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

Todd E. Clark
Federal Reserve Bank of Kansas City

Michael W. McCracken
Federal Reserve Bank of St. Louis

March 2009

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 pairwise 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 (corresponding author)*: Economic Research Dept.; Federal Reserve Bank of Kansas City; 1 Memorial Drive; Kansas City, MO 64198; `todd.e.clark@kc.frb.org`. *McCracken*: Research Division; Federal Reserve Bank of St. Louis; P.O. Box 442; St. Louis, MO 63166; `michael.w.mccracken@stls.frb.org`. This paper previously circulated with a title of "Reality Checks and Nested Forecast Model Comparisons." We gratefully acknowledge helpful comments from Bruce Hansen, Lutz Kilian, Norm Swanson, and seminar participants at the University of Michigan, Rutgers University, and the Midwest Econometrics Group. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Banks of Kansas City or St. Louis.

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 Clark and McCracken (2001, 2005) 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 (2006) and Clark and McCracken (2008), 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 evaluate various bootstrap-based methods 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 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 predictive ability, we are interested in using out-of-sample tests to test the null hypothesis of equal finite sample predictive ability. This seemingly minor 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.

Although our results apply only to a setup that some might see as restrictive — direct, multi–step (DMS) forecasts from nested models — the list of studies analyzing such forecasts suggests our results 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); 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); and Molodtsova and Papell (2008).

The remainder proceeds as follows. Section 2 introduces the notation and assumptions and presents our theoretical results. Section 3 characterizes the bootstrap-based methods we consider for testing the joint hypothesis of equal forecast accuracy. Section 4 presents Monte Carlo results on the finite–sample performance of the asymptotics and the bootstrap. Section 5 applies our tests to evaluate the predictability of U.S. stock returns and core PCE inflation. Section 6 concludes.

2 Theoretical results

We begin by laying out our testing framework when comparing the forecast accuracy of two nested models in the presence of weak predictive ability.

2.1 Environment

The possibility of weak predictors is modeled using a sequence of linear DGPs of the form (**Assumption 1**)[1]

$$\begin{aligned} y_{T,t+\tau} &= x'_{T,1,t}\beta^*_{1,T} + u_{T,t+\tau} = x'_{T,0,t}\beta^*_0 + x'_{T,12,t}(T^{-1/2}\beta^*_{12}) + u_{T,t+\tau}, \qquad (1) \\ Ex_{T,1,t}u_{T,t+\tau} &\equiv Eh_{T,1,t+\tau} = 0 \text{ for all } t = 1, ..., T, ...T + P - \tau. \end{aligned}$$

Note that we allow the dependent variable $y_{T,t+\tau}$, the predictors $x_{T,1,t}$ and the error term $u_{T,t+\tau}$ to depend upon T, the initial forecasting origin. This dependence is necessitated by the triangular array structure of the data. However, throughout much of the paper we omit the additional subscript T for ease of presentation.

At each origin of forecasting $t = T, ...T + P - \tau$, we observe the sequence $\{y_{T,s}, x'_{T,1,s}\}_{s=1}^t$. Forecasts of the scalar $y_{T,t+\tau}$, $\tau \geq 1$, are generated using a $(k \times 1, k = k_0 + k_1)$ vector of covariates $x_{T,1,t} = (x'_{T,0,t}, x'_{T,12,t})'$, and linear parametric models $x'_{T,i,t}\beta_i$, $i = 0, 1$. The parameters are estimated using OLS (**Assumption 2**) under either the recursive or rolling schemes. For the recursive scheme we have $\hat{\beta}_{i,t} = \arg\min_{\beta_i} t^{-1}\sum_{s=1}^{t-\tau}(y_{T,s+\tau} - x'_{T,i,s}\beta_i)^2$, $i = 0, 1.$ for the restricted and unrestricted, respectively. The rolling scheme is similar but the number of observations used for estimation is held constant as we proceed forward across forecast origins and hence $\hat{\beta}_{i,t} = \arg\min_{\beta_i} T^{-1}\sum_{s=t-\tau-T+1}^{t-\tau}(y_{T,s+\tau} - x'_{T,i,s}\beta_i)^2$, $i = 0, 1.$ We denote the loss associated with the τ-step ahead forecast errors as $\hat{u}^2_{i,t+\tau} = (y_{T,t+\tau} - x'_{T,i,t}\hat{\beta}_{i,t})^2$, $i = 0, 1$, for the restricted and unrestricted, respectively.

The following additional notation will be used. For the recursive scheme let $H_{T,i}(t) = (t^{-1}\sum_{s=1}^{t-\tau} x_{T,i,s}u_{T,s+\tau}) = (t^{-1}\sum_{s=1}^{t-\tau} h_{T,i,s+\tau})$ and $B_i(t) = (t^{-1}\sum_{s=1}^{t-\tau} x_{T,i,s}x'_{T,i,s})^{-1}$, and for the rolling case let $H_{T,i}(t) = (T^{-1}\sum_{s=t-\tau-T+1}^{t-\tau} x_{T,i,s}u_{T,s+\tau}) = (T^{-1}\sum_{s=t-\tau-T+1}^{t-\tau} h_{T,i,s+\tau})$ and $B_i(t) = (T^{-1}\sum_{s=t-\tau-T+1}^{t-\tau} x_{T,i,s}x'_{T,i,s})^{-1}$. In either case, define, for $i = 0, 1$, $B_i = \lim_{T\to\infty}(Ex_{T,i,s}x'_{T,i,s})^{-1}$. For $U_{T,t} = (h'_{T,1,t+\tau}, vec(x_{T,1,t}x'_{T,1,t})')'$, $V = \sum_{j=-\tau+1}^{\tau-1}\Omega_{11,j}$, where $\Omega_{11,j}$ is the upper block-diagonal element of Ω_j defined below. For any $(m \times n)$ matrix A let $|A|$ denote the max norm and $tr(A)$ denote the trace. For $H_{T,1}(t)$ defined above, J the selection matrix $(I_{k_0\times k_0}, 0_{k_0\times k_1})'$, $\sigma^2 = \lim_{T\to\infty} Eu^2_{T,t+\tau}$, and a $(k_1 \times k)$ matrix $\tilde{A}$ satisfying $\tilde{A}'\tilde{A} = B_1^{-1/2}(-J'B_0J + B_1)B_1^{-1/2}$, let $\tilde{h}_{T,1,t+\tau} = \sigma^{-1}\tilde{A}B_1^{1/2}h_{T,1,t+\tau}$ and $\tilde{H}_{T,1}(t) = \sigma^{-1}\tilde{A}B_1^{1/2}H_{T,1}(t)$. For the selection matrix $J_2 = (0_{k_1\times k_0}, I_{k_1\times k_1})'$ define $F_1 = J_2'B_1J_2$ and $F_1(t) = J_2'B_1(t)J_2$. If we define $\gamma_{\tilde{h}\tilde{h},1}(i) = \lim_{T\to\infty} E\tilde{h}_{T,1,t+\tau}\tilde{h}'_{T,1,t+\tau-i}$,

[1] The parameter $\beta^*_{1,T}$ does not vary with the forecast horizon τ since, in our analysis, τ is treated as fixed.

then $S_{\tilde{h}\tilde{h},1} = \gamma_{\tilde{h}\tilde{h},1}(0) + \sum_{i=1}^{\tau-1}(\gamma_{\tilde{h}\tilde{h},1}(i) + \gamma'_{\tilde{h}\tilde{h},1}(i))$. Let $W(s)$ denote a $k_1 \times 1$ vector standard Brownian motion and define the vector of weak predictor coefficients as $\delta = (0_{1\times k_0}, \beta^{*\prime}_{12})'$.

To derive our general results, we need three more assumptions (in addition to our assumptions (1 and 2) of a DGP with weak predictability and OLS–estimated linear forecasting models).

Assumption 3: (a) $T^{-1}\sum_{t=1}^{[rT]} U_{T,t}U'_{T,t-j} \Rightarrow r\Omega_j$ where $\Omega_j = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E(U_{T,t}U'_{T,t-j})$ for all $j \geq 0$, (b) $\Omega_{11,j} = 0$ all $j \geq \tau$, (c) $\sup_{T\geq 1, t\leq T+P} E|U_{T,t}|^{2q} < \infty$ some $q > 2$, (d) The zero mean triangular 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 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 L and constant $i \in (0, 0.5)$, $L = O(P^i)$, (c) For all $j > \tau - 1$, $Eh_{T,1,t+\tau}h'_{T,1,t+\tau-j} = 0$, (d) The number of covariance terms $\bar{j}$, used to estimate the long–run covariance S_{dd} defined in Section 2.2, satisfies $\tau - 1 \leq \bar{j} < \infty$.

Assumption 5: $\lim_{P,T\to\infty} P/T = \lambda_P \in (0, \infty)$.

Assumption 3 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 τ-step ahead forecast errors are MA($\tau - 1$). 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, $h_{T,1,t+\tau}$ 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 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.[2] We therefore impose conditions sufficient to cover applied practices. Parts (a) and (b) are not particularly controversial. Part (c), however, imposes the restriction that the orthogonality conditions used to identify the parameters form a moving average of finite order $\tau - 1$, while part (d) imposes the restriction that this fact is taken into account when constructing the MSE-t statistic discussed later in Section 2. Finally, in Assumption 5 we impose the requirement that $\lim_{P,T\to\infty} P/T = \lambda_P \in (0, \infty)$. This assumption implies that the duration of forecasting is finite but non-trivial.

This last 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.

2.2 Asymptotics for MSE-F, MSE-t with weak predictors

In the context of non-nested models, Diebold and Mariano (1995) propose a test for equal MSE based upon the sequence of loss differentials $\hat{d}_{t+\tau} = \hat{u}^2_{0,t+\tau} - \hat{u}^2_{1,t+\tau}$. If we define $MSE_i = (P - \tau + 1)^{-1} \sum_{t=T}^{T+P-\tau} \hat{u}^2_{i,t+\tau}$ $(i = 0, 1)$, $\bar{d} = (P - \tau + 1)^{-1} \sum_{t=T}^{T+P-\tau} \hat{d}_{t+\tau} = MSE_0 - MSE_1$, $\widehat{\gamma}_{dd}(j) = (P - \tau + 1)^{-1} \sum_{t=T+j}^{T+P-\tau} (\hat{d}_{t+\tau} - \bar{d})(\hat{d}_{t+\tau-j} - \bar{d})$, $\widehat{\gamma}_{dd}(-j) = \widehat{\gamma}_{dd}(j)$, and $\hat{S}_{dd} = \sum_{j=-\bar{j}}^{\bar{j}} K(j/M)\widehat{\gamma}_{dd}(j)$, the statistic takes the form

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

Under the null that $x_{12,t}$ has no population level predictive power for $y_{t+\tau}$, the population difference in MSEs, $Eu^2_{0,t+\tau} - Eu^2_{1,t+\tau}$, will equal 0 for all t. When $x_{12,t}$ has predictive power, the population difference in MSEs will be positive. Even so, the finite sample difference

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

need not be positive and in fact, for a given sample size (say, $t = T$) the difference in finite sample MSEs, $E\hat{u}^2_{0,T+\tau} - E\hat{u}^2_{1,T+\tau}$, may be zero, thus motivating a distinct null hypothesis of equal finite-sample predictive ability. Regardless of which null hypothesis we consider (equal population 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 $\hat{d}_{t+\tau}$ be 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 $d_{t+\tau} = 0$ for all t, which makes the corresponding variances also equal to 0. Because the sample analogues (for example, $\bar{d}$ 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 $d_{t+\tau}$ 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

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

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 parameter π, the number of exclusion restrictions k_1, and the unknown nuisance parameter $S_{\tilde{h}\tilde{h}}$. While this continues to hold in the presence of weak predictors, the asymptotic distributions now depend not only upon the unknown coefficients associated with the weak predictors but also upon other unknown second moments of the data. In the following, for the recursive scheme define $\Gamma_1 = \int_1^{1+\lambda_P} s^{-1}W'(s)S_{\tilde{h}\tilde{h}}dW(s)$, $\Gamma_2 = \int_1^{1+\lambda_P} s^{-2}W'(s)S_{\tilde{h}\tilde{h}}W(s)ds$, $\Gamma_5 = \int_1^{1+\lambda_P} s^{-2}W'(s)S^2_{\tilde{h}\tilde{h}}W(s)ds$, $\Gamma_6 = \int_1^{1+\lambda_P} s^{-1}(\delta' B_1^{-1/2}\tilde{A}'/\sigma)S^{3/2}_{\tilde{h}\tilde{h}}W(s)ds$. For the rolling scheme, define $\Gamma_1 = \int_1^{1+\lambda_P}(W(s)-W(s-1))'S_{\tilde{h}\tilde{h}}dW(s)$, $\Gamma_2 = \int_1^{1+\lambda_P}(W(s)-W(s-1))'S_{\tilde{h}\tilde{h}}(W(s)-W(s-1))ds$, $\Gamma_5 = \int_1^{1+\lambda_P}(W(s)-W(s-1))'S^2_{\tilde{h}\tilde{h}}(W(s)-W(s-1))ds$, and $\Gamma_6 = \int_1^{1+\lambda_P} s^{-1}(\delta' B_1^{-1/2}\tilde{A}'/\sigma)S_{\tilde{h}\tilde{h}}^{3/2}(W(s) - W(s-1))ds$. For both schemes define $\Gamma_3 = \int_1^{1+\lambda_P}(\delta' B_1^{-1/2}\tilde{A}'/\sigma)S_{\tilde{h}\tilde{h}}^{1/2}dW(s)$, $\Gamma_4 = \int_1^{1+\lambda_P}\delta' J_2 F_1^{-1}J_2'\delta/\sigma^2 ds = \lambda_P \delta' J_2 F_1^{-1} J_2'\delta/\sigma^2$ 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, 2, 3, and 5. MSE-$F \to_d \{2\Gamma_1 - \Gamma_2\} + 2\{\Gamma_3\} + \{\Gamma_4\}$.

Theorem 2.2: Maintain Assumptions 1−5. 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 use bootstrap methods to estimate the asymptotically valid critical values. Before describing our bootstrap approach, however, it is necessary to clarify the null hypothesis of interest.

2.3 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 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 $\beta_{12}^* = 0$. 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.[3] For example, under the recursive scheme we have

$$E\sum\nolimits_{t=T}^{T+P}(\hat{u}_{0,t+\tau}^2-\hat{u}_{1,t+\tau}^2)\approx\int_1^{1+\lambda_P}[-s^{-1}tr((-JB_0J'+B_1)V)+\delta' J_2F_1^{-1}J_2'\delta]ds$$

while under the rolling scheme we have

$$E\sum\nolimits_{t=T}^{T+P}(\hat{u}_{0,t+\tau}^2-\hat{u}_{1,t+\tau}^2)\approx\int_1^{1+\lambda_P}[-tr((-JB_0J'+B_1)V)+\delta' J_2F_1^{-1}J_2'\delta]ds.$$

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 $\frac{\delta' J_2F_1^{-1}J_2'\delta}{tr((-JB_0J'+B_1)V)}$ equals $\frac{\ln(1+\lambda_P)}{\lambda_P}$ and 1, respectively, for the recursive and rolling schemes. This ratio simplifies further when $\tau=1$ and the forecast errors are conditionally homoskedastic in which case $tr((-JB_0J'+B_1)V)=\sigma^2k_1$.

This marginal signal-to-noise ratio 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 $\beta_{12}^*=0$ – we test for equal average out-of-sample predictive ability $H_0: E(P^{-1}\sum_{t=T}^{T+P}(\hat{u}_{0,t+\tau}^2-\hat{u}_{1,t+\tau}^2))=0$ – for which it is approximately the case that $\beta_{12}^{*\prime}F_1^{-1}\beta_{12}^*=d$ where d equals $\frac{\ln(1+\lambda_P)}{\lambda_P}tr((-JB_0J'+B_1)V)$ or $tr((-JB_0J'+B_1)V)$, depending on whether the recursive or rolling scheme is used.

While we believe the result is intuitive, it is not immediately clear how such a restriction 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_{12}^*, 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. Rather than take either of these approaches, 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.

[3]By taking this approach we are using the fact that under our assumptions, notably the L^2-boundedness portion of Assumption 3, $\sum_{t=T}^{T+P}(\hat{u}_{0,t+\tau}^2-\hat{u}_{1,t+\tau}^2)$ is uniformly integrable and hence the expectation of its limit is equal to the limit of its expectation.

2.4 Bootstrap-based critical values with weak predictors

Our new, bootstrap-based method of approximating the asymptotically valid critical values for pairwise comparisons between nested models is different from that previously used in Kilian (1999) and Clark and McCracken (2005). In those applications, an appropriately dimensioned VAR was initially estimated by OLS imposing the restriction that β_{12}^* 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 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_{12}^* 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 generate the artificial samples.

For example, suppose we are interested in testing whether, under the recursive scheme, the two models have equal average predictive ability over the out-of-sample period and hence $\delta' J_2 F_1^{-1} J_2' \delta$ equals $\frac{\ln(1+\lambda_P)}{\lambda_P} tr((-JB_0 J' + B_1)V)$. 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 $\beta_{12}^{*\prime}\beta_{12}^* = c$ for some finite constant — as one would in a ridge regression — we instead impose the restriction that $\delta' J_2 F_1^{-1}(T) J_2' \delta$ 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: $\widehat{\lambda}_P = P/T, B_i(T) = (T^{-1}\sum_{s=1}^{T-\tau} x_{i,s}x_{i,s}')^{-1}$, $i = 0, 1$, $F_1(T) = J_2' B_1(T) J_2$, and $V(T)$ denotes an estimate of the long-run variance of $h_{1,t+\tau}$.[4] In addition, we estimate δ using the approximation $\widehat{\delta} = (0_{1\times k_0}, T^{1/2}\widetilde{\beta}_{12,t}')'$ where $\widetilde{\beta}_{12,T}$ denotes the restricted least squares

[4] 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.

estimator of the parameters associated with the weak predictors satisfying

$$\begin{aligned} \widetilde{\beta}_{1,T} &= (\widetilde{\beta}'_{11,T}, \widetilde{\beta}'_{12,T})' \\ &= \arg\min_{b_1} \sum\nolimits_{s=1}^{T-\tau} (y_{s+\tau} - x'_{1,s} b_1)^2 \text{ s.t. } b'_1 J_2 F_1^{-1}(T) J'_2 b_1 = \widehat{d}/T \end{aligned} \quad (4)$$

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 fixed regressor approach as discussed in Hansen (2000). In this 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.[5]

1. Estimate the parameter vector β^*_1 associated with the unrestricted model using the weighted ridge regression from equation (4) above. Note that the resulting parameter estimate will vary with the forecast horizon. If the recursive scheme is used, set $\widehat{d}$ to $\frac{\ln(1+\widehat{\lambda}_P)}{\widehat{\lambda}_P} tr((-JB_0(T)J' + B_1(T))V(T))$; if the rolling scheme is used, set $\widehat{d}$ to $tr((-JB_0(T)J' + B_1(T))V(T))$.

2. Using NLLS, estimate an $MA(\tau-1)$ model for the OLS residuals $\widehat{v}_{1,s+\tau}$ such that $v_{1,s+\tau} = \varepsilon_{1,s+\tau} + \theta_1\varepsilon_{1,s+\tau-1} + ... + \theta_{\tau-1}\varepsilon_{1,s+1}$. Let $\eta_{s+\tau}$, $s = 1, ..., T+P-\tau$, denote an $i.i.d$ $N(0,1)$ sequence of simulated random variables. 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 an estimate of the test statistics (e.g. MSE-F, MSE-t) as if this were the original data.

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

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

5. Reject the null hypothesis, at the $\alpha\%$ 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 $(T^{1/2}\widetilde{\beta}_{12,T})$ of the local-to-zero parameters β_{12}^* satisfy our approximation to the null hypothesis. This is despite the fact that the estimates of β_{12}^* 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_{12}^* 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_{12}^* 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 $\lambda_P \beta_{12}'^* J_2' V J_2 \beta_{12}^*$ that in general, need not have any relationship to the restriction $\beta_{12}'^* F_1^{-1} \beta_{12}^* = d$ implied by the null hypothesis. Hence, in general, the asymptotic distribution is an explicit function of the value of β_{12}^* implying that the null hypothesis itself does not imply a unique asymptotic distribution for either the MSE-F or MSE-t statistics.

Even so, as we discuss below, the bootstrap is asymptotically valid in two empirically relevant special cases. Before providing the result, however, we require a modest strengthening of the moment conditions on the model residuals.

<u>Assumption 3$'$</u>: (a) $T^{-1}\sum_{j=1}^{[rT]} U_{T,j}U_{T,j-l}' \Rightarrow r\Omega_l$ where $\Omega_l = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E(U_{T,j}U_{T,j-l}')$ for all $l \geq 0$, (b) $\mathrm{E}(\varepsilon_{1,s+\tau}|\varepsilon_{1,s+\tau-j}, x_{1,s-j}\ \ j \geq 0) = 0$, (c) Let $\gamma_T = (\beta_{2,T}', \theta_1, ..., \theta_{\tau-1})'$, $\widehat{\gamma}_T = (\hat{\beta}_{2,T}', \widehat{\theta}_1, ..., \widehat{\theta}_{\tau-1})'$, and define the function $\widehat{\varepsilon}_{2,s+\tau} = \widehat{\varepsilon}_{2,s+\tau}(\widehat{\gamma}_T)$ 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) $U_{T,j} - EU_{T,j} = (h_{T,1,j+\tau}', vec(x_{T,1,j}x_{T,1,j}' - Ex_{T,1,j}x_{T,1,j}')')'$ is a zero mean triangular array satisfying Theorem 3.2 of de Jong and Davidson (2000).

Assumption 3$'$ differs from Assumption 3 in two ways. First, in (b) it emphasizes the point that the forecast errors, and by implication $h_{1,t+\tau}$, form an $MA(\tau - 1)$. 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 3) but also the functions $\widehat{\varepsilon}_{1,s+\tau}(\gamma)x_{T,1,s}$, and $\nabla\widehat{\varepsilon}_{1,s+\tau}(\gamma)x_{T,1,s}$ 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 $\hat{\beta}_{i,T}$ and Assumption 3 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. 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 Theorems 2.1 and 2.2.

Theorem 2.3: Let $\beta_{12}'^{*} F_1^{-1} \beta_{12}^* = d$ and assume either (i) $\tau = 1$ and the forecast errors from the unrestricted model are conditionally homoskedastic, or (ii) $\dim(\beta_{12}^*) = 1$. (a) Given Assumptions $1, 2, 3'$, and 5, MSE-$F^* \to_d \{2\Gamma_1^* - \Gamma_2^*\} + 2\{\Gamma_3^*\} + \{\Gamma_4^*\}$. (b) Given Assumptions $1, 2, 3', 4$, and 5, 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, requires additional assumptions. In the first, we require that the forecast errors be one-step ahead and conditionally homoskedastic. In the latter we allow serial correlation and conditional heteroskedasticity but require that β_{12}^* is scalar. While neither case covers the broadest situation in which β_{12}^* 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 (Goyal and Welch (2008), Stock and Watson (2003), etc.), the unrestricted forecasts are made by simply adding one lag of a single predictor to the baseline restricted model.

By itself, however, 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 $\beta_{12}'^{*}F_1^{-1}\beta_{12}^{*} > d$. For example, rather than modeling the weak predictive ability in Assumption 1 as $T^{-1/2}\beta_{12}^{*}$ with $\beta_{12}'^{*}F_1^{-1}\beta_{12}^{*} = d$, one could model the predictive content as $T^{-a}C\beta_{12}^{*}$ for constants $C < \infty$ and $a \in (0, 1/2]$ satisfying $\beta_{12}'^{*}F_1^{-1}\beta_{12}^{*} > d$. 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 $J_2'\hat{\beta}_{1,T}$ is also a consistent estimator of $\beta_{12}^{*} \neq 0$. For this situation to hold we need the additional restriction that $a = 0$ and hence β_{12}^{*} is no longer interpretable as a local-to-zero parameter. With this modification (**Assumption 1$'$**) in hand, we address the validity of the bootstrap under the alternative in the following Proposition.

Theorem 2.4: Let $J_2'\hat{\beta}_{1,T} \to^p \beta_{12}^{*} \neq 0$ and assume either (i) $\tau = 1$ and the forecast errors from the unrestricted model are conditionally homoskedastic, or (ii) $\dim(\beta_{12}) = 1$. (a) Given Assumptions $1', 2, 3'$, and 5, MSE-$F^{*} \to_d \{2\Gamma_1^{*} - \Gamma_2^{*}\} + 2\{\Gamma_3^{*}\} + \{\Gamma_4^{*}\}$. (b) Given Assumptions $1', 2, 3', 4$, and 5, 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 $\beta_{12}'^{*}F_1^{-1}\beta_{12}^{*} = d$ against the alternative that the unrestricted model is more accurate (that $J_2'\hat{\beta}_{1,T} \to^p \beta_{12}^{*} \neq 0$). 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.

3 Bootstrap approaches

Drawing on the proceeding theoretical results, we use non–parametric and parametric bootstrap procedures and a fixed regressor bootstrap in testing for equal forecast accuracy, based on the above MSE-F and MSE-t tests.

3.1 Non-parametric bootstrap

Our non–parametric approach is 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 implementations, we follow the approach of White (2000) in centering the bootstrap distributions. 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. We report empirical rejection rates using a nominal size of 10%. Results using a nominal size of 5% are qualitatively similar.

3.2 Restricted VAR bootstrap

Our parametric bootstrap procedure broadly follows those of Kilian (1999) and Clark and McCracken (2005), among others. Vector autoregressive equations for y_t and x_t are estimated using the full sample of observations, with the residuals stored for sampling. Bootstrapped time series on y_t and x_t are generated by drawing with replacement from the sample residuals and using the autoregressive structures of the estimated VAR to iteratively construct data. The initial observations — observations preceding the sample of data used to estimate the models — necessitated by the lag structures of the estimated models, are selected by sampling from the actual data. In particular, following Stine (1987), among others, the initial observations are selected by picking one date at random and then taking the necessary number of initial observations in order from that date backward. For each sample of artificial data, we estimate forecasting models and generate forecasts and test statistics.

The VAR used in the bootstrap is restricted to impose the null that the x variables of interest have no predictive content — that is, coefficients of zero. The bootstrap DGP equation for y in the VAR takes the form of the null or benchmark forecasting model. For simplicity, the VAR equations for the x variables are specified as AR models. This basic approach is used in such studies as Mark (1995), Kilian (1999), Clark and McCracken (2005, 2006), and Clark and West (2006, 2007).

We use the restricted VAR bootstrap to construct critical values for tests of equal forecast accuracy based on the MSE-F and MSE-t tests. For all tests, because the null hypothesis of $\beta_{12}^* = 0$ is imposed in the data generation process, no adjustment of the sample test statistics is needed for inference. We simply compare the sample test statistics against the bootstrap draws, without any re-centering.

For comparison, we also use the restricted bootstrap to generate critical values for the adjusted $t-$test of equal MSE developed in Clark and West (2006, 2007). In the interest of obtaining a normally-distributed or nearly-normal test of equal MSE, Clark and West propose a simple adjustment to the MSE differential to account for the additional parameter estimation error of the larger model. When applied to a pair of rolling sample forecasts under a random walk null model, the adjusted test statistic has a standard normal distribution (asymptotically). With a null model that involves parameter estimation (as is the case in this paper), Clark and West (2007) argue that the limiting null distribution is approximately normal. Note, however, that in either case, the null hypothesis is that the smaller model is true, not that the null and alternative forecasts are equally accurate over the sample of interest.

3.3 Fixed regressor bootstrap

As outlined in section 2.4, we also consider a fixed regressor bootstrap under the null of equal forecast accuracy. Under this procedure, we re-estimate the alternative forecasting model subject to the constraint that implies the null and alternative model forecasts to be equally accurate. We take the residuals ($\widehat{v}_{t+\tau}$) and fitted values ($x_{1,t}'\hat{\beta}_{1,T}$) from this model. Following the algorithm outlined in section 2.4, we create artificial replicas of the residuals $\widehat{v}_{t+\tau}^*$ and add them to the fitted value to form artificial samples of y_{t+1}^*: $y_{t+1}^* = x_{1,t}'\hat{\beta}_{1,T} + \widehat{v}_{t+\tau}^*$. Using the artificial samples of data on y, we estimate the forecasting models (using actual data on all the variables on the right-hand side, rather than simulated data), generate samples of forecasts and forecast errors, and finally compute samples of test statistics. In particular, we use the fixed regressor bootstrap to construct critical values for the MSE-F and MSE-t tests. We compare the sample test statistics against the bootstrap draws, without any re-centering.

4 Monte Carlo Evidence

We use simulations of bivariate and multivariate DGPs 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 an autoregressive model of the predictand y; the alternative models add lags of various other variables of interest. The general null hypothesis is that the forecast from the alternative models is no more accurate than the benchmark forecast. This general null, however, can take different specific forms: either the variables in the alternative model have no predictive content, in that their coefficients are 0; or the variables have non–zero coefficients, but the coefficients are small enough that the benchmark and alternative models are expected to be equally more accurate over the forecast sample. We focus our presentation on recursive forecasts, but include some results for rolling forecasts.

4.1 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. The initial observations necessitated by the lag structure of each DGP are generated with draws from the unconditional normal distribution implied by the DGP. While our results can be generalized to any forecast horizon (with models of the direct multi–step form), for brevity we focus on one–step ahead forecasts. With quarterly data in mind, we also consider a range of sample sizes (R, P), reflecting those commonly available in practice: 80,40; 40,80; 80,80; and 80,120.

All of the DGPs are based on empirical relationships among U.S. inflation and a range of predictors, estimated with 1968-2007 data. In all cases, our reported results are based on 2000 Monte Carlo draws and 499 bootstrap replications.

4.1.1 DGPs

DGP 1 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):

$$\begin{aligned} y_t &= -0.4 y_{t-1} - 0.1 y_{t-2} + b_{11} x_{1,t-1} + u_t \\ x_{1,t} &= 0.8 x_{1,t-1} - 0.1 x_{1,t-2} + v_{1,t} \\ &\quad \text{var}\begin{pmatrix} u_t \\ v_{1,t} \end{pmatrix} = \begin{pmatrix} 0.8 & \\ 0.0 & 0.4 \end{pmatrix}. \end{aligned} \tag{5}$$

In the DGP 1 experiments, the alternative (unrestricted) forecasting model takes the form of the DGP equation for y (with constant added); the null or benchmark (restricted) model drops $x_{1,t-1}$:

$$\text{null:} \quad y_t = \beta_0 + \beta_1 y_{t-1} + \beta_1 y_{t-2} + u_{0,t}. \tag{6}$$

$$\text{alternative:} \quad y_t = \beta_0 + \beta_1 y_{t-1} + \beta_1 y_{t-2} + \beta_3 x_{1,t-1} + u_{1,t}. \tag{7}$$

We consider various experiments with different settings of b_{11}, the coefficient on $x_{1,t-1}$, which corresponds to the elements of our theoretical construct $\beta_{12}^{*}/\sqrt{T}$. In one set of simulations (Table 1), the coefficient is set to 0, such that the null forecasting model is expected to be more accurate than the alternative. In others (Tables 2 and 3), the coefficient is set to a value that makes the models equally accurate (in expectation) on average over the forecast sample. For example, with recursive forecasts and R and P both equal to 80 (this coefficient value changes with R and P), this value is 0.11, about 1/2 of the empirical estimate. In another set of experiments (Table 4), the coefficient is set to 0.3, such that the alternative model is expected to be more accurate than the null.

DGP 2 is based on the empirical relationship among the change in core PCE inflation (y_t), the CFNAI ($x_{1,t}$), growth in unit labor costs less core inflation ($x_{2,t}$), PCE food price inflation less core inflation ($x_{3,t}$), and PCE energy price inflation less core inflation ($x_{4,t}$). To simplify the lag structure necessary for reasonable forecasting models, the growth or inflation rates used in forming variables $x_{2,t}$, $x_{3,t}$, and $x_{4,t}$ are computed as two-quarter averages. Based on these data, DGP 2 takes the form

$$
\begin{aligned}
y_t &= -0.40y_{t-1} - 0.1y_{t-2} + b_{11}x_{1,t-1} + b_{21}x_{2,t-1} + b_{31}x_{3,t-1} + b_{41}x_{4,t-1} + u_t \\
x_{1,t} &= 0.8x_{1,t-1} - 0.1x_{1,t-2} + v_{1,t} \\
x_{2,t} &= 0.7x_{2,t-1} - 0.3x_{2,t-2} + v_{2,t} \\
x_{3,t} &= 0.9x_{3,t-1} - 0.2x_{3,t-2} + v_{3,t} \\
x_{4,t} &= 0.8x_{4,t-1} - 0.3x_{4,t-2} + v_{4,t}
\end{aligned}
\tag{8}
$$

$$
\mathrm{var}\begin{pmatrix} u_t \\ v_{1,t} \\ v_{2,t} \\ v_{3,t} \\ v_{4,t} \end{pmatrix} = \begin{pmatrix} 0.8 & & & & \\ 0.0 & 0.4 & & & \\ 0.2 & -0.1 & 3.8 & & \\ -0.1 & 0.0 & 0.3 & 2.2 & \\ 0.3 & -0.6 & 1.5 & 0.9 & 72.6 \end{pmatrix}.
$$

In DGP 2 experiments, the null (restricted) and alternative (unrestricted) forecasting models take the following forms, respectively:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_1 y_{t-2} + u_{0,t}. \tag{9}$$

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_1 y_{t-2} + \beta_3 x_{1,t-1} + \beta_4 x_{2,t-1} + \beta_5 x_{3,t-1} + \beta_6 x_{4,t-1} + u_{1,t}. \tag{10}$$

As with DGP 1, we consider experiments with three different settings of the set of b_{ij} coefficients, which correspond to the elements of $\beta_{12}^*/\sqrt{T}$. In one set of experiments (Table 1), all of the b_{ij} coefficients are set to zero, such that the null forecasting model is expected to be more accurate than the alternative. In others (Tables 2 and 3), empirically–based 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. With R and P at 80, this multiplying constant is 0.355. In another set of experiments (Table 4), the coefficients are set at empirically–based estimates: $b_{11} = 0.3$, $b_{12} = 0.1$, $b_{13} = 0.1$, and $b_{14} = .015$. With these values, the alternative model is expected to be more accurate than the null.

4.2 Results

Our interest lays in identifying those bootstrap 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 our 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 2, using $R = P = 80$. We provide three densities, for the cases in which the b_{ij} coefficients of the DGP (8) 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 above 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 about 1.0. Among our bootstrap procedures, the restricted VAR approach yields, by design, a distribution like that shown for the null best DGP. The non–parametric and fixed regressor bootstraps are intended to estimate a null distribution like that shown for the equally good models DGP. In all cases, the null will be rejected when the sample MSE ratio lays in the right tail of the bootstrapped distribution.

What, then, might we expect test rejection rates to look like across experiments and bootstraps? For DGPs in which the null model is best, tests compared against the restricted VAR bootstrap should have rejection rates of about 10%, the nominal size (prior studies such as Kilian (1999) and Clark and McCracken (2005) have shown this bootstrap approach to work reasonably well in this type of experiment). However, the same tests compared against the other bootstraps should have rejection rates below 10%, because given the DGP, the models should not be expected to be equally accurate. For DGPs with coefficients scaled such that the null and alternative models can be expected to be equally accurate, we want the tests compared against the non–parametric and fixed regressor bootstraps to have size of about 10%. That said, as indicated above, we shouldn't expect the non–parametric approach to perform well when applied to recursive forecasts from our nested models (based on the asymptotics of Giacomini and White (2006), the non-parametric bootstrap may perform better for rolling forecasts). We should expect the same tests compared to restricted VAR bootstrap critical values to yield rejection rates greater than 10%, because the restricted VAR distribution lays to the left of the equal accuracy distribution. Finally, with DGPs that imply the alternative model to be more accurate than the null, we should look for rejection rates that exceed 10%. Again, though, rejection rates based on the restricted VAR bootstrap should generally be higher than rejection rates based on the other approaches.

4.2.1 Null Model Most Accurate

Table 1 presents Monte Carlo results for DGPs in which, in truth, the x variables considered have no predictive content for y, such that the null forecasting model should be expected to be best. These results generally line up with the expectations described above. Comparing the MSE-F, MSE-t and CW t-test statistics against restricted VAR bootstrap critical values consistently yields rejection rates of about the nominal size of 10%. For example, across all the experiments, restricted VAR bootstrap rejection rates for the MSE-F test range from 9.3% to 11.4%.

Comparing the test statistics to other bootstrap distributions yields rejection rates typically well below 10%, and often close to 0. Non-parametric bootstrap rejection rates for the MSE-F test range from 0% to 4.1%. Using the fixed regressor bootstrap yields results qualitatively similar to those for the non–parametric approach, with a range of 0.3 to 3.8%. Under any bootstrap approach, results are qualitatively very similar for the MSE-F and MSE-t tests.

4.2.2 Null and Alternative Models Equally Accurate

Table 2 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 results also generally line up with the expectations described above, and show clearly that, for testing the null of equal forecast accuracy, the most reliable bootstrap method is our proposed fixed regressor procedure.

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, rejection rates range from 8.1% to 11.4%.

Tests based on the other bootstrap intended to test the null of equal accuracy, the non–parametric bootstrap, are less reliable indicators of equal accuracy. For the MSE-F test, using the non-parametric bootstrap to compute critical values leads to modest undersizing. In Table 2, the MSE-F test's rejection rate ranges from 3.2 to 8.8% based on non-parametric critical values, compared to a range of 8.1 to 11.4% based on fixed regressor bootstrap critical values.

Tests based on the restricted VAR bootstrap may also be seen as unreliable indicators of equal forecast accuracy — but in that they overstate, rather than understate, the likelihood of two models being equally accurate. Comparing test statistics against critical values from the restricted VAR bootstrap generally yields rejection rates far in excess of 10%. In the case of the MSE-F test, rejection rates range from 25.3% to 46.6% (Table 2). Similarly, rejection rates for the C-W t-test range from 23.7% to 54.1%.

Among the alternative tests for equal MSE, results are, for the most part, qualitatively similar for the MSE-F and MSE-t test. For example, with the fixed regressor bootstrap, MSE-t rejection rates range from 7.4 to 10.3%, compared to the MSE-F range of 8.1% to 11.4%.

Table 3 provides results for experiments using a rolling forecast scheme instead of the baseline recursive scheme, for models parameterized to make the null and alternative models equally accurate (the necessary scaling factor is a bit different in the rolling case than the recursive). In general, the results for the rolling scheme are very similar to those for the recursive. Under both schemes, tests based on the restricted VAR reject too often, while tests based on our fixed regressor bootstrap have size of about 10% (the nominal size). Tests based on the non-parametric bootstrap continue to be undersized, although the problem is a bit worse under the rolling scheme than the recursive. For example, with DGP 1 and $R = P = 80$, comparing the MSE-F test against critical values estimated with the non-parametric bootstrap yields a rejection rate of 7.0% for recursive forecasts (Table 2) and 5.2% for rolling forecasts (Table 3).

Our rolling scheme results on the behavior of the MSE-t test compared against non-parametric bootstrap critical values are somewhat at odds with the behavior of the test in Giacomini and White (2006).[6] The difference seems to stem from treating the test as one-sided rather than two-sided. In unreported results for our DGPs, with non-parametric bootstrap critical values, a two-sided MSE-t test ranged from being slightly undersized to slightly oversized, in contrast to the consistent undersizing of the one-sided test.

[6]Giacomini and White (2006) compare the MSE-t test against standard normal critical values, rather than bootstrap critical values. In our experiments, comparing the MSE-t test against normal critical values yields results very similar to those we report for the non-parametric bootstrap.

4.2.3 Alternative Model Most Accurate

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

As anticipated, comparing the test statistics against critical values estimated with the restricted VAR bootstrap yields the highest rejection rate. In the case of the MSE-F test, rejection rates range from 72.6 to 99.1%. Comparing the test statistics against critical values estimated with the fixed regressor bootstrap yields modestly lower rejection rates. For the MSE-F test, rejection rates range from 58.4 to 96.1%. 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 29.9 and 79.2%.

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. Finally, as noted above in other experiment settings, the power of the C-W t-test is broadly comparable to that of the MSE-F test compared against restricted VAR critical values.

4.2.4 Results summary

Overall, the Monte Carlo results show that, for testing equal forecast accuracy over a given sample, our proposed fixed regressor bootstrap works reasonably well. When the null of equal accuracy 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 have reasonable power. The non–parametric bootstrap procedure, which just re–samples the data without imposing the equal accuracy null in the data generation, is not as reliable when applied to nested forecasting models. Finally, in line with prior research, for the purpose of testing the null that certain coefficients are 0, a restricted VAR bootstrap is reliable. However, the null of 0 coefficients is not the same as the null of equal forecast accuracy.

5 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.[7] 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 VAR bootstrap, we use a null model with just a constant for excess returns and AR(6) models for each of the predictive variables. In the fixed regressor procedure, the bootstrap equation for excess returns takes the form of the unrestricted forecasting model from each application (with the coefficients rescaled to imply equal forecast accuracy).

In the inflation application, we examine 1-quarter ahead and 1-year ahead forecasts of core PCE inflation obtained from a few models. 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 and import price inflation less core inflation.[8] 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 one-quarter ahead forecasts of inflation defined as $\pi_t = 400 \ln(P_t/P_{t-1})$, using 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

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

[8] 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.

price inflation. In another, we consider one-year 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. The forecasts are generated recursively. In the restricted VAR bootstrap, the DGP for inflation takes the same form as the null forecasting model, and we use AR(2) models for each of the predictive variables. In the fixed regressor procedure, the bootstrap equation for inflation takes the form of the unrestricted forecasting model from each of the two applications (with the coefficients rescaled to imply equal forecast accuracy).

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 include p-values for the null that the benchmark model is true (restricted VAR bootstrap) or that the models are equally accurate (the non–parametric and fixed regressor bootstraps). In the interest of brevity, results are only presented for the MSE-F test. We use 9999 replications in computing the bootstrap p-values.

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 restricted VAR 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 weaker evidence of forecastability, with the null of equal forecast accuracy rejected for the cross-sectional premium, but not the Treasury return. 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 restricted VAR 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 restricted VAR and fixed regressor bootstrap reject the null. 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 no rejections.

6 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. For tests of equal population-level predictive ability, we find that, as suggested in Inoue and

Kilian (2004), a restricted VAR bootstrap provides accurately sized tests. However, this does not continue to hold when we consider tests of equal finite-sample predictive ability in the presence of weak predictors. Instead, 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) does not work so 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. In contrast, using non-parametric bootstrap methods that are technically invalid with nested models — methods that have relatively poor size and power properties — yields no evidence of predictability.

7 Appendix: Theory Details

In the following, in addition to the notation from Section 2, define $h^*_{T,1,s+\tau} = x_{T,1,s}v^*_{1,s+\tau}$ and $\hat{h}^*_{T,1,s+\tau} = x_{T,1,s}\hat{v}^*_{1,s+\tau}$. 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 $\hat{H}^*_{T,1}(t) = T^{-1}\sum_{s=t-T-\tau+1}^{t-\tau} \hat{h}^*_{T,1,s+\tau}$ Moreover let $\sup_t |.|$ denote $\sup_{T\leq t\leq T+P-\tau} |.|$.

Proof of Theorem 2.1: (a) The result is a special case of Theorem 1 of Clark and McCracken (2008) and as a result, we provide only an outline of the proof here. The proof consists of two steps. In the first we provide an asymptotic expansion. In the second we apply a functional central limit theorem and a weak convergence to stochastic integrals result, both from de Jong and Davidson (2000). Throughout we ignore the finite sample difference between P and $P-\tau+1$.

For the first step, straightforward algebra reveals that

$$\begin{aligned}
&\sum\nolimits_{t=T}^{T+P-\tau}(\hat{u}^2_{0,t+\tau} - \hat{u}^2_{1,t+\tau}) \qquad (11)\\
=\;& \{2\sum\nolimits_{t=T}^{T+P-\tau}(T^{-1/2}h'_{T,1,t+\tau})(-JB_0(t)J' + B_1(t))(T^{1/2}H_{T,1}(t))\\
&-T^{-1}\sum\nolimits_{t=T}^{T+P-\tau}(T^{1/2}H'_{T,1}(t))(-JB_0(t)x_{T,0,t}x'_{T,0,t}B_0(t)J'\\
&+B_1(t)x_{T,1,t}x'_{T,1,t}B_1(t))(T^{1/2}H_{T,1}(t))\}\\
&+2\{\sum\nolimits_{t=T}^{T+P-\tau}\delta' B_1^{-1}(t)(-JB_0(t)J' + B_1(t))(T^{-1/2}h_{T,1,t+\tau})\}\\
&+\{T^{-1}\sum\nolimits_{t=T}^{T+P-\tau}\delta'(x_{T,1,t}x'_{T,1,t} - 2x_{T,1,t}x'_{T,1,t}JB_0(t)J'B_1^{-1}(t)\\
&+B_1^{-1}(t)JB_0(t)x_{T,0,t}x'_{T,0,t}B_0(t)J'B_1^{-1}(t))\delta\}\\
&+2\{T^{-1}\sum\nolimits_{t=T}^{T+P-\tau}\delta'(B_1^{-1}(t)JB_0(t)x_{T,0,t}x'_{T,0,t}B_0(t)J'\\
&-x_{T,1,t}x'_{T,1,t}JB_0(t)J')(T^{1/2}H_{T,1}(t))\}.
\end{aligned}$$

Given Assumptions 3 (c) and 5, 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

$$\begin{aligned}
&\sum\nolimits_{t=T}^{T+P-\tau}(\hat{u}^2_{0,t+\tau} - \hat{u}^2_{1,t+\tau}) = \sigma^2\{2\sum\nolimits_{t=T}^{T+P-\tau}(T^{-1/2}\tilde{h}_{T,1,t+\tau})(T^{1/2}\tilde{H}_{T,1}(t))\\
&-T^{-1}\sum\nolimits_{t=T}^{T+P-\tau}(T^{1/2}\tilde{H}_{T,1}(t))(T^{1/2}\tilde{H}_{T,1}(t))\} + \sigma^2\{2\sum\nolimits_{t=T}^{T+P-\tau}(\delta' B_1^{-1/2}\tilde{A}/\sigma)(T^{-1/2}\tilde{h}_{T,1,t+\tau})\}\\
&+\sigma^2\{(P/T)(\delta' J_2 F_1^{-1}J_2'\delta/\sigma^2)\} + o_p(1).
\end{aligned}$$

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

$$\begin{aligned}
&\sum\nolimits_{t=T}^{T+P-\tau}(\hat{u}^2_{0,t+\tau} - \hat{u}^2_{1,t+\tau}) =\\
&\sigma^2\{2\int_1^{1+\lambda_P} s^{-1}W'(s)S_{\tilde{h}\tilde{h}}dW(s) - \int_1^{1+\lambda_P} s^{-2}W'(s)S_{\tilde{h}\tilde{h}}W(s)ds\}\\
&+\sigma^2\{\int_1^{1+\lambda_P}(\delta' B_1^{-1/2}\tilde{A}'/\sigma)S^{1/2}_{\tilde{h}\tilde{h}}dW(s)\} + \sigma^2\{\lambda_P\delta' J_2F_1^{-1}J_2'\delta/\sigma^2\}.
\end{aligned}$$

That $\mathrm{MSE}_2 \to_p \sigma^2$ then provides the desired result.

(b) The proof is largely the same as for the recursive 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 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: (a) Given Theorem 2.1(a) and the Continuous Mapping Theorem it suffices to show that $P\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)\widehat{\gamma}_{dd}(j) \to_d 4\sigma^4(\Gamma_5+\Gamma_6+\Gamma_7)$. Before doing so it is convenient to redefine the bracketed terms from (11) used in the primary decomposition of the loss differential in the proof of Theorem 2.1(absent the summations, but keeping the brackets) as

$$(\hat{u}_{0,t+\tau}^2 - \hat{u}_{1,t+\tau}^2) = \{2A_{1,t} - A_{2,t}\} + 2\{B_t\} + \{C_t\} + 2\{D_t\}. \tag{12}$$

With this in mind, if we ignore the finite sample difference between P and $P-\tau+1$, we obtain

$$P\sum_{j=-\bar{j}}^{\bar{j}} K(j/M)\widehat{\gamma}_{dd}(j) = \sum_{j=-\bar{j}}^{\bar{j}} K(j/M)\sum\nolimits_{t=T+j}^{T+P-\tau}(\hat{u}_{0,t+\tau}^2 - \hat{u}_{1,t+\tau}^2)(\hat{u}_{0,t-j+\tau}^2 - \hat{u}_{1,t-j+\tau}^2) \tag{13}$$

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

In the remainder we show that each of the 3 bracketed terms in (13) converges 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 (13), if we recall the definition of $\tilde{h}_{T,1,t+\tau}$ and that $\bar{j}$ is finite, algebra along the lines of Clark and McCracken (2005) gives us

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

Given Assumptions 3 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 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

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

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 (13), similar arguments give us

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

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

Given Assumptions 3 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 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

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

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 (13), similar arguments give us

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

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

Given Assumptions 3 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 5 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 (13) 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

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

Assumptions 3 and 5, along with de Jong and Davidson (2000) suffice for $\sup_{T\leq t\leq T+P-1}|T^{1/2}H_{T,1}(t)|=O_p(1)$. Assumption 3 along with Markov's inequality imply both

$$T^{-1}\sum\nolimits_{t=T}^{T+P-1}|h_{T,1,t+\tau}vec[-JB_1(t)x_{T,0,t}x'_{T,0,t}B_0(t)J'+B_1(t)x_{T,1,t}x'_{T,1,t}B_1(t)]'|=O_p(1)$$

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

(b) The proof is largely the same as for the recursive scheme. And as was the case for Theorem 2.1, 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 2, 3′, 4, and 5 as well as either Assumption 1 or 1′. (a) $T^{1/2}J_2'\widetilde{\beta}_{1,T}=O_p(1)$. (b) $\sup_{T\leq t\leq T+P-\tau}|T^{1/2}(\hat{H}^*_{T,1}(t)-H^*_{T,1}(t))|=o_p(1)$.

Proof of Lemma 1: (a) Let $\hat{\zeta}$ denote the Lagrange multiplier[9] associated with the ridge regression and define $C_{12}(T)=J'B_1^{-1}(T)J_2$ and $C_{12}=\lim_{T\to\infty}E(x_{T,0,t}x'_{T,12,t})$.

(a-i) Maintain Assumption 1. The definition of the ridge estimator implies that for $\frac{1}{1+\hat{\zeta}}=\sqrt{\frac{\hat{d}}{(T^{1/2}\hat{\beta}_{1,T})'J_2F_1^{-1}(T)J_2'(T^{1/2}\hat{\beta}_{1,T})}}$, the ridge estimator takes the form

$$\widetilde{\beta}_{1,T}=\begin{pmatrix} I & \frac{\hat{\zeta}}{1+\hat{\zeta}}B_0(T)C_{12}(T)\\ 0 & \frac{1}{1+\hat{\zeta}}I\end{pmatrix}\hat{\beta}_{1,T}=\begin{pmatrix} I & \frac{\hat{\zeta}}{1+\hat{\zeta}}B_0(T)C_{12}(T)\\ 0 & \frac{1}{1+\hat{\zeta}}I\end{pmatrix}(\beta^*+T^{-1/2}\delta+B_1(T)H_{T,1}(T)).$$

Hence

$$\begin{aligned}
T^{1/2}J_2'\widetilde{\beta}_{1,T}&=J_2'\begin{pmatrix} I & \frac{\hat{\zeta}}{1+\hat{\zeta}}B_0(T)C_{12}(T)\\ 0 & \frac{1}{1+\hat{\zeta}}I\end{pmatrix}[\delta+B_1(T)(T^{1/2}H_{T,1}(T))]\\
&\to_d J_2'\begin{pmatrix} I & \frac{\zeta^*}{1+\zeta^*}B_0C_{12}\\ 0 & \frac{1}{1+\zeta^*}I\end{pmatrix}N(\delta,B_1VB_1)
\end{aligned}$$

[9]This multiplier satisfies $(\frac{1}{1+\hat{\zeta}})^2=\frac{\hat{d}}{(T^{1/2}\hat{\beta}_{1,T})'J_2F_1^{-1}(T)J_2'(T^{1/2}\hat{\beta}_{1,T})}$ and hence $\hat{\zeta}$ is unique only up to its' sign. In all aspects of this paper we use the value satisfying $\frac{1}{1+\hat{\zeta}}=\sqrt{\frac{\hat{d}}{(T^{1/2}\hat{\beta}_{1,T})'J_2F_1^{-1}(T)J_2'(T^{1/2}\hat{\beta}_{1,T})}}$. Choosing the opposite sign is irrelevant since, in every case, what matters is not the value of $\frac{1}{1+\hat{\zeta}}$ but it's square.

where

$\zeta^* =^d (N(\delta, B_1 V B_1))' J_2 F_1^{-1} J_2'(N(\delta, B_1 V B_1))$ a mixed non-central chi-square variate, and the proof is complete.

(a-ii) Maintain Assumption 1′. The ridge estimator takes the form

$$\widetilde{\beta}_{1,T} = \begin{pmatrix} I & \frac{\hat{\zeta}}{1+\hat{\zeta}} B_0(T) C_{12}(T) \\ 0 & \frac{1}{1+\hat{\zeta}} I \end{pmatrix} \hat{\beta}_{1,T} = \begin{pmatrix} I & \frac{\hat{\zeta}}{1+\hat{\zeta}} B_0(T) C_{12}(T) \\ 0 & \frac{1}{1+\hat{\zeta}} I \end{pmatrix} (\beta_1^* + B_1(T) H_{T,1}(T)).$$

Hence

$$\begin{aligned} T^{1/2} J_2' \widetilde{\beta}_{1,T} &= \sqrt{\frac{\hat{d}}{\hat{\beta}_{1,T}' J_2 F_1^{-1}(T) J_2' \hat{\beta}_{1,T}}} J_2' [\beta_1^* + B_1(T) H_{T,1}(T)] \\ &\to_p \sqrt{\frac{d}{\beta_{12}^{*\prime} F_1^{-1} \beta_{12}^*}} \beta_{12}^* \end{aligned}$$

and the proof is complete.

(b) For ease of presentation, we show the result for the recursive scheme and 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,

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

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

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

and hence

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

Assumptions 1 or 1′, along with 3′ 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 6 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 $T^{-1} \sum_{s=1}^{T-\tau} \eta_{s+2} \nabla \widehat{\varepsilon}_{T,1,s+2}(\overline{\gamma}_T) x_{T,1,s}$, $T^{-1} \sum_{s=1}^{T-\tau} \eta_{s+1} \nabla \widehat{\varepsilon}_{T,1,s+1}(\overline{\gamma}_T) x_{T,1,s}$, and $T^{-1} \sum_{s=1}^{T-\tau} \eta_{s+1} \times \varepsilon_{T,1,s+1} x_{T,1,s}$ are all $o_{a.s.}(1)$. This in turn, (along with Assumption 5) implies that $sup_t|.|$ of each of the partial sums is $o_p(1)$ and the proof is complete.

Proof of Theorem 2.3: We provide details for the recursive scheme noting differences for the rolling later. Straightforward algebra implies that

$$\begin{array}{l}
\sum_{t=T}^{T+P-\tau}(\hat{u}_{0,t+\tau}^{*2}-\hat{u}_{1,t+\tau}^{*2})=\sum_{t=T}^{T+P-\tau}\{2h_{T,1,t+\tau}^{\prime *}(-JB_0(t)J'+B_1(t))H_{T,1}^{*}(t)\\
-H_{T,1}^{\prime *}(t)(-JB_0(t)J'x_{T,1,t}x_{T,1,t}'JB_0(t)J'+B_1(t)x_{T,1,t}x_{T,1,t}'B_1(t))H_{T,1}^{*}(t)\}\\
+T^{-1/2}\sum_{t=T}^{T+P-\tau}\{2h_{T,1,t+\tau}^{\prime *}(-JB_0(t)J'+B_1(t))B_1^{-1}(t)(T^{1/2}\widetilde{\beta}_{1,T})\}\\
+T^{-1}\sum_{t=T}^{T+P-\tau}\{(T^{1/2}\widetilde{\beta}_{1,T})'B_1^{-1}(t)(-JB_0(t)J'+B_1(t))x_{T,1,t}x_{T,1,t}'(-JB_0(t)J'+B_1(t))B_1^{-1}(t)(T^{1/2}\widetilde{\beta}_{1,T})\}\\
\\
+2\sum_{t=T}^{T+P-\tau}\{h_{T,1,t+\tau}^{\prime *}(-JB_0(t)J'+B_1(t))(\hat{H}_{T,1}^{*}(t)-H_{T,1}^{*}(t))\\
+(\hat{h}_{T,1,t+\tau}^{*}-h_{T,1,t+\tau}^{*})'(-JB_0(t)J'+B_1(t))H_{T,1}^{*}(t)\\
-H_{T,1}^{\prime *}(t)(-JB_0(t)J'x_{T,1,t}x_{T,1,t}'JB_0(t)J'+B_1(t)x_{T,1,t}x_{T,1,t}'B_1(t))(\hat{H}_{T,1}^{*}(t)-H_{T,1}^{*}(t))\\
+(\hat{h}_{T,1,t+\tau}^{*}-h_{T,1,t+\tau}^{*})'(-JB_0(t)J'+B_1(t))(\hat{H}_{T,1}^{*}(t)-H_{T,1}^{*}(t))\\
-(0.5)(\hat{H}_{T,1}^{*}(t)-H_{T,1}^{*}(t))'(-JB_0(t)J'x_{T,1,t}x_{T,1,t}'JB_0(t)J'+B_1(t)x_{T,1,t}x_{T,1,t}'B_1(t))(\hat{H}_{T,1}^{*}(t)-H_{T,1}^{*}(t))\\
-\widetilde{\beta}_{1,T}'B_1^{-1}(t)(-JB_0(t)J'+B_1(t))x_{T,1,t}x_{T,1,t}'JB_0(t)J'H_{T,1}^{*}(t)\\
+\widetilde{\beta}_{1,T}'B_1^{-1}(t)(-JB_0(t)J'+B_1(t))(\hat{h}_{T,1,t+\tau}^{*}-h_{T,1,t+\tau}^{*})\\
-\widetilde{\beta}_{1,T}'B_1^{-1}(t)(-JB_0(t)J'+B_1(t))x_{T,1,t}x_{T,1,t}'JB_0(t)J'(\hat{H}_{T,1}^{*}(t)-H_{T,1}^{*}(t))\}
\end{array} \tag{14}$$

Note that there are 4 bracketed $\{.\}$ terms in (14). The first three are directly analogous to the three bracketed terms in (11) from the proof of Theorem 2.1. We will show that these three terms have limits $\Gamma_i^* =^d \Gamma_i$, for Γ_i $i=1-4$ defined in the text. The additional assumption of either conditional homoskedasticity or $k_1=1$ are needed only in the proof for $\Gamma_3^* =^d \Gamma_3$. Finally, we then show that the remaining fourth bracketed term is $o_p(1)$.

<u>Proof of bracket 1</u>: The sole difference between this term and that in the proof of Theorem 2.1 is that they are defined in terms of $h_{1,t+\tau}^*$ rather than $h_{1,t+\tau}$. Since these terms have the same first and second moments, as well as the same mixing properties, the exact same proof is applicable and hence we have

$$\begin{array}{l}
\sum_{t=T}^{T+P-\tau}\{2h_{T,1,t+\tau}^{\prime *}(-JB_0(t)J'+B_1(t))H_{T,1}^{*}(t)\\
-H_{T,1}^{\prime *}(t)(-JB_0(t)J'x_{T,1,t}x_{T,1,t}'JB_0(t)J'+B_1(t)x_{T,1,t}x_{T,1,t}'B_1(t))H_{T,1}^{*}(t)\}\to_d 2\Gamma_1^*-\Gamma_2^*
\end{array}$$

where Γ_1^* and Γ_2^* denote independent replicas of Γ_1 and Γ_2 respectively. Independence follows from the fact that the $\eta_{t+\tau}$ are *i.i.d.* $N(0,1)$.

<u>Proof of bracket 2</u>: Rearranging terms gives us

$$\begin{array}{l}
T^{-1/2}2\sum_{t=T}^{T+P-\tau}h_{T,1,t+\tau}^{\prime *}(-JB_0(t)J'+B_1(t))B_1^{-1}(t)(T^{1/2}\widetilde{\beta}_{1,T})\\
=T^{-1/2}2\sum_{t=T}^{T+P-\tau}h_{T,1,t+\tau}^{\prime *}B_1(t)J_2F_1^{-1}(t)(T^{1/2}J_2'\widetilde{\beta}_{1,T})
\end{array}.$$

From Lemma 1 we know $T^{1/2}J_2'\widetilde{\beta}_{1,T}=O_p(1)$. Algebra along the lines of Clark and McCracken (2005) then gives us

$$T^{-1/2}2\sum_{t=T}^{T+P-\tau}h_{T,1,t+\tau}^{\prime *}B_1(t)J_2F_1^{-1}(t)(T^{1/2}J_2'\widetilde{\beta}_{1,T})=T^{-1/2}2\sum_{t=T}^{T+P-\tau}h_{T,1,t+\tau}^{\prime *}B_1J_2F_1^{-1}(T^{1/2}J_2'\widetilde{\beta}_{1,T})+o_p(1).$$

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 $T^{-1/2}2\sum_{t=T}^{T+P-\tau}h_{T,1,t+\tau}'B_1J_2F_1^{-1}\beta_{12}^*+o_p(1)$. What makes them different here is that since $T^{1/2}J_2'\widetilde{\beta}_{1,T}$ is not consistent for β_{12}^*, it is not the case that $T^{-1/2}2\sum_{t=T}^{T+P-\tau}h_{T,1,t+\tau}^{\prime *}B_1J_2F_1^{-1}(T^{1/2}J_2'\widetilde{\beta}_{1,T})$ equals $T^{-1/2}2\sum_{t=T}^{T+P-\tau}h_{T,1,t+\tau}^{\prime *}B_1J_2F_1^{-1}\beta_{12}^*+o_p(1)$.

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

$$T^{-1/2}2\sum_{t=T}^{T+P-\tau}h_{T,1,t+\tau}'B_1J_2F_1^{-1}\beta_{12}^*\to_d\Gamma_2\sim N(0,4\Omega)$$

where $\Omega=\lambda_P\beta_{12}^{*\prime}F_1^{-1}J_2'B_1VB_1J_2F_1^{-1}\beta_{12}^*$. But for the latter, due to the *i.i.d.* $N(0,1)$ (and strictly exogenous) nature of the $\eta_{t+\tau}$, we have

$$T^{-1/2}2\sum_{t=T}^{T+P-\tau}h_{T,1,t+\tau}^{\prime *}B_1J_2F_1^{-1}J_2'(T^{1/2}\widetilde{\beta}_{1,T})\to_d\Gamma_3^*\sim N(0,4W)$$

where

$$\begin{aligned}
&W = \lim Var\{T^{-1/2}\textstyle\sum_{t=T}^{T+P-\tau} h'^{*}_{T,1,t+\tau} B_1 J_2 F_1^{-1} J_2'(T^{1/2}\widetilde{\beta}_{1,T})\} \\
&= \lambda_P \lim E\{(T^{1/2}\widetilde{\beta}_{1,T})' J_2 F_1^{-1} J_2' B_1 \{\lim Var(P^{-1/2}\textstyle\sum_{t=T}^{T+P-\tau} h'^{*}_{T,1,t+\tau})\} B_1 J_2 F_1^{-1} J_2'(T^{1/2}\widetilde{\beta}_{1,T})\} \\
&= \lambda_P \lim E\{(T^{1/2}\widetilde{\beta}_{1,T})' J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2'(T^{1/2}\widetilde{\beta}_{1,T})\}
\end{aligned} .$$

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

$$\begin{aligned}
&W = \sigma^2 \lambda_P \lim E\{(T^{1/2}\widetilde{\beta}_{1,T})' J_2 F_1^{-1} J_2'(T^{1/2}\widetilde{\beta}_{1,T})\} \\
&= \sigma^2 \lambda_P \lim E\{(T^{1/2}\widetilde{\beta}_{1,T})' J_2 F_1^{-1}(T) J_2'(T^{1/2}\widetilde{\beta}_{1,T})\} \\
&= \sigma^2 \lambda_P \lim E\{\hat{d}\} = \sigma^2 \lambda_P d
\end{aligned} .$$

The result follows since under the null hypothesis, $\Omega = \lambda_P \beta_{12}^{*'} F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} \beta_{12}^* = \sigma^2 \lambda_P \beta_{12}^{*'} F_1^{-1} \beta_{12}^* = \sigma^2 \lambda_P d$.

(b) If we let $\dim(\beta_{12}^*) = 1, W$ simplifies to

$$\begin{aligned}
&W = \lambda_P \lim E\{(T^{1/2}\widetilde{\beta}_{1,T})' J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2'(T^{1/2}\widetilde{\beta}_{1,T})\} \\
&= \lambda_P \lim E\{(T^{1/2}\hat{\beta}_{12,T})^2 (F_1^{-1})^2 J_2' B_1 V B_1 J_2\}
\end{aligned} .$$

But $\hat{\beta}_{12,T}$ was estimated satisfying the restriction that $(T^{1/2}\hat{\beta}_{12,T})^2 = F_1(T)\hat{d}$ and hence $W = \lambda_P \lim E\{F_1(T)\hat{d}(F_1^{-1})^2 J_2' B_1 V B_1 J_2\} = \lambda_P F_1^{-1} d J_2' B_1 V B_1 J_2$. Following similar arguments, we also have $\Omega = \lambda_P (\beta_{12}^*)^2 (F_1^{-1})^2 J_2' B_1 V B_1 J_2$. But under the null, $(\beta_{12}^*)^2 = dF_1$ and the proof is complete.

Proof of bracket 3: Rearranging terms gives us

$$\begin{aligned}
&T^{-1}\textstyle\sum_{t=T}^{T+P-\tau} (T^{1/2}\widetilde{\beta}_{1,T})' B_1^{-1}(t)(-JB_0(t)J' + B_1(t)) x_{T,1,t} x_{T,1,t}' (-JB_0(t)J' + B_1(t)) B_1^{-1}(t)(T^{1/2}\widetilde{\beta}_{1,T}) \\
&= T^{-1}\textstyle\sum_{t=T}^{T+P-\tau} (T^{1/2}\widetilde{\beta}_{1,T})' J_2 F_1^{-1}(t) J_2' B_1(t) x_{T,1,t} x_{T,1,t}' B_1(t) J_2 F_1^{-1}(t) J_2'(T^{1/2}\widetilde{\beta}_{1,T})
\end{aligned} .$$

From Lemma 1 we know $T^{1/2} J_2'\widetilde{\beta}_{1,T} = O_p(1)$. From there, algebra long the lines of Clark and McCracken (2005) gives us

$$\begin{aligned}
&T^{-1}\textstyle\sum_{t=T}^{T+P-\tau} (T^{1/2}\widetilde{\beta}_{1,T})' J_2 F_1^{-1}(t) J_2' B_1(t) x_{T,1,t} x_{T,1,t}' B_1(t) J_2 F_1^{-1}(t) J_2'(T^{1/2}\widetilde{\beta}_{1,T}) \\
&= T^{-1}\textstyle\sum_{t=T}^{T+P-\tau} (T^{1/2}\widetilde{\beta}_{1,T})' J_2 F_1^{-1}(t) J_2' B_1(t) B_1^{-1} B_1(t) J_2 F_1^{-1}(t) J_2'(T^{1/2}\widetilde{\beta}_{1,T}) + o_p(1) \\
&= T^{-1}\textstyle\sum_{t=T}^{T+P-\tau} (T^{1/2}\widetilde{\beta}_{1,T})' J_2 F_1^{-1} J_2'(T^{1/2}\widetilde{\beta}_{1,T}) + o_p(1) \\
&= (P - \tau + 1/T)\hat{d} + o_p(1) \to_p \lambda_P d \equiv \Gamma_4^*
\end{aligned} .$$

The result follows since under the null hypothesis, $\Gamma_4 \equiv \beta_{12}^{*'} F_1^{-1} \beta_{12}^* = \lambda_P d$.

Proof of bracket 4: We must show that each of the eight components of the fourth bracketed term in (14) 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

$$\begin{aligned}
&|\textstyle\sum_{t=T}^{T+P-\tau} (\hat{h}_{T,1,t+\tau}^* - h_{T,1,t+\tau}^*)'(-JB_0(t)J' + B_1(t))(\hat{H}_{T,1}^*(t) - H_{T,1}^*(t))| \\
&\leq k_1^2 (T^{-1/2}\textstyle\sum_{t=T}^{T+P-\tau} |\hat{h}_{T,1,t+\tau}^* - h_{T,1,t+\tau}^*|)(\sup_t |-JB_0(t)J' + B_1(t)|)(\sup_t T^{1/2}|\hat{H}_{T,1}^*(t) - H_{T,1}^*(t)|)
\end{aligned}$$

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

$$\begin{aligned}
&\textstyle\sum_{t=T}^{T+P-\tau} \widetilde{\beta}_{1,T}' B_1^{-1}(t)(-JB_0(t)J' + B_1(t))(\hat{h}_{T,1,t+\tau}^* - h_{T,1,t+\tau}^*) \\
&= (T^{1/2} J_2'\widetilde{\beta}_{1,T})' F_1^{-1} J_2' B_1 (T^{-1/2}\textstyle\sum_{t=T}^{T+P-\tau} (\hat{h}_{T,1,t+\tau}^* - h_{T,1,t+\tau}^*)) + o_p(1).
\end{aligned}$$

Lemma 1 implies both $\sup_t T^{1/2}|\hat{H}_{T,1}^*(t) - H_{T,1}^*(t)| = o_p(1)$ and $T^{1/2} J_2'\widetilde{\beta}_{1,T} = O_p(1)$ while Assumption $3'$ 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 1b (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, 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 we have

$$
\begin{aligned}
&T^{-1/2}\sum\nolimits_{t=T}^{T+P-\tau}|\hat{h}^*_{T,1,t+\tau}-h^*_{T,1,t+\tau}|\leq\\
&2k_1T^{-1}\sum\nolimits_{t=T}^{T+P-\tau}|\eta_{t+2}\nabla\widehat{\varepsilon}_{T,1,t+2}(\overline{\gamma}_T)x_{T,1,t}||T^{1/2}(\widehat{\gamma}_T-\gamma_T)|\\
&+\theta 2k_1T^{-1}\sum\nolimits_{t=T}^{T+P-\tau}|\eta_{s+1}\nabla\widehat{\varepsilon}_{T,1,t+1}(\overline{\gamma}_T)x_{T,1,t}||T^{1/2}(\widehat{\gamma}_T-\gamma_T)|\\
&+(\widehat{\theta}-\theta)2k_1T^{-1}\sum\nolimits_{t=T}^{T+P-\tau}|\eta_{t+1}\nabla\widehat{\varepsilon}_{T,1,t+1}(\overline{\gamma}_T)x_{T,1,t}||T^{1/2}(\widehat{\gamma}_T-\gamma_T)|\\
&+(T^{1/2}(\widehat{\theta}-\theta))T^{-1}\sum\nolimits_{t=T}^{T+P-\tau}|\eta_{t+1}\varepsilon_{T,1,t+1}x_{T,1,t}|.
\end{aligned}
$$

Assumptions 1 or $1'$ and $3'$ 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 $3'$ 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 $i.i.d.N(0,1)$, $T^{-1}\sum_{s=1}^{T-\tau}|\eta_{s+2}\nabla\widehat{\varepsilon}_{T,1,s+2}(\overline{\gamma}_T)x_{T,1,s}|$, $T^{-1}\sum_{s=1}^{T-\tau}|\eta_{s+1}\nabla\widehat{\varepsilon}_{T,1,s+1}(\overline{\gamma}_T)x_{T,1,s}|$, and $T^{-1}\sum_{s=1}^{T-\tau}|\eta_{s+1}\varepsilon_{T,1,s+1}x_{T,1,s}|$ 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 2.4: Given Theorem 2.3 and the Continuous Mapping Theorem it suffices to show that $P\sum_{j=-\bar{j}}^{\bar{j}}K(j/M)\hat{\gamma}^*_{dd}(j)\rightarrow^d 4\sigma_u^4(\Gamma^*_5+\Gamma^*_6+\Gamma^*_7)$ 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 (14) used in the main decomposition of the loss differential in Theorem 2.3 (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, if we ignore the finite sample difference between P and $P-\tau+1$, we obtain

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

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: As was the case in the proof of Theorem 2.3, the sole difference between this term and that in the proof of Theorem 2.2 is that they are defined in terms of $h^*_{T,1,t+\tau}$ rather than $h_{T,1,t+\tau}$. Since these terms have the same first and second moments, as well as the same mixing properties, the exact same proof is applicable and hence we have

$$
\begin{aligned}
&\sum\nolimits_{j=-\bar{j}}^{\bar{j}}K(j/M)\sum\nolimits_{t=T+j}^{T+P-\tau}A^*_{1,t}A^*_{1,t-j}=\\
&\sigma^4\sum\nolimits_{j=-\bar{j}}^{\bar{j}}K(j/M)\sum\nolimits_{t=T+j}^{T+P-\tau}(T^{1/2}H'^*_{T,1}(t)B_1^{1/2}/\sigma^2)B_1^{-1/2}(-JB_0(t)J'+B_1(t))\times\\
&B_1^{-1/2}(B_1^{1/2}h^*_{T,1,t+\tau}h'^*_{T,1,t-j+\tau}B_1^{1/2}/\sigma^2)B_1^{-1/2}(-JB_0(t-j)J'+B_1(t-j))B_1^{-1/2}(T^{1/2}B_1^{1/2}H^*_{T,1}(t-j)/\sigma^2)\\
&\rightarrow^d\sigma^4\Gamma^*_5
\end{aligned}
$$

where Γ^*_5 denotes an independent replica of Γ_5. Independence follows from the fact that the $\eta_{t+\tau}$ are $i.i.d.$ $N(0,1)$.

Proof of bracket 2: After rearranging terms, the second bracketed term is

$$
\begin{aligned}
&\textstyle\sum_{j=-\bar{j}}^{\bar{j}} K(j/M) \sum_{t=T+j}^{T+P-\tau} A_{1,t}^{*} B_{1,t-j}^{*} \\
&= \textstyle\sum_{j=-\bar{j}}^{\bar{j}} K(j/M) T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2} H_{T,1}^{*}(t))'(-JB_0(t)J' + B_1(t)) \times \\
&h_{T,1,t+\tau}^{*} h_{T,1,t-j+\tau}^{\prime *}(-JB_0(t-j)J' + B_1(t-j))B_1^{-1}(t-j)(T^{1/2}\widetilde{\beta}_{1,T}) \\
&= \textstyle\sum_{j=-\bar{j}}^{\bar{j}} K(j/M) T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2} H_{T,1}^{*}(t))'(-JB_0(t)J' + B_1(t)) \times \\
&h_{T,1,t+\tau}^{*} h_{T,1,t-j+\tau}^{\prime *} B_1(t-j) J_2 F_{11}^{-1}(t-j) J_2'(T^{1/2}\widetilde{\beta}_{1,T})
\end{aligned}
.
$$

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, it differs because $J_2'(T^{1/2}\widetilde{\beta}_{1,T})$ is not consistent for β_{12}^{*}. However, it is true that both terms are asymptotically normal. To see this note that

$$
\begin{aligned}
&\textstyle\sum_{j=-\bar{j}}^{\bar{j}} K(j/M) T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2} H_{T,1}^{*}(t))'(-JB_0(t)J' + B_1(t)) \times \\
&h_{T,1,t+\tau}^{*} h_{T,1,t-j+\tau}^{\prime *} B_1(t-j) J_2 F_1^{-1}(t-j) J_2'(T^{1/2}\widetilde{\beta}_{1,T}) \\
&= \textstyle\sum_{j=-\bar{j}}^{\bar{j}} K(j/M) T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2} H_{T,1}^{*}(t))'(-JB_0J' + B_1) \times \\
&(Eh_{T,1,t+\tau}^{*} h_{T,1,t-j+\tau}^{\prime *}) B_1 J_2 F_1^{-1} J_2'(T^{1/2}\widetilde{\beta}_{1,T}) + o_p(1) \\
&= T^{-1} \textstyle\sum_{t=T+j}^{T+P-\tau} (T^{1/2} H_{T,1}^{*}(t))'(-JB_0J' + B_1) V B_1 J_2 F_1^{-1} J_2'(T^{1/2}\widetilde{\beta}_{1,T}) + o_p(1) \\
&\rightarrow^{d} \sigma^4 \Gamma_6^{*} \sim N(0, W)
\end{aligned}
$$

where $W = \ln(1+\lambda_P)\sigma^{-8} \lim E\{(T^{1/2}\widetilde{\beta}_{1,T})' J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2' (T^{1/2}\widetilde{\beta}_{1,T})\}$. 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 $T^{-1}\sum_{t=T+j}^{T+P-\tau}(T^{1/2}H_{T,1}^{*}(t)) \rightarrow^{d} \int_1^{1+\lambda_P} s^{-1} V^{1/2} W^{*}(s) ds \sim N(0, \ln(1+\lambda_P)V)$. As in the proof of Theorem 2.3, the exact relationship between Γ_6^{*} and Γ_6 depends upon the additional assumptions stated in the Theorem.

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

$$
\begin{aligned}
W &= \sigma^6 \ln(1+\lambda_P) \lim E\{(T^{1/2}\widetilde{\beta}_{1,T})' J_2 F_1^{-1} J_2' (T^{1/2}\widetilde{\beta}_{1,T})\} \\
&= \sigma^6 \ln(1+\lambda_P) \lim E\{(T^{1/2}\widetilde{\beta}_{1,T})' J_2 F_1^{-1}(T) J_2' (T^{1/2}\widetilde{\beta}_{1,T})\} \\
&= \sigma^6 \ln(1+\lambda_P) \lim E(\hat{d}) = \sigma^6 \ln(1+\lambda_P) d
\end{aligned}
.
$$

But from Theorem 2.2, the definition of Γ_6 gives us

$$\sigma^4\Gamma_6 = (\int_1^{1+\lambda_P} s^{-1} W(s) ds)' V^{1/2} B_1 J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2' \delta \sim N(0, \Omega)$$

where

$$\Omega = \ln(1+\lambda_P)\delta' J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2' \delta.$$

Assuming conditional homoskedasticity this simplifies to $\Omega = \sigma^6 \ln(1+\lambda_P)\beta_{12}^{*'} F_1^{-1} \beta_{12}^{*}$. The result then follows since under the null, $\beta_{12}^{*'} F_1^{-1} \beta_{12}^{*} = d$.

(b) If β_{12}^{*} is scalar we find that

$$
\begin{aligned}
W &= \ln(1+\lambda_P) \lim E\{(T^{1/2}\widetilde{\beta}_{12,T})^2 (F_1^{-1})^2 J_2' B_1 V B_1 J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2' B_1 V B_1 J_2\} \\
&= \ln(1+\lambda_P) \lim E\{\hat{d} F_1(T) (F_{11}^{-1})^4 (J_2' B_1 V B_1 J_2)^3\} \\
&= \ln(1+\lambda_P) d (F_1^{-1})^3 (J_2' B_1 V B_1 J_2)^3
\end{aligned}
.
$$

But from Theorem 2.2, the definition of Γ_6 gives us

$$\sigma_u^4\Gamma_6 = (\int_1^{1+\lambda_P} s^{-1} W'(s) V^{1/2} ds) B_1 J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2' \delta \sim N(0, \Omega)$$

where

$$\Omega = \ln(1+\lambda_P)(\beta_{12}^{*})^2 (F_1^{-1})^4 (J_2' B_1 V B_1 J_2)^3.$$

The result then follows since under the null, $(\beta_{12}^{*})^2 F_1^{-1} = d$.

Proof of bracket 3: After rearranging terms, the third bracketed term is

$$\begin{aligned}
&\textstyle\sum_{j=-\bar{j}}^{\bar{j}} K(j/M) \sum_{t=T+j}^{T+P-\tau} B_{1,t}^{*} B_{1,t-j}^{*} = \sum_{j=-\bar{j}}^{\bar{j}} K(j/M) T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2}\widetilde{\beta}_{1,T})' B_1^{-1}(t)(-JB_0(t)J' + B_1(t)) \times \\
&h_{T,1,t+\tau}^{*} h_{T,1,t-j+\tau}^{\prime *}(-JB_0(t-j)J' + B_1(t-j))B_1^{-1}(t-j)(T^{1/2}\widetilde{\beta}_{1,T}) \\
&= \textstyle\sum_{j=-\bar{j}}^{\bar{j}} K(j/M) T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2}\widetilde{\beta}_{1,T}) J_2 F_1^{-1}(t) J_2' B_1(t) h_{1,t+\tau}^{*} h_{1,t-j+\tau}^{\prime *} B_1(t-j) J_2 F_1^{-1}(t-j) J_2'(T^{1/2}\widetilde{\beta}_{1,T})
\end{aligned}.$$

This term is also different from that for the third bracketed term in Theorem 2.2. As in the proof of Lemma 2, it differs because $T^{1/2}J_2'\widetilde{\beta}_{1,T}$ is not consistent for β_{12}^{*}. Even so, since $T^{1/2}J_2'\widetilde{\beta}_{1,T} = O_p(1)$, the above term is also $O_p(1)$. To see this, algebra along the lines of Clark and McCracken (2005) gives us

$$\begin{aligned}
&\textstyle\sum_{j=-\bar{j}}^{\bar{j}} K(j/M) T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2}\widetilde{\beta}_{1,T})' J_2 F_1^{-1}(t) J_2' B_1(t) h_{T,1,t+\tau}^{*} h_{T,1,t-j+\tau}^{\prime *} B_1(t-j) J_2 F_1^{-1}(t-j) J_2'(T^{1/2}\widetilde{\beta}_{1,T}) \\
&= \textstyle\sum_{j=-\bar{j}}^{\bar{j}} K(j/M) T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2}\widetilde{\beta}_{1,T})' J_2 F_1^{-1}(t) J_2' B_1(t)(E h_{T,1,t+\tau}^{*} h_{T,1,t-j+\tau}^{\prime *}) B_1(t) J_2 F_1^{-1}(t) J_2'(T^{1/2}\widetilde{\beta}_{1,T}) + o_p(1) \\
&= \textstyle\sum_{j=-\bar{j}}^{\bar{j}} K(j/M) T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2}\widetilde{\beta}_{1,T})' J_2 F_1^{-1} J_2' B_1 (E h_{T,1,t+\tau}^{*} h_{T,1,t-j+\tau}^{\prime *}) B_1 J_2 F_1^{-1} J_2'(T^{1/2}\widetilde{\beta}_{1,T}) + o_p(1) \\
&= \textstyle T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2}\widetilde{\beta}_{1,T})' J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2'(T^{1/2}\widetilde{\beta}_{1,T}) + o_p(1) \\
&= \sigma^4 \Gamma_7^{*} \equiv \lim \lambda_P (T^{1/2}\widetilde{\beta}_{1,T})' J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2'(T^{1/2}\widetilde{\beta}_{1,T})
\end{aligned}$$

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 $V = \sigma^2 B_2^{-1}$, we immediately see that

$$\begin{aligned}
&\Gamma_7^{*} \equiv \lambda_P \sigma^{-4} \lim\{(T^{1/2}\widetilde{\beta}_{1,T})' J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2'(T^{1/2}\widetilde{\beta}_{1,T})\} \\
&= \lambda_P \sigma^{-2} \lim\{(T^{1/2}\widetilde{\beta}_{1,T}) J_2 F_1^{-1} J_2'(T^{1/2}\widetilde{\beta}_{1,T})\} = \lambda_P \sigma^{-2} \lim \hat{d} = \sigma^{-2}\lambda_P d
\end{aligned}.$$

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

$$\Gamma_7 \equiv \sigma^{-4}\lambda_P \beta_{12}^{*'} F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} \beta_{12}^{*} = \sigma^{-2}\lambda_P \beta_{12}^{*'} F_1^{-1} \beta_{12}^{*} = \sigma^{-2}\lambda_P d = \Gamma_7^{*}$$

and the proof is complete.

(b) If we let β_{12}^{*} be scalar we find that

$$\begin{aligned}
&\sigma^4 \Gamma_7^{*} \equiv \lim \lambda_P (T^{1/2}\widetilde{\beta}_{1,T})' J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2'(T^{1/2}\widetilde{\beta}_{1,T}) \\
&= \lambda_P \lim (T^{1/2}\widetilde{\beta}_{12,T})^2 (F_1^{-1})^2 J_2' B_1 V B_1 J_2 \\
&= \lambda_P \lim \hat{d} F_1^{-1}(T)(F_1^{-1})^2 J_2' B_1 V B_1 J_2 \\
&= \lambda_P d F_1^{-1} J_2' B_1 V B_1 J_2 + o_p(1)
\end{aligned}.$$

But under the null, and with the additional assumption of that β_{12}^{*} is scalar, from Theorem 2.2 we know that

$$\begin{aligned}
&\sigma^4 \Gamma_7 \equiv \lambda_P \beta_{12}^{*'} F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} \beta_{12}^{*} = \lambda_P (\beta_{12}^{*})^2 (F_1^{-1})^2 J_2' B_1 V B_1 J_2 \\
&= \lambda_P d F_1^{-1} J_2' B_1 V B_1 J_2 = \sigma^4 \Gamma_7^{*}
\end{aligned}$$

and the proof is complete.

Proof of bracket 4: 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 $h_{T,1,t+\tau}$. But since these terms have the same first and second moments, as well as the same mixing properties, nearly the same proof is applicable 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 2.5: Regardless of whether the recursive or rolling scheme is used, the proof follows very similar arguments to those used in Theorems 2.3 and 2.4. Any differences that arise come from differences in the asymptotic behavior of $T^{1/2}J_2'\widetilde{\beta}_{1,T}$ under Assumption $1'$ as compared to Assumption 1. Therefore, since the decomposition at the beginning of the proof of Theorem 2.3 is unaffected by whether Assumption 1 or $1'$ holds, and the first bracketed term does not depend upon the value of either β^*_{12} or $T^{1/2}J_2'\widetilde{\beta}_{1,T}$ the same proof can be applied to show $2\Gamma^*_1 - \Gamma^*_2 =^d 2\Gamma_1 - \Gamma_2$ and $\Gamma^*_5 =^d \Gamma_5$ under Assumption $1'$. For the third bracketed term, the asymptotic behavior of $T^{1/2}J_2'\widetilde{\beta}_{1,T}$ is also irrelevant – all that matters is that the ridge constraint is still imposed whether working under Assumption 1 or $1'$.

Differences 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, under Assumption 1, Lemma 1 implies $T^{1/2}J_2'\widetilde{\beta}_{1,T} = O_p(1)$. These arguments continue to hold since under Assumption $1'$, $T^{1/2}J_2'\widetilde{\beta}_{1,T}$ remains $O_p(1)$ – despite also having the property that $T^{1/2}J_2'\widetilde{\beta}_{1,T} \to^p \sqrt{\frac{d}{\beta^{*\prime}_{12}F_1^{-1}\beta^*_{12}}}\beta^*_{12}$.

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 $T^{1/2}J_2'\widetilde{\beta}_{1,T}$ under Assumption $1'$ 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.

Proof that $\Gamma^*_3 =^d \Gamma_3$: As in the proof for Theorem 2.3, the second bracketed term satisfies

$$T^{-1/2}2\sum_{t=T}^{T+P-\tau} h'^*_{T,1,t+\tau}B_1(t)J_2F_1^{-1}(t)J_2'(T^{1/2}\widetilde{\beta}_{1,T}) = T^{-1/2}2\sum_{t=T}^{T+P-\tau} h'^*_{T,1,t+\tau}B_1J_2F_1^{-1}J_2'(T^{1/2}\widetilde{\beta}_{1,T})+o_p(1).$$

What makes this different under Assumption $1'$ is that since $T^{1/2}J_2'\widetilde{\beta}_{1,T} \to^p \sqrt{\frac{d}{\beta^{*\prime}_{12}F_1^{-1}\beta^*_{12}}}\beta^*_{12}$ we also have

$$\begin{aligned} & T^{-1/2}2\sum_{t=T}^{T+P-\tau} h'^*_{T,1,t+\tau}B_1(t)J_2F_1^{-1}(t)J_2'(T^{1/2}\widetilde{\beta}_{1,T}) \\ = \; & T^{-1/2}2\sqrt{\frac{d}{\beta^{*\prime}_{12}F_1^{-1}\beta^*_{12}}}\sum_{t=T}^{T+P-\tau} h'^*_{T,1,t+\tau}B_1J_2F_1^{-1}\beta^*_{12} + o_p(1) \\ \to \; & {}^dN(0, 4W) \end{aligned}$$

where

$$W = (\tfrac{d}{\beta^{*\prime}_{12}F_1^{-1}\beta^*_{12}})\lambda_P\beta^{*\prime}_{12}F_1^{-1}J_2'B_1VB_1J_2F_1^{-1}\beta^*_{12} \;.$$

Since $\Gamma_3\tilde{}N(0, 4\Omega), \Omega = \lambda_P\beta^{*\prime}_{12}F_1^{-1}J_2'B_1VB_1J_2F_1^{-1}\beta^*_{12}$, 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 $V = \sigma^2B_1^{-1}$, W simplifies to

$$W = \sigma^2(\frac{d}{\beta^{*\prime}_{12}F_1^{-1}\beta^*_{12}})\lambda_P\beta^{*\prime}_{12}F_1^{-1}\beta^*_{12} = \sigma^2\lambda_Pd.$$

The result follows since under the null hypothesis, $\Omega = \lambda_P\beta^{*\prime}_{12}F_1^{-1}J_2'B_1VB_1J_2F_1^{-1}\beta^*_{12} = \sigma^2\lambda_P\beta^{*\prime}_{12}F_1^{-1}\beta^*_{12} = \sigma^2\lambda_Pd$.

(b) If we let $\dim(\beta^*_{12}) = 1$ and note that in this case $J_2'B_1VB_1J_2 = F_1 \cdot tr((-JB_0J' + B_1)V)$, W simplifies to

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

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

Proof that $\Gamma_6^* =^d \Gamma_6$: As in the proof for Theorem 2.4, the second bracketed term satisfies

$$\begin{array}{l}\sum_{j=-\bar{j}}^{\bar{j}} K(j/M) \sum_{t=T+j}^{T+P-\tau} A_{1,t}^* B_{1,t-j}^* \\ = \sum_{j=-\bar{j}}^{\bar{j}} K(j/M) T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2} H_{T,1}^*(t))'(-JB_0(t)J' + B_1(t)) \times \\ h_{T,1,t+\tau}^* h_{T,1,t-j+\tau}'^* (-JB_0(t-j)J' + B_1(t-j)) B_1^{-1}(t-j)(T^{1/2}\widetilde{\beta}_{1,T}) \\ = T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2} H_{T,1}^*(t))' B_1 J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} (T^{1/2} J_2' \widetilde{\beta}_{1,T}) + o_p(1)\end{array} .$$

What makes this different under Assumption $1'$ is that since $T^{1/2} J_2' \widetilde{\beta}_{1,T} \to^p \sqrt{\frac{d}{\beta_{12}^{*\prime} F_1^{-1} \beta_{12}^*}} \beta_{12}^*$ we also have

$$\begin{array}{l}T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2} H_{T,1}^*(t))' B_1 J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} (T^{1/2} J_2' \widetilde{\beta}_{1,T}) \\ = (\sqrt{\frac{d}{\beta_{12}^{*\prime} F_1^{-1} \beta_{12}^*}}) T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2} H_{T,1}^*(t))' B_1 J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} \beta_{12}^* + o_p(1) \\ \to_d N(0, W)\end{array}$$

where $W = \ln(1+\lambda_P)(\frac{d}{\beta_{12}^{*\prime} F_1^{-1} \beta_{12}^*})\{\beta_{12}^{*\prime} F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} \beta_{12}^*\}$. 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 $\Gamma_6 \tilde{} N(0,\Omega)$, $\Omega = \ln(1+\lambda_P)\delta' J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2' \delta$, 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,

$$\begin{array}{l}\Omega = \ln(1+\lambda_P)\delta' J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} J_2' \delta \\ = \sigma^6 \ln(1+\lambda_P) \beta_{12}^{*'} F_1^{-1} \beta_{12}^* = \sigma^6 \ln(1+\lambda_P) d.\end{array} .$$

(b) If we let $\dim(\beta_{12}^*) = 1$ and note that in this case $J_2' B_1 V B_1 J_2 = F_1 \cdot tr((-JB_0 J' + B_1)V)$, W simplifies to

$$W = \ln(1+\lambda_P) d \cdot tr((-JB_0 J' + B_1)V)^3 \ .$$

The result follows since under the null hypothesis, $\Omega = \ln(1+\lambda_P) d \cdot tr((-JB_0 J' + B_1)V)^3$ and the proof is complete.

Proof that $\Gamma_7^* =^d \Gamma_7$: As in the proof for Theorem 2.4, the third bracketed term satisfies

$$\begin{array}{l}\sum_{j=-\bar{j}}^{\bar{j}} K(j/M) \sum_{t=T+j}^{T+P-\tau} B_{1,t}^* B_{1,t-j}^* = \sum_{j=-\bar{j}}^{\bar{j}} K(j/M) T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2}\widetilde{\beta}_{1,T})' B_1^{-1}(t)(-JB_0(t)J' + B_1(t)) \times \\ h_{T,1,t+\tau}^* h_{T,1,t-j+\tau}'^* (-JB_0(t-j)J' + B_1(t-j)) B_1^{-1}(t-j)(T^{1/2}\widetilde{\beta}_{1,T}) \\ = T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2}\widetilde{\beta}_{1,T})' J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} (T^{1/2} J_2' \widetilde{\beta}_{1,T}) + o_p(1)\end{array} .$$

What makes this different under Assumption $1'$ is that since $T^{1/2} J_2' \widetilde{\beta}_{1,T} \to^p \sqrt{\frac{d}{\beta_{12}^{*\prime} F_1^{-1} \beta_{12}^*}} \beta_{12}^*$ we also have

$$\begin{array}{l}T^{-1} \sum_{t=T+j}^{T+P-\tau} (T^{1/2}\widetilde{\beta}_{1,T})' J_2 F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} (T^{1/2} J_2' \widetilde{\beta}_{1,T}) \\ = \lambda_P (\frac{d}{\beta_{12}^{*\prime} F_1^{-1} \beta_{12}^*}) \beta_{12}^{*\prime} F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} \beta_{12}^* \equiv \Gamma_7^*\end{array} .$$

In contrast, the associated term from Theorem 2.2 takes the value $\Gamma_7 = \lambda_P \beta_{12}^{*\prime} F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} \beta_{12}^*$. 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,$\Gamma_7 = \lambda_P \beta_{12}^{*\prime} F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} \beta_{12}^* = \lambda_P \sigma^2 d$ and the proof is complete.

(b) If we let $\dim(\beta_{12}^*) = 1$ and note that in this case $J_2' B_1 V B_1 J_2 = F_1 \cdot tr((-JB_0 J' + B_1)V)$, Γ_7^* simplifies to $\lambda_P d tr((-JB_0 J' + B_1)V)$. The result follows since under the null hypothesis, $\Gamma_7 = \lambda_P \beta_{12}^{*\prime} F_1^{-1} J_2' B_1 V B_1 J_2 F_1^{-1} \beta_{12}^* = \lambda_P d tr((-JB_0 J' + B_1)V)$ 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.

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 (2008), "Combining Forecasts From Nested Models," *Oxford Bulletin of Economics and Statistics*, forthcoming.

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.

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.

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.

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. (2000), "Testing for Structural Change in Conditional Models," *Journal of Econometrics* 97, 93-115.

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

Hjalmarsson, Erik (2006), "Should We Expect Significant Out-of-Sample Results when Predicting Stock Returns?" Federal Reserve Board International Finance Discussion Paper #855.

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 (2008), "Out-of-Sample Exchange Rate Predictability with Taylor Rule Fundamentals," manuscript, University of Houston.

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

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 2 experiments

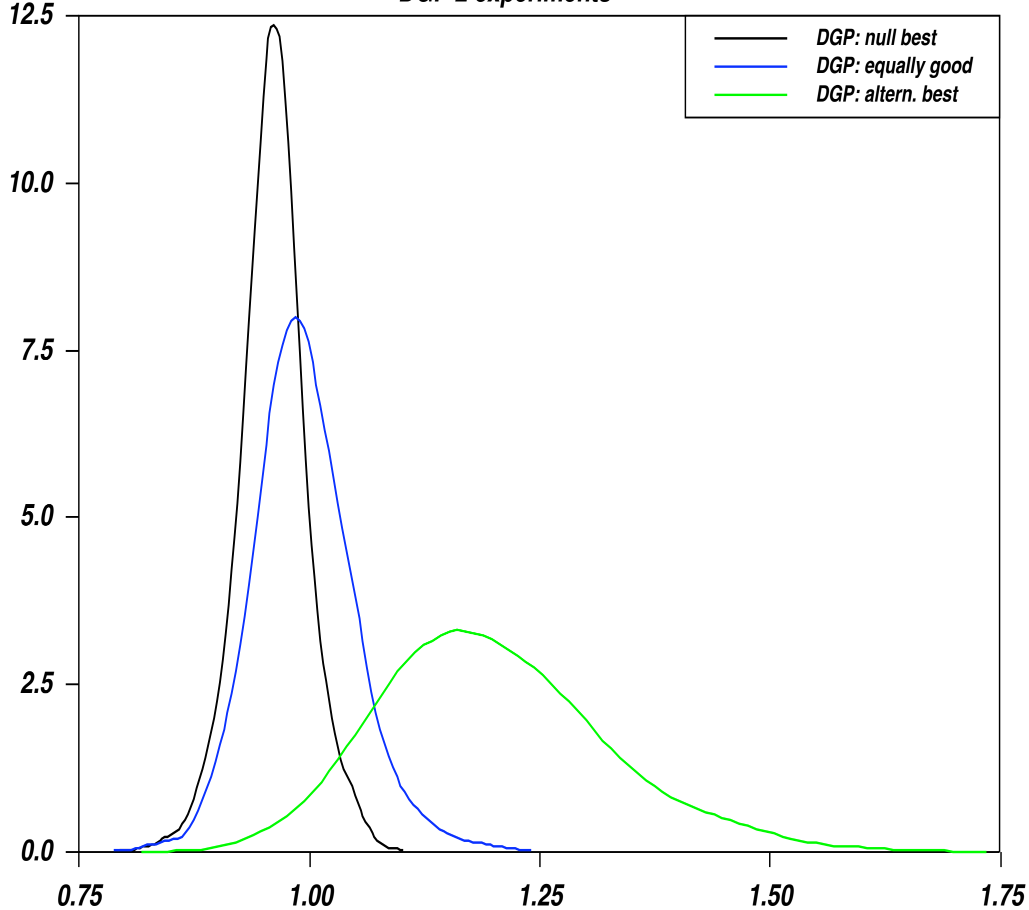

Table 1: Monte Carlo Rejection Rates, Null Model Best

(nominal size = 10%)

DGP 1					
statistic	*bootstrap approach*	R=40 P=80	R=80 P=40	R=80 P=80	R=80 P=120
MSE-F	non-parametric	.009	.041	.018	.015
MSE-F	restricted VAR	.099	.114	.095	.103
MSE-F	fixed regressor	.012	.038	.022	.018
MSE-t	non-parametric	.013	.052	.019	.018
MSE-t	restricted VAR	.095	.105	.099	.104
MSE-t	fixed regressor	.015	.047	.026	.024
CW t-test	restricted VAR	.094	.112	.097	.100
DGP 2					
statistic	*bootstrap approach*	R=40 P=80	R=80 P=40	R=80 P=80	R=80 P=120
MSE-F	non-parametric	.000	.013	.004	.003
MSE-F	restricted VAR	.096	.093	.097	.104
MSE-F	fixed regressor	.003	.010	.003	.005
MSE-t	non-parametric	.001	.019	.005	.004
MSE-t	restricted VAR	.096	.087	.095	.104
MSE-t	fixed regressor	.005	.019	.007	.005
CW t-test	restricted VAR	.098	.086	.102	.107

Notes:
1. The data generating processes are defined in equations (5) and (8). In these experiments, the coefficients $b_{ij} = 0$ for all i, j, such that the null forecasting model is expected to be most accurate.
2. For each artificial data set, one-step ahead forecasts of y_t are formed recursively using estimates of equations (6) and (7) in the case of the DGP 1 experiments and equations (9) and (10) in the case of the DGP 2 experiments. These forecasts are then used to form the indicated test statistics, defined in Section 2.2. R 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 3 describes the bootstrap procedures.
4. The number of Monte Carlo simulations is 2000; the number of bootstrap draws is 499.

Table 2: Monte Carlo Rejection Rates, Equally Accurate Models

(nominal size = 10%)

DGP 1					
statistic	*bootstrap approach*	R=40 P=80	R=80 P=40	R=80 P=80	R=80 P=120
MSE-F	non-parametric	.048	.088	.070	.068
MSE-F	restricted VAR	.296	.253	.256	.304
MSE-F	fixed regressor	.091	.114	.101	.108
MSE-t	non-parametric	.054	.098	.077	.073
MSE-t	restricted VAR	.283	.208	.237	.281
MSE-t	fixed regressor	.083	.100	.095	.103
CW t-test	restricted VAR	.310	.237	.270	.320
DGP 2					
statistic	*bootstrap approach*	R=40 P=80	R=80 P=40	R=80 P=80	R=80 P=120
MSE-F	non-parametric	.032	.060	.050	.064
MSE-F	restricted VAR	.466	.295	.372	.448
MSE-F	fixed regressor	.085	.088	.081	.099
MSE-t	non-parametric	.048	.080	.061	.072
MSE-t	restricted VAR	.480	.262	.356	.451
MSE-t	fixed regressor	.081	.087	.074	.095
CW t-test	restricted VAR	.541	.319	.432	.519

Notes:
1. See the notes to Table 1.
2. In these experiments, the coefficients $b_{ij} = 0$ are scaled such that the null and alternative models are expected to equally accurate (on average) over the forecast sample.

Table 3: Monte Carlo Rejection Rates, Equally Accurate Models
Rolling Forecasts
(nominal size = 10%)

DGP 1					
statistic	*bootstrap approach*	R=40 P=80	R=80 P=40	R=80 P=80	R=80 P=120
MSE-F	non-parametric	.034	.081	.052	.052
MSE-F	restricted VAR	.336	.245	.295	.340
MSE-F	fixed regressor	.097	.115	.105	.108
MSE-t	non-parametric	.043	.101	.065	.062
MSE-t	restricted VAR	.332	.206	.269	.328
MSE-t	fixed regressor	.081	.104	.097	.106
CW t-test	restricted VAR	.359	.241	.317	.378

DGP 2					
statistic	*bootstrap approach*	R=40 P=80	R=80 P=40	R=80 P=80	R=80 P=120
MSE-F	non-parametric	.011	.055	.032	.036
MSE-F	restricted VAR	.541	.323	.418	.524
MSE-F	fixed regressor	.074	.088	.084	.090
MSE-t	non-parametric	.022	.076	.047	.044
MSE-t	restricted VAR	.588	.287	.431	.546
MSE-t	fixed regressor	.072	.085	.077	.083
CW t-test	restricted VAR	.623	.358	.499	.612

Notes:
1. See the notes to Table 1.
2. In these experiments, the coefficients $b_{ij} = 0$ are scaled such that the null and alternative models are expected to equally accurate (on average) over the forecast sample.
3. In these experiments, the forecasting scheme is rolling, rather than recursive.

Table 4: Monte Carlo Rejection Rates, Alternative Model Best

(nominal size = 10%)

DGP 1					
		R=40	R=80	R=80	R=80
statistic	*bootstrap approach*	P=80	P=40	P=80	P=120
MSE-F	non-parametric	.348	.299	.424	.544
MSE-F	restricted VAR	.842	.726	.864	.930
MSE-F	fixed regressor	.636	.584	.738	.843
MSE-t	non-parametric	.406	.360	.466	.580
MSE-t	restricted VAR	.801	.559	.775	.882
MSE-t	fixed regressor	.508	.366	.541	.668
CW t-test	restricted VAR	.895	.753	.934	.978
DGP 2					
		R=40	R=80	R=80	R=80
statistic	*bootstrap approach*	P=80	P=40	P=80	P=120
MSE-F	non-parametric	.477	.330	.587	.792
MSE-F	restricted VAR	.948	.844	.964	.991
MSE-F	fixed regressor	.774	.676	.870	.961
MSE-t	non-parametric	.569	.452	.672	.843
MSE-t	restricted VAR	.960	.780	.957	.992
MSE-t	fixed regressor	.684	.488	.752	.896
CW t-test	restricted VAR	.992	.931	.998	1.000

Notes:
1. See the notes to Table 1.
2. In these experiments, the coefficients $b_{ij} = 0$ are set to empirically-based estimates, which are 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

		Bootstrap p–values		
alternative model variable	*MSE(null)/ MSE(altern.)*	*non-param.*	*restricted VAR*	*fixed regressor*
cross-sectional premium	1.009	.144	.011	.073
return on long-term Treasury	1.005	.380	.034	.170
BAA-AAA yield spread	.996	.691	.619	.492
BAA-AAA return spread	.995	.809	.775	.768
net equity expansion	.994	.658	.785	.656
CPI inflation	.993	.653	.871	.771
stock variance	.992	.736	.761	.242
dividend-payout ratio	.991	.677	.915	.724
term (yield) spread	.987	.737	.970	.987
earnings-price ratio	.985	.968	.919	.934
10-year earnings-price ratio	.983	.884	.969	.983
3-month T-bill rate	.982	.742	.991	.991
dividend-price ratio	.981	.836	.948	.995
dividend yield	.981	.832	.989	.997
yield on long-term Treasury	.978	.810	.995	.995
book-market ratio	.965	.998	.998	.996

Notes:

1. As described in section 5, 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 3 details the bootstrap methods. The RMSE of the null model is 0.046.

Table 6: Tests of Equal Accuracy for Core Inflation

		Bootstrap p–values		
alternative model variables	*MSE(null)/ MSE(altern.)*	*non-param.*	*restricted VAR*	*fixed regressor*
1-quarter horizon				
CFNAI	1.020	.376	.040	.251
CFNAI, food, imports	1.091	.156	.001	.084
4-quarter horizon				
CFNAI	.920	.711	.919	.894
CFNAI, food, imports	1.257	.235	.001	.053

Notes:
1. As described in section 5, 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 5). The 1-quarter forecasts are of quarterly inflation; the 4-quarter forecasts are of 4-quarter inflation. Forecasts from 1985:Q1 + $\tau - 1$ through 2007: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 3 details the bootstrap methods. The RMSE of the null model is 0.626 at the 1-quarter horizon and 0.451 at the 4-quarter horizon.