Journal of Econometrics 21 (1983) 161–194. North-Holland Publishing Company

# COMPARING ALTERNATIVE TESTS OF CAUSALITY IN TEMPORAL SYSTEMS

## Analytic Results and Experimental Evidence*

John GEWEKE

*University of Wisconsin, Madison, WI 53706, USA*

Richard MEESE

*University of California, Berkeley, CA 94720, USA*

Warren DENT

*Eli Lilly International Corporation, Indianapolis, IN 46206, USA*

Received February 1982, final version received May 1982

This paper discusses eight alternative tests of the absence of causal ordering, all of which are asymptotically valid under the null hypothesis in the sense that their limiting size is known. Their behavior under alternatives is compared analytically using the concept of approximate slope, and these results are supported by the outcomes of Monte Carlo experiments. The implications of these comparisons for applied work are unambiguous: Wald variants of a test attributed to Granger, and a lagged dependent variable version of Sims' test introduced in this paper, are equivalent in all relevant respects and are preferred to the other tests discussed.

## 1. Introduction

Granger (1969) has proposed a definition of 'causality' in economic systems which has been applied frequently in empirical work. Briefly, a time series $\{y_t\}$ 'causes' another time series $\{x_t\}$ if present $x$ can be predicted better by using past values of $y$ than by not doing so, other relevant information (including the past of $x$) being used in either case. When the domain of relevant information is restricted to past values of $x$ and $y$, $x$ and $y$ are jointly covariance stationary time series with autoregressive representation, the set of predictors is constrained to be linear in past $x$ and $y$, and one's

*Thanks to Susan Hudak for research assistance and R.R. Bahadur for helpful discussions on many points raised in the paper. Responsibility for remaining errors resides with the authors. Views expressed are not necessarily those of the Board of Governors or the Federal Reserve. Most of the work reported here was undertaken while the second author was a graduate student and the third was Visiting Associate Professor in the Economics Department, University of Wisconsin, Madison. The first author acknowledges financial support from the National Science Foundation.

0304–4076/83/0000–0000/$03.00 

criterion for comparison of predictors is mean square error, then the question of whether $\{y_t\}$ causes $\{x_t\}$ is equivalent to the question of whether there exists an autoregressive representation,

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{bmatrix} d(L) & e(L) \\ a(L) & b(L) \end{bmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \delta_t \end{pmatrix}, \qquad \begin{matrix} (1.1\text{a}) \\ (1.1\text{b}) \end{matrix}$$

in which $e(L)=0$.

Since (1.1) is a set of population regression equations with serially uncorrelated disturbances, a test of the proposition that $y$ does not cause $x$ is immediately suggested: estimate (1.1a) by ordinary least squares, and test the hypothesis $e(L)=0$ in the conventional way. We shall refer to such a test as a 'Granger test' since the restriction $e(L)=0$ stems directly from Granger's definition. Unless one has special knowledge which restricts $d(L)$ and $e(L)$ to be known functions of parameters whose number is reasonably small relative to sample size, however, this cannot be done. In actual applications [e.g., Sargent (1976) and Geweke (1978)] some arbitrary restrictions on the form of $d(L)$ and $e(L)$ must be made before estimation and testing can proceed; for example, $d(L)$ and $e(L)$ might be assumed to be polynomials of finite order, or ratios of polynomials of finite order. If the restrictions are true, then the asymptotic properties of conventional tests of $e(L)=0$ are well known. If they are false, such tests will in general reject the null hypothesis $e(L)=0$ asymptotically even if the null is true.

Sims (1972) considered the implications of the condition $e(L)=0$ for the projection of $y_t$ on current and past $x_t$,

$$\begin{aligned} y_t &= f_1(L)x_t + u_t = \sum_{s=0}^{\infty} f_{1s}x_{t-s} + u_t, \\ \operatorname{cov}(u_t, x_{t-s}) &= 0, \qquad s \geqq 0, \end{aligned} \qquad (1.2)$$

and the projection of $y_t$ on future, current and past $x_t$,

$$\begin{aligned} y_t &= f_2(L)x_t + v_t = \sum_{s=-\infty}^{\infty} f_{2s}x_{t-s} + v_t, \\ \operatorname{cov}(v_t, x_{t-s}) &= 0, \qquad \forall s. \end{aligned} \qquad (1.3)$$

He showed that $e(L)=0$ — i.e., $\{y_t\}$ does not cause $\{x_t\}$ — if and only if the second projection involves only current and past $x_t$, so that $f_1(L)=f_2(L)$. We shall refer to a test of the hypothesis $f_1(L)=f_2(L)$ as a 'Sims test' of the hypothesis that $\{y_t\}$ does not cause $\{x_t\}$. If such a test is to be implemented, $f_1(L)$ and $f_2(L)$ must be restricted for the same reasons that $d(L)$ and $e(L)$ were restricted in conducting a Granger test. In addition, because $u_t$ and $v_t$ are generally serially correlated, a feasible generalized least squares estimator which restricts autocovariance functions of $u_t$ and $v_t$ to known functions of a reasonable number of parameters must be employed in applications of a Sims test. Again, the asymptotic bias against the null hypothesis which would in general result from maintaining restrictions on $f_1(L)$ and $f_2(L)$ may be removed by allowing the number of estimated parameters to increase with sample size. As before, appropriate rates of parameter expansion are unknown.

There is a variant of the Sims test which avoids the need to use a feasible generalized least squares estimator or an *ad hoc* prefilter to cope with serial correlation in $\{u_t\}$. Premultiply the system of equations (1.1) by the matrix

$$\begin{bmatrix} 1 & 0 \\ -c/\sigma_\varepsilon^2 & 1 \end{bmatrix},$$

where

$$c=\operatorname{cov}(\varepsilon_t, \delta_t),$$

to yield a new system in which the first equation is (1.1a) and the second can be expressed as

$$\begin{aligned} y_t &= g^+(L)y_t + f_1^*(L)x_t + \eta_t \\ &= \sum_{s=1}^{\infty} g_s^+ y_{t-s} + \sum_{s=0}^{\infty} f_{1s}^* x_{t-s} + \eta_t, \end{aligned} \tag{1.4}$$

where $g^+(L)$ and $f_1^*(L)$ are functions of the parameters of (1.1). Since $\eta_t = \delta_t - (c/\sigma_\varepsilon^2)\varepsilon_t$, it is uncorrelated with $\varepsilon_t$, and hence $\eta_t$ is uncorrelated with current $x_t$ as well as past $x_t$ and $y_t$. The variables on the right side of (1.4) are therefore predetermined, and $\eta_t$ is serially uncorrelated. Now suppose that $v_t$ in (1.3) has autoregressive representation $h(L)v_t=\omega_t$, an assumption which is no stronger than those employed in rigorous justifications for the use of feasible generalized least squares in this equation. Let $h(L)=1-h^+(L)$ and $f_2^*(L)=h(L)f_2(L)$. Multiplying through (1.3) by $h(L)$ and rearranging results in

$$\begin{aligned} y_t &= h^+(L)y_t + f_2^*(L)x_t + \omega_t \\ &= \sum_{s=1}^{\infty} h_s^+ y_{t-s} + \sum_{s=-\infty}^{\infty} f_{2s}^* x_{t-s} + \omega_t. \end{aligned} \tag{1.5}$$

The disturbance $\omega_t$ is serially uncorrelated by construction. It is uncorrelated with all $x_{t-s}$ [because it is a linear combination of disturbances in the projection (1.3)] and with past values of $y_t$ (because these are linear combinations of all $x_{t-s}$ and past $\omega_t$). Since $h(L)$ is invertible, $f^*_{2s}=0$ for all $s>0$ if and only if $f_{2s}=0$ for all $s>0$. A 'Sims test' of the hypothesis that $\{y_t\}$ does not cause $\{x_t\}$ may therefore be conducted by basing a test of $f^*_{2s}=0$ for all $s<0$ on the estimates of some finite parameterization of (1.4) and (1.5) in the conventional way. Again, the number of parameters in $g^+(L)$, $h^+(L)$, $f^*_1(L)$ and $f^*_2(L)$ which are estimated must increase with sample size.

In this paper we report the results of an investigation into the actual properties of several tests of the hypothesis that $y$ does not cause $x$. This investigation was undertaken for three reasons. First, as is evident from the foregoing discussion, the theory which underlies these tests is only asymptotic. Because of the need to work with a finite number of parameters in a sample of finite size, the theory is even less operational than in the more familiar case in which the parameterization is known: in the latter case we know exactly how to construct the test statistic but are in doubt about its precise interpretation in a sample of a given size, whereas in the tests discussed above it is not known exactly how the test statistic should be constructed *or* interpreted in any given situation.

Second, once a parameterization has been adopted, there remain several ways in which linear restrictions may be tested. The most common are the Wald, likelihood ratio, and LaGrange-multiplier tests discussed by Silvey (1959); as Berndt and Savin (1977) have demonstrated, there are practical situations in which the outcome depends on the test used.

Finally, and most important, as the results of more and more tests for the absence of causal orderings are reported in the applied literature, it is becoming evident [see Pierce and Haugh (1977)] that these results may depend on which test is used.

The remainder of the paper is organized as follows. In the next section we introduce eight alternative tests of the hypothesis that $y$ does not cause $x$. All are asymptotically valid in the sense that under the null hypothesis the difference between the distribution of the test statistic and that of a sequence of chi-square distributions vanishes as sample size increases. In section 3 we compare the asymptotic behavior of all eight tests using the concept of approximate slope [Bahadur (1960) and Geweke (1981b)]. The design of a Monte Carlo experiment conducted to assess the behavior of these alternative tests in situations typical of those encountered in economic applications is discussed in section 4. The outcome of the experiment is reported in section 5, and a final section contains recommendations for empirical work based on the results of our investigation.

## 2. Alternative tests for the absence of a causal ordering

In what follows, we shall assume that $\{x_t\}$ and $\{y_t\}$ are jointly stationary, Gaussian, and have mean zero. The restriction that the set of predictors be linear in past $x$ and $y$ is then inconsequential, mean square error becomes the natural metric for comparison of forecasts, and the relevant likelihood functions are well known.

A Granger test of the hypothesis that $\{y_t\}$ does not cause $\{x_t\}$ is a test of the restriction $e(L)=0$ in (1.1a). If we write the univariate autoregressive representation of $x_t$ as

$$x_t = c(L)x_t + \zeta_t, \tag{2.1}$$

then $e(L)=0$ is equivalent to $\{\varepsilon_t\}=\{\zeta_t\}$ and $c(L)=d(L)$. If under the maintained hypothesis $d(L)$ were assumed to be a polynomial of order $l$ and $e(L)$ were assumed to be a polynomial of order $k$, then the test of $e(L)=0$ could be based on the sums of squared residuals from ordinary least squares estimates of the regression equations

$$x_t = C(L)x_t + \mathscr{X}_t = \sum_{s=1}^{l} C_s x_{t-s} + \mathscr{X}_t, \tag{2.2}$$

$$x_t = D(L)x_t + E(L)y_t + \mathscr{E}_t = \sum_{s=1}^{l} D_s x_{t-s} + \sum_{s=1}^{k} E_s y_{t-s} + \mathscr{E}_t. \tag{2.3}$$

If $\hat{\sigma}^2_{\mathscr{X}}$ is the maximum likelihood estimate of $\operatorname{var}(\mathscr{X}_t)$ in (2.2) and $\hat{\sigma}^2_{\mathscr{E}}$ is the maximum likelihood estimate of $\operatorname{var}(\mathscr{E}_t)$ in (2.3), then under the null hypothesis the distribution of each of the statistics,

$$T_n^{GW} = n(\hat{\sigma}^2_{\mathscr{X}} - \hat{\sigma}^2_{\mathscr{E}})/\hat{\sigma}^2_{\mathscr{E}}, \tag{2.4}$$

$$T_n^{GR} = n\log(\hat{\sigma}^2_{\mathscr{X}}/\hat{\sigma}^2_{\mathscr{E}}), \tag{2.5}$$

$$T_n^{GL} = n(\hat{\sigma}^2_{\mathscr{X}} - \hat{\sigma}^2_{\mathscr{E}})/\hat{\sigma}^2_{\mathscr{X}}, \tag{2.6}$$

converges uniformly to that of a $\chi^2(k)$ variate as the size of the sample, $n$, increases without bound. In the terminology of Silvey (1959), a test based on $T_n^{GW}$ and this convergence result is a 'Wald test', while one based on $T_n^{GR}$ or $T_n^{GL}$ and the convergence result is a 'likelihood ratio test' or a 'LaGrange-multiplier test', respectively.

In practice $k$ and $l$ are not known, nor even known to be finite. Furthermore, there are no known relations between $l$ and $n$ which guarantee the uniform convergence of (2.4)–(2.6) to a $\chi^2(k)$ variate under the null hypothesis. In practice, values of $k$ and $l$ are usually chosen according to procedures which are at best informal, if not arbitrary.

The restriction that $f_1(L)=f_2(L)$ in (1.2) and (1.3) on which Sims tests are based is equivalent to $\{u_t\}=\{v_t\}$. If under the maintained hypothesis $f_2(L)$ were restricted to a polynomial in the terms $L^{-p},\ldots,L^{-1},0,L,\ldots,L^q$, then a test could be based on the regression equations

$$y_t=F_1(L)x_t+U_t=\sum_{s=0}^{q}F_{1s}x_{t-s}+U_t, \tag{2.7}$$

$$y_t=F_2(L)x_t+V_t=\sum_{s=-p}^{q}F_{2s}x_{t-s}+V_t. \tag{2.8}$$

When an *ad hoc* prefilter [say, $R(L)$] is used to cope with serial correlation, $y_t$ and $x_t$ are replaced by $y_t^\dagger=R(L)y_t$ and $x_t^\dagger=R(L)x_t$, respectively, in (2.7) and (2.8). The equations are then estimated by ordinary least squares and the test $F_1(L)=F_2(L)$ is conducted using a conventional Wald test statistic. In more sophisticated feasible generalized least squares procedures, the serial correlation structure of the disturbances is estimated. In a typical two-step estimator [e.g., Hannan (1963) or Amemiya (1973)], (2.8) is first estimated by ordinary least squares. It is assumed that $\Omega_n^v=\operatorname{var}(V_1,\ldots,V_n)=\Omega_n^v(\boldsymbol{\alpha}_v)$, where the function $\Omega^v(\cdot)$ is known but the $m_v\times 1$ vector $\boldsymbol{\alpha}_v$ is unknown. The maximum likelihood estimate $\hat{\boldsymbol{\alpha}}_v$ (or its asymptotic equivalent) of $\boldsymbol{\alpha}_v$ is computed by ignoring differences between ordinary least squares residuals for (2.8) and $V_t$. The vector of coefficients $\hat{F}_2$ is then estimated by generalized least squares, replacing $\Omega_n^v$ with $\Omega_n^v(\hat{\boldsymbol{\alpha}}_v)$, which provides a vector of residuals $\hat{V}$. To construct the Wald test statistic, (2.7) is then estimated by generalized least squares, replacing $\Omega_n^u$ with $\Omega_n^v(\hat{\boldsymbol{\alpha}}_v)$, yielding a vector of residuals $\hat{U}_v$. The test statistic is

$$T^{(SW)}=\hat{U}_v'(\Omega_n^v(\hat{\boldsymbol{\alpha}}_v))^{-1}\hat{U}_v-\hat{V}'(\Omega_n^v(\hat{\boldsymbol{\alpha}}_v))^{-1}\hat{V}. \tag{2.9}$$

The LaGrange-multiplier test statistic is constructed by replacing (2.8) with (2.7) for purposes of obtaining an estimated serial correlation structure of the disturbances. In an obvious notation, the test statistic which results is

$$T^{(SL)}=\hat{U}'(\Omega_n^u(\hat{\boldsymbol{\alpha}}_u))^{-1}\hat{U}-\hat{V}_u'(\Omega_n^u(\hat{\boldsymbol{\alpha}}_u))^{-1}\hat{V}_u. \tag{2.10}$$

Under the null hypothesis the distributions of (2.9) and (2.10) converge to that of a $\chi^2(p)$ variate. When $m_U$, $m_V$, $p$ and $q$ are unknown and perhaps not even known to be finite, $m_U$, $m_V$, and $q$ may be increased with $n$: as functions of $n$, however, proper choices of $m_U$, $m_V$, and $q$ have not been worked out.

In implementations of (1.4) and (1.5) the parameterization problem may be managed just as it was in the Granger test based on (1.1). A test of $f_{2s}^*=0$ for all $s<0$ can be based on the sums of squared residuals from ordinary least squares estimates of the regression equations

$$y_t = G^+(L)y_t + F_1^*(L)x_t + H_t = \sum_{s=1}^{r} G_s^+ y_{t-s} + \sum_{s=0}^{q} F_{1s}^* x_{t-s} + H_t, \tag{2.11}$$

$$y_t = H^+(L)y_t + F_2^*(L)x_t + W_t = \sum_{s=1}^{r} H_s^+ y_{t-s} + \sum_{s=-p}^{q} F_{2s}^* x_{t-s} + W_t. \tag{2.12}$$

If $\hat{\sigma}_H^2$ is the maximum likelihood estimate of $\operatorname{var}(H_t)$ in (2.11) and $\hat{\sigma}_W^2$ is the maximum likelihood estimate of $\operatorname{var}(W_t)$ in (2.12), then Wald, likelihood ratio and LaGrange-multiplier tests may be based on the respective statistics

$$T_n^{SW} = n(\hat{\sigma}_H^2 - \hat{\sigma}_W^2)/\hat{\sigma}_W^2, \tag{2.13}$$

$$T_n^{SR} = n\ln(\hat{\sigma}_H^2/\hat{\sigma}_W^2), \tag{2.14}$$

$$T_n^{SL} = n(\hat{\sigma}_H^2 - \hat{\sigma}_W^2)/\hat{\sigma}_H^2, \tag{2.15}$$

all of which have a limiting $\chi^2(p)$ distribution under the null hypothesis.

## 3. The approximate slopes of the alternative tests

In the comparison of alternative tests of the same hypothesis we are concerned with adequacy of the limiting distributions under the null hypothesis and the ability to reject the null under various alternatives, in finite samples. In the absence of any analytical results for small samples, these questions may be addressed by sampling experiments, but without some sort of paradigm based on asymptotic considerations the results of sampling experiments are difficult, if not impossible, to interpret. The results of the previous section provide such a paradigm only when the null hypothesis is true. When the null is false, a convenient paradigm is the criterion of approximate slope, which was introduced by Bahadur (1960) and whose econometric applications were discussed by Geweke (1981b). We shall briefly recapitulate some of the points of the latter paper.

Suppose that test $i$ rejects the null in favor of the alternative in a sample of size $n$ if the statistic $T_n^i$ exceeds a critical value. When the limiting distribution of $T_n^i$ under the null is chi-square, the approximate slope of test $i$ is the almost sure limit of $T_n^i/n$, which we shall denote $T^i(\boldsymbol{\theta})$. The set $\boldsymbol{\theta}$ consists of all unknown parameters of the population distribution, perhaps countably infinite in number, and the approximate slope in general depends on the values of the elements of $\boldsymbol{\theta}$. Most test statistics are constructed in such a way that $T^i(\boldsymbol{\theta}) = 0$ for all $\boldsymbol{\theta}$ which satisfy the null hypothesis, and for all other $\boldsymbol{\theta}$, $T^i(\boldsymbol{\theta}) > 0$. The approximate slopes of alternative tests are related to their comparative behavior under the alternative in the following way. Let $n^i(t^*, \beta; \boldsymbol{\theta})$ denote the minimum number of observations required to insure that the probability that the test statistic $T_n^i$ exceeds a specified critical point $t^*$ is at least $1-\beta$. Then

$$\lim_{t^* \to \infty} n^1(t^*, \beta; \boldsymbol{\theta})/n^2(t^*, \beta; \boldsymbol{\theta}) = T^2(\boldsymbol{\theta})/T^1(\boldsymbol{\theta});$$

the ratio of the number of observations required to reject the alternative as $t^*$ is increased without bound (or alternatively, the asymptotic significance level of the test is reduced to zero) is inversely proportional to the ratio of their approximate slopes. Similarly, if $t^i(n\,\beta; \boldsymbol{\theta})$ indicates the largest non-rejection region (equivalently, smallest asymptotic significance level) possible in a sample of size $n$ if power $1-\beta$ is to be maintained against the alternative $\boldsymbol{\theta}$, then

$$\lim_{n \to \infty} t^1(n, \beta; \theta)/t^2(n, \beta; \theta) = T^1(\theta)/T^2(\theta).$$

In this section we confine consideration to the probability limits of $T_n^i/n$ for the various tests; demonstration of almost sure convergence appears to be confounded by the absence of a known, finite parameterization.

The approximate slopes of the three variants of the Granger test are readily calculated. If the parameters $j$ and $k$ in (2.3) increase suitably with sample size, then $\hat{\sigma}_{\varepsilon}^2 \xrightarrow{P} \sigma_{\varepsilon}^2$, where $\sigma_{\varepsilon}^2$ is the variance of the disturbance $\varepsilon_t$ in (1.1a). The increase in $l$ guarantees that when the null hypothesis $E(L)=0$ is imposed, the estimated variance of $Z_t$ in (2.2), $\hat{\sigma}_Z^2$, will converge in probability to the variance $\sigma_{\zeta}^2$ of the innovation in the univariate autoregressive representation of $\{x_t\}$. The latter variance is

$$\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln[S_x(\lambda)]\,\mathrm{d}\lambda\right\},$$

where $S_x(\lambda)$ denotes the spectral density of $\{x_t\}$ at frequency $\lambda$ [Whittle (1963, ch. 2)]. Consequently the approximate slopes of the tests on (2.4)–(2.6) are

$$T^{GW} = \sigma_{\zeta}^2/\sigma_{\varepsilon}^2 - 1 \tag{3.1}$$

for the Wald variant,

$$T^{GR} = \ln(\sigma_{\zeta}^2/\sigma_{\varepsilon}^2) \tag{3.2}$$

for the likelihood ratio variant, and

$$T^{GL} = 1 - \sigma_{\varepsilon}^2/\sigma_{\zeta}^2 \tag{3.3}$$

for the LaGrange-multiplier variant. The problem of deriving the limits of $T_n^{SW}/n$, $T_n^{SR}/n$, and $T_n^{SL}/n$ is formally the same. With suitable expansion of parameter spaces, we have [in view of (2.13)–(2.15) and the relation of (2.11) to (1.4) and (2.12) to (1.5)]

$$T^{SW}=\sigma_\eta^2/\sigma_\omega^2-1, \tag{3.4}$$

$$T^{SR}=\ln(\sigma_\eta^2/\sigma_\omega^2), \tag{3.5}$$

$$T^{SL}=1-\sigma_\omega^2/\sigma_\eta^2. \tag{3.6}$$

Since $x-1\geqq\ln(x)\geqq 1-(1/x)$ for all $x\geqq 1$, $T^{GW}\geqq T^{GR}\geqq T^{GL}$ and $T^{SW}\geqq T^{SR}\geqq T^{SL}$, the inequalities being strict under all alternative hypotheses. In view of the well-known relation which prevails among Wald, likelihood ratio, and LaGrange-multiplier test statistics [e.g., Breusch (1979)], these inequalities are not surpsing. To establish further orderings, it is convenient to use two results in spectral analysis. The first is that in the projection of $y_t$ on all $x_t$, eq. (1.3), the spectral density of the residual is $S_y(\lambda)-|S_{xy}(\lambda)|^2/S_x(\lambda)$ [e.g., Whittle (1963, p. 99)]. The second is that if $z_t$ is a multivariate stationary time series with autoregressive representation and $\varepsilon_t$ is the vector of innovations in the autoregressive representation, then

$$\ln|\mathrm{var}(\varepsilon_t)|=\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln|S_z(\lambda)|\,\mathrm{d}\lambda$$

[Rozanov (1967, p. 72)]. Since $\omega_t$ is the innovation in $v_t$ in (1.3),

$$\ln(\sigma_\omega^2)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln(S_y(\lambda)-|S_{xy}(\lambda)|^2/S_x(\lambda))\,\mathrm{d}\lambda.$$

We also have

$$\ln(\sigma_\zeta^2)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln(S_x(\lambda))\,\mathrm{d}\lambda,$$

whence

$$\ln(\sigma_\omega^2\sigma_\zeta^2)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln(S_x(\lambda)S_y(\lambda)-|S_{xy}(\lambda)|^2)\,\mathrm{d}\lambda=\ln(\sigma_\varepsilon^2\sigma_\delta^2-c^2).$$

But

$$\sigma_\varepsilon^2\sigma_\delta^2-c^2=\sigma_\varepsilon^2\sigma_\eta^2 \quad\text{and hence}\quad \sigma_\eta^2/\sigma_\omega^2=\sigma_\zeta^2/\sigma_\varepsilon^2.$$

It follows that

$$T^{GW}=T^{SW},\qquad T^{GR}=T^{SR},\qquad T^{GL}=T^{SL}.$$

Consider now $T^{(SW)}$ and $T^{(SL)}$, the probability limits of $T_n^{(SW)}/n$ and $T_n^{(SL)}/n$. It is instructive to consider successively sophisticated variants of tests in (2.7) and (2.8). Since there are fewer regressors in (1.2) than in (1.4), and fewer regressors in (1.3) than in (1.5), $\sigma_u^2 \geqq \sigma_\eta^2$ and $\sigma_v^2 \geqq \sigma_\omega^2$. Eqs. (1.2) and (1.4) coincide, and $\sigma_u^2 = \sigma_\eta^2$ if and only if $u_t$ is serially uncorrelated; similarly, $\sigma_v^2 = \sigma_\omega^2$ if and only if $v_t$ is serially uncorrelated. If $u_t$ and $v_t$ are serially uncorrelated, then a test of $F_1(L) = F_2(L)$ may be conducted in classical fashion

$$T_n^{(SW)} = n(\hat{\sigma}_u^2/\hat{\sigma}_v^2 - 1) \quad \text{and} \quad T_n^{(SL)} = n(1 - \hat{\sigma}_v^2/\hat{\sigma}_u^2),$$

$$T_n^{(SW)}/n \xrightarrow{\text{a.s.}} T^{(SW)} = \sigma_u^2/\sigma_v^2 - 1,$$

$$T^{(SL)}/n \xrightarrow{\text{a.s.}} T^{(SL)} = 1 - \sigma_v^2/\sigma_u^2.$$

In this case,

$$T^{(SW)} = T^{SW} \quad \text{and} \quad T^{(SL)} = T^{SL}.$$

Now suppose that $y_t$ is replaced by $y_t^\dagger = R(L)y_t$ and $x_t$ by $x_t^\dagger = R(L)x_t$, where $R(L)$ is an invertible one-sided filter, $R(0) = 1$. Transform all equations in sections 1 and 2 accordingly, affixing the superscript '†' to all variables and parameters. Because $R(L)$ is invertible and $R(0) = 1$, the parameters of the autoregressions (1.1), (1.4), and (2.1) and their disturbances are unaffected [e.g., $d^\dagger(L) = d(L)$, $\zeta_t = \zeta_t^\dagger$, and $\sigma_{\eta^\dagger}^2 = \sigma_\eta^2$], as is the case in (1.5). In (1.3), $f_2^\dagger(L) = f_2(L)$, but $\sigma_{v^\dagger}^2 \neq \sigma_v^2$ although $\omega_t$ is the innovation in both $v_t$ and $v_t^\dagger$. Unless $\{y_t\}$ does not cause $\{x_t\}$ the one-sided projection of $\{y_t\}$ on $\{x_t\}$ depends on $R(L)$: in (1.2), $f_1^\dagger(L) \neq f_1(L)$, $\sigma_{u^\dagger}^2 \neq \sigma_u^2$, and the innovations in $u_t^\dagger$ and $u_t$ will not be the same.

Consider a generalized least squares variant of the Wald Sims test, with $R(L)$ chosen so that $R(L)v_t$ is serially uncorrelated; then

$$T^{(SW)} = \sigma_{u^\dagger}^2/\sigma_{v^\dagger}^2 - 1 = \sigma_{u^\dagger}^2/\sigma_\omega^2 - 1.$$

Since $u_t^\dagger$ may be serially correlated, $\sigma_{u^\dagger}^2 \geqq \sigma_{\eta^\dagger}^2 = \sigma_\eta^2$. Consequently, $T^{(SW)} \geqq T^{SW}$, with equality if and only if the one-sided projection of $\{y_t^\dagger\}$ on $\{x_t^\dagger\}$ turns out to have a serially uncorrelated residual.

In a generalized least squares variant of the LaGrange-multiplier Sims test, with $R(L)$ chosen so that $R(L)u_t$ is serially uncorrelated, $T^{(SL)} = 1 - \sigma_{v^\dagger}^2/\sigma_{u^\dagger}^2$. In general, $R(L)$, $\sigma_{v^\dagger}^2$, and $\sigma_{u^\dagger}^2$ will differ from their counterparts in the Wald Sims test. Because $R(L)v_t \neq v_t^\dagger$, neither $u_t^\dagger$ nor $v_t^\dagger$ need be serially uncorrelated. The approximate slope of a Sims LaGrange-multiplier test is also sensitive to prefiltering. Treating $\{y_t\}$ and $\{x_t\}$ as the prefiltered series, we see that $f_1(L)$

and $\sigma_u^2$ vary with the prefilter, and consequently so will $\sigma_v^{2\dagger}$ and $\sigma_u^{2\dagger}$. [This problem does not arise in the Wald Sims test, in which $R(L)$ is always chosen so that the disturbance in the two-sided projection is $\omega_t$ after serial correlation correction.]

There appears to be no ordering between $T^{(SW)}$ and $T^{(SL)}$, and numerical examples in the next section provide instances in which $T^{(SL)} > T^{SW}$ and $T^{(SL)} < T^{SL}$.

In practice, $R(L)$ cannot be chosen so that $R(L)u_t$ or $R(L)v_t$ is known to be serially uncorrelated. Either $R(L)$ is chosen to be an arbitrary but reasonable approximation of the autoregressive representation of $u_t$ or $v_t$ [Sims (1972)] or more elaborate procedures are used to estimate $R(L)$ explicitly [Amemiya (1973)] or implicitly [Hannan (1963)]. In all instances the parameterization problem again arises. In the mathematical appendix of this paper we show that for sufficiently slow expansions of the relevant parameter spaces, sums of squared residuals (after correction for serial correlation) divided by sample size will converge to $\sigma_u^{2\dagger}$ and $\sigma_v^{2\dagger}$ in the restricted and unrestricted models, respectively. The approximate slopes of the tests based on feasible generalized least squares estimators are therefore the same as those based on generalized least squares estimators.

## 4. Experimental design

To supplement these analytic results, we conducted a series of Monte Carlo experiments. There are several interesting issues which might be addressed in such experiments; we have concentrated on the following:

(1) In a given population, what is the relative behavior of the tests discussed above? Are the asymptotic distributions under the null adequate? Under the alternative, do the approximate slopes of the different tests provide a reliable indication of their relative abilities to reject the null hypothesis?
(2) Under the null, or under different alternatives in which approximate slope is the same, is the performance of the tests sensitive to changes in population characteristics such as autocorrelation or the lengths of distributed lags?
(3) Under the alternative what is the effect of changing approximate slope on test performance?
(4) How are the tests affected when the assumed parameterization is not that of the population?
(5) In conducting a Sims test in which a correction for serial correlation is necessary, what is the effect of prefiltering?

It is impossible to deal with all of these issues systematically without conducting a very expensive study. In our experiments, we have examined the performance of all the tests, thus dealing thoroughly with the first question. These tests were applied in six different populations, providing some evidence on the second point. In a few cases the chosen parameterization was not compatible with the population parameterization, so limited information about point (5) may be forthcoming. All Sims tests requiring correlation for serial correlation were conducted with and without prefilters. In every case we used 100 observations.

The six models used in the experiments are summarized in table 1. The parameter $b$ was chosen in each case so that the common approximate slopes of the Wald variants of the Granger and Sims lagged dependent variable tests would be about 0.109. This value was selected to provide an interesting frequency of rejection for the hypothesis that $x$ does not cause $y$, using Wald's (1943) result on the distribution of Wald test statistics under the alternative hypothesis. In our notation, his result states that the asymptotic distribution of the test statistic under the alternative is non-central chi-square with degrees of freedom equal to the number of restricted parameters and non-centrality parameter equal to the product of the approximate slope and the number of observations. The implied rejection frequencies for the Granger and Sims lagged dependent variable Wald tests in our experiments are about 0.76 when four parameters are restricted under the null and about 0.55 when twelve parameters are restricted, if the significance level of the test is five percent. The approximate slopes of the Sims Wald and LaGrange-multiplier tests vary across the six experiments.

Table 1
Design of experiments.

Model A: $(1-0.75\,L)^2 y_t = 1.0 + bx_{t-1} + \delta_t^1$, $\delta_t^1 \sim \mathrm{IN}(0,1)$.

Model B: $y_t = 1.0 + b(3\,L + 2\,L^2 - L^3)x_t + u_t$, $(1-0.75\,L)^2 u_t = \delta_t^2$, $\delta_t^2 \sim \mathrm{IN}(0,1)$.

In both models, $(1-\theta L)x_t = 1.0 + \zeta_t$, $\zeta_t \sim \mathrm{IN}(0,1)$.

$\{\delta_t^1\}$, $\{\delta_t^2\}$, $\{\zeta_t\}$ are individually and mutually uncorrelated at all leads and lags.

| | Model A[a] | | | Model B[a] | | |
|---|---|---|---|---|---|---|
| | $\theta=0.5$ | $\theta=0.8$ | $\theta=0.99$ | $\theta=0.5$ | $\theta=0.8$ | $\theta=0.99$ |
| $b$ | 0.291 | 0.215 | 0.113 | 0.076 | 0.086 | 0.087 |
| $T^{GL}=T^{SL}$ | 0.0985 | 0.0985 | 0.0987 | 0.0978 | 0.0989 | 0.0985 |
| $T^{GR}=T^{SR}$ | 0.1037 | 0.1037 | 0.1037 | 0.1030 | 0.1041 | 0.1037 |
| $T^{GW}=G^{SW}$ | 0.1093 | 0.1093 | 0.1095 | 0.1085 | 0.1098 | 0.1093 |
| $T^{(SW)}$ | 0.1125 | 0.1240 | 0.1914 | 0.1118 | 0.1125 | 0.1155 |
| $T^{(SL)}$ | 0.1011 | 0.1103 | 0.1612 | 0.0944 | 0.0954 | 0.0976 |
| $T^{(SL)}$ (prefiltered) | 0.0713 | 0.0711 | 0.0767 | 0.1066 | 0.1016 | 0.1006 |

[a]Approximate slopes are given for respective tests of the hypothesis that $\{x_t\}$ does not cause $\{y_t\}$.

All artificial independent normal random variables were generated using the algorithm of Marsaglia [Knuth (1969)]. An autoregressive process of order $k$ was generated from a sequence of independent, identically distributed normal variables employing the unconditional distribution for the first observation, and using the appropriate different conditional distributions until the $(k-1)$th observation was generated, after which the conditional distributions remain the same. The random number generator was allowed to run continuously from one experiment to the next without restarting, so the results of the six experiments are quasi-independent. Within each experiment, the various tests and parameterizations were applied to the same data, which should allow for a more exact comparison of test behavior than would be obtained if the different tests were applied to quasi-independent data. All simulations were performed on a Univac 1110 in double precision. In order to calculate the test statistics reported in the text, 80 regressions with an average of 16 regressors were run for each replication. In every case, computations were performed using a Householder reduction technique so that the sum of squared residuals could be calculated without the need to solve for regression coefficients. In the computation of the Hannan efficient estimates, 240 calls of a fast Fourier transform routine were required each replication to transform the variables. Each replication required 75 CPU seconds.

In each of the six models the hypothesis that $\{y_t\}$ does not cause $\{x_t\}$ (which is true) and the hypothesis that $\{x_t\}$ does not cause $\{y_t\}$ (which is false) were tested. For the hypothesis that $y$ does not cause $x$ the truncated lag distribution on $x$ usually used in Sims tests requiring correction for autocorrelation of the disturbance term is exact in Model B (provided three or more lags are used) but in Model A it is not. When the (false) hypothesis that $x$ does not cause $y$ is tested using this method, the parameterization is not exact for either model. Three variants of each of Models A and B, corresponding to different degrees of serial correlation in $x$, were used. The values of the parameters of eqs. (1.1), relevant for the Granger tests, of eqs. (1.2) and (1.3), relevant for the Sims tests, and of eqs. (1.4) and (1.5), relevant for the Sims tests with lagged dependent variables, are shown in tables 2 and 3. Table 2 provides the parameters pertinent to tests of the true hypothesis $\{y_t\}$ does not cause $\{x_t\}$, and table 3 provides those for the tests of the false hypothesis that $\{x_t\}$ does not cause $\{y_t\}$. In table 3, $f_{1,s}$ is tabulated separately for the Wald and LaGrange-multiplier tests. These coefficients will differ in general because the coefficients in the one-sided projection of $\{y_t\}$ on $\{x_t\}$ depend on the filter used in forming $y_t^\dagger = R(L)y_t$ and $x_t^\dagger = R(L)x_t$. If the autocovariance function of the residual $u_t$ in the two-sided projection of $\{y_t\}$ on $\{x_t\}$ is proportional to the autocovariance function of the residual $v_t$ in the one-sided projection of $\{y_t\}$ on $\{x_t\}$, then the filters are the same and the $f_{1,s}$ coincide. That is the case of Experiment A, as may be seen by comparing $S_u(\lambda)$ and $S_v(\lambda)$.

Table 2
Parameterizations of (1.1)–(1.5) in the experiments for $H_0$: $\{y_t\}$ does not cause $\{x_t\}$.

| | Model A | | | Model B | | |
|---|---|---|---|---|---|---|
| | $\theta=0.50$ | $\theta=0.80$ | $\theta=0.99$ | $\theta=0.50$ | $\theta=0.80$ | $\theta=0.99$ |
| $c_0$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $c_1$ | −0.5000 | −0.8000 | −0.9900 | −0.5000 | −0.8000 | −0.9900 |
| $c_2$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $c_3$ | ↓ | ↓ | | ↓ | ↓ | |
| $f_{1,0}$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $f_{1,1}$ | 0.2910 | 0.2150 | 0.1130 | 0.2280 | 0.2580 | 0.2610 |
| $f_{1,2}$ | 0.4365 | 0.3225 | 0.1695 | 0.1520 | 0.1720 | 0.1740 |
| $f_{1,3}$ | 0.4911 | 0.3628 | 0.1907 | −0.0760 | −0.0860 | −0.0870 |
| $f_{1,4}$ | 0.4911 | 0.3628 | 0.1907 | 0.0000 | 0.0000 | 0.0000 |
| $f_{1,5}$ | 0.4604 | 0.3401 | 0.1788 | ↓ | ↓ | ↓ |
| $f_{1,6}$ | 0.4143 | 0.3061 | 0.1609 | | | |
| $f_{1,7}$ | 0.3625 | 0.2679 | 0.1408 | | | |
| $f_{1,8}$ | 0.3108 | 0.2296 | 0.1207 | | | |
| $f_{1,9}$ | 0.2622 | 0.1937 | 0.1018 | | | |
| $f_{1,10}$ | 0.2185 | 0.1614 | 0.0849 | | | |
| $f_{1,11}$ | 0.1803 | 0.1332 | 0.0700 | | | |
| $f_{1,12}$ | 0.1475 | 0.1090 | 0.0573 | | | |
| $g_0$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $g_1$ | −1.5000 | −1.5000 | −1.5000 | −1.5000 | −1.5000 | −1.5000 |
| $g_2$ | 0.5625 | 0.5625 | 0.5625 | 0.5625 | 0.5625 | 0.5625 |
| $g_3$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $g_4$ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ |
| $f^*_{1,0}$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $f^*_{1,1}$ | 0.2910 | 0.2150 | 0.1130 | 0.2280 | 0.2580 | 0.2610 |
| $f^*_{1,2}$ | 0.0000 | 0.0000 | 0.0000 | −0.1900 | −0.2150 | 0.2175 |
| $f^*_{1,3}$ | ↓ | ↓ | ↓ | −0.1758 | −0.1989 | −0.2012 |
| $f^*_{1,4}$ | | | | 0.1995 | 0.2258 | 0.2284 |
| $f^*_{1,5}$ | | | | −0.0428 | −0.0492 | −0.0489 |
| $f^*_{1,6}$ | | | | 0.0000 | 0.0000 | 0.0000 |
| $f^*_{1,7}$ | | | | ↓ | ↓ | ↓ |
| $S_u(0)$ | 256.0000 | 256.0000 | 256.0000 | 256.0000 | 256.0000 | 256.0000 |
| $S_u(0.25\pi)$ | 3.9707 | 3.9707 | 3.9707 | 3.9707 | 3.9707 | 3.9707 |
| $S_u(0.50\pi)$ | 0.4096 | 0.4096 | 0.4096 | 0.4096 | 0.4096 | 0.4096 |
| $S_u(0.75\pi)$ | 0.1453 | 0.1453 | 0.1453 | 0.1453 | 0.1453 | 0.1453 |
| $S_u(\pi)$ | 0.1066 | 0.1066 | 0.1066 | 0.1066 | 0.1066 | 0.1066 |

In the Granger tests, $l$ lagged values of the dependent variable and $k$ lagged values of the other variable (hypothesized not to cause the dependent variable) are used; in our experiments we examined the combinations $k=4$ and $l=4$, 8, and 12, respectively, as well as the combination $k=l=12$. From table 2 it may be seen that this parameterization is always exact for tests that $y$ does not cause $x$, but for tests that $x$ does not cause $y$, $k=4$ does not allow

Table 3
Parameterizations of (1.1)–(1.5) in the experiments for $H_0$: $\{x_t\}$ does not cause $\{y_t\}$.

| | Model A | | | Model B | | |
|---|---|---|---|---|---|---|
| | $\theta=0.50$ | $\theta=0.80$ | $\theta=0.99$ | $\theta=0.50$ | $\theta=0.80$ | $\theta=0.99$ |
| $c_0$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $c_1$ | −1.5493 | −1.5788 | −1.5977 | −1.4815 | −1.4884 | −1.5047 |
| $c_2$ | 0.6142 | 0.6239 | 0.6219 | 0.5855 | 0.5935 | 0.5973 |
| $c_3$ | −0.0044 | −0.0001 | −0.0020 | −0.0825 | −0.0790 | −0.0691 |
| $c_4$ | −0.0020 | −0.0001 | −0.0018 | 0.0550 | 0.0487 | 0.0410 |
| $c_5$ | −0.0009 | —[a] | −0.0016 | −0.0135 | −0.0116 | −0.0102 |
| $c_6$ | −0.0004 | —[a] | −0.0014 | 0.0040 | 0.0033 | 0.0018 |
| $c_7$ | −0.0001 | —[a] | −0.0013 | −0.0016 | −0.0012 | −0.0015 |
| $c_8$ | −0.0001 | —[a] | −0.0011 | 0.0004 | 0.0003 | −0.0004 |
| $d_0$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $d_1$ | −1.5000 | −1.5000 | −1.5000 | −1.5000 | −1.5000 | −1.5000 |
| $d_2$ | 0.5625 | 0.5625 | 0.5625 | 0.5625 | 0.5625 | 0.5625 |
| $d_3$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $d_4$ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ |
| $e_1$ | 0.2910 | 0.2150 | 0.1130 | 0.2280 | 0.2580 | 0.2610 |
| $e_2$ | 0.0000 | 0.0000 | 0.0000 | −0.1900 | −0.2150 | −0.2175 |
| $e_3$ | ↓ | ↓ | ↓ | −0.1757 | −0.1989 | −0.2012 |
| $e_4$ | | | | 0.1995 | 0.2258 | 0.2284 |
| $e_5$ | | | | −0.0427 | −0.0484 | −0.0489 |
| $e_6$ | | | | 0.0000 | 0.0000 | 0.0000 |
| $e_7$ | | | | ↓ | ↓ | ↓ |
| Wald: | | | | | | |
| $f_{1,0}$ | 0.0167 | 0.0361 | 0.0853 | −0.0039 | −0.0049 | 0.0034 |
| $f_{1,1}$ | −0.0100 | −0.0021 | 0.0243 | −0.0079 | −0.0080 | −0.0026 |
| $f_{1,2}$ | −0.0030 | −0.0015 | 0.0234 | 0.0145 | 0.0105 | 0.0119 |
| $f_{1,3}$ | −0.0007 | −0.0010 | 0.0224 | −0.0046 | −0.0027 | 0.0023 |
| $f_{1,4}$ | —[a] | −0.0007 | 0.0214 | 0.0014 | 0.0012 | 0.0054 |
| $f_{1,5}$ | 0.0001 | −0.0005 | 0.0203 | −0.0011 | −0.0005 | 0.0047 |
| $f_{1,6}$ | 0.0001 | −0.0004 | 0.0192 | 0.0002 | 0.0001 | 0.0054 |
| LaGrange: | | | | | | |
| $f_{1,0}$ | 0.0167 | 0.0361 | 0.0853 | 0.0064 | 0.0106 | 0.0448 |
| $f_{1,1}$ | −0.0100 | −0.0021 | 0.0243 | −0.0074 | −0.0085 | −0.0193 |
| $f_{1,2}$ | −0.0030 | −0.0015 | 0.0234 | 0.0014 | −0.0036 | −0.0068 |
| $f_{1,3}$ | −0.0007 | −0.0010 | 0.0224 | 0.0037 | 0.0041 | 0.0182 |
| $f_{1,4}$ | —[a] | −0.0007 | 0.0214 | −0.0011 | −0.0011 | 0.0026 |
| $f_{1,5}$ | 0.0001 | −0.0005 | 0.0203 | 0.0004 | 0.0005 | 0.0076 |
| $f_{1,6}$ | 0.0001 | −0.0004 | 0.0192 | −0.0003 | −0.0002 | 0.0060 |
| $f_{2,-12}$ | —[a] | 0.0023 | 0.0156 | —[a] | 0.0001 | 0.0215 |
| $f_{2,-11}$ | —[a] | 0.0032 | 0.0175 | —[a] | 0.0001 | 0.0220 |
| $f_{2,-10}$ | 0.0001 | 0.0045 | 0.0196 | 0.0001 | 0.0001 | 0.0226 |
| $f_{2,-9}$ | 0.0002 | 0.0062 | 0.0220 | −0.0001 | —[a] | 0.0230 |
| $f_{2,-8}$ | 0.0005 | 0.0086 | 0.0246 | 0.0009 | 0.0005 | 0.0239 |
| $f_{2,-7}$ | 0.0012 | 0.0120 | 0.0276 | −0.0014 | −0.0011 | 0.0234 |
| $f_{2,-6}$ | 0.0027 | 0.0166 | 0.0309 | 0.0092 | 0.0056 | 0.0287 |

Table 3 (continued)

| | Model A | | | Model B | | |
|---|---|---|---|---|---|---|
| | $\theta=0.50$ | $\theta=0.80$ | $\theta=0.99$ | $\theta=0.50$ | $\theta=0.80$ | $\theta=0.99$ |
| $f_{2,-5}$ | 0.0060 | 0.0230 | 0.0347 | −0.0121 | −0.0109 | 0.0173 |
| $f_{2,-4}$ | 0.0132 | 0.0319 | 0.0388 | 0.0775 | 0.0548 | 0.0715 |
| $f_{2,-3}$ | 0.0293 | 0.0443 | 0.0435 | −0.2560 | −0.2089 | −0.1507 |
| $f_{2,-2}$ | 0.0650 | 0.0614 | 0.0488 | 0.1229 | 0.1174 | 0.1296 |
| $f_{2,-1}$ | 0.1443 | 0.0851 | 0.0547 | 0.2492 | 0.2262 | 0.2230 |
| $f_{2,0}$ | −0.2620 | −0.1507 | −0.0529 | −0.1988 | −0.1518 | −0.1002 |
| $f_{2,1}$ | 0.0295 | 0.0003 | 0.0101 | 0.0153 | 0.0079 | 0.0324 |
| $f_{2,2}$ | 0.0133 | 0.0002 | 0.0090 | −0.0101 | −0.0101 | 0.0185 |
| $f_{2,3}$ | 0.0060 | 0.0002 | 0.0081 | 0.0063 | 0.0026 | 0.0271 |
| $f_{2,4}$ | 0.0027 | 0.0001 | 0.0072 | 0.0004 | −0.0004 | 0.0245 |
| $f_{2,5}$ | 0.0012 | 0.0001 | 0.0064 | 0.0011 | 0.0004 | 0.0244 |
| $f_{2,6}$ | 0.0005 | 0.0001 | 0.0057 | 0.0002 | 0.0001 | 0.0236 |
| $f_{2,7}$ | 0.0002 | 0.0000 | 0.0051 | 0.0002 | 0.0001 | 0.0231 |
| $f_{2,8}$ | 0.0001 | ↓ | 0.0046 | 0.0001 | 0.0001 | 0.0226 |
| $f_{2,9}$ | 0.0001 | | 0.0041 | —[a] | 0.0001 | 0.0220 |
| $f_{2,10}$ | —[a] | | 0.0036 | —[a] | —[a] | 0.0215 |
| $f_{2,11}$ | —[a] | | 0.0032 | —[a] | —[a] | 0.0210 |
| $f_{2,12}$ | —[a] | | 0.0029 | —[a] | —[a] | 0.0205 |
| $g_0$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $g_1$ | −0.5000 | −0.8000 | −0.9900 | −0.5000 | −0.8000 | −0.9900 |
| $g_2$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $g_3$ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ |
| $f^*_{1,0}$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $f^*_{1,1}$ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ |
| $h_0$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $h_1$ | −0.4507 | −0.7212 | −0.8923 | −0.5185 | −0.8116 | −0.9853 |
| $h_2$ | 0.0000 | 0.0000 | 0.0000 | −0.0412 | −0.0389 | −0.0324 |
| $h_3$ | ↓ | ↓ | ↓ | 0.0438 | 0.0527 | 0.0520 |
| $h_4$ | | | | −0.0088 | −0.0112 | −0.0115 |
| $h_5$ | | | | 0.0000 | 0.0000 | 0.0000 |
| $h_6$ | | | | ↓ | ↓ | ↓ |
| $f^*_{2,-12}$ | —[a] | 0.0011 | 0.0032 | —[a] | —[a] | 0.0010 |
| $f^*_{2,-11}$ | —[a] | 0.0016 | 0.0036 | —[a] | —[a] | 0.0010 |
| $f^*_{2,-10}$ | 0.0001 | 0.0022 | 0.0040 | 0.0001 | 0.0001 | 0.0010 |
| $f^*_{2,-9}$ | 0.0002 | 0.0030 | 0.0045 | −0.0002 | −0.0001 | 0.0009 |
| $f^*_{2,-8}$ | 0.0004 | 0.0041 | 0.0050 | 0.0009 | 0.0006 | 0.0014 |
| $f^*_{2,-7}$ | 0.0010 | 0.0057 | 0.0056 | −0.0018 | −0.0015 | —[a] |
| $f^*_{2,-6}$ | 0.0021 | 0.0080 | 0.0063 | 0.0099 | 0.0065 | 0.0058 |
| $f^*_{2,-5}$ | 0.0047 | 0.0111 | 0.0071 | −0.0168 | −0.0154 | −0.0107 |
| $f^*_{2,-4}$ | 0.0105 | 0.0153 | 0.0079 | 0.0833 | 0.0634 | −0.0545 |
| $f^*_{2,-3}$ | 0.0234 | 0.0212 | 0.0089 | −0.2953 | −0.2527 | −0.2205 |
| $f^*_{2,-2}$ | 0.0518 | 0.0295 | 0.0099 | 0.2518 | 0.2842 | 0.2763 |
| $f^*_{2,-1}$ | 0.1149 | 0.0409 | 0.0111 | 0.1995 | 0.1420 | 0.1037 |
| $f^*_{2,0}$ | −0.3270 | −0.2121 | −0.1017 | −0.3450 | −0.3516 | −0.3328 |
| $f^*_{2,1}$ | 0.1476 | 0.1090 | 0.0573 | 0.1157 | 0.1308 | 0.1324 |
| $f^*_{2,2}$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $f^*_{2,3}$ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ |

Table 3 (continued)

| | Model A | | | Model B | | |
|---|---|---|---|---|---|---|
| | $\theta=0.50$ | $\theta=0.80$ | $\theta=0.99$ | $\theta=0.50$ | $\theta=0.80$ | $\theta=0.99$ |
| $S_u(0)$ | 3.4096 | 14.5931 | 155.7753 | 3.9988 | 24.7180 | 2115.1299 |
| $S_u(0.25\pi)$ | 1.8183 | 2.2678 | 3.2567 | 1.7943 | 1.8741 | 1.9724 |
| $S_u(0.50\pi)$ | 0.8550 | 0.7462 | 0.9711 | 0.7884 | 0.6108 | 0.7255 |
| $S_u(0.75\pi)$ | 0.5589 | 0.4466 | 0.5716 | 0.5407 | 0.3968 | 0.5646 |
| $S_u(\pi)$ | 0.4888 | 0.3829 | 0.4886 | 0.4467 | 0.3130 | 0.3679 |
| $S_v(0)$ | 2.9879 | 11.5976 | 77.7061 | 3.9942 | 24.7144 | 1745.0116 |
| $S_v(0.25\pi)$ | 1.5934 | 1.8023 | 1.6869 | 1.7498 | 1.8341 | 1.6194 |
| $S_v(0.50\pi)$ | 0.7492 | 0.5930 | 0.5018 | 0.6527 | 0.4997 | 0.4256 |
| $S_v(0.75\pi)$ | 0.4898 | 0.3549 | 0.2947 | 0.4369 | 0.3129 | 0.2621 |
| $S_v(\pi)$ | 0.4283 | 0.3043 | 0.2517 | 0.4444 | 0.3086 | 0.2525 |

[a]Denotes coefficient less than 0.00005 in absolute value.

a complete parameterization of the effect of $x$ on $y$. In the Sims tests based on (1.2) and (1.3) $p$ leading and $q$ lagged values of the explanatory variable (hypothesized not to be caused by the dependent variable) are used; we examined the combinations $p=q=4$, $p=4$ and $q=8$, $p=4$ and $q=12$, and $p=q=12$. In the Sims tests based on (1.4) and (1.5) employing lagged dependent variables, four lagged values of the dependent variable were always used and combinations of leading and lagged values of the explanatory variable were the same as those in the tests based on (1.2) and (1.3). For tests that $y$ does not cause $x$ these parameterizations are always exact, and for tests that $x$ does not cause $y$, never.

Two techniques were used to correct for serial correlation of the disturbance in the first group of Sims tests. In both procedures, ordinary least squares residuals are used to estimate the variance matrix of the regression equation disturbances. In the Hannan efficient method [Hannan (1963)] the spectral density of the disturbances was estimated using an inverted '$V$' spectral window with a base of 19 ordinates, applied to the 100 residual periodogram ordinates. The estimated spectral density was then used to transform the data prior to a final estimation by ordinary least squares in the way discussed in Geweke (1978). In Amemiya's (1973) method an autoregression of ordinary least squares residuals of length 8 was estimated, and then used to transform the data prior to final estimation by ordinary least squares. It can be seen from table 1 that the parameterization implicit in Amemiya's method is always exact in tests of the hypothesis that $\{y_t\}$ does not cause $\{x_t\}$. For tests of $\{x_t\}$ not cause $\{y_t\}$, it is exact for Experiment $A$ ($\{u_t\}$ and $\{v_t\}$ are both first-order autoregressive) and for Experiment B it is inexact, but coefficients beyond the eighth in the autoregression are negligible. The parameterization in Hannan's method is never exact unless the disturbances are serially uncorrelated. It is also less flexible than the one for Amemiya's method, employing a spectral estimator with only about five effective degrees of freedom. The Hannan efficient estimator is especially ill-suited to cope with the large variation in $S_u(\lambda)$ and $S_v(\lambda)$ near zero.

Both methods were also used after application of the prefilter $(1-0.75L)^2$ to the data. For tests of the hypothesis that $\{y_t\}$ does not cause $\{x_t\}$ this renders the disturbance term serially uncorrelated in all models. For tests of $\{x_t\}$ not to cause $\{y_t\}$ the parameterization of Amemiya's test is then inexact in both experiments, but is a very good approximation. For those tests the prefilter removes the peaks in $S_u(\lambda)$ and $S_v(\lambda)$ at $\lambda=0$, whence the parameterization in the Hannan efficient estimator should be more suitable.

## 5. The outcome of the experiments

The results of our experiments are summarized in tables 4 through 7. The first two of these tables provide information about the distribution over the 100 replications of the various test statistics for the hypothesis that $\{y_t\}$ does not cause $\{x_t\}$, which is true. Tables 6 and 7 provide similar information for tests of the hypothesis that $\{x_t\}$ does not cause $\{y_t\}$. We shall discuss the performance of the tests in these two situations in turn.

Table 4 indicates the mean and standard deviation of the '$F$' statistic over the 100 replications for the Wald and LaGrange variants of the Granger, Sims (using both Hannan's and Amemiya's correction for serial correlation), and Sims lagged dependent variable tests. Table 5 reports the number of replications in which the hypothesis that $\{y_t\}$ does not cause $\{x_t\}$ was rejected at the 5% and 10% significance levels, respectively. If the asymptotic theory were exact, then with probability 0.95 the number of replications in which rejection occurs would lie between 1 and 9 for the 5% level and between 5 and 15 for the 10% level.

From these tables it can be seen that the experiments yield considerable information about differences in outcomes across tests, but much less about differences in outcomes across models; since the experiments were designed to do the former and not the latter, this is not surprising. Tests which require correction for serial correlation in general do not behave well under the null hypothesis. All Wald tests that involve a serial correlation correction yield statistics that are too large and reject the null hypothesis too often. As expected, prefiltering improves the performance of these tests when Hannan's method is used, but has little effect for Amemiya's procedure. The distributions of the LaGrange-multiplier test statistics are more reasonable. For the Hannan efficient estimator prefiltering is required for an adequate distribution, but again seems to have few consequences for tests conducted using Amemiya's procedure. Except for the Hannan efficient method without prefiltering, the distribution of the LaGrange-multiplier test statistics is close to that of the asymptotic distribution, rejection frequencies for some

Table 4

Mean values of $F$'s under null hypothesis.[a]

| Model and parameterization | Asymptotic $F$ distribution | | Granger tests | | Sims tests | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Lagged dependent variables | | Hannan efficient | | | | Amemiya GLS | | | |
| | | | | | | | Filtered | | Unfiltered | | Filtered | | Unfiltered | |
| | Granger | Sims | Wald | LaGrange | Wald | LaGrange | Wald | LaGrange | Wald | LaGrange | Wald | LaGrange | Wald | LaGrange |
| $A, \theta = 0.5$ | | | | | | | | | | | | | | |
| 4–4 | $F(4,91)$ | $F(4,90)$ | 1.11 (0.08) | 1.03 (0.07) | 1.20 (0.08) | 1.11 (0.07) | 1.22 (0.09) | 1.07 (0.07) | 2.25 (0.27) | 1.82 (0.15) | 1.92 (0.14) | 1.43 (0.09) | 1.73 (0.15) | 1.39 (0.12) |
| 4–8 | $F(4,87)$ | $F(4,86)$ | 1.12 (0.08) | 1.04 (0.07) | 1.22 (0.08) | 1.12 (0.07) | 1.11 (0.09) | 0.97 (0.07) | 2.15 (0.24) | 1.76 (0.15) | 1.37 (0.11) | 1.05 (0.07) | 1.68 (0.13) | 1.36 (0.09) |
| 4–12 | $F(4,83)$ | $F(4,82)$ | 1.15 (0.09) | 1.06 (0.07) | 1.24 (0.09) | 1.13 (0.07) | 1.16 (0.10) | 1.01 (0.08) | 1.87 (0.18) | 1.57 (0.12) | 1.48 (0.13) | 1.14 (0.08) | 1.69 (0.12) | 1.34 (0.09) |
| 12–12 | $F(12,75)$ | $F(12,74)$ | 1.06 (0.05) | 0.89 (0.04) | 1.19 (0.06) | 1.06 (0.04) | 1.12 (0.06) | 0.86 (0.03) | 1.55 (0.12) | 1.11 (0.05) | 1.47 (0.07) | 0.96 (0.03) | 1.57 (0.10) | 1.01 (0.04) |
| $B, \theta = 0.5$ | | | | | | | | | | | | | | |
| 4–4 | $F(4,91)$ | $F(4,90)$ | 1.00 (0.07) | 0.94 (0.06) | 0.96 (0.96) | 0.90 (0.06) | 1.02 (0.07) | 0.92 (0.06) | 1.61 (0.13) | 1.49 (0.11) | 1.26 (0.10) | 1.01 (0.07) | 1.31 (0.10) | 1.05 (0.07) |
| 4–8 | $F(4,87)$ | $F(4,86)$ | 0.98 (0.07) | 0.92 (0.06) | 0.95 (0.06) | 0.88 (0.05) | 1.04 (0.08) | 0.92 (0.06) | 1.40 (0.12) | 1.29 (0.09) | 1.19 (0.10) | 0.95 (0.07) | 1.26 (0.09) | 1.03 (0.07) |
| 4–12 | $F(4,83)$ | $F(4,82)$ | 0.98 (0.07) | 0.92 (0.06) | 0.97 (0.05) | 0.90 (0.05) | 1.06 (0.08) | 0.94 (0.06) | 1.25 (0.08) | 1.13 (0.07) | 1.21 (0.10) | 0.95 (0.07) | 1.25 (0.09) | 1.00 (0.06) |
| 12–12 | $F(12,75)$ | $F(12,74)$ | 0.95 (0.04) | 0.81 (0.03) | 1.01 (0.04) | 0.94 (0.03) | 1.07 (0.05) | 0.83 (0.03) | 1.32 (0.08) | 0.98 (0.04) | 1.28 (0.07) | 0.87 (0.03) | 1.37 (0.08) | 0.89 (0.03) |
| $A, \theta = 0.8$ | | | | | | | | | | | | | | |
| 4–4 | $F(4,91)$ | $F(4,90)$ | 1.11 (0.08) | 1.04 (0.07) | 1.09 (0.07) | 1.01 (0.06) | 1.17 (0.08) | 1.04 (0.07) | 2.35 (0.29) | 1.89 (0.17) | 1.63 (0.15) | 1.23 (0.09) | 1.45 (0.14) | 1.13 (0.09) |
| 4–8 | $F(4,87)$ | $F(4,86)$ | 1.09 (0.08) | 1.01 (0.07) | 1.08 (0.08) | 1.00 (0.07) | 1.18 (0.09) | 1.03 (0.07) | 1.91 (0.17) | 1.65 (0.13) | 1.33 (0.10) | 1.04 (0.07) | 1.38 (0.11) | 1.14 (0.09) |
| 4–12 | $F(4,83)$ | $F(4,82)$ | 1.14 (0.06) | 1.06 (0.07) | 1.13 (0.08) | 1.03 (0.07) | 1.22 (0.10) | 1.05 (0.08) | 1.70 (0.15) | 1.48 (0.11) | 1.40 (0.11) | 1.09 (0.08) | 1.46 (0.11) | 1.21 (0.08) |
| 12–12 | $F(12,75)$ | $F(12,74)$ | 1.09 (0.04) | 0.91 (0.03) | 1.10 (0.04) | 1.00 (0.03) | 1.20 (0.07) | 0.92 (0.04) | 1.46 (0.09) | 1.11 (0.05) | 1.41 (0.06) | 0.95 (0.03) | 1.44 (0.07) | 0.99 (0.03) |
| $B, \theta = 0.8$ | | | | | | | | | | | | | | |
| 4–4 | $F(4,91)$ | $F(4,90)$ | 1.19 (0.07) | 1.11 (0.06) | 1.09 (0.07) | 1.01 (0.06) | 1.21 (0.09) | 1.06 (0.07) | 1.96 (0.19) | 1.72 (0.14) | 1.48 (0.12) | 1.11 (0.07) | 1.49 (0.11) | 1.17 (0.08) |
| 4–8 | $F(4,87)$ | $F(4,86)$ | 1.19 (0.08) | 1.10 (0.07) | 1.05 (0.07) | 0.97 (0.06) | 1.22 (0.08) | 1.07 (0.07) | 1.69 (0.16) | 1.49 (0.12) | 1.37 (0.10) | 1.04 (0.07) | 1.42 (0.10) | 1.15 (0.07) |
| 4–12 | $F(4,83)$ | $F(4,82)$ | 1.17 (0.07) | 1.08 (0.06) | 1.07 (0.07) | 0.99 (0.06) | 1.27 (0.08) | 1.11 (0.07) | 1.46 (0.13) | 1.31 (0.10) | 1.46 (0.11) | 1.12 (0.07) | 1.45 (0.11) | 1.15 (0.08) |
| 12–12 | $F(12,75)$ | $F(12,74)$ | 1.15 (0.05) | 0.96 (0.03) | 1.06 (0.05) | 0.97 (0.03) | 1.27 (0.06) | 0.98 (0.04) | 1.52 (0.11) | 1.14 (0.05) | 1.50 (0.06) | 1.01 (0.03) | 1.49 (0.07) | 1.01 (0.04) |
| $A, \theta = 0.99$ | | | | | | | | | | | | | | |
| 4–4 | $F(4,91)$ | $F(4,90)$ | 1.26 (0.10) | 1.16 (0.08) | 1.17 (0.09) | 1.07 (0.08) | 1.57 (0.15) | 1.32 (0.11) | 1.47 (0.16) | 1.29 (0.12) | 1.61 (0.13) | 1.15 (0.09) | 1.68 (0.15) | 1.32 (0.10) |
| 4–8 | $F(4,87)$ | $F(4,86)$ | 1.26 (0.09) | 1.16 (0.08) | 1.18 (0.09) | 1.08 (0.07) | 1.43 (0.12) | 1.22 (0.10) | 1.41 (0.14) | 1.24 (0.11) | 1.52 (0.13) | 1.16 (0.09) | 1.60 (0.13) | 1.30 (0.10) |
| 4–12 | $F(4,83)$ | $F(4,82)$ | 1.27 (0.09) | 1.17 (0.08) | 1.19 (0.09) | 1.08 (0.08) | 1.26 (0.10) | 1.07 (0.08) | 1.46 (0.14) | 1.28 (0.11) | 1.51 (0.13) | 1.13 (0.09) | 1.59 (0.13) | 1.32 (0.10) |
| 12–12 | $F(12,75)$ | $F(12,74)$ | 1.14 (0.05) | 0.94 (0.04) | 1.31 (0.05) | 1.16 (0.04) | 1.30 (0.06) | 0.97 (0.04) | 1.65 (0.14) | 1.16 (0.06) | 1.68 (0.08) | 1.05 (0.04) | 1.90 (0.11) | 1.19 (0.04) |
| $B, \theta = 0.99$ | | | | | | | | | | | | | | |
| 4–4 | $F(4,91)$ | $F(4,90)$ | 1.08 (0.07) | 1.01 (0.07) | 1.05 (0.07) | 0.98 (0.06) | 1.96 (0.21) | 1.63 (0.15) | 1.71 (0.14) | 1.52 (0.12) | 1.95 (0.19) | 1.35 (0.10) | 1.18 (0.08) | 1.10 (0.07) |
| 4–8 | $F(4,87)$ | $F(4,86)$ | 1.10 (0.07) | 1.02 (0.07) | 1.09 (0.08) | 1.00 (0.07) | 1.25 (0.09) | 1.11 (0.07) | 1.63 (0.13) | 1.45 (0.11) | 1.48 (0.10) | 1.18 (0.08) | 1.26 (0.09) | 1.16 (0.08) |
| 4–12 | $F(4,83)$ | $F(4,82)$ | 1.10 (0.07) | 1.02 (0.07) | 1.12 (0.08) | 1.03 (0.07) | 1.22 (0.09) | 1.08 (0.07) | 1.53 (0.13) | 1.36 (0.11) | 1.41 (0.10) | 1.10 (0.07) | 1.27 (0.09) | 1.16 (0.08) |
| 12–12 | $F(12,75)$ | $F(12,74)$ | 1.09 (0.05) | 0.91 (0.03) | 1.17 (0.06) | 1.05 (0.04) | 1.25 (0.06) | 0.96 (0.03) | 1.44 (0.08) | 1.11 (0.05) | 1.59 (0.10) | 1.02 (0.04) | 1.30 (0.06) | 1.04 (0.04) |

[a]Standard errors of estimated means are shown parenthetically.

Table 5

Numbers of $F$'s in 5% and 10% critical regions under null hypothesis, in 100 replications.

| Model and parameterization | Asymptotic $F$ distribution | | Granger tests | | Sims tests | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Lagged dependent variables | | Hannan efficient | | | | Amemiya GLS | | | |
| | | | | | | | Filtered | | Unfiltered | | Filtered | | Unfiltered | |
| | Granger | Sims | Wald | LaGrange | Wald | LaGrange | Wald | LaGrange | Wald | LaGrange | Wald | LaGrange | Wald | LaGrange |
| $A, \theta=0.5$ | | | | | | | | | | | | | | |
| 4–4 | $F(4,91)$ | $F(4,90)$ | 7, 14 | 4, 11 | 5, 14 | 5, 9 | 10, 21 | 6, 14 | 25, 37 | 20, 32 | 29, 38 | 14, 23 | 28, 33 | 12, 25 |
| 4–8 | $F(4,87)$ | $F(4,86)$ | 7, 15 | 3, 11 | 7, 12 | 6, 8 | 7, 12 | 5, 7 | 24, 31 | 20, 24 | 12, 19 | 4, 11 | 24, 32 | 11, 20 |
| 4–12 | $F(4,83)$ | $F(4,82)$ | 10, 15 | 5, 12 | 9, 14 | 5, 10 | 9, 14 | 5, 9 | 21, 29 | 18, 27 | 15, 25 | 6, 15 | 19, 33 | 10, 15 |
| 12–12 | $F(12,75)$ | $F(12,74)$ | 8, 11 | 2, 3 | 7, 16 | 5, 7 | 10, 13 | 1, 6 | 27, 31 | 12, 15 | 24, 31 | 0, 5 | 24, 37 | 3, 6 |
| $B, \theta=0.5$ | | | | | | | | | | | | | | |
| 4–4 | $F(4,91)$ | $F(4,90)$ | 3, 8 | 2, 5 | 4, 7 | 2, 6 | 5, 11 | 2, 7 | 16, 29 | 14, 24 | 15, 19 | 5, 16 | 11, 17 | 5, 9 |
| 4–8 | $F(4,87)$ | $F(4,86)$ | 3, 4 | 2, 4 | 2, 6 | 1, 3 | 4, 8 | 4, 5 | 10, 20 | 10, 18 | 11, 17 | 7, 10 | 10, 16 | 6, 9 |
| 4–12 | $F(4,83)$ | $F(4,82)$ | 2, 4 | 2, 2 | 1, 7 | 1, 3 | 4, 8 | 4, 4 | 10, 16 | 9, 11 | 11, 14 | 4, 9 | 8, 14 | 3, 5 |
| 12–12 | $F(12,75)$ | $F(12,74)$ | 1, 5 | 1, 1 | 2, 9 | 0, 1 | 6, 8 | 0, 0 | 16, 22 | 5, 9 | 12, 17 | 0, 3 | 17, 22 | 0, 2 |
| $A, \theta=0.8$ | | | | | | | | | | | | | | |
| 4–4 | $F(4,91)$ | $F(4,90)$ | 7, 14 | 4, 11 | 6, 10 | 4, 9 | 6, 15 | 2, 8 | 26, 33 | 23, 30 | 18, 25 | 9, 15 | 15, 20 | 7, 14 |
| 4–8 | $F(4,87)$ | $F(4,86)$ | 5, 19 | 3, 11 | 7, 13 | 3, 11 | 7, 12 | 5, 7 | 24, 32 | 20, 26 | 13, 18 | 7, 13 | 16, 23 | 8, 18 |
| 4–12 | $F(4,83)$ | $F(4,82)$ | 9, 16 | 5, 13 | 9, 13 | 4, 12 | 8, 17 | 4, 9 | 21, 28 | 18, 27 | 16, 22 | 7, 15 | 18, 23 | 9, 17 |
| 12–12 | $F(12,75)$ | $F(12,74)$ | 8, 11 | 0, 2 | 6, 8 | 0, 6 | 9, 15 | 2, 4 | 26, 31 | 9, 17 | 17, 25 | 0, 3 | 23, 30 | 0, 3 |
| $B, \theta=0.8$ | | | | | | | | | | | | | | |
| 4–4 | $F(4,91)$ | $F(4,90)$ | 7, 16 | 4, 7 | 3, 7 | 1, 3 | 12, 14 | 4, 6 | 32, 38 | 18, 26 | 1, 4 | 7, 12 | 8, 18 | 6, 8 |
| 4–8 | $F(4,87)$ | $F(4,86)$ | 11, 19 | 5, 10 | 4, 6 | 1, 4 | 12, 21 | 3, 9 | 32, 41 | 9, 21 | 2, 5 | 3, 9 | 1, 6 | 5, 12 |
| 4–12 | $F(4,83)$ | $F(4,82)$ | 11, 22 | 4, 9 | 3, 8 | 3, 4 | 15, 24 | 4, 12 | 31, 38 | 8, 13 | 3, 6 | 5, 12 | 2, 6 | 5, 11 |
| 12–12 | $F(12,75)$ | $F(12,74)$ | 22, 36 | 1, 2 | 3, 5 | 1, 3 | 26, 35 | 2, 5 | 30, 34 | 8, 15 | 0, 4 | 0, 3 | 0, 5 | 1, 4 |
| $A, \theta=0.99$ | | | | | | | | | | | | | | |
| 4–4 | $F(4,91)$ | $F(4,90)$ | 10, 18 | 8, 14 | 7, 17 | 6, 10 | 16, 23 | 11, 18 | 16, 23 | 13, 19 | 19, 29 | 7, 12 | 17, 25 | 5, 18 |
| 4–8 | $F(4,87)$ | $F(4,86)$ | 9, 14 | 7, 12 | 8, 18 | 5, 10 | 16, 24 | 11, 17 | 13, 20 | 10, 18 | 15, 23 | 6, 15 | 17, 28 | 11, 20 |
| 4–12 | $F(4,83)$ | $F(4,82)$ | 11, 16 | 7, 11 | 8, 14 | 5, 11 | 11, 19 | 6, 11 | 16, 25 | 10, 19 | 16, 25 | 7, 17 | 19, 30 | 13, 19 |
| 12–12 | $F(12,75)$ | $F(12,74)$ | 7, 17 | 1, 2 | 15, 23 | 6, 13 | 14, 23 | 1, 8 | 32, 40 | 9, 17 | 34, 38 | 2, 4 | 41, 52 | 6, 14 |
| $B, \theta=0.99$ | | | | | | | | | | | | | | |
| 4–4 | $F(4,91)$ | $F(4,90)$ | 2, 14 | 2, 8 | 2, 11 | 2, 7 | 27, 31 | 24, 27 | 23, 28 | 17, 24 | 25, 31 | 17, 22 | 12, 21 | 5, 13 |
| 4–8 | $F(4,87)$ | $F(4,86)$ | 5, 15 | 3, 8 | 8, 13 | 5, 12 | 10, 16 | 7, 13 | 22, 29 | 14, 24 | 20, 29 | 7, 17 | 15, 21 | 8, 15 |
| 4–12 | $F(4,83)$ | $F(4,82)$ | 7, 16 | 2, 11 | 7, 13 | 5, 11 | 10, 14 | 6, 11 | 17, 25 | 11, 22 | 16, 22 | 5, 13 | 10, 23 | 6, 12 |
| 12–12 | $F(12,75)$ | $F(12,74)$ | 7, 13 | 0, 3 | 9, 19 | 3, 8 | 13, 21 | 5, 5 | 23, 34 | 9, 14 | 28, 38 | 2, 5 | 29, 39 | 3, 6 |

experiments being too large and for others too small but with no apparent pattern.

Overall, the results of the experiments reflect unfavorably on the performance under the null hypothesis of tests requiring correction for serial correlation. The Wald tests reject too frequently and are sensitive to prefiltering. LaGrange-multiplier tests are also sensitive to prefiltering. It is discouraging to note that the LaGrange-multiplier tests did not perform very well unless the parameterization of the distribution of the disturbance was exact (i.e., Amemiya's method) or the prefilter was used, in which case the disturbance is then serially uncorrelated before correction for serial correlation. If one of these conditions is indeed required for an adequate distribution of the test statistic then these methods are useless in applied work since then neither the serial correlation nor the functional form of its parameterization is known *a priori*. On the other hand, it should be noted that none of these tests are based on the exact maximum likelihood estimates presumed in the discussion in section 2. Hannan's and Amemiya's procedures in the Wald tests each constitute the first step in an iterative scheme leading to maximum likelihood estimates [Oberhofer and Kmenta (1974)], and in the sample size of 100 used in these experiments this first step may not be a very close approximation to exact maximum likelihood. This question might merit further investigation.

By comparison, the behavior of the Granger tests and Sims tests incorporating lagged dependent variables is excellent. Rejection frequencies for the Granger tests lie outside the 95% confidence intervals (constructed under the assumption that the asymptotic distribution theory is exact) in 25 of 96 cases, and for the Sims lagged dependent variable tests in 15 of 96 cases. There is some tendency for the LaGrange-multiplier to be more adequate than the Wald variant in the case of the Granger tests, and conversely for the Sims lagged dependent variable tests. In both instances rejection frequencies run a little too high for Wald tests and too low for LaGrange-multiplier tests. The distribution of these statistics is as good as or better than the distribution of the LaGrange-multiplier statistics after prefiltering in the tests requiring a serial correlation correction. These test statistics have the further advantages that they are cheaper to compute and their interpretation is unclouded by the prefiltering problem. However, there is an analog of the problem of parameterizing the serial correlation of the disturbance in choosing the number of lagged values of the dependent variable to be used. There is no indication that test results are sensitive to this choice as was the case in the tests requiring a serial correlation correction, even though in some cases the parameterization chosen was not the correct one (as may be seen by consulting table 2).

The distributions of the test statistics for the hypothesis that $\{x_t\}$ does not cause $\{y_t\}$ are summarized in tables 6 and 7. Given the results presented in tables 4 and 5, the behavior of the Granger and Sims lagged dependent

Table 6

Mean values of $F$'s under alternative hypotheses.[a]

| Model and parameterization | Asymptotic $F_h$ distribution[b] | | Wald tests | | | LaGrange tests | | Sims tests — Filtered Hannan efficient | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Wald | | LaGrange |
| | Granger | Sims | Granger | Sims (LDV) | Population[c] | Granger | Sims | Sample | Population | Sample |
| $A, \theta=0.5$ | | | | Slope = 0.1093 | | Slope = 0.0985 | | Slope = 0.1125 | | Slope = 0.0713 |
| 4–4 | $F(4, 91)$ | $F(4, 90)$ | 3.67 (0.19) | 3.75 (0.18) | 3.73 | 3.07 (0.13) | 3.10 (0.12) | 6.71 (0.30) | 3.81 | 4.63 (0.15) |
| 4–8 | $F(4, 87)$ | $F(4, 86)$ | 3.53 (0.18) | 3.63 (0.18) | 3.73 | 2.95 (0.12) | 2.99 (0.12) | 6.50 (0.29) | 3.81 | 4.48 (0.15) |
| 4–12 | $F(4, 83)$ | $F(4, 82)$ | 3.45 (0.17) | 3.58 (0.18) | 3.73 | 2.88 (0.12) | 2.93 (0.12) | 6.28 (0.28) | 3.81 | 4.30 (0.14) |
| 12–12 | $F(12, 75)$ | $F(12, 74)$ | 1.77 (0.07) | 1.87 (0.07) | 3.73 | 1.34 (0.04) | 1.54 (0.20) | 2.88 (0.11) | 1.94 | 1.78 (0.04) |
| $B, =0.5$ | | | | Slope = 0.1085 | | Slope = 0.0978 | | Slope = 0.1118 | | Slope = 0.1066 |
| 4–4 | $F(4, 91)$ | $F(4, 90)$ | 3.75 (0.19) | 3.64 (0.19) | 3.71 | 3.13 (0.13) | 3.02 (0.13) | 5.86 (0.31) | 3.80 | 4.15 (0.16) |
| 4–8 | $F(4, 87)$ | $F(4, 86)$ | 3.58 (0.19) | 3.51 (0.17) | 3.71 | 2.97 (0.13) | 2.90 (0.12) | 5.65 (0.30) | 3.80 | 3.97 (0.16) |
| 4–12 | $F(4, 83)$ | $F(4, 82)$ | 3.50 (0.19) | 3.56 (0.19) | 3.71 | 2.89 (0.13) | 2.90 (0.13) | 5.58 (0.30) | 3.80 | 3.88 (0.15) |
| 12–12 | $F(12, 75)$ | $F(12, 74)$ | 1.95 (0.08) | 1.90 (0.08) | 1.90 | 1.45 (0.04) | 1.55 (0.05) | 2.64 (0.11) | 1.93 | 1.66 (0.05) |
| $A, \theta=0.8$ | | | | Slope = 0.1093 | | Slope = 0.0985 | | Slope = 0.1240 | | Slope = 0.0711 |
| 4–4 | $F(4, 91)$ | $F(4, 90)$ | 3.48 (0.20) | 3.27 (0.20) | 3.73 | 2.91 (0.14) | 2.72 (0.14) | 5.37 (0.27) | 4.10 | 3.90 (0.15) |
| 4–8 | $F(4, 87)$ | $F(4, 86)$ | 3.43 (0.20) | 3.24 (0.19) | 3.73 | 2.85 (0.14) | 2.68 (0.14) | 5.25 (0.28) | 4.10 | 3.77 (0.15) |
| 4–12 | $F(4, 83)$ | $F(4, 82)$ | 3.35 (0.19) | 3.11 (0.18) | 3.73 | 2.78 (0.13) | 2.57 (0.13) | 5.06 (0.27) | 4.10 | 3.62 (0.14) |
| 12–12 | $F(12, 75)$ | $F(12, 74)$ | 1.83 (0.08) | 1.86 (0.07) | 1.91 | 1.37 (0.05) | 1.52 (0.05) | 2.73 (0.11) | 2.03 | 1.70 (0.44) |
| $B, \theta=0.8$ | | | | Slope = 0.1098 | | Slope = 0.0989 | | Slope = 0.1125 | | Slope = 0.1016 |
| 4–4 | $F(4, 91)$ | $F(4, 90)$ | 3.40 (0.17) | 3.54 (0.17) | 3.75 | 2.88 (0.12) | 2.95 (0.12) | 4.88 (0.25) | 3.81 | 3.63 (0.14) |
| 4–8 | $F(4, 87)$ | $F(4, 86)$ | 3.36 (0.17) | 3.50 (0.19) | 3.75 | 2.83 (0.12) | 2.90 (0.13) | 4.67 (0.24) | 3.81 | 3.47 (0.14) |
| 4–12 | $F(4, 83)$ | $F(4, 82)$ | 3.24 (0.15) | 3.32 (0.16) | 3.75 | 2.73 (0.12) | 2.75 (0.12) | 4.44 (0.23) | 3.81 | 3.32 (0.13) |
| 12–12 | $F(12, 75)$ | $F(12, 74)$ | 1.81 (0.06) | 1.89 (0.07) | 1.92 | 1.37 (0.38) | 1.55 (0.05) | 2.25 (0.08) | 1.94 | 1.52 (0.04) |
| $A, \theta=0.99$ | | | | Slope = 0.1095 | | Slope = 0.0987 | | Slope = 0.1914 | | Slope = 0.0767 |
| 4–4 | $F(4, 91)$ | $F(4, 90)$ | 3.62 (0.16) | 2.11 (0.13) | 3.74 | 3.05 (0.12) | 1.85 (0.10) | 3.85 (0.24) | 5.79 | 2.94 (0.15) |
| 4–8 | $F(4, 87)$ | $F(4, 86)$ | 3.55 (0.16) | 2.10 (0.14) | 3.74 | 2.98 (0.12) | 1.83 (0.11) | 3.72 (0.24) | 5.79 | 2.83 (0.15) |
| 4–12 | $F(4, 83)$ | $F(4, 82)$ | 3.38 (0.16) | 2.10 (0.13) | 3.74 | 2.84 (0.11) | 1.82 (0.10) | 3.66 (0.23) | 5.79 | 2.77 (0.14) |
| 12–12 | $F(12, 75)$ | $F(12, 74)$ | 1.76 (0.07) | 1.88 (0.08) | 1.91 | 1.34 (0.04) | 1.52 (0.05) | 2.29 (0.11) | 2.60 | 1.49 (0.05) |
| $B, \theta=0.99$ | | | | Slope = 0.1093 | | Slope = 0.0985 | | Slope = 0.1155 | | Slope = 0.1006 |
| 4–4 | $F(4, 91)$ | $F(4, 90)$ | 3.47 (0.17) | 3.31 (0.18) | 3.73 | 2.93 (0.13) | 2.77 (0.13) | 4.66 (0.25) | 3.89 | 3.49 (0.15) |
| 4–8 | $F(4, 87)$ | $F(4, 86)$ | 3.38 (0.17) | 3.23 (0.18) | 3.73 | 2.84 (0.13) | 2.69 (0.13) | 4.59 (0.24) | 3.89 | 3.37 (0.15) |
| 4–12 | $F(4, 83)$ | $F(4, 82)$ | 3.30 (0.17) | 3.18 (0.18) | 3.73 | 2.76 (0.12) | 2.63 (0.13) | 4.52 (0.24) | 3.89 | 3.31 (0.14) |
| 12–12 | $F(12, 75)$ | $F(12, 74)$ | 1.81 (0.07) | 1.82 (0.17) | 1.91 | 1.37 (0.04) | 1.50 (0.05) | 2.48 (0.11) | 1.96 | 1.54 (0.05) |

[a]Standard errors of estimated means are shown parenthetically.

[b]Asymptotic distribution under null hypothesis.

[c]Mean of non-central chi-square with degrees of freedom equal to that in numerator of '$F$' and non-centrality parameter 100 times the approximate slope.

Table 7

Numbers of $F$'s in 5% and 10% critical regions under alternative hypothesis, in 100 replications.

| Model and parameterization | Asymptotic $F$ distributions[a] | | Wald tests | | | LeGrange tests | | Sims tests — Filtered Hannan efficient | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Wald | | LaGrange |
| | Granger | Sims | Granger | Sims (LDV) | Population[b] | Granger | Sims (LDV) | Sample | Population | Sample |
| $A, \theta=0.5$ | | | | | | | | | | |
| 4–4 | $F(4,91)$ | $F(4,90)$ | 76, 82 | 75, 82 | 0.76, 0.85 | 66, 76 | 72, 79 | 97, 98 | 0.77, 0.86 | 93, 98 |
| 4–8 | $F(4,87)$ | $F(4,86)$ | 76, 85 | 73, 83 | 0.76, 0.85 | 62, 75 | 65, 77 | 96, 98 | 0.77, 0.86 | 93, 96 |
| 4–12 | $F(4,83)$ | $F(4,82)$ | 79, 86 | 75, 81 | 0.76, 0.85 | 56, 78 | 69, 79 | 96, 98 | 0.77, 0.86 | 91, 96 |
| 12–12 | $F(12,75)$ | $F(12,74)$ | 62, 72 | 45, 62 | 0.55, 0.68 | 8, 20 | 19, 38 | 82, 92 | 0.57, 0.69 | 38, 56 |
| $B, \theta=0.5$ | | | | | | | | | | |
| 4–4 | $F(4,91)$ | $F(4,90)$ | 73, 79 | 70, 81 | 0.76, 0.84 | 69, 78 | 66, 78 | 94, 96 | 0.77, 0.86 | 85, 94 |
| 4–8 | $F(4,87)$ | $F(4,86)$ | 72, 77 | 74, 79 | 0.76, 0.84 | 63, 72 | 64, 75 | 90, 94 | 0.77, 0.86 | 84, 93 |
| 4–12 | $F(4,83)$ | $F(4,82)$ | 68, 75 | 67, 76 | 0.76, 0.84 | 57, 72 | 59, 74 | 92, 95 | 0.77, 0.86 | 80, 94 |
| 12–12 | $F(12,75)$ | $F(12,74)$ | 50, 60 | 47, 61 | 0.55, 0.67 | 14, 31 | 26, 47 | 70, 85 | 0.56, 0.69 | 31, 42 |
| $A, \theta=0.8$ | | | | | | | | | | |
| 4–4 | $F(4,91)$ | $F(4,90)$ | 59, 76 | 60, 70 | 0.76, 0.85 | 55, 67 | 51, 65 | 90, 96 | 0.82, 0.89 | 82, 92 |
| 4–8 | $F(4,87)$ | $F(4,86)$ | 63, 71 | 64, 76 | 0.76, 0.85 | 54, 68 | 49, 64 | 88, 94 | 0.82, 0.89 | 78, 89 |
| 4–12 | $F(4,83)$ | $F(4,82)$ | 61, 70 | 58, 68 | 0.76, 0.85 | 55, 69 | 50, 64 | 85, 94 | 0.89, 0.89 | 77, 86 |
| 12–12 | $F(12,75)$ | $F(12,74)$ | 36, 50 | 41, 56 | 0.55, 0.68 | 15, 24 | 72, 37 | 81, 86 | 0.62, 0.73 | 32, 52 |
| $B, \theta=0.8$ | | | | | | | | | | |
| 4–4 | $F(4,91)$ | $F(4,90)$ | 69, 79 | 71, 79 | 0.76, 0.85 | 59, 74 | 63, 74 | 87, 83 | 0.78, 0.86 | 78, 88 |
| 4–8 | $F(4,87)$ | $F(4,86)$ | 64, 73 | 73, 77 | 0.76, 0.85 | 58, 70 | 65, 75 | 87, 91 | 0.78, 0.86 | 78, 89 |
| 4–12 | $F(4,83)$ | $F(4,82)$ | 64, 76 | 71, 76 | 0.76, 0.85 | 58, 70 | 65, 72 | 85, 90 | 0.78, 0.86 | 74, 84 |
| 12–12 | $F(12,75)$ | $F(12,74)$ | 39, 59 | 49, 61 | 0.55, 0.68 | 8, 24 | 25, 44 | 62, 79 | 0.57, 0.69 | 17, 33 |
| $A, \theta=0.99$ | | | | | | | | | | |
| 4–4 | $F(4,91)$ | $F(4,90)$ | 72, 80 | 32, 51 | 0.76, 0.85 | 67, 82 | 24, 43 | 69, 74 | 0.96, 0.98 | 55, 70 |
| 4–8 | $F(4,87)$ | $F(4,86)$ | 72, 81 | 33, 51 | 0.76, 0.85 | 65, 80 | 23, 37 | 61, 75 | 0.96, 0.98 | 55, 67 |
| 4–12 | $F(4,83)$ | $F(4,82)$ | 68, 83 | 30, 48 | 0.76, 0.85 | 61, 76 | 20, 39 | 63, 72 | 0.96, 0.98 | 56, 65 |
| 12–12 | $F(12,75)$ | $F(12,74)$ | 42, 53 | 48, 56 | 0.55, 0.68 | 8, 24 | 28, 46 | 64, 70 | 0.85, 0.91 | 17, 37 |
| $B, \theta=0.99$ | | | | | | | | | | |
| 4–4 | $F(4,91)$ | $F(4,90)$ | 68, 80 | 59, 72 | 0.76, 0.85 | 58, 73 | 54, 66 | 79, 87 | 0.78, 0.87 | 70, 83 |
| 4–8 | $F(4,87)$ | $F(4,86)$ | 66, 76 | 56, 70 | 0.76, 0.85 | 53, 69 | 51, 60 | 81, 89 | 0.78, 0.87 | 67, 84 |
| 4–12 | $F(4,83)$ | $F(4,82)$ | 62, 76 | 56, 68 | 0.76, 0.85 | 52, 65 | 47, 60 | 80, 89 | 0.78, 0.87 | 65, 81 |
| 12–12 | $F(12,75)$ | $F(12,74)$ | 40, 52 | 37, 52 | 0.55, 0.68 | 10, 26 | 22, 36 | 66, 78 | 0.58, 0.70 | 22, 40 |

[a]Asymptotic distribution under null hypothesis.

[b]Based on distribution of non-central chi-square with degrees of freedom equal to that in numerator of $F$ and non-centrality parameter 100 times the approximate slope.

variable tests are the most interesting. However, we also present some information on the distribution of the filtered Hannan efficient Sims test of the hypothesis that $\{x_t\}$ does not cause $\{y_t\}$, since its distribution under the null hypothesis was among the best of the tests involving serial correlation correction. The Wald variant of this test rejects the null very reliably, and its mean value is much higher than those of the other tests whose behavior is summarized in these two tables. The tendency for this test statistic to be 'too large' is again evident in the first four experiments where its rejection frequency is greater than the asymptotic distribution theory would lead one to expect, and its mean value exceeds the limiting value (calculated as the mean of a non-central chi-square distribution with degrees of freedom equal to '$F$' numerator degrees of freedom and non-centrality parameter equal to the approximate slope multiplied by 100 — the number of observations divided by the numerator degrees of freedom). In the fifth experiment the parameterization of the Hannan efficient estimator cannot approximate the behavior of the spectral density of the disturbance of (1.3) at low frequencies very well, and evidently this prevents the test statistic from attaining the large values suggested by the asymptotic theory set forth above. In the sixth experiment the distribution over the 100 replications matches the asymptotic distribution nicely, a result which is probably a fortuitous coincidence due to the tendency of Wald Hannan efficient test statistics to be too large combined with a downward bias due to an inability to parameterize the spectral density of the disturbance adequately.

The behavior of the LaGrange-multiplier variant of this test is more difficult to explain. This test rejects the null with a frequency greater than that of the Granger and Sims lagged dependent variable Wald tests in most instances when four restrictions are implied by the null, and usually with a frequency greater than that of the LaGrange-multiplier variants of those tests when twelve restrictions are implied. For the samples generated here, small sample considerations evidently dominate the asymptotic behavior demonstrated in section 2.

The behavior of both the Wald- and LaGrange-multiplier variants of the Granger and Sims lagged dependent variable tests under the alternative seems to be well described by asymptotic theory. In most cases rejection frequencies run a little less than their limiting values [Wald (1943)], a tendency which becomes more pronounced as serial correlation in $\{x_t\}$ increases. The parameterization of the Sims lagged dependent variable test is ill-suited in Model B, $\theta=0.99$ (see table 3) for essentially the same reason that the form of this test requiring correction for serial correlation was not well parameterized, a fact which again leads to a rate of rejection well below the asymptotic norm. No systematic differences between Granger and Sims lagged dependent variable tests emerge in the experiments beyond those which clearly seem due to the inadequacy of the parameterization of the latter in some experiments. For twelve restrictions LaGrange-multiplier test statistics are much smaller than their Wald counterparts, a phenomenon also observed when the null is true; this tendency seems likely to emerge whenever the number of parameters estimated under the alternative becomes a substantial fraction of the number of observations.

## 6. Conclusion

In this paper we have compared tests due to Granger (1969) and Sims (1972) for the absence of a Granger causal ordering in stationary Gaussian time series. Unlike previous efforts [e.g., Nelson and Schwert (1979)] our comparison is based on asymptotic properties of these tests under the alternative as well as the limiting behavior under the null and sampling experiments. The analytic results on limiting behavior under the alternative employ the concept of approximate slope, which is inversely proportional to the number of observations required for a given power against the alternative at small significance levels.

The implications of this study for empirical work are unambiguous: one ought to use the Wald variant of either the Granger or Sims lagged dependent variable tests described in the introduction. Both have three important advantages over all other tests for the absence of Granger causality which we have studied:

(1) The approximate slopes of these tests are at least as great as those of the other tests (with the exception of the Wald Sims test) under all alternatives.
(2) The sampling distribution of these tests under the null was very satisfactory in our experiments, and under the alternative rejection frequencies corresponded very closely to those indicated by the asymptotic theory. By contrast, the sampling distribution of statistics for Sims tests requiring correction for serial correlation conformed very poorly to its limiting distribution and was sensitive to prefiltering under the null hypothesis: under the alternative, rejection frequencies also departed from their limiting values.
(3) The Granger and Sims lagged dependent variable tests are inexpensive to compute and require only an ordinary least squares computer algorithm. The variants of the Sims test which require correction for serial correlation require much more computation time and more decisions about parameterizations.

Our results are, of course, confined to time series which are stationary and Gaussian, although the results of sections 2 and 3 will admit relaxation of the normality assumptions employed in their derivation. It is difficult even to conjecture the outcome of a similar comparison for nonstationary time series, for in that case Granger's definitions will not apply unless some care is taken with the kind of nonstationarity permitted. The results of sampling experiments are very limited if interpreted in isolation. In the present case, however, there exists asymptotic paradigms for the outcome of these experiments under both the null and all alternatives. The experimental results for our favored tests conform well to these paradigms, and we conjecture that this result would emerge under variations on the experiments which have been reported here.

## Mathematical appendix

### *A.1. Definitions and notation*

(i) $e_{in}$ is an $n \times 1$ vector with $i$th element unity and the rest zero.

(ii) $F_n$ is an $n \times n$ complex matrix, $[F_n]_{jk} = n^{-\frac{1}{2}} \exp(2\pi ijk/n)$, $i^2 = -1$. Note $F_n' F_n = F_n F_n' = I_n$.

(iii) The $n \times n$ matrix $T_n$ is a Toeplitz matrix if $[T_n]_{jk} = t_{j-k}$.

(iv) The eigenvalues of any $n \times n$ symmetric matrix $A$ are denoted

$\mu_1(A) \geqq \mu_2(A) \geqq \cdots \geqq \mu_n(A)$.

(v) If $A$ and $B$ are $n \times n$ symmetric matrices and $A - B$ is positive semidefinite we write $A \circledcirc B$.

(vi) For any $n \times n$ matrix $A$, define the norms

$$\|A\|_I = (\mu_1(A'A))^{\frac{1}{2}}, \qquad \|A\|_{II} = n^{-1} \sum_{j=1}^{n} (\mu_j(A'A))^{\frac{1}{2}}.$$

Note that if $A$ is symmetric,

$$\|A\|_I = \sup_j |\mu_j(A)|, \qquad \|A\|_{II} = n^{-1} \sum_{j=1}^{n} |\mu_j(A)|;$$

if $A$ is also positive semidefinite,

$$\|A\|_I = \mu_1(A), \qquad \|A\|_{II} = n^{-1} \sum_{j=1}^{n} \mu_j(A) = n^{-1} \operatorname{tr}(A).$$

(vii) If $R(L) = \sum_{s=0}^{\infty} r_s L^s$ is an absolutely summable lag operator, its Fourier transform is denoted $\tilde{R}(\lambda) = \sum_{s=0}^{\infty} r_s \exp(-i\lambda s)$.

(viii) For any function $f(\lambda) > 0$ defined on $[-\pi, \pi]$ and symmetric about zero, $\mathscr{D}_n[f(\lambda)]$ denotes the $n \times n$ diagonal matrix with $j$th entry $f(2\pi j/n)$.

### A.2. Preliminaries

*Lemma 1.* *If $A$ and $B$ are $n \times n$ symmetric matrices and $A \geqq B$, then $\mu_j(A) \geqq \mu_j(B)$, $j = 1, \ldots, n$.*

*Proof.* Bellman (1960, p. 115).

*Lemma 2.* $\|AB\|_{II} \leqq \|A\|_I \cdot \|B\|_{II}$.

*Proof.* For any $n \times 1$ vector $x$, $x'B'A'ABx \leqq x'B'Bx\|A\|_I^2$, so $B'A'AB \leqq B'B\|A\|_I^2$. From Lemma 1, $\mu_j(B'A'AB) \leqq \mu_j(B'B)\|A\|_I^2$, $j = 1, \ldots, n$, whence the result.

*Lemma 3.* *Let $\{A_n\}$ and $\{B_n\}$, $n = 1, 2, \ldots$, be two sequences of symmetric positive definite matrices. If $\mu_n(A) \geqq m_A > 0$ and $\mu_n(B_n) \geqq m_B > 0$ for all $n$ and $\lim_{n\to\infty} \|A_n - B_n\|_{II} = 0$, then $\lim_{n\to\infty} \|A_n^{-1} - B_n^{-1}\|_{II} = 0$.*

*Proof.* Note that $\|A_n^{-1}\|_I < m_A^{-1}$ and $\|B_n^{-1}\|_I < m_B^{-1}$, and apply Lemma 2 to $B_n^{-1}(B_n - A_n)A_n^{-1}$.

*Lemma 4.* *If $A$ and $B$ are symmetric, then $\|A + B\|_{II} \leqq \|A\|_{II} + \|B\|_{II}$.*

*Proof.* Define the symmetric matrix operators $[\;]^+$ and $[\;]^-$ as follows: For $n \times n$ diagonal $D$, $[D]^+ \equiv \operatorname{diag}\{|d_{11}|, \ldots, |d_{nn}|\}$ and $[D]^- \equiv -[D]^+$. For $n \times n$ symmetric $A$ with diagonalization, $A \equiv PMP'$ ($M$ diagonal and $P$ orthonormal), $[A]^+ = P[M]^+P'$ and $[A]^- \equiv P[M]^-P'$. For $n \times n$ symmetric $A$ and $B$, $[A]^+ + [B]^+ \geqq A + B \geqq [A]^- + [B]^-$. From Lemma 1, $\mu_j([A]^+ + [B]^+) \geqq \mu_j(A + B) \geqq \mu_j([A]^- + [B]^-)$, $j = 1, \ldots, n$. Hence

$$
\begin{aligned}
\|A + B\|_{II} &= n^{-1} \sum_{j=1}^{n} |\lambda_j(A + B)| \leqq n^{-1} \sum_{j=1}^{n} |\lambda_j([A]^+ + [B]^+)| \\
&= n^{-1} \operatorname{tr}([A]^+) + n^{-1} \operatorname{tr}([B]^+) \\
&= \|[A]^+\|_{II} + \|[B]^+\|_{II} \\
&= \|A\|_{II} + \|B\|_{II}.
\end{aligned}
$$

*Lemma 5.* *Let $\{Y_n\}$, $Y_n \geqq 0$, $\forall n$, be a random sequence. If $\lim_{n\to\infty} E(Y_n) = 0$, then* plim $Y_n = 0$.

*Proof.* Partition $[0, \infty)$ into the intervals $[0, \varepsilon]$ and $(\varepsilon, \infty)$, where $\varepsilon$ is any specified positive constant. Then $E(Y_n) \geqq \varepsilon \cdot P[Y_n > \varepsilon)$, whence $\operatorname{plim} Y_n = 0$.

*Lemma 6.* *Let $T_n$ be an $n \times n$ Toeplitz matrix, $[T_n]_{jk} = t_{j-k}$, for which*

*$\sum_{j=-\infty}^{\infty} t_j z^j$ is convergent in an open annulus of the unit circle. Let*

$$t_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ij\lambda} g(\lambda) d\lambda,$$

*and suppose*

$$m_g > \inf_\lambda g(\lambda) > -\infty, \qquad M_g = \sup_\lambda g(\lambda) < \infty.$$

*If $H(\cdot)$ is a continuous function defined on the interval $[m_g, M_g]$, then*

$$\lim_{n\to\infty} n^{-1} \sum_{j=0}^{n} H(\mu_j(T_n)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H[g(\lambda)] d\lambda.$$

*Proof.* Grenander and Szegö (1958, theorem 5.2).

*Lemma 7. Let $T_n$ be an $n \times n$ Toeplitz matrix, $[T_n]_{jk} = t_{j-k}$, for which $\sum_{j=-\infty}^{\infty} |t_j| < \infty$. Let*

$$t_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ij\lambda} g(\lambda) d\lambda,$$

*and let*

$$C_n = \mathscr{D}_n[g(\lambda)] \quad \text{and} \quad J_n = F_n' C_n F_n.$$

*Then*

$$\lim_{n\to\infty} \|T_n - J_n\|_{II} = 0.$$

*Proof.* We closely follow Grenander and Szegö (1958, pp. 112–113) who used weaker assumptions to show convergence in a weaker norm. Introduce the Cesáro sums

$$g_p(\lambda) = \sum_{j=-p}^{p} (1 - |j|/p) t_j e^{-ij\lambda} \equiv \sum_{j=-\infty}^{\infty} t_j' e^{-ij\lambda}.$$

Define the Toeplitz form $[K_n]_{jk} \equiv t_{j-k}'$, the diagonal matrix $E_n = \mathscr{D}_n[g_p(\lambda)]$, and the symmetric positive definite matrix $L_n = F_n' E_n F_n$.

Applying Lemma 6 with $H(x) = |x|$ to the Toeplitz form $T_n - K_n$,

$$\lim_{n\to\infty} \|T_n - K_n\|_{II} = \frac{1}{2\pi} \int_{-\pi}^{\pi} |g(\lambda) - g_p(\lambda)| d\lambda. \tag{A.1}$$

The right side of (A.1) is bounded above by $2\sum_{j=1}^{p}|jt_j|/p+2\sum_{j=p}^{\infty}|t_j|$, which can be made arbitrarily small by choosing $p$ sufficiently large.

Observe that

$$[L_n]_{jk}=\sum_{j=1}^{n}\exp(2\pi i(j-k)/n)g_p(2\pi j/n)=\sum_{l=-\infty}^{\infty}t'_{j-k+ln}.$$

The matrix $K_n-L_n$ therefore has typical element

$$[K_n-L_n]_{jk}=t'_{j-k}-\sum_{l=-\infty}^{\infty}t'_{j-k+ln}.$$

This expression is zero for $|j-k|<n-p$. Hence $K_n-L_n$ has at most $2p$ non-zero eigenvalues, and the absolute value of no eigenvalue can exceed $\sum_{l=-\infty}^{\infty}|t_l|$. Hence

$$\|K_n-L_n\|_{II}\leqq p\sum_{l=-\infty}^{\infty}|t_l|/n,$$

which vanishes as $n\to\infty$.

Finally,

$$\|L_n-J_n\|_{II}=\|F'_n(E_n-C_n)F_n\|_{II}=\|E_n-C_n\|_{II}$$

converges to zero because the right side of (A.1) converges to zero. The result follows from Lemma 4.

### A.3. Main results

*Lemma 8. Suppose $\{z_t\}$ is a zero-mean stationary Gaussian process with continuous spectral density $S_z(\lambda)$ bounded above and below by positive constants. Let $R(L)$ be an absolutely summable lag operator with all roots outside an open annulus of the unit circle, let $\tilde{R}(\lambda)$ be the Fourier transform of $R(L)$, and define $z_t^*=R(L)z_t$. Then*

$$\operatorname{plim} n^{-1}\sum_{t=1}^{n}z_t^{*2}=\frac{1}{2\pi}\int_{-\pi}^{\pi}|\tilde{R}(\lambda)|^2S_z(\lambda)\,d\lambda.$$

*Proof.* Since $\{z_t^*\}$ is ergodic, $\operatorname{plim} n^{-1}\sum_{t=1}^{n}z_t^{*2}=\operatorname{var}(z_t^*)$. The process $z_t^*$ has spectral density $|\tilde{R}(\lambda)|^2S_z(\lambda)$ [Fishman (1969, p. 41)], whence

$$\operatorname{var}(z_t^*)=\frac{1}{2\pi}\int_{-\pi}^{\pi}|\tilde{R}(\lambda)|^2S_z(\lambda)\,d\lambda.$$

*Lemma 9.* *Let $\{z_t\}$ be the process defined in Lemma 8 and denote $z_n = (z_1, \ldots, z_n)'$. Let $\hat{\Omega}_n$ be a sequence of $n \times n$ positive definite matrices, and $|\tilde{R}(\lambda)|^2$ a continuous function defined on $[-\pi, \pi]$ with positive lower and upper bounds. If*

$$\operatorname{plim} \|\hat{\Omega}_n^{-1} - F_n \mathscr{D}_n[|\tilde{R}(\lambda)|^2] F_n'\|_{II} = 0, \tag{A.2}$$

*then*

$$\operatorname{plim} n^{-1}[z_n'\hat{\Omega}_n^{-1} z_n - z_n' F_n \mathscr{D}_n[|\tilde{R}(\lambda)|^2] F_n' z_n] = 0. \tag{A.3}$$

*Proof.* Let the bracketed symmetric matrix in (A.2) have canonical factorization $P_n' \Lambda_n P_n$, where $P_n$ is orthonormal and $\Lambda_n$ is diagonal. Let $z_n^\dagger = P_n z_n$. Then

$$n^{-1} z_n' P_n' \Lambda_n P_n z_n = n^{-1} z_n^{\dagger\prime} \Lambda_n z_n^\dagger = n^{-1} \sum_{i=1}^{n} (z_{ni}^\dagger)^2 \lambda_i^n \leqq n^{-1} \sum_{i=1}^{n} (z_{ni}^\dagger)^2 |\lambda_i^n|.$$

Because $P_n$ is orthonormal, $\operatorname{var}(z_{ni}^\dagger) = \operatorname{var}(z_t)$. The expected value of the last term is $\operatorname{var}(z_t) n^{-1} \sum_{i=1}^n |\lambda_i^n|$. But this term is non-negative, so by Lemma 5 it converges in probability to zero.

*Lemma 10.* *Let $\{z_t\}$ and $R(L)$ be the process and lag operator defined in Lemma 8,*

$$\operatorname{plim} n^{-1} z_n' F_n \mathscr{D}_n[|\tilde{R}(\lambda)|^2] F_n' z_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\tilde{R}(\lambda)|^2 S_z(\lambda)\, \mathrm{d}\lambda.$$

*Proof.* Let $\tilde{z}_n = (\tilde{z}_1^n, \ldots, \tilde{z}_n^n)' = F_n' z_n$, the finite Fourier transform of $\{z_t\}$. Let $z_t$ have moving average representation $z_t = Q(L)\varepsilon_t$, in which the $\varepsilon_t$ are i.i.d. normal. Let $\tilde{\varepsilon}_n = (\tilde{\varepsilon}_1^n, \ldots, \tilde{\varepsilon}_n^n)' = F_n' \varepsilon_n$; the $\tilde{\varepsilon}_i^n$ are i.i.d. complex normal. We have

$$n^{-1} z_n' F_n' \mathscr{D}_n[|\tilde{R}(\lambda)|^2] F_n z_n = n^{-1} \sum_{j=1}^{n} |\tilde{R}(2\pi j/n)|^2 |\tilde{z}_j^n|^2.$$

By virtue of Hannan (1970, corollary 5, p. 214) and the fact that the $|\tilde{R}(2\pi j/n)|^2$ are bounded above and below by positive constants uniformly in $n$, the difference between the right side of (A.3) and

$$n^{-1} \sum_{j=1}^{n} |\tilde{R}(2\pi j/n)|^2 |\tilde{Q}(2\pi j/n)|^2\, |\tilde{\varepsilon}_j^n|^2$$

vanishes in probability. But the $\tilde{\varepsilon}_j^n$ are i.i.d. complex normal and $\tilde{R}$ and $\tilde{Q}$ are continuous, so the last expression converges in probability to

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}|\tilde{R}(\lambda)|^2 S_z(\lambda)\mathrm{d}\lambda.$$

*Theorem 1. Let $\{z_t\}$, $\{z_t^*\}$, and $R(L)$ be the processes and lag operator defined in Lemma 8, and define $\hat{\Omega}_n$ as in Lemma 9. Given (A.2),*

$$\operatorname{plim}\left\{n^{-1}\sum_{t=1}^{n} z_t^{*2} - n^{-1}z_n'\hat{\Omega}_n^{-1}z_n\right\}=0. \tag{A.4}$$

*If there exists $\varepsilon>0$ such that $P[\lambda_n(\hat{\Omega}_n)<\varepsilon]\to 0$, $P[\lambda_1(\hat{\Omega}_n)>\varepsilon^{-1}]\to 0$, then (A.2) may be replaced by*

$$\operatorname{plim}\|\hat{\Omega}_n - F_n\mathscr{D}_n[|\tilde{R}(\lambda)|^{-2}]F_n'\|_{II}=0. \tag{A.5}$$

*Proof.* The conditions on $\lambda_n(\hat{\Omega}_n)$ and $\lambda_1(\hat{\Omega}_n)$ guarantee that $\|\hat{\Omega}_n\|_I$ and $\|\hat{\Omega}_n^{-1}\|_I$ are bounded away from zero with probability 1, in the limit. The conditions on $R(L)$ guarantee that $|\tilde{R}(\lambda)|^2$ is bounded above and below by positive constants for $\lambda\in[-\pi,\pi]$. Lemma 4 then guarantees the equivalence of (A.2) and (A.5).

The result itself follows from Lemmas 8, 9, and 10, with application of Lemma 4.

*Remark 1.* Theorem 1 is sufficient for the claim of the text that tests based on a generalized least squares estimator with known variance matrix, and on a feasible generalized least squares estimator with estimated $\hat{\Omega}_n$ [satisfying (A.2) or (A.5)], have the same approximate slope; let $z_t=y_t-\sum_{j=1}^{k}b_jx_{jt}$, where $x_{1t},\dots,x_{kt}$ are the regressors and $b_j$ is any consistent estimator of $\beta_j$ in the projection $\sum_{j=1}^{k}\beta_jx_{jt}$ of $y_t$ on $x_{1t},\dots,x_{kt}$.

*Theorem 2* [Amemiya (1973) feasible generalized least squares]. *Suppose $\Omega_n$ is estimated as suggested by Amemiya (1973): the disturbances are assumed to follow an autoregressive process of order p, and the autoregression is estimated accordingly, using ordinary least squares residuals in lieu of the disturbances. If the true autoregressive representation of the disturbances is $R(L)u_t=\varepsilon_t$, $\operatorname{var}(\varepsilon_t)=1$, $R(L)$ satisfies the conditions of Lemma 8, and $\{u_t\}$ has absolutely summable autocovariance function, then there exists a function $p(n)$ such that if $p=p(n)$ then (A.2) is satisfied.*

*Proof.* Suppose initially that $p$ is fixed, and let $R_p(L)$ denote the lag operator from the autoregression of $u_t$ on $u_{t-1},\dots,u_{t-p}$. The roots of $R_p(L)=\sum_{s=0}^{p}r_s^pL^s$ lie outside an open annulus of the unit circle [Grenander and Szegö (1958, pp. 40–42)]. Let

$$\underset{(n-p)\times n}{G_n^p} = \begin{bmatrix} r_p^p & r_{p-1}^p & r_{p-2}^p & \cdots & r_1^p & r_0^p & 0 & \cdots & 0 & 0 & 0 \\ 0 & r_p^p & r_{p-1}^p & \cdots & r_2^p & r_1^p & r_0^p & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & r_2^p & r_1^p & r_0^p \end{bmatrix}.$$

We have

(i) $\lim \|\Omega_n^{-1} - G_n^{p\prime} G_n^p\|_{II} = 0,$

(ii) $\operatorname{plim} \|\hat{\Omega}_n^{-1} - \hat{G}_n^{p\prime} \hat{G}_n^p\|_{II} = 0,$

(iii) $\lim \|\Omega_n^{-1} - F_n \mathscr{D}_n[|\tilde{R}(\lambda)|^2] F_n'\|_{II} = 0,$

(iv) $\operatorname{plim} \|\hat{G}_n^{p\prime} \hat{G}_n^p - G_n^{p\prime} G_n^p\|_{II} = 0.$

The assertion (i) follows from the fact that $\Omega_n^{-1}$ and $G_n^{p\prime} G_n^p$ differ only in the $p \times p$ submatrix consisting of their first $p$ rows and columns. Hence $\sum_{j=1}^n |\mu_j(\Omega_n^{-1} - G_n^{p\prime} G_n^p)|$ is a fixed constant not depending on $n$. Similarly, (ii). From Lemma 7,

$$\lim_{n\to\infty} \|\Omega_n - F_n \mathscr{D}_n[|\tilde{R}(\lambda)|^2] F_n'\|_{II} = 0.$$

But

$$\mu_n[F_n \mathscr{D}_n[|\tilde{R}(\lambda)|^2] F_n'] \geqq \min_\lambda |\tilde{R}(\lambda)|^2 > 0,$$

and [Grenander and Szegö (1958, p. 66)]

$$\mu_n(\Omega_n) \geqq \min_\lambda |\tilde{R}(\lambda)|^2.$$

(iii) then follows from Lemma 3. To establish (iv), note

$$\|\hat{G}_n^{p\prime} \hat{G}_n^p - G_n^{p\prime} G_n^p\|_{II} \leqq \|\hat{G}_n^p - G_n^p\|_{II} \cdot \|G_n^p\|_I + \|\hat{G}_n^p - G_n^p\|_{II} \cdot \|G_n^p\|_I.$$

Let $x_m$ be any $m \times 1$ *vector with* $x_m' x_m = 1$.

$$\sup_{x_n} \|G_n^p x_n\|_I \leqq \left[ \sup_{x_{p+1}} \left( \sum_{s=0}^p r_s^p x_s^{p+1} \right)^2 \right]^{\frac{1}{2}},$$

which is bounded uniformly above for all $n$. Likewise $\operatorname{plim} \|\hat{G}_n^p\|_I$ is bounded above. Since $\operatorname{plim} \hat{r}_s^p = r_s^p$, $s=0,\ldots,p$,

$$\operatorname{plim}\left[\sup_{x_{p+1}}\left(\sum_{s=0}^{p}(\hat{r}_s^p - r_s^p)x_s^{p+1}\right)^2\right]^{\frac{1}{2}} = 0,$$

whence

$$\operatorname{plim}\|\hat{G}_n^p - G_n^p\|_I = \operatorname{plim}\|\hat{G}_n^p - G_n^p\|_{II} = 0.$$

The result for fixed $p$ follows by application of Lemma 4 to (i)—(iv).

The assumptions on $R(L)$ guarantee $\lim_{p\to\infty}|\tilde{R}_p(\lambda)|^2 = |\tilde{R}(\lambda)|^2$, whence

$$\lim_{n\to\infty}\|F_n\mathscr{D}_n[|\tilde{R}_p(\lambda)|^2]F_n' - F_n\mathscr{D}_n[|\tilde{R}(\lambda)|^2]F_n'\|_{II} = 0.$$

$p$ can be increased with $n$ at a rate such that (A.2) obtains.

*Theorem 3 (Hannan (1963) efficient estimation). Suppose* $\Omega_n = F_n\mathscr{D}_n[\hat{S}_u(\lambda)]F_n'$, *where* $\hat{S}_u(\lambda)$ *is a consistent estimator of* $S_u(\lambda)$, *the spectral density of the disturbance. If the true autoregressive representation of the disturbance is* $R(L)u_t = \varepsilon_t$, $\operatorname{var}(\varepsilon_t) = 1$, *and* $R(L)$ *satisfies the conditions of Lemma 7, then (A.5) is satisfied.*

*Proof.* Immediate from the fact that $S_u(\lambda) = |\tilde{R}(\lambda)|^{-2}$.

*Remark 2.* Theorems 2 and 3 illustrate applications of Theorem 1 to the two estimators investigated in this paper. Presumably this theorem could be applied in the case of other estimators as well.

## References


Amemiya, T., 1973, Generalized least squares with an estimated autocovariance matrix, Econometrica 41, 723–732.

Bahadur, R.R., 1960, Stochastics comparison of tests, Annals of Mathematical Statistics 31, 276–295.

Bellman, R.A., 1960, Introduction to matrix analysis (McGraw-Hill, New York).

Berndt, E.R. and N.E. Savin, 1977, Conflict among criteria for testing hypotheses in the multivariate linear regression model, Econometrica 45, 1263–1278.

Box, G.E.P. and M.E. Muller, 1958, A note on the generation of random normal deviates, Annals of Mathematical Statistics 27, 610–611.

Breusch, T.S., 1979, Conflict among criteria for testing hypothesis: Extension and comment, Econometrica 47, 203–208.

Fishman, G., 1969, Spectral methods of econometrics (Harvard University Press, Cambridge, MA).

Geweke, J., 1978, Testing the exogeneity specification in the complete dynamic simultaneous equation model, Journal of Econometrics 7, 163–185.

Geweke, J., 1981a, A comparison of tests of the independence of two covariance-stationary time series, Journal of the American Statistical Association 76, 363–373.

Geweke, J., 1981b, The approximate slopes of econometric tests, Econometrica 49, 1427–1442.

Granger, C.W.J., 1969, Investigating causal relations by econometric models and cross spectral methods, Econometrica 37, 428–438.
Grenander, U. and G. Szegö, 1958, Toeplitz forms and their applications (University of California Press, Berkekely, CA).
Hannan, E.J., 1963, Regression for time series, in: M. Rosenblatt, ed., Proceedings of a symposium on time series analysis (Wiley, New York).
Hannan, E.J., 1970, Multiple times series (Wiley, New York).
Knuth, D.E., 1969, The art of computer programming, Vol. 2 (Addison-Wesley, New York).
Nelson, C.R. and G.W. Schwert, 1979, Tests for Granger/Wiener causality: A Monte Carlo investigation, University of Rochester Graduate School of Management working paper 7905.
Oberhofer, W. and J. Kmenta, 1974, A general procedure for obtaining maximum likelihood estimates in generalized regression models, Econometrica 42, 579–590.
Pierce, D.A. and L.D. Haugh, 1977, Causality in temporal systems: Characterizations and a survey, Journal of Econometrics 5, 265–293.
Rozanov, Y.A., 1969, Stationary random processes (Holden-Day, San Francisco, CA).
Sargent, T.J., 1976, A classical macroeconometric model of the U.S., Journal of Political Economy 87, 207–237.
Silvey, N.E., 1959, The LaGrangian multiplier test, Annals of Mathematical Statistics 30, 389–407.
Sims, C., 1972, Money, income and causality, American Economic Review 62, 540–552.
Sims, C., 1974, Distributed lags, in: M.D. Intriligator and D.A. Kendrick, eds., Frontiers of quantitative economics (North-Holland, Amsterdam).
Wald, A., 1943, Tests of statistical hypotheses concerning several parameters when the number of observations is large, Transactions of the American Mathematical Society 54, 426–485.
Whittle, P., 1963, Prediction and regulation by linear least-squares method (Van Nostrand, Princeton, NJ).