Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 140 (2007) 719–752
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Asymptotics for out of sample tests of Granger causality

Michael W. McCracken*

Board of Governors of the Federal Reserve System, 20th and Constitution N.W., Washington, DC 20551, USA

Available online 30 August 2006

Abstract

This paper presents analytical, Monte Carlo and empirical evidence concerning out-of-sample tests of Granger causality. The environment is one in which the relative predictive ability of two nested parametric regression models is of interest. Results are provided for three statistics: a regression-based statistic suggested by Granger and Newbold [1977. Forecasting Economic Time Series. Academic Press Inc., London], a t-type statistic comparable to those suggested by Diebold and Mariano [1995, Comparing Predictive Accuracy. Journal of Business and Economic Statistics, 13, 253–263] and West [1996. Asymptotic Inference About Predictive Ability, Econometrica, 64, 1067–1084], and an F-type statistic akin to Theil's U. Since the asymptotic distributions under the null are nonstandard, tables of asymptotically valid critical values are provided. Monte Carlo evidence supports the theoretical results. An empirical example evaluates the predictive content of the Chicago Fed National Activity Index for growth in Industrial Production and core PCE-based inflation.
Published by Elsevier B.V.

JEL classification: C12; C32; C52; C53

Keywords: Granger causality; Forecast evaluation; Hypothesis testing; Model selection

1. Introduction

Evaluating a time series models' ability to forecast is one method of determining its usefulness. Swanson (1998), Sullivan et al. (1999), Lettau and Ludvigson (2001), Rapach and Wohar (2002) and Hong and Lee (2003) are just a few examples of applications that

*Tel.: +202 452 3369; fax: +202 452 5296.
E-mail address: michael.w.mccracken@frb.gov.

0304-4076/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.jeconom.2006.07.020

have determined the appropriateness of a model based on its ability to predict out-of-sample. When using this methodology a model is determined to be valuable if the resulting forecast errors are deemed small relative to some loss function. Typically this loss function is squared error though others such as absolute error and directional accuracy have been used by Leitch and Tanner (1991) and Breen et al. (1989), respectively. This methodology is in contrast to traditional methods that determine quality of the predictive model based on its ability to replicate or "fit" the same realizations used to estimate the model (see Inoue and Kilian, 2004).

This paper contributes to recent analytical work on out-of-sample model evaluation, specifically that of West (1996), by providing asymptotic results for out-of-sample tests of Granger causality—tests that compare the predictive ability of two nested models allowing for a wide range of loss functions including, but not limited to, squared error. Null asymptotic distributions are derived for three commonly used tests that compare the out-of-sample predictive ability of two nested models: a regression-based test for equal mean squared error (MSE) proposed by Granger and Newbold (1977), a similar t-type test commonly attributed to either Diebold and Mariano (1995) or West (1996), and an F-type test similar in spirit to Theil's U but perhaps closer to in-sample likelihood ratio tests. Since the asymptotic distributions of the former two tests are identical they will be referenced simultaneously as "OOS-t" tests; the latter test will be referenced as an "OOS-F" test.

The asymptotic null distributions of both the OOS-t and OOS-F tests are nonstandard. Each can be written as functions of stochastic integrals of Brownian motion. Tables are provided in order to facilitate the use of these distributions. Monte Carlo evidence provided in Section 4 suggests that the tests can be well sized with good power in samples of the size often found in economic applications. Supporting Monte Carlo evidence can also be found in Clark and McCracken (2001, 2005a, b) as well as Inoue and Kilian (2004).

The remainder of the paper proceeds as follows. Section 2 provides a brief background on out-of-sample methods with an emphasis on nested model comparisons. Section 3 and its subsections provide notation, assumptions and theorems regarding the null asymptotics of the test statistics. Section 4 provides Monte Carlo evidence on the size and power of the tests when accuracy is measured using square loss. Section 5 provides empirical evidence on the predictive content of the Chicago Fed's National Activity Index for growth in Industrial Production and core PCE-based inflation. Section 6 concludes. All proofs are presented within Appendix A.

2. Background

In recent work, West (1996) shows how to construct asymptotically valid out-of-sample tests of predictive ability when forecasts are generated using estimated parameters. Conditions are provided under which t-type statistics will be asymptotically standard normal. These conditions extend and clarify previous analytical work on out-of-sample inference made by Mincer and Zarnowitz (1969), Chong and Hendry (1986), Ghysels and Hall (1990), Fair and Shiller (1989, 1990), Mizrach (1992) and Diebold and Mariano (1995).

Building upon West (1996), even more recent work on out-of-sample inference has developed. West and McCracken (1998) construct regression-based tests of predictive ability. Corradi et al. (2001) allow for the comparison of nonnested models when cointegrating relationships exist. McCracken (2000a) provides analytical results for out-of-sample inference when the test involves nonsmooth functions such as the indicator or absolute value function. Harvey et al. (1999) construct tests of equal predictive ability in the presence of ARCH. Diebold et al. (1998) discuss the evaluation of density forecasts. Ashley (1998) considers a double bootstrap algorithm for comparing predictive ability. White (2000), Hansen (2005) and Inoue and Kilian (2004) propose asymptotic and bootstrap approaches of compensating for data-snooping biases when comparing the predictive ability of a large number of models. Chao et al. (2001) construct a test of predictive ability that permits both nested and nonnested comparisons. Clark and McCracken (2001, 2005a) establish results for tests of forecast encompassing when comparing nested models. Corradi and Swanson (2002, 2006) propose an integrated conditional moment type test of predictive ability that is consistent against generic nonlinear alternatives.

One question addressed by several of these authors is whether two models have the same predictive ability with respect to a loss function $L(.)$. In particular, Diebold and Mariano (1995) suggest a test of the form

$$\begin{aligned}\text{OOS-}t &= P^{-0.5}\hat{\Omega}^{-0.5}\sum_{t=R}^{T}[L(X_{t+1},\hat{\beta}_{1,t}) - L(X_{t+1},\hat{\beta}_{2,t})] \\ &= P^{-0.5}\hat{\Omega}^{-0.5}\sum_{t=R}^{T}[L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})], \qquad (1)\end{aligned}$$

where $\hat{\Omega} = P^{-1}\sum_{t=R}^{T}(L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t}) - \bar{d})^2$ denotes a consistent estimate[1] of the asymptotic variance of the scaled mean loss differential $P^{1/2}\bar{d} = P^{-0.5}\sum_{t=R}^{T}[L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})]$, $T+1 = P+R$; P is the number of out-of-sample observations and R is the number of observations used to construct the first forecast. When each forecast is constructed using $\hat{\beta}_{i,t}$, an estimate of the parameters associated with model $i = 1,2$, West (1996) shows that under certain conditions, the test statistic in (1) can be asymptotically standard normal.

Unfortunately, it is easy to overlook a crucial condition for asymptotic normality. For the results in West (1996) to apply it must be the case that Ω is positive. If Ω is zero then $P^{-0.5}\sum_{t=R}^{T}[L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})]\rightarrow_{p}0$. Since $\hat{\Omega}$ is still consistent for $\Omega = 0$, it is not immediately clear whether the OOS-t statistic in (1) is degenerate, divergent or bounded in probability. In particular, it is not clear that the statistic is asymptotically normal.

The failure of this condition may seem unlikely but indeed it is quite common. Using results in West (1996) one can easily show that Ω equals zero if the two parametric models are nested rather than nonnested. This has serious implications for out-of-sample tests of Granger causality and market efficiency for which the models are inherently nested.

For example, in testing for a causal relationship between aggregate advertising expenditure and aggregate consumption expenditure Ashley et al. (1980) construct an OOS-t statistic similar to that in (1). Using a method suggested by Granger and Newbold (1977) they test the null that advertising does not Granger cause consumption using the t-statistic (and standard normal tables) associated with α from the OLS estimated

[1] Under the conditions of this paper, the asymptotic distribution of (1) is invariant to whether $\bar{d}$ is included in the estimate of the variance.

regression $\hat{u}_{1,t+1} - \hat{u}_{2,t+1} = \alpha(\hat{u}_{1,t+1} + \hat{u}_{2,t+1}) +$ error term,

$$(P-1)^{1/2}\frac{P^{-1}\sum_{t=R}^{T}(\hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2)}{([P^{-1}\sum_{t=R}^{T}(\hat{u}_{1,t+1} + \hat{u}_{2,t+1})^2][P^{-1}\sum_{t=R}^{T}(\hat{u}_{1,t+1} - \hat{u}_{2,t+1})^2] - \bar{d}^2)^{1/2}}. \tag{2}$$

In (2) $\hat{u}_{1,t+1}$ is the one-step ahead forecast error from an autoregressive model for aggregate consumption, $\hat{u}_{2,t+1}$ is the one-step ahead forecast error from a bivariate vector autoregressive model for both aggregate consumption and aggregate advertising and $\bar{d} = P^{-1}\sum_{t=R}^{T}(\hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2)$.

There are also a number of potential applications to tests for the predictability of asset returns and more generally for tests of market efficiency. If the null is that asset returns form a martingale difference sequence then any parametric model for asset returns nests the null model within it. For example, building upon Meese and Rogoff (1983), Mark (1995) constructs an OOS-t statistics of the form in (1) to test the null that changes in exchange rates are unpredictable. If this is the case then the MSE using the null zero conditional mean model should equal the MSE using a linear model that depends upon certain fundamentals. Kilian (1999) constructs similar tests but under the null that changes in exchange rates form a martingale difference sequence around a nonzero unconditional mean.

It should be mentioned that Mark (1995) and Kilian (1999) each use the bootstrap when conducting inference. They do not reference standard normal tables per se. However, the reason they use the bootstrap is that they are concerned about finite sample size distortions relative to the (claimed) asymptotic standard normal distribution of (1). The results in Section 3 of this paper indicate that those distortions may also be because the asymptotic distribution is not standard normal nor is well approximated by a standard normal distribution.

This paper focuses on constructing asymptotically valid out-of-sample sample tests that compare the predictability of two nested parametric models. Three different statistics are considered. The first two, those from (1) and (2), are OOS-t statistics. The third is an OOS-F statistic taking the form

$$\text{OOS-}F = 2\cdot P\left[P^{-1}\sum_{t=R}^{T}[L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})]/\hat{c}\right]. \tag{3}$$

In the above formula, the constant $\hat{c}$ is a consistent estimate of a certain normalizing constant c that depends upon the choice of loss function and is discussed in the following section. When $L(.)$ denotes a negative log-likelihood and $\hat{c} = 1$, this statistic takes the form of a standard likelihood ratio statistic but adapted to an out-of-sample context. Under square loss and setting $\hat{c} = 2P^{-1}\sum_{t=R}^{T}\hat{u}_{2,t+1}^2$ this statistic takes the form of the standard F-test but again adapted to an out-of-sample context.

Before continuing we wish to be clear about what is and what is not accomplished within this paper. We derive the null asymptotic distributions of tests of equal forecast accuracy between two nested parametric models. In particular, we do so for test statistics with asymptotic distributions that are free of nuisance parameters. By doing so we feel we provide methods for inference in numerous applications. Even so, by restricting attention to parametric models and distributions that are free of nuisance parameters we acknowledge that there are applications for which the theory presented here is not applicable. Examples include tests of equal forecast accuracy for horizons greater than 1 and for the comparison of models that are misspecified in a sense made precise in Section three. Null asymptotics for long-horizons and misspecified linear models under square loss can be found in Clark and McCracken (2005a). We also exclude applications like those in Diebold and Nason (1990), Swanson and White (1997) and McCracken (2000b) who use local-regression, series-based and kernel-based nonparametric methods respectively to construct forecasts.

Notice also that we focus exclusively on *how* out-of-sample methods can be used to construct tests of Granger causality. We do not formally address *why* one might do so. Suggestions for why are scattered throughout the forecasting literature. Ashley et al. (1980) suggest that out-of-sample methods are more closely aligned with the spirit of the original concept of Granger causality.[2] They also, along with Fair and Shiller (1990), conjecture that out-of-sample inference may be more robust to model selection biases and overfitting. Formal arguments for this view, however, appear to be lacking. Inoue and Kilian (2004) provide analytical and Monte Carlo evidence that out-of-sample tests need not be less sensitive to model selection biases and overfitting, and that they may have lower power than in-sample tests in some cases. Clark and McCracken (2005b) provide analytical and Monte Carlo evidence suggesting that out-of-sample tests under certain conditions may be well suited for detecting predictive ability at the end of the sample. Regardless, we take the view that one has already chosen to use such methods and the only question is how to do so.

3. Theoretical results

This section provides the null asymptotic distributions of both the OOS-t tests in (1) and (2) and the OOS-F test in (3). It does so in three subsections. Section 3.1 presents the environment and assumptions. Section 3.2 presents the asymptotic null distributions of the statistics (1)–(3). Section 3.3 provides a discussion of the asymptotic results as well as tables of asymptotically valid critical values.

3.1. Environment and assumptions

The sample of observables $\{X_s\}_{s=1}^{T+1}$ is of length $T+1$. Using that sample, the researcher wishes to compare the one-step ahead predictive ability of two nested parametric regression models. These models are indexed as $i = 1, 2$ for the nested and nesting models, respectively. This structure allows for the applications discussed in Section 2.

Given the pair of nested models, two sequences of one-step ahead forecasts are constructed using one of three methods. We consider each of these schemes since each is prevalent within the forecasting literature. For example, the recursive, rolling and fixed schemes are used in Pagan and Schwert (1990), Swanson (1998) and Ashley et al. (1980), respectively.

Under the recursive scheme, the parameters are updated at each new forecast origin using the largest possible observation window. Specifically, at each time $t = R, \ldots, T$ the parameter estimate $\hat{\beta}_{i,t}$ depends explicitly on all information from $s = 1, \ldots, t$. Using these

[2] Similarly, Klein (1992) argues that the "…ability to make useful ex-ante forecasts is the *real* test of a model." In contrast, Clements (2002) emphasizes the point that models that forecast well need not be useful for policy and in particular need not imply anything about the validity of a particular economic theory.

parameter estimates, forecasts $\hat{y}_{i,t+1}(\hat{\beta}_{i,t})$, forecast errors $\hat{u}_{i,t+1} = y_{t+1} - \hat{y}_{i,t+1}(\hat{\beta}_{i,t})$ and losses $L(X_{t+1}, \hat{\beta}_{i,t}) \equiv L_{i,t+1}(\hat{\beta}_{i,t})$ are constructed for each model $i = 1, 2$.

The forecasts, forecast errors and losses for the rolling and fixed schemes differ from the recursive only in how the parameter estimates are constructed. For the rolling scheme, the parameter estimate is sequentially updated using an observation window of fixed width R ending at the forecast origin $s = t$ and beginning at time $s = t-R+1$. For the fixed scheme the parameter estimate is not updated at all. Instead, for all $t = R, \ldots, T$, $\hat{\beta}_{i,R} = \hat{\beta}_{i,t}$.

Using each of the two series of subsequent forecast errors and losses, one each from the nesting and nested models, a test statistic of the form (1)–(3) is constructed. Based upon the value of this statistic one either fails to reject or rejects the null of equal predictive ability. The null and alternative can be stated as

$$H_0 : \mathrm{E}L_{1,t+1}(\beta_1^*) = \mathrm{E}L_{2,t+1}(\beta_2^*) \quad \text{vs.} \quad H_A : \quad \mathrm{E}L_{1,t+1}(\beta_1^*) > \mathrm{E}L_{2,t+1}(\beta_2^*). \tag{4}$$

The alternative is one-sided rather than two sided because the two models are nested. Note that if a log-likelihood function is being used to evaluate predictive ability, (4) implies that $L(.)$ is defined as its negative.

Before discussing specific assumptions, some notation is required. For the loss function $L_{i,t}(\beta_i)$ let $h_{i,t}(\beta_i) = \partial L_{i,t}(\beta_i)/\partial\beta_i$, $h_{i,t}(\beta_i^*) = h_{i,t}$, $q_{i,t}(\beta_i) = \partial^2 L_{i,t}(\beta_i)/\partial\beta_i\partial\beta_i'$ and $q_{i,t}(\beta_i^*) = q_{i,t}$. For any matrix A with elements $a_{i,j}$ let $|A| = \max_{i,j}|a_{i,j}|$. Let J denote the selection matrix$(I_{k_1\times k_1},\ 0_{k_1\times k_2})'$. Without loss of generality, let $\beta_2^* = (\beta_{2,1}^*{}', \beta_{2,2}^*{}')' = (\beta_{2,1}^*{}', 0)' = (\beta_1^*{}', 0)' = (k_1 + k_2 = k \times 1)$. The following assumptions are not necessary and sufficient, only sufficient.

Assumption 1. The parameter estimates $\hat{\beta}_{i,t}$, $i = 1, 2$, $t = R, \ldots, T$, satisfy $\hat{\beta}_{i,t} - \beta_i^* = B_i(t)H_i(t)$, where $B_i(t)H_i(t)$ equals $(t^{-1}\sum_{s=1}^{t} q_{i,s}(\dot{\beta}_{i,t}))^{-1}(t^{-1}\sum_{s=1}^{t} h_{i,s})$, $(R^{-1}\sum_{s=t-R+1}^{t} q_{i,s}(\dot{\beta}_{i,t}))^{-1}(R^{-1}\sum_{s=t-R+1}^{t} h_{i,s})$, $(R^{-1}\sum_{s=1}^{R} q_{i,s}(\dot{\beta}_{i,t}))^{-1}(R^{-1}\sum_{s=1}^{R} h_{i,s})$ for the recursive, rolling and fixed schemes, respectively, some $\dot{\beta}_{i,t}$ contained in the closed cube with opposing vertices $\hat{\beta}_{i,t}$ and β_i^*.

Assumption 2. For $i = 1, 2$ and all t: (a) $\beta_i \in \Theta_i$, Θ_i compact; (b) $\mathrm{E}L_{i,t}(\beta_i)$ is uniquely minimized at $\beta_i^* \in \Theta_i$ with $\mathrm{E}q_{i,t}$ nonsingular; (c) in an open neighborhood N_i around $\beta_i^* L_{i,t}(\beta_i)$ is twice continuously differentiable; (d) for all $\beta_1 \in \Theta_1$ and $\tilde{\beta}_2 = (\beta_1', 0)'$, $L_{1,t}(\beta_1) = L_{2,t}(\tilde{\beta}_2)$, $J'h_{2,t}(\tilde{\beta}_2) = h_{1,t}(\beta_1)$ and $J'q_{2,t}(\tilde{\beta}_2)J = q_{1,t}(\beta_1)$; (e) In the open neighborhood N_i, and for all t there exists a finite positive constant φ and a uniformly L^v bounded positive random variable m_t such that $|q_{i,t}(\beta_i) - q_{i,t}(\beta_i^*)| \leqslant m_t|\beta_i - \beta_i^*|^{\varphi}$ with $2 \leqslant v\varphi$; (f) $\sup_{\beta_i\in\Theta_i}|t^{-1}\sum_{s=1}^{t} L_{i,t}(\beta_i) - \mathrm{E}L_{i,t}(\beta_i)| \to_{\text{a.s}} 0$.

Assumption 3. (a) $U_t = [h_{2,t}', \mathrm{vech}(q_{2,t} - \mathrm{E}q_{2,t})']'$ is covariance stationary, (b) $\mathrm{E}(h_{2,t}|h_{2,t-j}, q_{2,t-j}, j>0) = 0$, (c) $\mathrm{E}h_{2,t}h_{2,t}' = c\mathrm{E}q_{2,t} < \infty$, (d) for some $r>8$, U_t is uniformly L^r bounded, (e) for some $r>d>2$, U_t is strong mixing with coefficients of size $-rd/(r-d)$, (f) with $\tilde{U}_t$ denoting the vector of nonredundant elements of U_t, $\lim_{T\to\infty} T^{-1}\mathrm{E}(\sum_{s=1}^{T}\tilde{U}_s)(\sum_{s=1}^{T}\tilde{U}_s)' = \Omega < \infty$ is positive definite.

Assumption 4. $\lim_{P,R\to\infty} P/R = \pi \in (0, \infty)$ with $\lambda = (1+\pi)^{-1}$.

Assumption 4′. $\lim_{P,R\to\infty} P/R = 0$ with $\lambda = 1$.

Much of Assumption 1 simply clarifies how the parameter estimates are constructed relative to the recursive, rolling and fixed schemes. More importantly, it imposes the restriction that the loss function used to evaluate predictive ability is the same as that used to estimate the model parameters. For example, if MSE is the measure of predictive ability, parameters must be estimated using OLS, NLLS, or maximum likelihood under the additional assumption that the disturbances are normal. Note also that unless otherwise stated (e.g. Theorem 3.3 and Proposition 3.1) we only require that the loss function depends upon a parametric forecast. It need not depend directly upon a forecast error.

Assumption 2 insures that the parameters are identified and are consistently estimated. It is comparable to Theorem 2.1 of Newey and McFadden (1994). The substantive components of this assumption are the requirements that the loss function be twice continuously differentiable and that the models be nested. This allows for MSE, Linex and many log-likelihood type measures of predictive ability but eliminates applications, like that of Weiss and Andersen (1984), that estimate the parameters using LAD and then use mean absolute error as the measure of predictive ability.

In Assumption 3 we impose sufficient conditions for the application of an invariance principle. The details of Assumption 3 above are directly comparable to those for Theorems 2.1 and 3.1 in Hansen (1992). Another important aspect of this assumption arises in (b). In the terminology of Granger (1999), we require that the generalized forecast errors form a martingale difference sequence. The assumption has the side effect of imposing a type of correct model specification as discussed in detail by Patton and Timmermann (2003) and thus rules out applications where the predictive models are misspecified. Finally, note the definition of the normalizing constant c in (c). A discussion of this constant and estimating it consistently is discussed in Section 3.2.

The final two assumptions introduce the means by which the asymptotics are achieved. As in Ghysels and Hall (1990), West (1996), and White (2000), the asymptotic distributions are derived by imposing a slightly stronger condition than simply that the sample size T becomes arbitrarily large. Assumption 4 requires that both the number of in-sample (R) and out-of-sample (P) observations become arbitrarily large at the same rate. Assumption 4′ allows for the possibility that although P and R diverge with T, P is still arbitrarily small relative to R. By deriving the asymptotics under both scenarios, we feel we capture the majority of applications.[3]

We close this section with two additional assumptions. These are strengthened versions of Assumptions 2 and 3 in the context of squared loss. These modified assumptions are necessary because the OOS-t statistic in (2) is structurally distinct from the out-of-sample statistics in (1) and (3). In particular, the denominator component of (2) is distinct from that in (1) and as such slightly different assumptions are needed. In the following let $\nabla\hat{y}_{i,t+1}(\beta_i) = \partial\hat{y}_{i,t+1}(\beta_i)/\partial\beta_i$, $\nabla^2\hat{y}_{i,t+1}(\beta_i) = \partial^2\hat{y}_{i,t+1}(\beta_i)/\partial\beta_i\,\partial\beta_i'$ and note that under the null $u_{1t+1} = u_{2,t+1} \equiv u_{t+1}$.

Assumption 2′. Same as Assumption 2 but (e) in the open neighborhood N_i, and for all t there exists a finite positive constant φ and a uniformly L^v bounded positive random variable m_t such that for $2 \leqslant v\varphi$,

$$4 \quad \max[|\nabla\hat{y}_{i,t+1}(\beta_i)\nabla\hat{y}'_{i,t+1}(\beta_i) - \nabla\hat{y}_{i,t+1}(\beta_i^*)\nabla\hat{y}'_{i,t+1}(\beta_i^*)|,$$
$$|(y_{t+1} - \hat{y}_{i,t+1}(\beta_i))\nabla^2\hat{y}_{i,t+1}(\beta_i) - u_{t+1}\nabla^2\hat{y}_{i,t+1}(\beta_i^*)|] \leqslant m_t|\beta_i - \beta_i^*|^{\varphi}.$$

[3] Giacomini and White (2005) consider related asymptotics within which R is finite and P diverges.

Assumption 3′. Same as Assumption 3 but (a) $U_t = [u_{t+1}, \nabla\hat{y}_{2,t+1}(\beta_2^*), \text{vech}(\nabla^2\hat{y}_{2,t+1}(\beta_2^*) - \text{E}\nabla^2\hat{y}_{2,t+1}(\beta_2^*))']'$ is fourth order stationary, (c) $\text{E}h_{2,t}h_{2,t}' = 2\sigma^2\text{E}q_{2,t} < \infty$ and $\text{E}u_{t+1}\nabla^2\hat{y}_{2,t+1}(\beta_2^*) = 0$.

Assumption 2′ and 3′ impose detailed assumptions on u_{t+1}, $\nabla\hat{y}_{2,t+1}(\beta_2^*)$ and $\nabla^2\hat{y}_{2,t+1}(\beta_2^*)$ rather than $h_{2,t+1}$ and $q_{2,t}$. Assumption 2′ is a technical assumption necessary for showing that certain terms vanish asymptotically. Assumption 3′ imposes the restriction that the forecast errors are conditionally homoskedastic and uncorrelated with $\nabla^2\hat{y}_{2,t+1}(\beta_2^*)$.

3.2. Asymptotics under the null

This section presents the null asymptotic distributions of the OOS-t statistics in (1) and (2) as well as the OOS-F statistic in (3). When square loss is the measure of accuracy, we reference these statistics as MSE-t, MSE-Reg and MSE-F, respectively. In the following, $W(.)$ denotes a $(k_2 \times 1)$ standard Brownian motion. For the recursive, fixed and rolling schemes, let Γ_1 denote $\int_\lambda^1 s^{-1}W'(s)\,\text{d}W(s)$, $\lambda^{-1}\{W(1) - W(\lambda)\}'W(\lambda)$ and $\lambda^{-1}\int_\lambda^1\{W(s) - W(s-\lambda)\}'\,\text{d}W(s)$, respectively. Similarly, for the recursive, fixed and rolling schemes let Γ_2 denote $\int_\lambda^1 s^{-2}W'(s)W(s)\,\text{d}s$, $\pi\lambda^{-1}W'(\lambda)'W(\lambda)$ and $\lambda^{-2}\int_\lambda^1\{W(s) - W(s-\lambda)\}'\{W(s) - W(s-\lambda)\}\,\text{d}s$, respectively.

Theorem 3.1. (a) *Let* Assumption 1–4 *hold.* OOS-$t \to_\text{d} (\Gamma_1 - (0.5)\Gamma_2)/\Gamma_2^{1/2}$. (b) *Let* Assumption 1–4′ *hold and let* V_0 *and* V_1 *denote* $(k_2 \times 1)$ *independent standard normal vectors.* OOS-$t \to_\text{d} V_0'V_1/[V_0'V_0]^{1/2} \sim \text{N}(0,1)$.

Theorem 3.2. (a) *Let* Assumption 1–4 *hold.* OOS-$F \to_\text{d} 2\Gamma_1 - \Gamma_2$. (b) *Let* Assumption 1–4′ *hold and let* V_0 *and* V_1 *denote the independent standard normal vectors from* Theorem 3.1. $(R/P)^{1/2}$OOS-$F \to_\text{d} 2V'_0V_1$.

Theorem 3.3. *Let* Assumptions 1, 2′, 3′, and 4 or 4′ *hold.* MSE-t−MSE-Reg $= \text{o}_\text{p}(1)$.

Theorems 3.1 and 3.2 provide the asymptotic null distribution of the OOS-t and OOS-F statistics in (1) and (3), respectively. These distributions depend not only on which of the recursive, rolling or fixed forecasting schemes are used but also depend upon whether $\pi > 0$ or 0. Interestingly, when $\pi = 0$ the OOS-F requires rescaling in order to achieve a nondegenerate asymptotic distribution. Regardless of the sampling scheme or value of π, note that each of the asymptotic null distributions are free of nuisance parameters and in particular, do not depend upon the choice of loss function. As such we are able to construct estimates of asymptotically valid critical values without knowledge of the underlying data generating process. Tables of these critical values are provided and discussed in Section 3.3.

In Theorem 3.3 we show that the MSE-t and MSE-Reg statistics are asymptotically equivalent regardless of whether $\pi = 0$ or $\pi > 0$. We immediately obtain the result that they have the same asymptotic null distribution and hence the same tables can be used when conducting inference.

For the OOS-t and MSE-Reg statistics, using these critical values is straightforward. For the OOS-F, we must first provide a consistent estimate of the normalizing constant c noted in Section 3.1. As defined in Assumption 3, c is the finite nonzero constant satisfying $\text{E}h_{2,t}h_{2,t}' = c\text{E}q_{2,t}$. Fortunately for two leading cases, identifying c is trivial. For the case in which $L(.)$ denotes a negative log-likelihood (so that one is evaluating a predictive density), $c = 1$. For the case in which linear OLS estimated models with conditionally homoskedastic errors are evaluated using square loss, $c = 2\sigma_u^2$.

For other loss functions, identifying c is less straightforward. We discuss one particular approach noting that others are possible.[4] Consider the special case in which the loss functions only depend upon the parameters through a forecast error and hence $L_{i,t+1}(\hat{\beta}_{i,t}) = L(y_{t+1} - \hat{y}_{i,t+1}(\hat{\beta}_{i,t})) = L(\hat{u}_{i,t+1})$. In the notation of Huber (1972), $h_{i,t+1}(\beta_i) = \partial L(y_{t+1} - \hat{y}_{i,t+1}(\beta_i))/\partial\beta_i = -\psi(y_{t+1} - \hat{y}_{i,t+1}(\beta_i))\nabla\hat{y}_{i,t+1}(\beta_i)$, $h_{i,t+1} = -\psi(u_{t+1})\nabla\hat{y}_{i,t+1}(\beta_i^*)$, $q_{i,t+1}(\beta_i) = \partial^2 L(y_{t+1} - \hat{y}_{i,t+1}(\beta_i))/\partial\beta_i\,\partial\beta_i' = \psi'(y_{t+1} - \hat{y}_{i,t+1}(\beta_i))(\nabla\hat{y}_{i,t+1}(\beta_i)\nabla\hat{y}_{i,t+1}(\beta_i)) - \psi(y_{t+1} - \hat{y}_{i,t+1}(\beta_i))\nabla^2\hat{y}_{i,t+1}(\beta_i)$ and $q_{i,t+1} = \psi'(u_{t+1})(\nabla\hat{y}_{i,t+1}(\beta_i^*)\nabla\hat{y}_{i,t+1}'(\beta_i^*)) - \psi(u_{t+1})\nabla^2\hat{y}_{i,t+1}(\beta_i^*)$. The following proposition provides a method of identifying the constant c.

Proposition 3.1. *Let $\psi(u_{t+1})$ and $\psi'(u_{t+1})$ each be mean independent of $\nabla\hat{y}_{2,t+1}(\beta_2^*)$ and $\nabla^2\hat{y}_{2,t+1}(\beta_2^*)$. If $\mathrm{E}\psi(u_{t+1})\nabla^2\hat{y}_{2,t+1}(\beta_2^*) = 0$ then $c = \mathrm{E}\psi^2(u_{t+1})/\mathrm{E}\psi'(u_{t+1})$.*

Proposition 3.1 provides an easy to implement method of identifying the constant c using only the loss function and the forecast errors. The clearest case for which it will be applicable is when the population forecast errors u_{t+1} are independent of the information used to construct the forecasts. Note that for linear models the assumptions are a bit less strict since $\partial^2\hat{y}_{2,t+1}(\beta_2^*)/\partial\beta_2\,\partial\beta_2'$ equals zero. As an application of this proposition first consider the case of square loss. Here we find that $\psi(u_{t+1}) = 2u_{t+1}$ and $\psi'(u_{t+1}) = 2$ and hence we arrive at the conclusion from the previous paragraph whereby $c = 2\sigma_u^2$. Now consider Linex loss for which $L(u_{t+1}) = \exp(a \cdot u_{t+1}) - a \cdot u_{t+1} + 1$, where the finite nonzero hyperparameter a controls the degree and direction of skewness in the loss function. Here we find that $\psi(u_{t+1}) = a\exp(a \cdot u_{t+1}) - a$ and $\psi(u_{t+1}) = a^2\exp(a \cdot u_{t+1})$ and hence $c = \mathrm{E}(a\exp(a \cdot u_{t+1}) - a)^2/\mathrm{E}(a^2\exp(a \cdot u_{t+1}))$. Other loss functions can be handled similarly.

Having identified the constant c, estimation is straightforward. We can apply either the results in West (1996) or McCracken (2000a) depending upon the degree of smoothness of the loss function. For example, under square loss and conditional homoskedasticity $\hat{c} = 2P^{-1}\sum_{t=R}^{T}\hat{u}_{2,t+1}^2 \to_{\mathrm{p}} c = 2\sigma_u^2$. Similarly for Linex loss we find that

$$\hat{c} = P^{-1}\sum_{t=R}^{T}(a\exp(a \cdot \hat{u}_{2,t+1}) - a)^2/P^{-1}\sum_{t=R}^{T}(a^2\exp(a \cdot \hat{u}_{2,t+1}))$$
$$\to_{\mathrm{p}} c = \mathrm{E}(a\exp(au_{t+1}) - a)^2/\mathrm{E}(a^2\exp(au_{t+1})).$$

Certainly other estimators could be used. For instance, here we estimate the constant c using forecasts from the nesting model rather than the nested model. We do so simply because of the intuitive relationship between the MSE-F statistic and the standard in-sample F-test.

[4]For example, straightforward algebra reveals that $c = c = \mathrm{tr}[(\mathrm{E}q_{2,t})^{-1}(\mathrm{E}h_{2,t}h_{2,t}')]/k$. Applying the results of West (1996) or McCracken (2000a) we find that $\mathrm{tr}[(P^{-1}\sum_{t=R}^{T}q_{2,t+1}(\hat{\beta}_{2,t}))^{-1}(P^{-1}\sum_{t=R}^{T}h_{2,t+1}(\hat{\beta}_{2,t})h_{2,t+1}'(\hat{\beta}_{2,t}))]/k \to_{\mathrm{p}} c$. We do not focus on this case since in the absence of other assumptions, there is no obvious reason why Assumption 3 (c) would hold. The sole exception to this argument is the special case in which $k_1 = 0$ and $k_2 = k = 1$. In this situation the proportionality always holds.

3.3. Discussion of the null distributions

Though the null asymptotic distributions do not depend upon the loss function itself, the distributions are dependent upon two parameters. The first is the number of excess parameters k_2. We can see this in the dimension of the vector Brownian motion $W(s)$ or standard normal vectors V_0 and V_1. It is easier to see if we rewrite the asymptotic distribution of the OOS-F statistic. Consider the recursive sampling scheme and let $\pi > 0$. If we let $W_i(s)$ denote the ith element of $W(s)$ then

$$\text{OOS-}F \to_{\mathrm{d}} \sum_{i=1}^{k_2} \left[2 \int_{\lambda}^{1} s^{-1} W_i(s)\, \mathrm{d}W_i(s) - \int_{\lambda}^{1} s^{-2} W_i^2(s)\, \mathrm{d}s \right]. \tag{5}$$

This representation is useful for two purposes. First, it provides some insight into the effect that k_2 has on the mean of the asymptotic distribution. Taking expectations, and noting that each of the $i = 1, \ldots, k_2$ elements is independent and identically distributed, it is straightforward to show that the mean of the asymptotic distribution takes the value $-k_2 \ln(1+\pi)$ for the recursive scheme and $-k_2\pi$ for the rolling and fixed schemes. Hence as k_2 increases, for $\pi > 0$, we expect the distribution of the OOS-F statistic to drift into the negative orthant. In the $\pi = 0$ approximation, the mean is zero for all k_2. Second, it provides insight into the effect that k_2 has on the variance of the asymptotic distribution. Since the characterization of the asymptotic distribution in (5) can be written as the sum of k_2 i.i.d. terms we know that the variance is monotonically increasing in k_2. The same conclusion holds for the $\pi = 0$ case.

Although it is clear that k_2 has no effect on the asymptotic distribution of the OOS-t when $\pi = 0$, its effect when $\pi > 0$ is less clear. Since the characterization of the asymptotic distribution consists of the ratio of stochastic terms it is difficult to analytically derive properties concerning its mean and variance. Numerical results later within this section suggest that the mean does become increasingly negative in k_2 but the variance is relatively constant in k_2.

A second parameter, π, affects the null asymptotic distributions of the OOS-t and OOS-F statistics. It affects the asymptotic distributions in three ways. First, it directly affects the weights on each of the components of the statistics (recall that $\lambda = (1+\pi)^{-1}$). Second, it affects the range of integration on each of the stochastic integrals through λ. Finally, it can have an affect on the rate of convergence to the asymptotic distribution. In particular, when $\pi = 0$ the OOS-F requires rescaling by $(R/P)^{1/2}$ to obtain a nondegenerate asymptotic distribution.

Since the parameter π enters both asymptotic distributions nonlinearly its affect on their distributions is less clear than it was for k_2. Looking at (5) we can say with certainty that the asymptotic mean of the OOS-F statistic decreases with π just as it did with k_2. Numerical results indicate that the asymptotic variance of the OOS-F is also monotonically increasing in π for a fixed value of k_2. For the OOS-t, numerical results suggest that the mean is decreasing and the variance is increasing in π, but to a lesser extent than for the OOS-F.

Tables 1–3 provide critical values for the OOS-t statistic while those in 4–6 are for the OOS-F. These were generated numerically using the asymptotic distributions in Theorems 3.1 and 3.2 and hence can be considered estimates of the true asymptotic critical values. The critical values are the 90th, 95th and 99th percentiles of 5000 independent draws from

Table 1
Percentiles of OOS-t: recursive

k_2	%-ile	π 0.0	0.1	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
1	0.99	2.326	1.921	1.784	1.625	1.515	1.462	1.436	1.413	1.343	1.316	1.274	1.238
	0.95	1.645	1.245	1.111	0.994	0.971	0.863	0.771	0.740	0.705	0.671	0.638	0.610
	0.90	1.280	0.885	0.780	0.657	0.598	0.512	0.443	0.402	0.370	0.330	0.306	0.281
2	0.99	2.326	1.986	1.856	1.563	1.436	1.387	1.312	1.276	1.196	1.158	1.127	1.074
	0.95	1.645	1.274	1.140	0.986	0.868	0.782	0.704	0.623	0.596	0.537	0.507	0.478
	0.90	1.280	0.932	0.786	0.614	0.541	0.455	0.361	0.295	0.253	0.235	0.194	0.160
3	0.99	2.326	1.840	1.737	1.542	1.448	1.359	1.252	1.148	1.071	0.976	0.978	0.953
	0.95	1.645	1.300	1.120	0.968	0.808	0.685	0.610	0.552	0.496	0.438	0.419	0.386
	0.90	1.280	0.939	0.751	0.551	0.454	0.356	0.279	0.222	0.175	0.108	0.074	0.035
4	0.99	2.326	1.872	1.731	1.581	1.365	1.195	1.119	1.108	1.041	0.902	0.861	0.854
	0.95	1.645	1.264	1.101	0.914	0.772	0.609	0.502	0.419	0.345	0.285	0.239	0.221
	0.90	1.280	0.898	0.742	0.562	0.419	0.263	0.169	0.094	0.052	−0.014	−0.054	−0.106
5	0.99	2.326	1.849	1.679	1.468	1.242	1.095	0.995	0.979	0.913	0.795	0.732	0.677
	0.95	1.645	1.222	1.061	0.849	0.689	0.491	0.386	0.308	0.224	0.148	0.107	0.081
	0.90	1.280	0.866	0.694	0.461	0.315	0.179	0.062	−0.021	−0.083	−0.145	−0.174	−0.228
6	0.99	2.326	1.836	1.639	1.390	1.200	1.042	0.943	0.859	0.755	0.686	0.610	0.593
	0.95	1.645	1.192	0.998	0.768	0.615	0.429	0.328	0.259	0.141	0.078	0.055	−0.019
	0.90	1.280	0.823	0.642	0.394	0.256	0.108	−0.011	−0.101	−0.164	−0.218	−0.266	−0.319
7	0.99	2.326	1.836	1.649	1.341	1.154	0.994	0.872	0.810	0.637	0.549	0.476	0.438
	0.95	1.645	1.199	0.976	0.742	0.546	0.372	0.279	0.191	0.072	−0.002	−0.034	−0.105
	0.90	1.280	0.811	0.615	0.359	0.213	0.062	−0.088	−0.152	−0.230	−0.305	−0.363	−0.449
8	0.99	2.326	1.789	1.659	1.298	1.090	0.879	0.788	0.728	0.503	0.444	0.401	0.359
	0.95	1.645	1.193	0.928	0.677	0.462	0.302	0.198	0.105	0.020	−0.058	−0.101	−0.176
	0.90	1.280	0.773	0.574	0.329	0.139	0.003	−0.131	−0.203	−0.293	−0.383	−0.452	−0.516
9	0.99	2.326	1.813	1.607	1.268	1.112	0.804	0.724	0.634	0.523	0.427	0.391	0.305
	0.95	1.645	1.112	0.912	0.617	0.397	0.276	0.121	0.030	−0.055	−0.122	−0.193	−0.257
	0.90	1.280	0.733	0.561	0.273	0.096	−0.068	−0.187	−0.286	−0.377	−0.437	−0.518	−0.579
10	0.99	2.326	1.743	1.534	1.193	1.035	0.758	0.621	0.506	0.419	0.347	0.285	0.185
	0.95	1.645	1.082	0.890	0.566	0.358	0.205	0.043	−0.072	−0.162	−0.222	−0.296	−0.339
	0.90	1.280	0.749	0.529	0.226	0.032	−0.130	−0.248	−0.355	−0.454	−0.524	−0.591	−0.651

Notes: The critical values are those for the OOS-t when the recursive scheme is used for various values of k_2 and π. Note also that those critical values for $\pi = 0.0$ are identical to those in Tables 2 and 3. This arises since each of the recursive, rolling and fixed schemes are asymptotically equivalent in probability when $\pi = 0.0$.

the asymptotic distributions of OOS-F and OOS-t for a given sampling scheme and value of k_2 and π. For the $\pi > 0$ case, generating these draws proceeded as follows. Weights that depend upon π were estimated in the obvious fashion using $\hat{\pi} = P/R$. The necessary k_2 Brownian motions were simulated as random walks each using an independent sequence of 10,000 i.i.d. N(0, $T^{-0.5}$) increments. The integrals were emulated by summing the relevant weighted quadratics of the random walks from the $R+1$st observation to the Tth

Table 2
Percentiles of OOS-t: rolling

k_2	%-ile	π											
		0.0	0.1	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
1	0.99	2.326	1.875	1.799	1.604	1.447	1.340	1.221	1.179	1.098	1.021	0.969	0.882
	0.95	1.645	1.251	1.117	0.970	0.859	0.722	0.651	0.575	0.510	0.455	0.382	0.334
	0.90	1.280	0.903	0.776	0.637	0.530	0.401	0.317	0.246	0.180	0.136	0.116	0.078
2	0.99	2.326	1.959	1.757	1.504	1.325	1.180	1.165	0.996	0.953	0.883	0.744	0.640
	0.95	1.645	1.280	1.105	0.884	0.753	0.631	0.484	0.401	0.304	0.235	0.166	0.103
	0.90	1.280	0.915	0.755	0.569	0.425	0.280	0.155	0.111	0.026	−0.050	−0.094	−0.140
3	0.99	2.326	1.860	1.669	1.473	1.271	1.076	0.984	0.896	0.773	0.614	0.504	0.431
	0.95	1.645	1.274	1.088	0.842	0.667	0.490	0.381	0.251	0.146	0.066	−0.016	−0.084
	0.90	1.280	0.938	0.718	0.521	0.346	0.201	0.064	−0.042	−0.137	−0.224	−0.302	−0.346
4	0.99	2.326	1.905	1.700	1.503	1.183	1.003	0.903	0.755	0.656	0.455	0.342	0.234
	0.95	1.645	1.267	1.087	0.852	0.585	0.376	0.274	0.136	0.024	−0.080	−0.173	−0.222
	0.90	1.280	0.866	0.731	0.494	0.248	0.098	−0.047	−0.164	−0.262	−0.362	−0.434	−0.505
5	0.99	2.326	1.881	1.627	1.347	1.112	0.927	0.790	0.657	0.504	0.307	0.193	0.123
	0.95	1.645	1.229	1.034	0.716	0.479	0.280	0.155	−0.019	−0.090	−0.219	−0.329	−0.385
	0.90	1.280	0.825	0.694	0.402	0.154	−0.025	−0.168	−0.305	−0.399	−0.508	−0.589	−0.674
6	0.99	2.326	1.826	1.680	1.312	1.007	0.850	0.641	0.558	0.336	0.195	0.069	0.017
	0.95	1.645	1.176	0.966	0.621	0.407	0.225	0.058	−0.119	−0.218	−0.336	−0.428	−0.535
	0.90	1.280	0.811	0.602	0.319	0.088	−0.095	−0.262	−0.423	−0.523	−0.638	−0.732	−0.821
7	0.99	2.326	1.842	1.620	1.233	0.989	0.751	0.526	0.485	0.227	0.055	−0.039	−0.127
	0.95	1.645	1.154	0.936	0.628	0.346	0.171	−0.011	−0.182	−0.320	−0.433	−0.531	−0.663
	0.90	1.280	0.791	0.573	0.279	0.038	−0.157	−0.326	−0.497	−0.611	−0.750	−0.841	−0.933
8	0.99	2.326	1.819	1.582	1.178	0.918	0.702	0.466	0.349	0.132	−0.018	−0.176	−0.302
	0.95	1.645	1.157	0.924	0.562	0.258	0.081	−0.099	−0.281	−0.432	−0.552	−0.672	−0.785
	0.90	1.280	0.758	0.541	0.244	−0.042	−0.244	−0.408	−0.576	−0.727	−0.838	−0.957	−1.040
9	0.99	2.326	1.768	1.510	1.110	0.845	0.600	0.408	0.235	0.036	−0.099	−0.277	−0.407
	0.95	1.645	1.117	0.892	0.504	0.213	0.021	−0.156	−0.374	−0.529	−0.623	−0.785	−0.885
	0.90	1.280	0.742	0.520	0.193	−0.105	−0.322	−0.491	−0.657	−0.803	−0.951	−1.049	−1.153
10	0.99	2.326	1.713	1.428	1.075	0.808	0.536	0.298	0.122	−0.064	−0.248	−0.381	−0.482
	0.95	1.645	1.068	0.872	0.443	0.133	−0.038	−0.258	−0.466	−0.605	−0.765	−0.909	−1.011
	0.90	1.280	0.727	0.500	0.138	−0.144	−0.374	−0.568	−0.757	−0.902	−1.045	−1.167	−1.288

Notes: See the notes for Table 1.

observation. For the $\pi = 0$ case, draws from the asymptotic distribution of the OOS-F were generated by taking the product of independent standard normal ($k_2 \times 1$) vectors. For this case, recall that the OOS-t is asymptotically standard normal. We report these critical values for completeness.

Each table corresponds to either the recursive, rolling or fixed scheme. Within each table there are 360 critical values. Each of these correspond to one permutation of three parameters: $k_2 = \{1, 2, 3, \ldots, 9, 10\}$, $\pi = \{0.0, 0.1, 0.2, 0.4, \ldots, 1.0, 1.2, \ldots, 2.0\}$ and nominal size of the test {0.01, 0.05, 0.10}. Tables that allow for larger values of both k_2 and π are available from the author upon request.

Table 3
Percentiles of OOS-t: fixed

k_2	%-ile	π											
		0.0	0.1	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
1	0.99	2.326	2.201	2.051	1.974	2.061	2.037	2.024	1.992	2.018	1.996	2.016	1.993
	0.95	1.645	1.506	1.416	1.364	1.428	1.346	1.252	1.301	1.293	1.249	1.235	1.218
	0.90	1.280	1.149	1.079	1.042	1.040	0.976	0.917	0.896	0.893	0.908	0.834	0.862
2	0.99	2.326	2.145	2.089	1.923	1.947	1.964	1.749	1.751	1.665	1.725	1.646	1.613
	0.95	1.645	1.468	1.342	1.301	1.265	1.164	1.072	1.034	1.046	0.977	0.982	0.955
	0.90	1.280	1.096	0.999	0.901	0.873	0.798	0.711	0.680	0.639	0.578	0.556	0.520
3	0.99	2.326	2.045	1.977	1.957	1.805	1.739	1.602	1.520	1.597	1.463	1.513	1.407
	0.95	1.645	1.432	1.277	1.195	1.095	1.014	0.909	0.893	0.851	0.761	0.735	0.733
	0.90	1.280	1.063	0.922	0.793	0.705	0.621	0.540	0.511	0.455	0.386	0.373	0.306
4	0.99	2.326	2.013	1.883	1.829	1.687	1.528	1.467	1.475	1.422	1.318	1.255	1.277
	0.95	1.645	1.369	1.281	1.110	0.997	0.883	0.755	0.689	0.650	0.607	0.566	0.509
	0.90	1.280	1.004	0.895	0.764	0.575	0.476	0.367	0.340	0.273	0.204	0.171	0.081
5	0.99	2.326	1.930	1.878	1.716	1.596	1.405	1.254	1.301	1.230	1.171	1.115	1.034
	0.95	1.645	1.333	1.193	1.009	0.863	0.725	0.646	0.570	0.486	0.410	0.365	0.291
	0.90	1.280	0.945	0.838	0.636	0.487	0.374	0.258	0.193	0.115	0.020	−0.022	−0.085
6	0.99	2.326	1.933	1.874	1.628	1.481	1.382	1.146	1.188	1.091	1.016	1.007	0.878
	0.95	1.645	1.269	1.122	0.936	0.771	0.652	0.538	0.487	0.367	0.314	0.222	0.152
	0.90	1.280	0.912	0.764	0.552	0.400	0.299	0.169	0.103	0.003	−0.106	−0.146	−0.235
7	0.99	2.326	1.925	1.859	1.556	1.377	1.257	1.105	1.103	0.987	0.896	0.828	0.765
	0.95	1.645	1.263	1.086	0.878	0.692	0.557	0.446	0.346	0.254	0.191	0.074	0.014
	0.90	1.280	0.895	0.731	0.513	0.332	0.215	0.060	−0.003	−0.147	−0.252	−0.308	−0.386
8	0.99	2.326	1.856	1.827	1.467	1.245	1.146	1.029	0.980	0.860	0.786	0.762	0.666
	0.95	1.645	1.249	1.064	0.807	0.623	0.481	0.363	0.268	0.151	0.054	−0.042	−0.120
	0.90	1.280	0.868	0.663	0.467	0.247	0.153	−0.029	−0.115	−0.227	−0.343	−0.440	−0.502
9	0.99	2.326	1.878	1.697	1.440	1.198	1.124	0.902	0.791	0.683	0.644	0.595	0.507
	0.95	1.645	1.197	1.031	0.754	0.537	0.416	0.305	0.162	0.050	−0.067	−0.171	−0.242
	0.90	1.280	0.844	0.655	0.396	0.182	0.034	−0.111	−0.224	−0.303	−0.437	−0.543	−0.625
10	0.99	2.326	1.824	1.604	1.354	1.126	0.998	0.797	0.659	0.557	0.550	0.505	0.415
	0.95	1.645	1.143	1.007	0.688	0.455	0.337	0.167	0.040	−0.057	−0.174	−0.246	−0.358
	0.90	1.280	0.797	0.616	0.348	0.125	−0.055	−0.210	−0.305	−0.398	−0.559	−0.645	−0.729

Notes: See the notes for Table 1.

In each of Tables 1–6 it is useful to note that the critical values are not generally monotone in either k_2 or π. Earlier in this section we discussed the fact that the means of the OOS-F and OOS-t statistics were monotone decreasing in both k_2 and π and hence we expected the distributions to generally drift into the negative orthant as these parameters increased. That does not imply that the upper tails decrease monotonically. Consider the OOS-F statistic for a fixed value of π but allow k_2 to increase. As k_2 increases the mean of the asymptotic distribution becomes increasingly negative while at the same time the variance increases. Put together, these two forces imply that the upper tail need not be monotonically decreasing in k_2.

Table 4
Percentiles of OOS-F: recursive

k_2	%-ile	π											
		0.0	0.1	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
1	0.99	5.902	1.608	2.129	2.768	3.179	3.459	3.584	3.771	3.589	3.838	3.882	3.951
	0.95	3.270	0.850	1.038	1.298	1.554	1.567	1.548	1.583	1.623	1.599	1.553	1.518
	0.90	2.210	0.530	0.659	0.814	0.796	0.798	0.751	0.759	0.698	0.685	0.687	0.616
2	0.99	7.910	1.996	2.691	3.426	3.907	4.129	4.200	4.362	4.304	4.309	4.278	4.250
	0.95	4.826	1.184	1.453	1.733	1.891	1.820	1.802	1.819	1.752	1.734	1.692	1.706
	0.90	3.324	0.794	0.912	1.029	1.077	1.008	0.880	0.785	0.697	0.666	0.587	0.506
3	0.99	9.230	2.418	3.092	4.080	4.136	4.322	4.341	4.337	4.192	4.089	4.365	4.184
	0.95	5.946	1.434	1.710	2.062	2.073	1.978	1.909	1.930	1.795	1.715	1.710	1.612
	0.90	4.216	0.970	1.064	1.117	1.121	0.960	0.857	0.691	0.599	0.386	0.276	0.127
4	0.99	10.472	2.714	3.440	4.541	4.609	4.378	4.202	4.586	4.477	4.337	4.247	4.096
	0.95	6.712	1.566	1.964	2.246	2.194	1.900	1.809	1.578	1.376	1.256	1.122	1.029
	0.90	5.048	1.060	1.225	1.313	1.184	0.829	0.545	0.354	0.197	−0.058	−0.234	−0.456
5	0.99	11.398	2.902	3.673	4.466	4.434	4.249	4.351	4.349	4.187	3.945	3.783	3.783
	0.95	7.404	1.688	2.082	2.235	2.242	1.773	1.449	1.316	1.045	0.718	0.502	0.459
	0.90	5.568	1.130	1.277	1.228	0.958	0.614	0.241	−0.099	−0.361	−0.656	−0.820	−1.072
6	0.99	12.434	3.212	3.846	4.545	4.676	4.637	4.703	4.286	4.144	3.981	3.525	3.321
	0.95	8.164	1.828	2.124	2.217	2.121	1.660	1.360	1.181	0.761	0.413	0.299	−0.109
	0.90	6.216	1.220	1.313	1.164	0.890	0.419	−0.044	−0.405	−0.776	−1.072	−1.395	−1.664
7	0.99	13.212	3.450	4.098	4.508	4.419	4.271	4.312	4.150	3.677	3.155	3.090	2.880
	0.95	8.930	2.000	2.239	2.424	2.057	1.604	1.282	0.928	0.378	−0.008	−0.199	−0.591
	0.90	6.584	1.272	1.333	1.118	0.799	0.242	−0.363	−0.728	−1.194	−1.657	−2.033	−2.507
8	0.99	13.886	3.408	4.130	4.645	4.625	4.202	4.147	3.912	3.185	2.933	2.952	2.484
	0.95	9.346	2.136	2.312	2.373	1.895	1.390	0.943	0.587	0.131	−0.372	−0.680	−1.140
	0.90	7.014	1.338	1.369	1.058	0.552	0.014	−0.632	−1.076	−1.633	−2.174	−2.731	−3.160
9	0.99	15.010	3.540	4.388	4.703	4.873	4.122	4.066	3.753	3.027	2.925	2.802	2.186
	0.95	9.946	2.168	2.440	2.219	1.714	1.286	0.631	0.198	−0.356	−0.851	−1.241	−1.696
	0.90	7.480	1.354	1.432	0.920	0.393	−0.327	−1.007	−1.595	−2.229	−2.666	−3.250	−3.794
10	0.99	15.586	3.646	4.433	4.813	4.718	3.944	3.645	3.194	2.578	2.282	2.152	1.436
	0.95	10.414	2.202	2.489	2.157	1.536	1.055	0.205	−0.431	−1.071	−1.459	−1.988	−2.378
	0.90	7.862	1.458	1.401	0.884	0.155	−0.600	−1.341	−2.008	−2.782	−3.348	−3.839	−4.437

Notes: The critical values are those for the OOS-F when the recursive scheme is used for various values of k_2 and π. Note that the definition of the OOS-F changes when $\pi = 0$. In particular the usual statistic is rescaled by $(R/P)^{1/2}$. For this reason the critical values that change smoothly across columns for $\pi = 0.1$ to $\pi = 2.0$ do not do so between the columns for $\pi = 0.0$ to $\pi = 0.1$. Note also that those critical values for $\pi = 0.0$ are identical to those in Tables 5 and 6. This arises since each of the recursive, rolling and fixed schemes are asymptotically equivalent in probability when $\pi = 0.0$.

The density plots in Figs. 1 and 2 are intended to provide some feel for the behavior of the asymptotic distributions corresponding to the OOS-F and OOS-t statistics, respectively. In order to reduce the number of plots we focus exclusively on the recursive sampling scheme. Plots for the rolling and fixed schemes are qualitatively similar in shape. They do differ in location and scale. When the rolling and fixed schemes are used the

Table 5
Percentiles of OOS-F: rolling

k_2	%-ile	π 0.0	0.1	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
1	0.99	5.902	1.638	2.230	2.812	3.300	3.634	3.811	3.688	3.721	3.924	3.612	3.765
	0.95	3.270	0.854	1.112	1.394	1.644	1.627	1.583	1.574	1.469	1.488	1.378	1.215
	0.90	2.210	0.536	0.667	0.838	0.865	0.773	0.693	0.602	0.482	0.390	0.355	0.276
2	0.99	7.910	2.036	2.819	3.544	3.988	4.066	4.398	4.403	4.109	4.293	4.046	3.566
	0.95	4.826	1.232	1.481	1.802	1.889	1.841	1.695	1.495	1.264	1.015	0.783	0.504
	0.90	3.324	0.812	0.920	1.028	1.004	0.806	0.468	0.399	0.095	−0.198	−0.394	−0.623
3	0.99	9.230	2.476	3.128	4.135	4.120	4.264	4.519	4.386	4.123	3.373	3.089	2.685
	0.95	5.946	1.472	1.752	2.089	2.042	1.700	1.532	1.100	0.694	0.340	−0.071	−0.471
	0.90	4.216	1.006	1.074	1.135	0.944	0.617	0.224	−0.174	−0.600	−1.080	−1.529	−1.847
4	0.99	10.472	2.724	3.649	4.474	4.586	4.432	4.459	4.296	3.621	2.905	2.337	1.699
	0.95	6.712	1.600	2.078	2.332	1.979	1.536	1.228	0.701	0.116	−0.491	−1.112	−1.487
	0.90	5.048	1.096	1.284	1.263	0.777	0.356	−0.200	−0.788	−1.341	−1.973	−2.528	−3.182
5	0.99	11.398	3.008	3.721	4.504	4.710	4.508	4.199	4.042	3.216	2.167	1.370	1.055
	0.95	7.404	1.768	2.191	2.164	1.783	1.175	0.764	−0.121	−0.542	−1.454	−2.172	−2.765
	0.90	5.568	1.152	1.315	1.143	0.541	−0.114	−0.774	−1.583	−2.300	−3.102	−3.896	−4.649
6	0.99	12.434	3.214	4.093	4.532	4.786	4.456	3.899	3.473	2.324	1.500	0.615	0.159
	0.95	8.164	1.926	2.181	2.117	1.652	1.062	0.318	−0.745	−1.424	−2.302	−3.162	−4.256
	0.90	6.216	1.224	1.311	0.981	0.365	−0.428	−1.348	−2.392	−3.270	−4.292	−5.180	−6.114
7	0.99	13.212	3.552	4.340	4.568	4.566	4.181	3.450	3.167	1.696	0.442	−0.300	−1.151
	0.95	8.930	1.994	2.273	2.321	1.478	0.868	−0.076	−1.162	−2.243	−3.224	−4.261	−5.620
	0.90	6.584	1.292	1.355	0.939	0.151	−0.783	−1.833	−3.061	−4.153	−5.487	−6.474	−7.583
8	0.99	13.886	3.480	4.537	4.707	4.681	4.041	3.065	2.488	1.013	−0.144	−1.562	−2.729
	0.95	9.346	2.080	2.436	2.169	1.213	0.468	−0.626	−1.885	−3.233	−4.430	−5.681	−6.991
	0.90	7.014	1.386	1.357	0.886	−0.186	−1.278	−2.492	−3.824	−5.225	−6.556	−7.690	−8.939
9	0.99	15.010	3.552	4.438	4.711	4.443	3.754	2.622	1.723	0.327	−0.877	−2.543	−3.923
	0.95	9.946	2.164	2.518	1.966	1.075	0.138	−1.113	−2.620	−4.036	−5.258	−6.931	−8.345
	0.90	7.480	1.378	1.433	0.772	−0.521	−1.835	−3.139	−4.586	−6.133	−7.734	−9.173	−10.558
10	0.99	15.586	3.728	4.413	4.815	4.589	3.460	2.145	1.121	−0.624	−2.313	−3.795	−5.166
	0.95	10.414	2.224	2.520	1.893	0.730	−0.235	−1.733	−3.496	−4.940	−6.512	−8.255	−9.863
	0.90	7.862	1.456	1.411	0.555	−0.701	−2.189	−3.790	−5.566	−7.200	−9.046	−10.574	−12.294

Notes: See the notes for Table 4.

statistics have heavier tails and drift into the negative orthant much quicker than when the recursive scheme is used. For example when $k_2 = 4$ and $\pi = 2$ the 95th percentiles associated with the OOS-F statistic are 1.029, −1.487, and 1.784 for the recursive, rolling and fixed schemes, respectively.

Figs. 1 and 2 are each comprised of eight plots relating to the density of the OOS-F and OOS-t statistic, respectively. The first column of plots shows the effect on the density when π increases across 0.2, 1.0 and 2.0 holding k_2 constant at 1, 2, 5, and 10. In each plot, as π increases the probability that the statistic is negative increases. The second column of plots shows the effect on the density when k_2 increases across 1, 2, 5, 10 and 20 holding π

Table 6
Percentiles of OOS-F: fixed

k_2	%-ile	π 0.0	0.1	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
1	0.99	5.902	1.480	1.981	2.681	3.055	3.230	3.377	3.562	3.619	3.816	3.812	3.838
	0.95	3.270	0.784	1.015	1.345	1.534	1.677	1.667	1.738	1.807	1.812	1.857	1.862
	0.90	2.210	0.514	0.649	0.835	0.885	0.933	0.964	1.009	0.986	1.050	1.080	1.037
2	0.99	7.910	1.840	2.554	3.241	3.514	3.944	4.019	4.173	4.364	4.251	4.556	4.414
	0.95	4.826	1.132	1.421	1.765	1.999	2.077	2.116	2.169	2.232	2.275	2.260	2.195
	0.90	3.324	0.784	0.914	1.140	1.237	1.299	1.268	1.330	1.250	1.126	1.189	1.151
3	0.99	9.230	2.324	2.985	3.854	4.103	4.272	4.233	4.549	4.764	4.687	4.915	4.900
	0.95	5.946	1.408	1.653	2.050	2.322	2.308	2.319	2.325	2.336	2.187	2.283	2.275
	0.90	4.216	1.000	1.106	1.328	1.367	1.375	1.291	1.289	1.225	1.073	1.043	0.954
4	0.99	10.472	2.576	3.283	3.999	4.339	4.349	4.629	4.602	5.002	4.793	4.984	5.028
	0.95	6.712	1.536	1.947	2.374	2.478	2.426	2.238	2.310	2.175	2.063	1.891	1.784
	0.90	5.048	1.100	1.317	1.472	1.362	1.195	1.109	1.008	0.860	0.667	0.544	0.289
5	0.99	11.398	2.748	3.437	4.212	4.454	4.330	4.396	4.739	5.044	4.761	4.731	4.560
	0.95	7.404	1.664	2.018	2.368	2.504	2.337	2.167	2.109	1.862	1.593	1.540	1.249
	0.90	5.568	1.178	1.387	1.414	1.341	1.038	0.865	0.696	0.445	0.083	−0.075	−0.347
6	0.99	12.434	3.084	3.755	4.467	4.754	4.559	4.715	4.836	4.515	4.561	4.303	4.365
	0.95	8.164	1.826	2.164	2.406	2.422	2.267	2.010	1.995	1.654	1.302	1.107	0.744
	0.90	6.216	1.242	1.417	1.428	1.167	0.962	0.634	0.410	0.014	−0.449	−0.666	−1.113
7	0.99	13.212	3.294	3.980	4.599	4.683	4.704	5.000	4.828	4.667	4.489	4.367	4.155
	0.95	8.930	1.962	2.282	2.441	2.410	2.198	1.886	1.535	1.263	0.943	0.356	0.071
	0.90	6.584	1.342	1.536	1.457	1.057	0.817	0.254	−0.015	−0.668	−1.314	−1.610	−2.230
8	0.99	13.886	3.364	4.116	4.775	4.724	4.715	4.762	4.480	4.111	4.278	4.482	3.804
	0.95	9.346	2.078	2.394	2.580	2.244	2.007	1.666	1.266	0.744	0.293	−0.244	−0.658
	0.90	7.014	1.434	1.501	1.374	0.900	0.622	−0.146	−0.593	−1.157	−1.925	−2.678	−3.109
9	0.99	15.010	3.430	4.233	4.671	4.640	4.856	4.580	4.112	3.756	3.536	3.648	3.158
	0.95	9.946	2.090	2.525	2.533	2.064	1.881	1.434	0.892	0.292	−0.344	−0.960	−1.536
	0.90	7.480	1.474	1.564	1.329	0.727	0.181	−0.578	−1.168	−1.751	−2.653	−3.457	−4.122
10	0.99	15.586	3.582	4.232	4.750	4.674	4.489	4.251	3.643	3.467	3.373	3.342	3.036
	0.95	10.414	2.158	2.611	2.481	1.967	1.601	0.936	0.282	−0.360	−1.068	−1.467	−2.404
	0.90	7.862	1.504	1.583	1.176	0.520	−0.246	−1.125	−1.722	−2.449	−3.606	−4.312	−5.252

Notes: See the notes for Table 4.

constant at 0.2, 1, 2 and 50.0. In each plot, as k_2 increases the probability that the statistic is negative increases. Comparing plots across Figs. 1 and 2 it is clear that for any change in either k_2 or π holding the other constant, there is a much more dramatic effect on the density of the OOS-F statistic than that of the OOS-t statistic.

4. Monte Carlo evidence

In this section we provide simulation evidence on the finite sample size and power of the OOS-t and OOS-F tests discussed in Section 3. For brevity, attention is restricted to the

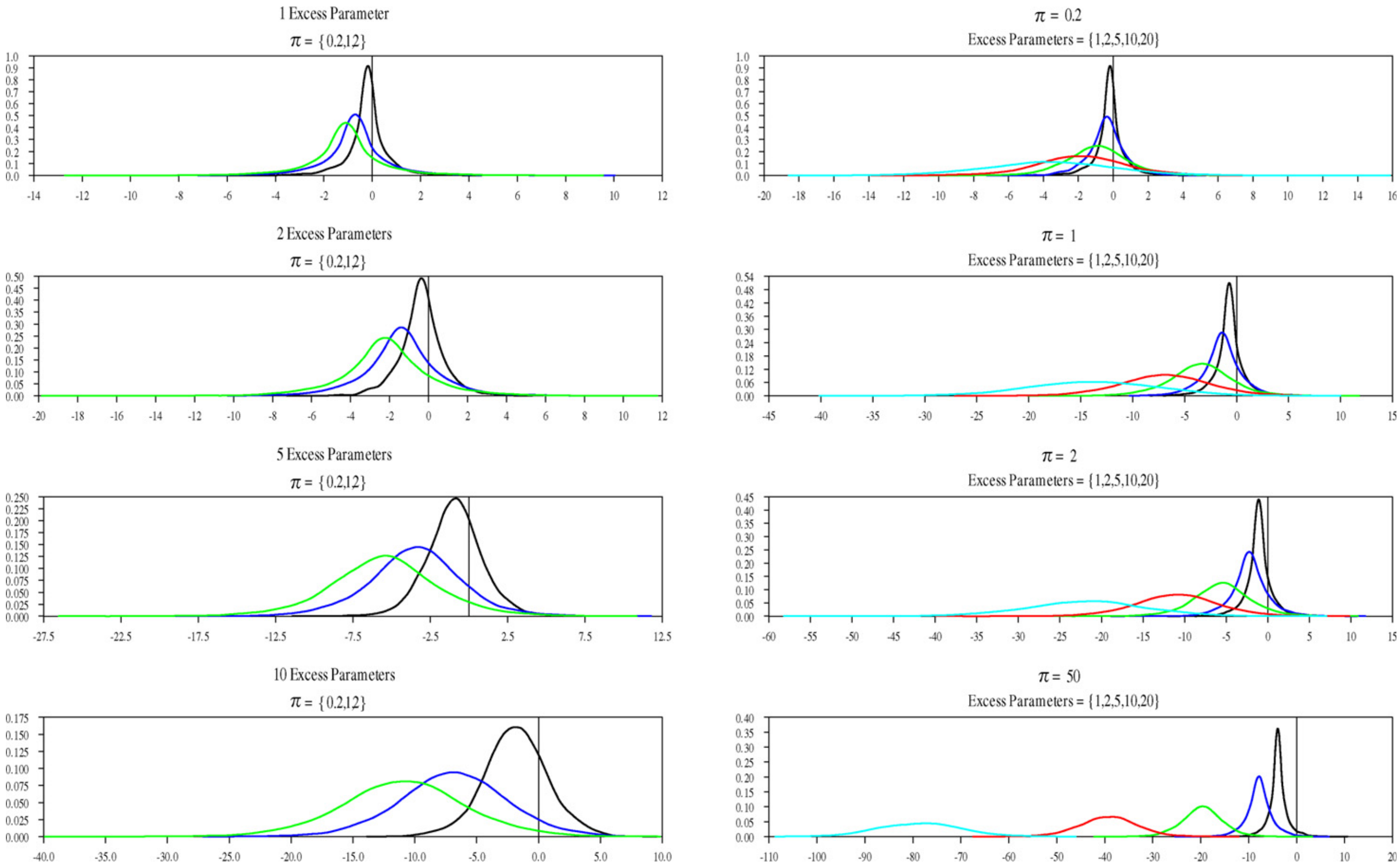

Fig. 1. Density plots for recursive OOS-*F*. *Notes*: The first column of panels provide the densities of the recursive OOS-*F* for a fixed value of k_2 as π increases across 0.2, 1, 2. The second column of panels provide the densities of the recursive OOS-*F* for a fixed value of π as k_2 increases across 1, 2, 5, 10, 20. In each case, the mode of the distribution shifts to the left as k_2 or π increases holding the other constant.

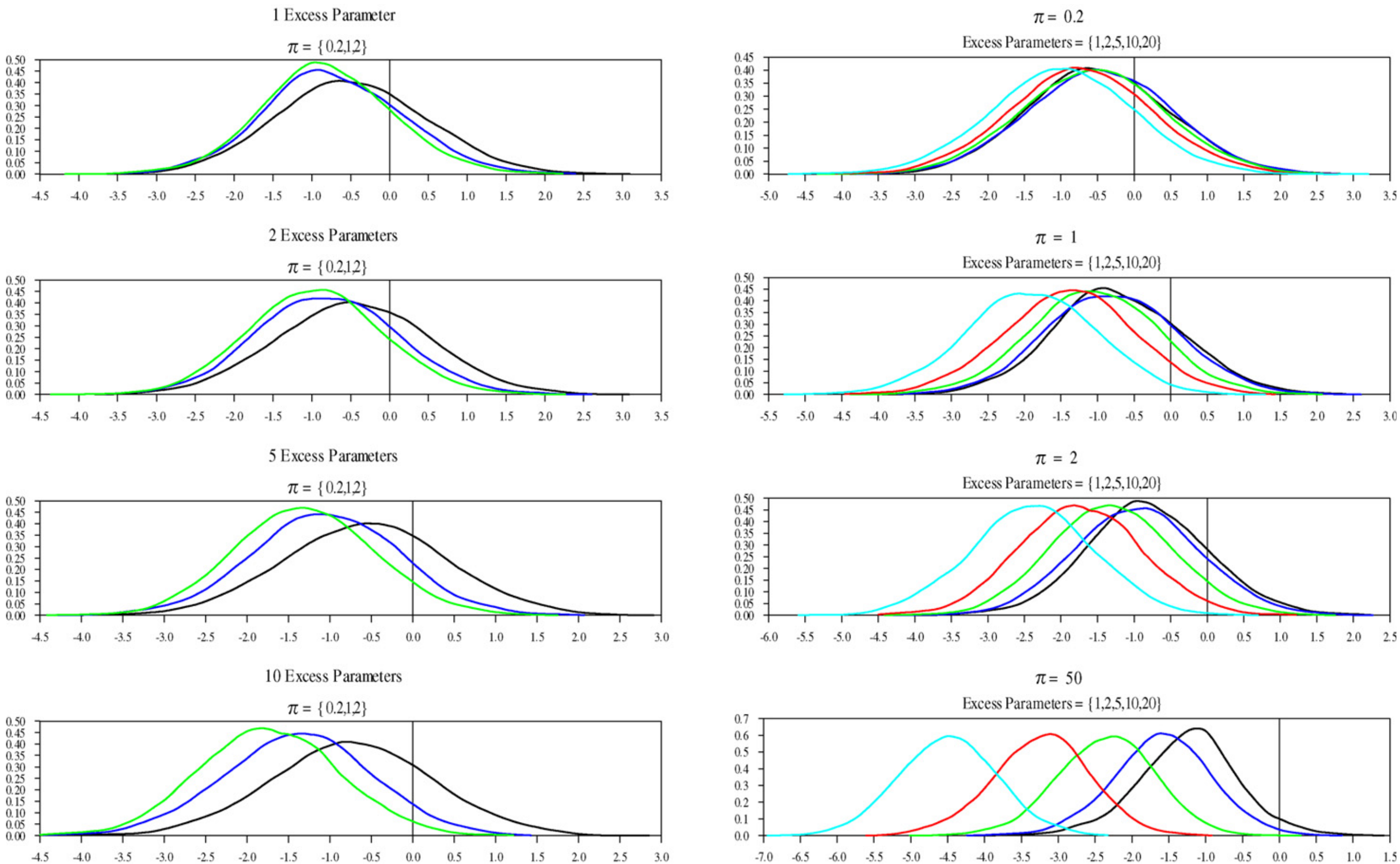

Fig. 2. Density plots for recursive OOS-*t*. *Notes*: The first column of panels provide the densities of the recursive OOS-*t* for a fixed value of k_2 as π increases across 0.2, 1, 2. The second column of panels provide the densities of the recursive OOS-*t* for a fixed value of π as k_2 increases across 1, 2, 5, 10, 20. In each case, the mode of the distribution shifts to the left as k_2 or π increases holding the other constant.

commonly used squared loss function and when the recursive scheme is used. For each of the MSE-t and MSE-F tests, and under both the null and alternative we conduct the test twice; once using the critical values implied by the theory when $\pi > 0$ and once using those when $\pi = 0$. For the sake of comparison, we also provide results for two other out-of-sample test statistics: a test of forecast encompassing suggested by Chao, Corradi and Swanson (CCS, 2001) and a modified version of a test of equal conditional predictive ability suggested by Giacomini and White (modified GW, 2005)[5].

In our Monte Carlo design we consider two distinct data-based data-generating processes. The first—DGP-1—is based upon OLS estimates of a bivariate VAR in IP growth (y_t) and the Chicago Fed National Activity Index (CFNAI) (x_t)

$$\begin{aligned} y_t &= 1.55 + .51y_{t-1} + \beta^*_{22}x_{t-1} + u_{y,t}, \\ x_t &= .12 + .02y_{t-1} + .41x_{t-1} - .06y_{t-2} + .69x_{t-2} + u_{x,t}, \\ &\mathrm{var}\begin{pmatrix} u_{y,t} \\ u_{x,t} \end{pmatrix} = \begin{pmatrix} 84.17 & \\ 6.68 & .70 \end{pmatrix} \end{aligned} \tag{6}$$

for i.i.d. normally distributed disturbances $u_{y,t}$ and $u_{x,t}$. The parameter $\beta^*_{2,2}$ takes the value 0 and 3.83 in the size and power experiments, respectively. The second—DGP-2—is constructed similarly but is based upon a bivariate model relating the change in core PCE-based inflation (y_t) and the CFNAI (x_t)

$$\begin{aligned} y_t &= .04 - .40y_{t-1} + \beta^*_{22}x_{t-1} + u_{y,t}, \\ x_t &= .07 + .69x_{t-1} + u_{x,t}, \\ &\mathrm{var}\begin{pmatrix} u_{y,t} \\ u_{x,t} \end{pmatrix} = \begin{pmatrix} 3.64 & \\ -.11 & .79 \end{pmatrix}, \end{aligned} \tag{7}$$

where the parameter $\beta^*_{2,2}$ takes the values 0 and 0.16 in the size and power experiments, respectively.

When choosing these empirically-based parameterizations, we were careful to avoid data-mining as much as possible. First, following the advice in Clark (2004), the models were selected using the initial (R) in-sample observations. Second, when parameterizing the restricted model for each of the y_t equations, we use BIC to choose a purely autoregressive model while always including an intercept; in each case an AR(1) was chosen. The unrestricted model is chosen by adding one lag of x_t to the restricted model. However, even this approach might be considered data-mining. To address this issue, in all of our simulation results we report the rejection frequencies not only based upon all of the replications but separately based upon only those replications for which BIC selects an AR(1) for the y_t equation. We reference these two sets of results as being 'unconditional' and 'conditional', respectively. The parameterizations for both x_t equations were selected using BIC to choose the number of lags for both x_t and y_t.

In each of the experiments we generate a sequence of 600 observations. Using these observations we consider a range of sample splits (R, P) corresponding to the values

[5]The CCS statistic is constructed using the fact that the errors are conditionally homoskedastic and the recursive scheme is being used. The modified GW statistic is constructed following their Corollary 2 with test function $(1, \hat{d}_t)$. To be absolutely clear, we use the qualifier 'modified' since their asymptotic theory requires the use of the rolling scheme with finite R while we implement it using the recursive scheme.

$\hat{\pi} = P/R = 0.1$, 0.4, 1.0 and 2.0 allowing R to increase across 50, 100, 150 and 200. We therefore have a total of 16 permutations of sample splits (R, P). Starting values for y_t and x_t were drawn from the unconditional distribution. The rejection frequency is estimated based upon 5000 replications.

Each of the statistics was constructed using the implied forecast errors from linear models that were used to forecast y_{t+1}. The restricted model takes the form $y_{t+1} = \beta_{1,1} + \beta_{1,2}y_t + u_{1,t+1} = x'_{1,t}\beta_1^* + u_{1,t+1}$ and the unrestricted model takes the form $y_{t+1} = \beta_{2,1} + \beta_{2,2}y_t + \beta_{2,3}x_t + u_{2,t+1} = x'_{2,t}\beta_2^* + u_{2,t+1}$. Note that under the null, each of these models is correctly specified in the sense that the implied population generalized forecast errors form a martingale difference sequence and hence the results of this paper are directly applicable. In particular, the conditions of Proposition 3.1 hold and hence when constructing the MSE-F statistic we let $\hat{c}$ equal $2P^{-1}\sum_{t=R}^{T}\hat{u}_{2,t+1}^2$.

Tables 7 and 8 provide the actual size and power of the tests, respectively. In each, the first column of panels corresponds to the unconditional size or power of the test while the second column of panels provides the same but conditional on BIC having selected a pure AR(1) for the y_t equation. For the MSE-t and MSE-F statistics the tests are constructed using the appropriate 95%-ile from Tables 1 and 4. Critical values for the CCS and modified GW statistics are the 95%-ile's of central chi-square distributions with 1 and 2 degrees of freedom, respectively.

First consider the size results from Table 7. For both DGP's, and using $\pi > 0$ critical values, the MSE-F and MSE-t statistics are well sized with perhaps a slight tendency to overreject when R and π are small. In contrast and consistent with the theory from Section 3, when using the $\pi = 0$ critical values both of these statistics are seriously undersized except when π is small. Of the remaining two statistics the CCS statistic is reasonably well sized in all situations with perhaps a slight tendency to underreject when R and π are small. In contrast the modified GW statistic is generally undersized except for those instances in which R and π are large. That the modified GW statistic is poorly sized except when P is large relative to R makes sense given that it is derived assuming that P is arbitrarily large relative to R. See below for more discussion in the context of power. Note that the relative impact of conditioning on the selected number of autoregressive lags equaling 1 is minimal.

Now consider the power results from Table 8. One immediately recognizes a sharp distinction between the ability to detect predictive ability in the first and second DGPs. For DGP-1, all of the tests have actual power of 1 for large enough values of π. For DGP-2 that is not the case. Here power is often quite low attaining a maximum of roughly 50% when the MSE-F statistic is used with $\pi > 0$ critical values. In many cases though, power is much lower sometimes hovering in the 5–10% level depending on the choice of R and π. Some patterns do emerge across the four permutations of (MSE-F, MSE-t) and ($\pi > 0$, and 0). In terms of power, it is uniformly the case that each MSE-F ($\pi > 0$ or 0) has weakly greater power than the corresponding MSE-t. Moreover, it is uniformly the case that the $\pi > 0$ asymptotic approximation (MSE-F or MSE-t) provides weakly greater power than the corresponding $\pi = 0$ approximation. Hence the MSE-F using the $\pi > 0$ approximation is most powerful and the MSE-t using the $\pi = 0$ approximation is least powerful. That the MSE-F statistic is more powerful than the corresponding MSE-t corroborates the evidence in Clark and McCracken (2001, 2005a, b). In terms of power, the CCS statistic behaves much like the MSE-t statistic using $\pi > 0$ critical values.

Among these statistics the modified GW statistic has the worst "power". Here we place quotation marks around the word power in order to emphasize that when using the term, it

Table 7
Size of tests

	Unconditional					Conditional				
	P/R					P/R				
	R	0.1	0.4	1.0	2.0	R	0.1	0.4	1.0	2.0
DGP-1										
MSE-*F*: $\pi>0$	50	0.076	0.057	0.049	0.052	50	0.072	0.054	0.049	0.053
	100	0.056	0.051	0.046	0.042	100	0.054	0.051	0.046	0.041
	150	0.051	0.053	0.045	0.048	150	0.049	0.050	0.045	0.049
	200	0.053	0.051	0.050	0.052	200	0.053	0.052	0.050	0.050
MSE-*F*: $\pi=0$	50	0.049	0.027	0.011	0.005	50	0.046	0.025	0.010	0.006
	100	0.036	0.024	0.010	0.006	100	0.034	0.024	0.011	0.006
	150	0.034	0.022	0.010	0.005	150	0.030	0.020	0.010	0.006
	200	0.036	0.023	0.015	0.005	200	0.035	0.023	0.015	0.005
MSE-*t*: $\pi>0$	50	0.114	0.074	0.062	0.051	50	0.110	0.072	0.063	0.052
	100	0.087	0.058	0.051	0.044	100	0.088	0.059	0.052	0.043
	150	0.077	0.064	0.051	0.049	150	0.075	0.062	0.052	0.050
	200	0.077	0.057	0.055	0.053	200	0.077	0.058	0.056	0.052
MSE-*t*: $\pi=0$	50	0.059	0.023	0.008	0.003	50	0.057	0.022	0.008	0.004
	100	0.046	0.015	0.007	0.002	100	0.047	0.015	0.007	0.002
	150	0.036	0.014	0.005	0.002	150	0.036	0.014	0.005	0.003
	200	0.036	0.014	0.007	0.003	200	0.037	0.015	0.007	0.002
CCS	50	0.021	0.044	0.043	0.044	50	0.021	0.043	0.041	0.042
	100	0.036	0.043	0.044	0.040	100	0.036	0.042	0.042	0.039
	150	0.045	0.043	0.039	0.044	150	0.044	0.041	0.039	0.044
	200	0.043	0.046	0.050	0.048	200	0.043	0.047	0.050	0.048
Modified GW	50	0.000	0.014	0.023	0.035	50	0.000	0.014	0.023	0.037
	100	0.006	0.025	0.035	0.042	100	0.006	0.024	0.035	0.042
	150	0.010	0.029	0.042	0.053	150	0.010	0.030	0.041	0.053
	200	0.018	0.040	0.046	0.053	200	0.017	0.040	0.046	0.054
DGP-2										
MSE-*F*: $\pi>0$	50	0.076	0.057	0.052	0.052	50	0.073	0.058	0.051	0.054
	100	0.056	0.055	0.051	0.054	100	0.054	0.054	0.051	0.053
	150	0.056	0.055	0.054	0.055	150	0.055	0.057	0.055	0.056
	200	0.058	0.055	0.054	0.054	200	0.059	0.055	0.054	0.053
MSE-*F*: $\pi=0$	50	0.052	0.026	0.015	0.005	50	0.049	0.025	0.016	0.004
	100	0.034	0.028	0.011	0.007	100	0.033	0.028	0.011	0.006
	150	0.039	0.025	0.012	0.006	150	0.039	0.026	0.012	0.006
	200	0.040	0.027	0.015	0.006	200	0.041	0.027	0.016	0.006
MSE-*t*: $\pi>0$	50	0.106	0.071	0.063	0.053	50	0.104	0.069	0.061	0.055
	100	0.083	0.064	0.055	0.053	100	0.082	0.064	0.055	0.052
	150	0.081	0.062	0.056	0.056	150	0.080	0.062	0.056	0.057
	200	0.080	0.059	0.060	0.052	200	0.080	0.058	0.059	0.051
MSE-*t*: $\pi=0$	50	0.055	0.021	0.009	0.004	50	0.053	0.021	0.009	0.005
	100	0.041	0.016	0.007	0.004	100	0.040	0.016	0.006	0.004
	150	0.041	0.016	0.006	0.005	150	0.041	0.017	0.006	0.005
	200	0.041	0.013	0.008	0.003	200	0.042	0.013	0.008	0.004

Table 7 (*continued*)

	Unconditional					Conditional				
	P/R					P/R				
	R	0.1	0.4	1.0	2.0	R	0.1	0.4	1.0	2.0
CCS	50	0.031	0.048	0.050	0.049	50	0.031	0.046	0.051	0.053
	100	0.038	0.048	0.048	0.051	100	0.039	0.049	0.049	0.052
	150	0.049	0.050	0.047	0.054	150	0.050	0.050	0.048	0.055
	200	0.045	0.044	0.051	0.054	200	0.045	0.043	0.052	0.053
Modified GW	50	0.000	0.016	0.024	0.033	50	0.000	0.016	0.024	0.032
	100	0.007	0.023	0.037	0.049	100	0.007	0.023	0.037	0.048
	150	0.012	0.028	0.045	0.053	150	0.011	0.030	0.047	0.053
	200	0.016	0.033	0.049	0.056	200	0.016	0.033	0.047	0.056

Notes: The first column of panels provides the actual unconditional size of nominal 5% tests while the second provides the size conditional on BIC selecting 1 AR lag for the conditioning variables All tests are constructed using the recursive scheme with $k_2 = 1$. Critical values for the MSE-t and MSE-F tests are taken from Tables 1 and 4. Critical values for the CCS and GW statistics are taken from $\chi^2(1)$ and $\chi^2(2)$ distributions respectively

is defined only relative to a particular alternative. In this paper, we derive asymptotic theory for testing the null hypothesis that, at the population level, the restricted and unrestricted models have equal predictive ability throughout the entire sample and hence $\mathrm{E}u_{1,t+1}^2 = \mathrm{E}u_{2,t+1}^2$ for all $t = 1, \ldots, T+1$. The relevant alternative is therefore that the unrestricted model exhibits better population level predictive ability than the restricted model at some point in the sample. In contrast, and among other things, the statistic proposed by Giacomini and White (2005) is designed to test the null hypothesis that, *in finite samples*, the restricted and unrestricted models have, on average, equal predictive ability throughout the out-of-sample period and hence $P^{-1}\sum_{t=R}^{T}\mathrm{E}\hat{u}_{1,t+1}^2 = P^{-1}\sum_{t=R}^{T}\mathrm{E}\hat{u}_{2,t+1}^2$. This permits the possibility that, despite having better population predictive ability, the unrestricted model might forecast as well as (or even worse than) the restricted model due to finite sample imprecision in the parameter estimates. As discussed in Giacomini and White (2005) as well as Clark and McCracken (2005c) this seemingly minor distinction in the statement of the null hypothesis can lead to sharp differences in the correct decision rule when deciding which forecasting model to use. As we will see in the following section, this theoretical distinction is empirically relevant as well.

5. Empirical evidence

In this section we follow a number of papers including Stock and Watson (2002) and Shintani (2005) that ask whether diffusion indices are useful for forecasting macroeconomic aggregates. In particular, as in Brave and Fisher (2004) and Clark and McCracken (2005c) we consider whether the CFNAI provides marginal predictive content for monthly growth in industrial production (IP) and monthly changes in core PCE-based inflation (PI). Following the theory in this paper, we focus on short term, one period ahead forecasts.

Table 8
Power of tests

	Unconditional					Conditional				
	P/R					P/R				
	R	0.1	0.4	1.0	2.0	R	0.1	0.4	1.0	2.0
DGP-1										
MSE-F: $\pi>0$	50	1.000	1.000	1.000	1.000	50	1.000	1.000	1.000	1.000
	100	1.000	1.000	1.000	1.000	100	1.000	1.000	1.000	1.000
	150	1.000	1.000	1.000	1.000	150	1.000	1.000	1.000	1.000
	200	1.000	1.000	1.000	1.000	200	1.000	1.000	1.000	1.000
MSE-F: $\pi=0$	50	1.000	1.000	1.000	1.000	50	1.000	1.000	1.000	1.000
	100	1.000	1.000	1.000	1.000	100	1.000	1.000	1.000	1.000
	150	1.000	1.000	1.000	1.000	150	1.000	1.000	1.000	1.000
	200	1.000	1.000	1.000	1.000	200	1.000	1.000	1.000	1.000
MSE-t: $\pi>0$	50	0.944	1.000	1.000	1.000	50	0.950	1.000	1.000	1.000
	100	1.000	1.000	1.000	1.000	100	1.000	1.000	1.000	1.000
	150	1.000	1.000	1.000	1.000	150	1.000	1.000	1.000	1.000
	200	1.000	1.000	1.000	1.000	200	1.000	1.000	1.000	1.000
MSE-t: $\pi=0$	50	0.708	1.000	1.000	1.000	50	0.739	1.000	1.000	1.000
	100	0.980	1.000	1.000	1.000	100	0.975	1.000	1.000	1.000
	150	0.997	1.000	1.000	1.000	150	0.996	1.000	1.000	1.000
	200	1.000	1.000	1.000	1.000	200	1.000	1.000	1.000	1.000
CCS	50	0.616	1.000	1.000	1.000	50	0.629	1.000	1.000	1.000
	100	1.000	1.000	1.000	1.000	100	1.000	1.000	1.000	1.000
	150	1.000	1.000	1.000	1.000	150	1.000	1.000	1.000	1.000
	200	1.000	1.000	1.000	1.000	200	1.000	1.000	1.000	1.000
Modified GW	50	0.000	0.918	1.000	1.000	50	0.000	0.915	1.000	1.000
	100	0.275	1.000	1.000	1.000	100	0.287	1.000	1.000	1.000
	150	0.755	1.000	1.000	1.000	150	0.760	1.000	1.000	1.000
	200	0.934	1.000	1.000	1.000	200	0.892	1.000	1.000	1.000
DGP-2										
MSE-F: $\pi>0$	50	0.109	0.110	0.132	0.180	50	0.109	0.112	0.135	0.185
	100	0.120	0.159	0.216	0.319	100	0.118	0.163	0.219	0.319
	150	0.146	0.208	0.293	0.427	150	0.144	0.208	0.291	0.426
	200	0.184	0.247	0.361	0.537	200	0.186	0.250	0.362	0.533
MSE-F: $\pi=0$	50	0.077	0.055	0.050	0.047	50	0.077	0.053	0.047	0.049
	100	0.089	0.092	0.091	0.106	100	0.088	0.093	0.090	0.107
	150	0.114	0.125	0.151	0.178	150	0.113	0.125	0.152	0.179
	200	0.145	0.158	0.199	0.254	200	0.148	0.157	0.198	0.250
MSE-t: $\pi>0$	50	0.124	0.106	0.131	0.161	50	0.121	0.107	0.129	0.164
	100	0.112	0.130	0.191	0.277	100	0.113	0.132	0.193	0.276
	150	0.117	0.167	0.248	0.373	150	0.118	0.169	0.246	0.373
	200	0.132	0.184	0.302	0.465	200	0.135	0.182	0.301	0.459
MSE-t: $\pi=0$	50	0.061	0.035	0.027	0.022	50	0.059	0.033	0.025	0.023
	100	0.057	0.040	0.035	0.042	100	0.058	0.042	0.037	0.042
	150	0.064	0.052	0.057	0.078	150	0.065	0.052	0.057	0.078
	200	0.067	0.063	0.075	0.108	200	0.069	0.063	0.074	0.106

Table 8 (*continued*)

	Unconditional					Conditional				
	P/R					P/R				
	R	0.1	0.4	1.0	2.0	R	0.1	0.4	1.0	2.0
CCS	50	0.032	0.059	0.096	0.147	50	0.031	0.057	0.094	0.149
	100	0.048	0.079	0.147	0.255	100	0.048	0.080	0.148	0.258
	150	0.061	0.108	0.206	0.363	150	0.062	0.110	0.206	0.363
	200	0.068	0.125	0.258	0.459	200	0.068	0.126	0.257	0.455
Modified GW	50	0.000	0.017	0.026	0.028	50	0.000	0.015	0.026	0.028
	100	0.007	0.026	0.029	0.034	100	0.007	0.026	0.028	0.033
	150	0.013	0.031	0.037	0.046	150	0.013	0.031	0.038	0.047
	200	0.020	0.034	0.047	0.057	200	0.019	0.035	0.047	0.056

Notes: The first column of panels provides the actual unconditional power of nominal 5% tests while the second provides the power conditional on BIC selecting 1 AR lag for the conditioning variables All tests are constructed using the recursive scheme with $k_2 = 1$. Critical values for the MSE-*t* and MSE-*F* tests are taken from Tables 1 and 4. Critical values for the CCS and GW statistics are taken from $\chi^2(1)$ to $\chi^2(2)$ distributions, respectively.

Data for core PCE and IP are obtained from the St. Louis Fed FRED II database while the CFNAI is obtained from the Chicago Fed website. Together, these variables are available on a monthly basis from 1967:03 to 2005:6. Annualized monthly IP growth is constructed as $y_{t+1} = 1200\Delta \ln(\mathrm{IP}_{t+1})$ while the annualized change in PCE-based inflation is constructed as $y_{t+1} = 1200\Delta^2 \ln(\mathrm{PI}_{t+1})$. The national activity index is left in levels and hence $x_t = \mathrm{CFNAI}_t$. For each dependent variable y_{t+1}, the restricted models were selected using BIC to choose an autoregressive model for y_{t+1} (with intercept). The unrestricted model is constructed by simply adding one lag of x_t to the restricted model. For the forecasting exercise we consider two separate out-of-sample periods. Since the first (nonhistoric) release of the CFNAI was for January 2001, it seems natural to consider the recent 2001:01–2005:06 ($P = 54$) period as one out-of-sample period. For the sake of comparison, we also consider a longer, but historic, out-of-sample period 1979:01–2000:12 ($P = 264$). In each instance, OLS-based forecasts are constructed recursively using data starting in 1968:01 and ending one month prior to the out-of-sample period. Hence R equals 132 ($\pi = 2$) and 342 ($\pi = .14$) for the historic and nonhistoric periods, respectively.[6]

As noted in Section 4, we attempt to avoid some of the obvious sources of data-mining by not using BIC to select the number of lags of x_t in the unrestricted model. In addition we choose the number of lags of y_t using only data prior to the first forecast in either forecasting exercise. This implies using BIC to choose a pure autoregressive model for both dependent variables using data from the 1967:3–1978:12 period. Since BIC selects 1 lag for both dependent variables, in each of our forecasting exercises we compare the predictive ability of forecasts from the restricted model $y_{t+1} = \beta_{1,1} + \beta_{1,2}y_t + u_{1,t+1}$ with those from the unrestricted model $y_{t+1} = \beta_{2,1} + \beta_{2,2}y_t + \beta_{2,3}x_t + u_{2,t+1}$.

The results are presented in Table 9. The first row of the panel contains the MSE of the restricted model over the out-of-sample period while the second row contains the ratio of

[6]We use the critical values associated with π equaling 2 and 0.1 when conducting the tests.

Table 9
Empirical results

	IP		PI	
	1979:1–2000:12	2001:1–2005:6	1979:1–2000:12	2001:1–2005:6
AR MSE	58.970	35.779	4.370	7.477
Factor/AR	0.877	0.793	1.003	0.998
MSE-t	3.100***	3.123***	−0.292	0.156
MSE-F	37.009***	13.863***	−0.766	0.090
CCS	0.050	1.937	0.399	0.276
Modified GW	13.760***	8.580**	1.575	2.298

Notes: The first row provides the MSE associated with the restricted, AR model while the second row provides the relative MSE of the unrestricted model to that of the restricted one. The remaining four rows provide values of the relevant test statistics. All tests are constructed using the recursive scheme with $k_2 = 1$. ***, ** and * denote significance at the 1, 5, and 10 % level. Critical values for the MSE-t and MSE-F tests are taken from Tables 1 and 4. Critical values for the CCS and modified GW statistics are taken from $\chi^2(1)$ to $\chi^2(2)$ distributions, respectively.

the two MSEs. In nominal terms it is clear that the CFNAI improves forecast accuracy for IP growth—providing 12% and 20% improvements. This is in contrast to the results for changes in inflation for which there is little difference in the MSEs. These observations by in large carry over to the test statistics. In the case of IP growth, each of the MSE-t, MSE-F and modified GW statistics provide strong evidence of predictive ability for both out-of-sample periods. This differs for the case of inflation for which there is no significant evidence of predictive ability from any of the tests.

The empirical results are largely consistent with the size and power results in Section 4. For the IP-based DGP-1, predictive ability was easily detected as indicated by all tests having power approaching 1 for most (R, π) combinations. Subsequently, in the empirical results we see that most of the tests strongly reject the null of equal predictive ability in favor of the unrestricted model—though the CCS statistic is a notable exception.

For the PI-based DGP-2 that is no longer the case. Here there was limited power among the statistics with the MSE-F attaining a maximum near 50%. Subsequently, in the empirical results we see that not only do none of the tests reject the null, the nominal MSE ratios are very close to 1, indicating equal predictive ability even accounting for parameter estimation error. One possible interpretation of this Phillips curve-type application (thinking of the index as a measure of real activity) is that there is no population level predictive ability of the index. Another interpretation is that while there is population level predictive ability, it is small enough to be overwhelmed by the incremental estimation uncertainty introduced by the unrestricted model relative to that from the restricted model.[7] If the former interpretation is correct, then for that hypothesis the "power" results in Table 8 indicate that most of the tests have little chance to detect deviations from that null with the modified GW statistic having the least chance. On the other hand, if the latter interpretation is correct then for that hypothesis the "power" results in Table 8 indicate that most of the tests, with the exception of the modified GW statistic, are actually

[7] See Clark and McCracken (2006) for more on this interpretation of the predictive content of the Phillip's curve for inflation.

oversized relative to that null hypothesis. Notably, when P is large relative to R—precisely the conditions under which the statistic is designed to be used—the modified GW statistic has "power" of roughly 5% and hence is correctly "sized" under the latter interpretation.

6. Conclusion

In this paper we provide the null asymptotic distributions of three statistics commonly used to test for equal predictive ability between two nested parametric regression models. The asymptotic null distributions of these three statistics are non-standard. Numerically, calculated critical values are provided so that asymptotically valid tests of equal predictive ability can be constructed. Monte Carlo and empirical evidence suggests that the critical values can be used to provide accurately sized and powerful tests of predictive ability.

A number of questions still remain concerning out-of-sample tests of Granger causality. As mentioned in the introduction the methods discussed here do not allow for nonparametric or semi-parametric forecasts nor do they permit the use of loss functions, such as absolute percentile loss, that are not twice continuously differentiable. Although these methods and loss functions can be incorporated into the procedures of Giacomini and White (2005) they impose the additional restriction that the forecasts be constructed using the rolling scheme with a fixed finite observation window. Since the recursive scheme is also prevalent, extending our results to less smooth loss functions and forecasts constructed using semi- and non-parametric methods would seem an important line of future work. Finally, the results of this paper require the absence of unit roots. Corradi et al. (2001) permit such when comparing the relative predictive ability of two models but require that they be nonnested. Since VARs with unit roots and their equilibrium correction equivalents are often used for forecasting, extending our results to predictive models that permit unit roots also seems an important line of future work.

Acknowledgements

We gratefully acknowledge the helpful comments of: Todd Clark, Ken West, Lutz Kilian, Norm Swanson, Bruce Hansen, Dek Terrell, the Associate Editor and two anonymous referees; seminar participants at UNLV, Queens University, University of Western Ontario, LSU, University of Missouri, 2000 Midwest Econometrics Group, 2000 World Congress of the Econometric Society and the 2000 NBER Summer Institute seminar on forecasting and time series methods. We would also like to thank LSU and the University of Missouri for providing financial support during work on a substantial portion of this paper.

Appendix A

In this appendix we provide proofs of Theorems 3.1–3.3 as well as three preliminary lemmas. In addition to the notation within the text let $\sup_t$ denote $\sup_{R\leqslant t\leqslant T}$, $B_i = (\mathrm{E}q_{i,t})^{-1}$ and for a $(k_2 \times k)$ matrix $\tilde{A}$ satisfying $\tilde{A}'\tilde{A} = B_2^{-1/2}(-JB_1J' + B_2)B_2^{-1/2}$, $\tilde{h}_{2,t+1} = c^{-1/2}\tilde{A}'B_2^{1/2}h_{2,t+1}$ and $\tilde{H}_2(t) = c^{-1/2}\tilde{A}'B_2^{1/2}H_2(t)$. Let $W(s)$ denote a $(k_2 \times 1)$ vector standard Brownian motion and let $\Rightarrow$ denote weak convergence. Throughout $\tilde{\beta}_{i,t}$ $\quad i = 1, 2$ denotes a $(k_i \times 1)$ vector on the closed cube with opposing vertices $\hat{\beta}_{i,t}$ and β_i^*. Note also that we maintain the existence of a known estimator $\hat{c}\rightarrow_{\mathrm{p}} c$. Such an estimator can be shown to be consistent using the results in either West (1996) or McCracken (2000a). Finally, note that since Assumptions 2′ and 3′ are stronger than Assumptions 2 and 3, Lemmas and Theorems that maintain 2′–3′ can reference those using 2–3 when necessary. Additional detail for the proofs can be found in a not-for-publication technical appendix McCracken (2005).

Lemma A1. *Let Assumptions* 1–4 *hold. For* $s \in [\lambda, 1]$, (a) $T^{-1/2}\sum_{j=1}^{t}\tilde{h}_{2,j} \Rightarrow W(s)$, (b) $(T/t)T^{-1/2}\sum_{j=1}^{t}\tilde{h}_{2,j} \Rightarrow s^{-1}W(s)$, (c) $(T/R)T^{-1/2}\sum_{j=t-R+1}^{t}\tilde{h}_{2,j} \Rightarrow \lambda^{-1}\{W(s) - W(s-\lambda)\}$.

Proof of Lemma A1. (a) Given Assumption 3 and the fact that $T^{-1}\mathrm{E}(\sum_{j=1}^{T}\tilde{h}_{2,j})(\sum_{j=1}^{T}\tilde{h}'_{2,j}) \to I$, the result follows from Corollary 29.19 of Davidson (1994). (b) Given (a) the result follows from the Continuous Mapping Theorem. (c) That $T/R \to \lambda^{-1}$ is immediate. Write $T^{-1/2}\sum_{j=t-R+1}^{t}\tilde{h}_{2,j}$ as $T^{-1/2}\sum_{j=1}^{t}\tilde{h}_{2,j} - T^{-1/2}\sum_{j=1}^{t-R}\tilde{h}_{2,j}$. That $T^{-1/2}\sum_{j=1}^{t}\tilde{h}_{2,j} \Rightarrow W(s)$ follows from (a). For the second piece, if we define $s' = s - \lambda$ then Corollary 29.19 of Davidson (1994) implies $T^{-1/2}\sum_{j=1}^{t-R}\tilde{h}_{2,j} \Rightarrow W(s')$. □

Lemma A2. *Let Assumptions* 1–4 *hold.* $\sum_{t=R}^{T}\tilde{H}'_2(t)\tilde{h}_{2,t+1} \to_d \Gamma_1$, *where* Γ_1 *equals* $\int_\lambda^1 s^{-1}W'(s)\,\mathrm{d}W(s)$, $\lambda^{-1}\{W(1) - W(\lambda)\}'W(\lambda)$ *and* $\lambda^{-1}\int_\lambda^1\{W(s) - W(s-\lambda)\}'\,\mathrm{d}W(s)$ *for the recursive, fixed and rolling schemes, respectively.*

Proof of Lemma A2. The result is simple for the fixed. For the rolling and recursive write $\sum_{t=R}^{T}\tilde{H}'_2(t)\tilde{h}_{2,t+1}$ as $(T/R)\sum_{t=R}^{T}(T^{-1/2}\sum_{j=1}^{t}\tilde{h}_{2,j} - T^{-1/2}\sum_{j=1}^{t-R}\tilde{h}_{2,j})'(T^{-1/2}\tilde{h}_{2,t+1})$ and $\sum_{t=R}^{T}(T/t)(T^{-1/2}\sum_{j=1}^{t}\tilde{h}_{2,j})'(T^{-1/2}\tilde{h}_{2,t+1})$, respectively. Given Assumption 2 and Lemma A1 the results follow from Corollary 29.19 of Davidson (1994) and Theorem 3.1 of Hansen (1992). □

Lemma A3. *Let Assumptions* 1–4 *hold.* $\sum_{t=R}^{T}\tilde{H}'_2(t)\tilde{H}_2(t) \to_d \Gamma_2$ *where* Γ_2 *equals* $\int_\lambda^1 s^{-2}W'(s)W(s)\,\mathrm{d}s$, $\lambda^{-2}\int_\lambda^1\{W(s) - W(s-\lambda)\}'\{W(s) - W(s-\lambda)\}\,\mathrm{d}s$ *and* $\pi\lambda^{-1}W'(\lambda)W(\lambda)$ *for the recursive, rolling and fixed schemes, respectively.*

Proof of Lemma A3. The result is immediate for the fixed. Given Lemma A1 the results for the recursive and rolling follow from the Continuous Mapping Theorem. □

Proof of Theorem 3.1. That $\sum_{t=R}^{T}(L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})) \to_d c(\Gamma_1 - 0.5\Gamma_2)$ and $P\bar{d}^2 = o_p(1)$ follows from Theorem 3.2. Given the Continuous Mapping Theorem it suffices to show that $\sum_{t=R}^{T}(L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t}))^2 = c^2\sum_{t=R}^{T}\tilde{H}'_2(t)\tilde{H}_2(t) + o_p(1)$. Expanding $L_{1,t+1}(\hat{\beta}_{1,t})$ and $L_{2,t+1}(\hat{\beta}_{2,t})$ around β_1^* and β_2^*, respectively we obtain

$$
\begin{aligned}
&\sum_{t=R}^{T}(L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t}))^2 \\
&= \left[\sum_{t=R}^{T} H'_1(t)B_1(t)h_{1,t+1}h'_{1,t+1}B_1(t)H_1(t) - 2\sum_{t=R}^{T} H'_1(t)B_1(t)h_{1,t+1}h'_{2,t+1}B_2(t)H_2(t)\right. \\
&\quad \left. + \sum_{t=R}^{T} H'_2(t)B_2(t)h_{2,t+1}h'_{2,t+1}B_2(t)H_2(t)\right] \\
&\quad + \left[-\sum_{t=R}^{T}(\hat{\beta}_{1,t} - \beta_1^*)'h_{1,t+1}\mathrm{vec}(q_{1,t}(\tilde{\beta}_{1,t}))'((\hat{\beta}_{1,t} - \beta_1^*) \otimes (\hat{\beta}_{1,t} - \beta_1^*))\right.
\end{aligned}
$$

$$
\begin{aligned}
&+ \sum_{t=R}^{T} (\hat\beta_{1,t} - \beta_1^*)' h_{1,t+1} \mathrm{vec}(q_{2,t}(\tilde\beta_{2,t}))'((\hat\beta_{2,t} - \beta_2^*) \otimes (\hat\beta_{2,t} - \beta_2^*)) \\
&+ \sum_{t=R}^{T} (\hat\beta_{2,t} - \beta_2^*)' h_{2,t+1} \mathrm{vec}(q_{1,t}(\tilde\beta_{1,t}))'((\hat\beta_{1,t} - \beta_1^*) \otimes (\hat\beta_{1,t} - \beta_1^*)) \\
&- \sum_{t=R}^{T} (\hat\beta_{2,t} - \beta_2^*)' h_{2,t+1} \mathrm{vec}(q_{2,t}(\tilde\beta_{2,t}))'((\hat\beta_{2,t} - \beta_2^*) \otimes (\hat\beta_{2,t} - \beta_2^*)) \\
&+ (0.25) \sum_{t=R}^{T} ((\hat\beta_{1,t} - \beta_1^*)' \otimes (\hat\beta_{1,t} - \beta_1^*)') \mathrm{vec}(q_{1,t}(\tilde\beta_{1,t})) \mathrm{vec}(q_{1,t}(\tilde\beta_{1,t}))'((\hat\beta_{1,t} - \beta_1^*) \otimes (\hat\beta_{1,t} - \beta_1^*)) \\
&- (0.5) \sum_{t=R}^{T} ((\hat\beta_{1,t} - \beta_1^*)' \otimes (\hat\beta_{1,t} - \beta_1^*)') \mathrm{vec}(q_{1,t}(\tilde\beta_{1,t})) \mathrm{vec}(q_{2,t}(\tilde\beta_{2,t}))'((\hat\beta_{2,t} - \beta_2^*) \otimes (\hat\beta_{2,t} - \beta_2^*)) \\
&+ (0.25) \sum_{t=R}^{T} ((\hat\beta_{2,t} - \beta_2^*)' \otimes (\hat\beta_{2,t} - \beta_2^*)') \mathrm{vec}(q_{2,t}(\tilde\beta_{2,t})) \mathrm{vec}(q_{2,t}(\tilde\beta_{2,t}))'((\hat\beta_{2,t} - \beta_2^*) \otimes (\hat\beta_{2,t} - \beta_2^*)) \Bigg] \\
&= \Bigg[\sum_{t=R}^{T} H_1'(t) B_1 \mathrm{E}(h_{1,t+1} h_{1,t+1}') B_1 H_1(t) - 2 \sum_{t=R}^{T} H_1'(t) B_1 \mathrm{E}(h_{1,t+1} h_{2,t+1}') B_2 H_2(t) \\
&+ \sum_{t=R}^{T} H_2'(t) B_2 \mathrm{E}(h_{2,t+1} h_{2,t+1}') B_2 H_2(t) \Bigg] + \mathrm{o_p}(1) \\
&= c^2 \sum_{t=R}^{T} \tilde H_2'(t) \tilde H_2(t) + \mathrm{o_p}(1) + \mathrm{o_p}(1), \qquad \text{(A11)}
\end{aligned}
$$

where the second equality follows from algebra along the lines of that in West (1996) and Clark and McCracken (2001) and the final equality uses the fact that $h_{1,t+1} = Jh_{2,t+1}$ and $\mathrm{E}(h_{i,t+1}h'_{i,t+1}) = cB_i^{-1}$ as well as the definition of $\tilde h_{2,t+1}$. The result follows by Lemma A3.

(b) We begin by providing expansions for the numerator and denominator of the OOS-t. Note that Theorem 3.2 (b) implies $P\bar d^2 = \mathrm{o_p}(P/R)$ and hence $\sum_{t=R}^{T}(L_{1,t+1}(\hat\beta_{1,t}) - L_{2,t+1}(\hat\beta_{2,t}) - \bar d)^2 = \sum_{t=R}^{T}(L_{1,t+1}(\hat\beta_{1,t}) - L_{2,t+1}(\hat\beta_{2,t}))^2 + \mathrm{o_p}(P/R)$. For both the numerator and denominator, if we expand $L_{1,t+1}(\hat\beta_{1,t})$ and $L_{2,t+1}(\hat\beta_{2,t})$ around β_1^* and β_2^*, respectively, and apply the identity $(\hat\beta_{i,t} - \beta_i^*) = (\hat\beta_{i,t} - \hat\beta_{i,R}) + (\hat\beta_{i,R} - \beta_i^*)$, $i = 1, 2$, we obtain

$$
\begin{aligned}
&\sum_{t=R}^{T} (L_{1,t+1}(\hat\beta_{1,t}) - L_{2,t+1}(\hat\beta_{2,t})) = \sum_{t=R}^{T} \{h_{2,t+1}'[-JB_1(R)J' + B_2(R)]H_2(R)\} + \sum_{t=R}^{T} \{-h_{1,t+1}'(\hat\beta_{1,t} - \hat\beta_{1,R}) \\
&\quad + h_{2,t+1}'(\hat\beta_{2,t} - \hat\beta_{2,R}) + (0.5)(\hat\beta_{1,t} - \beta_1^*)' q_{1,t}(\tilde\beta_{1,t})(\hat\beta_{1,t} - \beta_1^*) - (0.5)(\hat\beta_{2,t} - \beta_2^*)' q_{2,t}(\tilde\beta_{2,t})(\hat\beta_{2,t} - \beta_2^*)\} \\
&= c(P/R)^{1/2}[R^{1/2}\tilde H_2'(R)][P^{-1/2} \sum_{t=R}^{T} \tilde h_{2,t+1}] + \mathrm{o_p}((P/R)^{1/2}) \\
&= (P/R)^{1/2}[P^{-1/2} \sum_{t=R}^{T} h_{2,t+1}'][-JB_1J' + B_2][R^{1/2}H_2(R)] + \mathrm{o_p}((P/R)^{1/2})
\end{aligned}
$$

and

$$\sum_{t=R}^{T}(L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t}))^2 = \sum_{t=R}^{T}\{h'_{2,t+1}[-JB_1(R)J' + B_2(R)]H_2(R)$$
$$- h'_{1,t+1}(\hat{\beta}_{1,t} - \hat{\beta}_{1,R}) + h'_{2,t+1}(\hat{\beta}_{2,t} - \hat{\beta}_{2,R}) + (0.5)(\hat{\beta}_{1,t} - \beta_1^*)'q_{1,t}(\tilde{\beta}_{1,t})(\hat{\beta}_{1,t} - \beta_1^*)$$
$$- (0.5)(\hat{\beta}_{2,t} - \beta_2^*)'q_{2,t}(\tilde{\beta}_{2,t})(\hat{\beta}_{2,t} - \beta_2^*)\}^2$$
$$= (P/R)[R^{1/2}H'_2(R)][(-JB_1J' + B_2)E(h_{2,t+1}h'_{2,t+1})$$
$$\times(-JB_1J' + B_2)][R^{1/2}H_2(R)] + o_p(P/R)$$
$$= c^2(P/R)[R^{1/2}\tilde{H}'_2(R)][R^{1/2}\tilde{H}_2(R)] + o_p(P/R).$$

In each of the above expansions, the second equality follows from algebra along the lines of that in West (1996) and Clark and McCracken (2001) and the fact that given Assumption 4′, $\sup_t(R^{1/2}|\hat{\beta}_{i,t} - \hat{\beta}_{i,R}|) = o_p(1)$. The final equality uses the fact that $h_{1,t+1} = Jh_{2,t+1}$, $E(q_{i,t}) = B_i^{-1}$ and $E(h_{i,t+1}h'_{i,t+1}) = cB_i^{-1}$ as well as the definition of $\tilde{h}_{2,t+1}$.

Given the above expansions we can write OOS-t as

$$\text{OOS-}t = \frac{(P/R)^{1/2}[R^{1/2}\tilde{H}'_2(R)][P^{-1/2}\sum_{t=R}^{T}\tilde{h}_{2,t+1}] + o_p((P/R)^{1/2})}{[(P/R)[R^{1/2}\tilde{H}'_2(R)][R^{1/2}\tilde{H}_2(R)] + o_p(P/R)]^{1/2}}$$
$$= \frac{[R^{1/2}\tilde{H}'_2(R)][P^{-1/2}\sum_{t=R}^{T}\tilde{h}_{2,t+1}] + o_p(1)}{[[R^{1/2}\tilde{H}'_2(R)][R^{1/2}\tilde{H}_2(R)] + o_p(1)]^{1/2}}$$
$$= \frac{[R^{1/2}\tilde{H}'_2(R)][P^{-1/2}\sum_{t=R}^{T}\tilde{h}_{2,t+1}]}{[[R^{1/2}\tilde{H}'_2(R)][R^{1/2}\tilde{H}_2(R)]]^{1/2}} + o_p(1).$$

Given Assumption 3, Corollary 29.19 of Davidson (1994) suffices for $(P^{-1/2}\sum_{t=R}^{T}\tilde{h}'_{2,t+1}, R^{1/2}\tilde{H}'_2(R))' \to_d (V'_1, V'_0)'$ for independent $(k_i \times 1)$ standard normal vectors V_0 and V_1. The result follows immediately from the Continuous Mapping Theorem.

Proof of Theorem 3.2. (a) Expanding $L_{1,t+1}(\hat{\beta}_{1,t})$ and $L_{2,t+1}(\hat{\beta}_{2,t})$ around β_1^* and β_2^*, respectively we obtain

$$\sum_{t=R}^{T}(L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})) = \sum_{t=R}^{T}\{-h'_{1,t+1}B_1(t)H_1(t) + h'_{2,t+1}B_2(t)H_2(t)\}$$
$$- (0.5)\sum_{t=R}^{T}\{-H'_1(t)B_1(t)q_{1,t}(\tilde{\beta}_{1,t})B_1(t)H_1(t) + H'_2(t)B_2(t)q_{2,t}(\tilde{\beta}_{2,t})B_2(t)H_2(t)\}$$
$$= \sum_{t=R}^{T}\{-h'_{1,t+1}B_1H_1(t) + h'_{2,t+1}B_2H_2(t)\}$$
$$- (0.5)\sum_{t=R}^{T}\{-H'_1(t)B_1E(q_{1,t})B_1H_1(t) + H'_2(t)B_2E(q_{2,t})B_2H_2(t)\} + o_p(1)$$
$$+ c\sum_{t=R}^{T}\tilde{H}'_2(t)\tilde{h}_{2,t+1} - (0.5)c\sum_{t=R}^{T}\tilde{H}'_2(t)\tilde{H}_2(t) + o_p(1), \tag{A10}$$

where the second equality follows from algebra along the lines of that in West (1996) and Clark and McCracken (2001) and the final equality uses the fact that $h_{1,t+1} = Jh_{2,t+1}$ and $\mathrm{E}(q_{i,t}) = B_i^{-1}$ as well as the definition of $\tilde{h}_{2,t+1}$. The result follows from Lemmas A2, A3 and the fact that $\hat{c} \to_{\mathrm{p}} c$.

(b) If we expand $L_{1,t+1}(\hat{\beta}_{1,t})$ and $L_{2,t+1}(\hat{\beta}_{2,t})$ around β_1^* and β_2^*, respectively and apply the identity $(\hat{\beta}_{i,t} - \beta_i^*) = (\hat{\beta}_{i,t} - \hat{\beta}_{i,R}) + (\hat{\beta}_{i,R} - \beta_i^*)$ $i = 1, 2$, we obtain

$$
\begin{aligned}
\sum_{t=R}^{T}(L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})) &= \sum_{t=R}^{T}\{h'_{2,t+1}[-JB_1(R)J' + B_2(R)]H_2(R)\} \\
&\quad + \sum_{t=R}^{T}\{-h'_{1,t+1}(\hat{\beta}_{1,t} - \hat{\beta}_{1,R}) + h'_{2,t+1}(\hat{\beta}_{2,t} - \hat{\beta}_{2,R}) + (0.5)(\hat{\beta}_{1,t} - \beta_1^*)'q_{1,t}(\tilde{\beta}_{1,t})(\hat{\beta}_{1,t} - \beta_1^*) \\
&\quad - (0.5)(\hat{\beta}_{2,t} - \beta_2^*)'q_{2,t}(\tilde{\beta}_{2,t})(\hat{\beta}_{2,t} - \beta_2^*)\} \\
&= \sum_{t=R}^{T}\{h'_{2,t+1}[-JB_1J' + B_2]H_2(R)\} + \mathrm{o_p}((P/R)^{1/2}) \\
&= c(P/R)^{1/2}[R^{1/2}\tilde{H}'_2(R)]\left[P^{-1/2}\sum_{t=R}^{T}\tilde{h}_{2,t+1}\right] + \mathrm{o_p}((P/R)^{1/2}).
\end{aligned}
$$

The second equality follows from algebra along the lines of that in West (1996) and Clark and McCracken (2001) and the fact that given Assumption 4′, $\sup_t(R^{1/2}|\hat{\beta}_{i,t} - \hat{\beta}_{i,R}|) = \mathrm{o_p}(1)$. The final equality uses the fact that $h_{1,t+1} = Jh_{2,t+1}$ and $\mathrm{E}(q_{i,t}) = B_i^{-1}$ as well as the definition of $\tilde{h}_{2,t+1}$. Given Assumption 3, Corollary 29.19 of Davidson (1994) suffices for $(P^{-1/2}\sum_{t=R}^{T}\tilde{h}'_{2,t+1}, R^{1/2}\tilde{H}'_2(R))' \to_{\mathrm{d}} (V'_1, V'_0)'$ for the independent $(k_i \times 1)$ standard normal vectors V_0 and V_1 from Theorem 3.1. The result then follows from the fact that $\hat{c} \to_{\mathrm{p}} c$.

Proof of Theorem 3.3. The proof is conducted in two stages. In (a) we maintain Assumption 4 while in (b) we maintain Assumption 4′. Note also that the $(P-1)$ term in Eq. (2) of the text is replaced with P without consequence for the asymptotics. □

(a) We begin by providing expansions for both the numerator and denominator of the MSE-Reg. For the numerator note that Theorem 3.2 (a) implies $\sum_{t=R}^{T}(\hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2) = c\sum_{t=R}^{T}\tilde{H}'_2(t)\tilde{h}_{2,t+1} - (0.5)c\sum_{t=R}^{T}\tilde{H}'_2(t)\tilde{H}_2(t) + \mathrm{o_p}(1)$. For the denominator first note that Theorem 3.2 (a) implies $P\bar{d}^2 = \mathrm{o_p}(1)$ while Theorem 4.1 of West (1996) implies $P^{-1}\sum_{t=R}^{T}(\hat{u}_{1,t+1} + \hat{u}_{2,t+1})^2 \to_{\mathrm{p}} 4\sigma^2 = 2c$. If we then expand $\hat{u}_{1,t+1}$ and $\hat{u}_{2,t+1}$ around β_1^* and β_2^*, respectively we obtain

$$
\begin{aligned}
&\sum_{t=R}^{T}(\hat{u}_{1,t+1} - \hat{u}_{2,t+1})^2 \\
&= \left[\sum_{t=R}^{T} H'_1(t)B_1(t)\nabla\hat{y}_{1,t+1}(\beta_1^*)\nabla\hat{y}'_{1,t+1}(\beta_1^*)B_1(t)H_1(t)\right. \\
&\quad - 2\sum_{t=R}^{T} H'_1(t)B_1(t)\nabla\hat{y}_{1,t+1}(\beta_1^*)\nabla\hat{y}'_{2,t+1}(\beta_2^*)B_2(t)H_2(t)
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{t=R}^{T} H_2'(t)B_2(t)\nabla\hat{y}_{2,t+1}(\beta_2^*)\nabla\hat{y}_{2,t+1}'(\beta_2^*)B_2(t)H_2(t)\Bigg]\\
&+\Bigg[-\sum_{t=R}^{T}(\hat{\beta}_{1,t}-\beta_1^*)'\nabla\hat{y}_{1,t+1}(\beta_1^*)\mathrm{vec}(\nabla^2\hat{y}_{1,t+1}(\tilde{\beta}_{1,t}))'((\hat{\beta}_{1,t}-\beta_1^*)\otimes(\hat{\beta}_{1,t}-\beta_1^*))\\
&+\sum_{t=R}^{T}(\hat{\beta}_{1,t}-\beta_1^*)'\nabla\hat{y}_{1,t+1}(\beta_1^*)\mathrm{vec}(\nabla^2\hat{y}_{2,t+1}(\tilde{\beta}_{2,t}))'((\hat{\beta}_{2,t}-\beta_2^*)\otimes(\hat{\beta}_{2,t}-\beta_2^*))\\
&+\sum_{t=R}^{T}(\hat{\beta}_{2,t}-\beta_2^*)'\nabla\hat{y}_{2,t+1}(\beta_2^*)\mathrm{vec}(\nabla^2\hat{y}_{1,t+1}(\tilde{\beta}_{1,t}))'((\hat{\beta}_{1,t}-\beta_1^*)\otimes(\hat{\beta}_{1,t}-\beta_1^*))\\
&-\sum_{t=R}^{T}(\hat{\beta}_{2,t}-\beta_2^*)'\nabla\hat{y}_{2,t+1}(\beta_2^*)\mathrm{vec}(\nabla^2\hat{y}_{2,t+1}(\tilde{\beta}_{1,t}))'((\hat{\beta}_{2,t}-\beta_2^*)\otimes(\hat{\beta}_{2,t}-\beta_2^*))\\
&+(0.25)\sum_{t=R}^{T}((\hat{\beta}_{1,t}-\beta_1^*)'\otimes(\hat{\beta}_{1,t}-\beta_1^*)')\mathrm{vec}(\nabla^2\hat{y}_{1,t+1}(\tilde{\beta}_{1,t}))\\
&\times\mathrm{vec}(\nabla^2\hat{y}_{1,t+1}(\tilde{\beta}_{1,t}))'((\hat{\beta}_{1,t}-\beta_1^*)\otimes(\hat{\beta}_{1,t}-\beta_1^*))\\
&-(0.5)\sum_{t=R}^{T}((\hat{\beta}_{1,t}-\beta_1^*)'\otimes(\hat{\beta}_{1,t}-\beta_1^*)')\mathrm{vec}(\nabla^2\hat{y}_{1,t+1}(\tilde{\beta}_{1,t}))\\
&\times\mathrm{vec}(\nabla^2\hat{y}_{2,t+1}(\tilde{\beta}_{2,t}))'((\hat{\beta}_{2,t}-\beta_2^*)\otimes(\hat{\beta}_{2,t}-\beta_2^*))\\
&+(0.25)\sum_{t=R}^{T}((\hat{\beta}_{2,t}-\beta_2^*)'\otimes(\hat{\beta}_{2,t}-\beta_2^*)')\mathrm{vec}(\nabla^2\hat{y}_{2,t+1}(\tilde{\beta}_{2,t}))\\
&\times\mathrm{vec}(\nabla^2\hat{y}_{2,t+1}(\tilde{\beta}_{2,t}))'((\hat{\beta}_{2,t}-\beta_2^*)\otimes(\hat{\beta}_{2,t}-\beta_2^*))\\
=&\Bigg[\sum_{t=R}^{T} H_1'(t)B_1\mathrm{E}(\nabla\hat{y}_{1,t+1}(\beta_1^*)\nabla\hat{y}_{1,t+1}'(\beta_1^*))B_1H_1(t)\\
&-2\sum_{t=R}^{T} H_1'(t)B_1\mathrm{E}(\nabla\hat{y}_{1,t+1}(\beta_1^*)\nabla\hat{y}_{2,t+1}'(\beta_2^*))B_2H_2(t)\\
&+\sum_{t=R}^{T} H_2'(t)B_2\mathrm{E}(\nabla\hat{y}_{2,t+1}(\beta_2^*)\nabla\hat{y}_{2,t+1}'(\beta_2^*))B_2H_2(t)\Bigg]+\mathrm{o_p}(1)\\
=&(0.5)c\sum_{t=R}^{T}\tilde{H}_2'(t)\tilde{H}_2(t)+\mathrm{o_p}(1), \qquad \text{(A.12)}
\end{aligned}
$$

where the second equality follows from algebra along the lines of that in West (1996) and Clark and McCracken (2001) and the final equality uses the fact that $h_{1,t+1}=Jh_{2,t+1}$ and $\mathrm{E}\nabla\hat{y}_{i,t+1}(\beta_i^*)\nabla\hat{y}_{i,t+1}'(\beta_i^*)=(0.5)B_i^{-1}$ as well as the definition of $\tilde{h}_{2,t+1}$. The denominator term then satisfies $[P^{-1}\sum_{t=R}^{T}(\hat{u}_{1,t+1}+\hat{u}_{2,t+1})^2][\sum_{t=R}^{T}(\hat{u}_{1,t+1}-\hat{u}_{2,t+1})^2]-(P^{1/2}\bar{d})^2=c^2\sum_{t=R}^{T}\tilde{H}_2'(t)$ $\tilde{H}_2(t)+\mathrm{o_p}(1)$. Since each of the numerator and denominator expansions are identical to those for Theorem 3.1 (a), the Continuous Mapping Theorem provides the desired result.

(b) We begin by providing expansions for both the numerator and denominator of the MSE-Reg. For the numerator note that Theorem 3.2 (b) implies $\sum_{t=R}^{T}(\hat{u}_{1,t+1}^2-\hat{u}_{2,t+1}^2)=$

$c(P/R)^{1/2}[R^{1/2}\tilde{H}_2'(R)][P^{-1/2}\sum_{t=R}^{T}\tilde{h}_{2,t+1}] + o_p((P/R)^{1/2})$. For the denominator first note that Theorem 3.2 (b) implies $P\bar{d}^2 = o_p(P/R)$ while Theorem 4.1 of West (1996) implies $P^{-1}\sum_{t=R}^{T}(\hat{u}_{1,t+1} + \hat{u}_{2,t+1})^2 \to_p 4\sigma^2 = 2c$. If we then expand $\hat{u}_{1,t+1}$ and $\hat{u}_{2,t+1}$ around β_1^* and β_2^*, respectively, then we obtain

$$\begin{aligned}
\sum_{t=R}^{T}(\hat{u}_{1,t+1} - \hat{u}_{2,t+1})^2 &= \sum_{t=R}^{T}\{\nabla\hat{y}_{2,t+1}'(\beta_2^*)[-JB_1(R)J' + B_2(R)]H_2(R) - \nabla\hat{y}_{1,t+1}'(\beta_1^*)(\hat{\beta}_{1,t} - \hat{\beta}_{1,R}) \\
&\quad + \nabla\hat{y}_{2,t+1}'(\beta_2^*)(\hat{\beta}_{2,t} - \hat{\beta}_{2,R}) + (0.5)(\hat{\beta}_{1,t} - \beta_1^*)'\nabla^2\hat{y}_{1,t+1}(\beta_1^*) \\
&\quad \times(\hat{\beta}_{1,t} - \beta_1^*) - (0.5)(\hat{\beta}_{2,t} - \beta_2^*)'\nabla^2\hat{y}_{2,t+1}(\beta_2^*)(\hat{\beta}_{2,t} - \beta_2^*)\}^2 \\
&= (P/R)[R^{1/2}H_2'(R)][(-JB_1J' + B_2)\mathrm{E}(\nabla\hat{y}_{2,t+1}(\beta_2^*)\nabla\hat{y}_{2,t+1}'(\beta_2^*)) \\
&\quad \times(-JB_1J' + B_2)][R^{1/2}H_2(R)] + o_p(P/R) \\
&= (0.5)c(P/R)[R^{1/2}\tilde{H}_2'(R)][R^{1/2}\tilde{H}_2(R)] + o_p(P/R),
\end{aligned}$$

where the second equality follows from algebra along the lines of that in West (1996) and the final equality uses the fact that $h_{1,t+1} = Jh_{2,t+1}$ and $\mathrm{E}\nabla\hat{y}_{i,t+1}(\beta_i^*)\nabla\hat{y}_{i,t+1}'(\beta_i^*) = (0.5)B_i^{-1}$ as well as the definition of $\tilde{h}_{2,t+1}$. The denominator term then satisfies $[P^{-1}\sum_{t=R}^{T}(\hat{u}_{1,\ t+1} + \hat{u}_{2,t+1})^2][\sum_{t=R}^{T}(\hat{u}_{1,t+1} - \hat{u}_{2,t+1})^2] - (P^{1/2}\bar{d})^2 = c^2(P/R)[R^{1/2}\tilde{H}_2'(R)][R^{1/2}\tilde{H}_2(R)] + o_p(P/R)$. Since each of the numerator and denominator expansions are identical to those for Theorem 3.1 (b), the Continuous Mapping Theorem provides the desired result.

References

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

Ashley, R., Granger, C.W.J., Schmalensee, R., 1980. Advertising and aggregate consumption: an analysis of causality. Econometrica 48, 1149–1167.

Brave, S., Fisher, J.D.M., 2004. In search of a robust inflation forecast, economic perspectives (federal reserve bank of Chicago). Fourth Quarter 0, 12–13.

Breen, W., Glosten, L.R., Jagannathan, R., 1989. Economic significance of predictable variations in stock index returns. Journal of Finance 44, 1177–1189.

Chao, J., Corradi, V., Swanson, N.R., 2001. An out of sample test for granger causality. Macroeconomic Dynamics 5, 598–620.

Chong, Y.Y., Hendry, D.F., 1986. Econometric evaluation of linear macro-economic models. Review of Economic Studies 53, 671–690.

Clark, T.E., 2004. Can out-of-sample forecast comparisons help prevent overfitting? Journal of Forecasting 23, 115–139.

Clark, T.E., McCracken, M.W., 2001. Tests of equal forecast accuracy and encompassing for nested models. Journal of Econometrics 105, 85–110.

Clark, T.E., McCracken, M.W., 2005a. Evaluating long horizon forecasts. Econometric Reviews 24, 369–404.

Clark, T.E., McCracken, M.W., 2005b. The power of tests of predictive ability in the presence of structural breaks. Journal of Econometrics 124, 1–31.

Clark, T.E., McCracken, M.W., 2005c. Forecasting with weakly nested models. Federal Reserve Bank of Kansas City and the Federal Reserve Board of Governors.

Clark, T.E., McCracken, M.W., 2006. The predictive content of the output gap for inflation: resolving in–sample and out–of–sample evidence. Journal of Money, Credit, and Banking 38, 1127–1148.

Clements, M.P., 2002. Why forecast performance does not help us choose a model. University of Warwick.

Corradi, V., Swanson, N.R., 2002. A consistent test for nonlinear out of sample predictive accuracy. Journal of Econometrics 110, 353–381.

Corradi, V., Swanson, N.R., 2006. Nonparametric bootstrap procedures for predictive inference based on recursive estimation schemes. International Economic Review forthcoming.

Corradi, V., Swanson, N.R., Olivetti, C., 2001. Predictive ability with cointegrated variables. Journal of Econometrics 104, 315–358.

Davidson, J., 1994. Stochastic Limit Theory. Oxford University Press, New York.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–263.

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

Diebold, F.X., Gunther, T.A., Tay, A.S., 1998. Evaluating density forecasts, with applications to financial risk management. International Economic Review 39, 863–883.

Fair, R.C., Shiller, R., 1989. The informational content of ex ante forecasts. Review of Economics and Statistics 71, 325–331.

Fair, R.C., Shiller, R., 1990. Comparing information in forecasts from econometric models. American Economic Review 80, 375–389.

Ghysels, E., Hall, A., 1990. A test for structural stability of euler conditions with parameters estimated via the generalized method of moments estimator. International Economic Review 31, 355–364.

Giacomini, R., White, H., 2005. Tests of Conditional Predictive Ability. University of California, San Diego.

Granger, C.W.J., 1999. Outline of forecast theory using generalized cost functions. Spanish Economic Review 1, 161–173.

Granger, C.W.J., Newbold, P., 1977. Forecasting Economic Time Series. Academic Press Inc, London.

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

Hansen, P.R., 2005. A test for superior predictive ability. Journal of Business and Economic Statistics 23, 365–380.

Harvey, D.I., Leybourne, S.J., Newbold, P., 1999. Forecast evaluation tests in the presence of ARCH. Journal of Forecasting 18, 435–445.

Hong, Y., Lee, T., 2003. Inference on predictability of foreign exchange rates via generalized spectrum and nonlinear time series models. Review of Economics and Statistics 85, 1048–1062.

Huber, P.J., 1972. The 1972 Wald lecture robust statistics: a review. Annals of Statistics 43, 1041–1067.

Inoue, A., Kilian, L., 2004. In-sample or out-of-sample tests of predictability? which one should we use? Econometric Reviews 23, 371–402.

Kilian, L., 1999. Exchange rates and monetary fundamentals: what do we learn from long-horizon regressions? Journal of Applied Econometrics 14, 491–510.

Klein, L.R., 1992. The Test of a Model is it's Ability to Predict. University of Pennsylvania.

Leitch, G., Tanner, J.E., 1991. Economic forecast evaluation: profits versus the conventional error measures. American Economic Review 81, 580–590.

Lettau, M., Ludvigson, S., 2001. Consumption, aggregate wealth, and expected stock returns. Journal of Finance 3, 815–849.

Mark, N.C., 1995. Exchange rates and fundamentals: evidence on long-horizon predictability. The American Economic Review 85, 201–218.

McCracken, M.W., 2000a. Robust out of sample inference. Journal of Econometrics 99, 195–223.

McCracken, M.W., 2000b. An out of sample nonparametric test of the martingale difference hypothesis. In: Fomby, T., Hill, C. (Eds.), Advances in Econometrics: Applying Kernel and Nonparametric Estimation to Economic Topics, vol. 14. JAI Press, Stamford, CT, pp. 49–75.

McCracken, M.W., 2005. Technical Appendix to 'Asymptotics for Out-of-sample Tests of Granger Causality'. Federal Reserve Board of Governors.

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

Mincer, J., Zarnowitz, V., 1969. The evaluation of economic forecasts. In: Mincer, J. (Ed.), Economic Forecasts and Expectations. National Bureau of Economic Research, New York.

Mizrach, B., 1992. The distribution of the theil U-statistic in bivariate normal populations. Economic Letters 38, 163–167.

Newey, W.K., McFadden, D., 1994. Large sample estimation and hypothesis testing. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, vol. IV. North-Holland, Amsterdam.

Pagan, A.R., Schwert, G.W., 1990. Alternative models for conditional stock volatility. Journal of Econometrics 45, 267–290.

Patton, A., Timmermann, A., 2003. Properties of Optimal Forecasts. University of California-San Diego.

Rapach, D.E., Wohar, M.E., 2002. Testing the monetary model of exchange rate determination: new evidence from a century of data. Journal of International Economics 58, 359–385.

Shintani, M., 2005. Nonlinear forecasting analysis using diffusion indexes: an application to Japan. Journal of Money, Credit, and Banking 37, 517–538.

Stock, J.H., Watson, M.W., 2002. Macroeconomic forecasting using diffusion indexes. Journal of Business and Economic Statistics 20, 147–162.

Sullivan, R., Timmermann, A., White, H., 1999. Data-snooping, technical trading rule performance, and the bootstrap. Journal of Finance 54, 1647–1691.

Swanson, N.R., 1998. Money and output viewed through a rolling window. Journal of Monetary Economics 41, 455–473.

Swanson, N.R., White, H., 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.

Weiss, A.A., Andersen, A.P., 1984. estimating time series using the relevant forecast evaluation criterion. Journal of the Royal Statistical Society, Series A 147, 484–487.

West, K.D., 1996. Asymptotic inference about predictive ability. Econometrica 64, 1067–1084.

West, K.D., McCracken, M.W., 1998. Regression-based tests of predictive ability. International Economic Review 39, 817–840.

White, H., 2000. A reality check for data snooping. Econometrica 68, 1067–1084.