TESTS OF EQUAL FORECAST ACCURACY
AND ENCOMPASSING FOR NESTED MODELS

Todd E. Clark
Michael W. McCracken

OCTOBER 1999

RWP 99-11

Research Division
Federal Reserve Bank of Kansas City

Todd E. Clark is an assistant vice president and economist at the Federal
Reserve Bank of Kansas City. Michael W. McCracken is an assistant
professor of economics at Louisiana State University. The authors thank
Hilary Croke for valuable research assistance and Charles Engel, Lutz
Kilian, Norm Swanson, Dek Terrell, Ken West, and seminar participants
at Penn State and the University of Michigan for helpful comments. The
views expressed in this paper are those of the authors and do not
necessarily reflect the views of the Federal Reserve Bank of Kansas City
or the Federal Reserve System.

Clark e-mail: todd.e.clark@kc.frb.org
McCracken e-mail: mmccrac@unix1.sncc.lsu.edu

Abstract

We examine the asymptotic and finite-sample properties of tests for equal forecast accuracy and encompassing applied to 1-step ahead forecasts from nested parametric models. We first derive the asymptotic distributions of two standard tests and one new test of encompassing. Tables of asymptotically valid critical values are provided. Monte Carlo methods are then used to evaluate the size and power of the tests of equal forecast accuracy and encompassing. The simulations indicate that post-sample tests can be reasonably well sized. Of the post-sample tests considered, the encompassing test proposed in this paper is the most powerful. We conclude with an empirical application regarding the predictive content of unemployment for inflation.

1. Introduction

Since the influential work of Meese and Rogoff (1983, 1988), it has become common to use comparisons of out-of-sample forecasts to determine whether one variable has predictive power for another.[1] Typically, this out-of-sample comparison is made in two stages. First, forecasts of the variable of interest are constructed once using a model that includes a variable with putative predictive content and then a second time excluding that variable. Second, given the two sequences of forecast errors, tests of equal forecast accuracy or forecast encompassing are conducted. This out-of-sample approach is explicitly advocated by Ashley, Granger, and Schmalensee (1980), who argue that it is more in the spirit of the definition of Granger causality to employ post-sample forecast tests than to employ the standard full-sample causality test.

Although post-sample tests of this type are increasingly used, little is known about their effectiveness. Virtually all evidence on the asymptotic and finite-sample behavior of tests of equal forecast accuracy and encompassing pertain to forecasts from non-nested models. Diebold and Mariano (1995), West (1996, 1999), Harvey, Leybourne, and Newbold (1997, 1998), West and McCracken (1998), Clark (1999), Corradi, Swanson, and Olivetti (1999), and McCracken (1999a) each present results for non-nested forecasts. Yet when the forecasting models are nested rather than non-nested, many of the usual test statistics, such as the encompassing test of Ericsson (1992), fail to converge to the standard normal distribution.[2] This implies that critical values taken from the standard normal distribution are asymptotically invalid for testing equal accuracy or encompassing between forecasts from nested models.

---

[1] Examples of studies using this methodology include Diebold and Rudebusch (1991), Amano and van Norden (1995), Chinn and Meese (1995), Mark (1995), Krueger and Kuttner (1996), Blomberg and Hess (1997), Bram and Ludvigson (1998), Berkowitz and Giorgianni (1999), Evans and Lyons (1999), and Kilian (1999).

[2] One exception is the Chong and Hendry (1986) test of forecast encompassing. West and McCracken (1998) show that it can be asymptotically normal when applied to either nested or non-nested forecasts. In our simulations however, the power of this test was dominated by that of the other encompassing tests and hence has been excluded in order to limit the number of tables.

To fill the existing void, this paper examines the asymptotic and finite-sample properties of tests for equal accuracy and encompassing applied to 1-step ahead forecasts from nested parametric models. We first derive the asymptotic distributions of two standard tests and one new test of encompassing. The standard tests are those proposed by Ericsson (1992) and Harvey, Leybourne, and Newbold (1998). The new statistic we propose is a variant of these two tests. As in West (1996, 1999), West and McCracken (1998), Corradi, Swanson, and Olivetti (1999), and McCracken (1999a), the derived asymptotic distributions of the tests explicitly account for the uncertainty introduced by parameter estimation. To facilitate the use of the limiting distributions derived here, asymptotically valid critical values are generated numerically and reported in a set of tables. Using the same basic framework, McCracken (1999b) develops the asymptotic distributions and provides asymptotic critical values for tests of equal mean squared error (MSE) proposed by Granger and Newbold (1977) and Diebold and Mariano (1995), as well as a new F-type test.

We then evaluate the finite-sample size and size-adjusted power of these equal accuracy and encompassing tests using Monte Carlo simulations based upon different VAR data-generating processes. For comparison, the set of tests also includes a full-sample F-test of Granger causality. The post-sample tests are evaluated using tabulated asymptotic critical values provided in this paper and in McCracken (1999b). For those post-sample tests that would be asymptotically standard normal if the forecasting models were non-nested, we also compare the statistics against standard normal critical values in order to evaluate whether using incorrect critical values can yield misleading inferences. Finally, to illustrate how the tests perform in practical settings, each test is used to determine whether the unemployment rate has predictive content for inflation in quarterly U.S. data.

Our Monte Carlo analysis produces four key results. First, in most settings, each of the post-sample tests is reasonably well sized. In many instances, the size distortions associated with the post-sample tests are smaller than those associated with the full-sample F-test of causality. Second, comparing the post-sample forecast statistics against the inappropriate standard normal critical values makes the tests undersized. Third, when the features of the data-generating process make lag selection sufficiently imprecise, the post-sample tests suffer more substantial size distortions and lose some of their advantage over the full-sample F-test of causality. Accordingly, post-sample forecast tests are not necessarily a panacea for in-sample overfitting; many of the problems that lead to in-sample overfitting also lead to post-sample overfitting. Finally, the powers of the post-sample forecast tests permit some simple rankings, in which the new encompassing statistic proposed in this paper is most powerful. In some settings, the power of the new encompassing test rivals the power of the full-sample F-test of causality even though the full-sample test uses many more observations.

The remainder of the paper proceeds as follows. Section 2 introduces the notation and general environment under which the forecasts are generated and the tests of equal forecast accuracy and encompassing are constructed. Section 3 defines the test statistics considered and provides the null asymptotic results. In section 4 we present a Monte Carlo evaluation of the finite-sample size and power properties of the tests. Section 5 uses the tests to determine whether the unemployment rate has predictive power for inflation. Section 6 concludes. Proofs are contained within Appendix 1. Appendix 2 contains further detail on one of the Lemmas.

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 observations $\{z_t^{'}\}_{t=1}^{T+1} \equiv \{y_t, x_t^{'}\}_{t=1}^{T+1}$ includes a scalar random variable y_t to

be predicted and a vector of predictors x_t. The data sample 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, the out-of-sample observations span $R + 1$ through $R + P$. The total number

of observations in the sample is $R + P = T + 1$. The largest number of observations used to

estimate the model parameters is $T = R + P - 1$.

Forecasts of y_{t+1}, $t = R,\ldots,T$, are generated using two parametric models, $g_i(x_{t+1}, \beta_i^{*}) \equiv$

$g_{i,t+1}(\beta_i^{*})$, $i = 1,2$, each of which is estimated. Model 2 is unrestricted and nests the restricted

model 1. Under the null, 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.

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 through time. The first prediction, $g_{i,R+1}(\hat{\beta}_{i,R})$, is created using the

parameter estimate $\hat{\beta}_{i,R}$ based on data from 1 to R. The second prediction, $g_{i,R+2}(\hat{\beta}_{i,R+1})$, is

created using the parameter estimate $\hat{\beta}_{i,R+1}$ based on data from 1 to $R + 1$. In general, for t =

R,…,T, the prediction of y_{t+1}, $g_{i,t+1}(\hat{\beta}_{i,t})$, is created using the parameter estimate $\hat{\beta}_{i,t}$ based on

data from 1 to t.

Under the *rolling* scheme, model parameters are estimated using only the most recent R

observations. The first prediction, $g_{i,R+1}(\hat{\beta}_{i,R})$, is created using the parameter estimate $\hat{\beta}_{i,R}$

based on data from 1 to R. The second prediction, $g_{i,R+2}(\hat{\beta}_{i,R+1})$, is created using the parameter

estimate $\hat{\beta}_{i,R+1}$ based on data from 2 to R + 1. In general, for t = R,...,T, the prediction of y_{t+1},

$g_{i,t+1}(\hat{\beta}_{i,t})$, is created using the parameter estimate $\hat{\beta}_{i,t}$ based on data from t − R + 1 to t. Note

that under the rolling scheme the parameter estimate $\hat{\beta}_{i,t}$ should also be subscripted by R in order

to reflect the width of the sample window. To reduce notation we leave that subscript implicit.

Under the *fixed* scheme, each forecast is generated using parameters that are estimated

only once using data from 1 to R. Hence for each prediction of y_{t+1}, $g_{i,t+1}(\hat{\beta}_{i,t}) = g_{i,t+1}(\hat{\beta}_{i,R})$; the

prediction is created using the same parameter estimate $\hat{\beta}_{i,t} = \hat{\beta}_{i,R}$ based on data from 1 to R. As

was the case for the rolling scheme, under the fixed scheme the parameter estimates $\hat{\beta}_{i,t}$ should

be subscripted by R. To reduce notation we also leave this subscript implicit.

For each of the three forecasting schemes, the 1-step ahead forecast errors are

$\hat{u}_{1,t+1} = y_{t+1} - g_{1,t+1}(\hat{\beta}_{1,t})$ and $\hat{u}_{2,t+1} = y_{t+1} - g_{2,t+1}(\hat{\beta}_{2,t})$ 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+1} we let $\sum_t z_{t+1}$ denote the

summation $\sum_{t=R}^{T} z_{t+1}$. For example, the MSE for model i is $MSE_i \equiv P^{-1}\sum_{t=R}^{T} \hat{u}_{i,t+1}^2 = P^{-1}\sum_t \hat{u}_{i,t+1}^2$.

Before getting to the assumptions some final notation is needed. For any function f let

$f_{i,t+1} = f_{i,t+1}(\beta_i^*)$. Let $g_{i,\beta,t+1}(\beta_i) = \partial g_{i,t+1}(\beta_i)/\partial\beta_i$, $g_{i,\beta\beta,t+1}(\beta_i) = \partial g_{i,\beta,t+1}(\beta_i)/\partial\beta_i'$, $h_{i,t+1}(\beta_i) =$

$(y_{t+1} - g_{i,t+1}(\beta_i))g_{i,\beta,t+1}(\beta_i)$, $q_{i,t+1}(\beta_i) = g_{i,\beta,t+1}(\beta_i)g_{i,\beta,t+1}'(\beta_i) - (y_{t+1} - g_{i,t+1}(\beta_i))g_{i,\beta\beta,t+1}(\beta_i)$,

and $B_i = (Eq_{i,t+1})^{-1}$. Let W(s) denote a ($k_2\times1$) vector standard Brownian Motion. For any

(m×n) matrix A with elements $a_{i,j}$ and column vectors a_j let vec(A) denote the (mn×1) vector $[a_1^{'}, a_2^{'}, ..., a_n^{'}]^{'}$ and let |A| denote $\max_{i,j} |a_{i,j}|$. Finally, under the null $u_{1,t} = u_{2,t} \equiv u_t$.

Given the definitions and the three forecasting schemes described above, the following five assumptions are used to derive the limiting distributions of encompassing tests presented in Theorems 3.5, 3.6, and 3.7. The assumptions are also sufficient for the results of McCracken (1999b) when MSE is the measure of predictive ability. The assumptions are intended to be only sufficient, not necessary and sufficient.

<u>Assumption 1:</u> The parameter estimates $\hat{\beta}_{i,t}$, i = 1,2, t = R,...,T, satisfy $\hat{\beta}_{i,t} - \beta_i^* = B_i(t)H_i(t)$. For $\dot{\beta}_{i,t}$ on the line between $\hat{\beta}_{i,t}$ and β_i^*, $B_i(t)H_i(t)$ equals $(t^{-1}\sum_{j=1}^{t}q_{i,j}(\dot{\beta}_{i,t}))^{-1}(t^{-1}\sum_{j=1}^{t}h_{i,j})$,

$(R^{-1}\sum_{j=t-R+1}^{t}q_{i,j}(\dot{\beta}_{i,t}))^{-1}(R^{-1}\sum_{j=t-R+1}^{t}h_{i,j})$, 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.

The first assumption provides us with one primary piece of information. Analytically it tells us that the parameters must be estimated by OLS, NLLS, or maximum likelihood under normality assumptions. When a VAR is used, the system must be exactly identified. These restrictions are imposed to ensure that the statistics in Theorems 3.5-3.7 are pivotal. As in McCracken (1999b), 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 3.5-3.7 are functions of squared forecast errors. To achieve a pivotal statistic the parameters must then be estimated using mean square error as the loss function. Although this assumption restricts how the parameters are estimated, it does not otherwise restrict 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 random variable m_t such that

$$\max\{|g_{i,\beta,t}(\beta_i) - g_{i,\beta,t}|, |q_{i,t}(\beta_i) - q_{i,t})|\} \leq m_t |\beta_i - \beta_i^*|^\phi, \ Em_t < \infty \text{ and } 1 < \phi < \infty.$$

Most of Assumption 2 is imposed in order to ensure 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 it is twice continuously differentiable in the parameters.

Assumption 3: Let $U_t = [u_t, h_{2,t}', vec(h_{2,t}h_{2,t}' - Eh_{2,t}h_{2,t}')', vec(g_{2,\beta,t}g_{2,\beta,t}' - Eg_{2,\beta,t}g_{2,\beta,t}')', vec(u_t g_{2,\beta\beta,t})']'$.

(a) $EU_t = 0$, (b) U_t is uniformly L^8 bounded, (c) $Eu_t^2 = \sigma^2$, (d) For some $8 > d > 2$, U_t is strong

mixing with coefficients of size -8d/(8 - d), (e) $\lim_{T\to\infty} T^{-1}E\sum_{j=1}^{T} U_j U_j' < \infty$.

Assumption 4: (a) $Eh_{2,t}h_{2,t}' = \sigma^2 Eq_{2,t} \equiv \sigma^2 B_2^{-1}$, (b) $E(h_{2,t} | h_{2,t-j}, q_{2,t-j}, j = 1,2,...) = 0$.

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 ensure 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 Ghysels and Hall (1990), West (1996), and White (1999) 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 numbers of in-sample (R) and out-of-sample (P) observations also become arbitrarily large at the same rate. In this way we ensure that the parameters estimated in-sample and certain out-of-sample averages are both consistent estimators of their population level analogs.

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

3. Tests

While Ashley, Granger, and Schmalensee (1980) specifically advocate using tests of equal forecast accuracy to examine causality, given their definition of causality, any test designed to examine whether one variable carries information about another could reasonably be used. Accordingly, this paper considers the ability of full-sample 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

8

computational simplicity and performance in the non-nested investigations of Ericsson (1992), Diebold and Mariano (1995), Harvey, Leybourne, and Newbold (1997, 1998) and Clark (1999).

In the results below, the tests are applied to 1-step ahead forecasts. These findings should be widely useful because, in practice, most forecast comparisons include 1-step ahead results. Admittedly, many researchers are also interested in multi-step forecast results. We do not provide results for multi-step forecasts because the asymptotic distributions of the equal accuracy and encompassing tests appear to depend on the parameters of the data-generating process. For practical purposes, such dependence eliminates the possibility of using asymptotically pivotal approximations to test for equal accuracy or encompassing. Lutkepohl and Burda (1997) note similar difficulties associated with in-sample causality tests involving multi-step horizons. For those researchers interested in multi-step horizons, bootstrap procedures, such as those developed in Ashley (1998) and Kilian (1999), may yield accurate inferences.

3.1 Granger Causality (GC) Tests

In this paper we focus on testing ex-ante forecasts for equal accuracy and encompassing. However, for the sake of comparison we provide results for the commonly used full-sample F-test of Granger causality, which we refer to as the GC test. Letting $\hat{v}_{1,t} = y_t - g_{1,t}(\hat{\beta}_{1,R+P})$ and $\hat{v}_{2,t} = y_t - g_{2,t}(\hat{\beta}_{2,R+P})$ denote the residuals from two nested models estimated with the full sample of $R + P$ observations,

$$GC = \frac{R + P - k}{k_2} \cdot \frac{(R+P)^{-1}\sum_{t=1}^{R+P}\hat{v}_{1,t}^2 - (R+P)^{-1}\sum_{t=1}^{R+P}\hat{v}_{2,t}^2}{(R+P)^{-1}\sum_{t=1}^{R+P}\hat{v}_{2,t}^2}. \tag{1}$$

Under the null that the k_2 restrictions hold it is well known that, subject to certain conditions, the GC statistic has an exact $F(k_2, R + P - k)$ distribution. More generally, $k_2 \cdot GC$ converges in

9

distribution to a chi-square variate with k_2 degrees of freedom.[3]

We provide the formula in (1) for two key reasons. First, it helps motivate the tests of

equal MSE detailed below (equations (3), (4) and (6)). Second, it also helps motivate the tests of

encompassing considered below. To see this, decompose the numerator of (1) as

$$(R + P)^{-1} \sum_{t=1}^{R+P} \hat{v}_{1,t}^2 - (R + P)^{-1} \sum_{t=1}^{R+P} \hat{v}_{2,t}^2$$

$$= (R + P)^{-1} \sum_{t=1}^{R+P} (\hat{v}_{1,t}^2 - \hat{v}_{1,t} \hat{v}_{2,t}) - (R + P)^{-1} \sum_{t=1}^{R+P} (\hat{v}_{2,t}^2 - \hat{v}_{1,t} \hat{v}_{2,t}). \qquad (2)$$

If, for example, the two nested models are linear and estimated by OLS then the latter right-hand

side term in (2) is numerically zero and hence the numerator of (1) is identically

$(R + P)^{-1} \sum_{t=1}^{R+P} (\hat{v}_{1,t}^2 - \hat{v}_{1,t} \hat{v}_{2,t})$. This term is qualitatively similar to the orthogonality condition

used in the forecast encompassing tests (equations (7), (9) and (10)).[4]

3.2 The MSE-F Test

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

$$\text{MSE-F} = P \cdot \frac{P^{-1} \sum_t \hat{u}_{1,t+1}^2 - P^{-1} \sum_t \hat{u}_{2,t+1}^2}{P^{-1} \sum_t \hat{u}_{2,t+1}^2}. \qquad (3)$$

This statistic is comparable to the full-sample GC test in (1) 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 (1999b) shows that the

MSE-F statistic converges in distribution to a function of stochastic integrals of quadratics of

Brownian motion. Under the null, the limiting distribution, which varies with the forecasting

scheme, is a function of the limit of the ratio of post-sample to in-sample observations, π, and

---

[3] Comparing the statistic $k_2 \cdot GC$ against the chi-square distribution produces results very similar to those reported.

[4] Various discussions in the literatures on encompassing and artificial regression tests of non-nested hypotheses point out that encompassing tests are equivalent to F-type tests of exclusion restrictions. Davidson and MacKinnon (1993, pp. 386-87) summarize the basic point and relevant literature.

excess parameters, k_2, in model 2.

In the Monte Carlo experiments of section 4 the test statistic is compared against asymptotic critical values tabulated by McCracken (1999b). Since the models are nested, the null hypothesis is $Eu_{1,t+1}^2 \leq Eu_{2,t+1}^2$ and the alternative is $Eu_{1,t+1}^2 > Eu_{2,t+1}^2$. 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 MSE-T Test

Letting $d_{t+1} = \hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2$ and $\bar{d} = P^{-1}\sum_t d_{t+1} = MSE_1 - MSE_2$, Diebold and Mariano (1995) propose a t-statistic for equal MSE that, as calculated here, takes the form

$$MSE\text{-}T = (P-1)^{1/2}\frac{\bar{d}}{\sqrt{P^{-1}\sum_t(d_{t+1}-\bar{d})^2}} = (P-1)^{1/2}\frac{P^{-1}\sum_t(\hat{u}_{1,t+1}^2-\hat{u}_{2,t+1}^2)}{\sqrt{P^{-1}\sum_t(\hat{u}_{1,t+1}^2-\hat{u}_{2,t+1}^2)^2-\bar{d}^2}}. \quad (4)$$

Note that the term in front is $(P-1)^{1/2}$ rather than $P^{1/2}$ because, for computational convenience, we calculate the test using standard regression methods (we regress d_{t+1} on a constant) in which the estimated error variance incorporates a degrees-of-freedom adjustment.

While West (1996) proves that the MSE-T statistic can be asymptotically standard normal when applied to non-nested forecasts, the asymptotic distribution is non-normal when the forecasting models are nested under the null hypothesis. The root of the problem is that, under the null, $g_{1,t+1}(\beta_1^*) = g_{2,t+1}(\beta_2^*)$ and thus both $u_{1,t+1} = y_{t+1} - g_{1,t+1}(\beta_1^*) = u_{t+1}$ and $u_{2,t+1} = y_{t+1} - g_{2,t+1}(\beta_2^*) = y_{t+1} - g_{1,t+1}(\beta_1^*) = u_{t+1}$. Hence, at least heuristically, the squared loss differential d_{t+1} is exactly 0. McCracken (1999b) shows that, for forecasts from nested models, the MSE-T test statistic converges in distribution to a function of stochastic integrals of quadratics of Brownian motion. As was the case for the MSE-F statistic, the limiting distribution

11

depends on the forecasting scheme, π, and k_2.

In our Monte Carlo analysis, the MSE-T statistic is compared against the asymptotic critical values tabulated by McCracken (1999b). As with the MSE-F test, the alternative hypothesis is one-sided. To evaluate how using the standard, but asymptotically invalid, critical values would affect inference, results are also reported for a version of the test comparing the MSE-T statistic against the standard normal distribution.

3.4 The MSE-REG Test

Granger and Newbold (1977) also propose a test of equal MSE, referred to here as the MSE-REG statistic. It can be evaluated using the t-statistic associated with the coefficient α_1 from the OLS regression

$$\left(\hat{u}_{1,t+1} - \hat{u}_{2,t+1}\right) = \alpha_1 \left(\hat{u}_{1,t+1} + \hat{u}_{2,t+1}\right) + \text{error term},\tag{5}$$

which can be expressed as

$$\text{MSE-REG} = (P-1)^{1/2} \frac{\bar{d}}{\sqrt{P^{-1}\sum_t (\hat{u}_{1,t+1} - \hat{u}_{2,t+1})^2 P^{-1}\sum_t (\hat{u}_{1,t+1} + \hat{u}_{2,t+1})^2 - \bar{d}^2}}.\tag{6}$$

The covariance term in the numerator of (6) is equal to the difference in the MSEs for models 1 and 2. While West (1996) proves the MSE-REG statistic can be asymptotically standard normal when applied to non-nested forecasts, McCracken (1999b) shows that, for forecasts from nested models, the MSE-REG test statistic has the same limiting distribution as the MSE-T test.

In the Monte Carlo results of section 4, we compare the MSE-REG statistic against the asymptotic critical values tabulated by McCracken (1999b). As with the MSE-F and MSE-T tests, the alternative hypothesis is one-sided. To evaluate how using the standard, but asymptotically invalid, critical values would affect inference, results are also reported for a version of the test comparing the MSE-REG statistic against the standard normal distribution.

3.5 The ENC-T Test

Harvey, Leybourne, and Newbold (1998) develop a test of forecast encompassing based on the methodology of Diebold and Mariano (1995). Specifically, Harvey, Leybourne, and Newbold (1998) propose a test of encompassing that uses a t-statistic for the covariance between $\hat{u}_{1,t+1}$ and $\hat{u}_{1,t+1} - \hat{u}_{2,t+1}$. 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 $\bar{c} = P^{-1}\sum_t c_t$. Their encompassing test, denoted ENC-T, is formed as

$$\text{ENC-T} = (P-1)^{1/2} \frac{\bar{c}}{\sqrt{P^{-1}\sum_t (c_{t+1} - \bar{c})^2}} = (P-1)^{1/2} \frac{P^{-1}\sum_t (\hat{u}_{1,t+1}^2 - \hat{u}_{1,t+1}\hat{u}_{2,t+1})}{\sqrt{P^{-1}\sum_t (\hat{u}_{1,t+1}^2 - \hat{u}_{1,t+1}\hat{u}_{2,t+1})^2 - \bar{c}^2}} . \quad (7)$$

As with the MSE-T statistic, the term in front is $(P-1)^{1/2}$ rather than $P^{1/2}$ because we calculate the test using standard regression methods (we regress c_{t+1} on a constant). 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. Under the alternative that model 2 contains added information, the covariance should be positive. Hence the test is one-sided.

While West (1999) shows the ENC-T statistic can be 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 3.5.

Theorem 3.5: For ENC-T defined in (7), ENC-T $\rightarrow_d \chi_1 /(\chi_2)^{1/2}$ 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

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

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

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

According to Theorem 3.5, for each forecasting scheme the test statistic is pivotal. This permits the construction of estimates of asymptotically valid critical values without knowledge of the underlying data-generating process. With these critical values one can conduct an asymptotically valid test of the null. Morever, because the statistic is pivotal, the bootstrap procedures suggested by Ashley (1998) and Kilian (1999) may provide refinements to first-order asymptotics and thereby yield more accurate inference in finite samples.

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 because the vector Brownian Motion, $W(s)$, is ($k_2 \times 1$). The second parameter, π, also affects the null limiting distribution, in two ways. It directly affects the weights on each of the components of the statistics (recall that $\lambda = (1+\pi)^{-1}$). It also affects the range of integration on each of the stochastic integrals through λ.

We provide a selected set of asymptotic critical values for the ENC-T statistic in Appendix Tables 1-3.[5] These values were generated numerically using the limiting distribution in Theorem 3.5 and hence can be considered estimates of the true asymptotic critical values. The reported critical values are the 90th, 95th and 99th percentiles of 5000 independent draws from the distribution of $\chi_1/(\chi_2)^{1/2}$ for a given forecasting scheme and value of k_2 and π. Generating these

draws proceeded as follows. Weights that depend upon π are estimated in the obvious way using $\hat{\pi} \equiv P/R$. The necessary k_2 Brownian Motions are simulated as random walks each using an independent sequence of 10,000 i.i.d. $N(0,T^{-1/2})$ increments. The integrals are emulated by summing the relevant weighted quadratics of the random walks from the $R+1^{st}$ observation to the Tth observation. The random number generator is seeded so that all (k_2, π) pairs and all sampling schemes use the same 5000 draws of the k_2 random walks.

In the Monte Carlo results of section 4, we compare the ENC-T statistic against the asymptotic critical values tabulated in Appendix Tables 1-3, again using a one-sided alternative hypothesis. To evaluate how using the standard, but asymptotically invalid, critical values would affect inference, results are also reported for a version of the test comparing the ENC-T against the standard normal distribution.[6]

3.6 The ENC-REG Test

The forecast encompassing test proposed by Ericsson (1992) is a regression-based variant of the ENC-T test. The test statistic, denoted ENC-REG, is the t-statistic associated with the coefficient α_1 from the OLS regression

$$\hat{u}_{1,t+1} = \alpha_1 (\hat{u}_{1,t+1} - \hat{u}_{2,t+1}) + \text{error term,} \tag{8}$$

which can be expressed as

$$\text{ENC-REG} = (P-1)^{1/2} \frac{P^{-1}\sum_t \hat{u}_{1,t+1}(\hat{u}_{1,t+1} - \hat{u}_{2,t+1})}{\sqrt{P^{-1}\sum_t (\hat{u}_{1,t+1} - \hat{u}_{2,t+1})^2 P^{-1}\sum_t \hat{u}_{1,t+1}^2 - \bar{c}^2}}. \tag{9}$$

Under the null that model 1 forecast encompasses model 2, α_1 will be less than or equal to 0. Under the alternative that model 2 contains added information, α_1 should be positive.

---

[5] More detailed tables of critical values are available upon request.

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

Theorem 3.6: For ENC-REG defined in (9) and ENC-T defined in (7), ENC-REG - ENC-T = $o_p(1)$.

Theorem 3.6 states that ENC-REG and ENC-T are asymptotically equivalent under the null.[7] Hence we can use Appendix Tables 1-3 to construct asymptotically valid tests of forecast encompassing when the ENC-REG statistic is used. However, this does not imply that the two statistics will have similar finite sample properties. Accordingly, in the Monte Carlo experiments of section 4 we include both the ENC-T and ENC-REG statistics. As with the ENC-T test, in our Monte Carlo results we compare the ENC-REG test against the asymptotic critical values reported in Appendix Tables 1-3 and against critical values taken from the asymptotically invalid standard normal distribution.

3.7 A New Encompassing Test

Because the population prediction errors from models 1 and 2 are exactly the same under the null, the variances in the denominators of the ENC-T statistic (7) and the ENC-REG statistic (9) are, heuristically, 0. These denominators are estimates of the variance of c_{t+1} which, in population, is identically 0. This feature of the ENC-T and ENC-REG statistics may adversely affect the small-sample properties of the tests. Therefore, in parallel to the MSE-F test, this paper proposes a variant of the ENC-T and ENC-REG statistics in which $\bar{c}$ is scaled by the variance of one of the forecast errors rather than an estimate of the variance of $\bar{c}$.

---

[6] Incorporating the small-sample adjustments suggested by Harvey, Leybourne, and Newbold (1997, 1998) does not

This statistic, which we refer to as the ENC-NEW statistic, takes the form

$$\text{ENC-NEW} = P \cdot \frac{\bar{c}}{\text{MSE}_2} = P \cdot \frac{P^{-1} \sum_t (\hat{u}_{1,t+1}^2 - \hat{u}_{1,t+1}\hat{u}_{2,t+1})}{P^{-1} \sum_t \hat{u}_{2,t+1}^2}. \qquad (10)$$

The numerator is the object of interest in the ENC-NEW test (i.e., the covariance between

$\hat{u}_{1,t+1}$ and $\hat{u}_{1,t+1} - \hat{u}_{2,t+2}$). The denominator, MSE_2, serves as a scale correction. As with the

ENC-T and ENC-REG statistics, the limiting distribution is non-normal when the forecasts are

nested under the null. The actual limiting distribution is provided in Theorem 3.7.

Theorem 3.7: For ENC-NEW defined in (10) and χ_1 defined in Theorem 3.5, ENC-NEW $\to_d \chi_1$.

Given Theorem 3.5, this result is not surprising. The sole difference between the ENC-T

and ENC-NEW statistics is the denominator. Hence we expect their limiting distributions to be

somewhat related. As was the case for the ENC-T statistic, the limiting distribution is pivotal

and relies upon the forecasting scheme and the parameters k_2 and π.

In the Monte Carlo of section 4, the ENC-NEW statistic is compared against asymptotic

critical values tabulated in Appendix Tables 4-6. As with Appendix Tables 1-3, these values

were generated numerically using the limiting distribution in Theorem 3.7 and hence can be

considered estimates of the true asymptotic critical values.[8]

4. Monte Carlo Results

The small-sample properties of the tests described in section 3 are evaluated using a

bivariate VAR data-generating process. Specifically, we compare the predictive ability of an AR

---

alter the basic results for either the MSE-T or ENC-T tests.

[7] There is a parallel to this in McCracken (1999b). There it is shown that MSE-REG - MSE-T = $o_p(1)$.

[8] The random number generator was seeded so that the same χ_1 values were used in the construction of Appendix Tables 1-3 and 4-6.

model (model 1) with that from a VAR model (model 2). The presented results are based on

data generated using standard normal disturbances. The results are essentially unchanged when

the disturbances are drawn from the heavier-tailed t(6) distribution considered by Diebold and

Mariano (1995), Harvey, Leybourne, and Newbold (1997, 1998), and Clark (1999).

4.1 Experiment Design

In the presented results, data are generated using two different artificial VAR models.

The first, denoted VAR-1, takes the form

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} 0.3 & b \\ 0 & 0.5 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_{y,t} \\ u_{x,t} \end{pmatrix}. \tag{11}$$

The second, denoted VAR-2, takes the form

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} 0.3 & b \\ 0.7 & -0.5 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} 0.3 & 0 \\ 0.3 & 0 \end{pmatrix} \begin{pmatrix} y_{t-2} \\ x_{t-2} \end{pmatrix} + \begin{pmatrix} u_{y,t} \\ u_{x,t} \end{pmatrix}. \tag{12}$$

In both cases, y_t is the predictand, x_t is an auxiliary variable, and the disturbances are i.i.d.

standard normal random variates. To evaluate size in finite samples, the coefficient b is set at 0.

To evaluate power, b is set at 0.1 and 0.2. Simulations based on other VAR(2) models, the

trivariate stationary VAR(1) and VAR(3) models of Swanson, Ozyildirim, and Pisu (1996), and

the empirical inflation and unemployment model considered in section 5 produced results similar

to those from the VAR-1 and VAR-2 models in equations (11) and (12).[9]

In each Monte Carlo simulation we generate R + P + 4 observations. The additional four

observations allow for data-determined lag lengths in the forecasting models. Letting L denote

the lag length of the data-generating process, 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 standard normal distribution.

In the presented results, the lag length of the VAR model was chosen using Akaike's

information criterion; the same lag length was used for the AR model.[10] Only the in-sample

portion of the data was used to choose the lag length. Reserving observations 1 through 4 as

initial observations necessary to allow for a maximum of four lags in the estimated models, the

in-sample period spans observations 5 through R + 4. The estimated forecasting models are used

to form P 1-step ahead predictions, spanning observations R + 5 through R + P + 4. For brevity,

results are only presented for recursive forecasts, as the basic conclusions are essentially the

same for rolling and fixed forecasts.[11] Note that while the forecasting models use lags

determined from just the in-sample portion of the data, the GC test uses a lag length determined

from the full sample of R + P observations.

Results are reported for empirically relevant combinations of P and R such that $\hat{\pi} \equiv P/R$

takes the values 0.1, 0.2, 0.4, 0.6, and 1.0. Specifically, we use R = 100 with P = 10, 20, 40, 60,

and 100. We also use R = 200 with P = 20, 40, 80, 120, and 200.

4.2 Size Results

Table 1 presents the empirical sizes of Granger causality, equal forecast accuracy, and

---

[9] In addition, simulations based on a limited set of bivariate VARMA DGPs produced similar results. This is to be expected as long as the VARMAs can be reasonably well-approximated by a finite-order VAR.

[10] In computing power, 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 selected, 50,000 simulations were used in the size experiments to ensure the accuracy of the results with data-determined lags.

[11] While results for rolling forecasts are very similar to those for recursive forecasts, results for fixed forecasts do differ slightly. For example, the power of the MSE-T, MSE-REG, ENC-T, and ENC-REG tests is modestly lower for fixed forecasts than for recursive forecasts.

forecast encompassing tests for data from the VAR-1 and VAR-2 models of equations (11) and

(12), using a nominal size of 10%. The results are generally the same at a nominal size of 5%.

In these size experiments, the null is imposed by setting b = 0 in equations (11) and (12). Under

this null, the AR and VAR models have equal MSE and forecasts from the AR model encompass

those from the VAR. Three general results are evident from Table 1.

Size result 1. In most settings the post-sample tests have reasonable finite-sample size

properties, often suffering smaller size distortions than the full-sample F-test of causality.

More specifically, the MSE-F, MSE-REG, ENC-NEW, and ENC-REG tests perform

well, suffering only slight size distortions in finite samples. For example, when the DGP is

VAR-1, R = 100, and P = 20, these four tests have empirical sizes of 11.0%, 11.4%, 11.8%, and

11.9%, respectively. While the MSE-T and ENC-T statistics also perform reasonably well, when

P is small the tests suffer slightly greater distortions than do the MSE-REG and ENC-REG tests.

For instance, using the VAR-1 DGP, R = 100, and P = 10, the MSE-T test has an actual size of

14.8% while MSE-REG has an actual size of 12.6%. The better performance of MSE-REG and

ENC-REG likely stems from the regression forms of the tests using more precise variance

estimates. For example, the variance term in the denominator of the ENC-REG test (9) uses the

product of second moments, $P^{-1}\sum_t (\hat{u}_{1,t+1} - \hat{u}_{2,t+1})^2 P^{-1} \sum_t \hat{u}_{1,t+1}^2$, while the ENC-T test (7) uses the

sample fourth moment, $P^{-1}\sum_t (\hat{u}_{1,t+1}^2 - \hat{u}_{1,t+1}\hat{u}_{2,t+1})^2$. With normally distributed data, of course, in

population the fourth moment equals the product of the second moments.[12]

In general, given R, any size distortions of the post-sample tests fall as P rises. For

example, when data are generated using VAR-1 with R = 100 and P = 10, the actual sizes range

from 11.7% to 14.9%. When P increases to 100, the actual sizes range from 10% to 11%. The

improvement in size that comes with increases in P presumably stems from greater precision in the sample moments that enter the test statistics. Each of the sample moments is computed with P predictions.

By comparison, the GC test suffers a consistent size distortion that often exceeds (albeit slightly) that of the post-sample tests. For example, in the experiments using VAR-1 and R = 100, the GC test has empirical size of slightly more than 13% for all values of P. In contrast, the actual size of the ENC-NEW test ranges between 11.0% for P = 100 and 12.6% for P = 10. Similarly, in the experiments using VAR-2 and R = 200, the GC test consistently has size of about 13%, compared to roughly 11% or 12% in the ENC-NEW statistic. All of the size distortions in the GC test appear to stem from the pre-test bias associated with using the full sample of data to first determine the appropriate lag length and then test causality.[13] In unreported results, we find that if the GC test is calculated with the model lag length always set to the true order of the DGP (or at a longer lag length), the test is correctly sized.

In results not reported, some evidence suggests that the size advantage of post-sample tests may be larger than in Table 1 when more data mining is involved in choosing the lag length of the VAR. As noted above, in the Table 1 results, the lag length was set to minimize the AIC for the VAR. An alternative, more data-intensive approach to model selection is to allow the lags on y_t and x_t in the nesting equation for y_t (i.e., model 2) to differ, and then choose the lag combination that minimizes the AIC for that equation.[14] Using this approach to lag selection, when the DGP is VAR-1, R = 100, and P = 20, the GC test has actual size of 20.2%, while the

---

[12] We also find that MSE-REG and ENC-REG have better size than MSE-T and ENC-T in simulations with t(6)-distributed innovations.

[13] The same cannot be said about the size distortions of the post-sample tests. For these, results for simulations in which the true lag length is used are very similar to those reported (except for the experiments with VAR-2 and R = 100), in which the lag is data-determined.

MSE-F and ENC-NEW tests have size of 11.6% and 13.5%, respectively.

Size result 2. Comparing the MSE-T, MSE-REG, ENC-T, and ENC-REG tests against invalid asymptotic critical values generally leads to too-infrequent rejections.

While the ENC-T and ENC-REG tests are typically undersized when compared against standard normal critical values, the problem is more severe for the MSE-T and MSE-REG tests. For instance, using VAR-1 with $R = 100$ and $P = 20$, comparing the MSE-T and MSE-REG statistics against the standard normal distribution yields sizes of 5.8% and 4.7%, respectively. For a given R, using the wrong critical values typically causes the tests to become more undersized as P rises. When P increases to 60, the sizes of MSE-T and MSE-REG fall to 2.8% and 2.5%, respectively.

The MSE-T, MSE-REG, ENC-T, and ENC-REG tests are undersized when compared against standard normal critical values because the true asymptotic distributions of the statistics (and the empirical distributions) are shifted to the left relative to the standard normal. For a one-tailed test, the 10% critical value from the standard normal distribution is 1.282. In the case of the MSE-T and MSE-REG tests, for example, the 10% critical value from the true asymptotic distribution tabulated by McCracken (1999b) is 0.780 when $k_2 = 1$ and $\pi = 0.2$. For a given R, the undersizing becomes worse as P rises because the correct asymptotic distributions shift further to the left as π increases. With $k_2 = 1$, the 10% critical value from the true asymptotic distribution of the MSE-T and MSE-REG tests falls to 0.443 when π rises to 1.0.

Size result 3. When the length of the sample and certain features of the DGP combine to make data-based lag selection sufficiently imprecise, the size performance of all the tests deteriorates, more so for the post-sample tests than for the full-sample F-test of causality.

---

[14] Stock and Watson (1999), for example, take this approach to model selection. In our alternative simulations, we

In the case of the VAR-2 model, the true model for y_t is an AR(2). However, because the population correlation between x_{t-1} and y_{t-2} is large (0.57), in sufficiently small samples standard lag selection procedures cause a lag of 1 to be selected with some regularity.[15] When R = 100, the lag order is set at 1 in roughly 13% of the VAR-2 simulations; the lag order is correctly set at 2 in about 74% of the simulations. When R = 200, lag selection is much more precise. In this case, the lag order is set at 1 in only about 0.6% of the VAR-2 simulations. How often the lag of 1 is selected is crucial to the performance of the tests, because in an estimated model with just 1 lag, x_{t-1} will often spuriously appear to have predictive power for y_t.

Due to the difficulty in selecting lag length when the DGP is VAR-2 and R = 100, the size performance of each test deteriorates.[16] For example, when P = 20, the sizes of the MSE-F, ENC-NEW, and GC tests are 14.5%, 15.9%, and 16.2%, respectively. For the post-sample tests the size distortions generally do not fall as P rises. For instance, the size of the MSE-F test is 14.7% when P = 60. The size of the GC test does improve as P rises, because increases in P lead to greater precision in choosing the lag length of the model used to form the full sample-based GC test. In general, the deterioration in the performance of each test seems to be purely a function of the lag selection problem. In unreported results, we find that the sizes of the tests improve when the lag is fixed at the true order of 2 (or at a higher order).

This finding suggests that, contrary to the view some researchers may have, post-sample forecast tests are not necessarily a panacea for spurious in-sample or full-sample causality results. Many of the problems that lead to spurious in-sample results also seem likely to lead to

---

allow for lags between 1 and 4 of each variable.

[15] The problem is more severe if the BIC is used in lieu of the AIC.

[16] In addition to an overall deterioration in performance, the simulations based on VAR-2 and R = 100 produce a change in the performance of the ENC-T and ENC-REG tests compared against invalid critical values. While these encompassing tests are generally undersized if compared against the standard normal distribution, in the VAR-2 simulations with R = 100 the tests are slightly oversized.

spurious post-sample results. In the VAR-2 example, the basic problem is collinearity between x_{t-1} and y_{t-2}, which is a feature of both the in-sample and out-of-sample data. This collinearity can cause the in-sample-determined VAR lag to be incorrectly set at 1 with some frequency. When that occurs, both in-sample and post-sample tests will too frequently (relative to nominal size) find that x_{t-1} has predictive power for y_t.

4.3 Power Results

Tables 2 and 3 present results on the power of Granger causality, equal forecast accuracy, and forecast encompassing tests using the VAR-1 and VAR-2 DGPs in equations (11) and (12). In these power experiments, data are generated using $b = 0.1$ and $b = 0.2$, so the 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 some size distortions, the reported power figures are based on empirical critical values and therefore size-adjusted. The actual size of the tests is 10%; using 5% produces essentially the same results. For all of the tests, the null is rejected if the test statistic is greater than the 90th percentile of the statistic in the corresponding size experiment. Two general results are evident in Tables 2 and 3.[17]

Power result 1. The powers of the tests permit some simple rankings: ENC-NEW > MSE-F, ENC-T, ENC- REG > MSE-T, MSE-REG.

In our experiments, the ENC-NEW test is clearly the most powerful out-of-sample test of predictive ability. In some settings, the power of the ENC-NEW statistic rivals the power of the GC test, even though the GC test is based on many more observations (R + P rather than P). For example, as shown in the lower panel of Table 2, in simulations with VAR-2, $b = 0.1$, R = 100, and P = 40, the ENC-NEW test has power of 27.8%, little different from the GC test's power of

---

[17] As expected, it is also the case that power rises with the coefficient b defined in equations (11) and (12).

30.0%. The MSE-F, ENC-T, and ENC-REG tests are less powerful than the ENC-NEW test.

Using the experiment of the previous example, the MSE-F, ENC-T, and ENC-REG tests have

power of 24.2%, 23.1%, and 23.6%, respectively. The MSE-T and MSE-REG tests are less

powerful than the other tests.

There seem to be two factors underlying the power advantage of the ENC-NEW statistic.

First, as noted in section 3.7, the denominator of the statistic is simply MSE_2, rather than some

variance estimate that, heuristically, is 0 under the null. This feature seems to be the most likely

explanation of the ENC-NEW test's power advantage over the ENC-T and ENC-REG tests. The

greater power of the MSE-F test compared to the MSE-T and MSE-REG statistics lends further

support to this idea.

Second, encompassing tests like ENC-NEW are more powerful than their equal accuracy

counterparts, like MSE-F, because the equal accuracy statistics are essentially equal to an

encompassing test plus noise. Paralleling the decomposition of the GC test in equation (2), the

MSE-F statistic can be rewritten as

$$\text{MSE-F} = \text{P} \cdot \frac{\text{P}^{-1} \sum_t \hat{u}_{1,t+1}(\hat{u}_{1,t+1} - \hat{u}_{2,t+1})}{\text{P}^{-1} \sum_t \hat{u}_{2,t+1}^2} - \text{P} \frac{\text{P}^{-1} \sum_t \hat{u}_{2,t+1}(\hat{u}_{2,t+1} - \hat{u}_{1,t+1})}{\text{P}^{-1} \sum_t \hat{u}_{2,t+1}^2} \ .$$

The first term on the right-hand side of the above equation is simply the ENC-NEW statistic.

The second term is a test of the null that forecast 2 encompasses forecast 1 (in the full-sample

expression (2), by construction this second term is exactly 0). Because the models are nested,

forecast 2 encompasses forecast 1 under both the null and alternative hypotheses and,

heuristically, this second test statistic is 0. The MSE-F statistic has lower power because it is a

linear combination of the ENC-NEW test and a statistic testing a true hypothesis, which simply

adds noise to the test and thereby reduces its power relative to the ENC-NEW test.

Power result 2. Increasing the number of observations affects the powers of the tests

along several dimensions.

First, holding P fixed, the powers of the MSE-F and ENC-NEW tests rise with R, while the powers of the other tests are unaffected. For instance, as shown in the upper panel of Table 2, with the VAR-1 DGP and P = 40, the power of the ENC-NEW test rises from 31.2% when R = 100 to 39.7% when R = 200. Second, when R is held fixed, power rises with P. For example, in the VAR-1 experiment with R = 100 and b = 0.2, the power of the MSE-F test rises from 39.3% when P = 10 to 75.4% when P = 100. At the same time, the gap between the powers of (i) ENC-NEW and (ii) MSE-T, MSE-REG typically increases with P. Similarly, the difference in the powers of (i) ENC-T, ENC-REG and (ii) MSE-T, MSE-REG usually rises with P. These changes are mirrored by a reduction in the power difference between (i) MSE-F and (ii) MSE-T, MSE-REG and a reduction in the power difference between (i) ENC-T, ENC-REG and (ii) MSE-T, MSE-REG.

5. Empirical Example

In this section's example we use tests of Granger causality, equal forecast accuracy, and forecast encompassing to determine whether the prime-age male unemployment rate is useful in predicting core CPI inflation. Cecchetti (1995), Staiger, Stock, and Watson (1997), and Stock and Watson (1999) are recent examples of studies in the long literature on this basic question.

Our quarterly data, which begin in 1957:Q1, are divided into in-sample and out-of-sample portions so as to produce a $\hat{\pi} \equiv P/R$ value for which McCracken (1999b) and this paper report asymptotically valid critical values. After we allow for data differencing and a maximum of four data-determined lags, the in-sample period spans 1958:Q3-1987:Q1. This leaves 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, $\hat{\pi} = 0.4$.

26

Consistent with the results of augmented Dickey-Fuller tests for unit roots, our model

variables are the change in inflation and the change in the unemployment rate. Over the in-

sample period, the AIC for both the AR and the VAR is minimized at two lags. The sample test

statistics are compared against asymptotic critical values and empirical critical values generated

by Monte Carlo methods. The empirical critical values are based on 50,000 simulations of the

estimated inflation-unemployment model in which the null of no causality from unemployment

to inflation is imposed. The simulated model is constructed using the in-sample estimates of the

coefficients as the "true" values. Bootstrap methods produce similar critical values.

Table 4 presents results for the inflation-unemployment example. The upper panel

reports 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. In the in-sample model

estimates, unemployment clearly has predictive power for inflation. Moreover, the full-sample

GC test reported in the lower panel of the table strongly rejects the null of no causality from

unemployment to inflation.

Although weaker, the out-of-sample evidence indicates that unemployment has predictive

power for inflation.[18] As reported in the lower panel of Table 4, all of the encompassing tests

indicate that the change in unemployment has predictive content for the change in inflation. The

ENC-NEW test strongly rejects the null that the AR forecast encompasses the VAR forecast.

The ENC-T test clearly rejects, while the ENC-REG test marginally rejects. None of the tests for

equal MSE reject the null of equal accuracy.

Two factors may account for the difference in the in-sample and post-sample evidence.

One is simply power differences – some of the post-sample tests may not be powerful enough to

_____

[18] The forecasts are slightly biased. Demeaning the errors prior to calculating the test statistics actually strengthens
the evidence of unemployment's predictive power.

pick up unemployment's predictive content. The Monte Carlo results in section 4 indicate that the power of equal forecast accuracy tests such as MSE-F lag behind the power of encompassing counterparts like the ENC-NEW test, which has power rivaling that of the GC test. The second factor is model instabilities. Neither the AR model for inflation nor the VAR pass the supremum Wald or exponential Wald tests for stability developed in Andrews (1993) and Andrews and Ploberger (1994), respectively.[19]

6. Conclusions

In this paper we provide analytical, Monte Carlo, and empirical evidence on the properties of tests of equal forecast accuracy and encompassing applied to predictions from nested parametric models. We first derive the limiting distributions of two standard tests and one new test of forecast encompassing. Monte Carlo experiments are then used to provide evidence on the finite-sample size and power of equal accuracy and encompassing tests. These experiments yield four key results. First, the post-sample tests are, in general, reasonably well sized. Second, when compared against asymptotically invalid standard normal critical values the post-sample tests are undersized. Third, when lag selection is sufficiently imprecise, the post-sample tests suffer more substantial size distortions, with the implication that forecast tests are not necessarily a panacea for in-sample overfitting. Fourth, the encompassing test proposed in this paper (the ENC-NEW statistic defined in equation (10)) is most powerful. In the final part of our analysis, we find that the post-sample tests provide mixed evidence on the predictive content of unemployment for inflation. While each of the equal forecast accuracy tests fail to reject the null that unemployment has no predictive content for inflation, each of the

---

[19] The models do 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.

encompassing tests indicates that unemployment does have predictive power.

Although we extend the literature on out-of-sample tests of predictive ability there still remain a number of unanswered questions. Perhaps the most obvious is the optimal choice of the sample split parameter π. As reviewed by McCracken and West (1999), within the existing literature on out-of-sample hypothesis testing this seems to be an important determinant of both the size and power properties of tests of predictive ability.

Another important topic for future research is developing tests of equal forecast accuracy and encompassing that allow unit roots in the nested parametric models used to forecast. Corradi, Swanson and Olivetti (1999) show how tests of equal forecast accuracy can be constructed for predictions from non-nested parametric models in the presence of cointegrating relationships. Because tests for causality often, in practice, involve nonstationary variables (Stock and Watson, 1988) extending their results to a nested environment may prove useful.

A final topic for future research is developing tests of out-of-sample predictive ability for forecasts generated with nonparametric methods. Local-linear (Diebold and Nason, 1992), series-based (Swanson and White, 1997), and kernel-based (Chung and Zhou, 1996) nonparametric methods are frequently used to construct forecasts. Although McCracken (1999c) provides a limited set of results that are applicable to kernel-based forecasts, there do not exist a range of results that can be applied in other nonparametric environments.

References

Amano, R.A., and S. van Norden, 1995, Terms of trade and real exchange rates: The Canadian Evidence, Journal of International Money and Finance 14, 83-104.

Andrews, D.W.K., 1993, Tests for parameter instability and structural change with unknown change point, Econometrica 61, 821-56.

Andrews, D.W.K., and W. Ploberger, 1994, Optimal tests when a nuisance parameter is present only under the alternative, Econometrica 62, 1383-1414.

Ashley, R., 1998, A new technique for postsample model selection and validation, Journal of Economic Dynamics and Control 22, 647-665.

Ashley, R., C.W.J. Granger, and R. Schmalensee, 1980, Advertising and aggregate consumption: An analysis of causality, Econometrica 48, 1149-67.

Berkowitz, J. and L. Giorgianni, 1999, Long-horizon exchange rate predictability, Review of Economics and Statistics, Forthcoming.

Blomberg, S.B., and G.D. Hess, 1997, Politics and exchange rate forecasts, Journal of International Economics 43, 189-205.

Bram, J. and S. Ludvigson, 1998, Does consumer confidence forecast household expenditure? A sentiment horse race, Economic Policy Review: Federal Reserve Bank of New York June, 59-78.

Cecchetti, S.G., 1995, Inflation indicators and inflation policy, NBER Macroeconomics Annual, 189-219.

Chinn, M.D. and R.A. Meese, 1995, Banking on currency forecasts: How predictable is change in money?, Journal of International Economics 38, 161-178.

Chong, Y.Y. and D.F. Hendry, 1986, Econometric evaluation of linear macroeconomic models, Review of Economic Studies 53, 671-90.

Chung, Y.P. and Z.G. Zhou, 1996, The predictability of stock returns--A nonparametric approach, Econometric Reviews 15, 299-330.

Clark, T.E., 1999, Finite-sample properties of tests for equal forecast accuracy, Journal of Forecasting, Forthcoming.

Corradi, V., N.R. Swanson, and C. Olivetti, 1999, Predictive ability with cointegrated variables, manuscript, Texas A & M University.

Davidson, R., and J.G. MacKinnon, 1993, Estimation and inference in econometrics, Oxford University Press: New York.

Diebold, F.X. and C. Chen, 1996, Testing structural stability with endogenous breakpoint: A size comparison of analytic and bootstrap procedures, Journal of Econometrics 70, 221-41.

Diebold, F.X. and J. Nason, 1990, Nonparametric exchange rate prediction?, Journal of International Economics 28, 315-322.

Diebold, F.X., and R.S. Mariano, 1995, Comparing predictive accuracy, Journal of Business and Economic Statistics 13, 253-63.

Diebold, F.X. and G.D. Rudebusch, 1991, Forecasting output with the composite leading index: A real time analysis, Journal of the American Statistical Association 86, 603-610.

Ericsson, N.R., 1992, Parameter constancy, mean square forecast errors, and measuring forecast performance: An exposition, extensions, and illustration, Journal of Policy Modeling 14, 465-95.

Evans, M.D. and R.K. Lyons, 1999, Order flow and exchange rate dynamics, manuscript, Haas School of Business University of California - Berkeley.

Granger, C.W.J. and P. Newbold, 1977, Forecasting Economic Time Series, (Academic Press, Orlando, FL).

Hansen, B.E., 1992, Convergence to stochastic integrals for dependent heterogeneous processes, Econometric Theory 8, 489-500.

Harvey, D.I., S.J. Leybourne, and P. Newbold, 1997, Testing the equality of prediction mean squared errors, International Journal of Forecasting 13, 281-91.

Harvey, D.I., S.J. Leybourne, and P. Newbold, 1998, Tests for forecast encompassing, Journal of Business and Economic Statistics 16, 254-59.

Ghysels, E. and A. Hall, 1990, A test for structural stability of Euler conditions parameters estimated via the generalized method of moments estimator, International Economic Review 31, 355-64.

Kilian, L., 1999, Exchange rates and monetary fundamentals: What do we learn from Long-horizon regressions?, Journal of Applied Econometrics, Forthcoming.

Krueger, J.T. and K.N. Kuttner, 1996, The Fed funds futures rate as a predictor of Federal Reserve policy, Journal of Futures Markets 16, 865-79.

Lutkepohl, H. and M.M. Burda, 1997, Modified Wald tests under nonregular conditions, Journal of Econometrics 78, 315-332.

Mark, N.C., 1995, Exchange rates and fundamentals: Evidence on Long-horizon predictability, American Economic Review 85, 201-18.

McCracken, M.W., 1999a, Robust out of sample inference, manuscript, Louisiana State University.

McCracken, M.W., 1999b, Asymptotics for out-of-sample tests of causality, manuscript, Louisiana State University.

McCracken, M.W., 1999c, An out-of-sample, nonparametric test of the martingale difference hypothesis, manuscript, Louisiana State University.

McCracken, M.W. and K.D. West, 1999, Inference about predictive ability, manuscript, University of Wisconsin and Louisiana State University.

Meese, R.A. and K. Rogoff, 1983, Empirical exchange rate models of the seventies: Do they fit out of sample?, Journal of International Economics 14, 3-24.

Meese, R.A. and K. Rogoff, 1988, Was it real? The exchange rate-interest differential relation over the modern floating-rate period, Journal of Finance 43, 933-948.

Newey, W.K. and D. McFadden, 1994, Large sample estimation and hypothesis testing, in: R.F. Engle and D.L. McFadden, eds., Handbook of Econometrics, Vol. 4, (North-Holland, Amsterdam).

Nyblom, J., 1989, Testing for the constancy of parameters over time, Journal of the American Statistical Society 84, 223-30.

Staiger, D., J.H. Stock, and M.W. Watson, 1997, The NAIRU, unemployment and monetary policy, Journal of Economic Perspectives 11, 33-49.

Stock, J.H. and M.W. Watson, 1988, Interpreting the evidence on money-income causality, Journal of Econometrics 40, 161-182.

Stock, J.H. and M.W. Watson, 1999, Forecasting inflation, NBER Working Paper #7023.

Swanson, N.R., A. Ozyildirim and M. Pisu, 1996, A comparison of alternative causality and predictive accuracy tests in the presence of integrated and cointegrated economic variables, manuscript, Pennsylvania State University.

Swanson, N.R. and H. White, 1997, A model-selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks, The Review of Economics and Statistics 79, 265-275.

West, K.D., 1996, Asymptotic inference about predictive ability, Econometrica 64, 1067-84.

West, K.D., 1999, Tests for forecast encompassing when forecasts depend on estimated regression parameters, manuscript, University of Wisconsin.

West, K.D. and M.W. McCracken, 1998, Regression-based tests of predictive ability, International Economic Review 39, 817-40.

White, H., 1999, A reality check for data snooping, manuscript, University of California-San Diego and QuantMetrics R&D Associates, LLC.

Appendix 1

Lemmas A1 - A5 are abbreviated versions of Lemmas A1, 3.1, A9, A11 and A12 that appear

in McCracken (1999b). In order to facilitate reference, but also conserve space, these are

repeated below without proof. Lemmas A6 - A8 are new and hence their proofs are provided.

Throughout, the following notation will be used: J denotes the selection matrix $(I_{k_1 \times k_1}, 0_{k_1 \times k_2})$,

$\sup_t$ denotes $\sup_{R \leq t \leq T}$, for matrices A and C defined in Lemma A2 $\tilde{h}_{2,t+1}$ denotes

$s^{-1}A'CB_2^{1/2}h_{2,t+1}$ and $\tilde{H}_2(t)$ denotes $\sigma^{-1}A'CB_2^{1/2}H_2(t)$, $\nabla \dot{g}_{i,\beta,t+1}$ denotes $g'_{i,\beta,t+1}(\dot{\beta}_{i,t})(\hat{\beta}_{i,t} - \beta_i^*)$

for some $\dot{\beta}_{i,t}$ on the line between $\hat{\beta}_{i,t}$ and β_i^*. Note that, for simplicity, the P - 1 terms that

appear in the text formulas for ENC-T and ENC-REG are replaced by P in the theoretical results

below, without any consequence.

Lemma A1: For $a \in [0, 0.5)$ (a) $\sup_t t^a |\hat{\beta}_{i,t} - \beta_i^*| = o_p(1)$, (b) $\sup_t |T^{1/2}(vec[B_i(t)] - vec[B_i])|$

$= O_p(1)$.

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: $\sum_t (\hat{u}_{1,t+1} - \hat{u}_{2,t+1})^2 = \sigma^2 \sum_t \tilde{H}_2'(t)\tilde{H}_2(t) + o_p(1)$.

Lemma A4: $\sum_t \tilde{H}_2'(t)\tilde{h}_{2,t+1} \to_d \chi_1$ defined in Theorem 3.5 of the text.

Lemma A5: $\sum_t \tilde{H}_2'(t)\tilde{H}_2(t) \to_d \chi_2$ defined in Theorem 3.5 of the text..

Lemma A6: For $i = 1, 2$, $\sum_t u_{t+1} g_{i,\beta,t+1}'(\dot{\beta}_{i,t}) B_i(t) H_i(t) = \sum_t h_{i,t+1}' B_i H_i(t) + o_p(1)$.

Proof of Lemma A6: Add and subtract $g_{i,\beta,t+1}'$ and B_i to obtain

$$\sum_t u_{t+1} g_{i,\beta,t+1}'(\dot{\beta}_{i,t}) B_i(t) H_i(t) = \sum_t h_{i,t+1}' B_i H_i(t) \tag{13}$$

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

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

We must then show that the latter three right-hand side terms in (13) are $o_p(1)$. We do so for the recursive scheme; the arguments are similar for the other schemes. For the second right-hand side term in (13) note that

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

$$\le k^2 (\sup_t |u_{t+1}|) |B_i| (\sup_t T^{1/2} |H_i(t)|)(\sup_t T^{1/2} |g_{i,\beta,t+1}'(\dot{\beta}_{i,t}) - g_{i,\beta,t+1}'|).$$

That $\sup_t|u_{t+1}|$ is $O_p(1)$ follows from Assumption 3. To show that $\sup_t T^{1/2}|H_i(t)|$ is $O_p(1)$ note that $\sup_t T^{1/2}|H_i(t)| \le (T/R)(\sup_t |T^{-1/2} \sum_{s=1}^t h_{i,s}|)$. By Assumption 5 (T/R) is bounded; that $\sup_t |T^{-1/2} \sum_{s=1}^t h_{i,s}|$ is $O_p(1)$ follows from Theorem 3.1 of Hansen (1992). To see that $\sup_t T^{1/2} |g_{i,\beta,t+1}'(\dot{\beta}_{i,t}) - g_{i,\beta,t+1}'|$ is $o_p(1)$, note that by Assumption 2,

$$\sup_t T^{1/2} |g_{i,\beta,t+1}'(\dot{\beta}_{i,t}) - g_{i,\beta,t+1}'|$$

$$\le (\sup_t m_t)(\sup_t T^{1/2\varphi} |\dot{\beta}_{i,t} - \beta_i^*|)^\varphi \le (\sup_t m_t)(\sup_t T^{1/2\varphi} |\hat{\beta}_{i,t} - \beta_i^*|)^\varphi.$$

That $\sup_t m_t$ is $O_p(1)$ follows from Assumption 2. Since x^φ is a continuous function, it suffices to show that $\sup_t T^{1/2\varphi} |\hat{\beta}_{i,t} - \beta_i^*|$ is $o_p(1)$. Since $\varphi > 1$, the result follows from Lemma A1 (a).

To show that the third right-hand side term in (13) is $o_p(1)$ note that

$$\sum_t h_{i,t+1}'(B_i(t) - B_i) H_i(t)$$

35

$$= T^{-1/2} \sum_t (T/t) \text{vec}[T^{1/2} (B_i(t) - B_i)]'[T^{-1/2} h_{i,t+1} \otimes (T^{-1/2} \sum_{j=1}^t h_{i,j})].$$

Given Assumption 3 and Lemma A1 (b), Theorem 3.1 of Hansen (1992) implies that

$\sum_t (T/t) \text{vec}[T^{1/2} (B_i(t) - B_i)]'[T^{-1/2} h_{i,t+1} \otimes (T^{-1/2} \sum_{j=1}^t h_{i,j})]$ is $O_p(1)$. Since $T^{-1/2} = o(1)$ the desired

result follows.

To show that the fourth right-hand side term in (13) is $o_p(1)$ note that

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

$$\leq k^2 (\sup_t |u_{t+1}|)(\sup_t T^{1/2} |H_i(t)|)(\sup_t T^{1/2} |B_i(t) - B_i|)(\sup_t |g'_{i,\beta,t+1}(\dot{\beta}_{i,t}) - g'_{i,\beta,t+1}|).$$

That $\sup_t|u_{t+1}|$ and $\sup_t T^{1/2}|B_i(t) - B_i|$ are $O_p(1)$ follows from Assumption 3 and Lemma A1 (b).

That $\sup_t T^{1/2}|H_i(t)|$ is $O_p(1)$ and $\sup_t |g'_{i,\beta,t+1}(\dot{\beta}_{i,t}) - g'_{i,\beta,t+1}|$ is $o_p(1)$ were previously established.

Lemma A7: For $i,j = 1,2$, (a) $\sum_t u_{t+1}^2 H'_i(t)B_i(t)g_{i,\beta,t+1}(\dot{\beta}_{i,t})g'_{j,\beta,t+1}(\dot{\beta}_{j,t})B_j(t)H_j(t) =$

$\sum_t H'_i(t)B_i E(h_{i,t+1}h'_{j,t+1})B_j H_j(t) + o_p(1)$. (b) $\sum_t H'_i(t)B_i(t)g_{i,\beta,t+1}(\dot{\beta}_{i,t})g'_{j,\beta,t+1}(\dot{\beta}_{j,t})B_j(t)H_j(t) =$

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

Proof of Lemma A7: (a) We will show the result for the recursive scheme. Proofs for the

rolling and fixed schemes are similar. The proof is conducted in two stages. The first stage

consists of showing that

$$\sum_t u_{t+1}^2 H'_i(t)B_i(t)g_{i,\beta,t+1}(\dot{\beta}_{i,t})g'_{j,\beta,t+1}(\dot{\beta}_{j,t})B_j(t)H_j(t) = \sum_t H'_i(t)B_i h_{i,t+1}h'_{j,t+1}B_j H_j(t) + o_p(1).$$

In this proof only let $a_1 = B_i$, $a_2 = B_i(t) - B_i$, $b_1 = g_{i,\beta,t+1}$, $b_2 = g_{i,\beta,t+1} - g_{i,\beta,t+1}(\dot{\beta}_{i,t+1})$, $c_1 = g'_{j,\beta,t+1}$, c_2

$= g'_{j,\beta,t+1} - g'_{j,\beta,t+1}(\dot{\beta}_{j,t})$, $d_1 = B_j$, $d_2 = B_j(t) - B_j$. Using this notation, if we add and subtract $g_{i,\beta,t+1}$,

$g_{j,\beta,t+1}$, B_i and B_j we obtain the identity

$$\sum_t u_{t+1}^2 H'_i(t)B_i(t)g_{i,\beta,t+1}(\dot{\beta}_{i,t})g'_{j,\beta,t+1}(\dot{\beta}_{j,t})B_j(t)H_j(t) \qquad (14)$$

$$= \sum_{1 \le w, x, y, z \le 2} \sum_t u_{t+1}^2 H_i^{'}(t) a_w b_x c_y d_z H_j(t).$$

When $w = x = y = z = 1$ the argument of the outer summation takes the value

$\sum_t H_i^{'}(t) B_i h_{i,t+1} h_{j,t+1}^{'} B_j H_j(t)$. To obtain the result we must show that the remaining fifteen pieces

in (14) are each $o_p(1)$. The proof of each is very similar. Here we only show that the term

$\sum_t u_{t+1}^2 H_i^{'}(t) a_1 b_2 c_1 d_1 H_j(t) = \sum_t u_{t+1}^2 H_i^{'}(t) B_i (g_{i,\beta,t+1}(\dot\beta_{i,t}) - g_{i,\beta,t+1}) g_{j,\beta,t+1}^{'} B_j H_j(t)$ is $o_p(1)$. Taking

absolute values we immediately have

$$| \sum_t u_{t+1}^2 H_i^{'}(t) B_i (g_{i,\beta,t+1}(\dot\beta_{i,t}) - g_{i,\beta,t+1}) g_{j,\beta,t+1}^{'} B_j H_j(t) |$$

$$\le k^4 (\sup_t | u_{t+1}^2 g_{j,\beta,t+1}^{'} |) | B_i \| B_j | (\sup_t | T^{1/2} H_i(t) |) \times$$

$$(\sup_t | T^{1/2} H_j(t) |)(\sup_t | g_{i,\beta,t+1}(\dot\beta_{i,t}) - g_{i,\beta,t+1} |)$$

That $\sup_t | u_{t+1}^2 g_{j,\beta,t+1}^{'} |$ is $O_p(1)$ follows from Assumption 3. To show that $\sup_t T^{1/2}|H_i(t)|$ is $O_p(1)$

note that $\sup_t T^{1/2}|H_i(t)| \le (T/R)(\sup_t | T^{-1/2} \sum_{s=1}^t h_{i,s} |)$. By Assumption 5 (T/R) is bounded; that

$\sup_t | T^{-1/2} \sum_{s=1}^t h_{i,s} |$ is $O_p(1)$ follows from Theorem 3.1 of Hansen (1992). The argument is

identical for $\sup_t T^{1/2}|H_j(t)|$. It then suffices to show that $\sup_t | g_{i,\beta,t+1}^{'}(\dot\beta_{i,t}) - g_{i,\beta,t+1}^{'} |$ is $o_p(1)$. To

do so note that by Assumption 2,

$$\sup_t | g_{i,\beta,t+1}^{'}(\dot\beta_{i,t}) - g_{i,\beta,t+1}^{'} | \le (\sup_t m_t)(\sup_t | \dot\beta_{i,t} - \beta_i^* |)^\varphi \le (\sup_t m_t)(\sup_t | \hat\beta_{i,t} - \beta_i^* |)^\varphi.$$

That $\sup_t m_t$ is $O_p(1)$ follows from Assumption 2. Since x^φ is a continuous function, the result

follows from Lemma A1 (a).

The second stage of the proof consists of showing that

$$\sum_t H_i^{'}(t) B_i h_{i,t+1} h_{j,t+1}^{'} B_j H_j(t) = \sum_t H_i^{'}(t) B_i E h_{i,t+1} h_{j,t+1}^{'} B_j H_j(t) + o_p(1).$$

To do so add and subtract $E h_{i,t+1} h_{j,t+1}^{'}$ to obtain

$$\sum_t H_i^{'}(t)B_i h_{i,t+1} h_{j,t+1}^{'} B_j H_j(t) \tag{15}$$

$$= \sum_t H_i^{'}(t)B_i Eh_{i,t+1} h_{j,t+1}^{'} B_j H_j(t) \; + \; \sum_t H_i^{'}(t)B_i (h_{i,t+1} h_{j,t+1}^{'} - Eh_{i,t+1} h_{j,t+1}^{'})B_j H_j(t) \,.$$

It then suffices to show that the second right-hand side term in (15) is $o_p(1)$. Rearranging terms we obtain

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

$$= T^{-1/2} \sum_t (T/t)^2 [T^{-1/2} \sum_{s=1}^t h_{j,s}^{'} B_j \otimes T^{-1/2} \sum_{s=1}^t h_{i,s}^{'} B_i)] vec[T^{-1/2} (h_{i,t+1} h_{j,t+1}^{'} - Eh_{i,t+1} h_{j,t+1}^{'})]^{'}$$

That $\sum_t (T/t)^2 [T^{-1/2} \sum_{s=1}^t h_{j,s}^{'} B_j \otimes T^{-1/2} \sum_{s=1}^t h_{i,s}^{'} B_i)] vec[T^{-1/2} (h_{i,t+1} h_{j,t+1}^{'} - Eh_{i,t+1} h_{j,t+1}^{'})]^{'}$ is $O_p(1)$

follows from Assumption 3 and Theorem 3.1 of Hansen (1992). Since $T^{-1/2}$ is $o(1)$ the proof is complete.

(b) The proof of (b) follows essentially the same argument as that for (a) and hence is omitted to conserve space. Appendix 2 contains further detail.

Lemma A8: $\sum_t (\hat{u}_{1,t+1}^2 - \hat{u}_{1,t+1} \hat{u}_{2,t+1})^2 \rightarrow_d \sigma^4 \chi_2$.

Proof of Lemma A8: If we take first order Taylor expansions of both $\hat{u}_{1,t+1}$ and $\hat{u}_{2,t+1}$ around β_1^* and β_2^* respectively, we have

$$\sum_t (\hat{u}_{1,t+1}^2 - \hat{u}_{1,t+1} \hat{u}_{2,t+1})^2 \tag{16}$$

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

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

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

Consider the first bracketed right-hand side term in (16). If we note that $Eh_{1,t+1} h_{2,t+1}^{'} = \sigma^2 JB_2^{-1}$ and apply the definition of $\tilde{H}_2(t)$ then by Lemma A7 we obtain

$$[\sum_t u_{t+1}^2 \{\nabla \dot{g}_{1,t+1}\}^2 \ - \ 2\sum_t u_{t+1}^2 \{\nabla \dot{g}_{1,t+1}\}\{\nabla \dot{g}_{2,t+1}\} \ + \ \sum_t u_{t+1}^2 \{\nabla \dot{g}_{2,t+1}\}^2 \ ] \tag{17}$$

$$= \sigma^4 \sum_t \tilde{H}_2^{'}(t)\tilde{H}_2(t) \ + o_p(1).$$

Given Lemma A5 we now need only show that the second bracketed term on the right-hand side

of (16) is $o_p(1)$. To do so we show that $\sum_t u_{t+1}\{\nabla \dot{g}_{1,t+1}\}^3$ is $o_p(1)$. The remaining terms follow

similar arguments. Taking absolute values we obtain

$$| \sum_t u_{t+1}\{\nabla \dot{g}_{1,t+1}\}^3 \ | \ \leq \ k^3 (\sup_t | u_{t+1} |)(\sup_t T^{1/3} | \hat{\beta}_{1,t} - \beta_1^* |)^3 (\sup_t | g_{1,\beta,t+1}(\dot{\beta}_{1,t}) |)^3$$

That $\sup_t|u_{t+1}|$ is $O_p(1)$ follows from Assumption 3. That $(\sup_t T^{1/3} | \hat{\beta}_{1,t} - \beta_1^* |)^3$ is $o_p(1)$ follows

by continuity of the function x^3 and Lemma A1 (a). It then suffices to show that

$\sup_t | g_{1,\beta,t+1}(\dot{\beta}_{1,t}) |$ is $O_p(1)$. If we add and subtract $g_{1,\beta,t+1}$ we obtain

$$\sup_t | g_{1,\beta,t+1}(\dot{\beta}_{1,t}) | \ \leq \ \sup_t | g_{1,\beta,t+1}(\dot{\beta}_{1,t}) - g_{1,\beta,t+1} | \ + \ \sup_t | g_{1,\beta,t+1} |.$$

That $\sup_t | g_{1,\beta,t+1} |$ is $O_p(1)$ follows from Assumption 3. To show that $\sup_t | g_{1,\beta,t+1}(\dot{\beta}_{1,t}) - g_{1,\beta,t+1} |$

is $o_p(1)$ note that by Assumption 2,

$$\sup_t | g_{i,\beta,t+1}^{'}(\dot{\beta}_{i,t}) - g_{i,\beta,t+1}^{'} | \ \leq \ (\sup_t m_t)(\sup_t | \dot{\beta}_{i,t} - \beta_i^* |)^{\varphi} \ \leq \ (\sup_t m_t)(\sup_t | \hat{\beta}_{i,t} - \beta_i^* |)^{\varphi}.$$

That $\sup_t m_t$ is $O_p(1)$ follows from Assumption 2. Since x^{φ} is a continuous function, the result

follows from Lemma A1 (a).

Proof of Theorem 3.5: Given Theorem 3.7 and the Continuous Mapping Theorem it suffices to

show that $\sum_t (\hat{u}_{1,t+1}^2 - \hat{u}_{1,t+1}\hat{u}_{2,t+1})^2 \ - P\bar{c}^2 \rightarrow_d \sigma^4 \chi_2$. By Theorem 3.7 $\bar{c}$ is $O_p(P^{-1})$ and hence $P\bar{c}^2$

$= o_p(1)$. The result then follows from Lemma A8.

Proof of Theorem 3.6: Define $a_{0,T}$, $a_{1,T}$, $a_{2,T}$ and $a_{3,T}$ as $P^{-1}\sum_t \hat{u}_{1,t+1}^2 - \hat{u}_{1,t+1}\hat{u}_{2,t+1}$,

$P^{-1}\sum_t (\hat{u}_{1,t+1} - \hat{u}_{2,t+1})^2$, $P^{-1}\sum_t \hat{u}_{2,t+1}^2$, and $P^{-1}\sum_t (\hat{u}_{1,t+1}^2 - \hat{u}_{1,t+1}\hat{u}_{2,t+1})^2 - \bar{c}^2$ respectively. By

Theorems 3.5 and 3.7 we know both $Pa_{0,T}$ and $Pa_{0,T}/[Pa_{3,T}]^{1/2}$ are $O_p(1)$. By the proof of Lemma

A8 and Theorem 3.5 we know that from (17), $Pa_{3,T} = \sigma^4 \sum_t \tilde{H}_2'(t)\tilde{H}_2(t) + o_p(1)$. Given the

Continuous Mapping Theorem it then suffices to show that $P(a_{1,T}a_{2,T} - a_{0,T}^2) = \sigma^4 \sum_t \tilde{H}_2'(t)\tilde{H}_2(t)$

$+ o_p(1)$. That $a_{2,T} \to_p \sigma^2$ follows from Theorem 4.1 of West (1996). That $Pa_{0,T}^2 = o_p(1)$ follows

from Theorem 3.7. The result follows since by Lemma A3, $Pa_{1,T} = \sigma^2 \sum_t \tilde{H}_2'(t)\tilde{H}_2(t) + o_p(1)$.

Proof of Theorem 3.7: If we take first order Taylor expansions of $\hat{u}_{1,t+1}$ and $\hat{u}_{2,t+1}$ around β_1^*

and β_2^* respectively, we obtain

$$\sum_t \hat{u}_{1,t+1}^2 - \hat{u}_{1,t+1}\hat{u}_{2,t+1} \tag{18}$$

$$= \sum_t \{-u_{t+1}g_{1,\beta,t+1}'(\dot{\beta}_{1,t})B_1(t)H_1(t) + u_{t+1}g_{2,\beta,t+1}'(\dot{\beta}_{2,t})B_2(t)H_2(t)\}$$

$$- \sum_t \{-H_1'(t)B_1(t)g_{1,\beta,t+1}(\dot{\beta}_{1,t})g_{1,\beta,t+1}'(\dot{\beta}_{1,t})B_1(t)H_1(t)$$

$$+ H_1'(t)B_1(t)g_{1,\beta,t+1}(\dot{\beta}_{1,t})g_{2,\beta,t+1}'(\dot{\beta}_{2,t})B_2(t)H_2(t)\}$$

for $\dot{\beta}_{i,t}$ on the line between $\hat{\beta}_{i,t}$ and β_i^* respectively. Consider the first bracketed right-hand

side term in (18). If we note that both $h_{1,t+1} = Jh_{2,t+1}$ and $H_1(t) = JH_2(t)$ and apply the definition of

both $\tilde{h}_{2,t+1}$ and $\tilde{H}_2(t)$ then by Lemma A6 we obtain

$$\sum_t \{-u_{t+1}g_{1,\beta,t+1}'(\dot{\beta}_{1,t})B_1(t)H_1(t) + u_{t+1}g_{2,\beta,t+1}'(\dot{\beta}_{2,t})B_2(t)H_2(t)\}$$

$$= \sigma^2 \sum_t \tilde{H}_2'(t)\tilde{h}_2(t) + o_p(1).$$

Given Lemma A4 we now need only show that the second bracketed right-hand side term in (18)

is $o_p(1)$. If we note that $Eg_{1,\beta,t+1}'g_{2,\beta,t+1} = JEg_{2,\beta,t+1}'g_{2,\beta,t+1} = JB_2^{-1}$ the result follows by Lemma A7.

Appendix 2

This appendix contains a proof of Lemma A7 (b).

Proof of Lemma A7 (b): We will show the result for the recursive scheme. Proofs for the

rolling and fixed schemes are similar. The proof is conducted in two stages. The first stage

consists of showing that

$$\sum_t H_i^{'}(t)B_i(t)g_{i,\beta,t+1}(\dot\beta_{i,t})g_{j,\beta,t+1}^{'}(\dot\beta_{j,t})B_j(t)H_j(t) = \sum_t H_i^{'}(t)B_i g_{i,\beta,t+1}g_{j,\beta,t+1}^{'}B_j H_j(t) + o_p(1).$$

In this proof only let $a_1 = B_i$, $a_2 = B_i(t) - B_i$, $b_1 = g_{i,\beta,t+1}$, $b_2 = g_{i,\beta,t+1} - g_{i,\beta,t+1}(\dot\beta_{i,t+1})$, $c_1 = g_{j,\beta,t+1}^{'}$, c_2

$= g_{j,\beta,t+1}^{'} - g_{j,\beta,t+1}^{'}(\dot\beta_{j,t})$, $d_1 = B_j$, $d_2 = B_j(t) - B_j$. Using this notation, if we add and subtract $g_{i,\beta,t+1}$,

$g_{j,\beta,t+1}$, B_i and B_j we obtain the identity

$$\sum_t H_i^{'}(t)B_i(t)g_{i,\beta,t+1}(\dot\beta_{i,t})g_{j,\beta,t+1}^{'}(\dot\beta_{j,t})B_j(t)H_j(t) \qquad\qquad (AA1)$$

$$= \sum_{1\le w,x,y,z\le 2} \sum_t H_i^{'}(t)a_w b_x c_y d_z H_j(t).$$

When $w = x = y = z = 1$ the argument of the outer summation takes the value

$\sum_t H_i^{'}(t)B_i g_{i,\beta,t+1}g_{j,\beta,t+1}^{'}B_j H_j(t)$. To obtain the result we must show that the remaining fifteen

terms in (AA1) are each $o_p(1)$. The proof of each is very similar. Here we show that the term

$\sum_t H_i^{'}(t)a_1 b_2 c_1 d_1 H_j(t) = \sum_t H_i^{'}(t)B_i(g_{i,\beta,t+1}(\dot\beta_{i,t}) - g_{i,\beta,t+1})g_{j,\beta,t+1}^{'}B_j H_j(t)$ is $o_p(1)$. Taking absolute

values we immediately have

$$| \sum_t H_i^{'}(t)B_i(g_{i,\beta,t+1}(\dot\beta_{i,t}) - g_{i,\beta,t+1})g_{j,\beta,t+1}^{'}B_j H_j(t) |$$

$$\le k^4 (\sup_t | g_{j,\beta,t+1}^{'} |) | B_i \| B_j | (\sup_t | T^{1/2}H_i(t) |) \times$$

$$(\sup_t | T^{1/2}H_j(t) |)(\sup_t | g_{i,\beta,t+1}(\dot\beta_{i,t}) - g_{i,\beta,t+1} |)$$

That $\sup_t |g'_{j,\beta,t+1}|$ is $O_p(1)$ follows from Assumption 3. That $\sup_t T^{1/2}|H_i(t)|$ and $\sup_t T^{1/2}|H_j(t)|$ are

$O_p(1)$ was established in the proof of part (a). The result follows since

$\sup_t |g'_{i,\beta,t+1}(\dot\beta_{i,t}) - g'_{i,\beta,t+1}|$ was shown to be $o_p(1)$ in the proof of part (a).

The second stage of the proof consists of showing that

$$\sum_t H'_i(t)B_i g_{i,\beta,t+1}g'_{j,\beta,t+1}B_j H_j(t) = \sum_t H'_i(t)B_i Eg_{i,\beta,t+1}g'_{j,\beta,t+1}B_j H_j(t) + o_p(1).$$

To do so add and subtract $Eg_{i,\beta,t+1}g'_{j,\beta,t+1}$ to obtain

$$\sum_t H'_i(t)B_i g_{i,\beta,t+1}g'_{j,\beta,t+1}B_j H_j(t) \qquad\qquad (AA2)$$

$$= \sum_t H'_i(t)B_i Eg_{i,\beta,t+1}g'_{j,\beta,t+1}B_j H_j(t) + \sum_t H'_i(t)B_i(g_{i,\beta,t+1}g'_{j,\beta,t+1} - Eg_{i,\beta,t+1}g'_{j,\beta,t+1})B_j H_j(t).$$

It then suffices to show that the second r.h.s. term in (AA2) is $o_p(1)$. Rearranging terms we

obtain

$$\sum_t H'_i(t)B_i(g_{i,\beta,t+1}g'_{j,\beta,t+1} - Eg_{i,\beta,t+1}g'_{j,\beta,t+1})B_j H_j(t)$$

$$= T^{-1/2}\sum_t (T/t)^2[T^{-1/2}\sum_{s=1}^t h'_{j,s}B_j \otimes T^{-1/2}\sum_{s=1}^t h'_{i,s}B_i)]vec[T^{-1/2}(g_{i,\beta,t+1}g'_{j,\beta,t+1} - Eg_{i,\beta,t+1}g'_{j,\beta,t+1})]'$$

That $\sum_t (T/t)^2[T^{-1/2}\sum_{s=1}^t h'_{j,s}B_j \otimes T^{-1/2}\sum_{s=1}^t h'_{i,s}B_i)]vec[T^{-1/2}(g_{i,\beta,t+1}g'_{j,\beta,t+1} - Eg_{i,\beta,t+1}g'_{j,\beta,t+1})]'$ is $O_p(1)$

follows from Assumption 3 and Theorem 3.1 of Hansen (1992). Since $T^{-1/2}$ is o(1) the proof is

complete.

<table>
<tr><td colspan="11" align="center">**Table 1**
Empirical Size
Recursive Forecasts
Nominal Size = 10%</td></tr>
</table>

	$R = 100$					$R = 200$				
	$P = 10$	$P = 20$	$P = 40$	$P = 60$	$P = 100$	$P = 20$	$P = 40$	$P = 80$	$P = 120$	$P = 200$
VAR-1										
Tests Compared Against Valid Asymptotic Critical Values										
MSE-F	.117	.110	.102	.103	.103	.110	.107	.097	.103	.100
MSE-T	.149	.128	.116	.105	.102	.130	.118	.107	.101	.101
MSE-REG	.126	.114	.109	.099	.100	.116	.109	.102	.098	.099
ENC-NEW	.126	.118	.110	.111	.110	.118	.113	.105	.111	.106
ENC-T	.148	.134	.120	.111	.110	.129	.123	.108	.108	.105
ENC-REG	.127	.119	.108	.104	.105	.115	.113	.102	.104	.102
GC	.135	.134	.133	.133	.134	.133	.129	.129	.133	.128
Tests Compared Against Invalid Standard Normal Critical Values										
MSE-T	.085	.058	.039	.028	.018	.071	.050	.033	.026	.017
MSE-REG	.069	.047	.033	.025	.017	.060	.044	.030	.024	.015
ENC-T	.110	.091	.079	.073	.067	.094	.081	.070	.069	.062
ENC-REG	.093	.078	.071	.067	.062	.083	.073	.065	.065	.060
VAR-2										
Tests Compared Against Valid Asymptotic Critical Values										
MSE-F	.140	.145	.144	.147	.159	.111	.112	.107	.097	.103
MSE-T	.151	.144	.143	.141	.153	.123	.114	.109	.098	.102
MSE-REG	.130	.130	.136	.136	.151	.109	.106	.106	.096	.101
ENC-NEW	.158	.159	.162	.169	.181	.119	.118	.110	.107	.112
ENC-T	.154	.157	.157	.159	.172	.123	.120	.113	.103	.107
ENC-REG	.135	.142	.146	.151	.168	.110	.111	.107	.100	.105
GC	.174	.162	.153	.144	.132	.128	.130	.127	.127	.127
Tests Compared Against Invalid Standard Normal Critical Values										
MSE-T	.090	.066	.048	.042	.040	.068	.048	.030	.021	.014
MSE-REG	.075	.056	.043	.038	.037	.061	.042	.026	.019	.012
ENC-T	.128	.119	.117	.122	.131	.103	.090	.081	.076	.075
ENC-REG	.113	.106	.108	.115	.126	.092	.082	.077	.074	.072

Notes:
1. The VAR-1 and VAR-2 models are, respectively,

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

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} .3 & 0 \\ .7 & -.5 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} .3 & 0 \\ .3 & 0 \end{pmatrix} \begin{pmatrix} y_{t-2} \\ x_{t-2} \end{pmatrix} + \begin{pmatrix} u_{y,t} \\ u_{x,t} \end{pmatrix},$$

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 the order minimizing the AIC for the VAR.
3. R and P refer to the number of in–sample observations and post–sample predictions, respectively.
4. Section 3 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 3. The valid asymptotic critical values are taken from the appendix tables in this paper and in McCracken (1999b). 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>
<tr><th colspan="11">Table 2
Size–Adjusted Power, $b = .1$
Recursive Forecasts
(Empirical Size = 10%)</th></tr>
</table>

	$R = 100$					$R = 200$				
	$P = 10$	$P = 20$	$P = 40$	$P = 60$	$P = 100$	$P = 20$	$P = 40$	$P = 80$	$P = 120$	$P = 200$
VAR-1										
MSE-F	.202	.229	.268	.303	.365	.289	.329	.402	.462	.569
MSE-T	.139	.174	.216	.252	.328	.177	.227	.313	.375	.506
MSE-REG	.146	.181	.220	.253	.329	.183	.234	.318	.378	.507
ENC-NEW	.221	.255	.312	.353	.435	.331	.397	.493	.561	.678
ENC-T	.150	.198	.252	.301	.390	.208	.278	.396	.475	.621
ENC-REG	.163	.203	.258	.304	.393	.218	.287	.401	.480	.623
GC	.306	.322	.361	.393	.457	.483	.518	.571	.622	.711
VAR-2										
MSE-F	.178	.204	.242	.271	.333	.235	.275	.328	.385	.474
MSE-T	.136	.157	.197	.227	.287	.161	.206	.269	.332	.436
MSE-REG	.140	.164	.201	.229	.290	.169	.209	.271	.333	.436
ENC-NEW	.196	.232	.278	.316	.384	.275	.335	.408	.480	.590
ENC-T	.150	.176	.231	.271	.349	.185	.245	.331	.413	.535
ENC-REG	.156	.188	.236	.274	.351	.194	.251	.334	.414	.537
GC	.254	.276	.300	.331	.385	.416	.451	.479	.542	.631

Notes:
1. The VAR-1 and VAR-2 models are, respectively,

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

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} .3 & b \\ .7 & -.5 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} .3 & 0 \\ .3 & 0 \end{pmatrix} \begin{pmatrix} y_{t-2} \\ x_{t-2} \end{pmatrix} + \begin{pmatrix} u_{y,t} \\ u_{x,t} \end{pmatrix},$$

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 the order minimizing the AIC for the VAR.
3. R and P refer to the number of in–sample observations and post–sample predictions, respectively.
4. Section 3 in the text defines the test statistics. In each experiment, power is calculated by comparing the test statistics against empirical critical values, calculated as the 90th percentile of the distributions of the statistics in the corresponding size experiment (in which the DGP, R, and P are the same as in the power experiment, except $b = 0$). Because empirical critical values are used, in contrast to Table 1 results are not reported for tests compared against invalid standard normal critical values.
5. The number of simulations is 10,000.

	R = 100					R = 200				
	$P = 10$	$P = 20$	$P = 40$	$P = 60$	$P = 100$	$P = 20$	$P = 40$	$P = 80$	$P = 120$	$P = 200$
VAR-1										
MSE-F	.393	.463	.560	.640	.754	.535	.632	.753	.831	.920
MSE-T	.207	.288	.413	.513	.674	.273	.400	.588	.702	.863
MSE-REG	.228	.303	.425	.522	.678	.293	.413	.594	.707	.864
ENC-NEW	.467	.571	.700	.792	.891	.672	.799	.911	.960	.991
ENC-T	.257	.377	.555	.678	.832	.384	.583	.795	.899	.977
ENC-REG	.293	.403	.572	.690	.837	.416	.602	.806	.904	.979
GC	.711	.742	.811	.853	.911	.934	.952	.970	.981	.995
VAR-2										
MSE-F	.343	.419	.513	.594	.719	.468	.569	.685	.781	.894
MSE-T	.196	.277	.389	.486	.644	.261	.380	.549	.683	.852
MSE-REG	.211	.288	.397	.493	.648	.273	.390	.555	.688	.852
ENC-NEW	.408	.519	.641	.735	.849	.604	.746	.868	.936	.985
ENC-T	.244	.359	.513	.632	.793	.358	.537	.745	.867	.964
ENC-REG	.270	.380	.529	.646	.801	.379	.558	.755	.873	.965
GC	.640	.679	.743	.793	.868	.898	.926	.951	.971	.991

Table 3
Size–Adjusted Power, $b = .2$
Recursive Forecasts
(Empirical Size = 10%)

Notes:
1. The VAR-1 and VAR-2 models are, respectively,

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

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} .3 & b \\ .7 & -.5 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} .3 & 0 \\ .3 & 0 \end{pmatrix} \begin{pmatrix} y_{t-2} \\ x_{t-2} \end{pmatrix} + \begin{pmatrix} u_{y,t} \\ u_{x,t} \end{pmatrix},$$

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 the order minimizing the AIC for the VAR.
3. R and P refer to the number of in–sample observations and post–sample predictions, respectively.
4. Section 3 in the text defines the test statistics. In each experiment, power is calculated by comparing the test statistics against empirical critical values, calculated as the 90th percentile of the distributions of the statistics in the corresponding size experiment (in which the DGP, R, and P are the same as in the power experiment, except $b = 0$). Because empirical critical values are used, in contrast to Table 1 results are not reported for tests compared against invalid standard normal critical values.
5. The number of simulations is 10,000.

Table 4			
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	*Asymptotic*	*Empirical*
	statistics	*critical values*	*critical values*
MSE, AR	.420		
MSE, VAR	.412		
MSE-F	.839	1.029	1.115
MSE-T	.099	.614	.700
MSE-REG	.137	.614	.664
ENC-NEW	5.186	1.019	1.093
ENC-T	1.112	1.086	1.179
ENC-REG	1.698	1.086	1.144
GC	8.107	2.337	2.376

Notes:
1. The figures in parentheses in the upper panel of the table are standard errors for the reported coefficient estimates.
2. 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.
3. R and P refer to the number of in–sample observations and post–sample predictions, respectively.
4. The significance level of the tests is 10%.
5. Section 3 in the text defines the test statistics. The asymptotic critical values are taken from the appendix tables in this paper and in McCracken (1999b).
6. The empirical critical values are generated from a Monte Carlo experiment (using 50,000 simulations) in which the DGP is a VAR in the changes in inflation and unemployment imposing the null that unemployment not enter the inflation equation. The equations of the simulated model, estimated with just in–sample data, are given in columns 2 and 4 of the top panel. The covariance matrix of the residuals in the DGP is

$$\text{Var} \begin{pmatrix} u_{infl,t} \\ u_{unemp,t} \end{pmatrix} = \begin{pmatrix} 2.673 & -.081 \\ -.081 & .102 \end{pmatrix}.$$

Percentiles of the ENC-T and ENC-REG Statistics: Recursive Scheme

k_2	%-ile	.1	.2	.4	.6	.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	.995	.968	.955	.942	.939	.959	.956	.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

(The $\pi =$ header spans the .1 through 2.0 columns.)

Notes:
1. Appendix Table 1 reports estimates of the 90th, 95th and 99th percentiles of the asymptotic distribution of both the ENC-T and ENC-REG statistics when the recursive scheme is used.
2. The estimates were constructed based upon 5,000 simulated draws from the relevant distribution for a given value of both k_2 and π. See section 3.5 of the text for further detail on how the simulations were conducted.

Appendix Table 2

Percentiles of the ENC-T and ENC-REG Statistics: Rolling Scheme

k_2	%-ile						$\pi =$					
		.1	.2	.4	.6	.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	.999	.951	.949	.948	.964	.953	.965	.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

Notes:
1. Appendix Table 2 reports estimates of the 90th, 95th and 99th percentiles of the asymptotic distribution of both the ENC-T and ENC-REG statistics when the rolling scheme is used.
2. The estimates were constructed based upon 5,000 simulated draws from the relevant distribution for a given value of both k_2 and π. See section 3.5 of the text for further detail on how the simulations were conducted.

Appendix Table 3

Percentiles of the ENC-T and ENC-REG Statistics: Fixed Scheme

k_2	%-ile						$\pi =$					
		.1	.2	.4	.6	.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

Notes:
1. Appendix Table 3 reports estimates of the 90th, 95th and 99th percentiles of the asymptotic distribution of both the ENC-T and ENC-REG statistics when the fixed scheme is used.
2. The estimates were constructed based upon 5,000 simulated draws from the relevant distribution for a given value of both k_2 and π. See section 3.5 of the text for further detail on how the simulations were conducted.

Percentiles of the ENC-NEW Statistic: Recursive Scheme

k_2	%-ile						$\pi =$					
		.1	.2	.4	.6	.8	1.0	1.2	1.4	1.6	1.8	2.0
1	0.99	.954	1.397	2.098	2.662	2.904	3.209	3.436	3.611	3.722	3.920	4.134
	0.95	.520	.744	1.079	1.312	1.467	1.584	1.724	1.845	1.967	2.084	2.085
	0.90	.335	.473	.685	.791	.925	.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	.766	1.028	1.481	1.779	2.016	2.234	2.510	2.580	2.663	2.785	2.889
	0.90	.524	.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	.940	1.273	1.865	2.256	2.442	2.709	2.909	3.065	3.250	3.463	3.564
	0.90	.686	.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	.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	.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	.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

Notes:
1. Appendix Table 4 reports estimates of the 90th, 95th and 99th percentiles of the asymptotic distribution of the ENC-NEW statistic when the recursive scheme is used.
2. The estimates were constructed based upon 5,000 simulated draws from the relevant distribution for a given value of both k_2 and π. See section 3.5 of the text for further detail on how the simulations were conducted.

Appendix Table 5
Percentiles of the ENC-NEW Statistic: Rolling Scheme

k_2	%-ile	$\pi =$ .1	.2	.4	.6	.8	1.0	1.2	1.4	1.6	1.8	2.0
1	0.99	.997	1.466	2.278	2.854	3.264	3.676	3.852	4.410	4.614	4.835	5.064
	0.95	.534	.788	1.161	1.506	1.722	1.946	2.198	2.374	2.498	2.683	2.836
	0.90	.339	.502	.764	.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	.793	1.089	1.640	2.031	2.380	2.704	2.949	3.212	3.366	3.593	3.723
	0.90	.539	.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	.962	1.391	1.955	2.500	2.847	3.271	3.549	3.759	4.108	4.219	4.435
	0.90	.695	.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	.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	.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	.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

Notes:
1. Appendix Table 5 reports estimates of the 90th, 95th and 99th percentiles of the asymptotic distribution of the ENC-NEW statistic when the rolling scheme is used.
2. The estimates were constructed based upon 5,000 simulated draws from the relevant distribution for a given value of both k_2 and π. See section 3.5 of the text for further detail on how the simulations were conducted.

Appendix Table 6

Percentiles of the ENC-NEW Statistic: Fixed Scheme

k_2	%-ile	.1	.2	.4	.6	.8	1.0	1.2	1.4	1.6	1.8	2.0
							$\pi =$					
1	0.99	.908	1.300	1.941	2.421	2.813	3.069	3.307	3.550	3.821	4.016	4.196
	0.95	.489	.698	1.037	1.231	1.477	1.622	1.792	1.881	2.089	2.187	2.272
	0.90	.322	.449	.670	.795	.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	.729	1.002	1.486	1.790	2.049	2.307	2.545	2.733	2.948	3.071	3.249
	0.90	.500	.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	.901	1.227	1.799	2.234	2.618	2.925	3.240	3.477	3.742	3.808	4.042
	0.90	.663	.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	.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	.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	.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

Notes:

1. Appendix Table 5 reports estimates of the 90th, 95th and 99th percentiles of the asymptotic distribution of the ENC-NEW statistic when the fixed scheme is used.

2. The estimates were constructed based upon 5,000 simulated draws from the relevant distribution for a given value of both k_2 and π. See section 3.5 of the text for further detail on how the simulations were conducted.