Asymptotic Inference for Performance Fees and the Predictability of Asset Returns

Michael W. McCracken
and
Giorgio Valente

Working Paper 2012-049B
http://research.stlouisfed.org/wp/2012/2012-049.pdf

October 2012
Revised October 2014
Asymptotic Inference for Performance Fees and the Predictability of Asset Returns

Michael W. McCracken\textsuperscript{a} and Giorgio Valente\textsuperscript{b}
\textsuperscript{a}: Federal Reserve Bank of St. Louis
\textsuperscript{b}: City University of Hong Kong

October 2014

Abstract

In this paper we provide analytical, simulation, and empirical evidence on a test of equal economic value from competing predictive models of asset returns. We define economic value using the concept of a performance fee—the amount an investor would be willing to pay to have access to an alternative predictive model that is used to make investment decisions. We establish that this fee can be asymptotically normal under modest assumptions. Monte Carlo evidence shows that our test can be accurately sized but sometimes requires large samples. We apply the proposed test to predictions of the US equity premium.

JEL classification: C53, C12, C52

Keywords: Utility-based comparisons, economic value, out-of-sample forecasting, predictability.

\textsuperscript{*}This work was partly written while Giorgio Valente was visiting the Federal Reserve Bank of St Louis whose hospitality is gratefully acknowledged. We gratefully acknowledge helpful comments from Andrew Patton, Enrique Sentana, Nicholas Taylor and conference participants at the Society of Financial Econometrics Annual Meeting in Oxford and the seminar participants at Baptist University of Hong Kong, Chinese University of Hong Kong, and National Taiwan University. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of St. Louis, Federal Reserve System, or any of its staff.

\textsuperscript{1}McCracken (corresponding author): Research Division; Federal Reserve Bank of St. Louis; P.O. Box 442; St. Louis, MO 63166; michael.w.mccracken@stls.frb.org. Valente: Department of Economics and Finance; College of Business; City University of Hong Kong; Kowloon, Hong Kong; g.valente@cityu.edu.hk.
1 Introduction

A wide variety of studies have debated whether asset returns are, or should be, predictable using information available to investors. Many, including Barberis (2000), Lettau and Ludvigson (2001), Campbell and Thompson (2008), and Cochrane (2008, 2011) conclude that asset returns are predictable. Others, including Goyal and Welch (2008) and Boudoukh, Richardson, and Whitelaw (2008), remain skeptical.

In many instances, the evidence of predictability (or the lack thereof) is based upon out-of-sample conditional mean predictions of asset returns. These predictions are, in turn, evaluated using statistical measures of out-of-sample predictive accuracy. In most cases, mean squared error (MSE) is used (e.g. Campbell and Thompson, 2008, Goyal and Welch, 2008; Ferreira and Santa Clara, 2011 and the references therein). In many, though not all, cases the question of interest is whether a newly developed predictive model is better at guiding investment decisions than a pre-existing baseline model. As such, when statistical measures of predictive accuracy (such as MSE) are used to evaluate the performance of a new model, not only are MSEs reported but inference is conducted to determine whether any differences in the MSEs are statistically significant. Inference is often conducted using common approaches to out-of-sample inference (West, 1996; Clark and McCracken, 2001; McCracken, 2007).

While statistical metrics of evaluation are informative, there is increasing interest in evaluating the predictability of asset returns using economic value measures. Examples of economic measures of predictability include Sharpe ratios (Fleming et al., 2001), performance fees (Patton, 2004), certainty-equivalent returns (Ingersoll et al., 2007) and profits (Leitch and Tanner, 1991). As for the case of statistical measures, the most common question of interest is whether a newly developed predictive model is “better” at generating higher economic value than an established baseline model. However, despite their increasing use, these economic measures are typically reported without any indication of whether any empirical differences are statistically significant.

In this paper we develop an asymptotically valid approach to inference when performance fees
are used as the economic measure of predictive accuracy. We define performance fee as the amount an investor would be willing to pay to have access to an alternative predictive model that is used to make better investment decisions. We show that under the null hypothesis that an investor would not be willing to pay a performance fee ($\Phi$) to access a newly developed predictive model, the estimated performance fee is asymptotically normal with zero mean. In addition, following West (1996) and West and McCracken (1998), we are able to show how the asymptotic variance is affected by estimation error – additional variation induced by the fact that the models must be estimated prior to their use. Monte Carlo results suggest that our testing procedure can be reasonably well sized but sometimes requires large samples. In addition, the simulation evidence shows that there can be large differences in statistical versus economic measures of predictive ability. In fact, under some circumstances, the null hypothesis that $\Phi = 0$ can hold despite there being strong conditional mean predictive ability in asset returns.

We are, of course, not the first to emphasize that statistical evidence of predictability need not imply anything about economic predictability. In particular, Cochrane (1999), Sentana (1999, 2005), and Taylor (2013) have developed formulas linking $R^2$s of linear predictive equations to both Sharpe ratios and performance fees. Our main contribution is not in extending these formulas, but rather providing a method for conducting asymptotically valid inference on the null hypothesis that a performance fee is zero versus an alternative in which it is positive. Our results reinforce many of their findings by showing that the non-monotonic link between statistical and economic predictability leads to difficulties in conducting inference.

The paper is organized as follows. Section 2 provides a simple example of the type of application we have in mind. Section 3 develops the theoretical results and Section 4 discusses practical methods related to inference. Section 5 provides simulation results designed to investigate the size and power properties of the test statistic. In section 6 we apply our analytical results to predictions of the equity premium by means of various predictors (as in Goyal and Welch, 2008). A final section concludes.
2 A Simple Illustrative Example

In this section, we delineate a simple example of performance fees in the context of a portfolio of two assets: a single risk-free and a single risky asset (e.g. a stock index) that is predictable by means of a single variable. While simplistic, we use this example not only so that we can convey certain analytical details in closed form, but also since it is consistent with the existing literature that uses utility-based comparisons to assess asset returns predictability. It is worth emphasizing that our analytical results, discussed in the following section, are not restricted to this particular environment.

Let \( r_t \) denote the return on the risky asset and \( r_{ft} \) the rate of return on the risk-free asset. Define \( ep_{t+\tau} = r_{t+\tau} - r_{ft+\tau} \) the excess stock index return, or equity premium, in period \( t + \tau \) and let \( z_t \) denote a variable observed at time \( t \) that is believed to predict \( ep \) at a future time \( t + \tau \). The investor uses the predictive regression

\[
ep_{t+\tau} = \alpha_{1,0} + \alpha_{1,1} z_t + e_{1,t+\tau}, \tag{1}
\]

to make conditional mean forecasts of future stock index excess returns. The variable \( z_t \) has predictive content for \( ep_{t+\tau} \) if \( \alpha_{1,1} \neq 0 \). If \( z_t \) has no predictive content then \( \alpha_{1,1} = 0 \) and Equation (1) collapses to

\[
ep_{t+\tau} = \alpha_{0,0} + e_{0,t+\tau},
\]

where stock index excess returns are equal to their historical mean plus an unpredictable error term. Throughout the paper we denote the competing model as \( 1 \) and the baseline model as \( 0 \). In addition, the investor uses a parametric model to estimate the conditional variance of the excess returns. Define \( \sigma^2_{i,t+\tau}(\hat{\vartheta}_i) \) as the time \( t \) conditional variance of \( ep_{t+\tau} \) implied by model \( i = 0, 1 \) as a function of the finite dimensioned parameter estimates \( \hat{\vartheta}_{i,t} \).

At each forecast origin \( t = T, ..., T + P - \tau \), the investor uses the conditional mean \( (ep_{t+\tau}(\hat{\alpha}_{i,t})) \) and conditional variance \( (\sigma^2_{i,t+\tau}(\hat{\vartheta}_{i,t})) \) predictions of excess returns to decide how much of her wealth to invest in the risky and the risk-free assets. If the investor is endowed with mean-variance
preferences, the optimal allocation to the risky asset $w_{i,t}$ at any time $t$ from model $i = 0, 1$, is given by the conventional formula

$$w_{i,t}(\hat{\beta}_{i,t}) = \hat{w}_{i,t} = \frac{e^{P_{t+\tau}}(\hat{\alpha}_{i,t})}{\gamma \sigma_{i,t+\tau}^2 (\hat{\rho}_{i,t})}$$

(2)

where $\hat{\beta}_{i,t} = (\hat{\alpha}_{i,t}, \hat{\gamma}_{i,t})'$ and $\gamma$ is the investor’s known coefficient of relative risk aversion (RRA). If the investor is endowed with initial wealth $W = 1$ at each forecast origin, the time $t + \tau$ realized gross return implied by model $i = 0, 1$ equals

$$\hat{R}_{i,t+\tau} = 1 + r_{t+\tau}^f + \hat{w}_{i,t}(\hat{\beta}_{i,t})e^{P_{t+\tau}}.$$ 

Any improvements in predictive ability between models 0 and 1 are assessed using utility-based comparisons. That is, for a prespecified utility function $U(R)$, economic value is evaluated by comparing the average utility implied by models 0 and 1 across all forecast origins. If we let $\hat{P} = P - \tau$ this takes the form

$$\overline{U}(\hat{R}_i) = \hat{P}^{-1} \sum_{t=T}^{T+\tau} U(\hat{R}_{i,t+\tau}).$$

As in Fleming et al. (2001), the difference in utilities is characterized using the concept of a performance fee. This fee is the value of $\hat{\Phi}$ that satisfies

$$U(\hat{R}_1 - \hat{\Phi}) - U(\hat{R}_0) = 0.$$

We interpret $\Phi$ as the maximum fraction of wealth the investor would be willing to pay per period to switch from model 0 to model 1. If the two conditional variance models are identical, this criterion measures how much a risk-averse investor is willing to pay for conditioning on the information in the predictive variable $z_t$. It follows that, if there is no predictive power embedded in the variable $z_t$, then $\Phi = 0$; whereas, if $z_t$ helps to predict stock excess returns, one expects $\Phi > 0$.

We can better understand the behavior of $\Phi$ when $a_{1,1} \neq 0$ if a few more assumptions are made. In particular, let the forecast horizon be $\tau = 1$ and assume that $z_t$ follows a stationary AR(1) process of the form

$$z_t = \mu_z (1 - \rho) + \rho z_{t-1} + v_t,$$
where \((e_t, v_t)\) are i.i.d. normally distributed with zero means and variances \(\sigma_e^2\) and \(\sigma_v^2\). Let \(\mu_z\) and \(\sigma_z^2\) denote the unconditional mean and variance of \(z_t\). Finally, assume that the conditional mean models are estimated by OLS and the conditional variance models are identical so that 
\[
\sigma^2_{0,t+1}(\hat{\theta}_{0,t}) = \sigma^2_{1,t+1}(\hat{\theta}_{1,t}).
\]
More specifically, for ease of presentation, assume that the predictions of these conditional variances are obtained simply by using a consistent estimate of the unconditional variance.\(^1\)

Straightforward algebra shows that

\[
\Phi = \left( \frac{\alpha_{1,1}^2 \sigma_z^2}{\gamma (\alpha_{1,1}^2 \sigma_z^2 + \sigma_e^2)} \right) \left( \frac{\sigma_e^2 - 3(\alpha_{1,0} + \alpha_{1,1} \mu_z)^2}{2(\alpha_{1,1}^2 \sigma_z^2 + \sigma_e^2)} \right). \tag{3}
\]

Equation (3) is the product of two terms. If the first term is zero, as it is when \(\alpha_{1,1} = 0\), then \(\Phi = 0\). This first term can be interpreted as the (population) \(R^2\) from the predictive model 1, scaled by \(\gamma\). Using this interpretation, the smaller the \(R^2\) from model 1, the closer \(\Phi\) is to zero. The second term is less easily interpretable but arises due to the marginal differences in the variance components of the mean-variance utility functions.

Since the first term in parentheses increases monotonically with \(|\alpha_{1,1}|\), it seems likely that larger (absolute) values of \(\alpha_{1,1}\) imply larger values of \(\Phi\). In this case statistical measures of predictive accuracy coincide with utility-based economic measures of predictability. In fact, stronger evidence of statistical predictability, represented by large t-statistics on \(\alpha_{1,1}\) or large \(R^2\) recorded for the unrestricted regression, imply larger utility gains to investors, and subsequently larger values of \(\Phi > 0\).

However, it is worthwhile noting that this intuitive case need not be the only case in which \(\Phi\) equals zero nor is it trivially true that larger (absolute) values of \(\alpha_{1,1}\) imply larger values of \(\Phi\). Note that the second term in parentheses also depends upon \(\alpha_{1,1}\). Taken as a whole, this implies that as \(\alpha_{1,1}\) deviates from zero, \(\Phi\) can be positive, negative, or zero depending on the specifics of the data generating process. Clearly, large and statistically significant statistical measures of accuracy

\(^1\)While this may seem odd given our framework, rolling window estimates of the unconditional variance of \(e_{p,t+1}\) are often used as estimates of the conditional variance of excess returns in the empirical literature (Goyal and Welch, 2008; Campbell and Thompson, 2008 and Ferreira and Santa Clara, 2011).
like MSEs and $R^2$s need not provide any indication of economic performance when measured using performance fees.

As we will see in the following section, the fact that $\Phi$ can have multiple roots (as a function of $\alpha_{1,1}$) makes inference more complicated than we might want. In particular, while we are able to establish that $\tilde{P}^{1/2}\Phi$ is asymptotically normal with zero mean at each root, the asymptotic variance differs whether $\alpha_{1,1}$ is zero or non-zero. Fortunately, there is a straightforward estimator of the asymptotic variance that is robust to both instances. Simulation evidence suggests this estimator works reasonably well but sometimes requires large samples to provide accurately sized tests.

And finally, the fact that $\Phi$ can have multiple roots affects the choice of bootstrap one might use to conduct inference. For example, given the simple example above, one might be inclined to use a recursive VAR-based bootstrap in $(e_{P+1}, z_t)'$ with zero restrictions imposed on the first equation to impose the null hypothesis that $\Phi = 0$. While this might perform well when $\alpha_{1,1} = 0$ there is clearly no justification for it’s use at other roots of $\Phi$. Calhoun (2011) proposes a block bootstrap, designed explicitly for out-of-sample inference, that may be applicable as it imposes the null using recentering methods adapted from White (2000). We leave the question of bootstrap-based inference to further research.

## 3 Theoretical Results

This section provides the null asymptotic distribution of per-period performance fee measures $\hat{\Phi}$ constructed using pseudo-out-of-sample methods. The performance fee $\Phi$ is estimated as a function of two sequences of pseudo-out-of-sample forecasts; one each for models 0 and 1. In the context of the example from section 2 these forecasts consist of both conditional mean and conditional variance forecasts. In order to calculate the performance fee, we assume that the investor has access to the necessary observables over the time frame $s = 1, ..., T + P$. This sample is split into an in-sample period $s = 1, ..., T$ and an out-of-sample period $t = T + 1, ..., T + P$. At each forecast origin $t = T, ..., T + P - \tau$, both of the parametric $\tau$-period ahead investing models are estimated
and used to construct a forecast that is then used to construct portfolio weights. The assumptions
used to derive the asymptotic results are presented below and closely follow those in West (1996)
with some modest deviations.

1. Let \( \beta = (\beta_0', \beta_1')' \). (a) There exists a function \( f(X_{t+\tau}, \beta) = f_{t+\tau}(\beta) \), with \( f_{t+\tau}(\beta^*) = f_{t+\tau} \),
that is twice continuously differentiable in \( \beta \) and satisfies \( \hat{\Phi}^{1/2} = \tilde{P}^{1/2} = \sum_{t=0}^{T-P-\tau} \hat{f}_{t+\tau}(\hat{\beta}_t) + o_p(1) \). (b) For the max norm \( |.| \) and some open neighborhood \( N \) of \( \beta^* \),\( E(\sup_{\beta \in N} \partial^2 f_{t+\tau}(\beta)/\partial \beta \partial \beta') < D \) some finite scalar \( D \).

2. The parameters are estimated using one of two sampling schemes: the recursive or the rolling.
The recursive parameter estimates satisfy \( \hat{\beta}_{i,t} - \beta_i^* = B_i(t)H_i(t) \) where \( B_i(t) \rightarrow^{a.s.} B_i \) a non-

3. Define \( F = E f_{t+\tau, \beta} \). If the models are nested, \( F \beta \neq 0 \).

Before providing the main result, it is important to explain some of the key assumptions and
their implications for the validity of our testing procedure. Assumption 1 maps the problem of
inference on \( \hat{\Phi} \) into a framework in which the theoretical results in West (1996) can be applied
directly. While the assumption is stated at a very high level it is actually very simple to verify.
For example, in the context of the mean-variance example from Section 2, Assumption 1 is satisfied for the function

\[ f_{t+\tau}(\hat{\beta}_t) = (\hat{R}_{1,t+\tau} - \frac{\gamma}{2}(\hat{R}_{1,t+\tau} - ER_{1,t+\tau})^2) - (\hat{R}_{0,t+\tau} - \frac{\gamma}{2}(\hat{R}_{0,t+\tau} - ER_{0,t+\tau})^2) \] (4)

if \( \hat{R}_i \rightarrow^p ER_{i,t+\tau} \). As another example suppose that power utility is used and hence \( U(\hat{R}_{i,t+\tau}) = \hat{R}_1^{1-\rho} \). Since \( \hat{\Phi} \) is defined as a root and \( U(.) \) is continuously differentiable in its' argument we obtain

\[ \hat{\Phi} = (\hat{U}(\hat{R}_1) - \hat{U}(\hat{R}_0)) / \hat{P}^{-1} \sum_{t=T}^{T+P-\tau} \partial U(\hat{R}_{1,t+\tau} - \hat{\Phi}) / \partial \Phi \text{ some } \hat{\Phi} \text{ on the line between } \hat{\Phi} \text{ and } 0. \]

If \( \hat{P}^{-1} \sum_{t=T}^{T+P-\tau} \partial U(\hat{R}_{1,t+\tau} - \hat{\Phi}) / \partial \Phi \rightarrow^p E(\partial U(\hat{R}_{1,t+\tau}) / \partial \Phi) \neq 0 \), Assumption 1 is satisfied with

\[ f_{t+\tau}(\hat{\beta}_t) = (\hat{R}_{1,t+\tau}^{1-\rho} - \hat{R}_{0,t+\tau}^{1-\rho}) / (E(\partial U(\hat{R}_{i,t+\tau} / \partial \Phi)(1 - \rho)). \]

The requirement that \( f_{t+\tau}(\beta) \) is twice continuously differentiable in \( \beta \) is non-trivial for our results. In the vast majority of studies on economic value calculations, the utility function \( U(\cdot) \) itself is continuously differentiable in the gross return. Even so, there are cases where the assumption might fail because \( \hat{R}_{i,t+\tau} \) is not twice continuously differentiable in \( \beta \). For example, in some applications the estimated portfolio weights can be bounded (or winsorized) in order to limit the maximal amount of leverage in the constructed portfolio (Ferreira and Santa Clara, 2011). We pursue this issue further in simulations reported in section 5.

Given Assumption 1, Assumptions 2, 3, and 4 are nearly identical to those in West (1996). The predictive models must be parametric and estimated using a framework that can be mapped into GMM. The observables must have sufficient moments and mixing conditions to satisfy a Central Limit Theorem,\(^3\) and the number of in-sample and out-of-sample observations must be of the same order. Assumption 5 states that if the two models are nested under the null hypothesis, and hence \( R_{1,t+\tau} = R_{0,t+\tau} \), it must be the case that a certain product of two moments is nonzero. We do so since, as shown below, there is the potential for the asymptotic variance of \( \hat{P}^{1/2}\hat{\Phi} \) to be zero. To avoid this problem we state a high level assumption that insures that the asymptotic variance is\(^2\)

---

\(^2\)Under the null hypothesis that \( \Phi = 0. \)

\(^3\)West uses the central limit theorem of Wooldridge and White (1998).
non-zero. As a practical matter, the condition is likely to hold as long as the model parameters are not estimated using the utility function \( U(\cdot) \) as the objective function.\(^4\) Given the assumptions, our main result follows immediately from Theorem 4.1 of West (1996).

**Theorem 1** Maintain Assumptions 1-5. \( \tilde{\Phi}^{1/2} \tilde{\Phi} \to^d N(0, \Omega) \) with

\[
\Omega = S_{ff} + 2\Lambda_0(\pi)FBS_{fh}' + \Lambda_1(\pi)FBS_{hh}B'F'
\]

where \( S_{ff} = \lim_{T \to \infty} \text{Var} \left( T^{-1/2} \sum_{s=1}^{T} f_{s+\tau} \right), \ S_{hh} = \lim_{T \to \infty} \text{Var} \left( T^{-1/2} \sum_{s=1}^{T} h_{s+\tau} \right), \ S_{fh} = \lim_{T \to \infty} \text{Cov} \left( T^{-1/2} \sum_{s=1}^{T} f_{s+\tau}, T^{-1/2} \sum_{s=1}^{T} h_{s+\tau} \right) \), and

<table>
<thead>
<tr>
<th>scheme</th>
<th>( \Lambda_0(\pi) )</th>
<th>( \Lambda_1(\pi) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>recursive</td>
<td>((1 - \pi^{-1} \ln(1 + \pi)))</td>
<td>(2(1 - \pi^{-1} \ln(1 + \pi)))</td>
</tr>
<tr>
<td>rolling 0 &lt; ( \pi \leq 1 )</td>
<td>(\frac{\pi}{2})</td>
<td>(\pi - \frac{\pi^2}{3})</td>
</tr>
<tr>
<td>rolling 1 ≤ ( \pi &lt; \infty )</td>
<td>(1 - \frac{1}{2\pi})</td>
<td>(1 - \frac{1}{3\pi})</td>
</tr>
</tbody>
</table>

The Theorem shows that the estimated performance fee is asymptotically normal with an asymptotic variance that reflects not only variation in the difference in utilities, via \( S_{ff} \), but also the influence of estimation error via the remaining components of the variance, \( FBS_{fh}' \) and \( FBS_{hh}B'F' \), respectively.

When the models are nested there are two distinct cases in which the Theorem applies. The leading case is when \( \Phi = 0 \) because the models are identical under the null hypothesis and hence the “competing” and “baseline” models are better thought of as “unrestricted” and “restricted.” In this situation, there exists a selection matrix \( J \) such that \( J\beta^*_1 = \beta^*_0 \) and the asymptotic variance simplifies to

\[
\Omega = \Lambda_1(\pi)(E(\frac{\partial U_{1,t+\tau}}{\partial \Phi})^{-1}E(\frac{\partial U_{1,t+\tau}}{\partial \beta_1})(-JB_0J + B_1)Sh_{h_1}(-JB_0J + B_1)E(\frac{\partial U_{1,t+\tau}}{\partial \beta_1}). \quad (5)
\]

The asymptotic variance simplifies because in this specific case, \( R_{1,t+\tau} = R_{0,0,t+\tau} \) for all \( t \) and hence \( S_{ff} \) and \( S_{fh} \) are both trivially zero. In addition, the fact that the models are nested implies \( J'h_{1,t+\tau} = h_{0,t+\tau} \) and \( J'\frac{\partial U_{1,t+\tau}}{\partial \beta_1} = \frac{\partial U_{0,t+\tau}}{\partial \beta_1} \) and thus the asymptotic variance can be simplified even

\(^4\)This assumption precludes a few isolated applications including Cenizoglu and Timmermann (2011).
further. While not immediately obvious in the original Theorem, it is for this case that we have
Assumption 5. To achieve asymptotic normality in any useful sense we need \( \Omega \) to be positive.
When predictability exists and hence \( \beta_{1,1}^* \) is not zero, \( S_{ff} \) is non-zero and hence \( \Omega \) is non-zero so
long as \( S_{ff} \) does not happen to cancel with the remaining terms. By imposing Assumption 5 we
insure that \( \Omega \) is positive for the case in which the unrestricted model perfectly nests the restricted
model - a possibility that we must allow for under the null hypothesis.

A less intuitive case arises when the models are ostensibly nested and yet \( \Phi = 0 \) despite the
fact that \( \beta_{1,1}^* \) is non-zero. In this case, as stated in the Theorem, \( \Phi \) is asymptotically normal with
mean zero but with an asymptotic variance that does not simplify as it does when \( \beta_{1,1}^* = 0 \). Note
also that our results are applicable to the comparison of two models that are non-nested under the
null. While theoretically plausible, such a comparison does not appear to exist in the literature
and hence we do not pursue it any further herein.

4 Inference

The Theorem from the previous section provides a means of assessing the statistical significance
of performance fees for a given utility function. Specifically, if \( \hat{\Omega} \) is a consistent estimate of \( \Omega \)
it immediately follows that \( \tilde{P}^{1/2} \hat{\Phi} / \tilde{\Omega}^{1/2} \sim^d N(0, 1) \), and therefore one can use standard normal
critical values to test the null that \( \Phi = 0 \) against an alternative in which \( \Phi > 0 \). To facilitate
application of our results, and to emphasize some peculiar features of our results, in the following
we delineate three estimates of \( \Omega \) in the context of the simple environment discussed in Section 2.
Each of these three estimates are subsequently considered in the simulations of Section 5.

Recall that since utility is mean-variance, the percentage of wealth invested in the risky asset
takes the form

\[
 w_{i,t}(\hat{\beta}_{i,t}) = \hat{w}_{i,t} = \frac{ep_{t+1}(\hat{\alpha}_{i,t})}{\gamma \sigma_{i,t+1}^2 \left( \hat{\vartheta}_{i,t} \right)}
\]

for a known relative risk aversion parameter \( \gamma \), conditional mean prediction \( ep_{t+1}(\hat{\alpha}_{i,t}) \), and
conditional variance prediction \( \sigma_{i,t+1}^2 \left( \hat{\vartheta}_{i,t} \right) \). The baseline model 0 uses the historical mean of \( ep_{t+1} \)
\((\alpha_{0,0})\) as the conditional mean prediction and the historical unconditional variance of \(e_{p,t+1}\) as the conditional variance prediction. The competing model 1 forms the conditional mean prediction using the OLS-estimated regression \(e_{p,t+1} = \alpha_{1,0} + \alpha_{1,1}z_t + e_{1,t+1} = \alpha'_1 x_t + e_{1,t+1}\) and also uses the historical unconditional variance of \(e_{p,t+1}\) as the conditional variance prediction.

We allow for three distinct methods of estimating each of these predictions. Under the recursive scheme, both \(\alpha_i\) and \(\vartheta_i\) \(i = 0,1\) are estimated using all available observations from \(i = 1, ..., t\) for each forecast origin \(t = T, ..., T + P - 1\). Under the rolling scheme, \(\alpha_i\) and \(\vartheta_i\) \(i = 0,1\) are estimated using only the most recent \(T\) observations from \(i = t - T + 1, ..., t\) for each forecast origin \(t = T, ..., T + P - 1\). Under a “mixed” scheme, used in Goyal and Welch (2008), \(\alpha_i\) is estimated using the recursive scheme while \(\vartheta_i\) is estimated using a small rolling window of size \(M << T\). The first two schemes align with the maintained Assumptions of section 3 and, in particular, define \(\hat{\beta}_{i,t}\) as \((\hat{\alpha}'_{i,t}, \hat{\vartheta}'_{i,t})'\). The latter may align with the maintained Assumptions but only after reinterpreting the variance parameter \(\hat{\vartheta}_i = M^{-1} \sum_{s=t-M+1}^t (e_{p,s} - \bar{e}_{s,M})^2\) where \(\bar{e}_{s,M}\) denotes the sample mean of \(e_{p,s}\) over the sample \(s = t - M + 1, ..., t\). Since \(M << T\) it is not unrealistic to interpret the parameter estimates in the manner put forth by Giacomini and White (2006). There, asymptotics are developed whereby parameters are assumed to be estimated using a rolling window of fixed and finite length \(M\). By taking this approach, they effectively treat the parameter estimates as just another time series of observables in much the same way we are treating \(e_{p,t+1}\) and \(z_t\). If we take this view of the rolling window estimator of the unconditional variance, and redefine \(\hat{\beta}_{i,t}\) as \(\hat{\alpha}_{i,t}\), our theoretical results continue to be applicable with one distinction: Contributions to \(\Omega\) due to estimation error only arise via \(\hat{\alpha}_{i,t}\).

- \(\Omega_{0}\): The most complicated estimator is the one that estimates every component of \(\Omega\) using the formula in the Theorem. Under the recursive or rolling schemes, for which \(h_{s+1} = (e_{0,s+1}, (e_{p,s+1} - E e_{p,s+1})^2, E e_{p,s+1} - E e_{p,s+1}, e_{1,s+1}, z_s e_{1,s+1}, (e_{p,s+1} - E e_{p,s+1})^2, E e_{p,s+1} - E e_{p,s+1})'\), these elements take the form \(B = diag(B_0, B_1), B_0 = I_2, B_1 = diag((E x'_t)\gamma^{-1}, 1)\), and \(F = (-E_{\partial U_{i,t+1}}^{\partial \omega_{i,t+1}}, E_{\partial U_{i,t+1}}^{\partial \beta_{i,t+1}})'\) where \(E_{\partial U_{i,t+1}}^{\partial \omega_{i,t+1}} = E_{\partial u_{i,t+1}}^{\partial \beta_{i,t+1}} e_{p,t+1}(1 - \gamma(R_{i,t+1} - E R_{i,t+1}))\).

12
$$i = 0, 1.$$ Under the mixed scheme, for which $$h_{s+1} = (e_{0,s+1}, e_{1,s+1}, z_s e_{1,s+1})',$$ these elements take the form $$B = \text{diag}(B_0, B_1),$$ $$B_0 = 1,$$ $$B_1 = (E x_t x_t')^{-1},$$ and $$F = (-E \frac{\partial U_{0,t+1}'}{\partial \beta_0}, E \frac{\partial U_{1,t+1}'}{\partial \beta_1})$$ where $$E \frac{\partial U_{i,t+1}'}{\partial \beta_i} = E \frac{\partial U_{i,t+1}'}{\partial \beta_i} e_{p,t+1}(1 - \gamma(R_{t,t+1} - E R_{t,t+1}))$$ $$i = 0, 1$$ but with $$\beta_i$$ appropriately redefined. In either case, sample analogs can be used to estimate each component and a HAC estimator of the moment conditions $$h_{s+1}$$ and $$f_{s+1}$$ can be used to estimate $$S_{ff}, S_{fh}$$ and $$S_{hh}.$$ It’s important to note that in general, both $$h_{s+1}$$ and $$f_{s+1}$$ will be serially correlated even if $$e_{1,s+1}$$ is a martingale difference sequence: When $$\alpha_{1,1}$$ is not zero, the error term in model $$0$$ ($$e_{0,s+1}$$) will exhibit serial correlation induced by the omitted regressor $$z_s.$$ Finally, $$\Lambda_0(\pi)$$ and $$\Lambda_1(\pi)$$ can be estimated using $$\Lambda_0(P/T)$$ and $$\Lambda_1(P/T).$$ We describe this estimator as being “robust” because it is consistent for $$\Omega$$ whether or not $$\alpha_{1,1}$$ is zero. If $$\alpha_{1,1}$$ is zero then we have estimated additional terms, such as as $$S_{ff}$$ and $$S_{fh},$$ that are zero in population, but it is a consistent estimator nevertheless.

- $$\Omega_0$$: A simple way to estimate $$\Omega$$ is to ignore the contribution of estimation error completely. If we take this approach, the logical estimator ($$\hat{S}_{ff}$$) is a HAC estimator of the asymptotic variance associated with the moment condition $$f_{t+1} = (\hat{R}_{1,t+1} - \frac{\gamma}{2}(\hat{R}_{1,t+1} - \hat{R}_1)^2) - (\hat{R}_{0,t+1} - \frac{\gamma}{2}(\hat{R}_{0,t+1} - \hat{R}_0)^2).$$ One advantage of this approach is that it avoids the sometimes complicated estimation of $$F$$ and is thus the simplest of the three estimators. It has the disadvantage of being badly wrong in some circumstances under which the null holds. To see this consider the case in which $$\Phi = 0$$ because $$\alpha_{1,1} = 0.$$ The models are identical and hence $$R_{0,t+1} = R_{1,t+1}.$$ Since this implies $$f_{t+1} = 0$$ we find that $$\hat{S}_{ff} \to_p S_{ff} = 0$$ despite the fact that $$\Omega > 0.$$ Together this implies $$|\tilde{P}^{1/2} \hat{\Phi} / \Omega_0^{1/2}| = |\tilde{P}^{1/2} \hat{\Phi} / \hat{S}_{ff}^{1/2}|$$ diverges with probability 1 and hence we expect to reject in the upper tail 50% of the time under the null hypothesis that $$\Phi = 0$$ when $$\alpha_{1,1} = 0.$$ On the other hand, if $$\Phi = 0$$ and $$\alpha_{1,1}$$ is not zero, $$S_{ff}$$ is also not zero. In this case the estimator might work reasonably well if $$2\Lambda_0(\pi)F B S_{fh}' + \Lambda_1(\pi)F B S_{hh} B' F'$$ is small relative to $$S_{ff}.$$

- $$\Omega_c$$: Another approach to estimating $$\Omega$$ is to focus on the subset of the null in which $$\alpha_{1,1} =
0. Since this implies both $S_{ff}$ and $S_{fh}$ are zero we need only estimate the elements of the formula given in equation (5). Under the recursive or rolling schemes, for which $h_{1,s+1} = (e_{1,s+1}, z_se_{1,s+1}, (ep_{s+1} - Eep_{s+1})^2 - E(ep_{s+1} - Eep_{s+1})^2)'$, these elements take the form $B_0 = I_2, B_1 = diag((E_{xt}x_t')^{-1}, 1), J = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $E \frac{\partial u_{1,t+1}}{\partial \beta} = E \frac{\partial u_{1,t+1}}{\partial \beta} ep_{t+1}(1 - \gamma(R_{1,t+1} - ER_{1,t+1})), and \frac{\partial u_{1,t+1}}{\partial \phi} = -1$. Under the mixed scheme, for which $h_{1,s+1} = (e_{1,s+1}, z_se_{1,s+1})'$, these elements take the form $B_0 = 1, B_1 = (E_{xt}x_t')^{-1}, J = (1,0)'$, $E \frac{\partial u_{1,t+1}}{\partial \beta} = E \frac{\partial u_{1,t+1}}{\partial \beta} ep_{t+1}(1 - \gamma(R_{1,t+1} - ER_{1,t+1}))$ (but with $\beta$ appropriately redefined), and $\frac{\partial u_{1,t+1}}{\partial \phi} = -1$. In either case, sample analogs can be used to estimate each component and a HAC estimator of the moment condition $h_{1,s+1}$ can be used to estimate $S_{h_1 h_1}$. Finally, $\Lambda_1(\pi)$ can be estimated using $\Lambda_1(P/T)$. While easier to construct than $\Omega_a, \Omega_c$ suffers from the fatal flaw that it is an inconsistent estimate of $\Omega$ when $\alpha_{1,1}$ is not zero and yet $\Phi = 0$. As we will see in the simulations, this leads to large size distortions in much the same way when $\Omega_b$ is used and $\alpha_{1,1} = 0$.

5 Monte Carlo Evidence

In the previous two sections we show that $\hat{\Phi}$ is asymptotically normal and delineate an asymptotically valid approach to inference. In this section we provide Monte Carlo evidence on the finite sample properties of the asymptotic results. Specifically we provide simulation evidence on the efficacy of our approach to testing the null hypothesis $H_0 : \Phi = 0$ against the alternative $H_A : \Phi > 0$ across a variety of different parameterizations of the data generating process and methods for estimating the asymptotic variance.

5.1 Experiment design

In our experiments we consider the problem of a US investor who faces the problem of choosing how to optimally allocate her wealth between the value-weighted index of stocks traded on the NYSE and the 3-month T-bill. We assume that the investor is endowed with mean-variance preferences and the performance fees are computed using a mean-variance utility function, as outlined in Section
The experiments are conducted as follows. For all cases, we use data generating process loosely calibrated on the monthly estimates reported in Barberis (2000) relative to the empirical properties of excess returns to the NYSE value-weighted index and its dividend yield $z_t$:

\begin{align*}
ep_{t+1} &= a + bz_t + u_{t+1}z_t \\
z_t &= \mu_z(1 - \rho) + \rho z_{t-1} + v_t \\
r_{t+1}^f &= 0.036 \tag{6}
\end{align*}

with

\[
\begin{pmatrix}
u_t \\
v_t
\end{pmatrix} \sim i.i.d. N \left( 0, \begin{bmatrix} 0.0172 & -0.004 \\
-0.00004 & 0.0063(1 - \rho^2) \end{bmatrix} \right).
\]

We consider in-sample and out-of-sample periods such that $P/T = 1/3, 1, \text{ and } 3$ for overall sample sizes $T + P = 1024, 2048, \text{ and } 4096$. Note that we allow for the presence of conditionally heteroskedastic errors in the equity premium equation. The risk aversion parameter $\gamma$ is set to 3. For brevity we focus exclusively on the recursive and mixed schemes with $M = 60$. Results for the rolling scheme are broadly similar to those for the recursive.

For DGP 1, $a$ and $\mu_z$ take the data-based values 0.0252 and 0.32 respectively. For this design $\Phi$ has a single root at $b = 0$ and hence we parameterize the null hypothesis by setting $b$ equal to zero. To get a feel for the impact winsorizing may have on our asymptotics we conduct the experiment twice: once with and once without winsorized portfolio weights. In those experiments that winsorize the weights we restrict them to lie between $0.5 \text{ and } 1.5$. For DGP 2 we allow $a$ and $\mu_z$ to take the somewhat extreme values $-0.5 \text{ and } 0.5$. We do so to provide an example in which $\Phi$ has more than one root. Under the recursive scheme these roots are $b = 0$ and $b \simeq 1.16$

---

5 We use a burnout period of 500 observations to remove the effects of initial conditions.

6 In the current simulations, Assumption 5 requires that $FB = E(\frac{\partial U_1}{\partial \delta_1})(-JB_0 J + B_1)$ is non-zero. Straightforward algebra reveals that Assumption 5 fails if $ep_{t+1}$ and $ep_{1+1}$ are both mean independent of $(1, z_t)$. It is for this reason that we introduce conditionally heteroskedastic errors. The issue is only relevant when $\alpha_{t,1} = 0$. In unreported results for which $\alpha_{t,1} = 0$ and the errors are conditionally homoskedastic, the rejection frequencies are less than 1 percent for nominally 5 percent tests.
while under the mixed scheme the roots are $b = 0$ and $b \simeq 1.143$.\(^7\)

For the power experiments we focus exclusively on DGP 2 in order to highlight how power is affected by the presence of multiple roots. To do so we allow $b$ to range from 0 to 1.4 in increments of roughly 0.04. Finally, in both the size and power experiments we investigate the role of persistence in the predictor $z_t$ by considering both $\rho = 0.5$ and a much higher $\rho = 0.9$. All results are based on 2,000 Monte Carlo replications. In all simulations, Newey-West (1987) HAC estimators were used to estimate $S_{ff}$, $S_{hh}$, and $S_{fh}$ with a lag length fixed at 4.

5.2 Size Results

In Tables 1 and 2 we report the actual rejection frequencies of nominally 5 percent tests of the null hypothesis $\Phi = 0$. Table 1 corresponds to DGP 1 for which $b = 0$ while Table 2 corresponds to DGP 2 for which $b$ equals 0 and either 1.16 or 1.143 depending on which schemes used. Each table has two panels and three sub-panels with sub-panels corresponding to the total sample size increasing from 1024, to 2048, to 4096. Within each panel the columns are associated with varying sample splits and levels of persistence in $z$ while the rows correspond to various estimation schemes and various estimators of the asymptotic variance.

Consider the first panel of Table 1 within which we do not winsorize the portfolio weights. In these experiments, both $\Omega_a$ and $\Omega_c$ are valid estimators of the asymptotic variance though $\Omega_a$ estimates unnecessary terms. In almost all cases, and in accordance with the theory the tests associated with $\Omega_c$ are reasonably well-sized with rejection frequencies ranging from 0.064 to 0.093 with the greatest size distortions occurring when the persistence of the predictor is increased from 0.5 to 0.9. Despite being asymptotically valid, when $\Omega_a$ is used the tests tend to be seriously undersized with rejection frequencies as low as 0.003. Even so, as the sample size increases across the three sub-panels, the rejection frequencies typically improve with values approaching 0.03 when $T + P = 4096$. In contrast, and in accordance with the theoretical results, when $\Omega_b$ is used the

---

\(^7\)The non-zero values of $b$ were chosen based on extensive simulations. The formula for $\Phi$ in section 2 is not applicable due to the introduction of conditionally heteroskedastic errors.
rejection frequencies seem to be converging to 50% as the sample size increases. For example, when \( T + P = 1024 \) the rejection frequencies are generally in the range of 0.07 but as the overall sample size increases to 4096 we start to observe frequencies around 0.30.

In the second panel of Table 1 we replicate the experiments from the first but restricting the portfolio weights to lie between \(-0.5\) and 1.5. A casual glance at this portion of the Table reveals that there are very few deviations from the previous panel. In fact, the deviations that do exist typically occur only in the third decimal. For this experiment it appears that winsorizing has little-to-no impact on the asymptotics.

Consider the first panel of Table 2 for which the null holds because \( b = 0 \). In these experiments, as above, both \( \Omega_a \) and \( \Omega_c \) are valid estimators of the asymptotic variance though \( \Omega_a \) estimates unnecessary terms. In almost all cases, the tests are oversized and sometimes seriously so. This is particularly true when \( \Omega_b \) is used and matches the theory perfectly: The rejection frequencies converge to 50% as the sample size increases. Especially for \( \Omega_c \), the rejection frequencies tend to improve as the sample size increases though, admittedly, nominally 5% tests still reject at frequencies near 10%. Rejection frequencies tend to fall as the sample split increases from 1/3 to 3. Interestingly, in some cases, and particularly for the mixed scheme, \( \Omega_a \) seems to provide the most accurately sized tests so long as \( P/T \) is not too large. Increasing the persistence in \( z_t \) from 0.5 to 0.9 has little impact on the rejection frequencies. The one place it does seem to have an impact is that it tends to lower the rejection frequencies under the mixed scheme when \( \Omega_b \) is used.

The results in the second panel of Table 2, for which the null holds despite \( b > 0 \), are a bit better. In these experiments, \( \Omega_a \) is a valid estimator while \( \Omega_c \) is not. \( \Omega_b \) is also not a valid estimator but distortions may not be too bad because the additional variance terms driven by estimation error are not too large. Unsurprisingly, the experiments that use \( \Omega_c \) are hugely oversized. But when either \( \Omega_a \) or \( \Omega_b \) are used, and \( \rho = 0.5 \), rejection frequencies generally range from 0.04 to 0.10. Unfortunately, when we increase the persistence to \( \rho = 0.9 \), the rejection frequencies increase generally ranging from 0.10 to 0.20. Regardless of persistence, the rejection frequencies improve as
the sample size increases approaching nominally 5% tests at the largest sample size. In contrast to the results in the first panel of Table 2, rejection frequencies tend to rise as the sample split increases from $1/3$ to 3.

5.3 Power Results

In Figures 1 - 2 we plot rejection frequencies of nominally 5 percent tests as we increase the tuning parameter $b$ from 0 to 1.4. Despite the size distortions observed in the Tables, these plots give an indication of the power of the tests across different sampling schemes, sample sizes, and sample splits.

Figures 1 plots rejection frequencies when the recursive schemes is used. The four panels delineate power as we increase the sample size from 1024 to 2048 and increase the persistence from 0.5 to 0.9. In large part, each plot takes a similar shape. Power is non-monotonic in $b(=\alpha_{1,1})$ and exhibits two peaks: one at roughly 0.18 and one around 0.9. The first peak is distinctly lower with power ranging from 0.15 to 0.60 depending on the sample size and split. The latter peak has power ranging from 0.80 to 1.00. Unsurprisingly, the reason we observe non-monotonic power is that $\Phi$ appears to be non-monotonic in $\alpha_{1,1}$. In these simulations, the median simulated $\Phi$ is very close to zero (zero out to four decimals) when $\alpha_{1,1} = 0$. After that, the median simulated $\Phi$ increases, then decreases (but stays positive), and increases again before crashing to zero at roughly 1.16. At this point it turns and stays negative. Regardless of the non-monotonicity, power increases with the sample size and tends to be higher for smaller values of the sample split parameter. Increasing the degree of persistence from 0.5 to 0.9 has little affect on power.

In Figure 2 we report rejection frequencies when the mixed scheme is used. The plots are perhaps even more interesting than those in Figure 1. When $\rho = 0.5$, power appears to be monotonic in much the same way as Figure 1 but requires the larger sample size for that power to exhibit itself. The median simulated values of $\Phi$ reinforce this interpretation. When $\rho$ is increased to 0.9, rejection frequencies remain at nominal levels for all values of $\alpha_{1,1}$ less than roughly 0.6 at which point they sharply rise before again crashing to zero at roughly 1.143. As before, the
median simulated values align with the rejection frequencies: the estimated values of $\Phi$ remain very close to zero (both positive and negative) for $\alpha_{1,1}$ less than roughly 0.6 and hence we would expect little-to-no power.

When interpreting the mixed-scheme results, it is important to understand how our assumptions influence the null hypothesis. Under the recursive and rolling schemes, the estimated conditional variance in the denominator of $w_{i,t}, i = 0, 1$ converges to a constant in large samples. Under the mixed scheme, $M$ is treated as small and fixed and hence the estimated conditional variance never converges regardless of the sample size. This minor difference changes how we interpret $\Phi$ in much the same way as the asymptotic theory in Giacomini and White (2005) influences how we interpret (statistical, MSE-based) tests of equal predictive ability. In short, the null $\Phi = 0$ under the mixed scheme is distinct from that under either the recursive or rolling schemes (for which the null is the same).

6 The Economic Value of Predictions of the US Equity Premium

In this section we employ the testing procedure discussed in the previous sections to revisit the findings of the recent literature on the predictability of the US equity premium. Methodologically the framework we use is identical to the one highlighted in Section 2. We use monthly value-weighted returns from the S&P 500 index from January 1927 to December 2011 from the Centre for Research in Security Prices (CRSP) and Robert Shiller’s website. Stock returns are continuously compounded including dividends and the predictive variables $z_t$ are a selection of 14 variables from the ones used in Goyal and Welch (2008, and additional appendix).\(^8\) We begin forecasting in January 1965 and continue through December 2011 giving us $P = 564$ out-of-sample observations out of a total of $T + P = 1692$.

Following Ferreira and Santa-Clara (2011) we compute the weights $\hat{w}_{i,t}$ using the mixed scheme:

\(^8\) For further details on data construction, refer to Goyal and Welch (2008 and additional appendix). The full dataset used in this empirical exercise can be downloaded from Amit Goyal’s website http://www.bus.emory.edu/agoyal/Research.html.
a small rolling window of $M = 60$ observations is used to construct an estimate of the conditional variance while the recursive scheme is used to estimate the conditional means. The coefficient of relative risk aversion is set to 3 as in numerous existing studies (see, *inter alia*, Goyal and Welch, 2008; Campbell and Thompson, 2008 and Ferreira and Santa Clara, 2011). We compute the portfolio weights both unconstrained and winsorized by imposing a maximum value of the investment in the risky asset to 150% (i.e. $-0.5 \leq \hat{w}_{i,t} \leq 1.5$).

The results of our empirical exercise are reported in Table 3. When the performance fees are estimated using unconstrained weights (Panel A), virtually all of the predictive variables are unable to generate significant results. In fact only the default return spread exhibits small and positive performance fees that are significant at the 10% level and only when $\Omega$ is estimated one particular way ($\Omega_c$). The results are different when the portfolio weights are winsorized. In this case, five of the predictive variables imply positive performance fees that are statistically significant at the 10 percent level. Two of these, the term spread and inflation, are consistently significant across all three estimates of the asymptotic variance $\Omega$. Nonetheless, the size of the performance fees is small in value and does not exceed 0.18 percent per month. It is important to emphasize that the statistical significance of the term spread as a predictive variable for equity returns corroborates the early findings documented in Campbell (1987) and Fama and French (1989).

Overall the results reinforce the primary point of this paper. The mere evidence of a positive estimated performance fee does not provide conclusive evidence of superiority of a given predictive model against a given (in this case, no-predictability) benchmark. Both panels of Table 3 provide examples of estimated performance fees that are positive and yet are not significantly different from zero based on a standard z-score-based approach to inference. Since this level of rigor is commonly used when reporting evidence of statistical (MSE-based) predictability, it seems only fitting to do the same when reporting evidence of economic (performance fee-based) predictability. This is particularly true given that the data used to construct the MSEs is the same as that used to construct the performance fees!
7 Conclusion

Out-of-sample methods are a common approach to evaluating the predictive content of a model. As such, a healthy literature has developed that provides methods for conducting inference on measures of forecast accuracy. This literature is almost completely focused on statistical measures. Economic measures of predictive accuracy are becoming increasingly common and they are used to complement the evidence provided by statistical measure. In this paper we derive asymptotics that can be used to conduct inference on one economic measure of forecast quality – performance fees. In particular, building on the theoretical results in West (1996), we are able to establish that these performance fee measures are asymptotically normal with an asymptotic variance that is affected by parametric estimation error. Monte Carlo evidence suggests that the theoretical results can be useful but also suggest that large samples are sometimes required.
8 Appendix

To understand the results reported in Section 3, it is instructive to consider the forms of \( \Phi \) and \( f_{t+\tau}(\beta_t) \) for three commonly used functional forms for utility: mean-variance, quadratic, and power. In addition, we also characterize the moment \( F \) when the two models are nested under the null hypothesis and hence \( \beta_{1,1}^* = 0 \).

1. When utility is mean-variance the average utility obtained using model \( i = 0, 1 \) is

\[
\bar{U}(\bar{R}_i) = \bar{R}_i - \frac{\gamma}{2} \bar{P}^{-1} \sum_{t=1}^{T+P-\tau} (\bar{R}_{i,t+\tau} - \bar{R}_i)^2
\]

where \( \bar{R}_i = \bar{P}^{-1} \sum_{t=1}^{T+P-\tau} \bar{R}_{i,t+\tau} \) and \( \gamma \) is a known preference parameter. For this functional form we trivially obtain

\[
\Phi = \bar{U}(\bar{R}_1) - \bar{U}(\bar{R}_0).
\]

As stated in the text, for this utility function Assumption 1 is satisfied for the function

\[
f_{t+\tau}(\beta_t) = (\bar{R}_{1,t+\tau} - \frac{\gamma}{2}(\bar{R}_{1,t+\tau} - E\bar{R}_{1,t+\tau})^2) - (\bar{R}_{0,t+\tau} - \frac{\gamma}{2}(\bar{R}_{0,t+\tau} - E\bar{R}_{0,t+\tau})^2)
\]

if \( \bar{R}_i \to^p E\bar{R}_{i,t+\tau} \). When the models are nested straightforward algebra implies

\[
F = \left( -E \frac{\partial U_{0,t+\tau}}{\partial \beta_0}, E \frac{\partial U_{1,t+\tau}}{\partial \beta_1} \right)
\]

where \( E \frac{\partial U_{i,t+\tau}}{\partial \beta_i} = E \frac{\partial U_{i,t+\tau}}{\partial \beta_i} ep_{t+\tau}(1 - \gamma(R_{i,t+\tau} - E\bar{R}_{i,t+\tau})) \) \( i = 0, 1 \).

2. When utility is quadratic, the average utility obtained using models \( i = 0, 1 \) takes the similar but distinct form

\[
\bar{U}(\bar{R}_i) = \bar{R}_i - \frac{\gamma}{2} \bar{P}^{-1} \sum_{t=1}^{T+P-\tau} \bar{R}_{i,t+\tau}^2
\]

For this utility function there are actually two roots that satisfy the definition of \( \Phi \). If we use the larger of the two as our estimate of \( \Phi \) we obtain the following closed form for the performance fee

\[
\Phi = (\bar{R}_1 - (2\gamma)^{-1}) + [(\bar{R}_1 - (2\gamma)^{-1})^2 + \gamma^{-1}(\bar{U}(\bar{R}_1) - \bar{U}(\bar{R}_0))]^{1/2}.
\]
For this utility function Assumption 1 is satisfied for the function

\[ f_{t+\tau}(\beta_t) = ((\hat{R}_{1,t+\tau}^2 - \gamma \hat{R}_{1,t+\tau}^2) - (\hat{R}_{0,t+\tau}^2 - \gamma \hat{R}_{0,t+\tau}^2))/(1 + 2\gamma E(R_{1,t+\tau})) \]

if \(-1 + 2\gamma \hat{R}_1 - P - 1 + 2\gamma E(R_{1,t+\tau}) \neq 0\). When the models are nested straightforward algebra implies

\[ F = (-E \frac{\partial w'_{0,t+\tau}}{\partial \beta_0} e^{p_{t+\tau}}(1 - 2\gamma R_{0,t+\tau}), E \frac{\partial w'_{1,t+\tau}}{\partial \beta_1} e^{p_{t+\tau}}(1 - 2\gamma (R_{1,t+\tau}))')/(1 + 2\gamma E(R_{1,t+\tau})). \]

3. When utility is power the average utility obtained using model \(i = 0, 1\) is

\[ \hat{U}(\hat{R}_i) = \frac{\hat{P}^{-1} \sum_{t=T}^{T+P-\tau} \hat{R}_{i,t+\tau}^{1-\gamma}}{1 - \gamma}. \]

For this functional form of utility we do not obtain a closed form for the performance fee \(\hat{\Phi}\).

Estimating \(\hat{\Phi}\) is done numerically using the definition and hence we have

\[ \hat{\Phi} = \arg_{\Phi} \text{root}(\hat{P}^{-1} \sum_{t=T}^{T+P-\tau} (\hat{R}_{1,t+\tau} - \Phi)^{1-\gamma} - \hat{U}(\hat{R}_0)). \]

As stated in the text, for this utility function Assumption 1 is satisfied for the function

\[ f_{t+\tau}(\beta_t) = (\hat{R}_{1,t+\tau}^{1-P} - \hat{R}_{0,t+\tau}^{1-P})/(E(\partial U(R_{1,t+\tau})/\partial \Phi)(1 - \rho)) \]

if \(\hat{P}^{-1} \sum_{t=T}^{T+P-\tau} \partial U(\hat{R}_{1,t+\tau} - \Phi)/\partial \Phi \rightarrow -P E(\partial U(R_{1,t+\tau})/\partial \Phi) \neq 0\). When the models are nested straightforward algebra implies

\[ F = (-E \frac{\partial w_{0,t+\tau}}{\partial \beta_0} e^{p_{t+\tau}} R_{0,t+\tau}^{-\gamma}, E \frac{\partial w_{1,t+\tau}}{\partial \beta_1} e^{p_{t+\tau}} (R_{1,t+\tau})^{-\gamma})'/E(R_{1,t+\tau})^{-\gamma}. \]
9 References


Table 1. Actual Size of Nominal 5% tests for DGP 1

The data generating process is a bivariate VAR(1) with coefficients and error variance given in Section 5 of the main text. Notably, the intercept in the equation for the equity premium is 0.0252 and the slope coefficient on the dividend yield is 0. The left-hand set of three sub-panels corresponds to the case in which the portfolio weights are unconstrained while those on the right correspond to the case in which the weights are restricted between -0.5 and 1.5. In each set, the three subpanels report rejection frequencies when T+P=1024, 2048, and 4096 respectively. Columns denote different sample splits P/T and different levels of persistence in z. Rows denote different sampling schemes and different estimates of the asymptotic variance. Each value is the percentage of rejections out of 2,000 replications.

<table>
<thead>
<tr>
<th></th>
<th>Not winsorized</th>
<th></th>
<th>Winsorized</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho = 0.5$</td>
<td>$\rho = 0.9$</td>
<td>$\rho = 0.5$</td>
<td>$\rho = 0.9$</td>
</tr>
<tr>
<td>$T + P = 1024$</td>
<td>$P/T$</td>
<td>$1/3$</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Recursive</td>
<td>$\Omega_a$</td>
<td>0.006 0.004 0.005</td>
<td>0.007 0.008 0.008</td>
<td>0.006 0.004 0.005</td>
</tr>
<tr>
<td></td>
<td>$\Omega_b$</td>
<td>0.074 0.075 0.058</td>
<td>0.066 0.073 0.056</td>
<td>0.074 0.075 0.058</td>
</tr>
<tr>
<td></td>
<td>$\Omega_c$</td>
<td>0.069 0.073 0.069</td>
<td>0.072 0.072 0.072</td>
<td>0.069 0.073 0.069</td>
</tr>
<tr>
<td>Mixed</td>
<td>$\Omega_a$</td>
<td>0.006 0.006 0.003</td>
<td>0.008 0.005 0.006</td>
<td>0.005 0.006 0.003</td>
</tr>
<tr>
<td></td>
<td>$\Omega_b$</td>
<td>0.082 0.081 0.051</td>
<td>0.068 0.059 0.049</td>
<td>0.082 0.081 0.051</td>
</tr>
<tr>
<td></td>
<td>$\Omega_c$</td>
<td>0.065 0.075 0.071</td>
<td>0.078 0.065 0.076</td>
<td>0.065 0.075 0.071</td>
</tr>
<tr>
<td>$T + P = 2048$</td>
<td>$P/T$</td>
<td>$1/3$</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Recursive</td>
<td>$\Omega_a$</td>
<td>0.016 0.011 0.004</td>
<td>0.015 0.014 0.011</td>
<td>0.016 0.011 0.004</td>
</tr>
<tr>
<td></td>
<td>$\Omega_b$</td>
<td>0.192 0.200 0.110</td>
<td>0.186 0.177 0.121</td>
<td>0.192 0.200 0.110</td>
</tr>
<tr>
<td></td>
<td>$\Omega_c$</td>
<td>0.085 0.066 0.075</td>
<td>0.074 0.080 0.093</td>
<td>0.085 0.066 0.075</td>
</tr>
<tr>
<td>Mixed</td>
<td>$\Omega_a$</td>
<td>0.020 0.013 0.004</td>
<td>0.013 0.013 0.009</td>
<td>0.020 0.013 0.004</td>
</tr>
<tr>
<td></td>
<td>$\Omega_b$</td>
<td>0.215 0.207 0.119</td>
<td>0.171 0.164 0.111</td>
<td>0.215 0.207 0.119</td>
</tr>
<tr>
<td></td>
<td>$\Omega_c$</td>
<td>0.084 0.075 0.078</td>
<td>0.076 0.081 0.083</td>
<td>0.084 0.075 0.078</td>
</tr>
<tr>
<td>$T + P = 4096$</td>
<td>$P/T$</td>
<td>$1/3$</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Recursive</td>
<td>$\Omega_a$</td>
<td>0.033 0.028 0.017</td>
<td>0.027 0.022 0.014</td>
<td>0.033 0.028 0.017</td>
</tr>
<tr>
<td></td>
<td>$\Omega_b$</td>
<td>0.301 0.328 0.277</td>
<td>0.281 0.317 0.256</td>
<td>0.301 0.328 0.277</td>
</tr>
<tr>
<td></td>
<td>$\Omega_c$</td>
<td>0.080 0.079 0.079</td>
<td>0.066 0.064 0.075</td>
<td>0.080 0.079 0.079</td>
</tr>
<tr>
<td>Mixed</td>
<td>$\Omega_a$</td>
<td>0.036 0.031 0.019</td>
<td>0.026 0.021 0.013</td>
<td>0.036 0.031 0.019</td>
</tr>
<tr>
<td></td>
<td>$\Omega_b$</td>
<td>0.314 0.337 0.282</td>
<td>0.280 0.315 0.252</td>
<td>0.314 0.337 0.282</td>
</tr>
<tr>
<td></td>
<td>$\Omega_c$</td>
<td>0.081 0.082 0.087</td>
<td>0.067 0.064 0.067</td>
<td>0.081 0.082 0.087</td>
</tr>
</tbody>
</table>
Table 2. Actual Size of Nominal 5% tests for DGP 2

The data generating process is a bivariate VAR(1) with coefficients and error variance given in Section 5. Notably, the intercept in the equation for the equity premium is -0.5. The left-hand set of three sub-panels corresponds to the case in which the null holds because $b = 0$ while the right-hand side corresponds to the case in which the null holds despite $b > 0$. In each set, the three subpanels report rejection frequencies when $T+P = 1024$, 2048, and 4096 respectively. Columns denote different sample splits $P/T$ and different levels of persistence in $z$. Rows denote different sampling schemes and different estimates of the asymptotic variance. Each value is the percentage of rejections out of 2,000 replications.

<table>
<thead>
<tr>
<th></th>
<th>$b = 0$</th>
<th>$b &gt; 0$</th>
<th>$b = 0$</th>
<th>$b &gt; 0$</th>
<th>$b = 0$</th>
<th>$b &gt; 0$</th>
<th>$b = 0$</th>
<th>$b &gt; 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T + P = 1024$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recursive</td>
<td>$\Omega_a$</td>
<td>$0.107$</td>
<td>$0.060$</td>
<td>$0.022$</td>
<td>$0.112$</td>
<td>$0.071$</td>
<td>$0.023$</td>
<td>$0.101$</td>
</tr>
<tr>
<td></td>
<td>$\Omega_b$</td>
<td>$0.360$</td>
<td>$0.343$</td>
<td>$0.264$</td>
<td>$0.337$</td>
<td>$0.322$</td>
<td>$0.243$</td>
<td>$0.104$</td>
</tr>
<tr>
<td></td>
<td>$\Omega_c$</td>
<td>$0.149$</td>
<td>$0.135$</td>
<td>$0.137$</td>
<td>$0.169$</td>
<td>$0.160$</td>
<td>$0.169$</td>
<td>$0.223$</td>
</tr>
<tr>
<td>Mixed</td>
<td>$\Omega_a$</td>
<td>$0.047$</td>
<td>$0.026$</td>
<td>$0.009$</td>
<td>$0.042$</td>
<td>$0.032$</td>
<td>$0.012$</td>
<td>$0.079$</td>
</tr>
<tr>
<td></td>
<td>$\Omega_b$</td>
<td>$0.220$</td>
<td>$0.205$</td>
<td>$0.179$</td>
<td>$0.142$</td>
<td>$0.153$</td>
<td>$0.158$</td>
<td>$0.093$</td>
</tr>
<tr>
<td></td>
<td>$\Omega_c$</td>
<td>$0.155$</td>
<td>$0.153$</td>
<td>$0.143$</td>
<td>$0.148$</td>
<td>$0.130$</td>
<td>$0.150$</td>
<td>$0.203$</td>
</tr>
<tr>
<td>$T + P = 2048$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recursive</td>
<td>$\Omega_a$</td>
<td>$0.125$</td>
<td>$0.107$</td>
<td>$0.056$</td>
<td>$0.130$</td>
<td>$0.109$</td>
<td>$0.056$</td>
<td>$0.075$</td>
</tr>
<tr>
<td></td>
<td>$\Omega_b$</td>
<td>$0.456$</td>
<td>$0.464$</td>
<td>$0.400$</td>
<td>$0.454$</td>
<td>$0.456$</td>
<td>$0.396$</td>
<td>$0.082$</td>
</tr>
<tr>
<td></td>
<td>$\Omega_c$</td>
<td>$0.116$</td>
<td>$0.117$</td>
<td>$0.132$</td>
<td>$0.123$</td>
<td>$0.124$</td>
<td>$0.131$</td>
<td>$0.201$</td>
</tr>
<tr>
<td>Mixed</td>
<td>$\Omega_a$</td>
<td>$0.070$</td>
<td>$0.050$</td>
<td>$0.013$</td>
<td>$0.042$</td>
<td>$0.033$</td>
<td>$0.015$</td>
<td>$0.053$</td>
</tr>
<tr>
<td></td>
<td>$\Omega_b$</td>
<td>$0.323$</td>
<td>$0.314$</td>
<td>$0.233$</td>
<td>$0.172$</td>
<td>$0.187$</td>
<td>$0.163$</td>
<td>$0.066$</td>
</tr>
<tr>
<td></td>
<td>$\Omega_c$</td>
<td>$0.131$</td>
<td>$0.131$</td>
<td>$0.122$</td>
<td>$0.157$</td>
<td>$0.142$</td>
<td>$0.139$</td>
<td>$0.150$</td>
</tr>
<tr>
<td>$T + P = 4096$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recursive</td>
<td>$\Omega_a$</td>
<td>$0.108$</td>
<td>$0.107$</td>
<td>$0.086$</td>
<td>$0.107$</td>
<td>$0.109$</td>
<td>$0.082$</td>
<td>$0.058$</td>
</tr>
<tr>
<td></td>
<td>$\Omega_b$</td>
<td>$0.492$</td>
<td>$0.491$</td>
<td>$0.482$</td>
<td>$0.477$</td>
<td>$0.471$</td>
<td>$0.480$</td>
<td>$0.063$</td>
</tr>
<tr>
<td></td>
<td>$\Omega_c$</td>
<td>$0.089$</td>
<td>$0.099$</td>
<td>$0.100$</td>
<td>$0.096$</td>
<td>$0.102$</td>
<td>$0.108$</td>
<td>$0.188$</td>
</tr>
<tr>
<td>Mixed</td>
<td>$\Omega_a$</td>
<td>$0.091$</td>
<td>$0.064$</td>
<td>$0.034$</td>
<td>$0.049$</td>
<td>$0.032$</td>
<td>$0.012$</td>
<td>$0.025$</td>
</tr>
<tr>
<td></td>
<td>$\Omega_b$</td>
<td>$0.433$</td>
<td>$0.421$</td>
<td>$0.346$</td>
<td>$0.168$</td>
<td>$0.178$</td>
<td>$0.166$</td>
<td>$0.038$</td>
</tr>
<tr>
<td></td>
<td>$\Omega_c$</td>
<td>$0.106$</td>
<td>$0.107$</td>
<td>$0.109$</td>
<td>$0.150$</td>
<td>$0.133$</td>
<td>$0.122$</td>
<td>$0.118$</td>
</tr>
</tbody>
</table>
The Table reports performance fees, $\Phi$ based on out-of-sample forecasts of the conditional mean of stock index excess returns from predictive models with alternative predictive variables (Variables) against the benchmark represented by the historical mean of excess returns. Performance fees denote the amount investors with a mean-variance utility function and a coefficient of Relative Risk Aversion (RRA) $\gamma = 3$ would be willing to pay to switch from each one of the predictive models to the historical average benchmark. The allocation to the risky asset is computed using a mixed scheme where the conditional mean of excess returns is forecast using a recursive scheme and the variance of excess returns is estimated using the past 60 observations of the sample. The predictive variables are a sample of the ones used in Goyal and Welch (2008). The sample period is from January 1871 until December 2011 and the forecasts begin in January 1965. $\Phi$ are expressed as percentage per month. Values in brackets are the $p$-values for the null hypothesis that $H_0 : \Phi = 0$. The $p$-values are computed using different $\Omega$ schemes as discussed in Section 4 of the main text. When weights are winsorized, $-0.5 \leq \hat{w}_{i,t} \leq 1.5$ (i.e. a cap of 150%).

Panel A) Without winsorization

<table>
<thead>
<tr>
<th>Variables</th>
<th>$\Phi$</th>
<th>$\Omega_a$</th>
<th>$\Omega_b$</th>
<th>$\Omega_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>svar</td>
<td>-0.195</td>
<td>[0.59]</td>
<td>[0.86]</td>
<td>[0.59]</td>
</tr>
<tr>
<td>d/e</td>
<td>-0.192</td>
<td>[0.99]</td>
<td>[1.00]</td>
<td>[1.00]</td>
</tr>
<tr>
<td>lty</td>
<td>-0.055</td>
<td>[0.56]</td>
<td>[0.56]</td>
<td>[0.59]</td>
</tr>
<tr>
<td>tms</td>
<td>-0.005</td>
<td>[0.50]</td>
<td>[0.50]</td>
<td>[0.53]</td>
</tr>
<tr>
<td>ltr</td>
<td>-0.097</td>
<td>[0.72]</td>
<td>[0.72]</td>
<td>[1.00]</td>
</tr>
<tr>
<td>infl</td>
<td>0.015</td>
<td>[0.46]</td>
<td>[0.46]</td>
<td>[0.42]</td>
</tr>
<tr>
<td>tbl</td>
<td>0.030</td>
<td>[0.46]</td>
<td>[0.46]</td>
<td>[0.44]</td>
</tr>
<tr>
<td>dfr</td>
<td>0.051</td>
<td>[0.39]</td>
<td>[0.39]</td>
<td>[0.06]</td>
</tr>
<tr>
<td>dyf</td>
<td>-0.089</td>
<td>[0.65]</td>
<td>[0.67]</td>
<td>[0.69]</td>
</tr>
<tr>
<td>ntis</td>
<td>-1.476</td>
<td>[0.90]</td>
<td>[0.94]</td>
<td>[0.94]</td>
</tr>
<tr>
<td>d/p</td>
<td>0.056</td>
<td>[0.43]</td>
<td>[0.43]</td>
<td>[0.40]</td>
</tr>
<tr>
<td>d/y</td>
<td>-0.069</td>
<td>[0.56]</td>
<td>[0.56]</td>
<td>[0.59]</td>
</tr>
<tr>
<td>e/p</td>
<td>-0.126</td>
<td>[0.59]</td>
<td>[0.60]</td>
<td>[0.63]</td>
</tr>
<tr>
<td>b/m</td>
<td>-0.693</td>
<td>[0.90]</td>
<td>[0.91]</td>
<td>[0.97]</td>
</tr>
</tbody>
</table>
Panel B) With winsorization

<table>
<thead>
<tr>
<th>Variables</th>
<th>$\Phi$</th>
<th>$\Omega_a$</th>
<th>$\Omega_b$</th>
<th>$\Omega_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>svar Stock Variance</td>
<td>-0.010</td>
<td>0.50</td>
<td>0.70</td>
<td>0.50</td>
</tr>
<tr>
<td>d/e Dividend Payout Ratio</td>
<td>-0.038</td>
<td>0.81</td>
<td>0.89</td>
<td>0.93</td>
</tr>
<tr>
<td>lty Long-term Yield</td>
<td>0.107</td>
<td>0.32</td>
<td>0.31</td>
<td>0.22</td>
</tr>
<tr>
<td>tms Term Spread</td>
<td>0.177</td>
<td>0.05</td>
<td>0.03</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>ltr Long-term Return</td>
<td>0.104</td>
<td>0.20</td>
<td>0.15</td>
<td>0.03</td>
</tr>
<tr>
<td>infl Inflation</td>
<td>0.131</td>
<td>0.03</td>
<td>0.02</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>tbl T-Bill Rate</td>
<td>0.171</td>
<td>0.19</td>
<td>0.19</td>
<td>0.06</td>
</tr>
<tr>
<td>dfr Default Return Spread</td>
<td>0.069</td>
<td>0.34</td>
<td>0.22</td>
<td>0.28</td>
</tr>
<tr>
<td>dfy Default Yield Spread</td>
<td>0.074</td>
<td>0.20</td>
<td>0.05</td>
<td>0.18</td>
</tr>
<tr>
<td>ntis Net Equity Expansion</td>
<td>0.051</td>
<td>0.41</td>
<td>0.19</td>
<td>0.41</td>
</tr>
<tr>
<td>d/p Dividend-Price Ratio</td>
<td>-0.054</td>
<td>0.60</td>
<td>0.60</td>
<td>0.65</td>
</tr>
<tr>
<td>d/y Dividend Yield</td>
<td>0.018</td>
<td>0.47</td>
<td>0.47</td>
<td>0.45</td>
</tr>
<tr>
<td>e/p Earning-Price Ratio</td>
<td>0.152</td>
<td>0.26</td>
<td>0.25</td>
<td>0.16</td>
</tr>
<tr>
<td>b/m Book-Market Ratio</td>
<td>-0.130</td>
<td>0.69</td>
<td>0.70</td>
<td>0.75</td>
</tr>
</tbody>
</table>
Notes: 1. The data generating process is a bivariate VAR(1) with coefficients and error variance given in Section 5. Notably, the intercept in the equation for the equity premium is -0.5 while the slope coefficient on the dividend yield is allowed to increase from 0.0 to 1.4.  
2. For each artificial dataset, conditional mean and conditional variance forecasts are generated separately from models 0 and 1 using the recursive scheme.  
3. The first row of panels report rejection frequencies when $T+P=1024$. The second row reports the same but when $T+P=2048$. Columns denote different levels of persistence in $z$. Within each plot there are three different lines one each associated with $P/T = 1/3$, 1, and 3.  
4. Each value is the percent of rejections out of 2,000 replications.
Notes: 1. The data generating process is a bivariate VAR(1) with coefficients and error variance given in Section 5. Notably, the intercept in the equation for the equity premium is -0.5 while the slope coefficient on the dividend yield is allowed to increase from 0.0 to 1.4.
2. For each artificial dataset, conditional mean and conditional variance forecasts are generated separately from models 0 and 1 using the mixed scheme.
3. The first row of panels report rejection frequencies when $T+P=1024$. The second row reports the same but when $T+P=2048$. Columns denote different levels of persistence in $z$. Within each plot there are three different lines one each associated with $P/T = 1/3, 1,$ and $3$.
4. Each value is the percent of rejections out of 2,000 replications.