Research Division
Federal Reserve Bank of St. Louis
Working Paper Series

Testing the Economic Value of Asset Return Predictability

Michael W. McCracken
and
Giorgio Valente

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

October 2012

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

The views expressed are those of the individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

Federal Reserve Bank of St. Louis Working Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to Federal Reserve Bank of St. Louis Working Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors.

Testing the Economic Value of Asset Return Predictability*

Michael W. McCracken[a] and Giorgio Valente[b]

a: Federal Reserve Bank of St. Louis

b: Essex Business School

October 2012

Abstract

Economic value calculations are increasingly used to compare the predictive performance of competing models of asset returns. However, they lack a rigorous way to validate their evidence. This paper proposes a new methodology to test whether utility gains accruing to investors using competing predictive models are equal to zero. Monte Carlo evidence indicates that our testing procedure, that can account for estimation error in the asymptotic variance of the test statistic, provides accurately sized and powerful tests in empirically relevant sample sizes. We apply the test statistics proposed in the paper to revisit the predictability of the US equity premium by means of various predictors.

JEL classification: C53, C12, C52

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

*This work was partly written while Giorgio Valente was visiting the Federal Reserve Bank of St Louis and City University of Hong Kong, whose hospitality is gratefully acknowledged. We gratefully acknowledge helpful comments from Andrew Patton, Enrique Sentana 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.

†*McCracken*: Research Division; Federal Reserve Bank of St. Louis; P.O. Box 442; St. Louis, MO 63166; michael.w.mccracken@stls.frb.org. *Valente (corresponding author)*: Finance Group; Essex Business School, University of Essex; Wivenhoe Park, Colchester, CO4 3SQ, Essex, United Kingdom; gvalente@essex.ac.uk.

1 Introduction

The out-of-sample predictability of asset returns is a topic of particular interest among academics and market practitioners. In past decades, various studies have debated whether asset returns are or should be predictable by information available to investors. Although a large body of literature accepts that asset returns are predictable (see, *inter alia*, Barberis, 2000; Lettau and Ludvigson, 2001; Campbell and Thompson, 2008; Cochrane, 2008; 2011; Ferreira and Santa Clara; 2011 and the references therein) and defines this as a "new fact in finance" (Cochrane, 1999), others remain skeptical (see, Goyal and Welch, 2008 and the references therein). Even with the benefit of almost a century of hindsight, this issue has not been convincingly settled: evidence that asset returns predictions from empirical models are better than forecasts from naive benchmarks, such as an historical mean or a random walk model, is not conclusive.

Most of the existing research on the ability of empirical models to predict asset returns relies on *statistical* measures of out-of-sample predictive accuracy, mostly under the assumption of mean squared forecast error loss function.[1] However, in the past decade, some attention has been directed to assessing whether there is any *economic* value to asset return predictability. Leitch and Tanner (1991) are among the first to show that the use of different metrics of evaluation based on utility calculations provides an alternative way to analyze the predictability of asset returns. This may shed light on aspects of the relationship (or lack of it) between returns and predictors which cannot be captured by standard statistical criteria. Following the same line of reasoning, various studies have proposed the analysis of performance fees (Φ henceforth) in order to understand how much risk-averse investors would be willing to pay to switch from static benchmarks to predictive dynamic models (West *et al.*, 1993; Fleming *et al.*, 2001; 2003; Han, 2006; Della Corte *et al.*, 2008; 2009; 2010; Thornton and Valente, 2012).

[1] Although such assumption is routinely adopted because the resulting properties of the optimal forecasts, it also imposes a set of limiting restrictions on the assessment of the out-of-sample predictions of competing models of asset returns. On the properties of optimal forecasts under asymmetric or unknown loss functions see Patton and Timmermann (2007a, 2007b). See also Patton (2004) on the effect of higher moments of returns for asset allocation decisions in out-of-sample settings.

A cursory reading of the papers published in five premier outlets[2] over the period 2000-2012, lead us to identify more than 40 studies that report economic value assessments of asset returns in various contexts. One common feature of these economic value calculations is that, in contrast to statistical assessments of predictability, they lack a rigorous way to validate their evidence. Put differently, in virtually all cases, economic value calculations are never accompanied by a p-value associated with the null hypothesis that those numbers are genuinely zero. An illuminating example is represented by the assessment of the out-of-sample predictability of the US equity risk premium. Some recent influential studies show that across different sample periods, data frequencies and for a large menu of predictive variables, the average (median) utility gain of using the predictive model against a simple historical average equals a meager 0.38 (0.34) percent per annum inclusive of transaction costs.[3] Nonetheless, the findings are often interpreted as supportive of the predictive models only on the basis of the sign of the realized utility gains. This state of things begs the question as to whether those small utility gains are genuinely different from zero if we take into account the fact that economic value calculations are affected by the estimation risk associated with the parameters in the predictive models.

This paper provides a first answer to this important question. In fact, we propose an asymptotically valid approach to constructing p-values associated with the null hypothesis that Φ is zero against the alternative that it is positive.[4] Using arguments in line with West (1996) and West and McCracken (1998), we show that the estimated performance fees $\hat{\Phi}$ are asymptotically normal with variances that account for the effects of estimated parameters. Since the variances are often difficult to estimate we advocate the use of the bootstrap of Calhoun (2011) which is particularly

[2] Namely, *Journal of Finance*, *Journal of Financial Economics*, *Review of Financial Studies*, *Journal of Business* and *Journal of Financial and Quantitative Analysis*.

[3] The figure are calculated using the results reported in Goyal and Welch (2008 and appendix; Table 3 p.1484 and p. 7 respectively), Campbell and Thompson (2008, Table 4 p. 1527) and Ferreira and Santa Clara (2011, Table 5 p. 530).

[4] While we focus on a point null hypothesis in which $\Phi = 0$, and test it against a composite alternative in which $\Phi > 0$, it is possible for their to be data-generating processes in which $\Phi < 0$. In this environment it might be more appropriate to test the composite null hypothesis $\Phi \leq 0$ against the composite alternative in which $\Phi > 0$. Doing so is significantly more complicated than the approach we take and is left to future research. See Section 5 for a detailed discussion of this issue.

well suited to this context. Monte Carlo results suggest that our testing procedure has satisfactory size and power properties for parameterizations and sample sizes that are consistent with those recorded in existing empirical studies.

The rest of the paper is organized as follows. Section 2 provides a simple framework that highlights the main features of economic value calculations, while in Sections 3 and 4 we describe the asymptotic results for our testing procedure and the bootstrap approach required to implement it. Next, in Section 5, we describe the setup of the Monte Carlo experiments and investigate the size and power properties of the test statistics. In Section 6 we discuss the drawbacks associated with economic value calculations and the potential limitations of the proposed testing procedure. In section 7 we apply the test statistics proposed in the paper to revisit the predictability of the equity premium by means of various predictors (as in Goyal and Welch, 2008) and a final section concludes.

2 A Simple Illustrative Example

In this section, we discuss a simple example of a economic value calculations based on a portfolio of assets that comprise a risk-free asset and a single risky asset (e.g. a stock index) that is predictable by means of a single variable. This simple case is chosen because of its simplicity and intuitiveness, and also since it is consistent with a very large literature on utility-based comparisons and asset returns predictability. However, it is important to emphasize that the testing procedure proposed in this paper applies to, but it is not limited to, this case. In fact it can be easily applied to portfolios where multiple predictive variables and multiple risky assets are included in the relevant computations.

Let r_t denote the return on the risky asset and r_t^f the rate of return on the risk-free asset. Define $ep_{t+\tau} = r_{t+\tau} - r_{t+\tau}^f$ 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 can use 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 is said to have predictive power for $ep_{t+\tau}$ if $\alpha_{1,1} \neq 0$. $e_{1,t+\tau}$ is an unpredictable error term given information available at time t. If the variable z_t has no predictive content then $\alpha_{1,1} = 0$ and Equation (1) collapses to

$$ep_{t+\tau} = \alpha_{0,0} + e_{0,t+\tau}, \tag{2}$$

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 also uses a model to estimate the conditional variance of the excess returns. Define $\sigma^2_{i,t+\tau}(\vartheta_i)$ as the excess return conditional variance implied by model $i = 0, 1$ that is a function of a vector of some parameter estimates $\widehat{\vartheta}_i$.

The investor can use the conditional mean predictions $ep_{t+\tau}(\widehat{\alpha}_{i,t})$ from Equations (1) or (2) together with the conditional variance predictions $\sigma^2_{i,t+\tau}\left(\widehat{\vartheta}_{i,t}\right)$ of excess returns to decide how much of her wealth is invested in the risky and the risk-free assets. The conditional predictions are functions of the parameters $\widehat{\alpha}_{i,t}$ that are equal to $\begin{bmatrix} \widehat{\alpha}_{1,0} & \widehat{\alpha}_{1,1} \end{bmatrix}$ if $i = 1$ or $\widehat{\alpha}_{0,0}$ if $i = 0$, and $\widehat{\vartheta}_{i,t}$ that are estimated over time using information available up to time t.

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:

$$w_{i,t}(\widehat{\beta}_{i,t}) = \hat{w}_{i,t} = \frac{ep_{t+\tau}(\widehat{\alpha}_{i,t})}{\gamma \sigma^2_{i,t+\tau}\left(\widehat{\vartheta}_{i,t}\right)}, \tag{3}$$

where $\widehat{\beta}_{i,t} = \begin{bmatrix} \widehat{\alpha}'_{i,t} & \widehat{\vartheta}'_{i,t} \end{bmatrix}'$ and γ is the investor's coefficient of relative risk aversion (RRA).

Once the asset allocation is decided at time $t = T, ..., T + P - \tau$, the resulting portfolio's actual returns at time $t + \tau$ are equal to:

$$\widehat{r}^{port}_{i,t+\tau} = r^{port}_{i,t+\tau}(\widehat{\beta}_{i,t}) = r^f_{t+\tau} + \hat{w}_{i,t}(\widehat{\beta}_{i,t}) ep_{t+\tau}. \tag{4}$$

The additional predictive ability in the competing model can be assessed against the baseline model by carrying out utility-based comparisons. More specifically, the average realized utility for the investor with a given initial wealth $W_0 = 1$ is denoted by

$$\overline{U}(\hat{R}_i^{port}) = \tilde{P}^{-1} \sum_{t=T}^{T+P-\tau} U(\hat{R}_{i,t+\tau}^{port}), \tag{5}$$

where $\hat{R}_{i,t+\tau}^{port} = 1 + \widehat{r}_{i,t+\tau}^{port}$ is the gross return on the portfolio constructed using the forecast from model i and $\tilde{P} = P - \tau + 1$.[5]

As in Fleming *et al.* (2001) the measure of the economic value of alternative predictive models is obtained by equating average utilities in Equation (5) from selected pairs of portfolios. More specifically, the performance fee is the value of $\hat{\Phi}$ that satisfies

$$\overline{U}(\hat{R}_1 - \hat{\Phi}) - \overline{U}(\hat{R}_0) = 0. \tag{6}$$

Put differently, Φ can be interpreted 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 Φ when $\alpha_{1,1} \neq 0$ 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

$$z_t = a + \rho z_{t-1} + v_t, \tag{7}$$

where (e_t, v_t) are i.i.d. normally distributed with zero means and variances σ_e^2 and σ_v^2. Let μ and σ_z^2 denote the unconditional mean and variance of z_t, respectively. Finally, assume that the conditional mean models are estimated by OLS and the conditional variance models are identical so that $\sigma_{0,t+1}^2(\widehat{\vartheta}_{0,t}) = \sigma_{1,t+1}^2(\widehat{\vartheta}_{1,t})$. More specifically, for ease of presentation, assume that the

[5] In order to simplify notation, we omit the superscript *port* throughout the remainder of the text.

predictions of these conditional variances are obtained simply by using a consistent estimate of the unconditional variance.[6]

Straightforward algebra shows that

$$\Phi = \left(\frac{\alpha_{1,1}^2 \sigma_z^2}{\gamma(\alpha_{1,1}^2 \sigma_z^2 + \sigma_e^2)} \right) \left(\frac{\sigma_e^2 - 3(\alpha_{1,0} + \alpha_{1,1}\mu)^2}{2(\alpha_{1,1}^2 \sigma_z^2 + \sigma_e^2)} \right). \tag{8}$$

Equation (8) 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 γ. Using this interpretation, the smaller the R^2 from the model 1, the closer Φ is to zero. The second term is less easily interpretable but, in large part, 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 Φ. 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 Φ equals zero nor is it trivially true that larger (absolute) values of $\alpha_{1,1}$ imply larger values of Φ. In fact, Equation (8) shows that the second term in parentheses also depends upon $\alpha_{1,1}$, and does so in a nonlinear fashion. Hence, it is theoretically possible that larger values of $\alpha_{1,1}$ do not imply larger values of Φ.

In Figure 1 we plot Φ as a function of $\alpha_{1,1}$ on the basis of parameter values calibrated using the return on the NYSE value weighted index as our risky asset, its dividend yield as our predictor, and the yield on the 3-month T-bill as our return on the risk-free asset.[7] Given these parameter

[6] While this may seem odd given our framework, rolling window estimates of the unconditional variance of ep_{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).

[7] Specifically, we use this data to estimate empirically relevant values of μ, σ_z^2, σ_e^2, and $E(ep_{t+1})$. This latter term

values, Figure 1 shows the values of Φ obtained as a function of $\alpha_{1,1} \in [-2, 2]$ where the maximum (absolute) value of the parameter $\alpha_{1,1}$ is chosen to be consistent with reasonable responses of the US equity premium to shocks affecting dividend yields (Cochrane, 2011 and the references therein) and the negative values are reported for the sake of completeness. As expected, Φ is increasing in $|\alpha_{1,1}|$. The plot is symmetric about zero. The plot also highlights that for an empirical value of $\alpha_{1,1} = 1\%$, the implied value of $\Phi \simeq 0.4\%$ on a monthly basis (or 4.8% annualized). Note that while a *population* value of $\Phi = 0.4\%$ is economically large, it does not imply that an *empirical* value of $\hat{\Phi} = 0.4\%$ is also statistically significant from zero at, say, a five percent level of significance. In the following two sections we provide a method to assess the statistical significance of empirical estimates of Φ.

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.[8] In addition to stating the theorem, we also provide informal discussions to clarify the various assumptions and their implications. The performance fee Φ 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 τ-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

is used to parameterize $\alpha_{1,0}$ as $\alpha_{1,0} = E(ep_{t+1}) - \alpha_{1,1}\mu$. The coefficient of relative risk aversion is set to $\gamma = 5$. Full details of the DGP are provided in Section 5 where Monte Carlo evidence on the size and power of our testing procedure is discussed.

[8] It is important to reiterate that while we sometimes refer to the simple case of two assets discussed in Section 2, our analytical results are more general and can be applied to instances where the investor's asset allocation is made over several assets.

below and closely follow those in West (1996) with some modest deviations.

1. 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 β and satisfies $\tilde{P}^{1/2}\hat{\Phi} = \tilde{P}^{-1/2}\sum_{t=T}^{T+P-\tau} f_{t+\tau}(\hat{\beta}_t) + o_p(1)$.

2. The parameters are estimated using one of three sampling schemes: the recursive, the rolling, or the fixed. 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-stochastic matrix, $H_i(t) = t^{-1}\sum_{s=1}^{t-\tau} h_{i,s+\tau}$ with $Eh_{i,s+\tau} = 0$ and β_i^* denote the population counterparts of the parameter estimates $\hat{\beta}_{i,t}$.The rolling and fixed parameter estimates $\hat{\beta}_{i,t}$ are defined similarly but are constructed using data over the ranges $s = t-T-\tau, ..., t-\tau$ and $s = 1, ..., T-\tau$ respectively. Define $\hat{\beta}_t = (\hat{\beta}_{0,t}', \hat{\beta}_{1,t}')'$, $h_{t+\tau} = (h_{0,t+\tau}', h_{1,t+\tau}')'$, and $B = diag(B_0, B_1)$. All parameter estimates are constructed using the same sampling scheme.[9]

3. Define $f_{t+\tau,\beta} = \partial f_{t+\tau}(\beta^*)/\partial\beta$. (a) $(f_{t+\tau,\beta}', f_{t+\tau}, h_{t+\tau}')'$ satisfies the mixing and moment conditions in Assumption 3 of West (1996) and (b) for some open neighborhood N of β^*, $E(\sup_{\beta\in N} \partial^2 f_{t+\tau}(\beta)/\partial\beta\partial\beta') < D$ some finite scalar D.

4. The number of in-sample observations associated with the initial forecast origin T, and the number of predictions $\tilde{P} = P - \tau + 1$, are arbitrarily large and in particular, they satisfy the restriction that $\lim_{P,T\rightarrow\infty} \frac{P}{T} = \pi \in (0, \infty)$,

5. Define $F = Ef_{t+\tau,\beta}$. If the models are nested then $FB \neq 0$.

[9] In many applications (Barberis, 2000; Fleming *et al.*, 2001; 2003; Della Corte *et al.*, 2010; Valente and Thornton, 2012 and the references therein) the conditional mean parameters are estimated under the recursive scheme while the predictions of the conditional variance are estimated using a simple rolling scheme. Specifically, for both models $i = 0, 1$ the conditional variance is estimated nonparametrically using a rolling window estimator of the form $M^{-1}\sum_{s=t-M-\tau+1}^{t-\tau}\left(ep_{s+\tau} - \bar{d}_t\right)^2$ where $\bar{d}_t = M^{-1}\sum_{s=t-M-\tau+1}^{t-\tau} ep_{s+\tau}$ and $M << T$ denotes the number of observations in the rolling window. Note that the conditional variance specifications are the same for both models and hence $\hat{\sigma}_{0,t+\tau}^2 = \hat{\sigma}_{1,t+\tau}^2$. If we interpret $M^{-1}\sum_{s=t-M-\tau+1}^{t-\tau}(ep_{s+\tau} - \bar{d}_t)^2$ as a parameter estimate that is consistent for some underlying moment (such as the unconditional variance of $ep_{s+\tau}$), then the following theorem is not directly applicable since the conditional mean and conditional variance parameters are estimated using different sampling schemes. However, if we interpret $M^{-1}\sum_{s=t-M-\tau+1}^{t-\tau}(ep_{s+\tau} - \bar{d}_t)^2$ as an *inconsistent* estimate of some underlying moment, as suggested in Giacomini and White (2006), then the theoretical results can be applicable with a suitable reinterpretation of what the "parameters" are.

Before deriving 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. 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

$$f_{t+\tau}(\hat{\beta}_t) = (\hat{R}_{1,t+\tau} - \frac{\gamma}{2}(\hat{R}_{1,t+\tau} - ER_{1,t+\tau})^2) - (\hat{R}_{0,t+\tau} - \frac{\gamma}{2}(\hat{R}_{0,t+\tau} - ER_{0,t+\tau})^2) \tag{9}$$

if $\bar{R}_i \rightarrow^p ER_{i,t+\tau}$. As another example suppose that power utility is used and hence $U(\hat{R}_{i,t+\tau}) = \hat{R}_{i,t+\tau}^{1-\rho}/(1-\rho)$. Since $\hat{\Phi}$ is defined as a root and $U(.)$ is continuously differentiable in its' argument we obtain $\hat{\Phi} = (\bar{U}(\hat{R}_1) - \bar{U}(\hat{R}_0))/\tilde{P}^{-1}\sum_{t=T}^{T+P-\tau} \partial U(\hat{R}_{1,t+\tau} - \tilde{\Phi})/\partial\Phi$ some $\tilde{\Phi}$ on the line between $\hat{\Phi}$ and 0.[10] If $\tilde{P}^{-1}\sum_{t=T}^{T+P-\tau} \partial U(\hat{R}_{1,t+\tau} - \tilde{\Phi})/\partial\Phi \rightarrow^p E(\partial U(R_{1,t+\tau})/\partial\Phi) \neq 0$, Assumption 1 is satisfied with

$$f_{t+\tau}(\hat{\beta}_t) = (\hat{R}_{1,t+\tau}^{1-\rho} - \hat{R}_{0,t+\tau}^{1-\rho})/(E(\partial U(R_{i,t+\tau})/\partial\Phi)(1-\rho)). \tag{10}$$

The requirement that $f_{t+\tau}(\beta)$ is twice continuously differentiable in β 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 β. 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 and the references therein).[11]

Given Assumption 1, Assumptions 2, 3, and 4 are very closely related 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,[12] and the number of in-sample and out-of-sample observations must be of the same

[10] Under the null hypothesis that $\Phi = 0$.

[11] We explore the effect of portfolio weights' winsorization in Section 5.

[12] West uses the central limit theorem of Wooldridge and White (1998).

order.[13] 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 $\hat{\Phi}$ to be zero. To avoid this problem we state a relatively high level assumption that will insure that the asymptotic variance is 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.[14] Given the assumptions, our main result follows immediately from Theorem 4.1 of West (1996).

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

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

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

scheme	$\Lambda_0(\pi)$	$\Lambda_1(\pi)$
recursive	$(1 - \pi^{-1}\ln(1+\pi))$	$2(1 - \pi^{-1}\ln(1+\pi))$
rolling $0 < \pi \leq 1$	$\frac{\pi}{2}$	$\pi - \frac{\pi^2}{3}$
rolling $1 \leq \pi < \infty$	$1 - \frac{1}{2\pi}$	$1 - \frac{1}{3\pi}$
fixed	0	π.

.

The Theorem shows that the per-period 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. As stated, the Theorem permits both non-nested and nested model comparisons. But as a practical matter, the vast majority of applications involve the comparison of two nested models – as was the case of our simple illustrative example in section 2.

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

[13] This could be weakened to permit $\pi = 0$ or ∞ in some cases if we distinguish between nested and non-nested models.

[14] This assumption precludes a few isolated applications including Cenesizoglu and Timmermann (2011).

the "competing" and "baseline" models are better thought of as "unrestricted" and "restricted." In this situation, $\beta_1^* \equiv (\beta_{1,0}^{*\prime}, \beta_{1,1}^{*\prime})' = (\beta_0^{*\prime}, 0)'$ and the asymptotic variance simplifies to

$$\Omega = \Lambda_1(\pi)(E\frac{\partial U_{1,t+\tau}}{\partial \Phi})^{-2}E(\frac{\partial U_{1,t+\tau}}{\partial \beta_1})(-JB_0J + B_1)S_{h_1h_1}(-JB_0J + B_1)E(\frac{\partial U_{1,t+\tau}'}{\partial \beta_1}).$$

The asymptotic variance simplifies because in this specific case, $R_{1,t+\tau} = R_{0,t+\tau}$ 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 $J'\frac{\partial U_{1,t+\tau}}{\partial \beta_1} = \frac{\partial U_{0,t+\tau}}{\partial \beta_0}$ and thus the asymptotic variance can be simplified even 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 Ω to be positive. When predictability exists and hence $\beta_{1,1}^*$ is not zero, S_{ff} is non-zero and hence Ω is non-zero so long as S_{ff} does not happen to cancel with the remaining terms. By imposing Assumption 5 we insure that Ω 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 $\Phi = 0$, despite the fact that $\beta_{1,1}^*$ is non-zero. In this case, as stated in the Theorem, $\hat{\Phi}$ is asymptotically normal with mean zero but with an asymptotic variance that does not simplify as it does when $\beta_{1,1}^* = 0$.

4 A Bootstrap Approach to 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 Ω it immediately follows that $\tilde{P}^{1/2}\hat{\Phi}/\hat{\Omega}^{1/2} \rightarrow^d N(0,1)$, and therefore one can use standard normal critical values to test the null that $\Phi = 0$ against an alternative in which $\Phi > 0$.

In our simple illustrative example discussed in Section 2, forming a consistent estimate of Ω is not too difficult. However, the standard errors become increasingly complicated to estimate as the number of risky assets increases and we move from mean-variance utility to other utility functions.[15]

[15] See Appendix for a detailed discussion of the implications of our testing procedure for three commonly used

In order to provide a methodology that can be applied to the relevant cases reported in the literature, we consider a bootstrap approach to inference. The bootstrap used in this paper is consistent with the one recently developed by Calhoun (2011; Section 2). We have chosen this specific bootstrap since it is explicitly designed to be applicable in cases when out-of-sample methods are used to conduct inference and the relevant test statistic is asymptotically normal. Furthermore, the bootstrap is designed to allow for estimation error in the standard errors. This final feature is non-trivial for our results. To see this, recall that the Theorem implies that $\hat{\Phi}$ is asymptotically normal with an asymptotic variance that is a linear combination of three elements: S_{ff}, FBS_{fh}, and $FBS_{hh}B'F'$. The latter two of these terms exist because the portfolio weights are functions of estimated parameters and are not known *a priori*. As a consequence, accounting for estimation error in the standard errors is crucial in this context.

Let X_t denote the vector of time t observables. In the simple example discussed in Section 2, this consists of $X_t = (ep_t, z_t)'$. The first stage of the bootstrap consists of using the moving blocks, circular blocks, or stationary bootstrap with block lengths drawn from the geometric distribution to generate the time series of bootstrapped observations $X_1^*...X_{T+P}^*$. In each method the block length b satisfies $b/P \to 0$ as $b, P \to \infty$. In the second stage the bootstrapped data is used to construct the bootstrapped performance fee measure $\hat{\Phi}^*$. This process is repeated many times so that we have a collection of $\hat{\Phi}_j^*$ $j = 1, ..., N$ of bootstrapped performance fees that can be used to estimate asymptotically valid critical values.

As in White (2000), we recenter each of the bootstrapped $\hat{\Phi}_j^*$ statistics in order to ensure that the empirical critical values remain bounded under the alternative hypothesis, $\Phi > 0$. Unfortunately, as shown in Corradi and Swanson (2007), White's approach to recentering does not allow for estimation error in the asymptotic distribution. The issue is that White recommends recentering each $\hat{\Phi}_j^*$ by the empirical value of the performance fee $\hat{\Phi}$ which is a function of an entire sequence of parameter estimates $\hat{\beta}_t$, $t = T..., T + P - \tau$. In contrast, Calhoun recommends recentering by

utility functions: mean-variance, quadratic and power.

the constant $\hat{\Phi}(\hat{\beta}_{T+P-\tau})$ which is constructed exactly as was $\hat{\Phi}$ but where $\hat{\beta}_t = \hat{\beta}_{T+P-\tau}$ for all $t = T..., T+P-\tau$. Given the recentering constant $\hat{\Phi}(\hat{\beta}_{T+P-\tau})$ and the bootstrapped performance fees $\hat{\Phi}^*_j$, critical values are then estimated based upon the empirical distribution of $\hat{\Phi}^*_j - \hat{\Phi}(\hat{\beta}_{T+P-\tau})$.

5 Monte Carlo Evidence

In the previous section we show that $\hat{\Phi}$ is asymptotically normal and delineate a bootstrap procedure to conducting 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 the bootstrap procedure for testing the null hypothesis $H_0 : \Phi = 0$ against the alternative $H_A : \Phi > 0$.

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 2.

The experiments are conducted as follows. For all cases, we generate data using the following data generating process (DGP henceforth) 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:

$$\begin{aligned} ep_{t+1} &= r_{t+1} - r^f_{t+1} = (0.00252 - 0.032b) + bz_t + u_{t+1}\sigma_t \\ z_t &= 0.032(1-\rho) + \rho z_{t-1} + e_t \\ r^f_{t+1} &= 0.0036 \end{aligned} \tag{11}$$

with

$$\begin{pmatrix} u_t \\ e_t \end{pmatrix} \sim i.i.d.N\left(0, \begin{bmatrix} 0.00172 & -0.00004 \\ -0.00004 & 0.000063(1-\rho^2) \end{bmatrix}\right).$$

We consider sample splits between the observations used to estimate the model and the out-of-sample period, $P/T = 1/3, 1$, and 3 for overall sample sizes $T + P = 512$ and 1024.[16] Note that we have introduced the term σ_t in the equation for excess returns. By doing so we allow for the presence of conditionally heteroskedastic errors in the equity premium equation. For reasons discussed later in section 6, we set $\sigma_t = |a_0 + a_1 z_t|$ with $a_0 = 0.73$ and $a_1 = 2.5$ such that $E\sigma_t^2 = 1$. The risk aversion parameter is set to 5.

In the experiments aiming at assessing the empirical size of the test statistics, we fix $b = 0$ while in the experiments aiming at assessing the power properties of the test we allow b to range from 0 to 2. In order to assess the role of persistence in the predictor z_t on the asymptotic distribution, for both the size and power experiments results are provided when the persistence is a modest $\rho = 0.5$ and a much higher $\rho = 0.9848$.

As in much of the empirical literature (see, *inter alia*, Goyal and Welch, 2008 and the references therein) we assume that the two competing conditional mean specifications are recursive OLS estimated models taking the form

$$ep_{t+1} = \alpha_{0,0} + \varepsilon_{0,t+1},$$

and

$$ep_{t+1} = \alpha_{1,0} + \alpha_{1,1} z_t + \varepsilon_{1,t+1},$$

where z_t is the dividend yield on the stock index. In order to understand the magnitude of size distortions associated with using a small rolling window estimate of the conditional variance we conduct each of our simulations once using a 5 year rolling window estimate of the conditional variance $60^{-1} \sum_{s=t-60}^{t-1} (ep_{s+1} - \bar{d}_t)^2$ and once using the *known* unconditional variance $0.00172 + 0.000063b^2$. In addition, in order to understand the potential size distortions associated with winsorizing the portfolio weights, each simulation is conducted allowing i) a free estimation of the

[16] We use a burnout period of 500 observations to remove the effects of initial conditions.

portfolio weights and ii) bounded weights such that

$$\hat{w}_{i,t}^{b} = \begin{cases} -1 \text{ if } \hat{w}_{i,t} \leq -1 \\ \hat{w}_{i,t} \text{ if } -1 < \hat{w}_{i,t} < 2 \\ 2 \text{ if } \hat{w}_{i,t} \geq 2 \end{cases}$$

where $\hat{w}_{i,t}^{b}$ and $\hat{w}_{i,t}$ denote the winsorized and unconstrained weights, respectively. All results are based on $2,000$ Monte Carlo replications with 499 bootstrap replications of the block bootstrap.

5.2 Simulation Results

In Tables 1 and 2 we report the actual rejection frequencies of nominally 5 and 10 percent tests of the null hypothesis that $\Phi = 0$. Figures in Table 1 are computed using a rolling window to estimate σ_t^2 while figures in Table 2 are computed by estimating σ_t^2 by means of the known unconditional variance. The results across the two tables are fairly similar. In fact, for both sample sizes, both degrees of correlation in the predictor (Panels A and B), and regardless of whether or not we winsorize the portfolio weights, the actual rejection frequencies are satisfactorily close to their nominal 5 and 10 percent values, especially when the sample split takes the value $P/T = 1/3$. When the sample split parameter P/T increases deteriorates. More specifically, as P/T increases the actual rejection frequencies become smaller than their nominal values. This results is likely to be due to a finite sample problem since as the overall sample size increases, from $T = 512$ to $T = 1024$, the actual rejection frequencies improve holding the sample split P/T constant. It is worthwhile emphasizing that since both tables give similar results, it appears that the test is generally unaffected by whether or not we use the rolling window estimate of the conditional variance.

In Figures 2 and 3 we provide plots of actual power of the test. Figure 2 provides rejection frequencies when the portfolio weights are estimated without any boundaries, while those in Figure 3 are calculated with winsorized weights. In each figure, there are four panels: two associated with the sample sizes $T + P = 512$, $1,024$ and two associated with the correlation in the dividend yield set to $\rho = 0.5$, 0.9848. For each of these four panels there are 2 sets of three power functions. Each of these three functions correspond to a distinct sample split parameter $P/T = 1/3, 1$, or 3.

The 2 sets of power functions are the same values but defined on distinct scales. The black lines correspond to the upper scale while the red lines correspond to the lower scale. In the upper scale we show actual power as we increase the parameter b from 0 to 2. In the lower scale we show the same values of power but plotted as a function of the value of the implied Φ as we increase b from 0 to 2.[17] For both figures all simulations correspond to the experiment design used to compute the rejection frequencies in Table 1, for which the conditional variance is estimated using a 5-year rolling window.

In Figure 2, the red and black sets of three power functions tend to increase as Φ and b increase, and to a large extent, power is increasing in the sample split parameter P/T. This latter finding is consistent with the argument suggested by Hansen and Timmermann (2011) in the context of tests of statistical significance. The few exceptions arise for smaller values of b and Φ and may be explained in light of the simulation results in Tables 2 and 3. In fact, in the tables have recorded that for larger values of the sample split parameter, actual rejection frequencies were significantly smaller than the nominal size of the test. In Figure 3, we provide the same set of simulations as in Figure 2 but with the portfolio weights now winsorized. In almost all cases the empirical power of the test statistics is uniformly higher than the one plotted in Figure 2, where the portfolio weights are not winsorized. And as was the case in Figure 2, power tends to be higher for larger sample sizes, larger values of ρ, and larger values of the sample split parameter P/T.

6 Discussion and Caveats

In the previous sections we have provided both analytical and simulation-based evidence on the efficacy of a test of the null that performance fees are zero against the alternative that they are positive. In particular we show that for empirically relevant sample sizes and data-generating processes, a suitable bootstrap approach to inference exhibits satisfactory empirical size and power properties. That said, there are important caveats to our results.

[17] For each value of b the simulation is conducted 1000 times. The values of Φ on the lower axis are the median estimated values of these 1000 replications.

First, in the context of the example we have used throughout as a foil for our results, Assumption 5 requires that $FB = E(\frac{\partial U_{1,t+\tau}}{\partial \beta_1})(-JB_0J + B_1)$ is non-zero. If this condition fails, our theoretical results do not hold and in particular, since $\hat{\Phi}$ need not be asymptotically normal and our bootstrap approach to inference may not be valid. In fact, straightforward algebra reveals that Assumption 5 fails if ep_{t+1} and ep_{t+1}^2 are *both* mean independent of $(1, z_t)$ and, as a consequence, the equity premium is the realization of a conditionally homoskedastic martingale process around some constant unconditional mean. It is for this reason that we model the conditional variance of the equity premium as $\sigma_t^2 = (a_0 + a_1 z_t)^2$ in our simulations. In unreported simulation results in which we set $\sigma_t^2 = 1$, we find that the bootstrap does not fare well and in particular leads to rejection frequencies of less than 1 percent for nominally 5 and 10 percent tests.

Second, in Figure 1 we show that for empirically relevant values of $\alpha_{1,1} \in [-2, 2]$, Φ is an increasing function of $\alpha_{1,1}$. As such, we expect power to increase monotonically in both Φ and $\alpha_{1,1}$. This is the result we find in Figures 2 and 3. While this result is intuitive, it need not hold in general for all values of $\alpha_{1,1}$. To see this, in Figure 4 we plot Equation (8) as a function of $\alpha_{1,1}$ for a much broader, but implausible, range of values. As in Figure 1, Φ is an increasing function of $|\alpha_{1,1}|$ for all empirically relevant values of $\alpha_{1,1}$ but is not uniformly so. Specifically, Φ is a hump-shaped function of $|\alpha_{1,1}|$ achieving a maximum at $\alpha_{1,1}$ near ± 5 and then converges to zero for larger values of $|\alpha_{1,1}|$. Since larger values of $|\alpha_{1,1}|$ imply greater statistical predictability, we find that it is theoretically possible that statistical significance of predictability is associated with no economic value. This finding is line with the theoretical evidence provided in Sentana (2005) who shows that predictive improvements do not necessarily generate economic improvements. In fact, he shows that the unconditional Sharpe ratio of a dynamic asset allocation strategy, where a mean-variance investors uses the forecasts of linear models with multiple predictors, could be beaten by a simple buy-and-hold strategy (Sentana, 2005 p. 63).[18]

Third, in all of our results we have taken care to parameterize our DGP using first and second

[18] See also the example provided in Section 3.4 of Sentana (1999).

moments calibrated on data series routinely employed in various existing studies on asset returns predictability. However, it is important to emphasize that our simulation results may not apply in all contexts.

For example, we have parameterized the DGP in Section 5 so that the equity premium has a constant unconditional mean across all values of $\alpha_{1,1}$. If we bring this assumption back to Equation (8), we find that

$$\Phi = \left(\frac{\alpha_{1,1}^2 \sigma_z^2}{\gamma(\alpha_{1,1}^2 \sigma_z^2 + \sigma_e^2)} \right) \left(\frac{\sigma_e^2 - 3(Eep_{t+1})^2}{2(\alpha_{1,1}^2 \sigma_z^2 + \sigma_e^2)} \right). \tag{12}$$

Once again $\Phi = 0$ if $\alpha_{1,1} = 0$. However if $\alpha_{1,1} \neq 0$, the sign of Φ depends on the sign of $\sigma_e^2 - 3(Eep_{t+1})^2$. In our data-based simulation it is always the case that $\sigma_e^2 - 3(Eep_{t+1})^2$ is positive. However, we cannot rule out the case that $\sigma_e^2 - 3(Eep_{t+1})^2$ is negative, and hence the plot of Φ as a function of $\alpha_{1,1}$, would be the mirror image of the one reported in Figure 4. This implies that while Φ would be still equal to zero when $\alpha_{1,1} = 0$, $\Phi \leq 0$ for all other values of $\alpha_{1,1}$. Under these circumstances, the simple point null hypothesis presented in Section 3 becomes inappropriate. This problem could be solved by setting the null hypothesis as a composite hypothesis of the form $\Phi \leq 0$ with composite alternative $\Phi > 0$. However, constructing an asymptotically valid test of this composite null hypothesis is significantly harder than the simple conservative approach we take (see Hansen, 2005 and the references therein) and we leave it as avenue for future research.

7 The Economic Value of the 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. The framework we use is identical to the one highlighted in Section 2. We use monthly, quarterly and annual 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).[19]

We compute the weights $\hat{w}_{i,t}\left(\widehat{\beta}_{i,t}\right)$ using a rolling variance of excess returns, $\widehat{\sigma}^2_{i,t=j}$, estimated, as in Ferreira and Santa-Clara (2011), using the past j observations of the sample and we estimate the performance fee Φ assuming a RRA coefficient of 3 as in 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. $-0.5 \leq \hat{w}_{i,t}\left(\widehat{\beta}_{i,t}\right) \leq 1.5$).

The results of our empirical exercise are reported in Table 3. When the performance fees are estimated using monthly data (Panel A) with a rolling variance computed over the past 5 or 10 years and the weights are left unconstrained, the vast majority of the predictive variables is unable to generate significant results. In fact only 3 variables, namely the term spread (**tms**), default return spread (**dfr**) and default yield spread (**dfy**), exhibit small and positive performance fees. However, only the term spread is closer to significance at the 10% level across both rolling variance windows. The results are substantially different if the portfolio weights are winsorized. In fact, in this case, the number of predictive variables that are able to deliver positive performance fees that are statistically significant at 10 percent level increases to 8, and 5 of those are statistically significant at the same statistical level using both rolling variance windows. Nonetheless, the size of the performance fees is small in value and does not exceed 0.14 percent per month.

When the data frequency decreases from monthly to annual, the number of statistically significant positive performance fees decreases. The reduction is more pronounced when portfolio weights are not winsorized and they are computed using a rolling window volatility based on the past 10 years of observations. It is important to emphasize that the term spread is the only predictive variable that generates positive performance fees that are consistently statistically significant across different specifications and data frequencies. This result is particularly interesting and it

[19] 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.

corroborates the early findings documented in Campbell (1987) and Fama and French (1989) on the predictive power of the term spread for equity returns.

Overall the results of this exercise suggest two important prescriptions for any study interested in estimating and assessing the economic value of predictability: First, the mere evidence of positive performance fees does not provide conclusive evidence of superiority of a given predictive model against a given (no predictability) benchmark. In fact our results show that different windows used to estimate rolling conditional variances, constraints imposed on portfolio weights and, more importantly, different data frequencies can substantially affect the size and the sign of the performance fees over the very same sample period. Second, it is important to accompany any estimate of positive performance fees with a formal test for their statistical significance. This would bring the evidence of economic values from various predictive models at par with the context where statistical criteria (such as MSE or RMSE) are employed and for which the asymptotic theory is already fully developed.

8 Conclusion

In this paper we attempt to provide a first answer to the issue of testing the economic value of asset return predictability. Although several recent studies have begun to investigate the economic value associated with the predictions of asset returns from empirical models, they lack a rigorous assessment of the null hypothesis that the utility gains originated from competing predictive models are equal or smaller than zero. This is surprising since economic value calculations stem from predictions from empirical models, which in turn are associated with the uncertainty due to the fact that models parameters are to be estimated.

We propose a formal test of the null hypothesis that economic value gains equal to zero against the alternative that they are positive. Using asymptotic arguments we show that, under modest assumptions, the test statistics are normally distributed. Monte Carlo evidence indicates that our testing procedure, which can account for estimation error in the asymptotic variance of the test

statistic, can provide accurately sized and powerful tests in empirically relevant sample sizes. We apply the test statistics proposed in the paper to revisit the predictability of the US equity premium by means of various predictors.

9 Appendix A

To understand the results reported in Section 3, it is instructive to consider the forms of $\hat{\Phi}$ and $f_{t+\tau}(\hat{\beta}_t)$ 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

$$\bar{U}(\hat{R}_i) = \bar{R}_i - \frac{\gamma}{2}\tilde{P}^{-1}\sum\nolimits_{t=T}^{T+P-\tau}(\hat{R}_{i,t+\tau} - \bar{R}_i)^2 \tag{13}$$

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

$$\hat{\Phi} = \bar{U}(\hat{R}_1) - \bar{U}(\hat{R}_0). \tag{14}$$

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

$$f_{t+\tau}(\hat{\beta}_t) = (\hat{R}_{1,t+\tau} - \frac{\gamma}{2}(\hat{R}_{1,t+\tau} - ER_{1,t+\tau})^2) - (\hat{R}_{0,t+\tau} - \frac{\gamma}{2}(\hat{R}_{0,t+\tau} - ER_{0,t+\tau})^2) \tag{15}$$

if $\bar{R}_i \rightarrow^p ER_{i,t+\tau}$. When the models are nested straightforward algebra implies

$$F = (-E\frac{\partial U'_{0,t+\tau}}{\partial \beta_0}, E\frac{\partial U'_{1,t+\tau}}{\partial \beta_1})' \tag{16}$$

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

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

$$\bar{U}(\hat{R}_i) = \bar{R}_i - \gamma\tilde{P}^{-1}\sum\nolimits_{t=T}^{T+P-\tau}\hat{R}^2_{i,t+\tau}. \tag{17}$$

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

$$\hat{\Phi} = (\bar{R}_1 - (2\gamma)^{-1}) + [(\bar{R}_1 - (2\gamma)^{-1})^2 + \gamma^{-1}(\bar{U}(\hat{R}_1) - \bar{U}(\hat{R}_0))]^{1/2}. \tag{18}$$

For this utility function Assumption 1 is satisfied for the function

$$f_{t+\tau}(\hat{\beta}_t) = ((\hat{R}_{1,t+\tau} - \gamma\hat{R}^2_{1,t+\tau}) - (\hat{R}_{0,t+\tau} - \gamma\hat{R}^2_{0,t+\tau}))/(-1 + 2\gamma E(R_{1,t+\tau})) \tag{19}$$

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

$$F = (-E\frac{\partial w'_{0,t+\tau}}{\partial\beta_0}ep_{t+\tau}(1-2\gamma R_{0,t+\tau}), E\frac{\partial w'_{1,t+\tau}}{\partial\beta_1}ep_{t+\tau}(1-2\gamma(R_{1,t+\tau})))'/(-1+2\gamma E(R_{1,t+\tau})). \tag{20}$$

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

$$\bar{U}(\hat{R}_i) = \frac{\tilde{P}^{-1}\sum_{t=T}^{T+P-\tau}\hat{R}^{1-\gamma}_{i,t+\tau}}{1-\gamma}. \tag{21}$$

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

$$\hat{\Phi} = \arg_\Phi root(\frac{\tilde{P}^{-1}\sum_{t=T}^{T+P-\tau}(\hat{R}_{1,t+\tau} - \Phi)^{1-\gamma}}{1-\gamma} - \bar{U}(\hat{R}_0)). \tag{22}$$

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

$$f_{t+\tau}(\hat{\beta}_t) = (\hat{R}^{1-\rho}_{1,t+\tau} - \hat{R}^{1-\rho}_{0,t+\tau})/(E(\partial U(R_{1,t+\tau})/\partial\Phi)(1-\rho)) \tag{23}$$

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

$$F = (-E\frac{\partial w_{0,t+\tau}}{\partial\beta_0}ep_{t+\tau}R^{-\gamma}_{0,t+\tau}, E\frac{\partial w_{1,t+\tau}}{\partial\beta_1}ep_{t+\tau}(R_{1,t+\tau})^{-\gamma})'/E(R_{1,t+\tau})^{-\gamma}. \tag{24}$$

10 References

Barberis, N. (2000), "Investing for the Long Run When Returns Are Predictable," *Journal of Finance*, **55**, 225-264.

Calhoun, G. (2011), "An Asymptotically Normal Out-of-Sample Test of Equal Predictive Accuracy for nested Models," Iowa State University working paper.

Campbell, J.Y. (1987), "Stock Returns and the Term Structure", *Journal of Financial Economics* **18**, 373-399.

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

Cenesizoglu, T. and Timmermann, A. (2011), "Do Return Prediction Models Add Economic Value?", UCSD working paper.

Cochrane, J.H. (1999), "New Facts In Finance," *Economic Perspectives*, **23**, 36-58.

Cochrane, J.H. (2008), "The Dog that Did Not Bark: A Defense of Return Predictability," *Review of Financial Studies* **21**, 1533-1575.

Cochrane, J.H. (2011), "Presidential Address: Discount Rates," *Journal of Finance* **66**, 1047-1108.

Corradi, V. and Swanson, N.R. (2007), "Nonparametric Bootstrap procedures for Predictive Inference Based on Recursive Estimation Schemes," *International Economic Review*, **48**, 67-109.

Della Corte, P., Sarno, L. and Thornton, D.L. (2008), "The Expectations Hypothesis of the Term Structure of Very Short-Term Rates: Statistical Tests and Economic Value," *Journal of Financial Economics,* **89**, 158-174.

Della Corte, P., Sarno, L. and Tsiakas, I. (2009), "An Economic Evaluation of Empirical Exchange Rate Models," *Review of Financial Studies,* **22**, 3491-3530.

Della Corte, P., Sarno, L. and Valente, G. (2010), "A Century of Equity Premium Predictability and the Consumption-Wealth Ratio: An International Perspective," *Journal of Empirical Finance* **17**, 313-331.

Fama, E.F. and French, K.R. (1989), "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of Financial Economics* **22**, 3-25.

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

Fleming, J., Kirby, C. and Ostdiek, B. (2001), "The Economic Value of Volatility Timing," *Journal of Finance,* **56**, 329-352.

Fleming, J., Kirby, C. and Ostdiek, B. (2003), "The Economic Value of Volatility Timing Using 'Realized' Volatility," *Journal of Financial Economics,* **67**, 473-509.

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

Goyal, A. and Welch, I. (2008), "A Comprehensive Look at the Empirical Performance of Equity Premium Predictions," Review of Financial Studies, 21, 1455-1508; see *additional appendix* at http://www.hec.unil.ch/agoyal/docs/AllTables2009.pdf.

Kandel, S. and Stambaugh, R. F. (1996), "On the Predictability of Stock Returns: An Asset-Allocation Perspective," *Journal of Finance*, **51**, 385-424.

Han, Y., (2006), "Asset Allocation with a High Dimensional Latent Factor Stochastic Volatility Model," *Review of Financial Studies,* **19**, 237–71.

Leitch, G. and Tanner, J.E. (1991), "Economic forecast evaluation: profits versus the conventional measures," *American Economic Review,* **81**, 580–590.

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

Pastor, L. and Stambaugh, R. F. (2000), "Comparing Asset Pricing Models: An Investment Perspective," *Journal of Financial Economics*, **56**, 335-381.

Patton, A.J. (2004), "On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation," *Journal of Financial Econometrics* **2**, 130-168.

Patton, A.J. and Timmermann, A. (2007a), "Properties of Optimal Forecasts under Asymmetric Loss and Nonlinearities," *Journal of Econometrics* **140**, 884-918.

Patton, A.J. and Timmermann, A. (2007b), "Testing Forecast Optimality Under Unknown Loss," *Journal of the American Statistical Association* **102**, 1172-1184.

Sentana, E. (1999), "Least Square Predictions and Mean-Variance Analysis," *CEPR Discussion Paper No. 2088.*

Sentana, E. (2005), "Least Square Predictions and Mean-Variance Analysis," *Journal of Financial Econometrics* **1**, 56-78.

Thornton, D.L. and Valente, G. (2012), "Out-of-sample Predictions of Bond Excess Returns and Forward Rates: An Asset Allocation Perspective," *Review of Financial Studies* **25***,* 3141-3168.

West, K. D., Edison, H.J. and Cho, D. (1993), "A Utility-based Comparison of Some Models of Exchange Rate Volatility", *Journal of International Economics,* **35**, 23-45.

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

West, K. D. and McCracken, M.W. (1998), "Regression–based Tests of Predictive Ability," *International Economic Review,* **39**, 817-840

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

Wooldridge, J.M. and H. White (1998), "Central Limit Theorems for Dependent, Heterogeneous Processes with Trending Moments," in H. White ed., *Topics in Econometric Theory: The Selected Works of Halbert White*, (Edward Elgar, Cheltenham).

Table 1. Empirical Size: Rolling Window Conditional Variance

The Table reports the results of a Monte Carlo experiment where the null hypothesis of $\Phi = 0$ is true. In Panel A) the experiment is carried out by using the DGP discussed in Section 5 and setting the first-order serial correlation of the predictive variable z_t, $\rho = 0.5$. Panel B) reports the same exercise as in Panel A) but with a the first-order serial correlation of the predictive variable z_t, $\rho = 0.9848$. For both panels, the conditional variance of excess returns is estimated as 5 year rolling window variance of excess returns, $\widehat{\sigma}_t^2 = 60^{-1} \sum_{s=t-60}^{t-1} (ep_{s+1} - \bar{d}_t)^2$ and the values reported in the Table are rejection frequency for two significance levels (i.e. 5%, 10%) and different sample sizes $T = 512, 1024$. Rejection frequencies are constructed using $2,000$ replications with 499 bootstrap iterations of the block bootstrap with a block length b set to 40, 80, and 120 for sample splits $P/T = 1/3, 1$, and 3, respectively. Unconstrained portfolio weights denotes the case where the portfolio optimization is carried out without imposing any boundaries to the resulting portfolio weights. Winsorized portfolio weights denotes the case in which portfolio weights are assumed to be bounded $-1 \leq \hat{w}_{i,t} \leq 2$.

Panel A) Persistence of the predictive variable $\rho = 0.5$

		unconstrained portfolio weights		winsorized portfolio weights	
		T=512	*T=1024*	*T=512*	*T=1024*
signif = 5%	$P/T = 1/3$	0.067	0.045	0.066	0.047
	$P/T = 1$	0.028	0.035	0.027	0.034
	$P/T = 3$	0.012	0.020	0.010	0.018
signif = 10%	$P/T = 1/3$	0.133	0.129	0.130	0.129
	$P/T = 1$	0.086	0.090	0.081	0.087
	$P/T = 3$	0.048	0.060	0.042	0.054

Panel B) Persistence of the predictive variable $\rho = 0.9848$

		unconstrained portfolio weights		winsorized portfolio weights	
		T=512	*T=1024*	*T=512*	*T=1024*
signif = 5%	$P/T = 1/3$	0.057	0.050	0.058	0.044
	$P/T = 1$	0.034	0.041	0.029	0.040
	$P/T = 3$	0.014	0.022	0.010	0.019
signif = 10%	$P/T = 1/3$	0.113	0.114	0.113	0.113
	$P/T = 1$	0.085	0.097	0.082	0.092
	$P/T = 3$	0.051	0.062	0.041	0.048

Table 2. Empirical Size: Known Unconditional Variance

The Table reports the results of a Monte Carlo experiment where the null hypothesis of $\Phi = 0$ is true. In Panel A) the experiment is carried out by using the DGP discussed in Section 5 and setting the first-order serial correlation of the predictive variable z_t, $\rho = 0.5$. Panel B) reports the same exercise as in Panel A) but with a the first-order serial correlation of the predictive variable z_t, $\rho = 0.9848$. For both panels, the conditional variance of excess returns is estimated by means of the known unconditional variance of excess returns, $\widehat{\sigma}_t^2 = 0.00172 + 0.000063b^2$, and the values reported in the Table are rejection frequencies for two significance levels (i.e. 5%, 10%) and different sample sizes $T = 512, 1024$. Rejection frequencies are constructed using $2,000$ replications with 499 bootstrap iterations of the block bootstrap with a block length b set to 40, 80, and 120 for sample splits $P/T = 1/3, 1$, and 3, respectively. Unconstrained portfolio weights denotes the case where the portfolio optimization is carried out without imposing any boundaries to the resulting portfolio weights. Winsorized portfolio weights denotes the case in which portfolio weights are assumed to be bounded $-1 \leq \hat{w}_{i,t} \leq 2$.

Panel A) Persistence of the predictive variable $\rho = 0.5$

		unconstrained portfolio weights		winsorized portfolio weights	
		T=512	*T=1024*	*T=512*	*T=1024*
signif = 5%	$P/T = 1/3$	0.042	0.049	0.037	0.042
	$P/T = 1$	0.017	0.024	0.017	0.032
	$P/T = 3$	0.008	0.021	0.008	0.019
signif = 10%	$P/T = 1/3$	0.120	0.112	0.120	0.112
	$P/T = 1$	0.065	0.069	0.067	0.065
	$P/T = 3$	0.035	0.053	0.032	0.053

Panel B) Persistence of the predictive variable $\rho = 0.9848$

		unconstrained portfolio weights		winsorized portfolio weights	
		T=512	*T=1024*	*T=512*	*T=1024*
signif = 5%	$P/T = 1/3$	0.060	0.043	0.060	0.043
	$P/T = 1$	0.022	0.042	0.022	0.041
	$P/T = 3$	0.019	0.025	0.019	0.024
signif = 10%	$P/T = 1/3$	0.109	0.107	0.109	0.107
	$P/T = 1$	0.076	0.094	0.077	0.092
	$P/T = 3$	0.053	0.064	0.051	0.064

Table 3. US Equity Premium Predictability

The Table reports performance fees, Φ 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. $\widehat{\sigma}^2_{t=j}$ indicates that the allocation to the risky asset is computed using a variance of excess returns equal to the historical variance estimated using the past j 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 1927 until December 2011 and the forecasts begin in January 1965. Φ are expressed as percentage per period (i.e. month, quarter or year). Values in brackets are p-values for the null hypothesis that $H_0 : \Phi \leq 0$. The p-values are computed by bootstrap, as discussed in Section 4 of the main text with 499 replications and a block length of 80 (monthly), 20 (quarterly) and 5 (annual) observations. When weights are winsorized, $-0.5 \leq w_{i,t}\left(\widehat{\beta}_{i,t}\right) \leq 1.5$ (i.e. a cap of 150% investment in the risky asset is imposed).

Panel A) Monthly Frequency

		no winsorization				winsorized weights (*max 150%*)			
	Variables	$\Phi\left(\widehat{\sigma}^2_{t=60}\right)$		$\Phi\left(\widehat{\sigma}^2_{t=120}\right)$		$\Phi\left(\widehat{\sigma}^2_{t=60}\right)$		$\Phi\left(\widehat{\sigma}^2_{t=120}\right)$	
svar	Stock Variance	-0.132	[0.41]	-0.126	[0.43]	-0.009	[0.27]	-0.039	[0.36]
d/e	Dividend Payout Ratio	-0.178	[0.87]	-0.197	[0.87]	-0.043	[0.45]	-0.015	[0.32]
lty	Long-term Yield	-0.056	[0.77]	-0.098	[0.82]	0.073	[0.04]	0.052	[0.14]
tms	Term Spread	0.102	[0.14]	0.158	[0.05]	0.148	[0.03]	0.144	[0.02]
ltr	Long-term Return	-0.106	[0.49]	-0.056	[0.22]	0.076	[0.05]	0.082	[0.05]
infl	Inflation	0.043	[0.27]	0.108	[0.11]	0.113	[0.08]	0.086	[0.13]
tbl	T-Bill Rate	0.044	[0.34]	0.101	[0.11]	0.149	[0.02]	0.087	[0.06]
dfr	Default Return Spread	-0.001	[0.05]	-0.034	[0.16]	0.052	[0.04]	0.051	[0.02]
dfy	Default Yield Spread	-0.070	[0.47]	0.050	[0.04]	0.089	[0.01]	-0.002	[0.07]
ntis	Net Equity Expansion	-0.994	[0.98]	-0.446	[0.94]	0.020	[0.22]	0.021	[0.21]
d/p	Dividend-Price Ratio	-0.118	[0.68]	0.051	[0.24]	-0.086	[0.76]	-0.007	[0.28]
d/y	Dividend Yield	-0.263	[0.85]	-0.002	[0.40]	-0.025	[0.55]	0.012	[0.34]
e/p	Earning-Price Ratio	-0.256	[0.93]	-0.374	[0.96]	0.104	[0.09]	0.081	[0.11]
b/m	Book-Market Ratio	-0.801	[0.96]	-0.834	[0.95]	-0.190	[0.96]	-0.147	[0.86]

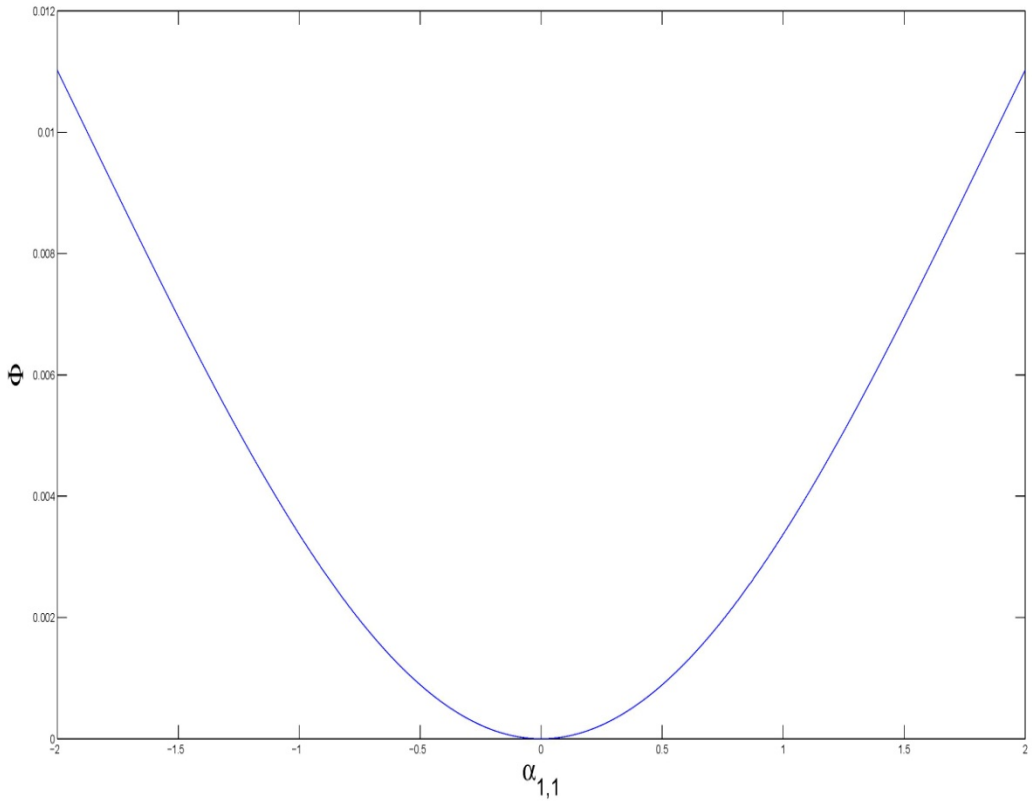
Panel B) Quarterly Frequency

	Variables	no winsorization				winsorized weights (*max 150%*)			
		$\Phi\left(\widehat{\sigma}^2_{t=15}\right)$		$\Phi\left(\widehat{\sigma}^2_{t=30}\right)$		$\Phi\left(\widehat{\sigma}^2_{t=15}\right)$		$\Phi\left(\widehat{\sigma}^2_{t=30}\right)$	
svar	Stock Variance	-0.143	[0.65]	-0.054	[0.39]	0.134	[0.04]	-0.061	[0.41]
d/e	Dividend Payout Ratio	-0.081	[0.38]	-0.151	[0.62]	-0.042	[0.41]	-0.039	[0.19]
lty	Long-term Yield	-0.182	[0.85]	-0.389	[0.94]	0.075	[0.14]	0.034	[0.30]
tms	Term Spread	0.386	[0.03]	0.462	[0.03]	0.438	[0.02]	0.234	[0.04]
ltr	Long-term Return	-0.940	[0.97]	-0.437	[0.93]	0.003	[0.23]	0.079	[0.05]
infl	Inflation	0.269	[0.05]	0.355	[0.02]	0.207	[0.04]	0.186	[0.04]
tbl	T-Bill Rate	-0.137	[0.61]	0.185	[0.12]	0.305	[0.02]	0.212	[0.08]
dfr	Default Return Spread	-3.161	[0.94]	-2.245	[0.95]	-0.425	[0.92]	-0.576	[0.95]
dfy	Default Yield Spread	-0.331	[0.70]	0.164	[0.05]	0.347	[< 0.01]	-0.020	[0.09]
ntis	Net Equity Expansion	-6.772	[0.96]	-2.391	[0.93]	-0.151	[0.58]	-0.228	[0.85]
d/p	Dividend-Price Ratio	-3.295	[0.95]	-1.503	[0.92]	-0.114	[0.72]	0.088	[0.36]
d/y	Dividend Yield	-1.488	[0.92]	-0.603	[0.86]	-0.057	[0.37]	0.193	[0.12]
e/p	Earning-Price Ratio	-0.796	[0.87]	-0.297	[0.81]	0.154	[0.17]	0.074	[0.41]
b/m	Book-Market Ratio	-8.958	[0.98]	-4.874	[0.98]	-0.788	[0.97]	-0.793	[0.95]

Panel C) Annual Frequency

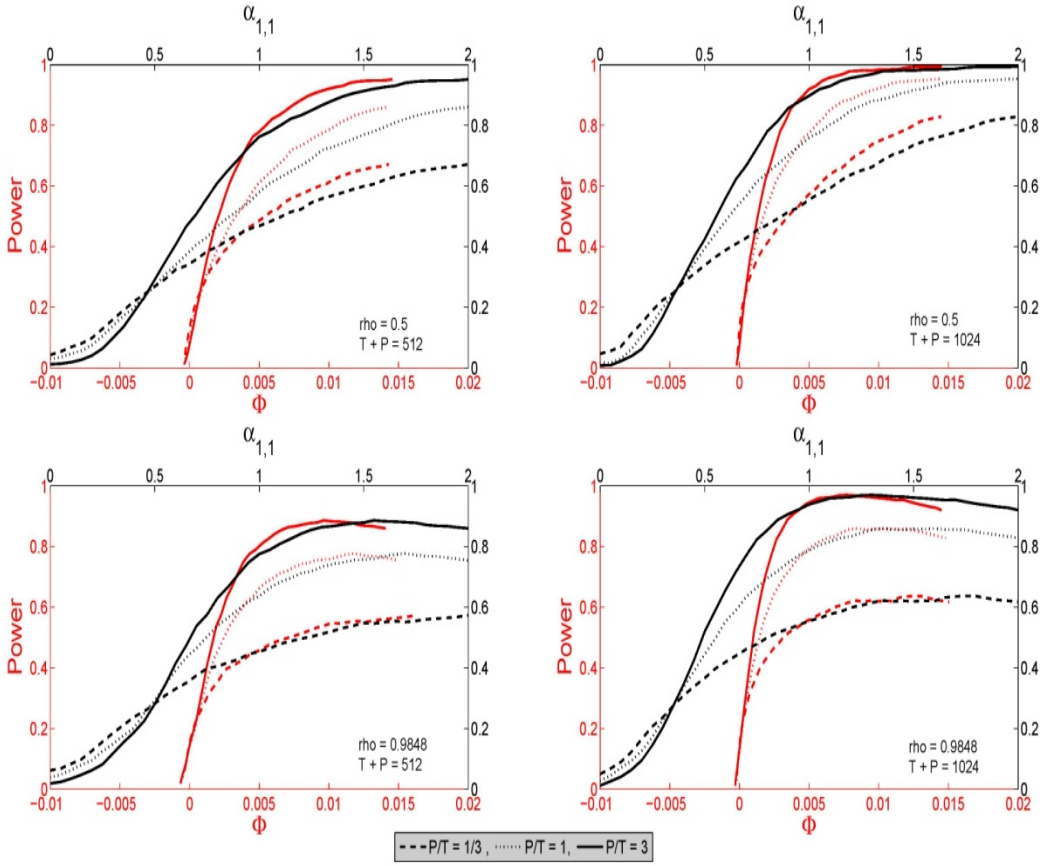
		no winsorization				winsorized weights (*max 150%*)			
	Variables	$\Phi\left(\widehat{\sigma}^2_{t=5}\right)$		$\Phi\left(\widehat{\sigma}^2_{t=10}\right)$		$\Phi\left(\widehat{\sigma}^2_{t=5}\right)$		$\Phi\left(\widehat{\sigma}^2_{t=10}\right)$	
svar	Stock Variance	-0.062	[0.29]	-0.039	[0.20]	-0.056	[0.27]	-0.009	[0.19]
d/e	Dividend Payout Ratio	-1.107	[0.78]	-1.174	[0.82]	-0.646	[0.75]	-0.736	[0.75]
lty	Long-term Yield	-1.403	[0.83]	-1.288	[0.89]	-0.296	[0.64]	-0.286	[0.64]
tms	Term Spread	1.797	[0.01]	1.358	[< 0.01]	1.295	[< 0.01]	0.864	[0.02]
ltr	Long-term Return	-2.515	[0.92]	-0.501	[0.46]	1.041	[0.01]	1.210	[< 0.01]
infl	Inflation	-0.544	[0.66]	-0.528	[0.73]	-0.372	[0.65]	-0.342	[0.64]
tbl	T-Bill Rate	-0.394	[0.41]	-0.120	[0.20]	0.555	[0.05]	0.556	[0.04]
dfr	Default Return Spread	-1.163	[0.83]	-1.341	[0.90]	-0.706	[0.85]	-0.677	[0.86]
dfy	Default Yield Spread	-0.044	[0.20]	0.027	[0.11]	-0.042	[0.18]	0.077	[0.09]
ntis	Net Equity Expansion	-4.311	[0.92]	-4.311	[0.96]	-2.281	[0.98]	-2.449	[0.99]
d/p	Dividend-Price Ratio	-0.547	[0.58]	0.122	[0.12]	-0.320	[0.57]	0.222	[0.10]
d/y	Dividend Yield	-1.334	[0.70]	-0.161	[0.08]	-0.801	[0.70]	-0.086	[0.12]
e/p	Earning-Price Ratio	-0.470	[0.75]	0.167	[0.21]	0.204	[0.40]	0.340	[0.16]
b/m	Book-Market Ratio	-2.847	[0.90]	-2.693	[0.91]	-1.966	[0.98]	-1.992	[0.96]

Figure 1: Φ as a function of $\alpha_{1,1}$



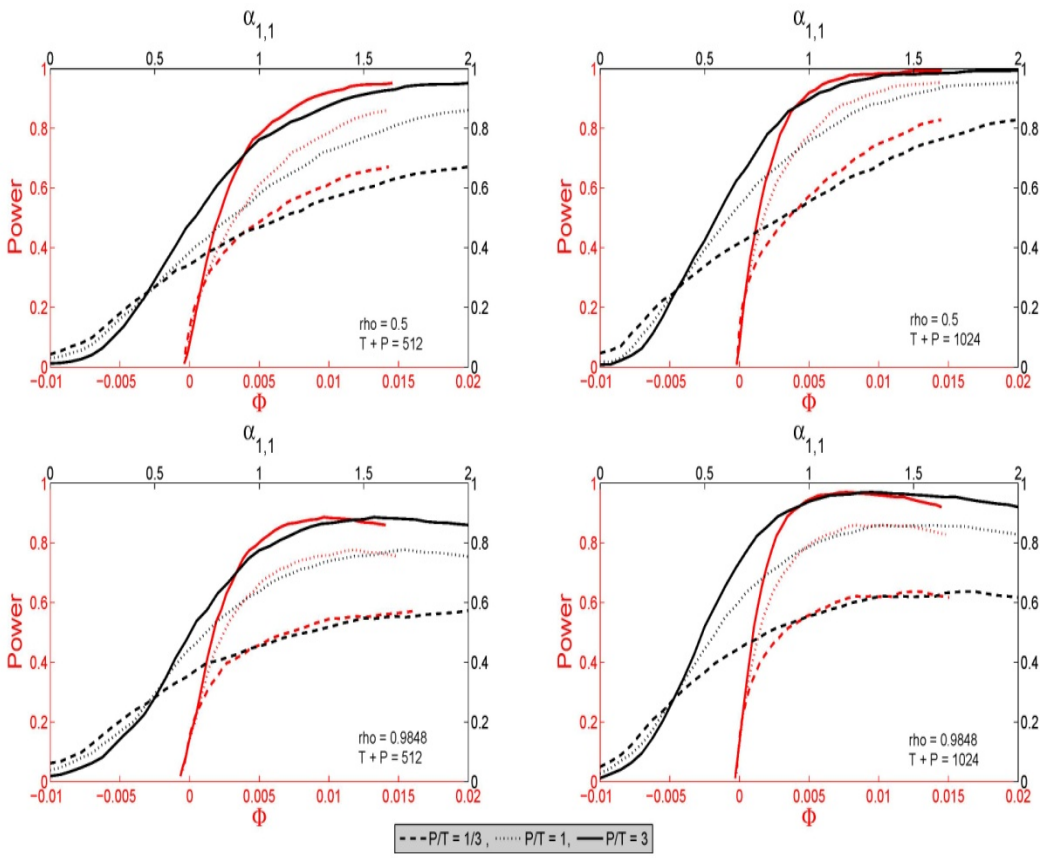
Notes: Figure 1 provides Φ as a function of $\alpha_{1,1}$ using the function in equation (11) of the text. The remaining moments in the formula are calibrated to the return on the NYSE value weighted index as our risky asset, its dividend yield as our predictor, and the yield on the 3-month T-bill as our return on the risk-free asset. See Section 5 for more details on this calibration.

Figure 2: Actual Power with rolling windows estimate of σ_t^2



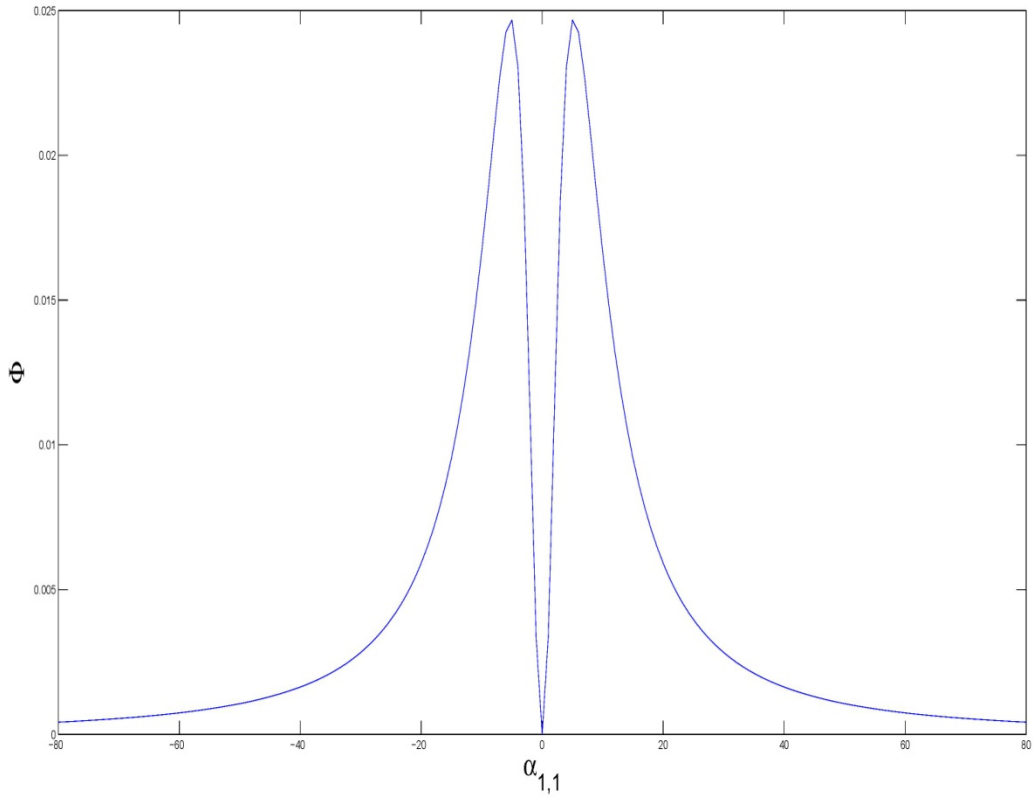
Notes: (a) Figure 2 provides rejection frequencies of nominally 5% tests of the null hypothesis that $\Phi = 0$. The simulation design parameterizes the alternative as a function of $\alpha_{1,1}$ with values ranging from 0 to 2. As a function of $\alpha_{1,1}$, power can be deduced using the black lines and the upper horizontal axis. For each value of $\alpha_{1,1}$ the simulation was conducted 2000 times and the median value of Φ was saved. The red lines and the lower horizontal axis can be used to deduce the power of the test as a function of these values of Φ. (b) The conditional variance is estimated using the rolling window estimator $\hat{\sigma}_t^2 = 60^{-1}\sum_{s=t-60}^{t-1}(ep_{s+1} - \overline{d}_t)^2$.

Figure 3: Actual Power with known σ_t^2



Notes: (a) See notes for Figure 2. (b) The simulations in Figure 3 estimate the conditional variance using the known unconditional variance of ep_{s+1}.

Figure 4: Φ as a function of $\alpha_{1,1}$



Notes: See notes to Figure 1. The only difference between the two plots is the range of $\alpha_{1,1}$.