# Department of Agricultural Economics & Agribusiness

## Monte Carlo Evidence on Cointegration and Causation

**Hector O. Zapata**

Department of Agricultural Economics and Agribusiness
Louisiana State University Agricultural Center
Baton Rouge, Louisiana 70803-5604
Email: xp2647@lsuvm.sncc.lsu.edu.

**Alicia N. Rambaldi**

Department of Econometrics
University of New England
Armidale, NSW 2351. Australia
Email: arambald@metz.une.edu.au.

Staff Paper
96-08

**Louisiana State University Agricultural Center**, H. Rouse Caffey, Chancellor
**Louisiana Agricultural Experiment Station**, R. Larry Rogers, Vice Chancellor and Director

# Monte Carlo Evidence on Cointegration and Causation

Hector O. Zapata
and
Alicia N. Rambaldi

## Abstract

The small sample performance of Granger causality tests under different model dimensions, degree of cointegration, direction of causality, and system stability are presented. Two tests based on maximum likelihood estimation of error-correction models (LR and WALD) are compared to a Wald test based on multivariate least squares estimation of a modified VAR (MWALD). In large samples all test statistics perform well in terms of size and power. For smaller samples, the LR and WALD tests perform better than the MWALD test. Overall, the LR test outperforms the other two in terms of size and power in small samples.

*Keywords*: Causality tests; Cointegration; Likelihood ratio; Wald statistic; Monte Carlo Experiments.

*JEL classification*: C32

This paper was presented at the 1996 American Agricultural Economics Association meeting is forthcoming in the Oxford Bulletin of Economics and Statistics, 1997. 

---

Hector Zapata. Department of Agricultural Economics and Agribusiness. Louisiana State University. Baton Rouge, LA 70803. USA. Ph.: 1 (504) 388 2766, Fax: 1 (504) 388 2716, Email: xp2647@lsuvm.sncc.lsu.edu.
Alicia Rambaldi. Department of Econometrics. University of New England. Armidale, NSW 2351. Australia. Ph.: 61 (67) 73 2386, Fax: 61 (67) 73 3607, Email: arambald@metz.une.edu.au.

# Monte Carlo Evidence on Cointegration and Causation

## 1. INTRODUCTION

Testing Granger non-causality in cointegrated time series has been the subject of considerable recent research. The first result that naturally emerged on this subject was the existence of "long-run" causality in at least one direction (Granger, 1988) where cointegration was represented by a bivariate error-correction model. The extension of this result to more than two variables was fairly straightforward under the existence of one cointegrating relation. In fact, the two-step procedure introduced by Engle and Granger (1987) was all that was needed to test non-causality hypotheses. As the literature presented below illustrates, the dimension of the cointegration space complicates this testing problem considerably. But recent developments in cointegration theory have solved important questions in what still is a somewhat controversial issue.

In the empirical literature the Wald test computed from an unrestricted vector autoregressive (VAR) model appears frequently. Toda and Phillips (1993) show that the asymptotic distribution of the test in the unrestricted case involves nuisance parameters and nonstandard distributions. An alternative procedure to the estimation of an unrestricted VAR consists of transforming an estimated error correction model (ECM) to its levels VAR form and then applying the Wald type test for linear restrictions to the resulting VAR model. Lütkepohl and Reimers (1992a) present the distribution of the Wald statistic for the bivariate case based on Johansen and Juselius' (1990) maximum likelihood estimator of ECMs. The limiting distribution of the statistic for the p-variates model is discussed in Toda and Phillips (1993). Toda and Yamamoto (1995) propose an interesting yet simple procedure requiring the estimation of an "augmented" VAR, even when there is cointegration, which guarantees the asymptotic distribution of the Wald statistic. An analysis and Monte Carlo results for cointegrated data is presented in Dolado and Lütkepohl (*forthcoming*).

Mosconi and Giannini (1992) suggest that it is possible to achieve an efficiency gain by imposing the cointegrating constraints under both the null and alternative hypotheses while testing for non-causality in cointegrated systems. The test statistic proposed by Mosconi and Giannini is a likelihood ratio (LR). The limited Monte Carlo evidence provided in their study lends support for this approach. However, in practice the computation of the LR test is considerably more cumbersome than any of the Wald versions of the test (detailed discussion in Section 3).

The paper presents results of a Monte Carlo experiment designed to study the performance of two Wald and a likelihood ratio tests for Granger non-causality in bivariate and trivariate cointegrated systems. Estimation and testing for two of the tests follows the maximum likelihood approach of Johansen (1988) and Johansen and Juselius (1990). The third test, which serves as a benchmark for comparing test performance, is computed from the multivariate least squares estimates of a VAR (Toda and Yamamoto (1995) and Dolado and Lütkepohl (1994)).

Section two introduces the model and establishes the notation. The alternative tests for Granger non-causality in cointegrated systems are the subject of Section three. Section four explains the experiment and presents the results. Section five contains the conclusions.

## 2. MODEL AND NOTATION

The basic VAR model for p variables and k lags with Gaussian errors is given by

$$\Phi(L)Z_t = \begin{bmatrix} \Phi_{11}(L) & \Phi_{12}(L) \\ \Phi_{21}(L) & \Phi_{22}(L) \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \mathbf{e_t} \qquad t=1,...,T \qquad (1)$$

where $e_1, ...., e_T$ are i.i.d. N ~ $(0,\Sigma)$, and the maximum lag in $\Phi(L)$ is k, $\mathbf{y}_t$ consists of $p_1$ variables and $\mathbf{x}_t$ of $p_2$ variables. We omit deterministic components for simplicity. In error-correction form this model can be expressed as

$$\Delta Z_t = \Gamma_1 \Delta Z_{t-1} + \ldots + \Gamma_{k-1} \Delta Z_{t-k+1} - \Pi Z_{t-k} + e_t \qquad (2)$$

where

$$\Gamma_i = -(I_p - \Phi_1 - ... - \Phi_i), \quad i = 1,...., k-1,$$

and

$$\Pi = I_p - \Phi_1 - ... - \Phi_K,$$

which using compact matrix notation reduces to

$$Z_o = \mathbf{\Gamma} Z_1 + \Pi Z_K + E \qquad (3)$$

where $Z_o$ is a p x T matrix of observations on first differences of $Z_t$, $Z_1$ contains the lagged differences, $Z_K$ is the kth lag of $Z_t$, $\mathbf{\Gamma}$ is a (p x (k-1)*p) matrix of the stacked $\Gamma_i$s, and E is the p x T matrix of disturbances for the p equations in the system.

The case of interest in this experiment is when there is cointegration, that is, when the rank of $\Pi$ equals r < p. This hypothesis is formulated as

$$H_1(r): \Pi = \alpha\beta' \tag{4}$$

where $\alpha$ and $\beta$ are p x r matrices, and r is the number of cointegrating relations $\beta' Z_t$. This restriction also provides some insight into the causality implications of cointegration whereby causality can come about through the cointegrating relations $\beta' Z_t$ or by conditioning on $\alpha$ such that a row of $\alpha$ equating to zero essentially excludes "long- run causality" in the corresponding equation.

The maximum likelihood estimation of this multivariate cointegration model follows a reduced rank regression (RRR) due to Johansen (1988) and Johansen and Juselius (1990, 1992), which for the concentrated (with respect to the parameters $\Gamma_i$, i=1, .., k-1) likelihood function can be expressed as

$$R_o = \alpha\beta' R_K + e_t \tag{5}$$

where $R_o$ are the residuals from the regression of $Z_o$ on $Z_1$, and $R_K$ are the residuals from the regression of $Z_K$ on $Z_1$ (refer to equation (3)).

In this paper we investigate Granger non-causality in bivariate and trivariate $Z_t$, i.e., $Z_t = [y_t\ x_{1t}]'$ and $Z_t = [y_t\ x_{1t}\ x_{2t}]'$ (refer to eq. (1)). In the trivariate case, the $Z_t$ are cointegrated if r = 1 or r = 2. We concentrate on r = 2, and more specifically, on the case when the cointegration space is made of two types of cointegrating vectors: those involving all variables, and those involving the **x**'s only. This setting has specific implications for the estimation of the model and the testing for Granger non-causality because of the presence of joint restrictions on $\alpha$ and $\beta$. Specific discussion is presented in Section 3. Following Mosconi and Giannini (1992), we define $p_1$ as the number of variables in $\mathbf{y}_t$ ($p_1$ = 1 in this study), and $p_2$ as the number of variables in $\mathbf{x}_t$.

## 3. TESTING GRANGER NON-CAUSALITY

This study explores three alternative tests for non-causality. Two of them are computed from the estimated parameters of the ECM representation (eq. (2)), while the third is computed using the estimated parameters from a VAR representation (eq.(1)). The latter is used as a benchmark for comparing test performance because of its simplicity.

## 3.1 NON-CAUSALITY TESTS USING AN ECM

### 3.1.1 Wald test

The first step towards testing causality in ECMs is to estimate the parameters in equation (2) by Johansen's maximum-likelihood. Using the parameterization in equation (1), $\mathbf{x}_t$ does not Granger-cause $\mathbf{y}_t$ if and only if the hypothesis (Lütkepohl (1993), p. 378):

$$H_o: \Phi_{12,i} = 0 \text{ for } i=1,2,...,k \tag{6}$$

is true, where $\Phi_{12}$ is the coefficient matrix on $\mathbf{x}_t$ in the $\mathbf{y}_t$ equations. In a bivariate system, for instance, $\Phi_{12,i}$ is the 1x1 coefficient on $x_t$ in the $y_t$ variable. Similarly, $\mathbf{y}_t$ does not Granger-cause $\mathbf{x}_t$ if and only if the corresponding $\Phi_{21,i}$ coefficients equal zero. Let $\phi = \text{vec}[\Phi_1,..., \Phi_k]$ be the vector of all VAR coefficients. Then, the test for the linear restrictions in eq. (6) is given by testing $H_o$: $R\phi = 0$ against $H_1$: $R\phi \neq 0$ for suitable chosen R. The Wald statistics for testing $H_o$ is

$$W = T\phi' R' (R \Sigma_\phi R')^{-1} R\phi \tag{7}$$

where R is N x $p^2k$, N is the rank of R and $\Sigma_\phi$ is the variance-covariance of $\phi$. W has a Chi-squared distribution with N degrees of freedom under $H_o$ if some conditions hold (refer to Toda and Phillips (1993), Section 4, Theorem 3). Suppose that we are interested in whether the $p_2$ elements of $\mathbf{x}_t$ are "not causing" the $p_1$ elements of $\mathbf{y}_t$. Then, for W to converge in distribution to a Chi-squared, it must be that rank($\beta_{p_2}$) = $p_2$ or rank($\alpha_{p_1}$) = $p_1$, where $\alpha$ and $\beta$ have been partitioned accordingly[1].

---

[1] Note that in practice this requires pre-testing of the ranks of $\beta_{p_2}$ and $\alpha_{p_1}$. The implementation of this test for cointegrated systems consists of: (a)Estimating $\alpha$, $\beta$, $\Gamma_i$ ( i=1, .., k-1), and $\Sigma_e$ by Johansen' s MLE, where r is taken to be the value determined by pre-testing for the rank of $\Pi$; (b)Transforming the ECM estimates to corresponding VAR in levels estimates. That is, reversing the transformation from equation (2) to equation (1). Lütkepohl and Reimers (1992b pp. 62-63) propose a one step formula to achieve this transformation;(c)Estimating $\Sigma_\phi$ consistently (see Lütkepohl and Reimers (1992b) pp. 63); (d) Computing W to accept or reject the null hypothesis of non-causality.

3.1.2 Likelihood ratio tests

A variety of tests of hypotheses on the parameters $\alpha$ and $\beta$ were introduced by Johansen and Juselius (1990,1992). Mosconi and Giannini (1992) proposed that there may be efficiency gains by applying the cointegration restrictions under both the null and the alternative hypotheses in testing for non-causality. The alternative hypothesis is that of cointegration $H_1(r)$ in equation (4). Under the null of non-causality ($\mathbf{y}_t$ does not Granger cause $\mathbf{x}_t$), we must consider not only $H_1(r)$ but also constraints on the parameter space defined by $\mathbf{\Gamma}$ and $\Pi$. The linear restrictions for non-causality are expressed as

$$H_G(r): B'\mathbf{\Gamma} V = 0, B'\Pi A = 0 \tag{8}$$

where $\mathbf{\Gamma}$ and $\Pi$ are the parameters of model (3), $B=[0' \ I_{p2}]'$ is p x $p_2$, $A=[I_{p1} \ 0']'$ is p x $p_1$, $V = (I_{(k-1)} \otimes A)$ is p(k-1) x $p_1$(k-1), $I_{pi}$ is an identity matrix of order $p_i$, and B'A = 0. The subscript G on H is used to indicate Granger causality. Note that the restrictions in (8) are equivalent to those in (6) except that the former are now expressed in terms of the ECM formulation. We define the known matrices as A and B to facilitate reference to Johansen and Juselius' notation (1990, 1992). The linear restrictions in (8) include both long- and short-run non-causality. Testing for long-run non-causality only (that is, causality through the error correction term only) could be expressed as:

$$H_{GL}(r): B'\Pi A = 0 \tag{9}$$

where the constraints are imposed on the parameter space defined by $\Pi$, and the subscript GL represents Granger non-causality in the long-run. Some of the specific implications for estimation and testing in this case are discussed below.

The hypothesis of Granger non-causality (G) conditional on cointegration (C) can be formulated as

$$H_{GC}(r): B'\mathbf{\Gamma} V = 0, B'\Pi A = 0, \Pi = \alpha\beta', \tag{10}$$

and the likelihood ratio (LR) test for this hypothesis is

$$LR = -2\ln\{L_{max}[H_{GC}(r)]/L_{max}[H_1(r)]\} \tag{11}$$

based on Johansen's results, Mosconi and Giannini (1992) indicate that (11) is asymptotically distributed as a Chi-squared and they compute the degrees of freedom to be: $pr - p_1r_1 - p_2r_2 - r_1r_2$ when k = 1 and $pr - p_1r_1 - p_2r_2 - r_1r_2 + p_1p_2(k - 1)$ when k > 1, where $r_1$ and $r_2$ are defined shortly.

Estimation of the ECM under $H_{GC}(r,r_1,r_2)$ follows classical restricted MLE estimation. This can be carried out using a switching algorithm proposed by Johansen and Juselius (1992) and Mosconi and Giannini (1992), whereby all restrictions are imposed. This estimation problem can be formulated by making use of Theorem 1 in Mosconi and Giannini (1992) which states that the hypothesis $B'\Pi A = 0$ holds true if and only if

$$\Pi = \alpha\beta' = \begin{bmatrix} \alpha_{11}\beta_{11}' & (\alpha_{11}\beta_{21}' + \alpha_{12}\beta_{22}') \\ \mathbf{0} & \alpha_{22}\beta_{22}' \end{bmatrix} = \begin{bmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{bmatrix} \quad (12).$$

The explicit partitioning of $\alpha$ and $\beta$ under the restriction $B'\Pi A = 0$ and $\Pi = \alpha\beta'$ is obtained by noting that[2] $B'A = 0$ implies

$$B'\Pi A = B'\alpha\beta' A = B'A\alpha_{11}\beta_1' A + B'\alpha_2\beta_{22}' B'A = 0$$

and $\alpha_2\beta_1' = 0$ implies

$$\alpha = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ 0 & \alpha_{22} \end{bmatrix} = [A\ \alpha_{11} \mid \alpha_2]$$

$$\beta = \begin{bmatrix} \beta_{11} & 0 \\ \beta_{21} & \beta_{22} \end{bmatrix} = [\beta_1 \mid B\beta_{22}]$$

where,

$\alpha_{11}$ is $p_1$ x $r_1$, $\alpha_2$ is p x $r_2$, $\beta_1$ is p x $r_1$, $\beta_{22}$ is $p_2$ x $r_2$, and $r = r_1 + r_2$. In this partition, $r_1 = \text{rank}(\Pi_{11})$, and $r_2 = \text{rank}(\Pi_{22})$. Using this partition and the compact notation from eq. (3), the model to be estimated under the null hypothesis of $H_{GC}(r)$, or more specifically $H_{GC}(r,r_1,r_2)$, is:

$$\mathbf{Z_o} = \mathbf{\Gamma Z_1} + A\alpha_{11}\beta_1'\mathbf{Z_K} + \alpha_2\beta_{22}'B'\mathbf{Z_K} + \mathbf{E}. \quad (13)$$

If there were no restrictions on $\mathbf{\Gamma}$, $\alpha$ and $\beta$ may be estimated using RRR as shown in equation (5). That is, if the hypothesis of interest were $H_{GLC}(r,r_1,r_2)$: $B'\Pi A = 0$, $\Pi = \alpha\beta'$, we can write the RRR (concentrating on $\mathbf{\Gamma}$) as

---

[2] We would like to thank an anonymous reviewer for pointing this out.

$$R_{ot} = A\alpha_{11}\beta_1' R_{kt} + \alpha_2\beta_{22}' B' R_{kt} + e_t \qquad (14).$$

The experiment presented in this paper refers to the case in eq. (13)[3].

Estimation of eq. (13) includes restrictions on $\alpha$ ($A\alpha_{11}$) and restrictions on $\beta$ ($B\beta_{22}$) and the problem does not easily reduce to the usual eigenvalue problem[4].

## 3.2 NON-CAUSALITY TESTS USING A VAR

### 3.2.1 Wald test

Toda and Yamamoto (1995) prove that the Wald test for restrictions on the parameters of a VAR(k) has an asymptotic $\chi^2$ distribution when a VAR(k + $d_{max}$) is estimated, where $d_{max}$ is the

---

3 An illustration of the estimation steps for the case in eq. (14) can be found in Zapata and Rambaldi (1995).

4 Adopting a modification of a switching algorithm proposed by Johansen and Juselius(1992), the estimation problem reduces to: (a)Initialization: Set $\alpha_{11}$ and $\beta_1$ and $\mathbf{\Gamma}$ equal to zero in : $(Z_o - \Gamma Z_1 - A\hat{\alpha}_{11}\hat{\beta}_1' Z_k) = \alpha_2\beta_{22}' B' Z_k + E$; (b) Solve for $\alpha_2$ and $\beta_{22}$. This requires solving the usual eigenvalue problem $|\lambda B' S_{kk} B - B' S_{k0} S_{00}^{-1} S_{0k} B|$. Calculate $\hat{\alpha}_2$ and $\hat{\beta}_{22}$ the usual way (refer to Johansen and Juselius (1990), p. 193); (c) Fix $\beta_{22}$ and $\alpha_2$ at the values in (b), condition on $\hat{\alpha}_2\hat{\beta}_{22}' B' \mathbf{Z}_k$ to obtain $(\mathbf{Z}_o - \mathbf{\Gamma Z}_1 - \hat{\alpha}_2\hat{\beta}_{22}' B' Z_k) = A\alpha_{11}\beta_1' Z_k + E$ and solve the eigenvalue problem:
$|\lambda S_{kk.b} - S_{ka.b} S_{aa.b}^{-1} S_{ak.b}| = 0$; The cross-product moment matrices conditioned on $\hat{\beta}_{22}$ are: $S_{bb} = B' S_{00} B$, $S_{ab} = A' S_{00} B$, $S_{kb} = S_{k0} B$, $S_{aa.b} = (A' R_0 - G_1 x_1)(A' R_0 - G_1 x_1)' / (T-k)$, $S_{ak.b} = ((A' R_0 - G_1 x_1)(R_k - G_2 x_1)' / (T-k)$, $S_{ka.b} = (R_k - G_2 x_2)(A' R_0 - G_1 x_1)' / (T-k)$, $S_{kk.b} = (R_k - G_2 x_1)(R_k - G_2 x_1)' / (T-k)$, where, $x_1 = B' R_0$, $G_1 = (A' R_0) x_1' (x_1 x_1')^{-1}$, $G_2 = R_k x_1' (x_1 x_1')^{-1}$. The solution to this eigenvalue problem generates $\hat{\beta}_1$ and $\hat{\alpha}_{11}$ . Note that $S_{kk.b}$ may be singular, hence Lemma 1 (Johansen and Juselius, 1992) shows how to solve the problem by first diagonalizing $S_{kk.b}$; (d) Finally, given estimates of $\beta_1$, $\alpha_{11}$, $\alpha_2$ and $\beta_{22}$, the estimator for $\mathbf{\Gamma}$ is given by:
$\hat{\Gamma} = \tilde{\Gamma} + \tilde{\Sigma} B (B' \tilde{\Sigma} B)^{-1} B' \tilde{\Gamma} V (V' (\mathbf{Z}_1 \mathbf{Z}_1')^{-1} V)^{-1} V' (\mathbf{Z}_1 \mathbf{Z}_1')^{-1}$, with definitions of $\tilde{\Gamma}$ and $\tilde{\Sigma}$ as in Mosconi and Giannini (1992, pp. 407). This estimator is derived following Spanos (1986) pp.583. We note that the formula for $\hat{\Gamma}$ shown in Mosconi and Giannini (1992) contains several typographical errors; (e) Iterate until convergence by repeating steps (b)-(d), where the convergence criterion is defined in terms of increments in the likelihood function.
Finally, note that in systems with cointegration rank r =1, the model in equation (13) reduces to

$$\mathbf{Z}_o = \mathbf{\Gamma Z}_1 + A\alpha_{11}\beta_1' \mathbf{Z}_K + E$$

and the restrictions enter only through $\alpha$. The estimation of $\alpha_{11}$ and $\beta_1$ in this case follows Johansen and Juselius (1990) pp. 199 - 200 since no conditioning on $\beta$ is necessary, and the estimator of $\mathbf{\Gamma}$ is defined as before. In our experiment this is the case for the bivariate models.

maximal order of integration suspected to occur in the process. Dolado and Lütkepohl (*forthcoming)* using a different approach, prove the same result and analyze the power properties of this test. The Wald statistic is computed using only the first k coefficient matrices. This procedure does not require knowledge of either the cointegration properties of the system or the order of integration of the variables. Thus, if there is uncertainty as to whether the variables are I(1) or I(0), the adding of an extra lag insures that the test is being performed on the safe side[5].

## 4. THE EXPERIMENT AND THE RESULTS

The criteria used in designing the DGPs were: model dimension, degree of cointegration, direction of causality, and stability. Six data generation processes (DGPs) were included in the experiment; four bivariate (DGP(1) to DGP(4)) and two trivariate (DGP(5) and DGP(6)). Bivariate models were included because their simplicity facilitates the study of test performance. They also appear frequently in applied work related to purchasing power parity, threshold cointegration, market efficiency, and other studies of economic dynamics with pairs of variables. Higher dimensional models have been used to study, for instance, exchange rate behavior and to test structural hypotheses of economic dynamics. In Monte Carlo work, however, higher dimensionality creates problems of experiment management and interpretation. Because of this we simulate trivariate models only. All six models meet the conditions of Theorem 3 from Toda and Phillips (1993) relative to the "degree of cointegration." The direction of causation is controlled through either the long-run or the short-run parameters, or both. In all models there is causality from **x** to **y**, and at the same time relative "power of the test" comparisons are permitted. For instance, models DGP(1) and DGP(2) are the same except that in model DGP(2) the speed of adjustment coefficient on $y_t$ in the $x_t$ equation is changed from 0 to 0.4. Table 1 presents the six models (data generation processes) used in this study.

---

5

The computation of this Wald test is very simple: (a) Estimate a VAR(k+$d_{max}$) process by multivariate least squares, where k is the known or pre-determined optimum lag of the system. Denote $\Phi(L)^{k}_{k+dmax}$ , the least squares estimator of the parameters in equation (1) with only the coefficients of the first k lags considered; (b) Let $\hat{\Sigma}^{k}_{k+dmax}$ be a consistent estimator of the variance-covariance of $\Phi(L)^{k}_{k+dmax}$ . Then, $\lambda_W = T(R\hat{\Phi}^{k}_{k+dmax})^{\prime}\,(R\hat{\Sigma}^{k}_{k+dmax}R^{\prime})^{-1}\,(R\hat{\Phi}^{k}_{k+dmax})$ has an asymptotic $\chi^2_{(N)}$ distribution, with N being equal to the rows of the restriction matrix R.

In all cases 1000 samples of size T + 50 were generated with the first 50 observations discarded. For each DGP, five sample sizes were included; T=25, 50, 100, 200, and 400.

The tabulated results of the experiment are presented in Tables 2 and 3. Table 2 contains the outcome for the *bivariate models*, while Table 3 presents the results for the *trivariate models*. In all cases the numbers in the body of the tables are the percentage of rejections at the 5% level, and the headings WALD, LR, and MWALD stand for Wald test computed from the estimated ECM, likelihood ratio computed from the estimated ECM, and Wald test computed from the estimated "augmented" VAR, respectively. The lag length for the estimated models is "T" for the true lag, "O" for overfitting by one lag, and "U" for underfitting by one lag, respectively.

The results for both hypotheses, **y -/-> x** and **x** -/-> **y** for the *bivariate models* are tabulated in the Tables. The results correspond to the experiments conducted when the variance-covariance is contemporaneously correlated. The results for the identity covariance matrix do not differ substantially and due to space constraints are not presented here[6]. All three tests suffer from size distortions in small samples. As the sample size increases they approximate the correct size (left block of Table 2). An exception is the underfitting of model DGP(3) where both the WALD and the LR suffer a substantial size distortion even for large samples (the percentage of rejections is 63% for both the LR and WALD cases, respectively). The MWALD test rejects 4.8%of the cases. A possible reason for this size distortion is that model DGP(3) is an ECM(2); that is, the true model contains one short-run lag, and the error correction term. By underfitting DGP(3), an ECM(1) is estimated omitting all short-run dynamics. When computing the MWALD test under the underfitting scenario, a VAR(2) is estimated and the test computed on the reduced model, a VAR(1). Clearly, the estimation of a VAR(2) does not omit important dynamics. For model DGP(2), where the null hypothesis is not consistent with the true, the MWALD test requires a sample size of at least 100 to achieve a relatively high power. The WALD and LR rejections approach 100% for sample sizes of at least 100 . These results indicate a substantially lower "power of the test" for MWALD in small samples. The percentage of rejections goes to 100% for all three tests as the sample size increases indicating local asymptotic power.

---

6 These results appear in Zapata and Rambaldi (1995).

The right side of Table 2 presents the tabulated results for the hypothesis that **x -/-> y** (the null hypothesis is not consistent with the true) for the *bivariate models*. In large samples (equal to or larger than 100), the three tests generate equivalent results. In small samples, however, the percentage of rejections of the LR is significantly higher than that of WALD and MWALD. The percentage of rejections of WALD is slightly higher than that of MWALD in samples of size 50. These results point to a lower power of the test for MWALD in small samples. The percentage of rejections for model DGP(2), where bidirectional causality is present, is similarly low for MWALD and WALD in small samples. The importance of this finding is that in small samples, when bidirectional causality is expected, the LR seems to be the only test with enough power to detect a false null hypothesis in the *bivariate models* studied here.

Table 3 contains the tabulated results for the *trivariate models*. The left block of the table presents the results for the hypothesis that **y -/-> x** (the hypothesis is consistent with the truth). Overall, results are similar to those for *bivariate models*. All three tests show significant size distortions in small samples. In large samples the MWALD test approximates the correct size, the WALD suffers in model DGP(6) "U" from the same size distortion as model DGP(3) when short-run dynamics are omitted in the estimation of the ECM. The LR shows the largest deviation from the theoretical size of 5% for model DGP(6). The size distortion occurs in both the "O" and "U" cases . The right block of the table shows the results for the hypothesis that **x -/-> y** (the hypothesis is not consistent with the null). For samples of size 25, the power of the LR test is very high in *trivariate models*. For samples of size 50 or smaller the MWALD test appears to be sensitive to model structure as a false null hypothesis is not rejected too often (DGP(5)). The power performance of the WALD test falls between the other two tests. For samples of size at least 50, the power approaches 100.

## 5. SUMMARY AND CONCLUSIONS

This paper has presented Monte Carlo evidence on the performance of two Wald and a likelihood ratio tests for non-causality in cointegrated bivariate and trivariate models. The models were estimated using Johansen's maximum likelihood approach and least squares estimation of an augmented VAR. The experiment was designed to a) address questions of test performance as related to sample size, lag structure, unidirectional and bidirectional causality; and b) to compare the small sample performance of a more efficient LR test to two versions of the Wald test, namely a Wald statistic computed from an estimated "augmented" VAR (MWALD) and a Wald statistic computed from an estimated ECM. The main conclusions that emerge from the experiment can be summarized as follows.

Long-run non-causality (that is, non-causality through the error-correction term (ECT) as in models 1, 2, and 5) is consistently detected by the three tests when the model is correctly specified. The MWALD test for non-causality approaches the nominal size as the sample increases. For small samples the empirical size is larger than the nomical size. The results at samples of size 100 and larger appear quite accurate. Overfitting or underfitting does not seem to affect the empirical size of the test in detecting non-causality. The MWALD test is based on an estimator that does not incorporate the information about the degree of integration and/or cointegration of the variables in the system. An advantage of the MWALD test is that it has a limiting chi-squared distribution even if there is no cointegration or the stability and rank conditions are not satisfied[7]. On the other hand, as the estimator (VAR) is less efficient than the maximum likelihood estimator for cointegrated systems, MWALD would be expected to have a lower power of the test than the WALD and the LR tests in all cases studied in this experiment. The results show this to be the case for small samples, 25 and 50, only. Therefore, it is important to note, given the power performance of the tests in larger samples, that the MWALD approach has much practical appeal because of its simplicity.

The empirical size of the WALD test approaches the nominal size when the model is estimated with the true or the overfitted lag structure. The results at samples of size 50, nonetheless, appear quite accurate. Underfitting (i.e., specification of the short-run dynamics) affects the test size. Both LR and WALD tests are very sensitive to the specification of the short-run dynamics in ECMs even in large samples. The size distortions were very apparent when an ECM(2) model was underfitted, as in DGP(3) and DGP(6). The effect of underfitting on the size of the tests is believed to be the result of the bias of the Johansen's maximum

---

7 If the simulation study had considered cases where there is no cointegration or the stability and rank conditions are not satisfied, the conclusions could have had a different flavour, as correctly indicated by an anonymous referee.

likelihood estimator when models are underfitted. Gonzalo (1994) finds OLS to be superior to the Johansen's maximum likelihood estimator when the ECM is underparametrized .

The empirical size of the LR test approaches the nominal size when the model is estimated with the true lag structure. For the bivariate models in the experiment, underfitting did affect test size. In trivariate models both underfitting and overfitting seem to affect the empirical size of the test. Alternative covariance structures appear to have a significant effect on empirical size when the estimated model is not the true lag structure.

One practical implication of these results is that in choosing the lag structure of ECMs, alternative selection criteria must be examined; it appears that in testing for directional causality in these systems, parsimony may not be the guiding principle as all three tests suffer from severe size distortions regardless of the sample size when important dynamics are omitted.

Overfitting affects the power of the three tests at small samples (25,50) and this effect appears to be stronger when there is bidirectional causality through the ECT (DGP(2)).

In summary, our experiments suggest that all three tests have a high power of the test in moderate to large samples regardless of model structure. In small samples (50 or less observations), the MWALD test suffers the most loss in power, with the LR performing best in terms of power. This is encouraging for practitioners who may often have limited data upon which to make inference about economic dynamics.

In closing, it must be pointed out that Phillips (1995, pp.1053) has advocated the use of the "Fully Modified VAR" (FMVAR) approach. It appears that, judging from the results for the MWALD test in these experiments, this new FMVAR approach has much to offer to applied researchers studying "causality" and other related dynamic questions. Using this estimator, Phillips shows that a Wald test for non-causality has a limiting distribution that is a linear combination of independent chi-squared variates (see Theorem 6.1 pp. 1054). Non-causality tests based on the FMVAR approach are expected to have higher power than those based on the Augmented VAR estimator (i.e MWALD) since the FMVAR approach does not involve the inefficiency of having to estimate coefficient matrices for surplus lags (see Phillips, 1995, pp. 1053).

References


Dolado, J. J. and H. Lütkepohl. Making Wald Tests Work for Cointegrated VAR Systems. *Econometric Reviews*, forthcoming.

Engle, R. and C.W.J.Granger (1987). Co-integration and Error Correction: Representation, Estimation, and Testing, *Econometrica*, Vol. 55 No. 2, 251-276

Granger, C.W.J. (1988). Some Recent Developments in the Concept of Causality, *Journal of Econometrics*, Vol. 39, 199-211.

Gonzalo, J. (1994). Five Alternative Methods For Estimating Long-Run Equilibrium Relationships, *Journal of Econometrics*, Vol. 60, 203-233.

Johansen, S. (1988). Statistical Analysis of Cointegration Vectors, *Journal of Economic Dynamics and Control*, Vol 12, 231-254.

Johansen, S. and K. Juselius (1990). Maximum Likelihood Estimation And Inference On Cointegration- With Applications To The Demand For Money, *Oxford Bulletin of Economics and Statistics*, Vol. 52, 169-210.

Johansen, S. and K. Juselius (1992). Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and the UIP for UK, *Journal of Econometrics*, Vol 53, 211-244.

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

Lütkepohl, H. and H. Reimers (1992a). Granger-Causality In Cointegrated Var Processes, *Economic Letters*, Vol. 40, 263-268.

Lütkepohl, H. and H. Reimers (1992b). Impulse Response Analysis of Cointegrated Systems, *Journal of Economic Dynamics and Control*, Vol. 16, 53-78.

Mosconi, R and C. Giannini (1992). Non-Causality in Cointegrated Systems: Representation, Estimation and Testing, *Oxford Bulletin of Economics and Statistics*, Vol. 54 No. 3, 399-417.

Phillips, Peter C.B. (1995). Fully Modified Least Squares and Vecotr Autoregression, *Econometrica*, Vol. 63 No. 5, 1023-1078.

Spanos, A. (1986). *Statistical Foundations of Econometric Modelling* (Cambridge University, Cambridge).

Toda, H. Y. and P. C. B. Phillips (1993). Vector Autoregressions and Causality, *Econometrica*, Vol. 61 No.6, 1367-1393.

Toda, H. Y. and T. Yamamoto (1995). Statistical Inference in Vector Autoregressions with Possibly Integrated Processes, *Journal of Econometrics,* Vol. 66, 225-250.

Zapata, H.O. and A.N. Rambaldi(1995). Monte Carlo Evidence on Cointegration and Causation, *DAE Research Report No. 700*. Department of Agricultural Economics and Agribusiness. Louisiana State University Agricultural Center, pp. 26.

**TABLE 1**. Data Generation Processes (DGP).

| Model | $\Gamma_1$ | | | $\Gamma_2$ | | $\alpha$ | | $\beta$ | |
|---|---|---|---|---|---|---|---|---|---|
| DGP(1) | | | | | | -0.25 | | 1.00 | |
| | | | | | | 0.00 | | -2.00 | |
| DGP(2) | | | | | | -0.25 | | 1.00 | |
| | | | | | | 0.40 | | -2.00 | |
| DGP(3) | 0.50 | 0.50 | | | | -0.25 | | 1.00 | |
| | 0.00 | 0.25 | | | | 0.00 | | -2.00 | |
| DGP(4) | 0.50 | 0.50 | | 0.00 | 0.25 | -0.25 | | 1.00 | |
| | 0.00 | 0.25 | | 0.00 | 0.00 | 0.00 | | -2.00 | |
| DGP(5) | | | | | | -0.68 | 0.10 | 1.00 | 0.00 |
| | | | | | | 0.00 | 0.31 | 0.50 | 0.75 |
| | | | | | | 0.00 | -0.38 | -0.50 | 1.00 |
| DGP(6) | -1.07 | -0.48 | 0.49 | | | -0.68 | 0.10 | 1.00 | 0.00 |
| | 0.00 | -0.46 | 0.02 | | | 0.00 | 0.31 | 0.50 | 0.75 |
| | 0.00 | 0.02 | -0.31 | | | 0.00 | -0.38 | -0.50 | 1.00 |

Note: Two cases of the covariance matrix $\Sigma$ were used: a) an identity matrix, and b) a symmetric matrix with 0.5 replacing the zeros in the identity matrix. In the bivariate models, $p_1$=1, $p_2$=1, $r_1$=1 and $r_2$=1, $\Sigma$ is a 2x2 matrix; and in the trivariate models, $p_1$=1, $p_2$=2, $r_1$=1 and $r_2$=1, and $\Sigma$ is a 3x3 matrix.

**TABLE 2**. Percentage Rejection of Non-Causality with Cointegration in ***contemporaneous Bivariate*** *Models*, Wald and LR Tests, 5 Percent Level.

| | | Sample Size | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 25 | 50 | 100 | 200 | 400 | 25 | 50 | 100 | 200 | 400 |
| | | Test: **y -/-> x** | | | | | Test: **x -/-> y** | | | | |
| **WALD** | | | | | | | | | | | |
| | T | 0.073 | 0.065 | 0.070 | 0.050 | 0.053 | 0.808 | 0.982 | 1.000 | 1.000 | 1.000 |
| DGP 1 | O | 0.119 | 0.079 | 0.065 | 0.055 | 0.063 | 0.716 | 0.951 | 1.000 | 1.000 | 1.000 |
| | T | 0.916 | 0.998 | 1.000 | 1.000 | 1.000 | 0.553 | 0.838 | 0.990 | 1.000 | 1.000 |
| DGP 2 | O | 0.607 | 0.851 | 0.991 | 1.000 | 1.000 | 0.513 | 0.773 | 0.976 | 1.000 | 1.000 |
| | T | 0.106 | 0.078 | 0.056 | 0.062 | 0.059 | 0.905 | 0.996 | 1.000 | 1.000 | 1.000 |
| DGP 3 | O | 0.165 | 0.105 | 0.066 | 0.060 | 0.048 | 0.818 | 0.987 | 1.000 | 1.000 | 1.000 |
| | U | 0.179 | 0.177 | 0.217 | 0.433 | 0.630 | 0.792 | 0.960 | 1.000 | 1.000 | 1.000 |
| | T | 0.159 | 0.076 | 0.059 | 0.055 | 0.047 | 0.924 | 0.999 | 1.000 | 1.000 | 1.000 |
| DGP 4 | O | 0.263 | 0.096 | 0.068 | 0.060 | 0.045 | 0.830 | 0.982 | 1.000 | 1.000 | 1.000 |
| | U | 0.097 | 0.065 | 0.053 | 0.057 | 0.048 | 0.955 | 0.998 | 1.000 | 1.000 | 1.000 |
| **LR** | | | | | | | | | | | |
| | T | 0.065 | 0.061 | 0.068 | 0.050 | 0.053 | 0.931 | 0.999 | 1.000 | 1.000 | 1.000 |
| DGP 1 | O | 0.096 | 0.069 | 0.063 | 0.053 | 0.062 | 0.968 | 1.000 | 1.000 | 1.000 | 1.000 |
| | T | 0.905 | 0.998 | 1.000 | 1.000 | 1.000 | 0.997 | 1.000 | 1.000 | 1.000 | 1.000 |
| DGP 2 | O | 0.506 | 0.823 | 0.990 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 |
| | T | 0.091 | 0.074 | 0.054 | 0.061 | 0.059 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| DGP 3 | O | 0.123 | 0.089 | 0.062 | 0.060 | 0.052 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | U | 0.163 | 0.158 | 0.211 | 0.433 | 0.629 | 0.794 | 0.972 | 1.000 | 1.000 | 1.000 |
| | T | 0.119 | 0.055 | 0.050 | 0.039 | 0.060 | 0.991 | 1.000 | 1.000 | 1.000 | 1.000 |
| DGP 4 | O | 0.187 | 0.085 | 0.052 | 0.058 | 0.072 | 0.998 | 1.000 | 1.000 | 1.000 | 1.000 |
| | U | 0.090 | 0.048 | 0.042 | 0.042 | 0.055 | 0.979 | 1.000 | 1.000 | 1.000 | 1.000 |
| **MWALD** | | | | | | | | | | | |
| | T | 0.096 | 0.069 | 0.067 | 0.078 | 0.056 | 0.508 | 0.799 | 0.982 | 0.999 | 1.000 |
| DGP 1 | O | 0.171 | 0.084 | 0.066 | 0.065 | 0.066 | 0.584 | 0.864 | 0.996 | 1.000 | 1.000 |
| | T | 0.390 | 0.608 | 0.914 | 0.995 | 1.000 | 0.541 | 0.833 | 0.988 | 1.000 | 1.000 |
| DGP 2 | O | 0.446 | 0.627 | 0.922 | 0.998 | 1.000 | 0.532 | 0.740 | 0.968 | 1.000 | 1.000 |
| | T | 0.229 | 0.108 | 0.070 | 0.052 | 0.051 | 0.645 | 0.937 | 0.998 | 1.000 | 1.000 |
| DGP 3 | O | 0.334 | 0.140 | 0.078 | 0.064 | 0.043 | 0.727 | 0.930 | 0.999 | 1.000 | 1.000 |
| | U | 0.146 | 0.087 | 0.062 | 0.054 | 0.048 | 0.477 | 0.835 | 0.988 | 1.000 | 1.000 |
| | T | 0.369 | 0.106 | 0.071 | 0.057 | 0.048 | 0.592 | 0.899 | 0.998 | 1.000 | 1.000 |
| DGP 4 | O | 0.546 | 0.151 | 0.080 | 0.067 | 0.051 | 0.723 | 0.907 | 0.998 | 1.000 | 1.000 |
| | U | 0.243 | 0.104 | 0.063 | 0.063 | 0.051 | 0.450 | 0.783 | 0.987 | 1.000 | 1.000 |

Note: T=true, O=overfit, U=underfit.

**TABLE 3**. Percentage Rejection of Non-Causality with Cointegration in ***contemporaneous Trivariate*** *Models*, Wald and LR Test, 5 Percent Level.

| | | Sample Size | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 25 | 50 | 100 | 200 | 400 | 25 | 50 | 100 | 200 | 400 |
| | | Test: **y -/-> x** | | | | | Test: **x -/-> y** | | | | |
| **WALD** | | | | | | | | | | | |
| DGP 5 | T | 0.073 | 0.037 | 0.023 | 0.020 | 0.011 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | O | 0.259 | 0.111 | 0.062 | 0.048 | 0.040 | 0.882 | 0.987 | 1.000 | 1.000 | 1.000 |
| DGP 6 | T | 0.155 | 0.054 | 0.029 | 0.018 | 0.010 | 0.941 | 0.998 | 1.000 | 1.000 | 1.000 |
| | O | 0.314 | 0.181 | 0.076 | 0.038 | 0.028 | 0.692 | 0.953 | 0.994 | 1.000 | 1.000 |
| | U | 0.279 | 0.434 | 0.664 | 0.930 | 0.998 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 |
| **LR** | | | | | | | | | | | |
| DGP 5 | T | 0.084 | 0.065 | 0.062 | 0.050 | 0.054 | 0.998 | 1.000 | 1.000 | 1.000 | 1.000 |
| | O | 0.276 | 0.143 | 0.105 | 0.129 | 0.186 | 0.991 | 1.000 | 1.000 | 1.000 | 1.000 |
| DGP 6 | T | 0.223 | 0.108 | 0.102 | 0.086 | 0.139 | 0.988 | 1.000 | 1.000 | 1.000 | 1.000 |
| | O | 0.614 | 0.378 | 0.473 | 0.742 | 0.979 | 0.983 | 1.000 | 1.000 | 1.000 | 1.000 |
| | U | 0.329 | 0.507 | 0.746 | 0.963 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| **MWALD** | | | | | | | | | | | |
| DGP 5 | T | 0.199 | 0.089 | 0.058 | 0.074 | 0.055 | 0.500 | 0.665 | 0.923 | 0.995 | 1.000 |
| | O | 0.432 | 0.144 | 0.082 | 0.069 | 0.067 | 0.747 | 0.853 | 0.981 | 1.000 | 1.000 |
| DGP 6 | T | 0.438 | 0.187 | 0.098 | 0.069 | 0.071 | 0.764 | 0.886 | 0.993 | 1.000 | 1.000 |
| | O | 0.740 | 0.260 | 0.119 | 0.083 | 0.059 | 0.926 | 0.972 | 1.000 | 1.000 | 1.000 |
| | U | 0.234 | 0.105 | 0.075 | 0.068 | 0.052 | 0.884 | 0.990 | 1.000 | 1.000 | 1.000 |

Note: T=true, O=overfit, U=underfit.