An Out of Sample Test for Granger Causality

John Chao(+), Valentina Corradi(++), and Norman Swanson(+++)*

(+) University of Maryland
(++) Queen Mary and Westfield College, University of London
(+++) Texas A&M University

this revision: April 2000

JEL classification: C22, C51.

Keywords: Granger causality, predictive ability, nonlinearity test.

*Corresponding Author: Norman R. Swanson, Department of Economics, Texas A&M University, College Station, Texas 77843-4228. The authors thank Clive W. J. Granger for helpful discussions during the writing of this paper. Philip Hans Franses and two anonymous referees provided insightful comments and suggestions that resulted in an improved revised version. Additionally, the authors wish to thank seminar participants at the conference on nonlinear modeling of multivariate macroeconomic relations in Rotterdam, the 2000 winter meetings of the Econometric Society in Boston, and the workshop on econometrics at Iowa State University for useful suggestions. Swanson thanks the Private Enterprise Research Center and the Bush Program in the Economics of Public Policy, both at Texas A&M University, for financial support.

Abstract

Granger (1980) summarizes his personal viewpoint on testing for causality, and outlines what he considers to be a useful operational version of his original definition of causality (Granger (1969)), which he notes was partially alluded to in Wiener (1958). This operational version is based on a comparison of the 1-step ahead predictive ability of competing models. However, Granger concludes his discussion by noting that it is common practice to test for Granger causality using in-sample F-tests. The practice of using in-sample type Granger causality tests continues to be prevalent. In this paper we develop simple (nonlinear) out-of-sample predictive ability tests of the Granger non-causality null hypothesis. In addition, Monte Carlo experiments are used to investigate the finite sample properites of the test. An empirical illustration shows that the choice of in-sample versus out-of-sample Granger causality tests can crucially affect the conclusions about the predictive content of money for output.

1 Introduction

Granger's original 1969 definition of noncausality (Granger (1969)) has received so much attention in economics that it scarcely needs any introduction (see e.g. the papers by Dufour and Renault (1998), Geweke, Meese and Dent (1983), Lütkepohl (1991), Newbold(1982), Pierce and Haugh (1977), and Sims (1972), for surveys, related results, and relevant references). One aspect of Granger's original definition which hasn't received as much attention, however, is the issue of whether or not standard in-sample implementations of Granger's definition are wholly in the spirit originally intended by Granger, and whether out-of-sample implementations might also be useful. Arguments in favor of using out-of-sample implementations are given in Granger (1980), and are summarized nicely in Ashley, Granger, and Schmalensee (1980), where it is stated on page 1149 that: " *... a sound and natural approach to such tests [Granger causality tests] must rely primarily on the out-of-sample forecasting performance of models relating the original (non-prewhitened) series of interest.*" In this paper we develop simple (nonlinear) out-of-sample predictive ability tests of the Granger noncausality null hypothesis. Our approach is to first study the asymptotic behavior of the tests, and then to investigate the finite sample behavior via a series of Monte Carlo experiments. Finally, an empirical illustration is used to show that the choice of in-sample versus out-of-sample Granger causality tests can crucially affect conclusions based on an empirical investigation of the marginal predictive content of money for output.

It is quite standard to say that x_t (Granger) causes y_t, if the past of x_t (or the present in the case of contemporaneous causality) helps to predict y_t. Thus, it is natural to perform causality tests before constructing forecasting models, and indeed causality tests can be viewed as tests of predictive ability. However, although it is true that both in-sample and out-of-sample lack of predictive ability hypotheses can be formulated in terms of zero restrictions, there is no reason why in-sample and out-of-sample tests should yield the same answers when moderate sample sizes are used. Thus, if we are interested in constructing forecasting models, for example, it is natural to compare out-of-sample predictive ability and hence to construct out-of-sample causality tests[1].

One of the most popular tests for evaluating the predictive ability of two competing forecasting models is the Diebold Mariano test (DM, 1995), and the version thereof which accounts for

[1]Meese and Rogoff (1983) is an important example of the application of out-of-sample model evaluation in the spirit of Granger's definition of non causality.

parameter estimation error (West (1996)). White (1999) further extends the DM test by allowing for the comparison of several models against one benchmark model. (For discussion of these and related tests, see Ashley (1998), Clark (2000), Harvey, Leybourne and Newbold (1998), Mizrach (1992), and the references contained therein.) However, all of these tests are constructed in a nonnested modelling framework, and in the strictly nested modelling framework associated with testing for Granger noncausality we cannot directly implement these tests. The reason for this is quite intuitive. Consider the DM test. In the context of strictly nested models, and when parameter estimation error vanishes, the DM test does not have a normal limiting distribution under the null of non causality, but instead approaches zero in probability. In addition, even when West's (1996) version of the test which accounts for parameter estimation error is used, then as long as the out-of-sample period, P, grows at the same rate as the in-sample period R (i.e. $0 < \pi < \infty$, where $\frac{P}{R} \to \pi$), Clark and McCracken (1999) and McCracken (1999) show that although various Granger causality type out-of-sample predictive ability test statistics can be constructed in the usual way (e.g. encompassing tests, DM tests, etc.), they no longer have normal limiting distributions, but instead converge to functionals of Brownian motion. We suggest a number of tests which have standard (normal) limiting distributions, which account for parameter estimation error when $\pi > 0$, *and which allow for the case where* $\pi = 0$. In addition, our tests are very easy to compute.

One feature of our tests is that they are formed using one-step ahead prediction errors. It should be noted, though, that in-sample implementations of the definition of noncausality to predictive ability at any period have been introduced by Lütkepohl (1993) and Dufour and Renault (1996). Dufour and Renault also provide a set of testable sufficient conditions for which non causality one-step ahead implies non causality at any period, and discuss implementing the test. While it is possible to extend our out-of-sample tests to the evaluation of noncausality at any period, this task is left for future research. In addition, model selection, such as the use of the AIC and SIC for "selecting" between alternative forecasting models offers an alternative to the tests considered here. Such approaches are discussed elsewhere (e.g. see Swanson (1998)).

The rest of the paper is organized as follows. Section 2 outlines the asymptotic properties of a simple linear out-of-sample Granger causality test. In addition, an extension of the test which allows for the evaluation of the linear and nonlinear-out-of sample predictive content of X_t for Y_t, and which is similar in spirit to the nonlinearity test of Lee, White and Granger (1993), is discussed. Section 3 reports the findings of a series of finite sample Monte Carlo experiments, where it is concluded that the simple tests perform reasonably well even when P and R are relatively small. In Section 4 an example in which we analyze the marginal predictive content of money for output is given. The example serves to illustrate the potential for in-sample and out-of-sample Granger causality tests to lead to different conclusions. All proofs are gathered in an Appendix.

2 Linear and Nonlinear Out- of-Sample Granger Causality Tests

Consider the restricted model[2],

$$y_t = \sum_{j=1}^{q} \beta_j^* y_{t-j} + \epsilon_t \tag{1}$$

and the unrestricted model[3],

$$y_t = \sum_{j=1}^{q} \beta_j^* y_{t-j} + \sum_{j=1}^{k} \alpha_j^* x_{t-j} + u_t \tag{2}$$

One implementation of Granger's definition of non causality involves forming a test of the following hypotheses,

$$H_0 : \ \alpha_j^* = 0, \forall j \text{ versus } H_A : \ \alpha_j^* \neq 0 \text{ for some } j$$

An approach in this context is to construct a Wald type statistic which has a limiting χ_k^2 distribution under the null and diverges under the alternative. For example, in the case of iid errors under the null, and given a maintained assumption of conditional homoskedasticity, one commonly constructs

$$F = \frac{(RRSS - URSS)/k}{URSS/(T-k)}$$

where $RRSS$ and $URSS$ are the sum of least squares residuals from the restricted and the unrestricted models, respectively, and $kF \xrightarrow{d} \chi_k^2$ under H_0, while it diverges under the alternative. In general, these type of tests are used to evaluate in sample predictive ability, although an out-of-sample analog is proposed by Clark and McCracken (1999).

Our objective is to construct a direct test for out of sample predictive ability. Suppose we estimate (1) and (2) using observations $t = 1, 2, \ldots R$, and compute $\widehat{\epsilon}_{R+1} = y_{R+1} - \sum_{j=0}^{q-1} \widehat{\beta}_{R,j} y_{t-j}$

[2]Hereafter β^* denotes the best linear predictor of y_t given its past history. Analogously, in the sequel, $\delta^* = (\beta^*, \alpha^*)'$ denotes the best linear predictor of y_t given its past and the past of x_{t-1}.

[3]All of our results generalize straightforwardly to the case where both the restricted and unrestricted models contain the past of other explanatory variables. Here, we simplify the exposition of the test, however, by focusing on the bivariate case.

and $\widehat{u}_{R+1} = y_{R+1} - \sum_{j=0}^{q-1} \widehat{\beta}_{R,j} y_{t-j} - \sum_{j=0}^{k-1} \widehat{\alpha}_{R,j} x_{t-j}$. We then re-estimate the model using $R+1$ observations and construct $\widehat{\beta}_{R+1,j}, \widehat{\alpha}_{R,j}, \widehat{\epsilon}_{R+2}$ and $\widehat{u}_{R+2}$. This procedure is repeated until sequences of P ex ante forecast errors (i.e. $(\widehat{\epsilon}_{R+1}, \widehat{\epsilon}_{R+2}, \ldots \widehat{\epsilon}_{R+P})$ and $(\widehat{u}_{R+1}, \widehat{u}_{R+2}, \ldots \widehat{u}_{R+P})$) have been constructed. Typically, tests for out of sample predictive ability (e.g. DM test) are based on

$$\frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(f(\widehat{\epsilon}_{t+1}) - f(\widehat{u}_{t+1}) \right), \tag{3}$$

where f is some given loss function, and the null hypothesis of equal predictive ability is formulated as

$$H_0' : E(f(\epsilon_{t+1})) - E(f(u_{t+1})) = 0$$

It follows immediately that if H_0 is true, then H_0' should also be true. In fact if $\alpha_j^* = 0, \forall j$, then $u_t = \varepsilon_t$, and so $E(f(\epsilon_{t+1})) - E(f(u_{t+1})) = 0$. In this sense, if X_t has in sample predictive power, it should also have out-of-sample predictive power. Thus, asymptotically we should obtain the same answer regardless of whether the test is performed in-sample or out-of-sample. However, analyses of finite samples may lead to different answers depending on whether in-sample or out-of-sample inference is carried out. This suggests that if we are interested in out-of-sample predictive ability, a natural approach is to construct an out-of-sample predictive ability test. If (3) is expanded around the "true" parameter values we obtain

$$\frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(f(\epsilon_{t+1}) - f(u_{t+1}) \right) + \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \nabla_\beta f|_{\overline{\beta}} (\widehat{\beta}_t - \beta^*) + \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \nabla_\delta f|_{\overline{\alpha}} (\widehat{\delta}_t - \delta^*), \tag{4}$$

where $\overline{\beta} \in (\widehat{\beta}_t, \beta^*)$, $\overline{\delta} \in (\widehat{\delta}_t, \delta^*)$, and $\beta^* = (\beta_1^*, \ldots \beta_q^*)'$, $\delta^* = (\beta^*, \alpha^*)'$. If the loss function is quadratic or if $P/R \to 0$, as $T \to \infty$, then the two last terms in (4) are $o_p(1)$, while the first term is zero under the null (given that the models are strictly nested.) Thus, we cannot use DM type predictive ability tests in the case of strictly nested models. McCracken (1999) proposes a DM type test for the case of nested models. In addition, he shows that if as $P/R \to \pi \neq 0$,as $T \to \infty$, then the parameter estimation error component does not vanish, even if the loss function is quadratic, and the limiting distribution of the DM test is non standard under the null hypothesis, and is dependent on the nuisance parameter π. One feature of the test which we propose is that it does not require $\pi > 0$.

In addition, our statistic has a standard limiting distribution. Consider the following statistic[4]:

$$m_P = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \widehat{\epsilon}_{t+1} X_t, \tag{5}$$

where $\widehat{\epsilon}_{t+1} = y_{t+1} - \sum_{j=0}^{q-1} \widehat{\beta}_{t,j} y_{t-j}$, $X_t = (x_t, x_{t-1}, \ldots x_{t-k-1})'$. We shall formulate the null and the alternative as

$$\widetilde{H}_0 : E(\epsilon_{t+1} x_{t-j}) = 0, j = 0, 1, \ldots k-1 \text{ and } \widetilde{H}_A : E(\epsilon_{t+1} x_{t-j}) \neq 0 \text{ for some } j, j = 0, 1, \ldots k-1$$

In the sequel we shall require the following assumption.

Assumption A. (y_t, x_t) are strictly stationary, strong mixing processes, with size $\frac{-4(4+\delta)}{\delta}$, for some $\delta > 0$, and $E(y_t)^8 < \infty, E(x_t)^8 < \infty, E(\epsilon_t y_{t-j}) = 0$, $j = 1, 2, \ldots q$.

Note that we require $E(\epsilon_t y_{t-j}) = 0$, $j = 1, 2, \ldots q$. Thus, even if we do not require correct dynamic specification, we need to choose q large enough so that the error is not correlated with the regressors. A natural approach is to estimate q using the model selection approach. Alternatively, we could require the lag order, q, to grow at an appropriate rate relative to the sample size. However, such an extension for the case of recursive parameter estimation is not straightforward.

Theorem 1. Let Assumption A hold. As $T \to \infty$, $P, R \to \infty, P/R \to \pi, 0 \leq \pi < \infty$; (i) under $\widetilde{H}_0$, for $0 < \pi < \infty$,

$$m_P \xrightarrow{d} N(0, S_{11} + 2(1 - \pi^{-1}\ln(1+\pi))F'MS_{22}MF - (1 - \pi^{-1}\ln(1+\pi))(F'MS_{12} + S_{12}'MF)).$$

In addition, for $\pi = 0, m_P \xrightarrow{d} N(0, S_{11})$, where $F = E(Y_t X_t')$, $M = \text{plim}\left(\frac{1}{t}\sum_{j=q}^{t} Y_j Y_j'\right)^{-1}$, $Y_j = (y_{j-1}, \ldots y_{j-q})'$, so that M is a $q \times q, F$ is a $q \times k, Y_j$ is a $k \times 1$, S_{11} is a $k \times k$, S_{12} is a $q \times k$, and S_{22} is a $q \times q$ matrix, with

$$S_{11} = \sum_{j=-\infty}^{\infty} E\left((X_t \varepsilon_{t+1} - \mu)(X_{t-j}\varepsilon_{t+1-j} - \mu)'\right),$$

[4] As we require neither the restricted model (under the null) nor the unrestricted to be dynamically correctly specified we need to allow for non martingale difference sequence scores. For the case of conditionally homoskedastic errors under the null, we could have used a regression based test (along the lines of West and McCracken, Theorem 7.1 (1998)). In particular, we could have regressed $\widehat{\epsilon}_{t+1}$ on past values of X_t and tested whether the regression coefficients are zero. In addition, Wooldridge (1990, 1991) proposes a regression based testing framework which allows for conditionally heteroskedasticity and/or non martingale difference errors. However, the extension of Wooldridge's set up to the case of recursively estimated parameters and hence out-of-sample predictive ability tests is not immediate.

where $\mu = E(X_t \epsilon_{t+1})$, $S_{22} = \sum_{j=-\infty}^{\infty} E\left((Y_{t-1}\varepsilon_t)(Y_{t-1-j}\varepsilon_{t-j})'\right)$ and
$S_{12}' = \sum_{j=-\infty}^{\infty} E\left((\epsilon_{t+1}X_t - \mu)(Y_{t-1-j}\epsilon_{t-j})'\right)$; (ii), $\lim_{P\to\infty} \Pr\left(\left|\frac{m_p}{\sqrt{P}}\right| > 0\right) = 1$, under $\widetilde{H}_A$.

Corollary 2. Let Assumption A hold. As $T \to \infty$, $P, R \to \infty, P/R \to \pi, 0 \leq \pi < \infty$, $l_T \to \infty, l_T/T^{1/4} \to 0$,(i) under $\widetilde{H}_0$, for $0 < \pi < \infty$,

$$m_p'(\widehat{S}_{11} + 2(1 - \pi^{-1}\ln(1+\pi))\widehat{F}'\widehat{M}\widehat{S}_{22}\widehat{M}\widehat{F} - (1 - \pi^{-1}\ln(1+\pi))(\widehat{F}'\widehat{M}\widehat{S}_{12} + \widehat{S}_{12}'\widehat{M}\widehat{F}))^{-1}m_p \xrightarrow{d} \chi_k^2,$$

where $\widehat{F} = \frac{1}{P}\sum_{t=R}^{T} Y_t X_t'$, $\widehat{M} = \left(\frac{1}{P}\sum_{t=R}^{T-1} Y_t Y_t'\right)^{-1}$, and

$$\begin{aligned}\widehat{S}_{11} &= \frac{1}{P}\sum_{t=R}^{T-1}(\widehat{\epsilon}_{t+1}X_t - \widehat{\mu}_1)(\widehat{\epsilon}_{t+1}X_t - \widehat{\mu}_1)' + \frac{1}{P}\sum_{t=\tau}^{l_T} w_\tau \sum_{t=R+\tau}^{T-1}(\widehat{\epsilon}_{t+1}X_t - \widehat{\mu}_1)(\widehat{\epsilon}_{t+1-\tau}X_{t-\tau} - \widehat{\mu}_1)' \\ &\quad + \frac{1}{P}\sum_{t=\tau}^{l_T} w_\tau \sum_{t=R+\tau}^{T-1}(\widehat{\epsilon}_{t+1-\tau}X_{t-\tau} - \widehat{\mu}_1)(\widehat{\epsilon}_{t+1}X_t - \widehat{\mu}_1)',\end{aligned}$$

where $\widehat{\mu}_1 = \frac{1}{P}\sum_{t=R}^{T-1}\widehat{\epsilon}_{t+1}X_t$,

$$\widehat{S}_{12}' = \frac{1}{P}\sum_{\tau=0}^{l_T} w_\tau \sum_{t=R+\tau}^{T-1}(\widehat{\epsilon}_{t+1-\tau}X_{t-\tau} - \widehat{\mu}_1)(Y_{t-1}\widehat{\epsilon}_t)' + \frac{1}{P}\sum_{\tau=1}^{l_T} w_\tau \sum_{t=R+\tau}^{T-1}(\widehat{\epsilon}_{t+1}X_t - \widehat{\mu}_1)(Y_{t-1-\tau}\widehat{\epsilon}_{t-\tau})', \text{ and}$$

$$\begin{aligned}\widehat{S}_{22} &= \frac{1}{P}\sum_{t=R}^{T-1}(Y_{t-1}\widehat{\epsilon}_t)(Y_{t-1}\widehat{\epsilon}_t)' + \frac{1}{P}\sum_{\tau=1}^{l_T} w_\tau \sum_{t=R+\tau}^{T-1}(Y_{t-1}\widehat{\epsilon}_t)(Y_{t-1-\tau}\widehat{\epsilon}_{t-\tau})' \\ &\quad + \frac{1}{P}\sum_{\tau=1}^{l_T} w_\tau \sum_{t=R+\tau}^{T-1}(Y_{t-1-\tau}\widehat{\epsilon}_{t-\tau})(Y_{t-1}\widehat{\epsilon}_t)',\end{aligned}$$

with $w_\tau = 1 - \frac{\tau}{l_T+1}$. In addition, for $\pi = 0$, $m_p'\widehat{S}_{11}m_p \xrightarrow{d} \chi_k^2$, and (ii) under the alternative (when $0 < \pi < \infty$),

$$\lim_{P\to\infty} \Pr \frac{m_p'(\widehat{S}_{11} + 2(1 - \pi^{-1}\ln(1+\pi))\widehat{F}'\widehat{M}\widehat{S}_{22}\widehat{M}\widehat{F} - (1 - \pi^{-1}\ln(1+\pi))(\widehat{F}'\widehat{M}\widehat{S}_{12} + \widehat{S}_{12}'\widehat{M}\widehat{F}))^{-1}m_p}{P} > 0$$
$$= 1,$$

while for $\pi = 0$, $\lim_{P\to\infty} \Pr\left(\frac{1}{P}m_{p'}\widehat{S}_{11}^{-1}m_p > 0\right) = 1$.

Thus far, we have focused on a test for the null of linear non causality. We can instead use a more general test function, such as the exponential (as in Bierens (1990)), a neural network with sigmoidal activation function, or a generically comprehensive function (as defined in Stinchcombe and White (1998)) and then construct a test for nonlinear out-of-sample predictive ability based on

$\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\widehat{\epsilon}_{t+1}h(\gamma' X_t)$, where $\gamma \in \Gamma$ is a nuisance parameter unidentified under the null hypothesis (for a detailed survey of nonlinearity tests used in economics, see Granger and Teräsvirta (1993)). Under mild conditions it is straightforward to establish that the statistic above converges to a Gaussian process, with covariance kernel that depends on γ, under the null hypothesis. However, it is not a trivial task to form bootstrap critical values which take parameter estimation error into account, particularly as the parameters are estimated recursively. Thus, we confine our attention to a finite grid of values for the nuisance parameter γ. More precisely we shall follow the approach suggested by Lee, White and Granger (1993, LWG) in the context of (in-sample) testing for neglected nonlinearities, and set

$$h(\gamma' X_t) = \gamma' X_t + (1 + \exp(c - \gamma' X_t))^{-1}, c \neq 0$$

where γ is a $k \times 1$ vector[5]. In this context, consider the following statistic[6],

$$s_P = \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\widehat{\epsilon}_{t+1}h(\gamma' X_t)$$

In the sequel, we specify the null and the alternative as,

$$H_0^* : E(\epsilon_{t+1}h(\gamma' X_t)) = 0, \text{ and } H_A^* : E(\epsilon_{t+1}h(\gamma' X_t)) \neq 0$$

Corresponding to the above results, we have that,

Proposition 3. Let Assumption A hold. As $T \to \infty$, $P, R \to \infty, P/R \to \pi, 0 \leq \pi < \infty$, $l_T \to \infty, l_T/T^{1/4} \to 0$, (i) under H_0^*, for $0 < \pi < \infty$, and for any given τ,

$$s_p^2/(\widehat{S}_{11} + 2(1 - \pi^{-1}\ln(1+\pi))\widehat{F}'\widehat{M}\widehat{S}_{22}\widehat{M}\widehat{F} - (1 - \pi^{-1}\ln(1+\pi))(\widehat{F}'\widehat{M}\widehat{S}_{12} + \widehat{S}_{12}'\widehat{F}\widehat{M})) \xrightarrow{d} \chi_1^2$$

where $\widehat{F} = \frac{1}{P}\sum_{t=R}^{T} Y_t h(\gamma' X_t)$, $\widehat{M} = \left(\frac{1}{P}\sum_{j=R}^{T} Y_j Y_j'\right)^{-1}$, and

$$\begin{aligned}
\widehat{S}_{11} &= \frac{1}{P}\sum_{t=R}^{T-1}(\widehat{\epsilon}_{t+1}h(\gamma' X_t) - \widehat{\mu}_1)(\widehat{\epsilon}_{t+1}h(\gamma' X_t) - \widehat{\mu}_1)' \\
&\quad + \frac{1}{P}\sum_{t=\tau}^{l_T} w_\tau \sum_{t=R+\tau}^{T-1}(\widehat{\epsilon}_{t+1}h(\gamma' X_t) - \widehat{\mu}_1)(\widehat{\epsilon}_{t+1-\tau}h(\gamma' X_{t-\tau}) - \widehat{\mu}_1)' \\
&\quad + \frac{1}{P}\sum_{t=\tau}^{l_T} w_\tau \sum_{t=R+\tau}^{T-1}(\widehat{\epsilon}_{t+1-\tau}h(\gamma' X_{t-\tau}) - \widehat{\mu}_1)(\widehat{\epsilon}_{t+1}h(\gamma' X_t) - \widehat{\mu}_1)',
\end{aligned}$$

[5]Different sets of weights, say γ_1 and γ_2 can be chosen for the linear and nonlinear components of the model.

[6]Lee, White and Granger (1993) construct their test statistic using the in-sample correlation of the estimated residuals from a linear model and a nonlinear (neural network) component.

where $\widehat{\mu}_1 = \frac{1}{P}\sum_{t=R}^{T-1}\widehat{\epsilon}_{t+1}h(\gamma' X_t)$,

$$\begin{aligned} \widehat{S}'_{12} &= \frac{1}{P}\sum_{\tau=0}^{l_T} w_\tau \sum_{t=R+\tau}^{T-1}\left(\widehat{\epsilon}_{t+1-\tau}h(\tau' X_{t-\tau}) - \widehat{\mu}_1\right)\left(Y_{t-1}\widehat{\epsilon}_t\right)' \\ &\quad + \frac{1}{P}\sum_{\tau=1}^{l_T} w_\tau \sum_{t=R+\tau}^{T-1}\left(\widehat{\epsilon}_{t+1}h(\tau' X_t) - \widehat{\mu}_1\right)\left(Y_{t-1-\tau}\widehat{\epsilon}_{t-\tau}\right)', \text{ and} \end{aligned}$$

$$\begin{aligned} \widehat{S}_{22} &= \frac{1}{P}\sum_{t=R}^{T-1}\left(Y_{t-1}\widehat{\epsilon}_t\right)\left(Y_{t-1}\widehat{\epsilon}_t\right)' + \frac{1}{P}\sum_{\tau=1}^{l_T} w_\tau \sum_{t=R+\tau}^{T-1}\left(Y_{t-1}\widehat{\epsilon}_t\right)\left(Y_{t-1-\tau}\widehat{\epsilon}_{t-\tau}\right)' \\ &\quad + \frac{1}{P}\sum_{\tau=1}^{l_T} w_\tau \sum_{t=R+\tau}^{T-1}\left(Y_{t-1-\tau}\widehat{\epsilon}_{t-\tau}\right)\left(Y_{t-1}\widehat{\epsilon}_t\right)', \end{aligned}$$

with $w_\tau = 1 - \frac{\tau}{l_T+1}$. In addition, for $\pi = 0$, $s_P^2/\widehat{S}_{11} \xrightarrow{d} \chi_k^2$, and (ii) under the alternative (when $0 < \pi < \infty$),

$$\lim_{P\to\infty} \Pr\left(\frac{s_p^2(\widehat{S}_{11} + 2(1-\pi^{-1}\ln(1+\pi))\widehat{F}'\widehat{M}\widehat{S}_{22}\widehat{M}\widehat{F} - (1-\pi^{-1}\ln(1+\pi))(\widehat{F}'\widehat{M}\widehat{S}_{12} + \widehat{S}'_{12}\widehat{M}\widehat{F}))^{-1}}{P} > 0\right)$$

$$= 1$$

while for $\pi = 0$, $\lim_{P\to\infty}\Pr\left(\frac{1}{P}m_{p'}\widehat{S}_{11}^{-1}m_p > 0\right) = 1$.

Note that both the finite sample size and power depend on the specific γ which is used. Following LWG however, we can randomly draw l different sets of γ and compute l different statistics, say. Let $PV_1, \ldots PV_l$ be the p-values associated with the l different statistics, so that $PV_1 \leq PV_2 \ldots \leq PV_l$. LWG suggest rejecting the null at 5% if there is a $j = 1, \ldots l$ such that $PV_j \leq 0.05/(l-j-1)$.

3 Monte Carlo Findings

In this section, we report results from a series of bivariate Monte Carlo experiments. Assume that:

$$y_t = \pi_1 + \pi_2 y_{t-1} + \pi_3 x_{t-1} + \varepsilon_{1,t} \ ,$$

$$x_t = a_1 + a_2 x_{t-1} + \varepsilon_{2,t} \ ,$$

where $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are $IN(0, \sigma_i^2)$, $i = 1, 2$. In order to change the predictive relevance of the past of x_t relative to the past of y_t in regression models of y_t, we focus on two quantities of interest when parameterizing the above DGP. In particular, and assuming that x_t and y_t are stationary, in our empirical power experiments we consider,

$$A = \pi_2^2 var(y_t)(\pi_3^2 var(x_t))^{-1} \text{ and}$$

$$B = \frac{\pi_2^2 var(y_t) + \pi_3^2 var(x_t)}{\pi_2^2 var(y_t) + \pi_3^2 var(x_t) + var(\varepsilon_{1,t})}.$$

Notice that A defines the magnitude of the explained variation in the model of y_t which is due the past of y_t relative to that due to the past of x_t. Thus, by changing A we can change the relative importance of the past of x_t for predicting y_t. Our other quantity of interest, B, is a measure of the goodness of fit of the model, and thus can be used as an indicator of how well we might expect to predict y_t given the past of both x_t and y_t. In order to parameterize our model using A and B, we assume that $\pi_2 = a_2$. In addition, and for simplicity, assume that $var(\varepsilon_{1,t}) = var(\varepsilon_{2,t}) = 1$, and that $\pi_1 = a_1 = 1$. Thus, given $|a_2| < 1$, $var(y_t) = \pi_2^2 var(y_t) + \pi_3^2 var(x_t) + var(\varepsilon_{1,t})$, and $var(x_t) = a_2^2 var(x_t) + var(\varepsilon_{2,t})$, it follows that

$$A = \frac{\pi_2^2(\pi_3^2 + (1-\pi_2^2)^3)}{\pi_3^2(1-\pi_2^2)}, \text{ and}$$

$$B = \frac{\pi_2^2(1-\pi_2^2)^3 + \pi_3^2}{\pi_2^2(1-\pi_2^2)^3 + (1-\pi_2^2)^2 + \pi_3^2},$$

so that by fixing A and π_2 it is possible to solve for π_3 and hence also for B. In the Monte Carlo experiments reported in Tables 4-6 (empirical power), we set $A = \{0.1, 0.5, 1.0, 5.0, 10.0\}$, and $\pi_2 = \{0.1, 0.3, 0.5, 0.7, 0.9\}$. In addition, P is set equal to $\{0.1T, 0.3T, 0.5T\}$, and samples of $\{250, 500, 1000\}$ observations are generated. When Wald F tests are constructed, the entire sample is used, while when the two versions of the properly scaled m_P statistics, which we shall call $v^{-1/2}m_P$ (one constructed based on the assumption that $\pi = 0$, say $v^{-1/2}m_P(\pi = 0)$, and the other constructed based on the assumption that $0 < \pi < \infty$, say $v^{-1/2}m_P(\pi > 0)$) are constructed, only P observations are used[7]. As discussed above, forecasts are generated recursively, with model parameters re-estimated before each new 1-step ahead forecast is constructed. However, for simplicity

[7]Note that for the cases in which the solutions for π_3 given A and π_2 are both complex, we do not generate data, and hence no results are reported in Tables 1-6. Largely, these cases involve small values of A together with small values of π_2. Also, in-sample tests are performed using the entire sample period. An alternative would be to use rolling windows of observations which correspond to the periods used to construct the one-step ahead forecasts, hence yielding sequences of in-sample tests at each simulation step. However, as empirical researchers often use the entire sample of data when constructing in-sample tests, we do likewise here.

it is assumed that the correct lag structure is known. All results are based on 10000 Monte Carlo iterations, and are rejection frequencies of the null hypothesis of non Granger causality. Needless to say, in corresponding empirical size experiments we set $\pi_3 = 0$ (Tables 1-3) so that the only parameter that matters is π_2.

Consider first the empirical size results reported in Tables 1-3. The rejection frequencies reported in Table 1 correspond to Wald F-tests run under the null of Granger noncausality, and as expected, empirical size is close to nominal size even for our smallest sample of 250 observations. Note also that in Table 2 empirical size of the $v^{-1/2}m_P(\pi = 0)$ test is slightly higher than nominal when $P = 0.1T$, for all T, and decreases when P increases, with T fixed. The results in Table 3 for $v^{-1/2}m_P(\pi > 0)$ suggest that empirical size is smaller than for the $v^{-1/2}m_P(\pi = 0)$ statistic, and analogous to the results of Table 2, the test is more undersized when $P = 0.5T$ than when $P = 0.3T$, so that empirical size appears to decrease when P is increased and R is decreased for fixed T, underscoring the importance of parameter estimation error.

Empirical power results reveal much more clearly the tradeoffs between out-of-sample and in-sample tests of Granger noncausality. Note in Table 4 that the Wald F-test is very powerful for all values of T, regardless of the magnitudes of A, B, and π. This means that even when the parameter associated with x_{t-1} is very small, and the relative importance of x_{t-1} in the overall regression model is very small, the Wald F-test favors a finding of Granger causality. This of course is expected, and certainly must be the case for large samples. Our evidence suggests that it also holds for small samples. However, in cases where A=10, say, and π_3 is between 0.03 and 0.13 (the last 5 rows of Table 4), it is clear that the marginal predictive content of x_{t-1} for y_t will be very low. In such cases, it is not clear whether a finding of Granger causality is desired, particularly if the objective of the modeler is to select variables for inclusion in a forecasting model for y_t. Note that in Tables 5 and 6, empirical power of the m_P statistics for these cases (again, see last 5 rows) is much lower than that based on the in-sample tests. Of course, power does increase as P, T increase, as expected. However, even for $P = 0.5T$ and $T = 1000$, power is still below 0.5. This suggests that even though the data are generated with nonzero π_3, x_{t-1} is nevertheless not always useful for predicting y_t, at least based on mean square error prediction loss. However, note that the $v^{-1/2}m_P$ statistics are powerful against alternatives where A values are 1 or below (equal predictive ability of x_{t-1} and y_{t-1}) and B values are higher than 0.5, even when P and T values are low, again as expected.

4 Empirical Illustration

In order to illustrate the potential for different empirical approaches to testing for Granger causality to lead to different conclusions, we consider the problem of assessing whether fluctuations in the money stock anticipate (or Granger cause) fluctuations in real output. This is a question which has received considerable attention in the applied macroeconomics literature (see e.g. Christiano and Ljungqvist (1988), Hafer and Jansen (1991), Stock and Watson (1989), Swanson (1998), Thoma (1994), and the references contained therein). Here we take as given the group of macroeconomic variables used by all of the above authors, and construct in- and out-of-sample tests of Granger non-causality.

To summarize, we fit $VEC(p)$ models of the form

$$\Delta Y_t = a + b(L)\Delta Y_{t-1} + cZ_{t-1} + \epsilon_t, \tag{6}$$

where $Y_t = (IP_t, M2_t, CPI_t, R_t)'$. The four elements of Y_t are monthly seasonally adjusted U.S. measures of industrial production, the nominal money stock (M2), the consumer price index, and the 3-month treasury bill return (secondary market) for the period 1961:1-1997:9. Based on the results obtained by forming augmented Dickey-Fuller test statistics, it was assumed that all variables are $I(1)$. In addition, $Z_{t-1} = dY_{t-1}$ is a $r \times 1$ vector of $I(0)$ variables, r is the rank of the cointegrating space, d is an $r \times 4$ matrix of cointegrating vectors, a is an 4×1 vector, $b(L)$ is a matrix polynomial in the lag operator L, with p terms, each of which is an 4×4 matrix, p is the order of the VEC model, c is an $4 \times r$ matrix, and ϵ_t is a vector error term. In order to ensure that the real time forecasting models which we construct are not affected by data revision problems, as discussed in Ghysels, Swanson, and Callan (1999), we use real-time versions of these variables, where by real-time we mean that at each point in time an entire vector of observations for each variable is constructed going back to the beginning of the sample. Each vector of observations is real-time because revisions and seasonal adjustment modifications which occurred $after$ the calendar date to which the real-time vector corresponds are not incorporated into the data.[8]

[8] As an example, consider downloading data on IP_t right now from CITIBASE. The data corresponds to observations available right now. However, if the last 50 observations were held back, and the first 150 observations, say, were used to form a forecast of the first observation in the out-of-sample period, then the forecast would not truly be real-time. The problem is that if one were to go back in time to the date of the last in-sample observation, then one would find that the data from CITIBASE do not correspond to the data that are actually available, as the CITIBASE

Using real time data, models of the form given by equation (6) were re-estimated 212 times using samples of observations beginning in 1961:1 and ending in 1980:1+x, for $x = 1, ..., 212$, so that the last sample of observations used was 1961:1-1997:8. Each re-estimation step involved fitting two different models - a bigger model (with money) and a smaller model (without money). The parameters r, p, a, and b, were re-estimated at each point in time using least squares and the SIC for selecting the number of lags[9]. As our forecasting results based on VEC models were never superior to those based on VAR models, we report here only results for the case where $r = 0$.

Our approach allowed us to compute sequences of 212 in-sample Wald F-tests of the null of Granger non-causality, for example. Of these, 94.8% resulted in rejection (at the 5% level), and hence in a finding that money is Granger causal for real output. This result is similar to that found by Swanson (1998). Our approach also allowed us to form sequences of 1-step ahead forecasts of the growth in industrial production using our smaller model and our bigger model, and to compare these forecasts with actual figures, thus forming sequences of real-time forecast errors along the lines discussed in Ashley, Granger, and Schmalensee (1980). Interestingly, the $MSFEs$[10] (reported as percentages) based on forecasts constructed using the bigger and smaller models were found to be 0.4084 and 0.4101, respectively. Thus, if point estimates are compared the model with money is preferred to the smaller model without money, in accord with our in-sample findings. However, note in Figure 1 (see right top panel) that there is a large outlier in the difference series of the absolute forecast errors from the two models (*bigger model forecast error minus smaller model forecast error)*. This outlier corresponds to a forecast for which the smaller model performed substantially worse the bigger model. On the other hand, note that most of the difference-forecast errors are above 0.0, corresponding to the observation that for most periods, the smaller model forecasted better than the bigger model. For this reason, our point $MSFEs$ may be misleading. Indeed, the properly rescaled m_P statistic values based on the two models are 0.0526 ($\pi = 0$) and 0.0476 ($\pi = 0.5$),

data have been revised, etc. This feature of macroeconomic data is well known, and is discussed by Diebold and Rudebusch (1991), for example.

[9] Amato and Swanson (1999), detail a thorough examination of the marginal predictive content of money for real output. In addition, as the SIC selected just over 1 lag, on average, across all samples for which models were estimated, we set $p = 1$. This allowed us to fix the regressor sets, X_t and Y_t used in the construction of the m_P statistics.

[10] Other loss functions may also be used to construct predictive ability tests, as discussed in Christoffersen and Diebold (1997), Clements and Hendry (1988a,b), and Weiss (1996), for example.

indicating that money is not causal for industrial production, at least in a predictive sense. The earliest period for which complete real-time vectors of data (back to 1961:1) could be constructed is 1978:1. In order to check the robustness of our finding that in- and out-of-sample analyses can lead to different conclusions, we also performed the above empirical investigation for the out-of-sample period 1978:2-1997:9. Based on this sample, our findings remained unchanged. Of course, it may be that for certain sub-samples the in-sample and out-of-sample results will match up, and in fact it would be surprising if this were not the case. In addition, it is important to note that there may be structural breaks in the underlying data generating process, and the degree to which different models are robust to such breaks might vary (e.g. the large outlier in the forecasts from the smaller model). It could thus be argued that structural breaks play a role in our finding that in-sample and out-of-sample tests yield contradictory findings. This and related issues are discussed in detail in Clements and Hendry (1999). Nevertheless, we can conclude that there are examples for which the decision between using in-sample versus out-of-sample inference is crucial. In particular, we have found that although in-sample tests suggest that there is Granger causality from money to output at least some of the time, predictive ability tests suggest that nothing is gained by using money in a forecasting model for output.

5 Conclusions

We discuss and implement a number of out-of-sample predictive ability tests in the spirit of Granger's original 1969 definition of noncausality. It is shown that in finite sample contexts our out-of-sample tests can lead to evidence that is more indicative of the true forecasting ability of one variable for another than when standard in-sample Wald type F-tests are used. In an empirical illustration, we show that in-sample and out of sample tests can lead to different conclusions.

6 Appendix

Proof of Theorem 1.(i)

$$\begin{aligned}
m_P &= \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\epsilon_{t+1}X_t - \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}X_tY_t'\left(\widehat{\beta}_t-\beta^*\right)\\
&= \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\epsilon_{t+1}X_t - \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}X_tY_t'M\left(\frac{1}{t}\sum_{j=q}^{t}Y_{j-1}\epsilon_j\right)+o_p(1)\\
&= I+II+o_p(1)
\end{aligned}$$

where $M=$plim $\left(\frac{1}{t}\sum_{j=q}^{t}Y_jY_j'\right)^{-1}$. Thus,

$$II=\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}F'M\left(\frac{1}{t}\sum_{j=q}^{t}Y_{j-1}\epsilon_j\right)+\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(X_tY_t'-F')M\left(\frac{1}{t}\sum_{j=q}^{t}Y_{j-1}\epsilon_j\right)+o_p(1) \tag{7}$$

where $F'=E(X_tY_t'), k\times q$. We want to show that the second term on the RHS of (7) is $o_p(1)$. We shall follow an argument similar to that used by West (1996). Let $v_t=(X_tY_t'-F')$, and $h_j=(Y_{j-1}\epsilon_j)$, so that the second term on the RHS of (7) can be written as $\frac{1}{\sqrt{P}}\sum_{t=R}^{T}v_tM\left(\frac{1}{t}\sum_{j=q}^{t}h_j\right)$. We begin by showing that the expectation of the last expression is $o_p(1)$. Let $\gamma_j=E(v_tMh_{t-j})$, where γ_j is $k\times 1$ and let γ_{ij} be the i-th component of γ_j. We shall show that each component is $o(1)$. So $\forall i=1,2,\dots k$,

$$\begin{aligned}
&E\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}v_tM\left(\frac{1}{t}\sum_{j=q}^{t}h_j\right)\right)_i\\
=&\ \frac{1}{\sqrt{P}}\left|\left(R^{-1}(\gamma_{i0}+\gamma_{i1}+\dots\gamma_{iR})+\dots(R+P-1)^{-1}(\gamma_{i0}+\dots+\gamma_{iR}+\dots\gamma_{iR+P-1})\right)\right|\\
\leq&\ \frac{1}{\sqrt{P}}\left|\left(R^{-1}+(R+1)^{-1}+\dots+(R+P-1)^{-1}\right)\right|\sum_{j=0}^{\infty}|\gamma_{ij}|
\end{aligned}$$

We begin by showing that $\forall i,\sum_{j=0}^{\infty}|\gamma_{ij}|<\infty$. Because of the covariance inequality for strong mixing processes, (e.g. Yokohama (1980)),

$$\sum_{j=0}^{\infty}|\gamma_{ij}|\leq 12E(|v_tMh_{t-j}|_i^3)^{1/3}\sum_{j=0}^{\infty}\alpha_j^3<\infty$$

where $E(|v_tMh_{t-j}|_i^3)^{1/3}<\infty$, and $\sum_{j=0}^{\infty}\alpha_j^3<\infty$, given the moment and mixing conditions in Assumption A. Also,

$$\frac{1}{\sqrt{P}}\left|\left(R^{-1}+(R+1)^{-1}+\dots+(R+P-1)^{-1}\right)\right|=O\left(\sum_{t=R+1}^{T-1}t^{-3/2}\right)=o(1)$$

Thus, the mean of the second term on the RHS of (7) is $o(1)$. By Chebyshev inequality, $\forall i = 1, \dots k$,

$$\begin{aligned}
\Pr\left(\left|\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1} v_t M\left(\frac{1}{t}\sum_{j=q}^{t} h_j\right)\right|_i > \epsilon\right) &\leq Var\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1} v_t M\left(\frac{1}{t}\sum_{j=q}^{t} h_j\right)\right)_i \\
&= E\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1} v_t M\left(\frac{1}{t}\sum_{j=q}^{t} h_j\right)\right)_i^2 + o(1)
\end{aligned}$$

$$E\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1} v_t M\left(\frac{1}{t}\sum_{j=q}^{t} h_j\right)\right)_i^2 = \frac{1}{P}\sum_{t=R+1}^{T-1} E\left(v_t M\left(\frac{1}{t}\sum_{j=q}^{t} h_j\right)\right)_i^2 \tag{8}$$

$$+\frac{2}{P}\sum_{t=R+1}^{T-1}\sum_{l=R+1}^{T-1} E\left(v_l M\left(\frac{1}{t}\sum_{j=q}^{t} h_j\right)\right)_i \tag{9}$$

Recalling that v is $k \times q$, M is $q \times q$, and h is $q \times 1$, $v_l M\left(\frac{1}{t}\sum_{j=q}^{t} h_j\right)_i$ can be written as (assuming for notational simplicity but without loss of generality that $k = q = 2$ and $i = 1$) $\sum_{s=1}^{2} v_{1s,l} M_{s1}\left(\frac{1}{t}\sum_{j=q}^{t} h_{1,j}\right) + \sum_{s=1}^{2} v_{1s,l} M_{s2}\left(\frac{1}{t}\sum_{j=q}^{t} h_{2,j}\right)$. Note also that as $|l-t| \to \infty$, $E\left(\sum_{s=1}^{2} v_{1s,l} M_{s1}\left(\frac{1}{t}\sum_{j=q}^{t} h_{1,j}\right)\right) \to 0$, so we can rewrite (9) as

$$\frac{2}{P}\sum_{\tau=R+1}^{l_T}\sum_{i=R+1}^{T-1} E\left(v_i M\left(\frac{1}{t}\sum_{j=q}^{t} h_j\right)\right)_i = \frac{2}{P} l_T o(1)$$

by the argument used above. Thus, the term on the RHS is $o(1)$ for $\frac{l_T}{P} \to 0$, as $T \to \infty$. Note also that the term on the RHS of (8) is $o(1)$, given the moment and mixing conditions in Assumption A, by the same argument used in the proof of Lemma 3.1 in Corradi (1999). Thus,

$$m_P = \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\epsilon_{t+1}X_t - \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1} F'M\left(\frac{1}{t}\sum_{j=q}^{t} Y_{j-1}\epsilon_j\right) + o_p(1)$$

From Lemma A5 in West (1996), we have that

$$E\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1} F'M\left(\frac{1}{t}\sum_{j=q}^{t} Y_{j-1}\epsilon_j\right)\left(\frac{1}{t}\sum_{j=q}^{t} Y_{j-1}\epsilon_j\right)' MF\right) \to 2(1-\pi^{-1}\ln(1+\pi))F'MS_{22}MF,$$

where S_{22} is defined as in the statement of the theorem. Also, from Lemma A6 in West (1996),

$$E\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\epsilon_{t+1}X_t\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\frac{1}{t}\sum_{j=q}^{t} Y_{j-1}\epsilon_j\right)' MF\right) \to (1-\pi^{-1}\ln(1+\pi))S_{12}'MF$$

where S_{12} is defined in the statement of the theorem, and finally

$$E\left(\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\epsilon_{t+1}X_t\right)\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\epsilon_{t+1}X_t\right)'\right) \rightarrow S_{11}$$

where S_{11} is defined in the statement. Thus, by the central limit theorem for stationary mixing processes,

$$\begin{pmatrix} \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\epsilon_{t+1}\gamma' X_t \\ \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}F'M\left(\frac{1}{t}\sum_{j=q}^{t}Y_{j-1}\epsilon_j\right) \end{pmatrix}$$
$$\xrightarrow{d} N\begin{pmatrix} S_{11} & (1-\pi^{-1}\ln(1+\pi))S_{12}'MF \\ 2(1-\pi^{-1}\ln(1+\pi))F'MS_{12} & 2(1-\pi^{-1}\ln(1+\pi))F'MS_{22}MF \end{pmatrix}$$

The result then follows for the case of $P/R \rightarrow \pi, 0 < \pi < \infty$.

For $\pi = 0$, it suffices to show that $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}X_tY_t'\left(\widehat{\beta}_t - \beta^*\right) = o_p(1)$, and so it suffices to show that $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}x_{t-j}y_{t-i}\left(\widehat{\beta}_{i,t} - \beta_i^*\right) = o_p(1)$, for $j = 0, 1, \ldots k-1$, $i = 1, \ldots q$.

$$\left|\frac{1}{\sqrt{P}}\sum_{t=R}^{T}x_{t-j}y_{t-i}\left(\widehat{\beta}_{i,t} - \beta_i^*\right)\right| \leq \sup_{t\geq R}\sqrt{P}|\widehat{\beta}_{i,t} - \beta_i^*|\frac{1}{P}\sum_{t=R}^{T}|x_{t-j}y_{t-i}|$$

Now, $\frac{1}{P}\sum_{t=R}^{T}(|x_{t-j}y_{t-i}|)$ converge in probability to a non random vector, while

$$\sup_{t\geq R}\sqrt{P}|\widehat{\beta}_t - \beta^*| \leq \left(\frac{1}{R}\sum_{j=q}^{R}y_{t-i}\overline{P}_y y_{t-i}\right)^{-1}\frac{\sqrt{P}}{\sqrt{R}}\left|\frac{1}{\sqrt{R}}\sum_{j=q}^{t}y_{j-i}\overline{P}_y\epsilon_j\right|$$

where $\overline{P}_y = I - P_y$ and P_y is the projection of y_t on $y_{t-l}, l = 1, \ldots i-1, i+1, \ldots q$. As $\frac{1}{\sqrt{R}}\left|\sum_{j=q}^{t}y_{j-i}\overline{P}_y\epsilon_j\right|$ satisfies an invariance principle and so is $O_p(1)$, the right hand side of the inequality above is $o_p(1)$ for $\frac{P}{R} = o(1)$.

(ii) $\frac{1}{\sqrt{P}}\sum_{t=R}^{T}X_tY_t'M\left(\frac{1}{t}\sum_{j=q}^{t}Y_{j-1}\epsilon_j\right) = O_p(1)$ by the same argument as above. On the other hand $E(\epsilon_{t+1}X_t) \neq 0$, and so $\left|\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\epsilon_{t+1}X_t\right|$ diverges at rate $\sqrt{P}$.

Proof of Corollary 2. $\widehat{M}, \widehat{F}, \widehat{S}_{ij}$ $i,j = 1,2$ are consistent for M, F, S_{ij}. The result follows immediately.

Proof of Proposition 3. Follows directly by the same arguments used in the proof of Theorem 1 and Corollary 2 when X_t is replaced with $h(\gamma' X_t)$.

7 References

Amato, J. and N.R. Swanson, (1999), The Real-time (In)Significance of M2, *Journal of Monetary Economics,* forthcoming.

Ashley, R., (1998), A New Technique for Postsample Model Selection and Validation, *Journal of Economic Dynamics and Control* 22, 647-665.

Ashley, R., C.W.J. Granger, and R. Schmalensee, (1980), Advertising and Aggregate Consumption: An Analysis of Causality, *Econometrica,* 48, 1149-1167.

Bierens, H.B., (1990), A Consistent Conditional Moment Test of Functional Form, *Econometrica,* 58, 1443-1458.

Christiano, L.J. and L. Ljungqvist, (1988), Money Does Granger-Cause Output in the Bivariate Money-Output Relation, *Journal of Monetary Economics*, 22, 217-235.

Christoffersen, P. and F.X. Diebold, (1997), Optimal Prediction Under Asymmetric Loss, *Econometric Theory*, 13, 808-817.

Clark, T., (2000), Finite-Sample Properties of Tests for Equal Forecast Accuracy, *Journal of Forecasting*, 18, 489-504.

Clark, T. and M.W. McCracken, M.W., (1999), Granger Causality and Tests of Equal Forecast Accuracy and Encompassing, Working Paper, Federal Reserve Bank of Kansas City.

Clements, M.P. and D.F. Hendry, (1998a), Which Methods Win Forecasting Competitions in Economics?, Working Paper, University of Oxford.

Clements, M.P. and D.F. Hendry, (1998b), *Forecasting Economic Time Series*, Cambridge: Cambridge University Press.

Clements, M.P. and D.F. Hendry, (1999), *Forecasting Non-Stationary Economic Time Series*, Cambridge: MIT Press.

Corradi, V., (1999), Deciding Between I(0) and I(1) via FLIL-Based Bounds, *Econometric Theory,* 15.

Diebold, F.X. and R.S. Mariano, (1995), Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*, 13, 253-263.

Diebold, F.X. and G.D. Rudebusch, (1991), Forecasting Output With the Composite Leading Index: A Real Time Analysis, *Journal of the American Statistical Association*, 86, 603-610.

Dufour, J.M. and E. Renault, (1998), Short Run and Long Run Causality in Time Series: Theory, *Econometrica,* 66, 1099-1126.

Geweke, J., R. Meese and W. Dent, (1983), Comparing Alternative Tests of Causality in Temporal Systems, *Journal of Econometrics*, 21, 161-194.

Ghysels, E., N.R. Swanson, and M. Callan, (1999), Monetary Policy Rules with Model and Data Uncertainty, Working Paper, Pennsylvania State University.

Granger, C.W.J., (1969), Investigating Causal Relations by Econometric Models and Cross Spectral Methods, *Econometrica,* 37, 424-438.

Granger, C.W.J., (1980), Testing for Causality: A Personal Viewpoint, *Journal of Economic Dynamics and Control,* 2, 329-352.

Granger, C.W.J. and T. Teräsvirta, 1993, *Modelling Nonlinear Economic Relationships*, New York: Oxford.

Hafer, R.W. and D.W. Jansen, (1991), The Demand for Money in the United States: Evidence from Cointegration Tests, *Journal of Money, Credit, and Banking*, 23, 155-168.

Harvey, D.I., S.J. Leybourne and P. Newbold, (1997), Tests for Forecast Encompassing, *Journal of Business and Economic Statistics*, 16, 254-259.

Lee, T.H., H. White and C.W.J. Granger, (1993), Testing for Neglected Nonlinearities in Time Series: A Comparison of Neural Network Methods and Alternative Tests, *Journal of Econometrics,* 56, 269-290.

Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis,* Springer and Verlag, Berlin.

Lütkepohl, H. (1993), Testing for Causation Between Two Variables in Higher Dimensional VAR Models, in *Studies in Applied Econometrics,* eds. H. Schneeweiss and K. Zimmerman, Heidelberg: Springer-Verlag.

McCracken, M.W., (1998), Asymptotics for Out of Sample Tests of Causality, Working Paper, Louisiana State University.

Meese, R.A. and K. Rogoff, (1983), Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?, *Journal of International Economics*, 14, 3-24.

Mizrach, B., (1992), The Distribution of the Theil U-Statistic in Bivariate Normal Populations, *Economics Letters*, 38, 163-167.

Newbold, P., (1982), Causality Testing in Economics, in: *Time Series Analysis: Theory and Practice I*, ed. O.D. Anderson, Amsterdam: North Holland.

Pierce, D.A. and L.D. Haugh, (1977), Causality in Temporal Systems: Characterization and a Survey, *Journal of Econometrics*, 5, 265-294.

Sims, C.A., (1972), Money, Income and Causality, *The American Economic Review*, 62, 540-552.

Stinchcombe, M.B. and H. White, (1998), Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative, *Econometric Theory,* 14, 3, 295-325.

Stock, J.H. and M.M. Watson, (1989), Interpreting the Evidence on Money-Income Causality, *Journal of Econometrics*, 40, 161-181.

Swanson, (1998), Money and Output Viewed Through a Rolling Window, *Journal of Monetary Economics* 41, 455-474.

Thoma, M.A., (1994), Subsample Instability and Asymmetries in Money-Income Causality, *Journal of Econometrics*, 64, 279-306.

Weiss, A.A., (1996), Estimating Time Series Models Using the Relevant Cost Function, *Journal of Applied Econometrics*, 11, 539-560.

West, K.D., (1996), Asymptotic Inference About Predictive Ability, *Econometrica*, 64, 1067-1084.

West, K. and M.W. McCracken, (1998), Regression Based Tests of Predictive Ability, *International Economic Review*, 39, 817-840.

White, H., (1999), A Reality Check for Data Snooping, *Econometrica*, forthcoming.

Wiener, N., (1958), The Theory of Prediction, in: *Modern Mathematics for Engineers,* Series 1, Chapter 8, eds. E.F. Bec.

Wooldridge, J.M., (1990), A Unified Approach to Robust Regression-Based Specification Tests, *Econometric Theory*, 6, 17-43.

Wooldridge, J.M., (1991), On the Application of Robust, Regression-Based Diagnostics to Models of Conditional Means and Conditional Variances, *Journal of Econometrics,* 47, 1, 5-46.

Yokohama, R. (1980), Moment Bounds for Stationary Mixing Sequences, *Probability Theory and Related Fields,* 45-57.

Table 1: Empirical Size of the Wald F Test*

In-Sample Wald F Test Results Based on Entire Sample			
	Sample Size (T)		
π_2	250	500	1000
0.100	0.118	0.094	0.099
0.300	0.118	0.090	0.095
0.500	0.113	0.098	0.108
0.700	0.107	0.096	0.117
0.900	0.117	0.117	0.109

* Notes: All entries are rejection frequencies of the null hypothesis of Granger noncausality based on 10% nominal size in-sample Wald F-tests. Data were generated as discussed in Section 3, with $\pi_3 = 0$ so that x_t is not Granger causal for y_t. All experiments are repeated for samples of 250, 500, and observations, and all entries are based on 10000 Monte Carlo replications.

Table 2: Empirical Size of the $v^{-1/2}m_P(\pi = 0)$ Test*

Out-of-Sample Predictive Ability Test Results Based on Sample Size P									
	Sample Size								
	$P = 0.1T$			$P = 0.3T$			$P = 0.5T$		
π_2	250	500	1000	250	500	1000	250	500	1000
0.100	0.151	0.131	0.131	0.127	0.118	0.102	0.107	0.107	0.100
0.300	0.143	0.125	0.127	0.124	0.116	0.108	0.101	0.111	0.102
0.500	0.144	0.124	0.124	0.117	0.118	0.106	0.095	0.110	0.093
0.700	0.146	0.117	0.128	0.123	0.120	0.108	0.096	0.114	0.101
0.900	0.165	0.128	0.129	0.127	0.126	0.116	0.126	0.121	0.111

* Notes: See notes to Table 1. All entries are rejection frequencies of the null hypothesis of Granger noncausality based on 10% nominal size out-of-sample predictive ability tests (i.e properly rescaled m_P statistics, say $v^{-1/2}m_P$). It is assumed that $\pi = 0$, so that parameter estimation error is not accounted for.

Table 3: Empirical Size of the $v^{-1/2}m_P(\pi > 0)$ Test*

Out-of-Sample Predictive Ability Test Results Based on Sample Size P									
	Sample Size								
	$P = 0.1T$			$P = 0.3T$			$P = 0.5T$		
π_2	250	500	1000	250	500	1000	250	500	1000
0.100	0.138	0.128	0.115	0.112	0.100	0.089	0.081	0.088	0.086
0.300	0.128	0.119	0.120	0.092	0.091	0.084	0.077	0.078	0.075
0.500	0.121	0.104	0.112	0.088	0.092	0.084	0.066	0.075	0.074
0.700	0.125	0.105	0.107	0.083	0.082	0.077	0.057	0.070	0.067
0.900	0.134	0.106	0.106	0.084	0.088	0.077	0.068	0.074	0.062

* Notes: See notes to Table 2. All entries are rejection frequencies of the null hypothesis of Granger noncausality based on 10% nominal size out-of-sample predictive ability tests (m_P). It is assumed that $\pi > 0$, so that parameter estimation error is accounted for.

Table 4: Empirical Power of the Wald F Test*

In-Sample Wald F Test Results Based on Entire Sample						
				Sample Size (T)		
A	B	π_2	π_3	250	500	1000
0.100	0.108	0.100	0.330	1.000	1.000	1.000
0.100	0.988	0.300	8.235	1.000	1.000	1.000
0.500	0.029	0.100	0.141	0.715	0.940	0.999
0.500	0.234	0.300	0.431	1.000	1.000	1.000
0.500	0.628	0.500	0.919	1.000	1.000	1.000
1.000	0.020	0.100	0.100	0.461	0.718	0.929
1.000	0.154	0.300	0.288	1.000	1.000	1.000
1.000	0.360	0.500	0.459	1.000	1.000	1.000
1.000	0.927	0.700	1.803	1.000	1.000	1.000
5.000	0.012	0.100	0.044	0.170	0.253	0.406
5.000	0.091	0.300	0.123	0.641	0.901	0.991
5.000	0.194	0.500	0.174	0.924	0.997	1.000
5.000	0.271	0.700	0.178	0.983	1.000	1.000
5.000	0.556	0.900	0.199	1.000	1.000	1.000
10.000	0.011	0.100	0.031	0.138	0.177	0.246
10.000	0.083	0.300	0.087	0.402	0.633	0.883
10.000	0.176	0.500	0.121	0.703	0.929	0.996
10.000	0.233	0.700	0.119	0.819	0.967	1.000
10.000	0.228	0.900	0.071	0.800	0.964	1.000

* Notes: See notes to Table 1. All entries are rejection frequencies of the null hypothesis of Granger noncausality based on 10% nominal size in-sample Wald F-tests. Values of the parameter B are constructed as discussed above by fixing A and π_2 and then solving for π_3.

Table 5: Empirical Power of the $v^{-1/2}m_P(\pi = 0)$ Test*

Out-of-Sample Predictive Ability Test Results Based on Sample Size P												
				Sample Size								
				$P = 0.1T$			$P = 0.3T$			$P = 0.5T$		
A	B	π_2	π_3	250	500	1000	250	500	1000	250	500	1000
0.100	0.108	0.100	0.330	0.310	0.462	0.692	0.560	0.844	0.977	0.753	0.955	0.999
0.100	0.988	0.300	8.235	0.952	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.500	0.029	0.100	0.141	0.180	0.177	0.281	0.218	0.311	0.507	0.275	0.444	0.686
0.500	0.234	0.300	0.431	0.353	0.544	0.788	0.660	0.914	0.998	0.863	0.985	1.000
0.500	0.628	0.500	0.919	0.498	0.807	0.975	0.916	0.999	1.000	0.988	1.000	1.000
1.000	0.020	0.100	0.100	0.164	0.148	0.205	0.164	0.220	0.323	0.199	0.301	0.448
1.000	0.154	0.300	0.288	0.245	0.343	0.546	0.414	0.697	0.905	0.580	0.859	0.980
1.000	0.360	0.500	0.459	0.287	0.475	0.722	0.592	0.883	0.991	0.800	0.975	1.000
1.000	0.927	0.700	1.803	0.266	0.536	0.911	0.764	0.991	1.000	0.950	1.000	1.000
5.000	0.012	0.100	0.044	0.151	0.125	0.139	0.130	0.141	0.161	0.128	0.149	0.195
5.000	0.091	0.300	0.123	0.160	0.152	0.225	0.184	0.246	0.379	0.212	0.346	0.527
5.000	0.194	0.500	0.174	0.172	0.195	0.266	0.226	0.323	0.525	0.281	0.453	0.702
5.000	0.271	0.700	0.178	0.159	0.170	0.239	0.192	0.291	0.451	0.244	0.401	0.623
5.000	0.556	0.900	0.199	0.201	0.134	0.134	0.137	0.149	0.217	0.140	0.206	0.344
10.000	0.011	0.100	0.031	0.143	0.127	0.130	0.131	0.124	0.131	0.121	0.123	0.150
10.000	0.083	0.300	0.087	0.157	0.135	0.175	0.148	0.176	0.243	0.162	0.244	0.330
10.000	0.176	0.500	0.121	0.161	0.153	0.198	0.180	0.227	0.323	0.185	0.327	0.454
10.000	0.233	0.700	0.119	0.157	0.139	0.181	0.161	0.214	0.284	0.171	0.268	0.386
10.000	0.228	0.900	0.071	0.180	0.126	0.145	0.141	0.139	0.152	0.138	0.155	0.188

* Notes: See notes to Table 2. All entries are rejection frequencies of the null hypothesis of Granger noncausality based on 10% nominal size out-of-sample predictive ability tests (m_P). It is assumed that $\pi = 0$, so that parameter estimation error is not accounted for. Values of the parameter B are constructed as discussed above by fixing A and π_2 and then solving for π_3.

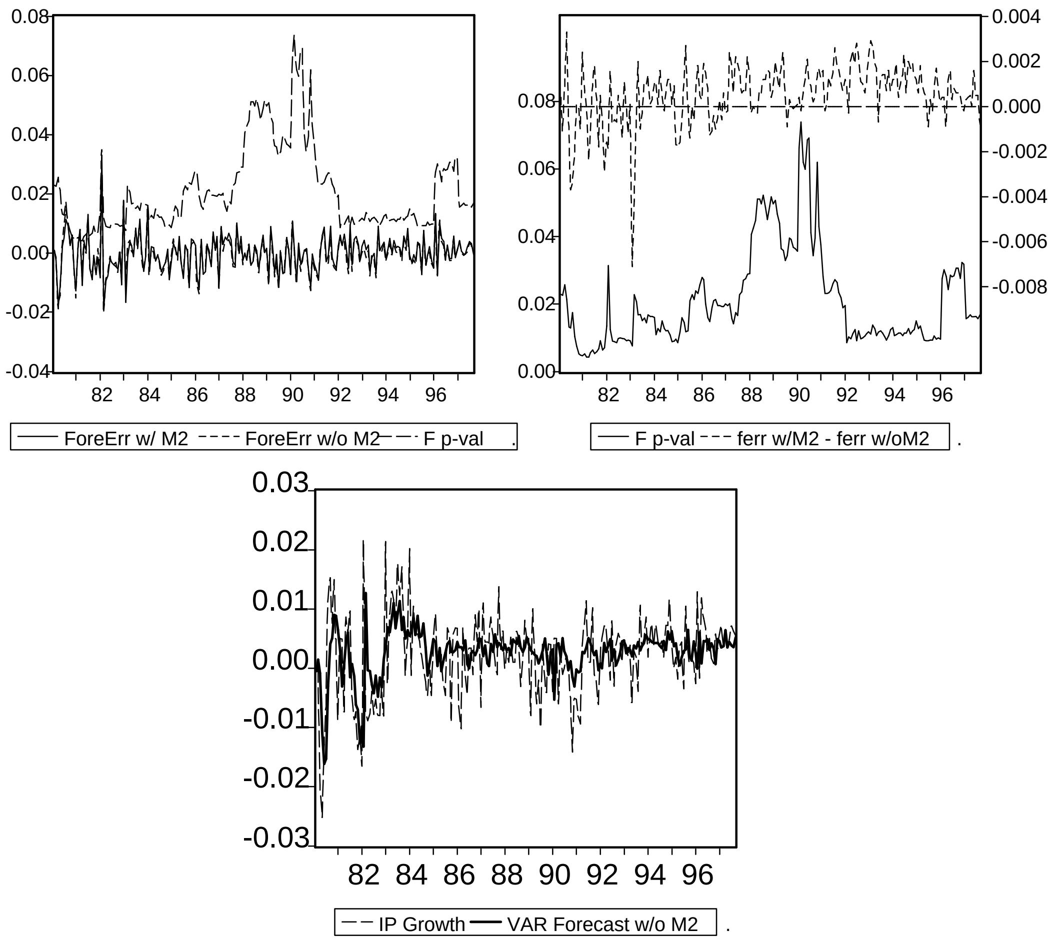
Table 6: Empirical Power of the $v^{-1/2}m_P(\pi > 0)$ Test*

Out-of-Sample Predictive Ability Test Results Based on Sample Size P												
				Sample Size								
				$P = 0.1T$			$P = 0.3T$			$P = 0.5T$		
A	B	π_2	π_3	250	500	1000	250	500	1000	250	500	1000
0.100	0.108	0.100	0.330	0.291	0.448	0.684	0.532	0.826	0.975	0.713	0.940	0.999
0.100	0.988	0.300	8.235	0.942	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.500	0.029	0.100	0.141	0.164	0.168	0.269	0.195	0.284	0.479	0.236	0.410	0.649
0.500	0.234	0.300	0.431	0.323	0.519	0.775	0.613	0.892	0.995	0.811	0.976	1.000
0.500	0.628	0.500	0.919	0.470	0.776	0.970	0.883	0.998	1.000	0.977	1.000	1.000
1.000	0.020	0.100	0.100	0.150	0.135	0.195	0.152	0.197	0.290	0.176	0.258	0.401
1.000	0.154	0.300	0.288	0.216	0.323	0.533	0.368	0.650	0.882	0.525	0.817	0.973
1.000	0.360	0.500	0.459	0.257	0.447	0.704	0.517	0.851	0.982	0.720	0.950	0.999
1.000	0.927	0.700	1.803	0.245	0.515	0.903	0.719	0.988	1.000	0.932	1.000	1.000
5.000	0.012	0.100	0.044	0.137	0.116	0.129	0.114	0.119	0.141	0.105	0.118	0.161
5.000	0.091	0.300	0.123	0.147	0.144	0.206	0.158	0.214	0.340	0.172	0.306	0.468
5.000	0.194	0.500	0.174	0.155	0.172	0.252	0.186	0.281	0.460	0.209	0.395	0.637
5.000	0.271	0.700	0.178	0.143	0.147	0.220	0.148	0.234	0.389	0.172	0.311	0.522
5.000	0.556	0.900	0.199	0.174	0.109	0.114	0.096	0.111	0.161	0.094	0.144	0.255
10.000	0.011	0.100	0.031	0.138	0.117	0.123	0.116	0.110	0.113	0.095	0.096	0.124
10.000	0.083	0.300	0.087	0.139	0.126	0.164	0.124	0.157	0.216	0.131	0.189	0.291
10.000	0.176	0.500	0.121	0.143	0.129	0.178	0.144	0.185	0.280	0.140	0.249	0.374
10.000	0.233	0.700	0.119	0.133	0.117	0.157	0.115	0.159	0.227	0.119	0.191	0.315
10.000	0.228	0.900	0.071	0.154	0.106	0.128	0.093	0.096	0.106	0.080	0.088	0.114

* Notes: See notes to Table 2. All entries are rejection frequencies of the null hypothesis of Granger noncausality based on 10% nominal size out-of-sample predictive ability tests (m_P). It is assumed that $\pi > 0$, so that parameter estimation error is accounted for.

Figure 1: Real Time Industrial Production Forecast Results

(p-values are for in-sample Granger noncausality tests)



Notes: p-values were calculated recursively starting with the sample 1961:1-1980:1 and ending with data for the period 1960:1-1997:8. All results are based on models and forecasts of the monthly growth rate in U.S. industrial production.