COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale University
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1001

VECTOR AUTOREGRESSION AND CAUSALITY:
A THEORETICAL OVERVIEW AND SIMULATION STUDY

by

Hiro Y. Toda and Peter C. B. Phillips

October 1991

VECTOR AUTOREGRESSION AND CAUSALITY: A THEORETICAL OVERVIEW AND SIMULATION STUDY*

by

Hiro Y. Toda

and

Peter C. B. Phillips

Cowles Foundation for Research in Economics

Yale University

* The authors would like to thank participants at the Australasian Meetings of the Econometric Society and seminar participants at the University of Western Australia and Curtin University of Technology for helpful comments on an earlier draft. The first author would like to acknowledge a travel grant from the University of Western Australia.

0 Abstract

This paper provides a theoretical overview of Wald tests for Granger causality in levels vector autoregressions (VAR's) and Johansen-type error correction models (ECM's). The theory is based on results in Toda and Phillips (1991) and allows for stochastic and deterministic trends as well as arbitrary degrees of cointegration. For VAR models the results for inference are not encouraging. The limit theory typically involves nonstandard distributions and nuisance parameters, and there is no sound statistical basis for testing causality in such a framework. Granger causality tests in ECM's also suffer from nuisance parameter dependencies asymptotically and nonstandard limit theory. But, in spite of these difficulties Johansen-type ECM's do offer a sound basis for empirical testing of the rank of the cointegration space and the rank of key submatrices that influence the asymptotics. In consequence, we recommend some operational procedures for conducting Granger causality tests in the important practical case of testing the causal effects of one variable on another group of variables and vice versa. This paper also investigates the sampling properties of these testing procedures through simulation exercises. Three sequential causality tests in ECM's are compared with conventional causality tests based on VAR's in levels and in differences. It is found that the sequential tests work reasonably well at least in large samples and that they generally outperform the conventional VAR causality tests.

1 Introduction

This paper provides a theoretical overview of Wald tests for Granger causality in levels vector autoregressions (VAR's) and Johansen-type error correction models (ECM's). The theory is based on results in Toda and Phillips (1991) and allows for stochastic and deterministic trends as well as arbitrary degrees of cointegration. For VAR models the theory extends earlier work by Sims, Stock and Watson (1990) on trivariate systems. In such models the results for inference are not encouraging. Explicit information on the number of unit roots in the system and the rank of a certain submatrix in the cointegrating space is needed to determine the appropriate limit theory in advance. Pretesting these conditions involves major complications in levels VAR's. But, even were the information to be available, the limit theory would frequently involve both nuisance parameters and nonstandard limit distributions, a situation where there is no satisfactory statistical basis for mounting the tests.

Granger causality tests in ECM's also suffer from nuisance parameter dependencies asymptotically and, in some cases, nonstandard limit theory. Both these results are somewhat surprising in the light of earlier research on the validity of asymptotic chi-square criteria in such systems. But, in spite of these difficulties Johansen-type ECM's do offer a sound basis for empirical testing of the rank of the cointegration space and the rank of key submatrices that influence the asymptotics. In consequence, we recommend some operational procedures for conducting Granger causality tests in the important practical case of testing the causal effects of one variable on another group of variables and vice versa. This paper also investigates the sampling properties of these testing procedures for Granger causality through simulation exercises. Three sequential causality tests in ECM's are compared with conventional causality tests based on VAR's in levels and in differences.

The plan of the paper is as follows. Section 2 reviews the theoretical results of Toda and Phillips (1991). Section 3 introduces the sequential causality tests and explains our experimental design for the Monte Carlo simulation. Section 4 reports the simulation results. Some concluding remarks are made in Section 5. A summary word on notation. We use $vec(M)$ to stack the *rows* of a matrix M. We use " $\xrightarrow{d}$ " and " $\equiv$ " to signify convergence in distribution and equality in distribution, respectively. The inequality " > 0 " denotes positive definite when applied to matrices. BM(Ω) denotes a multivariate Brownian motion with covariance matrix Ω. We write integrals with respect to Lebesgue measure such as $\int_0^1 W(s)ds$ more simply as $\int_0^1 W$ to achieve notational economy. All limits given in this paper are taken as the sample size $T \longrightarrow \infty$.

2 Theoretical Overview of Causality Tests

In this section we shall summarize the theoretical results of Toda and Phillips (1991). Consider the n-vector time series $\{y_t\}$ generated by the k-th order VAR model [1]

$$y_t = J(L)y_{t-1} + u_t \qquad t = -k+1, \ldots, T \tag{1}$$

where $J(L) = \sum_{i=1}^{k} J_i L^{i-1}$ and

(A1) $\{u_t = (u_{1t}, \ldots, u_{nt})'\}$ is an i.i.d. sequence of n dimensional random vectors with mean zero and covariance matrix $\Sigma_u > 0$ such that $E|u_{it}|^{2+\delta} < \infty$ for some $\delta > 0$.

We shall initialize (1) at $t = -k+1, \ldots, 0$ and allow the initial values $\{y_{-k+1}, \ldots, y_0\}$ to be any random vectors including constants. Define

$$J^*(L) = \sum_{i=1}^{k-1} J_i^* L^{i-1} \qquad \text{with} \qquad J_i^* = -\sum_{h=i+1}^{k} J_h.$$

We assume:

(A2) $|I_n - J(z)z| = 0$ implies $|z| > 1$ or $z = 1$.

(A3) $J(1) - I_n = \Gamma A'$ where Γ and A are $n \times r$ matrices of full column rank r, $0 \le r \le n-1$. (If $r = 0$, there is no Γ or A, and $J(1) = I_n$)

[1] For simplicity of exposition, we discuss in detail the case where there is no constant term. If a VAR has a constant term, y_t may have a deterministic time trend, and it affects the asymptotics for causality tests in levels VAR's. We will briefly discuss the results for such a case after the results for the model (1) are presented.

(A4) $\Gamma_\perp'(J^*(1) - I_n)A_\perp$ is nonsingular, where $\Gamma_\perp$ and $A_\perp$ are $n \times (n-r)$ matrices of full column rank such that $\Gamma_\perp'\Gamma = 0 = A_\perp' A$. (If $r = 0$, we take $A_\perp = I_n = \Gamma_\perp$.)

Under the above conditions y_t is CI(1,1) with r cointegrating vectors (if $r \geq 1$).[2] Condition (A2) precludes explosive processes but allows for the model (1) to have some unit roots. Condition (A3) defines the cointegrating space to be of rank r and A is a matrix whose columns span this space. Condition (A4) ensures that Δy_t is stationary. (See Theorem 3.1 of Johansen, 1989) Then, we can write (1) in the equivalent ECM format

$$\Delta y_t = J^*(L)\Delta y_{t-1} + \Gamma A' y_{t-1} + u_t. \tag{2}$$

Further, we need an additional assumption:

(A5) $E z_{1t} z_{1t}' > 0$ where $z_{1t} = (\Delta y_{t-1}', \ldots, \Delta y_{t-k+1}', (A' y_{t-1})')'$.

Note that $E z_{1t} z_{1t}'$ is the covariance matrix of the stationary component in the system, so this is a standard assumption.

Suppose that we want to test if there are causal effects from the last n_3 elements of y_t to the first n_1 elements of this vector, and accordingly partition y_t into three sub-vectors.

$$y_t = \begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} \begin{matrix} n_1 \\ n_2 \\ n_3 \end{matrix} .$$

Next, we introduce the selector matrices which will be used below:

$$S_1 = \begin{pmatrix} I_{n_1} \\ 0 \end{pmatrix}$$

and

$$S_3 = \begin{pmatrix} 0 \\ I_{n_3} \end{pmatrix}.$$

We shall first summarize the asymptotic results for causality tests in levels VAR's.

[2]The iid assumption (A1) is not, of course, necessary for y_t to be CI(1,1). In Section 3 we will discuss some cases where the u_t are MA(1) processes.

2.1 Causality Tests in Levels VAR's

The null hypothesis of noncausality can be formulated based on the model (1) as

$$\mathcal{H}: J_{1,13} = \cdots = J_{k,13} = 0 \tag{3}$$

where $J_{13}(L) = \sum_{i=1}^{k} J_{i,13}L^{i-1}$ is the $n_1 \times n_3$ upper-right submatrix of $J(L)$. Define

$$x_t = \left(y'_{t-1}, \ldots, y'_{t-k}\right)'$$

which is an nk-vector, and write (1) as

$$y_t = \Pi x_t + u_t$$

where $\Pi = (J_1, \ldots, J_k)$. Then the Wald statistic for noncausality can be written as

$$F = tr\left[S_1'\hat{\Pi}S[S'(X'X)^{-1}S]^{-1}S'\hat{\Pi}'S_1(S_1'\hat{\Sigma}_u S_1)^{-1}\right]$$

where $\hat{\Pi}$ and $\hat{\Sigma}_u$ are the least squares estimators of Π and Σ_u[3], $X' = (x_1, \ldots, x_T)$, and $S = I_k \otimes S_3$.

Asymptotic distributions for levels VAR causality tests are given by the following theorem.[4] Let A_3 denote the last n_3 rows of the matrix of cointegrating vectors A.

Theorem 1 *Suppose assumptions (A1) – (A5) are satisfied. If $rank(A_3) = g\,(\leq n_3)$, then under the null hypothesis (3)*

$$F \xrightarrow{d} \chi^2_{n_1[n_3(k-1)+g]} + tr\left[\int_0^1 dW_1\underline{W}_a'\left(\int_0^1 \underline{W}_a\underline{W}_a'\right)^{-1}\int_0^1 \underline{W}_a dW_1'\right]$$

where

$$\underline{W}_a(s) = W_a(s) - \int_0^1 W_a W_b'\left(\int_0^1 W_b W_b'\right)^{-1} W_b(s).$$

and

$$\begin{pmatrix} W_1(s) \\ W_a(s) \\ W_b(s) \end{pmatrix} \equiv BM(\Omega) \quad with \quad \Omega = \begin{pmatrix} I_{n_1} & \Omega_{a1}' & \Omega_{b1}' \\ \Omega_{a1} & I_{n_3-g} & 0 \\ \Omega_{b1} & 0 & I_{(n-r)-(n_3-g)} \end{pmatrix}.$$

□

[3] In this subsection " ˆ " on top of a parameter signifies the least squares estimator of the parameter.
[4] Proofs of all theorems and corollaries in the present paper are given in the companion paper Toda and Phillips (1991).

In the above theorem, Ω_{a1} and Ω_{b1} in general depend on the long-run covariance matrix of the system $(u_t', (A_\perp' \Delta y_t)')'$,[5] i.e., the limit distributions typically involve nuisance parameters. We have, however, two special cases that are noteworthy.

Corollary 1 *Suppose assumptions (A1) - (A5) are satisfied. If* $rank(A_3) = n_3$*, then under the null hypothesis (3)*

$$F \quad \xrightarrow{d} \quad \chi^2_{n_1 n_3 k}$$

□

Corollary 1 is a generalization of Sims, Stock and Watson's (1990) result from their analysis of trivariate VAR(k) systems with one cointegrating vector. Suppose that $n_1 = n_2 = n_3 = 1$ and the causal effect of y_3 on y_1 is being tested. Then, they conclude that if there is a linear combination involving y_3 which is stationary, the F-test will have an asymptotic χ_k^2/k distribution. In their example A_3 is nonzero scalar and $rank(A_3) = 1 = n_3$. So our Corollary 1 applies. But it should perhaps be noted that in view of Corollary 1 the situation concerning validity of chi-square asymptotics is more complex than their analysis of the trivariate example might suggest. For instance, if we wish to test the causal effects of two variables, say y_2 and y_3, on another, say y_1, then finding a cointegrating vector with nonzero coefficients for both y_2 and y_3 does not guarantee the usual chi-square asymptotics. Indeed, unless there are two cointegrating relations that involve both y_2 and y_3, the limit distribution will be nonstandard. Loosely put, we need "sufficient cointegration" with respect to the variables whose causal effects are being examined. Meanwhile, if there is no cointegration, we have nonstandard but nuisance parameter free limit distributions:

Corollary 2 *Suppose assumptions (A1) - (A5) are satisfied. If* y_t *is not cointegrated, i.e.,* $r = 0$*, then under the null hypothesis (3)*

$$F \xrightarrow{d} \chi^2_{n_1 n_3 (k-1)} + tr\left[\int_0^1 dW_1 \underline{W}_a' \left(\int_0^1 \underline{W}_a \underline{W}_a'\right)^{-1} \int_0^1 \underline{W}_a dW_1'\right]$$

[5]For the precise form of the dependence, see Theorem 1 of Toda and Phillips (1991).

where

$$\underline{W}_a(s) = W_a(s) - \int_0^1 W_a W_b' \left(\int_0^1 W_b W_b'\right)^{-1} W_b(s),$$

$$\begin{pmatrix} W_a(s) \\ W_b(s) \end{pmatrix} \begin{matrix} n_3 \\ n_1 + n_2 \end{matrix} \equiv BM(I_n),$$

and $W_1(s)$ *is the first* n_1 *elements of* $W_b(s)$. □

Corollary 1 is extended in a straightforward way to the case where the true model and the estimated equation have a constant term, while Corollary 2 is not. If the true model has a nonzero constant term and there is no cointegration in the system, then y_t contains a deterministic trend component. In order to obtain a nuisance parameter free limit distribution in such a case, we need to eliminate the deterministic trend by including not only a constant but also time as regressors in the estimated equation.[6] Then, the limit distribution component corresponding to the second term in Corollary 2 will be free of nuisance parameters but the Brownian motions in Corollary 2 will be replaced with "detrended Brownian motions". For example, $W_a(s)$ will be replaced by $\tilde{W}_a(s) = W_a(s) - \int_0^1 W_a \tau' \left(\int_0^1 \tau\tau'\right)^{-1} \tau(s)$ where $\tau(s) = (1, s)'$. (For further discussion see Section 3 of Toda and Phillips, 1991.)

Based on the foregoing formal results for causality tests in levels VAR's, Toda and Phillips (1991) conclude as follows:

> (i) Causality tests are valid asymptotically as chi-square criteria only when there is sufficient cointegration with respect to the variables whose causal effects are being tested. The precise condition for sufficiency involves a rank condition on a submatrix of the cointegrating matrix. Since the estimates of such matrices in levels VAR's suffer from simultaneous equation bias (as shown in Phillips, 1991), there is no valid statistical basis for determining whether the required sufficient condition applies.

[6]In other words, if the estimated equation has a constant term but not a time trend, then the distribution corresponding to the second term in Corollary 2 will be dependent on nuisance parameters in general.

(ii) When the rank condition for sufficiency fails, the limit distribution is more complex and involves a mixture of a chi-square distribution and a nonstandard distribution, which generally involves nuisance parameters. The precise form of the distribution depends on the actual rank of a submatrix of the cointegrating matrix and again no valid statistical basis for mounting a Wald test of causality applies.

(iii) If there is no cointegration, the Wald test statistic for causality has a nonstandard but nuisance parameter free limit distribution provided that the estimated equation is appropriately specified with regard to the presence of a deterministic time trend. This distribution could conceivably be used for tests when it is known that there are stochastic trends but no cointegration in the system.

2.2 Causality Tests in ECM's

Next, we discuss the asymptotics for causality tests in Johansen-type ECM's. The null hypothesis of noncausality can be formulated based on the model (2) as

$$\mathcal{H}^* : J^*_{1,13} = \cdots = J^*_{k-1,13} = 0 \qquad and \qquad \Gamma_1 A_3' = 0 \tag{4}$$

where $J^*_{13}(L) = \sum_{i=1}^{k-1} J^*_{i,13} L^{i-1}$ is the $n_1 \times n_3$ upper-right submatrix of $J^*(L)$, and Γ_1 is the first n_1 rows of the loading coefficient matrix Γ.

To test the hypothesis (4) we shall construct a Wald statistic. But before doing so, we need to introduce some more notation. First, let $\hat{A}$ denote the Gaussian estimator of A, i.e., the eigenvectors corresponding to the r largest eigenvalues that solve equation (9) of Johansen (1988) and let $\hat{A}_\perp$ be the eigenvectors corresponding to the $n-r$ smallest eigenvalues.[7] All the eigenvectors are normalized in the manner prescribed by Johansen (1988, p.235). Then the estimator of $(J_1^*, \ldots, J_{k-1}^*, \Gamma)$ is given by

$$(\hat{J}_1^*, \ldots, \hat{J}_{k-1}^*, \hat{\Gamma}) = \Delta Y' \hat{Z}_1 (\hat{Z}_1' \hat{Z}_1)^{-1}$$

[7] These $n-r$ eigenvectors do not provide a consistent estimator of the space spanned by $A_\perp$. But we call them $\hat{A}_\perp$ since their role in the derivation of the asymptotic distribution is the same as that of $A_\perp$.

where $\hat{Z}_1' = (\hat{z}_{11}, \ldots, \hat{z}_{1T})$ with $\hat{z}_{1t}' = (\Delta y_{t-1}, \ldots, \Delta y_{t-k+1}, (\hat{A}' y_{t-1})')$.[8] Also define $\hat{Z}_2' = (\hat{z}_{21}, \ldots, \hat{z}_{2T})$ with $\hat{z}_{2t} = \hat{A}_\perp' y_{t-1}$. Furthermore, let $\hat{\Sigma}_u$ be the Gaussian estimator of Σ_u, i.e.,

$$\hat{\Sigma}_u = T^{-1}\left[\Delta Y'\Delta Y - \Delta Y'\hat{Z}_1(\hat{Z}_1'\hat{Z}_1)^{-1}\hat{Z}_1'\Delta Y\right]$$

where $\Delta Y' = (\Delta y_1, \ldots, \Delta y_T)$ and let

$$\hat{\Omega}_c = (\hat{\Gamma}'\hat{\Sigma}_u^{-1}\hat{\Gamma})^{-1}.$$

Then we define

$$\hat{P}_* = \left(\begin{array}{c|c|c} I_{k-1}\otimes S_3'\otimes S_1' & 0 & 0 \\ \hline 0 & \hat{A}_3\otimes S_1' & \hat{A}_{\perp 3}\otimes\hat{\Gamma}_1 \end{array}\right)$$

where $\hat{A}_3$, $\hat{A}_{\perp 3}$, and $\hat{\Gamma}_1$ are the last n_3 rows of $\hat{A}$, the last n_3 rows of $\hat{A}_\perp$ and the first n_1 rows of $\hat{\Gamma}$, respectively, and let

$$\hat{\Omega}_* = \left(\begin{array}{c|c} (\hat{Z}_1'\hat{Z}_1)^{-1}\otimes\hat{\Sigma}_u & 0 \\ \hline 0 & (\hat{Z}_2'\hat{Z}_2)^{-1}\otimes\hat{\Omega}_c \end{array}\right).$$

Now we consider the following Wald statistic for testing the hypothesis (4)

$$F^* = vec(\hat{\Phi}_*')'\left(\hat{P}_*\hat{\Omega}_*\hat{P}_*'\right)^{-1} vec(\hat{\Phi}_*') \qquad (5)$$

where $\hat{\Phi}_* = (\hat{J}_{1,13}^*, \ldots, \hat{J}_{k-1,13}^*, \hat{\Gamma}_1\hat{A}_3')$ with $\hat{J}_{i,13}^*$ being the estimates for $J_{i,13}^*$ $(i = 1, \ldots, k-1)$.[9] We have the following asymptotic result for this statistic.

Theorem 2 *If assumptions (A1) - (A5) are satisfied, and if rank(Γ_1) = n_1 or rank(A_3) = n_3, then under the null hypothesis (4)*

$$F^* \xrightarrow{d} \chi^2_{n_1 n_3 k}.$$

□

As shown in Johansen (1989), the asymptotic distribution of the Gaussian estimator of A differs depending on whether the true model has a constant term or not, whether y_t

[8]In this subsection " ^ " on top of a parameter signifies the Gaussian estimator of the parameter.

[9]We cannot exclude the possibility that $\hat{P}_*\hat{\Omega}_*\hat{P}_*'$ is singular (even in the limit). But we ignore this problem here because the conditions given in Theorem 2 below ensure its nonsingularity in the limit and we will be interested only in such a situation.

actually contains a deterministic trend or not[10], and whether the presence or absence of the deterministic trend is taken into account in the estimation. But if one of the conditions in Theorem 2 is satisfied, Wald tests for causality will have asymptotic chi-squared distributions even in such cases (provided that the Wald statistic (5) is appropriately modified in obvious ways).

Theorem 2 shows that, as in levels VAR's, causality tests in ECM's are not in general valid asymptotic chi-square criteria since the conditions that guarantee the usual chi-square asymptotics do not always hold under the null. Suppose, for example, that there is only one cointegrating vector in a VAR(1) system, $n_1 = n_3 = 1$, and $A_3 = \Gamma_1 = 0$. Then, as proved in Example 3 of Toda and Phillips (1991), the Wald statistic for the noncausality hypothesis that $\Gamma_1 A_3 = 0$ has a limit distribution which is a nonlinear function of two independent chi-square variates, say χ_a and χ_b, viz.,

$$F^* \xrightarrow{d} \frac{\chi_a \chi_b}{\chi_a + \chi_b}.$$

The density of this distribution is more concentrated near the origin and has a thinner tail than χ_1^2, which is the limit distribution that we would obtain if either Γ_1 or A_3 is nonzero. (Figure 1 in Toda and Phillips, 1991)

Problems of both nuisance parameter dependencies[11] and nonstandard distributions enter the limit theory in the general case. These problems compromise the validity of conventional theory, and may be considered surprising and deserving of some emphasis in view of the fact that other types of Wald test in ECM's are known to be asymptotically valid chi-square tests. Thus, before we apply conventional asymptotic chi-square tests to noncausality hypotheses, we would have to test empirically whether $\text{rank}(\Gamma_1) = n_1$ or $\text{rank}(A_3) = n_3$ unless perhaps economic theory were to imply that one of them is of full row rank. Unlike the levels VAR approach, these conditions can, in principle, be tested using the Gaussian estimates of the submatrices of Γ_1 and A_3. Specifically, in the special but important case of testing causal effects of a group of variables on another variable and vice versa, the conditions can easily be tested. In the next section we discuss some

[10] A nonzero constant term does not always produce a deterministic trend if the system is cointegrated.

[11] Example 4 in Toda and Phillips (1991) illustrates nuisance parameter dependencies of the Wald tests.

operational procedures for testing causality in such cases. The following results as well as Theorem 2 provide the statistical basis for those sequential procedures.

Define

$$F_3^* = vec(\hat{A}_3)' \left[\hat{A}_{3\perp}(\hat{Z}_2'\hat{Z}_2)^{-1}\hat{A}_{3\perp}' \otimes \hat{\Omega}_c\right]^{-1} vec(\hat{A}_3),$$

$$F_1^* = vec(\hat{\Gamma}_1)' \left(S_1'\hat{\Sigma}_u S_1 \otimes \hat{\Sigma}_\gamma\right)^{-1} vec(\hat{\Gamma}_1)$$

where $\hat{\Sigma}_\gamma$ is the $r \times r$ lower-right block of $(\hat{Z}_1'\hat{Z}_1)^{-1}$,[12]

$$F_\dagger^* = vec(\hat{\Phi}_\dagger)' \left[S_1'\hat{\Sigma}_u S_1 \otimes (I_{k-1} \otimes S_3')\hat{\Sigma}_\dagger(I_{k-1} \otimes S_3)\right]^{-1} vec(\hat{\Phi}_\dagger)$$

where $\hat{\Phi}_\dagger = (\hat{J}_{1,13}^*, \ldots, \hat{J}_{k-1,13}^*)$ and $\hat{\Sigma}_\dagger$ is the $n(k-1) \times n(k-1)$ upper-left block of $(\hat{Z}_1'\hat{Z}_1)^{-1}$, and

$$F_{13}^* = vec(\hat{\Gamma}_1\hat{A}_3')' \left[S_1'\hat{\Sigma}_u S_1 \otimes \hat{A}_3\hat{\Sigma}_\gamma\hat{A}_3' + \hat{\Gamma}_1\hat{\Omega}_c\hat{\Gamma}_1' \otimes \hat{A}_{3\perp}(\hat{Z}_2'\hat{Z}_2)^{-1}\hat{A}_{3\perp}'\right]^{-1} vec(\hat{\Gamma}_1\hat{A}_3').$$

Then:

Proposition 1 *Suppose assumptions (A1) - (A5) are satisfied.*

(a) Under the null hypothesis that $A_3 = 0$,

$$F_3^* \xrightarrow{d} \chi^2_{n_3 r}.$$

(b) Under the null hypothesis that $\Gamma_1 = 0$,

$$F_1^* \xrightarrow{d} \chi^2_{n_1 r}.$$

(c) Under the null hypothesis that $J_{1,13}^* = \cdots = J_{k-1,13}^* = 0$,

$$F_\dagger^* \xrightarrow{d} \chi^2_{n_1 n_3 (k-1)}.$$

(d) Under the null hypothesis that $\Gamma_1 A_3' = 0$, *if* $rank(\Gamma_1) = n_1$ *or* $rank(A_3) = n_3$,

$$F_{13}^* \xrightarrow{d} \chi^2_{n_1 n_3}.$$

□

Proof: (a) follows from Lemma 4 of Toda and Phillips (1991), noting that $A_{3\perp}$ is of full row rank if $A_3 = 0$. (b) follows immediately from the same Lemma. (c) and (d) are just restatements of Theorem 1 above. □

[12] In fact, $\hat{\Sigma}_\gamma = T^{-1}I_r$ due to the normalization imposed by Johansen (1988, p.135)

3 Sequential Causality Tests and Experimental Design

According to Theorem 2, asymptotic chi-square criteria are applicable to causality tests based on (5) in ECM's only if (i) Γ_1 has full row rank or (ii) A_3 has full row rank. Hence we need to test these conditions empirically. But condition (i) or (ii) can be easily tested if $n_1 = 1$ or if $n_3 = 1$, respectively, since condition (i) is equivalent to $\gamma_1 \neq 0$ if $n_1 = 1$, and condition (ii) is equivalent to $\alpha_3 \neq 0$ if $n_3 = 1$.[13] For example, let $n_3 = 1$ and $k > 1$. Suppose that the pretest about the dimension of the cointegrating space has produced the estimate $\hat{r}$.[14] Then, Theorem 2 suggests that we first test whether (an $\hat{r}$ dimensional row vector) $\alpha_3 = 0$. If this is rejected, we may test noncausality using the Wald statistic (5). If it is accepted, we have only to test whether $J^*_{1,13} = \cdots = J^*_{k-1,13} = 0$ since α_3 being zero implies that $\gamma_1 \alpha'_3 = 0$. When $n_1 = 1$, we can proceed with a similar procedure. But if both n_1 and n_3 are one, a different testing strategy is also possible. Since in that case we can easily test both the hypotheses that $\gamma_1 = 0$ and that $\alpha_3 = 0$, it would be reasonable to proceed as follows. Begin with testing whether $J^*_{1,13} = \cdots = J^*_{k-1,13} = 0$. Suppose this is accepted. Then, we test whether $\alpha_3 = 0$ and whether $\gamma_1 = 0$. We accept the null of noncausality unless both are rejected. If both are rejected and $\hat{r} = 1$, then reject the null of noncausality. Otherwise test further whether $\gamma_1 \alpha'_3 = 0$.

To introduce the sequential procedures formally it is convenient to label each sub-hypothesis that appears in the sequential procedures. Let

$$\mathcal{H}^*_\dagger : J^*_{1,13} = \cdots = J^*_{k-1,13} = 0$$

$$\mathcal{H}^*_1 : \gamma_1 = 0$$

$$\mathcal{H}^*_3 : \alpha_3 = 0$$

$$\mathcal{H}^*_{13} : \gamma_1 \alpha'_3 = 0$$

and as in (4)

[13] In this section we use lower case letters to denote scalars and vectors. For example, α_3 corresponds to A_3 in the last section.

[14] In the simulation experiment below, the dimension of the cointegrating space will be estimated by the likelihood ratio test proposed by Johansen (1988, 1989).

$$\mathcal{H}^* : J^*_{1,13} = \cdots = J^*_{k-1,13} = 0 \qquad and \qquad \gamma_1 \alpha_3' = 0.$$

Now the sequential testing procedures to be considered in this paper are the following[15] :

(P1) Test $\mathcal{H}_1^*$. $\begin{cases} \text{If } \mathcal{H}_1^* \text{ is rejected, test } \mathcal{H}^*. \\ \text{Otherwise, test } \mathcal{H}_\dagger^*. \end{cases}$

(P2) Test $\mathcal{H}_3^*$. $\begin{cases} \text{If } \mathcal{H}_3^* \text{ is rejected, test } \mathcal{H}^*. \\ \text{Otherwise, test } \mathcal{H}_\dagger^*. \end{cases}$

(P3) Test $\mathcal{H}_\dagger^*$. $\begin{cases} \text{If } \mathcal{H}_\dagger^* \text{ is rejected, reject the null hypothesis of noncausality.} \\ \text{Otherwise, test } \mathcal{H}_1^* \text{ and } \mathcal{H}_3^*. \begin{cases} \text{If both } \mathcal{H}_1^* \text{ and } \mathcal{H}_3^* \text{ are rejected, test } \mathcal{H}_{13}^* \text{ if } \hat{r} > 1, \text{ or reject the null if } \hat{r} = 1. \\ \text{Otherwise, accept the null of noncausality.} \end{cases} \end{cases}$

where all the sub-hypotheses can be tested based on Theorem 2 and Proposition 1 in the last section. As stated above, (P3) differs from (P1) and (P2) because it takes advantage of the fact that both n_1 and n_3 are equal to one (i.e., both $\mathcal{H}_3^*$ and $\mathcal{H}_1^*$are tested in the second step). Thus, (P1), for example, is applicable when $n_1 = 1$ but $n_3 > 1$, while (P3) is applicable only when $n_1 = n_3 = 1$. Note that in (P3) it does not make any difference whether we start by testing $\mathcal{H}_\dagger^*$ or by testing $\mathcal{H}_1^*$ and $\mathcal{H}_3^*$ (and $\mathcal{H}_{13}^*$ if $\hat{r} > 1$), i.e., the results should be unchanged even though the order of testing is the other way around.

In the simulation experiment below, we set the nominal size of each sub-test to be 5 % in (P1) and (P2). But in (P3) the nominal size of each sub-test is 2.5 % if $k > 1$ and 5 % if $k = 1$. Though exact control of the overall size (i.e., the *maximal* probability of rejecting the null hypothesis under the null) of causality tests is not feasible, a heuristic

[15]The descriptions below assume that $k > 1$. If $k = 1$, obvious modifications should be made.

analysis suggests these choices of the size for each sub-test, and the overall size is expected to be approximately 5 % at least asymptotically.[16]

Now we explain our experimental design for investigating the sampling properties of the sequential test procedures introduced above. The prototype model for our simulation experiment is the trivariate VAR(1) :

$$y_t = J_1 y_{t-1} + u_t \tag{6}$$

or in its equivalent ECM format

$$\Delta y_t = \gamma\alpha' y_{t-1} + u_t. \tag{7}$$

where $y_t = (y_{1t}, y_{2t}, y_{3t})'$, $\alpha = (\alpha_1, \alpha_2, \alpha_3)'$ and $\gamma = (\gamma_1, \gamma_2, \gamma_3)'$ are 3×1 vectors. We shall later choose α and γ so as to satisfy conditions (A2) – (A4) of the last section.

We consider two different error processes $\{u_t\}$ in (7). Let

$$u_t = \epsilon_t - \Theta\epsilon_{t-1} \tag{8}$$

where $\epsilon_t \equiv iidN(0, I_3)$, and we consider the following Θ's:

(U1) $\quad \Theta = 0$

(U2) $\quad \Theta = \begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ \theta_{13} & 0 & 0.5 \end{pmatrix}.$

Some remarks are necessary about the error processes. First, (U2) appears to be inconsistent with our assumption (A1) in the last section. But since (8) is an invertible MA process when (U2) is employed as Θ, we can rewrite (7) as an ECM that has an infinite order AR lag polynomial:

$$\Delta y_t = J^*(L)\Delta y_{t-1} + \bar{\gamma}\alpha' y_{t-1} + \epsilon_t \tag{9}$$

where $J_i^* = -\sum_{h=i+1}^{\infty} \bar{J}_h$ with $\bar{J}_h = \Theta^{h-1}(J_1 - \Theta)$ $\;(i, h = 1, 2, \ldots)$ and $\bar{\gamma} = (I_3 - \Theta)^{-1}\gamma$. In view of this alternative expression (9) of the model (7), assumption (A1) is approximately

[16] Since the null of noncausality is consistent with different specifications of α and γ as we will see in (N1)-(N3) below, we cannot avoid relatively large distortions in such a case as (N3) to keep the overall size of causality tests approximately at 5 % level. See Section 4 below for more discussion.

satisfied at least when k is large enough. Second, note that if α and γ are chosen so that $\gamma_1\alpha_3 = 0$, then there is no causal effect from y_3 to y_1 in (9) since the (1,3) element of $\bar{J}_h$ is equal to $2^{1-h}\gamma_1\alpha_3$ and $\bar{\gamma}_1\alpha_3 = 2\gamma_1\alpha_3$. Third, simulations were run for different values of θ_{13} and Σ_ϵ (covariance matrix of ϵ_t). But the results were qualitatively the same in all cases except one, which we will discuss later. Therefore, in this paper we report mainly the results for $\theta_{13} = 1$ with $\Sigma_\epsilon = I_3$. In the following, we shall refer to the Θ with $\theta_{13} = 1$ as (U2a), i.e.,

$$\text{(U2a)} \quad \Theta = \begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 1 & 0 & 0.5 \end{pmatrix}.$$

Now we choose the values for α and γ as follows. If we set $J_1 = I_3 + \gamma\alpha'$ for any α and γ, condition (A3) in the last section is automatically satisfied. Further, it is easy to show that if α and γ satisfy

$$-2 < \alpha'\gamma < 0,$$

then condition (A2) is satisfied and the characteristic equation $|I_3 - J_1 z| = 0$ has two unit roots and one stable root equal to $(1 + \alpha'\gamma)^{-1}$. Thus, in our experiment we shall use the following values for α and γ :

$$\text{(N1)} \quad \alpha = (-0.5, 1, 0)' \quad \text{and} \quad \gamma = (1, 0, 1)', \quad \text{i.e.,} \quad J_1 = \begin{pmatrix} 0.5 & 1 & 0 \\ 0 & 1 & 0 \\ -0.5 & 1 & 1 \end{pmatrix}$$

$$\text{(N2)} \quad \alpha = (1, 0.5, -1)' \quad \text{and} \quad \gamma = (0, 1, 1)', \quad \text{i.e.,} \quad J_1 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1.5 & -1 \\ 1 & 0.5 & 0 \end{pmatrix}$$

$$\text{(N3)} \quad \alpha = (1, -0.5, 0)' \quad \text{and} \quad \gamma = (0, 1, 1)', \quad \text{i.e.,} \quad J_1 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0.5 & 0 \\ 1 & -0.5 & 1 \end{pmatrix}$$

$$\text{(L1)}\quad \alpha = (-0.5, 1, 0.1)' \quad \text{and} \quad \gamma = (1, 0, 1)', \quad \text{i.e.,} \quad J_1 = \begin{pmatrix} 0.5 & 1 & 0.1 \\ 0 & 1 & 0 \\ -0.5 & 1 & 1.1 \end{pmatrix}$$

$$\text{(L2)}\quad \alpha = (1, 0.5, -1)' \quad \text{and} \quad \gamma = (0.1, 1, 1)', \quad \text{i.e.,} \quad J_1 = \begin{pmatrix} 1.1 & 0.05 & -0.1 \\ 1 & 1.5 & -1 \\ 1 & 0.5 & 0 \end{pmatrix}$$

$$\text{(L3.1)}\quad \alpha = (1, -0.5, -0.1)' \quad \text{and} \quad \gamma = (0.1, 1, 1)', \quad \text{i.e.,} \quad J_1 = \begin{pmatrix} 1.1 & -0.05 & -0.01 \\ 1 & 0.5 & -0.1 \\ 1 & -0.5 & 0.9 \end{pmatrix}$$

$$\text{(L3.2)}\quad \alpha = (1, -0.5, -0.3)' \quad \text{and} \quad \gamma = (0.3, 1, 1)', \quad \text{i.e.,} \quad J_1 = \begin{pmatrix} 1.3 & -0.15 & -0.09 \\ 1 & 0.5 & -0.3 \\ 1 & -0.5 & 0.7 \end{pmatrix}$$

In (N1) - (N3) the values of α and γ were selected so that the stable root of the system is equal to 2. It is straightforward to show that each pair of α and γ above satisfies condition (A4) also. Hence y_t is CI(1,1) with one cointegrating vector α[17]. Observe that when the parameter values (N1) - (N3) are employed, there are no causal effects from y_3 to y_1 in (7) if u_t is iid, and in (9) if u_t is MA(1). Note also that (L1), (L2), (L3.1) and (L3.2) serve as corresponding " local " alternatives.

Next, in this study we concentrate on three different estimated equations which have lag lengths $k = 1, 2$, and 4, respectively. That is, the estimated systems of equations considered here are

$$\Delta y_t = \hat{\Gamma}\hat{A}' y_{t-1} + \hat{u}_t \tag{10}$$

if $k = 1$, and

$$\Delta y_t = \hat{J}_1^* \Delta y_{t-1} + \cdots + \hat{J}_{k-1}^* \Delta y_{t-k+1} + \hat{\Gamma}\hat{A}' y_{t-1} + \hat{u}_t \tag{11}$$

if $k = 2$ and 4. The lag lengths $k = 6$ and 8 were also tried for the combination of (N1) and (U2a), which we will discuss later.

[17] Again, the iid assumption (A1) is not necessary for y_t to be CI(1,1).

In our experiment we start by estimating r using the likelihood ratio test, specifically the "trace test", proposed by Johansen (1988, 1989). Then, having estimated $\hat{r}$, we proceed as follows. If $\hat{r} = 0$, a VAR in differences is estimated and causality is tested in the usual manner. If $\hat{r} = 3$, the data are regarded as stationary and causality is tested based on a levels VAR. If $0 < \hat{r} < 3$, we apply the sequential testing procedures (P1) – (P3). In this case the null hypothesis of noncausality is, if $k = 1$,

$$\mathcal{H}_{13}^{*} : \gamma_1 \alpha_3' = 0$$

and if $k > 1$,

$$\mathcal{H}^{*} : J_{1,13}^{*} = \cdots = J_{k-1,13}^{*} = 0 \qquad \textit{and} \qquad \gamma_1 \alpha_3' = 0$$

where γ_1 and α_3 are $\hat{r}$ dimensional vectors, and $J_{i,13}^{*}$ $(i = 1, \ldots, k-1)$ are scalars. Note that even though γ_1 and α_3 are scalars in the true model (7), they are not necessarily so in $\mathcal{H}_{13}^{*}$ and $\mathcal{H}^{*}$.

4 Simulation Results

For each combination of α, γ, Θ, lag length (k) in estimated equations, and a sample size (T), 5000 series of $T + k + 100$ observations were generated according to equation (6) with $y_0 = 0$. The innovation series $\{\epsilon_t\}$ were generated by the RNDN function of the GAUSS matrix programing language. The initial 100 observations were discarded, generating a series of length $T + k$, i.e., T observations for the dependent variables Δy_t in estimated equations (10) and (11). For each of those samples, the sequential testing procedures (P1) – (P3) described in the last section were applied and their performance was examined.

Simulation results are reported in Tables 1 – 16. Each of Tables 1 – 14 corresponds to one combination of α, γ, and Θ. For each k, the first column shows the results (%) of Johansen's likelihood ratio test about the dimension, r, of cointegrating spaces. The second through fourth columns show rejections (%) of noncausality conditioned on the estimated r's and in total replications. These tables also show the performance of causality tests in levels VAR's and differences VAR's, based on a common 5000 replications generated as above.[18] We note that testing causality in this conventional fashion does not yield a valid asymptotic chi-square criterion for all pairs of α and γ that are consistent with the null of noncausality. (See below for further discussion.)

Tables 1 – 3 show the simulation results of the tests under the null of noncausality when $\{u_t\}$ is an iid sequence[19]. Hence the correct specification of the estimated equation is (10). The testing procedures (P1) – (P3) perform similarly when $k = 1$, but (P3) appears to have less size distortions in (N2) and (N3) compared to the other two when $k > 1$. If we compare (P1) and (P2) when $k > 1$, (P1) seems better than (P2) in (N1) and (N3). All of (P1) – (P3) perform reasonably well when k is chosen correctly and/or the sample size is 100 or greater, though the results are rather sensitive to the values of α and γ.

Although size distortion due to wrong estimation of r is an inevitable nature of the sequential procedures, a case of notable size distortion under correct estimation of r (i.e., $\hat{r} = 1$) occurs when $k = 1$ and the true values of α_3 and γ_1 are both equal to zero (Table 3). The distortions in (P3) are due to the fact that we reject the null of noncausality only if the statistically independent sub-tests, $\mathcal{H}_1^*$ and $\mathcal{H}_3^*$, each of which has 5 % nominal size, are both rejected[20]. Thus the probability of rejecting the null of noncausality conditioned on $\hat{r} = 1$ is expected to be about 0.25 %. If we chose 22 % critical values for those sub-tests, then we would have approximately 5 % significance level for the overall causality test in this particular case of parameter values, but of course we cannot do so without allowing large upward size distortions in other cases where one of α_3 and γ_1 is not equal to zero. The tests (P1) and (P2) have the same distortional property in the case of (N3) with $k = 1$, and in fact the distortions are worse in (P1) and (P2) because in the case (N3) the limit distribution of the Wald test F_{13}^* is highly concentrated near the origin (see Example 3 and Figure 1 in Toda and Phillips, 1991) and the sub-test $\mathcal{H}_{13}^*$ almost never rejects the null.

[18]But those series were generated independently of the series which were used for the sequential procedures (P1) – (P3).

[19]If the pretest about the dimension of cointegrating spaces gives the estimate $\hat{r} = 0$ or 3, then there is no difference in (P1) – (P3).

[20]$\hat{\gamma}_1$ and $\hat{\alpha}_3$ have independent limit distributions by Lemma 4 of Toda and Phillips (1991)

This kind of downward size distortion can also explain why the test (P3) suffers from less size distortion in (N3) when $k > 1$. In fact, since we had to choose the nominal size of each sub-test to be 2.5 % when $0 < \hat{r} < n$ and $k > 1$, the probability of rejecting the noncausality null conditioned on $\hat{r} = 1$ in (N3) is expected to be about 2.56 % if the sub-test $\mathcal{H}_{\dagger}^{*}$ were independent of $\mathcal{H}_{1}^{*}$ and $\mathcal{H}_{3}^{*}$. Though $\mathcal{H}_{1}^{*}$ and $\mathcal{H}_{\dagger}^{*}$ are correlated in general, this probability is likely to be less than 5 % and it actually was for large samples as Table 3 shows. But, again, we cannot do better in (N3) without allowing large upward distortions in the cases of (N1) and (N2). Though, as Table 3 shows, this downward distortion of the conditional probability happened to contribute to "seemingly" less size distortions of the test (P3) in (N3), this might not be always the case. In (P1) and (P2), however, this sort of downward bias does not occur if $k > 1$ since the size of each sub-test can be selected to be 5 % even in (N3) without causing additional distortions in such cases as (N1) and (N2).

Note that in (N2) causality tests based on levels VAR's are valid since the cointegrating vector involves the variable y_3 whose causal effect is examined, i.e., there is "sufficient cointegration" with respect to y_3 (Corollary 1). Further, in (N2) and (N3) causality tests based on differences VAR's are valid since if $\gamma_1 = 0$, then the first equation of the ECM (7) dose not involve any level variables. Moreover, even when causality tests based on levels VAR's do not provide correct asymptotic chi-square tests, we expect that the more lags we include in estimated equations, the less serious the distortion becomes in general. This is because the limit distribution of the Wald statistic for testing causal effect, say, from one variable to another in levels VAR's has the form

$$\chi_{k-1}^{2} + \zeta$$

by Theorem 1, where the random variable ζ has some unit root type distribution. Hence the relative effect of the ζ term is expected to become smaller as the lag length k increases.

The figures in Tables 1 – 3 verify the above heuristic arguments. The performance of our sequential tests (P1) – (P3) and the tests based on levels VAR's seem similar in the case of (N2). Furthermore, Table 1 shows that, as predicted by the asymptotic theory, the advantage of the sequential procedures over the levels VAR based tests becomes smaller as k increases and, in fact, the tests (P1) - (P3) lose the advantage when $k = 4$ even in the sample size 200. Since in practice econometricians probably tend to include more lags than the true number of k (if k is finite), this result could be interpreted as supporting the use of levels VAR's even when the system is subject to "insufficient" cointegration with respect to the variable whose causality is tested. But Table 1 suggests that the performance of the sequential tests (P1) – (P3) is significantly better than that of the levels VAR based tests provided that the lag specification is correct. Moreover, Table 3 shows that in the case of (N3) the sequential tests are much better even when $k = 4$. Finally, the tests based on VAR's in differences perform better than our testing procedures in the cases (N2) and (N3) especially if sample sizes are small. However, the distortion in (N1) is enormous.

Tables 4 – 7 report the power of the tests under the "local" alternatives in the case of iid errors (U1). Note that (L2) and (L3.2) are comparable since under these settings the values of $\gamma_1\alpha_3$, i.e., the (1,3) element of J_1, are -0.1, and -0.09, respectively. Again, the sequential tests (P1) – (P3) perform rather similarly especially if $k = 1$, and the power of those tests significantly depends on the true values of α and γ. In (L2), (L3.1) and (L3.2) our testing procedures do not have much power unless the lag length k is specified correctly. This is not very surprising because if $k > 1$ the coefficients of the lagged differences of y_3 are all zero. For example, in (L2) with $k = 4$ the values of the coefficients which are tested for causality are : $(J^*_{1,13}, J^*_{2,13}, J^*_{3,13}, \gamma_1\alpha_3) = (0, 0, 0, -0.1)$. (Notice that if $k > 1$, the tests based on VAR's in levels and in differences do not have much power, either.) But at least for $k = 1$, Tables 4 - 7 show that the sequential procedures (P1) – (P3) have reasonable power and are in general more powerful than the tests based on levels VAR's. The comparison of the power among the tests (P1) – (P3) does not reveal any superiority of one of those sequential tests uniformly over the specifications of α, γ, and k. Thus, we have no strong reason to favor any particular one of (P1) – (P3) compared to the others in terms of its power.

Tables 8 – 10 show the performance of the tests when the error is an MA(1) process with (U2a). In the present case the (approximately) correct specification of the estimated equation is (11) with $k = 4$. As in the case of iid errors (U1), our sequential testing procedures perform rather well if k is specified "correctly" and the sample size is large. But they are not recommendable if the sample size is less than 100. In (N1) - (N3) with $k = 4$ the test (P3) seems to suffer from less size distortions than the tests (P1) and (P2). As we discussed in the iid error case, the "seemingly" less size distortion of the test (P3) in (N3) can be explained by the downward bias in the probability of rejecting the noncausality null conditioned on $\hat{r} = 1$. Comparing (P1) and (P2) when $k = 4$, (P1) seems to perform better than (P2) in all cases. Since the estimated equation with $k = 4$ in the case of (U2a) is only an approximation of the true model, the test performance is, of course, not as good as the iid case with $k = 1$.

Table 9 shows that when causality tests based on levels VAR's are asymptotically valid (i.e., α and γ are chosen as (N2)), their performance with $k = 4$ is similar to that of the sequential testing procedures (P1) - (P3). But in (N1) and (N3) with $k = 4$ they suffer from significantly more distortion compared to our sequential tests (Tables 8 and 10). When causality tests based on differences VAR's are valid (Tables 9 and 10), they outperform the sequential procedures (P1) - (P3), though the size distortion in the case of (N1) is enormous (Table 8). These tendencies are also observed when the error is an iid sequence and $k > 1$.

Tables 11 - 14 show, as in the case of iid errors (U1), that when $k = 4$, neither our sequential tests nor the conventional tests have much power under the "local alternatives" except for the case (L1). (Some high rejection rates in the case of $k = 1$ or $k = 2$ are, of course, due to misspecification of the estimated equations and hence do not mean that the tests are powerful.) Again this is rather an expected result. Though they are not strictly zero unlike the iid error case, the magnitude of the coefficients for the lagged differences Δy are very small since they are derived from inverting the MA(1) lag polynomial $I_3 - \Theta L$. See the model (9). (This fact reflects also in the results for levels VAR's and differences VAR's.) But we may still examine the relative power of the sequential tests (P1) - (P3). As in the case of iid errors (U1), comparing them when $k = 4$ suggests that any one of these tests does not outperform the others uniformly over the specifications of γ and α, e.g., (P1) and (P2) seem more powerful than (P3) in (L2) and vice versa in (L3.2).

In n-variate VAR (and ECM) frameworks the number of parameters increases by n^2 as the number of lags to be included in estimated equations increases by one. Hence, we would expect that the estimator may deteriorate if "too many" lags relative to the sample size are included in estimation. Moreover, as pointed out earlier, the asymptotic theory implies that the size distortion from which the tests in levels VAR's suffer becomes relatively small as k increases even though they do not yield correct asymptotic chi-square criteria. Therefore, it is of some interest to see how the test performances are affected by an increase in the number of lags included in estimation. Thus, we ran simulations with $k > 4$ for the combination of (N1) and (U2a). We chose (N1) rather than (N3) because levels VAR's are likely to perform better in (N1) than in (N3) though both (N1) and (N3) are the case where levels VAR's do not provide asymptotically chi-square tests. The results are shown in Table 15. For $T = 200$, the procedures (P1) - (P3) are still better than levels VAR's when $k = 6$, but when $k = 8$, only (P1) outperforms the levels VAR based tests. The table show that for the sequential procedures, $k = 4$ provides the best results for all sample sizes and as k increases the test performance deteriorates fairly quickly even though the true model has an infinite lag polynomial. The tests in levels VAR's reveal a similar tendency. (But for $T = 200$ the performance begins to worsen only as k exceeds 6.) However, in the sequential tests (P1) - (P3) this sort of deterioration is expected to be and actually was more serious since the Johansen-type ML method on which the sequential causality tests are based is more complicated than ordinary VAR estimation.

One interesting observation on levels VAR's which have been widely used in the econometric literature is that causality tests begin to deteriorate as k exceeds 6 even for the sample size equal to 200, which is relatively large in practice. Since $k = 8$, for example, is not an excessively long lag length compared with those used in practice (e.g., in investment studies), this finding is of some importance for interpreting existing empirical studies.

We also ran simulations for some parameter constellations of Θ and Σ_ϵ other than reported in Tables 1 - 14. For example, in addition to (U2a) the following Θ and Σ_ϵ were tried in the case (N1):

$$\text{(U2b)} \quad \Theta = \begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ -1 & 0 & 0.5 \end{pmatrix}$$

$$\text{(S1)} \quad \Sigma_\epsilon = \begin{pmatrix} 1.0 & 0.2 & 0.5 \\ 0.2 & 1.0 & 0.2 \\ 0.5 & 0.2 & 1.0 \end{pmatrix} \qquad \text{(S2)} \quad \Sigma_\epsilon = \begin{pmatrix} 1.0 & 0.2 & -0.5 \\ 0.2 & 1.0 & 0.2 \\ -0.5 & 0.2 & 1.0 \end{pmatrix}$$

For each combination of Θ and Σ_ϵ taken from (U2a) - (U2b) and (S1) - (S2), the test performance was examined. The results are basically the same as in Table 8 in all combinations except one, that is, (U2b) and (S2), as reported in Table 16. In this case levels VAR's suffer significantly less size distortion than the sequential tests (P1) - (P3).[21] Note that for this combination of Θ and Σ_ϵ, Johansen's likelihood ratio test does not work very well even for the sample size 200, and more importantly that the size distortion when $\hat{r} = 1$ is much larger than in other Tables, i.e., the Gaussian ML method provides poor estimates of the coefficients even though r is correctly specified. This is obviously because the inclusion of four lags in the estimated model could not remove the serial correlation in the errors effectively enough for the ML method to work well in the present case. Hence, we would have to include more lags in the estimated equations to improve the test performance. But Table 15 suggests that we must also have larger samples in order to allow more lags in the estimated equations without deteriorating the test performance.

5 Conclusions

This paper has provided a theoretical overview of Wald tests for Granger causality in levels VAR's and Johansen-type ECM's. In the ECM framework we have proposed some operational testing procedures that are applicable in the important practical case of testing

[21] But in the case of (N3) the sequential tests still performed significantly better than levels VAR based tests even when (U2b) and (S2) are used for the error process.

the causal effects of one variable on another group of variables and vice versa. We have also investigated the finite sample properties of these sequential causality tests through Monte Carlo simulations. Since the data generating processes we employed in this study are simple, it would be unwise to make strong general claims from this simulation study. But our findings may be summarized as follows:

(i) The sequential testing procedures perform well at least in large samples when the lag length is correctly specified.

(ii) The sequential tests outperform conventional VAR tests in the sense that the former tests work reasonably well for all specifications of cointegrating vectors and loading coefficients that are consistent with the null of noncausality, while the latter tests suffer from significant size distortion in cases where tests are not valid asymptotically as chi-square criteria.

(iii) For some types of serially correlated error processes the Johansen approach and hence our testing procedures that are based on it do not work well.

(iv) The simulation results do not support the use of either our sequential procedures or conventional causality tests in samples smaller than 100, at least if the system has three or more variables; and if these testing procedures are to be used in practice it is desirable that sample sizes be greater than 100 observations.

(v) Our simulations show the important role played by the choice of lag length in the performance of these tests.

Comparisons among the sequential test procedures themselves show similar performance and there is little evidence favoring any one of them. Hence it is not clear which one should be used if all of them are applicable, i.e., if $n_1 = n_3 = 1$. But one possible suggestion on this matter is the following. Since the null hypothesis of noncausality is consistent with different combinations of values of α and γ, it is impossible to choose the nominal size of each sub-test so that the probability of rejecting the null of noncausality is always, say 5 %, independent of the specifications of α_3 and γ_1, and hence we will not be able to avoid significant distortions in some cases. Therefore, we might want to proceed with the testing procedure for which the probability of rejecting the null depends on the true parameter values the least. Though a rigorous analysis of this problem would be very difficult, at least a heuristic analysis suggests that the tests (P1) and (P2) are less vulnerable with respect to this kind of distortion than (P3) because they have simpler structures than the test (P3). Thus, combining this observation and the simulation result that (P3) did not necessarily perform better than the others, it would seem reasonable to work with (P1) or (P2) rather than (P3).

If we exclude the test (P3), comparing the tests (P1) and (P2) when $k > 1$ shows that the test (P1) performs better than the test (P2). Furthermore, Table 15 suggests that the deterioration of the test performance associated with the increase in k seems less serious in (P1). Since we expect $k > 1$ in most practical applications, the test (P1) could be regarded as a better testing procedure. Therefore, if both n_1 and n_3 are equal to one, we conclude that a reasonable choice is to apply the sequential procedure (P1).

Table 1: (N1) $\alpha_3 = 0$ $\gamma_1 = 1$ (U1) $\Theta = 0$

	k=1				k=2				k=4			
		Rejections				Rejections				Rejections		
	LR Test	(P1)	(P2)	(P3)	LR Test	(P1)	(P2)	(P3)	LR Test	(P1)	(P2)	(P3)
					T = 50							
r = 0	0.0	–	–	–	0.9	95.6	95.6	95.6	30.9	60.2	60.2	60.2
r = 1	92.3	8.6	8.6	9.3	90.6	14.1	15.9	16.1	58.0	28.9	33.6	36.1
r = 2	6.9	35.3	34.4	34.4	7.5	29.6	30.1	29.3	9.3	35.1	34.8	36.3
r = 3	0.8	26.8	26.8	26.8	1.0	25.0	25.0	25.0	1.8	42.7	42.7	42.7
Total	–	10.6	10.5	11.1	–	16.1	17.8	17.9	–	39.4	42.1	43.7
					T = 100							
r = 0	0.0	–	–	–	0.0	–	–	–	1.2	88.7	88.7	88.7
r = 1	93.1	5.9	5.9	6.3	92.6	8.6	10.9	9.3	90.3	14.0	15.0	17.6
r = 2	6.2	29.2	28.9	28.9	6.8	26.0	26.3	28.4	7.7	24.0	24.3	26.1
r = 3	0.7	36.1	36.1	36.1	0.5	15.4	15.4	15.4	0.7	16.2	16.2	16.2
Total	–	7.6	7.6	7.9	–	9.9	12.0	10.6	–	15.7	16.7	19.2
					T = 200							
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	93.9	6.0	6.0	6.2	94.2	6.8	8.5	6.7	93.3	8.0	9.0	9.5
r = 2	5.6	29.5	28.8	28.8	5.2	26.6	27.4	26.6	6.2	24.9	24.6	27.8
r = 3	0.6	20.7	20.7	20.7	0.6	22.6	22.6	22.6	0.6	10.7	10.7	10.7
Total	–	7.4	7.3	7.5	–	7.9	9.6	7.9	–	9.1	9.9	10.6

VAR(k-1) in differences
Rejections

	k=1	k=2	k=4
T = 50	–	84.9	55.6
T = 100	–	98.9	85.6
T = 200	–	100.0	99.7

VAR(k) in levels
Rejections

	k=1	k=2	k=4
T = 50	13.7	14.0	19.6
T = 100	11.9	11.1	13.7
T = 200	12.1	10.4	10.1

Table 2: (N2) $\alpha_3 = -1$ $\gamma_1 = 0$ (U1) $\Theta = 0$

	k=1				k=2				k=4			
	LR Test	Rejections (P1)	(P2)	(P3)	LR Test	Rejections (P1)	(P2)	(P3)	LR Test	Rejections (P1)	(P2)	(P3)
T = 50												
r = 0	0.0	–	–	–	0.4	9.1	9.1	9.1	28.8	11.6	11.6	11.6
r = 1	92.8	5.4	5.4	5.4	90.0	10.0	9.1	8.5	60.4	17.3	18.7	17.9
r = 2	6.6	14.0	15.2	14.0	8.5	9.5	9.5	8.0	9.5	19.8	18.3	14.3
r = 3	0.7	17.6	17.6	17.6	1.1	5.6	5.6	5.6	1.3	26.9	26.9	26.9
Total	–	6.0	6.1	6.1	–	9.9	9.1	8.5	–	16.0	16.8	15.9
T = 100												
r = 0	0.0	–	–	–	0.0	–	–	–	1.0	3.8	3.8	3.8
r = 1	93.2	5.4	5.4	5.4	92.8	7.0	6.2	5.4	90.8	9.4	9.4	8.7
r = 2	6.2	11.0	12.0	11.0	6.4	6.9	7.5	6.2	7.1	10.1	10.4	6.4
r = 3	0.6	3.2	3.2	3.2	0.7	8.1	8.1	8.1	1.0	4.0	4.0	4.0
Total	–	5.8	5.8	5.8	–	7.0	6.3	5.5	–	9.4	9.3	8.5
T = 200												
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	93.6	5.5	5.5	5.5	93.7	6.1	5.6	4.6	93.3	6.8	6.9	6.2
r = 2	5.6	8.6	8.9	8.6	5.8	6.9	6.9	4.8	6.1	10.1	10.5	8.5
r = 3	0.8	18.4	18.4	18.4	0.5	8.7	8.7	8.7	0.6	10.0	10.0	10.0
Total	–	5.8	5.8	5.8	–	6.1	5.7	4.7	–	7.0	7.1	6.4

VAR(k-1) in differences
Rejections

	k=1	k=2	k=4
T = 50	–	6.0	10.5
T = 100	–	5.2	8.2
T = 200	–	5.2	6.1

VAR(k) in levels
Rejections

	k=1	k=2	k=4
T = 50	7.7	8.8	14.7
T = 100	5.7	6.4	9.4
T = 200	6.4	5.6	6.9

Table 3: (N3) $\alpha_3 = 0 \quad \gamma_1 = 0 \qquad$ (U1) $\Theta = 0$

	k=1				k=2				k=4			
		Rejections				Rejections				Rejections		
	LR Test	(P1)	(P2)	(P3)	LR Test	(P1)	(P2)	(P3)	LR Test	(P1)	(P2)	(P3)
T = 50												
r = 0	0.0	–	–	–	0.7	11.8	11.8	11.8	30.2	12.8	12.8	12.8
r = 1	92.0	0.0	0.0	0.5	90.2	6.7	7.4	5.0	58.9	16.3	16.7	15.4
r = 2	7.1	47.8	51.4	47.8	8.1	49.6	49.6	45.2	9.3	44.3	45.6	44.1
r = 3	0.9	51.2	51.2	51.2	1.0	42.9	42.9	42.9	1.6	41.5	41.5	41.5
Total	–	3.9	4.1	4.3	–	10.6	11.2	8.7	–	18.3	18.6	17.7
T = 100												
r = 0	0.0	–	–	–	0.0	–	–	–	1.6	8.5	8.5	8.5
r = 1	93.0	0.0	0.0	0.3	93.4	5.2	6.2	3.8	89.7	7.3	7.5	5.9
r = 2	6.4	46.4	49.8	46.4	6.0	48.3	47.7	43.7	7.6	42.1	42.4	45.3
r = 3	0.6	39.3	39.3	39.3	0.6	42.9	42.9	42.9	1.0	30.6	30.6	30.6
Total	–	3.2	3.4	3.5	–	8.0	8.9	6.4	–	10.2	10.4	9.2
T = 200												
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	93.8	0.0	0.0	0.4	93.3	4.4	5.2	3.2	92.9	5.7	6.0	3.8
r = 2	5.7	52.7	55.8	52.7	6.2	45.7	47.3	41.5	6.4	37.6	38.9	39.8
r = 3	0.6	32.1	32.1	32.1	0.5	26.9	26.9	26.9	0.7	17.1	17.1	17.1
Total	–	3.2	3.3	3.5	–	7.1	7.9	5.7	–	7.8	8.1	6.2

VAR(k-1) in differences

Rejections

	k=1	k=2	k=4
T = 50	–	5.9	11.2
T = 100	–	5.5	7.3
T = 200	–	5.7	5.9

VAR(k) in levels

Rejections

	k=1	k=2	k=4
T = 50	17.8	17.3	22.7
T = 100	17.2	14.5	15.9
T = 200	16.4	13.7	12.6

Table 4: (L1) $\alpha_3 = 0.1$ $\gamma_1 = 1$ (U1) $\Theta = 0$

	k=1 LR Test	k=1 Rejections (P1)	(P2)	(P3)	k=2 LR Test	k=2 Rejections (P1)	(P2)	(P3)	k=4 LR Test	k=4 Rejections (P1)	(P2)	(P3)
T = 50												
r = 0	0.0	–	–	–	0.5	95.8	95.8	95.8	28.5	92.5	92.5	92.5
r = 1	92.2	92.9	92.9	93.3	89.7	92.7	92.6	92.5	59.4	86.8	94.6	82.8
r = 2	6.8	95.6	95.3	95.3	8.6	94.2	94.0	91.4	10.1	91.7	93.3	87.7
r = 3	1.0	98.0	98.0	98.0	1.2	100.0	100.0	100.0	2.0	93.9	93.9	93.9
Total	–	93.2	93.1	93.5	–	92.9	92.8	92.5	–	89.0	93.8	86.3
T = 100												
r = 0	0.0	–	–	–	0.0	–	–	–	1.2	100.0	100.0	100.0
r = 1	92.6	99.7	99.7	99.7	92.1	99.8	99.8	99.8	89.7	99.4	99.7	98.6
r = 2	6.9	100.0	100.0	100.0	7.3	100.0	100.0	100.0	8.3	100.0	100.0	99.3
r = 3	0.5	100.0	100.0	100.0	0.7	100.0	100.0	100.0	0.9	100.0	100.0	100.0
Total	–	99.7	99.7	99.7	–	99.8	99.8	99.9	–	99.4	99.7	98.7
T = 200												
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	93.3	100.0	100.0	100.0	93.2	100.0	100.0	100.0	92.7	100.0	100.0	100.0
r = 2	6.1	100.0	100.0	100.0	5.9	100.0	100.0	100.0	6.4	100.0	100.0	100.0
r = 3	0.7	100.0	100.0	100.0	0.9	100.0	100.0	100.0	0.9	100.0	100.0	100.0
Total	–	100.0	100.0	100.0	–	100.0	100.0	100.0	–	100.0	100.0	100.0

VAR(k-1) in differences
Rejections

	k=1	k=2	k=4
T = 50	–	97.3	90.8
T = 100	–	100.0	99.8
T = 200	–	100.0	100.0

VAR(k) in levels
Rejections

	k=1	k=2	k=4
T = 50	97.1	95.0	92.4
T = 100	100.0	100.0	99.8
T = 200	100.0	100.0	100.0

Table 5: (L2) $\alpha_3 = -1$ $\gamma_1 = 0.1$ (U1) $\Theta = 0$

	k=1				k=2				k=4			
		Rejections				Rejections				Rejections		
	LR Test	(P1)	(P2)	(P3)	LR Test	(P1)	(P2)	(P3)	LR Test	(P1)	(P2)	(P3)
						T = 50						
r = 0	0.0	–	–	–	0.4	4.8	4.8	4.8	27.0	13.7	13.7	13.7
r = 1	91.6	27.8	27.8	27.8	89.8	15.5	14.4	12.9	61.1	23.2	23.2	22.3
r = 2	7.8	26.0	29.1	26.0	8.5	19.5	17.4	16.5	10.3	21.4	21.6	19.0
r = 3	0.7	30.3	30.3	30.3	1.3	15.9	15.9	15.9	1.5	25.0	25.0	25.0
Total	–	27.7	27.9	27.7	–	15.8	14.6	13.2	–	20.5	20.5	19.7
						T = 100						
r = 0	0.0	–	–	–	0.0	–	–	–	0.8	16.7	16.7	16.7
r = 1	92.7	45.6	45.6	45.6	92.1	20.7	16.1	16.5	90.0	16.9	16.6	14.2
r = 2	6.7	38.7	42.0	38.7	7.2	17.9	15.6	13.4	8.0	16.0	15.0	13.5
r = 3	0.6	46.9	46.9	46.9	0.8	10.5	10.5	10.5	1.1	17.9	17.9	17.9
Total	–	45.1	45.4	45.1	–	20.4	16.0	16.2	–	16.8	16.5	14.2
						T = 200						
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	93.2	72.1	72.1	72.1	93.3	30.0	23.9	24.7	93.4	22.0	20.3	17.8
r = 2	6.3	69.1	72.9	69.1	6.1	24.8	22.1	24.8	5.9	20.8	21.5	14.0
r = 3	0.5	60.0	60.0	60.0	0.6	16.7	16.7	16.7	0.7	10.8	10.8	10.8
Total	–	71.9	72.1	71.9	–	29.6	23.7	24.6	–	21.8	20.3	17.5

VAR(k-1) in differences
Rejections

	k=1	k=2	k=4
T = 50	–	10.6	15.2
T = 100	–	13.1	14.4
T = 200	–	19.1	19.2

VAR(k) in levels
Rejections

	k=1	k=2	k=4
T = 50	25.0	13.3	18.9
T = 100	43.0	15.4	16.3
T = 200	70.3	23.1	20.1

Table 6: (L3.1) $\alpha_3 = -0.1 \quad \gamma_1 = 0.1 \quad$ (U1) $\Theta = 0$

	k=1				k=2				k=4			
		Rejections				Rejections				Rejections		
	LR Test	(P1)	(P2)	(P3)	LR Test	(P1)	(P2)	(P3)	LR Test	(P1)	(P2)	(P3)
					T = 50							
r = 0	0.0	–	–	–	0.7	2.8	2.8	2.8	30.7	12.2	12.2	12.2
r = 1	92.2	12.3	12.3	15.8	90.6	9.7	8.9	8.9	57.7	18.4	18.5	20.2
r = 2	7.1	38.4	41.2	38.4	7.7	37.0	36.3	29.8	9.9	34.7	34.5	30.0
r = 3	0.7	32.4	32.4	32.4	1.0	32.0	32.0	32.0	1.6	44.4	44.4	44.4
Total	–	14.2	14.4	17.5	–	12.0	11.2	10.7	–	18.6	18.6	19.1
					T = 100							
r = 0	0.0	–	–	–	0.0	–	–	–	1.4	6.9	6.9	6.9
r = 1	92.9	23.8	23.8	25.6	92.5	9.7	9.4	8.6	90.5	10.3	9.8	9.9
r = 2	6.5	40.6	44.9	40.6	6.7	35.0	35.3	28.2	7.0	25.2	25.8	21.5
r = 3	0.6	32.3	32.3	32.3	0.8	27.5	27.5	27.5	1.1	17.0	17.0	17.0
Total	–	24.9	25.2	26.6	–	11.6	11.3	10.1	–	11.4	10.9	10.8
					T = 200							
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	94.1	43.2	43.2	43.8	93.9	12.7	14.2	11.3	93.0	8.5	8.6	8.7
r = 2	5.5	44.3	46.5	44.3	5.7	28.6	28.6	23.7	6.5	17.3	18.2	14.2
r = 3	0.5	41.7	41.7	41.7	0.4	10.0	10.0	10.0	0.5	22.2	22.2	22.2
Total	–	43.2	43.4	43.8	–	13.6	15.0	12.0	–	9.1	9.3	9.1

VAR(k-1) in differences

Rejections

	k=1	k=2	k=4
T = 50	–	6.3	11.4
T = 100	–	7.2	7.4
T = 200	–	8.4	6.8

VAR(k) in levels

Rejections

	k=1	k=2	k=4
T = 50	17.7	14.3	19.6
T = 100	19.5	12.7	11.6
T = 200	25.7	14.6	10.1

Table 7: (L3.2) $\alpha_3 = -0.3 \quad \gamma_1 = 0.3 \quad$ (U1) $\Theta = 0$

	k=1 LR Test	k=1 Rejections (P1)	k=1 Rejections (P2)	k=1 Rejections (P3)	k=2 LR Test	k=2 Rejections (P1)	k=2 Rejections (P2)	k=2 Rejections (P3)	k=4 LR Test	k=4 Rejections (P1)	k=4 Rejections (P2)	k=4 Rejections (P3)
T = 50												
r = 0	0.0	–	–	–	0.4	0.0	0.0	0.0	29.5	11.9	11.9	11.9
r = 1	92.1	73.9	73.9	74.7	91.0	23.8	23.2	23.7	58.8	26.3	26.7	27.1
r = 2	7.0	48.1	49.0	48.1	7.7	29.3	29.5	24.9	10.1	29.6	30.6	26.4
r = 3	0.9	27.9	27.9	27.9	0.8	19.0	19.0	19.0	1.6	22.0	22.0	22.0
Total	–	71.7	71.7	72.4	–	24.1	23.5	23.6	–	22.3	22.7	22.5
T = 100												
r = 0	0.0	–	–	–	0.0	–	–	–	1.1	12.3	12.3	12.3
r = 1	93.2	95.8	95.8	95.8	92.5	36.6	37.2	36.7	90.4	21.7	21.0	23.1
r = 2	6.1	65.0	65.4	65.0	6.9	36.4	35.9	32.4	7.8	22.9	21.4	19.8
r = 3	0.6	53.1	53.1	53.1	0.6	20.0	20.0	20.0	0.7	30.6	30.6	30.6
Total	–	93.6	93.6	93.7	–	36.5	37.0	36.3	–	21.7	21.0	22.8
T = 200												
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	94.3	99.9	99.9	99.9	94.0	62.7	63.2	64.1	93.0	30.7	29.6	35.5
r = 2	5.2	87.4	87.4	87.4	5.5	57.7	58.4	52.9	6.2	34.5	33.2	33.5
r = 3	0.5	100.0	100.0	100.0	0.5	63.0	63.0	63.0	0.8	19.0	19.0	19.0
Total	–	99.3	99.3	99.3	–	62.4	62.9	63.5	–	30.9	29.7	35.3

VAR(k-1) in differences

Rejections

	k=1	k=2	k=4
T = 50	–	6.3	12.0
T = 100	–	6.3	9.8
T = 200	–	6.7	9.7

VAR(k) in levels

Rejections

	k=1	k=2	k=4
T = 50	47.3	21.7	21.2
T = 100	77.6	33.7	20.5
T = 200	98.1	59.6	28.3

Table 8: (N1) $\alpha_3 = 0$ $\gamma_1 = 1$ (U2a) $\theta_{13} = 1$

	k=1				k=2				k=4			
		Rejections				Rejections				Rejections		
	LR Test	(P1)	(P2)	(P3)	LR Test	(P1)	(P2)	(P3)	LR Test	(P1)	(P2)	(P3)
T = 50												
r = 0	0.0	–	–	–	0.0	–	–	–	0.2	90.0	90.0	90.0
r = 1	87.7	5.3	5.3	5.8	90.4	11.4	15.7	10.8	87.8	15.9	19.0	15.7
r = 2	11.6	29.6	29.6	29.6	8.9	40.6	41.1	39.7	10.5	40.4	40.8	44.8
r = 3	0.6	25.8	25.8	25.8	0.7	40.5	40.5	40.5	1.5	48.6	48.6	48.6
Total	–	8.2	8.2	8.7	–	14.2	18.1	13.6	–	19.1	21.9	19.4
T = 100												
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	87.7	3.8	3.8	4.1	91.5	15.7	23.6	16.6	91.4	9.3	10.5	7.8
r = 2	11.6	25.7	25.7	25.7	7.8	49.0	50.3	47.2	7.7	34.6	34.6	40.4
r = 3	0.7	41.7	41.7	41.7	0.7	42.9	42.9	42.9	0.9	39.1	39.1	39.1
Total	–	6.6	6.6	6.9	–	18.5	25.8	19.2	–	11.5	12.6	10.6
T = 200												
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	85.3	3.4	3.4	3.6	91.9	26.9	38.1	28.4	92.7	6.8	8.6	5.5
r = 2	13.8	26.9	26.6	26.6	7.3	55.3	58.1	52.6	6.7	38.7	39.0	46.4
r = 3	0.9	28.3	28.3	28.3	0.8	47.4	47.4	47.4	0.5	44.4	44.4	44.4
Total	–	6.8	6.8	7.0	–	29.1	39.7	30.3	–	9.1	10.9	8.4

VAR(k-1) in differences

Rejections

	k=1	k=2	k=4
T = 50	–	85.9	90.2
T = 100	–	99.1	99.9
T = 200	–	100.0	100.0

VAR(k) in levels

Rejections

	k=1	k=2	k=4
T = 50	16.5	22.9	22.5
T = 100	16.4	27.7	16.2
T = 200	16.9	39.8	14.3

Table 9: (N2) $\alpha_3 = -1$ $\gamma_1 = 0$ (U2a) $\theta_{13} = 1$

	k=1				k=2				k=4			
	LR Test	Rejections (P1)	(P2)	(P3)	LR Test	Rejections (P1)	(P2)	(P3)	LR Test	Rejections (P1)	(P2)	(P3)
					T = 50							
r = 0	0.0	–	–	–	0.0	–	–	–	2.6	9.2	9.2	9.2
r = 1	24.0	36.7	36.7	36.8	62.0	22.8	18.4	18.8	84.5	14.4	15.9	13.7
r = 2	72.0	8.5	9.0	8.5	35.8	15.0	14.4	12.0	11.7	19.9	21.0	16.9
r = 3	3.9	14.2	14.2	14.2	2.2	13.5	13.5	13.5	1.1	15.8	15.8	15.8
Total	–	15.5	15.8	15.5	–	19.8	16.8	16.2	–	15.0	16.3	14.0
					T = 100							
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	18.2	66.2	66.2	66.3	56.3	31.2	24.3	25.7	88.0	8.8	9.7	8.9
r = 2	77.9	16.2	16.9	16.2	42.0	15.7	13.5	12.1	11.1	12.6	12.5	11.2
r = 3	3.9	27.5	27.5	27.5	1.7	8.1	8.1	8.1	0.9	4.3	4.3	4.3
Total	–	25.8	26.3	25.8	–	24.3	19.5	19.7	–	9.2	9.9	9.1
					T = 200							
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	15.5	96.4	96.4	96.4	54.1	50.6	42.0	45.4	89.1	7.2	7.2	7.2
r = 2	80.7	44.9	45.0	44.9	43.7	24.6	21.0	21.0	10.4	7.1	7.7	5.4
r = 3	3.8	56.8	56.8	56.8	2.2	35.8	35.8	35.8	0.6	10.7	10.7	10.7
Total	–	53.3	53.4	53.3	–	38.9	32.7	34.6	–	7.2	7.3	7.0

VAR(k-1) in differences
Rejections

	k=1	k=2	k=4
T = 50	–	10.8	11.1
T = 100	–	14.5	7.7
T = 200	–	21.7	6.9

VAR(k) in levels
Rejections

	k=1	k=2	k=4
T = 50	14.2	14.3	14.6
T = 100	25.5	16.5	9.4
T = 200	52.6	29.7	7.2

Table 10: (N3) $\alpha_3 = 0$ $\gamma_1 = 0$ (U2a) $\theta_{13} = 1$

	k=1				k=2				k=4			
		Rejections				Rejections				Rejections		
	LR Test	(P1)	(P2)	(P3)	LR Test	(P1)	(P2)	(P3)	LR Test	(P1)	(P2)	(P3)
T = 50												
r = 0	0.0	–	–	–	0.0	–	–	–	5.8	10.7	10.7	10.7
r = 1	81.2	2.6	2.6	5.7	91.3	7.4	7.9	5.4	83.9	12.2	12.2	9.6
r = 2	18.1	39.9	39.9	39.9	8.0	45.5	45.5	42.0	8.9	51.7	51.5	49.0
r = 3	0.7	22.2	22.2	22.2	0.7	45.9	45.9	45.9	1.4	45.8	45.8	45.8
Total	–	9.5	9.5	12.0	–	10.7	11.2	8.7	–	16.1	16.1	13.7
T = 100												
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	77.1	1.8	1.8	3.3	91.1	5.5	7.0	4.0	92.2	8.4	8.5	5.5
r = 2	21.6	39.0	39.0	39.0	8.3	43.2	44.0	41.1	7.0	41.6	41.9	44.4
r = 3	1.3	14.3	14.3	14.3	0.6	30.0	30.0	30.0	0.8	33.3	33.3	33.3
Total	–	10.0	10.0	11.2	–	8.8	10.2	7.2	–	10.9	11.0	8.4
T = 200												
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	75.2	1.9	1.9	2.8	90.7	5.2	7.5	5.2	93.1	6.7	6.9	4.1
r = 2	23.4	41.0	40.8	40.8	8.7	43.9	43.9	41.9	6.2	33.7	34.0	41.0
r = 3	1.3	13.4	13.4	13.4	0.6	28.6	28.6	28.6	0.7	36.4	36.4	36.4
Total	–	11.2	11.1	11.8	–	8.7	10.8	8.6	–	8.6	8.8	6.6

VAR(k-1) in differences
Rejections

	k=1	k=2	k=4
T = 50	–	6.8	11.7
T = 100	–	7.0	7.8
T = 200	–	8.1	6.5

VAR(k) in levels
Rejections

	k=1	k=2	k=4
T = 50	20.7	16.8	22.8
T = 100	20.8	16.1	15.2
T = 200	20.8	16.1	12.9

Table 11: (L1) $\alpha_3 = 0.1$ $\gamma_1 = 1$ (U2a) $\theta_{13} = 1$

	k=1				k=2				k=4			
		Rejections				Rejections				Rejections		
	LR Test	(P1)	(P2)	(P3)	LR Test	(P1)	(P2)	(P3)	LR Test	(P1)	(P2)	(P3)
T = 50												
r = 0	0.0	–	–	–	0.0	–	–	–	0.2	100.0	100.0	100.0
r = 1	86.4	99.4	99.4	99.5	89.9	100.0	100.0	100.0	86.6	97.0	100.0	92.6
r = 2	12.8	99.5	99.5	99.5	9.2	100.0	100.0	99.6	11.6	98.3	100.0	96.2
r = 3	0.8	100.0	100.0	100.0	0.9	100.0	100.0	100.0	1.7	100.0	100.0	100.0
Total	–	99.4	99.4	99.5	–	100.0	100.0	99.9	–	97.2	100.0	93.1
T = 100												
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	86.4	100.0	100.0	100.0	91.4	100.0	100.0	100.0	91.4	100.0	100.0	100.0
r = 2	13.1	100.0	100.0	100.0	7.9	100.0	100.0	100.0	7.6	100.0	100.0	100.0
r = 3	0.5	100.0	100.0	100.0	0.8	100.0	100.0	100.0	1.0	100.0	100.0	100.0
Total	–	100.0	100.0	100.0	–	100.0	100.0	100.0	–	100.0	100.0	100.0
T = 200												
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	85.1	100.0	100.0	100.0	92.1	100.0	100.0	100.0	93.1	100.0	100.0	100.0
r = 2	14.0	100.0	100.0	100.0	7.1	100.0	100.0	100.0	6.4	100.0	100.0	100.0
r = 3	0.9	100.0	100.0	100.0	0.9	100.0	100.0	100.0	0.5	100.0	100.0	100.0
Total	–	100.0	100.0	100.0	–	100.0	100.0	100.0	–	100.0	100.0	100.0

VAR(k-1) in differences
Rejections

	k=1	k=2	k=4
T = 50	–	95.8	99.9
T = 100	–	100.0	100.0
T = 200	–	100.0	100.0

VAR(k) in levels
Rejections

	k=1	k=2	k=4
T = 50	99.8	100.0	100.0
T = 100	100.0	100.0	100.0
T = 200	100.0	100.0	100.0

Table 12: (L2) $\alpha_3 = -1$ $\gamma_1 = 0.1$ (U2a) $\theta_{13} = 1$

	k=1				k=2				k=4			
	LR Test	Rejections (P1)	(P2)	(P3)	LR Test	Rejections (P1)	(P2)	(P3)	LR Test	Rejections (P1)	(P2)	(P3)
T = 50												
r = 0	0.0	–	–	–	0.0	–	–	–	1.6	18.3	18.3	18.3
r = 1	25.3	4.6	4.6	4.7	64.2	9.1	7.4	6.7	84.6	17.5	17.9	15.4
r = 2	69.8	18.3	20.0	18.3	33.8	16.8	16.4	14.2	12.3	18.5	19.0	14.1
r = 3	4.9	7.7	7.7	7.7	2.0	9.8	9.8	9.8	1.5	17.6	17.6	17.6
Total	–	14.3	15.5	14.4	–	11.7	10.5	9.3	–	17.6	18.1	15.3
T = 100												
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	21.2	5.3	5.3	5.3	60.1	8.0	6.8	6.3	88.9	15.3	15.2	13.0
r = 2	75.0	19.2	20.2	19.2	38.2	13.8	12.4	12.1	10.1	17.2	17.9	14.4
r = 3	3.8	8.5	8.5	8.5	1.7	5.7	5.7	5.7	0.9	12.8	12.8	12.8
Total	–	15.8	16.6	15.8	–	10.1	8.9	8.5	–	15.4	15.5	13.2
T = 200												
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	17.0	5.3	5.3	5.3	55.3	7.9	6.4	5.6	88.4	19.3	18.1	15.8
r = 2	79.5	17.2	17.7	17.2	42.9	12.3	10.4	9.5	10.9	20.6	20.2	16.4
r = 3	3.5	10.2	10.2	10.2	1.8	13.6	13.6	13.6	0.7	18.9	18.9	18.9
Total	–	14.9	15.3	14.9	–	9.9	8.2	7.4	–	19.5	18.3	15.9

VAR(k-1) in differences
Rejections

	k=1	k=2	k=4
T = 50	–	6.6	14.3
T = 100	–	6.4	12.6
T = 200	–	5.2	15.9

VAR(k) in levels
Rejections

	k=1	k=2	k=4
T = 50	14.7	11.2	17.7
T = 100	15.6	8.9	14.2
T = 200	14.3	8.5	17.8

Table 13: (L3.1) $\alpha_3 = -0.1$ $\gamma_1 = 0.1$ (U2a) $\theta_{13} = 1$

	k=1				k=2				k=4			
	LR Test	Rejections (P1)	(P2)	(P3)	LR Test	Rejections (P1)	(P2)	(P3)	LR Test	Rejections (P1)	(P2)	(P3)
T = 50												
r = 0	0.0	–	–	–	0.0	–	–	–	3.4	11.9	11.9	11.9
r = 1	81.0	22.0	22.0	34.8	91.1	9.4	8.1	8.8	85.2	13.5	12.9	12.6
r = 2	18.3	31.9	32.6	31.9	8.4	39.5	38.3	36.4	10.2	38.0	38.2	38.8
r = 3	0.7	8.8	8.8	8.8	0.5	26.9	26.9	26.9	1.2	29.5	29.5	29.5
Total	–	23.7	23.8	34.1	–	12.0	10.7	11.2	–	16.2	15.6	15.5
T = 100												
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	77.9	67.2	67.2	78.2	90.7	11.6	8.7	10.7	91.9	9.7	8.7	9.1
r = 2	21.2	32.0	32.1	32.0	8.6	40.2	39.7	34.3	7.5	31.0	30.7	28.3
r = 3	0.9	20.5	20.5	20.5	0.7	18.9	18.9	18.9	0.6	16.7	16.7	16.7
Total	–	59.3	59.3	67.9	–	14.1	11.4	12.7	–	11.4	10.4	10.6
T = 200												
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	74.6	99.4	99.4	99.7	91.4	15.9	13.4	14.6	93.5	8.3	8.1	8.1
r = 2	24.4	30.7	30.7	30.7	8.0	30.9	32.2	29.4	5.9	20.2	20.5	21.5
r = 3	1.0	14.6	14.6	14.6	0.5	22.2	22.2	22.2	0.5	14.8	14.8	14.8
Total	–	81.8	81.8	82.1	–	17.2	14.9	15.8	–	9.1	8.8	9.0

VAR(k-1) in differences
Rejections

	k=1	k=2	k=4
T = 50	–	7.4	10.4
T = 100	–	7.1	7.6
T = 200	–	7.2	5.9

VAR(k) in levels
Rejections

	k=1	k=2	k=4
T = 50	21.8	15.7	18.5
T = 100	28.3	14.2	12.6
T = 200	44.2	15.2	9.9

Table 14: (L3.2) $\alpha_3 = -0.3$ $\gamma_1 = 0.3$ (U2a) $\theta_{13} = 1$

	k=1				k=2				k=4			
		Rejections				Rejections				Rejections		
	LR Test	(P1)	(P2)	(P3)	LR Test	(P1)	(P2)	(P3)	LR Test	(P1)	(P2)	(P3)
T = 50												
r = 0	0.0	–	–	–	0.0	–	–	–	1.0	8.3	8.3	8.3
r = 1	72.0	7.1	7.1	9.3	88.5	18.0	16.2	17.1	86.6	23.4	22.3	23.0
r = 2	26.8	19.4	23.9	19.4	10.6	29.3	28.9	26.1	11.1	30.0	30.0	25.3
r = 3	1.2	8.2	8.2	8.2	0.9	15.9	15.9	15.9	1.4	26.5	26.5	26.5
Total	–	10.4	11.6	12.0	–	19.2	17.6	18.1	–	24.0	23.1	23.1
T = 100												
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	67.9	9.2	9.2	9.6	87.9	30.2	28.9	29.2	91.6	26.8	25.5	30.4
r = 2	30.5	16.4	20.9	16.4	11.4	29.4	29.4	26.2	7.5	26.3	26.3	26.3
r = 3	1.5	6.5	6.5	6.5	0.7	24.2	24.2	24.2	0.8	26.2	26.2	26.2
Total	–	11.4	12.8	11.6	–	30.1	28.9	28.8	–	26.7	25.5	30.1
T = 200												
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	62.8	12.4	12.4	12.6	86.7	52.2	52.0	53.0	93.6	41.5	40.3	50.6
r = 2	35.8	13.9	18.2	13.9	12.6	37.9	38.1	35.2	5.8	39.9	39.2	44.1
r = 3	1.4	8.8	8.8	8.8	0.7	40.0	40.0	40.0	0.6	25.8	25.8	25.8
Total	–	12.9	14.4	13.0	–	50.4	50.1	50.6	–	41.3	40.2	50.0

VAR(k-1) in differences
Rejections

	k=1	k=2	k=4
T = 50	–	6.4	11.7
T = 100	–	5.8	8.3
T = 200	–	6.4	7.1

VAR(k) in levels
Rejections

	k=1	k=2	k=4
T = 50	12.4	16.6	22.0
T = 100	11.9	24.7	23.5
T = 200	12.5	41.0	37.4

Table 15: (N1) $\alpha_3 = 0 \quad \gamma_1 = 1 \qquad$ (U2a) $\theta_{13} = 1$

	k=4 LR Test	k=4 Rejections (P1)	k=4 Rejections (P2)	k=4 Rejections (P3)	k=6 LR Test	k=6 Rejections (P1)	k=6 Rejections (P2)	k=6 Rejections (P3)	k=8 LR Test	k=8 Rejections (P1)	k=8 Rejections (P2)	k=8 Rejections (P3)
T = 50												
r = 0	0.2	90.0	90.0	90.0	8.0	75.8	75.8	75.8	19.9	76.0	76.0	76.0
r = 1	87.8	15.9	19.0	15.7	74.4	27.9	38.5	28.8	58.0	48.5	60.9	46.1
r = 2	10.5	40.4	40.8	44.4	15.5	51.3	52.2	52.6	18.2	65.6	66.0	63.2
r = 3	01.5	48.6	48.6	48.6	2.1	49.0	49.0	49.0	3.9	62.9	62.9	62.9
Total	–	19.1	21.9	19.4	–	35.8	43.8	36.6	–	57.6	64.9	55.8
T = 100												
r = 0	0.0	–	–	–	0.0	100.0	100.0	100.0	1.8	91.3	91.3	91.3
r = 1	91.4	9.3	10.5	7.8	89.5	14.3	18.6	14.6	86.7	18.7	28.9	19.7
r = 2	7.7	34.6	34.6	40.4	9.3	34.8	35.1	41.9	10.1	43.3	44.1	45.1
r = 3	0.9	39.1	39.1	39.1	1.2	35.6	35.6	35.6	1.3	36.9	36.9	36.9
Total	–	11.5	12.6	10.6	–	16.5	20.4	17.8	–	22.7	31.7	23.8
T = 200												
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	92.7	6.8	8.6	5.5	92.5	8.9	10.0	8.5	92.5	11.3	15.8	12.8
r = 2	6.7	38.7	39.0	46.4	6.7	30.3	30.0	41.1	6.7	33.6	32.7	42.6
r = 3	0.5	44.4	44.4	44.4	0.8	24.4	24.4	24.4	0.7	24.3	24.3	24.3
Total	–	9.1	10.9	8.4	–	10.4	11.5	10.8	–	12.9	17.0	14.8

VAR(k-1) in differences

Rejections

	k=4	k=6	k=8
T = 50	90.2	79.7	74.1
T = 100	99.9	98.3	89.8
T = 200	100.0	100.0	100.0

VAR(k) in levels

Rejections

	k=4	k=6	k=8
T = 50	22.5	34.2	50.5
T = 100	16.2	18.6	23.4
T = 200	14.3	13.2	14.1

Table 16 : (N1) $\alpha_3 = 0 \quad \gamma_1 = 1$ (U2b) $\theta_{13} = -1$ (S2) $\Sigma_\epsilon = \begin{pmatrix} 1.0 & 0.2 & -0.5 \\ 0.2 & 1.0 & 0.2 \\ -0.5 & 0.2 & 1.0 \end{pmatrix}$

	k=1				k=2				k=4			
		Rejections				Rejections				Rejections		
	LR Test	(P1)	(P2)	(P3)	LR Test	(P1)	(P2)	(P3)	LR Test	(P1)	(P2)	(P3)
T = 50												
r = 0	0.0	–	–	–	0.3	92.3	92.3	92.3	28.1	44.2	44.2	44.2
r = 1	24.9	31.1	31.1	32.6	65.2	42.0	47.1	45.7	61.1	50.2	52.3	56.5
r = 2	72.2	56.4	56.4	56.4	32.7	22.1	22.1	11.2	9.7	20.3	20.3	18.1
r = 3	2.9	46.2	46.2	46.2	1.8	19.8	19.8	19.8	1.0	17.6	17.6	17.6
Total	–	49.8	49.8	50.2	–	35.2	38.5	34.0	–	45.3	46.5	48.9
T = 100												
r = 0	0.0	–	–	–	0.0	–	–	–	0.4	85.0	85.0	85.0
r = 1	16.4	24.2	24.2	24.8	54.5	42.4	50.8	43.5	88.6	25.6	27.3	29.2
r = 2	80.0	66.9	66.9	66.9	43.7	27.1	27.2	14.8	10.1	15.2	15.2	15.0
r = 3	3.5	50.6	50.6	50.6	1.8	34.1	34.1	34.1	0.8	7.1	7.1	7.1
Total	–	59.3	59.3	59.4	–	35.5	40.2	30.8	–	24.6	26.1	27.8
T = 200												
r = 0	0.0	–	–	–	0.0	–	–	–	0.0	–	–	–
r = 1	12.0	21.2	21.2	21.5	51.1	55.2	65.6	55.9	88.7	16.4	18.3	17.7
r = 2	84.1	75.3	75.3	75.3	47.4	45.3	45.3	30.4	10.4	13.6	13.6	15.7
r = 3	3.9	61.5	61.5	61.5	1.5	35.5	35.5	35.5	0.9	4.7	4.7	4.7
Total	–	68.3	68.3	68.3	–	50.2	55.5	43.5	–	16.0	17.7	17.4

VAR(k-1) in differences

Rejections

	k=1	k=2	k=4
T = 50	–	70.7	40.1
T = 100	–	94.4	61.1
T = 200	–	99.9	90.6

VAR(k) in levels

Rejections

	k=1	k=2	k=4
T = 50	42.0	20.2	18.0
T = 100	55.6	27.6	11.7
T = 200	65.8	43.9	11.3

References

[1] Engle, R. F. and C. W. J. Granger, 1987, "Co-integration and Error Correction : Representation, Estimation, and Testing", *Econometrica*, 55, pp251–276

[2] Johansen, S., 1988, "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics and Control*, 12, pp231–254

[3] Johansen, S. , 1989, "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models", University of Copenhagen, Institute of Mathematical Statistics Preprint No. 3

[4] Phillips, P. C. B., 1991 , "Optimal Inference in Cointegrated Systems", *Econometrica*, 59, 283–306

[5] Sims, C. A., J. H. Stock and M. W. Watson, 1990, "Inference in Linear Time Series Models with Some Unit Roots", *Econometrica* 58, 113–144

[6] Toda, H. Y. and P. C. B. Phillips, 1991, "Vector Autoregressions and Causality", Cowles Foundation Discussion Paper No. 977