LACEA 2001
MONTEVIDEO URUGUAY

Fragiskos Archontakis

Testing the Order of Integration in a VAR Model for I(2) Variables

Testing the order of integration in a VAR model for I(2) variables*

Fragiskos Archontakis†
European University Institute

May 2001

Abstract

We propose a test for the order of integration of the univariate components of a vector process integrated of order two, i.e. an I(2) process, given by an autoregressive model. The null hypothesis of the test is that the particular univariate time series is an I(1) process. The hypotheses are formulated as linear restrictions on the directions orthogonal to the I(1) cointegration space. The statistic considered is the Wald test, which asymptotically follows a chi-squared distribution, such that standard inference can be applied. The theoretical results are illustrated by a Monte Carlo experiment.

1 Introduction

Cointegration theory has been vastly expanded the last two decades. For some representative references the reader is prompted to look into the seminal paper by Engle and Granger (1987), the collection of papers by Engle and Granger (1991), the textbook by Banerjee *et al* (1993) and the textbook by Johansen (1996). Various questions concerning the underlying economic theory can be tested through cointegration models. For instance, in the I(1) model hypotheses on the long-run coefficients β can be used to test that one of the components of a vector series is stationary.

In this paper it is our aim to construct Wald tests of linear restrictions on the cointegrating vectors β_2 of the I(2) model that are orthogonal to the I(1) cointegrating

*This work is part of my Ph.D. thesis under the patient guidance of Søren Johansen. Thanks are also due to N. Haldrup, D. O'Brien and A. Soro Bonmati and seminar participants of the European University Institute, the Universidad de Alicante, the ASSET 2000 Euroconference at Lisbon and The Spring Meeting of Young Economists 2001 at Copenhagen who commented on earlier drafts of the paper. Any remaining errors are mine.

†Address for correspondence: Department of Economics, European University Institute, Via dei Roccettini, 9 - I-50016 San Domenico di Fiesole (FI), Italy. E-mail: Fragiskos.Archontakis@iue.it

space, see Johansen (1995, 1996, 1997) and Paruolo (1996). In particular, the linear restriction of a zero row on β_2 is a test for I(1)-ness of a variable in the I(2) model. This is useful and extends the existing theory in the I(1) cointegrated model where, in a similar fashion, there are already tests detecting whether a univariate component of a vector time series is stationary or non-stationary, see for example Johansen (1996) and Paruolo (1997), while this is not yet the case for the I(2) cointegrated model.

The rest of the paper is as follows. Section 2 briefly discusses unit-root testing and sets the scene. Section 3 provides the I(2) cointegration model with the theory on Wald test for linear restrictions on β_2. Section 4 contains the simulation experiment. Section 5 concludes.

Finally, a word on notation. The backshift lag operator L is defined on a time series X_t (scalar or vector) as $LX_t = X_{t-1}$ and in general $L^k X_t = X_{t-k}$, for all $k = 0, 1, 2...$ Trivially $L^0 X_t = X_t$, that is, $L^0 = 1$ the identity operator. Moreover, the difference operator Δ, is defined as $\Delta = 1 - L$. $\mathcal{R}^n$ is the n-dimensional space of the real numbers. I_n denotes the $n \times n$ identity matrix and by defining $\bar{\beta} = \beta\left(\beta'\beta\right)^{-1}$, the matrix $P_\beta = \bar{\beta}\beta'$ is the projection matrix on the space spanned by the columns of the $n \times r$ full column rank matrix β, $r < n$. The orthogonal complement of β is an $n \times (n - r)$ full column rank matrix $\beta_\perp$ such that $\beta'\beta_\perp = 0$. By the symbol $\xrightarrow{w}$ we mean weak convergence.

2 Unit-root testing issues

This section eases the way of introducing the test in a multivariate context by discussing the issue in a univariate context and for a lower level of integration. The proposed test has, in a certain sense, a resemblance to the univariate KPSS test, by Kwiatkowski, Phillips, Schmidt and Shin (1992), which is introduced as a test of trend-stationarity against a unit root (difference stationarity) alternative; see also Hornok and Larsson (2000) for the finite sample distribution of the KPSS test and Choi and Ahn (1998) for a multivariate version of the test.

Just to fix ideas for unit-root testing, recall first Dickey and Fuller's DF test, see amongst others Dickey and Fuller (1981) and Sims, Stock and Watson (1990). Consider the following regression model for the stochastic process y_t:

$$y_t = c + \delta t + \rho y_{t-1} + \varepsilon_t,$$

where ε_t are i.i.d. $N\left(0, \sigma^2\right)$ and wish to test the null hypothesis of a unit-root series (with drift) $\mathcal{H}_0 : \rho = 1, \delta = 0$ against the alternative of trend-stationarity $\mathcal{H}_a : \rho < 1$. The model is reparameterised as follows

$$y_t = \theta' z_{t-1} + \varepsilon_t$$

where $z_t = (1, y_t - ct, t)'$, $\theta' = (c + \delta, \rho, \delta + \rho c)$ and the OLS estimator of the model is

$$\hat{\theta}_{\text{OLS}} = \left(\sum_{t=2}^{T} z_{t-1} z'_{t-1} \right)^{-1} \sum_{t=2}^{T} z_{t-1} y_t,$$

which follows a non-standard distribution and under the null is given by:

$$Q_T \left(\hat{\theta}_{\text{OLS}} - \theta \right) = V_T^{-1} \phi_T$$

where we have defined $Q_T = diag \left\{ T^{1/2}, T, T^{3/2} \right\}$, $V_T = Q_T^{-1} \left(\sum_{t=2}^{T} z_{t-1} z'_{t-1} \right) Q_T^{-1}$ and $\phi_T = Q_T^{-1} \sum_{t=2}^{T} z_{t-1} y_t$. The test statistic for $\hat{\rho}_{\text{OLS}}$ is

$$T \left(\hat{\rho}_{\text{OLS}} - 1 \right) \xrightarrow{w} f\left(W \right),$$

where $\hat{\rho}_{\text{OLS}}$ is the super-consistent OLS estimator of the model and $f\left(W\right)$ a function of Brownian motions[1].

On the other hand the KPSS test has the following structure:

$$\begin{aligned} y_t &= r_t + \xi t + \varepsilon_t \\ r_t &= r_{t-1} + u_t \end{aligned}$$

where ε_t are i.i.d. $N\left(0, \sigma_\varepsilon^2\right)$ and u_t are i.i.d. $N\left(0, \sigma_u^2\right)$ and r_0 a fixed initial value. The test has the null hypothesis of trend-stationarity $\mathcal{H}_0 : \sigma_u^2 = 0$ against the alternative of a unit root $\mathcal{H}_a : \sigma_u^2 > 0$. The statistic is a Lagrange-Multiplier (LM) one-sided test constructed as follows:

- let $\hat{\varepsilon}_t$ be the residuals of the regression of y_t on an intercept and time trend.
- define the partial sum process of the residuals: $S_t = \sum_{i=1}^{t} \hat{\varepsilon}_i$, $t = 1, 2, ..., T$.
- the LM statistic is $LM = T^{-2} \sum_{t=1}^{T} S_t^2 / \hat{\sigma}_\varepsilon^2$, and follows the tabulated KPSS asymptotic distribution, where $\hat{\sigma}_\varepsilon^2$ is a consistent estimate of the variance of the series.

To our knowledge there is no way of rigorously comparing the two test statistics, since they are formulated under different null (and alternative) hypotheses. One can think of the two tests being complementary to one another, as it is also argued by Kwiatkowski et al (1992), page 160. The gist of the argument is that one should perform both tests in order to finalize a decision on the process under investigation. See also the following table:

		KPSS	
		$\mathcal{H}_0$: (trend) stationarity	$\mathcal{H}_1 : I(1)$-ness
DF	$\mathcal{H}_0 : I(1)$-ness	inconclusive evidence	series I(1)
	$\mathcal{H}_1$:(trend) stationarity	series (trend) stationary	inconclusive evidence

[1] See Banerjee, Dolado, Galbraith and Hendry (1993) for more details.

Other papers arguing in a similar fashion are Carrion-i-Silvestre, Sansó-i-Rosselló and Artís Ortuño (2001) and Charemza and Syczewska (1998).

In the same way, if one wants to test a series for I(1)-ness within the I(2) model, one can perform univariate Dickey-Fuller tests but should also carry out the complementary test proposed in the present paper which is formulated as the null hypothesis of I(1)-ness against the alternative of I(2)-ness.

3 Analysis of the I(2) system

In this section we present the I(2) model. We give a short summary of the statistical analysis and finally we consider hypothesis testing on the cointegrating vectors of the I(2) system.

3.1 The Statistical Model

We consider the vector autoregressive (VAR) model in p dimensions with k lags:

$$A(L)\, X_t = \epsilon_t, \;\; t = 1, ..., T, \tag{1}$$

where $A(z) = I_p - \Pi_1 z - ... - \Pi_k z^k$ is a finite matrix polynomial of order k and the $p \times 1$ vector errors ϵ_t are i.i.d. Gaussian with zero mean and finite variance-covariance matrix Ω. We assume that $|A(z)| = 0$ has roots in the complex plane at $z = 1$ or $|z| > 1$, in order to exclude the existence of seasonal roots, i.e. roots with $|z| = 1$ and $z \neq 1$, and explosive roots, i.e. roots for $|z| < 1$. The VAR model in (1) can be written equivalently in the error-correction model (ECM) form:

$$\Delta^2 X_t = \Pi X_{t-1} - \Gamma \Delta X_{t-1} + \sum_{i=1}^{k-2} \Psi_i \Delta^2 X_{t-i} + \epsilon_t, \; t = 1, ..., T, \tag{2}$$

where we have omitted constant terms and dummies for simplicity. It is assumed that Π is a reduced rank matrix of rank r, i.e. the cointegration rank of the system is $r < p$, which implies that there exist $p \times r$ full rank matrices α and β such that $\Pi = \alpha\beta'$. The assumption that allows for order of integration higher than one is $\alpha'_\perp \Gamma \beta_\perp = \xi\eta'$, where ξ, η are full rank matrices $(p-r) \times s$ of rank $s < p - r$. Now, the direction matrices of full rank $\beta_1 = \bar{\beta}_\perp \eta$ and $\beta_2 = \beta_\perp \eta_\perp$ are of dimensions $p \times s$ and $p \times (p - r - s)$, respectively. Thus β, β_1, β_2 are mutually orthogonal and $(\beta, \beta_1, \beta_2)$ is also a full rank matrix of rank p, i.e. β, β_1 and β_2 span the whole space $\mathcal{R}^p$. Furthermore, the condition that assures $X_t \sim I(2)$ and not of higher order is that the matrix $\alpha'_2 M \beta_2$ is of full rank $p - r - s$, where $M = \Gamma\bar{\beta}\bar{\alpha}'\Gamma + I_p - \sum_{i=1}^{k-2} \Psi_i$. This is the I(2) cointegrated model.

The representation for X_t in case $\alpha'_2 M \beta_2$ has full rank is

$$X_t = C_2 \sum_{j=1}^{t} \sum_{i=1}^{j} \epsilon_i + C_1 \sum_{i=1}^{t} \epsilon_i + A + Bt + Y_t, \tag{3}$$

where A, B constants in $\mathcal{R}^p$ and Y_t a stationary process. It follows from (3) that $X_t \sim I(2)$ since it is a function of a cumulated random walk. We note that since $(\beta;\beta_1)'\beta_2 = 0$ we have that $(\beta;\beta_1)'C_2 = 0$ and hence $(\beta;\beta_1)'X_t \sim I(1)$. Here $C_2 = \beta_2(\alpha_2'M\beta_2)^{-1}\alpha_2'$, while for C_1 it holds that $\beta'C_1 = \bar{\alpha}'\Gamma C_2$ and $\beta_1'C_1 = \bar{\alpha}_1'(I - MC_2)$, so that

$$\beta'X_t - \bar{\alpha}'\Gamma\Delta X_t = \beta'C_1\sum_{i=1}^{t}\epsilon_i - \bar{\alpha}'\Gamma C_2\sum_{i=1}^{t}\epsilon_i + \text{stat.} = \text{stat. process.}$$

In this case X_t is called multicointegrated.

The statistical analysis of the model consists of either the two-step estimation procedure, see Johansen (1995), or the maximum likelihood procedure as developed in Johansen (1997). The former can be shown to be asymptotically equivalent to the MLE, see also Paruolo (2000). The procedure consists of two reduced rank regressions (RRR). After concentrating out the effects of the short-term dynamics Δ^2X_{t-i}, by regressing Δ^2X_t, X_{t-1} and ΔX_{t-1} on the Δ^2X_{t-i}, for all $i = 1, ..., k-2$, we work with the residuals of these regressions Z_{0t}, Z_{1t} and Z_{2t}, respectively, and have:

$$Z_{0t} = \alpha\beta'Z_{1t} - \Gamma Z_{2t} + u_t$$

Thus in Step 1 we apply RRR for Z_{0t} on Z_{1t} corrected for Z_{2t} and derive estimates for α, β and r. In Step 2 we consider α, β and r known and fixed equal to the parameter estimates. Then the model can be transformed into a partial system in the following way:

$$\begin{aligned}\alpha_\perp'Z_{0t} &= -\alpha_\perp'\Gamma Z_{1t} + \alpha_\perp'u_t\\ \bar{\alpha}'Z_{0t} &= \beta'Z_{1t} - \bar{\alpha}'\Gamma Z_{1t} + \bar{\alpha}'u_t\end{aligned}$$

and the first equation can be rewritten as

$$\alpha_\perp'Z_{0t} = -\alpha_\perp'\Gamma\left(\bar{\beta}\beta' + \beta_\perp\bar{\beta}_\perp'\right)Z_{1t} + \alpha_\perp'u_t$$

or

$$\alpha_\perp'Z_{0t} = -\alpha_\perp'\Gamma\bar{\beta}\left(\beta'Z_{1t}\right) - \xi\eta'\left(\bar{\beta}_\perp'Z_{1t}\right) + \alpha_\perp'u_t.$$

Then with a RRR of $\alpha_\perp'Z_{0t}$ on $\bar{\beta}_\perp'Z_{1t}$ corrected for $\beta'Z_{1t}$ we derive ξ, η and s.

A useful reparametrization of the model, due to Johansen (1997), can be achieved by choosing τ such that $sp\{\tau\} = sp\{\beta;\beta_1\}$ and $\tau_\perp = \beta_2$ with dimensions $p\times(r+s)$ and $p\times(p-r-s)$. We also define the matrix ρ with dimensions $(r+s)\times r$ so that τ and ρ are chosen to give $\beta = \tau\rho$ a $p\times r$ matrix of full rank r. Note that there are many different choices for ρ; an obvious one is $\rho = (I_r;0)'$ which applies when $\tau = (\beta;\beta_1)$. One can also derive $\beta_1 = \bar{\beta}_\perp\eta = \bar{\tau}\bar{\rho}_\perp(\bar{\rho}_\perp'\bar{\tau}'\bar{\tau}\bar{\rho}_\perp)^{-1}$. Finally define

$\psi' = -\left(\alpha'\Omega^{-1}\alpha\right)^{-1}\alpha'\Omega^{-1}\Gamma$ an $r \times p$ matrix, and $\kappa' = -\left(\alpha'_{\perp}\Gamma\bar{\beta}, \xi\right)$ a $(p-r)\times(r+s)$ matrix. In this way the I(2) model in (2) becomes

$$\Delta^2 X_t = \alpha\left(\rho'\tau' X_{t-1} + \psi'\Delta X_{t-1}\right) + \Omega\alpha_{\perp}\left(\alpha'_{\perp}\Omega^{-1}\alpha_{\perp}\right)^{-1}\kappa'\tau'\Delta X_{t-1} + \sum_{i=1}^{k-2}\Psi_i\Delta^2 X_{t-i} + \epsilon_t \tag{4}$$

and the multicointegration relation $\rho'\tau' X_{t-1} + \psi'\Delta X_{t-1} \sim I(0)$ enters directly into the model. This reparametrization has the additional advantage that the parameters vary freely so maximum likelihood analysis is easier to apply, see Johansen (1997).

Another, more complicated, way of estimation is via the maximum likelihood procedure as developed in Johansen (1997). It is based on the ECM representation of equation (4) which, after transformed in partial systems (as above), gives rise to a switching algorithm:

- for fixed $\tau = (\beta; \beta_1)$ estimate $\alpha_{\perp}$ (via an eigenvalue problem) and the other parameters (via regression)

- for fixed values of the parameters estimate τ (via generalized least squares).

Once τ is found, then $\beta_2 = \tau_{\perp}$ can be also determined. This is the method which we shall use in order to implement the Wald test for β_2 and simulate its behaviour.

One can find, see also Johansen (1997), that for β normalized on c, we have that $\hat{\beta}$ is asymptotically mixed Gaussian and hence can be tested with a Wald test using an estimate of the asymptotic conditional variance which is given by the observed information[2], see also Theorem 3 in Johansen (1997):

$$\left(I - \hat{\beta}c'\right)\hat{\beta}_{\perp}\left(\hat{\beta}'_{\perp}\sum_{t=1}^{T} Z_{1t}Z'_{1t}\hat{\beta}_{\perp}\right)^{-1}\hat{\beta}'_{\perp}\left(I - c\hat{\beta}'\right)\otimes\left(\hat{\alpha}'_c\hat{\Omega}^{-1}\hat{\alpha}_c\right)^{-1}.$$

Hypothesis testing on the vectors of the β_2 matrix is the purpose of the following section.

3.2 Hypothesis Testing for β_2 in the I(2) Model

In this subsection we consider hypothesis testing for β_2 in the I(2) cointegrated model. The paper is supplementary to the work by Johansen (1997) where hypothesis testing for β in the same model is considered. Using hypothesis testing for β_2 we can answer the following question: "Are there any I(2) trends in a given univariate component of the process?"

We want to find a Wald test for $\mathcal{H}_0 : R'\beta_2 = 0$, where R is a $p \times m$ matrix of m linear restrictions. In this way we can test that the i^{th}-row of the β_2 matrix is a zero row and hence deduce that the X_{it} variable does not contain any I(2) trends.

[2] The observed information is $-\partial^2 \log L(\beta)/\partial\beta^2$ to be distinguished from the Fisher's information $E_\theta\left[-\partial^2 \log L(\beta)/\partial\beta^2\right]$. With $L(\beta)$ we denote the likelihood function of the model.

Hence instead of I(2)-testing for unit roots in the univariate series by a Dickey-Fuller test or an alternative test and then proceeding modelling the series in a VAR, there is now the possibility of choosing the VAR framework and then testing for I(1). The difference between the two testing methods is the following: in a univariate framework test the null hypothesis is of I(2)-ness, while in a multivariate test the null hypothesis is that of I(1)-ness. By a multivariate test, one exploits more information, i.e. not only the data of the particular univariate series of interest but also the other series relevant, because of cointegration (for example), to the model. See also Hansen (1995) for a paper[3] where covariates are used in order to increase the power of unit root tests in a univariate approach.

Now define $\tau = (\beta; \beta_1)$ and since β, β_1 and β_2 span the whole space $\mathcal{R}^p$ it yields $\beta_2 = \tau_\perp$. Theorem 4 in Johansen (1997) states that if τ and $\hat{\tau}$ are normalized by the $p \times (r+s)$ full rank matrix b as $b'\tau = I_{r+s}$ and since

$$T\left(\hat{\tau} - \tau\right) \xrightarrow{w} \left(I - \tau b'\right) \beta_2 C^\infty \bar{\rho}'_\perp$$

and $\hat{\beta}_2 = \hat{\tau}_\perp = \left(I - b\hat{\tau}'\right) b_\perp$, we can derive that

$$T\left(\hat{\beta}_2 - \beta_2\right) \xrightarrow{w} -b\bar{\rho}_\perp \left(C^\infty\right)' \left(\beta_2'\beta_2\right),$$

where

$$C^\infty = \left[\int_0^1 H_0 H_0' dt\right]^{-1} \int_0^1 H_0 \left(dW_2\right)'$$

with H_0 being the limit of

$$T^{-\frac{1}{2}} \beta_2' Z_{2[Tu]} \xrightarrow{w} \beta_2' C_2 W\left(u\right) = H_0\left(u\right),$$

furthermore, see equation (25) in Johansen (1997)

$$T^{-2} \sum_{t=1}^{T} \left(\hat{\beta}_2' Z_{2t}\right) \left(\hat{\beta}_2' Z_{2t}\right)' \xrightarrow{w} \int_0^1 H_0 H_0' dt$$

and

$$W_2\left(u\right) = \left[\bar{\rho}'_\perp \kappa \left(\alpha'_\perp \Omega \alpha_\perp\right)^{-1} \kappa' \bar{\rho}_\perp\right]^{-1} \bar{\rho}'_\perp \kappa \left(\alpha'_\perp \Omega \alpha_\perp\right)^{-1} \alpha'_\perp W\left(u\right).$$

It holds that C^∞ is a mixed Gaussian distribution with mixing parameters

$$\left[\int_0^1 H_0 H_0' dt\right]^{-1} \otimes \left[\bar{\rho}'_\perp \kappa \left(\alpha'_\perp \Omega \alpha_\perp\right)^{-1} \kappa' \bar{\rho}_\perp\right]^{-1},$$

that is, if we condition upon $\mathcal{H}_0$ then C^∞ is a Gaussian.

[3] Niels Haldrup kindly pointed out this reference to me.

Remark. The distributions that were just derived are both singular. The distribution of $\hat{\tau}$ is singular in the directions b' from the left and ρ from the right, while the one of $\hat{\tau}_{\perp}$ is singular in the directions $b'_{\perp}$ and $\rho'\bar{b}'$ from the left.

The dimensions of the parameters entering the limiting distribution of interest are: $\beta_2 : p \times (p-r-s)$, $b : p \times (r+s)$ and $\bar{\rho}_{\perp} : (r+s) \times s$, hence C^{∞} is of $(p-r-s) \times s$ dimensions. This implies, see also the expression for the mixing parameters of C^{∞}, that the variance-covariance matrix will be of dimensions $[(p-r-s)s] \times [(p-r-s)s]$. We assume that the normalization is chosen such that the matrix $R'b$ is of full rank, because b is only a normalization parameter and hence it can be suitably chosen in order not to interfere with the hypothesis we want to test. We also need to assume that the $m \times s$ matrix $R'b\rho_{\perp}$ is of full rank $m \leq s$. Formally one should test for this before proceeding to hypothesis testing for the β_2 matrix. Nevertheless, the argument that the parameters satisfying this condition consist of a set with small measure will be applied here.

Hence, one should define the restriction matrices R_1 of dimensions $p \times m$, with $m \leq s$, and R_2 of $(p-r-s) \times n$, respectively; m and n determine the number of desired restrictions on β_2. Thus, the hypothesis $R'_1\beta_2 = 0$, or $R'_1\beta_2 R_2 = 0$, can be tested via Wald test. Because the variance of C^{∞} has the form $Var\,(C^{\infty}) = A \otimes B$, with $A = \left[\int_0^1 H_0 H'_0 dt\right]^{-1}$ of $(p-r-s) \times (p-r-s)$ dimensions and $B = \left[\bar{\rho}'_{\perp}\kappa\,(\alpha'_{\perp}\Omega\alpha_{\perp})^{-1}\kappa'\bar{\rho}_{\perp}\right]^{-1}$of $s \times s$ dimensions, the $Var\left(R'_1\hat{\beta}_2 R_2\right)$ is given by

$$Var\left(R'_1\hat{\beta}_2 R_2\right) = Var\left[\lambda'\,(C^{\infty})'\,\mu\right] = (\lambda' B\lambda) \otimes (\mu' A\mu)\,,$$

where we set $\lambda = -\bar{\rho}'_{\perp}b'R_1$ and $\mu = (\beta'_2\beta_2)\,R_2$. Hence the Wald test is constructed as[4]

$$W = T^2 tr\left\{\left(\hat{\lambda}'\hat{B}\hat{\lambda}\right)^{-1}\left(R'_1\hat{\beta}_2 R_2\right)\left(\hat{\mu}'\hat{A}\hat{\mu}\right)^{-1}\left(R'_1\hat{\beta}_2 R_2\right)'\right\} \tag{5}$$

and it holds that $W \xrightarrow{w} \chi^2_f$, for $T \to \infty$. The number of degrees of freedom $f = mn$ in the limiting chi-square distribution is equal to the rank of the matrix $Var\left(R'_1\hat{\beta}_2 R_2\right)$. We still have to find a way to estimate $Var\left(R'_1\hat{\beta}_2 R_2\right)$. It holds that $\hat{\beta}_2$ can be estimated from the two-step procedure. R_1 and R_2 are known and chosen by the investigator. Hence $\hat{\mu}$ can be found. The factor A can be substituted by the limit $\left[\int_0^1 H_0 H'_0 dt\right]^{-1}$. The rest of the parameters, i.e. κ, $\rho_{\perp}$, Ω, $\alpha_{\perp}$, can be estimated. Hence B can be found. The matrix b is chosen suitably for the normalization and is thus known, hence $\hat{\lambda}$ can be found.

Example. Consider the case of a VAR model in $p = 3$ dimensions, cointegration rank $r = 1$ and $s = 1$, hence there is only one I(2) trend $(p-r-s = 1)$. The

[4]Note here that formally speaking it is not A, but the estimate as given by the product moments of $\hat{\beta}'_2 Z_{2t}$ to enter the W test.

hypothesis to be tested is $(\beta_2)_3 = 0$. Hence β_2: 3×1, b: 3×2 and one can assume that b and ρ are such that the scalar $R'b\rho_\perp \neq 0$. Under these circumstances one can proceed with the hypothesis testing as described above. For testing the 3rd element of the β_2 vector one should apply $R_1 = R = (0, 0, 1)'$ and $R_2 = 1$. Because the mixed Gaussian limiting distribution C^∞ is, in this particular case, a 1×1 scalar we can impose only one linear restriction. We now have $\lambda = 1$, $\mu = \beta_2' \beta_2$ and the Wald test is

$$W_{test} = \frac{T^2 \left[(\hat{\beta}_2)_3 \right]^2}{\left(\hat{\lambda}' \hat{B} \hat{\lambda} \right) \left(\hat{\mu}' \hat{A} \hat{\mu} \right)} \xrightarrow{w} \chi_1^2, \text{ for } T \to \infty.$$

with notation as before for the rest of the parameter estimates, that is:

$$\hat{\lambda}' \hat{B} \hat{\lambda} = \left(-\widehat{\rho}_\perp' b' R_1 \right)' \left[\widehat{\rho}_\perp' \hat{\kappa} \left(\hat{\alpha}_\perp' \hat{\Omega} \hat{\alpha}_\perp \right)^{-1} \hat{\kappa}' \widehat{\rho}_\perp \right]^{-1} \left(-\widehat{\rho}_\perp' b' R_1 \right)$$

and

$$\hat{\mu}' \hat{A} \hat{\mu} = \left(\hat{\beta}_2' \hat{\beta}_2 \right) \left[T^{-2} \sum_{t=1}^{T} \left(\hat{\beta}_2' Z_{2t} \right) \left(\hat{\beta}_2' Z_{2t} \right)' \right]^{-1} \left(\hat{\beta}_2' \hat{\beta}_2 \right).$$

4 Simulations

In order to obtain a feeling of how the proposed Wald tests works in practice we have carried out a small Monte Carlo experiment. In particular, we examined the size properties of the Wald test, the rate of convergence of the test and got some critical values at 5% nominal level.

4.1 The Data Generation Process

The Data Generation Process (DGP) consists of a VAR with $p = 5$ and $k = 2$:

$$X_t = \Pi_1 X_{t-1} + \Pi_2 X_{t-2} + e_t,$$

where the error term e_t is iid and follows[5] $e_t \sim N(0, \Omega)$, $\Omega = 0.01 * I_5$ and we specified

$$\Pi_1 = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho & 0 \\ 0 & 0 & 0 & 0 & \varphi \end{pmatrix}, \quad \Pi_2 = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

[5] Variance scaling was applied because the $I(2)$ process was giving too large observation values at large samples.

Because of the presence of nonstationary processes in the model all initial values were set to be equal to zero in order to avoid having deterministic terms in the model, i.e. $X_{i,-1} = X_{i,0} = 0$. The autoregressive coefficients ρ and φ take the values 0.0 , 0.5 and 0.9. We consider the following collections of DGPs depending on the choice of ρ and φ:

$\varphi \setminus \rho$	0.0	0.5	0.9
0.0	$\mathcal{DGP}_1$	$\mathcal{DGP}_2$	$\mathcal{DGP}_4$
0.5		$\mathcal{DGP}_3$	$\mathcal{DGP}_5$
0.9			$\mathcal{DGP}_6$

Hence, the designed β-matrices are found to be

$$\beta = (0, I_3)'; \; \beta_1 = (0, 1, 0, 0, 0)'; \; \beta_2 = (1, 0, 0, 0, 0)'$$

and X_{1t} is an I(2) process, X_{2t} is an I(1) process, X_{3t} is an I(1) process that provides the multicointegration relationship in the model and X_{4t}, X_{5t} are stationary AR(1) processes. That is, the DGP has the specification $(p, r, s) = (5, 3, 1)$.

The null hypothesis is $\mathcal{H}_0 : (\beta_2)_2 = 0$, while the true $(\beta_2)_2 = 0$. The Wald test follows asymptotically a χ^2_1 distribution and the nominal 5% level critical value for the χ^2_1 is 3.84. Six sample sizes were used, that is, $T = 25$, 50, 75, 100, 200, 400. The number of replications was chosen to be $n = 10000$ for each specification[6].

4.2 Simulation Results

The 5% level empirical sizes and the 95% empirical critical values (as 95% quantiles in the simulations under the null hypothesis) of the Wald test statistic were calculated from the Monte Carlo rejection frequencies based on the asymptotic critical values of the χ^2_1 at the 5% level, that is, $\chi^2_1(0.95) = 3.84$. The results are reported in Tables 1 and 2.

Overall Comments:

¨ When departing from our DGP, that is, for a relatively high variance $\sigma^2_2 = var(e_{2t}) = 1$ of $X_{2t} \sim I(1)$ with respect to $\sigma^2_1 = var(e_{1t}) = 0.01$ of $X_{1t} \sim I(2)$ we get higher rejection frequencies for small and medium samples, while it affects less large samples. This is because the I(2)-variance is of order $O(\sigma^2_1 \frac{T^3}{3})$ relative to the I(1)-variance of order $O(\sigma^2_2 T)$. For instance, with $T = 100$, $\rho = \varphi = 0.5$ and 10000 replications we get an empirical size $\alpha^*_{\sigma^2_2 = 1} = 15\%$ with a 95% critical value of $\mathcal{F}_{\sigma^2_2 = 1}(0.95) = 9.76$, while the corresponding values for $\sigma^2_2 = \sigma^2_1 = 0.01$ are, see model $\mathcal{DGP}_3$ in Table 1, $\alpha^*_{\mathcal{DGP}_3} = 9\%$ and, see model $\mathcal{DGP}_3$ in Table 2, $\mathcal{F}_{\mathcal{DGP}_3}(0.95) = 5.24$, respectively.

Interpretation: the large variance $\sigma^2_2 >> \sigma^2_1$ implies that the effect of the I(2) process being "hidden" behind the I(1) process, particularly in small samples.

[6]Should anyone wish to replicate the results the simulations were carried out in RATS 4.30 with the seed chosen to be equal to 20.

¨ When $X_{it} \longrightarrow I(1)$, for $i = 4, 5$, instead of $X_{it} \sim I(0)$, it yields that we get higher rejection frequencies since for $X_{4t} \longrightarrow I(1)$ or $X_{5t} \longrightarrow I(1)$ the cointegration rank $r \longrightarrow 2$, instead of $r = 3$; when both $X_{4t} \longrightarrow I(1)$ and $X_{5t} \longrightarrow I(1)$ it even holds $r \longrightarrow 1$, instead of $r = 3$.

Interpretation: the test is unstable when $(r, s) \neq (r, s)_{true}$, which could be seen as a model misspecification problem.

¨ The Wald test is asymptotically a χ^2_f distribution, where f is the number of degrees of freedom. Hence one would expect that in small samples, like for $T = 25$ or $T = 50$ observations, the test would not work satisfactory. This is also supported by the simulation study. The simulated critical values imply that the distributions have very long tails in small samples.

As the sample size T increases from 25 to 400, we can see from Tables 1 and 2 that the empirical size decreases and the empirical critical values are closer to the asymptotic critical values:

Table 1: Empirial Size α^*(nominal size $\alpha = 5\%$) of the Wald test for β_2 for different DGPs and $n = 10000$ replications

	$\mathcal{DGP}_1$	$\mathcal{DGP}_2$	$\mathcal{DGP}_3$	$\mathcal{DGP}_4$	$\mathcal{DGP}_5$	$\mathcal{DGP}_6$
$T = 25$	0.31	0.35	0.39	0.42	0.45	0.47
$T = 50$	0.12	0.15	0.18	0.31	0.31	0.39
$T = 75$	0.09	0.10	0.11	0.26	0.26	0.34
$T = 100$	0.07	0.08	0.09	0.21	0.21	0.28
$T = 150$	0.07	0.07	0.07	0.16	0.16	0.21
$T = 200$	0.06	0.07	0.07	0.14	0.14	0.17
$T = 400$	0.06	0.06	0.06	0.10	0.10	0.11

and

Table 2: Empirical Critical Values $[\chi^2_1(0.95) = 3.84]$ of the Wald test for β_2 for different DGPs and $n = 10000$ replications

	$\mathcal{DGP}_1$	$\mathcal{DGP}_2$	$\mathcal{DGP}_3$	$\mathcal{DGP}_4$	$\mathcal{DGP}_5$	$\mathcal{DGP}_6$
$T = 25$	50.7	91.5	121	253	336	450
$T = 50$	6.39	9.03	12.2	94.8	99.8	179
$T = 75$	5.18	5.61	6.06	54.7	62.0	129
$T = 100$	4.50	5.00	5.24	26.4	28.1	70.1
$T = 150$	4.39	4.59	4.56	13.1	13.4	26.5
$T = 200$	4.16	4.34	4.39	9.72	9.38	13.8
$T = 400$	4.15	4.27	4.34	5.99	6.05	6.23

The ordering of the collection of DGPs according to their performance in the Wald test is

$$\mathcal{DGP}_1 \succ \mathcal{DGP}_2 \succ \mathcal{DGP}_3 \succ \mathcal{DGP}_4 \succ \mathcal{DGP}_5 \succ \mathcal{DGP}_6,$$

which is a "natural" result, given the comment on the correct specification of (r, s).

In order to have an idea of the precision of the empirical sizes α^* in the tables above we shall calculate their variance. For each simulation we assign probabilities for a nominal 5% test, assuming that $W \sim \chi^2_1$, as follows:

$$\begin{aligned} p &= \Pr[\text{Reject if } W > 3.84] = 0.05, \text{ say, and} \\ q &= 1 - p = \Pr[\text{Do not Reject if } W \leq 3.84] = 0.95, \end{aligned}$$

depending on whether the Wald test rejects the null hypothesis $\mathcal{H}_0$, or not. Hence the random variable Y, defined as $Y := \#$ of rejections of $\mathcal{H}_0$, follows a Binomial distribution. Thus $Y \sim B(n, p)$, where $n = \#$ of replications, and its variance is $Var(Y) = npq$, while the variance of the empirical size α^* will be

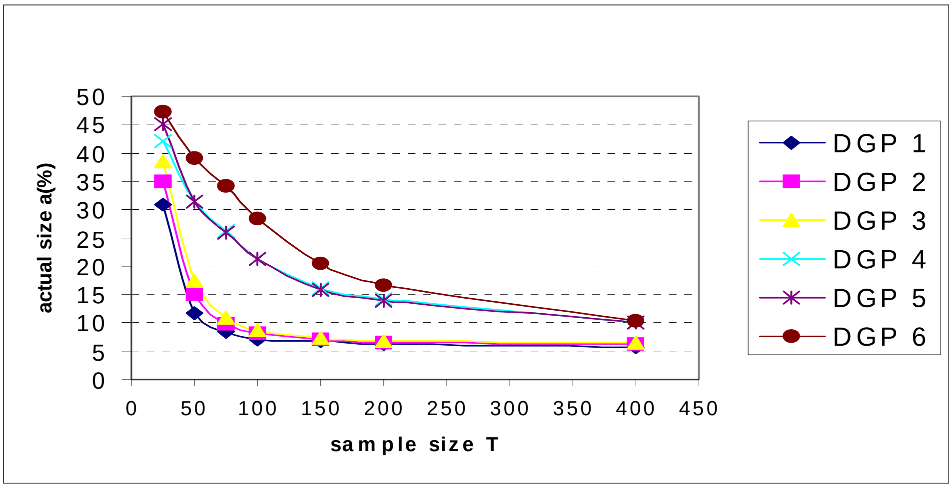
$$Var\left(\frac{1}{n}Y\right) = \frac{1}{n^2}Var(Y) = \frac{pq}{n} \Longrightarrow S.D.(\alpha^*) = \sqrt{\frac{pq}{n}}.$$

As a result, for a confidence interval for the empirical size α^* we have

$$\hat{\alpha}^* \in [\hat{\alpha}^* \pm 2 * 0.00218] \approx [\hat{\alpha}^* \pm 0.0044],$$

but only for $\hat{\alpha}^* \simeq 0.05$. Hence, for $\hat{\alpha}^* \simeq 0.05$ one would tend to keep two decimal points for $\hat{\alpha}^*$[7].

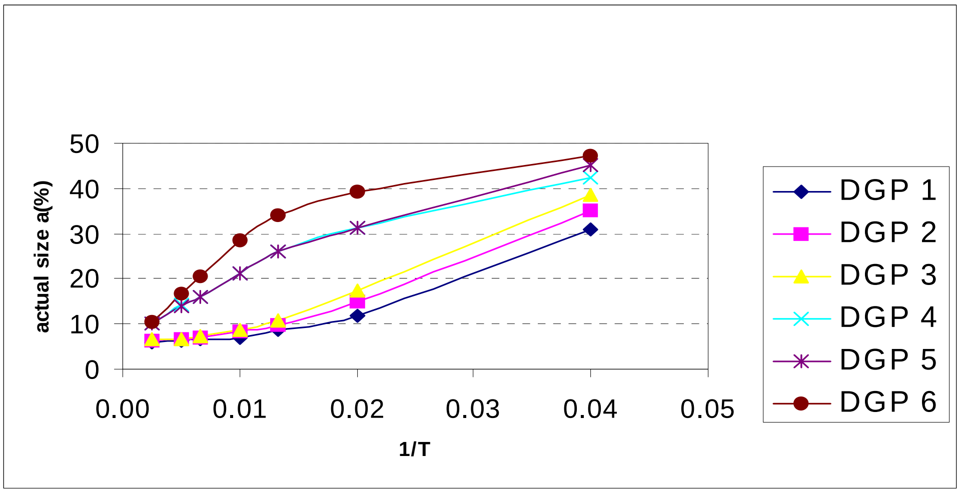
The graphs corresponding to the simulated empirical sizes are as follows:



Simulated actual size of the Wald test for β_2(# of replications = 10000).

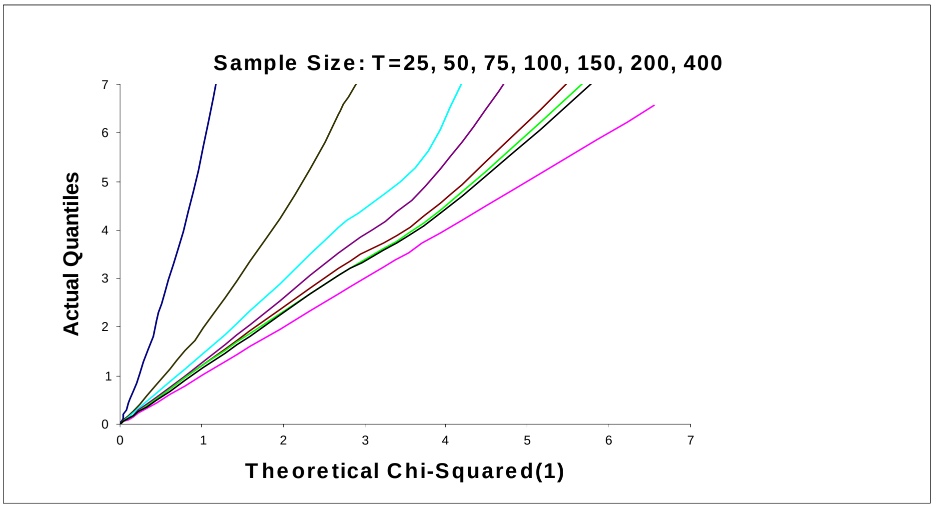
[7]Even when $\hat{\alpha}^* = 0.5$, the confidence interval has the form $[\hat{\alpha}^* \pm 0.01]$ which is farely accurate.

and



Rate of convergence of the Wald test for β_2.

Finally, for the "baseline DGP" $\mathcal{DGP}_3$ we provide the QQ-plot,



QQ-plot for $\mathcal{DGP}_3$: Wald test for β_2.

where it is clear that the Wald test overrejects in all samples. By using the conjecture that the order of convergence of the test is approximately a polynomial of T^{-h}, that is

$$T\left(\hat{\beta}_2 - \beta_2\right) = MG + \frac{c_1}{T} + \frac{c_2}{T^2} + ...,$$

where MG =Mixed Gaussian, we run a simple linear regression of $\log\left[\alpha^*(T) - 0.05\right]$ on a constant and the terms up to $\mathcal{O}\left(T^{-2}\right)$:

$$\log\left[\alpha^*(T) - 0.05\right] \approx c + \frac{c_1}{T} + \frac{c_2}{T^2},$$

where the full results of the regression where:

Variable	Coeff	Std Error	T-Stat	Signif
Constant	-5.019	0.109	-45.87	0.00000135
T^{-1}	200.748	14.761	13.599	0.00016926
T^{-2}	-2557.86	331.56	-7.714	0.00151978

hence we find that for rate of convergence of the "baseline DGP" $\mathcal{DGP}_3$ following formula holds approximately:

$$\log\left[\alpha^*(T) - 0.05\right] \approx -5 + 4 * \frac{50}{T} - \left(\frac{50}{T}\right)^2.$$

4.3 Pretesting

In order to determine the order of integration of the variable X_{it}, $i = 1, ..., p$, one has to carry out some preliminary test determine the cointegrating ranks r and s. In this subsection we briefly discuss the effect that this pretesting[8] may have upon a subsequent test for β_2.

The question we wish to answer is:

"What is the distribution of the test for the hypothesis $\mathcal{H}_0 : (\beta_2)_i = 0$, for some $i = 1, ..., p$, given that the cointegrating indices r and s are determined?"

Clearly the answer depends on the values of the cointegrating indices.

We have that the likelihood ratio test statistic for estimating the cointegrating indices r and s is given by the formula, see Johansen (1995):

$$S_{r,s} = Q_{r,s} + Q_r$$

where

$$Q_{r,s} = -2\log Q\left(H_{r,s} \mid H_r\right)$$

is a trace test statistic for the cointegrating index s when the rank r is fixed and

$$Q_r = -2\log Q\left(H(r) \mid H(p)\right)$$

is the I(1) trace test. The hypothesis $H_{r,s}$ is rejected if and only if the $H_{i,j}$ are rejected for all $i < r$ and for $i = r$ and $j \leq s$.

In order to estimate the cointegrating ranks from the data one has to carry out the following sequential procedure and choose the first non-rejected pair (r, s). In more detail:

1. Compare $S_{0,0}$ with its quantile c_{00}, say. If $S_{0,0} < c_{00}$, then choose $\hat{r} = 0$ and $\hat{s} = 0$, while if $S_{0,0} \geq c_{00}$, then

2. Calculate $S_{0,1}$ and compare it with its quantile c_{01}, say. If $S_{0,1} < c_{01}$, then choose $\hat{r} = 0$ and $\hat{s} = 1$, while if $S_{0,1} \geq c_{01}$, then

[8] Thanks are due to Niels Haldrup for pointing this out.

3. Calculate $S_{0,2}$ and compare it with its quantile c_{02}. If $S_{0,2} < c_{02}$, then choose $\hat{r} = 0$ and $\hat{s} = 2$, while if not compare $S_{0,3}$ with its quantile c_{03}, etc.

The above is a sequential decision process considering the Intersection-Union test (IUT). An IUT tests each of the hypotheses $H_{i,j}$ individually and rejects $H_{r,s}$ if and only if each of the sub-hypotheses $H_{i,j}$ is rejected because the test of $H_{r,s}$ concerns a union of the disjoint hypotheses $H_{i,j}$. Hence on an overall decision set D_{rs}, say, we decide that $r = \hat{r}$ and $s = \hat{s}$.

Define the non-critical region sets

$$\mathcal{A}_{r,s} = \{S_{0,0} \geq c_{00}, S_{0,1} \geq c_{01}, ..., S_{0,p-1} \geq c_{0,p-1}, S_{1,0} \geq c_{10}, ..., S_{r,s-1} \geq c_{r,s-1}, S_{r,s} \geq c_{rs}\},$$

for all possible pairs (r, s). They form a sequence of disjoint sets $\mathcal{A}_{0,0} = \{S_{0,0} \geq c_{00}\}$, $\mathcal{A}_{0,1} = \{S_{0,0} \geq c_{00}, S_{0,1} \geq c_{01}\}$, ..., etc. Thus the final decision is based on

$$D_{r,s} = \mathcal{A}_{r,s-1} \cap \mathcal{A}_{r,s}^{C} = \mathcal{A}_{r,s-1} \cap \{S_{r,s} < c_{rs}\},$$

while $D_{0,0} = \mathcal{A}_{0,0}^{C} = \{S_{0,0} < c_{00}\}$.

Hence for our hypothesis $\mathcal{H}_0$ it holds that the probability

$$\Pr\left[W_{(\beta_2)_i=0} \leq c \mid rank\,(\beta_2) = p - r - s\right] \longrightarrow \chi^2_{p-r-s}(c).$$

Let us assume that we decide based on a 10% test.

In our DGP specification, we know that $p = 5$ and the correct ranks are $r_0 = 3$ and $s_0 = 1$. That is, for the final test of the sequential procedure, we have

$$S_{3,1} = Q_{3,1} + Q_3 = -2\log Q\,(H_{3,1} \mid H_3) - 2\log Q\,(H_3 \mid H_5)$$

and $\Pr[S_{3,1} < c_{31}] = 0.9$, while one actually should find $\Pr[S_{3,1} < c_{31} \mid \mathcal{A}_{3,0}]$.

If $r = 3$ and $s = 1$, then all the test statistics that define $\mathcal{A}_{3,0}$ will diverge to ∞, so that for large T there is no difference between the two procedures

$$\lim_{T\to\infty} P\{S_{3,1} < c_{3,1} | A_{3,0}\} = \lim_{T\to\infty} P\{S_{3,1} < c_{3,1}\}.$$

When it comes to testing the order of integration of the i-th variable X_{it} we rely on having found the correct value of r and s. Asymptotically[9] we know that the correct value is attained with 90%, and values lower with zero probability. Thus there is a possibility that the I(2) rank will be incorrectly determined and that means that the test for I(1)-ness can be wrong. It is therefore a good idea to see how the conclusion changes for different choices of the I(2) rank $p - r - s$.

[9] In finite samples, we admit that the Wald test for the hypothesis $\mathcal{H}_0 : (\beta_2)_2 = 0$, which will follow a $\Pr\left[W_{(\beta_2)_2=0} \leq c \mid rank\,(\beta_2) = 1\right] = \chi^2_1(c)$ distribution, may be affected slightly from pretesting effects. However, to our knowledge, there is not yet developed an alternative way for carrying out a similar exercise.

4.4 Further investigation of the test

In this subsection we further consider the behaviour of the Wald test in two ways. Firstly, we use some simulation results carried out by Johansen (1995) to see how often the correct values of the pair (r, s) are achived and hence the test is valid. Secondly, we consider the effect to the test of a wrong rank r choice.

The Wald test considered in this paper is a test contingent on the choice of r and s since its value depends on the pair (r, s) of cointegrating indices. However, it is not always the case that the correct values of the pair (r, s) are achived with probability close to unity. We reproduce here Table 4 of page 44 in Johansen (1995) where the joint probability $\lim_T \Pr\left[Q_r \leq c_r\,(5\%)\,,\ Q_{r,s} \leq c_{r,s}\,(5\%)\right]$ is calculated via simulation for a time series with sample size $T = 200$:

$p-r-s \setminus s$	0	1	2	3
0		0.95	0.95	0.95
1	0.88	0.83	0.78	0.75
2	0.70	0.58	0.51	0.45
3	0.38	0.28	0.21	0.15

It is seen from the Table above that for the case $r = 3$, $(s, p-r-s) = (1, 1)$ that we are interested in due to our DGPs, the probability of attaining the correct pair of values is equal to 0.83. Hence, for the Wald test regarding $\mathcal{DGP}_1$, say, and by choosing $T = 200$ in order to be in accordance with the Table above, we have (see Table 1, paragraph 4.2) empirical size 0.06 and probability of correctly not rejecting the null 0.94:

$\mathcal{H}_0 : (\beta_2)_2 = 0$	$0.94 * 0.83 = 0.78$
$\mathcal{H}_1 : (\beta_2)_2 \neq 0$	$0.06 * 0.83 = 0.05$

That is, for a 94% test when r and s are the true ones we correctly do not reject $\mathcal{H}_0$ at 78% of the times when we don't know for sure that $(s, p-r-s) = (1, 1)$, because:

$$\Pr\left[W < c_\alpha\right] = \Pr\left[W < c_\alpha \mid (\hat{r}, \hat{s})\,\text{:correct}\right] * \Pr\left[(\hat{r}, \hat{s})\,\text{:correct}\right] = 0.94 * 0.83 = 0.78.$$

On the other hand at 17%, at worst, of the times with wrong choice of cointegrating indices we might conclude that there is not an I(2) trend in X_{2t} component while in reality there is not.

$$\Pr\left[W < c_\alpha\right] = \Pr\left[W < c_\alpha \mid (\hat{r}, \hat{s})\,\text{:correct}\right] * \Pr\left[(\hat{r}, \hat{s})\,\text{:wrong}\right] \leq 1.0 * (1 - 0.83) \leq 0.17$$

In order to investigate the robustness to the choice of the cointegrating rank r, we have carried out some additional simulations. We keep the specification of $p = 5$, $r_0 = 3$ and $s_0 = 1$, while we now assume that the initial choice of $r = 2$, instead of $r_0 = 3$, while keeping $s = 1$. The reason for picking up this specification is that all the other cases of $r = \{0, 1, 4, 5\}$ are not interesting for the I(2) model, because: for $r = 0$ the model is I(2) and exhibits no cointegration in levels; for $r = 1$ there is only one cointegrating vector β and hence we cannot observe the case of cointegration in levels and multicointegration at the same time; for $r = 4$ and because $s = 1$ it holds that there is no β_2 vector, i.e. there is no I(2) variable in the system; finally for $r = 5$ there is no further step for choosing s at all and the model reduces to I(1). Hence with this choice of $\{r = 2, s = 1\}$ the number of I(2) trends in the model are wrongly specified as being two $(p - r - s = 2)$ instead of one $(p - r_0 - s_0 = 1)$ which was the "true" DGP choice.

The DGPs and the sample sizes T of a small scale simulation study, here we chose $n = 1000$, were kept as before with $\rho = 0, 0.5$ and 0.9, while $\varphi = 0$ fixed. The test we calculate is based on the joint hypothesis that both coefficients in the second row of β_2 are equal to zero, that is $(\beta_2)_{21} = 0 = (\beta_2)_{22}$, which can be formulated in the following way: $\mathcal{H}_0 : R_1'\beta_2 R_2 = 0$, $R_1' = (0, 1, 0, 0, 0)$ and $R_2 = I_2$. The test statistic is given in equation (5) above and, in large samples, it follows a χ^2_2 distribution if in fact there were the case of $p - r - s = 2$, which is not. The critical value now used is $\chi^2_2(0.95) = 5.99$. Hence, we derive the following results for the empirical size α^* of the $\mathcal{DGP}_i$, $i = 1, 2, 4$:

Sample Size T	25	50	75	100	200	400
$\mathcal{DGP}_1$	60%	63%	62%	63%	65%	73%
$\mathcal{DGP}_2$	59%	62%	61%	65%	69%	71%
$\mathcal{DGP}_4$	55%	57%	59%	64%	67%	75%

A clear-cut result which comes out of this Monte Carlo simulation experiment is that by setting r too low then $p - r - s$ becomes too large and it is "too difficult" to accept zeros in β_2 overrejecting always. Furthermore, the higher the sample size T, the higher the empirical sizes α^*. One way of interpreting this is as follows: the more information one has about the DGP, the higher the number of rejections regarding two (zero) elements of β_2 where in practice there is only one (zero element). However, for higher sample sizes T there would be also lower the probability of attaining a wrong choice of indices before carrying out the test. On the other hand, also due to the small number of replicatations, it seems that the effect of the increasing ρ from $\rho = 0$ to 0.5 and 0.9 is rather insignificant.

5 Limitations of the test statistic

1. One limitation of the test occurs when one wants to test the restriction $R'\beta_2 = 0$ for all $(\beta_2)_i$. To be more precise consider the simple case (also example in paragraph 3.2) of $p = 3$, $r = s = 1$ and hence β_2 is a 3×1 vector.

It is possible to test that each $\mathcal{H}_{0i} : (\beta_2)_i = 0$, for $i = 1, 2, 3$, but it is not possible to make inference on say $i = 2$, conditioning on the fact that we could not reject the hypothesis for $i = 3$. This is because the asymptotic distribution of β_2 is singular and hence one can only by suitably rotating the distribution under the particular $\mathcal{H}_0$ to carry out hypothesis testing concerning some element of β_2, via the hypothesis $\mathcal{H}_{0i}$, $i = 1, 2, 3$. However, it is not possible to test the intersection of this sub-hypotheses, i.e. it is not possible to test the hypothesis $\mathcal{H}_{02} \cap \mathcal{H}_{03}$, say.

2. A second limitation concerns the specification of R. There is a simple example where the matrix $R'b\rho_\perp$ is of reduced rank, that is when the specification for $R = I_3$ then the hypothesis $\mathcal{H}_0 : R'\beta_2 = q$, where q is a 3×1 vector of constants, is equivalent to $\mathcal{H}_0 : \beta_2 = q$ and the $p \times s$ matrix $R'b\rho_\perp = b\rho_\perp$ has indeed reduced rank $s < p$. That is because the $p \times m$ matrix of restrictions R is now chosen to be a square matrix of $p \times p$ dimensions.

6 Conclusions

The purpose of this paper is to look into the I(2) cointegrated model and construct Wald tests in order to test hypotheses of linear restrictions on the directions β_2, that is, the directions orthogonal to the I(1) cointegrating space. This is important because it is in essence a multivariate alternative to the unit root Dickey-Fuller type univariate tests.

Our results are the following:

The limiting distribution of the Wald test for a hypothesis on the β_2 matrix $\mathcal{H}_0 : R_1'\beta_2 R_2 = 0$ follows a χ^2 with $f = mn$ degrees of freedom, where f equals to the number of imposed restrictions on β_2.

Testing on the β_2 matrix is possible but one has to"handle with care" the specification of the matrices of linear restrictions R_1 and R_2 of dimensions $p \times m$ and $(p - r - s) \times n$, respectively. This is because the full-rank condition of the matrix $R_1'b$ is crucial for the Wald test to make sense.

The Monte Carlo simulation experiment showed that the test performs relatively well, that is we have achieved actual size 6% and decreasing to 5% with increasing sample size. The only catch is when the cointegration indices r and s are not correctly specified. Then the Wald test performs poor.

References

[1] Anderson, T. W. (1951). "Estimating linear restrictions on regression coefficients for multivariate normal distributions". Annals of Mathematical Statistics 22: 327-351.

[2] Banerjee A., Dolado J., Galbraith J.W. and Hendry D. F. (1993). Cointegration, Error-Correction and the Econometric Analysis of Non-Stationary Data, Oxford, Oxford University Press.

[3] Carrion-i-Silvestre, J. L., Sansó-i-Rosselló, A. and Artís Ortuño, M.(2001). "Unit root and stationarity tests' wedding", Economics Letters 70: 1-8.

[4] Charemza, W. W. and Syczewska E. M. (1998). "Joint application of the Dickey-Fuller and KPSS tests", *Economics Letters* 61: 17-21.

[5] Choi, I. and Ahn, B. C. (1998). "Testing the null of stationarity for multiple time series", *Journal of Econometrics* 88: 41-77.

[6] Dickey D. A. and Fuller W. A. (1981). "Likelihood Ratio Statistics for Autoregressive Time Series With a Unit Root", *Econometrica* 49: 1057-1072.

[7] Engle R. F. and Granger C. W. J. (1987). "Cointegration and Error Correction: Representation, Estimation and Testing", *Econometrica* 55: 251-271.

[8] Engle R. F. and Granger C. W. J. (eds.)(1991). *Long-run economic relationships, readings in cointegration*, Oxford, Oxford University Press.

[9] Haldrup, N. (1998). "An Econometric Analysis of I(2) Variables", *Journal of Economic Surveys* 12, 5: 595-650.

[10] Hansen, B. (1995). "Rethinking the univariate approach to unit root tests: How to use covariates to increase power", *Econometric Theory* 11: 1148-1171.

[11] Hornok, A. and Larsson, R. (2000). "The finite sample distribution of the KPSS test", *Econometrics Journal* 3: 108-121.

[12] Johansen, S. (1995). "A Statistical Analysis of Cointegration for I(2) variables", *Econometric Theory* 11: 25-59.

[13] Johansen, S. (1996). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford, Oxford University Press.

[14] Johansen, S. (1997). "Likelihood Analysis of the I(2) Model". *Scandinavian Journal of Statistics* 24: 433-462.

[15] Kwiatkowski D., Phillips P.C.B., Schmidt P. and Shin Y. (1992). "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root", *Journal of Econometrics* 54: 159-178.

[16] Paruolo, P. (1996). "On the Determination of Integration Indices in I(2) Systems". *Journal of Econometrics* 72: 313-56.

[17] Paruolo, P. (1997). "Asymptotic Inference on the Moving Average Impact Matrix in Cointegrated VAR Systems", *Econometric Theory* 13: 79-118.

[18] Paruolo, P. (2000). "Asymptotic Efficiency of the Two Stage Estimator in I(2) Systems", *Econometric Theory* 16: 524-550.

[19] Sims, C., Stock, J. and Watson M. (1990). "Inference in Linear Time Series Models with some Unit Roots". *Econometrica* 58: 113-145.