A Comparison of Alternative Causality and Predictive Accuracy Tests in the Presence of Integrated and Co-integrated Economic Variables

Norman R. Swanson, Ataman Ozyildirim and Maria Pisu[1]

January 2001

Abstract

A number of variants of seven procedures designed to check for the absence of causal ordering are summarized. Five are based on classical hypothesis testing principles, including: Wald F-tests designed for stationary and difference stationary data; sequential Wald tests that account for cointegration; surplus lag regression type tests; and nonparametric fully modified vector autoregressive type tests. The other two are based on model selection techniques, and include: complexity penalized likelihood criteria; and ex-ante model selection based on predictive ability. In addition, various other approaches to checking for the causal order of economic variables are briefly discussed. A small set of Monte Carlo experiments is carried out in order to assess empirical size, and it is found that although all tests perform well in the environments where the true lag dynamics and cointegrating ranks are "accurately" estimated, simple surplus lag type tests of the variety discussed by Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996)), and model selection type procedures, retain good size properties in a variety of environments where the precise form of the true data generating process is unknown and difficult to ascertain, although analysis of the power of these tests is left to future research.

[1] Norman R. Swanson, Department of Economics, Texas A&M University, College Station, TX 77843-4228 (nswanson@econ.tamu.edu); Ataman Ozyildirim, The Conference Board, New York, NY; Maria Pisu, Department of Economics, Pennsylvania State University, University Park, PA 16802. The authors wish to thank Lutz Killian, seminar participants at the 1997 Winter Meetings of the Econometric Society and seminar participants at the University of Pennsylvania for useful comments on an earlier version of this paper. Swanson thanks the Private Enterprise Research Center at Texas A&M University for financial support.

1. Introduction

In a 1983 paper, Geweke, Meese, and Dent compared eight alternative tests of the absence of causal ordering, in the context of stationary variables. All were versions of the tests proposed by Granger (1969) and Sims (1972). Their unambiguous finding was that Wald variants of the tests are preferred. This paper is meant to extend the results of Geweke, Meese, and Dent (1983) in two directions. First, cases where the variables are not I(0) (in the sense of Engle and Granger (1987)), but also integrated of higher order and/or cointegrated, are considered. Second, in addition to classical in-sample tests of Granger noncausality, various model selection procedures for checking causal ordering are outlined and examined.

Our results, based on a series of Monte Carlo experiments, indicate that surplus lag Wald type tests due to Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996) (hereafter the TYDL test) perform well, regardless of integratedness and/or cointegratedness properties. However, when cointegrated data are examined, the sequential Wald type tests due Toda and Phillips (1993, 1994) come in a close second, and in some cases are superior. In addition, we find that model selection procedures based on the use of Schwarz Information Criteria as well as based on the ex ante forecasting approach suggested in Diebold and Mariano (1995) and Swanson (1998) perform surprisingly well in a variety of different contexts.

It should perhaps be noted that out-of-sample predictive ability type tests such as the model selection based ex ante procedure that we here consider are receiving considerable attention in the current causality literature. One reason for this may be an aspect of Granger's original definition which hadn't previously received much attention in the literature. In particular, at issue is whether or not standard in-sample implementations of Granger's definition are wholly in the spirit originally intended by Granger, and whether out-of-sample implementations might also be useful. Arguments in favor of using out-of-sample implementations are given in Granger (1980, 1988), and are summarized nicely in Ashley, Granger, and Schmalensee (1980), where it is stated on page 1149 that: "*... a sound and natural approach to such tests [Granger causality tests] must rely primarily on the out-of-sample forecasting performance of models relating the original (non-prewhitened) series of interest.*" Recent papers which develop and discuss predictive ability type causality tests include Clark and McCracken (1999), McCracken (1999), Chao, Corradi and Swanson (2000), and Corradi and Swanson (2000b). Other important papers in the burgeoning literature on predictive ability tests include Diebold and Mariano (1995), Harvey, Leybourne and Newbold (1997), West (1996), West and McCracken (1999), White (2000), Clark (2000), and Corradi and Swanson (2000a).

In addition to straightforward evaluation of the finite sample properties of various procedures, we address a number of additional empirical issues, such as: (1) Which model selection criteria perform best for choosing lags in finite sample scenarios, when the objective is to construct classical hypothesis tests of noncausality? (2) What is the relative impact of over- versus under-specification of the true lag order and/or the cointegration rank? (3) DGPs used include simple VAR(1) models and more complex VAR(3) models (possibly with MA errors). This allows us to examine not only which tests perform well in the context of simple models, but also in the context of more complex models. We ask the question: Is it true that a procedure which performs "best" in the context of a simple VAR(1) model may be dominated by another test when more complex DGPs describe the data. Further, we are able to examine the rate of increase in test size distortion as model complexity as well as sample size increases.

The rest of the paper is organized as follows: In Section 2, a variety of different procedures for ascertaining causal ordering are summarized and discussed. In Section 3, our experimental setup is discussed. Monte Carlo results are summarized in Section 4, and Section 5 concludes and offers a number of recommendations. Throughout, $\rightarrow_d$ signifies convergence in distribution.

2. Summary of Alternative Tests and Model Selection Procedures

For simplicity, we start by assuming that an n-vector time series, y_t, is generated by a K^{th} order VAR model:

$$y_t = J(L)y_{t-1} + u_t \ , \ t = -K+1, ..., T, \tag{2.1}$$

where $J(L) = \sum_{i=1}^{K} J_i \ L^{i-1}$. For simplicity of exposition, we leave the discussion of additional deterministic components in (2.1) until later. Assume also that

(A.1) $u_t = (u_{1t}, ..., u_{nt})$ 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 $\delta > 0$.

As we consider both integrated (integration of order 0 or higher) and cointegrated (by cointegrated we always refer to linear combinations of I(1) variables which are I(0) as discussed in Engle and Granger (1987)) economic variables, a number of alternative assumptions on the J matrices are in order. For the cases below where error correction models estimated, we assume that:

(A.2) $|I_n - J(z)z| = 0 => |z|>1$ or $z = 1$.

(A.3) $J(1) - I_n = \Gamma A'$ where Γ and A are nxr matrices of full column rank r, $0 \leq r \leq n-1$.

(A.4) $\Gamma'_P(J^*(1) - I_n)A_P$ is nonsingular, where Γ_P and A_P are nx$(n-r)$ matrices of full column rank such that $\Gamma_P\Gamma = 0 = A_P A$.

Assumptions (A.2)-(A.4) are standard, are taken from Toda and Phillips (1994), and ensure that (2.1) can be written in error-correction form:

$$\Delta y_t = J^*(L)\Delta y_{t-1} + \Gamma A' y_{t-1} + u_t, \tag{2.2}$$

where $J^*(L) = \sum_{i=1}^{K-1} J_i^* L^{i-1}$. Throughout, we assume also that the covariance matrix of the stationary component in the system is greater than zero. Finally, in the following discussion assume that we are interested in testing the hypothesis

$$H_0 : f(B) = 0 \text{ against } H_A : f(B) \neq 0,$$

where $f(\cdot)$ is an m-vector valued function satisfying standard assumptions. To simplify, we focus on the case where $f(\cdot) = C\beta$, C a standard Rx$(n^2$x$K)$ linear restriction matrix of rank R, where R is the number of restrictions (see Lütkepohl (1991) for further details).

2.1 Wald Type Tests

2.1.1 All Variables are Stationary

Assuming that y_t is generated by a stationary, stable VAR(K) process (see Lütkepohl (1991) for one definition of stability, e.g. assume (A.1) and (A.2) where $|z| > 1$), a frequently used version of the Wald test has

$$\lambda_{ST} = C\hat{\beta}'(C((ZZ')^{-1} \quad \hat{\Sigma}_u C')^{-1} C\hat{\beta} \rightarrow_d \chi^2_R,$$

where $\hat{\Sigma}_u$ is a consistent estimator of the covariance matrix of u (e.g. $\hat{\Sigma}_u = T^{-1}\hat{u}\hat{u}'$ or $\hat{\Sigma}_u = (T - nK)^{-1}\hat{u}\hat{u}'$), and $\hat{\beta}$ is the least squares estimate of vec(B) from levels regressions of the form

$$Y = BZ + U,$$

where $Y \equiv (y_1, ..., y_T)$ is the KxT matrix of data, $B \equiv (J_1, ..., J_K)$, $Z_t \equiv (y_t, ..., y_{t-K+1})'$, $Z \equiv (Z_0, ..., Z_{T-1})$, $\beta \equiv$ vec (B), and $U = (u_1, .., u_T)$ is an nxT matrix of errors. Of course, it has been known for quite some time that the χ^2_R distribution is not generally appropriate under the null hypothesis, when variables are integrated and/or cointegrated, so that in our experiments we do not expect particularly good results from this statistic when data are generated under assumptions of difference stationarity and/or cointegration. Indeed, the focus of our Monte Carlo experiments is on cases where variables are difference stationary and/or cointegrated. Finite sample properties of causal order tests under stationary data are well known, and are thus not included here. However, full results for stationary variable cases are available in an earlier working paper version (Swanson, Ozyildirim and Pisu (1996)). In addition, the reader is referred to Geweke, Meese, and Dent (1983).

2.1.2 All Variables Are Difference Stationary

Assuming that (A.1)–(A.4) hold and that there are n unit roots in the system (in which case we set $A_P = I_n = \Gamma_P$, and $J(1) = I_n$) it is reasonable to use λ_{ST} to test the null hypothesis, but with differenced data, and $K-1$ lags. We will call this statistic λ_{DS}.

2.1.3 Variables Are Integrated of Any Order, and Possibly Cointegrated

Assuming that (A.1) holds, consider the case where (A.2)–(A.4) may hold. Alternatively, given (A.1), assume that each element of $\{y_t\}$ may be integrated of any order (e.g. I(0) or I(1)), and that any combination of variables in $\{y_t\}$ may be cointegrated. Further, assume that we do not know, are unable to determine, or do not care what the orders of integratedness and/or cointegratedness are. Interestingly, in this case, Toda and Yamamoto (1995), and in independent work Dolado and Lütkepohl (1996), show that standard Wald tests can be applied to the coefficient matrices up to the correct lag order, when the system in levels (equation (2.1)) is artificially augmented by including at least as many extra lags of each variable as the highest order of integration of any of the variables in the system. (This test is hereafter referred to as the TYDL test.) In particular, it turns out that asymptotic chi–squared tests of causality restrictions can be applied to the submatrix of coefficients up to the correct lag order. The reason for this property is that all nonstandard asymptotics are essentially confined to the coefficient matrices beyond the correct lag order, and standard asymptotics apply with respect to the coefficient matrices up to the correct lag order. (Choi (1993) uses similar arguments to show that standard asymptotics can be applied to unit root tests which are suitably augmented.) Overall, the method of including surplus lags provides a very tractable, and hence appealing means with which to deal with many of the nonstandard distribution problems which characterize so many economic variables. However, it should be pointed out that the method is inefficient in the sense that extra coefficient matrices are estimated, so that the methods' power may be affected in finite samples. Indeed, it remains to see what the local power properties of such tests are. As the test is a standard "levels regression" type Wald test, it is clearly very easy to apply, and has the form

$$\lambda_{LW} = C\hat{\beta}'(C((ZQZ')^{-1} \quad \hat{\Sigma}_{\epsilon})C')^{-1}C\hat{\beta} \rightarrow_d \chi^2_R,$$

where all regressions are the same standard least squares regressions estimated using levels data (i.e. as used to construct λ_{ST}), except that extra lags of the regressors are added as surplus explanatory variables. That is:

$$Y = \hat{B}^* Z + \hat{\Psi} W + \hat{\epsilon}, \tag{2.3}$$

where $\hat{\Sigma}_{\epsilon} = T^{-1}\hat{\epsilon}\hat{\epsilon}'$. Note that $\hat{B}^*$, which is estimated using (2.3), is not the same as the estimator of B used in λ_{ST} and λ_{DS}, as W extra variables are added to the standard levels model, where $W_t \equiv (y_{t-K}, ..., y_{t-K-d+1})'$, and $W \equiv (W_0, ..., W_{T-1})$. Ψ is the coefficient matrix associated with these extra lags, and d is the highest order of integration of any of the variables in the system. The matrix Q is defined in the standard way as: $Q = Q_\tau - Q_\tau W'(WQ_\tau W')^{-1}WQ_\tau$, where $Q_\tau = I_T - S'(SS')^{-1}S$, with S is a matrix which contains constant and/or deterministic trend variables. In our case, we have set S to zero. However, if S is non-zero, then the variables can be added in the obvious way as additional regressors in (2.3), without affecting the distributional result under the null hypothesis.

One feature of the TYDL method which deserves further mention is that *any* form of Granger causal relation is captured by the test, regardless of whether it is "long–run", in the sense that error–correction terms are constructed using the variable(s) whose causal effect is being examined, or "short–run", in the sense that lagged differences of the relevant variable(s) appear in the error–correction model, or both. Alternatively, the sequential Wald tests discussed below have the feature that both "short–run noncausality" and "long–run noncausality" are sequentially tested for. However, as we will see below, if we incorrectly specify a difference stationary model (when the true model is cointegrated) then the estimated coefficient matrices become so "contaminated" that standard Wald tests (i.e. λ_{DS}) suffer from severe size distortion, often rendering the results of tests of "short–run noncausality" meaningless. Thus, it is not sufficient to simply model differenced data if all that we are interested in is "short–run noncausality". The reader is referred to Granger and Lin (1995) and Toda and Phillips (1994) for further discussion of short and long run causality.

2.1.4 Variables Are Integrated, and Possibly Cointegrated I

Toda and Phillips (1993) provide a theory for Wald tests of Granger noncausality in levels vector autoregressions that allow for stochastic and deterministic trends as well as arbitrary degrees of cointegration. The theory extends earlier work by Sims, Stock, and Watson (1990) on trivariate VARs and suggests a tractable strategy for implementing a variety of Wald type tests. An important feature of this approach is that tests are based on estimated versions of (2.2), the error correction model. Thus, while potentially offering gains in efficiency when estimating the system (relative to the TYDL approach of estimating the system in levels) the sequential Wald test adds a layer of estimation to the problem, in the sense that cointegrating ranks must be first estimated. Nevertheless, as shown in Toda and Phillips (1994), the sequential Wald tests appear to have good size properties, for a number of alternative specifications. To be more precise about the null hypothesis of interest for sequential Wald tests, we here consider variations of the joint null

$$H_0^* : J_{1,13}^* = ... = J_{K-1,13}^* = 0 \text{ and } \Gamma_1 A_3' = 0,$$

where $J_{13}^*(L) = \sum_{i=1}^{K-1} L^{i-1}$ is the n_1xn_3 upper right submatrix of $J^*(L)$, Γ_1 is the first n_1 rows of the loading coefficient matrix Γ, and $y_t = (y_{1t}, y_{2t}, y_{3t})'$ has been partitioned into three vectors of length n_1, n_2, and n_3, respectively.

Before discussing the sequential Wald tests any further, it is worth reiterating an important point which is made (and elaborated on) in Toda and Phillips (1994) (and also to some extent in Sims, Stock, and Watson (1990)). In the case of cointegrated economic variables, standard levels regression based causality tests are valid asymptotically as χ^2 criteria when there is sufficient cointegration among the variables whose causal effects are being tested. However, the rank condition associated with this criterion suffers from simultaneous equations bias when levels VARs are estimated, suggesting that there is no valid statistical basis in VARs for testing whether the condition holds. Furthermore, when the rank condition fails, there is no valid statistical basis for mounting a standard Wald test of causality. One interpretation of this point is simply that if we are using the "correct" specification of (2.1) (e.g. we have specified the correct number of lags) and there is cointegration among the variables, then standard Wald tests on levels are not directly feasible, and sequential Wald tests which are based on error–correction models are. Of final note is that if there is no cointegration, the standard levels regression based Wald statistic (i.e.λ_{ST}) for causality has a nuisance parameter free nonstandard limit distribution, and so can conceivably be used. These points are explained in some detail in Toda and Phillips (1994). We now state the three sequential procedures, and the testable hypotheses associated with them. The actual test statistics are given in Section 2 of Toda and Phillips (1994). For simplicity, assume that Γ_1 and A_3 are r, or $\hat{r}$-dimensional row vectors, and that $K > 1$ (with obvious modifications otherwise).

Then the tests are based on the null hypotheses

$$
\begin{aligned}
H_1^* &: J_{1,13}^* = ... = J_{K-1,13}^* = 0 \\
H_1^* &: J_{1,13}^* = ... = J_{K-1,13}^* = 0 \\
H_2^* &: \Gamma_1 = 0 \\
H_3^* &: A_3 = 0 \\
H_4^* &: \Gamma_1 A_3' = 0
\end{aligned}
$$

and as above

$$H_0^* : J_{1,13}^* = ... = J_{K-1,13}^* = 0 \text{ and } \Gamma_1 A_3' = 0$$

The sequential testing procedures are:

(P1) Test H_2^*. If H_2^* is rejected, test H_0^* using a $\chi^2_{n_3 K}$ critical value. Otherwise, test H_1^*.

(P2) Test H_3^*. If H_3^* is rejected, test H_0^* using a $\chi^2_{n_1 K}$ critical value. Otherwise, test H_1^*.

(P3) Test H_1^*. If H_1^* is rejected, reject the null hypothesis of noncausality. Otherwise, test H_2^* and H_3^*. If both are rejected, test H_4^* if $\hat{r} > 1$, or reject the null if $\hat{r} = 1$. Otherwise, accept the null of noncausality. These tests are easy to apply once standard procedures have been used to estimate the rank of the cointegrating space, and the number of cointegrating vectors in the system, although it should be noted that exact control of the overall size of these causality tests is not feasible. In what follows, we will refer to the test statistics associated with (P1)-(P3) as λ_{P1}, λ_{P2}, and λ_{P3}, respectively.

In closing, note that while the sequential Wald test discussed above is essentially designed to deal with I(1) variables which may or may not be cointegrated (of order (1,1) using the notation of Engle and Granger (1987)), the TYDL method allows for any mixtures of I(0), I(1), I(2), etc. variables which may or may not be cointegrated. Finally, the FM-VAR test (see below) also requires that (A.1)-(A.4) hold (e.g. $|I_n - J(L)L| = 0$ has roots on or outside the unit circle), with the modification that assumption (A.3) can be weakened so that $0 \leq r \leq n$, allowing for the possibility that all of the variables are stationary, for example. Of course, one reason why the sequential Wald test has $r < n$ stems from the fact that it is assumed that the variables are pre-tested to determine their order of integration, and if all are found to be I(0) then standard chi-squared asymptotics apply, and sequential Wald tests are not called for.

2.1.5 Variables Are Integrated, and Possibly Cointegrated II

Another type of Wald causality test which is in keeping with the unrestricted levels VAR method of TYDL is that based on fully modified vector autoregression (Phillips (1995)). Phillips proposes making corrections to standard OLS–VAR regression formulae which account for serial correlation effects as well as for regressor "endogeneity" that results from the existence of cointegrating relationship(s). The endogeneity in the context of cointegrated models arises for the following reason. Assume that (2.1) can be partitioned into two subsystems (i) $y_{1t} = Dy_{2t-1} + u_{1t}$, and (ii) $y_{2t} = y_{2t-1} + u_{2t}$. The standard procedure is to treat y_{2t-1} as predetermined, and regard the model as a reduced form. However, even though $E(u_{1t}y_{2t-1}) = 0$, the sample covariance, $T^{-1}\sum_{t=1}^{T} u_{1t}y'_{2t-1}$, does not converge to zero, and instead converges to a function of Brownian motions which are generally correlated in the limit.

The FM–VAR method is largely a generalization of work by Phillips and Hansen (1990), and has the feature that advance knowledge of which variables are stationary, nonstationary, and/or cointegrated is not required to achieve consistency of the fully modified VAR estimator. Given a Kernel condition (Phillips, pp. 1031), a bandwidth expansion rate condition (Phillips, pp. 1032), and a VAR condition (Phillips pp. 1044) which is essentially the same as assumptions (A.1)–(A.4) above, the relevant Wald type test statistic is

$$\lambda_{FM} = C\hat{\beta}^{+\prime}(C((XX')^{-1} \quad \hat{\Sigma}_{\epsilon})C')^{-1}C\hat{\beta}^{+},$$

where X is a matrix of observations with the same informational content of Z above, except that it is comprised of $K-1$ differences of the variables, as well as the lagged level of the variables. $\hat{\Sigma}_{\epsilon}$ is the usual OLS estimate of the error covariance matrix from a regression of Y on X. The test used here is actually distributed as a weighted sum of χ^2s, whose weights are difficult to estimate.

However, as pointed out by Phillips (1995), an asymptotically conservative test can by constructed by using the χ^2_R distribution as the limiting distribution for λ_{FM}. We adopt that approach in the sequel. Although the reader is referred to Phillips (1995) for further details, note that the FM-VAR estimator, $\hat{\beta}^+ = vec(\hat{F}^+)$, is

$$\hat{F}^+ = [YX_D' | YY_{-1}' - \hat{\Omega}_{\epsilon y}\hat{\Omega}_{yy}^{-1}(\Delta Y_{-1}Y_{-1}' - T\hat{\Delta}_{\delta y \Delta y})](XX')^{-1}$$

where X_D is the same as X, except that the levels variables are removed, Y_{-1} is Y lagged one period, and the Ω and Δ estimates are long-run covariance matrix estimates, and one-sided long-run covariance matrix estimates calculated in the usual way as the weighted sum of estimated autocovariance terms using some appropriate kernel function as the weighting scheme (see Phillips (1995) for details). Deterministic components can also be added to the FM-VAR regression model by augmenting the matrix of explanatory data used in the construction of λ_{FM} in the usual way.

2.2 Model Selection Type Causality Procedures

The next two procedures that can be used to examine causal ordering are not classical, in the sense that critical values are not used. In practice, these types of procedures involve choosing a "best" model. If that "best" model contains the variable(s) whose causal effect is being examined, then we "choose" the model which is not consistent with a null hypothesis of noncausality. The "best" model may be a single equation, or a model consisting of many equations, depending on how many variables are in y_1, when we are testing whether y_3 is noncausal for y_1. One advantage of model selection, for example based on complexity penalized likelihood criteria, is that the approach does not require specification of a correct model for its valid application. Another desirable feature is that the probability of selecting the truly best model approaches one as the sample size increases, if the model selection approach is properly designed. This is contrary to the standard practice of fixing a test size, and rejecting the null hypothesis at that fixed size, regardless of sample size. An interesting result of our simulations, which is in part a consequence of these features (see below) is that the empirical rejection frequencies based on both Akaike and Schwarz Information Criteria become vanishingly small (in many cases) for sample sizes of only 300 observations, regardless as to whether the "true" model is among those in the "selection" set.

As mentioned above, Granger causality tests have a natural interpretation as tests of predictive ability. In this sense, it may seem natural to use the results of causality tests when constructing forecasting models. However, it is commonly argued that any good forecasting model should be "tested" using some sort of ex ante forecasting experiment before being implemented in any practical setting. Along these lines, a reasonable alternative[1] procedure for checking causal ordering involves comparing competing forecasting models, both with and without the variable whose causal effect is being examined. These "competing" models could then be subjected to an ex ante forecasting analysis and the "best" model chosen using some sort of out–of–sample loss function (e.g. mean square prediction error), or by constructing an out-of-sample predictive ability test. If the "best" model contains the variable of interest, noncausality is rejected. In general, there are many model selection criteria to choose from, including forecast error summary measures, directional forecast accuracy measures, and profit measures, for example. Choice of criterion is naturally dependent upon the objective function of the practitioner. Diebold and Lopez (1995) and Swanson and White (1997a,b) discuss these and related issues, and contain other related references.

2.3 Complexity Based Likelihood Criteria

Granger, King, and White (1995) suggest that although standard hypothesis testing has a role to play in terms of testing economic theories, it is more difficult to justify using standard hypothesis tests for choosing between two competing models. One reason for their concern is that one model must be selected as the null, and this model is often the more parsimonious model. However, it is often difficult to distinguish between the two models (because of multicollinearity, near–identification, etc.), so that the null hypothesis may be unfairly favored. For example, it is far from clear that pre–test significance levels of 5 % and 1% are optimal, as pointed out Fomby and Guilkey (1978) in the context of Durbin– Watson tests. The use of model selection criteria neatly avoids related sticky issues associated with how to test theories and how to arbitrarily choose significance levels. Further, as discussed above, model selection procedures have a number of optimal properties if used to "select" among competing models, and seem a natural alternative to standard hypothesis testing when considering issues of Granger causality and/or predictive ability.

In a recent paper, Sin and White (1995) consider the use of penalized likelihood criteria for selecting models of dependent processes. In the context of strictly nested, overlapping or nonnested, linear or nonlinear, and correctly specified or misspecified models they provide sufficient conditions

[1] In our experiments, we use the predictive ability test due to Diebold and Mariano(1995) to "compare" our alternative models.

on the penalty to ensure that the model selected attains the lower average Kullback–Leibler Information Criterion, with probability (approaching) one. Further, as special cases, their results describe the Akaike and Schwarz Information Criteria (AIC and SIC respectively).

Our experiments into the use model selection criteria focus on their ability to "correctly" identify and unravel the causal relationships among stationary, integrated and cointegrated time series variables. If it turns out that these model selection criteria perform well, then we have direct evidence that model selection criteria offer a valid alternative to standard hypothesis testing in the context of uncovering causal relationships. Further, as model selection criteria are very easy to calculate, regardless of the properties of the data, evidence of the usefulness of AIC and SIC criteria might be of some interest to empirical economists. The two model selection criteria which we examine below are the Akaike Information Criterion (Akaike (1973, 1974)),

$$AIC = Tlog|\hat{\Sigma}| + 2f,$$

and the Schwarz Information Criterion (Rissanen (1978), Schwarz (1978)) criterion,

$$SIC = Tlog|\hat{\Sigma}| + flog(T),$$

where we define f to be the total number of parameters in the system (if we are measuring the causal effect of some variable(s) on a *group* of more than one other variable), or f is the number of parameters in the single equation of interest (when the *group* being examined is a single variable). Similarly, $\hat{\Sigma}$ is some standard estimate of the error covariance matrix, which is scalar if only one equation in the system is being examined.

2.4 Ex–Ante Forecast Comparison

Although AIC and SIC criteria may prove to be valuable when testing the null hypothesis of non– causality, note that they are calculated "in–sample", as are the standard Wald type tests discussed in Section 2.1. However, if, as is often the case, the goal of our analysis is to construct an "optimal" forecasting model, then causality might be better tested for in an ex ante rather than ex post setting, as discussed above. After all, such procedures would allow us to directly tackle the issue of predictive ability. In addition to those papers cited above, Engle and Brown (1986), Fair and Shiller (1990), and Swanson (1998) discuss ex ante model selection. However, these papers, as well as Diebold and Mariano (1995), West (1996), Swanson (1998), White (2000), and Corradi and

Swanson (2000a) consider the application of predictive accuracy tests which are not shown to be valid in the context of nested models (i.e. when forming Granger causality type predictive ability tests). For tests which are valid in the context of nested models, see Clark and McCracken (1999), McCracken (1999), Chao, Corradi and Swanson (2000), and Corradi and Swanson (2000b). The approach which we adopt involves constructing sequences of real–time one–step ahead forecasts of the variable of interest using a variety of models (levels VARs, differences VARs, and VECs) as well as a variety of different lag structures, cointegration assumptions, etc. The resultant sequences of forecasts are then used to construct three measures of out–of–sample forecast performance. The measures are mean squared forecast error ($MSE = \sum_{t=1}^{T^*} \widehat{fe}_t^2 / T^*$), and the well known related statistics: mean absolute forecast error deviation (MAD) and mean absolute forecast percentage error (MAPE), where T^* is the out-of-sample period (which here varies with sample size), while $\hat{f}e_t$ is the appropriate forecast error associated with the particular econometric model in question. The "best" model is not selected based on direct examination of the MSE, MAD, and MAPE criteria, however. Rather, we construct predictive accuracy tests described in Diebold and Mariano (1995), and implemented in Swanson and White (1997a,b). These tests statistics are constructed by first forming the mean $(\overline{d})$of a given loss differential series, d_t (e.g. for MSE we construct $d_t = \widehat{fe}_{1,t}^2 - \widehat{fe}_{2,t}^2$, where the subscript 1 corresponds to the null model, and subscript 2 denotes the same model as subscript 1, but with additional regressors corresponding to the variable whose causal effect is being examined). Then, the loss differential test statistic is constructed by dividing $(\overline{d})$by an estimate of the standard error of d_t, say $\hat{s}_d$. We use the parametric covariance matrix estimation procedure of Den Haan and Levin (1995) to estimate s_d, in part because the method is very easy to implement, and in part because Den Haan and Levin show that their procedure compares favorably with a number of nonparametric covariance matrix estimation techniques. Further details of these tests are given in Swanson and White (1995,1996a,b)). McCracken (1999) gives critical values which are valid when Diebold-Mariano tests are used to compare nested alternatives.

In our simulations we are able to implement a truly ex ante forecasting procedure. However, as is well known, the use of real–time forecasting methods becomes suspect when actual economic data which are used have been subjected to two–sided moving average filtering, periodic rebasing, and periodic revision, for example. Thus, we stress that these and related problems may make valid implementation of these types of tests quite difficult in practice. In the next section we discuss the design of a number of Monte Carlo experiments which were used to examine the finite sample properties of the above tests.

3. Experimental Design

Our Monte Carlo experiments are all based on 5000 simulations, and start with various difference stationary and cointegrated versions of the following two VAR models:

$$
\begin{aligned}
y_t &= J_1 y_{t-1} + u_t, t = 0, ..., T && (1) \\
y_t &= J_1 y_{t-1} + J_2 y_{t-2} + J_3 y_{t-3} + u_t, t = -2, ..., T && (2)
\end{aligned}
$$

We use a VAR(1) and a VAR(3) model because it is well known that causality tests often result in quite different conclusions, depending on how many lags are used to estimate a model. Clearly, one reason why this problem arises is that models which are misspecified in the sense that surplus lags are included, may suffer decreased test power. However, another possibility is that increasing the number of lags estimated results in what we will call "least squares confusion" (i.e. the tendency for least squares to yield poor parameter estimates), even if the "true" data are generated by a process with many lags. In practice, this type of situation may arise because of multicollinearity, and would probably manifest itself in the form of an increased incidence of rejecting the null hypothesis of Granger noncausality, even when the null is true. If this is the case, then we would expect experimental results based solely on VAR(1) processes to be conservative relative to those based on higher order VARs. As empirical models often contain many lags, "least squares confusion" may be damaging in the sense that we may later construct tests under the assumption that their size is close to the nominal test size, when in actuality the size is much higher than the nominal size. Further, the tests which perform "best" might change as we increase the number of lags in the system. In fact, we will see later that these issues are relevant, and have a non–negligible impact on our findings.

Our data generating processes can be summarized as follows:

Cointegrated Models (Cointegrating Rank $= 1$; $I_3 - \sum J_i = \Gamma A'$)

$$
\text{(VAR1 - CI1)} \quad J_1 = \begin{bmatrix} 0.5 & 0.5 & 0.0 \\ 0.0 & 1.0 & 0.0 \\ -1.0 & 1.0 & 1.0 \end{bmatrix} \quad \Gamma = \begin{bmatrix} 0.5 \\ 0.0 \\ 1.0 \end{bmatrix} \quad A = \begin{bmatrix} 1.0 \\ -1.0 \\ 0.0 \end{bmatrix}
$$

$$\text{(VAR3 - CI1)} \quad J_1 = J_2 = \begin{bmatrix} 0.17 & 0.17 & 0.0 \\ 0.0 & 0.4 & 0.0 \\ -0.4 & 0.4 & 0.4 \end{bmatrix} \quad J_3 = \begin{bmatrix} 0.16 & 0.16 & 0.0 \\ 0.0 & 0.2 & 0.0 \\ -0.2 & 0.2 & 0.2 \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} 0.5 \\ 0.0 \\ 1.0 \end{bmatrix} \qquad A = \begin{bmatrix} 1.0 \\ -1.0 \\ 0.0 \end{bmatrix}$$

$$\text{(VAR1 - CI2)} \quad J_1 = \begin{bmatrix} 1.0 & -1.0 & -1.0 \\ 0.0 & 0.5 & -0.5 \\ 0.0 & -1.0 & 0.0 \end{bmatrix} \quad \Gamma = \begin{bmatrix} 1.0 \\ 0.5 \\ 1.0 \end{bmatrix} \quad A = \begin{bmatrix} 0.0 \\ 1.0 \\ 1.0 \end{bmatrix}$$

$$\text{(VAR3 - CI2)} \quad J_1 = J_2 = \begin{bmatrix} 0.4 & -0.4 & -0.4 \\ 0.0 & 0.17 & -0.17 \\ 0.0 & -0.4 & 0.0 \end{bmatrix} \quad J_3 = \begin{bmatrix} 0.2 & -0.2 & -0.2 \\ 0.0 & 0.16 & -0.16 \\ 0.0 & -0.2 & 0.0 \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} 1.0 \\ 0.5 \\ 1.0 \end{bmatrix} \qquad A = \begin{bmatrix} 0.0 \\ 1.0 \\ 1.0 \end{bmatrix}$$

Difference Stationary Models

$$\text{(VAR1 - DS1)} \quad J_1 = I_3$$

$$\text{(VAR3 - DS1)} \quad J_1 = \begin{bmatrix} 0.2 & 0.1 & 0.0 \\ 0.0 & 0.4 & 0.0 \\ 0.4 & 0.5 & 0.5 \end{bmatrix} \quad J_2 = \begin{bmatrix} 0.2 & 0.1 & 0.0 \\ 0.0 & 0.4 & 0.0 \\ 0.2 & 0.1 & 0.3 \end{bmatrix}$$

$$J_3 = \begin{bmatrix} 0.6 & -0.2 & 0.0 \\ 0.0 & 0.2 & 0.0 \\ -0.6 & -0.6 & 0.2 \end{bmatrix}$$

It should perhaps be noted that the above parameterizations are *ad hoc*, in the sense that they are not calibrated by examining actual economic data, for example. In addition, economic data at most frequencies are usually modelled by including more than 3 lags. In this sense, the above DGPs are too simplistic. For these reasons, the Monte Carlo results reported below are meant only to be a guide to the sorts of relationships among the various causality tests which might be observed in practice. In addition, the following structure is placed on the errors of the above DGPs, where $u_t = \eta_t - \theta\eta_{t-1}$, and η_t is a unit normal random vector:

$$(MA1)\ \theta = 0$$

$$(MA2)\ \theta = \begin{bmatrix} 0.5 & 0.0 & 0.0 \\ 0.0 & 0.5 & 0.0 \\ 1.0 & 0.0 & 0.5 \end{bmatrix}$$

These values of θ are the same as those used by Toda and Phillips (1994). As discussed in Toda and Phillips (1994), even though (MA2) appears to be inconsistent with (A.1), u_t is an invertible MA process so that (VAR1) and (VAR3) can be rewritten as infinite order VARs. As the higher lags are exponentially decaying in the parameters, the actual number of lags needed for a "reasonable" approximation is low. Further, it will be interesting to see how well our various statistics perform when faced with inherently misspecified models (in the (MA1) models, the true specification is always included within the set of models which are fit). Finally, the addition of (MA2) to our experimental setup ensures that some of our models (i.e. the (VAR3) models) should ideally contain more than 3 lags when estimated.

In most cases we are testing the null hypothesis that y_3 is Granger noncausal for y_1 (we use the notation $y_3 \nrightarrow y_1$ for cases such as this), and thus we focus somewhat on the empirical size of the seven tests. The exception is (CI2), in which case we test the null hypothesis that $y_1 \nrightarrow y_2$. In all experiments, whenever a test of the null that $y_i \nrightarrow y_j$ is carried out, we also perform a test of the null that $y_j \nrightarrow y_i$. This allows us to tabulate power at the same time that we tabulate empirical size. It should be noted that in the case of our model selection procedures, "empirical size" is not the appropriate measure. Instead, we report the rejection frequency of the correct model in favor of the "incorrect" model. As the tests which we examine all have good power properties (albeit rejection frequencies under alternative hypotheses are not "size adjusted"), these findings leave little to choose between the tests. Not surprisingly, then, many analyses related to ours focus on empirical size. We follow suit, and report only empirical size (and its analog for our model selection procedures). Complete finding are available upon request from the authors.

In each of the experiments we use 100, 200, and 300 observations. Assuming that the true number of lags in each case is p, the number of lags used in each estimation is set at $\{p-1, p, p+1, p_{AIC}, p_{SIC}\}$, where p_{AIC} and p_{SIC} use the number of lags selected by implementation of the AIC and SIC model selection criteria, where the model selection criteria are constructed using estimates of levels VARs. Also, all lags are reported in terms of a VAR in levels, so that results for difference VAR Wald type tests (λ_{DS}) for one lag are not given. This is because a level VAR with one lag translates to a difference VAR with no lags.

In all experiments we calculate many difference stationary, stationary, and VEC systems, regardless of the true DGP. This is because the seven procedures considered variously involve estimation of all three types of VARs. For example, a cointegration model is estimated, and sequential Wald tests are constructed, even when the true data generating process is stationary. One reason for an extensive analysis of this type is that absolute size distortions associated with each test type are not necessarily as important as the size distortion of one test relative to the size distortion of other tests, regardless as to whether either test is theoretically valid. As an example of why this issue is perhaps important, note that Gonzalo and Lee (1996) show that while it is well known that augmented Dickey–Fuller tests suffer from low power when processes have near unit roots, and suffer from size distortions in other cases, it is less widely acknowledged that standard t–statistics also have similar power and size problems, even when the null hypothesis is that $\beta = 0.5$ and the true process is a stationary AR(1) with slope coefficient β. Extending this argument, it should be of interest to determine whether size and power distortions associated with using standard χ^2 asymptotics when they are not valid are big enough to warrant the use of more complex (albeit asymptotically valid) testing procedures which themselves may suffer from size and power distortions of other types. In particular, even though asymptotically valid, the finite sample performance of some of the more complex test statistics considered here is not certain.

We also used F–test versions of all Wald type tests, as it is well known that these often have better size properties in finite samples. To be more specific, we construct $F = \lambda/R$, where R is the number of restrictions. These statistics are assumed to be distributed as F, with T–f degrees of freedom in the denominator, where f is (i) the number of parameters in the equation being examined, and f is also defined as (ii) the number of parameters estimated in the entire system. Interestingly, the properties of these versions of the Wald tests were qualitatively the same as when the χ^2 versions of the tests were used. For this reason, these results are not reported in the tables, although complete results are available in an earlier working paper version of this piece.

Whenever VEC models are estimated, we use the maximum likelihood procedures developed by Johansen (1988, 1991) and Johansen and Juselius (1990). In particular, assume that the true cointegrating rank is r. Then, the cointegrating ranks used are usually $\{r-1, r, r+1, r_{JO1}, r_{JO2}\}$. The last two are rank estimates based on the well known likelihood ratio trace statistic using a 5% size, and with and without a constant in the estimated system, respectively. Results based on the related max statistic are not reported for the sake of brevity. The exception to this rule are cases where the true cointegrating rank is zero. In these cases, we use $\{r+1, r+2, r_{JO1}, r_{JO2}\}$, where r_{JO1} and r_{JO2} may be zero, in which case VEC models are not estimated, and corresponding sequential Wald test results are not reported. With the exception of the test statistic based on fully modified VAR estimates, all other estimations done in the experiments are multivariate least squares.

Overall, we calculate all Wald type tests whenever possible. However, it should be noted that there are some cases where Wald type tests are not applicable. As mentioned above, difference stationary versions of the standard Wald test are not calculated when the lag is one (as a one lag VAR in levels is a zero lag VAR in differences). Also, there are some lag structures and cointegrating rank cases (e.g. $r = n$) for which the sequential Wald type tests are not applicable, and are not calculated. (The reader is referred to Toda and Phillips (1994) for more details.) In such cases, the reported simulation results are calculated using possibly fewer than 5000 simulation trials. Instances where this situation arises are reported on in the next section.

The AIC and SIC type tests are implemented as follows. For all lags, cointegrating ranks, and alternative models (i.e. levels VAR, differences VAR, and VEC) AIC and SIC statistics are calculated. In particular, the statistics are calculated for versions of the models both with and without the variable(s) whose causal effect is being examined. As a summary test, the AIC and SIC–best models are chosen. If the "best" model contains any of the relevant variable(s), then the null hypothesis of Granger noncausality is rejected. We also track which *type* of model is chosen as "best" in each simulation. These types of tests are individually reported on for each lag and each model (e.g. cointegrating model with cointegrating space rank of unity), as well as across all models. The latter reporting style yields what we will call "global" tests, while the former is more in keeping with a comparison of only "two" alternative models when testing the null hypothesis of noncausality.

Ex ante forecasting tests are carried out the same way as the AIC and SIC type tests, with two exceptions. First, the statistics calculated are MSE, MAD, and MAPE. Second the statistics are constructed from sequences of one–step ahead forecast errors, and these forecast errors are in turn used to construct predictive accuracy test statistics (as discussed above). The procedure is implemented in the following way. A starting sample of $0.7T$ observations is used to construct a one–step ahead forecast, using all different model types, lags, etc., and for T = 100, 200, or 300. From this, the first forecast error is saved. Then one more observation is added to the sample, and another set of models are estimated, forecasts are constructed, and forecast errors are saved. This procedure continues until the entire sample is exhausted, and the resultant ex ante sequence of $(0.3T) - 1$ forecast errors is used to construct the model selection criteria. It should be noted that one case arises where the actual number of forecast errors used to construct the model selection criteria is not (0.3T)-1. For the VEC models which are estimated using cointegrating rank, $r = r_{JO1}$ or $r = r_{JO2}$, there may be some simulations where the estimated cointegrating rank is 0 or n (in which case the model is not estimated). This situation arises most frequently when the actual DGPs are stationary or difference stationary VARs. In these cases, we will report the average number of observations used in the construction of the selection criteria. In addition, it should be noted that due to possible model misspecification (see e.g. Corradi and Swanson (2000b)), finite sample results may not be the same for 1-step ahead versus multi-step ahead approached to ex ante forecasting implementations of Granger causality analysis.

4. Results

A summary of our experimental findings is given in Tables 1-12. Complete experimental results are contained in an earlier working paper version of this chapter (i.e. see Swanson, Ozyildirim and Pisu (1996)). Our findings from the Monte Carlo experiments are fairly clear cut, and can be summarized as follows: (1) In all cases (i.e. regardless of lag and model specification) the Toda and Yamamoto type Wald tests have empirical sizes close to nominal (nominal size is set at 5% throughout). (2) Sequential Wald tests, and standard Wald tests based on the use of differenced data only perform the best when correctly specified models (i.e. correct lags, cointegrating rank, data transformation) are estimated. (3) SIC type approaches, and to a lesser degree those based on the AIC perform quite well, even when the DGP has moving average (MA) error components (i.e. when the true lag order of the model is infinite and is hence only approximated using AIC and SIC measures). In many cases and for samples of moderate and large size (200 and 300 observations), the SIC type causality test has rejection frequency approaching zero very rapidly, and power close to unity. (4) As expected, test performance (with the exception of all of our model selection type causality "tests") worsens substantially when model complexity is increased (i.e. going from one to three lags and from no MA to an MA component in the true DGP.) (5) Lag selection based on the AIC and SIC does very well. In particular, the lag structure chosen using the SIC almost always yields the "best" results when the DGP is a VAR(1) with no MA component. However, for other DGPs, the lag structure given by the AIC generally results in better test performance. (6) Ex ante forecasting type approaches tend to perform well (see Chao, Corradi and Swanson (2000), and McCracken (1999) for further evidence). In particular, when models are estimated in levels, the tabulated size of the ex ante tests often approaches zero very quickly as T increases, regardless of the true DGP. (7) Difference stationary models perform exceptionally poorly, in all cointegrated cases, suggesting that difference type Wald tests do not provide a short–cut to testing for "short–run" noncausality when the true DGP is a VEC. This is because least squares becomes very confused in the context of this type of misspecification, and parameter estimates are hence very inaccurate. A final result worth noting is that even though the empirical sizes of many tests (with the exception of SIC and ex ante type model selection based tests) are often too high, suggesting that the null of Granger causality may be rejected too often, the use of F versions of the Wald–type tests does not appear to decrease the empirical sizes to any great extent, for the sample sizes which we consider. For this reason, we report rejection frequencies based only on the χ^2 versions of our test statistics.

4.1 Cointegration Models

In all of the cointegrated models (please see attached tables) the sequential Wald type tests, the Toda and Yamamoto Wald type tests, and the model selection based SIC type causality tests perform quite well. This is most evident when the true DGPs are (VAR1) and (VAR3) without an MA component. In these cases, λ_{LW} type tests have empirical sizes ranging from 0.049 to 0.056 for (VAR1) – T=200,300; and approximately 0.055 for (VAR3) – T=300 (see Table 1). In these cases, the best size is usually (except for (VAR1)–(MA1)) associated with lag selection based on the AIC. In fact, it turns out that the AIC is also useful for selecting lags in the context of our model selection based tests (see below). Thus, the more parsimonious method for choosing lags (SIC) performs better primarily for our least complex models. One reason for this finding may be that when complex DGPs are specified (as is usually the case when macroeconomic data are analyzed), the specification of surplus lags helps standard maximum likelihood estimation techniques to become less confused. This point is particularly important, given the tendency of our simplest TYDL type tests to perform so well. Put another way, if model complexity and estimation confusion is an issue, then avoiding the estimation of cointegrating vectors, ranks, etc., may be a useful approach – even if the true DGP is cointegrated. This indeed appears to be the case for more complex models, as the TYDL approach appears to perform as well as all of the other standard hypothesis testing approaches, at least in the context of empirical size. Of course, the cost of estimating models in levels, and ignoring cointegrating restrictions, is a reduction in efficiency.

Tests based on difference model estimations perform very poorly all the time. However, the Wald tests based on λ_{ST} have sizes which are usually Tests based on models estimated using differenced data perform very poorly whenever the true DGP is a VEC. However, λ_{ST} tests have sizes which are usually between 10 and 15 % , with the better sizes coming at the expense of bigger samples, but not necessarily less model complexity. For example, in Table 1, for T=300, the λ_{ST} test has size=0.108 in the VAR(3)-(MA1)-(CI-1) case when estimation is based on 4 lags, while the analogous test for the VAR(1) (i.e. 2 lags used in estimation) has size=0.128. Note that the reason for the relatively good performance of these two models, each of which is estimated with an extra lag, is probably related to the argument used to show that the λ_{LW} type tests are asymptotically valid. That is, even when we use the standard Wald test, as long as an extra lag is included, the size of the test is not too far from the nominal size, even though the asymptotic χ^2 distribution is clearly not valid.

The results for the tests based on the λ_{FM} statistic are somewhat less encouraging than those of the Toda and Yamamoto and sequential Wald type tests. In all of our experiments, the size of these tests appears to be in the 15–30 % range, regardless of sample size, model specification, etc. Although these results are not too bad, relatively, it seems apparent that a number of the other tests perform better. As the λ_{FM} test statistics rely on some optimal choice of bandwidth parameter as well as kernel function, for example, our results should be viewed with caution, as more suitable FM–VAR procedures may be available. To try and avoid this problem somewhat, we used three different lag truncation parameters to define the number of autocovariance terms used to compute the spectrum at zero frequency. These have been used elsewhere, and are: (i) $\mathrm{l} = 0$, (ii) $l = integer(4 * (T/100)^{0.25})$, and (iii) $l = integer(12 * (T/100)^{0.25})$. Further, we used the Parzen kernel function for the calculation of long–run covariance matrix estimates, as well as Andrews' (1991) automatic bandwidth selection procedures. Finally, an AR(1) filter was used to flatten the spectrum of the error around the zero frequency. Our results did not depend on the specification of l, however, and so in Table 1 we report results only for the case where $\mathrm{l} = 0$.

Sequential Wald test results are given in Tables 2–4. Single lag cases are not reported for λ_{P3}, as the test is not applicable in this case. In these tables, lag=1 corresponds to zero lags when the data are differenced, so that the corresponding error–correction models used to calculate λ_{P1} and λ_{P2} have no short–run dynamics. In general this has no effect on our results, as the DGPs are designed so as to impact causality both through the short–run dynamics (if there is any) as well as through the long–run dynamics (the error–correction part). In most cases, the best size tests are associated with one lag and with cointegrating rank either (i) selected to be the truth (one) or (ii) selected using a Johansen likelihood ratio trace test without a constant. In fact, the very similar sizes for the two "good" cointegrating (CI) rank cases arise because the Johansen method is usually able to "select" the true CI rank, even in the presence of substantial model complexity (i.e. for VAR(3) models with an MA component). However, of note is that when the (CI–1)–VAR(3)–(MA2) results are examined (see Table 2), the $\lambda_{P1} - \lambda_{P3}$ sizes range from "bests" of 0.194–0.234 (T=100); 0.118–0.161 (T=200); and 0.099–0.149 (T=300), when lags are selected using the AIC, and are always worse when lags are selected using the SIC. Interestingly, for the (CI-2) case (see Table 3), SIC lag selection leads to empirical sizes closer to nominal in the VAR(3)-MA(2) case, suggesting that the evidence in favor of using the AIC for more complex models is not always clearcut. Indeed, for corresponding λ_{LW} tests, size based on the SIC lag selection criterion is usually better than that based on the AIC criterion, and is: 0.130 (T=100), 0.109 (T=200), and 0.090 (T=300), for (CI–1). In fact, the only DGPs for which AIC is preferred over SIC for lag selection are VAR(3)-(MA2) specifications. This again underscored the usefulness of the AIC only in more complex models.

The ex ante forecasting based model selection type approaches perform quite well (see Tables 8,9, 13, and 14). In particular, for complex DGPs (e.g. (VAR3)–(MA2)) the reported sizes of the MSE, MAD, and MAPE tests are very small (from 0.002–0.015) for lags selected using either the AIC or the SIC. As is generally the case, though, lag selection based on the SIC is dominated by selection based on the AIC for more complex models. Even given the clear success of the ex-ante approaches, it should be noted that our experimental design is very simple. In addition, there is much research currently underway in the field of econometrics into the construction and performance of out of sample predictive ability tests, so that much remains to be learned.

The in–sample model selection approaches fare very well, also (see Tables 5-7 and 10-12). In particular, the SIC seems to provide a very reasonable alternative to standard hypothesis testing, when the null is Granger noncausality. In Tables 5 and 10 note that the number of "wins" (based on a comparison of model selection criteria for the different models) for a given "true" DGP and for each of the estimated cointegrating, difference stationary, and levels stationary models are reported. In this way, we can get an idea as to which type of model is preferred by AIC and SIC. As mentioned above, the way these numbers are reported is as follows. From amongst all lag and CI rank combinations, the AIC and SIC– best models are used in each simulation to "test" for noncausality. If the AIC–best model is a cointegration model, then a "win" is tallied for "cointegration", if the AIC–best model is a difference model, then a "win" is tallied for "differences", and so on. The rest of the tables on in-sample model selection report rejection frequencies. Notice in these tables that the AIC and SIC tend to select the "true" DGP (e.g. VEC models are found to be preferred whenever data are generated according to VECs). Even casual inspection of the results suggests that the SIC approach for checking Granger causality performs better than the AIC. In addition, both model selection criteria are somewhat sensitive to the combination of DGP and estimation method. For example, results are very poor when data are generated according to a VEC, but levels or difference stationary models are estimated, and fixing the CI rank at 2 when it is actually 1 also results in poor performance. Thus, although this sort of model selection appears promising, there do appear to be frailties with this approach (as with all approaches when used with misspecified models).

4.2 Difference Stationary Models

In order to directly examine the performance of Wald tests using differenced data (λ_{DS}), we also generated a number of difference stationary DGPs, and applied all seven types of causality tests. The results are omitted here for the sake of brevity, but are reported in Swanson, Ozyildirim and Pisu (SOP: 1996), Tables 17–23. Of note is that the test based on λ_{DS} is very precise for VAR(1) (T=100, 200, 300), and for the (VAR3) (T=300), when the correct number of lags is used (e.g. the sizes are 0.054, 0.051, 0.052, and 0.055, respectively). However, when an MA component (MA2) is introduced, the standard difference stationary type Wald tests are dominated by the levels type Toda and Yamamoto tests for all lags, as well as all sample sizes for the VAR1 case. For example, for VAR(1)–(MA2) sizes for the λ_{DS} test based on lags selected using the AIC are 0.176, 0.111, and 0.095, for T=100, 200, and 300, respectively. These are the "best" sizes available for this case. The corresponding values associated with the λ_{LW} type test are 0.122, 0.099, and 0.080. In addition, it is worth noting that sequential Wald tests yield rather uninspired results when faced with difference stationary data (see Tables 18-21 in SOP). Indeed, in most cases (except for VAR(3)–MA DGPs) the sizes of λ_{P1}–λ_{P3} are in the 10 – 20 % range.

Findings based on the use of in-sample model selection procedures are qualitatively the same as when VEC DGPs are generated. However, the SIC type causality test seems to perform even better, when lags are chosen using the AIC. This may be because AIC infrequently picks too few lags for inclusion in the estimated VAR model.

5. Conclusions

In this paper we have discussed and compared five different Wald type tests of the null hypothesis of noncausality, as well as two different tests based on a model selection approach to assessing the predictive ability of the variable(s) whose causal effect is of interest. Our main purpose was to gather together and discuss a reasonable number of recent tests, some of which are asymptotically valid under the null hypothesis regardless as to whether or not the data are integrated of order zero, integrated of higher order, or cointegrated of some order. We also examine a number of model selection based approaches to assessing causal ordering, including those based on in–sample complexity penalized likelihood criteria, and those based on the examination of ex ante forecasting ability. Finally, we obtain results which have some implications for empirical work. For example, our examination suggests that three types of tests perform reasonably well (in terms of size), regardless as to the integratedness and cointegratedness properties of the variables in the system, and regardless as to whether the model is correctly specified. The first is the standard Wald type test which is based on levels regressions, which adds "surplus" lags to account for integrated– and cointegratedness. This test is due to Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996). It should be noted, however, that the local power properties of such tests may indeed be suspect, although concrete results on this matter have yet to be published. The second consists of the model selection based approach of using the Schwarz Information Criterion (SIC) to select a "best" model, and then either "reject" or "accept" the noncausality null based on whether the causal variable of interest appears in the "best" model. In is perhaps worth stressing that model selection type approaches focus more on predictive ability, and as such may be particularly useful when the issue at hand is forecastability. In-sample hypothesis testing, on the other hand, may be preferable when the statistical significance of particular coefficients in a model is the issue being examined (e.g. when testing economic theories).

References

Ashley, R., C.W.J. Granger, and R. Schmalensee, 1980. Advertising and aggregate consumption: an analysis of causality. Econometrica 48, 1149-1167.

Akaike, Hirotuku, 1973. Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki, eds., 2nd International Symposium on Information Theory. Budapest: Akademiai Kiado, 267-281.

Akaike, Hirotuku, 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716-23.

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817-858.

Chao, J., V. Corradi and N.R. Swanson, 2000. An out of sample test for Granger causality. Macroeconomic Dynamics, forthcoming.

Choi, I., 1993. Asymptotic normality of least squares estimates for higher order autoregressive integrated processes with some applications. Econometric Theory.

Clark, T., 2000. Finite-sample properties of tests for equal forecast accuracy. Journal of Forecasting 18, 489-504.

Clark, T. and M.W. McCracken, M.W., 1999. Granger causality and tests of equal forecast accuracy and encompassing. Working Paper, Federal Reserve Bank of Kansas City.

Corradi, V. and N.R. Swanson, 2000a. Predictive ability with cointegrated variables. Working Paper, Texas A&M University.

Corradi, V. and N.R. Swanson, 2000b. A consistent test for nonlinear out of sample predictive accuracy. Working Paper, Texas A&M University.

Den Haan, W. and A. Levin, 1995. Inferences from parametric and non-parametric covariance matrix estimation procedures. NBER Technical Working Paper 195.

Diebold, F.X. and J.A. Lopez, 1995. Forecast evaluation and combination./ In G.S. Maddala and C.R. Rao, eds., Handbook of Statistics. Amsterdam: North Holland.

Diebold, F.X. and R.S. Mariano, 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253-263.

Dolado, J.J. and H. Lütkepohl, 1996. Making Wald tests work for cointegrated VAR systems, Econometric Reviews 15, 369-386.

Engle, Robert F. and Scott J. Brown, 1986. Model selection for forecasting. Applied Mathematics and Computation 20, 313-327.

Engle, R.F. and C.W.J. Granger, 1987. Co-integration and error correction: representation, estimation, and testing. Econometrica 55, 251-76.

Fair, Ray C. and Robert J. Shiller, 1990. Comparing information in forecasts from econometric models. American Economic Review 80, 375-389.

Fomby, T.B. and Guilkey, D.K., 1978. On choosing the optimal level of significance for the Durbin-Watson test and the Bayesian alternative. Journal of Econometrics 8, 203-213.

Geweke, J., Meese, R., and Dent, W., 1983. Comparing alternative tests of causality in temporal systems. Journal of Econometrics 21, 161-194.

Gonzalo, J. and Lee, T.-H., 1996. Relative Power of the t Type Tests for Stationary and Unit Root Processes. Journal of Time Series Analysis 17, 37-47.

Granger, C.W.J., 1969. Investigating causal relation by econometric models and cross spectral methods. Econometrica 37, 428-438.

Granger, C.W.J., 1980. Testing for causality: a personal viewpoint. Journal of Economic Dynamics and Control 2, 329-352.

Granger, C.W.J., 1988. Some recent developments in a concept of causality. Journal of Econometrics 39, 199-211.

Granger, C.W.J., King, M.L. and White, H., 1995. Comments on testing economic theories and the use of model selection criteria. Journal of Econometrics 67, 173-187.

Granger, C.W.J. and Lin, J.-L., 1995. Causality in the long run. Econometric Theory 11, 530-536.

Harvey, D.I., S.J. Leybourne and P. Newbold, 1997. Tests for forecast encompassing. Journal of Business and Economic Statistics 16, 254-259.

Johansen, S., 1988. Statistical analysis of cointegration vectors. Journal of Economics and Dynamics Control 12, 231-254.

Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica 59, 1551-1580.

Johansen, S. and Juselius, 1990. Maximum likelihood estimation and inference on cointegration - with applications to the demand for money. Oxford Bulletin of Economics and Statistics 52, 169-210.

Lütkepohl, H., 1991. Introduction to multiple time series analysis. Springer-Verlag: New York.

McCracken, M.W., 1999. Asymptotics for out of sample tests of causality. Working Paper, Louisiana State University.

Phillips, P.C.B., 1995. Fully modified least squares and vector autoregression. Econometrica 63, 1023-1078.

Phillips, P.C.B. and Hansen, B., 1990. Statistical inference in instrumental variables regressions with I(1) processes. Review of Economic Studies 53, 473-496.

Rissanen, J., 1978. Modeling by shortest data description. Automatica 14, 465-471.

Schwarz, G., 1978. Estimating the dimension of a model. The Annals of Statistics 6, 461-4.

Sims, C.A., 1972. Money, income, and causality. American Economic Review 62, 540-552.

Sims, C.A., Stock, J.H., and Watson, M.W., 1990. Inference in linear time series models with some unit roots. Econometrica 58, 113-144.

Sin, C.-Y. and White, H., 1995. Information criteria for selecting possibly misspecified parametric models. Journal of Econometrics 71, 207-225.

Swanson, N.R., 1998. Money and output viewed through a rolling window. Journal of Monetary Economics 41, 455-474.

Swanson, N.R., A. Ozyildirim, and M. Pisu, 1996. A comparison of alternative causality and predictive ability tests in the presence of integrated and co-integrated variables. Working Paper, Pennsylvania State University.

Swanson, N.R. and H. White, 1995. A model selection approach to assessing the information in the term structure using linear models and artificial neural networks. Journal of Business and Economic Statistics 13, 265-275.

Swanson, N.R. and H. White, 1997a. A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. Review of Economics and Statistics 79, 540-550.

Swanson, N.R. and H. White, 1997b. Forecasting economic time series using adaptive versus nonadaptive and linear versus nonlinear econometric models. International Journal of Forecasting 13, 439-461.

Toda, H.Y. and Phillips, P.C.B., 1993. Vector autoregressions and causality. Econometrica 61, 1367-1393.

Toda, H.Y. and Phillips, P.C.B., 1994. Vector autoregression and causality: a theoretical overview and simulation study. Econometric Reviews 13, 259-285.

Toda, H.Y. and Yamamoto, T., 1995. Statistical inference in vector autoregressions with possibly integrated processes. Journal of Econometrics 66, 225-250.

West, K.D., 1996. Asymptotic inference about predictive ability. Econometrica 64, 1067-1084.

West, K. and M.W. McCracken, 1998. Regression based tests of predictive ability. International Economic Review 39, 817-840.

White, H., 2000. A reality check for data snooping. Econometrica 68, 1097-1126.

Table 1: Summary of Wald Test Monte Carlo Results for $\lambda_{ST}, \lambda_{DS}, \lambda_{LW}$, and λ_{FM} Empirical Size[1], Data Generating Processes are (CI-1), (CI-2), and (DS1)

		(CI-1)				(CI-2)				(DS1)			
DGP/Sample	lag	λ_{ST}	λ_{DS}	λ_{LW}	λ_{FM}	λ_{ST}	λ_{DS}	λ_{LW}	λ_{FM}	λ_{ST}	λ_{DS}	λ_{LW}	λ_{FM}
VAR(1)-(MA1)	1	0.167	–	0.065	–	0.146	–	0.059	–	0.173	–	0.061	–
T = 100	2	0.142	0.999	0.074	0.279	0.127	0.651	0.068	0.800	0.152	0.054	0.070	0.329
	aic	0.175	0.990	0.064	0.408	0.154	0.571	0.059	0.660	0.186	0.234	0.062	0.335
	sic	0.167	–	0.065	–	0.146	–	0.059	–	0.173	1.00	0.061	–
	1	0.147	–	0.058	–	0.156	–	0.057	–	0.177	–	0.053	–
T = 200	2	0.133	1.00	0.064	0.247	0.129	0.911	0.064	0.938	0.146	0.051	0.061	0.299
	aic	0.155	1.00	0.057	0.336	0.160	0.788	0.057	0.788	0.185	0.211	0.053	0.338
	sic	0.147	–	0.058	–	0.156	–	0.057	–	0.177	–	0.053	–
	1	0.162	–	0.048	–	0.149	–	0.051	–	0.172	–	0.052	–
T = 300	2	0.128	1.00	0.058	0.256	0.119	0.979	0.053	0.975	0.137	0.052	0.058	0.295
	aic	0.169	1.00	0.049	0.366	0.155	0.912	0.052	0.837	0.180	0.259	0.054	0.384
	sic	0.162	–	0.048	–	0.149	–	0.051	–	0.172	–	0.052	–
VAR(1)-(MA2)	1	0.244	–	0.154	–	0.407	–	0.071	–	0.703	–	0.135	–
T = 100	2	0.217	1.00	0.108	0.784	0.215	0.964	0.074	0.948	0.344	0.298	0.108	0.821
	aic	0.215	0.996	0.114	0.353	0.207	0.941	0.116	0.507	0.282	0.189	0.135	0.483
	sic	0.238	1.00	0.104	0.741	0.213	0.967	0.075	0.779	0.582	0.328	0.096	0.718
	1	0.258	–	0.215	–	0.433	–	0.057	–	0.798	–	0.242	–
T = 200	2	0.274	1.00	0.106	0.861	0.217	1.00	0.063	0.996	0.504	0.505	0.134	0.876
	aic	0.169	1.00	0.079	0.260	0.159	0.994	0.082	0.719	0.188	0.117	0.104	0.485
	sic	0.251	1.00	0.093	0.724	0.170	1.00	0.069	0.406	0.413	0.363	0.105	0.710
	1	0.258	–	0.285	–	0.447	–	0.055	–	0.813	–	0.400	–
T = 300	2	0.324	1.00	0.119	0.908	0.226	1.00	0.059	1.00	0.656	0.681	0.165	0.918
	aic	0.140	1.00	0.066	0.266	0.140	1.00	0.073	0.687	0.158	0.099	0.084	0.587
	sic	0.217	1.00	0.078	0.560	0.151	1.00	0.061	0.384	0.321	0.278	0.090	0.656
VAR(3)-(MA1)	2	0.185	0.530	0.075	0.554	0.162	0.236	0.074	0.605	0.371	0.099	0.086	0.914
	3	0.139	0.874	0.085	0.229	0.132	0.276	0.088	0.165	0.145	0.066	0.103	0.408
T = 100	4	0.147	0.895	0.100	0.276	0.137	0.256	0.100	0.175	0.150	0.077	0.123	0.296
	aic	0.170	0.797	0.089	0.308	0.162	0.280	0.093	0.261	0.171	0.089	0.114	0.406
	sic	0.192	0.547	0.076	0.536	0.168	0.243	0.075	0.575	0.145	0.066	0.103	0.408
	2	0.169	0.821	0.061	0.565	0.150	0.399	0.060	0.814	0.382	0.106	0.070	0.924
	3	0.110	0.996	0.064	0.198	0.113	0.466	0.067	0.143	0.132	0.057	0.076	0.473
T = 200	4	0.107	0.997	0.070	0.285	0.112	0.426	0.077	0.152	0.132	0.065	0.086	0.271
	aic	0.117	0.993	0.066	0.209	0.118	0.466	0.070	0.155	0.142	0.066	0.078	0.465
	sic	0.162	0.866	0.060	0.480	0.148	0.422	0.060	0.638	0.132	0.057	0.076	0.473
	2	0.181	0.941	0.053	0.602	0.147	0.533	0.056	0.918	0.394	0.097	0.062	0.931
	3	0.107	1.00	0.062	0.213	0.110	0.653	0.063	0.141	0.128	0.054	0.067	0.555
T = 300	4	0.108	1.00	0.064	0.335	0.109	0.596	0.065	0.149	0.130	0.055	0.067	0.268
	aic	0.114	1.00	0.062	0.219	0.118	0.651	0.061	0.144	0.136	0.061	0.068	0.545
	sic	0.147	0.974	0.056	0.398	0.127	0.613	0.060	0.458	0.128	0.054	0.067	0.555
VAR(3)-(MA2)	2	0.448	0.932	0.197	0.978	0.464	0.149	0.114	0.626	0.865	0.493	0.227	0.963
	3	0.242	0.977	0.143	0.516	0.221	0.457	0.105	0.411	0.431	0.110	0.154	0.799
T = 100	4	0.187	0.985	0.129	0.564	0.177	0.812	0.114	0.480	0.226	0.152	0.144	0.815
	aic	0.254	0.964	0.161	0.585	0.239	0.835	0.161	0.600	0.326	0.178	0.205	0.846
	sic	0.295	0.971	0.142	0.613	0.221	0.570	0.109	0.441	0.437	0.123	0.151	0.810
	2	0.603	0.998	0.266	0.999	0.517	0.194	0.097	0.738	0.946	0.738	0.259	0.983
	3	0.293	1.00	0.146	0.598	0.229	0.733	0.082	0.477	0.553	0.131	0.194	0.881
T = 200	4	0.191	1.00	0.099	0.814	0.158	0.988	0.085	0.578	0.290	0.212	0.123	0.908
	aic	0.166	1.00	0.101	0.678	0.165	0.987	0.101	0.744	0.182	0.117	0.123	0.871
	sic	0.251	1.00	0.115	0.693	0.165	0.968	0.086	0.589	0.461	0.194	0.139	0.921
	2	0.718	1.00	0.352	1.00	0.539	0.244	0.095	0.792	0.969	0.858	0.268	0.989
	3	0.369	1.00	0.159	0.671	0.230	0.900	0.080	0.535	0.613	0.166	0.237	0.927
T = 300	4	0.200	1.00	0.090	0.953	0.155	1.00	0.073	0.672	0.372	0.302	0.139	0.946
	aic	0.136	1.00	0.075	0.714	0.149	0.998	0.092	0.835	0.169	0.110	0.117	0.887
	sic	0.223	1.00	0.094	0.899	0.152	1.00	0.074	0.733	0.349	0.220	0.130	0.974

[1] The tests examined are: λ_{ST} (Wald test based on levels OLS regression); λ_{DS} (Wald test based on differences OLS regression); λ_{LW} (Wald test based on levels OLS regression with surplus lags added to isolate nonstandard asymptotics); and λ_{FM} (Wald test based on fully modified vector autoregression). For each test, 5% χ^2 critical values are used. The lags used are listed, as is the number of observations in the sample, and the DGP used to generate the data. All results are based on 5000 Monte Carlo trials. Entries in the table correspond to the proportion of rejections of the null hypothesis of noncausality. For example, for the (CI 1)-VAR(1)-(MA1) with T = 100 observations, the upper left entry of the table is 0.167. Thus, a standard Wald test when applied to the data in levels has an empirical rejection rate of 16.7%.

Table 2: Summary of Sequential Wald Test Monte Carlo Results, Empirical Size[1], Data Generating Process is (CI-1)

		λ_{P1}				λ_{P2}				λ_{P3}			
Lag/ Sample	CI rank	VAR(1) (MA1)	VAR(1) (MA2)	VAR(3) (MA1)	VAR(3) (MA2)	VAR(1) (MA1)	VAR(1) (MA2)	VAR(3) (MA1)	VAR(3) (MA2)	VAR(1) (MA1)	VAR(1) (MA2)	VAR(3) (MA1)	VAR(3) (MA2)
lag = 1 T = 100	1	0.067	0.069			0.887	0.948			–	–		
	2	0.162	0.237			0.230	0.328			–	–		
	jo1	0.067	0.069			0.889	0.947			–	–		
	jo2	0.068	0.067			0.889	0.946			–	–		
lag = 2 T = 100	1	0.095	0.124	0.063	0.338	0.112	0.151	0.094	0.387	0.170	0.172	0.125	0.419
	2	0.150	0.208	0.181	0.448	0.155	0.227	0.191	0.468	0.229	0.285	0.272	0.495
	jo1	0.095	0.123	0.061	0.330	0.112	0.150	0.093	0.381	0.169	0.172	0.124	0.411
	jo2	0.094	0.123	0.060	0.328	0.111	0.150	0.091	0.380	0.168	0.172	0.122	0.410
lag = 3 T = 100	1			0.102	0.213			0.118	0.218			0.191	0.268
	2			0.140	0.242			0.147	0.251			0.235	0.324
	jo1			0.102	0.212			0.118	0.217			0.191	0.266
	jo2			0.102	0.212			0.117	0.216			0.190	0.265
lag = 4 T = 100	1			0.118	0.145			0.146	0.155			0.217	0.201
	2			0.152	0.190			0.157	0.198			0.253	0.285
	jo1			0.119	0.144			0.148	0.154			0.222	0.200
	jo2			0.117	0.143			0.145	0.153			0.217	0.199
lag = aic T = 100	1	0.082	0.144	0.114	0.201	0.628	0.174	0.137	0.241	0.376	0.187	0.202	0.268
	2	0.174	0.215	0.168	0.255	0.241	0.225	0.175	0.264	0.411	0.299	0.264	0.348
	jo1	0.082	0.141	0.113	0.194	0.631	0.170	0.134	0.234	0.370	0.184	0.200	0.262
	jo2	0.082	0.141	0.112	0.195	0.636	0.170	0.134	0.235	0.379	0.184	0.199	0.262
lag = sic T = 100	1	0.067	0.126	0.069	0.243	0.887	0.197	0.101	0.259	–	0.192	0.135	0.302
	2	0.162	0.231	0.187	0.293	0.230	0.263	0.196	0.304	–	0.299	0.273	0.367
	jo1	0.067	0.126	0.068	0.236	0.889	0.197	0.100	0.252	–	0.192	0.134	0.294
	jo2	0.068	0.125	0.066	0.234	0.889	0.196	0.098	0.250	–	0.191	0.132	0.291
lag = 1 T = 200	1	0.054	0.058			0.940	0.964			–	–		
	2	0.150	0.246			0.214	0.343			–	–		
	jo1	0.054	0.058			0.940	0.963			–	–		
	jo2	0.054	0.056			0.940	0.961			–	–		
lag = 2 T = 200	1	0.069	0.147	0.047	0.489	0.088	0.196	0.078	0.568	0.127	0.207	0.103	0.581
	2	0.131	0.267	0.169	0.595	0.142	0.288	0.176	0.619	0.202	0.334	0.250	0.641
	jo1	0.069	0.147	0.047	0.487	0.088	0.196	0.078	0.567	0.128	0.207	0.103	0.578
	jo2	0.069	0.146	0.047	0.486	0.087	0.196	0.077	0.566	0.127	0.207	0.103	0.577
lag = 3 T = 200	1			0.075	0.231			0.088	0.254			0.137	0.278
	2			0.114	0.287			0.122	0.303			0.203	0.353
	jo1			0.075	0.230			0.087	0.253			0.136	0.277
	jo2			0.075	0.228			0.087	0.252			0.136	0.276
lag = 4 T = 200	1			0.087	0.128			0.095	0.134			0.154	0.162
	2			0.110	0.182			0.115	0.190			0.207	0.279
	jo1			0.087	0.127			0.094	0.133			0.154	0.161
	jo2			0.087	0.127			0.094	0.133			0.154	0.161
lag = aic T = 200	1	0.062	0.103	0.083	0.120	0.649	0.118	0.096	0.137	0.341	0.144	0.144	0.173
	2	0.155	0.172	0.122	0.168	0.220	0.181	0.130	0.175	0.403	0.277	0.209	0.275
	jo1	0.062	0.102	0.082	0.119	0.650	0.117	0.096	0.136	0.342	0.143	0.144	0.173
	jo2	0.062	0.102	0.082	0.118	0.650	0.117	0.095	0.135	0.342	0.143	0.143	0.171
lag = sic T = 200	1	0.054	0.133	0.059	0.188	0.940	0.174	0.086	0.205	–	0.186	0.119	0.232
	2	0.150	0.243	0.162	0.246	0.214	0.260	0.168	0.258	–	0.320	0.241	0.325
	jo1	0.054	0.132	0.058	0.187	0.940	0.174	0.086	0.204	–	0.185	0.118	0.231
	jo2	0.054	0.132	0.058	0.186	0.940	0.174	0.085	0.203	–	0.185	0.118	0.230
lag = 1 T = 300	1	0.057	0.049			0.953	0.957			–	–		
	2	0.160	0.245			0.225	0.344			–	–		
	jo1	0.057	0.047			0.953	0.955			–	–		
	jo2	0.057	0.047			0.953	0.954			–	–		
lag = 2 T = 300	1	0.061	0.197	0.041	0.617	0.079	0.263	0.072	0.701	0.117	0.273	0.087	0.707
	2	0.123	0.311	0.178	0.716	0.130	0.342	0.185	0.741	0.201	0.389	0.260	0.763
	jo1	0.061	0.196	0.041	0.614	0.079	0.263	0.072	0.700	0.117	0.272	0.087	0.705
	jo2	0.061	0.196	0.041	0.617	0.079	0.263	0.071	0.703	0.116	0.272	0.087	0.709
lag = 3 T = 300	1			0.055	0.291			0.072	0.329			0.120	0.349
	2			0.107	0.361			0.112	0.381			0.203	0.423
	jo1			0.055	0.290			0.072	0.329			0.120	0.349
	jo2			0.055	0.290			0.072	0.329			0.120	0.348
lag = 4 T = 300	1			0.071	0.137			0.084	0.145			0.137	0.164
	2			0.111	0.194			0.116	0.203			0.212	0.285
	jo1			0.071	0.137			0.084	0.145			0.137	0.164
	jo2			0.071	0.137			0.084	0.145			0.137	0.164
lag = aic T = 300	1	0.064	0.082	0.062	0.099	0.648	0.093	0.080	0.112	0.291	0.119	0.127	0.149
	2	0.167	0.139	0.113	0.143	0.231	0.146	0.120	0.147	0.330	0.241	0.209	0.257
	jo1	0.064	0.082	0.062	0.099	0.646	0.093	0.080	0.112	0.288	0.119	0.127	0.149
	jo2	0.064	0.082	0.062	0.099	0.649	0.093	0.080	0.112	0.290	0.119	0.127	0.149
lag = sic T = 300	1	0.057	0.116	0.054	0.157	0.953	0.155	0.077	0.169	–	0.166	0.111	0.189
	2	0.160	0.209	0.146	0.217	0.225	0.226	0.152	0.227	–	0.292	0.234	0.306
	jo1	0.057	0.115	0.054	0.157	0.953	0.155	0.077	0.169	–	0.166	0.111	0.189
	jo2	0.057	0.115	0.053	0.156	0.953	0.155	0.076	0.169	–	0.166	0.110	0.189

[1] See notes to Table 1. The tests examined are: $\lambda_{P1} - \lambda_{P3}$ (sequential Wald tests based on VEC models estimated using Johansen's maximum likelihood procedure).

Table 3: Summary of Sequential Wald Test Monte Carlo Results, Empirical Size[1], Data Generating Process is (CI-2)

Lag/ Sample	CI rank	λ_{P1} VAR(1) (MA1)	VAR(1) (MA2)	VAR(3) (MA1)	VAR(3) (MA2)	λ_{P2} VAR(1) (MA1)	VAR(1) (MA2)	VAR(3) (MA1)	VAR(3) (MA2)	λ_{P3} VAR(1) (MA1)	VAR(1) (MA2)	VAR(3) (MA1)	VAR(3) (MA2)
lag = 1 T = 100	1	0.061	0.013	–	–	0.908	0.970	–	–	–	–	–	–
	2	0.147	0.409	–	–	0.213	0.530	–	–	–	–	–	–
	jo1	0.061	0.013	–	–	0.908	0.969	–	–	–	–	–	–
	jo2	0.061	0.012	–	–	0.910	0.968	–	–	–	–	–	–
lag = 2 T = 100	1	0.065	0.037	0.059	0.105	0.083	0.069	0.089	0.162	0.140	0.076	0.118	0.174
	2	0.137	0.211	0.163	0.470	0.141	0.221	0.174	0.480	0.206	0.295	0.253	0.536
	jo1	0.065	0.037	0.058	0.087	0.082	0.069	0.088	0.148	0.139	0.075	0.117	0.153
	jo2	0.065	0.037	0.058	0.087	0.082	0.068	0.088	0.148	0.139	0.075	0.117	0.152
lag = 3 T = 100	1	–	–	0.079	0.070	–	–	0.096	0.098	–	–	0.175	0.109
	2	–	–	0.137	0.224	–	–	0.146	0.232	–	–	0.227	0.330
	jo1	–	–	0.078	0.067	–	–	0.096	0.096	–	–	0.175	0.106
	jo2	–	–	0.078	0.067	–	–	0.096	0.095	–	–	0.174	0.104
lag = 4 T = 100	1	–	–	0.089	0.084	–	–	0.105	0.103	–	–	0.178	0.126
	2	–	–	0.144	0.180	–	–	0.151	0.187	–	–	0.236	0.291
	jo1	–	–	0.088	0.083	–	–	0.105	0.103	–	–	0.180	0.126
	jo2	–	–	0.089	0.082	–	–	0.105	0.101	–	–	0.179	0.124
lag = aic T = 100	1	0.071	0.115	0.093	0.163	0.603	0.151	0.119	0.203	0.316	0.171	0.186	0.222
	2	0.157	0.210	0.165	0.243	0.221	0.216	0.175	0.251	0.395	0.296	0.254	0.339
	jo1	0.070	0.112	0.090	0.156	0.604	0.148	0.115	0.195	0.312	0.168	0.183	0.214
	jo2	0.070	0.112	0.090	0.157	0.603	0.149	0.115	0.196	0.311	0.168	0.183	0.215
lag = sic T = 100	1	0.061	0.048	0.061	0.085	0.908	0.078	0.094	0.110	–	0.086	0.127	0.126
	2	0.147	0.208	0.172	0.221	0.213	0.215	0.183	0.228	–	0.286	0.253	0.328
	jo1	0.061	0.048	0.061	0.081	0.908	0.078	0.093	0.106	–	0.085	0.127	0.121
	jo2	0.061	0.047	0.061	0.081	0.910	0.077	0.093	0.106	–	0.084	0.126	0.119
lag = 1 T = 200	1	0.055	0.008	–	–	0.917	0.953	–	–	–	–	–	–
	2	0.157	0.428	–	–	0.225	0.562	–	–	–	–	–	–
	jo1	0.056	0.008	–	–	0.917	0.952	–	–	–	–	–	–
	jo2	0.056	0.008	–	–	0.917	0.950	–	–	–	–	–	–
lag = 2 T = 200	1	0.061	0.025	0.045	0.111	0.082	0.054	0.077	0.190	0.129	0.056	0.095	0.195
	2	0.129	0.209	0.157	0.525	0.137	0.218	0.236	0.539	0.207	0.299	1.00	0.590
	jo1	0.062	0.025	0.045	0.105	0.082	0.054	0.076	0.191	0.129	0.056	0.095	0.192
	jo2	0.062	0.025	0.044	0.104	0.082	0.054	0.076	0.191	0.130	0.056	0.094	0.192
lag = 3 T = 200	1	–	–	0.054	0.048	–	–	0.072	0.084	–	–	0.134	0.089
	2	–	–	0.112	0.228	–	–	0.119	0.236	–	–	0.206	0.336
	jo1	–	–	0.054	0.047	–	–	0.072	0.083	–	–	0.134	0.086
	jo2	–	–	0.054	0.047	–	–	0.071	0.083	–	–	0.133	0.086
lag = 4 T = 200	1	–	–	0.063	0.053	–	–	0.078	0.070	–	–	0.145	0.077
	2	–	–	0.116	0.155	–	–	0.118	0.161	–	–	0.207	0.271
	jo1	–	–	0.063	0.053	–	–	0.078	0.069	–	–	0.145	0.076
	jo2	–	–	0.064	0.052	–	–	0.078	0.069	–	–	0.145	0.076
lag = aic T = 200	1	0.062	0.084	0.061	0.098	0.610	0.117	0.081	0.126	0.306	0.135	0.142	0.150
	2	0.163	0.160	0.118	0.167	0.229	0.166	0.127	0.172	0.322	0.266	0.212	0.271
	jo1	0.063	0.084	0.061	0.098	0.612	0.117	0.081	0.125	0.309	0.135	0.141	0.150
	jo2	0.063	0.084	0.061	0.097	0.611	0.117	0.081	0.124	0.309	0.135	0.141	0.149
lag = sic T = 200	1	0.055	0.042	0.052	0.059	0.917	0.069	0.083	0.075	–	0.076	0.116	0.085
	2	0.157	0.172	0.151	0.167	0.225	0.179	0.163	0.174	–	0.271	0.239	0.279
	jo1	0.056	0.043	0.052	0.058	0.917	0.069	0.082	0.075	–	0.076	0.115	0.083
	jo2	0.056	0.042	0.051	0.057	0.917	0.069	0.082	0.073	–	0.076	0.115	0.083
lag = 1 T = 300	1	0.048	0.006	–	–	0.938	0.912	–	–	–	–	–	–
	2	0.147	0.447	–	–	0.209	0.581	–	–	–	–	–	–
	jo1	0.049	0.006	–	–	0.938	0.912	–	–	–	–	–	–
	jo2	0.049	0.006	–	–	0.941	0.912	–	–	–	–	–	–
lag = 2 T = 300	1	0.055	0.024	0.042	0.134	0.075	0.055	0.074	0.227	0.109	0.059	0.089	0.229
	2	0.122	0.220	0.156	0.549	0.130	0.228	0.165	0.567	0.195	0.313	0.238	0.608
	jo1	0.055	0.024	0.042	0.130	0.075	0.056	0.074	0.227	0.109	0.060	0.090	0.227
	jo2	0.055	0.023	0.042	0.131	0.075	0.055	0.074	0.227	0.109	0.059	0.089	0.227
lag = 3 T = 300	1	–	–	0.058	0.046	–	–	0.072	0.079	–	–	0.118	0.080
	2	–	–	0.113	0.230	–	–	0.119	0.238	–	–	0.204	0.343
	jo1	–	–	0.058	0.045	–	–	0.071	0.079	–	–	0.118	0.081
	jo2	–	–	0.058	0.046	–	–	0.071	0.080	–	–	0.118	0.080
lag = 4 T = 300	1	–	–	0.062	0.045	–	–	0.076	0.070	–	–	0.132	0.075
	2	–	–	0.113	0.151	–	–	0.116	0.158	–	–	0.214	0.277
	jo1	–	–	0.062	0.045	–	–	0.076	0.071	–	–	0.132	0.075
	jo2	–	–	0.062	0.045	–	–	0.076	0.071	–	–	0.132	0.075
lag = aic T = 300	1	0.054	0.074	0.065	0.090	0.602	0.098	0.079	0.114	0.286	0.116	0.126	0.136
	2	0.153	0.140	0.120	0.147	0.215	0.146	0.127	0.153	0.335	0.255	0.212	0.268
	jo1	0.054	0.074	0.065	0.090	0.601	0.097	0.079	0.114	0.283	0.116	0.126	0.135
	jo2	0.054	0.075	0.065	0.090	0.601	0.098	0.079	0.114	0.283	0.116	0.126	0.135
lag = sic T = 300	1	0.048	0.042	0.052	0.054	0.938	0.066	0.077	0.076	–	0.071	0.111	0.082
	2	0.147	0.148	0.133	0.147	0.209	0.156	0.141	0.154	–	0.261	0.219	0.272
	jo1	0.049	0.042	0.052	0.053	0.938	0.066	0.076	0.076	–	0.071	0.111	0.082
	jo2	0.049	0.042	0.052	0.054	0.941	0.066	0.076	0.076	–	0.070	0.110	0.082

[1] See notes to Table 2.

Table 4: Summary of Sequential Wald Test Monte Carlo Results, Empirical Size[1], Data Generating Process is (DS1)

		λ_{P1}				λ_{P2}				λ_{P3}			
Lag/ Sample	CI rank	VAR(1) (MA1)	VAR(1) (MA2)	VAR(3) (MA1)	VAR(3) (MA2)	VAR(1) (MA1)	VAR(1) (MA2)	VAR(3) (MA1)	VAR(3) (MA2)	VAR(1) (MA1)	VAR(1) (MA2)	VAR(3) (MA1)	VAR(3) (MA2)
lag = 1 T = 100	1	0.498	0.960	–	–	0.246	0.856	–	–	–	–	–	–
	2	0.452	0.799	–	–	0.195	0.724	–	–	–	–	–	–
	jo1	0.846	0.997	–	–	0.688	0.959	–	–	–	–	–	–
	jo2	0.759	0.995	–	–	0.537	0.947	–	–	–	–	–	–
lag = 2 T = 100	1	0.165	0.438	0.491	0.935	0.164	0.413	0.484	0.936	0.290	0.503	0.593	0.946
	2	0.157	0.364	0.382	0.870	0.157	0.346	0.382	0.869	0.199	0.390	0.441	0.895
	jo1	0.500	0.499	0.696	0.949	0.500	0.489	0.690	0.949	0.688	0.551	0.754	0.958
	jo2	0.407	0.495	0.668	0.950	0.407	0.482	0.661	0.950	0.574	0.546	0.727	0.959
lag = 3 T = 100	1	–	–	0.164	0.552	–	–	0.162	0.554	–	–	0.305	0.679
	2	–	–	0.155	0.432	–	–	0.154	0.436	–	–	0.214	0.521
	jo1	–	–	0.500	0.660	–	–	0.500	0.664	–	–	0.500	0.775
	jo2	–	–	0.425	0.633	–	–	0.375	0.637	–	–	0.525	0.754
lag = 4 T = 100	1	–	–	0.174	0.286	–	–	0.175	0.272	–	–	0.331	0.398
	2	–	–	0.166	0.237	–	–	0.166	0.233	–	–	0.237	0.292
	jo1	–	–	0.250	0.350	–	–	0.300	0.338	–	–	0.400	0.447
	jo2	–	–	0.269	0.339	–	–	0.288	0.324	–	–	0.500	0.435
lag = aic T = 100	1	0.478	0.329	0.187	0.382	0.259	0.322	0.186	0.383	0.487	0.430	0.325	0.513
	2	0.441	0.292	0.179	0.331	0.208	0.291	0.179	0.335	0.405	0.332	0.237	0.401
	jo1	0.750	0.524	0.579	0.634	0.667	0.515	0.553	0.626	0.600	0.546	0.605	0.730
	jo2	0.605	0.507	0.438	0.592	0.500	0.491	0.391	0.587	0.500	0.539	0.547	0.689
lag = sic T = 100	1	0.498	0.687	0.164	0.553	0.246	0.653	0.162	0.557	1.00	0.485	0.305	0.674
	2	0.453	0.611	0.155	0.438	0.195	0.588	0.154	0.444	1.00	0.388	0.214	0.522
	jo1	0.860	0.897	0.500	0.669	0.688	0.869	0.500	0.673	0.00	0.567	0.500	0.780
	jo2	0.759	0.869	0.425	0.646	0.537	0.837	0.375	0.650	0.00	0.546	0.525	0.762
lag = 1 T = 200	1	0.490	0.968			0.267	0.908	–	–	–	–	–	–
	2	0.451	0.875			0.203	0.816	–	–	–	–	–	–
	jo1	1.00	0.998			0.833	0.981	–	–	–	–	–	–
	jo2	0.818	0.996			0.500	0.974	–	–	–	–	–	–
lag = 2 T = 200	1	0.165	0.629	0.503	0.974	0.162	0.599	0.499	0.976	0.290	0.664	0.611	0.978
	2	0.154	0.547	0.398	0.949	0.157	0.515	0.393	0.949	0.194	0.570	0.460	0.958
	jo1	0.750	0.678	0.712	0.978	0.750	0.660	0.708	0.979	0.750	0.703	0.770	0.982
	jo2	0.500	0.670	0.684	0.981	0.533	0.648	0.679	0.981	0.700	0.696	0.751	0.984
lag = 3 T = 200	1	–	–	0.147	0.663	–	–	0.144	0.671	–	–	0.299	0.765
	2	–	–	0.136	0.552	–	–	0.140	0.560	–	–	0.205	0.624
	jo1	–	–	0.200	0.750	–	–	0.200	0.758	–	–	0.500	0.842
	jo2	–	–	0.387	0.733	–	–	0.355	0.741	–	–	0.548	0.827
lag = 4 T = 200	1	–	–	0.140	0.354	–	–	0.140	0.344	–	–	0.311	0.458
	2	–	–	0.135	0.299	–	–	0.137	0.298	–	–	0.212	0.356
	jo1	–	–	0.267	0.415	–	–	0.267	0.401	–	–	0.467	0.511
	jo2	–	–	0.293	0.406	–	–	0.293	0.391	–	–	0.488	0.499
lag = aic T = 200	1	0.465	0.213	0.158	0.218	0.269	0.213	0.154	0.213	0.383	0.364	0.308	0.387
	2	0.431	0.193	0.148	0.195	0.211	0.191	0.151	0.195	0.350	0.260	0.214	0.273
	jo1	0.833	0.373	0.182	0.447	0.714	0.373	0.182	0.433	0.000	0.452	0.545	0.538
	jo2	0.643	0.371	0.400	0.375	0.450	0.374	0.371	0.365	0.250	0.446	0.571	0.485
lag = sic T = 200	1	0.490	0.491	0.147	0.524	0.267	0.473	0.144	0.525	–	0.548	0.299	0.619
	2	0.451	0.432	0.136	0.456	0.203	0.419	0.140	0.463	–	0.449	0.205	0.515
	jo1	1.00	0.695	0.200	0.784	0.833	0.680	0.200	0.793	–	0.701	0.500	0.860
	jo2	0.818	0.678	0.387	0.751	0.500	0.658	0.355	0.758	–	0.688	0.548	0.825
lag = 1 T = 300	1	0.469	0.971	–	–	0.248	0.912	–	–	–	–	–	–
	2	0.458	0.886	–	–	0.202	0.834	–	–	–	–	–	–
	jo1	0.600	1.00	–	–	0.600	0.985	–	–	–	–	–	–
	jo2	0.696	0.997	–	–	0.533	0.979	–	–	–	–	–	–
lag = 2 T = 300	1	0.151	0.757	0.514	0.985	0.148	0.729	0.508	0.985	0.270	0.777	0.623	0.988
	2	0.145	0.696	0.408	0.970	0.146	0.664	0.407	0.970	0.185	0.716	0.464	0.977
	jo1	0.667	0.776	0.732	0.987	0.667	0.760	0.727	0.987	0.833	0.790	0.790	0.989
	jo2	0.364	0.776	0.699	0.987	0.333	0.756	0.695	0.987	0.576	0.790	0.765	0.989
lag = 3 T = 300	1	–	–	0.144	0.716	–	–	0.146	0.722	–	–	0.284	0.802
	2	–	–	0.135	0.613	–	–	0.138	0.620	–	–	0.190	0.675
	jo1	–	–	0.333	0.803	–	–	0.333	0.808	–	–	0.444	0.872
	jo2	–	–	0.296	0.783	–	–	0.296	0.789	–	–	0.370	0.858
lag = 4 T = 300	1	–	–	0.133	0.447	–	–	0.134	0.428	–	–	0.290	0.531
	2	–	–	0.125	0.381	–	–	0.127	0.377	–	–	0.198	0.429
	jo1	–	–	0.417	0.536	–	–	0.417	0.497	–	–	0.417	0.610
	jo2	–	–	0.355	0.519	–	–	0.387	0.485	–	–	0.452	0.592
lag = aic T = 300	1	0.453	0.179	0.152	0.202	0.253	0.177	0.153	0.198	0.482	0.338	0.289	0.373
	2	0.447	0.166	0.142	0.180	0.209	0.165	0.146	0.177	0.397	0.239	0.196	0.264
	jo1	0.571	0.371	0.333	0.328	0.667	0.371	0.333	0.310	0.667	0.427	0.444	0.474
	jo2	0.654	0.382	0.310	0.306	0.531	0.387	0.310	0.301	0.500	0.462	0.414	0.451
lag = sic T = 300	1	0.469	0.374	0.144	0.404	0.248	0.366	0.146	0.390	–	0.459	0.284	0.505
	2	0.458	0.336	0.135	0.357	0.202	0.329	0.138	0.356	–	0.373	0.190	0.408
	jo1	0.600	0.651	0.333	0.678	0.600	0.653	0.333	0.646	–	0.685	0.444	0.725
	jo2	0.696	0.593	0.296	0.641	0.533	0.587	0.296	0.613	–	0.637	0.370	0.691

[1] See notes to Table 2.

Table 5: Summary of SIC and AIC Criteria Approach to Causal Analysis[1], Data Generating Process is (CI-1)

DGP/ Sample	lag	AIC	SIC	Model Selected- aic			Model Selected - sic		
				CI	levels	difference	CI	levels	difference
VAR(1)-(MA1)	1	0.958	0.995	0.953	0.047	–	0.995	0.005	–
T = 100	2	0.358	0.113	0.889	0.105	0.006	0.929	0.052	0.019
	aic	0.945	0.960	0.946	0.048	0.006	0.984	0.008	0.008
	sic	0.958	0.995	0.953	0.047	–	0.995	0.005	–
	1	0.963	0.995	0.962	0.038	–	0.995	0.005	–
T = 200	2	0.380	0.072	0.907	0.093	0.000	0.966	0.034	0.000
	aic	0.945	0.955	0.957	0.042	0.001	0.991	0.008	0.001
	sic	0.963	0.995	0.962	0.038	–	0.995	0.005	–
	1	0.973	0.999	0.973	0.027	–	0.999	0.001	–
T = 300	2	0.343	0.057	0.901	0.099	0.000	0.977	0.023	0.000
	aic	0.950	0.957	0.967	0.033	0.000	0.998	0.002	0.000
	sic	0.973	0.999	0.973	0.027	–	0.999	0.001	–
VAR(1)-(MA2)	1	0.962	0.994	0.948	0.052	–	0.994	0.006	–
T = 100	2	0.450	0.143	0.919	0.080	0.001	0.972	0.023	0.005
	aic	0.365	0.051	0.829	0.134	0.037	0.920	0.063	0.017
	sic	0.583	0.359	0.922	0.077	0.001	0.974	0.021	0.005
	1	0.967	0.994	0.955	0.045	–	0.994	0.006	–
T = 200	2	0.505	0.136	0.936	0.064	0.000	0.986	0.014	0.000
	aic	0.261	0.015	0.866	0.117	0.017	0.955	0.039	0.006
	sic	0.459	0.106	0.934	0.066	0.000	0.981	0.017	0.002
	1	0.980	0.998	0.965	0.035	–	0.998	0.002	–
T = 300	2	0.586	0.162	0.966	0.034	0.000	0.994	0.006	0.000
	aic	0.239	0.003	0.897	0.098	0.005	0.966	0.033	0.001
	sic	0.425	0.056	0.943	0.057	0.000	0.985	0.015	0.000
VAR(3)-(MA1)	2	0.371	0.096	0.920	0.080	0.000	0.982	0.018	0.000
T = 100	3	0.275	0.047	0.835	0.163	0.002	0.907	0.089	0.004
	4	0.250	0.027	0.815	0.172	0.013	0.870	0.123	0.007
	aic	0.326	0.063	0.853	0.144	0.003	0.905	0.091	0.004
	sic	0.379	0.114	0.913	0.087	0.000	0.973	0.027	0.000
	2	0.383	0.058	0.933	0.067	0.000	0.986	0.014	0.000
T = 200	3	0.278	0.024	0.883	0.117	0.000	0.936	0.064	0.000
	4	0.232	0.013	0.860	0.139	0.001	0.918	0.082	0.000
	aic	0.290	0.029	0.886	0.114	0.000	0.934	0.065	0.001
	sic	0.381	0.054	0.929	0.071	0.00	0.970	0.030	0.000
	2	0.390	0.066	0.925	0.075	0.000	0.996	0.004	0.000
T = 300	3	0.260	0.018	0.873	0.127	0.000	0.950	0.050	0.000
	4	0.224	0.002	0.850	0.150	0.000	0.936	0.064	0.000
	aic	0.264	0.016	0.876	0.124	0.000	0.953	0.047	0.000
	sic	0.336	0.043	0.897	0.103	0.000	0.967	0.033	0.000
VAR(3)-(MA2)	2	0.669	0.364	0.956	0.044	0.000	0.975	0.025	0.000
T = 100	3	0.433	0.091	0.915	0.082	0.003	0.927	0.052	0.021
	4	0.304	0.032	0.854	0.119	0.027	0.925	0.056	0.019
	aic	0.323	0.022	0.814	0.150	0.036	0.910	0.080	0.010
	sic	0.504	0.147	0.925	0.068	0.007	0.943	0.043	0.014
	2	0.828	0.483	0.960	0.040	0.000	0.988	0.012	0.000
T = 200	3	0.485	0.062	0.936	0.064	0.000	0.960	0.039	0.001
	4	0.312	0.006	0.915	0.085	0.000	0.964	0.034	0.002
	aic	0.225	0.001	0.854	0.134	0.012	0.955	0.045	0.000
	sic	0.418	0.029	0.936	0.064	0.000	0.960	0.038	0.002
	2	0.895	0.585	0.976	0.024	0.000	0.992	0.008	0.000
T = 300	3	0.562	0.082	0.948	0.052	0.000	0.968	0.032	0.000
	4	0.312	0.001	0.922	0.078	0.000	0.971	0.029	0.000
	aic	0.185	0.00	0.860	0.131	0.009	0.952	0.048	0.000
	sic	0.350	0.011	0.928	0.072	0.000	0.975	0.025	0.000

[1] Results based on the model selection criteria type approaches discussed above are reported in this table. The lags used are listed, as is the number of observations in the sample, and the DGP used to generate the data. The proportion of "wins" accruing to each type of estimated model (VEC model (CI), stationary levels model (levels), and difference stationary model (difference) are listed in columns 5-10.

Table 6: SIC and AIC Approach Rejection Frequencies[1], Data Generating Process is (CI-1). All Estimations Based on Assumption that It Is Known that the Data Are Generated According to a VEC

Lag/	CI	VAR(1)-(MA1)		VAR(1)-(MA2)		VAR(3)-(MA1)		VAR(3)-(MA2)	
Sample	rank	AIC	SIC	AIC	SIC	AIC	SIC	AIC	SIC
lag = 1	1	0.076	0.076	0.023	0.023	–	–	–	–
T = 100	2	1.00	1.00	1.00	1.00	–	–	–	–
lag = 2	1	0.241	0.115	0.289	0.121	0.203	0.072	0.541	0.315
T = 100	2	1.00	1.00	1.00	1.00	1.00	0.999	1.00	1.00
lag = 3	1	–	–	–	–	0.233	0.096	0.337	0.104
T = 100	2	–	–	–	–	1.00	0.974	0.999	0.994
lag = 4	1	–	–	–	–	0.231	0.097	0.205	0.040
T = 100	2	–	–	–	–	0.999	0.724	0.998	0.914
lag = aic	1	0.104	0.091	0.260	0.059	0.252	0.108	0.240	0.041
T = 100	2	1.00	0.999	0.999	0.926	0.999	0.954	0.991	0.675
lag = sic	1	0.076	0.076	0.248	0.113	0.206	0.078	0.388	0.141
T = 100	2	1.00	1.00	1.00	1.00	1.00	0.997	0.998	0.991
lag = 1	1	0.061	0.061	0.023	0.023	–	–	–	–
T = 200	2	1.00	1.00	1.00	1.00	–	–	–	–
lag = 2	1	0.228	0.086	0.372	0.129	0.212	0.047	0.729	0.434
T = 200	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
lag = 3	1	–	–	–	–	0.212	0.070	0.403	0.082
T = 200	2	–	–	–	–	1.00	1.00	1.00	1.00
lag = 4	1	–	–	–	–	0.189	0.075	0.220	0.021
T = 200	2	–	–	–	–	1.00	0.996	1.00	1.00
lag = aic	1	0.084	0.067	0.177	0.026	0.223	0.074	0.154	0.024
T = 200	2	1.00	1.00	1.00	0.982	1.00	0.999	1.00	0.887
lag = sic	1	0.061	0.061	0.339	0.101	0.233	0.059	0.322	0.048
T = 200	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
lag = 1	1	0.058	0.058	0.023	0.023	–	–	–	–
T = 300	2	1.00	1.00	1.00	1.00	–	–	–	–
lag = 2	1	0.201	0.068	0.456	0.155	0.216	0.054	0.841	0.556
T = 300	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
lag = 3	1	–	–	–	–	0.182	0.057	0.488	0.100
T = 300	2	–	–	–	–	1.00	1.00	1.00	1.00
lag = 4	1	–	–	–	–	0.181	0.054	0.223	0.015
T = 300	2	–	–	–	–	1.00	1.00	1.00	1.00
lag = aic	1	0.079	0.064	0.157	0.020	0.188	0.052	0.135	0.023
T = 300	2	1.00	1.00	1.00	0.996	1.00	1.00	1.00	0.942
lag = sic	1	0.058	0.058	0.303	0.060	0.213	0.058	0.256	0.021
T = 300	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

[1] See notes to Table 5. The reported numerical values are the proportion of rejections of the "null hypothesis" of noncausality. All models are compared in pairwise fashion. For example, in this table each entry corresponds to the rejection frequency when two alternative cointegrating models are compared - one with the causal variable of interest included, and one without.

Table 7a: SIC and AIC Approach Rejection Frequencies [1], Data Generating Process is (CI-1). All Estimations Based on Assumption that It Is Known that the Data Are Generated According to a Levels VAR

Sample	Lag	VAR(1)-(MA1)		VAR(1)-(MA2)		VAR(3)-(MA1)			VAR(3)-(MA2)
		AIC	SIC	AIC	SIC	AIC	SIC	AIC	SIC
	1	0.329	0.113	0.410	0.205	–	–	–	–
T = 100	2	0.262	0.036	0.386	0.082	0.321	0.050	0.595	0.248
	3	–	–	–	–	0.212	0.010	0.370	0.037
	4	–	–	–	–	0.202	0.006	0.261	0.018
	aic	0.341	0.113	0.306	0.020	0.254	0.017	0.263	0.010
	sic	0.329	0.113	0.410	0.114	0.326	0.058	0.417	0.092
	1	0.359	0.081	0.420	0.156	–	–	–	–
T = 200	2	0.272	0.018	0.427	0.063	0.315	0.033	0.759	0.300
	3	–	–	–	–	0.205	0.012	0.432	0.034
	4	–	–	–	–	0.177	0.003	0.256	0.003
	aic	0.363	0.081	0.212	0.005	0.214	0.011	0.178	0.002
	sic	0.359	0.081	0.399	0.048	0.307	0.030	0.358	0.019
	1	0.349	0.089	0.458	0.168	–	–	–	–
T = 300	2	0.264	0.014	0.506	0.079	0.350	0.031	0.825	0.368
	3	–	–	–	–	0.193	0.003	0.496	0.040
	4	–	–	–	–	0.166	0.000	0.268	0.00
	aic	0.354	0.089	0.198	0.001	0.199	0.003	0.154	0.00
	sic	0.349	0.089	0.352	0.028	0.276	0.022	0.299	0.004

[1] See notes to Table 6.

Table 7b: SIC and AIC Approach Rejection Frequencies[1], Data Generating Process is (CI-1). All Estimations Based on Assumption that It Is Known that the Data Are Generated According to a VAR in Differences

Sample	Lag	VAR(1)-(MA1)		VAR(1)-(MA2)		VAR(3)-(MA1)			VAR(3)-(MA2)
		AIC	SIC	AIC	SIC	AIC	SIC	AIC	SIC
	1	–	–	–	–	–	–	–	–
T = 100	2	1.00	0.999	1.00	1.00	0.721	0.449	0.972	0.916
	3	–	–	–	–	0.949	0.668	0.990	0.905
	4	–	–	–	–	0.946	0.512	0.993	0.835
	aic	1.00	0.945	0.998	0.922	0.892	0.609	0.977	0.631
	sic	–	–	1.00	1.00	0.737	0.467	0.986	0.897
	1	–	–	–	–	–	–	–	–
T = 200	2	1.00	1.00	1.00	1.00	0.914	0.727	0.998	0.995
	3	–	–	–	–	1.00	0.954	1.00	0.998
	4	–	–	–	–	1.00	0.914	1.00	0.995
	aic	1.00	0.980	1.00	0.988	0.997	0.942	0.998	0.833
	sic	0.00	0.00	1.00	1.00	0.932	0.776	1.00	0.995
	1	–	–	–	–	–	–	–	–
T = 300	2	1.00	1.00	1.00	1.00	0.971	0.856	1.00	1.00
	3	–	–	–	–	1.00	0.998	1.00	1.00
	4	–	–	–	–	1.00	0.993	1.00	1.00
	aic	1.00	1.00	1.00	0.997	1.00	0.996	1.00	0.931
	sic	0.00	0.00	1.00	1.00	0.987	0.938	1.00	1.00

[1] See notes to Table 7a.

Table 8: MSE, MAD, and MAPE Type Ex Ante Predictive Ability[1], Data Generating Process is (CI-1). All Estimations Based on Assumption that It Is Known that the Data Are Generated According to a VEC

Lag/ Sample	CI rank	VAR(1)-(MA1)			VAR(1)-(MA2)			VAR(3)-(MA1)			VAR(3)-(MA2)		
		MSE	MAD	MAPE	MSE	MAD	MAPE	MSE	MAD	MAPE	MSE	MAD	MAPE
lag = 1	1	0.166	0.158	0.114	0.288	0.243	0.185	–	–	–	–	–	–
T = 100	2	0.147	0.144	0.105	0.260	0.228	0.178	–	–	–	–	–	–
lag = 2	1	0.013	0.016	0.014	0.014	0.019	0.016	0.011	0.017	0.015	0.024	0.025	0.016
T = 100	2	0.063	0.082	0.071	0.145	0.152	0.125	0.066	0.058	0.048	0.125	0.113	0.073
lag = 3	1	–	–	–	–	–	–	0.010	0.015	0.013	0.012	0.012	0.008
T = 100	2	–	–	–	–	–	–	0.043	0.047	0.033	0.055	0.060	0.050
lag = 4	1	–	–	–	–	–	–	0.008	0.010	0.014	0.012	0.010	0.006
T = 100	2	–	–	–	–	–	–	0.035	0.036	0.033	0.051	0.048	0.041
lag = aic	1	0.164	0.153	0.111	0.005	0.007	0.007	0.010	0.017	0.016	0.007	0.011	0.005
T = 100	2	0.144	0.142	0.105	0.080	0.071	0.065	0.048	0.051	0.037	0.045	0.044	0.031
lag = sic	1	0.166	0.158	0.114	0.077	0.077	0.058	0.014	0.018	0.016	0.023	0.018	0.010
T = 100	2	0.147	0.144	0.105	0.173	0.180	0.138	0.070	0.061	0.049	0.075	0.070	0.058
lag = 1	1	0.595	0.519	0.373	0.716	0.630	0.441	–	–	–	–	–	–
T = 200	2	0.564	0.495	0.357	0.713	0.625	0.431	–	–	–	–	–	–
lag = 2	1	0.025	0.017	0.011	0.017	0.022	0.014	0.013	0.012	0.009	0.063	0.061	0.047
T = 200	2	0.332	0.288	0.199	0.524	0.464	0.303	0.232	0.192	0.123	0.385	0.344	0.236
lag = 3	1	–	–	–	–	–	–	0.008	0.004	0.009	0.023	0.024	0.014
T = 200	2	–	–	–	–	–	–	0.224	0.171	0.117	0.304	0.248	0.157
lag = 4	1	–	–	–	–	–	–	0.009	0.012	0.014	0.011	0.012	0.006
T = 200	2	–	–	–	–	–	–	0.146	0.109	0.074	0.221	0.172	0.117
lag = aic	1	0.572	0.493	0.352	0.007	0.005	0.008	0.009	0.006	0.010	0.008	0.010	0.003
T = 200	2	0.550	0.476	0.346	0.238	0.219	0.155	0.220	0.163	0.117	0.138	0.124	0.078
lag = sic	1	0.595	0.519	0.373	0.015	0.022	0.016	0.014	0.009	0.009	0.019	0.018	0.010
T = 200	2	0.564	0.495	0.357	0.462	0.406	0.266	0.233	0.188	0.122	0.252	0.207	0.137
lag = 1	1	0.832	0.768	0.556	0.909	0.858	0.633	–	–	–	–	–	–
T = 300	2	0.806	0.750	0.554	0.902	0.851	0.626	–	–	–	–	–	–
lag = 2	1	0.030	0.033	0.023	0.039	0.040	0.032	0.019	0.015	0.011	0.074	0.072	0.048
T = 300	2	0.617	0.522	0.368	0.801	0.730	0.509	0.469	0.378	0.246	0.634	0.542	0.340
lag = 3	1	–	–	–	–	–	–	0.021	0.020	0.015	0.025	0.028	0.015
T = 300	2	–	–	–	–	–	–	0.417	0.343	0.218	0.477	0.417	0.245
lag = 4	1	–	–	–	–	–	–	0.020	0.019	0.011	0.005	0.008	0.006
T = 300	2	–	–	–	–	–	–	0.276	0.235	0.158	0.384	0.318	0.205
lag = aic	1	0.788	0.732	0.527	0.007	0.005	0.007	0.019	0.020	0.014	0.005	0.008	0.010
T = 300	2	0.788	0.735	0.544	0.394	0.312	0.224	0.407	0.332	0.215	0.214	0.180	0.108
lag = sic	1	0.832	0.768	0.556	0.017	0.024	0.023	0.024	0.018	0.014	0.008	0.012	0.005
T = 300	2	0.806	0.750	0.554	0.696	0.590	0.406	0.444	0.352	0.232	0.392	0.326	0.202

[1] The approaches reported on in this table are based on the use of ex ante MSE, MAD, and MAPE forecasting criteria. The lags used are listed, as is the number of observations in the sample. The DGPs used to generate the data are given across the columns of the table. The cointegrating rank of the estimated VEC model is given in column 2 of the table. Numerical entries are the proportion of rejections of the "null hypothesis" of noncausality. All models are compared in pairwise fashion. For example, in this table each entry corresponds to the rejection frequency when two alternative cointegrating models are compared - one with the causal variable of interest included, and one without. The choice of which model is "best" is based on an examination of the forecasts errors used in the calculation of the MSE, MAD, and MAPE. In particular, predictive accuracy tests in the spirit of Diebold and Mariano (1995) are carried out (see above discussion).

Table 9a: MSE, MAD, and MAPE Type Ex Ante Predictive Ability[1], Data Generating Process is (CI-1). All Estimations Based on Assumption that It Is Known that the Data Are Generated According to a Levels VAR

Sample	Lag	VAR(1)-(MA1)			VAR(1)-(MA2)			VAR(3)-(MA1)			VAR(3)-(MA2)		
		MSE	MAD	MAPE	MSE	MAD	MAPE	MSE	MAD	MAPE	MSE	MAD	MAPE
	1	0.018	0.012	0.024	0.016	0.018	0.021	–	–	–	–	–	–
T = 100	2	0.006	0.007	0.007	0.016	0.016	0.014	0.013	0.015	0.013	0.021	0.021	0.013
	3	–	–	–	–	–	–	0.015	0.012	0.016	0.008	0.008	0.009
	4	–	–	–	–	–	–	0.009	0.007	0.012	0.004	0.013	0.012
	aic	0.018	0.014	0.025	0.003	0.005	0.009	0.015	0.013	0.014	0.008	0.011	0.013
	sic	0.018	0.012	0.024	0.019	0.021	0.017	0.013	0.016	0.014	0.010	0.013	0.009
	1	0.017	0.017	0.025	0.022	0.020	0.028	–	–	–	–	–	–
T = 200	2	0.007	0.009	0.010	0.012	0.015	0.016	0.013	0.012	0.018	0.036	0.039	0.037
	3	–	–	–	–	–	–	0.004	0.006	0.015	0.012	0.010	0.011
	4	–	–	–	–	–	–	0.006	0.009	0.013	0.009	0.004	0.011
	aic	0.017	0.017	0.025	0.004	0.004	0.009	0.005	0.007	0.016	0.002	0.004	0.003
	sic	0.017	0.017	0.025	0.010	0.013	0.013	0.011	0.008	0.015	0.012	0.007	0.013
	1	0.027	0.020	0.023	0.027	0.031	0.039	–	–	–	–	–	–
T = 300	2	0.009	0.013	0.013	0.010	0.014	0.014	0.010	0.012	0.018	0.047	0.049	0.026
	3	–	–	–	–	–	–	0.006	0.011	0.017	0.013	0.014	0.015
	4	–	–	–	–	–	–	0.007	0.009	0.010	0.003	0.009	0.014
	aic	0.026	0.020	0.022	0.002	0.012	0.008	0.005	0.011	0.017	0.003	0.004	0.010
	sic	0.027	0.020	0.023	0.006	0.008	0.012	0.013	0.012	0.018	0.003	0.009	0.013

[1] See notes to Table 8.

Table 9b: MSE, MAD, and MAPE Type Ex Ante Predictive Ability[1], Data Generating Process is (CI-1). All Estimations Based on Assumption that It Is Known that the Data Are Generated According to a VAR in Differences

Sample	Lag	VAR(1)-(MA1)			VAR(1)-(MA2)			VAR(3)-(MA1)			VAR(3)-(MA2)		
		MSE	MAD	MAPE	MSE	MAD	MAPE	MSE	MAD	MAPE	MSE	MAD	MAPE
	1	–	–	–	–	–	–	–	–	–	–	–	–
T = 100	2	0.066	0.064	0.041	0.083	0.083	0.073	0.031	0.029	0.025	0.063	0.071	0.042
	3	–	–	–	–	–	–	0.043	0.039	0.039	0.048	0.045	0.028
	4	–	–	–	–	–	–	0.038	0.033	0.032	0.044	0.046	0.033
	aic	0.055	0.00	0.00	0.069	0.076	0.059	0.034	0.034	0.035	0.036	0.043	0.025
	sic	–	–	–	0.086	0.094	0.080	0.031	0.028	0.023	0.061	0.059	0.031
	1	0.00	0.00	0.00	0.00	0.00	0.00	–	–	–	–	–	–
T = 200	2	0.180	0.175	0.128	0.301	0.266	0.186	0.054	0.049	0.035	0.156	0.139	0.084
	3	–	–	–	–	–	–	0.081	0.077	0.047	0.142	0.142	0.077
	4	–	–	–	–	–	–	0.088	0.072	0.060	0.147	0.128	0.081
	aic	0.180	0.120	0.100	0.182	0.164	0.115	0.078	0.074	0.049	0.101	0.101	0.074
	sic	0.00	0.00	0.00	0.290	0.256	0.174	0.057	0.052	0.039	0.144	0.132	0.073
	1	–	–	–	0.00	0.00	0.00	–	–	–	–	–	–
T = 300	2	0.371	0.298	0.190	0.542	0.446	0.304	0.090	0.070	0.059	0.242	0.193	0.106
	3	–	–	–	–	–	–	0.162	0.141	0.092	0.257	0.207	0.122
	4	–	–	–	–	–	–	0.159	0.119	0.088	0.259	0.210	0.126
	aic	0.412	0.353	0.157	0.349	0.279	0.201	0.160	0.136	0.091	0.176	0.156	0.086
	sic	–	–	–	0.505	0.390	0.280	0.131	0.110	0.081	0.264	0.218	0.131

[1] See notes to Table 8.

Table 10: Summary of SIC and AIC Criteria Approach to Causal Analysis[1], Data Generating Process is (CI-2)

DGP/ Sample	lag	AIC	SIC	Model Selected- aic			Model Selected - sic		
				CI	levels	difference	CI	levels	difference
VAR(1)-(MA1)	1	0.962	0.994	0.961	0.039	–	0.994	0.006	–
T = 100	2	0.388	0.123	0.858	0.117	0.025	0.911	0.044	0.045
	aic	0.931	0.950	0.953	0.043	0.004	0.983	0.008	0.009
	sic	0.962	0.994	0.961	0.039	–	0.994	0.006	–
	1	0.951	0.994	0.949	0.051	–	0.994	0.006	–
T = 200	2	0.351	0.082	0.863	0.135	0.002	0.952	0.044	0.004
	aic	0.932	0.958	0.946	0.053	0.001	0.992	0.007	0.001
	sic	0.951	0.994	0.949	0.051	–	0.994	0.006	–
	1	0.946	0.997	0.943	0.057	–	0.997	0.003	–
T = 300	2	0.334	0.061	0.876	0.124	0.000	0.958	0.042	0.000
	aic	0.925	0.946	0.941	0.059	0.000	0.994	0.006	0.000
	sic	0.946	0.997	0.943	0.057	–	0.997	0.003	–
(VAR(1))-(MA2)	1	0.945	0.984	0.913	0.087	–	0.980	0.020	–
T = 100	2	0.440	0.123	0.909	0.091	0.000	0.986	0.014	0.000
	aic	0.353	0.046	0.804	0.134	0.062	0.919	0.056	0.025
	sic	0.436	0.121	0.892	0.101	0.007	0.971	0.019	0.010
	1	0.935	0.986	0.909	0.091	–	0.986	0.014	–
T = 200	2	0.457	0.082	0.938	0.062	0.000	0.990	0.010	0.000
	aic	0.233	0.007	0.839	0.132	0.029	0.971	0.027	0.002
	sic	0.335	0.030	0.916	0.084	0.000	0.989	0.011	0.000
	1	0.932	0.980	0.901	0.099	–	0.979	0.021	–
T = 300	2	0.465	0.063	0.915	0.085	0.000	0.990	0.010	0.000
	aic	0.225	0.002	0.835	0.145	0.020	0.959	0.041	0.000
	sic	0.294	0.011	0.895	0.105	0.000	0.989	0.011	0.000
(VAR(3))-(MA1)	2	0.388	0.094	0.885	0.115	0.00	0.955	0.043	0.002
T = 100	3	0.293	0.035	0.828	0.157	0.015	0.938	0.054	0.008
	4	0.248	0.013	0.798	0.165	0.037	0.941	0.055	0.004
	aic	0.333	0.049	0.832	0.152	0.016	0.936	0.055	0.009
	sic	0.402	0.118	0.881	0.116	0.003	0.953	0.044	0.003
	2	0.366	0.057	0.895	0.105	0.000	0.964	0.036	0.000
T = 200	3	0.287	0.024	0.873	0.126	0.001	0.933	0.064	0.003
	4	0.234	0.008	0.848	0.143	0.009	0.923	0.071	0.006
	aic	0.294	0.025	0.870	0.125	0.005	0.928	0.068	0.004
	sic	0.365	0.056	0.893	0.107	0.000	0.955	0.044	0.001
	2	0.402	0.041	0.886	0.114	0.000	0.967	0.033	0.000
T = 300	3	0.264	0.011	0.859	0.141	0.000	0.940	0.060	0.000
	4	0.215	0.001	0.840	0.159	0.001	0.932	0.068	0.000
	aic	0.274	0.010	0.865	0.135	0.000	0.940	0.060	0.000
	sic	0.330	0.026	0.874	0.126	0.000	0.947	0.053	0.000
(VAR(3))-(MA2)	2	0.671	0.381	0.912	0.088	0.000	0.932	0.068	0.000
T = 100	3	0.404	0.049	0.927	0.072	0.001	0.978	0.021	0.001
	4	0.276	0.017	0.908	0.087	0.005	0.975	0.019	0.006
	aic	0.293	0.015	0.834	0.130	0.036	0.936	0.055	0.009
	sic	0.376	0.051	0.921	0.077	0.002	0.977	0.021	0.002
	2	0.745	0.383	0.907	0.093	0.000	0.945	0.055	0.000
T = 200	3	0.392	0.041	0.924	0.076	0.000	0.988	0.012	0.000
	4	0.259	0.003	0.899	0.100	0.001	0.986	0.012	0.002
	aic	0.200	0.002	0.821	0.153	0.026	0.963	0.037	0.000
	sic	0.270	0.008	0.899	0.098	0.003	0.986	0.013	0.001
	2	0.783	0.404	0.905	0.095	0.000	0.914	0.086	0.000
T = 300	3	0.384	0.026	0.882	0.118	0.000	0.974	0.026	0.000
	4	0.238	0.001	0.868	0.132	0.000	0.981	0.019	0.000
	aic	0.174	0.000	0.810	0.174	0.016	0.959	0.041	0.000
	sic	0.224	0.002	0.860	0.139	0.001	0.981	0.018	0.001

[1] See notes to Table 5.

Table 11: Summary of SIC and AIC Criteria Approach to Causal Analysis[1], Data Generating Process is (CI-2). All Estimations Based on Assumption that It Is Known that the Data Are Generated According to a VEC

Lag/	CI	VAR(1)-(MA1)		VAR(1)-(MA2)		VAR(3)-(MA1)			VAR(3)-(MA2)
Sample	rank	AIC	SIC	AIC	SIC	AIC	SIC	AIC	SIC
lag = 1	1	0.064	0.064	0.003	0.003	–	–	–	–
T = 100	2	1.00	1.00	1.00	1.00	–	–	–	–
lag = 2	1	0.218	0.081	0.192	0.049	0.209	0.067	0.297	0.109
T = 100	2	0.983	0.757	1.00	1.00	0.996	0.915	1.00	0.999
lag = 3	1	–	–	–	–	0.193	0.041	0.196	0.022
T = 100	2	–	–	–	–	0.913	0.379	0.999	0.970
lag = 4	1	–	–	–	–	0.195	0.020	0.148	0.006
T = 100	2	–	–	–	–	0.714	0.096	0.997	0.773
lag = aic	1	0.086	0.074	0.220	0.026	0.218	0.049	0.190	0.010
T = 100	2	0.999	0.982	0.983	0.697	0.914	0.497	0.942	0.414
lag = sic	1	0.064	0.064	0.197	0.048	0.209	0.064	0.195	0.019
T = 100	2	1.00	1.00	1.00	0.991	0.991	0.893	0.998	0.914
lag = 1	1	0.068	0.068	0.00	0.00	–	–	–	–
T = 2 00	2	1.00	1.00	1.00	1.00	–	–	–	–
lag = 2	1	0.219	0.087	0.175	0.020	0.196	0.043	0.357	0.108
T = 2 00	2	1.00	0.965	1.00	1.00	1.00	0.999	1.00	1.00
lag = 3	1	–	–	–	–	0.193	0.058	0.181	0.012
T = 2 00	2	–	–	–	–	0.998	0.745	1.00	1.00
lag = 4	1	–	–	–	–	0.177	0.027	0.141	0.002
T = 2 00	2	–	–	–	–	0.943	0.248	1.00	0.992
lag = aic	1	0.085	0.073	0.147	0.013	0.202	0.058	0.142	0.010
T = 2 00	2	1.00	0.992	0.998	0.810	0.995	0.723	0.994	0.590
lag = sic	1	0.068	0.068	0.160	0.007	0.205	0.051	0.149	0.005
T = 2 00	2	1.00	1.00	1.00	1.00	1.00	0.947	1.00	0.981
lag = 1	1	0.069	0.069	0.002	0.002	–	–	–	–
T = 3 00	2	1.00	1.00	1.00	1.00	–	–	–	–
lag = 2	1	0.194	0.079	0.171	0.021	0.208	0.035	0.398	0.120
T = 3 00	2	1.00	0.998	1.00	1.00	1.00	1.00	1.00	1.00
lag = 3	1	–	–	–	–	0.172	0.061	0.171	0.008
T = 3 00	2	–	–	–	–	1.00	0.944	1.00	1.00
lag = 4	1	–	–	–	–	0.165	0.041	0.116	0.00
T = 3 00	2	–	–	–	–	0.989	0.451	1.00	1.00
lag = aic	1	0.083	0.069	0.142	0.022	0.180	0.059	0.118	0.011
T = 3 00	2	1.00	0.999	1.00	0.902	1.00	0.918	0.999	0.734
lag = sic	1	0.069	0.069	0.140	0.007	0.201	0.058	0.116	0.00
T = 3 00	2	1.00	1.00	1.00	1.00	1.00	0.966	1.00	0.997

[1] See notes to Table 5.

Table 12a: Summary of SIC and AIC Criteria Approach to Causal Analysis[1], Data Generating Process is (CI-2). All Estimations Based on Assumption that It Is Known that the Data Are Generated According to a Levels VAR

Sample	Lag	VAR(1)-(MA1) AIC	VAR(1)-(MA1) SIC	VAR(1)-(MA2) AIC	VAR(1)-(MA2) SIC	VAR(3)-(MA1) AIC	VAR(3)-(MA1) SIC	VAR(3)-(MA2) AIC	VAR(3)-(MA2) SIC
	1	0.341	0.115	0.554	0.349	–	–	–	–
T = 100	2	0.270	0.040	0.358	0.078	0.295	0.043	0.602	0.301
	3	–	–	–	–	0.221	0.014	0.346	0.037
	4	–	–	–	–	0.201	0.008	0.229	0.011
	aic	0.344	0.113	0.282	0.018	0.253	0.022	0.240	0.005
	sic	0.341	0.115	0.353	0.072	0.304	0.051	0.320	0.036
	1	0.301	0.080	0.563	0.329	–	–	–	–
T = 200	2	0.241	0.022	0.360	0.053	0.283	0.028	0.617	0.288
	3	–	–	–	–	0.202	0.006	0.327	0.031
	4	–	–	–	–	0.186	0.00	0.209	0.001
	aic	0.308	0.078	0.193	0.002	0.208	0.007	0.160	0.00
	sic	0.301	0.080	0.270	0.019	0.271	0.026	0.218	0.005
	1	0.310	0.057	0.599	0.342	–	–	–	–
T = 300	2	0.224	0.011	0.352	0.043	0.296	0.015	0.687	0.292
	3	–	–	–	–	0.182	0.002	0.312	0.020
	4	–	–	–	–	0.164	0.00	0.187	0.001
	aic	0.307	0.054	0.165	0.001	0.191	0.002	0.139	0.00
	sic	0.310	0.057	0.220	0.006	0.230	0.005	0.179	0.001

[1] See notes to Table 6a.

Table 12b: Summary of SIC and AIC Criteria Approach to Causal Analysis[1], Data Generating Process is (CI-2). All Estimations Based on Assumption that It Is Known that the Data Are Generated According to a VAR in Differences

Sample	Lag	VAR(1)-(MA1) AIC	VAR(1)-(MA1) SIC	VAR(1)-(MA2) AIC	VAR(1)-(MA2) SIC	VAR(3)-(MA1) AIC	VAR(3)-(MA1) SIC	VAR(3)-(MA2) AIC	VAR(3)-(MA2) SIC
	1	–	–	–	–	–	–	–	–
T = 100	2	0.829	0.567	0.995	0.937	0.385	0.164	0.309	0.109
	3	–	–	–	–	0.401	0.093	0.646	0.214
	4	–	–	–	–	0.364	0.047	0.884	0.388
	aic	0.828	0.609	0.968	0.702	0.406	0.111	0.870	0.277
	sic	0.00	0.00	0.994	0.924	0.386	0.164	0.724	0.282
	1	–	–	–	–	–	–	–	–
T = 200	2	0.978	0.828	1.00	0.999	0.594	0.280	0.332	0.108
	3	–	–	–	–	0.641	0.182	0.861	0.411
	4	–	–	–	–	0.550	0.068	0.996	0.797
	aic	0.930	0.512	0.997	0.874	0.642	0.186	0.993	0.548
	sic	0.00	0.00	1.00	0.996	0.601	0.260	0.982	0.762
	1	–	–	–	–	–	–	–	–
T = 300	2	0.995	0.948	1.00	1.00	0.722	0.382	0.403	0.131
	3	–	–	–	–	0.789	0.271	0.961	0.613
	4	–	–	–	–	0.718	0.111	1.00	0.947
	aic	0.963	0.870	1.00	0.930	0.787	0.264	0.998	0.747
	sic	–	–	1.00	1.00	0.761	0.316	1.00	0.944

[1] See notes to Table 6b.