Preliminary and incomplete

Tests of Equal Forecast Accuracy and Encompassing for Nested Models

Todd E. Clark and Michael W. McCracken*

April 1999

(current draft: April 7)

Abstract

This paper examines the asymptotic and finite-sample properties of out-of-sample tests for equal accuracy and encompassing as applied to nested models. With nested models, these tests can be viewed as Granger causality tests. Applied to nested models, however, the standard asymptotic critical values for many tests of equal accuracy and encompassing are invalid. Statistics such as those proposed by Diebold and Mariano (1995) and Harvey, et. al. (1998) fail to converge to the standard normal distribution when the models are nested rather than non-nested. Building on McCracken's (1999) results for equal accuracy tests, this paper derives the asymptotic distributions for a set of standard encompassing tests and one new encompassing test. Numerical simulations are used to generate the appropriate asymptotic critical values. Monte Carlo simulations are then used to evaluate the size and power of a battery of equal forecast accuracy and encompassing tests, as well as standard F-tests of causality. In these experiments, forecasts from an estimated VAR model are compared to those from a null estimated AR model. The simulation results indicate that McCracken's out-of-sample F-type test of equal accuracy and the encompassing test proposed in this paper can be more powerful than standard F-tests of causality. The Monte Carlo simulations also show that using invalid asymptotic critical values can produce misleading inferences.

JEL Nos.: C53, C12, C52

******Clark*: Economic Research Dept., Federal Reserve Bank of Kansas City, Kansas City, MO 64198, tclark@frbkc.org. *McCracken*: Dept. of Economics, Louisiana State University, 2107 CEBA Building, Baton Rouge, LA 70803, mmccrac@unix1.sncc.lsu.edu. The research assistance of Hilary Croke and the helpful comments of Lutz Kilian and Ken West are gratefully acknowledged. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Kansas City or the Federal Reserve System.

1. INTRODUCTION

As evident from recent studies such as Amano and van Norden (1995), Blomberg and Hess (1997), Bram and Ludvigson (1998), Krueger and Kuttner (1996), and Mark (1995), interest often lies in examining whether one variable helps predict another, both in-sample or out-of-sample. The standard in-sample metric is a simple Granger causality test. Out-of-sample predictive ability is usually gauged by first constructing forecasts from models that include and exclude the variable that may have predictive capacity and then testing for equal accuracy or encompassing. Typically, there is concern that in-sample Granger causality tests may lead to overfitting of the model. Out-of-sample forecast comparison is widely viewed as a more stringent test of the relationship between the variables. Moreover, Ashley, Granger, and Schmalensee (1980) argue that it is more in the spirit of their "reasonable" definition of Granger causality to employ post-sample forecast tests than to employ the standard in-sample tests of causality.[1] According to Ashley, et. al., if forecasts of y from a VAR in x and y are superior to forecasts from an AR model for y, then x carries information about y and hence x causes y.

This paper examines the ability of different post-sample forecast tests to determine whether one variable has predictive ability for another. Particularly, this paper examines the asymptotic and finite-sample properties of tests for equal accuracy and encompassing applied to nested models. Used with nested forecasts from models such as a VAR vs. an AR, the tests can be viewed as Granger causality tests. However, many of the standard tests of equal accuracy and encompassing are designed for forecasts from *non-nested*, rather than *nested,* models. Many of the standard test statistics – such as the Diebold and Mariano (1995) equal accuracy and the Harvey, Leybourne, and Newbold (1998) encompassing statistics – fail to converge to the

[1] Diebold and Mariano (1995) also suggest using out-of-sample forecast tests to examine Granger causality.

standard normal distribution when the models are nested rather than non-nested.[2] Therefore, the standard asymptotic critical values are invalid with nested models.

This study's analysis of the use of equal accuracy and encompassing tests for nested forecasts complements the non-nested forecast analyses of, among others, Corradi, Swanson, and Olivetti (1998), Diebold and Mariano (1995), Harvey, Leybourne, and Newbold (1997, 1998), West (1996), West and McCracken (1998) and McCracken (1998). In another related analysis, Swanson, Ozyildirim, and Pisu (1996) examine the finite-sample performance – principally, the size – of different Granger causality tests and Diebold-Mariano equal accuracy tests with both stationary and non-stationary data. The equal accuracy tests they consider, however, are all compared against standard asymptotic critical values that are invalid because the models are nested. The same problem applies to the Monte Carlo results of Corradi, et. al. (1998) on how Diebold and Mariano tests perform when applied to models with cointegrating relationships.

Building on McCracken's (1999) results for equal accuracy tests, this paper first derives the asymptotic distributions for a set of standard encompassing tests and one new encompassing test. The standard encompassing tests for which this paper derives asymptotic distributions are the Harvey, et. al. (1998) and Ericsson (1992) statistics. The set of standard statistics also includes the Chong and Hendry (1986) test, which remains asymptotically normal when applied to nested rather than non-nested forecasts. The new test proposed below is a variant of the Harvey, et. al. statistic. As in Corradi, et. al. (1998), McCracken (1999), West (1996), and West and McCracken (1998), the derived distributions of the tests explicitly account for the uncertainty introduced by model estimation. In order to facilitate the use of the limiting distributions derived here, asymptotically valid critical values are generated numerically and

[2] Most of the other available tests of forecast accuracy or encompassing – such as the Mizrach (1992) and Granger and Newbold (1977) tests of equal accuracy – are also asymptotically invalid when applied to nested forecasts.

reported in appendix tables. The equal accuracy tests include an out-of-sample F-type test of equal mean squared error (MSE) developed in McCracken (1999) and the Diebold and Mariano (1995) test of equal MSE, statistics for which McCracken develops the correct asymptotic distributions and provides critical values.

In order to evaluate the finite-sample size and size-adjusted power of these tests we conduct a series of Monte Carlo simulations based on VAR data-generating processes. For comparison, the set of tests considered also includes standard F-tests of Granger causality. In addition, in order to evaluate the extent to which using invalid critical values can produce misleading inferences, results are presented for Diebold and Mariano (1995), Harvey, et. al. (1998), and Ericsson (1992) statistics compared to the distributions that would be appropriate if the forecasts were from non-nested models. To further illustrate how the different tests perform in practical settings, the battery of tests is applied to determining whether the unemployment rate has predictive power for inflation in quarterly U.S. data.

Results summary. Asymptotics. Finite-sample. Monte Carlo simulations show that using invalid asymptotic critical values can produce misleading inferences in small samples. The simulations also indicate that out-of-sample F-type and encompassing tests can be more powerful than standard, in-sample F-tests of causality.

The remainder of the paper will proceed as follows. Section two introduces the notation and general environment under which the forecasts are generated and the tests of forecast accuracy and encompassing are constructed. Section three, and its subsections, introduce the test statistics considered and provide the asymptotic results under the null. In section four we present a collection of Monte Carlo experiments designed to determine the finite-sample size and power properties of the test statistics. Section five contains an empirical application of the tests to the problem of determining whether the unemployment rate has predictive power for inflation in quarterly U.S. data. Section six concludes. All proofs are contained within the Appendix.

2. General Environment

In order to present the tests considered we first provide some general notation, describe the forecasting schemes, and present the assumptions under which the asymptotic results are derived. The sample of data $\{y_j, x_j'\}_{j=1}^{T+1}$ is divided into in-sample and out-of-sample portions. The in-sample observations span 1 to *R*. Letting *P* denote the number of 1-step ahead predictions constructed, the out-of-sample observations span $R+1$ through $R+P$. The total number of observations in the sample is then $R+P=T+1$. The largest number of observations used to estimate the model under the forecast schemes considered is $T=R+P-1$.

The scalar variable to be predicted is y_{t+1}, t = R,…,T. Forecasts of y_{t+1} are generated using parametric models $g_i(x_{t+1}, \beta_i^*) \equiv g_{i,t+1}(\beta_i^*)$ denoted by i = 1 and 2, each of which is estimated. Model 2 is the unrestricted model, which nests the restricted model 1. Under the null hypothesis, model 2 includes k_2 excess parameters. Without loss of generality let $\beta_2^* = (\beta_{1\,1\times k_1}^{*'}, 0_{1\times k_2})'$ ($k_1 + k_2 = k\times 1$) such that for all t, $g_{1,t+1}(\beta_1^*) = g_{2,t+1}(\beta_2^*)$. Under the alternative hypothesis, the k_2 restrictions are not true, and model 2 is correct. Note that while models 1 and 2 take the form of AR and VAR models in the Monte Carlo analysis, the asymptotic results permit the use of nonlinear models.

Following West and McCracken (1998), three forecast schemes are considered. Under the *recursive* scheme, each model's parameters, β_i^* i = 1,2, are estimated with added data as forecasting moves forward. The first prediction, $g_{i,R+1}(\hat{\beta}_{i,R})$, is created using model parameter estimates $\hat{\beta}_{i,R}$ estimated using data from 1 to R, the second prediction $g_{i,R+2}(\hat{\beta}_{i,R+1})$ is created using model parameter estimates $\hat{\beta}_{i,R+1}$ estimated using data from 1 to $R+1$, etc. In general, for t = R,…,T, the prediction of y_{t+1}, $g_{i,t+1}(\hat{\beta}_{i,t})$, from time t is created using model parameter estimates $\hat{\beta}_{i,t}$ estimated using data from 1 to t.

Under the *rolling* forecast scheme, the model is estimated using only the most recent R observations. The first rolling prediction, $g_{i,R+1}(\hat{\beta}_{i,R})$, is created using model parameter estimates $\hat{\beta}_{i,R}$ estimated using data from 1 to R, the second prediction $g_{i,R+2}(\hat{\beta}_{i,R+1})$ is created using model parameter estimates $\hat{\beta}_{i,R+1}$ estimated using data from 2 to $R+1$, etc. In general, for t = R,…,T, the prediction of y_{t+1}, $g_{i,t+1}(\hat{\beta}_{i,t})$, from time t is created using model parameter estimates $\hat{\beta}_{i,t}$ estimated using data from t–R+1 to t. Note that under the rolling scheme the parameter estimates $\hat{\beta}_{i,t}$ should also be subscripted by R in order to reflect the size of the sample window. To reduce notation we leave that subscript implicit.

Under the *fixed* scheme, all forecasts are generated using models estimated with data from 1 to R. Hence for each prediction of y_{t+1}, $g_{i,t+1}(\hat{\beta}_{i,t}) = g_{i,t+1}(\hat{\beta}_{i,R})$, from time t = R,…,T, the prediction is created using the same model parameter estimate $\hat{\beta}_{i,t} = \hat{\beta}_{i,R}$ estimated using data from 1 to R. As was the case for the rolling scheme, the parameter estimates $\hat{\beta}_{i,t}$ under the fixed scheme should also be subscripted by R to reflect the sample window. To reduce notation we also leave this subscript implicit.

For each of the three forecasting schemes, the 1-step ahead forecast errors are $\hat{u}_{1,t+1} = y_{t+1} - g_{1,t+1}(\hat{b}_{1,t})$ and $\hat{u}_{2,t+1} = y_{t+1} - g_{2,t+1}(\hat{b}_{2,t})$ for models 1 and 2, respectively. Using the two sequences of P forecast errors the out-of-sample tests of forecast accuracy and encompassing are constructed. In all cases the out-of-sample statistics rely on sums of functions of these forecast errors. To simplify notation, for any variable z_t we will let $\sum_t z_t$ denote the summation $\sum_{t=R}^{T} z_t$. For example, the mean squared error (MSE) for model i is $MSE_i \equiv$

$P^{-1}\sum_{t=R}^{T}\hat{u}_{i,t}^2 = P^{-1}\sum_t \hat{u}_{i,t}^2$.

Before getting to the assumptions some final notation is needed. Let $g_{i,t+1,\beta}(\beta_i) =$

$\frac{\partial}{\partial\beta_i} g_{i,t+1}(\beta_i)$, $q_{i,t+1}(\beta_i) = g_{i,t+1,\beta}(\beta_i)g_{i,t+1,\beta}^{'}(\beta_i) - (y_{t+1} - g_{i,t+1}(\beta_i))\frac{\partial^2}{\partial\beta_i\partial\beta_i^{'}} g_{i,t+1}(\beta_i)$, $f_{i,t+1} =$

$f_{i,t+1}(\beta_i^*)$ and $f_{t+1} = f_{2,t+1}$ for any function f, $h_{i,t+1}(\beta_i) = (y_{t+1} - g_{i,t+1}(\beta_i))g_{i,t+1,\beta}(\beta_i)$, $B_i =$

$(Eq_{i,t+1})^{-1}$, W(s) is a $(k_2\times 1)$ vector standard Brownian Motion, and for any $(m\times n)$ matrix A with column vectors a_i let vec(A) denote the $(mn\times 1)$ vector $[a_1^{'}, a_2^{'}, ..., a_n^{'}]^{'}$.

Given the definitions and the three forecasting schemes described above, the following five assumptions are those used to derive the limiting distributions of encompassing tests presented in Theorems 3.4, 3.5, and 3.6. The assumptions are also sufficient for the results of McCracken (1999) when MSE is the measure of predictive ability. These assumptions are not intended to be necessary and sufficient, only sufficient. All proofs can be found within the Appendix.

Assumption 1: The parameter estimates $\hat{\beta}_{i,t}$, i = 1,2, t = R,…,T, satisfy $\hat{\beta}_{i,t} - \beta_i^* = B_i(t)H_i(t)$ where for $\dot{\beta}_{i,t}$ on the line between $\hat{\beta}_{i,t}$ and β_i^*, $B_i(t)H_i(t)$ equals

$(t^{-1}\sum_{j=1}^{t} q_{i,j}(\dot{\beta}_{i,t}))^{-1}(t^{-1}\sum_{j=1}^{t} h_{i,j})$, $(R^{-1}\sum_{j=t-R+1}^{t} q_{i,j}(\dot{\beta}_{i,t}))^{-1}(t^{-1}\sum_{j=t-R+1}^{t} h_{i,j})$ and $(R^{-1}\sum_{j=1}^{R} q_{i,j}(\dot{\beta}_{i,t}))^{-1}(R^{-1}\sum_{j=1}^{R} h_{i,j})$ respectively for the recursive, rolling and fixed schemes.

This first assumption provides us with one primary piece of information. Analytically it tells us that the parameters are estimated by OLS, NLLS, or maximum likelihood under normality assumptions. In the case where a VAR is being used, the system must be exactly identified and estimated by multivariate OLS. This type of restriction is imposed to insure that the statistics in Theorems 3.4-3.6 are pivotal. As in McCracken (1999), achieving a limiting distribution that does not depend upon the data-generating process requires that the loss function used to estimate the parameters be closely related to the loss function used to measure predictive ability. Each of the statistics in Theorems 3.4-3.6 is in one way or another testing whether the two models have equal mean square errors. In order then to achieve a pivotal statistic the parameters must be estimated using mean square error as the loss function. Although this assumption is restrictive in how the parameters are estimated, it otherwise does not place any restrictions on the type of model. Single and multiple equation models as well as linear and nonlinear models are permitted.

Assumption 2: For i = 1,2, (a) $\beta_i \in \Theta_i$, Θ_i compact, (b) $E[y_t - g_{i,t}(\beta_i)]^2$ is uniquely minimized at $\beta_i^* \in \Theta_i$ with $Eq_{i,t}$ nonsingular, (c) In some open neighborhood N_i around β_i^*, and with probability one $[y_t - g_{i,t}(\beta_i)]^2$ is twice continuously differentiable, (d) In the open neighborhood N_i, and for all t there exists a positive constant φ and a positive random variable m_t such that $|q_{i,t}(\beta_i) - q_{i,t}(\beta_i^*)| \le m_t |\beta_i - \beta_i^*|^{\varphi}$ with $Em_t < \infty$ and $\varphi < \infty$.

Most of Assumption 2 is imposed in order to insure that the parameters are identified and are consistently estimated. It is directly comparable to Theorem (2.1) of Newey and McFadden (1994). The substantive components of this assumption are that the predictive function, $g_{i,t}(\beta_i)$, is the conditional mean function and that the conditional mean function is twice continuously differentiable.

<u>Assumption 3:</u> Let $U_t \equiv [h_t^{'}, vec(h_t h_t^{'} - \sigma^2 B^{-1})^{'}, vec(q_t - B^{-1})^{'}]^{'}$. (a) $EU_t = 0$, (b) U_t is uniformly L^8 bounded, (c) For some $8 > d > 2$, U_t is strong mixing with coefficients of size $\frac{-8d}{8-d}$, (d) $\lim_{T\to\infty} T^{-1}E\sum_{j=1}^{T} U_j U_j^{'} = \Omega < \infty$.

<u>Assumption 4:</u> (a) $Eh_t h_t^{'} = \sigma^2 Eq_t \equiv \sigma^2 B^{-1}$, (b) $E(h_t \mid h_{t-j}, q_{t-j}, j = 1,2,...) = 0$.

Both Assumptions 3 and 4 largely consist of technical conditions sufficient for the application of an invariance principle. Moreover they are sufficient for joint weak convergence of partial sums and averages of these partial sums to Brownian Motion and integrals of these Brownian Motion. Assumption 3 is directly comparable to the assumptions in Hansen (1992) and hence we are able to apply his Theorems (2.1) and (3.1).

The reasons for imposing Assumption 4 are much the same as Assumption 1. In order to insure that the limiting distribution does not depend upon the underlying data generating process we must impose some extra conditions. Here we essentially require that the disturbances form a conditionally homoskedastic martingale difference sequence.

Assumption 5: $\lim_{T\to\infty} P/R = \pi$, $0 < \pi < \infty$, $\lambda \equiv (1+\pi)^{-1}$.

This final assumption introduces the means by which the asymptotics are achieved. As in Hoffman and Pagan (1989), West (1996), and White (1998) the limiting distribution results are derived by imposing a slightly stronger condition than simply that the sample size, T+1, becomes arbitrarily large. Here we impose the additional condition that both the number of in-sample (R) and out-of-sample (P) observations also become arbitrarily large at the same rate. In this way we insure that the parameters estimated in-sample and certain out-of-sample averages are both consistent estimators of their population level analogs.

It should be noted that the assumption that π is bounded from above and below is not trivial. Certainly in practice P/R will be bounded but whether it is near zero or much larger could affect how well the asymptotic approximation behaves in finite samples. If P/R is small then the parameters may be well estimated but, for example, the out-of-sample MSE will be estimated by too few observations for the empirical MSE to form a strong estimate of the population MSE. If P/R is large then we cannot expect the parameters to be well approximated, especially under the fixed scheme, and thus regardless of the out-of-sample size the empirical MSE may form a poor estimate of the population MSE. Hence when choosing how to split the sample into in-sample and out-of-sample portions one should consider choosing a split that leaves a sizable number of observations in each of the in-sample and out-of-sample portions.

Unless otherwise noted, the notation and assumptions presented in this section hold throughout the remainder of the paper.

3. TESTS

While Ashley, et. al. (1980) specifically advocate using tests of equal forecast accuracy to examine causality, given their definition of causality, any test designed to examine whether x carries information about y could reasonably be used. Accordingly, this paper considers the ability of simple Granger causality tests, equal forecast accuracy tests, and forecast encompassing tests to determine whether one variable has predictive power for another. Since a large number of tests for equal accuracy and encompassing already exist, for tractability the set examined is limited based on considerations of computational simplicity and performance in the non-nested investigations of Clark (1999), Diebold and Mariano (1995), and Harvey, Leybourne, and Newbold (1997, 1998). The set of tests includes: simple Granger causality statistics; McCracken's (1999) out-of-sample F-type test for equal MSE; Diebold and Mariano's statistic for equal MSE; the Harvey, et. al. (1998) encompassing test; the Ericsson (1992) encompassing statistic; a modified, asymptotically valid version of the Harvey, et. al. test developed below; and the Chong and Hendry (1986) encompassing statistic.

In the results below, the tests are applied to 1-step ahead forecasts. When multi-step forecasts from nested models are used, the asymptotic distributions of the tests appear to depend on the parameters of the data-generating process. For practical purposes, such dependence eliminates the possibility of using asymptotic approximations to test for equal accuracy or encompassing. Lutkepohl and Burda (1997) note similar difficulties associated with in-sample tests involving multi-step forecasts. Our belief is that researchers comfortable with assuming linear models should be adequately served by tests based on 1-step ahead forecasts. With linear models, multi-step forecasts are simply linear combinations of 1-step ahead forecasts. Hence there does not appear to be any reason to expect tests based on multi-step forecasts to be better at determining predictive power than tests based on 1-step ahead forecasts.

3.1 Simple Granger Causality (GC) Tests

The emphasis of this paper is on using ex-ante forecasts, rather than ex-post predictions, to test for forecast accuracy and encompassing. Even so, we include simple F-type Granger causality tests in the Monte Carlo simulations. We do so because they are the most commonly used statistics for testing for causality. We construct these statistics using both in-sample and out-of-sample data, using R observations in the former case and P observations in the latter.

In the results of section 4, the tests are computed as simple F-statistics for exclusion restrictions.[3] When lag lengths are set using data-based procedures, the out-of-sample GC tests rely on the lag order determined using the in-sample data. While standard GC tests are rarely applied to out-of-sample data, they are no less valid for the purpose of testing causality in out-of-sample data than are forecast accuracy or encompassing tests. Like out-of-sample accuracy and encompassing tests, simple out-of-sample GC tests may be less prone to spurious results due to overfitting than are in-sample causality tests.

3.2 The Out-of-Sample F (OOS F) Test

McCracken (1999) develops an out-of-sample F-type test of equal MSE, given by

$$OOS\ F = P \cdot \frac{MSE_1 - MSE_2}{MSE_2} = P \cdot \frac{P^{-1}\sum_t \hat{u}_{1,t+1}^2 - P^{-1}\sum_t \hat{u}_{2,t+1}^2}{P^{-1}\sum_t \hat{u}_{2,t+1}^2}. \qquad (1)$$

This statistic is comparable to the simple F-test form of the standard in-sample GC test and offers the advantage of being particularly simple to compute if forecast summary statistics are already available. Using assumptions broadly similar to those used in this paper, McCracken shows that the OOS F statistic converges to a function of stochastic integrals of quadratics of Brownian motion. The limiting distribution under the null, which varies with the forecasting scheme, is a function of the ratio of post-sample to in-sample observations, π, and excess parameters, k_2, in model 2.

In the results of section 4, the test statistic is compared against asymptotically valid critical values tabulated by McCracken. Since the models are nested, the null hypothesis is $\mathrm{MSE}_1 \leq \mathrm{MSE}_2$, and the alternative is $\mathrm{MSE}_1 > \mathrm{MSE}_2$. The alternative is one-sided because, if the restrictions imposed on model 1 are not true, there is no reason to expect forecasts from model 1 to be superior to those from model 2.

3.3 The Diebold-Mariano (DM) Test

Define $d_{t+1} = \hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2$ and $\bar{\mathrm{d}} = \mathrm{P}^{-1}\sum_t \mathrm{d}_{t+1} = \mathrm{MSE}_1 - \mathrm{MSE}_2$. The Diebold and Mariano (1995) test for equal MSE is formed as

$$DM = \frac{MSE_1 - MSE_2}{\sqrt{\hat{\mathrm{v}}\mathrm{ar}(MSE_1 - MSE_2)}} = \frac{\bar{d}}{\sqrt{\hat{\mathrm{v}}\mathrm{ar}(\bar{d})}} = \frac{\bar{d}}{\sqrt{P^{-2}\sum_t (d_{t+1} - \bar{d})^2}}. \qquad (2)$$

While the DM statistic is asymptotically standard normal when applied to non-nested forecasts (see Diebold and Mariano (1995) and West (1996)), the asymptotic distribution is non-normal when the forecasts are nested under the null hypothesis.

The root of the problem is that, under the null, $\mathrm{g}_{1,t+1}(\beta_1^*) = \mathrm{g}_{2,t+1}(\beta_2^*)$ and thus both $u_{1,t+1} = y_{t+1} - g_{1,t+1}(b_1^*) \equiv u_{t+1}$ and $u_{2,t+1} = y_{t+1} - g_{2,t+1}(b_2^*) = y_{t+1} - g_{1,t+1}(b_1^*) \equiv u_{t+1}$. Hence, at

[3] Computing the tests as Wald statistics and comparing them against the chi-square distribution would make the tests robust to non-normally distributed data. However, using the Wald statistics produces results very similar to those

least heuristically, asymptotically the difference in squared forecast errors is exactly 0, with 0 variance. The usual DM test – in which the statistic is compared against the standard normal distribution – is therefore asymptotically invalid. McCracken (1999) shows that, for forecasts from nested models, the DM test statistic converges to a function of stochastic integrals of quadratics of Brownian motion. This limiting distribution depends on the forecasting scheme, π, and k_2.

In the results of section 4, the test statistic is compared against asymptotically valid critical values tabulated by McCracken. Because models 1 and 2 are nested rather than non-nested, the alternative hypothesis is one-sided instead of two-sided. Under the null, $MSE_1 \leq MSE_2$; under the alternative, $MSE_1 > MSE_2$.

To evaluate how using invalid asymptotic critical values would affect inference, results are also reported for a version of the test comparing the DM statistic against the t_{P-1} distribution. While the DM statistic is asymptotically standard normal when the forecasts are non-nested, Harvey, et. al. (1997) find that comparing the DM statistic against the t_{P-1} distribution yields better small-sample properties. In the reported results, the invalid version of the test (but not the asymptotically valid version) also incorporates an adjustment, developed in Harvey, et. al (1997), to correct for bias in the estimated variance of $\bar{d}$. This adjustment takes the form of multiplying the test statistic (2) by $\sqrt{(P-1)/P}$.

3.4 The Harvey, et. al. (HLN) Test

Harvey, et. al. (1998) use the basic methodology of Diebold and Mariano (1995) to develop a test of forecast encompassing drawn from the conditional efficiency framework of

---

reported, with the difference that the empirical size is slightly higher.

Granger and Newbold (1973) and Nelson (1972).[4] Conditional efficiency is tested with the OLS-based regression

$$\hat{u}_{1,t+1} = l(\hat{u}_{1,t+1} - \hat{u}_{2,t+1}) + h_{t+1}. \tag{3}$$

If λ = 0, forecast 2 carries no useful information not already in forecast 1, so forecast 1 is conditionally efficient. If λ > 0, forecasts from model 2 do carry information not already in forecasts from model 1.

Harvey, et. al. (1998) propose testing encompassing with a t-statistic for the covariance between $\hat{u}_{1,t+1}$ and $\hat{u}_{1,t+1} - \hat{u}_{2,t+1}$ rather than with a t-statistic for the regression coefficient $\hat{l}$. Let $c_{t+1} = \hat{u}_{1,t+1}(\hat{u}_{1,t+1} - \hat{u}_{2,t+1}) = \hat{u}_{1,t+1}^2 - \hat{u}_{1,t+1}\hat{u}_{2,t+1}$ and $\bar{c} = P^{-1}\sum_t c_t$. The Harvey, et. al. encompassing test is formed as

$$HLN = \frac{\bar{c}}{\sqrt{\hat{\mathrm{v}}\mathrm{ar}(\bar{c})}} = \frac{\bar{c}}{\sqrt{P^{-2}\sum_t (c_t - \bar{c})^2}} = P^{1/2} \cdot \frac{P^{-1}\sum_t \hat{u}_{1,t+1}^2 - P^{-1}\sum_t \hat{u}_{1,t+1}\hat{u}_{2,t+1}}{\sqrt{P^{-1}\sum_t \left\{(\hat{u}_{1,t+1}^2 - \hat{u}_{1,t+1}\hat{u}_{2,t+1}) - \bar{c}\right\}^2}}. \tag{4}$$

Under the null that model 1 forecast encompasses model 2, the covariance between $u_{1,t}$ and $u_{1,t} - u_{2,t}$ will be less than or equal to 0, while under the alternative that model 2 contains added information, the covariance should be positive. The test is one-sided when applied to either nested or non-nested forecasts.

While the HLN statistic is asymptotically standard normal when applied to non-nested forecasts, the asymptotic distribution is non-normal when the forecasts are nested under the null. The actual limiting distribution is provided in Theorem 3.4.

---

[4] Condition efficiency in turn draws from the forecast combination literature, started by Bates and Granger (1969). The regression (5) can be used to determine the optimal combining weight. The linear combination of forecasts 1

Theorem 3.4: For HLN defined in (4), HLN $\rightarrow_d \frac{\chi_1}{(\chi_2)^{0.5}}$ where χ_1 equals

$\int_\lambda^1 s^{-1} W'(s) dW(s)$ for the recursive scheme,

$\lambda^{-1}\{W(1) - W(\lambda)\}' W(\lambda)$ for the fixed scheme,

$\lambda^{-1}\int_\lambda^1 \{W(s) - W(s-\lambda)\}' dW(s)$ for the rolling scheme,

and χ_2 equals

$\int_\lambda^1 s^{-2} W'(s) W(s) ds$ for the recursive scheme,

$\pi\lambda^{-1} W'(\lambda) W(\lambda)$ for the fixed scheme,

$\lambda^{-2}\int_\lambda^1 \{W(s) - W(s-\lambda)\}'\{W(s) - W(s-\lambda)\} ds$ for the rolling scheme.

There are a couple things to notice about Theorem 3.4. The first is that for each forecasting scheme the statistic is pivotal. This fact is not particularly useful if asymptotic critical values, associated with the limiting distributions, are used to construct asymptotically valid tests. If the bootstrap is used, as in Ashley (1998), then we know from Hall (1992) that the bootstrap provides refinements to first order asymptotics and hence in finite samples may provide more accurate inference.

Though the null limiting distributions do not depend upon the data generating process itself, the distributions are dependent upon two parameters. The first is the number of excess parameters k_2. It arises since the vector Brownian Motion, W(s), is $(k_2 \times 1)$. The second parameter, π, also affects the null limiting distribution. It affects the limiting distribution in two

and 2 will have a smaller MSE than forecast 1 unless the covariance between $\hat{e}_{1,t+1}$ and $\hat{e}_{1,t+1} - \hat{e}_{2,t+1}$ and, equivalently, the coefficient λ, are 0.

ways. It directly affects the weights on each of the components of the statistics (recall that λ = $(1+\pi)^{-1}$). It also affects the range of integration on each of the stochastic integrals through λ.

Since the limiting distribution of the HLN statistic is nonstandard (i.e. neither normal nor chi-square) we provide asymptotically valid critical values in Tables A1-A3. These were generated numerically using the limiting distribution in Theorem 3.4 and hence can be considered estimates of the true asymptotic critical values. The reported critical values are the 90^{th}, 95^{th} and 99^{th} percentiles of 5000 independent draws from the distribution of $\frac{\chi_1}{(\chi_2)^{0.5}}$ for a given value of k_2 and π. Generating these draws proceeded as follows. Weights that depend upon π were estimated in the obvious way using $\hat{\pi} = P/R$. The necessary k_2 Brownian Motions were simulated as random walks each using an independent sequence of 10,000 i.i.d. $N(0,T^{-0.5})$ increments. The integrals were emulated by summing the relevant weighted quadratics of the random walks from the $R+1^{st}$ observation to the T^{th} observation. The random number generator was seeded so that all k_2 and π pairs and all sampling schemes use the same 5000 draws of k_2 sequences of 10,000 i.i.d. $N(0,T^{-0.5})$ increments.

A brief listing of critical values is provided in Tables A1, A2, and A3. Each of the tables corresponds to either the recursive, rolling, or fixed forecasting scheme. Within each table there are 330 critical values. Each of these correspond to one permutation of three parameters: k_2 = {1, 2, 3,…, 9, 10}, π = {0.1, 0.2, 0.4,…, 1.0, 1.2,…, 2.0} and nominal size of the test = {0.01, 0.05, 0.10}. Tables that allow for larger values of both k_2 and π are available upon request.

In the results of section 4, the HLN statistic is compared against the asymptotically valid critical values tabulated in Tables A1-A3. To evaluate how using invalid asymptotic critical values would affect inference, results are also reported for a version of the test comparing the

HLN statistic against the t_{P-1} distribution. For non-nested models, Harvey, et. al. (1998) find that comparing the HLN statistic against the t_{P-1} distribution, rather than the standard normal, yields better small-sample properties. In the reported results, the invalid version of the test (but not the asymptotically valid version) also incorporates an adjustment, developed in Harvey, et. al (1997), to correct for bias in the estimated variance of $\bar{d}$. This adjustment takes the form of multiplying the test statistic (4) by $\sqrt{(P-1)/P}$.

3.5 A New Encompassing (CM) Test

As discussed below, Monte Carlo simulations suggest that the denominators of tests like the DM statistic (2) and the HLN statistic (4) adversely affect the small-sample properties of the tests. The denominator of the HLN statistic, for example, is the sample variance of c_t (normalized by P), which is asymptotically equal to 0. In parallel to the OOS F test, this paper proposes a variant of the HLN statistic in which $\bar{c}$ is scaled by the variance of one of the forecast errors rather than an estimate of the variance of $\bar{c}$.

This statistic, which we will refer to as the CM statistic, takes the form

$$CM = P \cdot \frac{\bar{c}}{MSE_2} = P \cdot \frac{P^{-1}\sum_t \hat{u}_{1,t+1}^2 - P^{-1}\sum_t \hat{u}_{1,t+1}\hat{u}_{2,t+1}}{P^{-1}\sum_t \hat{u}_{2,t+1}^2}. \tag{5}$$

The numerator is the object of interest in the HLN test – the covariance between $u_{1,t}$ and $u_{1,t} - u_{2,t}$. The denominator, MSE_2, serves as a scale correction. As was the case for the HLN statistic, the limiting distribution is non-normal when the forecasts are nested under the null. The actual limiting distribution is provided in Theorem 3.5.

Theorem 3.5: For CM defined in (5) and χ_1 defined in Theorem 3.4, CM $\rightarrow_d \chi_1$.

Given Theorem 3.4, this result is not surprising. The sole difference between the HLN and CM statistics is the denominator. Hence we expect their limiting distributions to be somewhat related. As was the case for the HLN statistic, the limiting distribution is pivotal and relies upon the parameters k_2 and π.

In the results of section 4, the CM statistic is compared against asymptotically valid critical values tabulated in Tables A4-A6. As was done for Tables A1-A3, these were generated numerically using the limiting distribution in Theorem 3.5 and hence can be considered estimates of the true asymptotic critical values. The numerical methods used to construct 5000 independent draws from the distribution of χ_1 were identical to those used to construct Tables A1-A3. Moreover the random number generator was seeded so that the same χ_1 values were used in the construction of both Tables A1-A3 and A4-A6. Tables A4-A6 contain the same 330 permutations of k_2, π and nominal size that are used in Tables A1-A3. Tables that allow for larger values of both k_2 and π are available upon request.

3.6 The Ericsson (ERIC) Test

Ericsson's (1992) forecast-differential encompassing test takes the same form as the conditional efficiency regression presented above:[5]

$$\hat{u}_{1,t+1} = l(\hat{u}_{1,t+1} - \hat{u}_{2,t+1}) + h_{t+1}. \qquad (6)$$

[5] As presented in Ericsson (1992) and used by others, the test often is expressed as a regression of the error from model 1 on the difference in forecasts rather than forecast errors. But that regression is equivalent to (6), with the appropriate sign change.

The test statistic is simply the t-statistic for the OLS-based regression coefficient $\hat{\lambda}$, which can be expressed as

$$\mathrm{ERIC} = \frac{\mathrm{P}^{1/2}\mathrm{a}_{0,\mathrm{T}}}{[\mathrm{a}_{1,\mathrm{T}}\mathrm{a}_{2,\mathrm{T}} - \mathrm{a}_{0,\mathrm{T}}^{2}]^{0.5}} \tag{7}$$

where $a_{0,T} = P^{-1}\sum_t \hat{u}_{1,t+1}(\hat{u}_{1,t+1} - \hat{u}_{2,t+1})$, $a_{1,T} = P^{-1}\sum_t (\hat{u}_{1,t+1} - \hat{u}_{2,t+1})^2$ and $a_{2,T} = P^{-1}\sum_t \hat{u}_{1,t+1}^2$.

Under the null that model 1 forecast encompasses model 2, the covariance between $u_{1,t}$ and $u_{1,t} - u_{2,t}$ will be less than or equal to 0, while under the alternative that model 2 contains added information, the covariance should be positive. The test is one-sided when applied to either nested or non-nested forecasts.

Once again the ERIC statistic is asymptotically standard normal when applied to non-nested forecasts but the asymptotic distribution is non-normal when the forecasts are nested under the null. The actual limiting distribution is provided in Theorem 3.6.

Theorem 3.6: For ERIC defined in (7) and $\frac{\chi_1}{(\chi_2)^{0.5}}$ defined in Theorem 3.4, ERIC $\to_d \frac{\chi_1}{(\chi_2)^{0.5}}$.

In Theorem 3.6 we find that the ERIC and HLN statistics have the same limiting distribution under the null.[6] Hence we can use Tables A1-A3 to construct asymptotically valid tests of forecast encompassing when the ERIC statistic is used. It should be mentioned however, that this does not imply that the two statistics will have similar finite sample properties. It is for this reason that we include both the HLN and ERIC statistics in the Monte Carlo experiments of section 4.

In the results of section 4, the ERIC statistic is compared against asymptotically valid critical values tabulated in Tables A1-A3. To evaluate how using invalid asymptotic critical values would affect inference, results are also reported for a version of the test comparing the ERIC statistic against the standard normal distribution.

3.7 The Chong-Hendry (CH) Test

Under the null that the restrictions on model 2 are correct, model 1 forecast encompasses model 2. The Chong and Hendry (1986) test of encompassing is formed as the t-statistic on $\hat{\alpha}$ from the OLS-based regression

$$\hat{u}_{1,t+1} = a g_{2,t+1}(\hat{b}_{2,t}) + n_{t+1}, \tag{8}$$

where $g_{2,t+1}(\hat{\beta}_{2,t})$ denotes the model 2 forecast.[7] It follows from West and McCracken (1998) that the CH statistic is asymptotically standard normal even when the models considered are nested. Accordingly, the t-statistic on $\hat{\alpha}$ is compared against the standard normal distribution. As with the DM, HLN, and ERIC statistics, to allow for non-normal data, the estimated variance of $\hat{\alpha}$ is robust to heteroskedasticity. Since the nesting of the models has no clear implications for the sign of $\hat{\alpha}$, the null $\alpha = 0$ is tested against a two-sided alternative.

4. MONTE CARLO RESULTS

Results on the small-sample properties of the tests described in section 3 are generated using a bivariate VAR data generating process, with forecasts of the variable of interest from an

---

[6] There is a parallel to this in McCracken (1999). There it is shown that the DM statistic has the same limiting distribution as the regression-based test for equal MSE by Granger and Newbold (1977).

[7] For the basic models considered below a modified CH test, that takes the form of the covariance between the LHS and RHS of (8) divided by an estimate of the standard error of the covariance, has modestly better size properties but

estimated AR model (model 1) compared to forecasts from a VAR (model 2). The presented results are based on data generated from the normal distribution. The results are essentially unchanged when data are generated from a heavier-tailed distribution, suggested by Diebold and Mariano (1995), in which one forecast error follows a t_6 distribution and the other is a linear combination of t_6 variables.

4.1 Experiment Design

In the presented results (currently), data are generated using one artificial VAR(1) model and one empirical VAR(2) model. The artificial VAR(1) takes the form

$$\begin{aligned} y_t &= 0.3y_{t-1} + bx_{t-1} + u_{y,t} \\ x_t &= 0.5x_{t-1} + u_{x,t}, \end{aligned} \tag{9-10}$$

where y_t is the variable to be forecast and x_t is an auxiliary variable. The error terms are independent standard normal variables. To evaluate size in finite samples, the coefficient b is set at 0. To evaluate power, b is set at 0.1, 0.2, and 0.4. Simulations based on VAR(2) models taking a comparable form, and simulations based on the trivariate stationary VAR(1) and VAR(3) models of Swanson, et. al. (1996), produced results in line with those from the bivariate VAR(1).

Alternative models. I. Models that generate larger size biases for in-sample GC tests due to pure pre-test bias effects (models with just much richer dynamics?). II. Misspecified models: (a) A VAR(2) in which the true model is an AR(2) in y and x(t-1) is strongly correlated with y(t-2). (b) Forecasts from bivariate VAR and AR when the true model is trivariate. (c) Forecasts from VAR and AR when the true model is a VARMA.

---

modestly lower power. This form of the test statistic parallels the Harvey, et. al. (1998) modification of the Ericsson

The empirical VAR is fit from quarterly data on the change in core CPI inflation and the change in the prime-age male unemployment rate, where the change in core inflation is the variable to be forecast. While the integration orders of these data are admittedly debatable, the use of changes is based on the results of unit root tests and produces autoregressive roots well within the unit circle. Consistent with the empirical application considered in section 5, the models are fit exclusively with in-sample data that span 1958:Q3-1987:Q1.[8] Under the null that unemployment does not affect inflation, the estimated model used in a size experiment is

$$\begin{aligned}
&\Delta \text{Infl} = .024 - .288\Delta \text{Infl}_{t-1} - .237\Delta \text{Infl}_{t-2} + u_t \\
&\Delta \text{Unemp} = -.009 + .057\Delta \text{Infl}_{t-1} + .015\Delta \text{Infl}_{t-2} + .703\Delta \text{Unemp}_{t-1} - .182\Delta \text{Unemp}_{t-2} + v_t \\
&\text{var}(u_t) = 2.795, \text{var}(v_t) = .107, \text{cov}(u_t, v_t) = -.084.
\end{aligned} \tag{11-12}$$

With unemployment affecting inflation, the estimated model used in a power experiment is

$$\begin{aligned}
&\Delta \text{Infl} = .033 - .391\Delta \text{Infl}_{t-1} - .266\Delta \text{Infl}_{t-2} - 1.207\Delta \text{Unemp}_{t-1} - .137\Delta \text{Unemp}_{t-2} + u_t \\
&\Delta \text{Unemp} = -.009 + .057\Delta \text{Infl}_{t-1} + .015\Delta \text{Infl}_{t-2} + .703\Delta \text{Unemp}_{t-1} - .182\Delta \text{Unemp}_{t-2} + v_t \\
&\text{var}(u_t) = 2.519, \text{var}(v_t) = .107, \text{cov}(u_t, v_t) = -.084.
\end{aligned} \tag{13-14}$$

The lag lengths of both models were selected to minimize the Akaike criterion, allowing a maximum of four lags.

Letting R denote the number of in-sample observations and P represent the number of predictions, Monte Carlo methods are used to generate a total of $R + P + 4$ observations.[9] The additional four observations generated allow for data-determined lag lengths in the forecasting models estimated in each simulation. Letting *L* denote the lag length of the data-generating

---

regression test.

[8] Using Kilian's (1998) bootstrap method to adjust the coefficients of the models produces essentially the same Monte Carlo results.

process (DGP), the first L observations are generated by drawing from the unconditional normal distribution implied by the model parameterization. The remaining $R + P + 4 - L$ observations are constructed using the autoregressive model structure and draws of the error terms from the normal distribution.

With observations 1 through 4 reserved as initial observations necessary to allow for as many as four lags in the estimated models, a VAR in y and x (or ΔInfl and ΔUnemp) and an AR model for y (or ΔInfl) are fit over the in-sample period spanning observations 5 through $R + 4$. In the interest of brevity the lag length of each estimated model was set at the "true" order, L, of the data generating process.[10] Results for forecasts based on lags set at the order minimizing the in-sample Akaike criterion for the VAR or the AR model are essentially the same.[11] In one unsurprising exception, the in-sample GC test tends to have slightly greater size and lower power than the presented results when the lag length is set to minimize the in-sample Akaike criterion for the VAR. Observations $R + 5$ through $R + P + 4$ are then used to form P 1-step ahead forecasts. The first forecast is for period $R + 4$; the last is for period $R + P + 4$. For brevity,

---

[9] For the empirical model, using bootstrap methods to generate artificial data produces results much like those reported for Monte Carlo-generated data.

[10] In the case of the simple VAR(1) model, when b is non-zero, the true VAR implies that the true univariate model for y_t is an ARMA(2,1). For the models considered here, however, inverting the MA component to rewrite the ARMA(2,1) in AR form yields very small coefficients on lags greater than 1, so the true (infinite order) AR model is very close to an AR(1). Accordingly, in the Monte Carlos the selected AR lag length is 1 in about 75 percent of the power simulations.

[11] In computing power when the lags are data-determined, the test statistic in simulation i, for which the selected lag is j, is compared against the distribution of test statistics from the set of simulations under the null in which the lag was selected to be j. For example, if lag j was selected in J of the 50,000 size simulations of a given experiment, empirical critical values for lag j were calculated from just those J simulated test statistics. In a corresponding power experiment, for those simulations in which the lag was selected to be j, the test statistics were compared against these critical values. Since longer lags tend to be somewhat infrequently rejected, 50,000 simulations were used in the size experiments to ensure the accuracy of the results with data-determined lags.

results are only presented for recursive forecasts, as the basic conclusions are essentially the same for rolling and fixed forecasts.[12]

Results are reported for simple, empirically relevant combinations of P and R such that $\hat{\pi} \equiv P/R$ takes a value of 0.1, 0.2, 0.4, 0.6, 1.0, or 2.0. Specifically, in one set of experiments based on the artificial VAR(1), R = 100 and P = 10, 20, 40, 60, 100, and 200. In another set using the VAR(1), R = 200 and P = 20, 40, 80, 120, 200, and 400. In experiments based on the inflation-unemployment model, as suggested above P and R are chosen to be consistent with the empirical application considered in section 5. Particularly, R = 115 and P = 46, for which $\hat{\pi} = 0.4$.

4.2 Size Results

Table 1 presents the empirical sizes of Granger causality, equal accuracy, and encompassing tests for data from the VAR(1) of equations (9)-(10), using a nominal size of 10%. The general results are the same at the nominal size of 5%. In these size experiments, data are generated imposing the null of $b = 0$, under which the AR and VAR forecasts of y have equal MSE (asymptotically), and the AR forecast encompasses the VAR forecast.

Four general results are evident from Table 1. First, OOS F and CM tests perform quite well, suffering only slight size distortions. For example, when $R = 100$ and $P = 20$, the empirical sizes of the OOS F and CM tests are 10.7% and 11.0% respectively. Given R, any distortions in the OOS F and CM tests generally decline as P rises. Second, the asymptotically valid versions of the DM and ERIC statistics and the CH test are modestly oversized in finite

[12] While results for rolling forecasts are very similar to those for recursive forecasts, results for fixed forecasts do differ slightly. For example, with fixed forecasts, the size distortions of the asymptotically valid DM test are smaller, while the size distortions of the CM test are a bit bigger.

samples. For example, when $R = 100$ and $P = 20$, the empirical sizes of the asymptotically valid DM, ERIC, and CH tests are 13.5%, 14.2%, and 15.5%, respectively. The distortions decline as P rises for a given R.[13] With $R = 100$ and $P = 200$, for instance, the ERIC test is essentially correctly sized, rejecting the null in 10.4% of the simulations.

The third general result is that comparing the DM, HLN, and ERIC tests against invalid asymptotic critical values will generally lead to too-infrequent rejections, the DM test more so than the HLN test and the HLN test somewhat more so than the ERIC test. In the cases of the DM and HLN statistics, using the incorrect asymptotic distribution causes the tests to be undersized for all sample sizes, and the tests become more undersized as P rises given R. For instance, with $R = 100$ and $P = 20$, the DM and HLN tests reject the null in, respectively, 5.5% and 8.3% of the simulations. With P increased to 40, the DM and HLN sizes fall to 3.8% and 7.4%, respectively. In the case of the ERIC statistic, the invalid version of the test is oversized for very small samples but undersized for larger samples. With $R = 100$, for instance, the size of the test falls from 13.5% when $P = 10$ to 6.6% when $P = 100$.

Finally, simple GC tests are sometimes slightly oversized when the number of observations is small, but about correctly sized otherwise. The out-of-sample GC test used with $P = 20$, for example, has an empirical size of 11.6%, and the in-sample GC test with $R = 100$ has size of essentially 10%. While not shown in the interest of brevity, when the model lags are chosen to minimize the Akaike criterion for the in-sample VAR, the in-sample GC test suffers modest size distortions, while the size of the out-of-sample GC test is about the same as when the

[13] Increasing P, however, does not necessarily make the empirical distribution of the test statistic quickly approach a normal distribution. CH test rejections occur more frequently in the left tail than in the right tail. When a nominal size α% is used, right tail rejections occur in less than $\alpha/2$% of the simulations. As P rises, right tail rejections occur even less frequently and account for most of the observed improvement in the empirical size of the CH test.

lag is fixed at the DGP order.[14] With $R = 100$ and $P = 20$, setting the lags of the AR and VAR models at the optimal VAR lag yields an in-sample GC test size of 13.7 percent, up from 10.4 percent in the fixed-lag results. The out-of-sample GC test size is 11.8% and 11.6% with the optimal VAR lag and fixed lag, respectively.

Table 2 presents size results based on the restricted VAR(2) model for the changes in core inflation and unemployment, equations (11) and (12). In this size experiment, data are generated imposing the null that unemployment does not affect inflation. This implies that AR and VAR forecasts of ΔInfl have equal MSE (asymptotically), and that the AR forecast encompasses the VAR forecast. Using this empirical model produces results the same as those for the artificial VAR(1). Again, the performance of the OOS F and CM tests is quite good, with both each essentially correctly sized. The asymptotically valid versions of the DM and ERIC statistics and the CH test are subject to modestly larger distortions. Comparing the DM, HLN, and ERIC tests against invalid asymptotic critical values continues to produce too-infrequent rejections, with the DM test more undersized than the HLN or ERIC tests. Finally, for the R and P combination used in this experiment, standard GC tests are about correctly sized when the lag length is set at the true order. In results not reported, however, the in-sample GC test is still subject to modest distortions when the lag length is determined with data-based criteria.

Interpretation/explanation of results.

4.3 Power Results

Tables 3-5 present results on the power of simple Granger causality, equal accuracy, and encompassing tests for data from the simple VAR(1), equations (9)-(10). In these power

[14] However, if the lags of the AR and VAR models are set at the order minimizing the Akaike criterion for the in-sample AR model, the in-sample GC test is essentially correctly sized, just as when the lag is set at the true order.

experiments, data are generated using b = 0.1, 0.2, and 0.4, so VAR forecasts of y have lower MSE than AR forecasts, and the AR forecast does not encompass the VAR forecast. Because the tests are, to varying degrees, subject to size distortions, the reported power figures are based on empirical critical values and therefore size-adjusted.[15] With empirical rather than asymptotic critical values used, there is no distinction between the valid and invalid versions of the DM, HLN, and ERIC tests. The size of the tests is 10%; using 5% produces essentially the same results. For the OOS F, DM, CM, HLN, and ERIC tests, which are one-sided, the null is rejected if the test statistic is greater than the 90% fractile of the statistic in the corresponding size experiment (for the same R and P) with b = 0. The same applies to the GC tests. For the CH test, which is two-sided, the null is rejected if the test statistic lies outside the 5% and 95% fractiles of the empirical distribution generated in the corresponding size experiment.

Several general results are evident in Tables 3-5. First, the powers of the tests permit some simple rankings: (1) CM > OOS F > DM > CH; (2) OOS F > out-of-sample GC; and (3) CM > ERIC $\geq$ HLN > DM. In the VAR experiment design, the CM test for encompassing is clearly the most powerful out-of-sample test of predictive power in finite samples. Both the CM encompassing and OOS F accuracy tests dominate GC tests applied to just the out-of-sample data. For example, as shown in Table 3, with b = 0.1, $R = 100$, and $P = 40$, the CM test rejects the null in 32.4% of the simulations, compared to 28.0% for the OOS F test and 18.1% for the out-of-sample GC test. Moreover, when the explanatory power of the Granger-causal variable is weak, as when b = 0.1, the power of the CM test sometimes exceeds that of the in-sample GC test. This holds even though the CM uses fewer observations than does the in-sample GC test. In the preceding example, the in-sample GC test rejects the null in 30.9% of the simulations.

[15] For those tests subject to small size distortions, particularly the OOS F and CM tests, power based on asymptotic critical values is very similar to the reported size-adjusted power.

The power of the DM and CH tests is lower, often substantially, than the other equal accuracy and encompassing tests. The HLN and ERIC encompassing tests have power somewhere in between the power of the CM and DM tests.

The second general result is that, given R, power rises with P. For example, as shown in Table 4, with b = 0.2 and $R = 100$, the power of the OOS F test increases from 48.4% when $P = 20$ to 58.1% when $P = 40$. Third and finally, power increases with the coefficient b, which determines the explanatory power of the causal variable. Power is systematically higher when $b = 0.4$ than when $b = 0.2$ and, in turn, than when $b = 0.1$.

Table 6 presents size-adjusted power based on the VAR(2) for the changes in core inflation and unemployment, equations (13)-(14). In the DGP underlying this power experiment, unemployment does affect inflation, so VAR forecasts of $\Delta \mathrm{Infl}$ have lower MSE than AR forecasts (asymptotically), and the AR forecast does not encompass the VAR forecast. The critical values used to evaluate power are calculated from the distribution of statistics generated in the Table 2 size experiment. Using the empirical model produces results the same as those for the artificial VAR(1). The rankings (1) CM > OOS F > DM > CH, (2) OOS F > out-of-sample GC, and (3) CM > ERIC ≥ HLN > DM still apply. In this example, the power of the CM test is not only greater than that of an out-of-sample GC test but also essentially the same as an in-sample GC test, despite the fact that $P < R$. These results suggest that, if the DGP were the true model, statistics like the CM test would be almost sure to correctly pick up the predictive power of unemployment.

Interpretations and explanations.

5. EMPIRICAL EXAMPLE

To provide a sense of the ability of the different tests to determine predictive power in practice, this section presents results on a question that many studies have examined: whether unemployment has predictive power, in-sample and post-sample, for inflation. Recent examples of studies addressing this question include Cecchetti (1995) and Staiger, Stock, and Watson (1997). In this analysis, tests on Granger causality, equal forecast accuracy, and forecast encompassing are applied to data on core CPI inflation and the prime-age male unemployment rate.

In the specification used here, both inflation and unemployment are differenced, consistent with the results of augmented Dickey-Fuller tests for unit roots. The quarterly data available from 1957:Q1-1998:Q4 are divided into an in-sample portion and an out-of-sample portion of modest length, so as to produce a $\hat{\pi} = P/R$ value for which McCracken (1999) reports asymptotic critical values. Allowing for data differencing and a maximum of four lags in determining the model's lag order, the in-sample period spans 1958:Q3-1987:Q1, a total of R = 115 observations. The out-of-sample period spans 1987:Q2-1998:Q3, yielding a total of P = 46 1-step ahead predictions. For this split, $\hat{\pi} = 0.4$. Over the in-sample period, the Akaike criterion for both the AR and the VAR is minimized at two lags.

Table 7 presents in-sample estimates of an AR(2) fit to changes in core CPI inflation and a VAR(2) fit to changes in core CPI inflation and prime-age male unemployment. The $\bar{R}^2$'s and GC tests reported in the table suggest that, over the in-sample period, unemployment has predictive power for inflation. The out-of-sample evidence, however, is weaker.[16] The CM encompassing test, shown above to be more powerful than all the other post-sample tests, indicates that unemployment has predictive power for inflation. The ERIC test, which has lower power than the CM test but higher power than the DM test, also rejects the null when compared against the empirical critical value. Surprisingly, the CH test, generally the weakest, also suggests unemployment has predictive power. None of the other tests reject the null of no predictive content in unemployment.

These results might lead to the conclusion that unemployment in fact has no predictive power for inflation. In this view the in-sample results, and out-of-sample test results suggesting otherwise, are spurious. The problem with this interpretation is that the Monte Carlo experiments of section 4 suggest that any size biases in in-sample Granger causality tests and, in particular, the out-of-sample CM test, should be small, even with data-determined lags. Accordingly, the in-sample GC and CM statistics strongly reject the null when compared against both asymptotic and empirical critical values.

A more reasonable interpretation of the results is that unemployment does have predictive power for inflation, power perhaps weakened by model instabilities. Both an in-sample GC test and the CM test for forecast encompassing overwhelmingly indicate that unemployment has predictive content. The Monte Carlo results in section 4 indicate these tests have the greatest power in finite samples. Nonetheless, the failure of other out-of-sample tests to detect any predictive power, despite the strong in-sample predictive power of unemployment, suggests there is a difference between the in-sample and out-of-sample predictive content of unemployment. This difference may reflect model instabilities. Neither the AR model nor the VAR pass the

[16] The forecasts are slightly biased. However, demeaning the errors prior to calculating the test statistics has little effect on the results.

supremum Wald or exponential Wald tests for stability developed in Andrews (1993) and Andrews and Ploberger (1994), respectively.[17]

6. CONCLUSIONS

[17] The models do, however, pass the Nyblom (1989) test for stability and Chow tests for a shift in the parameter estimates between 1958:Q3-87:Q1 and 1987:Q2-97:Q3. Following Diebold and Chen (1996), the stability test results are based on bootstrap critical values.

References

Amano, Robert A., and Simon van Norden, "Terms of Trade and Real Exchange Rates: The Canadian Evidence," *Journal of International Money and Finance*, v.14, no.1 (February 1995), pp. 83-104.

Andrews, Donald W.K., "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, v.61, no.4 (July 1993), pp. 821-56.

Andrews, Donald W.K., and Werner Ploberger, "Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative," *Econometrica*, v.62, no.6 (November 1994), pp. 1383-1414.

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

Ashley, R., C.W.J. Granger, and R. Schmalensee, "Advertising and Aggregate Consumption: An Analysis of Causality," *Econometrica*, v.48, no.5 (July 1980), pp. 1149-67.

Bates, J.M. and C.W.J. Granger, "The Combination of Forecasts," *Operational Research Quarterly*, v.20 (1969), pp. 451-68.

Blomberg, S. Brock, and Gregory D. Hess, "Politics and Exchange Rate Forecasts," *Journal of International Economics*, v.43, no.1-2 (August 1997), pp. 189-205.

Bram, Jason, and Sydney Ludvigson, "Does Consumer Confidence Forecast Household Expenditure? A Sentiment Horse Race," *Economic Policy Review*, Federal Reserve Bank of New York, June 1998, pp. 59-78.

Cecchetti, Stephen G., "Inflation Indicators and Inflation Policy," *NBER Macroeconomics Annual*, 1995, pp. 189-219.

Chong, Yock Y., and David F. Hendry, "Econometric Evaluation of Linear Macroeconomic Models," *Review of Economic Studies*, v.53, no.4 (August 1986), pp. 671-90.

Clark, Todd E., "Finite-Sample Properties of Tests for Equal Forecast Accuracy," forthcoming, *Journal of Forecasting* (1999).

Corradi, Valentina, Norman R. Swanson, and Claudia Olivetti, "Predictive Ability With Cointegrated Variables," manuscript, Pennsylvania State University, December 1998.

Diebold, Francis X., and Celia Chen, "Testing Structural Stability with Endogenous Breakpoint: A Size Comparison of Analytic and Bootstrap Procedures," *Journal of Econometrics*, v.70, no.1 (January 1996), pp. 221-41.

Diebold, Francis X., and Roberto S. Mariano, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, v.13, no.3 (July 1995), pp. 253-63.

Ericsson, Neil R., "On the Limitations of Comparing Mean Square Forecast Errors: Clarifications and Extensions," *Journal of Forecasting*, v.12, no.8 (December 1993), pp. 644-51.

Ericsson, Neil R., "Parameter Constancy, Mean Square Forecast Errors, and Measuring Forecast Performance: An Exposition, Extensions, and Illustration," *Journal of Policy Modeling*, v.14, no.4 (August 1992), pp. 465-95.

Granger, C.W.J., and Paul Newbold, "Some Comments on the Evaluation of Economic Forecasts," *Applied Economics*, v.5, no.1 (March 1973), pp. 35-47.

Granger, C.W.J., and Paul Newbold, *Forecasting Economic Time Series*, Orlando, FL: Academic Press, 1977.

Hall, Peter, *The Bootstrap and Edgeworth Expansion*, New York, Springer-Verlag, 1992.

Hansen, Bruce E., "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes," *Econometric Theory*, v.8, no.4 (December 1992), pp. 489-500.

Harvey, David I., Stephen J. Leybourne, and Paul Newbold, "Testing the Equality of Prediction Mean Squared Errors," *International Journal of Forecasting*, v.13, no.2 (June 1997), pp. 281-91.

Harvey, David I., Stephen J. Leybourne, and Paul Newbold, "Tests for Forecast Encompassing," *Journal of Business and Economic Statistics*, v.16, no.2 (April 1998), pp. 254-59.

Hoffman, Dennis, and Adrian Pagan, "Practitioners Corner: Post-Sample Prediction Tests for Generalized Method of Moments Estimators," *Oxford Bulletin of Economics and Statistics*, v. 51 (1989), pp. 333-343.

Kilian, Lutz, "Small-Sample Confidence Intervals for Impulse Response Functions," *Review of Economics and Statistics*, v.80, no.2 (May 1998), pp. 218-30.

Krueger, Joel T., and Kenneth N. Kuttner, "The Fed Funds Futures Rate as a Predictor of Federal Reserve Policy," *Journal of Futures Markets*, v.16, no.8 (December 1996), pp. 865-79.

Lutkepohl, Helmut and Maike M. Burda, "Modified Wald tests undre nonregular conditions," *Journal of Econometrics*, v. 78 (1997), pp. 315-332.

Mark, Nelson C., "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability," *American Economic Review*, v. 85, no. 1 (March 1995), pp. 201-18.

McCracken, Michael W., "Robust Out of Sample Inference," manuscript, Louisiana State

University, 1998.

McCracken, Michael W., “Asymptotics for Out-of-Sample Tests of Causality,” manuscript, Louisiana State University, January 1999.

Mizrach, Bruce, “The Distribution of the Theil U-Statistic in Bivariate Normal Populations,” *Economic Letters*, v. 38 (1992), pp. 163-167.

Nelson, Charles R., “The Prediction Performance of the FRB-MIT-Penn Model of the U.S. Economy,” *American Economic Review*, v.62, no.5 (December 1972), pp. 902-17.

Newey, Whitney K. and Daniel McFadden, “Large Sample Estimation and Hypothesis Testing,” in R.F. Engle and D.L. McFadden, ed., *Handbook of Econometrics, Volume IV*, Amsterdam, North-Holland, 1994.

Newey, Whitney K., and Kenneth D. West, “Automatic Lag Selection in Covariance Matrix Estimation,” *Review of Economic Studies*, v.61, no.4 (October 1994), pp. 631-53.

Nyblom, Jukka, “Testing for the Constancy of Parameters Over Time,” *Journal of the American Statistical Society*, v.84, no.405 (March 1989), pp. 223-30.

Staiger, Douglas, James H. Stock, and Mark W. Watson, “The NAIRU, Unemployment and Monetary Policy,” Journal of Economic Perspectives, v.11, no.1 (Winter 1997), pp. 33-49.

Swanson, Norman R., Ataman Ozyildirim, and Maria Pisu, “A Comparison of Alternative Causality and Predictive Accuracy Tests in the Presence of Integrated and Cointegrated Economic Variables,” manuscript, Pennsylvania State University, March 1996.

West, Kenneth D., “Asymptotic Inference About Predictive Ability,” *Econometrica*, v.64, no.5 (September 1996), pp. 1067-84.

West, Kenneth D., and Michael W. McCracken, “Regression-Based Tests of Predictive Ability,” *International Economic Review*, v.39, no.4 (November 1998), pp. 817-40.

White, Halbert, “A Reality Check for Data Snooping,” manuscript, University of California-San Diego and QuantMetrics R&D Associates, LLC, 1998.

Appendix

Lemmas A1 - A11 also appear in McCracken (1999) with a slightly different numerical ordering. In order to facilitate reference, but also conserve space, those Lemmas are repeated below without proof. Lemmas A12 - A14 are new and hence their proofs are provided.

Throughout the remainder the following notation will be used: J denotes the selection matrix $(I_{k_1\times k_1}, 0_{k_1\times k_2})$ $(k_1\times k,\ k > k_1)$, $\Rightarrow$ denotes weak convergence, Σ_t denotes the summation $\Sigma_{t=R}^{T}$, for any function $f_{i,t}(\beta)$ i =1,2 we will usually denote $f_{2,t}(\beta_2^*)$ as f_t, for matrices A and C defined in Lemma A2 $\tilde{h}_{t+1}$ denotes $\sigma^{-1}A'CB^{0.5}h_{t+1}$, $H_i(t)$ equals $t^{-1}\sum_{s=1}^{t}h_{i,s}$, $R^{-1}\sum_{s=t-R+1}^{t}h_{i,s}$ and $R^{-1}\sum_{s=1}^{R}h_{i,s}$ for the recursive, rolling and fixed schemes respectively, $\tilde{H}(t)$ denotes $\sigma^{-1}A'CB^{0.5}H(t)$, $\dot{\beta}_{i,t}$ is some vector on the line between $\hat{\beta}_{i,t}$ and β_i^*, $\nabla\dot{g}_{i,\beta,t+1}$ denotes $g'_{i,\beta,t+1}(\dot{\beta}_{i,t})(\hat{\beta}_{i,t}-\beta_i^*)$.

Lemma A1: For all $a\in[0,0.5)$ and each i = 1,2, (a) $\sup_t|\hat{\beta}_{i,t}-\beta_i^*| = o_p(1)$, (b) $\sup_t|B_i(t)-B_i| = o_p(1)$, (c) $\sup_t t^a|U(t)| = o_p(1)$, (d) $\sup_t t^a|\hat{\beta}_{i,t}-\beta_i^*| = o_p(1)$, (e) $\sup_t|T^{0.5}(\mathrm{vec}[B_i(t)]-\mathrm{vec}[B_i])| = O_p(1)$.

Lemma A2: (a) Let $-J'B_1J+B_2 = M$ and $B_2^{-1/2}MB_2^{-1/2} = Q$, then Q is idempotent. (b) Let A be a $(k\times k_2)$ matrix with $I_{k_2\times k_2}$ on the upper $(k_2\times k_2)$ block and zeroes elsewhere. There exists a symmetric orthonormal matrix C such that $Q = CAA'C$.

Lemma A3: For $s \in [\lambda,1]$, (a) $T^{-1/2}\sum_{j=1}^{t}\tilde{h}_j \Rightarrow W(s)$, (b) $(\frac{T}{t})T^{-1/2}\sum_{j=1}^{t}\tilde{h}_j \Rightarrow s^{-1}W(s)$, (c) $(\frac{T}{R})T^{-1/2}\sum_{j=t-R+1}^{t}\tilde{h}_j \Rightarrow \lambda^{-1}\{W(s)-W(s-\lambda)\}$.

Lemma A4: $\sum_t \tilde{H}'(t)\tilde{h}_{t+1} \rightarrow_d \chi_1$ where χ_1 equals

$\int_\lambda^1 s^{-1}W'(s)dW(s)$	for the recursive scheme,
$\lambda^{-1}\{W(1)-W(\lambda)\}'W(\lambda)$	for the fixed scheme,
$\lambda^{-1}\int_\lambda^1\{W(s)-W(s-\lambda)\}'dW(s)$	for the rolling scheme.

Lemma A5: $\sum_t \tilde{H}'(t)\tilde{H}(t) \rightarrow_d \chi_2$ where χ_2 equals

$\int_\lambda^1 s^{-2}W'(s)W(s)ds$	for the recursive scheme,
$\pi\lambda^{-1}W'(\lambda)W(\lambda)$	for the fixed scheme,
$\lambda^{-2}\int_\lambda^1\{W(s)-W(s-?)\}'\{W(s)-W(s-?)\}ds$	for the rolling scheme.

Lemma A6: $\sum_t\{-h_{t+1}'J'B_1(t)JH(t)+h_{t+1}'B(t)H(t)\}^2 = \sum_t\{-h_{t+1}'J'B_1JH(t)+h_{t+1}'BH(t)\}^2 +$ $o_p(1)$.

Lemma A7: $\sum_t\{-h_{t+1}'J'B_1JH(t)+h_{t+1}'BH(t)\}^2 = c^2\sum_t\{\tilde{H}(t)\tilde{h}_{t+1}\}^2$.

Lemma A8: $\sum_t\{\tilde{H}(t)\tilde{h}_{t+1}\}^2 \rightarrow_d \chi_2$ for χ_2 defined in Lemma A5.

Lemma A9: $\sum_t u_{t+1} g'_{i,\beta,t+1}(\dot{\beta}_{i,t}) B_i(t) H_i(t) = \sum_t h'_{i,t+1} B_i H_i(t) + o_p(1)$.

Lemma A10: For i,j = 1,2, $\sum_t H_i'(t) B_i(t) g_{i,\beta,t+1}(\dot{\beta}_{i,t}) g'_{j,\beta,t+1}(\dot{\beta}_{j,t}) B_j(t) H_j(t) =$

$\sum_t H_i'(t) B_i E(g_{i,\beta,t+1} g'_{j,\beta,t+1}) B_j H_j(t) + o_p(1)$.

Lemma A11: $\sum_t (u_{1,t+1}(\hat{\beta}_{1,t}) - u_{2,t+1}(\hat{\beta}_{2,t}))^2 \to_d \sigma^2 \chi_2$.

Lemma A12: For all $a \in [0,0.5)$, $\sup_t |T^a \nabla \dot{g}_{i,t+1}| = o_p(1)$.

Proof of Lemma A12: If we take a first order Taylor expansion of $g'_{i,\beta,t+1}(\dot{\beta}_{i,t})$ about β_i^* we immediately know that for some $\ddot{\beta}_{i,t}$ on the line between $\dot{\beta}_{i,t}$ and β_i^*,

$$\sup_t |T^a \nabla \dot{g}_{i,t+1}| \leq$$

$$\sup_t |T^a (\dot{\beta}_{i,t} - \beta_i^*)' q_{i,t+1}(\ddot{\beta}_{i,t})(\hat{\beta}_{i,t} - \beta_i^*)| + \sup_t |T^a g'_{i,\beta,t+1}(\hat{\beta}_{i,t} - \beta_i^*)|$$

$$\leq \sup_t |T^a (\dot{\beta}_{i,t} - \beta_i^*)' q_{i,t+1}(\ddot{\beta}_{i,t})(\hat{\beta}_{i,t} - \beta_i^*)| +$$

$$k(\sup_t |g_{i,\beta,t+1}|)(\sup_t |T^a (\hat{\beta}_{i,t} - \beta_i^*)|).$$

Adding and subtracting $q_{i,t+1}$, the r.h.s. of the final inequality is then less than or equal to

$$k^2(\sup_t |\dot{\beta}_{i,t} - \beta_i^*|)(\sup_t |q_{i,t+1}(\ddot{\beta}_{i,t}) - q_{i,t+1}|)(\sup_t |T^a (\hat{\beta}_{i,t} - \beta_i^*)|) +$$

$$k^2(\sup_t |\dot{\beta}_{i,t} - \beta_i^*|)(\sup_t |q_{i,t+1}|)(\sup_t |T^a (\hat{\beta}_{i,t} - \beta_i^*)|) +$$

$$k(\sup_t |g_{i,\beta,t+1}|)(\sup_t |T^a (\hat{\beta}_{i,t} - \beta_i^*)|).$$

That $\sup_t |q_{i,t+1}|$ and $\sup_t |g_{i,\beta,t+1}|$ are $O_p(1)$ follows from Assumption 3. That $\sup_t |T^a(\hat{\beta}_{i,t} - \beta_i^*)|$ is $o_p(1)$ follows from Lemma A1. Since $\sup_t |T^a(\dot{\beta}_{i,t} - \beta_i^*)| \leq \sup_t |T^a(\hat{\beta}_{i,t} - \beta_i^*)|$ it follows from Lemma A1 that $\sup_t |T^a(\dot{\beta}_{i,t} - \beta_i^*)|$ is $o_p(1)$. The result then follows since by Assumption 2 and Lemma A1, $\sup_t |q_{i,t+1}(\ddot{\beta}_{i,t}) - q_{i,t+1}| \leq$ $(\sup_t |m_t|)(\sup_t |T^a(\dot{\beta}_{i,t} - \beta_i^*)|)^{\varphi} \leq (\sup_t |m_t|)(\sup_t |T^a(\hat{\beta}_{i,t} - \beta_i^*)|)^{\varphi} = O_p(1)o_p(1)$.

Lemma A13: For i, j = 1,2, $\sum_t u_{t+1}^2 \nabla\dot{g}_{i,t+1} \nabla\dot{g}_{j,t+1} = \sum_t H_i(t)' B_i h_{i,t+1} h'_{j,t+1} B_j H_j(t) + o_p(1)$.

Proof of Lemma A13: First rewrite $\sum_t u_{t+1}^2 \nabla\dot{g}_{i,t+1} \nabla\dot{g}_{j,t+1}$ as

$$\sum_t H_i(t)'[(B_i(t) - B_i) + B_i][(u_{t+1} g_{i,\beta,t+1}(\dot{\beta}_{i,t}) - h_{i,t+1}) + h_{i,t+1}] \times$$

$$[(u_{t+1} g_{j,\beta,t+1}(\dot{\beta}_{j,t}) - h_{j,t+1}) + h_{j,t+1}]'[(B_j(t) - B_j) + B_j]H_j(t).$$

Expanding the above equation we then know that $\sum_t u_{t+1}^2 \nabla\dot{g}_{i,t+1} \nabla\dot{g}_{j,t+1}$ equals

$$\sum_t H_i'(t) B_i h_{i,t+1} h'_{j,t+1} B_j H_j(t) +$$

$$\sum_t H_i'(t)(B_i(t) - B_i) h_{i,t+1} h'_{j,t+1} B_j H_j(t) +$$

$$\sum_t H_i'(t) B_i (u_{t+1} g_{i,\beta,t+1}(\dot{\beta}_{i,t}) - h_{i,t+1}) h'_{j,t+1} B_j H_j(t) +$$

$$\sum_t H_i'(t) B_i h_{i,t+1} h'_{j,t+1} (B_j(t) - B_j) H_j(t) +$$

$$\sum_t H_i'(t) B_i h_{i,t+1} (u_{t+1} g_{j,\beta,t+1}(\dot{\beta}_{j,t}) - h_{j,t+1})' B_j H_j(t) +$$

$$\sum_t H_i'(t)(B_i(t) - B_i)(u_{t+1} g_{i,\beta,t+1}(\dot{\beta}_{i,t}) - h_{i,t+1}) h'_{j,t+1} B_j H_j(t) +$$

$$\sum_t H_i'(t)(B_i(t) - B_i) h_{i,t+1} (u_{t+1} g_{j,\beta,t+1}(\dot{\beta}_{j,t}) - h_{j,t+1})' B_j H_j(t) +$$

$$\sum_t H_i'(t)(B_i(t) - B_i) h_{i,t+1} h'_{j,t+1} (B_j(t) - B_j) H_j(t) +$$

$$\sum_t H_i'(t)B_i(u_{t+1}g_{i,\beta,t+1}(\dot{\beta}_{i,t})-h_{i,t+1})(u_{t+1}g_{j,\beta,t+1}(\dot{\beta}_{j,t})-h_{j,t+1})'B_jH_j(t) \ +$$

$$\sum_t H_i'(t)B_i(u_{t+1}g_{i,\beta,t+1}(\dot{\beta}_{i,t})-h_{i,t+1})h_{j,t+1}'(B_j(t)-B_j)H_j(t) \ +$$

$$\sum_t H_i'(t)B_ih_{i,t+1}(u_{t+1}g_{j,\beta,t+1}(\dot{\beta}_{j,t})-h_{j,t+1})'(B_j(t)-B_j)H_j(t) \ +$$

$$\sum_t H_i'(t)(B_i(t)-B_i)(u_{t+1}g_{i,\beta,t+1}(\dot{\beta}_{i,t})-h_{i,t+1})(u_{t+1}g_{j,\beta,t+1}(\dot{\beta}_{j,t})-h_{j,t+1})'B_jH_j(t) \ +$$

$$\sum_t H_i'(t)(B_i(t)-B_i)(u_{t+1}g_{i,\beta,t+1}(\dot{\beta}_{i,t})-h_{i,t+1})h_{j,t+1}'(B_j(t)-B_j)H_j(t) \ +$$

$$\sum_t H_i'(t)(B_i(t)-B_i)h_{i,t+1}(u_{t+1}g_{j,\beta,t+1}(\dot{\beta}_{j,t})-h_{j,t+1})'(B_j(t)-B_j)H_j(t) \ +$$

$$\sum_t H_i'(t)B_i(u_{t+1}g_{i,\beta,t+1}(\dot{\beta}_{i,t})-h_{i,t+1})(u_{t+1}g_{j,\beta,t+1}(\dot{\beta}_{j,t})-h_{j,t+1})'(B_j(t)-B_j)H_j(t) \ +$$

$$\sum_t H_i'(t)(B_i(t)-B_i)(u_{t+1}g_{i,\beta,t+1}(\dot{\beta}_{i,t})-h_{i,t+1})\times$$

$$(u_{t+1}g_{j,\beta,t+1}(\dot{\beta}_{j,t})-h_{j,t+1})'(B_j(t)-B_j)H_j(t).$$

The result will follow if the latter fifteen terms are each $o_p(1)$. The proof of each is largely the same hence we will do so only for the final term. The absolute value of the final term is less than or equal to

$$k^5(\sup_t|T^{0.5}H_i(t)|)(\sup_t|T^{0.5}H_j(t)|)(\sup_t|B_i(t)-B_i|)(\sup_t|B_j(t)-B_j|)\times$$

$$(\sup_t|u_{t+1}g_{i,\beta,t+1}(\dot{\beta}_{i,t})-h_{i,t+1}|)(\sup_t|u_{t+1}g_{j,\beta,t+1}(\dot{\beta}_{j,t})-h_{j,t+1}|).$$

That $\sup_t|T^{0.5}H_i(t)|$ and $\sup_t|T^{0.5}H_j(t)|$ are each $O_p(1)$ follows from Assumption 3 and Theorem 3.1 of Hansen (1992). That $\sup_t|B_i(t)-B_i|$ and $\sup_t|B_j(t)-B_j|$ are each $o_p(1)$ follows from Lemma A1. It remains to show that both $\sup_t|u_{t+1}g_{i,\beta,t+1}(\dot{\beta}_{i,t})-h_{i,t+1}|$ and $\sup_t|u_{t+1}g_{j,\beta,t+1}(\dot{\beta}_{j,t})-h_{j,t+1}|$ are $O_p(1)$. We will do so for the former. If we take a first order

Taylor expansion of $g'_{i,\beta,t+1}(\dot{\beta}_{i,t})$ about β_i^* we immediately know that for some $\ddot{\beta}_{i,t}$ on the line between $\dot{\beta}_{i,t}$ and β_i^*,

$$\sup_t |u_{t+1}g_{i,\beta,t+1}(\dot{\beta}_{i,t}) - h_{i,t+1}| \le (\sup_t |u_{t+1}|)(\sup_t |q_{i,t+1}(\ddot{\beta}_{i,t})(\dot{\beta}_{i,t} - \beta_i^*)|)$$

$$\le k(\sup_t |u_{t+1}|)(\sup_t |q_{i,t+1}|)(\sup_t |\dot{\beta}_{i,t} - \beta_i^*|) +$$

$$k(\sup_t |u_{t+1}|)(\sup_t |q_{i,t+1}(\ddot{\beta}_{i,t}) - q_{i,t+1}|)(\sup_t |\dot{\beta}_{i,t} - \beta_i^*|).$$

That $\sup_t |u_{t+1}|$ and $\sup_t |q_{i,t+1}|$ are $O_p(1)$ follows from Assumption 3. That $\sup_t |\dot{\beta}_{i,t} - \beta_i^*|$ is $o_p(1)$ follows from Lemma A1 since $\sup_t |\dot{\beta}_{i,t} - \beta_i^*| \le \sup_t |\hat{\beta}_{i,t} - \beta_i^*| = o_p(1)$. The result then follows from Assumption 2 and Lemma A1 since they imply that $\sup_t |q_{i,t+1}(\ddot{\beta}_{i,t}) - q_{i,t+1}| \le$ $(\sup_t |m_t|)(\sup_t |\ddot{\beta}_{i,t} - \beta_i^*|)^{\varphi} \le (\sup_t |m_t|)(\sup_t |\hat{\beta}_{i,t} - \beta_i^*|)^{\varphi} = O_p(1)o_p(1)$.

Lemma A14: $\sum_t (u_{1,t+1}^2(\hat{\beta}_{1,t}) - u_{1,t+1}(\hat{\beta}_{1,t})u_{2,t+1}(\hat{\beta}_{2,t}))^2 \to_d \sigma^4\chi_2$.

Proof of Lemma A14: If we take first order Taylor expansions of both $u_{1,t+1}(\hat{\beta}_{1,t})$ and $u_{2,t+1}(\hat{\beta}_{2,t})$ around β_1^* and β_2^* respectively, we have

$$\sum_t (u_{1,t+1}^2(\hat{\beta}_{1,t}) - u_{1,t+1}(\hat{\beta}_{1,t})u_{2,t+1}(\hat{\beta}_{2,t}))^2 =$$

$$[-2\sum_t u_{t+1}\{\nabla\dot{g}_{1,t+1}\}^3 + 4\sum_t u_{t+1}\{\nabla\dot{g}_{1,t+1}\}^2\{\nabla\dot{g}_{2,t+1}\} - 2\sum_t u_{t+1}\{\nabla\dot{g}_{1,t+1}\}\{\nabla\dot{g}_{2,t+1}\}^2$$

$$+ \sum_t\{\nabla\dot{g}_{1,t+1}\}^4 - 2\sum_t\{\nabla\dot{g}_{1,t+1}\}^3\{\nabla\dot{g}_{2,t+1}\} + 2\sum_t\{\nabla\dot{g}_{1,t+1}\}^2\{\nabla\dot{g}_{2,t+1}\}^2]$$

$$+ [\sum_t u_{t+1}^2\{\nabla\dot{g}_{1,t+1}\}^2 - 2\sum_t u_{t+1}^2\{\nabla\dot{g}_{1,t+1}\}\{\nabla\dot{g}_{2,t+1}\} + \sum_t u_{t+1}^2\{\nabla\dot{g}_{2,t+1}\}^2]$$

$$\equiv r_{1,T} + r_{2,T}.$$

That $r_{2,T} = \sigma^4 \sum_t \{\tilde{H}(t)\tilde{h}_{t+1}\}^2 \to_d \sigma^4 \chi_2$ follows from Lemmas A6-A8 and A13. The remainder of the proof then consists of showing that each of the six terms in $r_{1,T}$ are $o_p(1)$. We will do so for the first, the remaining terms follow from similar arguments. Taking absolute values we know that $|\sum_t u_{t+1}\{\nabla\dot{g}_{1,t+1}\}^3| \le (\sup_t |T^{1/3}\nabla\dot{g}_{1,t+1}|)^3(\sup_t |u_{t+1}|)$. That $\sup_t |T^{1/3}\nabla\dot{g}_{1,t+1}|$ is $o_p(1)$ follows from Lemma A12; that $\sup_t |u_{t+1}|$ is $O_p(1)$ follows from Assumption 3.

Proof of Theorem 3.4: (HLN) Given Lemma A14 and Theorem 3.5, the result follows from Theorem 2.1 of Hansen (1992) and the Continuous Mapping Theorem.

Proof of Theorem 3.5: (CM) If we take first order Taylor expansions of both $u_{1,t+1}(\hat{\beta}_{1,t})$ and $u_{2,t+1}(\hat{\beta}_{2,t})$ around β_1^* and β_2^* respectively, we have

$$\sum_t u_{1,t+1}^2(\hat{\beta}_{1,t}) - u_{1,t+1}(\hat{\beta}_{1,t})u_{2,t+1}(\hat{\beta}_{2,t}) = \tag{1}$$

$$\sum_t \{-u_{t+1}g'_{1,\beta,t+1}(\dot{\beta}_{1,t})B_1(t)JH(t) + u_{t+1}g'_{2,\beta,t+1}(\dot{\beta}_{2,t})B(t)H(t)\} -$$

$$\sum_t \{-H(t)'J'B_1(t)g_{1,\beta,t+1}(\dot{\beta}_{1,t})g'_{1,\beta,t+1}(\dot{\beta}_{1,t})B_1(t)JH(t) +$$

$$H(t)'J'B_1(t)g_{1,\beta,t+1}(\dot{\beta}_{1,t})g'_{2,\beta,t+1}(\dot{\beta}_{2,t})B(t)H(t)\}$$

for $\dot{\beta}_{i,t}$ on the line between $\hat{\beta}_{i,t}$ and β_i^* respectively. By Lemmas A2 and A9 we know that the first bracketed term on the r.h.s. of (1) equals $\sigma^2 \sum_t \tilde{H}'(t)\tilde{h}_{t+1}$ + $o_p(1)$. By Lemma A10 we know both

$$\sum_t -H(t)'J'B_1(t)g_{1,\beta,t+1}(\dot{\beta}_{1,t})g'_{1,\beta,t+1}(\dot{\beta}_{1,t})B_1(t)JH(t) = \sum_t -H(t)'J'B_1JH(t) + o_p(1)$$

and

$$\sum_t H(t)'J'B_1(t)g_{1,\beta,t+1}(\dot{\beta}_{1,t})g'_{2,\beta,t+1}(\dot{\beta}_{2,t})B(t)H(t) =$$

$$\sum_t H(t)'J'B_1 E(g_{1,\beta,t+1}g'_{2,\beta,t+1})BH(t) + o_p(1).$$

But $E(g_{1,\beta,t+1}g'_{2,\beta,t+1}) = JE(g_{2,\beta,t+1}g'_{2,\beta,t+1}) = JB^{-1}$. Hence the last bracketed term on the r.h.s. of (9) is $o_p(1)$. The result then follows from Lemma A4.

Proof of Theorem 3.6: (Eric) Given Theorem 3.5, the Continuous Mapping Theorem and Theorem 2.1 of Hansen (1992) it suffices to show $a_{1,T}a_{2,T} - a^2_{0,T} \to_d \sigma^4\chi_2$ for χ_2 defined in Lemma A5. That $a_{2,T} \to_p \sigma^2$ follows from Theorem 4.1 of West (1996). To show that $Pa^2_{0,T}$ = $o_p(1)$ note that Theorem 2.6 implies that $Pa_{0,T}$ = $O_p(1)$. The result follows since by Lemma A11, $Pa_{1,T} \to_d \sigma^2\chi_2$.

Table 1
Empirical Size
Artificial VAR(1)
Recursive Forecasts
Nominal Size = 10%

$R = 100$	$P = 10$	$P = 20$	$P = 40$	$P = 60$	$P = 100$	$P = 200$
Tests Compared Against Valid Asymptotic Critical Values						
OOS F	.113	.107	.100	.105	.106	.099
DM	.163	.135	.120	.110	.107	.099
CM	.118	.110	.103	.106	.105	.100
HLN	.162	.140	.121	.113	.111	.104
ERIC	.170	.142	.120	.113	.111	.104
CH	.200	.155	.132	.122	.116	.110
GC, OOS	.135	.116	.110	.108	.106	.103
GC, in-smpl.	.103	.104	.103	.105	.104	.104
Tests Compared Against Invalid Asymptotic Critical Values						
DM	.075	.055	.038	.029	.019	.010
HLN	.095	.083	.074	.069	.064	.057
ERIC	.135	.101	.080	.073	.066	.058

$R = 200$	$P = 20$	$P = 40$	$P = 80$	$P = 120$	$P = 200$	$P = 400$
Tests Compared Against Valid Asymptotic Critical Values						
OOS F	0.107	.104	.096	.104	.102	.097
DM	0.137	.122	.110	.105	.103	.097
CM	0.111	.106	.099	.106	.101	.098
HLN	0.136	.125	.109	.109	.105	.100
ERIC	0.140	.126	.109	.109	.104	.100
CH	0.152	.126	.113	.109	.105	.105
GC, OOS	0.116	.110	.103	.104	.102	.101
GC, in-smpl.	0.102	.099	.102	.102	.102	.102
Tests Compared Against Invalid Asymptotic Critical Values						
DM	0.067	.049	.034	.027	.018	.010
HLN	0.087	.076	.067	.066	.060	.056
ERIC	0.105	.084	.070	.068	.060	.056

Notes:
1. The DGP is

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} .3 & 0 \\ 0 & .5 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} e_{y,t} \\ e_{x,t} \end{pmatrix},$$

where the error terms are independent standard normal variables and, in these size experiments, $b = 0$. In each simulation, 1–step ahead forecasts of y are formed from an estimated AR model for y and an estimated VAR in y and x.
2. In each simulation, the lag lengths of the estimated models are set at one, the order of the DGP.
3. R and P refer to the number of in–sample observations and post–sample predictions, respectively.
4. Section 2 in the text defines the test statistics. The statistics included in the above set of asymptotically valid tests are compared to the correct asymptotic distributions, described in Section 2. The statistics in the set of asymptotically invalid tests are compared to the distributions that would be appropriate if the forecasting models were non–nested but are inappropriate for nested models.
5. The number of simulations is 50,000.

Table 2 Empirical Size VAR(2) in Inflation and Unemployment Recursive Forecasts Nominal Size = 10% $R = 115$, $P = 46$	
Tests Compared Against Valid Asymptotic Critical Values	
OOS F	.108
DM	.116
CM	.108
HLN	.118
ERIC	.120
CH	.139
GC, OOS	.104
GC, in-smpl.	.101
Tests Compared Against Invalid Asymptotic Critical Values	
DM	.032
HLN	.081
ERIC	.089

Notes:
1. The DGP is

$$\Delta Infl_t = -.024 - .288\Delta Infl_{t-1} - .237\Delta Infl_{t-2} + u_t$$

$$\Delta Unemp_t = -.009 + .057\Delta Infl_{t-1} + .015\Delta Infl_{t-2} + .703\Delta Unemp_{t-1} - .182\Delta Unemp_{t-2} + v_t$$

$$\mathrm{Var}(u_t) = 2.795, \ , \mathrm{Var}(v_t) = .107, \mathrm{Cov}(u_t, v_t) = -.084.$$

In each simulation, 1–step ahead forecasts of the change in inflation are formed from an estimated AR model for the change in inflation and an estimated VAR in the changes in inflation and unemployment.
2. In each simulation, the lag lengths of the estimated models are set at two, the order of the DGP.
3. R and P refer to the number of in–sample observations and post–sample predictions, respectively.
4. Section 2 in the text defines the test statistics. The statistics included in the above set of asymptotically valid tests are compared to the correct asymptotic distributions, described in Section 2. The statistics in the set of asymptotically invalid tests are compared to the distributions that would be appropriate if the forecasting models were non–nested but are inappropriate for nested models.
5. The number of simulations is 50,000.

Table 3
Size–Adjusted Power
Artificial VAR(1), $b = .1$
Recursive Forecasts
10% Signif. Empirical Critical Values

$R = 100$						
	$P = 10$	$P = 20$	$P = 40$	$P = 60$	$P = 100$	$P = 200$
OOS F	.212	.238	.280	.313	.382	.538
DM	.135	.176	.225	.257	.338	.509
CM	.230	.272	.324	.367	.455	.613
HLN	.149	.202	.266	.308	.405	.594
ERIC	.158	.201	.266	.308	.405	.595
CH	.111	.117	.124	.142	.157	.182
GC, OOS	.109	.138	.181	.208	.292	.482
GC, in-smpl.	.308	.312	.309	.301	.304	.297
$R = 200$						
	$P = 20$	$P = 40$	$P = 80$	$P = 120$	$P = 200$	$P = 400$
OOS F	.299	.345	.416	.486	.593	.757
DM	.181	.234	.315	.390	.526	.723
CM	.338	.413	.512	.584	.705	.854
HLN	.212	.287	.404	.493	.640	.829
ERIC	.215	.290	.404	.493	.639	.829
CH	.108	.116	.136	.144	.181	.227
GC, OOS	.141	.172	.260	.340	.501	.732
GC, in-smpl.	.476	.489	.482	.484	.488	.493

Notes:
1. The DGP is

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} .3 & b \\ 0 & .5 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} e_{y,t} \\ e_{x,t} \end{pmatrix},$$

where the error terms are independent standard normal variables and, in these power experiments, $b = .1$. In each simulation, 1–step ahead forecasts of y are formed from an estimated AR model for y and an estimated VAR in y and x.
2. In each simulation, the lag lengths of the estimated models are set at one, the order of the DGP.
3. R and P refer to the number of in–sample observations and post–sample predictions, respectively.
4. Section 2 in the text defines the test statistics.
5. The number of simulations is 5,000.

Table 4 Size–Adjusted Power Artificial VAR(1), $b = .2$ Recursive Forecasts 10% Signif. Empirical Critical Values						
	$R = 100$					
	$P = 10$	$P = 20$	$P = 40$	$P = 60$	$P = 100$	$P = 200$
OOS F	.405	.484	.581	.667	.780	.929
DM	.205	.293	.430	.523	.689	.897
CM	.486	.598	.728	.815	.910	.983
HLN	.260	.388	.577	.695	.855	.977
ERIC	.276	.395	.579	.699	.856	.977
CH	.140	.182	.240	.301	.401	.572
GC, OOS	.146	.227	.375	.508	.715	.938
GC, in-smpl.	.713	.712	.721	.714	.723	.719
	$R = 200$					
	$P = 20$	$P = 40$	$P = 80$	$P = 120$	$P = 200$	$P = 400$
OOS F	.558	.644	.766	.853	.932	.990
DM	.277	.408	.597	.718	.867	.984
CM	.692	.820	.925	.970	.994	1.000
HLN	.393	.588	.803	.914	.981	1.000
ERIC	.408	.594	.805	.917	.981	1.000
CH	.155	.206	.315	.394	.555	.770
GC, OOS	.226	.375	.627	.792	.937	.998
GC, in-smpl.	.937	.938	.936	.938	.938	.940

Notes:
1. The DGP is

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} .3 & b \\ 0 & .5 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} e_{y,t} \\ e_{x,t} \end{pmatrix},$$

where the error terms are independent standard normal variables and, in these power experiments, $b = .2$. In each simulation, 1–step ahead forecasts of y are formed from an estimated AR model for y and an estimated VAR in y and x.
2. In each simulation, the lag lengths of the estimated models are set at one, the order of the DGP.
3. R and P refer to the number of in–sample observations and post–sample predictions, respectively.
4. Section 2 in the text defines the test statistics.
5. The number of simulations is 5,000.

Table 5
Size–Adjusted Power
Artificial VAR(1), $b = .4$
Recursive Forecasts
10% Signif. Empirical Critical Values

	$R=100$					
	$P=10$	$P=20$	$P=40$	$P=60$	$P=100$	$P=200$
OOS F	.658	.767	.889	.943	.988	.999
DM	.317	.488	.715	.856	.968	.999
CM	.832	.930	.988	.999	1.000	1.000
HLN	.482	.738	.940	.988	.999	1.000
ERIC	.523	.754	.942	.988	.999	1.000
CH	.266	.400	.635	.776	.911	.991
GC, OOS	.282	.525	.820	.945	.995	1.000
GC, in-smpl.	.995	.995	.995	.996	.996	.997
	$R=200$					
	$P=20$	$P=40$	$P=80$	$P=120$	$P=200$	$P=400$
OOS F	.786	.902	.969	.992	.999	1.000
DM	.465	.684	.896	.970	.999	1.000
CM	.952	.994	1.000	1.000	1.000	1.000
HLN	.745	.939	.997	1.000	1.000	1.000
ERIC	.752	.943	.997	1.000	1.000	1.000
CH	.372	.590	.844	.936	.989	1.000
GC, OOS	.526	.831	.981	.999	1.000	1.000
GC, in-smpl.	1.000	1.000	1.000	1.000	1.000	1.000

Notes:
1. The DGP is

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} .3 & b \\ 0 & .5 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} e_{y,t} \\ e_{x,t} \end{pmatrix},$$

where the error terms are independent standard normal variables and, in these power experiments, $b = .4$. In each simulation, 1–step ahead forecasts of y are formed from an estimated AR model for y and an estimated VAR in y and x.
2. In each simulation, the lag lengths of the estimated models are set at one, the order of the DGP.
3. R and P refer to the number of in–sample observations and post–sample predictions, respectively.
4. Section 2 in the text defines the test statistics.
5. The number of simulations is 5,000.

Table 6 **Size–Adjusted Power** **VAR(2) in Inflation and Unemployment** **Recursive Forecasts** **10% Signif. Empirical Critical Values** $R = 115$, $P = 46$	
OOS F	.726
DM	.560
CM	.907
HLN	.783
ERIC	.785
CH	.388
GC, OOS	.547
GC, in-smpl.	.913

Notes:
1. The DGP is

$$\Delta Infl_t = -.033 - .391\Delta Infl_{t-1} - .266\Delta Infl_{t-2} - 1.207\Delta Unemp_{t-1} - .137\Delta Unemp_{t-2} + u_t$$

$$\Delta Unemp_t = -.009 + .057\Delta Infl_{t-1} + .015\Delta Infl_{t-2} + .703\Delta Unemp_{t-1} - .182\Delta Unemp_{t-2} + v_t$$

$$\text{Var}(u_t) = 2.519,\ ,\text{Var}(v_t) = .107, \text{Cov}(u_t, v_t) = -.084.$$

In each simulation, 1–step ahead forecasts of the change in inflation are formed from an estimated AR model for the change in inflation and an estimated VAR in the changes in inflation and unemployment.
2. In each simulation, the lag lengths of the estimated models are set at two, the order of the DGP.
3. R and P refer to the number of in–sample observations and post–sample predictions, respectively.
4. Section 2 in the text defines the test statistics.
5. The number of simulations is 5,000.

Table 7			
Testing the Predictive Content of Unemployment for Inflation			
Recursive Forecasts			
$R = 115$, $P = 46$			
In–Sample Model Estimates, 1958:Q3 to 1987:Q1			
Explanatory	*Dependent variable*		
variable	$Inflation_t$	$Inflation_t$	$Unemployment_t$
$Constant$	.024 (.154)	.033 (.148)	-.009 (.031)
$Inflation_{t-1}$	-.288 (.092)	-.391 (.093)	.057 (.019)
$Inflation_{t-2}$	-.237 (.092)	-.266 (.097)	.015 (.020)
$Unemployment_{t-1}$		-1.207 (.454)	.703 (.093)
$Unemployment_{t-2}$		-.137 (.457)	-.182 (.094)
$\bar{R}^2$	.092	.166	.356
Tests of Predictive Power of Unemployment for Inflation			
	Test statistics	*Asymptotic critical values*	*Empirical critical values*
MSE, AR	.420		
MSE, VAR	.412		
OOS F	.839	1.029	1.111
DM	.100	.614	.711
CM	5.186	1.019	1.080
HLN	1.124	1.282	1.191
ERIC	1.247	1.282	1.206
CH	1.853	-1.645,1.645	-2.114,1.345
GC, OOS	2.040	2.437	2.469
GC, in-smpl.	6.011	2.351	2.356

Notes:
1. 1–step ahead forecasts of the change in inflation are formed from an estimated AR model for the change in inflation and an estimated VAR in the changes in inflation and unemployment.
2. The significance level of the tests is 10%.
3. R and P refer to the number of in–sample observations and post–sample predictions, respectively.
4. Section 2 in the text defines the test statistics.

Table A1

Percentiles of the HLN and ERIC statistics: Recursive scheme

$k_2 \backslash \pi$		0.1	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
1	(0.99)	2.102	2.030	1.997	2.006	2.054	2.052	2.036	2.033	2.030	2.077	2.082
	(0.95)	1.422	1.360	1.338	1.399	1.350	1.331	1.350	1.360	1.332	1.329	1.322
	(0.90)	1.056	1.002	1.005	0.995	0.968	0.955	0.942	0.939	0.959	0.956	0.939
2	(0.99)	2.187	2.214	2.073	2.089	2.178	2.195	2.169	2.151	2.164	2.135	2.115
	(0.95)	1.505	1.467	1.445	1.441	1.413	1.413	1.425	1.439	1.427	1.440	1.443
	(0.90)	1.166	1.101	1.086	1.096	1.077	1.066	1.060	1.044	1.049	1.053	1.035
3	(0.99)	2.155	2.144	2.203	2.180	2.146	2.143	2.151	2.127	2.099	2.148	2.134
	(0.95)	1.574	1.525	1.529	1.496	1.462	1.476	1.491	1.475	1.478	1.501	1.473
	(0.90)	1.227	1.138	1.105	1.136	1.118	1.113	1.124	1.111	1.098	1.094	1.114
4	(0.99)	2.230	2.200	2.273	2.232	2.158	2.137	2.181	2.195	2.158	2.178	2.183
	(0.95)	1.594	1.596	1.552	1.532	1.469	1.463	1.482	1.472	1.483	1.473	1.481
	(0.90)	1.219	1.175	1.192	1.177	1.136	1.132	1.111	1.099	1.104	1.112	1.111
5	(0.99)	2.233	2.215	2.245	2.162	2.102	2.172	2.163	2.204	2.168	2.136	2.179
	(0.95)	1.567	1.583	1.544	1.557	1.470	1.460	1.474	1.475	1.464	1.459	1.472
	(0.90)	1.205	1.192	1.170	1.172	1.117	1.092	1.116	1.092	1.084	1.102	1.100
6	(0.99)	2.265	2.240	2.216	2.188	2.190	2.212	2.218	2.198	2.182	2.182	2.207
	(0.95)	1.576	1.548	1.536	1.546	1.498	1.491	1.477	1.509	1.507	1.496	1.478
	(0.90)	1.222	1.187	1.162	1.169	1.125	1.109	1.125	1.140	1.132	1.126	1.137
7	(0.99)	2.267	2.281	2.200	2.138	2.152	2.184	2.216	2.212	2.143	2.168	2.206
	(0.95)	1.630	1.577	1.574	1.551	1.543	1.492	1.525	1.527	1.500	1.514	1.498
	(0.90)	1.223	1.178	1.180	1.195	1.164	1.142	1.168	1.169	1.165	1.137	1.134
8	(0.99)	2.253	2.284	2.231	2.210	2.106	2.156	2.199	2.195	2.181	2.182	2.225
	(0.95)	1.653	1.566	1.592	1.548	1.513	1.519	1.554	1.534	1.522	1.534	1.530
	(0.90)	1.211	1.195	1.205	1.193	1.181	1.159	1.187	1.169	1.169	1.152	1.151
9	(0.99)	2.243	2.296	2.194	2.248	2.142	2.224	2.202	2.223	2.227	2.243	2.212
	(0.95)	1.613	1.595	1.557	1.532	1.561	1.530	1.543	1.546	1.539	1.549	1.544
	(0.90)	1.213	1.223	1.203	1.208	1.187	1.166	1.181	1.190	1.179	1.172	1.168
10	(0.99)	2.274	2.231	2.220	2.223	2.143	2.146	2.197	2.232	2.233	2.251	2.243
	(0.95)	1.602	1.609	1.556	1.510	1.546	1.496	1.514	1.519	1.539	1.531	1.518
	(0.90)	1.244	1.226	1.208	1.194	1.176	1.175	1.177	1.178	1.162	1.175	1.170

Table A2
Percentiles of the HLN and ERIC statistics: Rolling scheme

$k_2\backslash\pi$		0.1	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
1	(0.99)	2.069	2.055	1.967	2.016	2.066	2.049	1.998	1.978	2.039	2.028	1.996
	(0.95)	1.413	1.383	1.348	1.372	1.341	1.338	1.334	1.325	1.342	1.350	1.344
	(0.90)	1.070	1.028	1.019	0.999	0.951	0.949	0.948	0.964	0.953	0.965	0.992
2	(0.99)	2.207	2.117	2.059	2.051	2.100	2.140	2.151	2.149	2.161	2.126	2.109
	(0.95)	1.508	1.456	1.419	1.439	1.410	1.407	1.421	1.394	1.401	1.410	1.394
	(0.90)	1.150	1.091	1.067	1.083	1.064	1.022	1.044	1.059	1.029	1.053	1.073
3	(0.99)	2.143	2.138	2.182	2.126	2.107	2.208	2.197	2.194	2.093	2.112	2.127
	(0.95)	1.575	1.504	1.496	1.467	1.440	1.447	1.426	1.430	1.455	1.434	1.432
	(0.90)	1.223	1.121	1.104	1.110	1.082	1.089	1.086	1.067	1.062	1.075	1.089
4	(0.99)	2.226	2.184	2.242	2.188	2.150	2.194	2.226	2.222	2.130	2.175	2.152
	(0.95)	1.587	1.584	1.570	1.490	1.450	1.461	1.469	1.499	1.497	1.484	1.499
	(0.90)	1.203	1.184	1.185	1.104	1.100	1.100	1.092	1.081	1.103	1.123	1.117
5	(0.99)	2.195	2.181	2.197	2.198	2.168	2.157	2.226	2.218	2.148	2.162	2.167
	(0.95)	1.600	1.577	1.504	1.476	1.449	1.480	1.460	1.501	1.483	1.480	1.495
	(0.90)	1.177	1.201	1.160	1.125	1.063	1.106	1.105	1.127	1.110	1.119	1.127
6	(0.99)	2.296	2.316	2.169	2.211	2.186	2.199	2.216	2.191	2.118	2.194	2.163
	(0.95)	1.576	1.544	1.516	1.483	1.484	1.516	1.485	1.513	1.510	1.509	1.524
	(0.90)	1.214	1.170	1.153	1.149	1.140	1.119	1.117	1.137	1.125	1.115	1.161
7	(0.99)	2.326	2.300	2.192	2.193	2.215	2.210	2.252	2.184	2.146	2.207	2.184
	(0.95)	1.584	1.546	1.580	1.506	1.520	1.539	1.529	1.557	1.528	1.522	1.509
	(0.90)	1.218	1.201	1.198	1.187	1.171	1.141	1.132	1.152	1.153	1.167	1.161
8	(0.99)	2.249	2.268	2.208	2.221	2.218	2.213	2.215	2.227	2.191	2.183	2.145
	(0.95)	1.611	1.597	1.578	1.508	1.527	1.549	1.518	1.523	1.529	1.511	1.531
	(0.90)	1.232	1.198	1.177	1.150	1.168	1.175	1.163	1.187	1.169	1.177	1.184
9	(0.99)	2.262	2.217	2.203	2.199	2.221	2.191	2.201	2.250	2.225	2.222	2.161
	(0.95)	1.606	1.619	1.554	1.543	1.580	1.582	1.561	1.540	1.569	1.519	1.560
	(0.90)	1.221	1.228	1.204	1.186	1.185	1.216	1.183	1.198	1.190	1.192	1.198
10	(0.99)	2.266	2.212	2.236	2.261	2.221	2.173	2.233	2.232	2.179	2.216	2.199
	(0.95)	1.604	1.613	1.541	1.507	1.555	1.551	1.517	1.553	1.509	1.502	1.523
	(0.90)	1.240	1.241	1.206	1.168	1.190	1.181	1.169	1.183	1.170	1.179	1.190

Table A3
Percentiles of the HLN and ERIC statistics: Fixed scheme

$k_2\backslash\pi$		0.1	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
1	(0.99)	2.321	2.184	2.222	2.275	2.334	2.320	2.336	2.338	2.318	2.415	2.379
	(0.95)	1.622	1.570	1.598	1.685	1.681	1.606	1.663	1.672	1.664	1.682	1.658
	(0.90)	1.269	1.233	1.272	1.310	1.299	1.256	1.295	1.304	1.297	1.305	1.301
2	(0.99)	2.323	2.349	2.212	2.305	2.387	2.292	2.364	2.354	2.271	2.359	2.297
	(0.95)	1.673	1.604	1.659	1.677	1.680	1.645	1.666	1.680	1.684	1.681	1.679
	(0.90)	1.293	1.254	1.269	1.316	1.304	1.284	1.312	1.287	1.272	1.286	1.282
3	(0.99)	2.260	2.313	2.413	2.374	2.327	2.224	2.318	2.361	2.298	2.389	2.327
	(0.95)	1.665	1.629	1.666	1.684	1.675	1.666	1.681	1.693	1.659	1.694	1.686
	(0.90)	1.301	1.250	1.269	1.287	1.306	1.305	1.339	1.314	1.290	1.291	1.311
4	(0.99)	2.266	2.310	2.370	2.330	2.255	2.281	2.369	2.328	2.332	2.361	2.396
	(0.95)	1.647	1.697	1.684	1.700	1.672	1.616	1.667	1.641	1.650	1.633	1.640
	(0.90)	1.292	1.278	1.302	1.272	1.261	1.284	1.286	1.267	1.292	1.271	1.273
5	(0.99)	2.255	2.309	2.324	2.346	2.230	2.229	2.276	2.327	2.317	2.333	2.311
	(0.95)	1.645	1.632	1.650	1.667	1.632	1.608	1.644	1.612	1.629	1.644	1.625
	(0.90)	1.274	1.278	1.262	1.261	1.232	1.248	1.284	1.263	1.239	1.254	1.258
6	(0.99)	2.300	2.398	2.306	2.316	2.238	2.258	2.346	2.321	2.309	2.293	2.293
	(0.95)	1.610	1.630	1.632	1.632	1.648	1.630	1.678	1.637	1.662	1.694	1.634
	(0.90)	1.269	1.255	1.271	1.262	1.261	1.275	1.316	1.290	1.281	1.284	1.290
7	(0.99)	2.315	2.403	2.269	2.238	2.266	2.322	2.339	2.328	2.322	2.317	2.298
	(0.95)	1.649	1.624	1.632	1.646	1.650	1.638	1.649	1.646	1.627	1.655	1.623
	(0.90)	1.298	1.285	1.305	1.275	1.286	1.284	1.305	1.262	1.261	1.265	1.264
8	(0.99)	2.276	2.372	2.282	2.242	2.278	2.296	2.275	2.298	2.324	2.347	2.319
	(0.95)	1.660	1.642	1.656	1.634	1.648	1.645	1.675	1.646	1.648	1.632	1.617
	(0.90)	1.307	1.247	1.276	1.279	1.300	1.290	1.309	1.275	1.254	1.258	1.270
9	(0.99)	2.297	2.344	2.269	2.207	2.254	2.299	2.290	2.324	2.325	2.343	2.340
	(0.95)	1.646	1.653	1.655	1.623	1.652	1.653	1.668	1.646	1.643	1.628	1.619
	(0.90)	1.310	1.280	1.297	1.279	1.288	1.295	1.302	1.274	1.266	1.269	1.283
10	(0.99)	2.287	2.263	2.263	2.216	2.231	2.266	2.273	2.291	2.300	2.314	2.323
	(0.95)	1.631	1.671	1.636	1.618	1.628	1.625	1.631	1.638	1.612	1.635	1.606
	(0.90)	1.300	1.289	1.277	1.263	1.272	1.283	1.272	1.291	1.255	1.249	1.275

Table A4
Percentiles of the CM statistic: Recursive scheme

$k_2 \backslash \pi$		0.1	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
1	(0.99)	0.954	1.397	2.098	2.662	2.904	3.209	3.436	3.611	3.722	3.920	4.134
	(0.95)	0.520	0.744	1.079	1.312	1.467	1.584	1.724	1.845	1.967	2.084	2.085
	(0.90)	0.335	0.473	0.685	0.791	0.925	0.984	1.076	1.137	1.208	1.253	1.280
2	(0.99)	1.258	1.854	2.604	3.172	3.543	4.102	4.446	4.647	4.705	4.971	5.107
	(0.95)	0.766	1.028	1.481	1.779	2.016	2.234	2.510	2.580	2.663	2.785	2.889
	(0.90)	0.524	0.716	1.019	1.228	1.371	1.471	1.573	1.694	1.791	1.893	1.914
3	(0.99)	1.529	2.115	3.098	3.658	4.128	4.574	4.936	5.221	5.179	5.484	5.805
	(0.95)	0.940	1.273	1.865	2.256	2.442	2.709	2.909	3.065	3.250	3.463	3.564
	(0.90)	0.686	0.890	1.285	1.532	1.718	1.905	2.097	2.183	2.238	2.285	2.366
4	(0.99)	1.711	2.479	3.462	4.060	4.613	5.066	5.435	5.690	6.002	6.221	6.507
	(0.95)	1.060	1.526	2.181	2.576	2.743	3.007	3.313	3.463	3.636	3.769	3.894
	(0.90)	0.776	1.062	1.528	1.789	1.951	2.169	2.307	2.399	2.542	2.635	2.727
5	(0.99)	1.870	2.654	3.651	4.398	4.798	5.517	5.808	6.352	6.667	6.870	7.074
	(0.95)	1.186	1.671	2.349	2.788	2.972	3.283	3.592	3.808	4.009	4.121	4.307
	(0.90)	0.864	1.198	1.639	1.970	2.166	2.346	2.561	2.660	2.700	2.921	3.033
6	(0.99)	2.053	2.866	3.976	4.699	5.416	5.957	6.526	6.868	7.029	7.321	7.309
	(0.95)	1.281	1.790	2.514	3.008	3.260	3.586	3.966	4.236	4.407	4.524	4.701
	(0.90)	0.934	1.321	1.800	2.161	2.376	2.632	2.794	2.891	3.095	3.250	3.389
7	(0.99)	2.269	3.047	4.203	4.783	5.813	6.170	6.850	7.249	7.611	7.922	8.283
	(0.95)	1.417	1.936	2.768	3.280	3.606	3.892	4.270	4.617	4.788	4.910	5.103
	(0.90)	1.024	1.439	2.026	2.398	2.669	2.912	3.074	3.220	3.403	3.468	3.662
8	(0.99)	2.289	3.109	4.356	5.327	5.915	6.529	7.037	7.700	7.985	8.385	8.577
	(0.95)	1.518	2.121	2.988	3.470	3.821	4.226	4.594	4.842	4.999	5.231	5.289
	(0.90)	1.115	1.535	2.144	2.544	2.865	3.073	3.339	3.474	3.696	3.805	3.998
9	(0.99)	2.473	3.311	4.747	5.625	6.116	6.959	7.601	8.279	8.544	8.949	9.318
	(0.95)	1.584	2.300	3.100	3.667	4.144	4.555	4.886	5.159	5.435	5.556	5.630
	(0.90)	1.171	1.650	2.299	2.693	3.072	3.271	3.576	3.769	3.925	4.041	4.328
10	(0.99)	2.548	3.442	4.758	5.815	6.433	7.224	8.076	8.450	8.832	9.367	9.928
	(0.95)	1.695	2.394	3.218	3.729	4.294	4.576	4.949	5.283	5.522	5.728	5.923
	(0.90)	1.250	1.749	2.435	2.834	3.174	3.413	3.715	3.914	4.035	4.247	4.388

Table A5

Percentiles of the CM statistic: Rolling scheme

$k_2 \backslash \pi$		0.1	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
1	(0.99)	0.997	1.466	2.278	2.854	3.264	3.676	3.852	4.410	4.614	4.835	5.064
	(0.95)	0.534	0.788	1.161	1.506	1.722	1.946	2.198	2.374	2.498	2.683	2.836
	(0.90)	0.339	0.502	0.764	0.919	1.082	1.210	1.355	1.493	1.598	1.703	1.808
2	(0.99)	1.282	1.999	2.929	3.596	4.282	4.799	5.389	5.568	6.062	6.338	6.502
	(0.95)	0.793	1.089	1.640	2.031	2.380	2.704	2.949	3.212	3.366	3.593	3.723
	(0.90)	0.539	0.760	1.109	1.402	1.632	1.824	1.991	2.200	2.302	2.471	2.652
3	(0.99)	1.561	2.237	3.401	4.218	4.893	5.609	6.055	6.442	6.646	7.198	7.513
	(0.95)	0.962	1.391	1.955	2.500	2.847	3.271	3.549	3.759	4.108	4.219	4.435
	(0.90)	0.695	0.941	1.409	1.746	2.003	2.275	2.475	2.616	2.750	3.051	3.241
4	(0.99)	1.783	2.588	3.681	4.603	5.312	6.192	6.845	7.326	7.514	8.173	8.529
	(0.95)	1.087	1.607	2.310	2.779	3.220	3.681	4.079	4.510	4.811	5.051	5.289
	(0.90)	0.791	1.137	1.637	1.974	2.265	2.642	2.872	3.066	3.275	3.516	3.655
5	(0.99)	1.921	2.830	4.039	5.028	5.918	6.775	7.423	7.848	8.310	8.578	9.163
	(0.95)	1.211	1.773	2.545	3.073	3.522	4.082	4.468	4.899	5.199	5.470	5.840
	(0.90)	0.871	1.249	1.791	2.203	2.464	2.834	3.124	3.435	3.620	3.845	4.095
6	(0.99)	2.120	2.945	4.424	5.303	6.480	7.193	8.140	8.407	8.768	9.404	9.885
	(0.95)	1.334	1.870	2.766	3.413	3.939	4.473	4.881	5.326	5.675	5.979	6.288
	(0.90)	0.984	1.380	1.950	2.432	2.869	3.169	3.429	3.782	4.011	4.282	4.576
7	(0.99)	2.324	3.213	4.580	5.799	6.670	7.609	8.313	8.759	9.160	10.097	10.305
	(0.95)	1.465	2.082	3.029	3.696	4.299	4.904	5.364	5.805	6.199	6.389	6.763
	(0.90)	1.066	1.533	2.188	2.691	3.138	3.461	3.800	4.086	4.455	4.741	4.951
8	(0.99)	2.376	3.461	4.945	6.097	7.165	7.964	8.792	9.237	9.982	10.789	10.985
	(0.95)	1.557	2.241	3.217	3.865	4.610	5.306	5.644	6.049	6.484	6.759	7.231
	(0.90)	1.148	1.616	2.316	2.900	3.377	3.789	4.124	4.478	4.836	4.958	5.396
9	(0.99)	2.538	3.531	5.151	6.382	7.546	8.277	9.132	10.226	10.724	11.154	11.710
	(0.95)	1.657	2.403	3.397	4.223	4.972	5.585	5.996	6.432	6.860	7.111	7.670
	(0.90)	1.226	1.779	2.471	3.096	3.605	4.124	4.422	4.773	5.149	5.449	5.693
10	(0.99)	2.569	3.600	5.365	6.699	7.705	8.624	9.630	10.375	10.928	11.573	12.043
	(0.95)	1.709	2.526	3.519	4.382	5.138	5.719	6.129	6.694	6.975	7.349	7.867
	(0.90)	1.284	1.838	2.627	3.149	3.763	4.186	4.580	4.937	5.273	5.538	5.968

Table A6
Percentiles of the CM statistic: Fixed scheme

$k_2 \backslash \pi$		0.1	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
1	(0.99)	0.908	1.300	1.941	2.421	2.813	3.069	3.307	3.550	3.821	4.016	4.196
	(0.95)	0.489	0.698	1.037	1.231	1.477	1.622	1.792	1.881	2.089	2.187	2.272
	(0.90)	0.322	0.449	0.670	0.795	0.942	1.074	1.190	1.305	1.378	1.421	1.462
2	(0.99)	1.154	1.669	2.411	2.987	3.579	3.870	4.233	4.535	4.777	4.922	5.082
	(0.95)	0.729	1.002	1.486	1.790	2.049	2.307	2.545	2.733	2.948	3.071	3.249
	(0.90)	0.500	0.701	1.003	1.247	1.473	1.621	1.782	1.931	2.047	2.128	2.193
3	(0.99)	1.469	1.965	2.935	3.492	4.060	4.531	4.914	5.360	5.603	6.027	6.309
	(0.95)	0.901	1.227	1.799	2.234	2.618	2.925	3.240	3.477	3.742	3.808	4.042
	(0.90)	0.663	0.882	1.322	1.607	1.860	2.065	2.286	2.417	2.642	2.779	2.954
4	(0.99)	1.610	2.236	3.168	3.978	4.658	5.256	5.613	6.232	6.570	7.085	7.333
	(0.95)	1.015	1.433	2.119	2.527	2.886	3.284	3.561	3.856	4.161	4.341	4.429
	(0.90)	0.761	1.066	1.525	1.797	2.075	2.375	2.601	2.797	2.962	3.124	3.342
5	(0.99)	1.758	2.474	3.344	4.190	5.013	5.431	5.961	6.493	6.891	7.536	7.675
	(0.95)	1.135	1.582	2.295	2.807	3.185	3.520	3.999	4.267	4.524	4.808	5.044
	(0.90)	0.851	1.201	1.671	2.020	2.359	2.641	2.930	3.125	3.354	3.535	3.721
6	(0.99)	1.982	2.672	3.783	4.697	5.423	6.191	6.633	7.287	7.653	8.278	8.791
	(0.95)	1.251	1.738	2.506	3.051	3.509	3.980	4.409	4.690	5.033	5.393	5.658
	(0.90)	0.922	1.301	1.882	2.277	2.640	3.015	3.322	3.504	3.802	3.967	4.214
7	(0.99)	2.120	2.893	4.068	4.995	5.937	6.812	7.392	7.747	8.163	8.943	9.495
	(0.95)	1.395	1.910	2.729	3.303	3.872	4.290	4.847	5.098	5.393	5.861	6.048
	(0.90)	1.030	1.440	2.078	2.531	2.906	3.230	3.517	3.771	4.084	4.277	4.520
8	(0.99)	2.218	2.983	4.348	5.394	6.321	7.294	7.906	8.329	8.730	9.012	9.632
	(0.95)	1.497	2.032	2.905	3.548	4.223	4.696	5.121	5.421	5.800	6.178	6.495
	(0.90)	1.106	1.559	2.208	2.719	3.125	3.515	3.822	4.089	4.360	4.564	4.854
9	(0.99)	2.301	3.163	4.623	5.551	6.673	7.330	8.101	8.879	9.003	9.876	10.113
	(0.95)	1.552	2.205	3.145	3.744	4.511	5.013	5.433	5.937	6.243	6.620	6.829
	(0.90)	1.189	1.659	2.411	2.893	3.308	3.767	4.093	4.370	4.600	4.937	5.226
10	(0.99)	2.392	3.267	4.753	5.839	6.753	7.617	8.328	9.160	9.706	10.487	10.794
	(0.95)	1.607	2.312	3.238	3.896	4.602	5.213	5.679	6.099	6.412	6.778	7.084
	(0.90)	1.252	1.754	2.486	3.018	3.450	3.915	4.282	4.585	4.830	5.157	5.527