi}_{nt}\boldsymbol{\xi}'_{n,t-j}$. It follows that $S_{n1} \leq \Delta(A_{1n} + A_{2n} + A_{3n})$, where

$$A_{1n} = \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}\{\Delta\boldsymbol{C}_1(j)\Delta\boldsymbol{C}'_1(j)\},$$

$$A_{2n} = \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}\{\Delta\boldsymbol{C}_2(j)\Delta\boldsymbol{C}'_2(j)\},$$

$$A_{3n} = \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}\{\Delta\boldsymbol{C}_3(j)\Delta\boldsymbol{C}'_3(j)\}.$$

To study A_{ln}, $l = 1, 2, 3$, we introduce the following cross-covariance measures.

$$\boldsymbol{C}_{\boldsymbol{W}\boldsymbol{b}}(j-i) = n^{-1}\sum_{t=j+1}^{n} \boldsymbol{W}_{t-i}\boldsymbol{b}'_{t-j}, \qquad i = 1, \ldots, p-1,$$

$$\boldsymbol{C}_{\boldsymbol{Z}_2\boldsymbol{b}}(j-1) = n^{-1}\sum_{t=j+1}^{n} \boldsymbol{Z}_{2,t-1}\boldsymbol{b}'_{t-j}, \qquad \boldsymbol{C}_{\boldsymbol{Z}_1\boldsymbol{b}}(j-1) = n^{-1}\sum_{t=j+1}^{n} \boldsymbol{Z}_{1,t-1}\boldsymbol{b}'_{t-j},$$

$$\boldsymbol{C}_{\boldsymbol{X}\boldsymbol{b}}(j-i) = n^{-1}\sum_{t=j+1}^{n} \boldsymbol{X}_{t-i}\boldsymbol{b}'_{t-j}, \qquad i = 0, \ldots, s.$$

The following lemmas will be useful.

LEMMA A.2. $E[\operatorname{tr}\{\boldsymbol{C}_{\boldsymbol{W}\boldsymbol{b}}(j-i)\boldsymbol{C}'_{\boldsymbol{W}\boldsymbol{b}}(j-i)\}] \leq \Delta_1 n^{-1} + \Delta_2\rho^{2(j-i)}$, $i = 1, \ldots, p-1$, *where* $\|\boldsymbol{\Psi}_j\|_E \leq K\rho^j$, *for a constant* $K > 0$ *and* $\rho \in (0,1)$, $\boldsymbol{\Psi}(B) = (\boldsymbol{I}_d - \boldsymbol{P}_2\boldsymbol{Q}_2B)\cdot\{\boldsymbol{\Lambda}^{**}(B)\}^{-1} = \sum_{j=0}^{\infty}\boldsymbol{\Psi}_jB^j$.

LEMMA A.3. $E[\operatorname{tr}\{\boldsymbol{C}_{\boldsymbol{Z}_2\boldsymbol{b}}(j-1)\boldsymbol{C}'_{\boldsymbol{Z}_2\boldsymbol{b}}(j-1)\}] \leq \tilde{\Delta}_1 n^{-1} + \tilde{\Delta}_2\tilde{\rho}^{2(j-1)}$, *where* $\|\tilde{\boldsymbol{\Psi}}_j\|_E \leq \tilde{K}\tilde{\rho}^j$, *for a constant* $\tilde{K} > 0$ *and* $\tilde{\rho} \in (0,1)$, $\tilde{\boldsymbol{\Psi}}(B) = \boldsymbol{Q}_2\{\boldsymbol{\Lambda}^{**}(B)\}^{-1} = \sum_{j=0}^{\infty}\tilde{\boldsymbol{\Psi}}_jB^j$.

LEMMA A.4. $\operatorname{tr}\{\boldsymbol{C}_{\boldsymbol{Z}_1\boldsymbol{b}}(j-1)\boldsymbol{C}'_{\boldsymbol{Z}_1\boldsymbol{b}}(j-1)\} = O_p(1)$, *independently of* j.

LEMMA A.5. $\operatorname{tr}\{\boldsymbol{C}_{\boldsymbol{X}\boldsymbol{b}}(j-i)\boldsymbol{C}'_{\boldsymbol{X}\boldsymbol{b}}(j-i)\} = O_p(n^{-1})$, *independently of* i *and* j.

PROOF OF LEMMA A.2. Since $(\boldsymbol{I}_d - B) = (\boldsymbol{I}_d - \boldsymbol{P}_2\boldsymbol{Q}_2B)(\boldsymbol{I}_d - \boldsymbol{P}_1\boldsymbol{Q}_1B)$ and $\boldsymbol{\Lambda}(1)\boldsymbol{P}_1 = \boldsymbol{0}$, it follows that $\boldsymbol{\Lambda}(B)\boldsymbol{Y}_t = \boldsymbol{V}(B)\boldsymbol{X}_t + \boldsymbol{a}_t$ can be written as

$$\boldsymbol{W}_t = (\boldsymbol{I}_d - \boldsymbol{P}_2\boldsymbol{Q}_2B)\{\boldsymbol{\Lambda}^{**}(B)\}^{-1}\boldsymbol{V}(B)\boldsymbol{X}_t + (\boldsymbol{I}_d - \boldsymbol{P}_2\boldsymbol{Q}_2B)\{\boldsymbol{\Lambda}^{**}(B)\}^{-1}\boldsymbol{a}_t,$$

where $\boldsymbol{\Lambda}(B) = \boldsymbol{\Lambda}^{**}(B)(\boldsymbol{I}_d - \boldsymbol{P}_1\boldsymbol{Q}_1B)$, with $\boldsymbol{\Lambda}^{**}(B) = \boldsymbol{\Lambda}^{*}(B)(\boldsymbol{I}_d - \boldsymbol{P}_2\boldsymbol{Q}_2B) + \boldsymbol{\Lambda}(1)B$. See Pham *et al.* ((2003), p. 556). Consequently, we can write

$$\boldsymbol{W}_t = \boldsymbol{\Pi}(B)\boldsymbol{X}_t + \boldsymbol{\Psi}(B)\boldsymbol{a}_t,$$

where $\boldsymbol{\Pi}(B) = (\boldsymbol{I}_d - \boldsymbol{P}_2\boldsymbol{Q}_2 B)\{\boldsymbol{\Lambda}^{**}(B)\}^{-1}\boldsymbol{V}(B) = \sum_{j=0}^{\infty} \boldsymbol{\Pi}_j B^j$, and $\|\boldsymbol{\Pi}_j\| \leq K_1 \rho_1^j$ for a certain K_1 and $\rho_1 \in (0,1)$. Similarly, $\boldsymbol{\Psi}(B) = (\boldsymbol{I}_d - \boldsymbol{P}_2\boldsymbol{Q}_2 B)\{\boldsymbol{\Lambda}^{**}(B)\}^{-1} = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j B^j$, $\|\boldsymbol{\Psi}_j\| \leq K_2 \rho_2^j$ for a certain K_2 and $\rho_2 \in (0,1)$. For $i = 1$ (other cases are done similarly), we have that

$$n^{-1} \sum_{t=j+1}^{n} \boldsymbol{W}_{t-1}\boldsymbol{b}'_{t-j} = n^{-1} \sum_{t=j+1}^{n} \{\boldsymbol{\Pi}(B)\boldsymbol{X}_{t-1}\}\boldsymbol{b}'_{t-j} + n^{-1} \sum_{t=j+1}^{n} \{\boldsymbol{\Psi}(B)\boldsymbol{a}_{t-1}\}\boldsymbol{b}'_{t-j}.$$

In Lemma A.11 of Duchesne and Roy (2004), a bound is derived for a moment similar to $E[\mathrm{tr}\{\boldsymbol{C}_{\boldsymbol{Wb}}(j-i)\boldsymbol{C}'_{\boldsymbol{Wb}}(j-i)\}]$, but when the process admits an autoregressive component of order one. Here, we need to invert more general operators. To generalize their lemma is cumbersome but straightforward, since $\|\boldsymbol{\Psi}_j\|_E \leq K\rho^j$, $K > 0$ and $\rho \in (0,1)$. Consequently, proceeding as in the Lemma A.11 of Duchesne and Roy (2004), this shows Lemma A.2.

PROOF OF LEMMA A.3. Note that

$$(\boldsymbol{I}_d - \boldsymbol{P}_1\boldsymbol{Q}_1 B)\boldsymbol{Y}_t = \{\boldsymbol{\Lambda}^{**}(B)\}^{-1}\boldsymbol{V}(B)\boldsymbol{X}_t + \{\boldsymbol{\Lambda}^{**}(B)\}^{-1}\boldsymbol{a}_t. \tag{A.1}$$

Since $\boldsymbol{Q}_2\boldsymbol{P}_1 = \boldsymbol{0}$, then

$$\boldsymbol{Z}_{2t} = \boldsymbol{Q}_2\boldsymbol{Y}_t = \boldsymbol{Q}_2\{\boldsymbol{\Lambda}^{**}(B)\}^{-1}\boldsymbol{V}(B)\boldsymbol{X}_t + \boldsymbol{Q}_2\{\boldsymbol{\Lambda}^{**}(B)\}^{-1}\boldsymbol{a}_t = \tilde{\boldsymbol{\Pi}}(B)\boldsymbol{X}_t + \tilde{\boldsymbol{\Psi}}(B)\boldsymbol{a}_t,$$

where $\|\tilde{\boldsymbol{\Pi}}_j\| \leq \tilde{K}_1\tilde{\rho}_1^j$ for a certain $\tilde{K}_1$ and $\tilde{\rho}_1 \in (0,1)$, $\|\tilde{\boldsymbol{\Psi}}_j\| \leq \tilde{K}_2\tilde{\rho}_2^j$ for a certain $\tilde{K}_2$ and $\tilde{\rho}_2 \in (0,1)$. At this point, the proof is then the same as the proof of Lemma A.2.

PROOF OF LEMMA A.4. Multiplying (A.1) by $\boldsymbol{Q}_1$, since $\boldsymbol{Q}_1\boldsymbol{P}_1 = \boldsymbol{I}_{d_0}$, we find that

$$(\boldsymbol{I}_{d_0} - B)\boldsymbol{Z}_{1t} = \boldsymbol{Q}_1\{\boldsymbol{\Lambda}^{**}(B)\}^{-1}\boldsymbol{V}(B)\boldsymbol{X}_t + \boldsymbol{Q}_1\{\boldsymbol{\Lambda}^{**}(B)\}^{-1}\boldsymbol{a}_t,$$

and consequently $\{\boldsymbol{Z}_{1t}\}$ admits stationary increments. It follows that

$$(\boldsymbol{I}_{d_0} - B)\boldsymbol{Z}_{1t} = \tilde{\tilde{\boldsymbol{\Pi}}}(B)\boldsymbol{X}_t + \tilde{\tilde{\boldsymbol{\Psi}}}(B)\boldsymbol{a}_t = \boldsymbol{u}_t,$$

where $\{\boldsymbol{u}_t\}$ is a stationary process. Consequently, we deduce that $\boldsymbol{Z}_{1t} = \sum_{i=1}^{t} \boldsymbol{u}_i$. By a direct calculation:

$$\begin{aligned}
\mathrm{tr}\{\boldsymbol{C}_{\boldsymbol{Z}_1\boldsymbol{b}}(j-1)\boldsymbol{C}'_{\boldsymbol{Z}_1\boldsymbol{b}}(j-1)\} &= n^{-2} \sum_{t=j+1}^{n} \sum_{s=j+1}^{n} \langle \boldsymbol{Z}_{1,t-1}, \boldsymbol{Z}_{1,s-1}\rangle\langle \boldsymbol{b}_{t-j}, \boldsymbol{b}_{s-j}\rangle, \\
&= n^{-2} \sum_{t=j+1}^{n} \sum_{s=j+1}^{n} \sum_{i_1=1}^{t-1} \sum_{i_2=1}^{s-1} \boldsymbol{u}'_{i_1}\boldsymbol{u}_{i_2}\boldsymbol{b}'_{t-j}\boldsymbol{b}_{s-j}.
\end{aligned}$$

It follows that $E[\mathrm{tr}\{\boldsymbol{C}_{\boldsymbol{Z}_1\boldsymbol{b}}(j-1)\boldsymbol{C}'_{\boldsymbol{Z}_1\boldsymbol{b}}(j-1)\}] = B_{1n} + B_{2n}$, where

$$B_{1n} = n^{-2} \sum_{t=j+1}^{n} \sum_{s=j+1}^{n} \sum_{i_1=1}^{t-1} \sum_{i_2=1}^{s-1} E[\{\tilde{\tilde{\boldsymbol{\Pi}}}(B)\boldsymbol{X}_{i_1}\}'\{\tilde{\tilde{\boldsymbol{\Pi}}}(B)\boldsymbol{X}_{i_2}\}]E(\boldsymbol{b}'_{t-j}\boldsymbol{b}_{s-j}),$$

$$B_{2n} = n^{-2} \sum_{t=j+1}^{n} \sum_{s=j+1}^{n} \sum_{i_1=1}^{t-1} \sum_{i_2=1}^{s-1} E[\{\tilde{\tilde{\boldsymbol{\Psi}}}(B)\boldsymbol{a}_{i_1}\}'\{\tilde{\tilde{\boldsymbol{\Psi}}}(B)\boldsymbol{a}_{i_2}\}\boldsymbol{b}'_{t-j}\boldsymbol{b}_{s-j}].$$

It is easy to show that $B_{1n} = O(1)$. Now, we decompose B_{2n} as $B_{2n} = B_{21n} + 2B_{22n}$, where

$$B_{21n} = n^{-2} \sum_{t=j+1}^{n} \sum_{i_1,i_2=1}^{t-1} \sum_{l_1,l_2=0}^{\infty} E(\boldsymbol{a}'_{i_1-l_1} \tilde{\tilde{\boldsymbol{\Psi}}}'_{l_1} \tilde{\tilde{\boldsymbol{\Psi}}}_{l_2} \boldsymbol{a}_{i_2-l_2} \boldsymbol{b}'_{t-j} \boldsymbol{b}_{t-j}),$$

$$B_{22n} = n^{-2} \sum_{t>s} \sum_{i_1=1}^{t-1} \sum_{i_2=1}^{s-1} \sum_{l_1,l_2=0}^{\infty} E(\boldsymbol{a}'_{i_1-l_1} \tilde{\tilde{\boldsymbol{\Psi}}}'_{l_1} \tilde{\tilde{\boldsymbol{\Psi}}}_{l_2} \boldsymbol{a}_{i_2-l_2} \boldsymbol{b}'_{t-j} \boldsymbol{b}_{s-j}).$$

Since $\{\boldsymbol{a}_t\}$ is iid and $\sum_{l=0}^{\infty} \|\tilde{\tilde{\boldsymbol{\Psi}}}_l\| < \infty$, it follows using results on fourth order moments (see Hannan (1970)) that $|B_{21n}| \leq \Delta$ and $|B_{22n}| \leq \Delta$. This shows the Lemma A.4.

PROOF OF LEMMA A.5. The proof is obtained by developing directly $\boldsymbol{C}_{\boldsymbol{X}\boldsymbol{b}}(j-i)\boldsymbol{C}'_{\boldsymbol{X}\boldsymbol{b}}(j-i)$ as in the proof of Lemma A.4 and taking the trace. Using the strict exogeneity between $\{\boldsymbol{X}_t\}$ and $\{\boldsymbol{b}_t\}$, it is easily seen that the mathematical expectation of the resulting expression is $O(n^{-1})$, showing the result.

We now show that $A_{ln} = O_p(n^{-1})$, $l = 1, 2, 3$. We begin with A_{1n}. We can write

$$\begin{aligned} \Delta \boldsymbol{C}_1(j) = {} & \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1/2}(\boldsymbol{C} - \tilde{\boldsymbol{C}})\boldsymbol{P}_1 \boldsymbol{C}_{\boldsymbol{Z}_1\boldsymbol{b}}(j-1) + \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1/2}(\boldsymbol{C} - \tilde{\boldsymbol{C}})\boldsymbol{P}_2 \boldsymbol{C}_{\boldsymbol{Z}_2\boldsymbol{b}}(j-1) \\ & + \sum_{i=1}^{p-1} \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1/2}(\boldsymbol{\Lambda}_i^* - \tilde{\boldsymbol{\Lambda}}_i^*)\boldsymbol{C}_{\boldsymbol{W}\boldsymbol{b}}(j-i) \\ & + \sum_{i=0}^{s} \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1/2}(\boldsymbol{V}_i - \tilde{\boldsymbol{V}}_i)\boldsymbol{C}_{\boldsymbol{X}\boldsymbol{b}}(j-i). \end{aligned} \tag{A.2}$$

Consequently,

$$\begin{aligned} A_{1n} \leq \Delta \Bigg[& \sum_{i=1}^{p-1} \operatorname{tr}\{(\tilde{\boldsymbol{\Lambda}}_i^* - \boldsymbol{\Lambda}_i^*)'\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1}(\tilde{\boldsymbol{\Lambda}}_i^* - \boldsymbol{\Lambda}_i^*)\} D_{in} \\ & + \operatorname{tr}\{\boldsymbol{P}_2'(\tilde{\boldsymbol{C}} - \boldsymbol{C})'\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1}(\tilde{\boldsymbol{C}} - \boldsymbol{C})\boldsymbol{P}_2\} \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}[\boldsymbol{C}_{\boldsymbol{Z}_2\boldsymbol{b}}(j-1)\boldsymbol{C}'_{\boldsymbol{Z}_2\boldsymbol{b}}(j-1)] \\ & + \operatorname{tr}\{\boldsymbol{P}_1'(\tilde{\boldsymbol{C}} - \boldsymbol{C})'\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1}(\tilde{\boldsymbol{C}} - \boldsymbol{C})\boldsymbol{P}_1\} \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}\{\boldsymbol{C}_{\boldsymbol{Z}_1\boldsymbol{b}}(j-1)\boldsymbol{C}'_{\boldsymbol{Z}_1\boldsymbol{b}}(j-1)\} \\ & + \sum_{i=0}^{s} \operatorname{tr}\{(\tilde{\boldsymbol{V}}_i - \boldsymbol{V}_i)'\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1}(\tilde{\boldsymbol{V}}_i - \boldsymbol{V}_i)\} E_{in} \Bigg], \end{aligned}$$

where

$$D_{in} = \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}\{\boldsymbol{C}_{\boldsymbol{W}\boldsymbol{b}}(j-i)\boldsymbol{C}'_{\boldsymbol{W}\boldsymbol{b}}(j-i)\}, \qquad i = 1, \ldots, p-1,$$

$$E_{in} = \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}\{\boldsymbol{C}_{\boldsymbol{X}\boldsymbol{b}}(j-i)\boldsymbol{C}'_{\boldsymbol{X}\boldsymbol{b}}(j-i)\}, \qquad i = 0, \ldots, s.$$

Using Lemmas A.2–A.5 and Assumption 3.2, this shows the announced result for A_{1n}. Note that because of the nonstationarity, the rate of the moment $\operatorname{tr}\{\boldsymbol{C}_{\boldsymbol{Z}_1 \boldsymbol{b}}(j-1)\boldsymbol{C}'_{\boldsymbol{Z}_1 \boldsymbol{b}}(j-1)\}$ in Lemma A.4 appears somewhat larger than the rates of the moments obtained in the other lemmas. However, from Assumption 3.2, we have that $(\tilde{\boldsymbol{C}} - \boldsymbol{C})\boldsymbol{P}_1 = \boldsymbol{O}_p(n^{-1})$, which allows us to show the announced result for A_{1n}. The proof of A_{2n} is similar. The proof of A_{3n} follows by using Lemma A.1.

PROOF OF THEOREM A.2. We write $S_{2n} = A_{4n} + A_{5n} + A_{6n}$, where

$$A_{4n} = \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}\left\{ \boldsymbol{C}_{\boldsymbol{b}}(j) \left(n^{-1} \sum_{t=j+1}^{n} \boldsymbol{\xi}_{nt} \boldsymbol{b}'_{t-j} \right)' \right\},$$

$$A_{5n} = \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}\left\{ \boldsymbol{C}_{\boldsymbol{b}}(j) \left(n^{-1} \sum_{t=j+1}^{n} \boldsymbol{b}_t \boldsymbol{\xi}'_{n,t-j} \right)' \right\},$$

$$A_{6n} = \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}\left\{ \boldsymbol{C}_{\boldsymbol{b}}(j) \left(n^{-1} \sum_{t=j+1}^{n} \boldsymbol{\xi}_{nt} \boldsymbol{\xi}'_{n,t-j} \right)' \right\}.$$

For $A_{4n} = \sum_{j=1}^{n-1} k_{nj}^2 \operatorname{tr}\{\boldsymbol{C}_{\boldsymbol{b}}(j)\Delta \boldsymbol{C}'_1(j)\}$, we can decompose this expression using formula (A.2). Each term is treated separately: using Cauchy-Schwarz inequality for each term, Lemmas A.2–A.5, we obtain that globally $A_{4n} = o_p(p_n^{1/2}/n)$. Similarly, $A_{5n} = o_p(p_n^{1/2}/n)$ and $A_{6n} = o_p(p_n^{1/2}/n)$. Collecting the results, we obtain Theorem A.2.

PROOF OF THEOREM 4.1. Let $M_n(k)$ and $V_n(k)$, defined by (3.7) and (3.8), depend more explicitly on p_n : $M_n(k) \equiv M_n(k;p_n)$, $V_n(k) \equiv V_n(k;p_n)$. According to Lemma A.2 of Hong and Shehadeh (1999), if follows that

$$M_n(k;\hat{p}_n)/p_n = M_n(k;p_n)/p_n + o_p(p_n^{-1/2}),$$
$$V_n(k;\hat{p}_n)/p_n = V_n(k;p_n)/p_n + o_p(1).$$

The proof of Theorem 4.1 is completed provided the following result is proven:

$$\sum_{j=1}^{n-1} \{k^2(j/\hat{p}_n) - k^2(j/p_n)\} \operatorname{tr}\{\boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \boldsymbol{C}_{\tilde{\boldsymbol{a}}}(j) \boldsymbol{\Sigma}_{\boldsymbol{a}}^{-1} \boldsymbol{C}'_{\tilde{\boldsymbol{a}}}(j)\} = o_p(p_n^{1/2}/n). \tag{A.3}$$

We decompose the left hand of (A.3) as:

$$\sum_{j=1}^{n-1} \{k^2(j/\hat{p}_n) - k^2(j/p_n)\} \operatorname{tr}\{\boldsymbol{C}_{\tilde{\boldsymbol{b}}}(j) \boldsymbol{C}'_{\tilde{\boldsymbol{b}}}(j)\} = G_{1n} + G_{2n},$$

where

$$G_{1n} = \sum_{j=1}^{n-1} \{k^2(j/\hat{p}_n) - k^2(j/p_n)\} \operatorname{tr}\{\boldsymbol{C}_{\boldsymbol{b}}(j) \boldsymbol{C}'_{\boldsymbol{b}}(j)\},$$

$$G_{2n} = \sum_{j=1}^{n-1} \{k^2(j/\hat{p}_n) - k^2(j/p_n)\} [\operatorname{tr}\{\boldsymbol{C}_{\tilde{\boldsymbol{b}}}(j) \boldsymbol{C}'_{\tilde{\boldsymbol{b}}}(j)\} - \operatorname{tr}\{\boldsymbol{C}_{\boldsymbol{b}}(j) \boldsymbol{C}'_{\boldsymbol{b}}(j)\}].$$

We decompose further G_{1n} as $G_{1n} = H_{1n} + H_{2n} - H_{3n}$, where

$$H_{1n} = \sum_{j=1}^{l} \{k^2(j/\hat{p}_n) - k^2(j/p_n)\} \operatorname{tr}\{\boldsymbol{C}_b(j)\boldsymbol{C}'_b(j)\},$$

$$H_{2n} = \sum_{j=l+1}^{n-1} k^2(j/\hat{p}_n) \operatorname{tr}\{\boldsymbol{C}_b(j)\boldsymbol{C}'_b(j)\},$$

$$H_{3n} = \sum_{j=l+1}^{n-1} k^2(j/p_n) \operatorname{tr}\{\boldsymbol{C}_b(j)\boldsymbol{C}'_b(j)\},$$

where $l = [p_n^a]$ is a truncation point, $a > (2b - 1/2)/(2b - 1)$ ($[x]$ denotes the integer part of x). Since $|k(z)| \leq \Delta |z|^{-b}$ and $E[\operatorname{tr}\{\boldsymbol{C}_b(j)\boldsymbol{C}'_b(j)\}] = O(n^{-1})$, it can be shown that $H_{2n} = o_p(p_n^{1/2}/n)$. Similarly, $H_{3n} = o_p(p_n^{1/2}/n)$. Writing $k^2(j/\hat{p}_n) - k^2(j/p_n) = (k(j/\hat{p}_n) - k(j/p_n))^2 + 2k(j/p_n)(k(j/\hat{p}_n) - k(j/p_n))$, we can decompose H_{1n} in two additional terms that we study separately, say $H_{1n} = H_{11n} + H_{12n}$. Using the Lipschitz condition on $k(\cdot)$, it can be shown that $H_{11n} = o_p(p_n^{1/2}/n)$. For H_{12n}, the Cauchy-Schwarz inequality and the result for H_{11n} allow us to show that $H_{12n} = o_p(p_n^{1/2}/n)$ and it follows that $G_{1n} = o_p(p_n^{1/2}/n)$.

For G_{2n}, a reasoning similar to the proof of Lemma A.3 of Hong and Shehadeh (1999) allows us to show that

$$\frac{n}{p_n^{1/2}} \sum_{j=1}^{n-1} k^2(j/\hat{p}_n)[\operatorname{tr}\{\boldsymbol{C}_{\tilde{b}}(j)\boldsymbol{C}'_{\tilde{b}}(j)\} - \operatorname{tr}\{\boldsymbol{C}_b(j)\boldsymbol{C}'_b(j)\}] = o_p(1).$$

Using Theorem 3.1, it follows that $G_{2n} = o_p(p_n^{1/2}/n)$. Collecting the results, this yields the announced result.

References

Ahn, S. K. and Reinsel, G. C. (1988). Nested reduced-rank autoregressive models for multiple time series, *Journal of the American Statistical Association*, **83**, 849–856.

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

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

Duchesne, P. and Roy, R. (2004). On consistent testing for serial correlation of unknown form in vector time series models, *Journal of Multivariate Analysis*, **89**, 148–180.

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

Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification, *Journal of Econometrics*, **16**, 121–130.

Granger, C. W. J. and Weiss, A. A. (1983). Time series analysis of error-correction models, *Studies in Econometrics, Time Series, and Multivariate Statistics*, 255–278, Academic Press, New York.

Hannan, E. J. (1970). *Multiple Time Series*, Wiley, New York.

Hannan, E. J. and Deistler, M. (1988). *The Statistical Theory of Linear Systems*, Wiley, New York.

Hong, Y. and Shehadeh, R. D. (1999). A new test for ARCH effects and its finite-sample performance, *Journal of Business and Economic Statistics*, **17**, 91–108.

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

Judge, G. G., Hill, R. C., Griffiths, W. E., Lütkepohl, H. and Lee, T.-C. (1985). *The Theory and Practice of Econometrics*, 2nd ed., Wiley, New York.

Lütkepohl, H. (1982). Differencing multiple time series: Another look at Canadian money and income data, *Journal of Time Series Analysis*, **3**, 235–243.

Lütkepohl, H. (1993). *Introduction to Multiple Time Series Analysis*, 2nd ed., Springer-Verlag, Berlin.

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.

Phillips, P. C. B. and Durlauf, S. N. (1986). Multiple time series regression with integrated processes, *Review of Economic Studies*, **53**, 473–495.

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

Reinsel, G. C. and Ahn, S. K. (1992). Vector AR models with unit roots and reduced rank structure: Estimation, likelihood ratio test, and forecasting, *Journal of Time Series Analysis*, **13**, 353–375.

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

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

Sims, C. A., Stock, J. H. and Watson, M. W. (1990). Inference in linear time series models with some unit roots, *Econometrica*, **58**, 113–144.

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 Association*, **90**, 253–267.