ECONOMIC RESEARCH
FEDERAL RESERVE BANK OF ST. LOUIS

WORKING PAPER SERIES

Nested Forecast Model Comparisons: A New Approach to Testing Equal Accuracy

Authors	Todd E. Clark, and Michael W. McCracken
Working Paper Number	2009-050B
Revision Date	August 2012
Citable Link	https://doi.org/10.20955/wp.2009.050
Suggested Citation	Clark, T.E., McCracken, M.W., 2012; Nested Forecast Model Comparisons: A New Approach to Testing Equal Accuracy, Federal Reserve Bank of St. Louis Working Paper 2009-050. URL https://doi.org/10.20955/wp.2009.050

Published In	Journal of Econometrics
Publisher Link	https://doi.org/10.1016/j.jeconom.2014.06.016

Federal Reserve Bank of St. Louis, Research Division, P.O. Box 442, St. Louis, MO 63166

Nested Forecast Model Comparisons: A New Approach to Testing Equal Accuracy *

Todd E. Clark
Federal Reserve Bank of Cleveland

Michael W. McCracken
Federal Reserve Bank of St. Louis

August 2012

Abstract

This paper develops bootstrap methods for testing whether, in a finite sample, competing out-of-sample forecasts from nested models are equally accurate. Most prior work on forecast tests for nested models has focused on a null hypothesis of equal accuracy in population — basically, whether coefficients on the extra variables in the larger, nesting model are zero. We instead use an asymptotic approximation that treats the coefficients as non-zero but small, such that, in a finite sample, forecasts from the small model are expected to be as accurate as forecasts from the large model. Under that approximation, we derive the limiting distributions of tests of equal mean square error, and develop bootstrap methods for estimating critical values. Monte Carlo experiments show that our proposed procedures have good size and power properties for the null of equal finite-sample forecast accuracy. We illustrate the use of the procedures with applications to forecasting stock returns and inflation.

JEL Nos.: C53, C12, C52
Keywords: mean square error, prediction, reality check

* *Clark*: Economic Research Dept.; Federal Reserve Bank of Cleveland; P.O. Box 6387; Cleveland, OH 44101; `todd.clark@clev.frb.org`. *McCracken (corresponding author)*: Research Division; Federal Reserve Bank of St. Louis; P.O. Box 442; St. Louis, MO 63166; `michael.w.mccracken@stls.frb.org`. The research for this paper was completed while Clark was on the staff of the Federal Reserve Bank of Kansas City. We gratefully acknowledge excellent research assistance from Stephen Terry and helpful comments from Bruce Hansen, Lutz Kilian, Norm Swanson, and seminar or conference participants at the University of Kansas, University of Michigan, Rutgers University, HEC Montreal, Bank of Spain, Midwest Econometrics Group, Missouri Economics Conference, NBER Summer Institute, and the ECB/Bundesbank/CFS Joint Lunchtime Seminar. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Banks of Cleveland or St. Louis or the Federal Reserve Board of Governors.

1 Introduction

In this paper we examine the asymptotic and finite-sample properties of bootstrap-based tests of equal accuracy of out-of-sample forecasts from a baseline nested model and an alternative nesting model. In our analysis, we address two forms of the null hypothesis of equal predictive ability. One hypothesis, considered in such studies as Clark and McCracken (2001, 2005), Corradi and Swanson (2002), Inoue and Kilian (2004), and McCracken (2007), is that the models have equal population-level predictive ability. This situation arises when the coefficients associated with the additional predictors in the nesting model are zero and hence at the population level, the forecast errors are identical and thus the models have equal predictive ability.

However, this paper focuses on a different null hypothesis, one that arises when some of the additional predictors have non-zero coefficients associated with them, but the marginal predictive content is small. In this case, addressed in Trenkler and Toutenberg (1992), Hjalmarsson (2009) and Clark and McCracken (2009), the two models can have equal predictive ability at a fixed forecast origin (say time T) due to a bias-variance trade-off between a more accurately estimated, but misspecified, nested model and a correctly specified, but imprecisely estimated, nesting model. Building upon this insight, we derive the asymptotic distributions associated with standard out-of-sample tests of equal predictive ability between estimated models with weak predictors. We then develop a bootstrap-based method for imposing the null of equal predictive ability upon these distributions and conducting asymptotically valid inference. In our results, the forecast models may be estimated either recursively or with a rolling sample. Giacomini and White (2006) use a different asymptotic approximation to testing equal forecast accuracy in a given sample, but their asymptotics apply only to models estimated with a rolling window of fixed and finite width.

Our approach to modeling weak predictors is identical to the standard Pitman drift used to analyze the power of in-sample tests against small deviations from the null of equal population-level predictive ability. It has also been used by Inoue and Kilian (2004) in the context of analyzing the power of out-of-sample tests. In that sense, some (though not all) of our analytical results are quite similar to those in Inoue and Kilian (2004).

We differ, though, in our focus. While Inoue and Kilian (2004) are interested in examining the power of out-of-sample tests against the null of equal population-level predictive ability, we are interested in using out-of-sample tests to test the null hypothesis of equal finite sample predictive ability. This distinction arises because the estimation error associated with estimating unknown regression parameters can cause a misspecified, restricted model to be as accurate or more accurate than a correctly specified unrestricted model when the additional predictors are imprecisely estimated (or, in our terminology, are "weak"). We use Pitman drift simply as a tool for constructing an asymptotic approximation to the finite sample problem associated with estimating a regression coefficient when the marginal signal associated with it is small.

The lengthy literature evaluating direct, multi–step (DMS) forecasts from nested models indicates our results for these forecasts should be useful to many researchers. Applications considering DMS forecasts from nested linear models include, among others: many of the studies cited above; Diebold and Rudebusch (1991); Mark (1995); Kilian (1999); Lettau and Ludvigson (2001); Stock and Watson (2003); Bachmeier and Swanson (2005); Butler, Grullon and Weston (2005); Cooper and Gulen (2006); Giacomini and Rossi (2006); Guo (2006); Rapach and Wohar (2006); Bruneau, et al. (2007); Bordo and Haubrich (2008); Inoue and Rossi (2008); Molodtsova and Papell (2009); Chen, Rogoff, and Rossi (2010); and Ferreira and Santa-Clara (2011)

The remainder proceeds as follows. Section 2 introduces the notation, assumptions, and presents our theoretical results and bootstrap for testing the null of equal forecast accuracy in the finite sample. Section 3 presents Monte Carlo results on the finite–sample performance of the asymptotics and the bootstrap. Section 4 applies our tests to evaluate the predictability of U.S. stock returns and core PCE inflation. Section 5 concludes.

2 Theoretical results

We begin by using a simple example to illustrate our essential ideas and to clarify how our results differ from those obtained in Giacomini and White (2006). We then proceed to a more general setup. After detailing the necessary notation and test statistics, we present asymptotic results for tests of equal accuracy applied to forecasts from two nested models in the presence of weak predictive ability. The section then describes our proposed bootstrap and proves its validity. We focus all of this presentation on results for the recursive estimation and forecasting scheme. The last subsection summarizes the changes in moments and functions that apply under a rolling estimation and forecasting scheme.

2.1 A simple example

Suppose one wishes to compare the finite sample forecast accuracy of two simple nested models where accuracy is measured under quadratic loss. Model 0 is a simple no-change model and hence $\hat{y}_{0,t+1} = 0$ while Model 1 is an OLS-estimated location model and hence $\hat{y}_{0,t+1} = \bar{y}_t$ where $\bar{y}_t$ equals $T^{-1}\sum_{s=t-T+1}^{t} y_s$ or $t^{-1}\sum_{s=1}^{t} y_s$ if the rolling or recursive schemes are used, respectively. For ease of presentation assume that $y_{t+1} = \mu + u_{t+1}$ where μ is non-stochastic and u_{t+1} forms a homoskedastic martingale difference sequence. The difference in squared forecast errors associated with forecast origins $t = T, ..., T+P-1$ then takes the form

$$\hat{u}_{0,t+1}^2 - \hat{u}_{1,t+1}^2 = 2u_{t+1}\mu + \mu^2 + 2u_{t+1}H_1(t) - H_1^2(t),$$

where $H_1(t)$ equals $T^{-1}\sum_{s=t-T+1}^{t} u_s$ or $t^{-1}\sum_{s=1}^{t} u_s$ depending on whether the rolling or recursive schemes are used.

In Giacomini and White (2006) the rolling scheme is required and the null hypothesis takes the form $\lim_{P\to\infty} P^{-1/2}\sum_{t=T}^{T+P-1} E(\hat{u}_{0,t+1}^2 - \hat{u}_{1,t+1}^2) = 0$. In the context of our current example this hypothesis is equivalent to $\mu^2 - \sigma^2/T = 0$. The two sets of forecasts are equally accurate, on average, because the squared bias asociated with the misspecified restricted model 0 balances with the estimation error associated with the correctly specified, but less accurately estimated unrestricted model 1. Note that for this balance to occur the window size T used to estimate the parameters (in model 1) cannot be allowed to increase with the total sample size $T + P$. If it were allowed to do so the estimation risk component σ^2/T would converge to zero and no tradeoff could be made. Put differently, the estimation error must remain "large" under these asymptotics regardless of the sample size.

In the current paper we take a different approach that allows the estimation error to eventually become "small" in the sense that our parameter estimates converge in probability. To do so, and yet still allow a bias-variance tradeoff to exist, we model the bias as being equally small. In the context of our current example suppose that we model the unconditional mean μ as being local-to-zero such that $\mu = \mu_w/T^{1/2}$. If we then restate the null hypothesis as $\lim_{T,P\to\infty}\sum_{t=T}^{T+P-1} E(\hat{u}_{0,t+1}^2 - \hat{u}_{1,t+1}^2) = 0$, we find that this hypothesis is equivalent to $\mu_w^2 - \sigma^2 = 0$ if the rolling scheme is used and $\lambda_P\mu_w^2 - \ln(1+\lambda_P)\sigma^2 = 0$ if the recursive scheme is used.

In the following sections we generalize this asymptotic approximation to the null hypothesis of equal finite sample forecast accuracy to a broader range of nested model comparisons and to environments that permit multi-step ahead forecasts with conditionally heteroskedastic forecast errors. In all instances, for tractability, we restrict attention to linear models estimated by OLS. As we noted in the introduction, the linear-OLS framework covers a large fraction of the literature in applied forecasting and forecast evaluation.

2.2 Environment

The sample of observations $\{y_t, x'_{1,t}\}_{t=1}^{T}$ includes a scalar random variable y_t to forecast at a horizon of τ periods ahead and a $(k \times 1)$ vector of predictors $x_{1,t}$.[1] The vector of predictors contains one set of variables (denoted $x_{0,t}$, with k_0 elements) included in the null model and another distinct set of variables (denoted $x_{w,t}$, with k_w elements) with weak predictive content: $x_{1,t} = (x'_{0,t}, x'_{w,t})'$. The vector of predictors ($x_{1,t}$) may include lags of the dependent variable. At each origin of forecasting $t = T, ...T + P - \tau$, forecasts of $y_{t+\tau}$, $\tau \geq 1$, are generated from OLS-estimated linear parametric models of the form $x'_{i,t}\beta_i$, $i = 0, 1$, where the restricted model 0 includes only $x_{0,t}$ as predictors and the unrestricted model 1 includes both $x_{0,t}$ and $x_{w,t}$ as predictors. The forecast sample size is P.

The possibility of weak predictors is modeled using a sequence of linear data-generating processes (DGPs) of the form[2]

$$\begin{aligned} y_{t+\tau} &= x'_{1,t}\beta_1 + u_{t+\tau} = x'_{0,t}\beta_0 + x'_{w,t}(T^{-1/2}\beta_w) + u_{t+\tau}, \qquad (1) \\ Ex_{1,t}u_{t+\tau} &\equiv Eh_{1,t+\tau} = 0 \text{ for all } t = 1, ..., T, ...T + P - \tau. \end{aligned}$$

Turning to notation needed below, we denote the loss associated with the τ-step ahead forecast errors as $u^2_{i,t+\tau} = (y_{t+\tau} - x'_{i,t}\beta_{i,t})^2$, $i = 0, 1$, for the restricted and unrestricted models, respectively. Let $H_i(t) = (t^{-1}\sum_{s=1}^{t-\tau} x_{i,s}u_{s+\tau}) = (t^{-1}\sum_{s=1}^{t-\tau} h_{i,s+\tau})$, $B_i(t) = (t^{-1}\sum_{s=1}^{t-\tau} x_{i,s}x'_{i,s})^{-1}$, and $B_i = \lim_{T\to\infty}(Ex_{T,i,s}x'_{T,i,s})^{-1}$, for $i = 0, 1$. For $U_{T,t} = (h'_{T,1,t+\tau}, vec(x_{T,1,t}x'_{T,1,t})')'$, $\Omega_j = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E(U_{T,t}U'_{T,t-j})$ for all $j \geq 0$, $\Omega_{11,j}$ is the upper block-diagonal element of Ω_j, and $V = \sum_{j=-\tau+1}^{\tau-1}\Omega_{11,j}$. Define selection matrices that pull out the elements of x_1 associated with x_0 and x_w, respectively: $J = (I_{k_0\times k_0}, 0_{k_0\times k_w})'$ and $J_w = (0_{k_w\times k_0}, I_{k_w\times k_w})'$. The population residual variance is $\sigma^2 = \lim_{T\to\infty} Eu^2_{T,t+\tau}$. For a $(k_w \times k)$ matrix $\tilde{A}$ satisfying $\tilde{A}'\tilde{A} = B_1^{-1/2}(-J'B_0J + B_1)B_1^{-1/2}$, let $\tilde{h}_{1,t+\tau} = \sigma^{-1}\tilde{A}B_1^{1/2}h_{1,t+\tau}$ and $\tilde{H}_1(t) = \sigma^{-1}\tilde{A}B_1^{1/2}H_1(t)$. Let $F_1 = J'_wB_1J_w$ and $F_1(t) = J'_wB_1(t)J_w$. If we define

[1] For simplicity, we suppress the triangular array notation implied by our use of local to zero asymptotics. The triangular array structure of the data implies underlying notation of a predictand $y_{T,t+\tau}$, the predictors $x_{T,1,t}$ and the error term $u_{T,t+\tau}$. We omit the T subscript for readability unless necessary to convey concepts.

[2] The coefficient vector β_1 does not vary with the forecast horizon τ because, in our analysis, τ is treated as fixed.

$\gamma_{\tilde{h}\tilde{h}}(i) = \lim_{T\to\infty} E\tilde{h}_{T,1,t+\tau}\tilde{h}'_{T,1,t+\tau-i}$, then $S_{\tilde{h}\tilde{h}} = \gamma_{\tilde{h}\tilde{h}}(0) + \sum_{i=1}^{\tau-1}(\gamma_{\tilde{h}\tilde{h}}(i) + \gamma'_{\tilde{h}\tilde{h}}(i))$. For notational convenience in the presentation of the results, we define a vector containing 0's (for the $x_{0,t}$ variables) and the weak predictor coefficients: $\delta = (0_{1\times k_0}, \beta_w^{'})'$. Finally, for any $(m \times n)$ matrix A, let $|A|$ denote the max norm and $tr(A)$ denote the trace.

2.3 Test statistics

In the context of non-nested models, Diebold and Mariano (1995) propose a test for equal MSE based upon the sequence of loss differentials $\hat{L}_{t+\tau} = \hat{u}^2_{0,t+\tau} - \hat{u}^2_{1,t+\tau}$, where $\hat{u}^2_{i,t+\tau} = (y_{t+\tau} - x'_{i,t}\hat{\beta}_{i,t})^2$. If we define $MSE_i = (P-\tau+1)^{-1}\sum_{t=T}^{T+P-\tau}\hat{u}^2_{i,t+\tau}$ $(i = 0,1)$, $\bar{L} = (P-\tau+1)^{-1}\sum_{t=T}^{T+P-\tau}\hat{L}_{t+\tau} = MSE_0 - MSE_1$, $\widehat{\gamma}_{LL}(j) = (P-\tau+1)^{-1}\sum_{t=T+j}^{T+P-\tau}(\hat{L}_{t+\tau} - \bar{L})(\hat{L}_{t+\tau-j} - \bar{L})$, $\widehat{\gamma}_{LL}(-j) = \widehat{\gamma}_{LL}(j)$, and $\hat{S}_{LL} = \sum_{j=-\bar{j}}^{\bar{j}} K(j/M)\widehat{\gamma}_{LL}(j)$ for some kernel $K(\cdot)$ defined below, the statistic takes the form

$$\text{MSE-}t = (P-\tau+1)^{1/2} \times \frac{\bar{L}}{\sqrt{\hat{S}_{LL}}}. \tag{2}$$

Under the null that $x_{w,t}$ has no population-level predictive power for $y_{t+\tau}$, the population difference in MSEs, $E(u^2_{0,t+\tau} - u^2_{1,t+\tau})$, will equal 0 for all t. When $x_{w,t}$ has predictive power, the population difference in MSEs will be positive. Even so, the finite sample difference need not be positive and in fact, for a given sample size (say, $t = T$) the difference in finite sample MSEs, $E(\hat{u}^2_{0,T+\tau} - \hat{u}^2_{1,T+\tau})$, may be zero, thus motivating a distinct null hypothesis of equal finite-sample predictive ability. Regardless of which null hypothesis we consider (equal population-level or equal finite-sample predictive ability), the MSE-t test and the other equal MSE tests described below are one–sided to the right.

While West (1996) proves directly that the MSE-t statistic can be asymptotically standard normal when applied to non–nested forecasts, this is not the case when the models are nested. In particular, the results in West (1996) require that under the null, the population-level long run variance of $\hat{L}_{t+\tau}$ is positive. This requirement is violated with nested models regardless of the presence of weak predictors. Intuitively, with nested models (and for the moment ignoring the weak predictors), the null hypothesis that the restrictions imposed in the benchmark model are true implies the population errors of the competing forecasting models are exactly the same. As a result, in population $L_{t+\tau} = 0$ for all t, which makes the corresponding variance also equal to 0. Because the sample analogues (for example, $\bar{L}$ and its variance) converge to zero at the same rate, the test statistics have non–degenerate null distributions, but they are non–standard.

Motivated by (i) the degeneracy of the long-run variance of $L_{t+\tau}$ and (ii) the functional form of the standard in-sample F-test, McCracken (2007) develops an out–of–sample F-type test of equal MSE, given by

$$\text{MSE-}F = (P - \tau + 1) \times \frac{MSE_0 - MSE_1}{MSE_1} = (P - \tau + 1) \times \frac{\bar{L}}{MSE_1}. \tag{3}$$

Like the MSE-t test, the limiting distribution of the MSE-F test is non–standard when the forecasts are nested under the null. Clark and McCracken (2005) and McCracken (2007) show that, for τ–step ahead forecasts, the MSE-F statistic converges in distribution to functions of stochastic integrals of quadratics of Brownian motion, with limiting distributions that depend on the sample split (denoted λ_P below), the number of exclusion restrictions k_w, and the unknown nuisance parameter $S_{\tilde{h}\tilde{h}}$. While this continues to hold in the presence of weak predictors, the asymptotic distributions will depend not only upon the unknown coefficients associated with the weak predictors but also upon other unknown second moments of the data.

2.4 Asymptotic results

We use the following assumptions to derive the asymptotic distributions of the test statistics considered, for the case of weak predictors.

Assumption 1: The parameters of the forecasting models are estimated using OLS, yielding $\hat{\beta}_{i,t} = \arg\min_{\beta_i} t^{-1} \sum_{s=1}^{t-\tau} (y_{T,s+\tau} - x'_{T,i,s}\beta_i)^2$, $i = 0, 1$, for the restricted and unrestricted models, respectively.

Assumption 2: (a) $T^{-1} \sum_{t=1}^{[rT]} U_{T,t} U'_{T,t-j} \Rightarrow r\Omega_j$. (b) $\Omega_{11,j} = 0$ all $j \geq \tau$. (c) For some $q > 2$, $\sup_{T \geq 1, t \leq T+P} E|U_{T,t}|^{2q} < \infty$. (d) The zero mean array $U_{T,t} - EU_{T,t} = (h'_{T,1,t+\tau}, vec(x_{T,1,t}x'_{T,1,t} - Ex_{T,1,t}x'_{T,1,t})')'$ satisfies Theorem 3.2 of de Jong and Davidson (2000).

Assumption 3: $\lim_{P,T \to \infty} P/T = \lambda_P \in (0, \infty)$.

Assumption 4: (a) Let $K(x)$ be a continuous kernel such that for all real scalars x, $|K(x)| \leq 1$, $K(x) = K(-x)$ and $K(0) = 1$. (b) For some bandwidth M and constant $i \in (0, 0.5)$, $M = O(P^i)$. (c) The number of covariance terms $\bar{j}$, used to estimate the long–run covariance S_{LL} defined in Section 2.2, satisfies $\tau - 1 \leq \bar{j} < \infty$.

Assumption 2 imposes three types of conditions. First, in (a) and (c) we require that the observables, while not necessarily covariance stationary, are asymptotically mean square stationary with finite second moments. We do so in order to allow the observables to have marginal distributions that vary as the weak predictive ability strengthens along with the sample size but are 'well-behaved' enough that, for example, sample averages converge in probability to the appropriate population means. Second, in (b) we impose the restriction that the forecast errors form an MA($\tau - 1$) process and hence the model has sufficient lags to pick up all the autocorrelation in the errors other than that associated with the τ-step ahead nature of the forecasts. We do so in order to emphasize the role that weak predictors have on forecasting without also introducing other forms of model misspecification. Finally, in (d) we impose the high level assumption that, in particular, $h_{1,t+\tau}$ satisfies Theorem 3.2 of de Jong and Davidson (2000). By doing so we not only insure that certain weighted partial sums converge weakly to standard Brownian motion, but also allow ourselves to take advantage of various results pertaining to convergence in distribution to stochastic integrals.

Assumption 3's requirement on limiting sample sizes implies that the duration of forecasting is finite but non-trivial. This assumption, while standard in our previous work, differs importantly from that in Giacomini and White (2006). In their approach to predictive inference for nested models, they assume that a rolling window of fixed and *finite* width is used for estimation of the model parameters (hence $\lim_{P\to\infty} P/T = \infty$). While we allow rolling windows, our asymptotics assume that the window width is a non-trivial magnitude of the out-of-sample period and hence $\lim_{P,T\to\infty} P/T \in (0, \infty)$. This difference in the assumed window width, along with our assumption that the additional predictors in the nesting model are weak, is fundamentally what drives the difference in our results from theirs and in particular, allows us to derive results that permit the use of the recursive scheme.

Finally, Assumption 4 is necessitated by the serial correlation in the multi-step (τ-step) forecast errors — errors from even well-specified models exhibit serial correlation, of an MA($\tau - 1$) form. Typically, researchers constructing a t-statistic utilizing the squares of these errors account for serial correlation of at least order $\tau - 1$ in forming the necessary standard error estimates. Meese and Rogoff (1988), Groen (1999), and Kilian and Taylor (2003), among other applications to forecasts from nested models, use kernel-based methods to estimate the relevant long-run covariance.[3] We therefore impose conditions sufficient to cover applied practices. Parts (a) and (b) are not particularly controversial. Part (c), however, imposes the restriction that since the orthogonality conditions used to identify the parameters form a moving average of finite order $\tau - 1$, this fact is taken into account (in the sense of assuming a finite bandwidth) when constructing the MSE-t statistic.

Under these assumptions, the asymptotic distributions will depend on the following functions of a $k_w \times 1$ vector standard Brownian motion, denoted $W(s)$, with (as defined above) $\delta = (0_{1\times k_0}, \beta_w')'$: $\Gamma_1 = \int_1^{1+\lambda_P} s^{-1}W'(s)S_{\tilde{h}\tilde{h}}dW(s)$, $\Gamma_2 = \int_1^{1+\lambda_P} s^{-2}W'(s)S_{\tilde{h}\tilde{h}}W(s)ds$, $\Gamma_3 = \int_1^{1+\lambda_P}(\delta' B_1^{-1/2}\tilde{A}'/\sigma)S_{\tilde{h}\tilde{h}}^{1/2}dW(s)$, $\Gamma_4 = \int_1^{1+\lambda_P} \beta_w' F_1^{-1}\beta_w/\sigma^2 ds = \lambda_P \beta_w' F_1^{-1}\beta_w/\sigma^2$, $\Gamma_5 = \int_1^{1+\lambda_P} s^{-2}W'(s)S_{\tilde{h}\tilde{h}}^2 W(s)ds$, $\Gamma_6 = \int_1^{1+\lambda_P} s^{-1}(\delta' B_1^{-1/2}\tilde{A}'/\sigma)S_{\tilde{h}\tilde{h}}^{3/2}W(s)ds$, and $\Gamma_7 = \lambda_P(\delta' B_1^{-1/2}\tilde{A}'/\sigma)S_{\tilde{h}\tilde{h}}(\tilde{A}B_1^{-1/2}\delta/\sigma)$.

The following two Theorems provide the asymptotic distributions of the MSE-F and MSE-t statistics in the presence of weak predictors.

Theorem 2.1: Maintain Assumptions 1-3. MSE-$F \to^d \{2\Gamma_1 - \Gamma_2\} + 2\{\Gamma_3\} + \{\Gamma_4\}$.

Theorem 2.2: Maintain Assumptions 1-4. MSE-$t \to^d (\{\Gamma_1 - .5\Gamma_2\} + \{\Gamma_3\} + \{.5\Gamma_4\})/(\Gamma_5 + \Gamma_6 + \Gamma_7)^{.5}$.

Theorems 2.1 and 2.2 show that the limiting distributions of the MSE-t and MSE-F tests are neither normal nor chi-square when the forecasts are nested, regardless of the presence of weak predictors. Theorem 2.1 is very similar to Proposition 2 in Inoue and Kilian (2004) while Theorem 2.2 is unique. And again, the limiting distributions are free of nuisance parameters in only very special cases. In particular, the distributions here are free of nuisance parameters only if there are no weak predictors and if $S_{\tilde{h}\tilde{h}} = I$. If this is the case — if, for example, $\tau = 1$ and the forecast errors are conditionally homoskedastic — both representations simplify to those in McCracken (2007) and hence his critical values can be used for testing for equal population-level predictive ability. In the absence of weak predictors alone, the representation simplifies to that in Clark and McCracken (2005) and hence the asymptotic distributions still depend upon $S_{\tilde{h}\tilde{h}}$. In this case, and in the most general case where weak predictors are present, we will use bootstrap methods to estimate asymptotic critical values. Before describing our bootstrap approach, however, it will be helpful to clarify the null hypothesis of interest.

[3] For similar uses of kernel–based methods in analyses of non–nested forecasts, see, for example, Diebold and Mariano (1995) and West (1996).

2.5 A null hypothesis with weak predictors

The noncentrality terms, especially those associated with the asymptotic distribution of the MSE-F statistic (Γ_4), give some indication of the power that the test statistics have against deviations from the null hypothesis of equal population-level predictive ability H_0 : $E(u_{0,t+\tau}^2 - u_{1,t+\tau}^2) = 0$ for all t — for which it must be the case that $\beta_w = 0$. As noted earlier, it is in that sense that our analytical results are closely related to those in Inoue and Kilian (2004). Closer inspection, however, shows that the results provide opportunities for testing another form of the null hypothesis of equal predictive ability when weak predictors are present.

For example, under the assumptions made earlier in this section it is straightforward to show that the mean of the asymptotic distribution of the MSE-F statistic can be used to approximate the mean difference in the average out-of-sample predictive ability of the two models, as:[4]

$$E\sum\nolimits_{t=T}^{T+P}(\hat{u}_{0,t+\tau}^2 - \hat{u}_{1,t+\tau}^2) \approx \int_1^{1+\lambda_P} [-s^{-1}tr((-JB_0J' + B_1)V) + \beta_w' F_1^{-1}\beta_w]ds.$$

Intuitively, one might consider using these expressions as a means of characterizing when the two models have equal average finite-sample predictive ability over the out-of-sample period. For example, having set these two expressions to zero, integrating and solving for the marginal signal-to-noise ratio implies $\frac{\beta_w' F_1^{-1}\beta_w}{tr((-JB_0J'+B_1)V)}$ equals $\frac{\ln(1+\lambda_P)}{\lambda_P}$. This ratio simplifies further when $\tau = 1$ and the forecast errors are conditionally homoskedastic, in which case $tr((-JB_0J' + B_1)V) = \sigma^2 k_w$.

The marginal signal-to-noise ratio $\frac{\beta_w' F_1^{-1}\beta_w}{tr((-JB_0J'+B_1)V)}$ forms the basis of our new approach to testing for equal predictive ability. Rather than testing for equal population-level predictive ability H_0 : $E(u_{0,t+\tau}^2 - u_{1,t+\tau}^2) = 0$ for all t — for which it must be the case that $\beta_w = 0$ — we test for equal average out-of-sample predictive ability H_0 : $\lim_{P,T\to\infty} E(\sum_{t=T}^{T+P}(\hat{u}_{0,t+\tau}^2 - \hat{u}_{1,t+\tau}^2)) = 0$ – for which it is the case that $\beta_w' F_1^{-1}\beta_w = d$, where d equals $\frac{\ln(1+\lambda_P)}{\lambda_P} tr((-JB_0J' + B_1)V)$.[5]

While we believe the result is intuitive, it is not immediately clear how such a restriction

[4] By taking this approach we are using the fact that under our assumptions, notably the L^2-boundedness portion of Assumption 2, $\sum_{t=T}^{T+P}(\hat{u}_{0,t+\tau}^2 - \hat{u}_{1,t+\tau}^2)$ is uniformly integrable and hence the expectation of its limit is equal to the limit of its expectation.

[5] One could also derive a test for equal forecast accuracy at the end of the out-of-sample period. Using similar arguments, this hypothesis implies that $\beta_w' F_1^{-1}\beta_w = d$, where d equals $\frac{1}{1+\lambda_P} tr((-JB_0J' + B_1)V)$. Under this null hypothesis, our proposed bootstrap is valid so long as $\hat{d}$ (defined below) is modified appropriately.

on the regression parameters can be used to achieve asymptotically valid inference. If we look back at the asymptotic distribution of the MSE-F statistic, we see that in general it not only depends upon the unknown value of β_w, but also the asymptotic distribution is non-standard, thus requiring either extensive tables of critical values or simulation-based methods for constructing the critical values.

In the following section we develop a new bootstrap-based method for constructing asymptotically valid critical values that can be used to test the null of equal average finite-sample predictive ability.

2.6 Bootstrap-based critical values with weak predictors

Our new, bootstrap-based method of approximating the asymptotically valid critical values for comparisons between nested models is different from that previously used in studies such as Kilian (1999) and Clark and McCracken (2005). In those applications, an appropriately dimensioned VAR was initially estimated by OLS imposing the restriction that β_w was set to zero and the residuals saved for resampling. The recursive structure of the VAR was then used to generate a large number of artificial samples, each of which was used to construct one of the test statistics discussed above. The relevant sample percentile from this large collection of artificial statistics was then used as the critical value. Simulations show that this approach provides accurate inference for the null of equal population-level predictive ability not only for one-step ahead forecasts but also for longer horizons (in our direct multi-step framework).

However, there are two reasons we should not expect this bootstrap approach to provide accurate inference in the presence of weak predictors. First, imposing the restriction that β_w is set to zero implies a null of equal population — not finite-sample — predictive ability. Second, by creating the artificial samples using the recursive structure of the VAR we are imposing the restriction that equal one-step ahead predictive ability implies equal predictive ability at longer horizons. Our present framework in no way imposes that restriction. We therefore take an entirely different approach to imposing the relevant null hypothesis and generating the artificial samples.

To test whether the two models have equal average predictive ability over the out-of-sample period, we need to determine whether $\beta_w' F_1^{-1} \beta_w$ equals $\frac{\ln(1+\lambda_P)}{\lambda_P} tr((-JB_0J'+B_1)V)$. While this restriction is infeasible due to the various unknown moments and parameters, it suggests a closely related, feasible restriction quite similar to that used in ridge regression.

However, instead of imposing the restriction that $\beta_w'\beta_w = c$ for some finite constant — as one would in a ridge regression — we instead impose the restriction that $\delta' J_w F_1^{-1}(T) J_w' \delta$ equals $\frac{\ln(1+\widehat{\lambda}_P)}{\widehat{\lambda}_P} tr((-JB_0(T)J' + B_1(T))V(T))$, where the relevant unknowns are estimated using the obvious sample moments: $\widehat{\lambda}_P = P/T, B_i(T) = (T^{-1}\sum_{s=1}^{T-\tau} x_{i,s}x_{i,s}')^{-1}$ for $i = 0, 1$, $F_1(T) = J_w' B_1(T) J_w$, and $V(T) =$ an estimate of the long-run variance of $h_{1,t+\tau}$.[6] In addition, we estimate δ using the approximation $\widehat{\delta} = (0_{1\times k_0}, T^{1/2}\widetilde{\beta}_{w,T}')'$ where $\widetilde{\beta}_{w,T}$ denotes the restricted least squares estimator of the parameters associated with the weak predictors satisfying

$$\begin{aligned} \widetilde{\beta}_{1,T} &\equiv (\widetilde{\beta}_{0,T}', \widetilde{\beta}_{w,T}')' \\ &= \arg\min_{b_1} \sum\nolimits_{s=1}^{T+P-\tau} (y_{s+\tau} - x_{1,s}'b_1)^2 \text{ s.t. } b_1' J_w F_1^{-1}(T) J_w' b_1 = \widehat{d}/T \end{aligned} \tag{4}$$

where $\widehat{d}$ equals $\frac{\ln(1+\widehat{\lambda}_P)}{\widehat{\lambda}_P} tr((-JB_0(T)J' + B_1(T))V(T))$. For a given sample size, this estimator is equivalent to a ridge regression if the weak predictors are orthonormal. More generally, though, it lies in the class of asymptotic shrinkage estimators discussed in Hansen (2008).

Note that this approach to imposing the null hypothesis is consistent with the direct multi-step forecasting approach we assume is used to construct the forecasts and hence the restriction can vary with the forecast horizon τ. This approach therefore precludes using a VAR and its recursive structure to generate the artificial samples. Instead we use a variant of the wild fixed regressor bootstrap developed in Goncalves and Kilian (2007) that accounts for the direct multi-step nature of the forecasts. Specifically, in our framework the x's are held fixed across the artificial samples and the dependent variable is generated using the direct multi-step equation $y_{s+\tau} = x_{1,s}'\widetilde{\beta}_{1,T} + \widehat{v}_{s+\tau}^*$, $s = 1, ..., T + P - \tau$, for a suitably chosen artificial error term $\widehat{v}_{s+\tau}^*$ designed to capture both the presence of conditional heteroskedasticity and an assumed $MA(\tau-1)$ serial correlation structure in the τ-step ahead forecasts. Specifically, we construct the artificial samples and bootstrap critical values using the following algorithm.[7]

1. (a) Set $\widehat{d}$ to $\frac{\ln(1+\widehat{\lambda}_P)}{\widehat{\lambda}_P} tr((-JB_0(T)J' + B_1(T))V(T))$. Estimate the unrestricted model using the weighted ridge regression from equation (4) above and save the fitted values $x_{1,s}'\widetilde{\beta}_{1,T}$ $s = 1, ..., T + P - \tau$. Note that the resulting parameter estimate will vary

[6] In our Monte Carlo simulations and empirical work we use a Newey-West kernel with bandwidth 0 for horizon = 1 and bandwidth 1.5*horizon otherwise.

[7] Our approach to generating artificial samples of multi-step forecast errors builds on a sampling approach proposed in Hansen (1996).

with the forecast horizon. (b) Estimate the unrestricted model by OLS without imposing the restriction in (4) and save the residuals $\widehat{v}_{1,s+\tau}$, $s = 1, ..., T + P - \tau$.

2. Using NLLS, estimate an $MA(\tau - 1)$ model for the OLS residuals $\widehat{v}_{1,s+\tau}$ (from the unrestricted model without the restriction in (4)) such that $v_{1,s+\tau} = \varepsilon_{1,s+\tau} + \theta_1\varepsilon_{1,s+\tau-1} + \cdots + \theta_{\tau-1}\varepsilon_{1,s+1}$. Let $\eta_{s+\tau}$, $s = 1, ..., T+P-\tau$, denote an $i.i.d$ $N(0, 1)$ sequence of simulated random variables.[8] Define $\widehat{v}^*_{1,s+\tau} = (\eta_{s+\tau}\widehat{\varepsilon}_{1,s+\tau} + \widehat{\theta}_1\eta_{s-1+\tau}\widehat{\varepsilon}_{1,s+\tau-1} + ... + \widehat{\theta}_{\tau-1}\eta_{s+1}\widehat{\varepsilon}_{1,s+1})$, $s = 1, ..., T + P - \tau$. Form artificial samples of $y^*_{s+\tau}$ using the fixed regressor structure, $y^*_{s+\tau} = x'_{1,s}\widetilde{\beta}_{1,T} + \widehat{v}^*_{1,s+\tau}$.

3. Using the artificial data, construct forecasts and an estimate of the test statistics (MSE-F, MSE-t) as if this were the original data.

4. Repeat steps 2 and 3 a large number of times: $j = 1, ..., N$.

5. Reject the null hypothesis, at the α% level, if the test statistic is greater than the $(100 - \alpha)$%-ile of the empirical distribution of the simulated test statistics.

By using the weighted ridge regression to estimate the model parameters we are able, in large samples, to impose the restriction that the implied estimates $(T^{1/2}\widetilde{\beta}_{w,T})$ of the local-to-zero parameters β_w satisfy our approximation to the null hypothesis. This is despite the fact that the estimates of β_w are not consistent. While this estimator, along with the fixed regressor structure of the bootstrap, imposes the null hypothesis upon the artificial samples, it is not necessarily the case that the bootstrap is asymptotically valid in the sense that the estimated critical values are consistent for their population values. To see how this might happen, note that the asymptotic distributions from Theorem 2.1 depend explicitly upon the local-to-zero parameters β_w through the terms Γ_3 and Γ_4. In the case of Γ_4, this is not an issue because the null hypothesis imposes a restriction on the value of this term that does not depend upon β_w explicitly, just an appropriately chosen weighted quadratic that is known under the null. Γ_3 is a different story. This term is asymptotically normal with a zero mean and variance $\lambda_P\beta'_w V\beta_w$ that, in general, need not have any relationship to the restriction $\beta'_w F_1^{-1}\beta_w = d$ implied by the null hypothesis. Hence, in general, the asymptotic distribution is an explicit function of the value of β_w, implying that the null hypothesis itself does not imply a unique asymptotic distribution for either the MSE-F or MSE-t statistics.

Even so, we can show that the bootstrap is asymptotically valid in two empirically

[8] The assumption of Normality is not necessary for our results. Instead it is sufficient if η_{s+t} denotes an $i.i.d.$ zero-mean process with unit variance and is uniformly L^r-bounded for some $r > 2$.

relevant special cases. In both cases the bootstrap works despite the fact that we cannot consistently estimate β_w. The trick is to note that (i) while $T^{1/2}\widetilde{\beta}_{w,T}$ is not a consistent estimate of β_w, the first stage of the bootstrap insures that its probability limit $\tilde{\beta}_w$ is on the sphere defined by $\beta_w' F_1^{-1}\beta_w = d$ and (ii) in the two special cases discussed below, the null asymptotic distribution is invariant to the actual value of β_w so long as the relationship $\beta_w' F_1^{-1}\beta_w = d$ holds. To prove the result, however, we require a modest strengthening of the moment conditions on the model residuals.

Assumption 2$'$: (a) $T^{-1}\sum_{t=1}^{[rT]} U_{T,t}U_{T,t-j}' \Rightarrow r\Omega_j$. (b) $\mathrm{E}(\varepsilon_{1,s+\tau}|\varepsilon_{1,s+\tau-j}, x_{T,1,s-j}\ j \geq 0) = 0$. (c) Let $\gamma_T = (\beta_{1,T}', \theta_1, ..., \theta_{\tau-1})'$, $\widehat{\gamma}_T = (\hat{\beta}_{1,T}', \widehat{\theta}_1, ..., \widehat{\theta}_{\tau-1})'$, and define the function $\widehat{\varepsilon}_{1,s+\tau} = \widehat{\varepsilon}_{1,s+\tau}(\widehat{\gamma}_T)$ such that $\widehat{\varepsilon}_{1,s+\tau}(\gamma_T) = \varepsilon_{1,s+\tau}$. In an open neighborhood N_T around γ_T, there exists a finite constant c such that $\sup_{1\leq s\leq T, T\geq 1} || \sup_{\gamma\in N_T}(\widehat{\varepsilon}_{1,s+\tau}(\gamma), \nabla\widehat{\varepsilon}_{1,s+\tau}'(\gamma), x_{T,1,s})'||_4 \leq c$. (d) The zero mean array $U_{T,t} - EU_{T,t} = (h_{T,1,t+\tau}', vec(x_{T,1,t}x_{T,1,t}' - Ex_{T,1,t}x_{T,1,t}')')'$ satisfies Theorem 3.2 of de Jong and Davidson (2000).

Assumption 2$'$ differs from Assumption 2 in two ways. First, in (b) it emphasizes the point that the forecast errors, and by implication $h_{1,t+\tau}$, form an $MA(\tau-1)$ process. Second, in (c) it bounds the second moments not only of $h_{1,t+\tau} = (\varepsilon_{1,s+\tau} + \theta_1\varepsilon_{1,s+\tau-1} + ... + \theta_{\tau-1}\varepsilon_{1,s+1})x_{1,s}$ (as in Assumption 2) but also the functions $\widehat{\varepsilon}_{1,s+\tau}(\gamma)x_{1,s}$ and $\nabla\widehat{\varepsilon}_{1,s+\tau}(\gamma)x_{1,s}$ for all γ in an open neighborhood of γ_T. These assumptions are primarily used to show that the bootstrap-based artificial samples, which are a function of the estimated errors $\widehat{\varepsilon}_{1,s+\tau}$, adequately replicate the time series properties of the original data in large samples. Specifically we must insure that the bootstrap analog of $h_{1,s+\tau}$ is not only zero mean but has the same long-run variance V. Such an assumption is not needed for our earlier results since the model forecast errors $\widehat{u}_{i,s+\tau}$, $i = 0, 1$, are linear functions of $\hat{\beta}_i$ and Assumption 2 already imposes moment conditions on $\widehat{u}_{1,s+\tau}$ via moment conditions on $h_{1,s+\tau}$.

In the following let MSE-F^* and MSE-t^* denote statistics generated using the artificial samples from our bootstrap and let $=^{d^*}$ and $\rightarrow^{d^*}$ denote equality and convergence in distribution (more specifically, convergence occurs in a set with probability limiting to 1) with respect to the bootstrap-induced probability measure P^*. Similarly let Γ_i^*, $i = 1, ..., 7$, denote random variables generated using the artificial samples satisfying $\Gamma_i^* =^{d^*} \Gamma_i$, $i = 1, ..., 7$, for Γ_i defined in the discussion preceding the assumptions.

Theorem 2.3: Let $\beta_w' F_1^{-1}\beta_w = d$ and assume either (i) $\tau = 1$ and the forecast errors from the unrestricted model are conditionally homoskedastic, or (ii) $\dim(\beta_w) = 1$. (a) Given Assumptions 1, 2′, and 3, MSE-$F^* \to^{d^*} \{2\Gamma_1^* - \Gamma_2^*\} + 2\{\Gamma_3^*\} + \{\Gamma_4^*\}$. (b) Given Assumptions 1, 2′, 3, and 4, MSE-$t^* \to^{d^*} (\{2\Gamma_1^* - \Gamma_2^*\} + 2\{\Gamma_3^*\} + \{\Gamma_4^*\})/(\Gamma_5^* + \Gamma_6^* + \Gamma_7^*)^{.5}$.

In Theorem 2.3 we show that our fixed-regressor bootstrap provides an asymptotically valid method of estimating the critical values associated with the null of equal average finite sample forecast accuracy. The result, however, is applicable in only two special cases. In the first, we require that the forecast errors be one-step ahead and conditionally homoskedastic. In the second, we allow serial correlation and conditional heteroskedasticity but require that β_w is scalar. While neither case covers the broadest situation in which β_w is not scalar and the forecast errors exhibit either serial correlation or conditional heteroskedasticity, these two special cases cover a wide range of empirically relevant applications. Kilian (1999) argues that conditional homoskedasticity is a reasonable assumption for one-step ahead forecasts of quarterly macroeconomic variables. Moreover, in many applications in which a nested model comparison is made (examples include, among others, Chen, Rogoff, and Rossi (2010), Goyal and Welch (2008), and Stock and Watson (2003)), the unrestricted forecasts are made by simply adding one lag of a single predictor to the baseline restricted model. Of course, in more general settings that fall outside these two cases, it is possible that our proposed bootstrap will be reliable even if we can't prove its asymptotic validity. As we detail in section 4, some supplementary Monte Carlo experiments confirm this supposition on the broader reliability of our testing approach.

By itself, Theorem 2.3 is insufficient for recommending the use of the bootstrap: it does not tell us whether the proposed bootstrap is adequate for constructing asymptotically valid critical values under the alternative that the unrestricted model forecasts more accurately than the restricted model. Unfortunately, there are any number of ways to model the case in which $\beta_w' F_1^{-1} \beta_w > d$. For example, rather than modeling the weak predictive ability as $T^{-1/2}\beta_w$ with $\beta_w' F_1^{-1} \beta_w = d$, one could model the predictive content as $T^{-a} C \beta_w$ for constants $C < \infty$ and $a \in (0, 1/2]$ satisfying $\beta_w' F_1^{-1} \beta_w > d$. While mathematically elegant, this approach does not allow us to analyze the most intuitive alternative in which not only is the unrestricted model more accurate but also $J_w' \hat{\beta}_{1,T}$ is a consistent estimator of $\beta_w \neq 0$. For this situation to hold we need the additional restriction that $a = 0$ and hence β_w is no longer interpretable as a local-to-zero parameter. With this modification in hand, we address the validity of the bootstrap under the alternative in the following Theorem.

Theorem 2.4: Let $J_w'\hat{\beta}_{1,T} \to^p \beta_w \neq 0$ and assume either (i) $\tau = 1$ and the forecast errors from the unrestricted model are conditionally homoskedastic, or (ii) $\dim(\beta_w) = 1$. (a) Given Assumptions 1, 2′, and 3, MSE-$F^* \to^{d^*} \{2\Gamma_1^* - \Gamma_2^*\} + 2\{\Gamma_3^*\} + \{\Gamma_4^*\}$. (b) Given Assumptions 1, 2′, 3, and 4, MSE-$t^* \to^{d^*} (\{2\Gamma_1^* - \Gamma_2^*\} + 2\{\Gamma_3^*\} + \{\Gamma_4^*\})/(\Gamma_5^* + \Gamma_6^* + \Gamma_7^*)^{.5}$.

In Theorem 2.4 we see that indeed, the bootstrap-based test is consistent for testing the null hypothesis of equal finite sample predictive accuracy (that $\beta_w' F_1^{-1}\beta_w = d$) against the alternative that the unrestricted model is more accurate (that $J_w'\hat{\beta}_{1,T} \to^p \beta_w \neq 0$). This follows since under this alternative, the data-based statistics MSE-F and MSE-t each diverge to $+\infty$ while the the bootstrap-based statistics MSE-F^* and MSE-t^* each retain the same asymptotic distribution as they did under the null.

As we will show in section 3, our fixed regressor bootstrap provides reasonably sized tests in our Monte Carlo simulations, outperforming other bootstrap-based methods for estimating the asymptotically valid critical values necessary to test the null of equal average finite sample predictive ability.

2.7 Differences under the rolling scheme

The results presented above for the recursive estimation and forecasting scheme also apply under a rolling scheme, under which the number of observations used for estimation is held constant as we proceed forward across forecast origins. This subsection lists the changes that apply under a rolling scheme.

First, under the rolling scheme, the parameter estimates and associated moments are defined as $\hat{\beta}_{i,t} = \arg\min_{\beta_i} T^{-1}\sum_{s=t-\tau-T+1}^{t-\tau}(y_{s+\tau} - x_{i,s}'\beta_i)^2$, $H_i(t) = (T^{-1}\sum_{s=t-\tau-T+1}^{t-\tau} x_{i,s}u_{s+\tau})$ $= (T^{-1}\sum_{s=t-\tau-T+1}^{t-\tau} h_{i,s+\tau})$, and $B_i(t) = (T^{-1}\sum_{s=t-\tau-T+1}^{t-\tau} x_{i,s}x_{i,s}')^{-1}$, for models $i = 0, 1$.

Second, some of the functions entering the asymptotic distributions are slightly different under the rolling scheme: $\Gamma_1 = \int_1^{1+\lambda_P}(W(s) - W(s-1))' S_{\tilde{h}\tilde{h}} dW(s)$, $\Gamma_2 = \int_1^{1+\lambda_P}(W(s) - W(s-1))' S_{\tilde{h}\tilde{h}}(W(s) - W(s-1))ds$, $\Gamma_5 = \int_1^{1+\lambda_P}(W(s) - W(s-1))' S_{\tilde{h}\tilde{h}}^2(W(s) - W(s-1))ds$, and $\Gamma_6 = \int_1^{1+\lambda_P} s^{-1}(\delta' B_1^{-1/2}\tilde{A}'/\sigma) S_{\tilde{h}\tilde{h}}^{3/2}(W(s) - W(s-1))ds$.

Third, under the rolling scheme, the approximation of the mean difference in average forecast accuracy is

$$E\sum\nolimits_{t=T}^{T+P}(\hat{u}_{0,t+\tau}^2 - \hat{u}_{1,t+\tau}^2) \approx \int_1^{1+\lambda_P}[-tr((-JB_0J' + B_1)V) + \beta_w' F_1^{-1}\beta_w]ds.$$

Solving for the marginal signal-to-noise ratio that makes the models equally accurate implies $\frac{\beta_w' F_1^{-1}\beta_w}{tr((-JB_0J'+B_1)V)} = 1$. In turn, equal average out-of-sample predictive ability implies a condition $\beta_w' F_1^{-1} \beta_w = d$, where $d = tr((-JB_0 J' + B_1)V)$.

Finally, under the rolling scheme, the first step of the bootstrap consists of estimating the parameter vector β_1 associated with the unrestricted model (as detailed in equation (4)) subject to the restriction $\widehat{d} = tr((-JB_0(T)J' + B_1(T))V(T))$.

3 Monte Carlo evidence

We use simulations of bivariate and multivariate DGPs that are in most cases based on common macroeconomic applications to evaluate the finite sample properties of the above approaches to testing for equal forecast accuracy. In these simulations, the benchmark forecasting model is a univariate model of the predictand y; the alternative models add lags of various other variables of interest. The null hypothesis is that the forecast from the alternative model is no more accurate than the benchmark forecast, in the sense that the additional variables in the alternative model have non–zero coefficients, but the coefficients are small enough that the benchmark and alternative models are expected to be equally accurate over the forecast sample. We focus our presentation on recursive forecasts, but include some results for rolling forecasts. We report empirical rejection rates using a nominal size of 10%. Size results using nominal sizes of 5% and 1% are qualitatively the same and available upon request.

While our focus is on evaluating the size and power of MSE-F and MSE-t tests based on the asymptotic distributions and bootstrap developed in the last section, we also consider two other approaches to inference — that is, sources of critical values and tests. First, we include results for the MSE-t test compared against standard normal critical values. Second, we provide results based on a non–parametric bootstrap patterned on White's (2000) method: we create bootstrap samples of forecast errors by sampling (with replacement) from the time series of sample forecast errors, and construct test statistics for each sample draw. However, as noted above and in White (2000), this procedure is not, in general, asymptotically valid when applied to nested models. We include the method in part for its computational simplicity and in part to examine the potential pitfalls of using the approach.

In our non-parametric bootstrap implementation, we follow the approach of White (2000) in using the stationary bootstrap of Politis and Romano (1994) and centering the bootstrap distributions around the sample values of the test statistics. The stationary bootstrap is parameterized to make the average block length equal to twice the forecast horizon. As to centering of test statistics, under the non–parametric approach, the relevant null hypothesis is that the MSE difference (benchmark MSE less alternative model MSE) is at most 0, and the MSE ratio (benchmark MSE/alternative model MSE) is at most 1. Following White (2000), each bootstrap draw of a given test statistic is re-centered around the corresponding sample test statistic. Bootstrapped critical values are computed as percentiles of the resulting distributions of re–centered test statistics.

3.1 Monte Carlo design

For all DGPs, we generate data using independent draws of innovations from the normal distribution and the autoregressive structure of the DGP.[9] We consider forecast horizons of one and four steps. With quarterly data in mind, we also consider a range of sample sizes (T, P), reflecting those commonly available in practice: 40,80; 40,120; 80,40; 80,80; 80,120; 120,40; and 120,80.

The DGPs are in most cases generally based on empirical relationships among U.S. inflation and a range of predictors, estimated with 1968-2008 data.[10] We focus on results for four DGPs, all of which satisfy the assumptions necessary to prove the asymptotic validity of our proposed bootstrap. But to assess the reliability of our proposed approach in more general settings, we also include some results for two additional DGPs that feature conditional heteroskedasticity or serial correlation in the forecast error, with model 1 having three more variables than model 0. We also provide results for a DGP in which the forecasting models are misspecified. In all cases, our reported results are based on 5000 Monte Carlo draws and 499 bootstrap replications.

3.1.1 DGPs

DGP 1 takes a very simple form:

$$
\begin{aligned}
y_{t+1} &= 1.0 + b_{11}x_{1,t} + u_{t+1} \\
x_{1,t+1} &= v_{1,t+1} \\
\text{var}\begin{pmatrix} u_{t+1} \\ v_{1,t+1} \end{pmatrix} &= \begin{pmatrix} 1.0 & \\ 0.0 & 0.25 \end{pmatrix}.
\end{aligned} \tag{5}
$$

[9] In our baseline size experiments, using innovations drawn from a t distribution with 5 degrees of freedom yields very similar results and are available upon request.

[10] For simplicity, while the empirical models underlying the DGP parameterizations and the forecasting models used in our Monte Carlos include constants, for simplicity we leave constants out of most of our data-generating processes. Including non-zero intercepts in the DGPs yields results exactly the same as those reported, because all of the forecasting models include intercepts.

In the DGP 1 experiments, which focus on a forecast horizon of 1 step, the alternative (unrestricted) forecasting model takes the form of the DGP equation for y_{t+1}, while the null or benchmark (restricted) model includes just a constant:

$$\begin{aligned} \text{null:}\ y_{t+1} &= \beta_0 + u_{0,t+1} &(6)\\ \text{alternative:}\ y_{t+1} &= \beta_0 + \beta_1 x_{1,t} + u_{1,t+1}. &(7)\end{aligned}$$

In size experiments, the coefficient b_{11} on $x_{1,t}$, which corresponds to the elements of our theoretical construct $\beta_w/\sqrt{T}$, is set to a value that makes the models equally accurate (in expectation) on average over the forecast sample. We determined the appropriate value on the basis of the population moments implied by the model and our asymptotic approximations given in section 2.5. For example, with recursive forecasts and T and P both equal to 80 (this coefficient value changes with T and P), this value is 0.1862. In power experiments, the coefficient is set to 0.5, such that the alternative model is expected to be more accurate than the null.

DGP 2 is based on the empirical relationship between the change in core PCE inflation (y_t) and the Chicago Fed's index of the business cycle ($x_{1,t}$, the CFNAI):

$$\begin{aligned} y_{t+1} &= -0.4y_t - 0.1y_{t-1} + b_{11}x_{1,t} + u_{t+1}\\ x_{1,t+1} &= 0.7x_{1,t} + v_{1,t+1} \qquad (8)\\ \text{var}\begin{pmatrix} u_{t+1}\\ v_{1,t+1}\end{pmatrix} &= \begin{pmatrix} 0.8 & \\ 0.0 & 0.3\end{pmatrix}.\end{aligned}$$

In the DGP 2 experiments, which also focus on a forecast horizon of 1 step, the alternative (unrestricted) forecasting model takes the AR(2) form of the DGP equation for y_{t+1} (with constant added); the null or benchmark (restricted) model drops $x_{1,t}$:

$$\begin{aligned} \text{null:}\ y_{t+1} &= \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + u_{0,t+1} &(9)\\ \text{alternative:}\ y_{t+1} &= \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + \beta_3 x_{1,t} + u_{1,t+1}. &(10)\end{aligned}$$

In size experiments, the coefficient b_{11} on $x_{1,t}$ is set to a value that makes the forecasting models equally accurate (in expectation) on average over the forecast sample. As described above, we used our asymptotic approximations to determine the appropriate value. For example, with recursive forecasts and T and P both equal to 80, this value is 0.1086, about 1/2 of the empirical estimate. In power experiments, the coefficient is set to 0.3, such that the alternative model is expected to be more accurate than the null.

DGP 3 is based on the empirical relationship of the change in core PCE inflation (y_t) to the CFNAI ($x_{1,t}$), PCE food price inflation less core inflation ($x_{2,t}$), and import price inflation less core inflation ($x_{3,t}$). To simplify the lag structure necessary for reasonable forecasting models, the inflation rates used in forming variables $x_{2,t}$ and $x_{3,t}$ are computed as two-quarter averages. Based on these data, DGP 2 takes the form

$$\begin{aligned} y_{t+1} &= -0.4y_t - 0.1y_{t-1} + b_{11}x_{1,t} + b_{21}x_{2,t} + b_{31}x_{3,t} + u_{t+1} \\ x_{1,t+1} &= 0.7x_{1,t} + v_{1,t+1} \\ x_{2,t+1} &= 0.9x_{2,t} - 0.2x_{2,t-1} + v_{2,t+1} \\ x_{3,t+1} &= 1.1x_{3,t} - 0.3x_{3,t-1} + v_{3,t+1} \end{aligned} \tag{11}$$

$$\text{var}\begin{pmatrix} u_t \\ v_{1,t+1} \\ v_{2,t+1} \\ v_{3,t+1} \end{pmatrix} = \begin{pmatrix} 0.8 & & & \\ 0.0 & 0.3 & & \\ -0.1 & 0.0 & 2.2 & \\ 0.5 & 0.1 & 0.8 & 9.0 \end{pmatrix}.$$

In DGP 3 experiments, which also focus on a forecast horizon of 1 step, the null (restricted) and alternative (unrestricted) forecasting models take the following forms, respectively:

$$y_{t+1} = \beta_0 + \beta_1 y_t + \beta_1 y_{t-1} + u_{0,t+1} \tag{12}$$

$$y_{t+1} = \beta_0 + \beta_1 y_t + \beta_1 y_{t-1} + \beta_3 x_{1,t} + \beta_4 x_{2,t} + \beta_5 x_{3,t} + u_{1,t+1}. \tag{13}$$

In power experiments, the b_{ij} coefficients are set at $b_{11} = 0.3$, $b_w = 0.1$, and $b_{13} = .015$ (roughly their empirical values). With these values, the alternative model is expected to be more accurate than the null. In size experiments, these values of the b_{ij} coefficients are multiplied by a constant less than one, such that, in population, the null and alternative models are expected to be equally accurate, on average, over the forecast sample (we computed the scaling factor using the population moments implied by the model and section 2.5's asymptotic approximations). For example, with T and P at 80, this multiplying constant is 0.4118.

DGP 4, which incorporates a forecast horizon of four periods, is also based on the empirical relationship between the change in core PCE inflation (y_t) and the Chicago Fed's index of the business cycle. In this case, though, the model is based on empirical estimates using (changes in) the four-quarter rate of inflation:[11]

$$\begin{aligned} y_{t+4} &= b_{11}x_{1,t} + e_{t+4} \\ e_{t+4} &= u_{t+4} + .95u_{t+3} + .9u_{t+2} + .8u_{t+1} \\ x_{1,t+4} &= 0.7x_{1,t+3} + v_{1,t+4} \\ \text{var}\begin{pmatrix} u_{t+4} \\ v_{1,t+4} \end{pmatrix} &= \begin{pmatrix} 0.2 & \\ 0.0 & 0.3 \end{pmatrix}. \end{aligned} \tag{14}$$

In these experiments, the forecasting models are:

$$\text{null: } y_{t+4} = \beta_0 + u_{0,t+4} \tag{15}$$

$$\text{alternative: } y_{t+4} = \beta_0 + \beta_1 x_{1,t} + u_{1,t+4}. \tag{16}$$

Again, in size experiments, the coefficient b_{11} on $x_{1,t}$ is set to a value that makes the models equally accurate (in expectation) on average over the forecast sample (on the basis of the model-implied population moments and section 2.5's asymptotic approximations). For example, with recursive forecasts and T and P both equal to 80, this value is 0.1634. In power simulations, the coefficient is set to its empirical value of 0.4, such that the alternative model is expected to be more accurate than the null.

DGP 5 takes the same form as DGP 3 (equation (11)), except that it incorporates multiplicative conditional heteroskedasticity in the error term of the equation for y:

$$u_{t+1} = \frac{|x_{1,t}|}{\sigma_{x,1}} e_{t+1},$$

where $\sigma_{x,1}$ denotes the standard deviation of $x_{1,t}$ implied by the DGP, and the variance-covariance matrix of e_{t+1} and the innovations $v_{i,t}$, $i = 1, \ldots, 3$, is set to match the covariance matrix given in equation (11). The forecasting models in DGP 5 experiments are the same as in DGP 3 (equations (12) and (13)).

With DGP 5, in the interest of brevity we consider only size experiments. Because conditional heteroskedasticity makes it very difficult to compute the population moments needed to determine the coefficient settings that imply equal accuracy, we rely instead on preliminary rounds of Monte Carlo simulations to set the b_{ij} coefficients to yield equal accuracy of the competing models. We begin with (empirically-based) coefficients of $b_{11} = 0.3$, $b_{21} = 0.1$, $b_{31} = 0.015$. For each T, P combination, we use our asymptotic theory assuming conditional homoskedasticity to determine a preliminary re-scaling of the coefficient vector

[11] Specifically, in the empirical estimates underlying the DGP settings, we defined $y_{t+4} = 100\ln(p_{t+4}/p_t) - 100\ln(p_t/p_{t-4})$, where p denotes the core PCE price index.

to yield equal accuracy. For each T, P combination, we then conduct three sets of Monte Carlo experiments (with a large number of draws), searching across grids of the re-scaling of the coefficient vector to select a scaling of the set of b_{ij} coefficients that minimizes the average (across Monte Carlo draws) difference in MSEs from the competing forecasting models.[12]

DGP 6 extends DGP 4 to include more predictands for y:

$$\begin{aligned}
y_{t+4} &= b_{11}x_{1,t} + b_{21}x_{2,t} + b_{31}x_{3,t} + e_{t+4} \\
e_{t+4} &= u_{t+4} + .95u_{t+3} + .9u_{t+2} + .8u_{t+1} \\
x_{1,t+4} &= 0.7x_{1,t+3} + v_{1,t+4} \\
x_{2,t+4} &= 0.8x_{2,t+3} + v_{2,t+4} \\
x_{3,t+4} &= 0.8x_{3,t+3} + v_{3,t+4} \\
\text{var}\begin{pmatrix} u_{t+4} \\ v_{1,t+4} \\ v_{2,t+4} \\ v_{3,t+4} \end{pmatrix} &= \begin{pmatrix} 0.2 & & & \\ -0.01 & 0.3 & & \\ 0.03 & 0.03 & 2.2 & \\ -0.2 & 0.02 & 0.8 & 9.0 \end{pmatrix}.
\end{aligned}$$

In the DGP 6 experiments, the forecasting models are:

$$\text{null:} \quad y_{t+4} = \beta_0 + u_{0,t+4} \tag{17}$$

$$\text{alternative:} \quad y_{t+4} = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + u_{1,t+4}. \tag{18}$$

In the interest of brevity, we consider only size results for DGP 6. For these experiments, we use our asymptotic theory to determine an initial scaling of the set of b_{ij} coefficients that implies equal accuracy in the finite sample given T and P, with the coefficients initially set (before scaling) to $b_{11} = 0.4$, $b_{21} = 0.2$, $b_{31} = 0.05$ (based roughly on empirical estimates). Using the approach described for DGP 5, we conduct three preliminary rounds of Monte Carlo simulations to refine the coefficient settings to make the competing forecasts equally accurate in the finite sample.

Finally, we also consider size experiments with a **DGP 7** in which the competing forecasting models under consideration are both misspecified. The data-generating portion of

[12] Specifically, we first consider 11 different experiments, each using 20,000 draws and a modestly different set of coefficient values obtained by scaling the baseline values, using a grid of scaling factors. We then pick the coefficient scaling that yields the lowest (in absolute value) average (across draws) difference in MSEs. We then repeat the 11-experiment exercise. Finally, we consider a third set of 21 experiments, with a more refined grid of coefficient scaling values and 200,000 draws. The coefficient scaling value that yields the smallest (absolute) difference in MSEs in this third set of experiments is then used to set the coefficients in the DGP simulated for the purpose of evaluating test properties.

this DGP takes the same form as DGP 3:

$$
\begin{aligned}
y_{t+1} &= -0.4y_t - 0.1y_{t-1} + b_{11}x_{1,t} + b_{21}x_{2,t} + b_{31}x_{3,t} + u_{t+1} \\
x_{1,t+1} &= 0.7x_{1,t} + v_{1,t+1} \\
x_{2,t+1} &= 0.9x_{2,t} - 0.2x_{2,t-1} + v_{2,t+1} \\
x_{3,t+1} &= 1.1x_{3,t} - 0.3x_{3,t-1} + v_{3,t+1}
\end{aligned}
\tag{19}
$$

$$
\text{var}\begin{pmatrix} u_t \\ v_{1,t+1} \\ v_{2,t+1} \\ v_{3,t+1} \end{pmatrix} = \begin{pmatrix} 0.8 & & & \\ 0.0 & 0.3 & & \\ -0.1 & 0.0 & 2.2 & \\ 0.5 & 0.1 & 0.8 & 9.0 \end{pmatrix}.
$$

To make the 1-step ahead forecasting models misspecified, we use the forecasting equations of DGP 2. These forecasting equations both exclude the $x_{2,t}$ and $x_{3,t}$ that are in the data-generating process for the predictand y_t:

$$
\text{null:}\quad y_{t+1} = \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + u_{0,t+1} \tag{20}
$$

$$
\text{alternative:}\quad y_{t+1} = \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + \beta_3 x_{1,t} + u_{1,t+1}. \tag{21}
$$

In setting the b_{ij} coefficients to yield equal accuracy of the competing, misspecified forecasting models, we fixed the coefficients on $x_{2,t-1}$ and $x_{3,t-1}$ at $b_{21} = 0.1$ and $b_{31} = 0.015$. To set the value of the coefficient b_{11} on $x_{1,t-1}$, we relied on on preliminary rounds of Monte Carlo simulations. We conducted three sets of Monte Carlo experiments (with a large number of draws), searching across a grid of values for b_{11} and choosing the value that minimized the average (across Monte Carlo draws) difference in MSEs from the competing forecasting models. This value varied across experiments with different settings of T and P. For example, with T and P both equal to 80, this value is 0.1185.

3.2 Results

Our interest lays in identifying those testing approaches that yield reasonably accurate inferences on the forecast performance of models. At the outset, then, it may be useful to broadly summarize the forecast performance of competing models under various alternatives. Accordingly, Figure 1 shows estimated densities of the MSE ratio statistic (the ratio of the null model's MSE to the alternative model's MSE), based on experiments with DGP 3, using $T = P = 80$. We provide three densities, for the cases in which the b_{ij} coefficients of the DGP (11) are: (i) set to 0, such that the null model should be more accurate; (ii) set to non–zero values so as to make the null and alternative models (9) and (10) equally accurate over the forecast sample, according to our local–to–zero asymptotic results; and (iii) set at larger values, such that the alternative model is expected to be more accurate.

As the figure shows, for the DGP which implies the null model should be best, the MSE ratio distribution mostly lays below 1.0. For the DGP that implies the models can be expected to be equally accurate, the distribution is centered at about 1.0. Finally, for the DGP that implies the alternative model can be expected to be best, the distribution mostly lays above 1.0. Our proposed fixed regressor bootstrap is intended to estimate a null distribution like that shown for the equally good models DGP. In most of our results, the null will be rejected when the sample MSE ratio lays in the right tail of the bootstrapped distribution.

3.2.1 Size results: recursive forecasts

Table 1 presents results for DGPs in which the b_{ij} coefficients on some x variables are non–zero but small enough that, under our asymptotic approximation, the null and alternative forecasting models are expected to be equally accurate over the sample considered. These size results show that, for testing the null of equal forecast accuracy, our proposed fixed regressor procedure is quite reliable, in the sense of yielding correctly-sized tests.

Tests based on the fixed regressor bootstrap generally have rejection rates of about 10% (the nominal size). For example, in the case of the MSE-F test applied to 1-step ahead forecasts, rejection rates range from 8.3% to 10.7%. Admittedly, rejection rates for 4-step ahead forecast tests are modestly higher, ranging from 12.4% to 14.8% percent.[13] For multi-step horizons, using the fixed regressor bootstrap works better (yielding rates closer to nominal size) when T is relatively large than when T is relatively small. Rejection rates for the MSE-t test compared against critical values from the fixed regressor bootstrap are similar, although a bit lower, ranging from 7.7% to 9.6% at the 1-step horizon and from 11.3% to 13.6% at the 4-step horizon.

Tests based on the other bootstrap intended to test the null of equal accuracy, the non–parametric bootstrap, are somewhat — although not entirely — less reliable indicators of equal accuracy. With critical values from the non-parametric bootstrap, the MSE-F test is somewhat undersized at the 1-step horizon but correctly sized or somewhat oversized at

[13] The over-sizing of the fixed regressor bootstrap at the 4-step horizon most likely has to do with the HAC estimation of the variance matrix V that determines the coefficient rescaling factor.

the 4-step horizon. As shown in Table 1, the MSE-F test's rejection rate ranges from 4.1% to 8.4% at the 1-step horizon and from 9.1% to 16.2% at the 4-step horizon. With the non-parametric approach, empirical rejection rates generally rise as P/T falls. For example, with 4-step ahead forecasts (for DGP 4) and $T = 80$, the MSE-F rejection rate is 9.4% when $P = 120$ and 15.6% when $P = 40$. Rejection rates for the MSE-t test compared against critical values from the non-parametric bootstrap are similar, although typically a bit higher, ranging from 5.0% to 10.1% at the 1-step horizon and from 9.4% to 15.2% at the 4-step horizon.

In addition, comparing the MSE-t test against standard normal critical values (with a one-sided testing approach) yields results similar to those obtained by comparing the test statistic against critical values from the non-parametric bootstrap. For instance, at the 1-step horizon, MSE-t rejection rates range from 4.7% to 9.4% under standard normal critical values, compared to a range of 5.0% to 10.1% under the non-parametric bootstrap. Accordingly, the MSE-t test compared against standard normal critical values is somewhat undersized at the 1-step horizon but correctly or somewhat oversized at the 4-step horizon.

Table 2 presents results from some additional experiments, with DGPs 5-7, that address the effectiveness of our proposed bootstrap in forecasting applications for which we are unable to prove the asymptotic validity of the bootstrap or the forecasting models are misspecified. In the DGP 5 and 6 experiments, the larger forecasting model has three more variables than the smaller model (so β_w is not scalar), and the forecast errors are either conditionally heteroskedastic (DGP 5, which extends DGP 3 to include conditional heteroskedasticity) or serially correlated (DGP 6, which extends DGP 4 to include more variables). The DGP 7 experiments address the case of misspecification of the forecasting models, using a data-generating process taken from DGP 3 but (misspecified) forecasting models taken from DGP 2.

The results of the additional experiments with DGPs 5 and 6 provided in Table 2 indicate that our fixed regressor bootstrap continues to perform well, in line with the baseline results of Table 1. In the case of the MSE-F test applied to 1-step ahead forecasts (DGP 5), rejection rates range from 9.7% to 12.0%. For the MSE-t test applied to 1-step ahead forecasts (DGP 5), rejection rates range from 9.5% to 10.3%. Again, rejection rates for 4-step ahead forecast tests are somewhat higher than for 1-step ahead forecasts. With DGP 6, the sizes of the MSE-F and MSE-t tests range from 13.4% to 19.3% and from 12.0% to

17.0%, respectively. The oversizing is modestly greater with DGP 6 (for which $k_w = 3$) than DGP 4 (for which $k_w = 1$). With both of these DGPs, the size of the fixed regressor-based tests improves as T increases (in some unreported experiments with DGP 6, we have verified that size improves further with even larger sample sizes). The modest difference in results between DGP 4 and DGP 6 is most likely due to the additional regressors in DGP 6 (relative to DGP 4) further reducing the finite-sample precision of the HAC estimation of the variance matrix V that determines the coefficient rescaling factor used in the bootstrap.

The performance of tests based on other sources of critical values is qualitatively the same in DGPs 5 and 6 (Table 2) as in DGPs 1-4 (Table 1). The key difference across the sets of experiments is that conditional heteroskedasticity raises and, in most cases, improves the sizes of MSE-F and MSE-t tests based on the non-parametric bootstrap and MSE-t tests based on standard normal critical values.[14] The rejection rate for the MSE-F test compared against critical values from the non-parametric bootstrap ranges from 7.8% to 11.2% in DGP 5, compared to 4.1% to 8.0% in DGP 3. The rejection rate for the MSE-t test compared against critical values from the non-parametric bootstrap ranges from 9.4% to 14.3% in DGP 5, compared to 5.0% to 10.0% in DGP 3 (the pattern is similar with standard normal critical values).

Introducing misspecification of the forecast models, as in the experiments with DGP 7, doesn't significantly affect the performance of our proposed testing approach. The size of the MSE-F and MSE-t tests based on our bootstrapped critical values (obtained with our fixed regressor approach) is very similar to their performance in correctly specified forecasting models. The tests are roughly correctly sized, with rejection rates for the MSE-F test across T, P combinations ranging from 10.6% to 12.0%, and rates for the MSE-t test ranging from 9.2% to 10.4%. Similarly, using critical values obtained with the non-parametric bootstrap or from the normal distribution yields results qualitatively very similar to those in the baseline experiments.

3.2.2 Size results: rolling forecasts

Table 3 provides size results for experiments using a rolling forecast scheme instead of the baseline recursive scheme, based on models parameterized to make the null and alternative forecasting models equally accurate (the necessary scaling factor in the DGP is a bit different

[14]This finding does not appear to be dependent on having k_w exceed 1. In unreported experiments, we obtained a similar result in a version of DGP 2 with conditional heteroskedasticity.

in the rolling case than the recursive). In general, the results for the rolling scheme are very similar to those for the recursive. Tests based on our fixed regressor bootstrap have size of about 10% (the nominal size), although with some slight to modest oversizing at the 4-step horizon. Tests based on the non-parametric bootstrap or standard normal critical values continue to be undersized at the 1-step horizon, although the problem is a bit worse under the rolling scheme than the recursive.[15] For example, with DGP 2, $T = 40$, and $P = 80$, comparing the MSE-t test against critical values estimated with the non-parametric bootstrap yields a rejection rate of 6.5% for recursive forecasts (Table 1) and 4.9% for rolling forecasts (Table 3); comparing the test against fixed regressor bootstrap critical values yields corresponding rejection rates of 8.8% (recursive) and 8.6% (rolling). At the 4-step horizon, tests based on the non-parametric bootstrap or standard normal critical values continue to range from correctly sized to oversized, with oversizing that is sharpest when P is small.

Our rolling scheme results on the behavior of the MSE-t test compared against non-parametric bootstrap and standard normal critical values are somewhat at odds with the behavior of the test in Giacomini and White (2006). Giacomini and White (2006) compare the MSE-t test against standard normal critical values, and find a two-sided test to be roughly correctly sized at the one-step forecast horizon, with small-to-modest undersizing for some sample sizes and comparable oversizing for others. One source of differences in results is our treatment of the test as one-sided rather than two-sided. Giacomini and White (2006) permit rejections of the alternative model in favor of the null and conduct two-sided tests; we prefer to take the small model as the null and only consider rejections of the null in favor of the alternative, or one-sided tests. When we use a two-sided MSE-t test and standard normal critical values (while not shown in the interest of brevity, the same applies with critical values from the non-parametric bootstrap), the test is roughly correctly sized at the 1-step horizon and correctly sized to somewhat oversized at the 4-step horizon (the same applies in the recursive forecast results of Table 1). The increase in rejection rates that occurs with the move from a one-sided to two-sided test likely reflects an empirical distribution that is shifted to the left relative to the standard normal.

Admittedly, though, other aspects of our Monte Carlo results seem to be at odds with the asymptotic results of Giacomini and White (2006), if not their Monte Carlo results.

[15] The rise in rejection rates that occurs as P/T falls is a bit sharper in the rolling case than the recursive. As a consequence, the differences in rejection rates (based on the non-parametric bootstrap or standard normal critical values) across the recursive and rolling forecasting schemes are larger when P/T is relatively big than when it is relatively small.

Their asymptotics imply the MSE-t test has an asymptotic distribution that is standard normal for rolling forecasts but not recursive forecasts, suggesting the test should have better size properties in the rolling case than the recursive. But in our Monte Carlo results, the standard normal approximation for MSE-t seems to work better with recursive forecasts than rolling, yielding 1-step ahead rejection rates closer to nominal in the former case than the latter. In addition, their theory rests on asymptotics that treat T as fixed and P as limiting to infinity, which suggests the test should behave better when P is large relative to T than when P is relatively small. In fact, in our Monte Carlo results, rejection rates based on the non-parametric bootstrap and standard normal critical values tend to be farther from nominal size when P is large than when it is small. In the case of the second issue, the Monte Carlo results in Giacomini and White (2006) seem to yield a similar pattern, with rejection rates falling as the forecast sample increases relative to the estimation sample, often to levels consistent with the undersizing we have reported.

3.2.3 Power results: recursive forecasts

Table 4 provides results for DGPs in which the b_{ij} coefficients on some x variables are large enough that, under our finite-sample asymptotics, the alternative model is expected to be more accurate than the null model in the finite sample.

Comparing the test statistics against critical values estimated with the fixed regressor bootstrap seems to yield relatively good power. In the case of the MSE-F test, rejection rates range from from 42.8% to 82.1%. Comparing tests against distributions estimated with the non–parametric bootstrap yields materially lower power. In Table 4's results, using the non–parametric bootstrap for the MSE-F test yields a rejection rate between 25.0% and 56.9%.

Rejection rates for the MSE-t test are broadly similar to those for the MSE-F test, although with some noticeable differences. In most cases in Table 4's results, the MSE-t test is less powerful than the MSE-F test (as with the fixed regressor bootstrap), but in some cases (as with the non–parametric bootstrap), the MSE-t test is more powerful.

3.2.4 Results summary

Overall, the Monte Carlo results show that, for testing equal forecast accuracy over a given sample, our proposed fixed regressor bootstrap works well. When the null of equal accuracy in the finite sample is true, the testing procedures yield approximately correctly sized tests.

When an alternative model is, in truth, more accurate than the null, the testing procedures seem to have reasonable power. The non–parametric bootstrap procedure, which just re–samples the data without imposing the equal accuracy null in the data generation, tends to be less reliable when applied to nested forecasting models.

4 Applications

In this section we use the tests and inference approaches described above in forecasting excess stock returns and core inflation, both for the U.S. Some recent examples from the long literature on stock return forecasting include Rapach and Wohar (2006), Goyal and Welch (2008), and Campbell and Thompson (2008). Some recent inflation examples include Atkeson and Ohanian (2001) and Stock and Watson (2003).

More specifically, in the stock return application, we use the data of Goyal and Welch (2008) and examine forecasts of monthly excess stock returns (CRSP excess returns measured on a log basis) from a total of 17 models. The null model includes just a constant. The alternative models add in one lag of a common predictor, taken from the set of variables in the Goyal-Welch data set available over all of our sample.[16] These include, among others, the dividend-price ratio, the earnings-price ratio, and the cross-sectional premium. The full set of 16 predictive variables is listed in Table 5, with details provided in Goyal and Welch (2008). Following studies such as Pesaran and Timmermann (1995), we focus on the post-war period. Our model estimation sample begins with January 1954, and we examine recursive 1-month ahead forecasts (that is, our estimation sample expands as forecasting moves forward in time) for 1970 through 2002.

In the inflation application, we examine 1-quarter ahead and 4-quarter ahead forecasts of core PCE inflation obtained from a few models, over a sample of 1985:Q1+horizon-1 to 2008:Q2. The null model includes a constant and lags of the change in inflation. One alternative model adds one lag of the CFNAI to the baseline model. Another includes one lag of the CFNAI, PCE food price inflation less core inflation, and import price inflation less core inflation.[17] We specify the models in terms of the change in inflation, following, among others, Stock and Watson (1999, 2003) and Clark and McCracken (2006). In one application, we consider 1-quarter ahead forecasts of inflation defined as $\pi_t = 400 \ln(P_t/P_{t-1})$, using

[16]We obtained the data from Amit Goyal's website.

[17]We obtained the CFNAI data from the Chicago Fed's website and the rest of the data from the FAME database of the Federal Reserve Board of Governors.

models relating $\Delta\pi_{t+1}$ to a constant, $\Delta\pi_t$, $\Delta\pi_{t-1}$, and the period t values of the CFNAI, relative food price inflation, and relative import price inflation. In another, we consider 4-quarter ahead forecasts of inflation defined as $\pi_t^{(4)} = 100\ln(P_t/P_{t-4})$, using models relating $\pi_{t+4}^{(4)} - \pi_t^{(4)}$ to a constant, $\pi_t^{(4)} - \pi_{t-4}^{(4)}$, and the period t values of the CFNAI, relative food price inflation, and relative import price inflation. To simplify the lag structure necessary for reasonable forecasting models, the (relative) food and import price inflation variables are computed as two-period averages of quarterly (relative) inflation rates. For both inflation forecast horizons, our model estimation sample uses a start date of 1968:Q3.

Results for the stock return and inflation forecast applications are reported in Tables 5 and 6. The tables provide, for each alternative model, the ratio of the MSE of forecasts from the benchmark model to the alternative model's forecast MSE. The tables also provide, for the MSE-F test, p-values obtained under three different approaches.[18] Two of these approaches are the same ones included in last section's Monte Carlo analysis: the non-parametric bootstrap and our proposed fixed regressor bootstrap.

A third approach — a no-predictability version of the fixed regressor bootstrap — is included to illustrate some differences in testing equal accuracy in the finite sample versus equal accuracy in population. The theoretical results in Clark and McCracken (2012) on comparisons of forecasts from multiple nested models establish the asymptotic validity of the no-predictability fixed regressor bootstrap for the null of equal forecast accuracy in population, which is equivalent to a null hypothesis of $\beta_w = 0$. This no-predictability fixed regressor bootstrap takes the same form as described in sections 2, with the sole difference being that in step 1, $\hat{d} = 0$, which is equivalent to simply estimating the null forecasting model by OLS (model 0, which includes only the variables $x_{0,t}$) rather than the alternative model (model 1, which includes the variables $x_{0,t}$ and $x_{w,t}$).

In the case of excess stock returns, the evidence in Table 5 is consistent with much of the literature: return predictability is limited. Of the 16 alternative forecasting models, only two — the first two in the table — have MSEs lower than the benchmark (that is, MSE ratios greater than 1). The no-predictability fixed regressor bootstrap p-values reject the null model in favor of the alternative for each of these two models. These test results indicate the predictor coefficients on the cross-sectional premium and return on long-term Treasuries are non–zero. However, p-values based on the fixed regressor bootstrap imply

[18] In all three approaches, we use 9999 replications in computing the bootstrap p-values.

weaker evidence of forecastability, with the null of equal forecast accuracy rejected for the cross-sectional premium, but not the Treasury return (at a 10% significance level). This pattern suggests that, while the coefficient on the Treasury return may differ from zero, the coefficient is not large enough that a model including the Treasury return would be expected to forecast better than the null model over a sample of the size considered. Critical values based on the non-parametric bootstrap yield no rejections, presumably (given our Monte Carlo evidence) reflecting lower power.

The inflation results reported in Table 6 yield similarly mixed evidence of predictability. By itself, the CFNAI improves the accuracy of 1-quarter ahead forecasts but not 4-quarter ahead forecasts. At the 1-step horizon, the no-predictability fixed regressor bootstrap p-values reject the null model in favor of the alternative — indicating the predictor coefficients on the CFNAI to be non–zero. However, p-values based on the fixed regressor bootstrap fail to reject the null of equal accuracy. So while the coefficient on the CFNAI may differ from zero, it is not large enough that a model including the CFNAI would be expected to forecast better than the null model in a sample of the size considered. Including not only the CFNAI but also relative food and import price inflation yields larger gains in forecast accuracy, at both horizons. In this case, critical values from both the no-predictability fixed regressor and fixed regressor bootstrap reject the null (at a 10% significance level). This suggests the relevant coefficients are non-zero and large enough to make the alternative model more accurate than the null. Here, too, critical values based on the non-parametric bootstrap yield fewer rejections.

5 Conclusion

This paper develops bootstrap methods for testing whether, in a finite sample, competing out-of-sample forecasts from nested models are equally accurate. Most prior work on forecast tests for nested models has focused on a null hypothesis of equal accuracy in population — basically, whether coefficients on the extra variables in the larger, nesting model are zero. We instead use an asymptotic approximation that treats the coefficients as non-zero but small, such that, in a finite sample, forecasts from the small model are expected to be as accurate as forecasts from the large model. While an unrestricted, correctly specified model might have better population-level predictive ability than a misspecified restricted model, it need not do so in finite samples due to imprecision in the additional parameter estimates.

In the presence of these "weak" predictors, we show how to test the null of equal average predictive ability over a given sample size.

Under our asymptotic approximation of weak predictive ability, we first derive the asymptotic distributions of two tests for equal out-of-sample predictive ability. We then develop a parametric bootstrap procedure — a fixed regressor bootstrap — for testing the null of equal finite-sample forecast accuracy. We next conduct a range of Monte Carlo simulations to examine the finite–sample properties of the tests and bootstrap procedures. Our proposed fixed regressor bootstrap works reasonably well: When the null of equal finite-sample predictive ability is true, the testing procedure yields approximately correctly sized tests. Moreover when an alternative model is, in truth, more accurate than the null, the testing procedure has reasonable power. In contrast, when applied to nested models, the non-parametric method of White (2000) often does not work as well, in a size or power sense.

In the final part of our analysis, we apply our proposed methods for testing equal predictive ability to forecasts of excess stock returns and core inflation, using U.S. data. In both applications, our methods for testing equal finite sample accuracy yield weaker evidence of predictability than do methods for testing equal population-level accuracy. There remains some evidence, but only modest. Using non-parametric bootstrap methods that are technically invalid with nested models — methods that have poorer size and power properties — yields much less evidence of predictability.

6 Appendix: Theory Details

In this section we provide proofs of the theorems described in the text. We also include a lemma, and its proof, that is used in the proofs of Theorems 2.3 and 2.4.

In the following, in addition to the notation from Section 2, let $\sup_t |.|$ denote $\sup_{T\leq t\leq T+P-\tau} |.|$ and define $v^*_{T,1,s+\tau} = (\eta_{s+\tau}\varepsilon_{T,1,s+\tau} + \theta_1\eta_{s+\tau-1}\varepsilon_{T,1,s+\tau-1} + \cdots + \theta_{\tau-1}\eta_{s+1}\varepsilon_{T,1,s+1})$, $\widehat{v}^*_{T,1,s+\tau} = (\eta_{s+\tau}\widehat{\varepsilon}_{T,1,s+\tau}+\widehat{\theta}_1\eta_{s+\tau-1}\widehat{\varepsilon}_{T,1,s+\tau-1}+\cdots+\widehat{\theta}_{\tau-1}\eta_{s+1}\widehat{\varepsilon}_{T,1,s+1})$, $h^*_{T,1,s+\tau} = x_{T,1,s}v^*_{T,1,s+\tau}$, and $\widehat{h}^*_{T,1,s+\tau} = x_{T,1,s}\widehat{v}^*_{T,1,s+\tau}$. For the recursive scheme define $H^*_{T,1}(t) = t^{-1}\sum_{s=1}^{t-\tau} h^*_{T,1,s+\tau}$ and $\hat{H}^*_{T,1}(t) = t^{-1}\sum_{s=1}^{t-\tau} \hat{h}^*_{T,1,s+\tau}$ while for the rolling scheme define $H^*_{T,1}(t) = T^{-1}\sum_{s=t-T-\tau+1}^{t-\tau} h^*_{T,1,s+\tau}$ and $\hat{H}^*_{T,1}(t) = T^{-1}\sum_{s=t-T-\tau+1}^{t-\tau} \hat{h}^*_{T,1,s+\tau}$. More generally, we let the superscript * denote a property, such as convergence in distribution $\to^{d^*}$, defined with respect to the bootstrap-induced probability measure (again, more specifically, convergence occurs in a set with probability limiting to 1). Throughout we ignore the finite sample difference between P and $P-\tau+1$.

Proof of Theorem 2.1: The result is a special case of Theorem 1 of Clark and McCracken (2009) and as a result, we provide only an outline of the proof here. The proof consists of two steps. For the first step, straightforward algebra reveals that

$$\begin{aligned}
&\sum\nolimits_{t=T}^{T+P-\tau}(\hat{u}^2_{0,t+\tau}-\hat{u}^2_{1,t+\tau}) \qquad (22)\\
=\ &\{2\sum\nolimits_{t=T}^{T+P-\tau}(T^{-1/2}h'_{T,1,t+\tau})(-JB_0(t)J'+B_1(t))(T^{1/2}H_{T,1}(t))\\
&-T^{-1}\sum\nolimits_{t=T}^{T+P-\tau}(T^{1/2}H'_{T,1}(t))(-JB_0(t)x_{T,0,t}x'_{T,0,t}B_0(t)J'\\
&+B_1(t)x_{T,1,t}x'_{T,1,t}B_1(t))(T^{1/2}H_{T,1}(t))\}\\
&+2\{\sum\nolimits_{t=T}^{T+P-\tau}\delta' B_1^{-1}(t)(-JB_0(t)J'+B_1(t))(T^{-1/2}h_{T,1,t+\tau})\}\\
&+\{T^{-1}\sum\nolimits_{t=T}^{T+P-\tau}\delta'(x_{T,1,t}x'_{T,1,t}-2x_{T,1,t}x'_{T,1,t}JB_0(t)J'B_1^{-1}(t)\\
&+B_1^{-1}(t)JB_0(t)x_{T,0,t}x'_{T,0,t}B_0(t)J'B_1^{-1}(t))\delta\}\\
&+2\{T^{-1}\sum\nolimits_{t=T}^{T+P-\tau}\delta'(B_1^{-1}(t)JB_0(t)x_{T,0,t}x'_{T,0,t}B_0(t)J'\\
&-x_{T,1,t}x'_{T,1,t}JB_0(t)J')(T^{1/2}H_{T,1}(t))\}.
\end{aligned}$$

Given Assumptions 2 and 3, straightforward moment-based bounding arguments, along with the definitions of $\tilde{A}$, $\tilde{h}_{T,1,t+\tau}$, and $\tilde{H}_{T,1}(t)$ imply

$$\begin{aligned}
&\sum\nolimits_{t=T}^{T+P-\tau}(\hat{u}^2_{0,t+\tau}-\hat{u}^2_{1,t+\tau}) = \sigma^2\{2\sum\nolimits_{t=T}^{T+P-\tau}(T^{-1/2}\tilde{h}_{T,1,t+\tau})(T^{1/2}\tilde{H}_{T,1}(t))\\
&-T^{-1}\sum\nolimits_{t=T}^{T+P-\tau}(T^{1/2}\tilde{H}_{T,1}(t))(T^{1/2}\tilde{H}_{T,1}(t))\} + \sigma^2\{2\sum\nolimits_{t=T}^{T+P-\tau}(\delta' B_1^{-1/2}\tilde{A}/\sigma)(T^{-1/2}\tilde{h}_{T,1,t+\tau})\}\\
&+\sigma^2\{(P/T)(\beta'_w F_1^{-1}\beta'_w/\sigma^2)\} + o_p(1).
\end{aligned}$$

For the second step we apply weak convergence results from de Jong and Davidson (2000), notably Theorems 3.2 and 4.1. Taking limits, and noting that $T^{1/2}\tilde{H}_{T,1}(t) \Rightarrow s^{-1}S_{\tilde{h}\tilde{h}}^{1/2}W(s)$ we obtain

$$\begin{aligned}
&\sum\nolimits_{t=T}^{T+P-\tau}(\hat{u}_{0,t+\tau}^2 - \hat{u}_{1,t+\tau}^2) \rightarrow^d \\
&\sigma^2\{2\int_1^{1+\lambda_P} s^{-1}W'(s)S_{\tilde{h}\tilde{h}}dW(s) - \int_1^{1+\lambda_P} s^{-2}W'(s)S_{\tilde{h}\tilde{h}}W(s)ds\} \\
&+\sigma^2\{\int_1^{1+\lambda_P}(\delta' B_1^{-1/2}\tilde{A}'/\sigma)S_{\tilde{h}\tilde{h}}^{1/2}dW(s)\} + \sigma^2\{\lambda_P\beta_w' F_1^{-1}\beta_w'/\sigma^2\}.
\end{aligned}$$

That $\text{MSE}_2 \rightarrow^p \sigma^2$ then provides the desired result.

To establish a similar result for the rolling, rather than recursive, estimation scheme, the only important difference is that instead of $H_{T,1}(t) = (t^{-1}\sum_{s=1}^{t-\tau} h_{T,1,s+\tau})$ for the recursive scheme we now have $H_{T,1}(t) = (T^{-1}\sum_{s=t-\tau-T+1}^{t-\tau} h_{T,1,s+\tau})$ for the rolling scheme. Hence in the final step of the proof for the recursive scheme we have $T^{1/2}\tilde{H}_{T,1}(t) \Rightarrow s^{-1}S_{\tilde{h}\tilde{h}}^{1/2}W(s)$ whereas for the rolling scheme we would have $T^{1/2}\tilde{H}_{T,1}(t) \Rightarrow S_{\tilde{h}\tilde{h}}^{1/2}(W(s) - W(s-1))$. Other differences are minor and omitted for brevity.

Proof of Theorem 2.2: Given Theorem 2.1(a) and the Continuous Mapping Theorem it suffices to show that $P\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)\widehat{\gamma}_{LL}(j) \rightarrow^d 4\sigma^4(\Gamma_5 + \Gamma_6 + \Gamma_7)$. Before doing so it is convenient to redefine the bracketed terms from (14) used in the primary decomposition of the loss differential in the proof of Theorem 2.1(absent the summations, but keeping the brackets) as

$$(\hat{u}_{0,t+\tau}^2 - \hat{u}_{1,t+\tau}^2) = \{2A_{1,t} - A_{2,t}\} + 2\{B_t\} + \{C_t\} + 2\{D_t\}. \tag{23}$$

With this in mind we obtain

$$\begin{aligned}
&P\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)\widehat{\gamma}_{LL}(j) = \sum_{j=-\bar{j}}^{\bar{j}} K(j/M)\sum\nolimits_{t=T+j}^{T+P-\tau}(\hat{u}_{0,t+\tau}^2 - \hat{u}_{1,t+\tau}^2)(\hat{u}_{0,t-j+\tau}^2 - \hat{u}_{1,t-j+\tau}^2) \qquad (24)\\
= \; &4\{\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)\sum\nolimits_{t=T+j}^{T+P-\tau} A_{1,t}A_{1,t-j}\} + 4\{\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)\sum\nolimits_{t=T+j}^{T+P-\tau} A_{1,t}B_{t-j}\} \\
&+4\{\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)\sum\nolimits_{t=T+j}^{T+P-\tau} B_tB_{t-j}\} \\
&+ \text{other cross products of } A_{1,t},\ A_{2,t},\ B_t,\ C_t,\ D_t \text{ with } A_{1,t-j},\ A_{2,t-j},\ B_{t-j},\ C_{t-j},\ D_{t-j}.
\end{aligned}$$

In the remainder we show that each of the 3 bracketed terms in (16) converge to σ^4 times Γ_5, Γ_6, and Γ_7, respectively, and that the other cross product terms are each $o_p(1)$.

For the first bracketed term in (16), if we recall the definition of $\tilde{h}_{T,1,t+\tau}$ and note that $\bar{j}$ is finite, algebra along the lines of Clark and McCracken (2005) gives us

$$\sum_{j=-\bar{j}}^{\bar{j}} K(j/M) \sum_{t=T+j}^{T+P-\tau} A_{1,t} A_{1,t-j}$$

$$= \sigma^4 \sum_{j=-\bar{j}}^{\bar{j}} K(j/M) T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2} H'_{T,1}(t) B_1^{1/2}/\sigma) B_1^{-1/2}(-JB_0(t)J' + B_1(t)) B_1^{-1/2} \times$$

$$(B_1^{1/2} h_{T,1,t+\tau} h'_{T,1,t-j+\tau} B_1^{1/2}/\sigma^2) B_1^{-1/2}(-JB_0(t-j)J' + B_1(t-j)) B_1^{-1/2}(T^{1/2} B_1^{1/2} H_{T,1}(t-j)/\sigma)$$

$$= \sigma^4 \sum_{j=-\bar{j}}^{\bar{j}} K(j/M) T^{-1} \sum_{t=T}^{T+P-\tau} (T^{1/2} H'_{T,1}(t) B_1^{1/2}/\sigma) B_1^{-1/2}(-JB_0 J' + B_1) B_1^{-1/2} \times$$

$$(B_1^{1/2} E h_{T,1,t+\tau} h'_{T,1,t-j+\tau} B_1^{1/2}/\sigma^2) B_1^{-1/2}(-JB_0 J' + B_1) B_1^{-1/2}(T^{1/2} B_1^{1/2} H_{T,1}(t)/\sigma) + o_p(1)$$

$$= \sigma^4 \sum_{j=-\bar{j}}^{\bar{j}} K(j/M) T^{-1} \sum_{t=T}^{T+P-\tau} (T^{1/2} \tilde{H}_{T,1}(t))(E\tilde{h}_{T,1,t+\tau} \tilde{h}'_{T,1,t-j+\tau})(T^{1/2} \tilde{H}_{T,1}(t)) + o_p(1)$$

$$= \sigma^4 (T^{-1} \sum_{t=T}^{T+P-\tau} [T^{1/2} \tilde{H}'_{T,1}(t) \otimes T^{1/2} \tilde{H}'_{T,1}(t)]) vec[\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)(E\tilde{h}_{T,1,t+\tau} \tilde{h}'_{T,1,t-j+\tau})] + o_p(1).$$

Given Assumptions 2 and 4, $\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)(E\tilde{h}_{T,1,t+\tau}\tilde{h}'_{T,1,t-j+\tau}) \to S_{\tilde{h}\tilde{h}}$. Since Assumption 2 and Theorem 3.2 of de Jong and Davidson (2000) suffice for $T^{1/2}\tilde{H}_{T,1}(t) \Rightarrow s^{-1} S_{\tilde{h}\tilde{h}}^{1/2} W(s)$, the Continuous Mapping Theorem implies

$$T^{-1} \sum_{t=T}^{T+P-\tau} T^{1/2} \tilde{H}'_{T,1}(t) \otimes T^{1/2} \tilde{H}'_{T,1}(t) \to^d \int_1^{1+\lambda_P} s^{-2}[W'(s) S_{\tilde{h}\tilde{h}}^{1/2} \otimes W'(s) S_{\tilde{h}\tilde{h}}^{1/2}] ds.$$

Since $(\int_1^{1+\lambda_P} s^{-2}[W'(s)S_{\tilde{h}\tilde{h}}^{1/2} \otimes W'(s)S_{\tilde{h}\tilde{h}}^{1/2}]ds) vec[S_{\tilde{h}\tilde{h}}] = \Gamma_5$, we obtain the desired result.

For the second bracketed term in (16), similar arguments give us

$$\sum_{j=-\bar{j}}^{\bar{j}} K(j/M) \sum_{t=T+j}^{T+P-\tau} A_{1,t} B_{1,t-j} =$$

$$\sigma^4 \sum_{j=-\bar{j}}^{\bar{j}} K(j/M) T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2} H'_{T,1}(t) B_1^{1/2}/\sigma) B_1^{-1/2}(-JB_0(t)J' + B_1(t)) B_1^{-1/2} \times$$

$$(B_1^{1/2} h_{T,1,t+\tau} h'_{T,1,t-j+\tau} B_1^{1/2}/\sigma^2) B_1^{-1/2}(-JB_0(t-j)J' + B_1(t-j)) B_1^{-1/2}(t-j)(B_1^{1/2}(t-j)\delta/\sigma)$$

$$= \sigma^4 \sum_{j=-\bar{j}}^{\bar{j}} K(j/M) T^{-1} \sum_{t=T}^{T+P-\tau} (T^{1/2} H'_{T,1}(t) B_1^{1/2}/\sigma) B_1^{-1/2}(-JB_0 J' + B_1) B_1^{-1/2} \times$$

$$(B_1^{1/2} E h_{T,1,t+\tau} h'_{T,1,t-j+\tau} B_1^{1/2}/\sigma^2) B_1^{-1/2}(-JB_0 J' + B_1) B_1^{-1/2}(B_1^{1/2}\delta/\sigma) + o_p(1)$$

$$= \sigma^4 \sum_{j=-\bar{j}}^{\bar{j}} K(j/M) T^{-1} \sum_{t=T}^{T+P-\tau} (T^{1/2} \tilde{H}_{T,1}(t))(E\tilde{h}_{T,1,t+\tau} \tilde{h}'_{T,1,t-j+\tau})(\tilde{A} B_1^{1/2}\delta/\sigma) + o_p(1)$$

$$= \sigma^4 (T^{-1} \sum_{t=T}^{T+P-\tau} [(\tilde{A} B_1^{1/2}\delta/\sigma)' \otimes T^{1/2} \tilde{H}'_{T,1}(t)]) vec[\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)(E\tilde{h}_{T,1,t+\tau} \tilde{h}'_{T,1,t-j+\tau})] + o_p(1).$$

Given Assumptions 2 and 4, $\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)(E\tilde{h}_{T,1,t+\tau}\tilde{h}'_{T,1,t-j+\tau}) \to S_{\tilde{h}\tilde{h}}$. Since Assumption 2 and Theorem 3.2 of de Jong and Davidson (2000) suffice for $T^{1/2}\tilde{H}_{T,1}(t) \Rightarrow s^{-1}S_{\tilde{h}\tilde{h}}^{1/2}W(s)$, the Continuous Mapping Theorem implies

$$T^{-1}\sum_{t=T}^{T+P-\tau}[(\tilde{A}B_1^{1/2}\delta/\sigma)' \otimes T^{1/2}\tilde{H}'_{T,1}(t)] \to^d \int_1^{1+\lambda_P} s^{-1}[(\tilde{A}B_1^{1/2}\delta/\sigma)' \otimes W'(s)S_{\tilde{h}\tilde{h}}^{1/2}]ds.$$

Since $(\int_1^{1+\lambda_P} s^{-1}[(\tilde{A}B_1^{1/2}\delta/\sigma)' \otimes W'(s)S_{\tilde{h}\tilde{h}}^{1/2}]ds)vec[S_{\tilde{h}\tilde{h}}] = \Gamma_6$, we obtain the desired result.

For the third bracketed term in (16), similar arguments give us

$$\begin{aligned}
&\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)\sum_{t=T+j}^{T+P-\tau} B_{1,t}B_{1,t-j} = \\
&\sigma^4\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)T^{-1}\sum_{t=T}^{T+P-\tau}(\delta' B_1^{1/2}(t)/\sigma)B_1^{-1/2}(t)(-JB_0(t)J'+B_1(t))B_1^{-1/2}\times \\
&(B_1^{1/2}h_{T,1,t+\tau}h'_{T,1,t-j+\tau}B_1^{1/2}/\sigma^2)B_1^{-1/2}(-JB_0(t-j)J'+B_0(t-j))B_1^{-1/2}(t-j)(B_1^{1/2}(t-j)\delta/\sigma) \\
=\;&\sigma^4\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)T^{-1}\sum_{t=T}^{T+P-\tau}(\delta' B_1^{1/2}/\sigma)B_1^{-1/2}(-JB_0J'+B_1)B_1^{-1/2}\times \\
&(B_1^{1/2}Eh_{T,1,t+\tau}h'_{T,1,t-j+\tau}B_1^{1/2}/\sigma^2)B_1^{-1/2}(-JB_0J'+B_1)B_1^{-1/2}(B_1^{1/2}\delta/\sigma)+o_p(1) \\
=\;&\sigma^4\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)T^{-1}\sum_{t=T}^{T+P-\tau}(\delta' B_1^{1/2}\tilde{A}'/\sigma)(E\tilde{h}_{T,1,t+\tau}\tilde{h}'_{T,1,t-j+\tau})(\tilde{A}B_1^{1/2}\delta/\sigma)+o_p(1) \\
=\;&\sigma^4((P/T)[(\delta' B_1^{1/2}\tilde{A}'/\sigma)\otimes(\delta' B_1^{1/2}\tilde{A}'/\sigma)])vec[\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)(E\tilde{h}_{T,1,t+\tau}\tilde{h}'_{T,1,t-j+\tau})]+o_p(1).
\end{aligned}$$

Given Assumptions 2 and 4, $\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)(E\tilde{h}_{T,1,t+\tau}\tilde{h}'_{T,1,t-j+\tau}) \to S_{\tilde{h}\tilde{h}}$. Since Assumption 3 implies $P/T \to \lambda_P$ and $(\lambda_P[(\delta' B_1^{1/2}\tilde{A}'/\sigma)\otimes(\delta' B_1^{1/2}\tilde{A}'/\sigma)])vec[S_{\tilde{h}\tilde{h}}] = \Gamma_7$, we obtain the desired result.

There are twelve remaining terms in (16) that are cross products of $A_{1,t}$, $A_{2,t}$, B_t, C_t, and D_t with $A_{1,t-j}$, $A_{2,t-j}$, B_{t-j}, C_{t-j}, and D_{t-j} for each j. That each are $o_p(1)$ follow comparable arguments. For brevity we show this for the term comprised of $A_{1,t}$ and $A_{2,t-j}$. For this term we have

$$\begin{aligned}
&|\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)\sum_{t=T}^{T+P-\tau} A_{1,t}A_{2,t-j}| = \\
&|\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)T^{-3/2}\sum_{t=T}^{T+P-\tau}(T^{1/2}H'_{T,1}(t))(-JB_0(t)J'+B_1(t))\times \\
&(h_{T,1,t+\tau}vec[-JB_0(t)x_{T,0,t}x'_{T,0,t}B_0(t)J'+B_1(t)x_{T,1,t}x'_{T,1,t}B_1(t)]'(T^{1/2}H_{T,1}(t-j)\otimes T^{1/2}H_{T,1}(t-j))| \\
\leq\;&2\bar{j}k^4T^{-1/2}(T^{-1}\sum_{t=T}^{T+P-\tau}|h_{T,1,t+\tau}vec[-JB_0(t)x_{T,0,t}x'_{T,0,t}B_0(t)J'+B_1(t)x_{T,1,t}x'_{T,1,t}B_1(t)]'|)\times \\
&(\sup_t|T^{1/2}H_{T,1}(t)|)^3(\sup_t|-JB_0(t)J'+B_1(t)|).
\end{aligned}$$

Assumptions 2 and 3, along with Theorem 3.2 of de Jong and Davidson (2000) suffice for $\sup_t|T^{1/2}H_{T,1}(t)| = O_p(1)$. Assumption 2 along with Markov's inequality imply both

$$T^{-1}\sum_{t=T}^{T+P-1} |h_{T,1,t+\tau}vec[-JB_0(t)x_{T,0,t}x'_{T,0,t}B_0(t)J' + B_1(t)x_{T,1,t}x'_{T,1,t}B_1(t)]'| = O_p(1)$$

and $\sup_t| - JB_0(t)J' + B_1(t)| = O_p(1)$. Since $\bar{j}$ and k are finite and $T^{-1/2} = o_p(1)$, the proof is complete.

(b) To establish a similar result for the rolling, rather than recursive, estimation scheme, the primary difference is that instead of $H_{T,1}(t) = (t^{-1}\sum_{s=1}^{t-\tau} h_{T,1,s+\tau})$ for the recursive scheme we now have $H_{T,1}(t) = (T^{-1}\sum_{s=t-\tau-T+1}^{t-\tau} h_{T,1,s+\tau})$ for the rolling scheme. Hence in each step of the proof for the recursive scheme where the fact that $T^{1/2}\tilde{H}_{T,1}(t) \Rightarrow s^{-1}S_{\tilde{h}\tilde{h}}^{1/2}W(s)$ is used, we instead use the fact that for the rolling scheme $T^{1/2}\tilde{H}_{T,1}(t) \Rightarrow S_{\tilde{h}\tilde{h}}^{1/2}(W(s) - W(s-1))$. Other differences are minor and omitted for brevity.

Lemma 1: Maintain Assumptions $1, 2$, and 3. (a) $\sup_t|T^{1/2}(B_j(t)-B_j)| = O_p(1)$ $j = 0, 1$, (b) If either (i) β_w is a local to zero parameter or (ii) $J'_w\hat{\beta}_{1,T} \to^p \beta_w \neq 0$, $T^{1/2}J'_w\widetilde{\beta}_{1,T} = O_p(1)$. Maintain Assumptions $1, 2'$ and 3. (c) $T^{-1/2}\sum_{s=1}^{t-\tau} h^*_{T,1,s+\tau} \Rightarrow^* \Omega_{11}^{1/2}W^*(s)$, (d) $\sum_{t=T}^{T+P-\tau}(T^{-1/2}\tilde{h}'^*_{T,1,t+\tau})(T^{1/2}\tilde{H}^*_{T,1}(t)) \to^{d^*} \Gamma^*_1$, (e) $\sup_t|T^{1/2}H^*_{T,1}(t))| = O_{p^*}(1)$, (f) $\sup_t|T^{1/2}(\hat{H}^*_{T,1}(t)-H^*_{T,1}(t))| = o_{p^*}(1)$, (g) $MSE^*_1 = \hat{\sigma}^{2*} \to^{p^*} \sigma^2$,

Proof of Lemma 1: (a) The proof is given in Lemma A1 of Clark and McCracken (2005).

(b) Let $\hat{\zeta}$ denote the Lagrange multiplier[19] associated with the ridge regression and define $C_w(T) = J'B_1^{-1}(T)J_w$ and $C_w = \lim_{T\to\infty} E(C_w(T))$.

(b-i) The definition of the ridge estimator implies that for $\frac{1}{1+\hat{\zeta}} = \sqrt{\frac{\hat{d}}{(T^{1/2}\hat{\beta}_{1,T})'J_wF_1^{-1}(T)J'_w(T^{1/2}\hat{\beta}_{1,T})}}$, the ridge estimator takes the form

$$\widetilde{\beta}_{1,T} = \begin{pmatrix} I & \frac{\hat{\zeta}}{1+\hat{\zeta}}B_0(T)C_w(T) \\ 0 & \frac{1}{1+\hat{\zeta}}I \end{pmatrix}\hat{\beta}_{1,T} = \begin{pmatrix} I & \frac{\hat{\zeta}}{1+\hat{\zeta}}B_0(T)C_w(T) \\ 0 & \frac{1}{1+\hat{\zeta}}I \end{pmatrix}(\beta_1 + T^{-1/2}\delta + B_1(T)H_{T,1}(T)).$$

Hence

$$T^{1/2}J'_w\widetilde{\beta}_{1,T} = J'_w\begin{pmatrix} I & \frac{\hat{\zeta}}{1+\hat{\zeta}}B_0(T)C_w(T) \\ 0 & \frac{1}{1+\hat{\zeta}}I \end{pmatrix}[\delta + B_1(T)(T^{1/2}H_{T,1}(T))]$$
$$\to^d J'_w\begin{pmatrix} I & \frac{\zeta}{1+\zeta}B_0C_w \\ 0 & \frac{1}{1+\zeta}I \end{pmatrix}N(\delta, B_1VB_1)$$

where

$\zeta =^d (N(\delta, B_1VB_1))'J_wF_1^{-1}J'_w(N(\delta, B_1VB_1))$ a mixed non-central chi-square variate, and the proof is complete.

[19] This multiplier satisfies $(\frac{1}{1+\hat{\zeta}})^2 = \frac{\hat{L}}{(T^{1/2}\hat{\beta}_{1,T})'J_wF_1^{-1}(T)J'_w(T^{1/2}\hat{\beta}_{1,T})}$ and hence $\hat{\zeta}$ is unique only up to its sign. In all aspects of this paper we use the value satisfying $\frac{1}{1+\hat{\zeta}} = \sqrt{\frac{\hat{L}}{(T^{1/2}\hat{\beta}_{1,T})'J_wF_1^{-1}(T)J'_w(T^{1/2}\hat{\beta}_{1,T})}}$. Choosing the opposite sign is irrelevant since, in every case, what matters is not the value of $\frac{1}{1+\hat{\zeta}}$ but its square.

(b-ii) The ridge estimator takes the form

$$\widetilde{\beta}_{1,T} = \begin{pmatrix} I & \frac{\hat{\zeta}}{1+\hat{\zeta}} B_0(T) C_w(T) \\ 0 & \frac{1}{1+\hat{\zeta}} I \end{pmatrix} \hat{\beta}_{1,T} = \begin{pmatrix} I & \frac{\hat{\zeta}}{1+\hat{\zeta}} B_0(T) C_w(T) \\ 0 & \frac{1}{1+\hat{\zeta}} I \end{pmatrix} (\beta_1 + B_1(T) H_{T,1}(T)).$$

Hence

$$\begin{aligned} T^{1/2} J_w' \widetilde{\beta}_{1,T} &= \sqrt{\tfrac{\hat{d}}{\hat{\beta}_{1,T}' J_w F_1^{-1}(T) J_w' \hat{\beta}_{1,T}}} J_w' [\beta_1 + B_1(T) H_{T,1}(T)] \\ &\to^p \sqrt{\tfrac{d}{\beta_w' F_1^{-1} \beta_w'}} \beta_w \end{aligned}$$

and the proof is complete.

(c) First note that relative to the bootstrap-induced probability measure, $h^*_{T,1,s+\tau}$ is a heteroskedastic vector $MA(\tau-1)$ triangular array with independent, zero mean, Normally distributed increments. Second, note that since the increments η_s are $i.i.d.$ $N(0,1)$, we obtain $E^*(T^{-1/2}\sum_{s=1}^{t-\tau} h^*_{T,1,s+\tau})$ $(T^{-1/2}\sum_{s=1}^{t-\tau} h^*_{T,1,s+\tau})' = (T^{-1/2}\sum_{s=1}^{t-\tau} h_{T,1,s+\tau})\ (T^{-1/2}\sum_{s=1}^{t-\tau} h_{T,1,s+\tau})'$ Finally, since Assumption 2$'$ implies $\lim_{T\to\infty}(T^{-1/2}\sum_{s=1}^{T} h_{s+\tau})\ (T^{-1/2}\sum_{s=1}^{T} h_{s+\tau})' \to^p \Omega_{11} < \infty$, Theorem 3.2 of de Jong and Davidson (2000) implies $T^{-1/2}\sum_{s=1}^{t-\tau} h^*_{T,1,s+\tau} \Rightarrow^* \Omega_{11}^{1/2} W^*(s)$ and the proof is complete.

(d) Given the proof of (c) (notably the delineation of the properties associated with $h^*_{T,1,s+\tau}$), Theorem 4.1 of de Jong and Davidson (2000) implies $\sum_{t=T}^{T+P-\tau}(T^{-1/2}\tilde{h}'^*_{T,1,t+\tau})(T^{1/2}\tilde{H}^*_{T,1}(t)) \to^{d^*} \Gamma_1^*$. Note that the typical drift term, associated with a stochastic integral based on correlated increments, is zero because of the $\tau-$lag between $\tilde{h}^*_{T,1,t+\tau}$ and $\tilde{H}^*_{T,1}(t)$ – that is, $E^*(\tilde{h}^*_{T,1,t+\tau}|\tilde{H}^*_{T,1}(t)) = 0$ for all t. A detailed argument is given in Lemma A1 of Clark and McCracken (2005).

(e) First note that since $T \le t \le T+P-\tau$, $\sup_t |T^{1/2} H^*_{T,1}(t))| \le \sup_t |T^{-1/2}\sum_{s=1}^{t-\tau} h^*_{T,1,s+\tau}|$. The Continuous Mapping Theorem, Lemma 1 (c), and Assumption 3 then imply $\sup_t |T^{-1/2}\sum_{s=1}^{t-\tau} h^*_{T,1,s+\tau}| \to^{d^*}$ $\sup_{1\le s\le 1+\lambda_P} |\Omega_{11}^{1/2} W^*(s)| = O_{p^*}(1)$ and the proof is complete.

(f) For ease of presentation, we show the result for the recursive scheme assuming $\tau = 2$ and hence $\widehat{v}^*_{T,1,s+2} = \eta_{s+2}\widehat{\varepsilon}_{T,1,s+2} + \widehat{\theta}\eta_{s+1}\widehat{\varepsilon}_{T,1,s+1}$ and $v^*_{T,1,s+2} = \eta_{s+2}\varepsilon_{T,1,s+2} + \theta\eta_{s+1}\varepsilon_{T,1,s+1}$. (a) Rearranging terms gives us,

$$\begin{aligned} &T^{1/2}(\hat{H}^*_{T,1}(t) - H^*_{T,1}(t)) = T^{-1/2}\sum\nolimits_{s=1}^{t-\tau}(\widehat{v}^*_{T,1,s+2} - v_{T,1,s+2}) x_{T,1,s} = \\ &T^{-1/2}\sum\nolimits_{s=1}^{t-\tau}(\eta_{s+2}(\widehat{\varepsilon}_{T,1,s+2} - \varepsilon_{T,1,s+2}) + \theta\eta_{s+1}(\widehat{\varepsilon}_{T,1,s+1} - \varepsilon_{T,1,s+1}) + \\ &(\widehat{\theta} - \theta)\eta_{s+1}(\widehat{\varepsilon}_{T,1,s+1} - \varepsilon_{T,1,s+1}) + (\widehat{\theta} - \theta)\eta_{s+1}\varepsilon_{T,1,s+1}) x_{T,1,s}. \end{aligned}$$

If we take a first order Taylor expansion of both $\widehat{\varepsilon}_{T,1,s+2}$ and $\widehat{\varepsilon}_{T,1,s+1}$, then for some $\overline{\gamma}_T$ in the closed cube with opposing vertices $\widehat{\gamma}_T$ and γ_T we obtain

$$\begin{aligned} &T^{1/2}(\hat{H}^*_{T,1}(t) - H^*_{T,1}(t)) = \\ &T^{-1/2}\sum\nolimits_{s=1}^{t-\tau}(\eta_{s+2}\nabla\widehat{\varepsilon}_{T,1,s+2}(\overline{\gamma}_T)(\widehat{\gamma}_T - \gamma_T) + \theta\eta_{s+1}\nabla\widehat{\varepsilon}_{T,1,s+1}(\overline{\gamma}_T)(\widehat{\gamma}_T - \gamma_T) \\ &+(\widehat{\theta} - \theta)\eta_{s+1}\nabla\widehat{\varepsilon}_{T,1,s+1}(\overline{\gamma}_T)(\widehat{\gamma}_T - \gamma_T) + (\widehat{\theta} - \theta)\eta_{s+1}\varepsilon_{T,1,s+1}) x_{T,1,s} \end{aligned}$$

and hence

$$\begin{aligned}
&\sup_t |T^{1/2}(\hat{H}^*_{T,1}(t) - H^*_{T,1}(t))| \leq \\
&2k \sup_t |T^{-1} \sum\nolimits_{s=1}^{t-\tau} \eta_{s+2} \nabla\widehat{\varepsilon}_{T,1,s+2}(\overline{\gamma}_T) x_{T,1,s}||T^{1/2}(\widehat{\gamma}_T - \gamma_T)| \\
&+\theta 2k \sup_t |T^{-1} \sum\nolimits_{s=1}^{t-\tau} \eta_{s+1} \nabla\widehat{\varepsilon}_{T,1,s+1}(\overline{\gamma}_T) x_{T,1,s}||T^{1/2}(\widehat{\gamma}_T - \gamma_T)| \\
&+(\widehat{\theta} - \theta) 2k \sup_t |T^{-1} \sum\nolimits_{s=1}^{t-\tau} \eta_{s+1} \nabla\widehat{\varepsilon}_{T,1,s+1}(\overline{\gamma}_T) x_{T,1,s}||T^{1/2}(\widehat{\gamma}_T - \gamma_T)| \\
&+(T^{1/2}(\widehat{\theta} - \theta)) \sup_t |T^{-1} \sum\nolimits_{s=1}^{t-\tau} \eta_{s+1} \varepsilon_{T,1,s+1} x_{T,1,s}|.
\end{aligned}$$

Assumptions 1 and $2'$ suffice for both $T^{1/2}(\widehat{\gamma}_T - \gamma_T)$ and $T^{1/2}(\widehat{\theta} - \theta)$ to be $O_p(1)$. In addition since, for large enough samples, Assumption $2'$ bounds the second moments of $\nabla\widehat{\varepsilon}_{T,1,s+2}(\overline{\gamma}_T)$ and $\nabla\widehat{\varepsilon}_{T,1,s+1}(\overline{\gamma}_T)$ as well as $x_{T,1,s}$, the fact that the $\eta_{s+\tau}$ are *i.i.d.* $N(0,1)$ then implies $T^{-1}\sum_{s=1}^{t-\tau} \eta_{s+1}\varepsilon_{T,1,s+1}x_{T,1,s}$, $T^{-1}\sum_{s=1}^{t-\tau} \eta_{s+2}\nabla\widehat{\varepsilon}_{T,1,s+2}(\overline{\gamma}_T)x_{T,1,s}$, and $T^{-1}\sum_{s=1}^{t-\tau} \eta_{s+1}\nabla\widehat{\varepsilon}_{T,1,s+1}(\overline{\gamma}_T)x_{T,1,s}$ are all $o_{a.s.^*}(1)$. This in turn, (along with Assumption 3) implies that $sup_t|.|$ of each of these partial sums is $o_{p^*}(1)$ and the proof is complete.

(g) Straightforward algebra shows that

$$\begin{aligned}
\hat{\sigma}^{2*} \quad = \quad & P^{-1}\sum\nolimits_{t=T}^{T+P-\tau} \hat{u}^{*2}_{1,t+\tau} = \{P^{-1}\sum\nolimits_{t=T}^{T+P-\tau} \hat{v}^{*2}_{T,1,t+\tau}\} \\
& -\{2P^{-1}\sum\nolimits_{t=T}^{T+P-\tau} \hat{h}'^*_{T,1,t+\tau}(-JB_0(t)J' + B_1(t))B_1^{-1}(t)J\widetilde{\beta}_{1,T} - 2P^{-1}\sum\nolimits_{t=T}^{T+P-\tau} \hat{h}'^*_{T,1,t+\tau}B_1(t)\hat{H}^*_{T,1}(t) \\
& +\widetilde{\beta}'_{1,T}J[P^{-1}\sum\nolimits_{t=T}^{T+P-\tau} B_1^{-1}(t)(-JB_0(t)J' + B_1(t))x_{T,1,t}x'_{T,1,t}(-JB_0(t)J' + B_1(t))B_1^{-1}(t)]J\widetilde{\beta}_{1,T} \\
& -2\widetilde{\beta}'_{1,T}J[P^{-1}\sum\nolimits_{t=T}^{T+P-\tau} B_1^{-1}(t)(-JB_0(t)J' + B_1(t))x_{T,1,t}x'_{T,1,t}B_1(t)\hat{H}^*_{T,1}(t)] \\
& +P^{-1}\sum\nolimits_{t=T}^{T+P-\tau} \hat{H}^{*\prime}_{T,1}B_1(t)x_{T,1,t}x'_{T,1,t}B_1(t)\hat{H}^*_{T,1}(t)\}.
\end{aligned}$$

We first show that $P^{-1}\sum_{t=T}^{T+P-\tau} \hat{v}^{*2}_{T,1,t+\tau} \to^{p^*} \sigma^2$. If we take a first order Taylor expansion of $\hat{v}^{*2}_{T,1,t+\tau}$ then for some $\bar{\gamma}_T$ in the closed cube with opposing vertices $\hat{\gamma}_T$ and γ we obtain $\hat{v}^{*2}_{T,1,t+\tau} = v^{*2}_{T,1,t+\tau} + 2\hat{v}^*_{T,1,t+\tau}(\bar{\gamma}_T)(\partial\hat{v}^*_{T,1,t+\tau}(\bar{\gamma}_T)/\partial\gamma)(\hat{\gamma}_T - \gamma_T)$. That $P^{-1}\sum_{t=T}^{T+P-\tau} v^{*2}_{T,1,t+\tau} \to^{p^*} \sigma^2$ follows from the fact that $E^*(P^{-1}\sum_{t=T}^{T+P-\tau} v^{*2}_{T,1,t+\tau}) = P^{-1}\sum_{t=T}^{T+P-\tau} v^2_{T,1,t+\tau} \to^p \sigma^2$ and $\lim_{T\to\infty} V^*(P^{-1}\sum_{t=T}^{T+P-\tau} v^{*2}_{T,1,t+\tau}) = 0$. Since Assumptions 1 and $2'$ suffice for both $P^{-1}\sum_{t=T}^{T+P-\tau} \hat{v}^*_{T,1,t+\tau}(\bar{\gamma}_T)(\partial\hat{v}^*_{T,1,t+\tau}(\bar{\gamma}_T)/\partial\gamma) = O_{p^*}(1)$ and $\hat{\gamma}_T - \gamma_T = o_p(1)$, the proof is complete.

We now must show that each element of the second bracketed right-hand side term is $o_{p^*}(1)$. For brevity we only show the result for the first and third terms. For the first bracketed term note that since the η_t's are *i.i.d.* zero mean increments, conditional on the observables $\hat{h}'^*_{T,1,t+\tau}(-JB_0(t)J' + B_1(t))B_1^{-1}(t)J$ is a heteroskedastic $MA(\tau-1)$ process with finite variance and hence $P^{-1}\sum_{t=T}^{T+P-\tau} \hat{h}'^*_{T,1,t+\tau}(-JB_0(t)J' + B_1(t))B_1^{-1}(t)J$ $= o_{p^*}(1)$. Since $\widetilde{\beta}_{1,T} = O_p(1)$ the result is complete. For the third bracketed term, algebra along the lines of that in Clark and McCracken (2005) implies that

$$P^{-1}\sum_{t=T}^{T+P-\tau} B_1^{-1}(t)(-JB_0(t)J'+B_1(t))x_{T,1,t}x'_{T,1,t}(-JB_0(t)J'+B_1(t))B_1^{-1}(t)$$
$$\to {}^pB_1^{-1}[-JB_0J'+B_1]B_1^{-1}$$

Since $\widetilde{\beta}_{1,T}= O_p(1)$ and $B_1^{-1}[-JB_0J'+B_1]B_1^{-1}= 0$ the proof is complete.

Proof of Theorem 2.3: (a) We provide details for the recursive scheme noting differences for the rolling later. Given Lemma 1(g), throughout we will ignore the denominator term $\hat{\sigma}^{2*}$. Straightforward algebra implies that

$$\begin{aligned}
&\sum_{t=T}^{T+P-\tau}(\hat{u}^{*2}_{0,t+\tau}-\hat{u}^{*2}_{1,t+\tau}) = \sum_{t=T}^{T+P-\tau}\{2h'^{*}_{T,1,t+\tau}(-JB_0(t)J'+B_1(t))H^*_{T,1}(t)\\
&-H'^{*}_{T,1}(t)(-JB_0(t)J'x_{T,1,t}x'_{T,1,t}JB_0(t)J'+B_1(t)x_{T,1,t}x'_{T,1,t}B_1(t))H^*_{T,1}(t)\}\\
&+T^{-1/2}\sum_{t=T}^{T+P-\tau}\{2h'^{*}_{T,1,t+\tau}(-JB_0(t)J'+B_1(t))B_1^{-1}(t)(T^{1/2}\widetilde{\beta}_{1,T})\}\\
&+T^{-1}\sum_{t=T}^{T+P-\tau}\{(T^{1/2}\widetilde{\beta}_{1,T})'B_1^{-1}(t)(-JB_0(t)J'+B_1(t))x_{T,1,t}x'_{T,1,t}(-JB_0(t)J'+B_1(t))B_1^{-1}(t)(T^{1/2}\widetilde{\beta}_{1,T})\}\\
\\
&+2\sum_{t=T}^{T+P-\tau}\{h'^{*}_{T,1,t+\tau}(-JB_0(t)J'+B_1(t))(\hat{H}^*_{T,1}(t)-H^*_{T,1}(t))\\
&+(\hat{h}^*_{T,1,t+\tau}-h^*_{T,1,t+\tau})'(-JB_0(t)J'+B_1(t))H^*_{T,1}(t)\\
&-H'^{*}_{T,1}(t)(-JB_0(t)J'x_{T,1,t}x'_{T,1,t}JB_0(t)J'+B_1(t)x_{T,1,t}x'_{T,1,t}B_1(t))(\hat{H}^*_{T,1}(t)-H^*_{T,1}(t))\\
&+(\hat{h}^*_{T,1,t+\tau}-h^*_{T,1,t+\tau})'(-JB_0(t)J'+B_1(t))(\hat{H}^*_{T,1}(t)-H^*_{T,1}(t))\\
&-(0.5)(\hat{H}^*_{T,1}(t)-H^*_{T,1}(t))'(-JB_0(t)J'x_{T,1,t}x'_{T,1,t}JB_0(t)J'+B_1(t)x_{T,1,t}x'_{T,1,t}B_1(t))(\hat{H}^*_{T,1}(t)-H^*_{T,1}(t))\\
&-\widetilde{\beta}'_{1,T}B_1^{-1}(t)(-JB_0(t)J'+B_1(t))x_{T,1,t}x'_{T,1,t}JB_0(t)J'H^*_{T,1}(t)\\
&+\widetilde{\beta}'_{1,T}B_1^{-1}(t)(-JB_0(t)J'+B_1(t))(\hat{h}^*_{T,1,t+\tau}-h^*_{T,1,t+\tau})\\
&-\widetilde{\beta}'_{1,T}B_1^{-1}(t)(-JB_0(t)J'+B_1(t))x_{T,1,t}x'_{T,1,t}JB_0(t)J'(\hat{H}^*_{T,1}(t)-H^*_{T,1}(t))\}
\end{aligned} \tag{25}$$

Note that there are 4 bracketed $\{.\}$ terms in (17). The first three are directly analogous to the first three bracketed terms in (14) from the proof of Theorem 2.1. We will show that these three terms provide the limits $\Gamma_i^* =^{d^*} \Gamma_i$, for Γ_i $i = 1-4$ defined in the text. The additional assumption of either conditional homoskedasticity or $\dim(\beta_w) = 1$ are needed only in the proof for $\Gamma_3^* =^{d^*} \Gamma_3$. Finally, we then show that the fourth bracketed term in (17) is $o_{p^*}(1)$.

Proof of bracket 1 in (17): Consider the first part of the bracket. Rearranging terms gives us

$$\begin{aligned}
&\sum_{t=T}^{T+P-\tau}2h'^{*}_{T,1,t+\tau}(-JB_0(t)J'+B_1(t))H^*_{T,1}(t) = 2\sum_{t=T}^{T+P-\tau}h'^{*}_{T,1,t+\tau}(-JB_0J'+B_1)H^*_{T,1}(t)\\
&+\sum_{t=T}^{T+P-\tau}h'^{*}_{t+\tau}\{(-JB_0(t)J'+B_1(t))-(-JB_0J'+B_1)\}H^*_{T,1}(t)\\
&=2\sigma^2\sum_{t=T}^{T+P-\tau}(T^{-1/2}\tilde{h}'^{*}_{T,1,t+\tau})(T^{1/2}\tilde{H}^*_{T,1}(t)+\\
&T^{-1}\sum_{t=T}^{T+P-\tau}h'^{*}_{T,1,t+\tau}\{T^{1/2}((-JB_0(t)J'+B_1(t))-(-JB_0J'+B_1))\}(T^{1/2}H^*_{T,1}(t)),
\end{aligned}$$

where $\tilde{h}^*_{T,1,t+\tau}$ and $\tilde{H}^*_{T,1}(t)$ are the bootstrap equivalents of $\tilde{h}_{T,1,t+\tau}$ and $\tilde{H}_{T,1}(t)$ defined in section 2.1. That $\sigma^2\sum_{t=T}^{T+P-\tau}(T^{-1/2}\tilde{h}'^{*}_{T,1,t+\tau})(T^{1/2}\tilde{H}^*_{T,1}(t)) \to^{d^*} \sigma^2\Gamma_1^*$ follows from Lemma 1 (d). To show that the remaining term is $o_{p^*}(1)$ note that Lemmas 1 (e) and (a) imply $\sup_t|T^{1/2}H^*_{T,1}(t)| = O_{p^*}(1)$ and $\sup_t|T^{1/2}((-JB_0(t)J'+B_1(t))-(-JB_0J'+B_1))| = O_p(1)$. Since $E^*(h^*_{T,1,t+\tau}|H^*_{T,1}(t),B_0(t),B_1(t)) = 0$ for all t and $Var^*(T^{-1}\sum_{t=T}^{T+P-\tau}h'^{*}_{T,1,t+\tau}\{T^{1/2}((-JB_0(t)J'+B_1(t))-(-JB_0J'+B_1))\}(T^{1/2}H^*(t))) = o_{p^*}(1)$, the proof is complete. The same qualifier applies to similar results used in the bootstrap proofs.)

Now consider the second part of the bracket. Rearranging terms gives us

$$\begin{aligned}
&\textstyle\sum_{t=T}^{T+P-\tau} H_{T,1}^{*'}(t)(-JB_0(t)J'x_{T,1,t}x'_{T,1,t}JB_0(t)J' + B_1(t)x_{T,1,t}x'_{T,1,t}B_1(t))H_{T,1}^*(t) \\
&= \textstyle\sum_{t=T}^{T+P-\tau} H_{T,1}^{*'}(t)(-JB_0J' + B_1)H_{T,1}^*(t) \\
&+ \textstyle\sum_{t=T}^{T+P-\tau} H_{T,1}^{*'}(t)\{(-JB_0(t)J'x_{T,1,t}x'_{T,1,t}JB_0(t)J' + B_1(t)x_{T,1,t}x'_{T,1,t}B_1(t)) - (-JB_0J' + B_1)\}H_{T,1}^*(t) \\
&= \sigma^2 \textstyle\sum_{t=T}^{T+P-\tau} \tilde{H}_{T,1}^{*\prime}(t)\tilde{H}_{T,1}^*(t) \\
&+ \textstyle\sum_{t=T}^{T+P-\tau} H_{T,1}^{*'}(t)\{(-JB_0(t)J'x_{T,1,t}x'_{T,1,t}JB_0(t)J' + B_1(t)x_{T,1,t}x'_{T,1,t}B_1(t)) - (-JB_0J' + B_1)\}H_{T,1}^*(t).
\end{aligned}$$

That $\sigma^2 \sum_{t=T}^{T+P-\tau} \tilde{H}_{T,1}^{*\prime}(t)\tilde{H}_{T,1}^*(t) \to^{d^*} \sigma^2\Gamma_2^*$ follows from the Continuous Mapping Theorem and Lemma 1 (c). To show that the remaining term is $o_{p^*}(1)$ note that by adding and subtracting terms we obtain

$$\begin{aligned}
&\textstyle\sum_{t=T}^{T+P-\tau} H_{T,1}^{*'}(t)\{(-JB_0(t)J'x_{T,1,t}x'_{T,1,t}JB_0(t)J' + B_1(t)x_{T,1,t}x'_{T,1,t}B_1(t)) - (-JB_0J' + B_1)\}H_{T,1}^*(t) \\
&= \textstyle\sum_{(\text{all } m,n,o)=1,2} \sum_{t=T}^{T+P-\tau} H_{T,1}^{*'}(t)\{(-Ja_{m,t}J'b_{n,t}Ja_{o,t}J' + c_{m,t}b_{n,t}c_{o,t}) - (-JB_0J' + B_1)\}H_{T,1}^*(t),
\end{aligned}$$

where $a_{1,t} = B_0$, $a_{2,t} = B_0(t) - B_0$, $b_{1,t} = B_1^{-1}$, $b_{2,t} = x_{T,1,t}x'_{T,1,t} - B_1^{-1}$, $c_{1,t} = B_1$, and $c_{2,t} = B_1(t) - B_1$. If the indices m, n, o are all 1 then the remainder term is numerically zero and hence it suffices to show that for all other permutations of the indices that the elements of the remainder term are $o_{p^*}(1)$. The proofs of each are very similar and hence we show the result for the case when the indices are all equal to 2. To do so note that

$$\begin{aligned}
&|\sum_{t=T}^{T+P-\tau} H_{T,1}^{*'}(t)(-Ja_{2,t}J'b_{2,t}Ja_{2,t}J' + c_{2,t}b_{2,t}c_{2,t})H_{T,1}^*(t)| \\
\leq\ & 2k^4T^{-1}(\sup_t |T^{1/2}H_{T,1}^*(t)|)^2(\max_{i=0,1}\sup_t |T^{1/2}(B_i(t) - B_i)|)^2(T^{-1}\sum_{t=T}^{T+P-\tau} |x_{T,1,t}x'_{T,1,t} - B_1^{-1}|)
\end{aligned}$$

Lemma 1 (e) implies $\sup_t |T^{1/2}H_{T,1}^*(t)|=O_{p^*}(1)$ while Lemma 1 (a) implies $\max_{i=0,1}\sup_t |T^{1/2}(B_i(t)-B_i)=O_p(1)$. Since Assumption 2′ is sufficient for $T^{-1}\sum_{t=T}^{T+P-\tau} |x_{T,1,t}x'_{T,1,t} - B_1^{-1}|=O_p(1)$ the result follows from the fact that T^{-1} is $o(1)$.

Proof of bracket 2 in (17): Rearranging terms gives us

$$\begin{aligned}
&T^{-1/2}2\textstyle\sum_{t=T}^{T+P-\tau} h_{T,1,t+\tau}^{\prime *}(-JB_0(t)J' + B_1(t))B_1^{-1}(t)(T^{1/2}\widetilde{\beta}_{1,T}) \\
&= T^{-1/2}2\textstyle\sum_{t=T}^{T+P-\tau} h_{T,1,t+\tau}^{\prime *}B_1(t)J_wF_1^{-1}(t)(T^{1/2}J'_w\widetilde{\beta}_{1,T})
\end{aligned}.$$

From Lemma 1(b) we know $T^{1/2}J'_w\widetilde{\beta}_{1,T} = O_p(1)$. Algebra along the lines of Clark and McCracken (2005) then gives us

$$T^{-1/2}2\sum_{t=T}^{T+P-\tau} h_{T,1,t+\tau}^{\prime *}B_1(t)J_wF_1^{-1}(t)(T^{1/2}J'_w\widetilde{\beta}_{1,T}) = T^{-1/2}2\sum_{t=T}^{T+P-\tau} h_{T,1,t+\tau}^{\prime *}B_1J_wF_1^{-1}(T^{1/2}J'_w\widetilde{\beta}_{1,T})+o_{p^*}(1).$$

This term is a bit different from that for the second bracketed term in Theorem 2.1. There, the second bracketed term takes the form $T^{-1/2}2\sum_{t=T}^{T+P-\tau} h'_{T,1,t+\tau}B_1J_wF_1^{-1}\beta_w+o_p(1)$. What makes them different here is that since $T^{1/2}J'_w\widetilde{\beta}_{1,T}$ is not a consistent estimate of β_w, it is not the case that $T^{-1/2}2\sum_{t=T}^{T+P-\tau} h_{T,1,t+\tau}^{\prime *}B_1J_wF_1^{-1}(T^{1/2}J'_w\widetilde{\beta}_{1,T})$ equals $T^{-1/2}2\sum_{t=T}^{T+P-\tau} h_{T,1,t+\tau}^{\prime *}B_1J_wF_1^{-1}\beta_w+o_p^*(1)$.

However, it is true that both terms are asymptotically normal. For the former, clearly

$$T^{-1/2}2\sum_{t=T}^{T+P-\tau} h'_{T,1,t+\tau}B_1J_wF_1^{-1}\beta_w \to^d \Gamma_2 \sim N(0, 4\Omega)$$

where $\Omega = \lambda_P\beta_w^{'}F_1^{-1}J'_wB_1VB_1J_wF_1^{-1}\beta_w$. But for the latter, due to the *i.i.d.* $N(0,1)$ (and strictly exogenous) nature of the $\eta_{t+\tau}$, we have

$$T^{-1/2}2\sum_{t=T}^{T+P-\tau} h_{T,1,t+\tau}^{\prime *}B_1J_wF_1^{-1}J'_w(T^{1/2}\widetilde{\beta}_{1,T}) \to^{d^*} \Gamma_3^* \sim N(0, 4W)$$

where

$$\begin{aligned} W &= \lim Var\{T^{-1/2}\textstyle\sum_{t=T}^{T+P-\tau} h'^{*}_{T,1,t+\tau} B_1 J_w F_1^{-1} J'_w (T^{1/2}\widetilde{\beta}_{1,T})\} \\ &= \lambda_P \lim E\{(T^{1/2}\widetilde{\beta}_{1,T})' J_w F_1^{-1} J'_w B_1 \{\lim Var(P^{-1/2}\textstyle\sum_{t=T}^{T+P-\tau} h'^{*}_{T,1,t+\tau})\} B_1 J_w F_1^{-1} J'_w (T^{1/2}\widetilde{\beta}_{1,T})\} \\ &= \lambda_P \lim E\{(T^{1/2}\widetilde{\beta}_{1,T})' J_w F_1^{-1} J'_w B_1 V B_1 J_w F_1^{-1} J'_w (T^{1/2}\widetilde{\beta}_{1,T})\} \end{aligned}.$$

The precise relationship between Γ_3^* and Γ_3 depends on the relationship between Ω and W. This in turn depends upon the additional restrictions in the statement of the Theorem.

(a) If we let $V = \sigma^2 B_1^{-1}$, W simplifies to

$$\begin{aligned} W &= \sigma^2 \lambda_P \lim E\{(T^{1/2}\widetilde{\beta}_{1,T})' J_w F_1^{-1} J'_w (T^{1/2}\widetilde{\beta}_{1,T})\} \\ &= \sigma^2 \lambda_P \lim E\{(T^{1/2}\widetilde{\beta}_{1,T})' J_w F_1^{-1}(T) J'_w (T^{1/2}\widetilde{\beta}_{1,T})\} \\ &= \sigma^2 \lambda_P \lim E\{\hat{d}\} = \sigma^2 \lambda_P d \end{aligned}.$$

The result follows since under the null hypothesis, $\Omega = \lambda_P \beta_w' F_1^{-1} J'_w B_1 V B_1 J_w F_1^{-1} \beta_w = \sigma^2 \lambda_P \beta_w' F_1^{-1} \beta_w = \sigma^2 \lambda_P d$.

(b) If we let $\dim(\beta_w) = 1, W$ simplifies to

$$\begin{aligned} W &= \lambda_P \lim E\{(T^{1/2}\widetilde{\beta}_{1,T})' J_w F_1^{-1} J'_w B_1 V B_1 J_w F_1^{-1} J'_w (T^{1/2}\widetilde{\beta}_{1,T})\} \\ &= \lambda_P \lim E\{(T^{1/2}\tilde{\beta}_{w,T})^2 (F_1^{-1})^2 J'_w B_1 V B_1 J_w\} \end{aligned}.$$

But $\tilde{\beta}_{w,T}$ was estimated satisfying the restriction that $(T^{1/2}\tilde{\beta}_{w,T})^2 = F_1(T)\hat{d}$ and hence $W = \lambda_P \lim E\{F_1(T)\hat{d}(F_1^{-1})^2 J'_w B_1 V B_1 J_w\} = \lambda_P F_1^{-1} d J'_w B_1 V B_1 J_w$. Following similar arguments, we also have $\Omega = \lambda_P (\beta_w)^2 (F_1^{-1})^2 J'_w B_1 V B_1 J_w$. But under the null, $(\beta_w)^2 = dF_1$ and the proof is complete.

Proof of bracket 3 in (17): Rearranging terms gives us

$$\begin{aligned} & T^{-1}\textstyle\sum_{t=T}^{T+P-\tau} (T^{1/2}\widetilde{\beta}_{1,T})' B_1^{-1}(t)(-JB_0(t)J' + B_1(t)) x_{T,1,t} x'_{T,1,t} (-JB_0(t)J' + B_1(t)) B_1^{-1}(t)(T^{1/2}\widetilde{\beta}_{1,T}) \\ &= T^{-1}\textstyle\sum_{t=T}^{T+P-\tau} (T^{1/2}\widetilde{\beta}_{1,T})' J_w F_1^{-1}(t) J'_w B_1(t) x_{T,1,t} x'_{T,1,t} B_1(t) J_w F_1^{-1}(t) J'_w (T^{1/2}\widetilde{\beta}_{1,T}) \end{aligned}.$$

From Lemma 1(b) we know $T^{1/2} J'_w \widetilde{\beta}_{1,T} = O_p(1)$. From there, algebra long the lines of Clark and McCracken (2005) gives us

$$\begin{aligned} & T^{-1}\textstyle\sum_{t=T}^{T+P-\tau} (T^{1/2}\widetilde{\beta}_{1,T})' J_w F_1^{-1}(t) J'_w B_1(t) x_{T,1,t} x'_{T,1,t} B_1(t) J_w F_1^{-1}(t) J'_w (T^{1/2}\widetilde{\beta}_{1,T}) \\ &= T^{-1}\textstyle\sum_{t=T}^{T+P-\tau} (T^{1/2}\widetilde{\beta}_{1,T})' J_w F_1^{-1}(t) J'_w B_1(t) B_1^{-1} B_1(t) J_w F_1^{-1}(t) J'_w (T^{1/2}\widetilde{\beta}_{1,T}) + o_p(1) \\ &= T^{-1}\textstyle\sum_{t=T}^{T+P-\tau} (T^{1/2}\widetilde{\beta}_{1,T})' J_w F_1^{-1} J'_w (T^{1/2}\widetilde{\beta}_{1,T}) + o_p(1) \\ &= (P - \tau + 1/T)\hat{d} + o_p(1) \to^p \lambda_P d \equiv \Gamma_4^* \end{aligned}.$$

The result follows since under the null hypothesis, $\Gamma_4 \equiv \beta_w' F_1^{-1} \beta_w = \lambda_P d$.

Proof of bracket 4 in (17): We must show that each of the eight components of the fourth bracketed term in (17) are $o_{p^*}(1)$. The proofs of each are similar and as such we show the results only for the fourth and seventh components. If we take absolute value of the former we find that

$$\begin{aligned} & |\textstyle\sum_{t=T}^{T+P-\tau} (\hat{h}^*_{T,1,t+\tau} - h^*_{T,1,t+\tau})'(-JB_0(t)J' + B_1(t))(\hat{H}^*_{T,1}(t) - H^*_{T,1}(t))| \\ &\leq k^2 (T^{-1/2}\textstyle\sum_{t=T}^{T+P-\tau} |\hat{h}^*_{T,1,t+\tau} - h^*_{T,1,t+\tau}|)(\sup_t |-JB_0(t)J' + B_1(t)|)(\sup_t T^{1/2}|\hat{H}^*_{T,1}(t) - H^*_{T,1}(t)|) \end{aligned}$$

while for the latter straightforward algebra along the lines of Clark and McCracken (2005) gives us

$$\begin{aligned} & \textstyle\sum_{t=T}^{T+P-\tau} \widetilde{\beta}'_{1,T} B_1^{-1}(t)(-JB_0(t)J' + B_1(t))(\hat{h}^*_{T,1,t+\tau} - h^*_{T,1,t+\tau}) \\ &= (T^{1/2} J'_w \widetilde{\beta}_{1,T})' F_1^{-1} J'_w B_1 (T^{-1/2}\textstyle\sum_{t=T}^{T+P-\tau} (\hat{h}^*_{T,1,t+\tau} - h^*_{T,1,t+\tau})) + o_{p^*}(1). \end{aligned}$$

Lemma 1(f) and (b) imply both $\sup_t T^{1/2}|\hat{H}^*_{T,1}(t) - H^*_{T,1}(t)| = o_{p^*}(1)$ and $T^{1/2}J_w'\widetilde{\beta}_{1,T} = O_p(1)$ while Assumption 2′ suffices for $\sup_t |-JB_0(t)J' + B_1(t)| = O_p(1)$. That $T^{-1/2}\sum_{t=T}^{T+P-\tau}(\hat{h}^*_{T,1,t+\tau} - h^*_{T,1,t+\tau}) = o_{p^*}(1)$ follows an almost identical line of proof to that in Lemma 1(f) (without the $\sup_t |.|$ component) but with a different range of summation.

The result will follow if $T^{-1/2}\sum_{t=T}^{T+P-\tau} |\hat{h}^*_{T,1,t+\tau} - h^*_{T,1,t+\tau}| = o_{p^*}(1)$. For simplicity we assume, as in the proof of Lemma 1(f), that $\tau = 2$ and hence the forecast errors form an $MA(1)$. If we then take a Taylor expansion in precisely the same fashion as in the proof of Lemma 1(f) we have

$$
\begin{aligned}
&T^{-1/2}\sum\nolimits_{t=T}^{T+P-\tau} |\hat{h}^*_{T,1,t+\tau} - h^*_{T,1,t+\tau}| \leq \\
&2k_1 T^{-1}\sum\nolimits_{t=T}^{T+P-\tau} |\eta_{t+2}\nabla\widehat{\varepsilon}_{T,1,t+2}(\overline{\gamma}_T)x_{T,1,t}||T^{1/2}(\widehat{\gamma}_T - \gamma_T)| \\
&+\theta 2k_1 T^{-1}\sum\nolimits_{t=T}^{T+P-\tau} |\eta_{s+1}\nabla\widehat{\varepsilon}_{T,1,t+1}(\overline{\gamma}_T)x_{T,1,t}||T^{1/2}(\widehat{\gamma}_T - \gamma_T)| \\
&+(\widehat{\theta} - \theta)2k_1 T^{-1}\sum\nolimits_{t=T}^{T+P-\tau} |\eta_{t+1}\nabla\widehat{\varepsilon}_{T,1,t+1}(\overline{\gamma}_T)x_{T,1,t}||T^{1/2}(\widehat{\gamma}_T - \gamma_T)| \\
&+(T^{1/2}(\widehat{\theta} - \theta))T^{-1}\sum\nolimits_{t=T}^{T+P-\tau} |\eta_{t+1}\varepsilon_{T,1,t+1}x_{T,1,t}|.
\end{aligned}
$$

Assumptions 1 and 2′ suffice for both $T^{1/2}(\widehat{\gamma}_T - \gamma_T)$ and $T^{1/2}(\widehat{\theta} - \theta)$ to be $O_p(1)$. Since, for large enough samples, Assumption 2′ bounds the second moments of $\nabla\widehat{\varepsilon}_{T,1,s+2}(\overline{\gamma}_T)$ and $\nabla\widehat{\varepsilon}_{T,1,s+1}(\overline{\gamma}_T)$ as well as $\mathrm{x}_{T,1,s}$; with $\eta_{s+\tau}$ distributed $i.i.d.N(0,1)$, $T^{-1}\sum_{s=1}^{T-\tau} |\eta_{s+2}\nabla\widehat{\varepsilon}_{T,1,s+2}(\overline{\gamma}_T)x_{T,1,s}|$, $T^{-1}\sum_{s=1}^{T-\tau} |\eta_{s+1}\varepsilon_{T,1,s+1}x_{T,1,s}|$, and $T^{-1}\sum_{s=1}^{T-\tau} |\eta_{s+1}\nabla\widehat{\varepsilon}_{T,1,s+1}(\overline{\gamma}_T)x_{T,1,s}|$ and are all $O_{p^*}(1)$, and the proof is complete.

Proof for the rolling scheme: Results for the rolling scheme differ only in the definition of $H^*_{T,1}(t) = T^{-1}\sum_{s=t-T+1}^{t} h^*_{T,1,s+\tau}$ (and to a lesser extent $\hat{H}^*_{T,1}(t) = T^{-1}\sum_{s=t-T+1}^{t} \hat{h}^*_{T,1,s+\tau}$). In particular, if we substitute $T^{1/2}H^*_{T,1}(t) \Rightarrow^* V^{1/2}(W^*(s) - W^*(s-1))$ for $T^{1/2}H^*_{T,1}(t) \Rightarrow^* V^{1/2}s^{-1}W^*(s)$ as used above and in the proof of Theorem 2.1, we obtain the desired conclusion.

Proof of Theorem 2.3 (b): Given Theorem 2.3(a) and the Continuous Mapping Theorem it suffices to show that $P\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)\hat{\gamma}^*_{LL}(j) \to^{d^*} 4\sigma_u^4(\Gamma_5^* + \Gamma_6^* + \Gamma_7^*)$ where $\Gamma_i^* =^{d^*} \Gamma_i$ for Γ_i $i = 5-7$ defined in the text. Before doing so it is convenient to redefine the four bracketed terms terms from (17) used in the main decomposition of the loss differential in Theorem 2.3(a) (absent the summations, but keeping the brackets) as

$$(\hat{u}^{*2}_{0,t+\tau} - \hat{u}^{*2}_{1,t+\tau}) = \{2A^*_{1,t} - A^*_{2,t}\} + 2\{B^*_{1,t}\} + \{C^*_t\} + \{D^*_t\}.$$

With this in mind we obtain

$$\begin{aligned}
& P\sum\nolimits_{j=-\bar{j}}^{\bar{j}} K(j/M)\hat{\gamma}^*_{LL}(j) \qquad (26)\\
= & \sum\nolimits_{j=-\bar{j}}^{\bar{j}} K(j/M)\sum\nolimits_{t=T+j}^{T+P-\tau} (\hat{u}^{*2}_{0,t+\tau}-\hat{u}^{*2}_{1,t+\tau})(\hat{u}^{*2}_{0,t-j+\tau}-\hat{u}^{*2}_{1,t-j+\tau})\\
= & 4\{\sum\nolimits_{j=-\bar{j}}^{\bar{j}} K(j/M)\sum\nolimits_{t=T+j}^{T+P-\tau} A^*_{1,t}A^*_{1,t-j}\}+4\{\sum\nolimits_{j=-\bar{j}}^{\bar{j}} K(j/M)\sum\nolimits_{t=T+j}^{T+P-\tau} A^*_{1,t}B^*_{1,t-j}\}\\
& +4\{\sum\nolimits_{j=-\bar{j}}^{\bar{j}} K(j/M)\sum\nolimits_{t=T+j}^{T+P-\tau} B^*_{1,t}B^*_{1,t-j}\}\\
& +\text{ other cross products of } A^*_{1,t},A^*_{2,t},B^*_{1,t},C^*_t,D^*_t \text{ with } A^*_{1,t-j},A^*_{2,t-j},B^*_{1,t-j},C^*_{t-j},D^*_{t-j}
\end{aligned}$$

In the remainder we show that each of the three bracketed terms converges to σ^4 times $\Gamma^*_i =^{d^*} \Gamma_i$ $i=5-7$ respectively and that each of the cross product terms are each $o_{p^*}(1)$.

Proof of bracket 1 in (18): Straightforward algebra implies that

$$\begin{aligned}
&\sum\nolimits_{j=-\bar{j}}^{\bar{j}} K(j/M)\sum\nolimits_{t=T+j}^{T+P-\tau} A^*_{1,t}A^*_{1,t-j} =\\
&\sigma^4\sum\nolimits_{j=-\bar{j}}^{\bar{j}} K(j/M)\sum\nolimits_{t=T+j}^{T+P-\tau}\{(T^{1/2}\tilde{H}^{*'}_{T,1}(t))E^*\tilde{h}^*_{T,1,t+\tau}\tilde{h}^{*\prime}_{T,1,t-j+\tau}(T^{1/2}\tilde{H}^*_{T,1}(t-j))\}\\
&+\sigma^4\sum\nolimits_{j=-\bar{j}}^{\bar{j}} K(j/M)\sum\nolimits_{t=T+j}^{T+P-\tau}\{H^{*'}_{T,1}(t)((-JB_0(t)J'+B_1(t))-(-JB_0J'+B_1))\times\\
&E^*h^*_{T,1,t+\tau}h^{*\prime}_{T,1,t-j+\tau}(-JB_0J'+B_1)H^*_{T,1}(t-j)\\
&+H^{*'}_{T,1}(t)((-JB_0(t)J'+B_1(t))-(-JB_0J'+B_1))(h^*_{T,1,t+\tau}h^{*\prime}_{T,1,t-j+\tau}-E^*h^*_{T,1,t+\tau}h^{*\prime}_{T,1,t-j+\tau})\times\\
&(-JB_0J'+B_1)H^*_{T,1}(t-j)\\
&+H^{*'}_{T,1}(t)((-JB_0(t)J'+B_1(t))-(-JB_0J'+B_1))(h^*_{T,1,t+\tau}h^{*\prime}_{T,1,t-j+\tau}-E^*h^*_{T,1,t+\tau}h^{*\prime}_{T,1,t-j+\tau})\times\\
&((-JB_0(t-j)J'+B_1(t-j))-(-JB_0J'+B_1))H^*_{T,1}(t-j)\\
&+H^{*'}_{T,1}(t)(-JB_0J'+B_1)(h^*_{T,1,t+\tau}h^{*\prime}_{T,1,t-j+\tau}-E^*h^*_{T,1,t+\tau}h^{*\prime}_{T,1,t-j+\tau})(-JB_0J'+B_1)H^*_{T,1}(t-j)\\
&+H^{*'}_{T,1}(t)(-JB_0J'+B_1)(h^*_{T,1,t+\tau}h^{*\prime}_{T,1,t-j+\tau}-E^*h^*_{T,1,t+\tau}h^{*\prime}_{T,1,t-j+\tau})\times\\
&((-JB_0(t-j)J'+B_1(t-j))-(-JB_0J'+B_1))H^*_{T,1}(t-j)\\
&+H^{*'}_{T,1}(t)(-JB_0J'+B_1)E^*h^*_{T,1,t+\tau}h^{*\prime}_{T,1,t-j+\tau}((-JB_0(t-j)J'+B_1(t-j))-(-JB_0J'+B_1))H^*_{T,1}(t-j)\\
&+H^{*'}_{T,1}(t)((-JB_0(t-j)J'+B_1(t-j))-(-JB_0J'+B_1))E^*h^*_{T,1,t+\tau}h^{*\prime}_{T,1,t-j+\tau}\times\\
&((-JB_0(t-j)J'+B_1(t-j))-(-JB_0J'+B_1))H^*_{T,1}(t-j)\},
\end{aligned} \qquad (27)$$

where $\tilde{H}^*_{T,1}(t)$ is the bootstrap equivalent of $\tilde{H}_{T,1}(t)$ defined in section 2.1. Since j is finite, the Continuous Mapping Theorem and Lemma 1(c) imply

$$\sigma^4\sum\nolimits_{j=-\bar{j}}^{\bar{j}} K(j/M)\sum\nolimits_{t=T+j}^{T+P-\tau}(T^{1/2}\tilde{H}^{*'}_{T,1}(t))(E^*\tilde{h}^*_{T,1,t+\tau}\tilde{h}^{*\prime}_{T,1,t-j+\tau})(T^{1/2}\tilde{H}^*_{T,1}(t-j))\to^{d^*}\sigma^4\Gamma^*_3.$$

We must now show that each element of the second bracketed right-hand side term in (19) is $o_{p^*}(1)$. The proof of each is similar and as such we provide the result for the first and fourth elements.

For the first, after taking the absolute value we obtain

$$\begin{aligned}
&|\sum_{j=-\bar{j}}^{\bar{j}} K(j/M) \sum_{t=T+j}^{T+P-\tau} H_{T,1}^{*'}(t)((-JB_0(t)J'+B_1(t)) - (-JB_0J'+B_1))(E^* h_{T,1,t+\tau}^{*} h_{T,1,t-j+\tau}^{*\prime})\times \\
&(-JB_0J'+B_1)H_{T,1}^{*}(t-j)| \\
\leq\ & 2T^{-1/2}\bar{j}k^4(\sup_t T^{1/2}|H_{T,1}^{*}(t)|)^2(|-JB_0J'+B_1|)(\sup_t T^{1/2}|(-JB_0(t)J'+B_1(t)) - (-JB_0J'+B_1)|)\times \\
&(\max_{|j|\leq\bar{j}}(T^{-1}\sum_{t=T+j}^{T+P-\tau}|E^* h_{T,1,t+\tau}^{*} h_{T,1,t-j+\tau}^{*\prime}|)).
\end{aligned}$$

Lemmas 1(e) and (a) imply both $\sup_t T^{1/2}|H_{T,1}^{*}(t)| = O_{p^*}(1)$ and $\sup_t T^{1/2}|(-JB_0(t)J'+B_1(t))-(-JB_0J'+B_1)| = O_p(1)$. Since for each $|j| \leq \bar{j}$, Assumption 2$'$ is sufficient for $T^{-1}\sum_{t=T+j}^{T+P-\tau}|E^* h_{T,1,t+\tau}^{*} h_{T,1,t-j+\tau}^{*\prime}| = T^{-1}\sum_{t=T+j}^{T+P-\tau}|h_{T,1,t+\tau} h_{T,1,t-j+\tau}^{'}|$ to be $O_p(1)$ the result follows from the fact that $T^{-1/2} = o(1)$.

For the fourth term note that after rearranging terms we obtain

$$\begin{aligned}
&\sum_{j=-\bar{j}}^{\bar{j}} K(j/M) \sum_{t=T+j}^{T+P-\tau} H_{T,1}^{*'}(t)(-JB_0J'+B_1)(h_{T,1,t+\tau}^{*} h_{T,1,t-j+\tau}^{\prime *} - E^* h_{T,1,t+\tau}^{*} h_{T,1,t-j+\tau}^{\prime *})\times \\
&(-JB_0J'+B_1)H_{T,1}^{*}(t-j) \\
=\ & T^{-1/2}\sum_{j=-\bar{j}}^{\bar{j}} K(j/M) \sum_{t=T+j}^{T+P-\tau}(T^{1/2}H_{T,1}^{*'}(t-j)(-JB_0J'+B_1)\otimes T^{1/2}H_{T,1}^{*'}(t)(-JB_0J'+B_1))\times \\
&(T^{-1/2}vec(h_{T,1,t+\tau}^{*} h_{T,1,t-j+\tau}^{*\prime} - E^* h_{T,1,t+\tau}^{*} h_{T,1,t-j+\tau}^{*\prime})).
\end{aligned}$$

Recall that by Lemma 1 (c), $T^{1/2}H_{T,1}^{*}(t) \Rightarrow^* \Omega_{11}^{1/2}W^*(s)$. Moreover, note that conditional on the observables, $h_{T,1,t+\tau}^{*} h_{T,1,t-j+\tau}^{*\prime} - E^* h_{T,1,t+\tau}^{*} h_{T,1,t-j+\tau}^{*\prime}$ forms a heteroskedastic L^2-bounded $MA(\tau-1)$ process with increments that are uncorrelated with $H_{T,1}^{*}(t)$ and $H_{T,1}^{*}(t-l)$. Hence conditional on the observables and for each $|j| \leq \bar{j}$, Theorem 4.1 of de Jong and Davidson (2000) suffices for

$$\begin{aligned}
&\sum_{t=T+j}^{T+P-\tau}(T^{1/2}H_{T,1}^{*'}(t-j)(-JB_0J'+B_1)\otimes T^{1/2}H_{T,1}^{*'}(t)(-JB_0J'+B_1))\times \\
&(T^{-1/2}vec(h_{T,1,t+\tau}^{*} h_{T,1,t-j+\tau}^{*\prime} - E^* h_{T,1,t+\tau}^{*} h_{T,1,t-j+\tau}^{*\prime}) = O_{p^*}(1).
\end{aligned}$$

The result follows since $T^{-1/2} = o(1)$ and $K(x) < 1$ for all x.

Proof of bracket 2 in (18): After rearranging terms, the second bracketed term is

$$\begin{aligned}
&\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)\sum_{t=T+j}^{T+P-\tau} A_{1,t}^{*}B_{1,t-j}^{*} \\
&= \sum_{j=-\bar{j}}^{\bar{j}} K(j/M)T^{-1}\sum_{t=T+j}^{T+P-\tau}(T^{1/2}H_{T,1}^{*}(t))'(-JB_0(t)J' + B_1(t))\times \\
&h_{T,1,t+\tau}^{*} h_{T,1,t-j+\tau}^{\prime *}(-JB_0(t-j)J' + B_1(t-j))B_1^{-1}(t-j)(T^{1/2}\widetilde{\beta}_{1,T}) \\
&= \sum_{j=-\bar{j}}^{\bar{j}} K(j/M)T^{-1}\sum_{t=T+j}^{T+P-\tau}(T^{1/2}H_{T,1}^{*}(t))'(-JB_0(t)J' + B_1(t))\times \\
&h_{T,1,t+\tau}^{*} h_{T,1,t-j+\tau}^{\prime *}B_1(t-j)J_wF_1^{-1}(t-j)J_w'(T^{1/2}\widetilde{\beta}_{1,T})
\end{aligned}$$

This term is a bit different from that for the second bracketed term in Theorem 2.2. As in the proof of Theorem 2.3(a), it differs because $J_w'(T^{1/2}\widetilde{\beta}_{1,T})$ is not a consistent estimate of β_w. However, it is true that both terms are asymptotically normal. To see this note that similar arguments to those used immediately above in the proof for bracket 1 imply

$$\begin{aligned}
&\textstyle\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)T^{-1}\sum_{t=T+j}^{T+P-\tau}(T^{1/2}H^*_{T,1}(t))'(-JB_0(t)J'+B_1(t))\times \\
&h^*_{T,1,t+\tau}h'^*_{T,1,t-j+\tau}B_1(t-j)J_wF_1^{-1}(t-j)J'_w(T^{1/2}\widetilde{\beta}_{1,T}) \\
&=\textstyle\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)T^{-1}\sum_{t=T+j}^{T+P-\tau}(T^{1/2}H^*_{T,1}(t))'(-JB_0J'+B_1)\times \\
&(E^*h^*_{T,1,t+\tau}h'^*_{T,1,t-j+\tau})B_1J_wF_1^{-1}J'_w(T^{1/2}\widetilde{\beta}_{1,T})+o_{p^*}(1) \\
&=T^{-1}\textstyle\sum_{t=T+j}^{T+P-\tau}(T^{1/2}H^*_{T,1}(t))'(-JB_0J'+B_1)VB_1J_wF_1^{-1}J'_w(T^{1/2}\widetilde{\beta}_{1,T})+o_{p^*}(1) \\
&\to^{d^*}\ \sigma^4\Gamma_6^*\sim N(0,W)
\end{aligned}$$

where $W=\ln(1+\lambda_P)\sigma^{-8}\lim E\{(T^{1/2}\widetilde{\beta}_{1,T})'J_wF_1^{-1}J'_wB_1VB_1J_wF_1^{-1}J'_wB_1VB_1J_wF_1^{-1}J'_wB_1VB_1J_wF_1^{-1}J'_w(T^{1/2}\widetilde{\beta}_{1,T})\}$. The asymptotic normality follows from the fact that $H^*_{T,1}(t)$ is independent of $T^{1/2}\widetilde{\beta}_{1,T}$ and moreover that $T^{-1}\sum_{t=T+j}^{T+P-\tau}(T^{1/2}H^*_{T,1}(t))\to^{d^*}\int_1^{1+\lambda_P}s^{-1}V^{1/2}W^*(s)ds\sim N(0,\ln(1+\lambda_P)V)$. As in the proof of Theorem 2.3(a), the exact relationship between Γ_6^* and Γ_6 depends upon the additional assumptions stated in the Theorem.

(a) If we let $V=\sigma^2B_1^{-1}$, W simplifies to

$$\begin{aligned}
W&=\sigma^6\ln(1+\lambda_P)\lim E\{(T^{1/2}\widetilde{\beta}_{1,T})'J_wF_1^{-1}J'_w(T^{1/2}\widetilde{\beta}_{1,T})\} \\
&=\sigma^6\ln(1+\lambda_P)\lim E\{(T^{1/2}\widetilde{\beta}_{1,T})'J_wF_1^{-1}(T)J'_w(T^{1/2}\widetilde{\beta}_{1,T})\}\ . \\
&=\sigma^6\ln(1+\lambda_P)\lim E(\hat{d})=\sigma^6\ln(1+\lambda_P)d
\end{aligned}$$

But from Theorem 2.2, the definition of Γ_6 gives us

$$\sigma^4\Gamma_6=(\int_1^{1+\lambda_P}s^{-1}W(s)ds)'V^{1/2}B_1J_wF_1^{-1}J'_wB_1VB_1J_wF_1^{-1}\beta_w\sim N(0,\Omega)$$

where

$$\Omega=\ln(1+\lambda_P)\beta'_wF_1^{-1}J'_wB_1VB_1J_wF_1^{-1}J'_wB_1VB_1J_wF_1^{-1}J'_wB_1VB_1J_wF_1^{-1}\beta_w.$$

Assuming conditional homoskedasticity this simplifies to $\Omega=\sigma^6\ln(1+\lambda_P)\beta'_wF_1^{-1}\beta_w$. The result then follows since under the null, $\beta'_wF_1^{-1}\beta_w=d$.

(b) If β_w is scalar we find that

$$\begin{aligned}
W&=\ln(1+\lambda_P)\lim E\{(T^{1/2}\widetilde{\beta}_{w,T})^2(F_1^{-1})^2J'_wB_1VB_1J_wF_1^{-1}J'_wB_1VB_1J_wF_1^{-1}J'_wB_1VB_1J_w\} \\
&=\ln(1+\lambda_P)\lim E\{\hat{d}F_1(T)(F_{11}^{-1})^4(J'_wB_1VB_1J_w)^3\}\qquad . \\
&=\ln(1+\lambda_P)d(F_1^{-1})^3(J'_wB_1VB_1J_w)^3
\end{aligned}$$

But from Theorem 2.2, the definition of Γ_6 gives us

$$\sigma_u^4\Gamma_6=(\int_1^{1+\lambda_P}s^{-1}W'(s)V^{1/2}ds)B_1J_wF_1^{-1}J'_wB_1VB_1J_wF_1^{-1}J'_w\delta\sim N(0,\Omega)$$

where

$$\Omega=\ln(1+\lambda_P)(\beta_w)^2(F_1^{-1})^4(J'_wB_1VB_1J_w)^3.$$

The result then follows since under the null, $(\beta_w)^2F_1^{-1}=d$.

Proof of bracket 3 in (18): After rearranging terms, the third bracketed term is

$$\begin{aligned}
&\textstyle\sum_{j=-\bar{j}}^{\bar{j}}K(j/M)\sum_{t=T+j}^{T+P-\tau}B^*_{1,t}B^*_{1,t-j}=\sum_{j=-\bar{j}}^{\bar{j}}K(j/M)T^{-1}\sum_{t=T+j}^{T+P-\tau}(T^{1/2}\widetilde{\beta}_{1,T})'B_1^{-1}(t)\times \\
&(-JB_0(t)J'+B_1(t))h^*_{T,1,t+\tau}h'^*_{T,1,t-j+\tau}(-JB_0(t-j)J'+B_1(t-j))B_1^{-1}(t-j)(T^{1/2}\widetilde{\beta}_{1,T}) \\
&=\textstyle\sum_{j=-\bar{j}}^{\bar{j}}K(j/M)T^{-1}\sum_{t=T+j}^{T+P-\tau}(T^{1/2}\widetilde{\beta}_{1,T})J_wF_1^{-1}(t)J'_wB_1(t)h^*_{T,1,t+\tau}h'^*_{T,1,t-j+\tau}B_1(t-j)\times\quad . \\
&J_wF_1^{-1}(t-j)J'_w(T^{1/2}\widetilde{\beta}_{1,T})
\end{aligned}$$

This term is also different from that for the third bracketed term in Theorem 2.2. As in the proof of Theorem 2.3(a), it differs because $T^{1/2}J_w'\widetilde{\beta}_{1,T}$ is not a consistent estimate of β_w. Even so, since $T^{1/2}J_w'\widetilde{\beta}_{1,T} = O_p(1)$, the above term is also $O_p(1)$. To see this, note that similar arguments to those used in the proof for bracket 1 of (18) imply

$$\begin{aligned}
&\textstyle\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)T^{-1}\sum_{t=T+j}^{T+P-\tau}(T^{1/2}\widetilde{\beta}_{1,T})' J_w F_1^{-1}(t)J_w' B_1(t)h_{T,1,t+\tau}^* h_{T,1,t-j+\tau}'^* \times \\
&B_1(t-j)J_w F_1^{-1}(t-j)J_w'(T^{1/2}\widetilde{\beta}_{1,T}) \\
&= \textstyle\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)T^{-1}\sum_{t=T+j}^{T+P-\tau}(T^{1/2}\widetilde{\beta}_{1,T})' J_w F_1^{-1}(t)J_w' B_1(t)(E^* h_{T,1,t+\tau}^* h_{T,1,t-j+\tau}'^*) \times \\
&B_1(t)J_w F_1^{-1}(t)J_w'(T^{1/2}\widetilde{\beta}_{1,T}) + o_{p^*}(1) \\
&= \textstyle\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)T^{-1}\sum_{t=T+j}^{T+P-\tau}(T^{1/2}\widetilde{\beta}_{1,T})' J_w F_1^{-1}J_w' B_1(E^* h_{T,1,t+\tau}^* h_{T,1,t-j+\tau}'^*) \times \\
&B_1 J_w F_1^{-1}J_w'(T^{1/2}\widetilde{\beta}_{1,T}) + o_{p^*}(1) \\
&= T^{-1}\textstyle\sum_{t=T+j}^{T+P-\tau}(T^{1/2}\widetilde{\beta}_{1,T})' J_w F_1^{-1}J_w' B_1 V B_1 J_w F_1^{-1}J_w'(T^{1/2}\widetilde{\beta}_{1,T}) + o_{p^*}(1) \\
&\to^{p^*} \lim \lambda_P (T^{1/2}\widetilde{\beta}_{1,T})' J_w F_1^{-1}J_w' B_1 V B_1 J_w F_1^{-1}J_w'(T^{1/2}\widetilde{\beta}_{1,T}) = \sigma^4\Gamma_7^*.
\end{aligned}$$

As in the proof for bracket 2 above, the exact relationship between Γ_7^* and Γ_7 depends upon the additional assumptions stated in the the Theorem.

(a) If we let $V = \sigma^2 B_1^{-1}$, we immediately see that

$$\begin{aligned}
&\Gamma_7^* \equiv \lambda_P\sigma^{-4}\lim\{(T^{1/2}\widetilde{\beta}_{1,T})' J_w F_1^{-1}J_w' B_1 V B_1 J_w F_1^{-1}J_w'(T^{1/2}\widetilde{\beta}_{1,T})\} \\
&= \lambda_P\sigma^{-2}\lim\{(T^{1/2}\widetilde{\beta}_{1,T})J_w F_1^{-1}J_w'(T^{1/2}\widetilde{\beta}_{1,T})\} = \lambda_P\sigma^{-2}\lim\hat{d} = \sigma^{-2}\lambda_P d
\end{aligned}.$$

But under the null, and with the additional assumption of conditional homoskedasticity, from Theorem 2.2 we know that

$$\Gamma_7 \equiv \sigma^{-4}\lambda_P\beta_w' F_1^{-1}J_w' B_1 V B_1 J_w F_1^{-1}\beta_w = \sigma^{-2}\lambda_P\beta_w' F_1^{-1}\beta_w = \sigma^{-2}\lambda_P d = \Gamma_7^*$$

and the proof is complete.

(b) If we let β_w be scalar we find that

$$\begin{aligned}
&\sigma^4\Gamma_7^* \equiv \lim\lambda_P(T^{1/2}\widetilde{\beta}_{1,T})' J_w F_1^{-1}J_w' B_1 V B_1 J_w F_1^{-1}J_w'(T^{1/2}\widetilde{\beta}_{1,T}) \\
&= \lambda_P\lim(T^{1/2}\widetilde{\beta}_{w,T})^2(F_1^{-1})^2 J_w' B_1 V B_1 J_w \\
&= \lambda_P\lim\hat{d}F_1(T)(F_1^{-1})^2 J_w' B_1 V B_1 J_w \\
&= \lambda_P d F_1^{-1}J_w' B_1 V B_1 J_w + o_p(1)
\end{aligned}.$$

But under the null, and with the additional assumption of that β_w is scalar, from Theorem 2.2 we know that

$$\begin{aligned}
&\sigma^4\Gamma_7 \equiv \lambda_P\beta_w' F_1^{-1}J_w' B_1 V B_1 J_w F_1^{-1}\beta_w = \lambda_P(\beta_w)^2(F_1^{-1})^2 J_w' B_1 V B_1 J_w \\
&= \lambda_P d F_1^{-1}J_w' B_1 V B_1 J_w = \sigma^4\Gamma_7^*
\end{aligned}$$

and the proof is complete.

Proof of bracket 4 in (18): We must show each of the remaining cross-products of $A_{1,t}^*$, $A_{2,t}^*$, B_t^*, C_t^*, and D_t^* with $A_{1,t-j}^*$, $A_{2,t-j}^*$, B_{t-j}^*, C_{t-j}^*, and D_{t-j}^* are $o_{p^*}(1)$. The proof is nearly identical to that for the fourth bracketed term from the proof of Theorem 2.2. The primary difference is that the relevant moment conditions are all defined in terms of $h_{T,1,t+\tau}^*$ rather than $h_{T,1,t+\tau}$. The distinction does not appreciably alter the proof and hence for brevity we do not repeat the details.

Proof for the rolling scheme: Results for the rolling scheme differ only in the definition of $H_{T,1}^*(t) = T^{-1}\sum_{s=t-T+1}^{t} h_{T,1,s+\tau}^*$ (and to a lesser extent $\hat{H}_{T,1}^*(t) = T^{-1}\sum_{s=t-T+1}^{t}\hat{h}_{T,1,s+\tau}^*$). In particular, if we substitute $T^{1/2}H^*_{T,1}(t) \Rightarrow^* V^{1/2}(W^*(s) - W^*(s-1))$ for $T^{1/2}H^*_{T,1}(t) \Rightarrow^* V^{1/2}s^{-1}W^*(s)$ as used above, we obtain the desired conclusion.

Proof of Theorem 2.4: Regardless of whether the recursive or rolling scheme is used, the proof follows very similar arguments to those used in Theorems 2.3(a) and (b). Any differences that arise come from differences in the asymptotic behavior of $T^{1/2}J_w'\widetilde{\beta}_{1,T}$ as delineated in Lemma 1 (b-i) versus Lemma 1 (b-ii). Since the decomposition at the beginning of the proof of Theorem 2.3(a) is unaffected by the asymptotic behavior of $T^{1/2}J_w'\widetilde{\beta}_{1,T}$ and the first bracketed term does not depend upon the value of either β_w or $T^{1/2}J_w'\widetilde{\beta}_{1,T}$ the same proof can be applied to show $2\Gamma_1^* - \Gamma_2^* =^{d^*} 2\Gamma_1 - \Gamma_2$ and $\Gamma_5^* =^{d^*} \Gamma_5$. For the third bracketed term, the asymptotic behavior of $T^{1/2}J_w'\widetilde{\beta}_{1,T}$ is also irrelevant – all that matters is that the ridge constraint is imposed.

Differences in the proof arise for the second, and fourth bracketed terms. For the fourth bracketed term, the differences remain minor since we need only show that the relevant components are all $o_{p^*}(1)$ and the corresponding proofs only make use of the fact that Lemma 1(b) implies $T^{1/2}J_w'\widetilde{\beta}_{1,T} = O_p(1)$. These arguments continue to hold here since $T^{1/2}J_w'\widetilde{\beta}_{1,T}$ remains $O_p(1)$ – despite also having the property that $T^{1/2}J_w'\widetilde{\beta}_{1,T} \to^p \sqrt{\frac{d}{\beta_w' F_1^{-1}\beta_w}}\beta_w$.

We therefore focus attention on showing that $\Gamma_i^* =^{d^*} \Gamma_i$ for $i = 3, 6, 7$. In each case, the different asymptotic behavior of $T^{1/2}J_w'\widetilde{\beta}_{1,T}$ does impact the proofs directly. And as we saw earlier, in each case the proof also requires additional assumptions as noted in the statement of the theorem.

<u>Proof that $\Gamma_3^* =^{d^*} \Gamma_3$:</u> As in the proof for Theorem 2.3(a), the second bracketed term satisfies

$$T^{-1/2}2\sum_{t=T}^{T+P-\tau} h'^*_{T,1,t+\tau}B_1(t)J_wF_1^{-1}(t)J_w'(T^{1/2}\widetilde{\beta}_{1,T}) = T^{-1/2}2\sum_{t=T}^{T+P-\tau} h'^*_{T,1,t+\tau}B_1J_wF_1^{-1}J_w'(T^{1/2}\widetilde{\beta}_{1,T})+o_{p^*}(1).$$

However, since $T^{1/2}J_w'\widetilde{\beta}_{1,T} \to^p \sqrt{\frac{d}{\beta_w' F_1^{-1}\beta_w}}\beta_w$ we also have

$$\begin{aligned} & T^{-1/2}2\sum_{t=T}^{T+P-\tau} h'^*_{T,1,t+\tau}B_1(t)J_wF_1^{-1}(t)J_w'(T^{1/2}\widetilde{\beta}_{1,T}) \\ = \; & T^{-1/2}2\sqrt{\frac{d}{\beta_w' F_1^{-1}\beta_w}}\sum_{t=T}^{T+P-\tau} h'^*_{T,1,t+\tau}B_1J_wF_1^{-1}\beta_w + o_{p^*}(1) \to^{d^*} N(0,4W) \end{aligned}$$

where

$$W = (\tfrac{d}{\beta_w' F_1^{-1}\beta_w})\lambda_P\beta_w' F_1^{-1}J_w'B_1VB_1J_wF_1^{-1}\beta_w \quad .$$

Since $\Gamma_3 \sim N(0, 4\Omega), \Omega = \lambda_P\beta_w' F_1^{-1}J_w'B_1VB_1J_wF_1^{-1}\beta_w$, the precise relationship between Γ_3^* and Γ_3 depends on the relationship between Ω and W. This in turn depends upon the additional restrictions in the statement of the Theorem.

(a) If we let $V = \sigma^2 B_1^{-1}$, W simplifies to

$$W = \sigma^2(\frac{d}{\beta_w' F_1^{-1}\beta_w})\lambda_P\beta_w' F_1^{-1}\beta_w = \sigma^2\lambda_P d.$$

The result follows since under the null hypothesis, $\Omega = \lambda_P\beta_w' F_1^{-1}J_w'B_1VB_1J_wF_1^{-1}\beta_w = \sigma^2\lambda_P\beta_w' F_1^{-1}\beta_w = \sigma^2\lambda_P d$.

(b) If we let $\dim(\beta_w) = 1$ and note that in this case $J_w' B_1 V B_1 J_w = F_1 tr((-JB_0J' + B_1)V)$, W simplifies to

$$W = d\lambda_P tr((-JB_0J' + B_1)V).$$

The result follows since under the null hypothesis, $\Omega = d\lambda_P tr((-JB_0J' + B_1)V)$ and the proof is complete.

Proof that $\Gamma_6^* =^{d^*} \Gamma_6$: As in the proof for Theorem 2.3(b), the second bracketed term satisfies

$$\begin{array}{l}\sum_{j=-\bar{j}}^{\bar{j}} K(j/M) \sum_{t=T+j}^{T+P-\tau} A_{1,t}^* B_{1,t-j}^* \\ = \sum_{j=-\bar{j}}^{\bar{j}} K(j/M) T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2} H_{T,1}^*(t))' (-JB_0(t)J' + B_1(t)) \times \\ h_{T,1,t+\tau}^* h_{T,1,t-j+\tau}'^* (-JB_0(t-j)J' + B_1(t-j)) B_1^{-1}(t-j)(T^{1/2}\widetilde{\beta}_{1,T}) \\ = T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2} H_{T,1}^*(t))' B_1 J_w F_1^{-1} J_w' B_1 V B_1 J_w F_1^{-1} (T^{1/2} J_w' \widetilde{\beta}_{1,T}) + o_{p^*}(1)\end{array} .$$

However, since $T^{1/2} J_w' \widetilde{\beta}_{1,T} \to^p \sqrt{\frac{d}{\beta_w' F_1^{-1} \beta_w}} \beta_w$ we also have

$$\begin{array}{l}T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2} H_{T,1}^*(t))' B_1 J_w F_1^{-1} J_w' B_1 V B_1 J_w F_1^{-1} (T^{1/2} J_w' \widetilde{\beta}_{1,T}) \\ = (\sqrt{\frac{d}{\beta_w' F_1^{-1} \beta_w}}) T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2} H_{T,1}^*(t))' B_1 J_w F_1^{-1} J_w' B_1 V B_1 J_w F_1^{-1} \beta_w + o_p^*(1) \\ \to^{d^*} N(0, W)\end{array}$$

where $W = \ln(1 + \lambda_P)(\frac{d}{\beta_w' F_1^{-1} \beta_w})\{\beta_w' F_1^{-1} J_w' B_1 V B_1 J_w F_1^{-1} J_w' B_1 V B_1 J_w F_1^{-1} J_w' B_1 V B_1 J_w F_1^{-1} \beta_w\}$. The asymptotic normality follows from the fact that $T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2} H_{T,1}^*(t)) \to^{d^*} \int_1^{1+\lambda_P} s^{-1} V^{1/2} W^*(s) ds \sim N(0, \ln(1+\lambda_P)V)$. Since $\Gamma_6 \sim N(0, \Omega), \Omega = \ln(1+\lambda_P)\beta_w' F_1^{-1} J_w' B_1 V B_1 J_w F_1^{-1} J_w' B_1 V B_1 J_w F_1^{-1} J_w' B_1 V B_1 J_w F_1^{-1} \beta_w$, the precise relationship between Γ_6^* and Γ_6 depends on the relationship between Ω and W. This in turn depends upon the additional restrictions in the statement of the Theorem.

(a) If we let $V = \sigma^2 B_1^{-1}$, W simplifies to

$$W = \sigma^6 \ln(1 + \lambda_P) d \ .$$

The result follows since under the null hypothesis,

$$\begin{array}{l}\Omega = \ln(1 + \lambda_P)\beta_w' F_1^{-1} J_w' B_1 V B_1 J_w F_1^{-1} J_w' B_1 V B_1 J_w F_1^{-1} J_w' B_1 V B_1 J_w F_1^{-1} \beta_w \\ = \sigma^6 \ln(1 + \lambda_P)\beta_w' F_1^{-1} \beta_w = \sigma^6 \ln(1 + \lambda_P) d.\end{array} .$$

(b) If we let $\dim(\beta_w) = 1$ and note that in this case $J_w' B_1 V B_1 J_w = F_1 tr((-JB_0J' + B_1)V)$, W simplifies to

$$W = \ln(1 + \lambda_P) d \cdot tr((-JB_0J' + B_1)V)^3 \ .$$

The result follows since under the null hypothesis, $\Omega = \ln(1 + \lambda_P) d \cdot tr((-JB_0J' + B_1)V)^3$ and the proof is complete.

Proof that $\Gamma_7^* =^{d^*} \Gamma_7$: As in the proof for Theorem 2.3(b), the third bracketed term satisfies

$$\begin{array}{l}\sum_{j=-\bar{j}}^{\bar{j}} K(j/M) \sum_{t=T+j}^{T+P-\tau} B_{1,t}^* B_{1,t-j}^* = \sum_{j=-\bar{j}}^{\bar{j}} K(j/M) T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2}\widetilde{\beta}_{1,T})' B_1^{-1}(t)(-JB_0(t)J' + B_1(t)) \times \\ h_{T,1,t+\tau}^* h_{T,1,t-j+\tau}'^* (-JB_0(t-j)J' + B_1(t-j)) B_1^{-1}(t-j)(T^{1/2}\widetilde{\beta}_{1,T}) \\ = T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2}\widetilde{\beta}_{1,T})' J_w F_1^{-1} J_w' B_1 V B_1 J_w F_1^{-1} (T^{1/2} J_w' \widetilde{\beta}_{1,T}) + o_{p^*}(1)\end{array} .$$

However, since $T^{1/2}J_w'\widetilde{\beta}_{1,T} \to^p \sqrt{\frac{d}{\beta_w' F_1^{-1}\beta_w}}\beta_w$ we also have

$$\begin{array}{l} T^{-1}\sum_{t=T+j}^{T+P-\tau}(T^{1/2}\widetilde{\beta}_{1,T})' J_w F_1^{-1} J_w' B_1 V B_1 J_w F_1^{-1}(T^{1/2}J_w'\widetilde{\beta}_{1,T}) \\ \to^p \lambda_P(\frac{d}{\beta_w' F_1^{-1}\beta_w})\beta_w' F_1^{-1} J_w' B_1 V B_1 J_w F_1^{-1}\beta_w \equiv \Gamma_7^* \end{array} .$$

In contrast, the associated term from Theorem 2.2 takes the value $\Gamma_7 = \lambda_P \beta_w' F_1^{-1} J_w' B_1 V B_1 J_w F_1^{-1}\beta_w$. The exact relationship between these two terms depends upon the additional assumptions stated in the Theorem.

(a) If we let $V = \sigma^2 B_1^{-1}$, Γ_7^* simplifies to $\lambda_P \sigma^2 d$. The result follows since under the null hypothesis, $\Gamma_7 = \lambda_P \beta_w' F_1^{-1} J_w' B_1 V B_1 J_w F_1^{-1}\beta_w = \lambda_P \sigma^2 d$ and the proof is complete.

(b) If we let $\dim(\beta_w) = 1$ and note that in this case $J_w' B_1 V B_1 J_w = F_1 tr((-JB_0J' + B_1)V)$, Γ_7^* simplifies to $\lambda_P d \cdot tr((-JB_0J' + B_1)V)$. The result follows since under the null hypothesis, $\Gamma_7 = \lambda_P \beta_w' F_1^{-1} J_w' B_1 V B_1 J_w F_1^{-1}\beta_w = \lambda_P d \cdot tr((-JB_0J' + B_1)V)$ and the proof is complete.

References

Atkeson, Andrew, and Lee E. Ohanian (2001), "Are Phillips Curves Useful for Forecasting Inflation?" Federal Reserve Bank of Minneapolis *Quarterly Review* 25, 2-11.

Bachmeier, Lance, and Norman R. Swanson (2005), "Predicting Inflation: Does The Quantity Theory Help?" *Economic Inquiry* 43, 570-585

Bordo, Michael, and Joseph G. Haubrich (2008), "The Yield Curve as a Predictor of Growth: Long-Run Evidence, 1875-1997," *Review of Economics and Statistics* 90, 182-185.

Bruneau, C., O. De Bandt, A. Flageollet, and E. Michaux (2007), "Forecasting Inflation Using Economic Indicators: The Case of France," *Journal of Forecasting* 2, 1-22.

Butler, Alexander W., Gustavo Grullon, and James P. Weston (2005), "Can Managers Forecast Aggregate Market Returns?" *Journal of Finance* 60, 963-986.

Campbell, John Y., and Samuel B. Thompson (2008), "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?" *Review of Financial Studies* 21, 1509-1531.

Chen, Yu-Chin, Kenneth S. Rogoff, and Barbara Rossi (2010), "Can Exchange Rates Forecast Commodity Prices?" *Quarterly Journal of Economics* 125, 1145-1194.

Clark, Todd E., and Michael W. McCracken (2001), "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics* 105, 85-110.

Clark, Todd E., and Michael W. McCracken (2005), "Evaluating Direct Multistep Forecasts," *Econometric Reviews* 24, 369-404.

Clark, Todd E., and Michael W. McCracken (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.

Clark, Todd E., and Michael W. McCracken (2009), "Combining Forecasts From Nested Models," *Oxford Bulletin of Economics and Statistics* 71, 303-329.

Clark, Todd E., and Michael W. McCracken (2012), "Reality Checks and Comparisons of Nested Predictive Models," *Journal of Business and Economic Statistics* 30, 53-66.

Clark, Todd E., and Kenneth D. West (2006), "Using Out-of-Sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis," *Journal of Econometrics* 135, 155-186.

Clark, Todd E., and Kenneth D. West (2007), "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models," *Journal of Econometrics* 138, 291-311.

Cooper, Michael, and Huseyin Gulen (2006), "Is Time-Series-Based Predictability Evident in Real Time?" *Journal of Business* 79, 1263-1292.

Corradi, Valentina, and Norman R. Swanson (2002), "A Consistent Test for Nonlinear Out of Sample Predictive Accuracy," *Journal of Econometrics* 110, 353-381.

de Jong, Robert M., and James Davidson (2000), "The Functional Central Limit Theorem and Weak Convergence to Stochastic Integrals I: Weakly Dependent Processes," *Econometric Theory* 16, 621-42.

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

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

Ferreira, Miguel A., and Pedro Santa-Clara (2011), "Forecasting Stock Market Returns: The Sum of the Parts is More than the Whole," *Journal of Financial Economics* 100, 514-537.

Giacomini, Raffaella, and Barbara Rossi (2006), "How Stable Is the Forecasting Performance of the Yield Curve for Output Growth?" *Oxford Bulletin of Economics and Statistics* 68, 783-795.

Giacomini, Rafaella, and Halbert White (2006), "Tests of Conditional Predictive Ability," *Econometrica* 74, 1545-1578.

Goncalves, Sylvia, and Lutz Kilian (2007), "Asymptotic and Bootstrap Inference for AR(∞) Processes with Conditional Heteroskedasticity," *Econometric Reviews* 26, 609-641.

Goyal, Amit, and Ivo Welch (2003), "Predicting the Equity Premium with Dividend Ratios," *Management Science* 49, 639-54.

Goyal, Amit, and Ivo Welch (2008), "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction, *Review of Financial Studies* 21, 1455-1508.

Groen, Jan J. (1999), "Long Horizon Predictability of Exchange Rates: Is it for Real?" *Empirical Economics* 24, 451-469.

Guo, Hui (2006), "On the Out-of-Sample Predictability of Stock Market Returns," *Journal of Business* 27, 645-670.

Hansen, Bruce E. (1996), "Erratum: The Likelihood Ratio Test Under Nonstandard Conditions: Testing the Markov Switching Model of GNP," *Journal of Applied Econometrics* 11, 195-198.

Hansen, Bruce E. (2008), "Generalized Shrinkage Estimators," manuscript, University of Wisconsin.

Hjalmarsson, Erik (2009), "Should We Expect Significant Out-of-Sample Results when Predicting Stock Returns? in G.I. Ellison (ed.), *Stock Returns: Cyclicity, Prediction and Economic Consequences*, Nova Science Publishers, New York, pp. 269-274.

Inoue, Atsushi, and Lutz Kilian (2004), "In–Sample or Out–of–Sample Tests of Predictability? Which One Should We Use?" *Econometric Reviews* 23, 371-402.

Inoue, Atsushi, and Barbara Rossi (2008), "Monitoring and Forecasting Financial Crises," *Journal of Money, Credit, and Banking* 40, 523-534

Kilian, Lutz (1999), "Exchange Rates and Monetary Fundamentals: What Do We Learn from Long-Horizon Regressions?" *Journal of Applied Econometrics* 14, 491-510.

Kilian, Lutz, and Mark P. Taylor (2003), "Why Is it So Difficult to Beat the Random Walk Forecast of Exchange Rates?" *Journal of International Economics* 60, 85-107.

Lettau, Martin, and Sydney Ludvigson (2001), "Consumption, Aggregate Wealth, and Expected Stock Returns," *Journal of Finance* 56, 815-849.

Mark, Nelson C. (1995), "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability," *American Economic Review* 85, 201-218.

McCracken, Michael W. (2007), "Asymptotics for Out-of-Sample Tests of Granger Causality," *Journal of Econometrics* 140, 719-752.

Meese, Richard, and Kenneth Rogoff (1988), "Was it Real? The Exchange Rate–Interest Differential Relation Over the Modern Floating-Rate Period," *Journal of Finance* 43, 933-948.

Molodtsova, Tanya, and David H. Papell (2009), "Out-of-Sample Exchange Rate Predictability with Taylor Rule Fundamentals," *Journal of International Economics* 77, 167-180.

Pesaran, M. Hashem, and Allan Timmermann (1995), "Predictability of Stock Returns: Robustness and Economic Significance," *Journal of Finance* 50, 1201-1228.

Politis, Dimitris N., and Joseph P. Romano (1994), "The Stationary Bootstrap," *Journal of the American Statistical Association* 89, 1303-1313.

Rapach, David, and Mark E. Wohar (2006), "In-Sample vs. Out-of-Sample Tests of Stock Return Predictability in the Context of Data Mining," *Journal of Empirical Finance* 13, 231-247.

Stine, Robert A. (1987), "Estimating Properties of Autoregressive Forecasts," *Journal of the American Statistical Association* 82, 1072-1078.

Stock, James H., and Mark W. Watson (1999), "Forecasting Inflation," *Journal of Monetary Economics* 44, 293-335.

Stock, James H., and Mark W. Watson (2003), "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature* 41, 788-829.

Trenkler, G., and H. Toutenburg (1992), "Pre-test Procedures and Forecasting in the Regression Model Under Restrictions," *Journal of Statistical Planning and Inference* 30, 249-56.

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

White, Halbert (2000), "A Reality Check For Data Snooping," *Econometrica* 68, 1097-1127.

Figure 1: Densities of MSE(null model)/MSE(alt. model), R = 80, P = 80
DGP 3 experiments

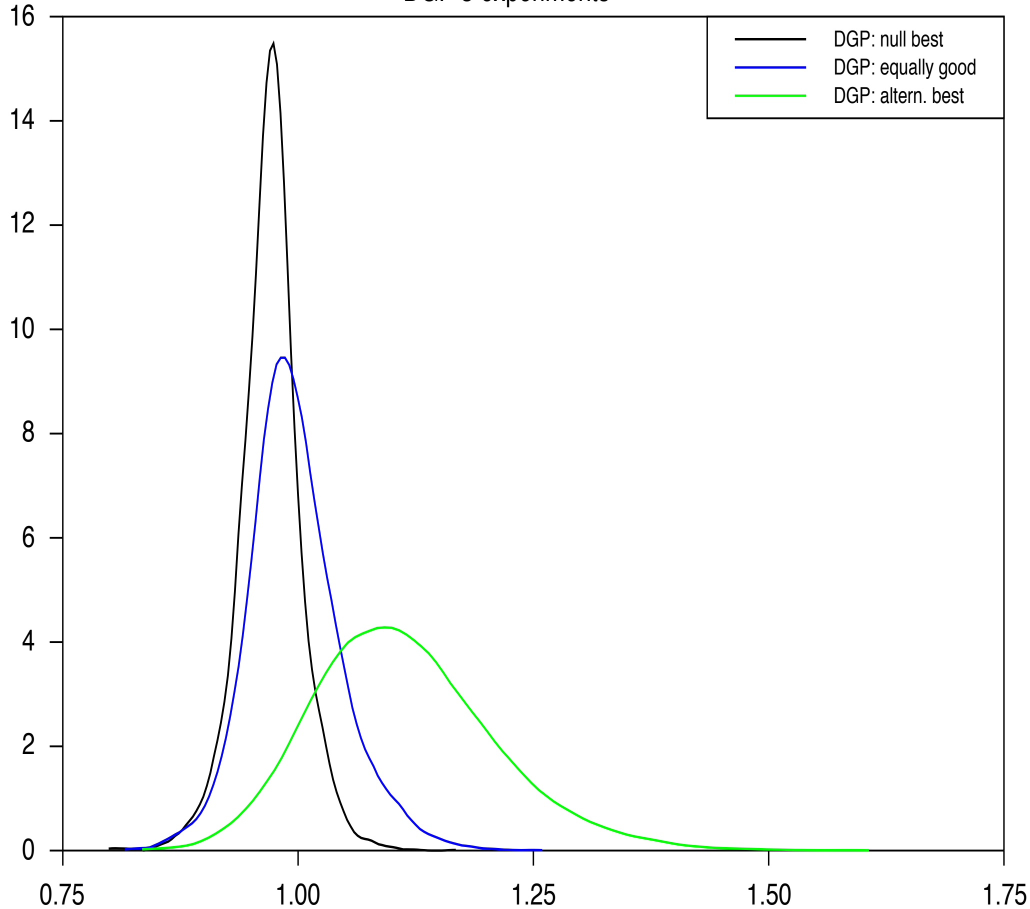

Table 1: Monte Carlo Results on Size
(nominal size = 10%)

DGP 1, 1-step forecasts								
statistic	*source of critical values*	T=40 P=80	T=40 P=120	T=80 P=40	T=80 P=80	T=80 P=120	T=120 P=40	T=120 P=80
MSE-F	non-parametric	0.062	0.058	0.084	0.067	0.063	0.083	0.068
MSE-F	fixed regressor	0.107	0.107	0.106	0.099	0.103	0.096	0.095
MSE-t	non-parametric	0.070	0.068	0.101	0.073	0.069	0.095	0.078
MSE-t	fixed regressor	0.089	0.093	0.096	0.088	0.094	0.086	0.085
MSE-t	normal	0.066	0.064	0.094	0.069	0.067	0.084	0.070
MSE-t, 2-sided	normal	0.097	0.105	0.113	0.106	0.100	0.109	0.106
DGP 2, 1-step forecasts								
statistic	*source of critical values*	T=40 P=80	T=40 P=120	T=80 P=40	T=80 P=80	T=80 P=120	T=120 P=40	T=120 P=80
MSE-F	non-parametric	0.054	0.048	0.080	0.062	0.057	0.083	0.070
MSE-F	fixed regressor	0.101	0.096	0.101	0.102	0.096	0.099	0.103
MSE-t	non-parametric	0.065	0.055	0.094	0.074	0.064	0.097	0.079
MSE-t	fixed regressor	0.088	0.088	0.092	0.089	0.085	0.091	0.093
MSE-t	normal	0.059	0.053	0.085	0.068	0.058	0.086	0.076
MSE-t, 2-sided	normal	0.098	0.100	0.113	0.114	0.099	0.115	0.112
DGP 3, 1-step forecasts								
statistic	*source of critical values*	T=40 P=80	T=40 P=120	T=80 P=40	T=80 P=80	T=80 P=120	T=120 P=40	T=120 P=80
MSE-F	non-parametric	0.041	0.044	0.068	0.060	0.055	0.080	0.072
MSE-F	fixed regressor	0.083	0.094	0.089	0.097	0.090	0.084	0.093
MSE-t	non-parametric	0.055	0.050	0.092	0.075	0.064	0.100	0.084
MSE-t	fixed regressor	0.077	0.087	0.086	0.089	0.082	0.088	0.088
MSE-t	normal	0.047	0.049	0.081	0.070	0.061	0.085	0.078
MSE-t, 2-sided	normal	0.093	0.098	0.108	0.094	0.093	0.102	0.099
DGP 4, 4-step forecasts								
statistic	*source of critical values*	T=40 P=80	T=40 P=120	T=80 P=40	T=80 P=80	T=80 P=120	T=120 P=40	T=120 P=80
MSE-F	non-parametric	0.102	0.091	0.156	0.111	0.094	0.162	0.114
MSE-F	fixed regressor	0.148	0.142	0.131	0.132	0.131	0.128	0.124
MSE-t	non-parametric	0.110	0.094	0.152	0.114	0.097	0.152	0.115
MSE-t	fixed regressor	0.133	0.136	0.122	0.117	0.123	0.117	0.113
MSE-t	normal	0.115	0.103	0.158	0.115	0.105	0.162	0.119
MSE-t, 2-sided	normal	0.154	0.150	0.209	0.161	0.152	0.209	0.165

Notes:
1. The data generating processes are defined in equations (6), (9), (12), and (15). In these experiments, the coefficients b_{ij} are scaled such that the null and alternative models are expected to be equally accurate (on average) over the forecast sample.
2. For each artificial data set, forecasts of $y_{t+\tau}$ (where τ denotes the forecast horizon) are formed recursively using estimates of equations (7) and (8) in the case of the DGP 1 experiments, equations (10) and (11) in the case of the DGP 2 experiments, equations (13) and (14) in the case of the DGP 3 experiments, and equations (16) and (17) in the case of the DGP 3 experiments. These forecasts are then used to form the indicated test statistics, defined in Section 2.3. T and P refer to the number of in–sample observations and 1-step ahead forecasts, respectively.
3. In each Monte Carlo replication, the simulated test statistics are compared against bootstrapped critical values, using a significance level of 10%. Section 2.6 describes the bootstrap procedures.
4. The number of Monte Carlo simulations is 5000; the number of bootstrap draws is 499.

Table 2: Monte Carlo Results on Size
Additional DGPs
(nominal size = 10%)

DGP 5, 1-step forecasts								
statistic	*source of critical values*	T=40 P=80	T=40 P=120	T=80 P=40	T=80 P=80	T=80 P=120	T=120 P=40	T=120 P=80
MSE-F	non-parametric	0.092	0.078	0.111	0.095	0.081	0.112	0.099
MSE-F	fixed regressor	0.120	0.110	0.107	0.104	0.104	0.097	0.103
MSE-t	non-parametric	0.111	0.094	0.143	0.110	0.095	0.137	0.117
MSE-t	fixed regressor	0.103	0.102	0.099	0.096	0.098	0.098	0.095
MSE-t	normal	0.096	0.080	0.109	0.092	0.082	0.114	0.098
MSE-t, 2-sided	normal	0.090	0.084	0.106	0.098	0.093	0.112	0.092
DGP 6, 4-step forecasts								
statistic	*source of critical values*	T=40 P=80	T=40 P=120	T=80 P=40	T=80 P=80	T=80 P=120	T=120 P=40	T=120 P=80
MSE-F	non-parametric	0.102	0.089	0.157	0.112	0.101	0.168	0.113
MSE-F	fixed regressor	0.193	0.188	0.159	0.158	0.169	0.137	0.134
MSE-t	non-parametric	0.120	0.106	0.169	0.123	0.112	0.169	0.118
MSE-t	fixed regressor	0.168	0.170	0.139	0.140	0.149	0.125	0.120
MSE-t	normal	0.123	0.110	0.171	0.128	0.117	0.169	0.122
MSE-t, 2-sided	normal	0.141	0.115	0.218	0.148	0.132	0.213	0.150
DGP 7, 1-step forecasts								
statistic	*source of critical values*	T=40 P=80	T=40 P=120	T=80 P=40	T=80 P=80	T=80 P=120	T=120 P=40	T=120 P=80
MSE-F	non-parametric	0.065	0.061	0.088	0.071	0.067	0.088	0.070
MSE-F	fixed regressor	0.120	0.120	0.109	0.113	0.111	0.106	0.108
MSE-t	non-parametric	0.077	0.068	0.103	0.081	0.075	0.105	0.080
MSE-t	fixed regressor	0.104	0.104	0.100	0.099	0.103	0.098	0.092
MSE-t	normal	0.072	0.067	0.094	0.077	0.072	0.095	0.073
MSE-t, 2-sided	normal	0.098	0.095	0.123	0.111	0.109	0.116	0.118

Notes:
1. See the notes to Table 1.
2. These experiments address the properties of our proposed testing procedures in DGP and forecasting settings in which the assumptions necessary to prove the validity of the bootstrap are not satisfied. As detailed in section 3.1, in DGP 5 the number of variables included in model 1 and not model 0 is 3 and the forecast errors are conditionally heteroskedastic. In DGP 6, the number of variables included in model 1 and not model 0 is 3 and the forecast errors are serially correlated. In DGP 7, the forecasting models are misspecified, in the sense that neither model includes two variables that appear in the data-generating process for the predictand. In each DGP, the coefficients b_{ij} are scaled such that the null and alternative models are equally accurate (on average) over the forecast sample.

Table 3: Monte Carlo Results on Size
Rolling Forecasts
(nominal size = 10%)

DGP 1, 1-step forecasts								
statistic	*source of critical values*	T=40 P=80	T=40 P=120	T=80 P=40	T=80 P=80	T=80 P=120	T=120 P=40	T=120 P=80
MSE-F	non-parametric	0.047	0.038	0.078	0.055	0.049	0.081	0.063
MSE-F	fixed regressor	0.106	0.107	0.105	0.100	0.105	0.096	0.097
MSE-t	non-parametric	0.059	0.052	0.100	0.065	0.059	0.094	0.072
MSE-t	fixed regressor	0.088	0.091	0.094	0.088	0.089	0.085	0.085
MSE-t	normal	0.053	0.047	0.088	0.057	0.052	0.084	0.067
MSE-t, 2-sided	normal	0.082	0.084	0.114	0.098	0.092	0.106	0.094
DGP 2, 1-step forecasts								
statistic	*source of critical values*	T=40 P=80	T=40 P=120	T=80 P=40	T=80 P=80	T=80 P=120	T=120 P=40	T=120 P=80
MSE-F	non-parametric	0.036	0.032	0.078	0.052	0.039	0.080	0.065
MSE-F	fixed regressor	0.097	0.099	0.103	0.097	0.098	0.102	0.103
MSE-t	non-parametric	0.049	0.041	0.092	0.063	0.049	0.096	0.076
MSE-t	fixed regressor	0.086	0.088	0.092	0.089	0.088	0.092	0.093
MSE-t	normal	0.044	0.036	0.083	0.060	0.043	0.086	0.067
MSE-t, 2-sided	normal	0.100	0.105	0.112	0.108	0.091	0.123	0.110
DGP 3, 1-step forecasts								
statistic	*source of critical values*	T=40 P=80	T=40 P=120	T=80 P=40	T=80 P=80	T=80 P=120	T=120 P=40	T=120 P=80
MSE-F	non-parametric	0.020	0.018	0.062	0.044	0.034	0.080	0.060
MSE-F	fixed regressor	0.074	0.080	0.087	0.090	0.088	0.084	0.094
MSE-t	non-parametric	0.030	0.027	0.086	0.058	0.044	0.098	0.076
MSE-t	fixed regressor	0.068	0.076	0.084	0.087	0.080	0.086	0.091
MSE-t	normal	0.028	0.023	0.076	0.053	0.039	0.085	0.070
MSE-t, 2-sided	normal	0.124	0.141	0.103	0.093	0.095	0.107	0.099
DGP 4, 4-step forecasts								
statistic	*source of critical values*	T=40 P=80	T=40 P=120	T=80 P=40	T=80 P=80	T=80 P=120	T=120 P=40	T=120 P=80
MSE-F	non-parametric	0.112	0.103	0.146	0.104	0.091	0.165	0.110
MSE-F	fixed regressor	0.160	0.162	0.132	0.137	0.140	0.129	0.126
MSE-t	non-parametric	0.132	0.127	0.151	0.114	0.101	0.162	0.118
MSE-t	fixed regressor	0.142	0.148	0.120	0.126	0.132	0.116	0.114
MSE-t	normal	0.128	0.123	0.156	0.115	0.102	0.165	0.115
MSE-t, 2-sided	normal	0.158	0.147	0.198	0.166	0.143	0.208	0.153

Notes:
1. See the notes to Table 1.
2. In these experiments, the forecasting scheme is rolling, rather than recursive.

Table 4: Monte Carlo Results on Power
(nominal size = 10%)

DGP 1, 1-step forecasts								
statistic	*source of critical values*	T=40 P=80	T=40 P=120	T=80 P=40	T=80 P=80	T=80 P=120	T=120 P=40	T=120 P=80
MSE-F	non-parametric	0.272	0.354	0.248	0.330	0.423	0.264	0.352
MSE-F	fixed regressor	0.486	0.599	0.451	0.589	0.705	0.523	0.649
MSE-t	non-parametric	0.313	0.385	0.299	0.365	0.459	0.308	0.388
MSE-t	fixed regressor	0.361	0.464	0.288	0.404	0.529	0.293	0.419
MSE-t	normal	0.304	0.384	0.277	0.354	0.450	0.287	0.379
MSE-t, 2-sided	normal	0.185	0.247	0.175	0.222	0.300	0.182	0.239
DGP 2, 1-step forecasts								
statistic	*source of critical values*	T=40 P=80	T=40 P=120	T=80 P=40	T=80 P=80	T=80 P=120	T=120 P=40	T=120 P=80
MSE-F	non-parametric	0.263	0.351	0.250	0.335	0.422	0.269	0.363
MSE-F	fixed regressor	0.481	0.609	0.445	0.593	0.715	0.518	0.659
MSE-t	non-parametric	0.296	0.385	0.295	0.372	0.457	0.311	0.397
MSE-t	fixed regressor	0.360	0.487	0.280	0.412	0.534	0.294	0.425
MSE-t	normal	0.282	0.374	0.270	0.352	0.448	0.285	0.380
MSE-t, 2-sided	normal	0.178	0.233	0.172	0.232	0.300	0.184	0.251
DGP 3, 1-step forecasts								
statistic	*source of critical values*	T=40 P=80	T=40 P=120	T=80 P=40	T=80 P=80	T=80 P=120	T=120 P=40	T=120 P=80
MSE-F	non-parametric	0.282	0.434	0.268	0.429	0.569	0.319	0.484
MSE-F	fixed regressor	0.527	0.697	0.491	0.685	0.821	0.579	0.763
MSE-t	non-parametric	0.349	0.485	0.346	0.497	0.616	0.396	0.547
MSE-t	fixed regressor	0.426	0.601	0.329	0.533	0.680	0.366	0.568
MSE-t	normal	0.331	0.474	0.319	0.476	0.606	0.368	0.527
MSE-t, 2-sided	normal	0.200	0.322	0.207	0.320	0.451	0.241	0.370
DGP 4, 4-step forecasts								
statistic	*source of critical values*	T=40 P=80	T=40 P=120	T=80 P=40	T=80 P=80	T=80 P=120	T=120 P=40	T=120 P=80
MSE-F	non-parametric	0.290	0.349	0.315	0.347	0.421	0.342	0.383
MSE-F	fixed regressor	0.467	0.563	0.428	0.557	0.649	0.509	0.611
MSE-t	non-parametric	0.324	0.379	0.328	0.375	0.442	0.366	0.406
MSE-t	fixed regressor	0.360	0.440	0.270	0.380	0.487	0.286	0.399
MSE-t	normal	0.332	0.393	0.339	0.385	0.460	0.373	0.419
MSE-t, 2-sided	normal	0.244	0.279	0.281	0.284	0.336	0.299	0.302

Notes:
1. See the notes to Table 1.
2. In these experiments, the coefficients b_{ij} are set to values (given in section 3.1) large enough that the alternative model is expected to be more accurate than the null model.

Table 5: Tests of Equal Accuracy for Monthly Stock Returns

		MSE-F Bootstrap p–values		
alternative model variable	*MSE(null)/ MSE(altern.)*	*non-param.*	*no predictability fixed regressor*	*fixed regressor*
cross-sectional premium	1.009	0.136	0.001	0.071
return on long-term Treasury	1.005	0.381	0.024	0.177
BAA-AAA yield spread	0.996	0.688	0.828	0.487
BAA-AAA return spread	0.995	0.824	0.933	0.779
net equity expansion	0.994	0.648	0.358	0.659
CPI inflation	0.993	0.646	0.587	0.776
stock variance	0.992	0.773	0.512	0.230
dividend-payout ratio	0.991	0.681	0.572	0.724
term (yield) spread	0.987	0.724	0.939	0.984
earnings-price ratio	0.985	0.938	0.383	0.933
10-year earnings-price ratio	0.983	0.876	0.985	0.984
3-month T-bill rate	0.982	0.739	0.952	0.993
dividend-price ratio	0.981	0.843	0.550	0.993
dividend yield	0.981	0.836	0.436	0.996
yield on long-term Treasury	0.978	0.796	0.988	0.995
book-market ratio	0.965	0.996	0.967	0.994

Notes:
1. As described in section 4, monthly forecasts of excess stock returns in period $t+1$ are generated recursively from a null model that includes just a constant and 15 alternative models that include a constant and the period t ($t-1$ in the case of CPI inflation) value of each of the variables listed in the first column. Forecasts from January 1970 to December 2002 are obtained from models estimated with a data sample starting in January 1954.
2. For each alternative model, the table reports the ratio of the null model's forecast MSE to the alternative model's MSE and bootstrapped p-values for the null hypothesis of equal accuracy, based on the MSE-F statistic. Section 2.6 details the bootstrap methods. The RMSE of the null model is 0.046.

Table 6: Tests of Equal Accuracy for Core Inflation

		MSE-F Bootstrap p–values		
alternative model variables	*MSE(null)/ MSE(altern.)*	*non-param.*	*no predictability fixed regressor*	*fixed regressor*
1-quarter horizon				
CFNAI	1.016	0.343	0.092	0.293
CFNAI, food, imports	1.098	0.100	0.001	0.062
4-quarter horizon				
CFNAI	0.921	0.675	0.881	0.915
CFNAI, food, imports	1.279	0.317	0.000	0.031

Notes:
1. As described in section 4, 1-quarter and 4-quarter ahead forecasts of core PCE inflation (specified as a period $t + \tau$ predictand) are generated recursively from a null model that includes a constant and lags of inflation (from period t and earlier) and alternative models that include one lag (period t values) of the variables indicated in the table (defined further in section 4). The 1-quarter forecasts are of quarterly inflation; the 4-quarter forecasts are of 4-quarter inflation. Forecasts from 1985:Q1 + $\tau - 1$ through 2008:Q2 are obtained from models estimated with a data sample starting in 1968:Q3.
2. For each of the alternative models, the table reports the ratio of the null model's forecast MSE to the alternative model's MSE and bootstrapped p-values for the null hypothesis of equal accuracy, based on the MSE-F statistic. Section 2.6 details the bootstrap methods. The RMSE of the null model is 0.613 at the 1-quarter horizon and 0.444 at the 4-quarter horizon.