November 3, 1994

MAKING WALD TESTS WORK FOR COINTEGRATED VAR SYSTEMS

Juan J. Dolado

CEMFI

Casado del Alisal, 5

28014 Madrid

and

Helmut Lütkepohl

Humboldt Universität zu Berlin

Spandauer Strasse 1

10178 Berlin, Germany

September 1994

Abstract

Wald tests of restrictions on the coefficients of vector autoregressive (VAR) processes are known to have nonstandard asymptotic properties for I(1) and cointegrated systems of variables. A simple device is proposed which guarantees that Wald tests have asymptotic χ^2–distributions under general conditions. If the true generation process is a VAR(p) it is proposed to fit a VAR(p+1) to the data and perform a Wald test on the coefficients of the first p lags only. The power properties of the modified tests are studied both analytically and numerically by means of simple illustrative examples.

(*) The research for this paper was supported by funds from Sonderforschungsbereich 373 at Humboldt Universitt Berlin. We are indebted to R. Mestre for invaluable research assistance. We have also benefited from comments by Peter C.B. Phillips, Enrique Sentana and participants at the European Meeting of the Econometric Society, Maastricht 1994.

1 Introduction

Wald tests are standard tools for testing restrictions on the coefficients of vector autoregressive (VAR) processes. Their conceptual simplicity and easy applicability make them attractive for applied work to carry out statistical inference on hypotheses of interest. For instance, a typical example is the test of Granger-causality in the VAR framework where the null hypothesis is formulated as zero restrictions on the coefficients of the lags of a subset of the variables.

Unfortunately, those tests may have nonstandard asymptotic properties if the variables considered in the VAR are integrated or cointegrated. The difficulties in dealing with the levels estimation of such time series are well known and they have been illustrated by means of the general asymptotic theory for inference in multiple linear regressions with integrated processes recently developed by Park and Phillips (1988, 1989), Sims, Stock and Watson (1990) and Toda and Phillips (1993 a,b) among others. As a byproduct of the analysis it has been found that, for instance, Wald tests for Granger-causality are known to result in nonstandard limiting distributions depending on the cointegration properties of the system and possibly on nuisance parameters. This means that to test such hypotheses, the limiting distributions under the null hypothesis need to be simulated in each relevant case, depending on the number of variables, cointegration rank, the number of lags and possibly unknown nuisance parameters. This can be computationally burdensome and may be impossible if the required information is unavailable.

Faced with that problem, a possible solution which has been usually adopted in applied work is to condition the testing procedure on the estimation of unit roots, cointegration rank and cointegrating vectors. Thus, for instance, a first order differenced VAR could be estimated if variables were known to be I(1) with no cointegration, or an error correction model (ECM) could be specified if they were known to be cointegrated. Of course, a priori, it is hardly the case that such a knowledge exists with certainty. Consequently, a pretesting sequence is usually needed before estimating the VAR model in which inference is conducted. Given the low power of those tests and their dependence on nuisance paramenters in finite samples, that testing sequence has typically unknown overall properties, leaving open the possibility of severe distortions in the inference procedure.

To overcome these difficulties, we propose in this paper an extremely simple method which leads to Wald tests with standard asymptotic χ^2–distributions and which avoids possible pretest biases. With this device the tests may be performed directly on the least squares (LS) estimators of the coefficients of the VAR process specified in the levels of the variables. Note that although the variables are allowed to be potentially cointegrated it is not assumed that the cointegration structure of the system under investigation is known. Hence, preliminary unit root tests are not necessary and, therefore, the testing procedure is robust to the integration and cointegration properties of the process.

The idea underlying the procedure is based on the following argument. It is well known that the nonstandard asymptotic properties of the Wald test on the coefficients of cointegrated VAR processes are due the singularity of the asymptotic distribution of the LS estimators. Then, the simple device presented here is to get rid of the singularity by fitting a VAR process whose order exceeds the true order. It can be shown that this device leads to a nonsingular asymptotic distribution of the relevant coefficients, overcoming the problems associated with standard tests and their complicated nonstandard limiting properties. In what follows, the test based upon the estimated coefficients of the augmented VAR process will be denoted as modified Wald test.

In independent work Choi (1993) and Toda & Yamamoto (1993) have proposed a similar device for univariate and multivariate processes, respectively. However, their analysis of the power properties of the modified tests is rather limited. This is an important issue since the modified approach uses the sample inefficiently and thereby may result in severe reductions of power. Thus, in this paper we pay particular attention to analysing those cases in which the inefficiency is likely to be more important. Also, we feel that our proof of the aymptotic distribution of the Wald statistic is more transparent than that of Toda & Yamamoto. From our result it is apparent when it is actually necessary to add an extra lag and when standard asymptotic results make that device unnecessary.

The rest of the paper is planned as follows. First, in Section 2, the simple idea underlying the testing procedure is illustrated by means of a unit root test in an AR(1) model. Section 3 extends the previous analysis to the general VAR case with I(1) variables, since this is the most important one in practice. The power properties of the modified test are analysed in Section 4. Some illustrating Monte Carlo simulations are offered in Section 5. Finally, some conclusions are drawn in Section 6.

2 Testing for a Unit Root with Standard Asymptotics

Let the univariate time series $\{y_t\}$ be generated by the random walk:

$$\Delta y_t = \rho y_{t-1} + \varepsilon_t; \quad \rho = 0, \quad (t = 1, \ldots, T) \tag{1}$$

where $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with mean zero and variance σ_ε^2 such that $E|\varepsilon_t|^{2+\tau} < \infty$ for some $\tau > 0$. Further, the initial condition y_0 is assumed to be any random variable whose distribution is fixed and independent of the sample size T. Suppose that one wishes to test the null hypothesis $\mathrm{H}_0 : \rho = 0$. The standard approach is the use of the Dickey-Fuller test based on $T\hat{\rho}$ or the t-ratio of $\hat{\rho}$. However, if the model is augmented by one lag, as in the Augmented Dickey-Fuller test[1], that is:

$$\Delta y_t = \rho y_{t-1} + \theta y_{t-2} + \varepsilon_t \tag{2}$$

with $\theta = 0$, and (2) is estimated by OLS, then, the estimated ρ is:

$$\hat{\rho} = (y_{-1}' M_2 y_{-1})^{-1} (y_{-1}' M_2 \Delta y)$$

where $y_{-i} = [y_{2-i}, \ldots, y_{T-i}]'$, $i = 1, 2$, M_2 is the projection matrix spanned by y_{-2}, that is $M_2 = I - y_{-2}(y_{-2}' y_{-2}) y_{-2}'$ and $\Delta y = [\Delta y_2, \ldots, \Delta y_T]'$. Using a similar notation for the ε_t's and defining $\varepsilon = [\varepsilon_2, \ldots, \varepsilon_T]'$ we have:

$$y_{-1}' M_2 \Delta y = (y_{-2} + \varepsilon_{-1})' M_2 \varepsilon = \varepsilon_{-1}' M_2 \varepsilon$$

$$y_{-1}' M_2 y_{-1} = (y_{-2} + \varepsilon_{-1})' M_2 (y_{-2} + \varepsilon_{-1}) = \varepsilon_{-1}' M_2 \varepsilon_{-1}$$

Then, applying the convergence results for I(1) variables (see Park and Phillips (1989)) to the previous expressions, with the symbol "$\Rightarrow$" denoting weak convergence in distribution, we get:

$$\begin{aligned} T^{1/2}\hat{\rho} &= \frac{(T^{-1/2}\varepsilon_{-1}'\varepsilon)(T^{-2}y_{-2}'y_{-2}) - (T^{-1}\varepsilon_{-1}'y_{-2})(T^{-1}y_{-2}'\varepsilon)T^{-1/2}}{(T^{-1}\varepsilon_{-1}'\varepsilon_{-1})(T^{-2}y_{-2}'y_{-2}) - (T^{-2}\varepsilon_{-1}'y_{-2})^2 T^{-1}} \\ &= (T^{-1}\varepsilon_{-1}'\varepsilon_{-1})^{-1}(T^{-1/2}\varepsilon_{-1}'\varepsilon) + O_p(T^{-1/2}) \Rightarrow N(0,1) \end{aligned} \tag{3}$$

from the application of standard Central Limit Theorems for i.i.d. random variables. Thus, the modified test based on $T^{1/2}\hat{\rho}$ (or the t-ratio of $\hat{\rho}$) could be used to test for a unit root,

[1] Note that (2) can be reparameterised as $\Delta y_t = \rho \Delta y_{t-1} + (\rho + \theta) y_{t-2} + \varepsilon_t$.

yielding standard asymptotics. Of course, the disadvantage of this procedure is that it uses the sample information inefficiently, leading to a loss in power. As mentioned earlier, it is well known that $T\hat{\rho}$ (or the t-ratio) in the estimation of (1) follows the Dickey-Fuller distribution (see Fuller (1976)), so the modified test is clearly disadvantageous once the true distribution is known[2,3]. However, in more general cases within a VAR framework, the correct nonstandard distributions depend on the size of the system, the number of lags and the cointegration structure (see Table 1 in Toda and Phillips (1993a)) making it difficult or even impossible to use the correct tables in each specific case faced in applied work. Thus, it may be sensible to make a sacrifice in terms of power and gain the correct size in terms of an asymptotic χ^2–distribution. We devote the next section to analyzing this simple idea in a more general setup.

3 Main Result for VAR Systems

Consider the k-dimensional multiple time series generated by a VAR (p) process:

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + \varepsilon_t \tag{4}$$

where $\varepsilon = (\varepsilon_{1t}, \ldots, \varepsilon_{kt})'$ is a zero mean independent white noise process with nonsingular covariance matrix Σ_ε and, for $j = 1, \ldots, k, \quad E|\varepsilon_{jt}|^{2+\tau} < \infty$ for some $\tau > 0$. The order p of the process is assumed to be known or alternatively it may be estimated by some consistent model selection criterion (see, e.g. Paulsen (1984) or Lütkepohl (1991, Chapter 11)) [4].

Let a_p =vec$[A_1 \ldots A_p]$ where vec denotes the vectorization operator that stacks the columns of the argument matrix and suppose that we are interested in testing q independent

[2] Choi (1993), using a similar approach, finds that the t-ratio for $\hat{\rho}$ suffers from low power relative to the Dickey-Fuller test. Moreover, for small samples, it suffers from size distortions. However, it has reasonable properties in constructing confidence intervals for the sum of AR coefficients possibly in the presence of unit roots.

[3] A similar result will hold for $T^{1/2}\hat{\theta}$. However, as is well known, the joint distribution of $T^{1/2}\hat{\rho}$ and $T^{1/2}\hat{\theta}$ has a singular asymptotic covariance matrix; thus joint tests do not follow standard limiting distributions (see Banerjee and Dolado (1988) and Sims, Stock and Watson (1990)).

[4] Of course this involves some pretesting bias, but it is also involved in the standard procedure. Paulsen (1984) and Toda & Yamamoto (1994) prove that if y_t is I(d) (integrated of order d) the usual selection procedures are consistent if $p \geq d$. Thus, if $d = 1$, the lag selection procedures are always valid. In Section 5, we examine the consequences of overestimating the true VAR order.

linear restrictions:

$$H_0 : Ra_p = s \quad vs. \quad H_1 : Ra_p \neq s \tag{5}$$

where R is a known $(q \times k^2 p)$ matrix of rank (rk) q and s is a known $(q \times 1)$ vector. For example, if y_t is partitioned in m and $(k-m)$–dimensional subvectors y_t^1 and y_t^2 and the A_i matrices are partitioned conformably, then y_t^2 does not Granger-cause y_t^1 iff the hypothesis $\mathrm{H}_0 : A_{12,i} = 0$ for $i = 1 \ldots p$ is true. The standard Wald test is as follows. Get an asymptotically normal estimator $\hat{a_p}$ satisfying:

$$T^{1/2}(\hat{a}_p - a_p) \Rightarrow N(0, \Sigma_p)$$

and use the statistic:

$$\lambda_w = T(R\hat{a}_p - s)'(R\hat{\Sigma}_p R')^{-1}(R\hat{a}_p - s) \tag{6}$$

where $\hat{\Sigma}_p$ is some consistent estimator of Σ_p. The Wald statistic λ_w has an asymptotic χ^2–distribution with q degrees of freedom if Σ_p is nonsingular. If the VAR(p) process $\{y_t\}$ is I(0), invertibility holds for the usual estimators (LS or ML) and Wald tests may be applied in the usual manner. However, this is not true if $\{y_t\}$ is I(d), $d > 0$. An exposition of the previous result for I(1) processes can be found in Lütkepohl (1991, Chapter 11) and we summarize the main arguments in what follows.

Consider the EC (error correction) representation of (4):

$$\Delta y_t = D_1 \Delta y_{t-1} + \ldots + D_{p-1} \Delta y_{t-p+1} - \Pi y_{t-p} + \varepsilon_t \tag{7}$$

where $D_i = -(I_k - A_1 - \ldots - A_i)$, $\quad i = 1, \ldots, p-1$, and $\Pi = (I_k - A_i - \ldots - A_p)$ with $\mathrm{rk}(\Pi) = r$. Therefore, Π can be written as the product $\Pi = BC$ where B is $(k \times r)$ and C is $(r \times k)$ with $\mathrm{rk}(B) = \mathrm{rk}(C) = r$.

Defining:

$$\Delta Y = [\Delta y_1 \ldots \Delta y_T], \quad \Delta X_t = \begin{bmatrix} \Delta y_t \\ \vdots \\ \Delta y_{t-p+2} \end{bmatrix}, \quad \Delta X = [\Delta X_0 \ldots \Delta X_{T-1}]$$

$$D = [D_1 \ldots D_{p-1}], \quad Y_{-p} = [Y_{1-p}, \ldots, Y_{T-p}], \quad E = [\varepsilon_1, \ldots, \varepsilon_T]$$

(7) can be rewritten in compact form as:

$$\Delta Y = D \Delta X - BCY_{-p} + E \tag{8}$$

Then, denoting as $\tilde{D}$ and $\tilde{B}\tilde{C}$ the ML estimators of D and BC (see Johansen (1991)) we have:

$$T^{1/2} vec(\left[\tilde{D}, -\tilde{B}\tilde{C}\right] - [D, -BC]) \Rightarrow N(0, \Sigma_{co})$$

with

$$\Sigma_{co} = \begin{pmatrix} I_{k(p-1)} & 0 \\ 0 & C' \end{pmatrix} \Omega^{-1} \begin{pmatrix} I_{k(p-1)} & 0 \\ 0 & C \end{pmatrix} \otimes \Sigma_{\varepsilon} \quad (9)$$

and

$$\Omega = \text{plim } T^{-1} \begin{pmatrix} \Delta X \Delta X' & \Delta X Y_{-p}' C' \\ C Y_{-p} \Delta X' & C Y_{-p} Y_{-p}' C' \end{pmatrix} \quad (10)$$

Note that the dimension of Σ_{co} is $(k^2 p \times k^2 p)$, whereas the dimension of Ω is $[(k(p-1)+r) \times (k(p-1)+r)]$. Thus, the rank of Σ_{co} cannot be greater than $k(k(p-1)+r)$ which is smaller than $k^2 p$ unless $r = k$ (the stationary case). Hence, if $\{y_t\}$ is I(1), Σ_{co} is singular.

It has been shown by various authors (Park and Phillips (1989) and Sims, Stock and Watson (1990)) that the ML estimators $\tilde{A}_i$ of A_i obtained via the EC representation (7) with known cointegrating matrix C and the unrestricted LS estimators $\hat{A}_i$ obtained from (4) have the same asymptotic distribution. Therefore we assume that C is known and that $\tilde{B}$ and $\tilde{D}_i$ are the ML estimators of the remaining parameters in (7). Then it follows that:

$$T^{1/2} vec\left[(\tilde{D}, -\tilde{B}) - (D, -B)\right] \Rightarrow N(0, \Sigma_{EC})$$

where Σ_{EC} is a nonsingular $[k(k(p-1)+r)]$ covariance matrix. Note that $\tilde{B}$ and B disappear for $r = 0$.

From (7) we get the parameters of the VAR representation (4) as

$$\begin{aligned} A_1 &= I_k + D_1 \\ A_i &= D_i - D_{i-1} \qquad (i = 2, \ldots, p-1) \\ A_p &= -D_{p-1} - \Pi \end{aligned}$$

Hence,

$$[A_1, \ldots, A_p] = [B, D_1, \ldots, D_{p-1}]\, W + [I_k, 0 \ldots 0] \quad (11)$$

where

$$W = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & -C \\ I_k & -I_k & 0 & \dots & 0 & 0 \\ 0 & I_k & -I_k & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -I_k & 0 \\ 0 & 0 & 0 & \dots & I_k & -I_k \end{bmatrix} \qquad [(r + k(p-1)) \times kp]$$

Since $\text{rk}(W) = r + k(p-1)$ and

$$T^{1/2}(\hat{a}_p - a_p) \Rightarrow N(0, \Sigma_p)$$

it follows that the $[k^2p \times k^2p]$ covariance matrix $\Sigma_p = (W' \otimes I_k)\Sigma_{EC}(W \otimes I_k)$ has obviously rank $k[k(p-1)+r]$. Thus, it is singular for $r < k$. However, if A_i is omitted from (11), k columns of W are deleted which results in a $[(r + k(p-1)) \times k(p-1)]$ matrix W^* with rank $k(p-1)$.

Denoting by $\hat{a}_p^{p-1}$ the estimator of the remaining $[k^2(p-1)]$ elements of a_p obtained from $\hat{a}_p$, we get:

$$T^{1/2}(\hat{a}_p^{p-1} - a_p^{p-1}) \Rightarrow N\left[0, (W^{*\prime} \otimes I_k)\Sigma_{EC}(W^* \otimes I_k)\right] \equiv N\left[0, \Sigma_p^{p-1}\right] \qquad (12)$$

where the $[k^2(p-1) \times k^2(p-1)]$ covariance matrix Σ_p^{p-1} has now full rank and the Wald test can be implemented in the usual way.

Therefore, the following theorem holds:

Theorem 1

Let the k–dimensional I(1) process $\{y_t\}$ be generated by the VAR(p) process in (7) and let $\hat{A}_i$ $(i = 1, \dots, p)$ be the LS estimators and $\hat{a}_p^{p-1}$ the $[k^2(p-1)]$–dimensional vector consisting of the $k^2(p-1)$ elements of $\hat{a}_p = \text{vec}\left[\hat{A}_{1p}, \dots, \hat{A}_p\right]$ that are obtained by deleting one of the $\hat{A}_i$ matrices. Then:

$$T^{1/2}(\hat{a}_p^{p-1} - a_p^{p-1}) \Rightarrow N(0, \Sigma_p^{p-1})$$

where the $[k^2(p-1) \times k^2(p-1)]$ covariance matrix Σ_p^{p-1} is nonsingular. Moreover, given a consistent estimator $\hat{\Sigma}_p^{p-1}$, the Wald test of the null hypothesis $\text{H}_0 : Ra_p^{p-1} = s$,

$$\lambda_w = T(R\hat{a}_p^{p-1} - s)'(R\hat{\Sigma}_p^{p-1}R')^{-1}(R\hat{a}_p^{p-1} - s) \qquad (13)$$

has an asymptotic $\chi^2(q)$–distribution.□

The theorem implies that whenever the elements in at least one of the complete coefficient matrices A_i are not restricted under H_0, the Wald statistic has its usual χ^2–distribution. Thus, if elements from all A_i, $i = 1, \ldots, p$, are involved in the restrictions as, for instance, in noncausality hypotheses, we may just add an extra lag in estimating the parameters of the process and thereby ensure standard asymptotics for the Wald test. Of course, if the true DGP is a VAR(p) process, then a VAR($p+1$) with $A_{p+1} = 0$ is also an appropriate model. Using the previous notation, in this case the modified Wald test will be based on the estimator $\hat{a}^p_{p+1}$, namely the first $[k^2p]$ elements of $vec\left[\hat{A}_1 \ldots \hat{A}_{p+1}\right]$.

Notice that for this procedure to work it is obviously neither necessary to know the cointegration properties of the system nor the order of integration of the variables. Thus, if there is uncertainty whether the variables are I(1) or I(0), one may simply add the extra lag and then perform the test to make sure to be on the safe side. Of course there will be a loss of power, given that in the nonstationary case some VAR coefficients or linear combinations of them can be estimated more effectively with larger rate of convergence than in the I(0) case. Nevertheless, one may argue about the acceptability of the resulting loss in power. In general, we will expect the loss in power to be of little relevance if the true order p is large and the dimension k is small or moderate, since in this case the relative reduction in the estimation precision due to one extra VAR coefficient matrix will be small. However, if the true order is small and k is large, an extra lag of all variables may lead to a sizeable decline in the power of the modified Wald test. To get a feeling for the trade-off between size and power of the proposed procedure, a small Monte-Carlo analysis is carried out in Section 5.

It may be worth noting that the theorem remains valid if an intercept term or other deterministic terms, like seasonal dummies or time trends, are included in the VAR model. This follows from the results in Park and Phillips (1989) and Sims, Stock & Watson (1990) who demonstrate that the asymptotic properties of the VAR coefficients are essentially unaffected by such terms. Moreover, a similar result can be obtained for VAR systems with I(d) variables where $d > 1$. In that case, d coefficient matrices A_i must be unrestricted under H_0. Alternatively, d lags must be added if all parameter matrices of the original process are restricted. This is also a consequence of results given in Sims, Stock & Watson (1990).

4 Power Properties

To analyse the power properties of the modified Wald test, we first notice that it is consistent. Suppose, for instance, that the alternative hypothesis is:

$$H_1 : Ra_p = s + \delta, \quad \delta \neq 0 \tag{14}$$

Then, under H_1, we have:

$$\begin{aligned}
\lambda_w &= T(R\hat{a}^p_{p+1} - s)'(R\hat{\Sigma}^p_{p+1}R')^{-1}(R\hat{a}^p_{p+1} - s) \\
&= T(R\hat{a}^p_{p+1} - s - \delta)'(R\hat{\Sigma}^p_{p+1}R')^{-1}(R\hat{a}^p_{p+1} - s - \delta) \\
&\quad + T\delta'(R\hat{\Sigma}^p_{p+1}R')^{-1}\delta + 2T^{1/2}\delta'(R\hat{\Sigma}^p_{p+1}R')^{-1}T^{1/2}(R\hat{a}^p_{p+1} - s - \delta) \\
&= O_p(1) + O_p(T) + O_p(T^{1/2}) = O_p(T)
\end{aligned}$$

Thus, for any positive number M, $\Pr[\lambda_w > M] \to 1$ as $T \uparrow \infty$, i.e. the test is consistent.

To study the local power properties, consider the local alternative:

$$H_1 : Ra_p = s + T^{-1/2}\delta \quad \text{for fixed } \delta \tag{15}$$

Then, $\lambda_w \Rightarrow \chi^2(q, \mu^2)$, i.e. a non-central χ^2–distribution with non-centrality parameter given by:

$$\mu^2 = \delta'(R\Sigma^p_{p+1}R')^{-1}\delta \tag{16}$$

Following Kendall and Stuart (1961, Chapter 24) (see Mizon and Hendry (1980)), the first two moments of the non-central χ^2–distribution can be approximated by a central χ^2 (with different degrees of freedom). More precisely:

$$\chi^2(q, \mu^2) \simeq h\chi^2(m, 0) \tag{17}$$

where $h = (q + 2\mu^2)/(q + \mu^2)$ and $m = (q + \mu^2)^2/(q + 2\mu^2)$. Consequently, for any M, the approximate and large sample power P^* of λ_w is given by:

$$P^* = \Pr[\lambda_w > M] \simeq Pr\left[\chi^2\left(\frac{(q + \mu^2)^2}{q + 2\mu^2}\right) > M\frac{q + \mu^2}{q + 2\mu^2}\right] \tag{18}$$

Note that if H_0 is true, $\mu = 0$, so that

$$Pr[\lambda_w > M] \to Pr\left[\chi^2(q) > M\right]$$

confirming the appropriate nominal and large sample size of the test. Moreover, since $\mu^2 = T(Ra^p_{p+1} - s)'(R\Sigma^p_{p+1}R')^{-1}(Ra^p_{p+1} - s)$, $\mu^2 \uparrow \infty$ with T so that $h \uparrow 2$ and $m \uparrow \infty$, i.e. $P^* \uparrow 1$. Similarly, if δ takes higher values, for fixed T, μ^2 and m increase and so does the power. To summarise, equation (18) offers an analytical formula to examine the effects of the factors $(a^p_{p+1}, \delta, T, k) = \psi$ on the large sample power of λ_w to reject H_0 against the sequence (15). We devote the next section to analysing some of those effects in finite samples.

5 A Small Monte-Carlo Analysis

To illustrate the previous discussion on the use of Granger-causality tests in VAR systems with I(1) variables, we have generated 1000 replications of the bivariate VAR(2) cointegrated process $y_t = (y_{1t}, y_{2t})'$ given by:

$$\Delta y_t = \begin{bmatrix} -\beta & \beta \\ 0 & 0 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0.5 & 0.3 \\ T^{-1/2}\delta & 0.5 \end{bmatrix} \Delta y_{t-1} + \varepsilon_t \qquad (19)$$

where $\varepsilon_t \sim N(0, I_2)$. The process has cointegration rank $r = 1$ $(= 0)$ iff $\beta \neq 0$ $(\beta = 0)$. If $\delta = 0$, y_{1t} is Granger noncausal for y_{2t} and if $\delta \neq 0$, y_{1t} causes y_{2t}. Therefore, $\delta = 0$ is used to study the size of the test and $\delta = 1, 2$ are used to analyse power.

For each time series 50 presample values are generated with zero initial conditions, taking net sample sizes of $T = 50$, 100 and 200. The fitted processes include a constant term, that is the model $y_t = \nu + A_1 y_{t-1} + A_2 y_{t-2} + \varepsilon_t$ is fitted for the standard procedure and an analogous VAR(3) process for the modified procedure.

Table 1(a) presents the relative rejection frequencies for tests with asymptotic 5% significance level of a $\chi^2(2)$–distribution when $\beta = 1$, i.e. there is cointegration. In this case it is not difficult to see that the standard Wald test has an aymptotic $\chi^2(2)$– distribution under H_0. Thus, this case is favourable for the standard test. To assess whether the rejection rates are significantly different from the theoretical rate the following 95% confidence interval is useful: $[3.6\%, 6.4\%]$. The test rejects slightly too often for small and moderate samples ($T =$ 50 and 100)[5]. With respect to the power, it is clear that it is higher when the true VAR(2) process is estimated. In other words, the modified test wastes information by estimating

[5]This is in agreement with the slow convergence of the standard t-ratio in the univariate case analysed by Choi (1993).

extra coefficients. However, the assumption that the true order is known might be too optimistic, so in Table 1(b) we pretend that the data are generated by a VAR(3) process and repeat the tests which now have asymptotic $\chi^2(3)$–null–distributions. The corresponding modified Wald test is obtained from a VAR(4) process. In this case the powers of the two tests are found to be almost identical. Thus, even under this minor deviation from the ideal conditions for the standard test, the loss in efficiency for the modified procedure almost disappears.

Table 1(c) reports the size and power for $\beta = 0$, i.e. the case where there is no cointegration. In practice, the cointegration rank is unknown and has to be determined in a pretesting procedure. In this case the standard test does not have an asymptotic $\chi^2(2)$–distribution under the null hypothesis. Hence, this example illustrates the consequences of using the standard Wald test incorrectly with a 5% critical value from a $\chi^2(2)$–distribution. As in the first example, VAR(2) and VAR(3) processes are fitted to the variables in levels. We find that the standard test rejects too often under H_0 even for large samples (see Ohanian (1988), and Toda and Phillips (1993a)) while the modified test converges to its correct nominal size for T= 200. Hence, the standard test is clearly misleading while the modified test maintains roughly the same properties in large samples as for the cointegrated process (19) with $\beta \neq 0$. Consequently, in terms of size, the modified procedure is clearly preferable if the cointegration rank is unknown[6].

Next, in order to check the loss in power of the modified Wald test for given values of the dimension k of the process and the true order p of the VAR, we carry out two types of experiments. First, to analyse the effect of enlarging p for given k, the DGP(19) is generalised to:

$$\Delta y_t = \begin{bmatrix} -\beta & \beta \\ 0 & 0 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0.5 & 0.3 \\ T^{-1/2}\delta & 0.5 \end{bmatrix} \Delta y_{t-p+1} + \varepsilon_t \qquad (20)$$

where $\varepsilon \sim N(0, I_2)$, $\beta = 1$, $\delta = 1$ and $p = (2, 3, \ldots, 6)$. The empirical powers were calculated out of 1000 replications for a net sample size of 100 and are reported in Table 2. The null hypothesis is again $H_0 : \delta = 0$. The figures without parentheses and those with parentheses denote the relative inefficiency (measured by the ratio of powers) of the modified with respect

[6]Note that the power of the standard test in this case is upwards biased since it has a larger size than the nominal 5% level. Computation of the size adjusted power for T=200 and $\delta = 1, 2$ yields rejection frequencies 23.4% and 61.3% for the standard test and 18.7% and 54.6% for the modified test, respectively.

to the standard Wald test and the absolute empirical power of the latter, respectively. In agreement with the conjecture offered in Section 3 we find that, for $k = 2$, the relative inefficiency of the modified test, based upon the estimation of a VAR($p+1$) rather than a VAR(p), decreases with the true order p. For instance, we find that, for $p > 3$, the loss in power becomes less than 10%. Hence, if a VAR system has a small number of variables with a long lag length, as is often the case in practice, then the inefficiency caused by adding a few more lags would be relatively small.

Second, to examine the effect of enlarging k for given p, the DGP in (19) is generalised to:

$$\Delta y_t = \begin{bmatrix} -\beta & \beta & \beta & \dots & \beta \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0.5 & a_{12} & a_{13} & \dots & a_{1k} \\ T^{-1/2}\delta & 0.5 & a_{23} & \dots & a_{2k} \\ 0 & 0 & 0.5 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 0.5 \end{bmatrix} \Delta y_{t-1} + \varepsilon_t \qquad (21)$$

where now $y_t = (y_{1t}, y_{2t}, \dots, y_{kt})'$, $k = (2, 3, \dots, 6)$, $\varepsilon_t \sim N(0, I_k)$, $\beta = 1$, $\delta = 1$, $a_{1l} = 0.3/(l-1)$ $(l = 2, \dots, k)$ and $a_{2l} = 0.3/(l-2)$ $(l = 3, \dots, k)$. Having generated 1000 replications for $T = 100$, the numbers in Table 3 have the same meaning as in Table 2, with the null hypothesis being again $\mathrm{H}_0 : \delta = 0$. We conclude from this experiment that if the VAR system has many variables and the true lag length is short (p=2 in this case), then the inefficiency caused by adding even one extra lag would be relatively big. For instance, for $k = 6$, the modified Wald test has only a little more than one-fourth of the power of the standard test. However, given that the absolute power of the latter is around 20%, the absolute loss of power is not that large after all.

Finally, in order to make analytical comparisons of the relative power properties of both tests by means of the approximate power function derived in (18), we have used a simpler illustrative bivariate DGP based upon a VAR(1) system with I(1) variables. In this way, the analysis becomes tractable and it can be used to shed light on the effect of some of the incidental parameters of the DGP. In particular, we focus attention on the following set of parameters $\psi = [\beta, \delta, V(\varepsilon_{1t}), V(\varepsilon_{2t}), Cov(\varepsilon_{1t}, \varepsilon_{2t})]$.

We consider the following DGP:

$$\Delta y_t = \begin{bmatrix} -\beta & \beta \\ \alpha & -\alpha \end{bmatrix} y_{t-1} + \varepsilon_t; \quad \varepsilon_t \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \begin{pmatrix} 1 & \theta \\ \theta & \lambda \end{pmatrix}\right] \tag{22}$$

with $\alpha = T^{-1/2}\delta$. As in the DGP's considered above, $\delta = 0$ corresponds to the case where y_{1t} is Granger noncausal for y_{2t}.

Given the simplicity of the DGP, it is easy to compute the non-centrality parameter μ^2 in the VAR(1) system (standard procedure) which is given by (see Appendix):

$$\mu_1^2 = \delta^2(1 + \lambda - 2\theta)/\lambda(1 - \rho^2) \tag{23}$$

where $\rho = 1 - \alpha - \beta$. Similarly, in the VAR(2) model (modified procedure), the corresponding expression is (see Appendix):

$$\mu_2^2 = \delta^2(1 - \theta^2/\lambda)/\lambda \tag{24}$$

For $|\rho| < 1$, it is easy to show that $\mu_1^2 > \mu_2^2$, as expected[7]. Moreover, since h and m are increasing in μ^2 this means that the power of the standard test is larger than the power of the modified test. Note, also, that for $\beta = 0$ and $\alpha = 0$, i.e. $r = 1$, μ_1^2 is not defined, reflecting the non-standard distribution of the standard Wald test in the absence of cointegration. Nevertheless, the modified test has a non-centrality parameter which does not depend upon ρ, reflecting that it has the correct size under the null hypothesis and that its limiting distribution is a non-central χ^2 even when cointegration does not exist.

To check how well the analytical approximate large sample power compares to the empirical rejection frequencies, 2000 replications were conducted for $T = 100$ of the following four experiments, (parameter configurations in parentheses): Experiment 1 ($\lambda = 1, \delta = 1, \beta = 1$); Experiment 2 ($\lambda = 0.2, \delta = 1, \beta = 1$); Experiment 3 ($\lambda = 1, \delta = 2, \beta = 1$); and Experiment 4 ($\lambda = 1, \delta = 1, \beta = 0.1$). For each experiment, the correlation between ε_{1t} and ε_{2t} (corr $= \theta/\lambda^{1/2}$) takes three values, i.e. corr $= (0.0, 0.5$ and $-0.5)$. This is done to control for the dependence of the power functions on the covariance θ as exemplified by expressions (23) and (24). Thus, Experiment 1 is the base experiment; Experiment 2 examines the effect of a reduction in λ with respect to the base experiment. Similarly, Experiments 3 and 4 examine the effect of an increase in δ and a decrease in β, respectively.

[7]Since $\mu_1^2 > \delta_1^2 > \delta^2(1 + \lambda - 2\theta)/\lambda$ and $(1 + \lambda - 2\theta) \geq 1 - \theta^2/\lambda$. Thus, $\mu_1^2 > \mu_2^2$.

Table 4 reports the results of the previous set of experiments in terms of analytical (P^*) and empirical (P) rejection frequencies, together with the values of the proportion factor ($h^{-1} = (q+\mu^2)/(q+2\mu^2)$), the number of degrees of freedom (m) and the relative power (R) computed in terms of the ratio of empirical rejections. To compute the analytical power, the degrees of freedom of the approximate central χ^2–distributions were proxied by the closest integer part of m.

Several results are worth mentioning. First, the analytical and empirical rejection frequencies yield broadly similar results with their differences never exceeding 10 percentage points in the least favourable cases. Thus, the asymptotic local power analysis proves to be useful in interpreting the relative power outcomes in finite samples.

Second, within each experiment, the power of the standard test is highest for corr $= -0.5$ and lowest for corr $= 0.5$, reflecting the fact that μ_1^2 decreases with increasing correlation between the error terms. At the same time, the power of the modified Wald test does not depend on the sign of the correlation coefficient, as shown in (24). Therefore, the more negative is the correlation coefficient the larger will be the relative inefficiency of the modified test, i.e. the smaller is R. The intuition behind this result lies in the form of the cointegrating vector in DGP (22), i.e. (1,-1). This implies that the variance of deviations from the cointegrating relationship, $(y_{1t} - y_{2t})$, depends upon $V(\varepsilon_{1t} - \varepsilon_{2t})$ (see Appendix). Thus if $\theta < 0$, $V(y_{1t} - y_{2t})$ will increase. Since in the standard Wald test the null hypothesis $\alpha = 0$ can be solely expressed as a restriction on the coefficient of $(y_{1t-1} - y_{2t-1})$, the higher the variance of that variable, the more efficiently the coefficient will be estimated and, hence, the larger will be the power of the test. Once we condition on further lags of y_{1t} and y_{2t}, as in the modified procedure, that direct effect disappears. This is reflected by the dependence of μ_2^2 on θ^2 rather than θ. Had the cointegrating vector been (1,1), the "residual" $(y_{1t} + y_{2t})$ would have a variance which depends on $V(\varepsilon_{1t} + \varepsilon_{2t})$. Therefore, in this case, the opposite result holds, that is, $\theta > 0$ will increase μ_1^2 and the power of the standard test.

Third, the powers of the two tests decrease with λ, reflecting the fact that a lower variance of the error term in the equation of interest results in a higher power. Fourth, the powers of the two tests obviously increase towards unity as δ increases. Lastly, the lower is β, namely, the less cointegrated are the variables and the higher is the variance of $(y_{1t} - y_{2t})$, the larger is the power of the standard test relative to the power of the modified test, since μ_2^2 does not depend on β.

Overall, we conclude that the loss in power entailed by the use of the modified procedure, for the particular DGP under study, will be larger the more negative is the correlation coefficient between the error terms and the less cointegrated are the variables. Note, however, that low values of β could lead to potential size distortions (over-rejections) of the standard test and thereby exaggerate the loss of power of the modified test (see footnote 4)[8].

6 Concluding Remarks

In this paper a device is proposed that guarantees standard χ^2 asymptotics for Wald tests performed on the coefficients of cointegrated VAR processes with I(1) variables if at least one coefficient matrix is unrestricted under the null hypothesis. By the same token, if all the matrices are restricted, it is shown that adding one extra lag to the process and concentrating on the original set of coefficients results in Wald tests with standard asymptotic distributions. This leads to a number of interesting implications which stem from the possibility of expressing null hypotheses as restrictions on coefficients of stationary variables (see Sims, Stock and Watson (1990)). First, for I(1) variables (with or without cointegration), if a VAR(2) is fitted, all t-ratios are asymptotically normal. Second, a VAR(p) can be tested against a VAR($p+1$), $p \geq 1$, with a standard Wald test. Third, if the true DGP is a VAR(p) and a VAR($p+1$) is fitted, standard Wald tests can be applied to the first p VAR coefficient matrices. These results do not depend on the presence of deterministic terms in the DGP as long as the restrictions are confined to the VAR coefficients. Furthermore, nonlinear restrictions can be tested in the same way.

As regards the reduction in power entailed by the inefficient use of the sample in the modified procedure, our Monte Carlo simulations show that it will be more severe in high dimensional VARs with a small true lag length. Moreover, in bivariate systems, possibly cointegrated, we find that a negative correlation between the error terms in the equations seems to cause larger inefficiency when the cointegrating relationship is of the form (1,-1), while a positive correlation causes larger inefficiency if it is of the form (1,1).

[8]As in DGP(19), we pretended that the data were generated by a VAR(2) and repeated the tests with $\chi^2(2)$ critical values in VAR(2) and VAR(3) models. As in the previous case, we found that the relative inefficiency in terms of power was minor, with $R \leq 0.90$ in all cases.

However, we find that when there are serious doubts about the series being cointegrated, the size distortions of the modified procedure are much smaller in finite samples. Thus, the power disadvantage is likely to be outweighed by the ease of applicability of the modified procedure. Finally, it is important to note that the previous results could be generalised to VAR systems with I(d) variables, $d > 1$. In that case, the modified procedure involves adding d extra lags.

APPENDIX

Given the DGP(22), the univariate representations of y_{1t} and y_{2t} are given by:

$$\Delta y_{1t} = (\varepsilon_{1t} - (1-\alpha)\varepsilon_{1t-1} + \beta\varepsilon_{2t-1})/1 - \rho L \qquad (A.1)$$

$$\Delta y_{2t} = (\varepsilon_{2t} - (1-\beta)\varepsilon_{2t-1} + \alpha\varepsilon_{1t-1})/1 - \rho L \qquad (A.2)$$

while the deviation from the cointegrating relationship, u_t, follows the process:

$$u_t = (y_{1t} - y_{2t}) = (\varepsilon_{1t} - \varepsilon_{2t})/1 - \rho L \qquad (A.3)$$

where $\rho = 1 - \beta - \alpha$, such that $|\rho| < 1$. Here L is the lag operator.

Then, the standard test is based upon the regression model:

$$y_t = \hat{A}_1 y_{t-1} + \hat{e}_t; \qquad E(e_t e_t') = \Sigma_e$$

or

$$\Delta y_t = \hat{B} y_{t-1} + \hat{e}_t.$$

In particular, the second equation of the system, to which the noncausality test is applied, can be written as:

$$\Delta y_{2t} = b_{21} y_{1t-1} + b_{22} y_{2t-1} + e_{2t} \qquad (A.4)$$

Using (A.3), (A.4) can be reparameterised as:

$$\Delta y_{2t} = b_{21} u_{t-1} + (b_{21} + b_{22}) y_{2t-1} + e_{2t} \qquad (A.5)$$

That is, the reparameterisation makes it possible to express the parameter of interest, b_{21}, as a coefficient on an I(0) variable. Obviously, estimation of (A.4) by OLS yields consistent estimators of α and λ in DGP(22), such that plim $\hat{b}_{21} = \alpha$ and plim $\hat{\sigma}^2_{e2} = \lambda$. Moreover, since u_{t-1} is asymptotically orthogonal to y_{2t-1}, (being I(0) and I(1) variables, respectively), the asymptotic variance of $\hat{b}_{21}$, $V(\hat{b}_{21})$ depends only on $E(u_2^2)$. Indeed, $V(\hat{b}_{21}) = \lambda/E(u_t^2) = \lambda(1-\rho^2)/(1+\lambda-2\theta)$. Thus, the non-centrality parameter of the standard test is given by:

$$\mu_1^2 = \delta^2/V(\hat{b}_{21}) = \delta^2\left[1 + \lambda - 2\theta\right]/\lambda(1-\rho^2) \qquad (A.6)$$

In the modified Wald test, the regression model is:

$$y_t = \hat{A}_1 y_{t-1} + \hat{A}_2 y_{t-2} + \hat{\varepsilon}_t$$

or

$$\Delta y_t = \hat{B}\Delta y_{t-1} + \hat{C}y_{t-2} + \hat{\varepsilon}_t$$

In particular, the second equation of the system will be:

$$\Delta y_{2t} = b_{21}y_{1t-1} + b_{22}y_{2t-1} + c_{21}y_{1t-2} + c_{22}y_{2t-2} + e_{2t} \tag{A.7}$$

which can be reparameterised as:

$$\begin{aligned}\Delta y_{2t} &= b_{21}u_{t-1} + c_{21}u_{t-2} + (b_{21} + b_{22})\Delta y_{2t-1} \\ &\quad +(b_{21} + b_{22} + c_{21} + c_{22})y_{2t-2} + e_{2t}\end{aligned} \tag{A.8}$$

Using similar arguments as in the VAR(1) case, plim $\hat{b}_{21} = \alpha$, plim $\hat{\sigma}^2_{e2} = \lambda$ and the I(0) regressors $\{u_{t-1}, u_{t-2}, \Delta y_{2t-1}\}$ are asymptotically orthogonal to y_{2t-2}.

Thus, in this case the asymptotic variance of $\hat{b}_{21}$, $V(\hat{b}_{21})$ is given by the (1,1) element of:

$$\lambda \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \cdot & \gamma_{22} & \gamma_{23} \\ \cdot & \cdot & \gamma_{33} \end{pmatrix}^{-1}$$

where $\{\gamma_{ij}\}$ is the covariance matrix of $\{u_t, u_{t-1}, \Delta y_{2t}\}$. From (A.1) - (A.3), we get:

$$\begin{aligned}
\gamma_{11} &= E(u_t^2) = \gamma_{22} = (1 + \lambda - 2\theta)/(1 - \rho^2) \\
\gamma_{12} &= E\left[u_t, u_{t-1}\right] = \rho\gamma_{11} \\
\gamma_{13} &= E\left[u_t, \Delta y_{2t}\right] = \left[\theta - (1-\beta)\theta\rho + \alpha\rho - \lambda + (1-\beta)\rho\lambda - \alpha\theta\rho\right]/(1-\rho^2) \\
\gamma_{23} &= E\left[u_{t-1}, \Delta y_{2t}\right] = \left[\theta\rho - (1-\beta)\theta + \alpha - \lambda\rho + (1-\beta)\lambda - \alpha\theta\right]/(1-\rho^2) \\
\gamma_{33} &= E\left[\Delta y_{2t}^2\right] = \left[(1 + (1-\beta)^2)\lambda + \alpha^2 - 2\alpha(1-\beta)\theta - 2\rho(1-\beta)\lambda + 2\rho\alpha\theta\right]/(1-\rho^2)
\end{aligned}$$

From these results we can obtain the much simplified expression $V(\hat{b}_{21}) = \lambda^2/(\lambda - \theta^2)$. Thus, the non-centrality parameter of the modified Wald test, is given by:

$$\mu_2^2 = \delta^2/V(\hat{b}_{21}) = \delta^2(1 - \theta^2/\lambda)/\lambda \tag{A.9}$$

Table 1

Relative Rejection Frequencies (%)

(a) ($\beta = 1$, AM = VAR(2), 5% CV = 5.99)

		Standart Test δ			Modified Test δ		
		0	1	2	0	1	2
	50	7.6	41.4	89.7	8.7	20.5	55.8
T	100	7.1	40.9	91.9	7.1	19.5	58.5
	200	5.8	40.4	93.8	4.7	19.0	57.6

(b) ($\beta = 1$, AM = VAR(3), 5% CV = 7.81)

		Standart Test δ			Modified Test δ		
		0	1	2	0	1	2
	50	8.7	27.2	72.8	10.8	29.2	72.0
T	100	5.8	24.8	73.6	6.7	26.0	72.6
	200	5.4	23.9	72.9	5.1	23.1	71.6

(c) ($\beta = 0$, AM = VAR(2), 5% CV = 5.99)

		Standart Test δ			Modified Test δ		
		0	1	2	0	1	2
	50	21.5	36.1	70.4	11.5	24.0	57.8
T	100	16.7	36.2	68.8	8.4	22.9	58.0
	200	16.7	32.1	68.2	6.2	19.7	56.1
			(23.4)	(61.3)		(18.7)	(54.6)

Note: AM denotes assumed model; the 5% C.V. in parts (a) and (c) correspond a $\chi^2(2)$–distribution while that in part (b) corresponds to a $\chi^2(3)$–distribution; Figures in parenthesis in block (c) correspond to size-adjusted powers; Number of replications = 1000; Computations performed using MATLAB.

Table 2

Rejection Frequencies (%)

(DGB:(20) $T = 100, \ \delta = 1, \ k = 2$)

Lag p/Power	2	3	4	5	6
	0.47	0.73	0.91	0.95	0.98
	(40.9)	(35.6)	(33.2)	(30.5)	(28.2)

Note: Numbers without parentheses and those within parentheses denote the relative power of the modified to the standard Wald test and the empirical power of the latter, respectively.

Table 3

Rejection Frequencies (%)

(DGB:(21) $T = 100, \ \delta = 1, \ p = 2$)

Dimension k/Power	2	3	4	5	6
	0.47	0.43	0.38	0.34	0.28
	(40.9)	(37.1)	(34.3)	(29.8)	(23.2)

Note: See Note in Table 2.

Table 4

Analytical and Empirical Power (%)

(DGP:(22) T=100)

Standard Test [$VAR(1)$]					Modified Test [$VAR(2)$]				
Experiment 1 [$\lambda = 1, \delta = 1, \beta = 1$]									
Corr	h^1	m	P^*	P	h^1	m	P^*	P	R
0.0	0.60	1.81	32.60	25.72	0.67	1.33	13.22	17.15	0.67
0.5	0.67	1.34	13.22	15.10	0.70	1.22	13.08	14.35	0.95
-0.5	0.57	2.30	34.50	34.90	0.70	1.22	13.08	14.35	0.41
Experiment 2 [$\lambda = 0.2, \delta = 1, \beta = 1$]									
0.0	0.53	3.80	72.36	64.70	0.54	3.27	52.66	58.55	0.90
0.5	0.56	2.68	52.46	48.32	0.56	2.65	52.44	48.00	0.99
-0.5	0.53	4.92	83.76	75.10	0.56	2.65	52.44	48.70	0.65
Experiment 3 [$\lambda = 1, \delta = 2, \beta = 1$]									
0.0	0.53	4.93	83.76	76.22	0.56	2.78	52.36	50.70	0.66
0.5	0.55	2.86	52.43	49.43	0.57	2.29	34.50	39.38	0.80
-0.5	0.52	7.01	96.02	87.75	0.57	2.29	34.50	39.75	0.39
Experiment 4 [$\lambda = 1, \delta = 1, \beta = 0.1$]									
0.0	0.54	3.54	72.66	66.34	0.67	1.33	13.22	17.35	0.26
0.5	0.58	2.18	34.53	34.67	0.70	1.22	13.08	14.42	0.41
-0.5	0.53	4.93	83.76	77.43	0.70	1.22	13.08	14.38	0.19

Note: P^* and P are the analytical and empirical rejection frequencies, respectively; R is the ratio between the empirical powers of the modified and standard tests.

References

Banerjee, A. and J. Dolado (1988) "Tests of the Life Cycle - Permanent Income Hypothesis in the Presence of Random Walks: Asymptotic Theory and Small-Sample Interpretations", *Oxford Economic Papers* 40, 610-633.

Choi, I. (1993) "Asymptotic Normality of the Least-Squares Estimates for Higher Order Autoregressive Integrated Processes with some Applications", *Econometric Theory* 9, 263-282.

Fuller, W. (1976) *Introduction to Statistical Time Series*, John Wiley, New York.

Johansen, S. (1991) "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models", *Econometrica* 59, 1551-1580.

Kendall, M. and A. Stuart (1961) *The Advanced Theory of Statistics*, vol. 2 (New York: Charles Griffin)

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

Mizon, G. and D.F. Hendry (1980) "An Empirical Application and Monte Carlo Analysis of Tests of Dynamic Specification", *Review of Economic Studies* 47, 21-45.

Ohanian, L. (1988) "The Spurious Effects of Unit Roots on Vector Autoregressions: A Monte Carlo Analysis", *Journal of Econometrics* 39, 251-266.

Park, J. and P.C.B. Phillips (1988) "Statistical Inference in Regressions with Integrated Processes: Part 1", *Econometric Theory* 4, 468-497.

Park, J. and P.C.B. Phillips (1989) "Statistical Inference in Regressions with Integrated Processes: Part 2", *Econometric Theory* 5, 95-131.

Paulsen, J. (1984), "Order Determination of Multivariate Autoregressive Time Series with Unit Roots", *Journal of Time Series Analysis* 5, 115-127.

Sims, C., Stock, J. and M. Watson (1990) "Inference in Linear Time Series Models with Some Unit Roots", *Econometrica* 58, 113-144.

Toda, H. and P.C.B. Phillips (1993a) "The Spurious Effect of Unit Roots on Vector Autoregressions: An Analytical Study", *Journal of Econometrics* 59, 229-256.

Toda, H. and P.C.B. Phillips (1993b) "Vector Autoregressions and Causality", *Econometrica* 61, 1367-1394.

Toda, H. and T. Yamamoto (1993) "Statistical Inference in Vector Autoregressions with Possibly Integrated Processes" University of Tsubuka (mimeo).