Econometrics Working Paper EWP0305
ISSN 1485-6441

SOME FINITE SAMPLE RESULTS ON TESTING FOR GRANGER NONCAUSALITY

Judith A. Clarke and Sadaf Mirza

Department of Economics, University of Victoria,
PO Box 1700, STN CSC, Victoria, BC, V8W 3P5
Email: jaclarke@uvic.ca

Key Words and Phrases: cointegration; error correction model; lag length selection; sequential testing; testing for Granger noncausality; Monte Carlo simulation.

JEL Classifications: C32, C51, C52.

ABSTRACT

We compare testing strategies for Granger noncausality in vector autoregressions (VARs) that may or may not have unit roots and cointegration. Sequential testing methods are examined; these test for cointegration and use either a differenced VAR or a vector error correction model (VECM), in which to undertake the main noncausality test. Basically, these strategies attempt to verify the validity of appropriate standard limit theory. We contrast such methods with an augmented lag approach that ensures the limiting χ^2 null distribution irrespective of the data's nonstationarity characteristics. Our simulations involve bivariate and trivariate VARs in which we allow for the lag order to be selected by general to specific testing as well as by model selection criteria. We find that the current practice of pretesting for cointegration can result in severe over-rejections of the noncausal null while overfitting results in better control of the Type I error probability with often little loss in power.

1. INTRODUCTION

This paper examines the properties of Wald-tests for Granger noncausality (GNC)[1], when applied in finite-order vector autoregressive (VAR) models with nonstationary variables. Several preliminary steps affect the finite-sample properties of the GNC statistic: estimation of the VAR lag order and determination of nonstationarity characteristics. We use Monte Carlo experiments to determine the impact of these preliminary steps on the empirical probability of a Type I error and the associated "power" of GNC statistics, allowing for the VAR lag order and cointegrating rank to be estimated. We surveyed over two hundred applied studies so we could examine methods

[1] Granger (1969), which builds on earlier work of Wiener (1956).

system of order at most three, allowing for three alternative error covariance matrices and three GNC testing strategies. In the first, the test is based on a vector autoregression in levels, denoted as a VARL, that ignores the impact of nonstationarity on the test statistic's asymptotic null distribution. The second strategy, suggested by Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996)[2], tests for GNC using an adjusted VARL model so that the test statistic's limiting null distribution is standard, irrespective of the system's nonstationarity properties. The third approach, applied most often, uses preliminary cointegration tests to basically ensure that the GNC test is undertaken in an appropriately specified model; usually either a vector error-correction model (VECM) or a VAR specified in terms of first differences; a VARD model.

As we allow for estimation of the dynamics of the system, our work substantially elaborates on Toda and Phillips (1994), Dolado and Lütkepohl (1996), Zapata and Rambaldi (1997), and Yamada and Toda (1998). The simulation experiments of Toda and Phillips (1994), though extensive, are limited to trivariate VAR(1) DGPs with the lag order either specified correctly or overestimated by a fixed order. Dolado and Lütkepohl (1996) undertake a small Monte Carlo involving a bivariate VAR(2) system with iid errors; they assume that the VAR order is either known or over-specified. Zapata and Rambaldi (1997) examine testing for GNC within trivariate systems, but they limit attention to DGPs that are "sufficiently" cointegrated in the sense of Toda and Phillips (1993, 1994) so the statistic has a standard limiting distribution; we consider situations in which nonstandard asymptotic distributions result. Zapata and Rambaldi also assume that the lag order is either correctly specified or over/under specified by one lag. Finally, Yamada and Toda (1998) investigate a wide range of DGPs, but limit attention to a bivariate VAR(1) model that is always assumed to be a VAR(2). They show that the finite sample distributions of their GNC statistics, with any innovation covariance matrix and a known lag order, are equivalent to those of a transformed model with an identity covariance. However, this result no longer holds once we allow for estimation of the lag order, as the error covariance matrix affects lag order choice and, therefore, the finite sample distributions of the GNC statistics.

The rest of this paper is organized as follows. In section 2 we outline the tests for Granger noncausality, the tests for cointegration and noncointegration, and the methods we consider for selecting the lag order. Section 2 also provides information from our survey of the empirical GNC literature that motivates some of our procedural choices. The simulation study is described in section 3, with results from the experiments presented and examined in section 4. We consider an empirical illustration in section 5 to highlight the differences that can result between the approaches we have studied. Some concluding remarks are given in section 6.

---

[1] Granger (1969), which builds on earlier work of Wiener (1956).

[2] Saikkonen and Lütkepohl (1996) extend this to infinite order cointegrated VARs.

2. THE TESTING/SELECTION PROCEDURES

There are three main testing and selection issues: testing for GNC, testing for the cointegrating rank and selection of lag length; augmentation in the cointegrating regressions also needs determining. We briefly detail each below and then outline the pretesting strategies.

2.1. Tests for Granger Noncausality

Each of the statistics used to test for GNC is a Wald statistic from an appropriate model: a VARL model, an augmented VARL model, a VECM, and a VARD model. In each of these models, the lag length k must be determined; the selection of k is considered in section 2.3 below.

In general, consider a model where θ is an $m\times 1$ vector of parameters and let R be a known nonstochastic $q\times m$ matrix with rank q. To test H_0: $R\theta=0$, a Wald statistic is

$$W = T\,\hat{\theta}^T R^T \{R\,\hat{V}[\hat{\theta}]R^T\}^{-1} R\,\hat{\theta} \qquad (1)$$

where $\hat{\theta}$ is a consistent estimator of θ, $\hat{V}[\hat{\theta}]$ is a consistent estimator of the asymptotic variance-covariance matrix of $\sqrt{T}(\hat{\theta}-\theta)$, and T is the number of observations. Given appropriate conditions, W is asymptotically distributed as a $\chi^2(q)$ variate under H_0.

For an $n\times 1$ vector time series $\{x_t: t=1,2,...,T\}$, the VARL model of order k is

$$x_t = \sum_{i=1}^{k} \Pi_i x_{t-i} + \varepsilon_t \qquad (2)$$

where Π_i is an $n\times n$ matrix of parameters, $\varepsilon_t \sim IN(0, \Sigma)$, and (2) is initialized at t=-k+1...0; the initial values can be any random vectors including constants. Let $x_t = \left(x_{1t}^T, x_{2t}^T, x_{3t}^T\right)^T$ where x_{st} is an $n_s\times 1$ vector for s=1,2,3 with $n=n_1+n_2+n_3$. Also, with Π_i conformably partitioned, let $\Pi_{i,13}$ be the $n_1\times n_3$ top-right partition of Π_i. Suppose we wish to determine whether or not x_{3t} Granger causes x_{1t}. Then, in this levels model, the null hypothesis of GNC is H_0^L: $P_{13}=0$ where $P_{13}= [\Pi_{1,13}, \Pi_{2,13},..., \Pi_{k,13}]$. This null hypothesis can also be written in the form H_0^L: $R\theta=0$ so the Wald statistic from (2), denoted WL, is then given by (1) where $\hat{\theta}$ is the LS estimator of $\theta=vec[\Pi_1, \Pi_2,...,\Pi_k]$ and R is a selector matrix such that $R\theta=vec[P_{13}]$.

Sims et al. (1990) and Toda and Phillips (1993, 1994), show that WL is asymptotically a $\chi^2(n_1n_3k)$ variate under H_0^L when each series is either stationary or nonstationary with "sufficient" cointegration, which involves a rank condition on a submatrix of the cointegrating matrix. The statistic WL has a nonstandard, but free of nuisance parameters, limiting distribution for a non-intercept or an intercept/time trend VARL model; a nonstandard distribution involving nuisance parameters arises for an intercept/no time trend VARL model when there is no cointegration; and WL has a nonstandard limiting distribution that may depend on nuisance parameters when the relevant nonstationary series are "insufficiently" cointegrated.

These results lead Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996) to propose a method that gives an asymptotic χ^2 null distribution for the GNC statistic in the VARL model, irrespective of the system's integration or cointegration properties. They use that the covariance matrix singularity of the LS estimator in a nonstationary system can be removed by fitting a VARL process whose order exceeds the true order by the highest degree of integration in the system. For instance, when the true DGP is a VARL(k) with I(1) variables, we estimate a VARL(k+1) model (irrespective of whether cointegration exists or not) and test for GNC using the first k coefficients. Consider the augmented VARL model

$$x_t = \sum_{i=1}^{k} \Pi_i x_{t-i} + \sum_{i=1}^{d} \Pi_{k+i} x_{t-k-i} + \varepsilon_t \tag{3}$$

where d is the highest order of integration for any element of x_t. In this augmented levels model, the null hypothesis of GNC (between x_{3t} and x_{1t}) is H_0^{AL} : $P_{13}=0$ where P_{13} is as above; note that H_0^{L} and H_0^{AL} test the same set of restrictions in the VARL and augmented VARL models, respectively. The Wald statistic from (3) is then given by (1) where $\hat{\theta}$ is the LS estimator of $\theta=vec[\Pi_1, \Pi_2, ..., \Pi_{k+d}]$ and R is a selector matrix such that $R\theta=vec[P_{13}]$. We denote the resulting Wald statistics as WALd. Under our assumptions, WALd is asymptotically distributed as a $\chi^2(n_1n_3k)$ variate under H_0^{AL} , irrespective of the nonstationarity properties of x_t.

Alternatively, using the operator $\Delta x_t=x_t-x_{t-1}$, the VARL can be written as a VECM

$$\Delta x_t = \Pi x_{t-1} + \sum_{j=1}^{k-1} \Gamma_j \Delta x_{t-j} + \varepsilon_t \tag{4}$$

where $\Pi = -(I_n - \sum_{i=1}^{k} \Pi_i)$ and $\Gamma_j = -\sum_{i=j+1}^{k} \Pi_i$ for j=1,2,..., k-1. We assume that all the roots of $\left|I_n - \sum_{i=1}^{k} \Pi_i z^i\right| = 0$ lie outside the complex unit circle except for possibly some unit roots. Since the rank of Π, denoted r, equals the dimension of the cointegrating space, the statistic used to test for noncausality depends on the value of r. There are three possible cases to consider: r=n, 0<r<n, and r=0. The vector x_t is believed stationary when r=n; then WL is used to test for GNC.

If 0<r<n then Π can be decomposed as $\Pi=\alpha\beta^T$, where α and β are $n\times r$ matrices of rank r, β is the cointegrating matrix and α contains the error correction or adjustment vectors. With Γ_j conformably partitioned with Δx_t, let $\Gamma_{j,13}$ be the $n_1\times n_3$ top-right partition of Γ_j. In this error-

correction model, the null hypothesis of GNC (between x_{3t} and x_{1t}) is H_0^{EC} : $G_{13}=0$ and $\alpha_1\beta_3^T=0$ where $G_{13}=[\Gamma_{1,13}, \Gamma_{2,13}, ..., \Gamma_{k-1,13}]$, α_1 contains the first n_1 rows of α, and β_3^T contains the last n_3 columns of β^T. The Wald statistic from (4) with $\Pi=\alpha\beta^T$, denoted WEC, is then given by (1) where $\hat{\theta}$ is the estimator of $\theta = \text{vec}[\Gamma_1, \Gamma_2, ..., \Gamma_{k-1}, \alpha\beta^T]$ and R is a selector matrix such that $R\theta=\text{vec}[G_{13}, \alpha_1\beta_3^T]$. We estimate system (4) by maximum likelihood as outlined in Johansen (1988) using the normalization suggested by Johansen (1988, p.235); the sample value of WEC is then obtained using the transformations given in Lütkepohl (1993).

Under our assumptions, Toda and Phillips (1993, Theorem 2) show that the statistic WEC has a limiting $\chi^2(n_1n_3k)$ distribution when $\text{rank}(\alpha_1)=n_1$ or $\text{rank}(\beta_3)=n_3$. If either of the rank conditions is not satisfied then the problems of nuisance parameters, and nonstandard distributions, enter the limit theory. Incorporating testing empirically whether these rank conditions hold gives rise to the additional pretesting in Toda and Phillips' (1994) strategies P1-P3. Researchers are not currently applying the additional pretest step; they assume validity of Theorem 2 of Toda and Phillips (1993, 1994) when cointegration is identified. As our aim is to replicate applied practice, we also do not test the validity of the rank conditions.

If r=0 then $\Pi=0$ in which case (4) reduces to the VARD model

$$\Delta x_t = \sum_{j=1}^{k-1}\Gamma_j\Delta x_{t-j} + \varepsilon_t \qquad (5)$$

with the GNC null (between x_{3t} and x_{1t}) being H_0^D : $G_{13}=0$ where G_{13} is as above so the Wald statistic from (5), denoted WD, is then given by (1) where $\hat{\theta}$ is the LS estimator of $\theta=\text{vec}[\Gamma_1, \Gamma_2, ..., \Gamma_{k-1}]$ and R is a selector matrix such that $R\theta=\text{vec}[G_{13}]$. The results of Toda and Phillips (1994, Proposition 1) ensure that WD is asymptotically distributed as a $\chi^2(n_1n_3(k-1))$ variate under H_0^D.

2.2. Tests for Noncointegration & Cointegration

Since Granger's (1981) seminal paper on cointegration and the work of Engle and Granger (1987), this topic has received enormous attention. Numerous procedures exist for testing whether nonstationary series are cointegrated or noncointegrated. Some methods are based on single equation analysis, while others use a systems approach, which requires solving an identification problem, typically by principal components, canonical correlations, or restrictions on the parameter space. They also differ on the null hypothesis examined- a null of noncointegration versus a null of cointegration. This is an important distinction under classical hypothesis testing, as traditional choices of the level imply that the null will be rejected only for extreme samples.

Which are the most commonly applied cointegration tests? We provide information in Table 1 from an ad hoc survey we conducted based on 218, post 1992, publications or discussion papers to help answer this question. The table reports the cointegration test applied; the method employed to choose the augmentation or bandwidth parameter (as necessary); and the strategy adopted to determine the lag order for any estimated autoregressive models (as applicable). Our aim in preparing this table was not to be exhaustive, but rather to cover a wide range of sources of applied papers. We list the journals and origins of the discussion papers in the Appendix. Of the 218 papers we sampled, 173 applied one or more cointegration tests, and, in total, 226 cointegration tests were performed with 180 of the 226 tests requiring choice of an augmentation or bandwidth parameter. An autoregressive model was estimated by 135 of the 218 papers surveyed employing one or more ways to select the lag order. Table 1 gives relative percentages on the methods applied at each stage. For example, 47% of the 226 cointegration tests apply Johansen's ML test (JJ), while the Engle-Granger ADF test (EG-ADF) was employed on 32% of occasions. The coverage of "OTHER" in each case is available on request from the first author.

Table 1. Results of survey of empirical literature on cointegration test; bandwidth/augmentation procedure and lag order selection process

Cointegration Test		Augmentation or bandwidth parameter method, applied at the cointegration stage		Lag Order Procedure	
JJ	47%	NS	26%	LR	19%
EG-ADF	32%	PRESET	16%	PRESET	19%
PO	4%	LR	14%	AIC	14%
PP	3%	RANGE	13%	NS	14%
NS	2%	SC	10%	SC	13%
OTHER 12%		AIC	9%	FPE	12%
		BPQ	5%	RANGE	4%
		FPE	2%	OTHER	5%
		OTHER	5%		

Notes: PO = Phillips and Ouliaris (1990); PP = Phillips and Perron (1988); NS = Not Specified; PRESET = preset to an arbitrary number; LR = Likelihood Ratio general to specific; RANGE = range of numbers arbitrarily selected; SC = Schwarz's (1978) Criterion; AIC = Akaike's (1973) Information Criterion; BPQ = Box and Pierce's (1970) Q-test; FPE = Akaike's (1969) Final Prediction Error Criterion.

Given this, we consider three tests: the EG-ADF and JJ procedures, which dominate applied studies, and the McCabe et al. (1997) (MLS) test. The EG-ADF and JJ methods analyze a null hypothesis of noncointegration and, given classical testing procedures, we only reject in favor of cointegration if there is substantial evidence to do so. However, as the hypothesis of cointegration is often of primary interest we include the MLS residual based test also. As the EG-ADF and JJ methods are common tools, we do not detail them here. In our application of the EG-ADF test, we follow the general recommendations of Hall (1994), and Ng and Perron (1995), for the ADF nonstationarity test, by estimating the number of augmentation terms via general-to-specific modeling. We use the critical values from MacKinnon (1991). For the JJ test, we examine a VECM with no deterministic trends: $\Delta x_t = \alpha\beta^T x_{t-1} + \sum_{j=1}^{k-1}\Gamma j\,\Delta x_{t-j} + \varepsilon_t$. We consider the λ_{max} (maximum eigenvalue) test to illustrate the procedure, using the asymptotic critical values estimated by MacKinnon et al. (1999).

The MLS approach is an extension of the Kwiatkowski et al. (1992) stationarity test. As this cointegration test is not commonly used in applied work, we provide salient details here. The underlying framework is the unobserved components model: $y_t = w_{1t}+\varepsilon_{1t}$, $w_{1t}=w_{1,t-1}+\mu_{1t}$, where ε_{1t} and μ_{1t} are white noise, mutually independent iid processes with zero mean and constant, finite variances. Cointegrating series have the same common component whereas they "differ" by a process that has the same structure as themselves when they noncointegrate. Specifically, we consider the normalized relationship: $y_t = \upsilon + \omega^T z_t + u_t$, for t=1...T, with z_t being an (n-1)-variate I(1) process. If u_t is an I(0) process then $[y_t, z_t^T]$ cointegrate, while an I(1) u_t implies noncointegration. We investigate this via an unobserved components process for u_t: $\phi(L)u_t=\gamma_t+\eta_t$; $\gamma_t=\gamma_{t-1}+\nu_t$, $\gamma_0=0$, where $\phi(L)=1-\sum_{j=1}^{p}\phi j\,L^j$ is a pth order autoregressive polynomial in the lag operator L with roots outside the unit circle. We assume η_t is iid $(0, \sigma^2_\eta)$, ν_t is iid $(0, \sigma^2_\nu)$ and mutually independent of η_t. If $\sigma^2_\nu=0$ then u_t is an I(0) stationary AR(p) process. Alternatively, if $\sigma^2_\nu>0$ then u_t is I(1), with an ARIMA(p,1,1) representation: $\phi(L)(1-L)u_t = (1-\vartheta L)\xi_t$, and $\xi_t \sim$ iid $(0, \sigma^2_\xi)$, $\sigma^2_\xi = \sigma^2_\eta\,\vartheta^{-1}$. The cointegration score-based test statistic is $S = \overline{\sigma}_\eta^{-2}\,T^{-2}e^TVe$ with $\overline{\sigma}_\eta^{-2}$ being a consistent estimator of σ^2_η, V is a T×T matrix whose ij'th element is equal to minimum (i,j), (i,j = 1,2,...,T) , and e is an appropriately defined residual vector. The test statistic has a nonstandard nuisance parameter free limiting null distribution.

To proceed we estimate the normalized regression $y_t = \upsilon + \omega^T z_t + \varphi t + \sum_{j=-s}^{s}\Delta z^T_{t-j}\,\pi_j + u_t$; the augmentation terms include current, lead and lag terms to parametrically allow for serial correlation in u_t, and contemporaneous correlation between z_t and u_t. A linear time trend is included, as then the regression residuals become invariant to Δz_t having a (possibly) non-zero mean. We follow MLS by choosing s to be the highest integer value satisfying that s grows at a rate below $T^{1/3}$; i.e., $s=[T^{1/3}]$. Let $\tilde{u}_t$ be the t'th residual. We then fit an ARIMA(p,1,1) to $\tilde{u}_t$. Let the ARIMA(p,1,1) parameter estimates be $\tilde{\phi}_1,..., \tilde{\phi}_p, \tilde{\vartheta}$. We now form $\tilde{e}_t = \tilde{u}_t - \sum_{i=1}^{p} \tilde{u}_{t-i}$ and $\tilde{\sigma}_\eta^2 = \tilde{\sigma}_\xi^2 \tilde{\vartheta}$, with $\tilde{\sigma}_\xi^2$ being the pseudo MLE of σ_ξ^2. The test statistic is then given by $\tilde{S} = \tilde{\sigma}_\eta^{-2} T^{-2} \tilde{e}^T V \tilde{e}$, which we compare with a 10% critical value; we use the simulated values for the asymptotic distribution provided by MLS.

MLS use a general to specific approach to consistently estimate p; we apply this here. Let $\hat{p}$ be our final estimate of p, ρ be an upper bound such that $\rho > p$, which we use to estimate an ARIMA(ρ,1,1) by pseudo ML. We test the significance of the ρ'th lag using the statistic $z(\rho) = T^{1/2} \bar{\phi}_\rho \bar{\vartheta}$, with $\bar{\phi}_\rho, \bar{\vartheta}$ being the MLE's; $z(\rho) \underset{d}{\rightarrow} N(0,1)$ when $\phi_\rho = 0$. This choice of the test statistic, rather than the usual t-ratio, avoids using computed standard errors, which may be unreliable near or on the boundary of invertibility. If we support $\phi_\rho = 0$, then we estimate an ARIMA(ρ-1,1,1) model and so on. In the event of rejection for $p=\rho-j$, we set $\hat{p} = \rho-j$ and terminate the procedure. Otherwise, we set $\hat{p}=0$.

2.3. Selection of Lag Order

Lag length choice in the VARL, VECM or VARD is important to avoid spurious causality (or spurious absence of causality), and for obtaining an estimate of the cointegrating rank via the JJ approach. We use two goodness-of-fit criteria and one sequential testing procedure based on our survey of the empirical literature. The LR general to specific approach is the most commonly applied procedure in determining the lag order, k, followed by arbitrarily setting k. A group of information criteria (IC) follow next: Akaike's (1973) information criterion (AIC), Schwarz's (1978) Bayesian criterion (SC), and Akaike's (1969) final prediction error criterion (FPE).

The AIC and FPE (which are asymptotically equivalent) do not provide consistent estimators of the lag order (e.g., Nishi, 1988; Lütkepohl, 1993), while the SC is strongly consistent. The FPE and AIC both have a positive probability of overestimating the true lag order, which perhaps suggests that we should not apply these selection procedures. However, as argued by Cheung and Lai (1993) and Gonzalo (1994) for example, overestimation of the lag order may be preferable, given the problems that can occur with wrong inferences about r in an under-specified model. Further, it has been empirically established that the SC produces overly parsimonious estimates in finite samples that may impact on the performance of subsequent hypothesis tests. Given this and the practices of applied researchers, we examine the FPE and SC information criteria in our study.

2.4. Pretest Strategies

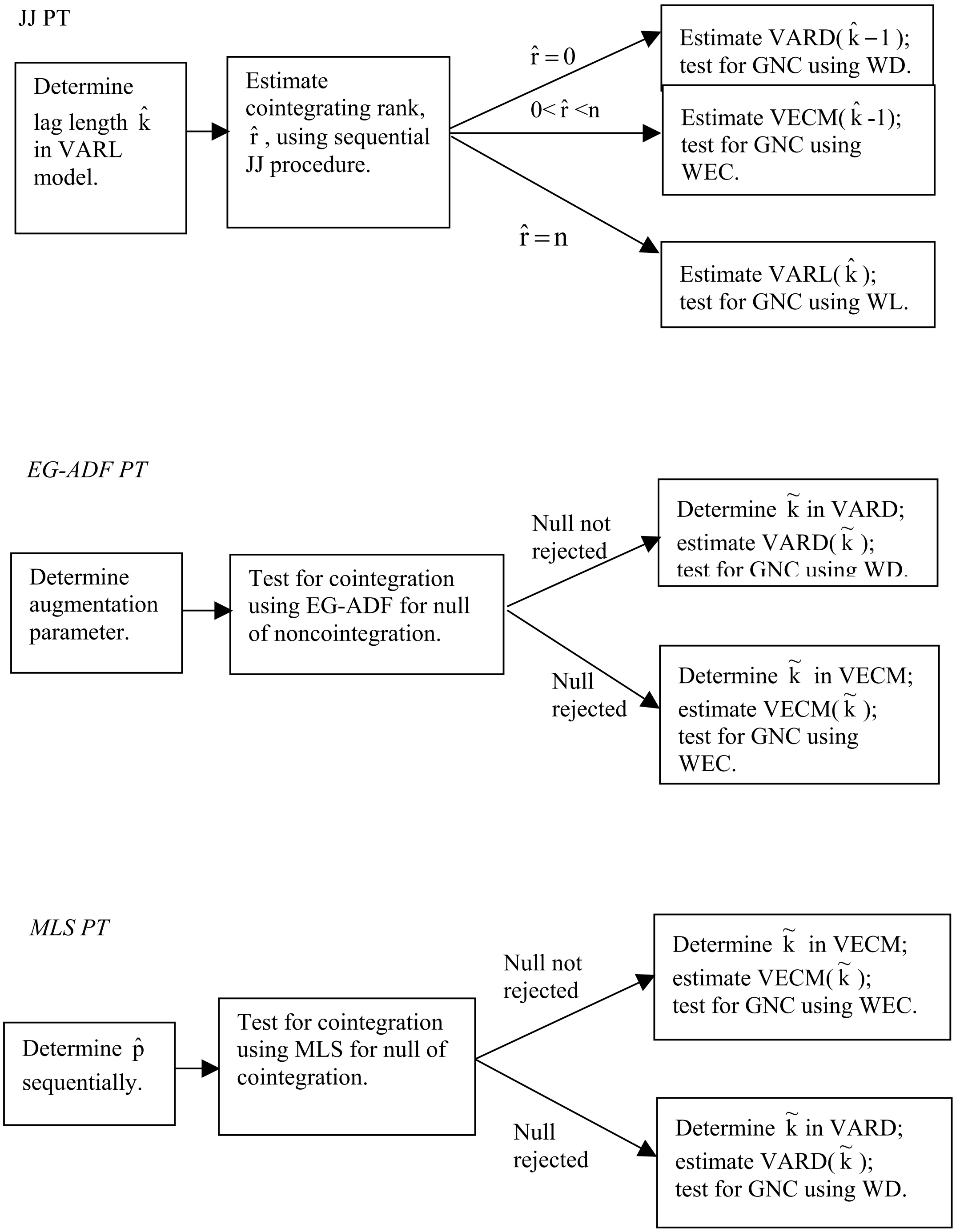
Basically, the pretesting strategies use prior cointegration tests to attempt to verify the validity of the appropriate conditions for the asymptotic null distribution of the Wald statistics presented in section 2.1, with the exception of the WALd where no pre-testing for cointegration is required. We consider three pretest (PT) strategies for testing for GNC based on the three cointegration tests outlined in section 2.2; we denote these as EG-ADF PT, JJ PT and MLS PT. We briefly outline the PT approaches below and schematically show them in Figure 1.

The first step of JJ PT, based on the JJ cointegration test, is to estimate the lag order from the VARL model, denoted $\hat{k}$, using the FPE or SC goodness-of-fit measures, or the LR sequential testing strategy. We next determine the cointegrating rank, denoted $\hat{r}$, within the n-dimensional VARL($\hat{k}$) model using the JJ λ_{max} sequential test. Although the limiting distribution of the JJ statistic does not depend on the lag length, the choice of k may result in a misspecified VARL, which will affect the inferences for r in finite samples. Further, if the errors from the misspecified VARL are no longer iid, then asymptotic inference is also invalid the limiting distribution of the JJ statistic does not depend on the lag length, the choice of k may result in a misspecified VARL, which will affect the inferences for r in finite samples. Further, if the errors from the misspecified VARL are no longer iid, then asymptotic inference is also invalid when based on the usual iid tabulated critical values. Having determined $\hat{k}$ and $\hat{r}$, the final step in the JJ PT strategy is to test for GNC using WD if $\hat{r}$=0, WEC if 0<$\hat{r}$<n, or WL if $\hat{r}$=n.

The EG-ADF PT procedure begins by determining the augmentation parameter for the ADF test on the cointegrating regression residuals using a sequential general-to-specific testing strategy, followed by a test for the null hypothesis of noncointegration. Nonrejection of this null suggests using a VARD model for the I(1) data, from which we estimate the lag order (say $\tilde{k}$) via either the SC, AIC or LR. When we reject the noncointegration null, we use a VECM, denoted as VECM($\tilde{k}$), which implies an underlying VARL($\tilde{k}$+1) DGP. In this case the error correction term is formed from the cointegrating regression residuals. The final task is to test for GNC using either WD for the VARD($\tilde{k}$) model or WEC for the VECM($\tilde{k}$).

The MLS PT strategy is identical to the EG-ADF PT procedure except for the first two steps. Given s=[$T^{1/3}$], we first determine $\hat{p}$ via the general-to-specific approach outlined in section 2.2, and we then use the MLS test to examine the null hypothesis of cointegration. Next, depending on the outcome of this test, we either model the nonstationary data as a VARD or a VECM selecting the lag order by the SC, AIC or LR methods. As with the EG-ADF PT strategy, the last step is to test for GNC using either WD or WEC appropriately.

Figure 1. Schematic outline of the pretest strategies



3. THE MONTE CARLO DESIGN

Our simulation experiment uses a trivariate VARL(3) as its basis: $x_t = \Pi_1 x_{t-1} + \Pi_2 x_{t-2} + \Pi_3 x_{t-3} + \varepsilon_t$, with $x_t^T = [x_{1t}\ x_{2t}\ x_{3t}]$ and Π_i a square matrix of dimension 3. The corresponding VECM(2) is: $\Delta x_t = \Gamma_1 \Delta x_{t-1} + \Gamma_2 \Delta x_{t-2} + \Pi x_{t-1} + \varepsilon_t$. We use the coefficient values in Π_i to set the system's dimension, lag order, integration properties, cointegrating rank, and whether Corollary 1 and Theorem 2 of Toda and Phillips (1993, 1994) are satisfied. Recall that these specify rank conditions under which the GNC Wald statistics WL and WEC are asymptotic χ^2 variates. We consider ten families of DGPs in an attempt to cover a range of processes and to enable comparisons between our results and those published elsewhere in the literature. We provide a brief outline of each DGP family in Table 2, denoted as F1 to F10. For each DGP family we examine four sample sizes (T=50, 100, 200, 400) and three error variance-covariance matrices, denoted S1, S2 and S3; the error variance-covariance choices are considered below.

The bivariate DGP families F1, F6, and F7 have one cointegrating vector with unidirectional Granger causality from x_2 to x_1. The cointegration is "sufficient" in terms of Corollary 1 and Theorem 2 of Toda and Phillips (1993, 1994); this is always so for bivariate cointegrated processes, so that GNC statistics WL and WEC are asymptotic χ^2 variates. We test for GNC from x_2 to x_1 and from x_1 to x_2, denoted respectively as $x_2 \not\rightarrow x_1$ and $x_1 \not\rightarrow x_2$, which enables us to determine rejection frequencies when the GNC null hypothesis is true and false. The dynamic order differs between the families: F1 is a VARL(1), F6 is a VARL(2), and F7 is a VARL(3). There is also cointegration with the bivariate VARL(1) family F5, which differs from F1, F6 and F7 due to there being bidirectional Granger causality.

There are two cointegrating vectors in the trivariate DGP families F8 (a VARL(1)) and F10 (a VARL(2)), and we examine for GNC between x_1 and (x_2 x_3), which ensures satisfaction of Corollary 1 and Theorem 2 of Toda and Phillips (1993, 1994); depending on the causal variable, n_1 (n_3) is either 1 or 2. The causality is unidirectional, so our experiments provide rejection proportions when the null hypothesis is true and false. The DGP family F9 is also a trivariate VARL(1), but with only one cointegrating relationship. We test whether $x_1 \not\rightarrow x_3$ (giving n_1=n_3=1) and $x_1 \not\rightarrow x_3$; the latter hypothesis is false and the former hypothesis is true. We regard this as a worst-case scenario as neither Corollary 1 nor Theorem 2 of Toda and Phillips is satisfied since the cointegration is "insufficient" for the causal null under investigation.

The families F1, F5, F6, F7, F8 and F10 are also used by Zapata and Rambaldi (1997), whose Monte Carlo study compares WALd, JJ PT and Mosconi and Giannini's (1992) LR method of testing for GNC. It is instructive to contrast our results with those of Zapata and Rambaldi as they assume that the lag order is either correctly specified or over/under specified by one lag whereas our lag order is estimated. Toda and Phillips (1994) examine F9 (it corresponds to their N3) in their investigation of testing for GNC using the JJ PT, allowing for tests for "sufficient" cointegration, along with the Wald statistics WD and WL. They also assume that the lag order is either known or overestimated by a fixed order. We compare some of our results with those of Zapata and Rambaldi and Toda and Phillips in section 4.

There is no cointegration present in DGP families F2, F3 and F4. The family F2 involves two independent random walks so the GNC hypothesis is true in both directions. There is also no Granger causality present within F3 or F4: DGP family F3 is a stationary bivariate VARL(2) and a random walk. For these three DGPs we only generate estimates of probabilities of Type I errors. Corollary 1 and Theorem 2 of Toda and Phillips (1993, 1994) are not met for F2 or F4, while Corollary 2 of Toda and Phillips (1993, 1994) applies; WL is not asymptotically χ^2. Appropriate (in terms of standard limit theory) GNC Wald statistics for these two families are WD and WALd, with WAL1 preferred to WAL2 since the data are I(1). Standard distributional results hold for WL with DGP F3, though the near integratedness of the data may cause difficulties. There are no gains in testing with WALd over WL for F3 so this family provides us with some information on the empirical distortions that may occur when the augmentation method is applied unnecessarily. Freeman et al. (1998) also use DGP families F2, F3 and F4 in their limited Monte Carlo study of GNC testing using Phillips' (1995) Fully Modified VAR approach. We term their investigation "limited" with respect to ours as they assume correctly specified DGPs with iid error processes; we relax both assumptions in our experiments. We presuppose that all series are I(1); i.e., we do not undertake preliminary unit root tests (aside from that implicit within the JJ PT strategy with the full rank case). Adding in a level for pretesting for unit roots severely complicates the number of alternative models that would need to be incorporated. Such a task entails its own paper.

The maximum value of the lag order employed to estimate k is 4. We generate 5000 series of (T+100+4) observations for each experiment then we discard the first 100 to remove the effect of the zero starting values that we used; the other 4 observations are needed for lagging. We study three error variance-covariance matrices: S1 is a symmetric matrix with contemporaneous covariance parameter 0.5 across the errors, S2 is an identity matrix, and S3 is such that the errors are independent across equations, but are moving average processes of order one within equations with a moving average parameter equal to 0.8.[3] We employed the following nominal significance

[3] The introduction of moving average errors can potentially alter the causality structure. This is unlikely here due to our assumption of uncorrelated errors across equations. Our DGPs under S3 satisfy the

Table 2. DGP descriptions and null hypotheses examined.

F1: $\Pi_1=\begin{bmatrix}0.75 & 0.5 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{bmatrix}$, $\Pi_2=\Pi_3=0$; r=1; $x_1 \nrightarrow x_2$; $x_2 \rightarrow x_1$. H1 : $x_1 \nrightarrow x_2$; H2 : $x_2 \nrightarrow x_1$.
F2: $\Pi_1=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{bmatrix}$, $\Pi_2=\Pi_3=0$; r=0; $x_1 \nrightarrow x_2$; $x_2 \nrightarrow x_1$. H3 : $x_1 \nrightarrow x_2$; H4 : $x_2 \nrightarrow x_1$.
F3: $\Pi_1=\begin{bmatrix}0.95 & 0 & 0\\ 0 & 0.95 & 0\\ 0 & 0 & 0\end{bmatrix}$, $\Pi_2=\begin{bmatrix}-0.75 & 0 & 0\\ 0 & -0.25 & 0\\ 0 & 0 & 0\end{bmatrix}$, $\Pi_3=0$; r=0; $x_1 \nrightarrow x_2$; $x_2 \nrightarrow x_1$. H5 : $x_1 \nrightarrow x_2$; H6 : $x_2 \nrightarrow x_1$.
F4: $\Pi_1=\begin{bmatrix}0.75 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{bmatrix}$, $\Pi_2=\Pi_3=0$; r=0; $x_1 \nrightarrow x_2$; $x_2 \nrightarrow x_1$. H7 : $x_1 \nrightarrow x_2$; H8 : $x_2 \nrightarrow x_1$.
F5: $\Pi_1=\begin{bmatrix}0.75 & 0.5 & 0\\ 0.4 & 0.2 & 0\\ 0 & 0 & 0\end{bmatrix}$, $\Pi_2=\Pi_3=0$; r=1; $x_2 \leftrightarrow x_1$. H9 : $x_1 \nrightarrow x_2$; H10 : $x_2 \nrightarrow x_1$.
F6: $\Pi_1=\begin{bmatrix}1.5 & 0.5 & 0\\ 0 & 1.25 & 0\\ 0 & 0 & 0\end{bmatrix}$, $\Pi_2=\begin{bmatrix}-0.75 & 0 & 0\\ 0 & -0.25 & 0\\ 0 & 0 & 0\end{bmatrix}$, $\Pi_3=0$; r=1; $x_1 \nrightarrow x_2$; $x_2 \rightarrow x_1$. H11 : $x_1 \nrightarrow x_2$; H12 : $x_2 \nrightarrow x_1$.
F7: $\Pi_1=\begin{bmatrix}1.5 & 0.5 & 0\\ 0 & 1.25 & 0\\ 0 & 0 & 0\end{bmatrix}$, $\Pi_2=\begin{bmatrix}-0.5 & -0.25 & 0\\ 0 & -0.25 & 0\\ 0 & 0 & 0\end{bmatrix}$, $\Pi_3=\begin{bmatrix}-0.25 & 0.25 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}$; r=1; $x_1 \nrightarrow x_2$; $x_2 \rightarrow x_1$. H13 : $x_1 \nrightarrow x_2$; H14 : $x_2 \nrightarrow x_1$.
F8: $\Pi_1=\begin{bmatrix}0.32 & -0.265 & 0.44\\ 0 & 1.2325 & 0.31\\ 0 & -0.285 & 0.62\end{bmatrix}$, $\Pi_2=0$, $\Pi_3=0$; r=2; $x_1 \nrightarrow x_2$; $x_1 \nrightarrow x_3$; $x_2 \rightarrow x_1$; $x_2 \rightarrow x_3$; $x_3 \rightarrow x_1$; $x_3 \rightarrow x_2$. H15 : $x_1 \nrightarrow (x_2, x_3)$ **;** H16 : $(x_2, x_3) \nrightarrow x_1$ **.**
F9: $\Pi_1=\begin{bmatrix}1 & 0 & 0\\ 1 & 0.5 & 0\\ 1 & -0.5 & 1\end{bmatrix}$, $\Pi_2=\Pi_3=0$; r=1; $x_1 \rightarrow x_2$; $x_1 \rightarrow x_3$; $x_2 \nrightarrow x_1$; $x_2 \rightarrow x_3$; $x_3 \nrightarrow x_1$; $x_3 \nrightarrow x_2$. H17 : $x_1 \nrightarrow x_3$ **;** H18 : $x_3 \nrightarrow x_1$ **.**
F10: $\Pi_1=\begin{bmatrix}-0.07 & -0.48 & 0.49\\ 0 & 0.54 & 0.02\\ 0 & 0.02 & 0.69\end{bmatrix}$, $\Pi_2=\begin{bmatrix}0.39 & 0.215 & -0.05\\ 0 & 0.6925 & 0.29\\ 0 & -0.305 & -0.07\end{bmatrix}$, $\Pi_3=0$; r=2; $x_1 \nrightarrow x_2$; $x_1 \nrightarrow x_3$; $x_2 \rightarrow x_1$; $x_2 \rightarrow x_3$; $x_3 \rightarrow x_1$; $x_3 \rightarrow x_2$. H19 : $x_1 \nrightarrow (x_2, x_3)$; H20 : $(x_2, x_3) \nrightarrow x_1$.

---

noncausality Theorems 3 and 4 of Boudjellaba et al. (1994) but these theorems require invertible ARIMA processes, which precludes cointegrating relations. It would seem highly likely that the theorems would extend in this case, but the proofs are as yet unavailable; see Boudjellaba et al. (1994, p.278) for discussion.

levels: 5% for the GNC test, 10% for the cointegration tests, 1% for the general-to-specific LR based lag selection, 5% for the general-to-specific method of choosing the EG-ADF augmentation parameter, and 5% for the standard normal test used to estimate p in the MLS PT procedure. We realize the potential impact of our choices; it remains to explore the extent of this.

4. SIMULATION RESULTS

It is a challenge to present the results from our extensive set of experiments in a way that is both compact and useful. The usual tabular approach, aside from taking up over ten pages, does not allow much information to be conveyed or assimilated. Accordingly we use several different graphical methods that communicate the relevant information compactly and that enable the impact of changes to estimated rejection proportions from the sample size, the covariance matrix, testing methods, and other factors to be easily noted[4].

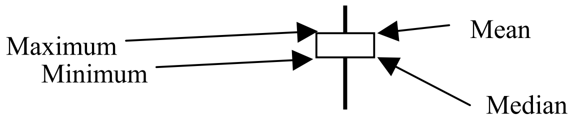
In section 4.1 we provide summary information on estimated Type I error probabilities. This discussion highlights the difficulties that can occur with the PT methods. Accordingly, we follow, in section 4.2, with a more detailed examination of the PT methods. We show that problems in GNC testing usually arise when the cointegration pretest does not work well. In section 4.3 we compare empirical powers and we examine, in section 4.4, the importance of estimating the lag order, rather than pre-specifying it, in Monte Carlo experiments.

4.1 Summary Estimated Type I Error Probabilities

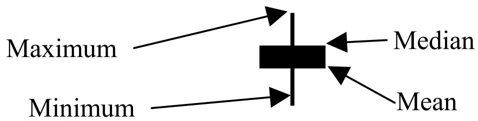
In this section we identify the tendency of the various GNC testing methods to systematically over-reject or under-reject over the range of cases examined that satisfy the GNC null and how this is influenced by sample size, covariance matrix and lag selection tool. We use a graphical tool, which we call a "4M Chart", to provide the information; "4M" denotes "Maximum", "Minimum", "Mean" and "Median", formed as statistics from the twelve cases for which the GNC null hypothesis is true. That is, of the twenty null hypotheses examined for our ten DGPs, as detailed in Table 1, twelve of these nulls are true: H1, H3, H4, H5, H6, H7, H8, H11, H13, H15, H18, H19. A 4M Chart is comprised of 4M bars, each bar provides the minimum, maximum, median and mean estimated Type I error probabilities, which we denote as ETIE, of the twelve cases that satisfy the GNC null hypothesis, with testing method, covariance matrix, lag selection tool and sample size fixed for a particular bar. A 4M bar takes one of two forms, depending on whether the mean ETIE is less than or greater than the median ETIE. When the

[4] The usual tabulated results are available from the first author.

mean is greater than the median the bar is of the form:



Alternatively, when the mean is less than the median, the bar's form is:



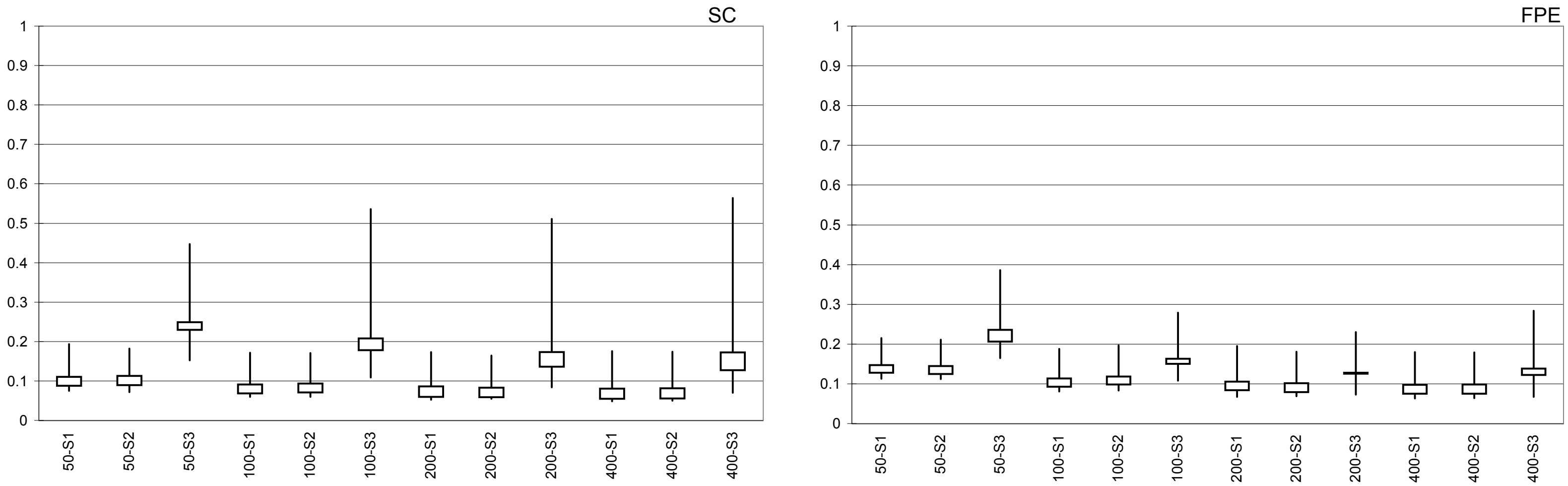
The range of ETIEs can be quickly gleaned from the bars, in addition to the tendency of a GNC testing strategy to over- or under-reject. At the same time, it is straightforward to ascertain the impact of sample size, covariance matrix and lag selection tool on the rejection proportions as many bars can be presented in one chart. The 4M charts are shown in Figure 2 in panels labelled A to F, with each panel associated with one of the six GNC testing procedures: WL, WAL1, WAL2, EG-ADF PT, MLS PT and JJ PT. Each panel provides a 4M Chart for SC and FPE; the charts for LR are similar to those for FPE and are available on request. Rejections proportions are provided on the "y-axis" and the "x-axis" denotes the sample size and covariance matrix used in that particular set of simulations; e.g., "100-S3" signifies that the 4M bar arises from the twelve cases when the sample size is 100 and the S3 covariance matrix is used to generate the errors. The nominal significance level of the GNC test is 0.05 and the standard error associated with the frequencies is approximately 0.003, though it can be as high as 0.009.

The following comments follow from the 4M charts given in Figure 2:

1. The WALd results in the narrowest range of ETIEs, irrespective of lag selection method, followed by WL. The PT methods can obtain ETIEs approaching one; that is, the GNC null will always be rejected even when it is true, which suggests a preference for WALd over the PT methods when interest lies only on ensuring consistent support for a true GNC hypothesis.
2. All methods over-reject, though sample size can assist to mitigate this effect.
3. The rejection proportions are similar for covariance matrices S1 and S2, while the presence of moving average errors usually results in greater over-rejection.
4. Typically, the addition of the extra lag for WAL2 from WAL1 results in little change in the ETIEs, though the differences are more noticeable for T=50. The addition of the extra lag can be helpful in reducing the degree of over-rejection with moving average errors.
5. The minimum and median ETIEs, compared to the maximum and mean ETIEs, for the PT approaches show that for at least half of the cases examined the over-rejection problem is not

Figure 2. 4M Charts

Panel A. WL



Panel B. WAL1

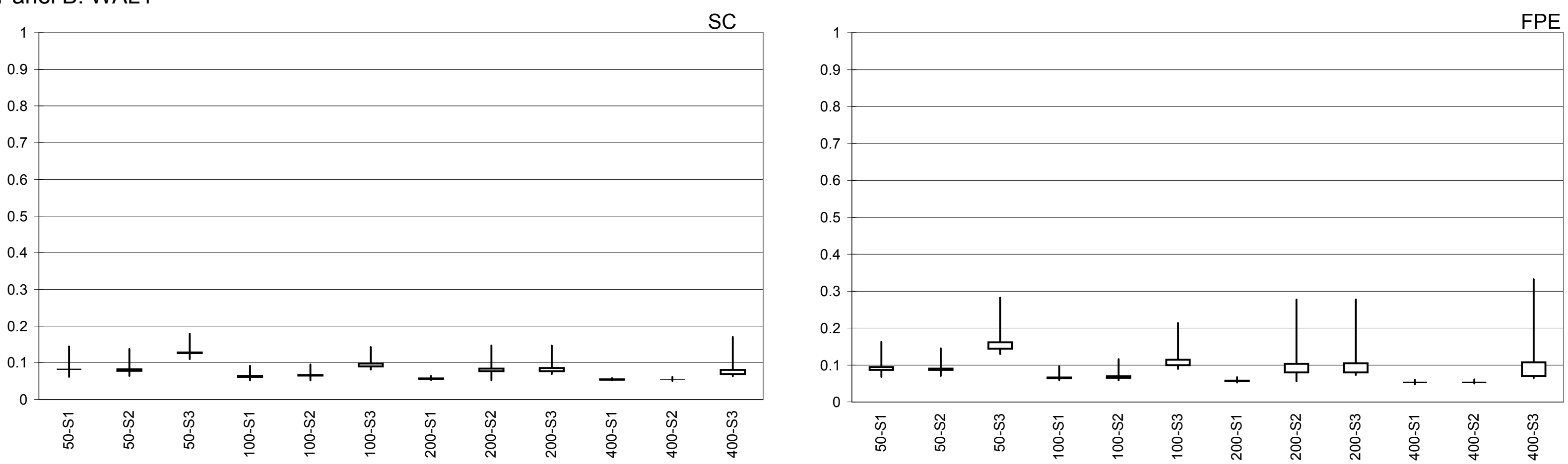
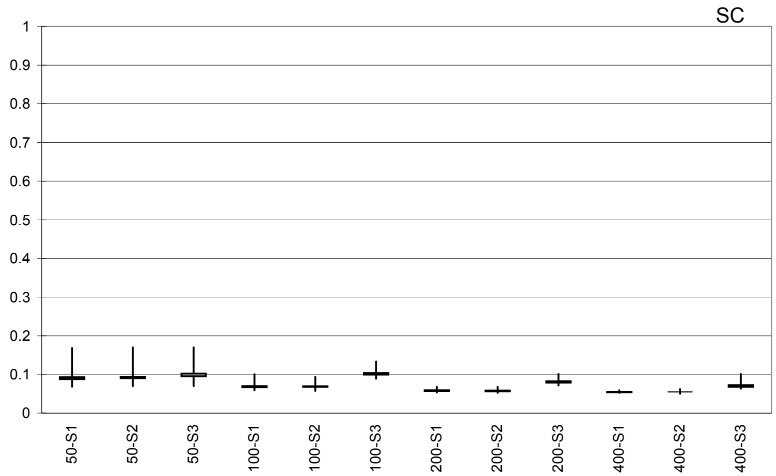
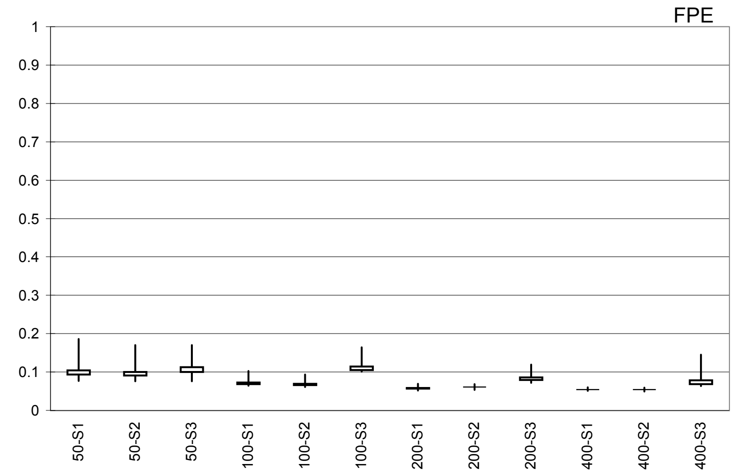


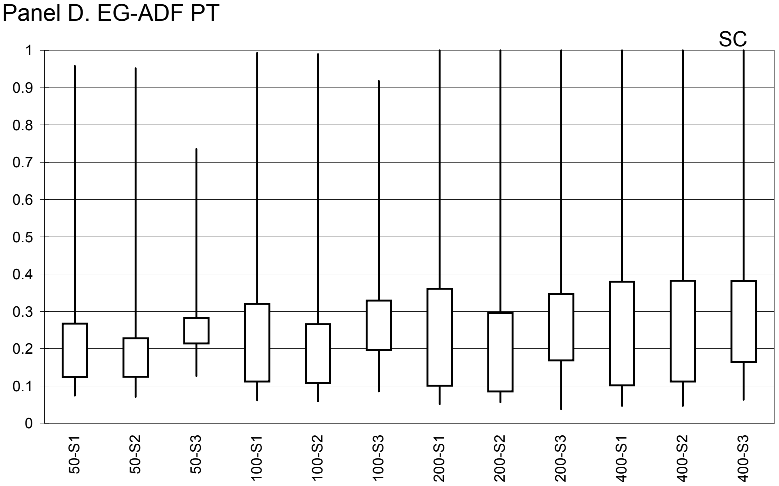
Figure 2 (ctd). 4M Charts

Panel C. WAL2





Panel D. EG-ADF PT



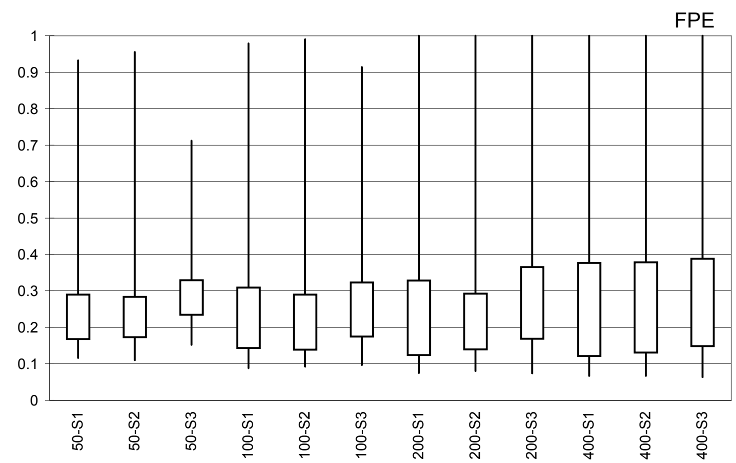
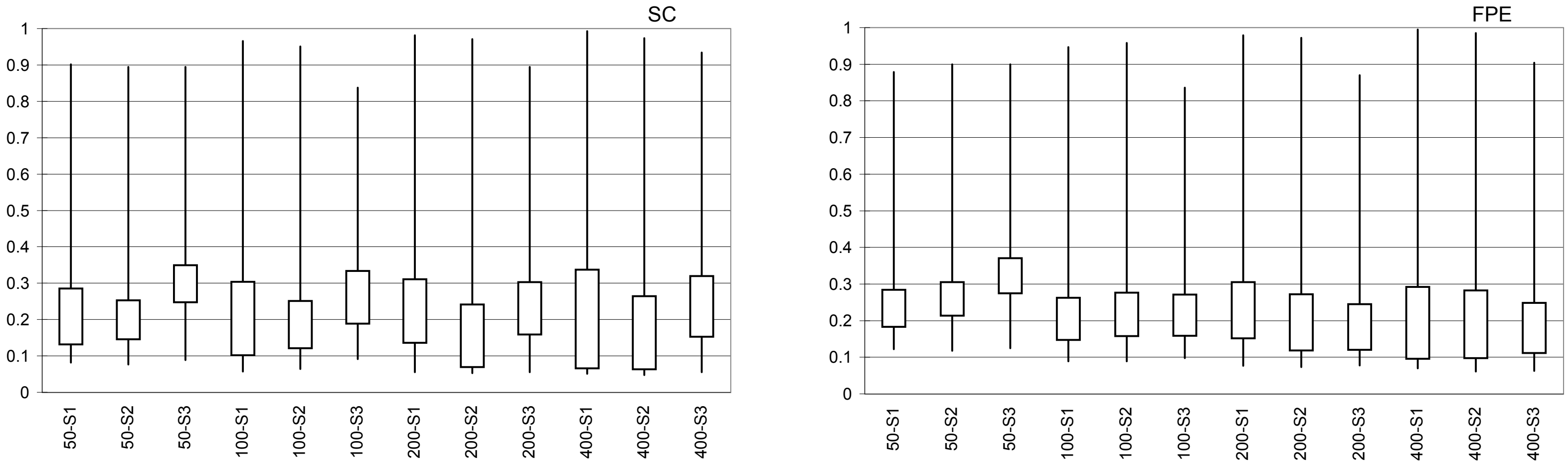
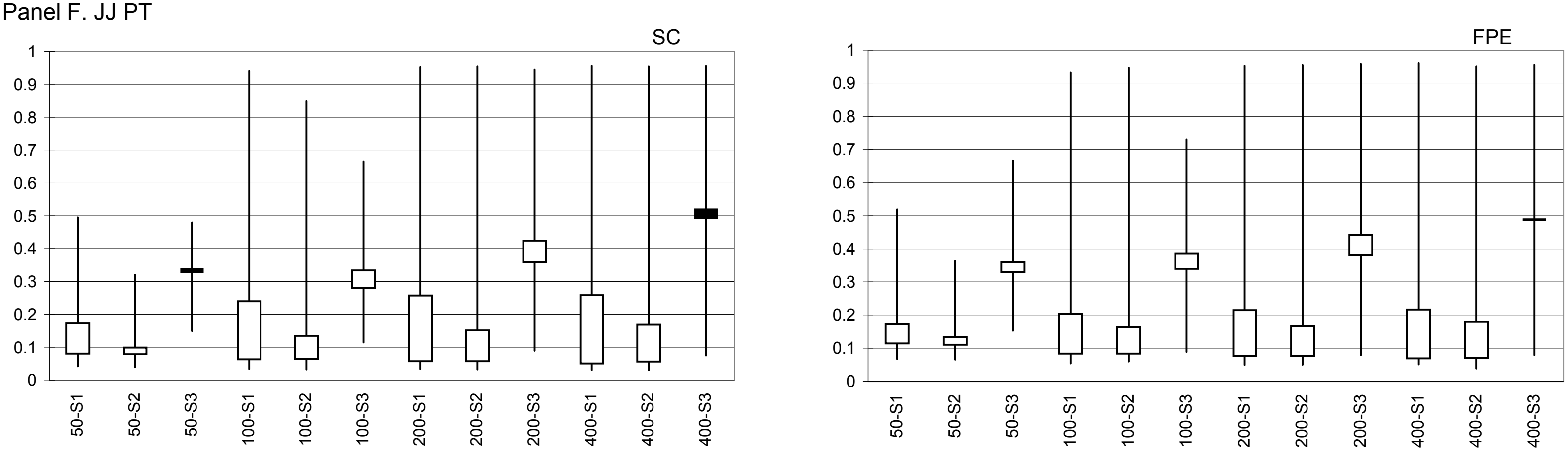


Figure 2 (ctd). 4M Charts

Panel E. MLS PT



Panel F. JJ PT



too serious and is comparable in magnitude to that for WL, WAL1 and WAL2, at least when the errors are not serially correlated. This suggests that the form of the DGP is far more crucial for ensuring control of the Type I error probability for the PT methods than for WL, WAL1 or WAL2. We explore this issue further in section 4.2.

6. At least for WAL1 and WAL2, our results tend to support use of the SC criterion for selecting the lag order over the FPE (and LR) method.
7. The spread between the mean and median ETIEs is typically narrower for the FPE than the SC, for the PT methods and the WL approach, especially with moving average errors.

The 4M Charts show the consistent performance of the WALd method in terms of ETIEs over the PT approaches across the wide range of DGPs, covariance matrices, lag selection methods and sample sizes that we examined. The charts also show that the problems with the PT methods are DGP dependent, as the ETIEs for some DGPs are comparable, and may indeed be preferable, to those of WAL1 and WAL2. We explore this issue in the next section.

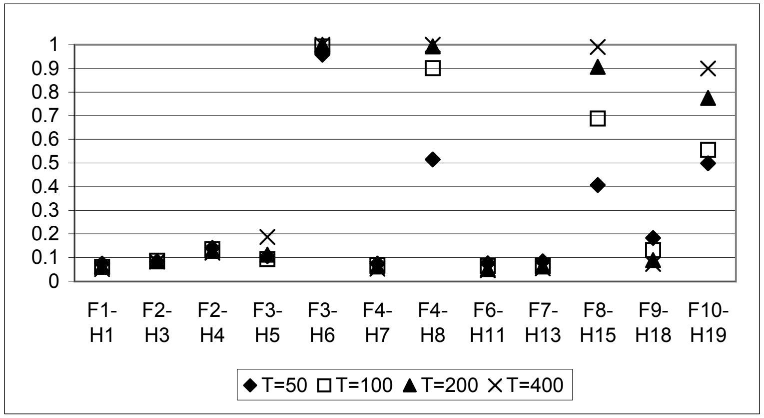
4.2 More on Estimated Type I Error Probabilities for the PT Methods

We illustrate the impact of DGPs on the ETIEs for the PT methods in Figure 3, which provides scatter plots of ETIEs, when T=50, 100, 200 and 400, for EG-ADF PT, MLS PT and JJ PT when S1 is used to generate the errors and SC is the chosen lag order tool; the plots for the other covariance matrices and lag selection tools are qualitatively similar. We include a scatter plot for WAL1 for comparative purposes. The "x-axis" for each panel in Figure 3 is the DGP family and null hypothesis under test; e.g., "F3-H6" denotes DGP F3 when null hypothesis H6 is examined. The "y-axis" in each case is the estimated Type I error probability.

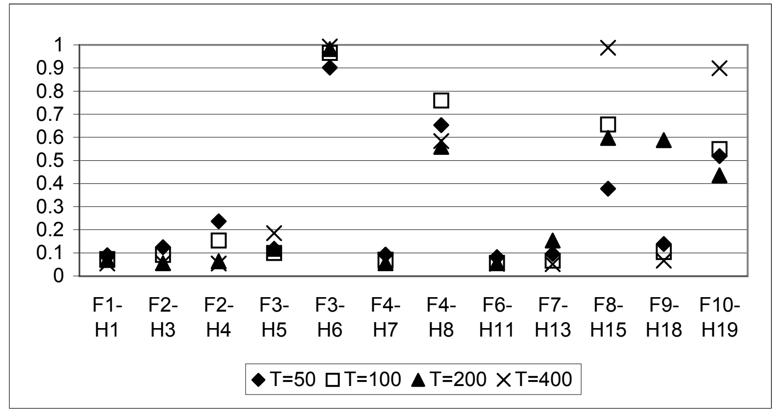
One obvious feature is that F4-H8 and F10-H19 are "problem" cases for all three PT methods; rejection probabilities are over 0.40 irrespective of sample size. The residual based PT methods also have difficulties with F3-H6 and F8-H15, but not so JJ PT. Conversely, JJ PT suffers serious over-rejection for F7-H13 while this case is not of concern for EG-ADF or MLS-PT. Serious over-rejection does not occur for the other null hypotheses; for these cases, often JJ PT results in a Type I error that is closer to 0.05 than for WAL1 or WAL2, but this does not occur for either of the residual based PT procedures. The "problem" DGPS are quite different in nature; e.g., DGP F4 is comprised of a stationary series and a unit root process, while DGP F10 has two cointegrating vectors and "sufficient" cointegration in the sense of Toda and Phillips (1993, 1994). However, the common thread is that we observe an over-rejection of the GNC null hypothesis for the PT methods when there is concurrently a higher than expected detection of one or more

Figure 3. PT & WAL1 ETIEs for S1 and SC

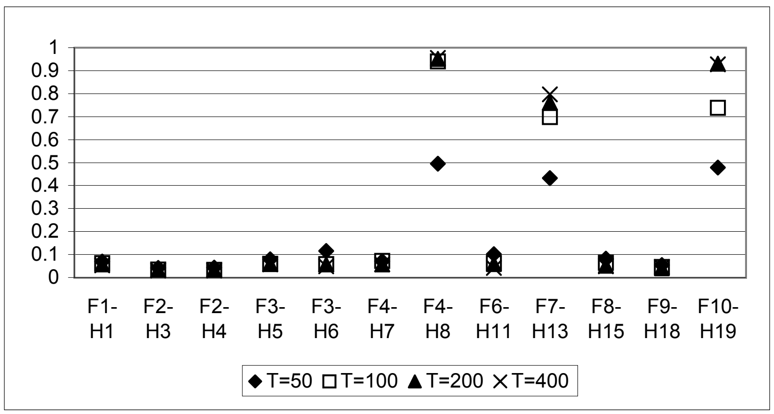
Panel A. EG-ADF PT



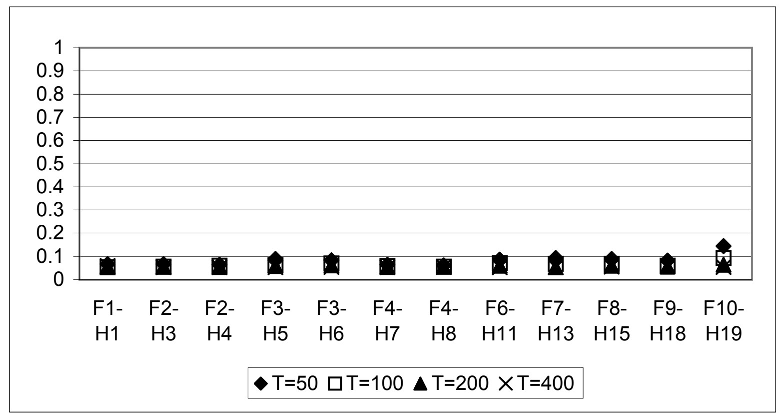
Panel B. MLS PT



Panel C. JJ PT



Panel D. WAL1



cointegrating relationships. This is inevitable, as cointegration implies Granger causality.

We illustrate this problem in Figure 4, where we provide stacked bar charts that compare the estimated Type 1 error probabilities with the corresponding proportion of trials for which an outcome of cointegration is found. We examine the same cases presented in Figure 3, limiting attention to T=100 as qualitatively similar figures result for the other sample sizes, covariance matrices and lag selection methods. The lower part of each bar, shaded as , provides the ETIEs, which ideally would be close to 0.05, and the upper part of each bar, shaded as reports the proportion of trials for which cointegration is detected, denoted as CF for "cointegration frequencies". There is no cointegration with DGPs F2, F3 and F4, yet the residual based tests often detect such for F3 and F4 with a consequential finding of at least unidirectional Granger causality, while JJ PT does not have trouble with F3 but does so with F4; recall that F3 is a near integrated case and F4 consists of a stationary series and an integrated process. These results highlight the importance for noncausality testing of accurately determining the cointegrating rank.

The DGP F8 is quite different. Then, the pretest methods do well at detecting cointegration, with the JJ method reasonably accurate at detecting two cointegrating vectors. However, the residual based PT approaches suffer from extreme over-rejection of the GNC null due to their limitation of only being able to detect one cointegrating vector. The resulting omission of a relevant error correction term in the VECM leads to lack of control of the GNC Type I error probability. This suggests that residual based approaches should not be used to form VECMs for GNC testing when there is a likelihood of more than one cointegrating relationship.

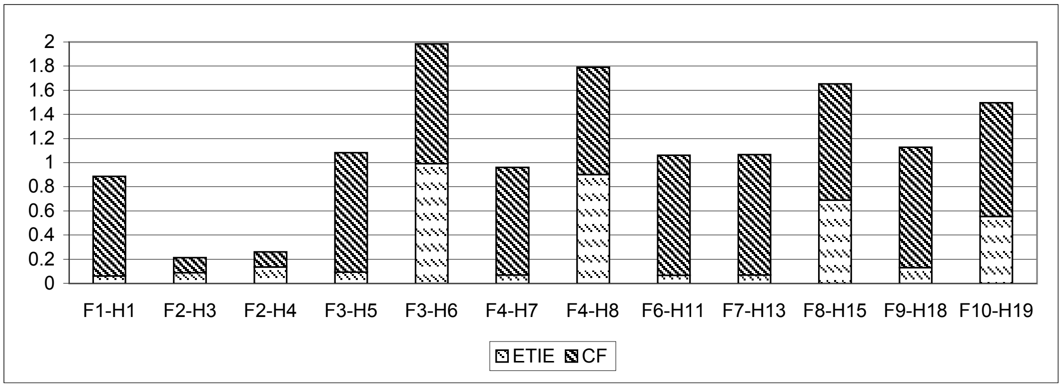
The residual based tests outperform JJ PT, on the other hand, for DGP F7 for H13; this DGP is a bivariate VAR(3) with one cointegrating relationship. Though all PT methods do well at accurately detecting cointegration, JJ PT more often than not finds bidrectional causality rather than the actual unidirectional causality – this is not a problem for the residual based PT methods.

The DGP F10 is an enigma in some sense. Despite there being two cointegrating relationships, JJ PT does not detect cointegration, while the residual based tests do reasonably well at finding cointegration. Nevertheless, the PT methods strongly reject both H19 and H20 when H19 is true. We suspect that the problem is one of complexity – the PT methods find it difficult to accurately estimate the cointegrating vectors and Granger causal relationships in higher dimensional systems. We conjecture poor performance for the PT methods in systems comprised of four or more variables, though this remains to be verified in future research.

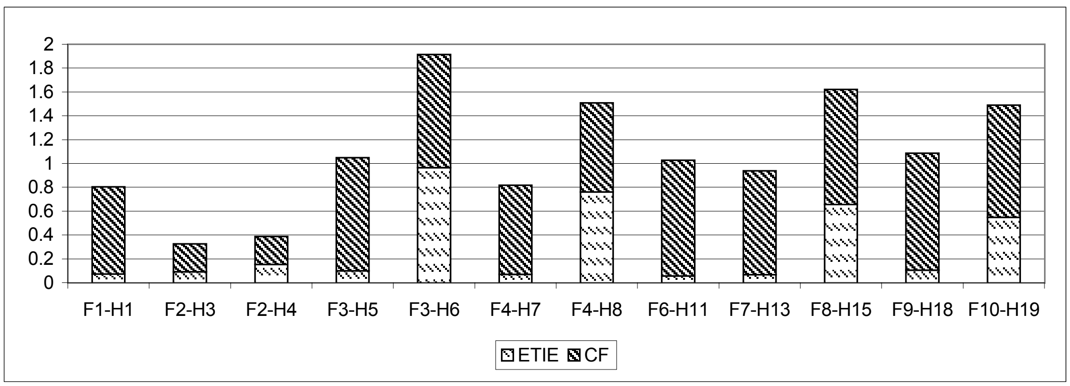
Our examination of the PT methods in this section provides evidence that the key to adequate control of the GNC Type I error probability is accurate assessment of the cointegrating

Figure 4. PT Stacked Bar Chart of ETIEs & CFs for S1 and SC when T=100

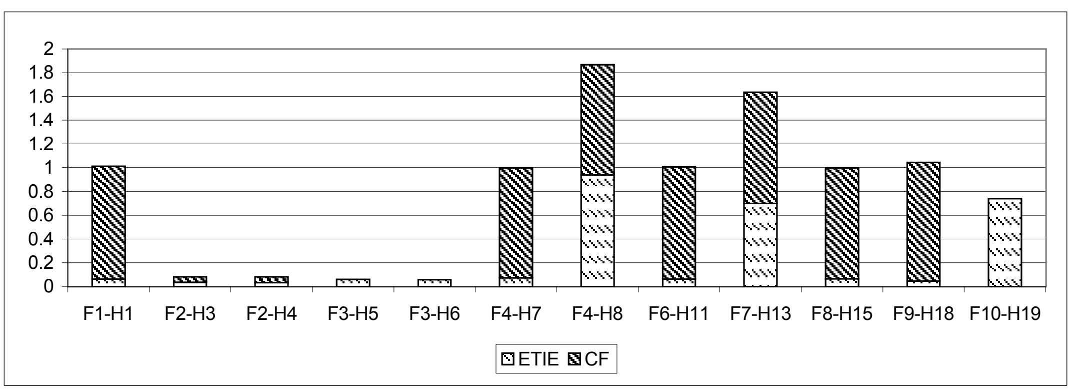
Panel A. EG-ADF PT



Panel B. MLS PT



Panel C. JJ PT



rank, known to often be a difficulty. We now turn our attention to power comparisons.

4.3 Empirical Power Comparisons

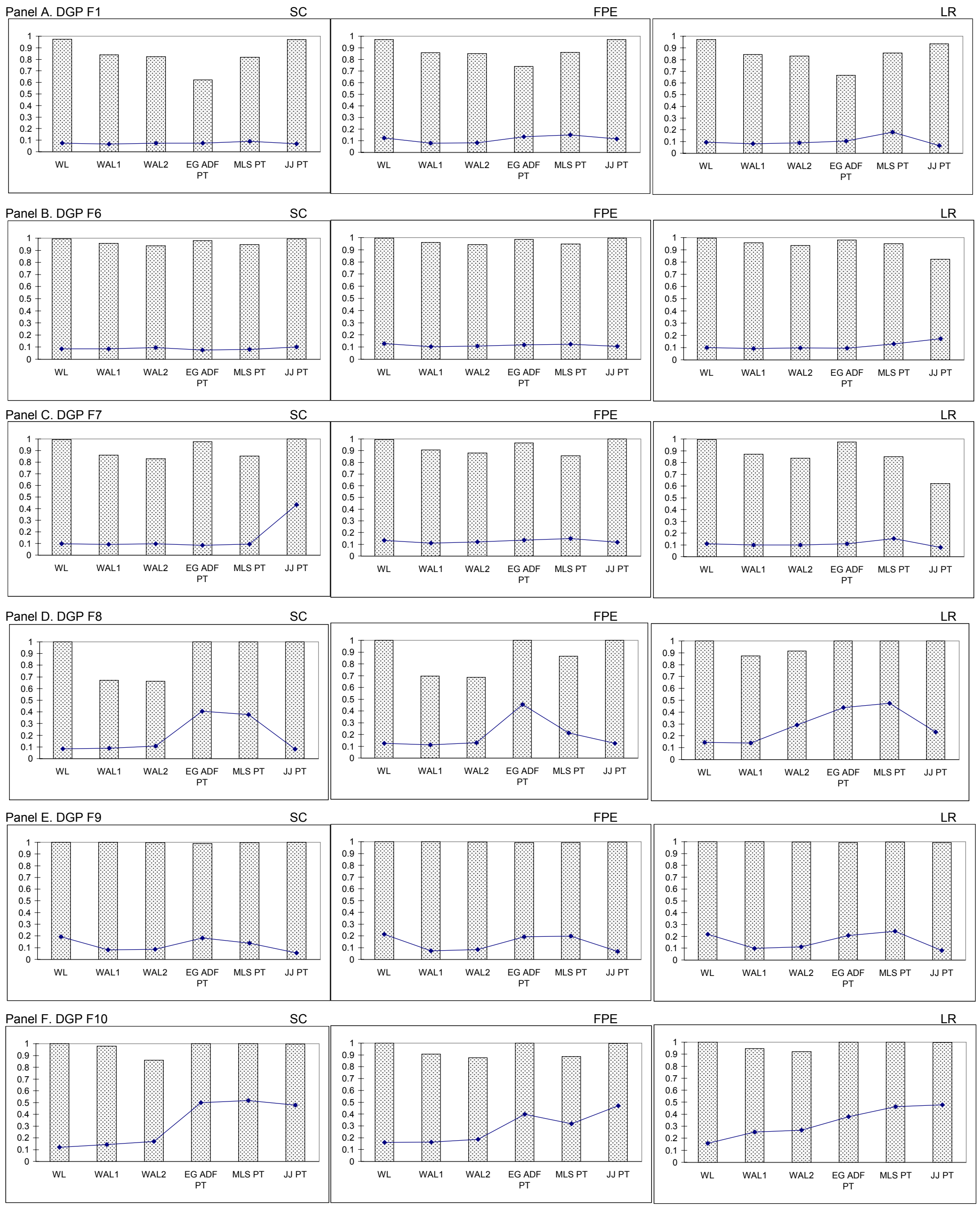
Our experiments provide empirical "powers" - the proportion of trials for which we reject a false null hypothesis – when testing hypotheses H2, H9, H10, H12, H14, H16, H17 and H20. The estimated "powers", denoted by EP, are size uncorrected. We are aware of the debate on reporting such estimates, but for many of our cases there is often strict dominance; i.e., a procedure which has smaller empirical probability of a Type I error and higher empirical 'power'. We believe that for the other cases, the results are informative enough to enable us to offer advice to practitioners, without the need for providing size corrected power information, a task that would be highly difficult here given the number of DGPs that satisfy the GNC null.

Figure 5 shows EPs for each GNC testing strategy for T=50 when S1 is the error covariance matrix. Though quantitatively different, the figures for S2 and S3 are qualitatively similar. As our results suggest little difference in EPs when T=100 and no significant difference for the larger sizes, we limit attention to the smaller sample size. The figure is divided into six panels – one for each DGP – and each panel has three bar charts – one for each lag selection method. The height of the bars in each chart provides EPs; the graphical representation, rather than the usual table format, makes it far easier to compare across procedures, DGPs and lag order criterion. Each chart also provides the associated ETIEs, which is useful when comparing EPs, given that these are not size corrected. The "y-axis" in each case is the estimated proportion of trials for which we reject the relevant GNC null hypothesis and the "x-axis" provides the testing method.

The following points arise from the charts in Figure 5:

1. WL is, as expected, preferred when the rank conditions hold to ensure its standard limiting null distribution. Though the unadjusted EP is comparable when the rank conditions fail (see DGP F9), Type I error problems suggest that WL would then be dominated by WALd and JJ PT.
2. There is an anticipated loss in EP from using WALd over WL when the latter's null limiting distribution is indeed χ^2 (F1, F6, F7, F8, F10); e.g., the reduction in EP is approximately 17% for F1 while up to 52% for F8. The loss in power reduces significantly with increases in the true lag order when T=50 and there is minimal difference for samples greater than 50. Typically, the addition of the extra lag for WAL2 from WAL1 results in little change in the estimated rejection proportions, though it can be helpful with moving average errors.
3. Rarely does MLS PT and EG-ADF PT result in a higher EP than WAL1 or WAL2, especially once allowance is made for differences in Type I error probabilities.

Figure 5. Empirical Powers for S1 and T=50



4. EP for the JJ PT method is typically higher than that for WALd when T=50, though the differences disappear for higher sample sizes. However, this advantage must be weighed against the issues that can occur with Type I error probabilities with JJ PT.
5. The lag order method adopted results in some quantitative differences in EP but similar qualitative comparisons occur. Interestingly, the EPs for JJ PT are more sensitive to the lag selection criteria than are the other approaches; e.g., with LR, JJ PT usually has the lowest EP but this is rarely so with SC and FPE, while, allowing for Type I error probability differences, the EP for the other methods varies less with lag method.

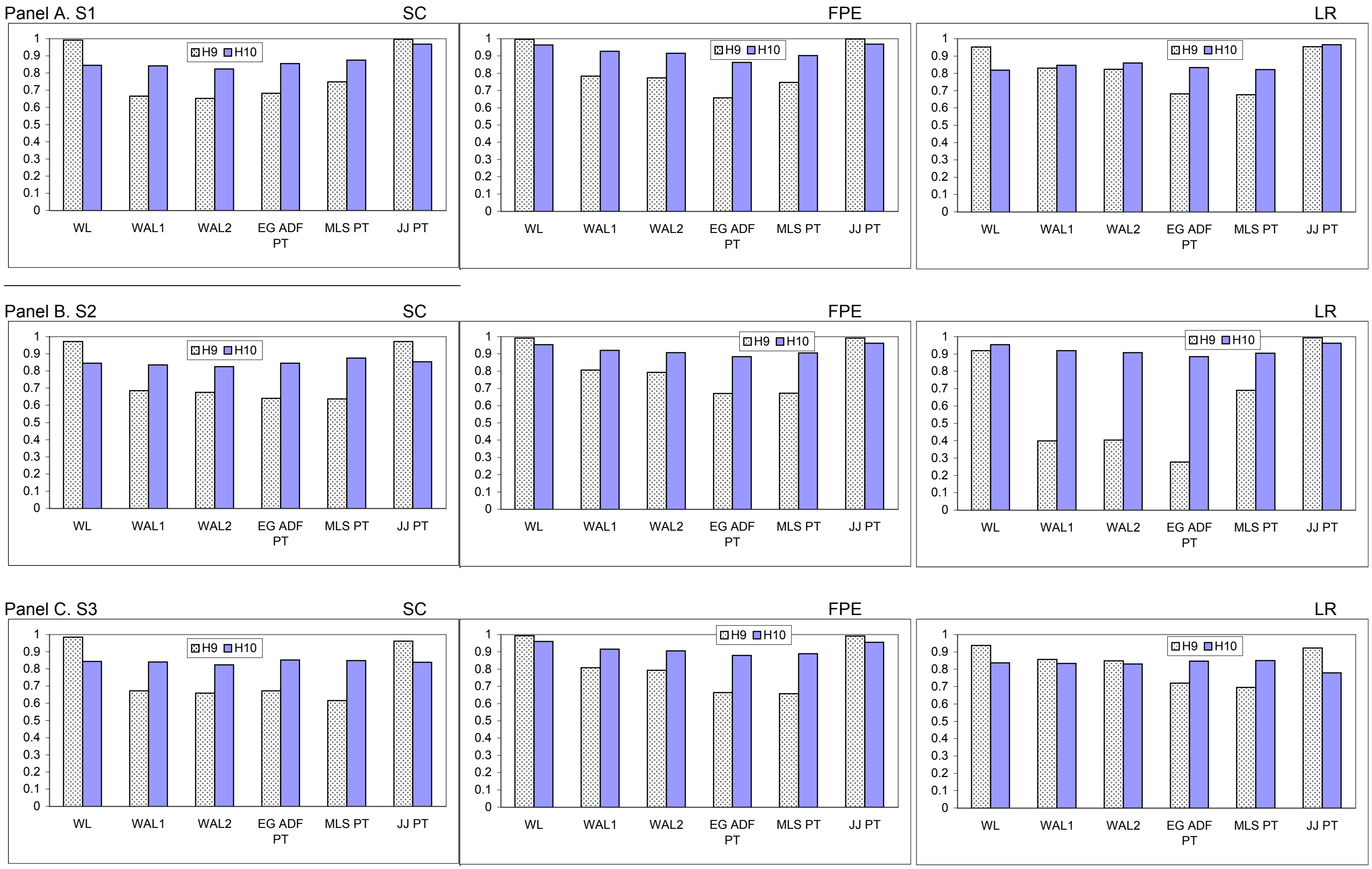
The DGP F5 has bidirectional causality so that power is only of interest. Figure 6 provides bar charts of the EPs that result after testing H9 and H10 when T=50. The figure is divided into three panels – one for each error covariance matrix; each panel reports three charts – one for each lag approach. The JJ PT procedure outperforms the other methods when there is bidirectional causality. The loss in EP for WALd over WL, which is valid for this DGP, and JJ PT can be quite substantial in small samples, but the differences in EP are minor (typically) when $T \geq 200$. The residual based PT approaches rarely do well, usually being outperformed by WALd, JJ PT and WL. There are only minor differences in EP across the lag methods, for a given error variance-covariance matrix, though LR is less consistent across the three covariance matrices.

4.4 Impact of Estimating the Lag Order on the Monte Carlo Results

The accuracy of lag selection methods has been well explored; we can add little to this debate. Of importance is whether we gain in our learning about the properties of GNC statistics by substantially expanding the Monte Carlo experiments to incorporate lag selection, as opposed to assuming a correctly specified structure or a VAR process that is always over/under specified.

We address this by comparing some of our results with those available from studies that assume a given lag structure; e.g., Toda and Phillips (1994), Dolado and Lütkepohl (1996), Zapata and Rambaldi (1997), and Yamada and Toda (1998). As we described in section 3, several of our DGP families match those used by Zapata and Rambaldi (1997), who assume that the lag-order is "true", or "under"-specified by one, or "over"-specified by one and examine (in our terminology) WAL1 and JJ PT. Comparing our results to those reported by Zapata and Rambaldi (1997) we find that the cases of "correct", "under", and "over" often do not reflect the frequencies found when the lag structure is estimated, especially for small sample sizes (e.g., T=50). For instance, for our DGP F8, which corresponds to DGP5 of Zapata and Rambaldi (1997), we find that their

Figure 6. Empirical Powers for H9 and H10 (DGP F5), T=50



“over” frequencies substantially overstate the distortion in the estimated Type I error probability for WAL1 when T=50 with the lag dynamics estimated using SC or FPE; our results correspond more closely to those reported by Zapata and Rambaldi for “true”. However, our LR results for this example are more similar to Zapata and Rambaldi’s “over” frequencies.

In contrast, for the same DGP, their “true” frequencies for the JJ PT method understate the distortion we observe when the lag-order is estimated by SC and FPE; our results are closer to their “over” outcome for this comparison. Moreover, our LR results with JJ PT exhibit much greater distortion than suggested by even their “over” case.

These comparative results need not hold for the other DGP families. The DGP family F7, which is Zapata and Rambaldi’s DGP4, provides another example. Comparing results, their “under” and “true” cases reflect the size distortion outcomes likely when estimating the lag order when using WAL1 with T≤100. However, their “under” frequencies substantially underestimate the uncorrected powers when the lag structure is estimated; these empirical powers are closer to those observed for “true” or “over”, depending on lag selection method.

Toda and Phillips (1994, Table III, p276), who assume that the lag-order is either “true” or “over” by one or three, provide rejection proportions comparable to those for our DGP F9. We find that the rejection proportions for their pre-specified lag results are typically smaller than ours.

Another relevant analysis is that of Yamada and Toda (1998), who compare the empirical “size” and “size-unadjusted” power of, amongst others, the WAL1 and JJ PT (with the additional P1 test of Toda and Phillips, 1994) methods, assuming a known lag length for a bivariate system. They limit attention to an identity matrix error variance-covariance as they show that the finite sample distributions of their Wald test statistics for any choice of innovation variance-covariance matrix are equivalent to those of a transformed model with an identity covariance matrix, when the lag structure is known. However, this no longer follows once the lag dynamics is estimated; our results illustrate that the form of the error variance-covariance matrix can substantially change the finite sample performance of the statistics and the preference for one method over another.

In summary, the finite sample properties of GNC statistics can be markedly different when we estimate the lag structure than when the dynamics is pre-specified, either correctly or incorrectly, so that the additional effort in coding for the Monte Carlo experiments is worthwhile in order learn more about the procedures applied by practitioners.

5. AN EMPIRICAL APPLICATION

Many have examined for the existence of a long-run stable relationship between real money balances (*mb*), real income (*gdp*) and nominal interest rates (*ir*). The results generally provide support for the stationarity of a money demand relation, which implies Granger causality in at least one direction. We examine these three time series to test for GNC using each procedure and lag selection method examined in our Monte Carlo study. We then discuss the outcomes in light of our simulation experiments.

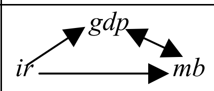
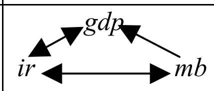
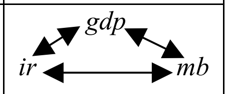
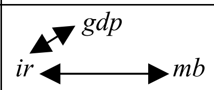
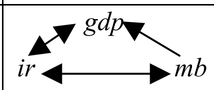
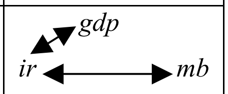
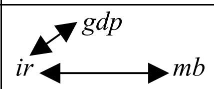
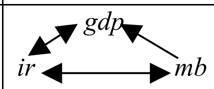
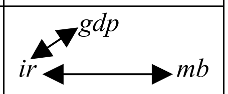
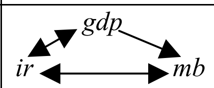
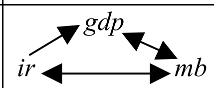
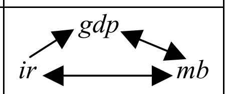
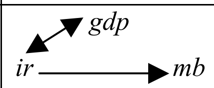
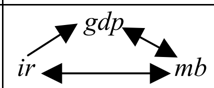
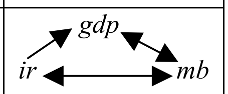
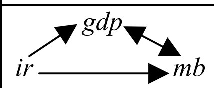
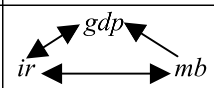
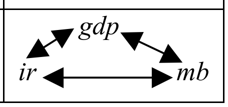
We use Hoffman and Rasche's 1996 data for our study downloaded from the *Journal of Applied Econometrics* Data Archive. The data are quarterly, seasonally adjusted U.S. time series originally obtained from the Citibase data set. Real money balances are calculated by deflating the nominal series by the GDP deflator. Both real balances and real GDP are expressed as natural logarithms, and the Treasury bill rate is used as the interest rate variable. Allowing for appropriate lagging, estimation for all equations is undertaken over the period 1954:1 to 1994:4 (164 observations). We considered up to ten lags when choosing the systems lag order with the lag order selected from a VARL model for WAL1, WAL2, JJ PT and WL, while we employ the VECM or VARD model with appropriate residual vectors for the EG-ADF PT and MLS PT procedures. In the VARL model the SC criterion results in a parsimonious two lags while the FPE and LR criteria lead to six and eight lags respectively. Cointegration is detected from the EG-ADF test with a VECM(2) being selected by SC, and a VECM(8) is selected by both the FPE and LR. Conversely, noncointegration is suggested by the MLS test with lag orders for the resulting VARD model being 2, 8 and 7 for the SC, FPE and LR respectively.

We follow Hoffman and Rasche's (1996) analysis, and assume that each of our time series is integrated of order one with no deterministic trends. The JJ λ_{max} test supports one cointegrating vector, irrespective of lag method. This aligns with the research on this question. Regarding cointegration via the EG-ADF method we required one augmentation term in the auxiliary regression, with resulting support for the existence of a stable relationship. We found no support for cointegration from the MLS test.

Our final task is to undertake appropriate GNC Wald tests for each of our models and lag selection methods; we use a nominal 10% level of significance. We report the results in Table 3 using Granger causal maps. The causal map is a "directed graph", with an arrow that leads from one variable to another indicating support for Granger causality; e.g., $mb \rightarrow gdp$ implies that *mb* Granger causes *gdp*. We interpret these maps in light of our simulation experiments.

As presented, the maps show no unanimity between lag methods and procedures, except for the question of GNC from *ir* to *gdp* and *ir* to *gdp*. We consider first the causality direction between *gdp* and *ir*. All methods support Granger causality from *ir* to *gdp* and, given our

Table 3. Granger causal maps

Lag/Procedure	SC	FPE	LR
WL			
WAL1			
WAL2			
EG-ADF PT			
MLS PT			
JJ PT			

simulation experiments, it seems reasonable to assume that this exists. The sample values for WAL1 and WAL2, and (applying a relatively liberal significance level) EG-ADF PT and MLS PT, support bidirectional causality here; i.e., the methods also indicate causality from *gdp* to *ir.* The JJ PT method supports bidirectional causality when using the FPE and LR method, but not with SC. When there is bidirectional causality, our simulations suggested that JJ PT is most likely to detect it (size control is not of issue here), and so the JJ PT results we have here are consistent with these observations, at least for the FPE and LR methods. What about SC? The outcome for this selection criterion probably reflects the known slow rate at which the SC moves away from the low lag orders. Further, as the probability of overfitting using FPE decreases exponentially with the number of freely estimated parameters in the system, it would seem that the FPE (and so also LR here) results are more likely than the SC. That is, the SC is likely to be underestimating the lag order, and may not be detecting causality from the JJ PT method, when in fact it exists. Consequently we can reconcile the observed differences and support bidirectional causality between *ir* and *gdp*, using the results from our simulations and those from other studies on lag order performance.

Turning now to GNC between *gdp* and *mb*, the results typically suggest Granger causality from *mb* to *gdp* with FPE and LR, but only WL and JJ PT support causality when using SC. There is no support for GNC from *gdp* to *mb,* using WAL1 and WAL2. The statistic WL with SC and LR, but not with FPE, implies Granger causality. The pretest approach JJ PT indicates bidirectional causality with SC and LR, but not with FPE; then, we find unidirectional causality from *mb* to *gdp*. The EG-ADF PT method, on the other hand, suggests bidirectional causality with FPE and LR, but only unidirectional causality from *gdp* to *mb* with SC; MLS PT suggests similar outcomes. These results illustrate the potential conflicts that can occur, demonstrate that researchers would be wise to apply more than one approach when testing for GNC, and to consider simulation results to try to interpret the differences that may arise. We now attempt this.

First, our simulation experiments suggest that LR is more likely to over-reject GNC than is FPE and SC. Second, our earlier discussion also indicates likelihood that the SC may be underestimating the lag order leading us to support a false GNC null. However, this is unlikely with the FPE criterion, and given the simulation performance of the augmented lags approach across many different DGPs, we proceed with the FPE WAL1 (and WAL2) result of GNC from *gdp* to *mb*. This is supported by the JJ PT FPE conclusion. We know that WL can overreject a true null more than WAL1 or WAL2, even when WL is valid, and this may be driving the WL FPE GNC result. If we proceed with GNC from *gdp* to *mb*, then the EG-ADF PT and the MLS PT are supporting causality when it does not exist. This often occurred in our simulation experiments. Finally, can we conclude unidirectional causality from *mb* to *gdp*? Typically, our simulations indicate that the JJ PT method has the highest size uncorrected power when causality exists. This lends support for unidirectional causality from *mb* to *gdp* as this outcome results irrespective of lag method with JJ PT, EG-ADF PT and MLS PT, except for SC with the residual based tests. Moreover, our simulations indicate that WAL1 and WAL2 sometimes have reduced power (relative to the PT methods) for detecting causality. The FPE WAL1 and WAL2 P-values imply unidirectional causality, but not nearly as strongly as for JJ PT, while the SC P-value points to GNC from *mb* to *gdp*; these outcomes may be a reflection of our simulation results and the underestimation of the lag structure by the SC.

The methods suggest bidirectional causality between *ir* and *mb,* except for WL with SC, MLS PT and JJ PT where the support is for unidirectional causality from *ir* to *mb*. Parsimonious lag order selection could be causing the SC outcome, giving a spurious finding of noncausality. Thus, using our experiments and the observed P-values, we conclude unidirectional causality from *mb* to *gdp,* bidirectional causality between *ir* and *gdp* and between *ir* and *mb*.

6. RECOMMENDATIONS FOR APPLIED RESEARCH AND CONCLUDING REMARKS

In this paper we compared several methods for testing for Granger noncausality over a wide range of data generating processes, allowing for the lag order to be selected by either Schwarz's (1978) criterion, Akaike's (1973) FPE criterion or a likelihood ratio general to specific hypothesis testing approach. Overall, our simulation results suggest that the augmented lag method of Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996) exhibits consistent performance over the wide range of investigated DGP families. We prefer this method to the pretesting approaches, whether applying Johansen's (1988) method of testing for cointegration, or the residual based procedures of Engle and Granger (1987) and McCabe et al. (1997), when there is noncausality or unidirectional Granger causality. The cointegration pretest techniques can exhibit serious overrejection frequencies of a true noncausal null. Typically these problems arise from inaccurate determination of the cointegrating rank at the prior testing stage, which may be exacerbated by the form of error variance-covariance matrix. When causal relationships exist, we found that Johansen's cointegration maximum likelihood pretest method often exhibits the highest empirical power, though the difference in empirical power between the various pretest approaches is usually insignificant in large samples.

Thus, we recommend a two-prong strategy to applied researchers interested in Granger causal relationships. We suggest that the Granger noncausality test be undertaken using both the augmented lag method and Johansen's cointegration pretest procedure, with an outcome of noncausality being supported when suggested by the WALd statistic, but a result of causality if indicated by the WALd statistic and confirmed, probably even more strongly, by Johansen's pretest method. Our results do not show a preference for the residual based cointegration pretest procedures. We illustrated many of the distortions that we observed in our simulation experiments with a money demand example. Then, our experiments helped us reconcile several different causality outcomes.

Our simulation experiments generally indicated preference for the Schwarz (1978) criterion for lag order determination, in terms of GNC statistic properties, particularly for samples less than 100 observations. We observed little difference in GNC frequencies between the performances of lag selection criteria in larger samples. Rarely, though, was the likelihood ratio strategy a preferred choice.

However, if the true lag order is relatively large, SC may underestimate it, missing a causal relationship. Our empirical example illustrated this possibility. Accordingly, we advise applied researchers to apply two lag selection methods; e.g., SC and FPE (or AIC). When the difference between the resulting lag orders is large, and if there are sufficient observations, we recommend researchers proceed with the FPE noncausality results when there is a conflict with the SC outcome. However, when the two criteria suggest a similar lag order, SC usually exhibits better control of the Type I error probability and empirical power than does FPE.

Much remains for future research. The fully modified VAR (FM-VAR) estimation procedures (Phillips, 1995 and Quintos, 1998) offer an alternative route to those explored here. Though they were found to suffer from some overrejection frequencies in the known lag case studied by Yamada and Toda (1998), the FM-VAR strategies may do better when selection criteria are used. Jointly determining the cointegrating rank and lag order (e.g., Gonzalo and Pitarakis, 1998) may provide better finite sample performance when testing for Granger noncausality than the sequential procedures we examine. The application of Monte Carlo or bootstrap tests may also be beneficial here. We have also limited our attention to Wald tests for noncausality; a useful avenue of research would be to explore whether LR test statistics and/or out of sample prediction criteria do better at accurately determining Granger noncausality.

ACKNOWLEDGEMENTS

This research was supported by a grant from the Social Sciences and Humanities Council of Canada for the first author. We benefited greatly from Naorayex Dastoor's careful reading and suggestions. We thank Jean-Marie Dufour, Wayne Fuller, John Galbraith, René Garcia, David Giles, Stephen Gordon, Lynda Khalaf, Brendan McCabe, Nilanjana Roy, Pierre Siklos, Ken Stewart, and participants at several conferences and seminars. We also appreciated the thoughtful and helpful comments and suggestions from the referees, associate editor and editor.

REFERENCES

Akaike, H. Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics* **1969**, *21*, 243-247.

Akaike, H., 1973, Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory;* Petrov, B.N., Csáki,F., Eds.; Académiai Kiadó: Budapest, **1973**; 267-281.

Boudjellaba, H.; Dufour, J.-M.; Roy, R. Simplified conditions for noncausality between vectors in multivariate ARMA models. *Journal of Econometrics*, **1994**, *63*, 271-287.

Box, G.E.P.; Pierce, D. Distribution of residual autocorrelations in autoregressive moving average time series models. *Journal of the American Statistical Association*, **1970**, *65*, 1509-1526.

Cheung, Y. -W.; Lai, K.S. Finite-sample sizes of Johansen's likelihood ratio tests for Dickey cointegration. *Oxford Bulletin of Economics and Statistics*, **1993**, *55*, 313-327.

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

Engle, R.F.; Granger, C.W.J. Co-integration and error correction: Representation, estimation and testing. *Econometrica*, **1987**, *55*, 251-276.

Freeman, J.; Houser, D.; Kellstedt, P.M.; Williams, J.T. Long-memoried processes, unit roots and causal inference in political science. *American Journal of Political Science*, **1998**, *42*, 1289-1327.

Gonzalo, J.; Five alternative methods of estimating long run equilibrium relationships. *Journal of Econometrics*, **1994**, *60*, 423-433.

Gonzalo, J.; Pitarakis, J.-Y. Specification via model selection in vector error correction models. *Economics Letters*, **1998**, *60*, 321-328.

Granger, C.W.J., Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **1969**, *36*, 424-438.

Granger, C.W.J. Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, **1981**, *16*, 121-130.

Hall, A. Testing for a unit root in time series with pretest data-based model selection. *Journal of Business and Economic Statistics*, **1994**, *12*, 461-470.

Hoffman, D.L.; Rasche, R.H. Assessing forecast performance in a cointegrated system. *Journal of Applied Econometrics*, **1996**, *11*, 495-517.

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

Kwiatkowski, D.; Phillips, P.C.B.; Schmidt, P.; Shin, Y. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, **1992**, *54*, 159-178.

Lütkepohl, H. *Introduction to Multiple Time Series Analysis*, 2nd Edition; Berlin: Springer-Verlag: Berlin, **1993**.

McCabe, B.P.M.; Leybourne, S.J.; Shin, Y. A parametric approach to testing the null of cointegration. *Journal of Time Series Analysis,* **1997**, *18*, 395-413.

MacKinnon, J.G. Critical values for cointegration tests. In *Long-run Economic Relationships: Readings in Cointegration*, Engle, R.F.; Granger, C.W.J., Eds.; Oxford University Press: Oxford, **1991**; 267-276.

MacKinnon, J.G.; Haug, A.A.; Michelis, L. Numerical distribution functions of likelihood ratio tests for cointegration. *Journal of Applied Econometrics*, **1999**, *14*, 563-577.

Mosconi, R.; Giannini, C. Non-causality in cointegrated systems: Representation, estimation and testing. *Oxford Bulletin of Economics and Statistics*, **1992**, *54*, 399-417.

Ng, S.; Perron, P. Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association*, **1995**, *90*, 268-281.

Nishi, R. Maximum likelihood principle and model selection when the true model is unspecified. *Journal of Multivariate Analysis*, **1988**, *27*, 392-403.

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

Phillips, P.C.B.; Ouliaris, S. Asymptotic properties of residual based tests for cointegration. *Econometrica*, **1990**, *58*, 165-193.

Phillips, P.C.B.; Perron, P. Testing for a unit root in time series regression. *Biometrika*, **1988**, *75*, 335-346.

Quintos, C.E. Fully modified VAR inference in partially nonstationary models. *Journal of the American Statistical Association*, **1998**, *93*, 783-795.

Saikkonen, P.; Lütkepohl, H. Infinte-order cointegrated vector autoregressive processes: estimation and inference. *Econometric Theory*, **1996**, *12*, 814-844.

Schwarz, G. Estimating the dimension of a model. *Annals of Statistics*, **1978**, *6*, 461-464.

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

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

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

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

Wiener, N. The theory of prediction. In *Modern Mathematics for Engineers (Series 1)*; Beckenback, E.F., Ed.; McGraw-Hill: New York, **1956**; Chapter 8.

Yamada, H.; Toda, H.Y. Inference in possibly integrated vector autoregressive models: Some finite sample evidence. *Journal of Econometrics*, **1998**, *86*, 55-95.

Zapata, H.O.; Rambaldi, A.N. Monte Carlo evidence on cointegration and causation. *Oxford Bulletin of Economics and Statistics*, **1997**, *59*, 285-298.

APPENDIX

A.1 Journals

Applied Economics; Applied Economics Letters; Australian Economic Papers; Canadian Journal of Economics; Econometric Theory; Econometrica; Economics Letters; Empirical Economics; European Economic Review; International Journal of Forecasting; Japan and the World Economy; Journal of Business & Economic Statistics; Journal of Development Economics; Journal of Applied Econometrics; Journal of Econometrics; Journal of Economic Studies; Journal of Finance; Journal of International Money and Finance; Journal of International Trade and Economic Development; Journal of Monetary Economics; Journal of Money, Credit and Banking; Oxford Bulletin of Economics and Statistics; Oxford Economic Papers; Public Finance; Quarterly Journal of Economics; Quarterly Review of Economics and Finance; Review of Economics and Statistics; Review of World Economics; Southern Economic Journal; The Indian Journal of Economics; The Singapore Economic Review; The Manchester School.

A.2 Discussion Paper Sources

Australian National University; Bank of Canada; Bank of Spain; Board of Governors of the Federal Reserve System; Boston College; City University of Hong Kong; Curtin University; East Carolina University; Erasmus University of Rotterdam; Federal Reserve Bank of New York; Humboldt-Universität zu Berlin; Latrobe University; London Business School; Monash University; Queen's University; Simon Fraser University; South Langston University; Stockholm University; The Kiel Institute of World Economics; Université Catholique de Louvain; Université de Montreal; University of Aarhus; University of California; University of Edinburgh; University of Groningen; University of Melbourne; University of New England; University of Queensland; University of Texas at Arlington; University of Victoria; University of Wales; University of Western Australia; University of Western Sydney.