Testing in vector autoregressions with possibly seasonally and non-seasonally (co-)integrated processes

Stéphane Gregoir*
EDHEC Business School

January 2010

Abstract

To avoid possible distortions introduced by seasonal adjustment, econometricians may want to work with non seasonally adjusted data. Without a prioris on the DGP of the seasonal features, this involves numerous seasonal and non seasonal unit root and cointegration rank tests that may imply pretest biases. Drawing on Toda and Yamamoto (1995) lag-augmented approach, we show how we can estimate VAR's formulated in levels and test for general restrictions on the matrix parameters when the processes may be integrated or cointegrated at various frequencies and of various orders. When introducing in a levels VAR for each variable at least as many additional lags as the number of unit roots present in its individual data generating process, we show that the Wald test statistic associated to nonlinear constraints on the initial VAR parameters is asymptotically chi-squared distributed. Size and power properties of this approach are illustrated with a Monte Carlo exercise

Key-words : test, seasonal unit roots, seasonal cointegration, seasonal adjustment
JEL code :C12, C32, E32

1 Introduction

Many macroeconomic and financial intra-annual time series present regular patterns associated to seasonal causes (biological, meteorological or institutional ones). In their attempt to model these time series and test for hypotheses about their Data Generating Process, econometricians can adopt two approaches. They can seasonally adjust them to remove the seasonal patterns from the data or they can try to explicitly model the seasonal features. Both approaches have their limits and advantages, particularly when the econometrician wants to model simultaneously a set of variables with VAR-type models.

Most practitioners use seasonally adjusted data. They pile up in a vector separately seasonally adjusted components since the usual seasonal adjustment statistical procedures are univariate. This treatment nevertheless may have altered the original relationships between

*I thank participants to EC2-2007 conference (Faro, University of Algarve), ESEM2008 (Milan), Toulouse University seminar for their useful comments. Remaining errors are my sole responsibility.

the variables. This is a long-standing issue. It was a subject of research in the seventies (Porter (1975), Geweke (1979), Plosser (1979) and Wallis (1976)) but then sank into oblivion. Today seasonal adjustment softwares offer a large number of optional statistical treatments that are applied independently on each time series as well as several different or data-dependent filters. For instance, seasonally adjusted data in the middle of the sample are obtained by various linear combinations of past and future values. According to the selected statistical approach, the choice of these linear combinations can be data driven or data dependent. Sims (1974) and Wallis (1974) pleaded for the use of the same linear filter to all series appearing in a multiple regression to avoid distortion in their relationships. Similarly, occurrence of outliers at the same dates or around the same dates in different time series is an informative feature that may not be preserved by automatic statistical detection and correction of outliers implemented in these softwares. At last, most of the filters that are used in this statistical treatment can lead to over differenced processes at the seasonal frequencies and may lead to the non-existence of finite order VAR approximation (Maravall (1993)).

When the econometricians work on non-seasonally adjusted data, they have to model their seasonal features. They can resort to a deterministic modeling of the intra-annual movements or a periodic stochastic accumulation of shocks, in other words, the introduction of unit-roots at seasonal frequencies related to the observation frequency. Specification tests have been designed to discriminate situations in which one approach is more in concordance with the data than the other one (Canova and Hansen (1995), Caner (1998), Hasza and Fuller (1982), Said and Dickey (1984), Hylleberg, Engle, Granger and Yoo (1990), Smith and Taylor (1998) among others). In practice from empirical studies, it seems that numerous time series can be parsimoniously described as integrated ones but at a subset of the eligible frequencies. Unfortunately, these tests as well as tests for unit root at zero frequency (Dickey and Fuller (1979), Phillips and Perron (1988) inter alios) are known to have a low power against their respective alternative hypothesis. These power properties can be improved (Elliott, Rothenberg and Stock (1996), Ng and Perron (2001), Gregoir (2006), Rodrigues and Taylor (2007)), but the gain remains limited. When working simultaneously with several seasonally integrated processes in a multivariate set-up, the practitioner is naturally confronted to the problem of cointegration (Granger (1983)) and seasonal cointegration (Engle, Granger and Hallman (1989)). (S)He must determine the presence of seasonal cointegration at each possible and reasonable frequency and the dimension of each cointegration space to be able to specify and estimate a VECM that involves various error correction terms. Tests and estimations procedure have been developed (e.g. Johansen (1988,1991), Harris (1997), Phillips and Ouliaris (1990), Stock and Watson (1989), Lee (1992), Johansen and Schaumburg (1999), Gregoir (1999), Cubbada (2001)). Simulation studies (Reimers (1992) and Toda (1995) inter alios), have shown that at frequency 0, these tests for cointegrating ranks in the maximum likelihood framework are sensitive to the values of nuisance parameters in sample sizes that are typical for economic time series. This means that standard approach consisting of testing for economic hypotheses conditionally on tests for the presence of various unit roots and cointegration ranks may suffer from pretest biases.

However the alternative of working directly on level VAR model is not straightforward. Sims, Stock and Watson (1990) and Toda and Phillips (1993) have shown that the Wald test statistic of linear constraints based on levels estimation may have non-standard asymptotic distributions and may depend on nuisance parameters. Toda and Yamamoto (1995), Dolado and Lütkepohl (1996) and Yamamoto (1996) have proposed a simple way to overcome these problems in hypothesis testing when working with levels VAR for VAR processes that may have unit roots at frequency 0. Toda and Yamamoto (1995) propose to introduce at least as many additional lags as the highest order of integration at frequency 0 in the model at hand. This allows for testing for linear and nonlinear restrictions with standard Wald test statistic on the coefficients by estimating a levels VAR. This is nevertheless at the cost of a possible inefficient use of the information and may have consequences in terms of size and power of the hypothesis tests in finite samples (Kurozumi and Yamamoto (2000)).

We propose to extend their approach to the situation of a DGP with possibly different roots on the unit circle and cointegration. We first give a simple example with seasonal unit roots which leads to a situation similar to those illustrated by Sims, Stock and Watson (1990) and Toda and Phillips (1993). This motivates our interest in a procedure that extends Toda and Yamamoto (1995) approach to more general DGPs. We then state an algebraic property that allows us to rewrite a levels VAR model in terms of covariance stationary variables and a set of integrated variables at each frequency involved in the DGP. This allows us to derive the asymptotic distribution of a Wald test statistic of non linear restriction on the VAR matrix coefficients in a lag-augmented regression. As soon as the number of lags is larger than the original one augmented for each component of the number of unit roots present in their individual generating process, the Wald statistic is asymptotically chi-squared distributed. This thus allows the practitioner to run test in a VAR model of non seasonally adjusted data without testing for the presence of seasonal and non-seasonal unit roots and seasonal and non-seasonal cointegrating relationships.

2 Introductory example

We first illustrate that the Wald test statistic of linear constraints based on levels estimation may have non-standard asymptotic distribution and depend on nuisance parameters when some seasonal unit roots are present in the Data Generating Process. This is quite similar to Sims, Stock and Watson (1990) and Toda and Phillips (1993) results. We focus on a simple case. Let us consider the following bivariate VAR(1) model:

$$\begin{aligned} y_t &= \phi_\alpha y_{t-1} + \varepsilon_t \\ &= \begin{pmatrix} 0 & \alpha \\ -\frac{1}{\alpha} & 0 \end{pmatrix} y_{t-1} + \varepsilon_t \end{aligned} \tag{1}$$

where $\alpha \neq 0$ and $\{\varepsilon_t\}$ is a bivariate strong white noise with $V\varepsilon_t = I_2$. $\phi_\alpha(L) = I_2 - \phi_\alpha L$ is such that $\det \phi_\alpha(L)$ has two unit roots $\{i, -i\}$. It can be shown that the process $\{y_t\}$ is integrated of order 1 at the two frequencies $\frac{\pi}{2}$ and -$\frac{\pi}{2}$ and cointegrated at these frequencies (cf. section 6). We are interested for instance in testing for $H_0 : \phi_{\alpha,12} = \alpha$ and propose to consider the associated Student test statistic. We have the following result:

Lemma 1 *Let $\widehat{\phi}_{\alpha,12}$ be the OLS estimate of $\phi_{\alpha,12}$ in* (1), *then*

$$T\left(\widehat{\phi}_{\alpha,12} - \alpha\right) \Longrightarrow \frac{\Re\left\{ \begin{pmatrix} \alpha & 0 \end{pmatrix} \int dW(s)\,\overline{W}(s)'\,ds \begin{pmatrix} 1 \\ -\alpha i \end{pmatrix} \right\}}{\frac{1}{2} \begin{pmatrix} 1 & \alpha i \end{pmatrix} \int W(s)\,\overline{W}(s)'\,ds \begin{pmatrix} 1 \\ -\alpha i \end{pmatrix}}$$

and

$$\frac{\widehat{\phi}_{\alpha,12}-\alpha}{\sqrt{\widehat{V}_{\widehat{\phi}_{\alpha,12}}}} \Longrightarrow \frac{\Re\left\{\begin{pmatrix} \alpha & 0 \end{pmatrix} \int dW(s)\,\overline{W}(s)'\,ds \begin{pmatrix} 1 \\ -\alpha i \end{pmatrix}\right\}}{\sqrt{\frac{1}{2}\begin{pmatrix} 1 & \alpha i \end{pmatrix} \int W(s)\,\overline{W}(s)'\,ds \begin{pmatrix} 1 \\ -\alpha i \end{pmatrix}}}$$

where $W(s) = W_R(s) + iW_I(s)$ with $W_R(s)$ and $W_I(s)$ two independent bivariate real Wiener processes whose variance-covariance matrix is equal to I_2.

The OLS estimator is superconsistent but its asymptotic distribution is not Gaussian and depends on the value of α. The Student test statistic is not asymptotically normally distributed and depends on the value of α. In section 6, we give the form of the VECM satisfied by this process and use it to study the small sample properties of the test procedure we are now introducing.

3 General objective

In this section, we introduce the general objective of this paper as well as our framework. This requires stating some definitions and introducing some notations and assumptions. We start from the well-known definition of integratedness, then describe the data generating process of the process under study and the problem we are interested in. We work with complex processes, this simplifies the notations. The case of real processes is derived in this framework by adding constraints that will be detailed in a set of footnotes when necessary as we go along.

A purely non deterministic process integrated of order d at the only frequency ω is such that when we apply the first difference operator at this frequency raised at the power d, we get a covariance-stationary process (that is not overdifferenced). We denote this first difference operator $\delta_\omega(L) = (1 - e^{-i\omega}L)$. This is a complex operator. The definition of a univariate integrated of order d process at the only frequency ω takes then the following form:

Definition 2 *A univariate purely non deterministic process $\{z_t\}_{t\in\mathbb{Z}}$ is said to be integrated of order $d \in \mathbb{N}$ at the only frequency $\omega \in\,]-\pi, \pi]$, if it is such that*

$$\left(1 - e^{-i\omega}L\right)^d z_t \;=\; \eta_t$$

where $\{\eta_t\}_{t\in\mathbb{Z}}$ is a (complex) purely non deterministic covariance stationary process such that its spectral density is strictly positive at frequency ω.

In practice, univariate processes can be integrated of various orders at various frequencies, in particular non-seasonally adjusted processes may parsimoniously be described as processes simultaneously integrated at various seasonal frequencies. This is the framework used in standard seasonal adjustment procedure such as X12 or TRAMO-SEATS. The appropriate set of first difference operators must then be applied to get a covariance stationary process that is not overdifferenced[1].

[1]When the process under study is real, if $\omega \notin \{0, \pi\}$ is a frequency of integration of order d, then necessary so is $-\omega$ with the same order d (see Gregoir (1999a)). In these circumstances, to obtain a stationary process we must apply to the integrated process the first difference operator at both frequencies raised at the power d. The corresponding real first difference operator is equal to $\delta_\omega(L)\delta_{-\omega}(L) = \left(1 - 2\cos\omega L + L^2\right)$.

In this paper, we consider a n-dimensional process $\{y_t\}_{t\in\mathbb{Z}}$ whose data generating process is given by: for $t > -p$,

$$y_t = d(t) + x_t \tag{2}$$

where

(*i*) $d(t)$ is a deterministic function that involves trend and sinusoidal polynomials of various degrees at D different frequencies: for each frequency $\widetilde{\omega}_j$ in $\{\widetilde{\omega}_1, \ldots, \widetilde{\omega}_D\} \in \left]-\pi,\pi\right]^D$ we denote q_j the degree of its polynomial,

$$d(t) = \sum_{j=1}^{D}\left(\sum_{k=0}^{q_j} \beta_{jk} t^k e^{-i\widetilde{\omega}_j t}\right) \tag{3}$$

with $\left\{(\beta_{jk})_{k\in\{0,\ldots,q_j\}}\right\}_{j\in\{1,\ldots,D\}}$ a set of n-dimensional vectors

(*ii*) $\{x_t\}_{t\in\mathbb{Z}}$ is a n-dimensional process that satisfies a p^{th}-order vector autoregression

$$\begin{aligned} x_t &= \phi(L)\,x_t + \varepsilon_t \\ &= \sum_{j=1}^{p} \phi_j x_{t-j} + \varepsilon_t \end{aligned} \tag{4}$$

where p is supposed to be known, we initialize (4) at $t = -p+1, \ldots, 0$ with $O_p(1)$ random vectors and the process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ satisfies the following property:

Assumption 3 *:* $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is n-dimensional martingale difference satisfying $E(\varepsilon_t\,|\mathcal{F}_{t-1}) = 0$, $E(\varepsilon_t\varepsilon_t'\,|\mathcal{F}_{t-1}) = \Omega > 0$ and $\sup_t E\left(\max_{a\in\{1,..,n\}} |\varepsilon_{a,t}|^{2+\delta}\,|\mathcal{F}_{t-1}\right) < +\infty$ a.s. for some $\delta > 0$, where $\mathcal{F}_{t-1}$ is the σ-field generated by $\{\varepsilon_{t-\tau}, \tau = 1, 2, \ldots\}$.

(*iii*) $\{x_t\}_{t\in\mathbb{Z}}$ is integrated of various orders at S different frequencies $\{\omega_1, \ldots, \omega_S\} \in \left]-\pi,\pi\right]^S$ and may be cointegrated at each frequency. This means that the polynomial $\det(I - \phi(u))$ has several roots on the unit circle.

For sake of simplicity, we assume that all the components of x_t are integrated of the same order at the same frequency. It is possible to consider situations in which each component is integrated of different orders at each frequency, this makes the notations more cumbersome. In the sequel, we detail in a set of remarks how the results can be generalized to this situation since we refer to this situation in the introduction and abstract. For each frequency $\omega_j \in \left]-\pi,\pi\right]$, we denote d_j the order of integration common to all the components. The integration structure is therefore summarized by the set $I_x = \{(\omega_j, d_j) \in \left]-\pi,\pi\right] \times \mathbb{N}, j = 1, \ldots, S\}$. For the process $\{x_t\}_{t\in\mathbb{Z}}$ characterized by I_x, we introduce the generalized difference operator $\Delta_x(L) = \prod_{j=1}^{S} \delta_{\omega_j}^{d_j}(L)$ that is such that $\Delta_x(L)\,x_t$ is a n-dimensional covariance stationary process whose spectral density of each component at each frequency ω_j, $j = 1, \ldots, S$ is strictly positive. We denote $d_x = \sum_{j=1}^{S} d_j$ the degree of this polynomial $\Delta_x(L)$.

Substituting $x_t = y_t - d(t)$ into (4), we get

$$y_t = \widetilde{d}(t) + \phi(L)\,y_t + \varepsilon_t \tag{5}$$

where the vector coefficients of $\widetilde{d}(t)$ are functions of $\{\beta_{jk}\}_{k\in\{0,\dots,q_j\}}$, $j \in \{1,\dots,D\}$ and ϕ_l, $l \in \{1,\dots,p\}$. Note that if some frequency $\widetilde{\omega}_j$ in the deterministic part $d(t)$ is a frequency of integration of $\{x_t\}_{t\in\mathbb{Z}}$, the degree $\widetilde{q}_j$ in $\widetilde{d}(t)$ of the associated deterministic sinusoidal polynomial might be lower than the degree q_j in (3). For instance, when $D = 1$, $S = 1$, $\widetilde{\omega}_1 = \omega_1$ and the process is integrated of order d_1, but not cointegrated at this frequency, we have $I_n - \phi(L) = \delta_{\omega}^{d_1}(L)(I_n - \phi_1(L))$ and the degree of the deterministic function at this frequency is $\widetilde{q}_1 = q_1 - d_1$ (when $q_1 - d_1 \geq 0$). In the sequel, we denote $\widetilde{q} = \sum_{j=1}^{D} \widetilde{q}_j$ the number of the linearly independent deterministic functions that span the vector space in which $\widetilde{d}(t)$ takes its values

$$\widetilde{d}(t) = \sum_{j=1}^{D} \left(\sum_{k=0}^{\widetilde{q}_j} \widetilde{\beta}_{jk} t^k e^{-i\widetilde{\omega}_j t} \right)$$

Our interest does not lie in whether the process $\{y_t\}_{t\in\mathbb{Z}}$ is integrated or cointegrated at the different frequencies, but in testing the following hypothesis:

$$H_0 : f(vec\phi) = 0 \tag{6}$$

where $vec\phi$ is the vector obtained in stacking the columns of the $(n \times np)$ matrix $\begin{pmatrix} \phi_1 & \dots & \phi_p \end{pmatrix}$ from equation (5) and $f(.)$ is a m-vector valued function satisfying the following assumptions

Assumption 4 *: $f(.)$ is a twice continuous differentiable function which is such that in the neighborhood of the true value of $vec\phi$, the jacobian matrix $\frac{\partial f}{\partial vec\phi}$ is full rank.*

Drawing on Toda and Yamamoto (1995) lag-augmented VAR approach, we consider estimating by ordinary least squares (OLS) a levels VAR

$$y_t = \widetilde{d}(t) + \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{k=p+1}^{p_a} \phi_k y_{t-k} + \varepsilon_t \tag{7}$$

where $p_a \geq p + d_x$ to test for H_0 with the p OLS estimates of the $\{\phi_j\}_{j=1\dots p}$ in (7). Our main objective is therefore to show that the standard Wald statistic of H_0 based on OLS estimates in (7) is asymptotically chi-square distributed with m degrees of freedom as soon as $p_a \geq p + d_x$. Our result relies on algebraic and statistical properties. The next section is devoted to algebraic properties of matrix polynomials that allow us to reformulate the test problem under study and the following one deals with statistical properties of the OLS estimators. For sake of simplicity, in this latter section, we limit our attention to at most $I(1)$ processes at a set of a priori known frequencies with linear trend and seasonal dummies. $I(2)$ processes at frequency 0 were considered by Toda and Yamamoto (1995) as they sometimes are used to describe some macroeconomic nominal variables, but $I(2)$ processes at seasonal frequencies do not seem to be used in practice.

4 Algebraic results and reformulation of the null hypothesis

We first work on a reparameterization of the matrix polynomial associated to the VAR(p) model and then illustrate this reparameterization in some standard examples. Finally, we show that the Wald statistic of H_0 can take two equivalent forms, one of them involving matrix coefficients associated to covariance stationary regressors, which is an indication that standard asymptotics applies.

4.1 Polynomial factorization

We first rewrite the matrix polynomial $\phi(L)$ under a form that involves p lagged values of covariance-stationary processes and a set of d_x processes that are integrated at each frequency and each intermediate order of integration of the data generating process. Furthermore, we show that the p matrix coefficients associated to the covariance-stationary processes obtained in this reparameterization are in a one-to-one correspondence with the p initial matrix coefficients $\{\phi_j\}_{j=1,\ldots,p}$, so that it is possible to translate a set of constraints on the latter in an equivalent one on the former. These results are presented in a Theorem and a Lemma. Their proofs are given in Appendix. Toda and Yamamoto (1995) approach corresponds to a particular case of this algebraic framework.

Theorem 5 *Let $\phi(L)$ be a $(n \times n)$ matrix polynomial of degree p*

$$\phi(L) = \phi_1 L + \ldots + \phi_p L^p$$

where the $(\phi_j)_{j=1,\ldots,p}$ are $(n \times n)$ possibly complex matrices. There exist two $(n \times n)$ matrix polynomials $\psi(L)$ (the quotient) and $R(L)$ (the remainder) of respective degree p and $d_x - 1$ such that

$$\begin{aligned} \phi(L) &= \Delta_x(L)\psi(L) + L^{p+1}R(L) \\ \psi(L) &= \psi_1 L + \ldots + \psi_p L^p \end{aligned}$$

and there exists M a full-rank upper triangular $(p \times p)$ matrix such that

$$\begin{pmatrix} \psi_1 & \ldots & \psi_p \end{pmatrix} = \begin{pmatrix} \phi_1 & \ldots & \phi_p \end{pmatrix} M \otimes I_n$$

We emphasized that the matrix coefficients $\{\psi_j\}_{j=1,\ldots,p}$ are linear functions of the initial matrix coefficients $\{\phi_j\}_{j=1,\ldots,p}$, but this also holds for the matrix coefficients of $R(L)$. The relationship between the two sets of parameters is linear.

Remark 6 *When the components of $\{x_t\}_{t\in\mathbb{Z}}$ have various orders of integration at various frequencies, we have to introduce the sets $I_{x,a}$, $a \in \{1,\ldots n\}$, associated to the integration structure of each component $\{x_{a,t}\}_{a\in\{1,\ldots,n\},t\in\mathbb{Z}}$ and the related generalized difference operators $\left\{\Delta_x^{(a)}(L)\right\}_{a\in\{1,\ldots,n\}}$ such that $\Delta_x^{(a)}(L)\,x_{a,t}$ is a univariate covariance stationary process with a nonzero spectral density at the associated frequencies. We note $d_x^{(a)}$ the degree of $\Delta_x^{(a)}(L)$. We then apply the above Theorem when $n = 1$ on each polynomial coefficient of the matrix polynomial $\phi(L) = [\phi_{ba}(L)]_{(b,a)\in\{1,\ldots,n\}^2}$ in the following way:*

$$\phi_{ba}(L) = \psi_{ba}(L)\,\Delta_x^{(a)}(L) + L^{p+1}R_{ba}(L)$$

where $\psi_{ba}(L)$ and $R_{ba}(L)$ are of respective degree p and $d_x^{(a)}-1$. We set all these equations in a matrix equation where $\Delta_x^{(0)}(L)$ stands for the diagonal matrix whose elements are equal to $\Delta_x^{(a)}(L)$ as follows

$$\phi(L)=\psi(L)\,\Delta_x^{(0)}(L)+L^{p+1}R(L)$$

where the degree of the polynomials in the a^{th} column of $R(L)$ is $d_x^{(a)}-1$.

We now introduce a set of difference operators that when applied on $\{x_t\}_{t\in\mathbb{Z}}$ give integrated processes at a single frequency. For $j=1,\ldots S$, we put $\Delta_{x,-j}(L)=\prod_{k=1,k\neq j}^{S}\delta_{\omega_k}^{d_k}(L)$, that is such that $\{\Delta_{x,-j}x_t\}_{t\in\mathbb{Z}}$ is $I_{\omega_j}(d_j)$. An algebraic result allows us to decompose the remainder $R(L)$ into the sum of polynomials that when applied on $\{y_t\}_{t\in\mathbb{Z}}$ involve processes that are integrated at an only frequency and are therefore constructed with the set of operators $\{\Delta_{x,-j}\}_{j\in\{1,\ldots S\}}$. This corresponds to the property that $\left\{\left(\Delta_{x,-j}(L)\,\delta_{\omega_j}(L)^k\right)_{k=0,\ldots d_j-1}\right\}_{j=1,\ldots S}$ is a basis of the polynomials of degree at most d_x-1. We state a general result.

Lemma 7 *For any polynomial matrix $Q(L)$ of degree d_x-1, there exist S unique matrix polynomials of respective degree (d_j-1), $j=1,\ldots,S$*

$$Q_j(L)=\sum_{k=0}^{d_j-1}Q_{j,k}L^k$$

such that the matrix polynomial $Q(L)$ can be rewritten under the following form:

$$Q(L)=\sum_{j=1}^{S}Q_j\left(\delta_{\omega_j}(L)\right)\Delta_{x,-j}(L)$$

It follows from Lemma 7 that the polynomial matrix $R(L)$ introduced in Theorem 5 can be rewritten under the following form:

$$R(L)=\sum_{j=1}^{S}R_j\left(\delta_{\omega_j}(L)\right)\Delta_{x,-j}(L)$$

where the R_j's are polynomial matrices of respective degree d_j-1. When $R(L)$ is applied to y_t, the vector space spanned by y_t and the d_x-1 lagged variables can be decomposed into the direct sum of S subspaces, each of them generated by processes integrated at one of the frequencies under study, say ω_j, with orders from 1 to d_j.

Remark 8 *When the components of $\{x_t\}_{t\in\mathbb{Z}}$ have various orders of integration at various frequencies, we have to complete the notations introduced in Remark 6 . Let S_a the number of integration frequencies of the a^{th} component and $d_j^{(a)}$ their respective order of integration, $j\in\{1,\ldots S_a\}$. We then introduce the sets of operators $\left\{\left(\Delta_{x,-j}^{(a)}(L)=\prod_{k\in I_{x,a},k\neq j}^{S_a}\delta_{\omega_k}^{d_k}(L)\right)_{j\in I_{x,a}}\right\}_{a\in\{1,\ldots n\}}$*

, that is such that $\left\{\Delta_{x,-j}^{(a)} x_{a,t}\right\}_{t\in\mathbb{Z}}$ is $I_{\omega_j}\left(d_j^{(a)}\right)$. Lemma 7 can be used but on the column structure of $R(L)$. For the a^{th} column, there exist S_a unique vector polynomials of respective degree $\left(d_j^{(a)}-1\right)$, $j=1,...,S_a$

$$R_{j,a}(L)=\sum_{k=0}^{d_j^{(a)}-1} R_{j,ka}L^k$$

such that the a^{th} column of the matrix polynomial $R(L)$ can be rewritten under the following form:

$$R_{.,a}(L)=\sum_{j\in I_{x,a}} R_{j,a}\left(\delta_{\omega_j}(L)\right)\Delta_{x,-j}^{(a)}(L)$$

4.2 Examples

We illustrate the above results with simple examples associated to quarterly and monthly data. First, we consider a real n-dimensional process whose components are supposed to be integrated of order 1 at each seasonal frequencies associated to quarterly observations, namely $\left\{0,\frac{\pi}{2},-\frac{\pi}{2},\pi\right\}=\{\omega_1,\omega_2,\omega_3,\omega_4\}$. We limit the deterministic terms to seasonal dummies and a constant term. It is well-known that the space generated by these deterministic functions is also spanned by the four real functions $\left\{1,\cos\frac{\pi}{2}t,\sin\frac{\pi}{2}t,\cos\pi t\right\}$. Using the relationships $\cos\frac{\pi}{2}t=\frac{1}{2}\left(e^{i\frac{\pi}{2}t}+e^{-i\frac{\pi}{2}t}\right)$ and $\sin\frac{\pi}{2}t=\frac{1}{2i}\left(e^{i\frac{\pi}{2}t}-e^{-i\frac{\pi}{2}t}\right)$, we parameterize the deterministic function as follows

$$d(t)=\beta_0+\beta_{-\frac{\pi}{2}}e^{i\frac{\pi}{2}t}+\beta_{\frac{\pi}{2}}e^{-i\frac{\pi}{2}t}+\beta_\pi(-1)^t$$

with $\beta_{-\frac{\pi}{2}}=\overline{\beta_{\frac{\pi}{2}}}$ to ensure that $d(t)$ is real. We start from

$$\begin{aligned} y_t &= d(t)+x_t \\ x_t &= \sum_{j=1}^{p}\phi_j x_{t-j}+\varepsilon_t \end{aligned}$$

The generalized first difference operator is the usual seasonal first difference operator

$$\begin{aligned} \Delta_x(L) &= (1-L)\left(1-e^{-i\pi}L\right)\left(1-e^{i\frac{\pi}{2}}L\right)\left(1-e^{-i\frac{\pi}{2}}L\right) \\ &= \left(1-L^2\right)\left(1+L^2\right) \\ &= \left(1-L^4\right) \end{aligned}$$

It is such that $\Delta_x(L)\,d(t)=0$. Equation (5) takes the following form

$$y_t=(I-\phi(L))\,d(t)+\sum_{j=1}^{p}\phi_j y_{t-j}+\varepsilon_t$$

Theorem 5 and Lemma 7 can be applied to $\phi(L)$ and $(I-\phi(L))$. We keep to the notations introduced above for $\phi(L)$ and set

$$I-\phi(L)=\zeta(L)\,\Delta_x(L)+L^{p+1}\sum_{j=1}^{4}Q_j\Delta_{x,-j}$$

It is such that $Q_j \Delta_{x,-j}\left(e^{-i\omega_j}\right) = I - \phi\left(e^{-i\omega_j}\right)$ and therefore

$$
\begin{aligned}
(I - \phi(L))\, d(t) &= (I - \phi(1))\, \beta_0 + (I - \phi(-1))\, \beta_\pi (-1)^{t-p-1} \\
&\quad + (I - \phi(i))\, \beta_{-\frac{\pi}{2}} e^{i\frac{\pi}{2}(t-p-1)} + (I - \phi(-i))\, \beta_{\frac{\pi}{2}} e^{-i\frac{\pi}{2}(t-p-1)} \\
&= \widetilde{d}(t)
\end{aligned}
$$

This allows us to rewrite the DGP equation as follows

$$
\begin{aligned}
y_t &= \sum_{j=1}^{p} \psi_j \left(y_{t-j} - y_{t-j-4}\right) + \Big(R_1 \left(1 + L + L^2 + L^3\right) y_{t-p-1} \\
&\quad + R_2 \left(1 - iL - L^2 + iL^3\right) y_{t-p-1} + R_3 \left(1 + iL - L^2 - iL^3\right) y_{t-p-1} \\
&\quad + R_4 \left(1 - L + L^2 - L^3\right) y_{t-p-1}\Big) + \varepsilon_t + \widetilde{d}(t)
\end{aligned}
$$

where
$\left\{\left(1 + L + L^2 + L^3\right) y_t\right\}_{t\in\mathbb{Z}}$,
$\left\{\left(1 - iL - L^2 + iL^3\right) y_t\right\}_{t\in\mathbb{Z}}$,
$\left\{\left(1 + iL - L^2 - iL^3\right) y_t\right\}_{t\in\mathbb{Z}}$
and
$\left\{\left(1 - L + L^2 - L^3\right) y_t\right\}_{t\in\mathbb{Z}}$
are respectively integrated of order 1 at frequencies 0, $\frac{\pi}{2}$, $-\frac{\pi}{2}$ and π. For instance, when $p = 1$, we have

$$
\begin{aligned}
y_t &= \phi y_{t-1} + \widetilde{d}(t) + \varepsilon_t \\
&= \phi\left(y_{t-1} - y_{t-5}\right) + \phi y_{t-5} + \widetilde{d}(t) + \varepsilon_t \\
&= \phi\left(y_{t-1} - y_{t-5}\right) + \varepsilon_t + \widetilde{d}(t) \\
&\quad + \frac{1}{4}\phi\left(y_{t-2} + y_{t-3} + y_{t-4} + y_{t-5}\right) \\
&\quad - \frac{i}{4}\phi\left(y_{t-2} - iy_{t-3} - y_{t-4} + iy_{t-5}\right) \\
&\quad + \frac{i}{4}\phi\left(y_{t-2} + iy_{t-3} - y_{t-4} - iy_{t-5}\right) \\
&\quad - \frac{1}{4}\phi\left(y_{t-2} - y_{t-3} + y_{t-4} - y_{t-5}\right)
\end{aligned}
$$

and when $p=4$, we have

$$\begin{aligned}
y_t &= \sum_{j=1}^{4}\phi_j\left(y_{t-j}-y_{t-j-4}\right)+\sum_{j=1}^{4}\phi_j y_{t-j-4}+\widetilde{d}\left(t\right)+\varepsilon_t \\
&= \sum_{j=1}^{4}\phi_j\left(y_{t-j}-y_{t-j-4}\right)+\widetilde{d}\left(t\right)+\varepsilon_t \\
&\quad +\frac{1}{4}\left(\sum_{j=1}^{4}\phi_j\right)\left(y_{t-5}+y_{t-6}+y_{t-7}+y_{t-8}\right) \\
&\quad +\frac{1}{4}\left(\sum_{j=1}^{4}e^{i(j-1)\frac{\pi}{2}}\phi_j\right)\left(y_{t-5}-iy_{t-6}-y_{t-7}+iy_{t-8}\right) \\
&\quad +\frac{1}{4}\left(\sum_{j=1}^{4}e^{-i(j-1)\frac{\pi}{2}}\phi_j\right)\left(y_{t-5}+iy_{t-6}-y_{t-7}-iy_{t-8}\right) \\
&\quad +\frac{1}{4}\left(\sum_{j=1}^{4}e^{i(j-1)\pi}\phi_j\right)\left(y_{t-5}-y_{t-6}+y_{t-7}-y_{t-8}\right)
\end{aligned}$$

Second, we consider a real n-dimensional process whose components are supposed to be integrated of order 1 at each seasonal frequencies associated to monthly observations, namely $\left\{-\frac{5\pi}{6},-\frac{2\pi}{3},-\frac{\pi}{2},-\frac{\pi}{3},-\frac{\pi}{6},0,\frac{\pi}{6},\frac{\pi}{3},\frac{\pi}{2},\frac{2\pi}{3},\frac{5\pi}{6},\pi\right\}=\left\{\omega_j\right\}_{j=1,\ldots,12}$. We limit our attention to a DGP without deterministic terms

$$y_t=\sum_{j=1}^{p}\phi_j y_{t-j}+\varepsilon_t$$

The generalized first difference operator is the usual seasonal first difference operator

$$\begin{aligned}
\Delta_x\left(L\right) &= \prod_{j=1}^{12}\left(1-e^{-i\omega_j}L\right) \\
&= \left(1-L^{12}\right)
\end{aligned}$$

and the family of difference operators that define processes integrated of order 1 at each particular frequency, say ω_j, is given by

$$\Delta_{x,-j}\left(L\right)=\sum_{k=0}^{11}e^{-ik\omega_j}L^k$$

Theorem 5 and Lemma 7 allow us to rewrite this equation as follows

$$\begin{aligned}
y_t &= \sum_{j=1}^{p}\psi_j\left(y_{t-j}-y_{t-j-12}\right)+\sum_{j=1}^{12}R_j\left(\sum_{k=0}^{11}e^{-ik\omega_j}L^k\right)y_{t-p-1}+\varepsilon_t \\
&= \sum_{j=1}^{p}\psi_j\left(y_{t-j}-y_{t-j-12}\right)+L^{p+1}\sum_{j=0}^{11}\left(\sum_{k=1}^{12}e^{-i(j-1)\omega_k}R_k\right)L^j y_t+\varepsilon_t
\end{aligned}$$

When $p=12$, we get

$$y_t=\sum_{j=1}^{12}\phi_j\left(y_{t-j}-y_{t-j-12}\right)+\sum_{j=1}^{12}\frac{1}{12}\left(\sum_{l=1}^{12}e^{i(l-1)\omega_j}\phi_l\right)\left(\sum_{k=0}^{11}e^{-ik\omega_j}L^k\right)y_{t-13}+\varepsilon_t$$

4.3 General test procedure

Testing for H_0 with OLS estimates derived from (7) will correspond to testing with standard Wald statistic ignoring the under the null zero matrix coefficients introduced in the lag augmented regression. The Wald test statistic can be derived from iterative applications of Frisch-Waugh Theorem. It is convenient to write (7) with matrix notations as follows: Let τ be the $\widetilde{d}\times T$ matrix whose columns τ_t are equal to $\begin{pmatrix}\tau_{1,t} & \ldots & \tau_{D,t}\end{pmatrix}'$ where $\tau_{j,t}=\begin{pmatrix}e^{-i\widetilde{\omega}_jt} & te^{-i\widetilde{\omega}_jt} & \ldots & t^{\widetilde{d}_j}e^{-i\widetilde{\omega}_jt}\end{pmatrix}$ for $j=1,...D$, y_{-1} the $np\times T$ matrix whose columns are equal to $\begin{pmatrix}y'_{t-1} & \ldots & y'_{t-p}\end{pmatrix}'$ and y_{-p_a} the $n\left(p_a-p\right)\times T$ matrix whose columns are equal to $\begin{pmatrix}y'_{t-p-1} & \ldots & y'_{t-p-p_a}\end{pmatrix}'$ and y and ε the $n\times T$ matrices whose columns are equal to y_t and ε_t, then

$$y=\widetilde{\beta}\tau+\phi y_{-1}+\Phi y_{-p_a}+\varepsilon \tag{8}$$

with $\underset{\left(n\times\widetilde{d}\right)}{\widetilde{\beta}}=\begin{pmatrix}\widetilde{\beta}_{10} & \ldots & \widetilde{\beta}_{D\widetilde{d}_D}\end{pmatrix}$, $\underset{(n\times np)}{\phi}=\begin{pmatrix}\phi_1 & \ldots & \phi_p\end{pmatrix}$ and $\underset{(n\times n(p_a-p))}{\Phi}=\begin{pmatrix}\phi_{p+1} & \ldots & \phi_{p_a}\end{pmatrix}$.

Let $P_\tau=I_T-\overline{\tau}'\left(\tau\overline{\tau}'\right)^{-1}\tau$ and

$$P_{y_{-p_a}}=P_\tau-P_\tau\overline{y}_{-p_a}{}'\left(y_{-p_a}P_\tau\overline{y}_{-p_a}{}'\right)^{-1}y_{-p_a}P_\tau$$

the Wald statistic is equal to

$$\xi_W=\overline{f\left(vec\widehat{\phi}\right)}'\left[\frac{\partial f}{\partial vec\phi'}\left\{\left(y_{-1}P_{y_{-p_a}}\overline{y}'_{-1}\right)^{-1}\otimes\widehat{\Omega}_\varepsilon\right\}\overline{\frac{\partial f}{\partial vec\phi}}'\right]^{-1}f\left(vec\widehat{\phi}\right) \tag{9}$$

with $\widehat{\phi}=yP_{y_{-p_a}}\overline{y}'_{-1}\left(y_{-1}P_{y_{-p_a}}\overline{y}'_{-1}\right)^{-1}$ and $\widehat{\Omega}_\varepsilon=\frac{1}{T}\widehat{\varepsilon}\overline{\widehat{\varepsilon}}'$.

From Theorem 5, we know that we can rewrite (5) under the form

$$y_t=\widetilde{d}\left(t\right)+\psi\left(L\right)\Delta_xy_t+R\left(L\right)y_{t-p-1}+\varepsilon_t$$

where $\psi\left(L\right)=\psi_1L+\ldots+\psi_pL^p$ and

$$\begin{pmatrix}\psi_1 & \ldots & \psi_p\end{pmatrix}=\begin{pmatrix}\phi_1 & \ldots & \phi_p\end{pmatrix}M\otimes I_n$$

with M a full ranked matrix. It follows that $f\left(vec\phi\right)=0$ can be expressed with this new set of parameters under the form $f\left(\left(\left(M^{-1}\right)'\otimes I_n\right)\otimes I_nvec\psi\right)=g\left(vec\psi\right)=0$. In the neighborhood of the true value of $vec\phi$, the jacobian matrix $\frac{\partial f}{\partial vec\phi}$ is assumed to be full rank which ensures that the jacobian matrix $\frac{\partial g}{\partial vec\psi}=\frac{\partial f}{\partial vec\phi}\left(\left(\left(M^{-1}\right)'\otimes I_n\right)\otimes I_nvec\psi\right)\left(\left(M^{-1}\right)'\otimes I_n\right)\otimes I_n$ is also full rank.

From Lemma 7, we can rewrite (5) under the form

$$y_t=\widetilde{d}\left(t\right)+\psi\left(L\right)\Delta_xy_t+\sum_{j=1}^{S}R_j\left(\delta_{\omega_j}\right)\Delta_{x,-j}y_{t-p-1}+\varepsilon_t \tag{10}$$

which we also write with matrix notations as follows : Let $\Delta_x y$ be the $np \times T$ matrix whose columns are equal to $\Delta_x y_{t-p}^{t-1} = \begin{pmatrix} \Delta_x y'_{t-1} & \dots & \Delta_x y'_{t-p} \end{pmatrix}'$ and z_{-p} the $nd_x \times T$ matrix whose columns z_t are equal to $\begin{pmatrix} z_{1,t} & \dots & z_{S,t} \end{pmatrix}'$ with

$$z_{j,t} = \begin{pmatrix} \Delta_{x,-j} y'_{t-p-1} & \Delta_{x,-j} \delta_{\omega_j} y'_{t-p-1} & \dots & \Delta_{x,-j} \delta_{\omega_j}^{d_j-1} y'_{t-p-1} \end{pmatrix},$$

then

$$y = \widetilde{\beta}\tau + \psi \Delta_x y + R z_{-p} + \varepsilon \tag{11}$$

with $\underset{(n \times np)}{\psi} = \begin{pmatrix} \psi_1 & \dots & \psi_p \end{pmatrix}$ and $\underset{(n \times nd_x)}{R} = \begin{pmatrix} R_1 & \dots & R_S \end{pmatrix}$ with $\underset{(n \times nd_j)}{R_j} = \begin{pmatrix} R_{j,0} & \dots & R_{j,d_j-1} \end{pmatrix}$.
When $p_a \geq p + d_x$, we apply the algebraic reparameterization of Theorem 5 and Lemma 7 to a matrix polynomial of degree $p_a - d_x \geq p$ obtained by introducing $p_a - p - d_x$ additional lags with zero matrix coefficients. Equation (10) takes the following form

$$y_t = \widetilde{d}(t) + \psi(L) \Delta_x y_t + L^{p+1} \xi(L) \Delta_x y_t + \sum_{j=1}^{S} R_j \left(\delta_{\omega_j}\right) \Delta_{x,-j} y_{t-p_a-1} + \varepsilon_t \tag{12}$$

where $\xi(L)$ is a polynomial matrix of degree $(p_a - p - d_x - 1)$. Since the matrix M in Theorem 5 is an upper triangular matrix, the linear relationship between $\begin{pmatrix} \psi_1 & \dots & \psi_p \end{pmatrix}$ and $\begin{pmatrix} \phi_1 & \dots & \phi_p \end{pmatrix}$ is preserved. Similarly, the space spanned by the regressors in (7) is the same one as that spanned by the regressors in (12), it follows that the OLS residuals and the estimate $\widehat{\Omega}_\varepsilon$ are equal in both regressions. With obvious matrix notations, equation (11) takes the following form

$$y = \widetilde{\beta}\tau + \psi \Delta_x y + \xi \Delta_x y_{-p} + R z_{-p_a} + \varepsilon$$

or equivalently

$$y = \widetilde{\beta}\tau + \psi \Delta_x y + \widetilde{R}\widetilde{z} + \varepsilon \tag{13}$$

with $\widetilde{z} = \begin{pmatrix} \Delta_x y'_{-p} & z'_{-p_a} \end{pmatrix}'$ and $\widetilde{R} = \begin{pmatrix} \xi & R \end{pmatrix}$.

We can now show that testing for $H_0 : f(vec\phi) = 0$ is equivalent to testing for $H'_0 : g(vec\psi) = 0$. The Wald test statistic can again be derived from iterative applications of Frisch-Waugh Theorem. This gives:

Lemma 9 *Let* $P_\tau = I_T - \overline{\tau}' (\tau \overline{\tau}')^{-1} \tau$ *and* $P_{\widetilde{z}} = P_\tau - P_\tau \overline{\widetilde{z}}' \left(\widetilde{z} P_\tau \overline{\widetilde{z}}'\right)^{-1} \widetilde{z} P_\tau$ *, we get that the standard Wald test statistic in* (9) *is such that*

$$\overline{g\left(vec\widehat{\psi}\right)}' \left[\frac{\partial g}{\partial vec\psi'} \left\{ \left(\Delta_x y P_{\widetilde{z}} \overline{\Delta_x y}' \right)^{-1} \otimes \widehat{\Omega}_\varepsilon \right\} \overline{\frac{\partial g}{\partial vec\psi}}' \right]^{-1} g\left(vec\widehat{\psi}\right) = \xi_W \tag{14}$$

with $\widehat{\psi} = y P_{\widetilde{z}} \overline{\Delta_x y}' \left(\Delta_x y P_{\widetilde{z}} \overline{\Delta_x y}' \right)^{-1}$.

We now turn to the analysis of the asymptotic distribution of (14) which is also that of (9).

5 Asymptotic analysis

The formal asymptotic analysis of H_0 testing in VAR's relies on the asymptotic behavior of the partial sums of the process

$$\widetilde{z}_t = \left(\begin{array}{cccc} \Delta_x y'_{t-p-1} & \dots & \Delta_x y'_{t-p_a} & z'_{t-p_a+p} \end{array} \right)'$$

introduced in (13) and its cross-product. We limit our attention to processes that are at most integrated of order one at the various frequencies present in the DGP, possibly cointegrated at each frequency in presence of a deterministic trend and a set of seasonal dummies. Toda and Yamamoto (1995) consider VAR processes that are integrated at most of order two at frequency 0, which allows for polynomial cointegration and makes the analysis more worked out. Empirically, cases of polynomial seasonal cointegration introduced by Gregoir (1999) do not seem frequent and for sake of simplicity we do not consider this situation. The basic arguments are nevertheless quite similar and could be extended to deal with this situation. We start by describing the DGP of the multivariate process under study, we then introduce some operators that allow us to summarize under a convenient form in three Lemmata the asymptotic properties of regressor cross-products. At last, we establish that as soon as $p_a \geq p+S$, the Wald test statistic (14) is χ_2-distributed with the usual degrees of freedom, invariant to whether the process is covariance stationary, integrated of order one or cointegrated at the set of frequencies under scrutiny.

5.1 The data generating process

The DGP we deal with in this section is given by (3) with $D = S$, $\{\widetilde{\omega}_1, \dots, \widetilde{\omega}_D\} = \{\omega_1, \dots, \omega_S\}$, $\omega_1 = 0$, $q_1 = 1$, $\forall j \neq 1$, $q_j = 0$

$$d(t) = \beta_{10} + \beta_{11} t + \sum_{j=2}^{S} \beta_{j0} e^{-i\omega_j t}$$

where $d(t)$ is real so that $\forall \omega_j \in \left]0, \pi\right[$, $\exists k \in \{2, ..., S\}$, such that $\omega_k = -\omega_j$ and $\beta_{k0} = \overline{\beta}_{j0}$ and (4) where the real process $\{x_t\}_{t \in \mathbb{Z}}$ may be $I(0)$, integrated and cointegrated at the frequencies $\{\omega_1, \dots, \omega_S\}$, technically speaking we write

$$x_t = \sum_{j=1}^{p_a} \phi_j x_{t-j} + \varepsilon_t \tag{15}$$

where the roots of the polynomial $\det\left(I_n - \sum_{j=1}^{p_a} \phi_j u^j\right)$ have moduli larger than one or are in $\{e^{i\omega_j}\}_{j=1,\dots S}$. This last equation can be written in a VECM format in applying Theorem 5 to the matrix polynomial $I - \sum_{j=1}^{p_a-S} \phi_j L^j = \psi(L) \Delta_x + L^{p_a-S+1} R(L)$ with $\Delta_x = 1 - L^S$. This gives

$$\Delta_x x_t = \psi(L) \Delta_x x_t + L^{p_a-S+1} R(L) x_t + \sum_{j=p_a-S+1}^{p_a} \phi_j x_{t-j} + \varepsilon_t$$

with

$$L^{p_a-S+1} R(L) + \sum_{j=p_a-S+1}^{p_a} \phi_j L^j = L^{p_a-S+1} \Pi(L)$$

where $\Pi(L)$ is a matrix polynomial of degree $S-1$ than can be decomposed by Lemma 7 into the basis of polynomials $\{\Delta_{x,-j}\}_{j=1,\ldots S}$. This gives the following VECM equation:

$$\Delta_x x_t = \sum_{j=1}^{p_a-S} \psi_j \Delta_x x_{t-j} + \sum_{k=1}^{S} \pi_k \Delta_{x,-k} x_{t-p_a+S-1} + \varepsilon_t$$

where all the matrix coefficients $\{\psi_j\}_{j=1,\ldots p_a-S}$ are real and some matrices π_k associated to frequencies different from 0 and π are complex and such that if π_k is associated to ω_k then $\overline{\pi}_k$ is associated to $-\omega_k$. This equation is similar to those proposed by Johansen and Schaumburg (1999) or Gregoir (1999a) with different notations. In particular, since the last equation is just a rewriting of (15), we have

$$\pi_k = -\frac{e^{-i(p_a-S+1)\omega_k}}{\Delta_{x,-k}\left(e^{i\omega_k}\right)} \left(I_n - \sum_{j=1}^{p_a} \phi_j e^{ij\omega_k} \right)$$

When $e^{i\omega_k}$ is a root of the polynomial $\det\left(I_n - \sum_{j=1}^{p_a} \phi_j u^j\right)$, the rank of π_k or equivalently that of $\left(I_n - \sum_{j=1}^{p_a} \phi_j e^{ij\omega_k}\right)$ is necessarily less than n. The matrix π_k is such that $\pi_k = \alpha_k \overline{\beta}'_k$ for some α_k and β_k that are $n \times r_k$ matrices of rank r_k. Notice that π_k can be equal to 0; in this case, we set $r_k = 0$ and there is no cointegration at this frequency. To ensure that the order of integration is at most 1 at each frequency, we must rule out polynomial cointegration, this corresponds to the following set of conditions (Johansen and Schaumburg (1999)): for all k

$$\overline{\alpha}'_{k,\perp} \left(\sum_{j=1}^{p_a} j\phi_j e^{ij\omega_k} \right) \beta_{k,\perp} \text{ is full rank}$$

where $\alpha_{k,\perp}$ and $\beta_{k,\perp}$ are full rank $(n \times n - r_k)$ matrices such that $\overline{\alpha}'_{k,\perp}\alpha_k = 0$ and $\overline{\beta}'_{k,\perp}\beta_k = 0$[2]. In the sequel, we assume that the above conditions are satisfied. In the matrix notations of (13), for the DGP under study, we have $\widetilde{z} = \begin{pmatrix} \Delta_x y'_{-p} & z'_{-p_a} \end{pmatrix}'$ where columns of $\Delta_x y_{-p}$ are equal to $\begin{pmatrix} \Delta_x y'_{t-p-1} & \ldots & \Delta_x y'_{t-p_a+S} \end{pmatrix}'$ and those of z_{-p_a} to $\begin{pmatrix} \Delta_{x,-1} y'_{t-p_a-1+S} & \ldots & \Delta_{x,-S} y'_{t-p_a-1+S} \end{pmatrix}'$.

5.2 Invariance principle and asymptotic distributions of cross-products

To present the asymptotic convergence of the regressor cross-products, we introduce now an integral operator which plays the role of the inverse of the first difference operator at frequency ω, namely $\delta_\omega = (1 - e^{-i\omega}L)$.

Definition 10 *The integral operator S_ω associates to any sequence $\varepsilon_t = (\varepsilon_t,\ t = \ldots -1,\ 0, 1, \ldots)$ of real numbers a (complex) sequence $S_\omega \varepsilon_t$ defined by :*

$$S_\omega \varepsilon_t = \begin{cases} \sum_{\tau=1}^{t} \varepsilon_\tau e^{-i\omega(t-\tau)} & \text{for } t > 0 \\ 0 & \text{for } t = 0 \\ -\sum_{\tau=0}^{t+1} \varepsilon_\tau e^{-i\omega(t-\tau)} & \text{for } t < 0 \end{cases}$$

[2]This can also be characterized in this case by assuming that the order of multiplicity of each unit root $e^{i\omega_j}$ in $\det\left(I_n - \sum_{j=1}^{p_a} \phi_j u^j\right)$ is equal to $n - rank(\pi_j)$ (Gregoir (1999a)).

Its algebraic properties are summed up in the following statement (Gregoir (1999a)):

Corollary 11

$$(i)\delta_\omega S_\omega = Id, \tag{16}$$

$$(ii)\forall \left(y_t\right)_t, S_\omega \delta_\omega y_t = y_t - y_0 e^{-i\omega t} \tag{17}$$

We stress that the operator δ_ω and S_ω do not commute. The constant term which appears in (ii) in the above Corollary is due to our definition of the integral operator that constructs integrated of order 1 processes that are always equal to 0 at $t = 0$. Alternative definitions with different conventions are possible.

Under assumption 3, Chan and Wei (1988) and Tsay and Tiao (1990) have shown that when T goes to infinity and $\omega \notin \{0, \pi\}$, the following multivariate convergence in distribution holds :

$$\frac{1}{\sqrt{T}} \sum_{\tau=1}^{[Tt]} e^{i\omega\tau} \varepsilon_\tau = \frac{e^{i\omega[Tt]}}{\sqrt{T}} S_\omega \varepsilon_{[T\tau]} \Longrightarrow \frac{1}{\sqrt{2}} W_\omega(t)$$

where $[Tt]$ is equal to the integer part of Tt and $W_\omega(t) = W_{\omega,R}(t) + iW_{\omega,I}(t)$ and $W_{\omega,R}(t)$ and $W_{\omega,I}(t)$ are two real independent Wiener processes whose variance-covariance matrix is Ω. This presentation with complex number is a convenient way to state the joint convergence of $\frac{1}{\sqrt{T}} \sum_{\tau=1}^{[Tt]} \varepsilon_\tau \cos \omega\tau$ and $\frac{1}{\sqrt{T}} \sum_{\tau=1}^{[Tt]} \varepsilon_\tau \sin \omega\tau$. When $\omega \in \{0, \pi\}$, we have

$$\frac{1}{\sqrt{T}} \sum_{\tau=1}^{[Tt]} e^{i\omega\tau} \varepsilon_\tau = \frac{e^{i\omega[Tt]}}{\sqrt{T}} S_\omega \varepsilon_{[T\tau]} \Longrightarrow W_\omega(t)$$

where $W_\omega(t)$ is a real Wiener process whose variance-covariance matrix is Ω.

We now partition the set of regressors $\widetilde{z}$ in two sets, the first one is composed of covariance stationary processes, the second one of integrated of order one processes: for each frequency, we have

$$\begin{aligned}
\Delta_{x,-j} x_t &= \begin{pmatrix} \beta_j \left(\overline{\beta}'_j \beta_j\right)^{-1} & \beta_{j,\perp} \left(\overline{\beta}'_{j,\perp} \beta_{j,\perp}\right)^{-1} \end{pmatrix} \begin{pmatrix} \overline{\beta}'_j \Delta_{x,-j} x_t \\ \overline{\beta}'_{j,\perp} \Delta_{x,-j} x_t \end{pmatrix} \\
&= A_j \begin{pmatrix} \overline{\beta}'_j \Delta_{x,-j} x_t \\ \overline{\beta}'_{j,\perp} \Delta_{x,-j} x_t \end{pmatrix} \\
&= A_j \begin{pmatrix} w_{0,j,t} \\ w_{1,j,t} \end{pmatrix}
\end{aligned}$$

where $\Delta_{x,-j} x_t$ is integrated of order one at the frequency ω_j, $\overline{\beta}'_j \Delta_{x,-j} x_t$ is a complex r_j-dimensional covariance stationary process and $\overline{\beta}'_{j,\perp} \Delta_{x,-j} x_t$ is integrated of order one at ω_j. When the process is not cointegrated, the covariance-stationary term collapses. Let A be a $n(p_a - p) \times n(p_a - p)$ block diagonal matrix whose $p_a - p - S$ first $(n \times n)$ blocks are equal to I_n and for $j = 1, ..., S$, the $p_a - p + j^{th}$ one to A_j, P be a $n(p_a - p) \times n(p_a - p)$ matrix that reorders the rows of $\widetilde{z}$ to collect in the $n(p_a - p - S) + \sum_{j=1}^{S} r_j$ first positions the covariance-stationary components, i.e. $\begin{pmatrix} \Delta_x x'_{t-p-1} & \ldots & \Delta_x x'_{t-p_a+S} \end{pmatrix}'$ and the $w_{0,j,t}$, and in the $nS - \sum_{j=1}^{S} r_j$ last ones, the non-stationary components $w_{1,j,t}$. We note

$$PA\widetilde{z}_t = \begin{pmatrix} w_{0,t} \\ w_{1,t} \end{pmatrix}$$

It is relatively cumbersome to work with this vector completed with $\left(\begin{array}{ccc} \Delta_x x'_{t-1} & ... & \Delta_x x'_{t-p} \end{array} \right)'$ to get the set of regressors. We propose in the statement of the properties used in the derivation of the asymptotic distribution to work with a set of S covariance stationary processes composed of the $n(p_a - p - S) + \sum_{j=1}^{S} r_j + n - r_k$-dimensional processes

$$u_{k,t} = \left(\begin{array}{ccc} \Delta_x x_{t-p}^{t-1\prime} & w'_{0,t} & \delta_{\omega_k} w'_{1,k,t} \end{array} \right)'$$

for $k = 1, ...S$. These processes satisfy a FCLT as follows (cf. Gregoir (2010)): when $\omega_k \notin \{0, \pi\}$

$$\frac{e^{i\omega_k [Tt]}}{\sqrt{T}} S_{\omega_k} u_{k,[T\tau]} \Longrightarrow \frac{1}{\sqrt{2}} B_k(t)$$

where B_k is a complex Wiener process and when $\omega_k \in \{0, \pi\}$

$$\frac{e^{i\omega_k [Tt]}}{\sqrt[2]{T}} S_{\omega_k} u_{k,[T\tau]} \Longrightarrow B_k(t)$$

where B_k is a real Wiener process. In both cases, the variance of B_k is equal to the spectral density matrix of $\{u_{k,t}\}$ at frequency ω_k

$$\begin{aligned} d_k(\omega_k) &= \frac{1}{2\pi} \sum_{j=-\infty}^{+\infty} e^{-ij\omega} E u_{k,t} \overline{u}'_{k,t+j} \\ &= \frac{1}{2\pi} \left(\Sigma_k + \Lambda_{k,\omega_k} + \overline{\Lambda}'_{k,\omega_k} \right) \end{aligned}$$

with $\Sigma_k = E u_{k,t} \overline{u}'_{k,t}$ and $\Lambda_{k,\omega_k} = \sum_{j=1}^{+\infty} e^{-i\omega j} E u_{k,t} \overline{u}'_{k,t+j}$. We partition B_k, Σ_k, Λ_{k,ω_k} et $d_k(\omega_k)$ conformally with $u_{k,t}$ with indexes $x, 0$ and 1, for instance

$$B_k = \left(\begin{array}{ccc} B'_{k,x} & B'_{k,0} & B'_{k,1} \end{array} \right)'$$

Furthermore, from Chan and Wei (1988), an orthogonality property between two integrated processes at two different frequencies

$$\frac{1}{T^2} \sum w_{1,k,t} \overline{w}'_{1,j,t} \to 0$$

allows us to derive the asymptotic behavior of the regressor cross-products.We now summarize the asymptotic behavior of the sample moment matrices that appear in the Wald test statistic in three lemmata. The first one is similar to Lemma 2 in Toda and Yamamoto (1995).

Lemma 12 *Under assumption 3, for the DGP under study, for all $k \in \{1, ..., S\}$*

$$\frac{1}{T} \sum_{t=1}^{T} \left(\begin{array}{c} \Delta_x x_{t-p}^{t-1} \\ w_{0,t} \end{array} \right) \left(\begin{array}{c} \Delta_x x_{t-p}^{t-1} \\ w_{0,t} \end{array} \right)' \to_P \Sigma_{x0,x0} \equiv \left(\begin{array}{cc} \Sigma_{xx} & \Sigma_{x0} \\ \Sigma_{0x} & \Sigma_{00} \end{array} \right)$$

and when $\omega_k \notin \{0, \pi\}$

$$\left(\begin{array}{c} \frac{e^{i\omega_k [Tt]}}{\sqrt{T}} S_{\omega_k} \varepsilon_{[Ts]} \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\begin{array}{c} \Delta_x x_{t-p}^{t-1} \\ w_{0,t} \end{array} \right) \otimes \varepsilon_t \end{array} \right) \Longrightarrow \left(\begin{array}{c} \frac{\mathbf{1}}{\sqrt{2}} W_{\omega_k}(t) \\ \zeta_x \\ \zeta_0 \end{array} \right)$$

and when $\omega_k \in \{0, \pi\}$

$$\left(\begin{array}{c} \frac{e^{i\omega_k [Tt]}}{\sqrt{T}} S_{\omega_k} \varepsilon_{[Ts]} \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\begin{array}{c} \Delta_x x_{t-p}^{t-1} \\ w_{0,t} \end{array} \right) \otimes \varepsilon_t \end{array} \right) \Longrightarrow \left(\begin{array}{c} W_{\omega_k}(t) \\ \zeta_x \\ \zeta_0 \end{array} \right)$$

where ζ *is a normal random vector with mean zero and covariance matrix* $\Sigma_{x0,x0} \otimes \Omega$ *and* ζ, W_{ω_k} *and* W_{ω_j} $(j \neq k)$ *are independent.*

The second Lemma is a restatement of a particular case of Theorem 6 in Gregoir (2010).

Lemma 13 *Under assumption 3, for the DGP under study, for all* $k \in \{1, ..., S\}$, *when* $\omega_k \notin \{0, \pi\}$

$$\frac{1}{T} \sum_{t=1}^{T} w_{1,k,t} \overline{w}'_{0,t} \Longrightarrow \frac{1}{2} \int_0^1 B_{k,1}(t)\, d\overline{B}_{k,0}(t) + \Sigma_{k,10} + \Lambda_{k,10}$$

and when $\omega_k \in \{0, \pi\}$

$$\frac{1}{T} \sum_{t=1}^{T} w_{1,k,t} \overline{w}'_{0,t} \Longrightarrow \int_0^1 B_{k,1}(t)\, dB_{k,0}(t) + \Sigma_{k,10} + \Lambda_{k,10}$$

The third Lemma summarizes the asymptotic behavior of the sample moments.

Lemma 14 *Under assumption 3, for the DGP under study,*

- *(i) when* $\omega_j \notin \{0, \pi\}$,
  - $\frac{1}{T^{1/2}} \sum_{t=1}^{T} e^{i\omega_j t} w_{0,t} \Longrightarrow \frac{1}{\sqrt{2}} B_{j,0}(1)$
  - $\frac{1}{T^{3/2}} \sum_{t=1}^{T} e^{i\omega_j t} w_{1,j,t} \Longrightarrow \frac{1}{\sqrt{2}} \int_0^1 B_j(s)\, ds$
  - $j \neq k$, $\frac{1}{T^{3/2}} \sum_{t=1}^{T} e^{i\omega_k t} w_{1,j,t} \Longrightarrow 0$
  - $\frac{1}{T^2} \sum_{t=1}^{T} w_{1,j,t} \overline{w}'_{1,j,t} \Longrightarrow \frac{1}{2} \int_0^1 B_j(s)\, \overline{B_j}(s)'\, ds$
  - $j \neq k, \frac{1}{T^2} \sum_{t=1}^{T} w_{1,j,t} \overline{w}'_{1,k,t} \Longrightarrow 0$
- *(ii) when* $\omega_j \in \{0, \pi\}$,
  - $\frac{1}{T^2} \sum_{t=1}^{T} w_{1,j,t} w'_{1,j,t} \Longrightarrow \int_0^1 B_j(s)\, B_j(s)'\, ds$
  - $j \neq k, \frac{1}{T^2} \sum_{t=1}^{T} w_{1,j,t} \overline{w}'_{1,k,t} \Longrightarrow 0$
- *(iii) for* $\omega_1 = 0$,
  - $\frac{1}{T^{3/2}} \sum_{t=1}^{T} t\varepsilon_t \Longrightarrow \int s dW_0(s)$
  - $\frac{1}{T^{3/2}} \sum_{t=1}^{T} t w_{0,t} \Longrightarrow \int s dB_0(s)$
  - $\frac{1}{T^{5/2}} \sum_{t=1}^{T} t w_{1,1,t} \Longrightarrow \int_0^1 s B_1(s)\, ds$
  - $j \neq 1$, $\frac{1}{T^{5/2}} \sum_{t=1}^{T} t w_{1,j,t} \Longrightarrow 0$

5.3 Asymptotic distribution of the Wald statistic

We now turn to the analysis of the asymptotic distribution of ξ_W in (14). We first note that

$$\begin{aligned}
\Delta_x y_t &= S\beta_{11} + \Delta_x x_t \\
\Delta_{x,-1} y_t &= S\left[\beta_{10} + \beta_{11}\left(t - \frac{S-1}{2}\right)\right] + \Delta_{x,-1} x_t \\
\Delta_{x,-j} y_t &= S\beta_{j,0} e^{-i\omega_j t} + \beta_{11}\sum_{k=0}^{S-1} k e^{-i\omega_j k} + \Delta_{x,-j} x_t
\end{aligned}$$

whence $\Delta_x y P_\tau = \Delta_x x P_\tau$ and $\forall j \in \{1, ..S\}, \Delta_{x,-j} y P_\tau = \Delta_{x,-j} x P_\tau$. From usual algebra and the standard OLS estimator definition in a complex number framework

$$\widehat{\psi} = y P_{\widetilde{z}} \Delta_x y' \left(\Delta_x y P_{\widetilde{z}} \Delta_x y'\right)^{-1} = y P_{\widetilde{\widetilde{z}}} \Delta_x x' \left(\Delta_x x P_{\widetilde{\widetilde{z}}} \Delta_x x'\right)^{-1}$$

with

$$\widetilde{\widetilde{z}} = \begin{pmatrix} \Delta_x x'_{-p} & z'_{x,-p_a} \end{pmatrix}'$$

and

$$z_{x,-p_a} = \begin{pmatrix} \Delta_{x,-1} x'_{t-p_a-1+S} & \ldots & \Delta_{x,-S} x'_{t-p_a-1+S} \end{pmatrix}'$$

we get that

$$\widehat{\psi} - \psi = \varepsilon P_{\widetilde{\widetilde{z}}} \Delta_x x' \left(\Delta_x x P_{\widetilde{\widetilde{z}}} \Delta_x x'\right)^{-1}$$

and

$$\begin{aligned}
P_{\widetilde{z}} &= P_{\widetilde{\widetilde{z}}} \\
&= P_\tau - P_\tau \overline{\widetilde{\widetilde{z}}}' \left(\widetilde{\widetilde{z}} P_\tau \overline{\widetilde{\widetilde{z}}}'\right)^{-1} \widetilde{\widetilde{z}} P_\tau \\
&= P_\tau - P_\tau \overline{\widetilde{\widetilde{z}}}' \overline{A}' P' \left(P A \widetilde{\widetilde{z}} P_\tau \overline{\widetilde{\widetilde{z}}}' \overline{A}' P'\right)^{-1} P A \widetilde{\widetilde{z}} P_\tau
\end{aligned}$$

where $PA\widetilde{\widetilde{z}}$ is the $T \times n(p_a - p)$ matrix whose elements are $\begin{pmatrix} w'_{0,t} & w'_{1,t} \end{pmatrix}'$. To obtain the limit distribution of the OLS estimator of ψ we use the limits described in the previous set of lemmata. We introduce two block diagonal matrices with the rates of convergence of the different terms, $D_{1,T}$ is associated to the deterministic terms and $D_{2,T}$ to the stochastic ones:

$$D_{1,T} = \begin{pmatrix} T^{-1} & 0 \\ 0 & T^{-1/2} I_S \end{pmatrix}$$

and

$$D_{2,T} = \begin{pmatrix} T^{-1/2} I_{n(p_a-p-S)+\sum_{j=1}^S r_j} & 0 \\ 0 & T^{-1} I_{nS-\sum_{j=1}^S r_j} \end{pmatrix}$$

We have

$$D_{1,T} \tau \overline{\tau}' D_{1,T} \longrightarrow \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & I_S \end{pmatrix}$$

Relying on the three previous lemmata, we can state the following convergences

Lemma 15 *Under assumption 3, for the DGP under study,*

$$D_{2,T} P A \widetilde{\widetilde{z}} P_{\tau} \overline{\widetilde{\widetilde{z}}}' \overline{A}' P' D_{2,T} \Longrightarrow \begin{pmatrix} \Sigma_{00} & 0 \\ 0 & diag\left(\int_0^1 B_j^* \overline{B}_j^{*\prime} ds\right) \end{pmatrix}$$

where $diag\left(\int_0^1 B_j^* \overline{B}_j^{*\prime} ds\right)$ *is a* $\left(nS - \sum_{j=1}^S r_j\right) \times \left(nS - \sum_{j=1}^S r_j\right)$ *block diagonal matrix whose blocks are equal to* $\frac{1}{2}\int_0^1 B_j^*(s)\, \overline{B_j^*}(s)' ds$ *when* $\omega_j \notin \{0,\pi\}$ *and to* $\int_0^1 B_j^*(s)\, B_j^*(s)' ds$ *when* $\omega_j \in \{0,\pi\}$ *with* $B_j^*(s) = B_j(s) - \int_0^1 B_j$ *when* $j \neq 1$ *and* $B_1^*(s) = B_1(s) - (4-6s)\int B_1 - 6(2s-1)\int tB_1$

$$T^{-1/2} vec\left(\varepsilon P_{\tau} \Delta_x x'\right) \Longrightarrow \zeta_x$$

$$vec\left(\varepsilon P_{\tau} \overline{\widetilde{\widetilde{z}}}' \overline{A}' P' D_{2,T}\right) \Longrightarrow \begin{pmatrix} \zeta_0 \\ vec\left(\int dW_k \overline{B}_k^{*\prime}\right) \end{pmatrix}$$

where $\int dW_k \overline{B}_k^{*\prime}$ *is a* $nS \times \left(nS - \sum_{j=1}^S r_j\right)$ *block matrix whose the* k^{th} *block is a* $n \times (n - r_k)$ *matrix equal to* $\frac{1}{2}\int dW_k(s)\, \overline{B}_k^*(s)' ds$ *when* $\omega_j \notin \{0,\pi\}$ *and to* $\int dW_k(s)\, B_k^*(s)' ds$ *when* $\omega_j \in \{0,\pi\}$

$$\begin{aligned} T^{-1} \Delta_x x P_{\tau} \Delta_x x' &\longrightarrow \Sigma_{xx} \\ T^{-1} \Delta_x x P_{\widetilde{\widetilde{z}}} \Delta_x x' &\longrightarrow \Sigma_{xx} - \Sigma_{x0}\Sigma_{00}^{-1}\Sigma_{0x} \end{aligned}$$

$$T^{-1/2} \Delta_x x P_{\tau} \overline{\widetilde{\widetilde{z}}}' \overline{A}' P' D_{2,T} \longrightarrow \begin{pmatrix} \Sigma_{x0} & 0 \end{pmatrix}$$

We now can derive the asymptotic distribution of $\widehat{\psi}$:

$$\begin{aligned} \sqrt{T} vec\left(\widehat{\psi} - \psi\right) &= \sqrt{T} vec\left[\varepsilon P_{\widetilde{\widetilde{z}}} \Delta_x x' \left(\Delta_x x P_{\widetilde{\widetilde{z}}} \Delta_x x'\right)^{-1}\right] \\ &= \left(\frac{1}{T}\left(\Delta_x x P_{\widetilde{\widetilde{z}}} \Delta_x x'\right)^{-1} \otimes I_n\right) \frac{1}{\sqrt{T}} vec\left(\varepsilon P_{\widetilde{\widetilde{z}}} \Delta_x x'\right) \\ &\Longrightarrow \left(\left(\Sigma_{xx} - \Sigma_{x0}\Sigma_{00}^{-1}\Sigma_{0x}\right)^{-1} \otimes I_n\right) \begin{pmatrix} I_{n^2 p} & -\Sigma_{x0}\Sigma_{00}^{-1} \otimes I_n \end{pmatrix} \begin{pmatrix} \zeta_x \\ \zeta_0 \end{pmatrix} \\ &\Longrightarrow \mathcal{N}\left(0, \left(\Sigma_{xx} - \Sigma_{x0}\Sigma_{00}^{-1}\Sigma_{0x}\right)^{-1} \otimes \Omega\right) \end{aligned}$$

We use the standard argument of applying a Taylor expansion to $g(.)$ in the neighborhood of $vec\psi$ under the null $H_0' : g(vec\psi) = 0$. This gives

$$\sqrt{T} g\left(vec\widehat{\psi}\right) \Longrightarrow \mathcal{N}\left(0, \frac{\partial g}{\partial vec\psi'}\left[\left(\Sigma_{xx} - \Sigma_{x0}\Sigma_{00}^{-1}\Sigma_{0x}\right)^{-1} \otimes \Omega\right] \frac{\partial g}{\partial vec\psi}'\right)$$

From Lemma 15, we conclude that $\widehat{\psi}$ is consistent and so is $\widehat{\Sigma}_{\varepsilon}$. It follows that

$$\xi_W \Longrightarrow \chi_2(\dim g(.))$$

From Lemma 9, we can claim that testing H_0 with a Wald test statistics based on OLS estimates in levels VAR is $\chi_2(\dim f(.))$ as soon as $p_a \geq p + S$.

Remark 16 *This result can be extended to situations in which the variables are integrated of different orders at different frequencies. If we introduce for a given variable at least as many additional lags as the number of unit roots present in its individual data generating process, we preserve the fact that the Wald statistic is asymptotically chi-squared distributed.*

We can now turn to the finite sample properties of this test procedure.

6 Finite-sample properties

We propose to illustrate the size and power properties of this approach on a particular DGP. We first consider the size and power properties of a standard lag order test procedure and then turn to the size and power properties of Student and Fisher test statistics to test for linear constraints. The lag selection procedure we consider is based on the usual sequence of nested Fisher tests for the nullity of the matrices associated to the largest lags. In our approach, these tests are carried out in a model in which for each variable, additional lags equal to the number of unit roots present in their individual data generating process have been introduced. In a model of order p, we first test for null hypothesis that the p^{th} matrix is equal to 0, then when accepted, we test for the null hypothesis that the p^{th} and $(p-1)^{th}$ matrices are jointly equal to zero and so on.

We consider a particular case of the DGP introduced in the second section, we set $\alpha = 1:$

$$\begin{pmatrix} 1 & -L \\ L & 1 \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

The determinant of the matrix polynomial associated to this VAR process is equal to $1 + L^2$, whose roots are the two unit roots $\{i, -i\}$. Theorem 5 and Lemma 7 allow us to rewrite the DGP as follows

$$\begin{aligned} \begin{pmatrix} 1 & -L \\ L & 1 \end{pmatrix} &= \begin{pmatrix} 1 & -L \\ L & 1 \end{pmatrix} \begin{pmatrix} 1+L^2 & 0 \\ 0 & 1+L^2 \end{pmatrix} + L^2 \begin{pmatrix} -1 & L \\ L & -1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & -L \\ L & 1 \end{pmatrix} \begin{pmatrix} 1+L^2 & 0 \\ 0 & 1+L^2 \end{pmatrix} \\ &\quad + L^2 \left\{ \frac{1}{2} \begin{pmatrix} -1 & -i \\ i & -1 \end{pmatrix} (1+iL) + \frac{1}{2} \begin{pmatrix} -1 & i \\ -i & 1 \end{pmatrix} (1-iL) \right\} \end{aligned}$$

Since

$$\begin{pmatrix} -1 & -i \\ i & -1 \end{pmatrix} = \begin{pmatrix} -1 \\ i \end{pmatrix} \begin{pmatrix} 1 & i \end{pmatrix}$$

and

$$\begin{pmatrix} -1 & i \\ -i & -1 \end{pmatrix} = \begin{pmatrix} -1 \\ -i \end{pmatrix} \begin{pmatrix} 1 & -i \end{pmatrix}$$

this last matrix polynomial corresponds to the VECM representation of $\{y_t\}$ whose components are integrated of order 1 at frequencies $\frac{\pi}{2}$ and $-\frac{\pi}{2}$ and cointegrated at these frequencies:

$$\begin{pmatrix} 1 & -L \\ L & 1 \end{pmatrix} \begin{pmatrix} (1+L^2)\, y_{1t} \\ (1+L^2)\, y_{2t} \end{pmatrix} = \frac{1}{2} \left\{ \begin{pmatrix} 1 \\ -i \end{pmatrix} \begin{pmatrix} 1 & i \end{pmatrix} \begin{pmatrix} (1+iL)\, y_{1,t-2} \\ (1+iL)\, y_{2,t-2} \end{pmatrix} \right.$$

$$+\left(\begin{array}{c}1\\ i\end{array}\right)\left(\begin{array}{cc}1 & -i\end{array}\right)\left(\begin{array}{c}(1-iL)\,y_{1,t-2}\\ (1-iL)\,y_{2,t-2}\end{array}\right)\Bigg\}+\varepsilon_t$$

In short, the unconstrained VAR representation of this DGP involves one lag. Its VECM representation shows that the associated processes are integrated of order 1 at frequency $\frac{\pi}{2}$ and $-\frac{\pi}{2}$ and cointegrated. There are two unit roots in their individual DGP and two additional lags must be introduced in the model to test for the lag length of the VAR representation.

Starting from a lag order equal at most to 8, we run a Monte Carlo exercise to measure the size of each Fisher tests involved in the sequential testing (i.e. sequence of null hypotheses: $H_{0,k}\,:\,\forall j\in\{k,k+1,...,8\}\,,\ \Phi_j=0,\ ;\ H_{a,k}\,:\,\exists\ j\in\{k,k+1,...,p\}\,,\ \Phi_j\neq 0$. We consider Gaussian innovations with a variance-covariance matrice equal to

$$\left(\begin{array}{cc}1 & 0.5\\ 0.5 & 1\end{array}\right)$$

and various specifications without or with deterministic terms (constant, sine and cosine functions at frequency $\frac{\pi}{2}$, both). Results are presented in Table 1. When the sample size is small (T=50), the size of each test is significantly larger than the nominal level (5%) and this distortion is larger when deterministic terms are introduced in the specification. When the sample size gets larger, the empirical size gets closer to the nominal one. The empirical size is close to the nominal one when T is larger than 250. We can also in this exercise produce the empirical distribution of the lag lengths obtained in this iterative procedure. They are given in Table 2. When the sample size is small, we observe that we obtain the correct lag order in at most 75% of the simulations when there is no deterministic term in the regression. Due to the dependence of the tests, the decisions we take at various steps of the procedure lead to an underrepresentation of the correct lag length. When the sample size is larger, the correct lag order is more frequently selected (about 87% when T=500).

We then propose to test for the following set of null hypotheses on this DGP for various sample sizes and values of the variance-covariance matrix of the error terms:

$$\begin{aligned}H_{01} &: \phi_{12}=1\\ H_{02} &: \phi_{21}=-1\\ H_{03} &: \phi_{12}=1 \text{ and } \phi_{21}=-1\end{aligned}$$

We first simulate the size of these tests in the same situations as those considered above and then move to the power properties. The empirical sizes are given in Table 3. They have been computed without imposing the lag length but with the lag length selected with the above iterative procedure. We observe that the empirical size is larger than the nominal one when T is small and in this case, it is more over-sized in presence of deterministic terms in the regressions. When T gets larger, the empirical size gets closer to the nominal one. It can be over or under the nominal one.

We then turn to the analysis of the power properties when the lag length has been selected with the above iterative procedure. Nevertheless to keep to a finite order VAR representation, we keep to processes that are cointegrated at frequencies $\frac{\pi}{2}$ and $-\frac{\pi}{2}$ but change the value of the cointegrating vector. To do so, we consider the DGPs introduced in the introductory example

Table 1: Lag selection procedure ($p_{max} = 8$, 5000 simulations)

	$H_{0,2}$	$H_{0,3}$	$H_{0,4}$	$H_{0,5}$	$H_{0,6}$	$H_{0,7}$	$H_{0,8}$
T=500							
no deterministic term	0.0478	0.0476	0.0524	0.0520	0.0518	0.0516	0.0546
constant term	0.0488	0.0476	0.0532	0.0510	0.0530	0.0512	0.0532
sine and cosine functions	0.0502	0.0494	0.0542	0.0528	0.0534	0.0520	0.0552
all terms	0.0512	0.0510	0.0540	0.0534	0.0540	0.0522	0.0538
T=250							
no deterministic term	0.0592	0.0558	0.0498	0.0498	0.0510	0.0498	0.0512
constant term	0.0620	0.0588	0.0546	0.0538	0.0528	0.0540	0.0504
sine and cosine functions	0.0612	0.0578	0.055	0.0548	0.0540	0.0516	0.0532
all terms	0.0614	0.0606	0.0596	0.0568	0.0558	0.0538	0.0578
T=100							
no deterministic term	0.0754	0.0710	0.0688	0.0660	0.0606	0.0586	0.0616
constant term	0.0786	0.0746	0.0714	0.0686	0.0634	0.0616	0.0642
sine and cosine functions	0.0906	0.0876	0.0822	0.0762	0.0670	0.0640	0.0690
all terms	0.0876	0.0852	0.0822	0.0770	0.0666	0.0668	0.0770
T=50							
no deterministic term	0.1526	0.1412	0.1324	0.1210	0.1010	0.0892	0.0830
constant term	0.1634	0.1502	0.1452	0.1320	0.1140	0.1006	0.0890
sine and cosine functions	0.2246	0.2088	0.1886	0.1708	0.1432	0.1220	0.1034
all terms	0.2190	0.2024	0.1880	0.1720	0.1352	0.1254	0.1238

Table 2: Empirical distribution of the selected lag orders.

Proportion of lag order selected	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=6$	$p=7$	$p=8$
T=500								
no deterministic term	0.8684	0.0054	0.0066	0.0086	0.0138	0.0170	0.0256	0.0546
constant term	0.8714	0.0050	0.0064	0.0090	0.0130	0.0162	0.0258	0.0532
sine and cosine functions	0.8658	0.0058	0.0068	0.0092	0.0140	0.0172	0.0260	0.0552
all terms	0.8694	0.0050	0.0068	0.0088	0.0132	0.0174	0.0256	0.0538
T=250								
no deterministic term	0.8532	0.0104	0.0138	0.0128	0.0132	0.0186	0.0268	0.0512
constant term	0.8492	0.0096	0.0142	0.0144	0.0136	0.0186	0.0300	0.0504
sine and cosine functions	0.8466	0.0110	0.0134	0.0138	0.0146	0.0200	0.0274	0.0532
all terms	0.8422	0.0076	0.0140	0.0158	0.0156	0.0202	0.0268	0.0578
T=100								
no deterministic term	0.8344	0.0124	0.0116	0.0156	0.0194	0.0208	0.0242	0.0616
constant term	0.8296	0.0118	0.0130	0.0160	0.0190	0.0208	0.0256	0.0642
sine and cosine functions	0.8144	0.0154	0.0134	0.0176	0.0222	0.0220	0.0260	0.0690
all terms	0.8124	0.0124	0.0120	0.0172	0.0232	0.0208	0.0250	0.0770
T=50								
no deterministic term	0.7592	0.0148	0.0162	0.0252	0.0294	0.0302	0.0420	0.0830
constant term	0.7436	0.0154	0.0170	0.0266	0.0296	0.0338	0.0450	0.0890
sine and cosine functions	0.6820	0.0214	0.0234	0.0296	0.0390	0.0436	0.0576	0.1034
all terms	0.6826	0.0194	0.0222	0.0296	0.0396	0.0350	0.0478	0.1238

Table 3: Empirical Size.

	t_{α}	$t_{1/\alpha}$	F
T=500			
no deterministic term	0.0518	0.0550	0.0500
constant term	0.0528	0.0542	0.0500
sine and cosine functions	0.0514	0.0544	0.0496
all terms	0.0514	0.0532	0.0506
T=250			
no deterministic term	0.0534	0.0586	0.0584
constant term	0.0532	0.0584	0.0558
sine and cosine functions	0.0518	0.0578	0.0572
all terms	0.0562	0.0614	0.0576
T=100			
no deterministic term	0.0632	0.0690	0.0686
constant term	0.0640	0.0682	0.0686
sine and cosine functions	0.0594	0.0662	0.0662
all terms	0.0674	0.0760	0.0786
T=50			
no deterministic term	0.0846	0.0940	0.1026
constant term	0.0878	0.0940	0.1104
sine and cosine functions	0.0790	0.0860	0.0948
all terms	0.0992	0.1044	0.1230

in Section 2 associated to various values of α and test for the capacity to reject the null that $\alpha = 1$. The results are given in Figure 1. The results are not size-adjusted. We observe that the power properties are relatively satisfactory, increasing with the size sample and somewhat asymmetric.

7 Conclusion

We extend the lag-augmented approach introduced by Toda and Yamamoto (1995), Dolado and Lütkepohl (1996) and Yamamoto (1996) to the situation in which seasonal unit roots are present in the data generating process. This allows the econometrician to test in a VAR framework for linear constraints such as for instance Granger-Causality relationships with non-seasonally adjusted data. This is of practical interest as we know that separate seasonal adjustment may introduce some distortion in the relationships between the variables the economist wants simultaneously study. The rule we obtain is that if we introduce for a given variable at least as many additional lags as the number of unit roots present in its individual data generating process, we preserve the fact that the Wald statistic is asymptotically chi-squared distributed. The simulation exercise illustrates that as soon as the sample size is large enough (more than 250), size and power properties are satisfactory.

8 References

ANDERSON T.W. [1971], *The Statistical Analysis of Time Series*, J. Wiley.

CANOVA F. and **B.E. HANSEN** [1995], 'Are Seasonal Patterns Constant over time? A Test for Seasonal Stability', *Journal of Business and Economic Statistics*, 13, 237-252.

CHAN N.H. and **C.Z. WEI** [1988], 'Limiting Distribution of Least Squares Estimates of Unstable Auto regressive Processes', *The Annals of Statistics*, 16 (1), 367-401.

CUBADDA G. [2001], 'Complex Reduced Rank models for seasonally cointegrated time series', *Oxford Bulletin of Economics and Statistics,* 63, 497-511.

DICKEY D.A. and **W. FULLER** [1979], 'Distribution of the estimators for autoregressive time series with a unit root', *Journal of the American Statistical Association*, 74, 427-431

DICKEY D.A., **D.P. HASZA** and **W. FULLER** [1984], 'Testing for Unit Roots in Seasonal Time Series', *Journal of the American Statistical Association*, 79, 355-367.

DOLADO J.J and **H. LÜTKEPOHL** [1996], 'Making Wald tests work for cointegrated VAR systems', *Econometric Reviews*, 15, 369-386

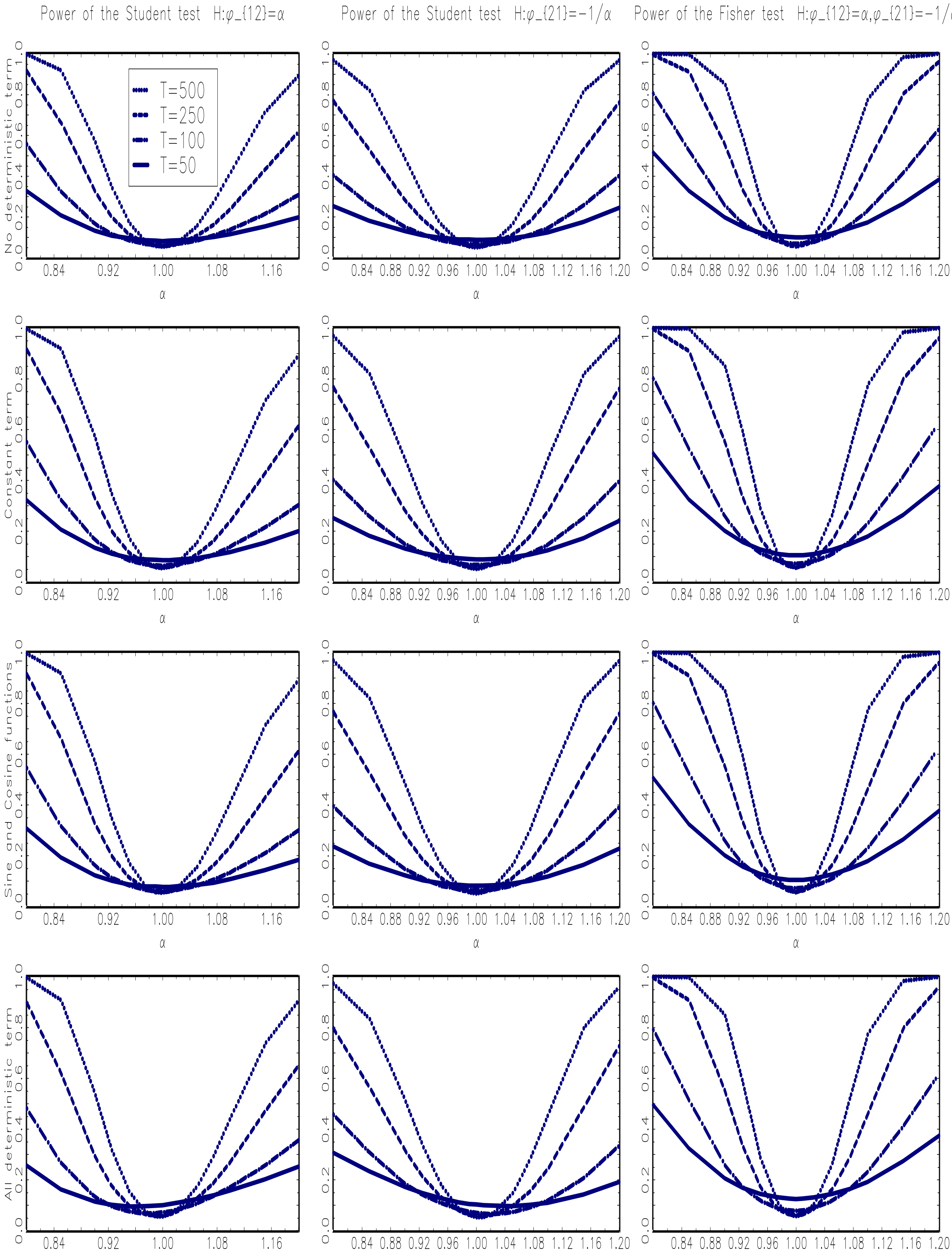
ENGLE R.F , C.W.J. GRANGER and **J.HALLMAN** [1989], 'Merging short- and long-run forecasts: An application of seasonal co-integration to monthly electricity sales forecasting', *Journal of Econometrics,* 40, 45-62.

ENGLE R.F , C.W.J. GRANGER, S.HYLLEBERG and **H.S.LEE** [1993], 'Seasonal cointegration: The Japanese consumption function', *Journal of Econometrics,* 55, 275-298.

GEWEKE, J. [1979], 'The Temporal and Sectoral Aggregation of Seasonally Adjusted Time Series', in *Seasonal Analysis of Economic Time Series*, Ed. A. Zellner, Washington D.C., U.S. Bureau of Census, 411-427

GOODMAN N.R. [1963], 'Statistical analysis based on a multivariate complex Gaussian distribution (an introduction)', *Annals of Statistics*, 34, 152-177.

Figure 1: Non adjusted power curves of t and F tests ($\sigma_1 = \sigma_2 = 1, \rho = 0.5$)



GREGOIR S. [1999a], 'Multivariate Time Series with Various Hidden Unit Roots: Part I : Integral Operator Algebra and Representation Theory', *Econometric Theory*,15, 435-468.

GREGOIR S. [1999b], 'Multivariate Time Series with Various Hidden Unit Roots: Part II : Estimation and Testing', *Econometric Theory*,15, 469-518.

GREGOIR S. [2006], 'Efficient tests for the presence of a pair of complex conjugate unit roots in real time series', *Journal of Econometrics, 130* (1) , 45-100.

GREGOIR S. [2010], 'Fully Modified Estimation of Seasonally cointegrated processes', *Econometric Theory*, forthcoming.

HANSEN B.E. [1992], 'Convergence to Stochastic Integrals for Dependent Heterogeneous Processes', *Econometric Theory*, 8,(4),489-500.

HARRIS D. [1997], 'Principal components analysis of cointegrated time series', *Econometric Theory,* 13, 529-557

HYLLEBERG S., **R.F.ENGLE**, **C.W.J.GRANGER** and **B.S.YOO** [1991], 'Seasonal Integration and Cointegration', *Journal of Econometrics*, 44, 215-238.

JOHANSEN S. [1988], 'Statistical analysis of cointegration vectors', *Journal of Economics Dynamics and Control*, 12, 231-254

JOHANSEN S. [1991], 'Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models', *Econometrica*, 59, 1551-1580

JOHANSEN S. and **E.SCHAUMBURG** [1999] 'Likelihood Analysis of Seasonal Cointegration', *Journal of Econometrics,* 88(2), 301-339

KUROZUMI E. and **T.YAMAMOTO** [2000], 'Modified lag augmented vector autoregressions', *Econometric Reviews*, 19, 207-231

LEE J.H. [1992], 'Maximum likelihood inference on cointegration and seasonal cointegration', *Journal of Econometrics*, 54, 1-48

PHILLIPS P.C.B. and **P. PERRON** [1988], 'Testing for a unit root in time series regression', *Biometrika*, 75 (2), 971-1001.

PORTER R.D. [1975] 'Multiple time series containing unobserved components', Special Studies Paper, n°65, Federal Reserve Board

REIMERS H-E [1992] 'Comparisons of tests for multivariate cointegration', *Statistical Papers*, 33, 335-346

RODRIGUES P.M.M. and **A.M.R. TAYLOR** [2007], 'Efficient test of the seasonal unit root hypothesis', *Journal of Econometrics*

SIMS, C.A. [1974], 'Seasonality in Regression', *Journal of the American Statistical Association*, 69, 618-626.

SIMS, C.A. , J.H. STOCK and **M.W.WATSON** [1990], 'Inference in linear time series with some unit roots', *Econometrica*, 58, 113-144

STOCK J.H and **M.W. WATSON** [1988], 'Testing for common trends', *Journal of the American Statistical Association*, 83, 1097-1107.

TODA H.Y. [1995] 'Finite sample performance of likelihood ratio tests for cointegrating ranks in vector autoregressions', *Econometric Theory*, 11, 1015-1032

TODA H.Y. and **P.C.B. PHILLIPS** [1993], 'Vector autoregressions and causality', *Econometrica*, 61, 1367-1393

TODA H.Y. and **T. YAMAMOTO** [1995], 'Statistical infrence in vector autoregressions with possibly integrated processes', *Journal of Econometrics*, 66, 225-250

TSAY R.S. and **G.C. TIAO** [1990],'Asymptotic properties of multivariate nonstationary processes with applications to auto regressions', *Annals of Statistics*, 18, 220-250.

WALLIS, K.F. [1974], 'Seasonal Adjustment and the Relation between Variables', *Journal of the American Statistical Association*, 69, 18-32.

WALLIS, K.F. [1976], 'Seasonal Adjustment and Multiple Time Series Analysis',in *Seasonal Analysis of Economic Time Series*, Ed. A.Zellner, Washington D.C., U.S. Department of Commerce, Bureau of Census, 366-397

YAMADA Y. and **H.Y.TODA** [1998], 'Inference in possibly integrated vector autoregressive models: some finite sample evidence', *Journal of Econometrics,* 86, 55-95

YAMAMOTO T. [1996], 'A simple approach to the statistical inference in linear time series models which may have some unit roots', *Hitotsubashi Journal of Economics*, 37, 87-100

9 Appendix

Proof of Theorem 5: The proof is by induction and relies on a polynomial division by ascending order. We denote $\Delta_{x,j}$ the coefficients of the polynomial $\Delta_x(L)$

$$\Delta_x(L) = \sum_{j=0}^{d_x} \Delta_{x,j} L^j$$

where $\Delta_{x,0} = 1$. By convention, when the index j is larger than d_x, $\Delta_{x,j} = 0$. We set $\phi^{(1)}(L) = \phi(L)$ and $R^{(1)}(L) = 0$. The first step consists of computing the first order polynomial remainder defined by

$$\begin{aligned}
\phi(L) - \phi_1 \Delta(L) L &= \sum_{j=1}^{p} \phi_j L^j - \phi_1 \sum_{l=0}^{d_x} \Delta_{x,l} L^{l+1} \\
&= \sum_{j=1}^{p} (\phi_j - \phi_1 \Delta_{x,j-1}) L^j - \sum_{l=p+1}^{d_x+1} \phi_1 \Delta_{x,l-1} L^l \\
&= \sum_{j=2}^{p} (\phi_j - \phi_1 \Delta_{x,j-1}) L^j - L^{p+1} \left(\sum_{l=0}^{d_x-p} \phi_1 \Delta_{x,p+l} L^l \right)
\end{aligned}$$

where the second term in the right-hand side of the last equation is equal to 0 if $d_x \leq p$. Let us denote

$$\begin{aligned}
\phi^{(2)}(L) &= \sum_{j=2}^{p} (\phi_j - \phi_1 \Delta_{x,j-1}) L^j \\
&= \sum_{j=2}^{p} \phi_j^{(2)} L^j
\end{aligned}$$

a polynomial matrix whose minimal degree is equal to 2 and maximal one to p, and

$$R^{(2)}(L) = -\left(\sum_{l=0}^{d_x-p} \phi_1 \Delta_{x,p+l} L^l \right)$$

a polynomial matrix of degree at most $d_x - p - 1$, we have

$$\begin{aligned}
\phi(L) &= \phi^{(1)}(L) + L^{p+1} R^{(1)}(L) \\
&= \phi_1^{(1)} \Delta(L) L + \phi^{(2)}(L) + L^{p+1} R^{(2)}(L)
\end{aligned}$$

and

$$\begin{pmatrix} \phi_1^{(1)} & \phi_2^{(2)} & \ldots & \phi_p^{(2)} \end{pmatrix} = \begin{pmatrix} \phi_1^{(1)} & \phi_2^{(1)} & \ldots & \phi_p^{(1)} \end{pmatrix} N^{(1)} \otimes I_n$$

where

$$\underset{(p\times p)}{N^{(1)}} = \begin{pmatrix} 1 & -\Delta_{x,1} & & -\Delta_{x,p-1} \\ 0 & 1 & 0 & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

We now define (i) a sequence of polynomial matrices $\left\{\phi^{(j)}(L)\right\}_{j=2,\ldots,p}$ whose matrix coefficients are given by the following recurrence equation

$$\begin{aligned}
\phi_k^{(j)} &= 0 \text{ for } 1 \le k \le j-1 \\
\phi_k^{(j)} &= \phi_k^{(j-1)} - \phi_{j-1}^{(j-1)} \Delta_{x,k-j+1} \text{ for } j \le k \le p
\end{aligned}$$

or similarly

$$\begin{pmatrix} \phi_{j-1}^{(j-1)} & \phi_j^{(j)} & \ldots & \phi_p^{(j)} \end{pmatrix} = \begin{pmatrix} \phi_{j-1}^{(j-1)} & \phi_j^{(j-1)} & \ldots & \phi_p^{(j-1)} \end{pmatrix} N^{(j-1)} \otimes I_n$$

where

$$\underset{(p-j+2\times p-j+2)}{N^{(j-1)}} = \begin{pmatrix} 1 & -\Delta_{x,1} & & -\Delta_{x,p-j+1} \\ 0 & 1 & 0 & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

(ii) a sequence of polynomial matrices $\left\{R^{(j)}(L)\right\}_{j=2,\ldots,p}$ given by the following recurrence equation

$$R^{(j)}(L) = R^{(j-1)}(L) - \sum_{k=0}^{d_x+j-p-2} \phi_{j-1}^{(j-1)} \Delta_{x,p-j+k} L^k$$

where the degree of the second term of the right-hand side of the last equation is at most $d_x + j - p - 2$, which is increasing with j.

We claim that for $j = 1, \ldots, p$

$$\phi(L) = \sum_{k=1}^{j-1} \phi_k^{(k)} \Delta(L) L^k + \phi^{(j)}(L) + L^{p+1} R^{(j)}(L)$$

This holds for $j=1$ and $j=2$. Let assume that it holds up to order $j<p-1$. We then have

$$\begin{aligned}
\phi^{(j)}(L)-\phi_j^{(j)}\Delta_x(L)L^j &= \sum_{k=j}^{p}\phi_k^{(j)}L^k-\phi_j^{(j)}\sum_{l=0}^{d_x}\Delta_{x,l}L^{l+j} \\
&= \sum_{k=j}^{p}\left(\phi_k^{(j)}-\phi_j^{(j)}\Delta_{x,k-j}\right)L^k-\sum_{l=p+1}^{d_x+j}\phi_j^{(j)}\Delta_{x,l-j}L^l \\
&= \sum_{k=j}^{p}\phi_k^{(j+1)}L^j-L^{p+1}\left(\sum_{l=0}^{d_x+j-p-1}\phi_j^{(j)}\Delta_{x,p+l+1-j}L^l\right) \\
&= \phi^{(j+1)}(L)+L^{p+1}\left(-\sum_{l=0}^{d_x+j-p-1}\phi_j^{(j)}\Delta_{x,p+l+1-j}L^l\right)
\end{aligned}$$

therefore

$$\begin{aligned}
\phi(L) &= \sum_{k=1}^{j}\phi_k^{(k)}\Delta(L)L^k+\phi^{(j+1)}(L) \\
&\quad +L^{p+1}\left(R^{(j)}(L)-\sum_{l=0}^{d_x+j-p-1}\phi_j^{(j)}\Delta_{x,p+l+1-j}L^l\right) \\
&= \sum_{k=1}^{j}\phi_k^{(k)}\Delta(L)L^k+\phi^{(j+1)}(L)+L^{p+1}R^{(j+1)}(L)
\end{aligned}$$

When $j=p$, since $\phi^{(p+1)}=0$, we get

$$\phi(L)=\sum_{k=1}^{p}\phi_k^{(k)}\Delta(L)L^k+L^{p+1}\left(R^{(p)}(L)-\sum_{l=0}^{d_x-1}\phi_p^{(p)}\Delta_{x,l+1}L^l\right)$$

We set $\psi(L)=\sum_{k=1}^{p}\phi_k^{(k)}L^k$ and $R(L)=R^{(p)}(L)-\sum_{l=0}^{d_x-1}\phi_p^{(p)}\Delta_{x,l+1}L^l$ whose degree is d_x-1. If we denote $M^{(1)}=N^{(1)}$ and for $j>2$

$$M^{(j)}=\begin{pmatrix} I_{j-1} & 0 \\ 0 & N^{(j-1)} \end{pmatrix}$$

we set $M=M^{(1)}...M^{(p-1)}$, so that

$$\begin{pmatrix} \phi_1^{(1)} & \phi_2^{(2)} & \dots & \phi_p^{(p)} \end{pmatrix}=\begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_p \end{pmatrix}M\otimes I_n$$

where M is a full rank matrix as a upper triangular matrix with diagonal terms equal to 1.
Q.E.D. ■

Proof of Lemma 7: The proof amounts to show that the set of d_x polynomials

$$\left(\cup_{j=1}^{S}\left\{\Delta_{x,-j},\delta_{\omega_j}\Delta_{x,-j},\delta_{\omega_j}^2\Delta_{x,-j},\dots\delta_{\omega_j}^{d_j-1}\Delta_{x,-j}\right\}\right)$$

is a basis of the polynomials of degree less or equal to $d_x - 1$. Indeed, if this is true, any polynomial $q(L)$ of degree less or equal to $d_x - 1$ can be decomposed in these basis elements, *i.e.* there exists a set of numbers $\left\{(q_{jk})_{k=0,\ldots d_j-1}\right\}_{j=1,\ldots S}$ such that

$$\begin{aligned} q(L) &= \sum_{j=1}^{S}\left(\sum_{k=0}^{d_j-1} q_{jk}\delta_{\omega_j}^{k}\Delta_{x,-j}\right) \\ &= \sum_{j=1}^{S}\left(\sum_{k=0}^{d_j-1} q_{jk}\delta_{\omega_j}^{k}\right)\Delta_{x,-j} \end{aligned}$$

This is true for each coefficient of a polynomial matrix, so it holds for the polynomial matrix itself. To show that the above set of polynomials is a basis we establish it is a set of linearly independent elements that generates polynomials of degree less or equal to $d_x - 1$. We start with the independence property. Let us consider a set of numbers $\left\{(q_{jk})_{k=0,\ldots d_j-1}\right\}_{j=1,\ldots S}$ such that $q(L) = 0$. We partition the set $\{1, ..., S\}$ into subsets of index related to frequencies that have the same order of integration:

$$\{1, ..., S\} = \bigcup_{k=1}^{\max_j d_j} J_k$$

with

$$J_k = \{j \in \{1, ..., S\} \,|d_j = k\}$$

and

$$q(L) = \sum_{k=1}^{\max_j d_j} \sum_{j\in J_k}\left(\sum_{l=0}^{k-1} q_{jl}\delta_{\omega_j}^{l}\right)\Delta_{x,-j}$$

We denote $K_j = \cup_{k=1}^{j} J_k$. We proceed by induction on k. First, we compute the value of $q(L)$ at each unit root and get $\forall j \in \{1, \ldots, S\}$

$$\begin{aligned} q\left(e^{i\omega_j}\right) &= q_{j0}\Delta_{x,-j}\left(e^{i\omega_j}\right) = 0 \\ &\implies q_{j0} = 0 \end{aligned}$$

since $\forall k \neq j$, $\Delta_{x,-k}\left(e^{i\omega_j}\right) = 0$, then

$$q(L) = \sum_{k=2}^{\max_j d_j} \sum_{j\in J_k}\left(\sum_{l=1}^{k-1} q_{jl}\delta_{\omega_j}^{l-1}\right)\delta_{\omega_j}\Delta_{x,-j}$$

If we compute the derivative of this polynomial, we get

$$q'(L) = \sum_{k=2}^{\max_j d_j} \sum_{j\in J_k}\left(\sum_{l=2}^{k-1} (l-1)\, q_{jl}\delta_{\omega_j}^{l-2}\right)\delta_{\omega_j}\Delta_{x,-j} + \sum_{k=2}^{\max_j d_j} \sum_{j\in J_k}\left(\sum_{l=1}^{k-1} q_{jl}\delta_{\omega_j}^{l-1}\right)\left[\delta_{\omega_j}\Delta'_{x,-j} - e^{i\omega_j}\Delta_{x,-j}\right]$$

which is equal to zero. We compute the value of $q'(L)$ at each unit root whose index is in $J\backslash K_2$ we get that $\forall j \in J\backslash K_2$

$$\begin{aligned} q'\left(e^{i\omega_j}\right) &= -q_{j1}e^{i\omega_j}\Delta_{x,-j}\left(e^{i\omega_j}\right) = 0 \\ &\Longrightarrow q_{j1} = 0 \end{aligned}$$

and then

$$\begin{aligned} q(L) &= \sum_{k=3}^{\max_j d_j}\sum_{j\in J_k}\left(\sum_{l=2}^{k-1} q_{jl}\delta_{\omega_j}^{l-2}\right)\delta_{\omega_j}^2\Delta_{x,-j} \\ &\quad + \sum_{j\in J_2} q_{j1}\delta_{\omega_j}\Delta_{x,-j} \end{aligned}$$

This result holds because the value of the derivative of $\Delta_{x,-k}$ at $e^{i\omega_j}$ when j is in $J\backslash K_2$ is zero due to fact that the monomial associated to this unit root is raised at a power strictly larger than 1 in $\Delta_{x,-k}$. Let us assume that when we repeat this kind of operations h times, we get the following form for $q(L)$:

$$\begin{aligned} q(L) &= \sum_{k=h+1}^{\max_j d_j}\sum_{j\in J_k}\left(\sum_{l=h}^{k-1} q_{jl}\delta_{\omega_j}^{l-h}\right)\delta_{\omega_j}^h\Delta_{x,-j} \\ &\quad + \sum_{k=2}^{h}\sum_{j\in J_k} q_{j,k-1}\delta_{\omega_j}^{k-1}\Delta_{x,-j} \end{aligned} \tag{18}$$

We compute the derivative of order h. We take in turn each polynomial in this sum and start by those in the second term. We use the property that when $j \in J\backslash K_h$, the value of the derivative of order h of $\Delta_{x,-k}$ at $e^{i\omega_j}$ for any $k \neq j$ is zero due to fact that the monomial associated to this unit root is raised at a power strictly larger than h in $\Delta_{x,-k}$. By Leibnitz rule, we get that for $j \in K_h$,

$$\frac{d^h}{dL^h}\left(q_{j,k-1}\delta_{\omega_j}^{k-1}\Delta_{x,-j}\right) = q_{j,k-1}\sum_{l=0}^{h}\binom{h}{l}\frac{d^l}{dL^l}\delta_{\omega_j}^{k-1}\frac{d^{h-l}}{dL^{h-l}}\Delta_{x,-j}$$

and conclude that its value at each unit root whose index is in $J\backslash K_h$ is 0 by the above property. For $j \in J\backslash K_h$, we get

$$\frac{d^h}{dL^h}\left[\left(\sum_{g=h}^{d_j-1} q_{jg}\delta_{\omega_j}^{g-h}\right)\delta_{\omega_j}^h\Delta_{x,-j}\right] = \sum_{l=0}^{h}\binom{h}{l}\frac{d^l}{dL^l}\left\{\left(\sum_{g=h}^{d_j-1} q_{jg}\delta_{\omega_j}^{g-h}\right)\delta_{\omega_j}^h\right\}\frac{d^{h-l}}{dL^{h-l}}\Delta_{x,-j}$$

When we compute the value of the term at a unit root whose index $j' \in J\backslash K_h$, by the above property, the only non zero term is the one when $l = h$ and $j' = j$. This implies that for $j \in J\backslash K_h$

$$\begin{aligned} \frac{d^h q\left(e^{i\omega_j}\right)}{dL^h} &= h!\,(-1)^h e^{ih\omega_j}\Delta_{x,-j}\left(e^{i\omega_j}\right) = 0 \\ &\Longrightarrow q_{jh} = 0 \end{aligned}$$

This ensures that (18) holds at the order $h+1$. By induction, we conclude that

$$q(L)=\sum_{k=2}^{\max_j d_j}\sum_{j\in J_k} q_{j,k-1}\delta_{\omega_j}^{k-1}\Delta_{x,-j}$$

or

$$q(L)=\left(\sum_{k=1}^{\max_j d_j}\sum_{j\in J_k} q_{j,k-1}\prod_{l\neq j}\delta_{\omega_l}\right)\frac{\Delta_x}{\prod_{j=1}^{S}\delta_{\omega_j}}=0$$

In this product, the second polynomial is different from 0, the first one is therefore exactly equal to 0. If we compute its value in each unit root ω_j we conclude that $q_{j,d_j-1}=0$. This family is linearly independent. The dimension of the vector space of polynomials of degree less that d_x-1 is exactly equal to d_x . The above family is composed of d_x linearly independent elements, it is therefore a basis of the space of polynomials of degree less that d_x-1.

Q.E.D. ■

Proof of Lemma 9: We start from the equivalent representations of the generating equation with additional lags written with matrix notations in equations (4.3,13)

$$\begin{aligned}
y &= \widetilde{\beta}\tau+\phi y_{-1}+\Phi y_{-p_a}+\varepsilon\\
y &= \widetilde{\beta}\tau+\psi\Delta_x y+\xi\Delta_x y_{-p}+Rz_{-p}+\varepsilon\\
y &= \widetilde{\beta}\tau+\psi\Delta_x y+\widetilde{R}\widetilde{z}_{-p}+\varepsilon
\end{aligned}$$

where according Theorem 5, Lemma 7 and the attached comment we know there exist two matrices M and K of respective dimensions $(p_a-d_x-1\times p_a-d_x-1)$ and $(p\times d_x-1)$ such that

$$\begin{aligned}
\begin{pmatrix}\phi & \Phi\end{pmatrix}(M\otimes I_n) &= \begin{pmatrix}\psi & \xi\end{pmatrix}\\
\begin{pmatrix}\phi & \Phi\end{pmatrix}(K\otimes I_n) &= R
\end{aligned}$$

and M is upper triangular so that there exists with obvious notations a $(p\times p)$ matrix M_{11} such that

$$\phi(M_{11}\otimes I_n)=\psi$$

or

$$vec\psi=\left[(M_{11}'\otimes I_n)\otimes I_n\right]vec\phi$$

The test statistics have the following form

$$\zeta_W=\overline{f\left(vec\widehat{\phi}\right)}'\left[\frac{\partial f}{\partial vec\phi'}\left(\left(y_{-1}P_{y_{-p_a}}\overline{y}_{-1}'\right)^{-1}\otimes\widehat{\Omega}_\varepsilon\right)\overline{\frac{\partial f'}{\partial vec\phi}}\right]^{-1}f\left(vec\widehat{\phi}\right)$$

and

$$\zeta_W'=\overline{g\left(vec\widehat{\psi}\right)}'\left[\frac{\partial g}{\partial vec\psi'}\left(\left(\Delta_x yP_{\widetilde{z}_{-p}}\overline{\Delta_x y}'\right)^{-1}\otimes\widehat{\Omega}_\varepsilon\right)\overline{\frac{\partial g'}{\partial vec\psi}}\right]^{-1}g\left(vec\widehat{\psi}\right).$$

On the one hand,

$$\frac{\partial g}{\partial vec\psi'}=\frac{\partial f}{\partial vec\phi'}\left[\left(M_{11}^{-1}\otimes I_n\right)\otimes I_n\right]$$

on the other hand, with obvious notations

$$
\begin{aligned}
\phi y_{-1}+\Phi y_{-p_a} &= \psi\Delta_x y+\xi\Delta_x y_{-p}+Rz_{-p}\\
&= \begin{pmatrix}\phi & \Phi\end{pmatrix}(M\otimes I_n)\begin{pmatrix}\Delta_x y\\ \Delta_x y_{-p}\end{pmatrix}+\begin{pmatrix}\phi & \Phi\end{pmatrix}(K\otimes I_n)\,z_{-p}\\
&= \phi\,(M_{11}\otimes I_n)\,\Delta_x y+\begin{pmatrix}(\phi\,(M_{12}\otimes I_n)+\Phi\,(M_{22}\otimes I_n)) & \begin{pmatrix}\phi & \Phi\end{pmatrix}(K\otimes I_n)\end{pmatrix}\begin{pmatrix}\Delta_x y_{-p}\\ z_{-p}\end{pmatrix}\\
&= \phi\,(M_{11}\otimes I_n)\,\Delta_x y+\begin{pmatrix}(\phi\,(M_{12}\otimes I_n)+\Phi\,(M_{22}\otimes I_n)) & \begin{pmatrix}\phi & \Phi\end{pmatrix}(K\otimes I_n)\end{pmatrix}\widetilde{z}_{-p}
\end{aligned}
$$

and since each component at date t of $\widetilde{z}_{-p}$ is a linear combination of $y_{t-p-1},..\ y_{t-p_a}$, there exists a $(d_x-1+p_a-p)\times(p_a-p)$ matrix H such that

$$Hy_{-p_a}=\widetilde{z}_{-p},$$

it follows that

$$\phi y_{-1}P_{y_{-p_a}}=\phi\,(M_{11}\otimes I_n)\,\Delta_x yP_{\widetilde{z}_{-p}}$$

This holds for any value of ϕ where $y_{-1}P_{y_{-p_a}}\overline{y}'_{-1}$ and $\Delta_x yP_{\widetilde{z}_{-p}}\overline{\Delta_x y}'$ are symmetric definite positive matrices whence the result.

Q.E.D. ■

Proof of Lemma 15: It is direct application of the joint weak convergences summarized in the Lemmata of section 5 and the algebraic results. ■