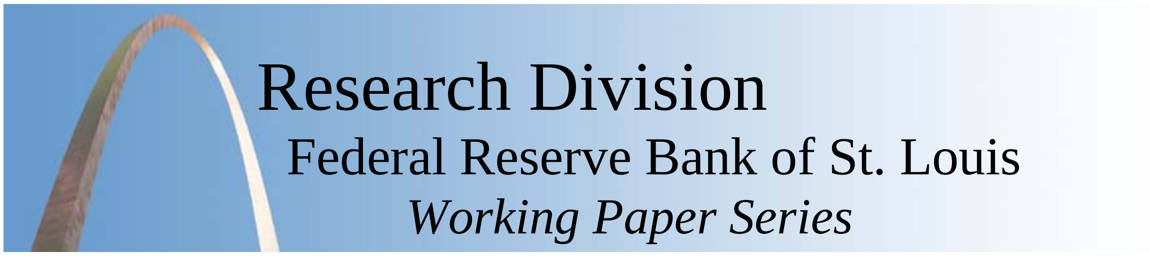

In-Sample Tests of Predictive Ability: A New Approach

Todd E. Clark
and
Michael W. McCracken

Working Paper 2009-051A
http://research.stlouisfed.org/wp/2009/2009-051.pdf

October 2009

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
P.O. Box 442
St. Louis, MO 63166

In-Sample Tests of Predictive Ability: A New Approach *

Todd E. Clark
Federal Reserve Bank of Kansas City

Michael W. McCracken
Federal Reserve Bank of St. Louis

July 2009

Abstract

This paper presents analytical, Monte Carlo, and empirical evidence linking in-sample tests of predictive content and out-of-sample forecast accuracy. Our approach focuses on the negative effect that finite-sample estimation error has on forecast accuracy despite the presence of significant population-level predictive content. Specifically, we derive simple-to-use in-sample tests that test not only whether a particular variable has predictive content but also whether this content is estimated precisely enough to improve forecast accuracy. Our tests are asymptotically non-central chi-square or non-central normal. We provide a convenient bootstrap method for computing the relevant critical values. In the Monte Carlo and empirical analysis, we compare the effectiveness of our testing procedure with more common testing procedures.

JEL Nos.: C53, C52, C12
Keywords: predictability, forecast accuracy, in-sample

* *Clark*: Economic Research Dept.; Federal Reserve Bank of Kansas City; 1 Memorial Drive; Kansas City, MO 64198; todd.e.clark@kc.frb.org. *McCracken (corresponding author)*: Economic Research Dept.; Federal Reserve Bank of St. Louis; P.O. Box 442 ; St. Louis, MO 63166; michael.w.mccracken@stls.frb.org. We gratefully acknowledge Lutz Kilian for helpful comments and Amit Goyal for making data available. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Kansas City, Federal Reserve Bank of St. Louis, or the Federal Reserve System.

1 Introduction

It is well known that in-sample evidence of predictive content of a particular variable – for example the significance of a t-statistic – frequently does not imply an improvement in forecast accuracy. Often it is the case that a model that excludes that variable provides lower mean square errors in an out-of-sample exercise. There are many possible reasons for such to occur, including the presence of unmodeled structural change (Clark and McCracken, 2005) and data-snooping (White, 2000).

Another well-known reason is overfitting, or to be more precise, the absence of enough data to precisely estimate model parameters prior to forecasting. In fact, this is the primary motivation for the development of information criteria. One punchline of this literature is that in many practical situations, estimating additional parameters can raise the forecast error variance above what might be obtained with a simple model. Such is clearly true when the additional parameters have population values of zero. But the same can apply even when the population values of the additional parameters are non–zero, if the additional explanatory power associated with the additional parameters is low enough. In such cases, in finite samples the additional parameter estimation noise may raise the forecast error variance more than including information from additional variables lowers it.

As this discussion suggests, parameter estimation noise creates a forecast accuracy trade-off. Excluding some variables that truly belong in the model could adversely affect forecast accuracy. Yet including the variables could raise the forecast error variance if the associated parameters are estimated sufficiently imprecisely. Surprisingly, however, this well-known fact is rarely taken into account when examining putative predictive content. In other words, while the t-statistics that are commonly reported based upon in-sample fit provide information regarding whether or not the variable has a non-zero signal, it does not neccesarily convey whether or not that signal is estimated precisely enough to improve forecast accuracy relative to a model that excludes it.

Accordingly, this paper presents analytical, Monte Carlo, and empirical evidence on in-sample tests of predictive content from nested models — for tests that take account of the bias-variance trade-off described above. These tests are straightforward extensions of the standard F-, t- and HAC-robust Wald tests for predictability. In applications when the forecast errors are 1-step ahead and form a conditionally homoskedastic martingale difference sequence, the F- and HAC-robust Wald tests can be used directly, but with critical values taken from the non-central — rather than the more typical central — χ^2 distribution. Compared to the usual approach based on standard (central) distributions, our suggested approach raises the bar for including a predictor in the estimated forecasting model. In the simplest case where one additional predictor is being considered at a 10 percent significance level, the usual Wald test statistic would be compared against a non–central χ^2 critical value of 5.217 instead of the central χ^2 critical value of 2.706 ($= 1.645^2$). In the same environment, when using the standard t-statistic, critical values can be taken from standard normal tables after centering the t-statistic based upon the assumed sign of the relevant coefficient.

In more general environments that allow multi-step forecasts and that have forecast errors that are conditionally heteroskedastic, we are able to establish a similar result in the scalar case where there is a single additional predictor in the unrestricted model. We are unable to establish a comparable result in the non-scalar case when either conditional heteroskedasticity or serial correlation are present in the model errors. In those cases for which our theory applies, we are able to establish the asymptotic validity of a novel, simple bootstrap procedure that approximates the relevant critical values. Our Monte Carlo results indicate that, in some practical settings, bootstrap inference is more reliable than inference based on asymptotic critical values (from non-central distributions).

Our results are most closely related to those in Trenkler and Toutenberg (1992). These authors begin by deriving the difference in mean square forecast error, $E(\hat{u}_{1,T+1}^2 - \hat{u}_{2,T+1}^2)$, between two nested classical normal linear regression models, each estimated by OLS using data available at time T. They proceed to show that for this difference to be zero, not only do the additional predictors in the unrestricted model have to have some predictive content (i.e., not all the additional coefficients are zero), but also the non-centrality parameter from the associated F-statistic has to equal 1. They conclude that if one wants to construct an in-sample test of equal out-of-sample forecast accuracy, the standard F-statistic can be constructed as is typical but critical values need to be taken from a non-central F-distribution with non-centrality parameter equal to 1 for accurate inference. A related result is discussed in Toro-Vizcarrondo and Wallace (1968).

Note, however, that this existing result requires strong assumptions on the type of data being used. The predictors must be strictly exogenous, and the model errors must be conditionally homoskedastic, serially uncorrelated, and normally distributed. Any one of these assumptions precludes application to a wide range of macroeconomic and financial data. To get around this issue, we provide asymptotically valid results that can be used to approximate the finite-sample problem. Our analytics are based on models we characterize as "weakly nested": the unrestricted model is the true model, but as the sample size grows large, the DGP converges to the restricted model. This analytic approach captures the practical reality that, in many instances, the predictive content of some variables of interest is quite low. Admittedly, by taking this approach, we lose the "exact" finite-sample size results available from the Trenkler and Toutenberg result. However, we gain applicability to enviroments that are relevant to most applied research. Our Monte Carlos experiments suggest that our asymptotic approximation yields accurate inference in many cases.

Another closely related result is discussed in Torous and Valkanov (2000). There the authors provide analytical evidence on the "right" way to model the signal-to-noise ratio in a noisy predictive regression: one in which the dependent variable is highly variable and the conditioning variable has a very low signal. Loosely speaking, they argue that the signal-to-noise ratio should be modeled as being on the order of T^{-2a} for some $a \geq 0$. With this parameterization in hand they show that if $0 \leq a < 1/2$ we should expect the predictor to be useful for forecasting in finite samples.[1] But if $a > 1/2$, we should not expect the predictor to be useful. The case in which $a = 1/2$ implies that, for the given sample size, the predictor is on the boundary of predictability. For a range of common predictors used to forecast monthly excess stock returns, they find that median unbiased estimates of a tend to lie very close to, or above, 1/2 and hence we should not be surprised that out-of-sample predictions of excess stock returns tend to do no better or worse than a simple random walk. In the language of their paper, our analytical results can be interpreted as providing a formal test for whether or not the predictor is on the boundary of predictability. Our approach though, is different from theirs – as will become clear in the following section.

The paper proceeds as follows. Section 2 provides theoretical results on in-sample tests of equal out-of-sample forecast accuracy as well as our suggested bootstrap approach to constructing critical values. In section 3 we present Monte Carlo evidence on the finite-sample effectiveness of our proposed testing procedure. Section 4 illustrates the use of the testing procedures in determining the predictability of stock returns. Section 5 concludes.

[1] For more detail the reader should reference the Torous and Valkanov paper directly. In this discussion, we are modifying their notation, and a bit of their interpretation, in order to simplify the comparison of their results to ours. In particular, we should note that they focus on the case in which the predictor is highly persistent whereas we restrict attention to predictors that are covariance stationary.

Proofs are provided in the Appendix.

2 Theory

We begin by laying out the necessary notation and assumptions sufficient for our results. We then provide an analytical characterization of the bias-variance tradeoff, created by weak predictability, involved in choosing between restricted and unrestricted forecasts. Given that tradeoff, we then derive a test of equal predictive ability based on a null hypothesis under which the competing models yield equally accurate forecasts, in light of the parameter estimation error associated with estimating the coefficients on the weak predictors.

2.1 Environment

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

$$\begin{aligned} y_{T,t+\tau} &= x'_{T,2,t}\beta_{2,T} + u_{T,t+\tau} = x'_{T,1,t}\beta_1 + x'_{T,22,t}(T^{-1/2}\beta_{22}) + v_{T,t+\tau}, \quad (1) \\ Ex_{T,2,t}u_{T,t+\tau} &\equiv Eh_{T,t+\tau} = 0 \text{ for all } t = 1, ...T. \end{aligned}$$

Note that we allow the dependent variable $y_{T,t+\tau}$, the predictors $x_{T,2,t}$ and the error term $v_{T,t+\tau}$ to depend upon T, the final forecast origin. We make this explicit in the notation to emphasize that as the overall sample size is allowed to increase in our asymptotics, this parameterization affects their marginal distributions.

At the fixed forecast origin T, our forecasting agent observes the sequence $\{y_{T,t}, x'_{T,2,t}\}_{t=1}^{T}$. Forecasts of the scalar $y_{T,T+\tau}$, $\tau \geq 1$, are to be generated using a $(k \times 1, k = k_1 + k_2)$ vector of covariates $x_{T,2,t} = (x'_{T,1,t}, x'_{T,22,t})'$ and one of the linear parametric models $x'_{T,i,t}\beta_i$, $i = 1, 2$. The parameters are estimated using OLS (**Assumption 2**) and hence $\hat{\beta}_{i,T} = \arg\min_{\beta_i} \sum_{t=1}^{T-\tau} (y_{T,t+\tau} - x'_{T,i,t}\beta_i)^2$, $i = 1, 2$, for the restricted and unrestricted models, respectively. We denote the loss associated with the τ-step ahead forecast error as $\hat{u}^2_{T,i,T+\tau} = (y_{T,T+\tau} - x'_{T,i,T}\hat{\beta}_{i,T})^2$, $i = 1, 2$, for the restricted and unrestricted models, respectively. The model residuals, $\hat{v}_{T,i,t+\tau}$, $i = 1, 2$, $t = 1, T - \tau$, associated with the time T estimated restricted and unrestricted models are defined similarly.

The following additional notation will be used. Let $H_i(T) = (T^{-1} \sum_{t=1}^{T-\tau} x_{T,i,t}v_{T,t+\tau}) = (T^{-1} \sum_{t=1}^{T-\tau} h_{T,i,t+\tau})$, $B_i(T) = (T^{-1} \sum_{t=1}^{T-\tau} x_{T,i,t}x'_{T,i,t})^{-1}$, and $B_i = \lim_{T\to\infty}(Ex_{T,i,t}x'_{T,i,t})^{-1}$ for $i = 1, 2$. For $U_{T,t} = (h'_{T,2,t+\tau}, vec(x_{T,2,t}x'_{T,2,t})')'$, let $V = \sum_{l=-\tau+1}^{\tau-1} \Omega_{11,l}$, where $\Omega_{11,l}$ is the upper block-diagonal element of Ω_l defined below. We define the selection matrices $J = (I_{k_1 \times k_1}, 0_{k_1 \times k_2})'$ and $J_2 = (0_{k_2 \times k_1}, I_{k_2 \times k_2})'$ as well as the second moment matrices $F_2(T) = J_2' B_2(T) J_2$ and $F_2 = \lim_{T\to\infty}(Ex_{T,22,t}x'_{T,22,t} - Ex_{T,22,t}x'_{T,1,t}(Ex_{T,1,t}x'_{T,1,t})^{-1}Ex_{T,1,t}x'_{T,22,t})^{-1}$. Finally, let $W(1)$ denote a vector standard Normal random variate and $\hat{\sigma}_T^2 = T^{-1}\sum_{t=1}^{T-\tau}\hat{v}_{T,2,t+\tau}^2$.

To derive our results, we need two 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) $T^{-1}\sum_{j=1}^{[rT]} U_{T,j}U'_{T,j-l} \Rightarrow r\Omega_l$ 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) $\Omega_{11,l} = 0$ all $l \geq \tau$. (c) $\sup_{T\geq 1, s\leq T} E|U_{T,s}|^{2q} < \infty$ for some $q > 1$. (d) $U_{T,j} - EU_{T,j} = (h'_{T,2,j+\tau}, vec(x_{T,2,j}x'_{T,2,j} - Ex_{T,2,j}x'_{T,2,j})')'$ is a zero mean triangular array satisfying Theorem 1 of de Jong (1997).

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 $i \in (0, 0.5)$, $L = O(T^i)$. (c) The number of covariance terms $\bar{j}$ used to estimate the long-run variance V defined above satisfies $\tau - 1 \leq \bar{j} < \infty$.[2]

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 model 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. In (d) we impose the high level assumption that, in particular, $h_{T,2,t+\tau}$ satisfies Theorem 1 of De Jong (1997). By doing so we insure (results needed in the Appendix) that certain scaled sample averages converge in distribution to normal random variates. Finally, Assumption 4 simply provides primitive conditions under which a nonparametric, kernel-based estimator $V(T) = \sum_{j=-\bar{j}}^{\bar{j}} K(j/M)(T^{-1}\sum_{s=1+j}^{T-\tau}\hat{v}_{T,i,s+\tau}\hat{v}_{T,i,s+\tau-j}x_{T,2,s}x'_{T,2,s-j})$ of the long-run variance matrix V will be consistent

With these assumptions in hand we first lay the groundwork for our tests. As noted in the introduction, Trenkler and Toutenberg (1992) derive an exact, finite-sample result

[2]In our Monte Carlo simulations and empirical work we use a Newey and West (1987) kernel with bandwidth 0 for horizon = 1 and bandwidth 1.5*horizon otherwise.

relating the non-centrality parameter associated with the F-statistic to the null hypothesis H_0: $E(\hat{u}^2_{1,T+1} - \hat{u}^2_{2,T+1}) = 0$. However, their result requires classical normal regression assumptions in order to hold. Our goal is to obtain a similar result but under significantly weaker assumptions on the data. To simplify our notation, we omit the additional "T" subscript, associated with the triangular array nature of the observables, unless necessary to avoid confusion.

2.2 Testing for equal forecast accuracy

We first establish an asymptotic approximation to the expected loss differential, in the following Proposition.

Proposition 1: Maintain Assumptions 1-3. Then $\lim_{T\to\infty} T \cdot E(\hat{u}^2_{1,T+\tau} - \hat{u}^2_{2,T+\tau}) = \beta'_{22}F_2^{-1}\beta_{22} - tr((-JB_1J' + B_2)V)$.

The expected loss differential is comprised of two components. The first, $\beta'_{22}F_2^{-1}\beta_{22}$, captures the marginal increase in mean square error that arises due to the omitted variable bias when estimating the restricted model. The second, $tr((-JB_1J' + B_2)V)$, captures the marginal increase in mean square error that arises due to the imprecise estimation of the weak predictor in the unrestricted model. The tradeoff between these two components suggests a test for equal predictive ability that accounts for the weak predictive ability associated with the additional predictors in the unrestricted model. To see this, note that if we set this expectation to zero and rearrange terms we obtain

$$\beta'_{22}F_2^{-1}\beta_{22} = tr((-JB_1J' + B_2)V), \tag{2}$$

which simply states that the non-centrality parameter associated with the asymptotic distribution of the standard F-test (essentially the LHS of the equation) takes a particular value that depends upon the second moments of the data through B_1, B_2, and V – each of which is consistently estimable. Note that for conditionally homoskedastic 1-step ahead forecasts, this restriction simplifies to

$$\beta'_{22}F_2^{-1}\beta_{22} = \sigma^2 k_2 \tag{3}$$

With this in mind, we consider whether three commonly used tests of predictive ability can be used to test the null of equal predictive ability accounting for parameter estimation error: the standard F-test $GC(T) = T\hat{\beta}'_{2,T}J_2(\hat{\sigma}_T^{-2}F_2^{-1}(T))J'_2\hat{\beta}_{2,T}$, its HAC-robust Wald version $GC'(T) = T\hat{\beta}'_{2,T}J_2(J'_2B_2(t)V(T)B_2(T)J_2)^{-1}J'_2\hat{\beta}_{2,T}$, and the HAC-robust t-test $t(T) = T^{1/2}J'_2\hat{\beta}_{2,T}/(J'_2B_2(t)V(T)B_2(T)J_2)^{1/2}$. In this analysis, note that the Spectral Decomposition Theorem implies $M = (-F_2^{1/2}Ex_{22,t}x'_{1,t}B_1, F_2^{1/2})V(-F_2^{1/2}Ex_{22,t}x'_{1,t}B_1, F_2^{1/2})' = DAD'$, for a $(k_2 \times k_2)$ orthonormal matrix D and diagonal matrix A of eigenvalues associated with M.

If we define $\lim_{T\to\infty} Eu^2_{T,t+1} = \sigma^2$ and continue to maintain Assumptions 1-4 we obtain the following Proposition.

Proposition 2: Maintain Assumptions 1-4. Then $GC(T) \to^d (W(1) - A^{-1/2}D'F_2^{-1/2}\beta_{22})' \times A(W(1) - A^{-1/2}D'F_2^{-1/2}\beta_{22})$ and $GC'(T) \to^d \chi^2(k_2, \Lambda)$, $\Lambda = \beta'_{22}(J'_2B_2VB_2J_2)^{-1}\beta_{22}$.

Proposition 2 characterizes the limiting distribution of the standard F-test, as well as the autocorrelation and (conditional) heteroskedasticity-robust Wald statistic, allowing for weak predictors. When we allow the model errors to be conditionally heteroskedastic or serially correlated, $GC(T)$ is asymptotically mixed non-central χ^2, while $GC'(T)$ is asymptotically non-central χ^2. Unfortunately, in both cases, this result is not immediately useful for testing the null $\lim_{T\to\infty} T \cdot E(\hat{u}^2_{1,T+\tau} - \hat{u}^2_{2,T+\tau}) = 0$ or, to be precise, $\beta'_{22}F_2^{-1}\beta_{22} = tr((-JB_1J' + B_2)V)$. The problem is that the null implies a restriction that we cannot directly impose on either of the limiting distributions. In the former case, the distribution depends explicitly on each of the separate non-centrality parameters from the mixture rather than (say) some aggregate of these noncentrality parameters (like $\beta'_{22}F_2^{-1}\beta_{22}$). In the latter case, the non-centrality parameter bears no obvious relationship with $\beta'_{22}F_2^{-1}\beta_{22}$. In the following Corollary we see that under some important special cases, these distributions are in fact useful for testing the null hypothesis.

Corollary 1: Maintain Assumptions 1-4. (a) Let $u_{T,t+1}$ be conditionally homoskedastic. If $\lim_{T\to\infty} T \cdot E(\hat{u}^2_{1,T+1} - \hat{u}^2_{2,T+1}) = 0$ then both $GC(T)$ and $GC'(T) \to^d \chi^2(k_2, k_2)$. (b) Let $k_2 = 1$. If $\lim_{T\to\infty} T \cdot E(\hat{u}^2_{1,T+\tau} - \hat{u}^2_{2,T+\tau}) = 0$ then $GC(T) \to^d (tr((-JB_1J' + B_2)V)/\sigma^2) \cdot \chi^2(1, 1)$ and $GC'(T) \to^d \chi^2(1, 1)$.

In part (a) of the corollary, when we impose the restriction that the model errors are conditionally homoskedastic and from 1-step ahead forecasts, both asymptotic distributions depend explicitly upon $\beta_{22}'F_2^{-1}\beta_{22}$ — a quantity whose value is restricted to be $tr((-JB_1J'+B_2)V)=\sigma^2k_2$ under the null. Similarly, in part (b) of the corollary, each of the distributions in Proposition 2 simplify significantly to ones that can easily be used to provide critical values for asymptotically valid inference.

Accordingly, motivated by the principle of parsimony that often applies in effective forecasting, we suggest a non–central testing approach that takes equal predictive ability as the null hypothesis, and selects the unrestricted model if the test rejects in the positive tail but otherwise selects the restricted model.[3] In particular, when the forecast errors are 1-step ahead and conditionally homoskedastic we suggest an approach of rejecting (at, say, the 10% level) the restricted model in favor of the unrestricted if $GC(T)$ (or $GC'(T)$) is larger than the 90%-ile associated with a $\chi^2(k_2,k_2)$ distribution rather than the more typical central $\chi^2(k_2)$ distribution. Moreover, when β_{22} is scalar and either the forecast errors are conditionally heteroskedastic or the forecast horizon is greater than one, we suggest an approach of rejecting the restricted model in favor of the unrestricted if $GC'(T)$ (or $(\hat{\sigma}_T^2/tr((-JB_1(T)J'+B_2(T))V(T)))\cdot GC(T))$ is larger than 5.217 – the 90%-ile associated with a $\chi^2(1,1)$ distribution rather than the more typical value of 2.706 – the 90%-ile associated with the central $\chi^2(1)$ distribution.

Critical values and p-values from the noncentral χ^2 distribution — while not as common as its centralized version — are available in many econometric software packages (including Matlab, Rats, R, and SAS). Another alternative is to directly simulate the critical values using the formulas provided in sources such as Imhof (1961). While the necessary asymptotic critical values are straightforward to obtain or construct, below we also discuss an alternative, bootstrap-based method for constructing estimates of the critical values. In finite samples, this bootstrap method may yield more accurate inference.

Although each of the $GC(T)$ and $GC'(T)$ statistics from Proposition 2 are commonly used to test for predictive ability, perhaps the most common is the simple t-statistic. This is certainly true when β_{22} is scalar but particularly true when one is willing to impose

[3] Depending on what alternatives are of interest, many other non–centrality parameters and model selection rules could be justified. For example, based on the null of equal predictive ability, which implies the MSE–minimizing forecast to be a simple average of the restricted and unrestricted forecasts, one might use the simple average as the default forecast and select the unrestricted forecast if the non–central test rejects. We considered some such alternatives in our Monte Carlo analysis, but from a forecast accuracy perspective, none seemed to offer any general advantages over the approaches for which we report results.

the additional restriction that β_{22} takes a particular sign, since by doing so there may be a gain in the power of the test.[4] Consider then the autocorrelation and (conditional) heteroskedasticity robust t-statistic.

Proposition 3: Maintain Assumptions 1-4. Then (a) $t(T) \to^d N(\Lambda^{1/2}, 1)$, where $\Lambda^{1/2} = \beta_{22}(J_2' B_2 V B_2 J_2)^{-1/2}$. (b) If $\lim_{T\to\infty} T \cdot E(\hat{u}^2_{1,T+\tau} - \hat{u}^2_{2,T+\tau}) = 0$, $t(T) \to^d N(sign(\beta_{22}), 1)$.

Proposition 3 characterizes the limiting distribution of the HAC–robust t-statistic. It is asymptotically normal with unit variance and non-zero mean that, when squared, gives us the noncentrality parameter associated with $GC'(T)$. While part (a) is unsurprising, in part (b) we find that regardless of the presence of conditional heteroskedasticity or autocorrelation in the forecast errors, if we are willing to assume the sign of β_{22} we can easily impose the null hypothesis. In particular, part (b) of Proposition 3 implies that $t(T) - sign(\beta_{22})$ is asymptotically standard normal under the null hypothesis. We can then readily test for equal predictive ability by rejecting the null hypothesis when $t(T) - sign(\beta_{22})$ is sufficiently small (when $\beta_{22} < 0$) or large (when $\beta_{22} > 0$) relative to standard normal critical values.

We should stress that our proposed tests based on non-central distributions address a null hypothesis that differs from the usual null hypothesis associated with tests compared against central distributions. Standard tests pertain to a null hypothesis of $\beta_{22} = 0$, which implies that, at the population level, the restricted and unrestricted models will be equally accurate for forecasting. Instead, our tests based on non-central distributions allow β_{22} to be non-zero, such that, in a finite sample, forecasts from the restricted and unrestricted models can be expected to be equally accurate.

Regardless of this non-standard null hypothesis – that $\lim_{T\to\infty} T \cdot E(\hat{u}^2_{1,T+\tau} - \hat{u}^2_{2,T+\tau}) = 0$ – we do expect our tests to have good power against the alternative that $\lim_{T\to\infty} T \cdot E(\hat{u}^2_{1,T+\tau} - \hat{u}^2_{2,T+\tau}) > 0$. To see this note that both the non–central chi–square and normal distributions have the monotone likelihood ratio property[5] and hence we can conclude that $GC'(T)$, $(\hat{\sigma}^2_T / tr((-JB_1(T)J' + B_2(T))V(T))) \cdot GC(T)$, and $t(T)$ are (asymptotically) uniformly most powerful against this alternative.

[4] The power advantages of such one-sided tests of predictive ability is compared to out-of-sample tests of predictive ability in Inoue and Kilian (2004).

[5] See Eaton (2007, pp. 465-470) for a proof using the non-central chi-square distribution and DeGroot (1984, p. 468) for a proof using the normal distribution.

2.3 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 parametric methods previously used in studies such as Kilian (1999). In Kilian's (1999) application, an appropriately dimensioned VAR was initially estimated by OLS imposing the restriction that β_{22} 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 the test statistic. The relevant sample percentile from this large collection of artificial statistics was then used as the critical value.

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_{22} 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 1-step ahead predictive ability implies equal predictive ability at longer horizons. Our present framework in no way imposes that restriction. We take an entirely different approach to imposing the relevant null hypothesis and generate the artificial samples.

To see the basis of our bootstrap, first note that with a minor rearrangement of terms, the null hypothesis imposes the restriction $\delta' J_2 F_2^{-1} J_2' \delta = tr((-JB_1J' + B_2)V)$ where $\delta = (0, \beta_{22}')'$. 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 $\beta_{22}'\beta_{22} = c$ for some finite constant — as one would in a ridge regression — we instead impose the restriction that $\delta' J_2 F_2^{-1}(T) J_2' \delta$ equals $tr((-JB_1(T)J' + B_2(T))V(T))$. In addition, we estimate δ using the approximation $\widehat{\delta} = (0, T^{1/2}\widetilde{\beta}_{22,t})'$ where $\widetilde{\beta}_{22,T}$ denotes the restricted least squares estimator of the parameters associated with the weak predictors satisfying

$$\begin{aligned} \widetilde{\beta}_{2,T} &= (\widetilde{\beta}_{1,T}', \widetilde{\beta}_{22,T}')' \\ &= \arg\min_{b_2} \sum\nolimits_{s=1}^{T-\tau} (y_{s+\tau} - x_{2,s}' b_2)^2 \qquad (4) \\ \text{s.t.} \quad & b_2' J_2 F_2^{-1}(T) J_2' b_2 = tr((-JB_1(T)J' + B_2(T))V(T))/T. \end{aligned}$$

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).

This approach to imposing the null hypothesis is directly comparable to the direct multi-step forecasting approach we assume is used to construct the forecasts, so the restriction can vary with the forecast horizon τ. This approach therefore precludes using a VAR and its recursive structure to generate the artificial samples. Instead we use a variant of the wild fixed regressor bootstrap developed in Goncalves and Kilian (2007) that accounts for the direct multi-step nature of the forecasts. Specifically, in our framework the x's are held fixed across the artificial samples and the dependent variable is generated using the direct multi-step equation $y^*_{s+\tau} = x'_{2,s}\widetilde{\beta}_{2,T} + \widehat{v}^*_{s+\tau}$, $s = 1, ..., T-\tau$, 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.[6]

1. (a) For the $GC(T)$ or $GC'(T)$ statistics, construct the parameter vector $\widetilde{\beta}_{2,T}$ associated with the unrestricted model using the weighted ridge regression from equation (4) above.

(b) For the $t(T)$ statistic, do the same but also imposing the restriction that $sign(\tilde{\beta}_{22,T}) = sign(\beta_{22})$.[7]

2. Using nonlinear least squares, estimate an $MA(\tau-1)$ model for the OLS residuals (from the unrestricted model) $\widehat{v}_{2,s+\tau}$ such that $v_{2,s+\tau} = \varepsilon_{2,s+\tau} + \theta_1\varepsilon_{2,s+\tau-1} + ... + \theta_{\tau-1}\varepsilon_{2,s+1}$. Let $\eta_{s+\tau}$, $s = 1, ..., T-\tau$, denote an $i.i.d$ $N(0,1)$ sequence of simulated random variables. Define $\widehat{v}^*_{2,s+\tau} = (\eta_{s+\tau}\widehat{\varepsilon}_{2,s+\tau} + \widehat{\theta}_1\eta_{s-1+\tau}\widehat{\varepsilon}_{2,s+\tau-1} + ... + \widehat{\theta}_{\tau-1}\eta_{s+1}\widehat{\varepsilon}_{2,s+1})$, $s = 1, ..., T-\tau$. Form artificial samples of $y^*_{s+\tau}$ using the fixed regressor structure $y^*_{s+\tau} = x'_{2,s}\widetilde{\beta}_{2,T} + \widehat{v}^*_{s+\tau}$.

3. Using the artificial data, construct an estimate of the various test statistics (e.g., $GC(T)$, $GC'(T)$, and $t(T)$) as if this were the original data.

[6] Our approach to generating artificial samples of multi-step forecast errors builds on a sampling approach proposed in Hansen (1996). Note also that while we use nonlinear least squares estimates of the MA model (for computational speed in our Monte Carlo), it would also be valid to use maximum likelihood estimation.

[7] Straightforward algebra provides a closed form solution for this estimator. If we let $\hat{\lambda}$ denote an estimate of the Lagrange multiplier and $C'_{12}(T) = J'B_2^{-1}(T)J_2$, we find that it satisfies $\tilde{\beta}_{2,T} = \begin{pmatrix} I & \frac{\hat{\lambda}}{1+\hat{\lambda}}B_1(T)C_{12}(T) \\ 0 & \frac{1}{1+\hat{\lambda}}I \end{pmatrix}\hat{\beta}_{2,T}$, where $\frac{1}{1+\hat{\lambda}} = \pm\sqrt{\frac{tr((-JB_1(T)J'+B_2(T))V(T))}{(T^{1/2}\hat{\beta}_{2,T})'J_2F_2^{-1}(T)J'_2(T^{1/2}\hat{\beta}_{2,T})}}$. For the $GC(T)$ and $GC'(T)$ statistics the sign of $\frac{1}{1+\hat{\lambda}}$ is asymptotically irrelevant since the asymptotic distribution only depends upon $(\frac{1}{1+\hat{\lambda}})^2$. For the $t(T)$ statistic the sign of $\frac{1}{1+\hat{\lambda}}$ matters for the asymptotic distribution. Valid inference requires using the value of $\frac{1}{1+\hat{\lambda}}$ that implies $sign(\tilde{\beta}_{22,T}) = sign(\beta_{22})$. Note that this implies that the correct sign of $\frac{1}{1+\hat{\lambda}}$ depends upon the sign of $\hat{\beta}_{22,T}$.

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 $(T^{1/2}\widetilde{\beta}_{22,T})$ of the local-to-zero parameters β_{22} satisfy our approximation to the null hypothesis. This is despite the fact that the estimates of β_{22} are not consistent. As we see below, our bootstrap is asymptotically valid in precisely those cases for which our theory applies despite the fact that we are unable to consistently estimate the local-to-zero parameters β_{22}. Before providing the result, we require a modest strengthening of the moment conditions in Assumption 3.

Assumption 3$'$: (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) $\mathrm{E}(\varepsilon_{2,s+\tau}|\varepsilon_{2,s+\tau-j}, x_{2,s-j}\ j \geq 0) = 0$. (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 $\widehat{\varepsilon}_{2,s+\tau}(\gamma_T) = \varepsilon_{2,s+\tau}$. In an open neighborhood N_T around γ_T, there exists a finite constant c such that $\sup_{1\leq s\leq T, T\geq 1} || \sup_{\gamma\in N_T}(\widehat{\varepsilon}_{2,s+\tau}(\gamma), \nabla\widehat{\varepsilon}'_{2,s+\tau}(\gamma), x_{T,2,s})'||_4 \leq c$. (d) $U_{T,j} - EU_{T,j} = (h'_{T,2,j+\tau}, vec(x_{T,2,j}x'_{T,2,j} - Ex_{T,2,j}x'_{T,2,j})')'$ is a zero mean triangular array satisfying Theorem 1 of de Jong (1997).

Assumption 3$'$ differs from Assumption 3 in two ways. First, in (b) it emphasizes the point that the forecast errors, and by implication $h_{2,t+\tau}$, form an $MA(\tau - 1)$ process. Second, in (c) it bounds the second moments not only of $h_{2,t+\tau} = (\varepsilon_{2,s+\tau} + \theta_1\varepsilon_{2,s+\tau-1} + ... + \theta_{\tau-1}\varepsilon_{2,s+1})x_{2,s}$ (as in Assumption 3) but also the functions $\widehat{\varepsilon}_{2,s+\tau}(\gamma)x_{T,2,s}$, and $\nabla\widehat{\varepsilon}_{2,s+\tau}(\gamma)x_{T,2,s}$ for all γ in an open neighborhood of γ_T. These assumptions are primarily used to show that the bootstrap estimator $V^*(T)$, which is a function of the estimated errors $\widehat{\varepsilon}_{2,s+\tau}$, is a consistent estimate of V. Such an assumption is not needed for showing that $V(T)$ is a consistent estimate of V since the model residuals $\widehat{v}_{2,s+\tau}$ are linear functions of $\hat{\beta}_{2,T}$ and Assumption 3 already imposes moment conditions on $\widehat{v}_{2,s+\tau}$ via moment conditions on $h_{2,s+\tau}$.

Proposition 4: Maintain Assumptions 1, 2, 3$'$, and 4. (a) Let $u_{T,t+1}$ be conditionally homoskedastic. If $\lim_{T\to\infty} T \cdot E(\hat{u}^2_{1,T+1} - \hat{u}^2_{2,T+1}) = 0$ then both $GC^*(T)$ and $GC'^*(T) \to^d \chi^2(k_2, k_2)$. (b) Let $k_2 = 1$. If $\lim_{T\to\infty} T \cdot E(\hat{u}^2_{1,T+\tau} - \hat{u}^2_{2,T+\tau}) = 0$ then $GC^*(T) \to^d (tr((-JB_1J' + B_2)V)/\sigma^2) \cdot \chi^2(1,1)$, $GC'^*(T) \to^d \chi^2(1,1)$, and $t^*(T) \to^d N(sign(\beta_{22}), 1)$.

In both (a) and (b) of Proposition 4, our fixed-regressor bootstrap provides an asymptotically valid method of estimating the critical values associated with the null of equal finite-sample forecast accuracy. In (a), we require that the forecast errors be 1-step ahead and conditionally homoskedastic. In (b), we allow serial correlation and conditional heteroskedasticity but require that β_{22} is scalar. While neither case covers the broadest situation in which β_{22} 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 1-step ahead forecasts of quarterly macroeconomic variables. Moreover, in many applications in which a nested model comparison is made (e.g., Goyal and Welch (2008)) the unrestricted forecasts are made by simply adding one lag of a single predictor to the baseline restricted model.

What Proposition 4 does not tell us is whether the proposed bootstrap is adequate for constructing asymptotically valid critical values under the alternative: that the unrestricted model will forecast more accurately than the restricted model. Unfortunately, there are any number of ways to model the case in which $\lim_{T\to\infty} T \cdot E(\hat{u}^2_{1,T+\tau} - \hat{u}^2_{2,T+\tau})$ is greater than zero. For example, rather than modeling the weak predictive ability in Assumption 1 as $T^{-1/2}\beta_{22}$ with $\lim_{T\to\infty} T \cdot E(\hat{u}^2_{1,T+\tau} - \hat{u}^2_{2,T+\tau}) = 0$, one could model the predictive content as $T^{-a}C\beta_{22}$ for constants $C < \infty$ and $a \in (0, 1/2]$ satisfying $\lim_{T\to\infty} T \cdot E(\hat{u}^2_{1,T+\tau} - \hat{u}^2_{2,T+\tau}) > 0$. While mathematically elegant, this approach does not allow us to analyze the most intuitive alternative in which not only $\lim_{T\to\infty} T \cdot E(\hat{u}^2_{1,T+\tau} - \hat{u}^2_{2,T+\tau}) > 0$, but also $J_2'\hat{\beta}_{2,T}$ is a consistent estimator of $\beta_{22} \neq 0$. For this situation to hold we need the additional restriction that $a = 0$ and hence β_{22} 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.

Proposition 5: Maintain Assumptions 1′, 2, 3′, and 4. (a) Let $u_{T,t+1}$ be conditionally homoskedastic. If $J_2'\hat{\beta}_{2,T} \to^p \beta_{22} \neq 0$ then both $GC^*(T)$ and $GC'^*(T) \to^d \chi^2(k_2, k_2)$. (b) Let $k_2 = 1$. If $J_2'\hat{\beta}_{2,T} \to^p \beta_{22} \neq 0$ then $GC^*(T) \to^d (tr((-JB_1J' + B_2)V)/\sigma^2) \cdot \chi^2(1, 1)$, $GC'^*(T) \to^d \chi^2(1, 1)$, and $t^*(T) \to^d N(sign(\beta_{22}), 1)$.

In Proposition 5 we see that, indeed, the bootstrap-based test is consistent for testing the null hypothesis $\lim_{T\to\infty} T \cdot E(\hat{u}^2_{1,T+\tau} - \hat{u}^2_{2,T+\tau}) = 0$ against the alternative that $J_2'\hat{\beta}_{2,T} \to^p \beta_{22} \neq 0$ (and hence $\lim_{T\to\infty} T \cdot E(\hat{u}^2_{1,T+\tau} - \hat{u}^2_{2,T+\tau}) > 0$). This follows since under this alternative, the data-based statistics $GC(T)$, $GC'(T)$, or $t(T)$ each diverge, while the bootstrap-based statistics $GC^*(T)$, $GC'^*(T)$, or $t^*(T)$ each retain the same asymptotic distribution they followed under the null.

3 Monte Carlo Evidence

To evaluate the finite-sample performance of the testing methods described above, we use Monte Carlo simulations of data-generating processes based on finance applications. In these experiments, the benchmark restricted predictive model for y includes just a constant; the unrestricted model adds a lag of other variables with potential predictive content. The general null hypothesis is that the unrestricted model won't predict $y_{T+\tau}$ more accurately than the restricted model does. This general null, however, can take different specific forms: either the additional variables in the unrestricted model have no predictive content, in that their coefficients are 0, such that, in a finite sample, the restricted forecast of $y_{T+\tau}$ is more accurate; or the coefficients are non-zero but small enough that restricted and restricted models are expected to predict $y_{T+\tau}$ equally accurately.

3.1 Monte Carlo design

For all DGPs, we generate data using independent draws of innovations from the normal distribution and the autoregressive structure of the DGP. We also consider a range of sample sizes (T), reflecting those commonly available in applications to (post-war) monthly financial data: 240, 360, 480, and 600. We consider prediction horizons of $\tau = 1$ and 12 months ahead. The DGPs are based on empirical relationships among excess stock returns and various predictors (over a sample of 1956 to 2002), taken from the data set of Goyal and Welch (2008).[8] In all cases, our reported results are based on 20,000 Monte Carlo draws and 999 bootstrap replications.

[8]We obtained the data from Amit Goyal's website. For convenience in reporting DGP coefficients, all variables were multiplied by 100 prior to estimation of the DGP.

3.1.1 DGPs

DGP 1 is based on the empirical relationship between excess returns ($y_{t+\tau}$) and net issuance of equity (x_t = NTIS in Goyal and Welch (2008)):

$$\begin{aligned} y_{t+\tau} &= bx_t + v_{t+\tau} \\ v_{t+\tau} &= \varepsilon_{t+\tau} + \theta_1\varepsilon_{t+\tau-1} + \cdots + \theta_{\tau-1}\varepsilon_{t+1} \\ x_{t+\tau} &= 0.95x_{t+\tau-1} + u_{t+\tau} \\ \text{var}\begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix} &= \begin{pmatrix} 18.0 & \\ 0.18 & 0.14 \end{pmatrix} \text{for } \tau = 1, \quad \begin{pmatrix} 25.0 & \\ 0.20 & 0.14 \end{pmatrix} \text{for } \tau = 12. \end{aligned} \tag{5}$$

In 1-step ahead experiments ($\tau = 1$), the residual in the DGP for $y_{t+\tau}$ is serially uncorrelated, so the MA coefficients $\theta_i = 0 \; \forall \; i$. In 12-step ahead experiments ($\tau = 12$), the residual in the DGP for $y_{t+\tau}$ follows an MA(11) process, with coefficients declining gradually from $\theta_1 = 0.95$ to $\theta_{11} = 0.6$, taken from empirical estimates of equations corresponding to the DGPs. Our specification of the 12-step model is based on a 12-month return (computed as a simple sum of 1-month returns from $t+1$ through $t+12$).

In DGP 1 experiments, the restricted and unrestricted prediction models take the following forms, respectively:

$$y_{t+\tau} = \alpha_0 + v_{1,t+\tau} \tag{6}$$

$$y_{t+\tau} = \alpha_0 + \alpha_1 x_t + v_{2,t+\tau} \tag{7}$$

We consider various experiments with different settings of b, the coefficient on x_t, which corresponds to the elements of our theoretical construct $\beta_{22}/\sqrt{T}$. In one set of simulations (Table 1), the coefficient is set to 0, such that the restricted model is expected to predict $y_{T+\tau}$ better than the unrestricted model. In others (Table 2), the coefficient is set to a value that means the models can be expected to be equally accurate for forecasting $y_{T+\tau}$. For example, with $T = 360$, we use $b = 0.187$ (compared to the empirical estimate of 0.23) at a prediction horizon of 1 month and $b = 1.987$ (compared to the empirical estimate of 1.81) at a prediction horizon of 12 months. In another set of experiments (Table 3), the coefficient is set to 0.3 in 1-step experiments and 2.0 in 12-step experiments, such that the unrestricted forecast is expected to be more accurate than the restricted forecast.

To verify that our parameterizations yield the intended patterns in predictive content, in unreported results we have checked the average across Monte Carlo draws of the difference in the squared forecast errors for period $T+\tau$ — specifically, the average across draws of

$\hat{v}^2_{1,T+\tau} - \hat{v}^2_{2,T+\tau}$. Consider the example of DGP 1 with $\tau = 1$ and $T = 240$, for which the population forecast error variance is 18.0. In the experiment with $b = 0$, the restricted model is, on average, more accurate than the restricted: the mean difference in squared forecast errors is -0.094, with a t-statistic (from the Monte Carlo sample) of -5.05. In the experiment with b set at 0.229 to make the models forecast equally well, the models are indeed equally accurate, on average: the mean difference in squared forecast errors is 0.019, with a t-statistic of 0.78. Finally, in the experiment with $b = 0.3$, the unrestricted model forecasts more accurately than the restricted model: the mean difference in squared forecast errors is 0.086, with a t-statistic of 4.03.

DGP 2 is based on the empirical relationship among excess returns ($y_{t+\tau}$), net issuance of stock ($x_{1,t}$), long-term bond returns ($x_{2,t}$), and the term spread ($x_{3,t}$) (respectively, NTIS, LTR, and TMS in Goyal and Welch (2008)):

$$\begin{aligned}
y_{t+\tau} &= b_1 x_{1,t} + b_2 x_{2,t} + b_3 x_{3,t} + v_{t+\tau} \\
v_{t+\tau} &= \varepsilon_{t+\tau} + \theta_1 \varepsilon_{t+\tau-1} + \cdots + \theta_{\tau-1}\varepsilon_{t+1} \\
x_{1,t+\tau} &= 0.95_1 x_{1,t+\tau-1} + u_{1,t+\tau} \\
x_{2,t+\tau} &= 0.07_2 x_{2,t+\tau-1} + u_{2,t+\tau} \\
x_{3,t+\tau} &= 0.95 x_{3,t+\tau-1} + u_{3,t+\tau}
\end{aligned} \tag{8}$$

$$\text{var}\begin{pmatrix} \varepsilon_t \\ u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{pmatrix} = \begin{pmatrix} 18.0 & & & \\ 0.18 & 0.14 & & \\ 2.24 & 0.03 & 7.29 & \\ 0.01 & 0.01 & -0.32 & 0.2 \end{pmatrix} \text{for } \tau = 1, \quad \begin{pmatrix} 25.0 & & & \\ 0.20 & 0.14 & & \\ 2.18 & 0.03 & 7.29 & \\ 0.03 & 0.01 & -0.32 & 0.2 \end{pmatrix} \text{for } \tau = 12.$$

In DGP 2 experiments, the restricted and unrestricted prediction models take the following forms:

$$y_{t+\tau} = \alpha_0 + v_{1,t+\tau} \tag{9}$$

$$y_{t+\tau} = \alpha_0 + \alpha_1 x_{1,t} + \alpha_2 x_{2,t} + \alpha_3 x_{3,t} + v_{2,t+\tau} \tag{10}$$

As in the case of DGP 1, in 1-step ahead experiments ($\tau = 1$), the residual in the DGP for $y_{t+\tau}$ is serially uncorrelated, so $\theta_i = 0 \; \forall \; i$. In 12-step ahead experiments ($\tau = 12$), the residual in the DGP for $y_{t+\tau}$ follows an MA(11) process, with coefficients declining gradually, taking the values described above for DGP 1.

As with DGP 1, we consider experiments with three different settings of the set of b_i coefficients, which correspond to the elements of $\beta_{22}/\sqrt{T}$. In one set of experiments (Table 1), all of the b_i coefficients are set to zero, such that the restricted model is expected to predict $y_{T+\tau}$ better than the unrestricted model does. In others (Table 2), empirically–based values of the b_i coefficients are multiplied by a constant less than one, such that, in population, the restricted and unrestricted models are expected to be equally accurate for forecasting $y_{T+\tau}$. For example, with $T = 360$, we use $b_1 = 0.18$, $b_2 = 0.09$, and $b_3 = 0.15$ at a horizon of 1 month (compared to empirical estimates of, respectively, 0.2, 0.14, and 0.25) and $b_1 = 1.24$, $b_2 = 0.50$, and $b_3 = 1.99$ at a horizon of 12 months (compared to empirical estimates of, respectively, 1.6, 0.7, and 2.5). In another set of experiments (Table 3), in which the unrestricted model is expected to forecast more accurately than the restricted model does, the coefficients are set to $b_1 = 0.3$, $b_2 = 0.15$, and $b_3 = 0.25$ in 1-step experiments and $b_1 = 1.5$, $b_2 = 0.6$, and $b_3 = 2.4$ in 12-step experiments.

3.2 Results

Tables 1 through 3 present results for our various Monte Carlo experiments. For DGP 1, for which the unrestricted model has one more variable than the restricted, we report results for both a two-sided Wald test and a one-sided t-test. For DGP 2, for which the unrestricted model has three extra variables, we only report results for the Wald test. The variances entering the test statistics incorporate the White (1980) correction for heteroskedasticity at the 1-month horizon and the Newey and West (1987) correction at the 12-month horizon (using a bandwidth of 18 lags). For each test statistic, we report rejection rates based on the central χ^2 (Wald) or normal (t-test) distributions, non-central χ^2 (Wald) or normal (t-test) distributions, and bootstrap (non-central) distributions. In light of the well-known problems in long-horizon regression inference associated with estimation of HAC variances (see, e.g., Hodrick (1992), Nelson and Kim (1993), and Kirby (1997)), we also report rejection rates based on a bootstrap of the central distributions.[9] We use a fixed regressor bootstrap, under the null that the coefficients on the additional variables in the larger model are zero. This bootstrap takes the same form as the non-central bootstrap detailed in section 2, modified to impose zero coefficients on the additional variables. We focus our presented results on 10 percent critical values; results are qualitatively similar at the 5 percent level.

[9]In applications satisfying the martingale difference sequence assumptions of Hodrick (1992), replacing the Newey and West (1987) variance estimate with that proposed by Hodrick (1992) should significantly improve the size performance of the tests based on asymptotic critical values. In practice, though, most researchers seem to use the Newey-West estimator.

3.2.1 DGPs with no predictive ability

Table 1 presents Monte Carlo results for DGPs in which, in truth, the x variables considered have no predictive content for y, such that, in a finite sample, the restricted forecasting model should be expected to forecast $y_{T+\tau}$ most accurately. These results generally line up with the expectations described above. At a prediction horizon of one month (left half of table), comparing Wald and t-tests against conventional critical values from the central χ^2 and normal distributions yields rejection rates of roughly the nominal size (10 percent), with rejection rates ranging from 9.2 to 12.2 percent across experiments and tests. The bootstrap of the central distribution yields very similar rejection rates. In contrast, comparing the Wald and t-tests against non-central χ^2 and normal distributions or our proposed bootstrap distribution yields much lower rejection rates, ranging from 0.7 to 2.5 percent. In the non-central case, results are very similar under the asymptotic (χ^2 and normal) and bootstrap distributions. Our proposed test approach lowers the rejection rates because the null of equal predictive ability implies (for a finite sample) non-zero coefficients on the x variables, and in fact the coefficients are zero in these experiments.

At a prediction horizon of 12 months, there remains a large qualitative difference in inference based on central distributions and inference based on non-central distributions. However, the well-known problems in long-horizon regression inference (see, e.g., Hodrick (1992), Nelson and Kim (1993), and Kirby (1997)) increases asymptotic rejection rates, particularly for tests compared to the usual central distributions. Using bootstrap critical values yields rejection rates significantly below those based on asymptotic critical values. Specifically, the Wald test compared against central χ^2 critical values is significantly oversized, with size ranging from 18.0 to 37.4 percent. The t-test compared against central normal critical values is also oversized, although not as badly as the Wald test. For example, in the DGP 1 experiment with $T = 360$, the sizes of the Wald and t-tests are, respectively, 20.3 and 14.4 percent. In testing against the central distributions, a bootstrap yields more accurate inference — indeed, tests that are about correctly sized. Across the Wald and t tests, central bootstrap–based size ranges from 8.6 to 11.1 percent.

Again, comparing the test statistics against non-central χ^2 and normal distributions yields much lower rejection rates. At the 12-step horizon, the tests are typically, although not always, undersized. For example, in the DGP 1 experiment with $T = 360$, comparing the Wald and t-tests against non-central χ^2 and normal distributions yields rejection rates of 8.6 and 3.7 percent, respectively. Comparing the test statistics against our non-central bootstrap approximation yields still-lower rejection rates, of 4.6 and 2.0 percent in the same example. With our non-central bootstrap, test rejection rates are below 10 percent in all of the 12-step experiments in Table 1, ranging from 1.9 to 5.1 percent (compared to the central bootstrap range of 8.6 to 11.1 percent).

3.2.2 DGPs with equal predictive ability

Table 2 presents results for DGPs in which the b_i coefficients on the x variables are non–zero but small enough that, under our asymptotic approximation, the restricted and unrestricted forecasting models are expected to be equally accurate for forecasting $y_{T+\tau}$. These results also generally line up with the expectations described above, and show clearly that, for testing the null of equal predictive ability, using our proposed bootstrap yields the most reliable inference. At the 1-month prediction horizon, comparing Wald and t-tests against critical values from the non-central bootstrap distribution yields rejection rates close to, although a bit below, the nominal size — specifically, rejection rates ranging from 7.9 to 9.0 percent.

At the 12-month horizon, using the bootstrap yields rejection rates that are modestly oversized, more so for DGP 2 than DGP 1 and more so for small samples than large. Across the 12-step experiments in Table 2, bootstrap rejection rates range from 12.6 to 18.1 percent. Much of the oversizing of the bootstrap with DGP 2 seems to stem from the persistence of some of the regressors. In unreported results in which we replaced the AR(1) coefficients of 0.95 with AR(1) coefficients of 0.5, the bootstrap yielded better-sized tests, ranging from 11.1 to 14.2 percent for the modified version of DGP 2 instead of the range of 16.0 to 18.1 percent shown in Table 2 for DGP 2. We should acknowledge, though, that our asymptotic results do not establish the validity of a non-central test for multi-step forecasts from DGP 2. With multi-step forecasts, our proposed tests are technically only valid with one extra variable in the unrestricted model; the unrestricted model in the DGP 2 experiments has three extra variables. Our test applied to multi-step forecasts from DGP 2 — compared against bootstrap critical values — seems to perform adequately, but our theoretical results do not provide a formal basis.

In the same experiments with DGPs satisfying the null of equal predictive ability, using the non-central χ^2 and normal distributions yields inference about as accurate as the bootstrap at the 1-month horizon, but significantly less accurate than the bootstrap at the

12-month horizon. At the 1-month prediction horizon, comparing Wald and t-tests against critical values from the non-central asymptotic distributions yields rejection rates ranging from 8.9 to 11.3 percent (compared to the bootstrap range of 7.9 to 9.0 percent). At the 12-month horizon, using critical values from the non-central asymptotic distributions yields rejection rates ranging from 18.2 to 49.3 percent (compared to the bootstrap range of 12.6 to 18.1 percent). In the asymptotic case, the over-sizing is much more severe for DGP 2 than DGP 1. The better performance of the bootstrap likely reflects well-known difficulties in estimating the HAC variance in finite samples (see, e.g., Hodrick (1992)), and the success of the bootstrap in capturing the finite-sample impact of HAC imprecision.[10]

In contrast, comparing the Wald and t-tests against critical values from the central χ^2 and Wald distributions consistently and significantly overstates the evidence of predictability: even though the models are equally accurate in forecasting $y_{T+\tau}$, these inference approaches reject the null with a frequency well in excess of 10 percent.[11] At the 1-month prediction horizon (left half of Table 2), rates of rejection based on central bootstrap distributions range from 22.5 to 37.6 percent (rates based on the central asymptotic distributions are very similar). At the 12-month prediction horizon, central bootstrap-based rejection rates are similarly high, ranging from 22.6 to 42.6 percent. At the 12-month horizon, using asymptotic critical values from the central distributions yields rejection rates significantly higher than those obtained with bootstrapped critical values, peaking at 75.8 percent, due to the HAC estimation-related size distortions discussed above.

3.2.3 DGPs with strong predictive ability

Table 3 provides results for DGPs in which the b_i coefficients on the x variables are large enough that, under our asymptotics, the unrestricted model is expected to predict $y_{T+\tau}$ more accurately than the restricted model does. These results confirm that, when the unrestricted model is the more accurate, tests based on non-central distributions have power, more so the larger the sample size.[12] For example, with DGP 1 and a 1-month ahead

[10] The modest oversizing of our non-central bootstrap-based tests in the multi-step case likely stems from a tendency to understate the HAC variance V that enters the rescaling calculations (the oversizing of tests compared against central asymptotic distributions indicates V is biased downward). Understating V causes the signal-noise ratio to be over-estimated, which results in the bootstrap DGP coefficients being scaled down modestly too much (in finite samples).

[11] Consistent with the findings of Inoue and Kilian (2004), in the DGP 1 experiments the one-sided t-test compared against central distributions rejects the null of no predictability at a higher frequency than does the (two-sided) Wald test.

[12] In the case of non-central tests, the one-sided t-test and two-sided Wald test have very similar power because the squared percentiles of the non-central normal and the percentiles of the chi square distributions

prediction horizon, t-test rejection rates based on the non-central bootstrap distribution rise from 13.0 percent at $T = 240$ to 36.7 percent at $T = 600$. As expected, tests based on central distributions yield higher rejection rates, rising (in the case of the t-test and the bootstrap distribution) from 43.9 percent at $T = 240$ to 73.0 percent at $T = 600$.

4 Applications

To illustrate the use of our proposed testing methods, in this section we apply them to models of excess stock returns. Some recent examples from the long literature on stock return forecasting include Rapach and Wohar (2006), Ang and Bekaert (2007), Campbell and Thompson (2008), and Goyal and Welch (2008).

In our application, we use the data of Goyal and Welch (2008) to examine the predictability of excess returns (from CRSP, measured on a log basis) at horizons of 1 and 12 months. We construct the 12-month return as a simple sum of the monthly return variable provided in the Goyal-Welch dataset. The restricted model includes just a constant. The unrestricted models add in one lag of a predictor, taken from the set of variables in the Goyal-Welch data set. We consider 16 possible predictors of returns, including, among others, net equity expansion, an interest rate term spread, the dividend-price ratio, and the cross-sectional premium. The full set of 16 predictive variables is listed in Table 4, 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 is January 1956 to December 2002.

In our assessment of predictability, we use the coefficient signs considered in Campbell and Thompson (2008) and Goyal and Welch (2008). For simplicity, we have done so by multiplying by -1 those variables for which the coefficients should be expected to be negative (Amit Goyal kindly provided us with the list: NTIS, LTY, TBL, INFL, and D/E). As a result, all of the estimated coefficients and t-statistics should be expected to be positive. In turn, we use one-sided (to the right) t-tests of the null of equal predictive ability. But using (two-sided) Wald tests yields very similar results.

Results for tests of stock return predictability are reported in Table 4. The second column reports t-statistics, which incorporate the White (1980) correction for heteroskedasticity at the 1-month horizon and the Newey-West variance estimate at the 12-month horizon

turn out to be virtually the same (with $k_2 = 1$ as it is in DGP 1, the sample Wald test is the square of the sample t-test).

(using a bandwidth of 18 lags). The remaining columns report p-values computed under alternative distributions (all one-sided): (1) the usual, central, normal distribution that is appropriate under the null of no predictive ability; (2) a fixed regressor bootstrap of the central distribution (computed with 9999 replications); (3) the non-central normal distribution that is appropriate under our asymptotics and the null of equal predictive ability; and (4) our bootstrap of the non-central distribution that applies under our asymptotics (computed with 9999 replications).

Based on conventional tests of the null of no predictability, the evidence in Table 4 is broadly consistent with much of the literature: based on regression tests, there appears to be some evidence of predictability. At the 1-month horizon, seven variables appear to have predictive content for excess returns (based on 10 percent significance). At the 12-month horizon, seven variables are significant under asymptotic critical values, and three are significant under the more reliable (central) bootstrap critical values. Under bootstrap critical values, at both horizons, net equity expansion, the long-term bond return, and the term spread appear to have significant predictive content for excess returns.

However, using our proposed testing methodology yields very little evidence of predictability. That is, there is little evidence to reject the null of equal predictive ability. At the 1-month horizon, none of the t-statistics are significant when compared against non-central normal or bootstrap distributions. At the 12-month horizon, only the long-term bond return appears to have significant predictive content when inference is based on the non-central normal or bootstrap distributions. The contrast in results based on our (non-central) asymptotic distributions versus conventional (central) distributions suggests that while several predictive variables might have regression coefficients that, in population, differ from 0, the coefficients are sufficiently small that, in the samples of data available, the coefficients are estimated with sufficient imprecision so as to make forecasts from a restricted model at least as accurate as forecasts from a true, unrestricted model.

5 Conclusion

As reflected in the principle of parsimony, when some variables are truly but weakly related to the variable being forecast, having the additional variables in the model may detract from forecast accuracy, because of parameter estimation error. Focusing on such cases of weak predictability, we show that standard in-sample tests of equal predictive ability fail to take account of estimation error.

We first derive, theoretically, tests that take account of the bias-variance trade-off associated with including the additional variables in the model. These tests are straightforward extensions of the standard F-, t- and HAC-robust Wald tests for predictability. In applications when the forecast errors are 1-step ahead and form a conditionally homoskedastic martingale difference sequence, the F- and HAC-robust Wald tests can be used directly, but with critical values taken from the non-central — rather than the more typical central — χ^2 distribution. Compared to the usual approach based on standard (central) distributions, our suggested approach raises the bar for including a predictor in the estimated forecasting model. In more general environments that allow multi-step forecasts and that have forecast errors that are conditionally heteroskedastic, we are able to establish a similar result in the scalar case where there is a single additional predictor in the unrestricted model. In those cases for which our theory applies, we are able to establish the asymptotic validity of a novel, simple bootstrap procedure that approximates the relevant critical values.

Monte Carlo experiments generally confirm our theoretical results. Specifically, in DGPs based on empirical finance applications, using our proposed non-central testing methods yield reliable inference in tests of the null of equal predictive ability. Our Monte Carlo results also indicate that, in some practical settings, bootstrap inference is more reliable than inference based on asymptotic critical values (from non-central distributions). An application to prediction of stock returns shows that conventional tests based on central distributions indicate some predictability to stock returns — in the sense that some predictors appear to have non-zero coefficients. However, our proposed tests systematically fail to reject the null of equal predictive ability: while some coefficients may be non-zero, they are small enough that a model restricted to include just a constant can be expected to forecast at least as well as any model including another one of the variables considered.

6 Appendix: Theory Details

Proof of Proposition 1: Straightforward algebra reveals that

$$
\begin{aligned}
&T(\hat{u}^2_{1,T+\tau} - \hat{u}^2_{2,T+\tau}) = \\
&2(T^{1/2}h'_{T,2,T+\tau})(-JB_1(T)J' + B_2(T))(T^{1/2}H_{T,2}(T)) \\
&-(T^{1/2}H'_{T,2}(T))(-JB_1(T)x_{T,1,T}x'_{T,1,T}B_1(T)J' + B_2(T)x_{T,2,T}x'_{T,2,T}B_2(T))(T^{1/2}H_{T,2}(T)) \\
&+2\delta' B_2^{-1}(T)(-JB_1(T)J' + B_2(T))(T^{1/2}h_{T,2,T+\tau}) \\
&+\delta'(x_{T,2,T}x'_{T,2,T} - 2x_{T,2,T}x'_{T,2,T}JB_1(T)J'B_2^{-1}(T) + B_2^{-1}(T)JB_1(T)x_{T,1,T}x'_{T,1,T}B_1(T)J'B_2^{-1}(T))\delta \\
&+2\delta'(B_2^{-1}(T)JB_1(T)x_{T,1,T}x'_{T,1,T}B_1(T)J' - x_{T,2,T}x'_{T,2,T}JB_1(T)J')(T^{1/2}H_{T,2}(T)).
\end{aligned}
$$

Taking expectations then gives

$$
\begin{aligned}
&TE(\hat{u}^2_{1,T+\tau} - \hat{u}^2_{2,T+\tau}) = \\
&E\{-(T^{1/2}H'_{T,2}(T))(-JB_1(T)x_{T,1,T}x'_{T,1,T}B_1(T)J' + B_2(T)x_{T,2,T}x'_{T,2,T}B_2(T))(T^{1/2}H_{T,2}(T)) \\
&+\delta'(x_{T,2,T}x'_{T,2,T} - 2x_{T,2,T}x'_{T,2,T}JB_1(T)J'B_2^{-1}(T) + B_2^{-1}(T)JB_1(T)x_{T,1,T}x'_{T,1,T}B_1(T)J'B_2^{-1}(T))\delta \\
&+2\delta'(B_2^{-1}(T)JB_1(T)x_{T,1,T}x'_{T,1,T}B_1(T)J' - x_{T,2,T}x'_{T,2,T}JB_1(T)J')(T^{1/2}H_{T,2}(T))\}.
\end{aligned}
$$

If we then note that Assumption 3 suffices for the observables to be uniformly integrable, taking limits gives us

$$
\begin{aligned}
&\lim_{T\to\infty} TE(\hat{u}^2_{1,T+\tau} - \hat{u}^2_{2,T+\tau}) = -tr[(-JB_1J' + B_2)\lim_{T\to\infty} E(T^{1/2}H_{T,2}(T))(T^{1/2}H'_{T,2}(T))] \\
&+\delta' B_2^{-1}(-JB_1J' + B_2)B_2^{-1}\delta \\
= \quad & -tr[(-JB_1J' + B_2)V] + \delta' B_2^{-1}(-JB_1J' + B_2)B_2^{-1}\delta.
\end{aligned}
$$

and the proof is complete.

Proof of Proposition 2: Assumption 3 suffices for $B_i(T) \to_p B_i$ $i = 1, 2$ and, along with de Jong (1997), is sufficient for $T^{1/2}H_{T,2}(T) \to_d V^{1/2}W(1)$. Continuity then implies

$$
\begin{aligned}
&T^{1/2}\widehat{\beta}_{22,T} \to_d N(\beta_{22}J_2'B_2VB_2J_2) \\
&T^{1/2}(J_2'\hat{\sigma}^2B_2(T)J_2)^{-1/2}\hat{\beta}_{22,T} \to_d N((J_2'\sigma^2B_2J_2)^{-1/2}\beta_{22}, (J_2'\sigma^2B_2J_2)^{-1/2}(J_2'B_2VB_2J_2)(J_2'\sigma^2B_2J_2)^{-1/2}) \\
&T^{1/2}(J_2'B_2(T)V(T)B_2(T)J_2)^{-1/2}\hat{\beta}_{22,T} \to_d N((J_2'B_2VB_2J_2)^{-1/2}\beta_{22}, I).
\end{aligned}
$$

Taking the inner product of the final two terms, as well as the definition of $M = DAD$, provides the desired result.

Proof of Corollary 1: Under the null hypothesis we know $tr((-JB_1J'+B_2)V) = \beta_{22}'F_2^{-1}\beta_{22}$. (a) Under conditional homoskedasticity however, since both $\beta_{22}'(J_2'B_2VB_2J_2)^{-1}\beta_{22} = \beta_{22}'F_2^{-1}\beta_{22}/\sigma^2$ and $tr((-JB_1J'+B_2)V) = \sigma^2k_2$ the proof is complete. (b) In the scalar case, note that $J_2'B_2VB_2J_2 = F_2 \cdot tr((-JB_1J'+B_2)V)$. But under the null hypothesis, $\beta_{22}^2F_2^{-1} = tr((-JB_1J'+B_2)V)$ and hence $\beta_{22}^2 = J_2'B_2VB_2J_2$ and the proof is complete.

Proof of Proposition 3: (a) This is immediate from the proof of Proposition 2. (b) As in the proof of Corollary 1 (b), under the null hypothesis we know $\beta_{22}^2 = J_2' B_2 V B_2 J_2$. Hence $\beta_{22} = sign(\beta_{22})(J_2' B_2 V B_2 J_2)^{1/2}$ and the proof is complete.

Throughout the remainder of the proofs define $v^*_{T,2,s+\tau} = (\eta_{s+\tau}\varepsilon_{T,2,s+\tau} + \theta_1\eta_{s-1+\tau}\varepsilon_{T,2,s+\tau-1} + ... + \theta_{\tau-1}\eta_{s+1}\varepsilon_{T,2,s+1})$, $\widehat{v}^*_{T,2,s+\tau} = (\eta_{s+\tau}\widehat{\varepsilon}_{T,2,s+\tau} + \widehat{\theta}_1\eta_{s-1+\tau}\widehat{\varepsilon}_{T,2,s+\tau-1} + ... + \widehat{\theta}_{\tau-1}\eta_{s+1}\widehat{\varepsilon}_{T,2,s+1})$, $h^*_{T,2,s+\tau} = x_{T,i,s}v^*_{T,2,s+\tau}$, $\widehat{h}^*_{T,2,s+\tau} = x_{T,i,s}\widehat{v}^*_{T,2,s+\tau}$, $H^*_{T,2}(T) = (T^{-1}\sum_{t=1}^{T-\tau} h^*_{T,2,s+\tau})$, and $\widehat{H}^*_{T,2}(T) = (T^{-1}\sum_{t=1}^{T-\tau}\widehat{h}^*_{T,2,s+\tau})$.

Lemma 1: Maintain Assumptions 1 or 1′, 2, 3′, and 4. (a) $T^{1/2}(\widehat{H}^*_{T,2}(T) - H^*_{T,2}(T)) = o_p(1)$. (b) $V^*(T) \to^p V$.

Proof of Lemma 1: For ease of presentation, we show both results assuming $\tau = 2$ and hence $\widehat{v}^*_{T,2,s+2} = \eta_{s+2}\widehat{\varepsilon}_{T,2,s+2} + \widehat{\theta}\eta_{s+1}\widehat{\varepsilon}_{T,2,s+1}$ and $v^*_{T,2,s+2} = \eta_{s+2}\varepsilon_{T,2,s+2} + \theta\eta_{s+1}\varepsilon_{T,2,s+1}$. (a) Rearranging terms gives us,

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

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

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

or equivalently

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

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 $O_p(1)$. In addition, since for large enough samples Assumption 3$'$ bounds the second moments of $\nabla\widehat{\varepsilon}_{T,2,s+2}(\overline{\gamma}_T)\mathrm{x}_{2,s}$ and $\nabla\widehat{\varepsilon}_{T,2,s+1}(\overline{\gamma}_T)\mathrm{x}_{2,s}$, the fact that the $\eta_{s+\tau}$ are $iidN(0,1)$ then implies $T^{-1}\sum_{s=1}^{T-\tau}\eta_{s+2}\nabla\widehat{\varepsilon}_{T,2,s+2}(\overline{\gamma}_T)x_{T,2,s}$, $T^{-1}\sum_{s=1}^{T-\tau}\eta_{s+1}\nabla\widehat{\varepsilon}_{T,2,s+1}(\overline{\gamma}_T)x_{T,2,s}$, and $T^{-1}\sum_{s=1}^{T-\tau}\eta_{s+1}\varepsilon_{T,2,s+1}x_{T,2,s}$ are all $o_p(1)$ and the proof is complete.

(b) First note that under our assumptions,

$$
\begin{aligned}
V &= \Gamma_0 + (\Gamma_1 + \Gamma_1') = \lim_{T\to\infty} E(\varepsilon_{T,2,s+2}^2 x_{T,2,s}x_{T,2,s}' + \theta^2\varepsilon_{T,2,s+1}^2 x_{T,2,s}x_{T,2,s}') \\
&+ (\theta E(\varepsilon_{T,2,s+1}^2 x_{T,2,s}x_{T,2,s-1}' + \varepsilon_{T,2,s+1}^2 x_{T,2,s-1}x_{T,2,s}')).
\end{aligned}
$$

Since $\overline{j}$ is finite and $\lim_{T\to\infty} K(j/L) = 1$ for each j, it suffices to show that

$$T^{-1}\sum_{s=1+j}^{T-\tau}\widehat{v}_{T,2,s+\tau}^*\widehat{v}_{T,2,s+\tau-j}^* x_{T,2,s}x_{T,2,s-j}' \to_p \Gamma_j \ \forall\ j$$

For all $j > \tau - 1$ this is trivial given the $iidN(0,1)$ nature of the increments $\eta_{s+\tau}$ and the fact that under our assumptions, $\Gamma_j = 0$ for these values of j. We will show this for $j = 1$, the case for $j = 0$ is similar.

Straightforward algebra, along with a first order Taylor expansion gives us

$$
\begin{aligned}
& T^{-1}\sum_{s=2}^{T-1}\widehat{v}_{T,2,s+2}^*\widehat{v}_{T,2,s+1}^* x_{T,2,s}x_{T,2,s-1}' \\
= \; & T^{-1}\sum_{s=2}^{T-1}(\eta_{s+2}\eta_{s+1}\widehat{\varepsilon}_{T,2,s+2}\widehat{\varepsilon}_{T,2,s+1})x_{T,2,s}x_{T,2,s-1}' + \widehat{\theta}T^{-1}\sum_{s=2}^{T-1}(\eta_{s+2}\eta_s\widehat{\varepsilon}_{T,2,s+2}\widehat{\varepsilon}_{T,2,s})x_{T,2,s}x_{T,2,s-1}' \\
& +\widehat{\theta}^2 T^{-1}\sum_{s=2}^{T-1}(\eta_{s+1}\eta_s\widehat{\varepsilon}_{T,2,s+1}\widehat{\varepsilon}_{T,2,s})x_{T,2,s}x_{T,2,s-1}' + \widehat{\theta}T^{-1}\sum_{s=2}^{T-1}\eta_{s+1}^2\varepsilon_{T,s+1}^2 x_{T,2,s}x_{T,2,s-1}' \\
& +\widehat{\theta}[(T^{1/2}(\widehat{\gamma}_T - \gamma_T))' \otimes (T^{1/2}(\widehat{\gamma}_T - \gamma_T))']T^{-2}\sum_{s=2}^{T-1}\eta_{s+1}^2 vec[\nabla\widehat{\varepsilon}_{T,2,s+1}(\overline{\gamma}_T)\nabla\widehat{\varepsilon}_{T,2,s+2}(\overline{\gamma}_T)']x_{T,2,s}x_{T,2,s-1}'
\end{aligned}
$$

for some $\overline{\gamma}_T$ in the closed cube with opposing vertices $\widehat{\gamma}_T$ and γ_T. Since $\widehat{\theta} \to^p \theta$, and Assumption 3$'$ suffices for $T^{-1}\sum_{s=2}^{T-1}\eta_{s+1}^2\varepsilon_{T,s+1}^2 x_{T,2,s}x_{T,2,s-1}' \to^p E\eta_{s+1}^2\varepsilon_{T,s+1}^2 x_{T,2,s}x_{T,2,s-1}' = E\varepsilon_{T,s+1}^2 x_{T,2,s}x_{T,2,s-1}'$, we know $\widehat{\theta}T^{-1}\sum_{s=2}^{T-1}\eta_{s+1}^2\varepsilon_{T,s+1}^2 x_{T,2,s}x_{T,2,s-1}' \to_p \Gamma_1$. The result will follow if the first three right-hand side terms, as well as the last, are all $o_p(1)$. In each case Assumption 3$'$ implies that the arguments of the summation are L^2-bounded and hence the iid $N(0,1)$ nature of the $\eta_{s+\tau}$ imply $T^{-1}\sum_{s=2}^{T-1}(\eta_{s+2}\eta_{s+1}\widehat{\varepsilon}_{T,2,s+2}\widehat{\varepsilon}_{T,2,s+1})x_{T,2,s}x_{T,2,s-1}'$, $T^{-1}\sum_{s=2}^{T-1}(\eta_{s+2}\eta_s\widehat{\varepsilon}_{T,2,s+2}\widehat{\varepsilon}_{T,2,s})x_{T,2,s}x_{T,2,s-1}'$, $T^{-1}\sum_{s=2}^{T-1}(\eta_{s+1}\eta_s\widehat{\varepsilon}_{T,2,s+1}\widehat{\varepsilon}_{T,2,s})x_{T,2,s}x_{T,2,s-1}'$, and $T^{-2}\sum_{s=2}^{T-1}\eta_{s+1}^2 vec[\nabla\widehat{\varepsilon}_{T,2,s+1}(\overline{\gamma}_T)\nabla\widehat{\varepsilon}_{T,2,s+2}(\overline{\gamma}_T)']$ $\times\ x_{T,2,s}x_{T,2,s-1}'$ are all $o_p(1)$. Since $\widehat{\theta} \to^p \theta$ and $T^{1/2}(\widehat{\gamma}_T - \gamma_T) = O_p(1)$ the proof is complete.

Proof of Proposition 4: We will provide the results for $GC'^*(T)$. The proofs for $GC^*(T)$ and $t^*(T)$ are very similar. Let $\hat{\beta}_{2,T}^*$ denote the OLS estimate of β_2 in the bootstrap sample. We begin by noting that $T^{1/2}\hat{\beta}_{2,T}^* = T^{1/2}J_2'[B_2(T)H_{T,2}^*(T) + \widetilde{\beta}_{2,T} + B_2(T)(\widehat{H}_{T,2}^*(T) - H_{T,2}^*(T))]$. Assumption 3$'$

and Lemma 1 imply $T^{1/2}(\widehat{H}^*_{T,2}(T) - H^*_{T,2}(T)) = o_p(1)$ and $J_2'B_2(T)V^*(T)B_2(T)J_2 \to^p J_2'B_2VB_2J_2$. $T^{1/2}J_2'B_2(T)H^*_{T,2}(T) \to^d N(0, J_2'B_2VB_2J_2)$, $T^{1/2}J_2'\widetilde{\beta}_{2,T} = O_p(1)$, and both of the latter two terms are asymptotically independent of one another due to the $i.i.d.N(0,1)$ nature of the bootstrap increments $\eta_{t+\tau}$. Hence conditional on $T^{1/2}J_2'\widetilde{\beta}_{2,T}$, $(J_2'B_2(T)V^*(T)B_2(T)J_2)^{-1/2}T^{1/2}J_2'[B_2(T)H^*_{T,2}(T) + \widetilde{\beta}_{2,T}]$ is asymptotically normal with unit variance and mean equal to $(J_2'B_2VB_2J_2)^{-1/2}[\lim_{T\to\infty} T^{1/2} \times J_2'\widetilde{\beta}_{2,T}]$. The result will follow if $[\lim_{T\to\infty} T^{1/2}J_2'\widetilde{\beta}_{2,T}]'(J_2'B_2VB_2J_2)^{-1}[\lim_{T\to\infty} T^{1/2}J_2'\widetilde{\beta}_{2,T}] = k_2$. We will show this under two sets of assumptions.

(a) Assume conditional homoskedasticity. Note that with this assumption, $(J_2'B_2VB_2J_2)^{-1} = \sigma^{-2}F_2^{-1}$. But the estimator for the bootstrap replication was constructed to satisfy the property $(T^{1/2}J_2'\widetilde{\beta}_{2,T})F_2^{-1}(T)(T^{1/2}J_2'\widetilde{\beta}_{2,T}) = \widehat{\sigma}^2 k_2$. Since $F_2^{-1}(T) \to^p F_2^{-1}$ and $\widehat{\sigma}^2_T \to^p \sigma^2$ the proof is complete.

(b) Assume that β_{22} is scalar. Note that under this assumption, $J_2'B_2VB_2J_2 = F_2 tr((-JB_1J' + B_2)V)$. But the estimator for the bootstrap replication was constructed to satisfy the property $(T^{1/2}J_2'\widetilde{\beta}_{2,T})^2 = F_2(T)tr((-JB_1(T)J' + B_2(T))V(T))$. Since $F_2(T) \to^p F_2$, $(J_2'B_2(T)V^*(T)B_2(T)J_2) \to^p (J_2'B_2VB_2J_2)$, and $tr((-JB_1(T)J' + B_2(T))V(T)) \to^p tr((-JB_1J' + B_2)V)$ we find

$$[\lim_{T\to\infty} T^{1/2}J_2'\widetilde{\beta}_{2,T}]'(J_2'B_2VB_2J_2)^{-1}[\lim_{T\to\infty} T^{1/2}J_2'\widetilde{\beta}_{2,T}] = 1$$

and the proof is complete.

Proof of Proposition 5: We will start by providing the results for $GC'^*(T)$ and then provide the result for $t^*(T)$. The proof for $GC^*(T)$ is very similar to that of $GC'^*(T)$. First note that by the definition of the ridge estimator (without the sign restriction) we have

$$\begin{aligned} T^{1/2}J_2'\widetilde{\beta}_{2,T} &= \pm\left(\frac{tr[(-JB_0(T)J' + B_1(T))V(T)]}{(T^{1/2}\hat{\beta}_{2,T})'J_2F_2^{-1}(T)J_2'(T^{1/2}\hat{\beta}_{2,T})}\right)^{1/2} J_2'(T^{1/2}\hat{\beta}_{2,T}) \\ &= \pm\left(\frac{tr[(-JB_0(T)J' + B_1(T))V(T)]}{\hat{\beta}_{22,T}'F_2^{-1}(T)\hat{\beta}_{22,T}}\right)^{1/2} \hat{\beta}_{22,T} \\ &\to^p \pm\left(\frac{tr[(-JB_0J' + B_1)V]}{\beta_{22}'F_2^{-1}\beta_{22}}\right)^{1/2} \beta_{22}. \end{aligned}$$

Since $T^{1/2}(\widehat{H}^*_{T,2}(T) - H^*_{T,2}(T)) = o_p(1)$ by Lemma 1 and $(J_2'B_2(T)V^*(T)B_2(T)J_2) \to^p (J_2'B_2VB_2J_2) > 0$ we also have

$$\begin{aligned} & T^{1/2}(J_2'B_2(T)V^*(T)B_2(T)J_2)^{-1/2}J_2'\hat{\beta}^*_{2,T} \\ = \; & T^{1/2}(J_2'B_2(T)V^*(T)B_2(T)J_2)^{-1/2}J_2'[B_2(T)H^*_{T,2}(T) + \widetilde{\beta}_{2,T} + B_2(T)(\widehat{H}^*_{T,2}(T) - H^*_{T,2}(T))] \\ \to^d \; & N(\pm(J_2'B_2VB_2J_2)^{-1/2}\left(\frac{tr[(-JB_0J' + B_1)V]}{\beta_{22}'F_2^{-1}\beta_{22}}\right)^{1/2} \beta_{22}, I). \end{aligned}$$

The result will follow if the inner product of the mean of the asymptotic distribution takes the value k_2.

(a) Assume conditional homoskedasticity. Note that with this assumption, $(J_2'B_2VB_2J_2)^{-1} = \sigma^{-2}F_2^{-1}$. Taking the inner product of the mean and noting $tr[(-JB_0J' + B_1)V] = \sigma^2 k_2$ gives us

$$\begin{aligned}
& \left(\frac{tr[(-JB_0J' + B_1)V]}{\beta_{22}'F_2^{-1}\beta_{22}}\right)\beta_{22}'(J_2'B_2VB_2J_2)^{-1}\beta_{22} \\
= & \left(\frac{\sigma^2 k_2}{\beta_{22}'F_2^{-1}\beta_{22}}\right)\beta_{22}'F_2^{-1}\beta_{22}/\sigma^2 = k_2
\end{aligned}$$

and the proof is complete.

(b) Assume that β_{22} is scalar. Note that under this assumption, $(J_2'B_2VB_2J_2)^{-1} = F_2^{-1}/tr((-JB_1J' + B_2)V)$. Taking the inner product of the mean then gives us

$$\begin{aligned}
& \left(\frac{tr[(-JB_0J' + B_1)V]}{\beta_{22}'F_2^{-1}\beta_{22}}\right)\beta_{22}'(J_2'B_2VB_2J_2)^{-1}\beta_{22} \\
= & \left(\frac{tr[(-JB_0J' + B_1)V]}{\beta_{22}'F_2^{-1}\beta_{22}}\right)\beta_{22}'F_2^{-1}\beta_{22}/tr((-JB_1J' + B_2)V) = 1
\end{aligned}$$

and the proof is complete.

We now show the result for $t^*(T)$. Note that with the sign restriction, the ridge estimator now takes the form $T^{1/2}J_2'\widetilde{\beta}_{2,T} = sign(\beta_{22})\left(\frac{tr[(-JB_0(T)J'+B_1(T))V(T)]}{(T^{1/2}\hat{\beta}_{2,T})'J_2F_2^{-1}(T)J_2'(T^{1/2}\hat{\beta}_{2,T})}\right)^{1/2}|J_2'(T^{1/2}\hat{\beta}_{2,T})|$. As above, this estimator is asymptotically normal with unit variance and mean equal to $(J_2'B_2VB_2J_2)^{-1/2} \times \left(\frac{tr[(-JB_0J'+B_1)V]}{\beta_{22}'F_2^{-1}\beta_{22}}\right)^{1/2}\beta_{22}$. The result will follow if this is equal to $sign(\beta_{22})$. But, as in case (b) above, we know that when β_{22} is scalar $(J_2'B_2VB_2J_2)^{-1} = F_2^{-1}/tr((-JB_1J' + B_2)V)$ and hence $(J_2'B_2VB_2J_2)^{-1/2}\left(\frac{tr[(-JB_0J'+B_1)V]}{\beta_{22}'F_2^{-1}\beta_{22}}\right)^{1/2}\beta_{22} = \left(\frac{F_2 tr[(-JB_0J'+B_1)V]}{J_2'B_2VB_2J_2}\right)^{1/2}\frac{\beta_{22}}{|\beta_{22}|} = sign(\beta_{22})$ and the proof is complete.

References

Ang, Andrew, and Geert Bekaert, (2007), "Stock Return Predictability: Is It There?" *Review of Financial Studies* 20, 651-707.

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 (2005), "The Power of Tests of Predictive Ability in the Presence of Structural Breaks", *Journal of Econometrics*, 124, 1-31.

DeGroot, Morris H. (1984), *Probability and Statistics*, Reading, MA: Addison-Wesley.

de Jong, Robert M. (1997), "Central Limit Theorems for Dependent Heterogeneous Random Variables", *Econometric Theory*, 13, 353-367.

Eaton, Morris L. (2007), *Multivariate Statistics: A Vector Space Approach*, Beachwood, OH: Institute of Mathematical Statistics.

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

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

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

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

Hodrick, Robert J. (1992), "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," *Review of Financial Studies* 5, 357-386.

Imhof, J.P. (1961), "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika* 48, 419-26.

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

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

Kirby, Chris (1997), "Measuring the Predictable Variation in Stock and Bond Returns," *Review of Financial Studies* 10, 579-630.

Nelson, Charles R., and Myung J. Kim (1993), "Predictable Stock Returns: The Role of Small Sample Bias," *Journal of Finance* 48, 641-61.

Newey, Whitney K., and Kenneth D. West (1987), "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703-08.

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.

Touros, Walter and Rossen Valkanov (2000), "Boundaries of Predictability: Noisy Predictive Regressions", manuscript, UCLA.

Toro-Vizcarrondo, Carlos and T. D. Wallace (1968), "A Test of the Mean Square Error Criterion for Restrictions in Linear Regression," *Journal of the American Statistical Association*, 63, 558-572.

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.

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

White, Halbert (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and Direct Test for Heteroskedasticity," *Econometrica* 48, 817-38.

Table 1: Monte Carlo Rejection Rates, Restricted Model Best

(nominal size = 10%)

		horizon = 1 month				horizon = 12 months			
test	*distribution*	T=240	T=360	T=480	T=600	T=240	T=360	T=480	T=600
DGP 1									
Wald	central chi-square	.103	.108	.100	.105	.229	.203	.193	.180
Wald	central bootstrap	.097	.103	.097	.103	.100	.099	.102	.099
Wald	non-central chi-square	.023	.024	.024	.025	.105	.086	.076	.068
Wald	non-central bootstrap	.020	.022	.021	.024	.051	.046	.045	.041
t	central normal	.092	.093	.092	.096	.157	.144	.138	.135
t	central bootstrap	.088	.089	.090	.094	.086	.088	.090	.093
t	non-central normal	.010	.009	.011	.011	.045	.037	.033	.031
t	non-central bootstrap	.009	.008	.010	.010	.023	.020	.020	.019
DGP 2									
Wald	central chi-square	.122	.115	.111	.104	.374	.312	.271	.249
Wald	central bootstrap	.109	.105	.104	.098	.107	.111	.107	.106
Wald	non-central chi-square	.014	.012	.010	.008	.154	.107	.081	.067
Wald	non-central bootstrap	.007	.008	.007	.007	.040	.035	.030	.028

Notes:
1. The data generating processes are defined in equations (5) and (8). In these experiments, the coefficients $b_i = 0$ for all i, such that the restricted forecasting model is expected to be more accurate for forecasting $y_{T+\tau}$.
2. For each artificial data set, we estimate equation (7) (DGP 1) or equation (9) (DGP 2) by OLS and form Wald and/or t-tests of the explanatory power of the x variables. T and refers to the size of the estimation sample.
3. In each Monte Carlo replication, the simulated test statistics are compared against critical values from central chi square and normal distributions, non-central chi square and normal distributions, and the bootstrap distribution generated as described in section 2.
4. The number of Monte Carlo simulations is 20,000; the number of bootstrap draws is 999.

Table 2: Monte Carlo Rejection Rates, Equally Accurate Models

(nominal size = 10%)

		horizon = 1 month				horizon = 12 months			
test	*distribution*	T=240	T=360	T=480	T=600	T=240	T=360	T=480	T=600
DGP 1									
Wald	central chi-square	.233	.241	.243	.248	.397	.372	.359	.350
Wald	central bootstrap	.225	.234	.237	.243	.226	.234	.233	.237
Wald	non-central chi-square	.089	.093	.091	.095	.235	.211	.192	.185
Wald	non-central bootstrap	.080	.085	.085	.090	.134	.131	.126	.126
t	central normal	.341	.350	.358	.363	.479	.463	.456	.448
t	central bootstrap	.330	.343	.352	.356	.341	.350	.354	.360
t	non-central normal	.089	.092	.090	.095	.229	.207	.188	.182
t	non-central bootstrap	.079	.084	.085	.090	.138	.133	.127	.128
DGP 2									
Wald	central chi-square	.386	.381	.389	.375	.758	.704	.668	.644
Wald	central bootstrap	.361	.365	.376	.365	.404	.418	.425	.426
Wald	non-central chi-square	.113	.106	.108	.098	.493	.412	.368	.334
Wald	non-central bootstrap	.081	.085	.089	.084	.181	.172	.162	.160

Notes:
1. See the notes to Table 1.
2. In these experiments, the coefficients $b_i = 0$ are scaled such that the restricted and unrestricted models are expected to equally accurate for forecasting $y_{T+\tau}$.

Table 3: Monte Carlo Rejection Rates, Unrestricted Model Best

(nominal size = 10%)

test	*distribution*	horizon = 1 month T=240	T=360	T=480	T=600	horizon = 12 months T=240	T=360	T=480	T=600
DGP 1									
Wald	central chi-square	.319	.431	.530	.613	.473	.546	.621	.688
Wald	central bootstrap	.309	.423	.524	.607	.290	.388	.482	.563
Wald	non-central chi-square	.144	.219	.297	.379	.301	.359	.426	.488
Wald	non-central bootstrap	.130	.208	.286	.366	.180	.248	.312	.378
t	central normal	.449	.566	.665	.736	.566	.648	.718	.783
t	central bootstrap	.439	.558	.657	.730	.424	.535	.628	.708
t	non-central normal	.144	.220	.298	.380	.298	.359	.426	.488
t	non-central bootstrap	.130	.208	.286	.367	.186	.253	.318	.382
DGP 2									
Wald	central chi-square	.570	.730	.847	.915	.750	.808	.860	.908
Wald	central bootstrap	.545	.714	.839	.911	.395	.553	.678	.778
Wald	non-central chi-square	.237	.390	.546	.676	.483	.549	.623	.698
Wald	non-central bootstrap	.188	.343	.510	.649	.176	.252	.336	.422

Notes:
1. See the notes to Table 1.
2. In these experiments, the coefficients b_i are set to values large enough that the unrestricted model is expected to forecasting $y_{T+\tau}$ more accurately than the restricted model does.

Table 4: Tests of Predictability of Excess Stock Returns, 1956-2002

1-month horizon

		p–values			
explanatory variable	*t-statistic*	*central normal*	*central bootstrap*	*non-central normal*	*non-central bootstrap*
net equity expansion (NTIS)	2.139	.016	.018	.127	.135
long-term return (LTR)	2.132	.017	.018	.129	.139
term spread (TMS)	1.990	.023	.023	.161	.166
dividend payout ratio (D/E)	-.543	.706	.705	.939	.940
stock variance (SVAR)	-1.313	.905	.843	.990	.979
default return spread (DFR)	.563	.287	.283	.669	.673
long-term yield (LTY)	.272	.393	.392	.767	.768
inflation (INFL)	.920	.179	.186	.532	.539
Treasury bill rate (TBL)	1.274	.101	.110	.392	.398
default yield spread (DFY)	1.761	.039	.041	.223	.232
dividend-price ratio (D/P)	1.334	.091	.097	.369	.387
dividend yield (D/Y)	1.387	.083	.085	.349	.363
earning-price ratio (E/P)	1.080	.140	.141	.468	.478
book to market (B/M)	.429	.334	.339	.716	.720
earning (10 year)-price ratio (E^{10}/P)	1.089	.138	.143	.465	.467
cross-sectional premium (CSP)	1.513	.065	.065	.304	.317

12-month horizon

		p–values			
explanatory variable	*t-statistic*	*central normal*	*central bootstrap*	*non-central normal*	*non-central bootstrap*
net equity expansion (NTIS)	1.711	.044	.096	.239	.303
long-term return (LTR)	4.023	.000	.000	.001	.002
term spread (TMS)	2.221	.013	.035	.111	.171
dividend payout ratio (D/E)	-.581	.719	.683	.943	.922
stock variance (SVAR)	.707	.240	.308	.615	.607
default return spread (DFR)	-.897	.815	.807	.971	.974
long-term yield (LTY)	-.427	.665	.627	.923	.872
inflation (INFL)	1.353	.088	.148	.362	.443
Treasury bill rate (TBL)	.669	.252	.305	.630	.601
default yield spread (DFY)	1.056	.145	.209	.477	.517
dividend-price ratio (D/P)	1.394	.082	.158	.347	.501
dividend yield (D/Y)	1.412	.079	.162	.340	.470
earning-price ratio (E/P)	1.378	.084	.139	.353	.427
book to market (B/M)	.517	.303	.340	.685	.718
earning (10 year)-price ratio (E^{10}/P)	1.208	.114	.175	.418	.469
cross-sectional premium (CSP)	-2.040	.979	.951	.999	.992

Notes:
1. As described in section 4, the null model for excess stock returns includes just a constant; the alternative models include a constant and the variable listed in the first column. The estimation sample is January 1956 to December 2002.
2. For each variable or alternative model, the table reports the HAC-robust t-statistic on the variable given in the first column and one-sided (to the right) p-values based on central and non-central distributions. Those variables for which the coefficients should be expected to be negative have been multiplied by -1, such that all estimated coefficients and t-statistics should be expected to be positive.