# Predictability of Equity Returns over Different Time Horizons: A Nonparametric Approach

Qingqing Chen[a] Yongmiao Hong[b,c]

We thank Guofu Zhou and the seminar participants at Rochester, SMU-ESSEC Symposium on Empirical Finance and Financial Econometrics, and University of California at Riverside for comments and suggestions. The views expressed in this paper are those of the authors alone and do not necessarily reflect those of the Office of the Comptroller of the Currency (OCC). Any remaining errors are solely ours.

a: Office of the Comptroller of the Currency, 400 7th Street SW, Mail Stop 6E-3, Washington, DC 20024. Email address: qingqing.chen@occ.treas.gov

b: Department of Economics & Department of Statistical Science, Cornell University, Ithaca, NY 14850. Email address: yh20@cornell.edu

c: Wang Yanan Institute for Studies in Economics (WISE) & MOE Key Laboratory in Econometrics, Xiamen University, Xiamen 361005, China.

# Predictability of Equity Returns over Different Time Horizons: A Nonparametric Approach

*Abstract:* This paper aims to test whether equity returns are predictable over various horizons. We propose a reliable and powerful nonparametric test to examine the predictability of equity returns, which can be interpreted as a signal-to-noise ratio test. Our comprehensive in-sample and out-of-sample analysis shows that the commonly used predictive variables such as short rate, dividend yields and earnings yields have good predictability power at both short and long horizons, different from both the conventional wisdom and Ang and Bekaert (2007). Contrary to Goyal and Welch (2007), an out-of-sample nonparametric forecast outperforms the historical mean model and linear predictive models.

*Key Words*: Asset Return Predictability, bootstrap, Hypothesis Testing, Kernel, Nonlinearity, Signal-to-Noise Ratio, Time Horizons, Out-of-sample Inference, Pricing Error

*JEL Classifications*: G17 C14 C53

# 1. INTRODUCTION

There is a long tradition to study the predictability of equity returns in finance. Despite an enormous amount of past efforts, whether equity returns can be meaningfully predicted over different time horizons remains a subject of ongoing debates and intensive empirical research. Long-horizon asset returns are more informative than their shorter-horizon counterparts. Thus random walk models, and martingale models based on past asset returns are statistically weak to explain the real data. The widely used present value model assumes that the expected stock return is constant over time. It makes no assumption about equity repurchases by firms which affect the time pattern of expected future dividends (Rozeff (1984), Campbell and Shiller (1987), Campbell and Shiller (1988a, 1988b), and West (1988)). While stock prices and dividends appear to grow exponentially over time rather than linearly, the linear model is less appropriate than a nonlinear model which can better capture the properties of asset returns across time.Thereafter, researchers have proposed several nonlinear models to predict asset returns. One is the dividend model with rational bubble in which the bubble is a nonlinear function of the stock's dividends (Froot and Obstfeld (1991)). This model has its limitation in explaining the observed predictability of stock returns. Another nonlinear model is a loglinear present-value model (Campbell (1991), Ang and Bekaert (2007)), which suggests a nonlinear relation among equity returns, dividend ratio, and interest rates.

The existing literature has showed that there exists strong nonlinearity in the models of predicting asset returns, and that expected asset returns and dividend ratios are highly persistent and time-varying (Froot and Obstfeld (1991), Campbell (1991), Ang and Bekaert (2007)). However, there is no consensus in the literature where the nonlinearities come from. Nonlinearities can arise for several reasons: structural breaks in the mean of the dividend price ratio (Lettau and Van Nieuwerburgh (2008)), time-varing dividend growth rates (Cochrane, 2008; Binsbergen and Koijen, 2010; Golez, 2014), a log-linear approximation of present value model (Campbell and Shiller (1988), Binsbergen and Koijen (2011)), and the correlation of the expected returns and business conditions (Henkel et al. (2011)). Therefore it is important to propose a better method to predict the equity returns incorporating the unknown nonlinearities and the predictive relationship between asset returns and time horizons.

Empirical studies increasingly cast doubt on the forecasting power of price-based predictors of equity returns. There are two recent debates on the predictability of equity returns in the literature. The first debate is to question whether the predictability of equity returns exists at short horizons or longer horizons.[1] The conventional wisdom in the literature is that aggregate dividend yields strongly predict excess returns, and the predictability is stronger at longer horizons (Fama and French (1988),[2] Campbell (1991), and Cochrane (1992)). If daily returns can be predicted by a slow-moving or persistent variable, then the predictability adds up over the long horizons. In contrast, Ang and Bekaert (2007) find that dividend yields, together with the short rate, predict excess returns only at short horizons and do not have any long-horizon predictive power. The second debate is whether the existing price-based predictors have better predictive power than the historical average model of equity returns. Goyal and Welch (2007) argue that the historical average excess stock return forecasts future excess stock returns better than the regressions of excess returns on predictive variables. In response to their arguments, Campbell and Thompson (2007) show that many predictive regressions beat historical average returns by imposing restrictions on the signs of coefficients and return forecasts, or the coefficients relating valuation ratios to future returns based on steady-state models. The conclusions of the two debates are controversial and inconclusive.

In this paper, we undertake an analysis of both in-sample and out-of-sample tests in an

[1] Fama and French (1988), Campbell and Shiller (1988a,b), Goetzmann and Jorion (1993, 1995), Hodrick (1992), Stambaugh (1999), Wolf (2000), Goyal and Welch (2003, 2007), Valkanov (2003), Lewellen (2004), Campbell and Yogo (2006), Campbell and Thompson(2007), and Ang and Bekaert (2007)

[2] Fama and French (1988) provide the strongest evidence in support of the dividend yield effect by using overlapping multiple-year horizon returns. They observe that the explanatory power of the dividend yield increases with the time horizon of the returns; over 4-year horizons, the $R^2$'s reach an astonishing high value of 64%.

effort to better understand the empirical evidence on return predictability. We are particularly interested in investigating the following problems: (i) Does the predictability of valuation ratios such as dividend yields exist at various horizons? (ii) Do the linear predictive models suffer from neglected nonlinear predictability? In particular, is the poor out-of-sample performance of most linear predictive models due to the limitation of linear models or due to the nonexistence of predictability of equity returns? (iii) Do the predictive models beat the historical average excess stock return (historical mean model)?

The existing economic theory and literature has suggested the linear predictive model is not optimal to predict equity returns due to the strong existing nonlinearities. However, there is no evidence or support for the deterministic functional form of nonlinearities in the literature. Therefore, we first develop a reliable out-of-sample nonparametric model-free predictability test, which has several appealing features. First, the nonparametric method can capture a wide variety of linearities and nonlinearities without assuming any parametric forms. Thus, it can directly assess the predictability of equity return data itself rather than the predictability of a specific model for equity return. Second, the nonparametric predictability test can be interpreted as a signal-to-noise ratio, because it is based on the average of the squared predictable components over the sample variance of pricing errors. Third, we propose to use a conditional bootstrap procedure which maintain the original dynamics of predictive variables and serial dependence structure of the multi-step-ahead forecast errors. Such a bootstrap procedure provides reliable statistical inference for different sample sizes typically used in the literature. Simulation studies show that it has reasonable size and power in finite samples even when the regressors are highly persistent and the forecast horizon is relatively long.

We apply the proposed test to examine the predictability of equity returns at short or longer horizons. Our empirical results show that the short rate, dividend yields and earnings yields have good predictability power at both short and long horizons, which is different from both the conventional wisdom and Ang and Bekaert (2007). Second, the comprehensive in-sample and out-of-sample analysis suggests that the variables such as dividend yields, earnings yields, dividend payout ratio, short rate, inflation, book-to-market ratio, investment to capital ratio, corporate issuing activity, and consumption, wealth, and income ratio have predictability power for equity returns, but it cannot be fully captured by the prevailing linear predictive models. Our conclusion is in contrast to Goyal and Welch (2007) and consistent with Campbell and Thompson (2007). Campbell and Thompson (2007) find that predictive variables have better out-of-sample performance than the historical average return forecasts when imposing the restrictions on the coefficients. In fact, the restriction on coefficients is a form of nonlinearity. Our nonparametric test shows that the prevailing predictive model beats the historical mean model without any restriction because there is a higher neglected signal-to-noise ratio from the historical mean model.

Most of the existing empirical studies use linear regressions to forecast asset returns. There are a number of pitfalls applying those models to predict asset returns or evaluate the predictability power. First, the apparent predictability of stock returns might be spurious since many predictive variables, such as valuation ratios, are highly persistent (Nelson and Kim (1993), Stambaugh (1999), and Sarkissian, and Simin (2003)). An active recent literature discusses alternative econometric methods or proposes new statistical tests for correcting the bias and conducting valid inference on estimation of long-horizon predictive regression models with persistent variables and errors.[3] These studies have emphasized the bias toward rejection of the null hypothesis of no predictability. In particular, the usual corrections to the standard errors are only valid asymptotically, and there is some question as to whether "asymptotic" should be measured in terms of years, decades, or even centuries, especially for long-horizon forecasts.

[3] See Cavanagh et al., 1995; Goetzmann and Jorion, 1993; Mark, 1995; Kilian, 1999; Wolf, 2000; Lewellen, 2004; Campbell and Yogo, 2006; Polk et al., 2006; Ang and Bekaert, 2007; Valkanov, 2003; Maynard, Shimotsu, and Wang, 2011

Second, there exists an "errors-in-variables" problem that the explanatory variable like the dividend yield is not properly exogenous, but rather contains a price level variable that also appears in the regression (Stambaugh (1986)). The regression coefficient in the dividend yield regression may be downward biased because the yields contain forecasts of future returns and dividend growth (Fama and French (1988), Fama (1990), and Kothari and Shanken (1992)). Therefore there exists serial correlation in the forecast error particularly when the time horizon $h$ is large relative to the sample size. As a result, there also exist some finite sample problems for reliable statistical inference (Hodrick (1992), Nelson and Kim (1993), Mankiw and Shapiro (1986)). Under the same conditions, the standard t-test for predictability has incorrect sizes in finite samples (Cavanagh et al. (1995), Campbell and Yogo (2006)). These problems become more serious if applied econometricians are data mining, considering large numbers of variables, and reporting only those results that are apparently statistically significant (Foster et al. (1997), Ferson, Sarkissian, and Simin (2003)).

Third, some empirical studies use a finite-order VAR system to model returns and dividend yields (Hodrick (1992), Campbell and Shiller (1988a,b), Stambaugh (1999)). The VAR model cannot fully capture the nonlinear dynamics of dividend yields implied by the present value model. For a linear predictive regression model, when a price-based estimator or regressor appeals to be statistically insignificant, one cannot conclude that the null hypothesis of no predictability holds, because there may exist neglected nonlinear predictability.

Fourth, a different critique[4] emphasizes that most linear predictive regressions have often performed poorly out-of-sample (Goyal and Welch (2003, 2007), Campbell and Thompson (2007)). It is well-known that it may cause overfitting and capture spurious predictability even though in-sample diagnostic analysis is significant and can reveal useful information on possible sources

[4]This critique had a particular force during the bull market of the late 1990s, when low valuation ratios predicted extraordinarily low stock returns that did not materialize until the early 2000s (Campbell and Shiller, 1998).

of model misspecification. Out-of-sample evaluation can capture the true predictability of a model or the data generating process.[5] The disparities between in-sample and out-of-sample results of return predictability documented in the literature make an overall assessment of return predictability difficult. In particular, it is unclear whether the poor out-of-sample performance of linear prediction models is due to the limitation of linear models or due to the nonexistence of predictability of equity returns. Many earlier out-of-sample tests have focused on the dividend ratios. Fama and French (1988) interpret the out-of-sample performance of dividend ratios to have been a success. Bossaerts and Hillion (1999) interpret the out-of-sample performance of the dividend yield to be a failure. Torous and Valkanov (2000) find that a low signal-noise ratio of many predictive variables makes a spurious relation between returns and persistent predictive variables unlikely and would lead to a low out-of-sample forecasting power. Rapach and Wohar (2006) find that certain financial variables display significant in-sample and out-of-sample predictive ability for stock returns. Neely, Rapach, Tu, and Zhou (2011) shows that utilizing information from both technical indicators and macroeconomic variables substantially increases the out-of-sample gains relative to using either macroeconomic variables or technical indicators alone. Goyal and Welch (2007) argue that the poor out-of-sample performance of predictive regressions is a systemic problem. They compare predictive regressions with historical average returns and find that historical average returns almost always generate superior return forecasts. They conclude that "the profession has yet to find some variable that has meaningful and robust empirical equity premium forecasting power."

The use of model-based predictors facilitates a better understanding of specific aspects of the

[5] Here are several important reasons why out-of-sample predictability check is important. First, the usual practice of extensive search for more complicated models using the same or similar data set may suffer from the so-called data snooping bias, as pointed out by Lo and MacKinlay (1989) and White (2000). A more complicated model may overfit idiosyncratic features of the data without capturing the true data generating process. Out-of-sample prediction evaluation will alleviate, if not eliminate completely, such data snooping bias. Second, a model that fits in-sample data well may not predict the future well because of unforeseen structural changes or regime shifts in the data generating process.

economic mechanism, however, these predetermined variables may not be enough to capture all information required in decision making. Forecast combination has recently received renewed attention in the forecasting literature. Stock and Watson (1999, 2003, 2004) use the combination forecast for inflation and real output growth. Rapach, Strauss, and Zhou (2010) propose a combination approach to improve the out-of-sample equity premium forecasting problem. Tae-Hwy Lee, Yundong Tu, and Aman Ullah (2014) shows imposing the positive constraint and bagging method can help reduce the asymptotic variance and improve the out-of-time performance. Ferreira and Santa-Clara (2011) develop an intriguing "sum-of-the-parts" (SOP) approach to forecast the market return as the sum of a 20-year moving average of earnings growth rates and the current dividend price ratio (minus the risk-free rate). Specifically, they decompose the log market return into the sum of the growth in the price-earnings ratio, growth in earnings, and the dividend-price ratio. Previous studies suggest that there exists strong nonlinearity in the predictive models, and that expected asset returns and dividend ratios are highly persistent and time-varying. The poor out-of-sample performance of most linear predictive models is due to the limitation of linear models. The lack of consistent out-of-sample evidence in Goyal and Welch (2008) indicates the need for improved forecasting methods to better establish the empirical reliability of equity return predictability.

In order to shed light on the recent debate, we evaluate the out-of-sample forecast of different models used in the literature. We propose nonparametric estimators to forecast the equity returns using both individual and combined forecast. We choose the same 15 economic variables in Goyal and Welch (2008) to predict the equity returns. The benchmark model is historical mean model. The alternative models are linear predictive model and two nonparametric predictive models. We find that the combined forecast methods outperform the individual forecast methods. Fama and French (1989) and others show that these variables can detect changes in economic conditions that potentially signal fluctuations in the equity risk premium. But the dividend yield or term spread alone could capture different components of business conditions, and a given individual economic variable may give a number of "false signals" and/or imply an implausible equity risk premium during certain periods. Rapach, Strauss, and Zhou (2010) argue that if individual forecasts based on the predictors are weakly correlated, combined forecast should be less volatile and more reliable to track movements in the equity risk premium. Our results are consistent with their findings. Combining forecast incorporates information from a host of economic variables and it helps to reduce forecast variability. Since the historical average ignores economic variables, combined forecasts have a substantially smaller bias than the historical average.

Our nonparametric predictive models have lower RMSE than the historical mean model at both short-horizon and long-horizon and significantly improve the out-of-sample performance without any restrictions. To evaluate the economic significance, we calculate realized utility gains for a mean-variance investor on a real-time basis, following Marquering and Verbeek (2004), Campbell and Thompson (2008), Welch and Goyal (2008), Rapach, Strauss, and Zhou (2010), and Wachter and Warusawitharana (2009). The nonparametric forecast successfully produces a positive utility gain for all the predictors. Using our nonparametric methods, both combined and individual forecast outperform the historical average. It holds across a number of historical periods using both statistical and economic criteria, even for the more recent periods when the out-of-sample predictive ability of many individual variables is relatively poor. One reasonable explanation is that the historical mean estimator is a simple average of past equity returns while our kernel estimator is a weighted average of past equity returns, with the weights depending on the values of predictive variables. The predictive variable indeed provides useful information for out-of-sample forecasts. It is not restricted to the parametric forms. It can fit the data better than simply the linear or nonlinear parametric model. Nonparametric prediction generates a forecast with a variance near that of the smooth real equity return data, thereby reducing the noise in the individual predictive model.

Section 2 proposes a nonparametric predictability test. Section 3 describes the data. Section 4 discusses the empirical results. Section 5 reports the out-of-sample performance of the individual and combined forecasts and its economic implications. Section 6 concludes the paper.

## 2. NONPARAMETRIC TEST FOR PREDICTABILITY

### 2.1 Hypotheses of Interests and Nonparametric Test

We are interested in whether the predictability of excess returns depends on time horizons. If future excess returns cannot be predicted by the past dividend yield or other variables over any time horizon, then the null hypothesis holds.

Specifically, suppose $\{Y_t, X_t'\}'$ is a stationary time series process where $Y_t$ is a scalar, and $X_t$ is a $d$-dimensional vector. We are interested in testing the predictability of $Y_{t+h}$ using $X_t$, where the integer $h$ is the time horizon index for a multi-step ahead prediction. In our applications below, $X_t$ is, for example, the dividend yield in period $t$, and $Y_{t+h}$ is the asset return $h$ periods ahead. Different $h$'s will allow us to examine the relationship between asset return predictability and time horizons. Formally, the null and alternative hypothesis can be written respectively as

$$H_0 : E(Y_{t+h}|X_t) = E(Y_{t+h}) \tag{2.1}$$

$$H_A : E(Y_{t+h}|X_t) \neq E(Y_{t+h}). \tag{2.2}$$

The null hypothesis $H_0$ is characterized by the horizon index $h$. It is possible that $H_0$ holds for a relatively long horizon but it does not hold for a relatively short horizon. This is one of our focuses in this paper, namely we will investigate the relationship between predictability of excess asset returns and the time horizon $h$, which has been a long-standing problem in empirical finance.

In empirical finance, a linear predictive regression model

$$Y_{t+h} = X_t'\beta + \varepsilon_{t+h}, \tag{2.3}$$

is used to check predictability of excess asset returns. When an estimator for $\beta$ is statistically insignificant, one does not find evidence for predictability power of $X_t$ for $Y_{t+h}$. Strictly speaking, one cannot conclude that $H_0$ holds. This is because a zero parameter value for $\beta$ is a necessary condition for $H_0$ but it is not a sufficient condition. A zero $\beta$ implies that there is no linear predictive power of $X_t$ for $Y_{t+h}$, but there may exist a nonlinear predictive power of $X_t$ for $Y_{t+h}$. For example, suppose $Y_{t+h} = X_t^2 + \varepsilon_{t+h}$, where $X_t$ is normally distributed with zero mean and the disturbance $\varepsilon_{t+h}$ is independent of $X_t$. Then a linear regression coefficient $\beta$ will be exactly zero although $E(Y_{t+h}|X_t) = X_t^2$.

When an estimator for $\beta$ is statistically significant, there exists evidence of the predictive power of $X_t$ for $Y_{t+h}$. In this case, one may be interested in testing whether the linear regression model has the optimal predictive power for $Y_{t+h}$. Put it differently, one may be interested in testing whether there exists any nonlinear predictive power of $X_t$ for $Y_{t+h}$, in addition to the documented linear predictability. In this case, the null and alternative hypothesis are written respectively as

$$H_0 : E(\varepsilon_{t+h}|X_t) = 0 \tag{2.4}$$

$$H_A : E(\varepsilon_{t+h}|X_t) \neq 0, \tag{2.5}$$

where $\varepsilon_{t+h}$ is the prediction error from the linear regression model in (2.3). The null hypothesis $H_0$ in (2.4) implies that the linear regression model in (2.3) has optimal predictive power. When $\mathrm{H}_A$ in (2.5) holds, there exists a nonlinear predictive relationship between $\mathrm{X}_t$ and $\mathrm{Y}_{t+h}$, and a suitbale nonlinear predictive model will outperform the linear regression model in (2.3). Becasue $\varepsilon_{t+h}$ is unobservable, we need to use an estimated residual $\hat{\varepsilon}_{t+h} = Y_{t+h} - X_t'\hat{\beta}$, where $\hat{\beta}$ is an estimator for $\beta$. Note that when $H_0$ holds, $\{\varepsilon_{t+h}\}$ may not be a martingale difference sequence unless $h = 1$. In general, $H_0$ allows $\{\varepsilon_{t+h}\}$ to follow a MA($h$) dependence. This has important implication on inference, particularly when $h$ is relatively large.

In this section, we develop a unified nonparametric testing framework which is applicable to test hypotheses in (2.1) and (2.4). The basic idea is to use a nonparametric estimator for $E(Y_{t+h}|X_t)$ or $E(\varepsilon_{t+h}|X_t)$ and check if the estimator is close to constant or zero. As is well-known, the nonparametric method has an advantage that it does not require an ex ante model specification and can capture any predictive relationship no matter whether it is linear or nonlinear (c.f. Härdle (1993), Pagan and Ullah (1999)). Thus, it is quite suitable for our purpose here.

To avoid capturing spurious predictability due to in-sample overfitting, we consider out-of-sample predictability check. There are several reasons why out-of-sample predictability check is important. First, the usual practice of extensive search for more complicated models using the same or similar data set may suffer from the so-called data snooping bias, as pointed out by Lo and MacKinlay (1989) and White (2000). A more complicated model may overfit some idiosyncratic features of the data without capturing the true data generating process. Out-of-sample prediction evaluation will alleviate, if not eliminate completely, such data snooping bias. Second, a model that fits in-sample data well may not predict the future well because of unforeseen structural changes or regime shifts in the data generating process. Therefore, in-sample analysis is not adequate and it is important to examine out-of-sample prediction. Third, out-of-sample prediction is more relevant to most economic applications in practice.

Specifically, suppose we have an observed sample $\{Y_t, X_t'\}_{t=1}^T$ of size $T$. We first split the sample into two parts: the first subsample contains $R$ observations, and the second subsample contains $n = T - R$ observations. We will use the first subsample or a modification of it to estimate model parameter $\beta$ and use the second subsample to check predictability. There are various methods to estimate parameter $\beta$. One simple method is to use the first subsample $\{Y_{t+h}, X_t'\}_{t=1}^R$ to estimate $\beta$. Another method is to use $\{Y_{t+h}, X_t\}_{t=i+1}^{R+i}$ to estimate $\beta$ when predicting $Y_{R+h+1+i}$, for $0 \leq i \leq n-h-1$. This is called the rolling estimation. One can also use the recursive estimation method, which uses the subsample $\{Y_{t+h}, X_t\}_{t=1}^{R+i}$ to estimate $\beta$ when predicting $Y_{R+h+1+i}$. Generally, we use the notation $\hat{\beta}_t$ to denote an estimator for $\beta$ when predicting $Y_{t+h}$ in an out-of-sample context. The resulting estimated out-of-sample residual from a linear model (2.3) is

$$\hat{\varepsilon}_{t+h} = Y_{t+h} - X_t'\hat{\beta}_t, t = R + 1, \cdots T - h$$

To capture potentially neglected nonlinear predictable component in $\varepsilon_{t+h}$, we use a smoothed kernel method to estimate $E(\varepsilon_{t+h}|X_t)$. Put

$$\begin{aligned}
\widehat{m}_h(x) &= \frac{1}{n-h}\sum_{s=R+1}^{T-h}\hat{\varepsilon}_{s+h}K_b(x-X_s), \\
\widehat{g}_h(x) &= \frac{1}{n-h}\sum_{s=R+1}^{T-h}K_b(x-X_s),
\end{aligned}$$

where $x=(x_1,x_2,\cdots,x_d)'$, $y=(y_1,y_2,\cdots,y_d)'$, and $K_b(x-y)=\Pi_{i=1}^{d}b^{-1}K[(x_i-y_i)/b]$. The kernel function $K(\cdot)$ is is a prespecified symmetric probability density function. Examples include a Gaussian kernel $K(u)=(2\pi)^{-1/2}\exp(-u^2/2)$ and a quatic kernel $K(u)=\frac{3}{4}(1-u^2)\mathbf{1}(|u|\leq 1)$, where $\mathbf{1}(\cdot)$ is the indicator function, giving value 1 if $|u|\leq 1$ and value 0 otherwise. The bandwidth $b=b(n)$ vanishes to zero as the sample size $n\to\infty$, but at a slower rate. For simplicity, we use the same bandwidth for each components of $X_t$. In practice, one can first standardize each component of the vector $X_t$ by its sample standard deviation. The regression estimator for $E(\varepsilon_{t+h}|X_t)$ is then defined as follows:

$$\widehat{r}_h(x)=\frac{\widehat{m}_h(x)}{\widehat{g}_h(x)}.$$

This is called the Nadaraya-Watson regression estimator. The estimator $\widehat{g}_h(x)$ in the denominator is a kernel estimator for the marginal density $g_h(x)$ of $\{X_t\}$. Under regularity conditions, $\widehat{r}_h(x)\to r_h(x)=E(\varepsilon_{t+h}|X_t=x)$ in probability as both $R,n\to\infty$.

Under $H_0$, $\widehat{r}_h(x)$ is close to zero for all $x$. Under the alternative hypothesis $H_A$, $\hat{r}_h(x)$ is not a zero function but is a nontrivial function of $x$ subject to sampling variation. To measure the departure of $\hat{r}_h(x)$ from zero over all $x$, we use the following global measure

$$\widehat{Q}(h)=\frac{1}{n-h}\sum_{t=R+1}^{T-h}\widehat{r}_h^2(X_t)w(X_t),$$

where the positive weighting function $w(\cdot)$ can be chosen to trim the extreme observations where the estimation of $\hat{r}(x)$ is not reliable due to sparse observations (we allow the distribution of $X_t$ has unbounded support). It can also be used to direct power of the proposed test to the region of interest, such as predictability when $X_t$ is negative (in this case, we choose $w(x) = \mathbf{1}(x \leq 0)$. The statistic $\widehat{Q}(h)$ can be viewed as a measure of the magnitude of the "signal" that can be extracted to predict asset returns if (and only if) it contains no systematic predictable component in $E(\varepsilon_{t+h}|X_t)$, the estimator $\widehat{r}_h(X_t)$ and therefore $\widehat{Q}(h)$ will be close to zero.

Alternatively, we can directly use an integrated global measure $\widetilde{Q}(h) = \int \widehat{r}_h^2(x)\widehat{g}(x)w(x)dx$, where the integral is over the support of $w(x)$, and it can be computed using either a numerical integration method (e.g. the Gauss-Newton method) or a Monte Carlo simulation method.[6]

The asymptotic behaviors of $\widehat{Q}(h)$ and $\widetilde{Q}(h)$ are similar, so we can focus on $\widetilde{Q}(h)$. To gain insight, we consider the heuristic decomposition

$$\widetilde{Q}(h) = \int \widehat{m}_h^2(x)a(x)dx + \widehat{R},$$

where $a(x) = w(x)/g(x)$, and $\widehat{R}$ is a reminder term dominated by the first (leading) term under suitable regularity conditions. Thus, we can focus on the first term, which will determine the asymptotic distribution of the statistic $\widehat{Q}(h)$. For the first term, we have

$$\begin{aligned}
\int \widehat{m}_h^2(x)a(x)dx &= \frac{1}{(n-h)^2}\sum_{|t-s|>h} \hat{\varepsilon}_{t+h}\hat{\varepsilon}_{s+h}\int K_b(x-X_t)K_b(x-X_s)a(x)dx \\
&\quad + \frac{1}{(n-h)^2}\sum_{|t-s|\leq h} \hat{\varepsilon}_{t+h}\hat{\varepsilon}_{s+h}\int K_b(x-X_t)K_b(x-X_s)a(x)dx \\
&= \widehat{A}(h) + \widehat{B}(h),
\end{aligned}$$

where the term $\hat{A}(h)$ is a sum over $(t,s)$ with $|t-s| > h$, and the term $\hat{B}(h)$ is a sum over $(t,s)$ with $|t-s| \leq h$. For the term $\hat{B}(h)$, we have

$$\widehat{B}(h) = \frac{1}{(n-h)b}\sigma_\varepsilon^2 \int w(x)dx \int K^2(u)du + \frac{2}{(n-h)}\sum_{j=1}^{h}\gamma(j)E[a(X_t)f_j(X_t,X_t)] + O_p((nb)^{-1}),$$

where $\sigma_\varepsilon^2 = var(\varepsilon_{t+h})$, $\gamma(j) = cov(\varepsilon_t, \varepsilon_{t-j})$, and $f_j(\cdot,\cdot)$ is the joint probability density of $(X_t, X_{t-j})$.

[6]The Monte Carlo method can be implemented as follows. Without loss of generality assume that $w(\cdot)$ is a prespecified probability density function.Then we can generate a large $i.i.d.$ sample $\{X_i^*\}_{i=1}^N$ from the probability distribution $w(\cdot)$. Then the average $\widehat{Q}^*(h) = N^{-1}\Sigma_{i=1}^N \widehat{r}_h^2(X_i^*)$ will be arbitrarily close to $\widehat{Q}(h)$ if $N$ is sufficiently large (much larger than the sample size $n$) by the law of large numbers.

Note that generally $\gamma(j) \neq 0$ for $0 \leq j \leq h$ in a $h$-step ahead prediction model (2.3), even when $H_0$ holds. As noted earlier, $\{\varepsilon_{t+h}\}$ generally displays a MA$(h-1)$ structure under $H_0$.

Thus, $\hat{B}(h)$ depends on the serial dependence of $\{\varepsilon_{t+h}\}$ due to the existence of the second term. The effect of serial dependence in $\{\varepsilon_{t+h}\}$ on $\widehat{B}(h)$ is generally larger when the horizon $h$ is larger. In our construction of a test statistic, we could subtract the original form of $\hat{B}(h)$ directly from the global measure $\widehat{Q}(h)$, rather than use the asymptotic approximation of $\hat{B}(h)$. This will make the proposed test robust to the effect of serial dependence contained in $\hat{B}(h)$. The term $\hat{A}(h)$ can be written as

$$\widehat{A}(h) = \frac{2}{(n-h)^2} \sum_{t=R+2}^{n-h} \sum_{s=R+1}^{t-h-1} \hat{\varepsilon}_{t+h}\hat{\varepsilon}_{s+h} \int K_b(x - X_t)K_b(x - X_s)a(x)dx.$$

Under $H_0$, $\widehat{A}(h)$ has an approximately zero mean. Its variance $var(\widehat{A}(h))$ depends on serial dependence in $\{\varepsilon_{t+h}\}$. However, when $\{\varepsilon_{t+h}\}$ has a MA$(h-1)$ structure where $h$ is a fixed integer, the effect of serial dependence in $\{\varepsilon_t\}$ on $var[\widehat{A}(h)]$ is an asymptotically negligible higher order term, and it can be shown that the asymptotic variance of $b^{d/2}(n-h)\hat{A}(h)/\sigma_\varepsilon^2$ is given by

$$V = 8\int w^2(x)dx \int \left[\int K(u)K(u+v)du\right]^2 dv.$$

Using the central limit theorem for degenerate $U$-statistics, we can show $b^{\frac{d}{2}}(n-h)\widehat{A}(h)/\sigma_\varepsilon^2 \xrightarrow{d} N(0, V)$ as $n \to \infty$, as stated below:

**Theorem 1** *Suppose Assumptions A.1–A.6 in the Appendix hold. Then as $n \to \infty$,*

*(i) under $H_0$, we have*

$$\hat{\mathbf{Q}}_h = \frac{\sqrt{b^d}(n-h)\hat{Q}(h)/\widehat{\sigma}_\varepsilon^2 - C/\sqrt{b^d}}{\sqrt{V}} \xrightarrow{d} N(0,1)$$

*where* $C = \int w(x)dx \int K^2(u)du$, $\widehat{\sigma}_\varepsilon^2 = (n-h)^{-1} \sum_{t=R+1}^{T-h} e_{t+h}^2$, *and* $e_{t+h} = \widehat{\varepsilon}_{t+h} - \widehat{r}_h(X_t)$.

*(ii) under $H_A$,*

$$\frac{\hat{\mathbf{Q}}_h}{\sqrt{b^d}(n-h)} \to \frac{V^{-1/2}\int r_h^2(x)g(x)w(x)dx}{\sigma_\varepsilon^2}.$$

The proof of this theorem is given in the Appendix. Among other things, the $\hat{\mathbf{Q}}_h$ test allows the serial correlation of $\{\varepsilon_{t+h}\}$. The $\hat{\mathbf{Q}}_h$ test statistic has an appealing interpretation. Ignoring the centering and scaling factors, the $\hat{\mathbf{Q}}_h$ test statistic is essentially based on the ratio $\hat{Q}(h)/\widehat{\sigma}_\varepsilon^2$. Here, the denominator $\widehat{\sigma}_\varepsilon^2$ is the sample variance of pricing errors, and the numerator $\hat{Q}(h)$ is the average of the squared predictable components neglected by the linear regression model (2.3). Therefore, the ratio $\hat{Q}(h)/\widehat{\sigma}_\varepsilon^2$ can be viewed as an estimator for the neglected signal-to-noise ratio of the linear model. If the neglected pricing signal $\hat{Q}(h)$ is weak relative to the pricing noise $\widehat{\sigma}_\varepsilon^2$, the $\hat{\mathbf{Q}}_h$ test will not reject the null hypothesis $H_0$. If the neglected pricing signal $\hat{Q}(h)$ is sufficiently large relative to the pricing noise $\widehat{\sigma}_\varepsilon^2$, the $\hat{\mathbf{Q}}_h$ test will reject the null hypothesis $H_0$. How large the signal-to-noise ratio should be in order to be considered as sufficiently large is determined by the critical value of the test statistic.

Theorem 1$(ii)$ shows that under $H_A$, the $\hat{\mathbf{Q}}_h$ statistic diverges to infinity at rate $\sqrt{b^d}(n-h)$. Thus, as long as $r_h(x)$ is not zero over the support of the weighting function $w(x)$ under $H_A$ , the $\hat{\mathbf{Q}}_h$ test will be able to reject $H_0$ at any given level with probability approaching one as the sample sizes $R, n \to \infty$.

In computing the neglected pricing signal-to-noise ratio, we have used a nonparametric estimator for $\sigma_\varepsilon^2$. The variance estimator $\widehat{\sigma}_\varepsilon^2$ is based on the nonparametric residual $e_{t+h}$ which is always consistent for the true pricing error $\varepsilon_t^o \equiv Y_{t+h} - E(Y_{t+h}|X_t)$ under both $H_0$ and $H_A$. One could also use the parametric variance estimator $\widetilde{\sigma}_\varepsilon^2 = 1/(n-h)\sum_{t=R+1}^{T-h}\widehat{\varepsilon}_{t+h}^2$ using the estimated residuals from the linear regression model. This estimator is simpler than $\widehat{\sigma}_\varepsilon^2$, and may give better sizes in finite samples, because it is a better estimator for $\sigma_\varepsilon^2$ than $\widehat{\sigma}_\varepsilon^2$ under $H_0$. However, $\widetilde{\sigma}_\varepsilon^2$ is not consistent for the true error variance $Var(\varepsilon_t^o)$ under $H_A$, because it contains the neglected signals. Consequently, it may give a lower power in finite samples.

The test statistic $\hat{\mathbf{Q}}_h$ is constructed to check the out-of-sample predictability of the residual $\hat{\varepsilon}_{t+h}$ using $X_t$. It can also be used to test the null hypothesis $H_0$ in (2.1), namely the predictability of $X_t$ for $Y_{t+h}$. This can be done by replacing the sample size $n$ with $T$, and replacing the estimated residual $\hat{\varepsilon}_{t+h}$ with $Y_{t+h} - \bar{Y}$, where $\bar{Y} = (T-h)^{-1} \sum_{t=1}^{T-h} Y_{t+h}$ is the sample mean of $\{Y_{t+h}\}_{t=1}^{T-h}$. The resulting test statistic is still asymptotically N(0,1) under $H_0$ in (2.1).

Theorem 1$(i)$ implies that approximately $\gamma(n-h)\hat{Q}(h)/\hat{\sigma}_{\varepsilon}^2 \sim \chi^2_{\lambda_n}$ as $R, n \to \infty$ where the constant $\gamma = 2C/V$ and the degree of freedom $\lambda_n = 2C^2/bV$. Here, both constants $\gamma$ and $\lambda_n$ do not depend on any nuisance parameters or nuisance functions, such as the error distribution and density function of $X_t$. In fact, they are independent of the data generating process. Therefore, the asymptotic null distribution of the scaled signal-to-noise ratio statistic $\gamma(n-h)\hat{Q}(h)/\hat{\sigma}_{\varepsilon}^2$ is independent of nuisance parameters or nuisance functions, and approximately $\gamma(n-h)\hat{Q}(h)/\hat{\sigma}_{\varepsilon}^2$ is distributed as $N(\lambda_n, 2\lambda_n)$ where $\lambda_n$ is known. This is the so-called Wilks' phenomena in statistics. One important implication of Wilks' phenomena is that one can simply simulate the null distributions by setting the nuisance parameters under the null hypothesis at reasonable values or estimates.

The asymptotic normality is quite convenient to use in practice. However, several reasons suggest that the asymptotic normal approximation may not work well in finite samples. First, the nonparametric estimator $\widehat{r}_h(x)$ converges slowly to the true function $r_h(x)$ particularly when the dimension $d$ of $X_t$ is relatively large. As it turns out, the neglected reminder terms in the asymptotic expansion of $\hat{Q}(h)/\widehat{\sigma}_{\varepsilon}^2$ are quite close to in order of magnitude to the dominating term which determines the asymptotic normal distribution of $\hat{\mathbf{Q}}_h$. Evidence in related literature shows that the size of nonparametric test statistics is generally very poor in finite samples. Second, in the present framework, $\{\varepsilon_{t+h}\}$ is not an i.i.d. or martingale difference sequence under the null hypothesis. Instead, it follows an MA$(h-1)$ structure in $\varepsilon_{t+h}$ under the null hypothesis $H_0$ due to the $h$-step ahead prediction. Asymptotic analysis shows that the serial dependence in $\{\varepsilon_{t+h}\}$ has no impact on the asymptotic mean $C/\sqrt{b^d}$and the asymptotic variance $V$, but it may substantially affect the finite sample mean and variance of the test statistic $\hat{Q}(h)/\widehat{\sigma}_{\varepsilon}^2$, particularly when $h$ is relatively large. Third, our asymptotic analysis shows that parameter estimation uncertainty in

$\hat{\beta}_t$ has an asymptotically negligible impact on the asymptotic distribution of the proposed test, but the impact depends on the relative magnitude between two sample sizes $R, n$. When the ratio $n/R$ is large (i.e., when $n$ is large relative to $R$) , the impact of parameter estimation uncertainty of $\hat{\beta}_t$ may be substantial in finite samples.[7]

**2.2 Simulation Design and Monte Carlo Evidence**

It is well-known that there exist two well-documented sources of size distortion that may arise in long-horizon regressions if the inference procedures are based on linear prediction models. First, many predictors, such as dividends and earning price ratios, interest rates are highly persistent and only predetermined, rather than fully exogenous. Second, standard test-statistics based on prediction regressions do not have their usual limiting distribution (Cavanagh et al., 1995). The use of standard critical values is known to generate severe size distortion. These problems may carry over to the proposed nonparametric predictability test, particularly when $h$ is large. In order to check the reliability of the proposed test, we investigate the finite performance (both size and power) of the proposed test using data-generating processes that could potentially be employed to capture the persistent behavior commonly observed in predictive regressors. To obtain a reliable reference based on the proposed test in finite samples, we propose the following conditional bootstrap procedure which preserves the MA$(h)$ structure in $\varepsilon_{t+h}$ among other things:

Step 1: Use the first subsample $\{Y_{t+h}, X_t'\}_{t=1}^{R}$ to estimate the linear regression model $Y_{t+h} = X_t'\beta + \varepsilon_{t+h}, t = 1, ..., R$.Obtain the parameter estimator $\hat{\beta}$. Alternatively, rolling estimation or recursive estimation could also be used.

Step 2: Use $\hat{\beta}$ to compute the out-of-sample residual $\hat{\varepsilon}_{t+h} = Y_t - X_t'\hat{\beta}$ for $t = R+1, ..., T-h$.

Step 3: Compute the nonparametric estimates $\widehat{r}_h(X_t)$ and the nonparametric residual $\widehat{e}_{t+h} = \widehat{\varepsilon}_{t+h} - \widehat{r}_h(X_t)$ for $t = R+1, ..., T-h$.

Step 4: Compute the signal-to-noise ratio $\hat{Q}(h)/\widehat{\sigma}_{\varepsilon}^2$ using a prespecified kernel $k(\cdot)$ and bandwidth

[7] One implication of this result is that one should use a large $R$ relative to $n$ in practice to alleviate the impact of parameter estimation in $\widehat{\beta}$.

$b = (n-h)^{1/5}$. In practice, data-driven methods can be used to choose the bandwidth $b$.

Step 5: Estimate an MA($h-1$) model for the nonparametric residual

$$\widehat{e}_{t+h} = \sum_{j=1}^{h-1} \alpha_j v_{t+h-j} + v_{t+h}, t = R+1, ..., T-h.$$

This can be done by the conditional quasi-maximum likelihood estimation. Save the moving average parameter estimates $\{\hat{\alpha}_j\}_{j=1}^{h}$ and estimated residual $\{\hat{v}_{t+h}\}_{t=R+1}^{T-h}$ in the MA($h-1$) model.

Step 6: Draw a bootstrap residual sample $\{\hat{v}_{t+h}^*\}_{t=R+1}^{T-h}$ from the centered empirical distribution of $\{\hat{v}_{t+h}\}_{t=R+1}^{T-h}$. Then obtain a bootstrap residual sample $\{\hat{\varepsilon}_{t+h}^*\}_{t=R+1}^{T-h}$ by the MA($h-1$) model

$$\hat{\varepsilon}_{t+h}^* = \sum_{j=1}^{h-1} \hat{\alpha}_j \hat{v}_{t+h-j}^* + \hat{v}_{t+h}^*, t = R+1, ..., T-h,$$

where the parameter estimates $\{\hat{\alpha}_j\}_{j=1}^{h}$ are obtained in step 5. The bootstrap residual $\{\hat{\varepsilon}_{t+h}^*\}_{t=R+1}^{T-h}$ approximately preserves the MA($h-1$) structure of $\{\varepsilon_{t+h}\}$ under $H_0$.

Step 7: Use the bootstrap sample $\{\varepsilon_{t+h}^*, X_t\}_{t=R+1}^{T-h}$ to compute the bootstrap signal-to-noise ratio $\hat{Q}^*(h)/\widehat{\sigma}_\varepsilon^{*2}$ using the same kernel $k(\cdot)$ and bandwidth $b$ as in Step 4.

Step 8: Repeat Steps 6 and 7 for a total of $B$ times where $B$ is a large number. Denote the obtained $B$ bootstrap test statistics as $\{\hat{Q}_l^*(h)/\widehat{\sigma}_{\varepsilon l}^{*2}\}_{l=1}^{B}$.

Step 9: Compute the bootstrap $p$-value of the $\hat{\mathbf{Q}}_h$:

$$p^* = \frac{1}{B} \sum_{l=1}^{B} \mathbf{1}\left[\frac{\hat{Q}(h)}{\widehat{\sigma}_\varepsilon^2} < \frac{\hat{Q}_l^*(h)}{\widehat{\sigma}_{\varepsilon l}^{*2}}\right],$$

where $\mathbf{1}(\cdot)$ is the indicator function. Reject the null hypothesis $H_0$ at level $\alpha$ if and only if $p^* < \alpha$.

The above resampling approximation is a wild bootstrap. Here, one only need to calculate the signal-to-noise ratio $\hat{Q}(h)/\widehat{\sigma}_\varepsilon^2$ using the observed sample and bootstrap samples. There is no need to compute the original test statistic $\hat{\mathbf{Q}}_h$ which involves calculation of centering and scaling parameters. This follows because computing the bootstrap $p$-value involves ranking $\hat{\mathbf{Q}}_h$ and $\hat{\mathbf{Q}}_h^*$, which is equivalent to ranking the pricing signal-to-noise ratios $\hat{Q}(h)/\widehat{\sigma}_\varepsilon^2$ and $\hat{Q}^*(h)/\widehat{\sigma}_\varepsilon^{*2}$, given the fact that the centering and scaling factors do not depend on nuisance parameters and the data generating process. This greatly simplifies the computation of the test statistic.

When testing predictability of $X_t$ for $Y_{t+h}$ (i.e., testing $H_0$ in (2.1)), Steps 1 and 2 are not needed, the nonparametric residual in step 3 is replaced with $\widehat{e}_{t+h} = Y_{t+h} - \bar{Y}$, and the MA$(h-1)$ models in Steps 6 should be changed to the following:

$$Y^*_{t+h} = \overline{Y} + \sum_{j=1}^{h-1} \hat{\alpha}_j v^*_{t+h-j} + v^*_{t+h}, t = 1, ..., T-h.$$

where $\bar{Y}$ is the sample mean of $\{Y_{t+h}\}_{t=1}^{T-h}$.

We will examine the finite sample performance of the above conditional bootstrap procedure via simulation studies. Table 2.0 summarizes the five data generating processes we use to investigate the empirical size of the tests for both linear and nonlinear predictability check.

Under $A.0(h)$, $X_t$ has no predictive power for $Y_{t+h}$. This allows us to examine the size of the nonparametric test under $H_1 : E(Y_{t+h}|X_t) = E(Y_{t+h})$. Under $A.1(h)$ and $A.2(h)$, there exist linear and nonlinear predictability of $X_t$ for $Y_{t+h}$.This allows us to examine the power of the test under the alternatives. Next, under $B.0(h)$, there is no neglected nonlinear predictability of $X_t$ for $Y_{t+h}$. This allows us to examine the size of the test for the null hypothesis $H_2 : E(\varepsilon_{t+h}|X_t) = 0$. Under $B.1(h)$, there exists neglected nonlinear predictability, which allows us to examine the power of the test.

Tables 2.1a and 2.1b report empirical rejection rates of the test at the 1%, 5%, and 10% nominal levels, for sample sizes of $T = 250, 500, 1000$ and time horizons of $h = 1, 4, 12, 20$. The nonparametric test with the bootstrap procedure has reasonable sizes in finite samples under both the null hypotheses $H_1$ and $H_2$, which are robust to the time horizon $h$ and the persistence of regressor $X_t$ (as measured by the large value of the autoregressive coefficient $\rho$). Moreover, the proposed test has power under various alternatives to $H_1$ and $H_2$ respectively.

There is an upward bias in the predictive coefficient on the regressors due to both long-horizon returns and persistence of the regressors(Stambaugh 1999, Amihud and Hurvich 2004, Lewellen 2004). The existing long-horizon tests with robust Newey-West standard errors suffer from substantial overrejection.[8] Our proposed test has reasonable sizes and robust power performance to investigate the predictability and neglected nonlinear components over different time horizons.

## 3. DATA AND LONG-HORIZON PREDICTABILITY

### 3.1 The Long-horizon Framework and Predictability Regression

We now use the proposed test to investigate the predictability of equity returns over different horizons. Denote the gross return on equity by $G_{t+1} = (P_{t+1} + D_{t+1})/P_{t+1}$ and the continuously compounded return by $\widetilde{y}_{t+1} = log(G_{t+1})$. The long-horizon predictability regression considered is

$$Y_{t+h} = \alpha_h + \beta_h' X_t + \varepsilon_{h,t+h} \tag{3.1}$$

where $Y_{t+h} = (\tau/h)[(\widetilde{y}_{t+1} - r_t) + \cdots + ( \widetilde{y}_{t+h} - r_{t+h-1})$ is the annualized $h$-period excess return for the aggregate stock market, $r_t$ is the risk-free rate from $t$ to $t+1$, and $\widetilde{y}_{t+1} - r_t$ is the one period excess return from time $t$ to $t+1$. The constant $\tau$ is different, depending on the frequency of the data, i.e., $\tau = 1$ (annually), $\tau = 4$ (quarterly), and $\tau = 12$ (monthly). All returns are continuously compounded. The error term $\varepsilon_{h,t+h}$ follows a $MA(h-1)$ process under the null hypothesis of no predictability $H_0 : E(Y_{t+h}|X_t) = E(Y_{t+h})$ and $H_0 : E(\varepsilon_{t+h}|X_t) = 0$. We will use different predictors as instruments in $X_t$ and estimate the regression (3.1) by OLS and compute standard errors of the parameters using the Newey and West(1987) and Hodrick (1992) standard error formula.[9] We use the test proposed in section 2.2 to check the predictability of different variables using the regression framework in (3.1).

### 3.2 Data

[8] Ang and Bekaert (2007) point out that the univariate dividend yield regression displays negligible size distortions in the shortest sample for the one-quarter horizon, but for the bivariate regressions, all tests slightly over-reject at asymptotic critical values with longer horizons.

[9] Using generalized method of moments, (GMM) has an asymptotic distribution $\sqrt{T}(\widehat{\theta} - \theta) \overset{a}{\sim} N(0, \Omega)$ where $\Omega = Z_0^{-1} S_0 Z_0^{-1}$, $Z_0 = E(x_t' x_t)$, and $x_t = (1\ z_t')'$. Hodrick(1992) sums $x_t' x_{t-j}$ into the past and estimates $S_0$ by $\widehat{S}_0 = \frac{1}{T}\sum_{t=h}^{T} wh_t wh_t'$, $wh_t = \varepsilon_{1,t+1}\sum_{i=0}^{h-1} x_{t-i}$.

Following Goyal and Welch (2008), we choose fifteen economic variables to examine predictability of the equity returns using annual, quarterly, and monthly data.

*Stock Returns*: Stock returns are continuously compounded returns on the S&P 500 index, including dividends. Our quarterly data consist of price return (capital gain only), total returns (capital gain plus dividend), and dividends on the Standard & Poor's Composite Index from March 1936 to December 2001. This data is obtained from the Security Price Index Record, published by Standard & Poor's Statistical Service. For monthly data, we use S&P 500 index returns from January 1970 to December 2006 from CRSP's monthend values. Monthly dividends on the S&P 500 index are from Standard & Poor's Statistical Service. For annual data, we get data from 1872 to 2005 provided in Robert Shiller's personal website.

*Risk-free Rate*: The risk-free rate is the T-bill rate from 1920 to 2005. We follow the methods by Goyal and Welch (2007) to estimate T-bill rate prior to the 1920's.[10] For quarterly and monthly data, T-bill rates from 1934 to 2005 are the 3-Month Treasury Bill: the Secondary Market Rate from the economic research data base at the Federal Reserve Bank at St. Louis (FRED).

*Dividend Yields, Earnings Yields, and Dividend Payout Ratio*: Dividends and Earnings are the twelve-month moving sums of dividends and earnings paid on the S&P 500 index. The data from 1871 to 1970 are available from Robert Shiller's website. Quarterly dividends and earnings from 1936 to 2005 and monthly dividends and earnings from 1970 to 2006 are from the S&P Corporation. Dividends and Earnings are summed up over the past year. Monthly or quarterly dividends and earnings are impossible to use because they are dominated by seasonal components. The dividend yield ($d/y$) is defined as $D_t^4/\ P_t$ with the superscript 4 to denote

[10] Commercial paper rates for New York City are from the NBER's Macrohistory data base. These are available from 1871 to 1970. We estimated a regression from 1920 to 1971, which yielded $T - billRate = -0.004 + 0.886 * CommercialPaperRate$ , with an $R^2$ of 95.7% according to Goyal and Welch (2007). Therefore, we instrumented the risk-free rate from 1871 to 1919 with the predicted regression equation. The correlation for the period 1920 to 1971 between the equity premium computed using the actual T-bill rate and that computed using the predicted T-bill rate (using the commercial paper rate) is 99.8%.

that it is constructed using dividends summed up over the past year (four quarters), where $D_t^4 = D_t + D_{t+1} + D_{t+2} + D_{t+3}$ represents dividends summed over the past year and $P_t$ is the price level on S&P 500.[11] We also define the monthly dividend yield with a superscript of 12 to indicate that dividends have been summed over the past 12 months using the same method. We also denote log dividend yields as $dy_t^4 = \log(D_t^4/\ P_t)$ for quarterly data and $dy_t^{12} = \log(D_t^{12}/\ P_t)$ for monthly data. We use the similar definitions for log earnings yields for both quarterly and monthly. The Dividend Payout Ratio ($d/e$) is the difference between the log of dividends and the log of earnings.

*Stock Variance (svar)*: Stock Variance is computed as sum of squared daily returns on the S&P 500. G. William Schwert provided daily returns from 1871 to 1926; data from 1926 to 2005 are from CRSP.

*Book to Market Ratio*: The Book to Market Ratio ($b/m$) is the ratio of book value to market value for the Dow Jones Industrial Average.[12] Book values from 1920 to 2005 are from Value Line's website, specifically their Long-Term Perspective Chart of the Dow Jones Industrial Average.

*Corporate Issuing Activity*: We follow the two measures of corporate issuing activity in Goyal and Welch (2007). Net Equity Expansion ($ntis$) is the ratio of twelve-month moving sums of net issues by S&P listed stocks divided by the total end-of-year market capitalization of S&P stocks. This dollar amount of net equity issuing activity (IPOs, SEOs, stock repurchases, less dividends) for NYSE listed stocks is computed from the CRSP data as $NetIssue_t = Mcap_t - Mcap_{t-1} \cdot (1 + vwretx_t)$, where $Mcap$ is the total market capitalization, and $vwretx$ is the value weighted return (excluding dividends) on the S&P 500 index. These data are available from 1926 to 2005. The second measure, Percent Equity Issuing ($eqis$), is the ratio of equity issuing activity as a fraction

[11] See, e.g., Ball (1978), Campbell (1987), Campbell and Shiller (1988a, 1988b), Campbell and Viceira (2002), Campbell and Yogo (2006), the survey in Cochrane (1997), Fama and French (1988), Hodrick (1992), Lewellen (2004), Menzly, Santos, and Veronesi (2004), and Ang and Bekaert(2007).

[12] See Kothari and Shanken (1997) and Ponti and Schall (1998).

of total issuing activity. This is the variable proposed in Baker and Wurgler (2000).[13] The first equity issuing measure is relative to the aggregate market cap, while the second is relative to the aggregate corporate issuing.

*Long Term Yield (lty)*: The data is from Goyal and Welch (2008). The long-term government bond yield data from 1919 to 1925 is the U.S. Yield On Long-Term United States Bonds series in the NBER's Macrohistory data base. Yields from 1926 to 2005 are from Ibbotson's Stocks, Bonds, Bills and Inflation Yearbook, the same source that provided the *Long Term Rate of Returns (ltr). The Term Spread (tms)* is the difference between the long term yield on government bonds and the T-bill. (See, e.g., Campbell (1987) and Fama and French (1989).)

Corporate Bond Returns: Long-term corporate bond returns from 1926 to 2005 are again from Ibbotson's Stocks, Bonds, Bills and Inflation Yearbook. Corporate Bond Yields on AAA and BAA-rated bonds from 1919 to 2005 are from FRED. The Default Yield Spread (dfy) is the difference between BAA and AAA-rated corporate bond yields. The Default Return Spread (dfr) is the difference between long-term corporate bond and long-term government bond returns. (See, e.g., Fama and French (1989) and Keim and Stambaugh (1986).)

*Inflation* ($infl$): Inflation is the Consumer Price Index (All Urban Consumers) from 1919 to 2005 from the Bureau of Labor Statistics.

*Investment to Capital Ratio* ($i/k$): The investment to capital ratio is the ratio of aggregate (private nonresidential fixed) investment to aggregate capital for the whole economy.

*Consumption, wealth, income ratio* ($cay$): The variable $cay$ is proposed by Lettau and Ludvigson (2001). Data for $cay$'s construction at quarterly frequency from the second quarter of 1952 to the fourth quarter of 2005 are available from Martin Lettau's website. The annual data from 1948 to 2001 is available from Martin Lettau's website.

Table 3.1 summarizes the descriptive statistics of the predictors. Panels (a), (b), and (c)

[13] We get the data from http://pages.stern.nyu.edu/ jwurgler/.

report the results for quarterly, monthly, and annual data respectively. We summarize the test statistics of the predictors under the null hypothesis of a unit root. Short rates, dividend and earnings yields, book-to-market ratio, and inflation are all highly persistent at different frequencies.

## 4. IS THE PREDICTABILITY THERE?

In this section, we first apply the nonparametric test to examine whether there exists the predictability of equity returns in both short and long horizons. Then we test the predictability power of the conventional predictive models and the historical mean model.

### 4.1 Short-Horizon and Long-Horizon Predictability

In this section, we use the linear predictive regression $Y_{t+h} = \alpha_h + \beta_h' X_t + \varepsilon_{h,t+h}$ in (3.1) to check the predictability of equity returns by using quarterly, monthly, and annual data. For quarterly and monthly data, we report the results of four sample periods (1936-2001, 1952-2001, 1936-1990, and 1952-1990), which are the same sample periods used in Ang and Bekaert (2007).[14]

Table 4.1 summarizes the results of the 1-quarter, 1-year, 3-year, and 5-year-ahead predictability of excess return predictability. Table 4.1a focuses on the univariate regression with dividend yields and earnings yields as the regressor. The t-statistics in parentheses are computed using the Newey and West (1987) and Hodrick (1992) standard error respectively. The parameter estimates have similar patterns over the four periods, but the estimated coefficients are twice as large as for the period omitting the 1990s from the sample. The Hodrick standard errors are smaller than the Newey-West standard errors. During the 1936-2001 and 1952-2001 periods, there is no evidence of predictability for dividend yields for both short and long horizons. For the

[14] Interest rate data are hard to interpret before the 1951 Treasury Accord, as the Federal Reserve pegged interest rates during the 1930s and the 1940s. Hence, we examine the post-Accord period, starting in 1952. Second, the majority of studies establishing strong evidence of predictability use data before or up to the early 1990s. Studies by Lettau and Ludvigson (2001) and Goyal and Welch (2003) point out that predictability by the dividend yield is not robust to the addition of the 1990s decade. Hence, we separately consider the effect of adding the 1990s to the sample.

1936-1990 periods, there is strong predictability for dividend yields over the 1-quarter, 1-year, 3-year, and 5-year ahead time horizons respectively. For the 1952-1990 period, there exists only the predictability for 1-quarter and 1-year shorter horizons.

Table 4.1a also reports the bootstrap $p$-value for the predictability test under two hypotheses $H1$ and $H2$. Hypothesis $H1$ is $H_0 : E(Y_{t+h}|X_t) = E(Y_{t+h})$, namely that $X_t$ has no predictive power for $Y_{t+h}$. Hypothesis $H2$ is $H_0 : E(\varepsilon_{t+h}|X_t) = 0$, namely that $X_t$ has no neglected nonlinear predictive power for $Y_{t+h}$ beyond the linear model (3.1). As mentioned in Section 2.1, the $\hat{\mathbf{Q}}_h$ test has an appealing interpretation: it is essentially based on the ratio $\hat{Q}(h)/\widehat{\sigma}_{\varepsilon}^2$, where the denominator $\widehat{\sigma}_{\varepsilon}^2$ is the sample variance of pricing errors, and the numerator $\hat{Q}(h)$ is the average of the squared predictable components neglected by the linear regression model. Therefore, the ratio $\hat{Q}(h)/\widehat{\sigma}_{\varepsilon}^2$ can be viewed as an estimator for the neglected signal-to-noise ratio of the linear prediction model (3.1). If the neglected pricing signal $\hat{Q}(h)$ is weak relative to the pricing noise $\widehat{\sigma}_{\varepsilon}^2$, the $\hat{\mathbf{Q}}_h$ test will not reject the null hypothesis $H_0$. If the neglected pricing signal $\hat{Q}(h)$ is strong relative to the pricing noise $\widehat{\sigma}_{\varepsilon}^2$, the $\hat{\mathbf{Q}}_h$ test will reject the null hypothesis $H_0$. The results for testing $H1$ show that the $\hat{\mathbf{Q}}_h$ test strongly rejects the null hypothesis $H1$ for dividend yields over the four sample periods. This implies that dividend yield is a significant predictor of excess returns at all time horizon $h$, which is consistent with Campbell and Shiller (1988a,b). We also examine whether there exists neglected nonlinear predictability of dividend yield for equity returns. Table 4.1a show that the $\hat{\mathbf{Q}}_h$ test strongly rejects the null hypothesis $H2$ for all four sample periods. It implies that there exists a nonlinear predictive relationship between $X_t$ and $Y_{t+h}$. The right four columns of Table 4.1a also report a univariate regression with the earnings yield as regressor. The t-statistics suggest that there is no strong evidence for linear predictability of earnings yields over the four sample periods. However, the nonparametric tests for hypothesis $H1$ and $H2$ show that earnings yield is a good predictor for equity returns over all the different time horizons $h$.

Table 4.1b summarizes the bivariate regression with log dividend yields and short rate together as regressors. It reports the bootstrap $p$-value of the predictability test for six various hypotheses $H1-H6$, where $X_1$ represents the short rate $r$ and $X_2$ the dividend yield. The six hypotheses are, respectively, $H1 : E(Y_{t+h}|X_{1t}) = E(Y_{t+h})$, $H2 : E(\varepsilon_{t+h}|X_{1t}) = 0$, $H3 : E(Y_{t+h}|X_{2t}) = E(Y_{t+h})$, $H4 : E(\varepsilon_{t+h}|X_{2t}) = 0$, $H5 : E(Y_{t+h}|X_{1t}, X_{2t}) = E(Y_{t+h})$, and $H6 : E(\varepsilon_{t+h}|X_{1t}, X_{2t}) = 0$. Hypotheses $H1$ and $H3$ are to test the predictability of the short rate or dividend yield separately and Hypothesis $H5$ is to test the joint predictability of the short rate and dividend yield together. The short rate has strong predictability over the four periods based on the Newey-West standard errors but the predictability only exists at short horizons when using the Hodrick (1992) standard errors. In the bivariate regression, there is evidence of predictability of dividend yields for equity returns when the sample period excludes the 1990s. The coefficient on the dividend yield is larger using bivariate regression than the univariate regression. This suggests that the univariate regression suffers from an omitted variable bias that lowers the marginal impact of dividend yields on expected excess returns.[15] It is consistent with Ang and Bekaert (2007). They find that dividend yields predict excess returns only at short horizons together with the short rate and do not have any long-horizon predictive power. At short horizons, the short rate strongly negatively predicts returns. However, our $\hat{\mathbf{Q}}_h$ test significantly rejects the hypotheses $H1 - H4$ for all the four sample periods. It indicates that short rate and dividend yield are two good predictors for equity returns but it cannot be fully captured by linear predictive regressions.

The $\hat{\mathbf{Q}}_h$ test rejects the hypothesis $H5$ only at the 1-quarter-ahead time horizon and fails to reject at the 1-year, 3-year, and 5-year-ahead time horizon of the 1936-2001 period. The $\hat{\mathbf{Q}}_h$ test rejects hypothesis $H5$ for the 1952-2001, 1936-1990, and 1952-1990 periods. There is evidence of joint predictability for the short rate and dividend yield together for the three sample period of 1952-2001, 1936-1990, and 1952-1990. The predictability of the short rate and dividend yield

[15] Engstrom (2003), Menzly, Santos, and Veronesi (2004), and Lettau and Ludvigson (2005) also note that a univariate dividend yield regression may understate the dividend yield's ability to forecast returns.

for equity returns is only identified at the 1-quarter-ahead short horizon of the 1936-2001 period. The $\hat{\mathbf{Q}}_h$ test rejects hypothesis $H6$ for the four time periods except for the 5-year-ahead time horizon of the 1936-2001 and 1952-1990 periods. The bivariate linear regression does not have the optimal predictive power for equity returns and there are neglected nonlinear components that are not captured by the linear regression models. Nevertheless, there may exist the long-horizon predictability for the 5-year time horizon in the 1936-2001 and 1952-1990 periods which can be captured in the linear regression model since there is no strong evidence to reject hypothesis $H6$.

To compare with Lamont (1998) and Ang and Bekaert (2007), we report a bivariate regression of excess returns on log dividend and log earnings yields. Lamont (1998) finds a positive coefficient on the dividend yield and a negative coefficient on the earnings yield. He argues that the predictive power of the dividend yield stems from the role of dividends in capturing permanent components of prices, whereas the negative coefficient on the earnings yield is due to earnings being a good measure of business conditions. Ang and Bekaert (2007) finds that dividend and earnings yields do not have a strong predictive power and only when the 1990s are excluded they find significant coefficients for dividend and earnings yields. Table 4.1c summarizes the bivariate regression with the log dividend yields and log earnings yields together as regressors. The dividend yields and earnings yields have a strong predictive power for equity returns over the four time periods when using the Newey-West (1987) standard errors. The results using the Hodrick (1992) standard errors are similar to Ang and Bekaert (2007). The $\hat{\mathbf{Q}}_h$ test rejects the six hypotheses over all the time horizons and for all 4 time periods. It supports Lamont (1998)'s arguments. Dividend yields and earnings yields have the predictability power for equity returns but the bivariate linear regression model cannot fully capture such predictability.

Table 4.1d summarize the test results of the trivariate regression with the short rate, log dividend yields, and log earnings yields together as regressors. When we add the short rate as a predictor in a trivariate regression of excess returns on risk-free rates, dividend and earnings yields, the coefficients on dividend and earnings yields remain insignificantly different from zero, and the sign on the earnings yield is fragile. For the post-1952 samples, the short rate, and dividend yields have predictive power in the presence of the earnings yield. The results for the $\hat{\mathbf{Q}}_h$ test show that the three variables short rate, dividend yields, and earnings yields do have the predictability power for the equity returns. The $\hat{\mathbf{Q}}_h$ test for the joint predictability of the three variables rejects the hypothesis $H7(H_0 : E(Y_{t+h}|X_{1t}, X_{2t}, X_{3t}) = E(Y_{t+h}))$ for most of the cases except the 1-year and 3-year ahead forecasts in the 1936-2001 and 1936-1990 periods and the 5-year ahead forecast in the 1952-2001 period. The trivariate regression does not capture the true equity returns and it needs a better nonlinear model to capture it.

We also use the monthly data from January 1970 to December 2006 to test the predictability of the short rate, dividend yields, and earnings yields in univariate, bivariate, and trivariate regressions respectively.[16] We get similar results using monthly data. Using the Hodrick (1992) standard errors, our results suggest that the short rate has strong predictability. The nonparametric predictability tests show that the three variables are good candidates to predict equity returns but it cannot be fully captured by the linear predictive models.

**4.2 Does the prevailing models beat the historical mean?**

Goyal and Welch (2007) reexamine the performance of predictive variables commonly used in the academic literature. They find that the historical mean model outperforms the predictor-based models in terms of both in-sample and out-of-sample performance. We consider both In-Sample (IS) and Out-of-Sample (OOS) tests. Following Goyal and Welch (2007), the OOS forecasts use only the data available up to the time at which the forecast is made. Let $e_N$ denote the vector of rolling OOS errors from the historical mean model and $e_A$ denote the vector of rolling OOS errors from the OLS model. The OOS statistics are computed as $R^2 = 1 - \frac{MSE_A}{MSE_N}$, $\overline{R}^2 = R^2 - (1 - R^2) \cdot \left|\frac{T-k}{T-1}\right|$ , $\Delta RMSE = \sqrt{MSE_N} - \sqrt{MSE_A}$. It is important but difficult for

[16] The results for the regressions and predictability tests are summarized in Table 4.2. Please check the supplementary document for details.

OOS tests to choose the periods over which a regression model is estimated and subsequently evaluated. In this section we consider the annual prediction with similar data used in Goyal and Welch (2007). For the OOS test, we use the time period twenty years after data are available as the out-of-sample validation period.

We estimate the predictive regressions $Y_{t+h} = \alpha_h + \beta_h' X_t + \varepsilon_{h,t+h}$ in (3.1). The predictive variables $X_t$ are log dividend yields, log earnings yields, dividend payout ratio, short rate, inflation, book-to-market ratio(b/m), investment to capital ratio(i/k), corporate issuing activity (Eqis and Ntis), and consumption, wealth, and income ratio(cay). The results are summarized in Table 4.3. The t-statistics in parentheses are computed using the Newey and West (1987) and Hodrick (1992) standard errors. We report the bootstrap $p$-value for the predictability test under two hypotheses $H1$ and $H2$. Hypothesis $H1$ is $E(Y_{t+h}|X_t) = E(Y_{t+h})$ and Hypothesis $H2$ is $E(\varepsilon_{t+h}|X_t) = 0$. Table 4.3 summarizes both in-sample and out-of-sample results. We use the difference of the signal-to-noise ratios as a criterion to evaluate the predictability power: $\Delta(\frac{Q_h}{\sigma^2}) = \hat{Q}_N(h)/\widehat{\sigma}_\varepsilon^2 - \hat{Q}_A(h)/\widehat{\sigma}_\varepsilon^2$, where $\hat{Q}_N(h)/\widehat{\sigma}_\varepsilon^2$ and $\hat{Q}_A(h)/\widehat{\sigma}_\varepsilon^2$ are the signa-to-noise ratios of the historical mean model and the prevailing predictive model respectively. If $\Delta(\frac{Q_h}{\sigma^2}) > 0$, there is more neglected signal which cannot be explained by the historical mean model and thus the prevailing predictive model performs better. If $\Delta(\frac{Q_h}{\sigma^2}) < 0$, there is more neglected signal which cannot be captured by the prevail predictive model and so the historical mean model performs better.

Table 4.3 shows that with a linear predictive model, all variables considered are insignificant and only several variables (dividend yield, short rate, eqis, and cay) are significant for 1-year ahead forecast using the Newey-West standard errors. However, the results for both in-sample and out-of-sample nonparametric tests show that all variables(i.e., log dividend yields, log earnings yields, dividend payout ratio, short rate, inflation, book-to-market ratio(b/m), investment to capital ratio (i/k), corporate issuing activity (Eqis and Ntis), and consumption, wealth and income ratio(cay)) have predictability power for equity returns. The historical meam model has a higher signal-to-noise ratio than the prevailing predictive models for both in-sample and out-of-sample tests. There exists more neglected signals which cannot be explained by the historical mean model and the prevailing predictive model performs better. This conclusion differs from Goyal and Welch (2007) and supports Campbell and Thompson (2007).

## 5. OUT-OF-SAMPLE FORECASTING OF EQUITY RETURNS

As mentioned in the previous sections, Goyal and Welch (2008) create enough of a controversy within the profession and argue that the historical average equity return model produces better forecasts than regressions of excess returns on predictive variables. However, Campbell and Thompson (2008) argue that the empirical models can have a better out-of-sample forecast if one restricts their parameters in the economically justified ways. Cochrane (2008) argues that the out-of-sample tests performed by Goyal and Welch are relatively weak, and have a better in-sample predictability power. The literature emphasizes that most of the linear predictive regressions have poor out-of-sample performance (Goyal and Welch (2003, 2007); Campbell and Thompson (2007)). The lack of consistent out-of-sample performance in Goyal and Welch (2008) indicates the need for improved forecasting methods to better establish the empirical reliability of equity premium predictability. Rapach, Strauss, and Zhou (2010) propose a combination approach to improve the out-of-sample equity premium forecasting problem. In this section, we propose nonparametric estimators to forecast the equity returns and compare the out-of-sample performance of the different models.

### 5.1 Nonparametric forecast, linear predictive model, and Historical Mean Model

In the previous sections, our nonparametric test has showed that there exists the predictability of equity returns at the short and long horizons. The predictors such as dividend yields, earnings yields, dividend payout ratio, short rate, inflation, book-to-market ratio, investment to capital ratio, corporate issuing activity, and consumption, wealth, and income ratio have predictability power for equity returns, but this cannot be fully captured by popular linear regression models. We find that the poor out-of-sample performance of most linear prediction models is due to the limitation of linear models. To find the better fit of the equity returns, we use two nonparametric estimators to forecast the equity returns following the section 2.1.

The first estimator is to use a kernel method to estimate $E(\varepsilon_{t+h}|X_t)$ and capture potentially neglected nonlinear predictable component in $\varepsilon_{t+h}$. The expected equity returns can be estimated by:

$$\begin{aligned}
\widehat{E}(Y_{t+h}|X_t) &= X_t'\widehat{\beta} + \widehat{E}(\varepsilon_{t+h}|X_t) = X_t'\widehat{\beta} + \widehat{r}_h(x). \\
&= X_t'\widehat{\beta} + \frac{\widehat{m}_h(x)}{\widehat{g}(x)} \\
\widehat{m}_h(x) &= \frac{1}{t-h}\sum_{s=1}^{t-h}\hat{\varepsilon}_{s+h}K_b(x-X_s), \\
\widehat{g}_h(x) &= \frac{1}{t-h}\sum_{s=1}^{t-h}K_b(x-X_s)
\end{aligned}$$

where $x = (x_1, x_2, \cdots, x_d)'$, $y = (y_1, y_2, \cdots, y_d)'$, and $K_b(x-y) = \Pi_{i=1}^d b^{-1}K[(x_i - y_i)/b]$. The kernel function $K(\cdot)$ is is a prespecified symmetric probability density function. The second estimator is to use a kernel method to estimate $E(Y_{t+h}|X_t)$ directly. We can predict the equity returns by

$$\widehat{E}(Y_{t+h}|X_t) = \frac{\frac{1}{t-h}\sum_{s=1}^{t-h}Y_{s+h}K_b(x-X_s)}{\widehat{g}_h(x)}$$

Our out-of-time kernel estimator here is a weighted average of the past equity returns, where the weights depend on the values of the predictive variable. For comparison, the historical mean estimator used by Goyal and Welch (2008) is a simple average of the past equity returns. The weights in our methods can provide more useful information in out-of-sample forecasts then the historical mean model. In our paper, we choose the bandwith which is correlated with the size of the out-of-sample forecasting period. There are several reasons that we use nonparametric models to detect the nonlinear predictive components other than nonlinear models. First, the existing economic theory in the literature can not give a concrete form of the nonlinear predictive model because we don't know where the nonlinearity exactly comes from. Second, nonlinear models, such as cubic or quadratic functions, may misspecify the nonlinearity of the true data. There may exist outliers and it will cause the spurious identification for the predictability. Third, nonparametric model can capture both the linear and nonlinear component without the model specification. It is not restricted to the parametric forms. It can fit the data better than simply the linear or nonlinear parametric model.

We want to compare the out-of-sample forecast results of four models: historical mean model, linear predictive model, and two nonlinear predictive models. The three measures we use are MSE (Mean squared error), MAE (Mean absolute error), and RMSE (Root mean squared error) defined as below.

$$\begin{aligned} MSE &= \frac{1}{n-h}\sum_{s=R+1}^{T-h}(Y_{s+h}-\widehat{Y}_{s+h})^2, \\ MAE &= \frac{1}{n-h}\sum_{s=R+1}^{T-h}\left|Y_{s+h}-\widehat{Y}_{s+h}\right|, \\ RMSE &= \sqrt{\frac{1}{n-h}\sum_{s=R+1}^{T-h}(Y_{s+h}-\widehat{Y}_{s+h})^2}. \end{aligned}$$

Table 5.1 show the out-of-sample results of the univariate linear predictive models. Table 5.1a summarize the MSE, MAE, and RMSE of univariate linear predictive regression for dividend yield during the period 1936-2001, 1952-2001, 1936-1990, and 1952-1990. The benchmark model is the historical average equity returns. The alternative models are linear predictive model and two nonparametric predictive models. Both the first and second nonparametric predictive models have the lower RMSE than the historical mean model for the period 1936-1990, and 1952-1990. The second nonparametric predicitve model can do a better job than historical mean model in both short horizon and long horizon across the different time periods. The linear predictive model has a higher RMSE than the historical mean model which is consistent with the results in Goyal and Welch (2007). The results show that the two nonparametric predicitve models can improve the out-of-sample performance at both short horizon and long horizon compared to the linear predictive model and the historical mean model. We also get the similar results for earning yield by comparing the MSE, MAE, and RMSE of the four models across the different time periods.[17]

Table 5.2 reports out-of-sample bivariate regression results with short rate as an additional regressor using quarterly data. Table 5.2a summarize the MSE, MAE, and RMSE of the bivariate predictive regression using dividend yield and short rate during the period 1936-2001, 1952-2001, 1936-1990, and 1952-1990. The benchmark model is historical average equity returns. The alternative models are linear predictive model and two nonparametric predictive models. For the period of 1936-2001, 1952-2001, 1936-1990, and 1952-1990, the second nonparametric predictive regression model has the smallest RMSE. The second nonparametric predicitve model can do better job than historical mean model in both short horizon and long horizon. The first nonparametric predictive model has a lower RMSE than the historical mean model except for the period of 1952-2001. For the post-Treasury Accord 1952–2001 sample, linear predictive model and the first nonparametric predictive model has higher RMSE than the historical mean model. In the bivariate regression with earning yield and short rate, the second nonparametric predictive regression model is superior to the other three models across the different time periods.[18] Ang and Bekaert (2007) find that dividend yields, together with the short rate, predict excess returns only at short horizons. Our nonparametric predictive models shows higher prediction power for equity returns and the short rate, dividend yields, and earnings yields have good predictability power at both short and long horizons. The results of our nonparametric predictive models are

[17] The results for the regressions and predictability tests are summarized in Table 5.1b. Please check the supplementary document for details.

[18] The results for the regressions and predictability tests are summarized in Table 5.2b. Please check the supplementary document for details.

robust for the four subsamples.

Goyal and Welch (2007) argue that the historical average excess stock return forecasts future excess stock returns better than the predictive regressions. In this paper we choose fifteen variables used in Goyal and Welch (2008). They are dividend-price ratio ($D/P$), dividend yield ($D/Y$), earnings-price ratio ($E/P$), dividend-payout ratio ($D/E$), stock variance ($SVAR$), book-to-market ratio ($B/M$), net equity expansion ($NTIS$), treasure bill rate ($TBL$), long-term yield ($LTY$), long-term return ($LTR$), term spread ($TMS$), default yield spread ($DFY$), default return yield ($DFR$), inflation ($INFL$), and investment-to-capital ratio ($I/K$). The benchmark model is historical average equity returns. The alternative models are linear predictive model and two nonparametric predictive models. Table 5.3 report the out-of-sample forecasting results of equity premium using the annual data from 1872 to 2005. Consistent with the previous results, the second nonparametric predicitve model can do better job than historical mean model and linear predictive model at both short horizon and longer horizon. Table 5.5 report the out-of-sample forecasting results of equity returns using the quarterly data from 1947:1 to 2007:4. We consider the out-of-sample forecast evaluation periods covering from 1965:1 to 2007:4 consistent with Goyal and Welch (2008). The statistical results show that the second nonparametric predicitve model outperform the historical mean model and linear predictive model at both short horizon and long horizon. For most predictors except dividend-price ratio ($D/P$), dividend yield ($D/Y$), earnings-price ratio ($E/P$), and book-to-market ratio ($B/M$), the two nonparametric models outperform the historical mean model and linear regression model.

The out-of-sample forecasting results show that linear predictive model has higher RMSE than historical mean model, and apparently it is consistent with Goyal and Welch (2008). However, from our nonparametric test results it shows that the linear predictive regression models have a lower signal-to-noise ratio and can beat the historical mean model without any restrictions. Our nonparametric test can detect the nonlinear predictive component of the equity returns which contains more information that the historical mean model can not provide. Campbell and Thompson (2008) show the similar results when imposing some restrictions on the predictors. Following the same logic, our nonparametric predictive model can directly capture both the linear and nonlinear predictive components of the equity returns and it has a better out-of-sample forecasting performance. It is consistent with our nonparametric results in the previous sections.

The out-of-sample forecasting performance shows that the predictability power of equity returns increases with lower RMSEs when the forecasting horizon $h$ increases. Ang and Bekaert (2007) find that dividend yields, together with the short rate, predict excess returns only at short horizons and do not have any long-horizon predictive power. Goyal and Welch (2008), and Campbell and Thomason (2008) do not find the relationship between the predictability and time horizons. However, Fama and French (1987a) show that the autocorrelations of portfolio returns imply that time-varying expected returns explain 25-40% of 3- to 5-year return variances. Poterba and Summers (1987) find that long-horizon stock returns have large predictable components using variance-ratio tests. Economic theory has shown that there exists nonlinear relationship between equity returns and the predictors such as dividend yield. If expected returns have strong autocorrelation, the 1- to 4-year ahead forecast of equity returns are highly correlated. As a consequence, the variance of expected returns grows faster with the time horizon than the variance of unexpected returns. The variation of expected returns becomes a larger fraction of the variation of returns. In the short run, the nonleariy is relatively weak. When time accumulates, the nonlinear relationship becomes stronger in the long run. Our nonparametric method has its advantage to detect the nonlinearity. That explains why the RMSE becomes smaller when the time horizon $h$ becomes larger. In other words, the predictability of the linear predictive models performs better in the short run than in the long run. The difference of the RMSE between nonparametric model and linear predictive model is relatively small when forecasting horizon $h$

is small and it becomes bigger when the forecasting horizon $h$ increases.

**5.2 Individual Forecast and Combined Forecast**

In the literature, most papers focus on a set of predictors based on theoretical models. From an academic viewpoint, the use of model-based predictors facilitates an understanding of specific aspects of the economic mechanism. From an investor's viewpoint, however, these predetermined variables may not be enough to capture all information required in decision making. Forecast combination has recently received renewed attention in the forecasting literature (Stock and Watson (1999, 2003, 2004), Rapach, Strauss, and Zhou (2010)). In this section, we use both the individual forecast and the combined forecast to examine the out-of-sample forecasting performance of equity returns.

We follow the definition of the combined forecast by Rapach, Strauss, and Zhou (2010). The combination forecasts of $Y_{t+1}$ made at time $t$ are weighted averages of the $M$ individual forecasts based on $\widehat{Y}_{c,t+h} = \sum_{i=1}^{M} \omega_{i,t} \widehat{Y}_{i,t+h}$ where $\{\omega_{i,t}\}_{i=1}^{M}$ are the ex ante combining weights formed at time $t$, and $\widehat{Y}_{i,t+h}$ is the out-of-sample forecast of the equity premium based on the individual predictive models[19]. For the individual predictors, we choose the 15 predictors used in the previous sections. We calculate five different combining methods based on the definition of the weights. The first three methods use the simple averaging schemes: mean, median, and trimmed mean. The mean combination forecast sets $w_{i,t} = 1/M$ for $i = 1, \cdots, M$. The median combination forecast is the median of $\{\widehat{Y}_{i,t+h}\}_{i=1}^{M}$, and the trimmed mean combination forecast sets $w_{i,t} = 0$ for the individual forecasts with the smallest and largest values and $w_{i,t} = 1/(M-2)$ for the remaining individual forecasts. The other two combining methods are based on Stock and Watson (2004) and Rapach, Strauss, and Zhou (2010), where the combining weights formed at time $t$ are functions of the historical forecasting performance of the individual models over the holdout out-of-sample period. Their discount mean square prediction error ($DMSPE$) combining method employs the following

[19] $Y_{t+h} = \alpha_h + \beta_h' X_t + \varepsilon_{h,t+h}$.

weights: $w_{i,t} = \phi_{i,t}^{-1} / \sum_{j=1}^{M} \phi_{j,t}^{-1}$, $\phi_{i,t} = \sum_{s=R}^{t-1} \theta^{t-1-s}(Y_{i,t+h} - \widehat{Y}_{i,t+h})^2$ and $\theta$ is a discount factor. The DMSPE method thus assigns greater weights to individual predictive regression model forecasts that have lower MSPE values (better forecasting performance) over the holdout out-of-sample period. We consider the two values of 1.0 and 0.9 for $\theta$.

Table 5.4 report the out-of-sample combined forecasting results of equity returns using the annual data. Consistent with the previous results, the two nonparametric predicitve models have lower $RMSE$ and can do a better job than historical mean model and linear predictive model in both short horizon and long horizon. Furthermore, the linear predictive model can outperform the historical mean model by using combined method. Table 5.6 report the out-of-sample combined forecasting results of equity premium using the quarterly data from 1947:1 to 2007:4. We choose the out-of-sample forecasting periods from 1965:1 to 2007:4 consistent with Goyal and Welch (2008). The statistical results show that the two nonparametric predicitve models outperform the historical mean model and linear predictive model at both short horizon and long horizon. Rapach, Strauss, and Zhou (2010) find that forecast combination outperforms the historical mean model by statistically and economically meaningful margins for out-of-sample period. Our results are consistent with their conclusion. Using our nonparametric methods, both combined and individual forecast outperform the historical average. The combined forecast methods outperform the individual forecast methods with lower $RMSE$.

Figure 5.1 and 5.3 illustrate the out-of-sample forecasting performance for individual predictor-based methods using annual data over 1-year and 5-year rolling windows. The black dotted line is realized equity returns and the red dotted line is the unconditional historical average. The red and green solid line are the forecasted returns by the first and second nonparametric models respectively. For individual predictor-based models, the second nonparametric prediction is below the unconditional historical average line for most of the cases. Figure 5.2 and 5.4 illustrate the out-of-sample forecasting performance for combined methods using annual data over 1-year and

5-year rolling windows. For combined predictor-based models, the two nonparametric prediction models are below the unconditional historical average line. On average the nonparametric method outperforms the historical average. Campbell and Thompson (2008) show that imposing theoretically motivated restrictions on individual predictive regression models can improve their out-of-sample performance. Our nonparametric prediction can improve the out-of-sample performance without restrictions. We also get the similar results using the quarterly data. Figure 5.5, 5.7 and 5.9 illustrate the out-of-sample performance for individual methods using quarterly data over 1-quarter, 1-year, and 3-year rolling windows.[20] Figure 5.6, 5.8 and 5.10 illustrate the out-of-sample performance for combined methods using quarterly data over 1-quarter, 1-year, and 3-year rolling windows.

Compared the individual forecast with the combined forecast, we find that combined predictive models have lower RMSE than individual predictive models for the same forecasting horizon $h$. Fama and French (1989) and others show that the existing predictor variables can detect changes in economic conditions that potentially signal fluctuations in the equity risk premium. But the dividend yield or term spread alone could capture different components of business conditions, and a given individual economic variable may give a number of "false signals" and/or imply an implausible equity risk premium during certain periods. Rapach, Strauss, and Zhou (2010) argue that if individual forecasts based on the predictors are weakly correlated, forecast combinatio should be less volatile and more reliably track movements in the equity risk premium. This is one explanation why the combined forecast methods outperform the individual forecast methods.

On the other hand, the nonparametric predictive model can fit the equity return better based on the predictors. First, nonparametric prediction generates a forecast with a variance near that of the smooth real equity return data, thereby reducing the noise in the individual predictive

[20] Figure 5.5, 5.7 and 5.9 can be found in the supplementary document for details.

regression model forecasts. Second, combining forecast incorporates information from a host of economic variables while the historical average ignores economic variables. Combined forecasts have a substantially smaller bias than the historical average. Combining individual forecasts helps to reduce forecast variability.

### 5.3 Economic Implication

In this section, we investigate how well our nonparametric predictive models capture true expected returns implied by the models. Campbell and Thompson (2008) argue that even very small positive $R^2_{OS}$ values, such as 0.5% for monthly data and 1% for quarterly data, can signal an economically meaningful degree of return predictability in terms of increased annual portfolio returns for a mean-variance investor. This provides a simple assessment of forecastability in practice. To evaluate the economic significance, we calculate realized utility gains for a mean-variance investor on a real-time basis, following Marquering and Verbeek (2004), Campbell and Thompson (2008), Welch and Goyal (2008), Rapach, Strauss, and Zhou (2010), and Wachter and Warusawitharana (2009). More specifically, we first compute the average utility for a mean-variance investor with relative risk aversion parameter $\gamma$ who allocates her portfolio monthly between stocks and risk-free bonds using forecasts of the equity premium based on the historical average[21]. A mean-variance investor who forecasts the equity premium using the historical average will decide how to allocate the share of her portfolio to equities $w_{0,t} = (\frac{1}{\gamma})(\frac{\overline{r}_{t+1}}{\widehat{\sigma}^2_{t+1}})$ in period $t+1$ at the end of period t, where $\widehat{\sigma}^2_{t+1}$ is the rolling-window estimate of the variance of stock returns. Over the out-of-sample period, the investor realizes an average utility level of:

$$\widehat{v}_0 = \widehat{\mu}_0 - (\frac{1}{2})\gamma\widehat{\sigma}^2_0 \tag{5.1}$$

where $\widehat{\mu}_0$ and $\widehat{\sigma}^2_0$ are the sample mean and variance, respectively, over the out-of-sample period for the return on the benchmark portfolio formed using forecasts of the equity premium based on the historical average. We then compute the average utility for the same investor when she forecasts the equity premium using an individual predictive regression model or combining method. She will choose an equity share of $w_{j,t} = (1/\gamma)(\overline{r}_{t+1}/\widehat{\sigma}^2_{t+1})$ and realizes an average utility level of:

$$\widehat{v}_j = \widehat{\mu}_j - (\frac{1}{2})\gamma\widehat{\sigma}^2_j \tag{5.2}$$

[21] We assume that the investor estimates the variance using a ten-year rolling window of quarterly returns.

where $\widehat{\mu}_j$ and $\widehat{\sigma}_j^2$ are the sample mean and variance, respectively, over the out-of-sample period for the return on the portfolio formed using forecasts of the equity premium based on an individual predictive regression model or combining method indexed by $j$. We measure the utility gain $\Delta$ as the difference between (5.2) and (5.1), and the utility gain (or certainty equivalent return) can be interpreted as the portfolio management fee that an investor would be willing to pay to access the additional information available in a predictive regression model or combination forecast relative to the information in the historical equity premium alone. We report results for $\gamma$= 3; the results are qualitatively similar for other reasonable $\gamma$ values. Table 5.5 report the average utility gains for individual predictive models by using different methods. 13 of the 15 predictors produce positive utility gains relative to the historical average for all three forecast models except for $LTY$ and Book-to-Market ratio. The average utility gains shows that our nonparametric forecast successfully produces a positive utility gain for all the predictors. Table 5.6 report the average utility gains for combined forecast. The utility gains associated with the combined forecasts are sizable and positive and greater than the utility gains using the individual methods. The key finding is that our nonparametric predictive models outperform the historical average with statistically and economically meaningful margins for the out-of-sample periods.

## 6. CONCLUSION

The predictability of equity returns has been a long-standing problem in finance over decades. In this paper, we develop a reliable and powerful nonparametric predictability test to examine whether there exists the predictability of equity returns at short and long horizons. The prevailing predictive variables, such as log dividend yields, log earnings yields, dividend payout ratio, short rate, inflation, book-to-market ratio, investment to capital ratio, corporate issuing activity, and consumption, wealth, and income ratio, have predictability power for equity returns at both short and long horizons. The popular linear predictive regression models cannot fully capture the predictability due to the neglected nonlinear predictable components. We also compare the in-sample and out-of-sample forecast performance of the conventional predictive regression models with the historical mean model. We find that the prevailing predictive model outperforms the historical mean model because it yields a smaller neglected signal-to-noise ratio based on our test, which is different from the conclusion of Goyal and Welch (2007).

The poor out-of-sample performance of most linear predictive models is due to the limitation of linear models. We propose two nonparametric estimators to forecast the equity returns. Our nonparametric predictive models have lower RMSE than the historical mean model at both short-horizon and long-horizon. Our nonparametric prediction can improve the out-of-sample performance without restrictions. Using our nonparametric methods, both combined and individual forecast outperform the historical average statistically and economically. The combined forecast methods outperform the individual forecast methods for the out-of-sample periods.

Although motivated by investigating the predictability of equity returns, our econometric test is applicable to examine predictability of other asset returns over the different time horizons. For example, there are subjects of futher research such as bond returns and inflation rates.

## REFERENCES


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

Andrew Ang, and Jun Liu, 2007, Risk, return, and dividends, Journal of Financial Economics, 85, 1- 38

Baker, M., and J. Wurgler. 2000. The Equity Share in New Issues and Aggregate Stock Returns, Journal of Finance 55:2219–57.

Binsbergen, J. H. van, and R. S. J. Koijen. 2010, Predictive regressions: A present-value approach, Journal of Finance 65: 1439-1471

Binsbergen, J. H. van, and R. S. J. Koijen. 2011. Likelihood-Based Estimation of Exactly-Solved Present Value Models, Working Paper, Stanford University

Campbell, J. Y., 1987, Stock returns and the Term Structure, Journal of Financial Economics, 18(2), 373- 399.

Campbell, J. Y., 1991, A Variance Decomposition for Stock Returns, Economic Journal, 101, 157-179

Campbell, J. Y., and R. J. Shiller, 1987, Cointegration and Tests of Present Value Models, Journal of Political Economy, 95, 1067-1087.

Campbell, J. Y., and R. J. Shiller, 1988a, Stock Prices, Earnings, and Expected Dividends, Journal of Finance, 43(3), 661- 676.

Campbell, J. Y., and R. J. Shiller,1988b, The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors, Review of Financial Studies, 1(3), 195- 227.

Campbell, John Y., and S. Thompson, 2007, Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average?, Forthcoming, Review of Financial Studies.

Campbell, J. Y., and M. Yogo, 2006, Efficient Tests of Stock Return Predictability, Journal of Financial Economics, 81(1), 27- 60.

Cavanagh, C. L., G. Elliott and J. H. Stock (1995), Inference in models with nearly integrated regressors. Econometric Theory 11, 1131–1147.

Cochrane, J.H., 1999a, New Facts in Finance, Economic Perspectives 23, 36-58.

Cochrane, J.H., 1999b. Portfolio Advice for a Multifactor World, Economic Perspectives 23, 59-78.

Cochrane, J. H., 1992, Explaining the Variance of Price-Dividend Ratios, Review of Financial Studies, 5, 243- 280.

Cochrane, John H., 2007, Financial Markets and the Real Economy, in John H. Cochrane, ed., Financial Markets and the Real Economy, Volume 18 of the International Library of Critical Writings in Financial Economics, London: Edward Elgar, p. xi-lxix.

Cochrane, J.H., 2008, The Dog That Did Not Bark: A Defense of Return Predictability, Review of Financial Studies, 21, 1533–1575.

Dickey, D., and Fuller, W. , 1979, Distribution of estimators for autoregressive time series with a unit root. Journal of American Statistical Association, 74, 427-31.

Fama, E., 1990, Stock returns, expected returns, and real activity. Journal of Finance, 45 (September), 1089-1108.

Fama, E., and F. French, 1988, Dividend Yields and Expected Stock Returns, Journal of Financial Economics, 22, 3- 26.

Fama, Eugene F. and Kenneth R. French, 1989, Business Conditions and Expected Returns on Stocks and Bonds, Journal of Financial Economics, 25, 23- 49.

Ferreira, M.A., Santa-Clara, P., 2011, Forecasting stock market returns: The sum of the parts is more than the whole. Journal of Financial Economics, forthcoming.

Ferson,Wayne E. and Merrick John Jr., 1987, Non-stationarity and stage-of-the-businesscycle effects in consumption-based asset pricing relations, Journal of Financial Economics, 18(1), 127-146.

Ferson, W. E., S. Sarkissian, and T. T. Simin. 2003. Spurious Regressions in Financial Economics? Journal of Finance 58:1393–413.

Froot, K., and M. Obstfeld, 1991, Instrinsic Bubbles: The Case of Stock Prices, American Economic Review, 81, 1189-1217

Foster, F. D., T. Smith, and R. E. Whaley. 1997. Assessing Goodness-of-Fit of Asset Pricing Models: The Distribution of the Maximal R2. Journal of Finance 52:591–607.

Goetzmann, W. N., and P. Jorion, 1993, Testing the Predictive Power of Dividend Yields, Journal of Finance, 48(2), 663- 679.

Goetzmann, W. N., and P. Jorion, 1995, A Longer Look at Dividend Yields, Journal of Business, 68, 483- 508.

Golez, B. 2014. Expected returns and dividend growth rates implied by derivative markets, Review of Financial Studies 27: 790-822.

Goyal, A., and I.Welch, 2003, The Myth of Predictability: Does the Dividend Yield Forecast the Equity Premium?, Management Science, 49(5), 639- 654.

Goyal, Amit and Ivo Welch, 2007, A Comprehensive Look at the Empirical Performance of Equity Premium Prediction, Forthcoming, Review of Financial Studies.

Henkel, S. J., J. S. Martin, and F. Nardari, 2011. Time-varying short-horizon predictability, Journal of Financial Economics 99: 560-580

Hodrick, R. J., 1992, Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement, Review of Financial Studies, 5(3), 257- 286.

Hong Y. and Lee T., 2003, Inference on via Genaralized Spectrum and Nonlinear Time Series

Models, The Review of Economics and Statistics, November 2003, 85(4), 1048-1062
Hong.Y. and Lee. Y., 2013, A Loss Function Approach to Model Specification Testing and Its Relative Efficiency to the GLR Test, Annals of Statistics, 41(3), 1166-1203.
Kothari, S., and J. Shanken, 1997, Book-to-market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis, Journal of Financial Economics, 44(2), 169- 203.
Lewellen, J., 2004, Predicting Returns with Financial Ratios, Journal of Financial Economics, 74, 209- 235.
Lamont, O., 1998, Earnings and Expected Returns, Journal of Finance, 53(5), 1563- 1587.
Lettau, Martin, and Sydney Ludvigson, 2005, Measuring and Modeling Variation in the Risk-Return Tradeoff, Forthcoming in the Handbook of Financial Econometrics, edited by Yacine Ait-Shalia and Lars-Peter Hansen.
Lettau, M. and S. Van Nieuwerburgh, 2008, Reconciling the Return Predictability Evidence, Review of Financial Studies, 21, 1607–1652.
Mankiw, N.G. and M. Shapiro, 1986, Do we reject too often? Small sample properties of tests of rational expectations models. Economics Letters 20, 139-145.
Maynard, A, Shimotsu, K. and Wang, 2011, Inference in predictive quantile regressions, working paper
Neely, C. J., Rapach, D. E., Tu J., and Zhou, 2011, Forecasting the Equity Risk Premium: The Role of Technical Indicators, working paper.
Nelson, C., and Kim, M., 1993, Predictable stock returns: Reality or statistical illusion?, Journal of Finance , 48(June), 641-61.
Polk, C., S. Thompson, and T. Vuolteenaho. 2006. Cross-Sectional Forecasts of the Equity Premium. Journal of Financial Economics 81:101–41.
Ponti , J., and L. D. Schall, 1998, Book-to-Market Ratios as Predictors of Market Returns, Journal of Financial Economics, 49(2), 141-160.
Rapach, D. E., and M. E. Wohar, 2006, In-Sample vs. Out-of-Sample Tests of Stock Return Predictability in the Context of Data Mining, Journal of Empirical Finance, 13(2), 231–247.
Rapach, Strauss, and Zhou, 2010, Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy, Review of Financial Studies, 23, 821–862.
Rozeff, M. ,1984, Dividend yields are equity risk premiums. Journal of Portfolio Management, 11 (Fall), 68-75.
Stambaugh, R. ,1986, Biases in regressions with lagged stochastic regressors. Working Paper no. 156. Chicago: University of Chicago, Graduate School of Business.
Stambaugh, R. F., 1999, Predictive Regressions, Journal of Financial Economics, 54, 375- 421.
Stock, J.H. and M.W. Watson, 1999, Forecasting Inflation, Journal of Monetary Economics, 44, 293–335.

Stock, J.H. and M.W.Watson, 2003, Forecasting Output and Inflation: The Role of Asset Prices, Journal of Economic Literature, 41, 788–829.

Stock, J.H. and M.W.Watson, 2004, Combination Forecasts of Output Growth in a Seven-Country Data Set, Journal of Forecasting, 23, 405–430.

Stock, J.H. and M.W.Watson, 2006, Forecasting with Many Predictors, In G. Elliott, C.W.J. Granger, and A. Timmermann (Eds.), Handbook of Economic Forecasting, Amsterdam: Elsevier, pp. 515–554.

Tae-Hwy Lee, Yundong Tu, and Aman Ullah, 2014, Forecasting Equity Premium: Global Historical Average versus Local Historical Average and Constraints, Working paper

Torous, W., and R. Valkanov, 2000, Boundaries of Predictability: Noisy Predictive Regressions, Working Paper, UCLA.

Valkanov, R., 2003, Long-Horizon Regressions: Theoretical Results and Applications, Journal of Financial Economics, 68(2), 201- 232.

West, K., 1988, Dividend Innovations and Stock Price Volatility, Econometrica, 56, 37-61

Wolf, M., 2000, Stock Returns And Dividend Yields Revisited: A New Way To Look At An Old Problem, Journal of Business and Economic Statistics, 18(1), 18- 30.

# APPENDIX

**Proof of Theorem 1:** To prove theorem $1(i)$, we impose the following assumptions:

**Assumption A.1**: $\{Y_{t+h}, X_t\}$ is a stationary time series process with mixing condition. The marginal density function $g(x)$ of $X_t$ is twice continuous differentiable with bounded second derivatives and $g(x)$ is strictly positive over the support of weighting function $w(\cdot)$ given in Assumption A.5. The dimension of $X_t$ is $d$.

**Assumption A.2**: $\varepsilon_{t+h}$ is a $h$-dependent process and $\varepsilon_{t+h}$ is independent of $X_s, s \leq t.(a)$ $0 < E(\varepsilon_{t+h}^2|X_t) = \sigma_\varepsilon^2$ $a.s.$; $(b) 0 < E(\varepsilon_{t+h}^4) = D$

**Assumption A.3**: $\sqrt{R}(\hat{\beta} - \beta) = O_P(1)$, where $\beta = p\lim \hat{\beta}$.

**Assumption A.4**: The kernel function $k : \mathbb{R} \to [0, 1]$ is a symmetic, and twice continuously differentiable probability density with bounded second derivatives.

**Assumption A.5**: $w(\cdot)$ is a positive continuous function over its support with $\int w(x)dx < \infty$ and $\int w^2(x)dx < \infty$.

**Assumption A.6**: $(i) b = b(n) = n^{-\alpha} \to \infty$,where $\alpha \in (0, 1/d)$ and $n = T - R$. $(ii) n^\lambda/R \to 0$, where $\lambda < \max\{1 - \alpha d, \frac{1}{2}(1 + \alpha d)\}$.

Assumption A.1 and A.2 are regularity conditions on the data generating process (DGP). Given $E(Y_{t+h}^2) < \infty$, there exists a measurable function $r_h(x) = E(\varepsilon_{t+h}|X_t = x)$ which is twice continuously differentiable with bounded second derivatives. Assumption A.3 allows for any in-sample $\sqrt{R}$-consistent estimator for $\beta$, which need not be asymptotically most efficient. Assumption A.4 is a standard regularity condition on kernel function $k(\cdot)$. Assumption A.5 is the regularity condition on the positive weighting function $w(\cdot)$. Assumption A.6 provides conditions on the bandwidth $b$ and the relative speed between $R$ and $n$, the sizes of the estimation sample and the prediction sample, respectively. Moreover, we allow the size of the prediction sample, $n$, to be larger or smaller than or the same as the size of the estimation sample, $R$. This offers a wide scope of applicability of our procedure, particularly when the whole sample $\{Y_t\}_{t=1}^T$ is relatively small.

Under the above regularity conditions, we have the following asymptotic results for the $\hat{\mathbf{Q}}_h$ statistics.

To measure the departure of $\hat{r}_h(x)$ from zero over all $x$, we use the following global measure $\hat{Q}(h) = \frac{1}{n-h} \sum_{t=R+1}^{T-h} \widehat{r}_h^2(X_t)w(X_t)$. Define $\hat{Q}^*(h) = \int \widehat{r}_h^2(x)g(x)w(x)dx$. We first show that $\hat{Q}(h)$ and $\hat{Q}^*(h)$ are asymptotically equivalent under $\mathbb{H}_0$.

**Lemma 1.1:** *Under the conditions of Theorem 1,* $(n-h)\hat{Q}(h) - (n-h)\hat{Q}^*(h) = o_p(b^{-d/2})$ *under* $\mathbb{H}_0$.

**Proof of Lemma 1.1:** Because $\hat{G}(x) - G(x) = O_p(n^{-1/2} (\ln n)^2)$ where $\hat{G}(x)$ is the empirical distribution function of $X_t$, and $(n-h) \int \widehat{r}_h^2(x)w(x)g(x)dx = O_p(b^{-d})$. Here we have made use of the well-known fact that

$$\sup_{x\in\mathbb{G}}\left|\hat{G}(x)-G(x)\right|=O_p(n^{-1/2}(\ln n)^2)$$

(see, e.g., Bentkus, Gotse and Tikhomirov (1997)) under Assumption A.1 and $\int\widehat{r}_h^2(x)w(x)g(x)dx = O_p(n^{-1}b^{-d})$ by Markov's inequality. We have

$$\begin{aligned}(n-h)\hat{Q}(h) &= \sum_{t=R+1}^{T-h}\widehat{r}_h^2(X_t)w(X_t)\\ &= (n-h)\int\widehat{r}_h^2(x)w(x)dG(x)+(n-h)\int\widehat{r}_h^2(x)w(x)d[\hat{G}(x)-G(x)]\\ &= (n-h)\int\widehat{r}_h^2(x)g(x)w(x)dx+O_p(\sup_{x\in\mathbb{G}}|\hat{G}(x)-G(x)|)\\ &= (n-h)\int\widehat{r}_h^2(x)g(x)w(x)dx+O_p(n^{-1/2}(\ln n)^2)O_p(b^{-d})\\ &= (n-h)\int\widehat{r}_h^2(x)g(x)w(x)dx+o_p(b^{-d/2})\end{aligned}$$

given $b\propto n^{-\alpha}$ for $\alpha\in(0,1/d)$.This completes the proof of Lemma 1.1. ■

Next we show that $\hat{Q}^*(h)$ and $\widetilde{Q}(h)$ are asymptotically equivalent under $\mathbb{H}_0$, where $\widetilde{Q}(h) = \int\widetilde{r}_h^2(x)g(x)w(x)dx$ and $\widetilde{r}_h^2(x)$ is defined in the same way as $\widehat{r}_h^2(x)$, with $\varepsilon_s$ replacing $\widehat{\varepsilon}_s$.

**Lemma 1.2:** *Under the conditions of Theorem 1,* $(n-h)\hat{Q}^*(h)-(n-h)\widetilde{Q}(h) = (n-h)\int\widehat{r}_h^2(x)g(x)w(x)dx-(n-h)\int\widetilde{r}_h^2(x)g(x)w(x)dx = o_p(b^{-d/2})$*,under* $\mathbb{H}_0$.

**Proof of Lemma 1.2:** We decompose

$$\begin{aligned}&(n-h)\int\widehat{r}_h^2(x)w(x)dG(x)-(n-h)\int\widetilde{r}_h^2(x)w(x)dG(x)\\ =\;&(n-h)\int\left[\widehat{r}_h(x)-\widetilde{r}_h(x)\right]^2w(x)dG(x)+2(n-h)\int\widetilde{r}_h(x)\left[\widehat{r}_h(x)-\widetilde{r}_h(x)\right]w(x)dG(x)\\ =\;&\hat{J}_1+2\hat{J}_2,\ \text{say.}\end{aligned}\qquad\text{(A.1)}$$

For $\hat{J}_1$, we further decompose

$$
\begin{aligned}
\hat{J}_1 &= (n-h)\int \left[\frac{(n-h)^{-1}\sum_{s=R+1}^{T-h}(\hat{\varepsilon}_{s+h}-\varepsilon_{s+h})\mathbf{K}_b(x-X_s)}{(n-h)^{-1}\sum_{s=R+1}^{T-h}\mathbf{K}_b(x-X_s)}\right]^2 dG(x) \\
&= (n-h)\int \left[\frac{(n-h)^{-1}\sum_{s=R+1}^{T-h}(\hat{\varepsilon}_{s+h}-\varepsilon_{s+h})\mathbf{K}_b(x-X_s)}{\hat{g}(x)}\right]^2 dG(x) \\
&= (n-h)\int \frac{\left[(n-h)^{-1}\sum_{s=R+1}^{T-h}(\hat{\varepsilon}_{s+h}-\varepsilon_{s+h})\mathbf{K}_b(x-X_s)\right]^2}{g(x)}dx \\
&\quad +(n-h)\int \left[(n-h)^{-1}\sum_{s=R+1}^{T-h}(\hat{\varepsilon}_{s+h}-\varepsilon_{s+h})\mathbf{K}_b(x-X_s)\right]^2 \left[\frac{1}{\hat{g}^2(x)}-\frac{1}{g^2(x)}\right] g(x)dx \\
&= \hat{J}_{11}+\hat{J}_{12}, \text{ say.} \qquad\qquad (A.2)
\end{aligned}
$$

It suffices to consider the first term $\hat{J}_{11}$ in (A.2), since the second term $\hat{J}_{12}$ is a smaller order given $\sup_{x\in\mathbb{G}}|\hat{g}(x)-g(x)| \xrightarrow{p} 0$.Under the null hypothesis $E(\varepsilon_{t+h}|X_t)=0$, noting that $\hat{\varepsilon}_{s+h}-\varepsilon_{s+h}=Y_{t+h}-X_t'\widehat{\beta}-(Y_{t+h}-X_t'\beta)=X_t'(\beta-\widehat{\beta})$, we have

$$
\begin{aligned}
\hat{J}_{11} &= (n-h)\int \frac{\left[(n-h)^{-1}\sum_{s=R+1}^{T-h}X_t(\beta-\widehat{\beta})\mathbf{K}_b(x-X_s)\right]^2}{g(x)}dx \\
&\leq ||\widehat{\beta}-\beta||^2(n-h)\int \frac{\left[(n-h)^{-1}\sum_{s=R+1}^{T-h}||X_t||^2\mathbf{K}_b(x-X_s)\right]^2}{g(x)}dx \\
&= ||\widehat{\beta}-\beta||^2(n-h)=O_p(R^{-1/2})^2 n \qquad\qquad (A.3)
\end{aligned}
$$

And $\left|\hat{J}_{12}\right| \leq \hat{J}_{11}\left|g^2(x)/\hat{g}^2(x)-1\right| = O_p(\hat{J}_{11}\cdot Sup_{x\in G}|\widehat{g}(x)-g(x)|)=o_p(\hat{J}_{11})=o_p(n/R)$ given $Sup_{x\in G}|\widehat{g}(x)-g(x)|=op(1)$.

Next, we consider $\hat{J}_2$ in (A.1). By the second order Taylor series expansion, we have

$$
\begin{aligned}
\hat{J}_2 &= (n-h)\int \widetilde{r}_h(x)\left[\widehat{r}_h(x)-\widetilde{r}_h(x)\right]w(x)dG(x) \\
&= (n-h)\int \widetilde{r}_h(x)\frac{(n-h)^{-1}\sum_{s=R+1}^{T-h}(\hat{\varepsilon}_{s+h}-\varepsilon_{s+h})\mathbf{K}_b(x-X_s)}{\hat{g}(x)}w(x)dG(x) \\
&= (\widehat{\beta}-\beta)'(n-h)\int \widetilde{r}_h(x)\frac{(n-h)^{-1}\sum_{s=R+1}^{T-h}X_{s+h}\mathbf{K}_b(x-X_s)}{\hat{g}(x)}w(x)dG(x) \\
&= O_p(R^{-1/2})nO_p(n^{-1/2})=O_p(R^{-1/2}n^{1/2}) \qquad\qquad (A.4)
\end{aligned}
$$

This, together with (A.1) and (A.4), yields the desired result. The proof of Lemma 1.2 is completed. ■

**Lemma 1.3:** *Under the conditions of Theorem 1,* $\widehat{\sigma}_\epsilon^2=\sigma_\epsilon^2+O_p(n^{-1/2})$ *under* $\mathbb{H}_0$.

**Proof of Lemma 1.3:** Since $\hat{\sigma}_\varepsilon^2 = (n-h)^{-1}\sum_{s=R+1}^{T-h}\hat{e}_{t+h}^2$, and $\hat{e}_{t+h} = \hat{\varepsilon}_{t+h} - \hat{r}_h(X_t)$, and the same definition for $\widetilde{\sigma}_\epsilon^2 = (n-h)^{-1}\sum_{t=R+1}^{T-h}\tilde{e}_{t+h}^2$ and $\tilde{e}_{t+h} = \varepsilon_{t+h} - \tilde{r}_h(X_t)$. Put $\hat{A}_t = \hat{\varepsilon}_{t+h} - \varepsilon_{t+h}$ $= X_t(\beta - \widehat{\beta})$ and $\hat{B}_t = \hat{r}_h(X_t) - \tilde{r}_h(X_t)$, where $\tilde{r}_h(X_t)$ is defined in the same way as $\hat{r}_h(X_t)$ with $\varepsilon_{t+h}$ replacing $\hat{\varepsilon}_{t+h}$. Then we have

$$\begin{aligned}
\hat{\sigma}_\epsilon^2 &= (n-h)^{-1}\sum_{t=R+1}^{T-h}\left[\hat{\varepsilon}_{t+h} - \hat{r}_h(X_t)\right]^2 \\
&= (n-h)^{-1}\sum_{t=R+1}^{T-h}\left\{(\hat{\varepsilon}_{t+h} - \varepsilon_{t+h}) - [\hat{r}_h(X_t) - \tilde{r}_h(X_t)] + [\varepsilon_{t+h} - \tilde{r}_h(X_t)]\right\}^2
\end{aligned}$$

and we can write

$$\hat{\sigma}_\epsilon^2 = \tilde{\sigma}_\epsilon^2 + (n-h)^{-1}\sum_{t=R+1}^{T-h}(\hat{A}_t - \hat{B}_t)^2 + 2(n-h)^{-1}\sum_{t=R+1}^{T-h}[\varepsilon_{t+h} - \tilde{r}_h(X_t)](\hat{A}_t - \hat{B}_t), \tag{A.5}$$

For the second term in (A.5), we first put

$$m_{st} = \frac{\mathbf{K}_b(x - X_s)}{\sum_{s=R+1}^{T-h}\mathbf{K}_h(x - X_s)}.$$

Then $\sum_{s=R+1}^{T-h} m_{st} = 1$ for all $s$ and $\hat{B}_t = \sum_{s=R+1}^{T-h} m_{st}\hat{A}_s$. Under $\mathbb{H}_0$, we have

$$\hat{A}_t = X_t'(\beta - \widehat{\beta}) \tag{A.6}$$

It follows that

$$\begin{aligned}
(n-h)^{-1}\sum_{t=R+1}^{T-h}(\hat{A}_t - \hat{B}_t)^2 &= (n-h)^{-1}\sum_{t=R+1}^{T-h}\left(\hat{A}_t - \sum_{s=1}^{n} m_{st}\hat{A}_s\right)^2 \\
&\leq 4(n-h)^{-1}\sum_{t=R+1}^{T-h}\hat{A}_t^2 \\
&\leq 4(n-h)^{-1}\sum_{t=1}^{n}\left[X_t'(\beta - \widehat{\beta})\right]^2 \\
&\leq 4\|\widehat{\beta} - \beta\|^2(n-h)^{-1}\sum_{t=1}^{n}\|X_t\|^2 = O_p(R^{-1})
\end{aligned} \tag{A.7}$$

where the first term is $O_p(R^{-1})$ by the mean-value theorem, and Assumptions A.2 and A.3.

For the third term in (A.5), we have

$$
\begin{aligned}
& (n-h)^{-1} \sum_{t=R+1}^{T-h} [\varepsilon_{t+h} - \tilde{r}_h(X_t)](\hat{A}_t - \hat{B}_t) \\
= \; & (n-h)^{-1} \sum_{t=R+1}^{T-h} \varepsilon_{t+h}(\hat{A}_t - \hat{B}_t) - (n-h)^{-1} \sum_{t=R+1}^{T-h} \tilde{r}_h(X_t)(\hat{A}_t - \hat{B}_t) \\
= \; & \hat{T}_1 - \hat{T}_2. \qquad (A.8)
\end{aligned}
$$

For the $\hat{T}_2$ term, by the Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
|\hat{T}_2| & \leq \left[(n-h)^{-1} \sum_{t=R+1}^{T-h} \tilde{r}_h^2(X_t)\right]^{\frac{1}{2}} \left[(n-h)^{-1} \sum_{t=R+1}^{T-h} (\hat{A}_t - \hat{B}_t)^2\right]^{\frac{1}{2}} \\
& = O_p(n^{-1/2} b^{-d/2}) O_p(R^{-1/2}) \\
& = O_p(n^{-1/2} b^{-d/2} R^{-1/2}) = o_p(n^{-1/2}), \qquad (A.9)
\end{aligned}
$$

given $Rb^d \to \infty$ and (A.13), where $(n-h)^{-1} \sum_{t=R+1}^{T-h} \tilde{r}_h^2(X_t) = O_p(n^{-1}b^{-d})$ by Markov's inequality, $E(\varepsilon_{t+h}|X_t) = 0$ *a.s.* and Assumption A.1.

For the $\hat{T}_1$ term, we decompose

$$
\begin{aligned}
\hat{T}_1 & = (n-h)^{-1} \sum_{t=R+1}^{T-h} \varepsilon_{t+h}(\hat{A}_t - \hat{B}_t) \\
& = (n-h)^{-1} \sum_{t=R+1}^{T-h} \varepsilon_{t+h}\hat{A}_t - (n-h)^{-1} \sum_{t=R+1}^{T-h} \varepsilon_{t+h}\hat{B}_t \\
& = \hat{T}_{11} - \hat{T}_{12}, \text{ say.} \qquad (A.10)
\end{aligned}
$$

Here, using (A.6), we have

$$
\hat{T}_{11} = (n-h)^{-1} \sum_{t=R+1}^{T-h} \varepsilon_{t+h} X_t'(\beta - \widehat{\beta}) = O_p(n^{-1/2} R^{-1/2}) \qquad (A.11)
$$

where the first term is $O_p(n^{-1/2}R^{-1/2})$ by a second order Taylor series expansion, Chebyshev's inequality, the Cauchy-Schwarz inequality, and Assumptions A.2 and A.3; the second term is 0.

For the $\hat{T}_{12}$ term in (A.10), recalling $\hat{B}_t = \sum_{s=R+1}^{T-h} m_{st}\hat{A}_s$ and using (A.2), we have

$$\begin{aligned}
\hat{T}_{12} &= (n-h)^{-1}\sum_{t=R+1}^{T-h}\varepsilon_{t+h}\left[(n-h)^{-1}\sum_{s=R+1}^{T-h} m_{st}\hat{A}_s\right] \\
&= (n-h)^{-1}\sum_{t=R+1}^{T-h}\varepsilon_{t+h}\left\{(n-h)^{-1}\sum_{s=R+1}^{T-h} m_{st}\left[X_t'(\beta-\widehat{\beta})\right]\right\} \\
&= O_p(n^{-1/2}R^{-1/2}) \qquad \text{(A.12)}
\end{aligned}$$

Finally, for the first term $\tilde{\sigma}_\epsilon^2$ in (A.5), we have

$$\begin{aligned}
\tilde{\sigma}_\epsilon^2 &= (n-h)^{-1}\sum_{t=R+1}^{T-h}\varepsilon_{t+h}^2 - (n-h)^{-1}2\sum_{t=R+1}^{T-h}\varepsilon_{t+h}\tilde{r}_h(X_t) + (n-h)^{-1}\sum_{t=R+1}^{T-h}\tilde{r}_h^2(X_t) \\
&= [\sigma_\epsilon^2 + O_p(n^{-1/2})] + O_p(n^{-1}b^{-d/2}) + O_p(n^{-1}b^{-d}) \\
&= \sigma_\epsilon^2 + O_p(n^{-1/2}),
\end{aligned}$$

given $\sum_{t=R+1}^{T-h}\tilde{r}_h^2(X_t) = O_p(n^{-1}b^{-d})$ by Markov's inequality and $\sum_{t=R+1}^{T-h}\varepsilon_{t+h}\tilde{r}_h(X_t) = O_p(n^{-1}b^{-d})$. Collecting (A.5) and (A.7)–(A.12) yields the desired result of this lemma. ■

With Lemma 1.3, we have $(n-h)\widetilde{Q}(h)/\hat{\sigma}_\varepsilon^2 - (n-h)\widetilde{Q}(h)/\sigma_\varepsilon^2 = (n-h)\widetilde{Q}(h)/\sigma_\varepsilon^2(\sigma_\varepsilon^2/\hat{\sigma}_\varepsilon^2 - 1) = O_p(b^{-d})O_p(n^{-1/2}) = o_p(b^{-d/2})$ given $(n-h)\widetilde{Q}(h) = O_p(b^{-d})$ and $nb^d \to \infty$. Therefore, we can focus on $(n-h)\widetilde{Q}(h)/\sigma_\varepsilon^2$.

We need to prove $\hat{\mathbf{Q}}_h = \frac{\sqrt{b^d}(n-h)\widetilde{Q}(h)/\hat{\sigma}_\varepsilon^2 - C/\sqrt{b^d}}{\sqrt{V}} \xrightarrow{d} N(0,1)$.

**Proof of Theorem 1:** The proof follows from Theorem 3.4 (Hong and Lee (2013)) with suitable modifications with $\epsilon_t$ replaced by $\epsilon_{t+h}$. ■

# TABLES AND GRAPHICS

## Table 2.0 Summary of Simulation DGPs and Predictability Check

The data-generating processes are summarized in the table below, where $\{v_{t+h}\}$ and $\{u_t\}$ are mutually independent and $\varepsilon_{t+h} = \sum_{j=1}^{h} \alpha_j v_{t+h-j} + v_{t+h}$, $v_{t+h} \sim i.i.d.N(0,1)$, $u_t \sim i.i.d.N(0,1)$. Examine the case where $h = 1, 4, 12, 20$. Sample size $T = 250, 500, 1000$.

| | DGP | $Y_{t+h}$ | $X_t$ | $\beta_1, \beta_2, \rho$ |
|---|---|---|---|---|
| (1) Linear | $A.0(h)$ | $Y_{t+h} = \alpha_0 + \varepsilon_{t+h}$ | $X_t = \rho X_{t-1} + u_t$ | $(0.1, 0.3, 0.5, 0.7, 0.9)$ |
| | $A.1(h)$ | $Y_{t+h} = \beta_0 + \beta_1 X_t + \varepsilon_{t+h}$ | $X_t = \rho X_{t-1} + u_t$ | $(0.1, 0.3, 0.5, 0.7, 0.9)$ |
| | $A.2(h)$ | $Y_{t+h} = \beta_0 + \beta_1 X_t^2 + \varepsilon_{t+h}$ | $X_t = \rho X_{t-1} + u_t$ | $(0.1, 0.3, 0.5, 0.7, 0.9)$ |
| (2) Nonlinear | $B.0(h)$ | $Y_{t+h} = \beta_0 + \beta_1 X_t + \varepsilon_{t+h}$ | $X_t = \rho X_{t-1} + u_t$ | $(0.1, 0.3, 0.5, 0.7, 0.9)$ |
| | $B.1(h)$ | $Y_{t+h} = \beta_0 + \beta_1 X_t + \beta_2 X_t^2 + \varepsilon_{t+h}$ | $X_t = \rho X_{t-1} + u_t$ | $(0.1, 0.3, 0.5, 0.7, 0.9)$ |

## Table 2.1a Bootstrap Results for Size and Power of Predictability Test

Panel a: DGP $A.0(h)$ follows $Y_{t+h} = \alpha_0 + \varepsilon_{t+h}$. $\varepsilon_{t+h}$ has a MA$(h-1)$ property. $\rho$ is the autocoefficient of $X_t$ and $\varepsilon_{t+h} = \sum_{j=1}^{h} \alpha_j v_{t+h-j} + v_{t+h}$.And $\{v_{t+h}\}$ and $\{u_t\}$ are mutually independent. The table presents the rejection rate of bootstrap results for different $h$, and autocoefficient $\rho$ under the significance level of 10%, 5%, and 1% respectively with sample size$T = 250, 500, 1000$.

Test $H1_0 : E(Y_{t+h}|X_t) = E(Y_{t+h})$, and $h = 1, 4, 12, 20$

| DGP | | | | 10% | | | 5% | | | 1% | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $A.0$ | $h$ | $\beta$ | $\rho$ | $T=250$ | 500 | 1000 | $T=250$ | 500 | 1000 | $T=250$ | 500 | 1000 |
| | 1 | 0 | 0.1 | 0.077 | 0.110 | 0.114 | 0.040 | 0.065 | 0.062 | 0.012 | 0.018 | 0.015 |
| | | | 0.3 | 0.089 | 0.124 | 0.098 | 0.050 | 0.071 | 0.045 | 0.015 | 0.024 | 0.010 |
| | | | 0.5 | 0.084 | 0.110 | 0.102 | 0.040 | 0.066 | 0.047 | 0.010 | 0.025 | 0.011 |
| | | | 0.7 | 0.085 | 0.091 | 0.098 | 0.051 | 0.059 | 0.043 | 0.012 | 0.015 | 0.014 |
| | | | 0.9 | 0.091 | 0.097 | 0.100 | 0.058 | 0.052 | 0.062 | 0.011 | 0.018 | 0.019 |

Panel b: DGP $A.1(h)$ follows $Y_{t+h} = \beta_0 + \beta_1 X_t + \varepsilon_{t+h}$ and DGP $A.2(h)$ follows $Y_{t+h} = \beta_0 + \beta_1 X_t^2 + \varepsilon_{t+h}$. $\varepsilon_{t+h}$ has a MA$(h-1)$ property $\varepsilon_{t+h} = \sum_{j=1}^{h} \alpha_j v_{t+h-j} + v_{t+h}$. $\rho$ is the autocoefficient of $X_t$. And $\{v_{t+h}\}$ and $\{u_t\}$ are mutually independent. The table presents the rejection rate of bootstrap results for different $h$, and autocoefficient $\rho$ under the significance level of 10%, 5%, and 1% respectively with sample size$T = 250, 500, 1000$.

Test $H1_0 : E(Y_{t+h}|X_t) = E(Y_{t+h})$, and $h = 1, 4, 12, 20$

| DGP | | | | 10% | | | 5% | | | 1% | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $A.1$ | $h$ | $\beta$ | $\rho$ | $T=250$ | 500 | 1000 | $T=250$ | 500 | 1000 | $T=250$ | 500 | 1000 |
| | 1 | 0.1 | 0.1 | 0.107 | 0.183 | 0.310 | 0.062 | 0.121 | 0.218 | 0.010 | 0.034 | 0.083 |
| | | 0.3 | 0.1 | 0.649 | 0.955 | 0.998 | 0.526 | 0.910 | 0.997 | 0.249 | 0.729 | 0.981 |
| | | 0.5 | 0.5 | 0.993 | 1.000 | 1.000 | 0.984 | 1.000 | 1.000 | 0.913 | 0.999 | 1.000 |
| | | 0.9 | 0.7 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 4 | 0.1 | 0.1 | 0.104 | 0.175 | 0.282 | 0.054 | 0.099 | 0.193 | 0.007 | 0.028 | 0.060 |
| | | 0.3 | 0.1 | 0.620 | 0.917 | 0.999 | 0.498 | 0.869 | 0.997 | 0.217 | 0.642 | 0.974 |
| | | 0.5 | 0.5 | 0.993 | 1.000 | 1.000 | 0.980 | 1.000 | 1.000 | 0.880 | 0.998 | 1.000 |
| | | 0.9 | 0.5 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 12 | 0.1 | 0.1 | 0.098 | 0.174 | 0.294 | 0.052 | 0.112 | 0.194 | 0.012 | 0.038 | 0.075 |
| | | 0.3 | 0.1 | 0.587 | 0.910 | 1.000 | 0.472 | 0.855 | 0.999 | 0.221 | 0.628 | 0.972 |
| | | 0.5 | 0.5 | 0.987 | 1.000 | 1.000 | 0.971 | 1.000 | 1.000 | 0.863 | 1.000 | 1.000 |
| | | 0.9 | 0.5 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 20 | 0.1 | 0.1 | 0.094 | 0.167 | 0.303 | 0.056 | 0.074 | 0.217 | 0.020 | 0.020 | 0.077 |
| | | 0.3 | 0.1 | 0.559 | 0.889 | 0.998 | 0.424 | 0.808 | 0.994 | 0.194 | 0.572 | 0.961 |
| | | 0.5 | 0.5 | 0.975 | 1.000 | 1.000 | 0.957 | 1.000 | 1.000 | 0.819 | 0.999 | 1.000 |
| | | 0.9 | 0.5 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $A.2$ | 1 | 1 | 0.1 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.985 | 1.000 | 1.000 |
| | | | 0.3 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.989 | 1.000 | 1.000 |
| | | | 0.5 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.986 | 1.000 | 1.000 |
| | | | 0.9 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.991 | 0.999 | 1.000 |

## Table 2.1b. Bootstrap Results for Size and Power of Predictability Test

Panel c: DGP $B.0(h)$ follows $Y_{t+h}=\beta_0+\beta_1 X_t+\varepsilon_{t+h}$. $\varepsilon_{t+h}$ has a MA$(h-1)$ property $\varepsilon_{t+h}=\sum_{j=1}^{h}\alpha_j v_{t+h-j}+v_{t+h}$. $\rho$ is the autocoefficient of $X_t$. And $\{v_{t+h}\}$ and $\{u_t\}$ are mutually independent. The table presents the rejection rate of bootstrap results for different $h$, and autocoefficient $\rho$ under the significance level of 10%, 5%, and 1% respectively with sample size$T=250,500,1000$.

Test $H2_0: E(\varepsilon_{t+h}|X_t)=0$, and $h=1,4,12,20$

| DGP | | | | 10% | | | 5% | | | 1% | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $B.0$ | $h$ | $\beta$ | $\rho$ | $T=250$ | 500 | 1000 | $T=250$ | 500 | 1000 | $T=250$ | 500 | 1000 |
| | 1 | 0.5 | 0.1 | 0.091 | 0.108 | 0.097 | 0.051 | 0.060 | 0.055 | 0.014 | 0.015 | 0.011 |
| | | | 0.3 | 0.099 | 0.117 | 0.089 | 0.050 | 0.068 | 0.051 | 0.012 | 0.012 | 0.013 |
| | | | 0.5 | 0.101 | 0.110 | 0.100 | 0.046 | 0.064 | 0.058 | 0.014 | 0.013 | 0.011 |
| | | | 0.9 | 0.111 | 0.107 | 0.086 | 0.069 | 0.059 | 0.046 | 0.019 | 0.015 | 0.016 |
| | 4 | 0.5 | 0.1 | 0.107 | 0.102 | 0.089 | 0.054 | 0.039 | 0.045 | 0.012 | 0.006 | 0.012 |
| | | | 0.3 | 0.093 | 0.092 | 0.095 | 0.054 | 0.052 | 0.050 | 0.015 | 0.015 | 0.008 |
| | | | 0.5 | 0.098 | 0.099 | 0.104 | 0.055 | 0.053 | 0.049 | 0.020 | 0.012 | 0.010 |
| | | | 0.9 | 0.113 | 0.122 | 0.093 | 0.059 | 0.056 | 0.048 | 0.017 | 0.016 | 0.010 |
| | 12 | 0.5 | 0.1 | 0.081 | 0.090 | 0.091 | 0.045 | 0.052 | 0.052 | 0.012 | 0.009 | 0.013 |
| | | | 0.3 | 0.082 | 0.092 | 0.101 | 0.041 | 0.046 | 0.058 | 0.012 | 0.010 | 0.015 |
| | | | 0.5 | 0.091 | 0.085 | 0.103 | 0.042 | 0.046 | 0.061 | 0.011 | 0.014 | 0.013 |
| | | | 0.9 | 0.088 | 0.108 | 0.088 | 0.040 | 0.062 | 0.049 | 0.009 | 0.017 | 0.014 |
| | 20 | 0.5 | 0.1 | 0.085 | 0.085 | 0.119 | 0.049 | 0.041 | 0.077 | 0.010 | 0.011 | 0.022 |
| | | | 0.5 | 0.082 | 0.092 | 0.097 | 0.041 | 0.052 | 0.049 | 0.009 | 0.017 | 0.013 |
| | | | 0.9 | 0.097 | 0.093 | 0.090 | 0.048 | 0.051 | 0.058 | 0.012 | 0.019 | 0.019 |

Panel d: DGP $B.1(h)$ follows $Y_{t+h}=\beta_0+\beta_1 X_t+\beta_2 X_t^2+\varepsilon_{t+h}$. $\varepsilon_{t+h}$ has a MA$(h-1)$ property $\varepsilon_{t+h}=\sum_{j=1}^{h}\alpha_j v_{t+h-j}+v_{t+h}$. $\rho$ is the autocoefficient of $X_t$. And $\{v_{t+h}\}$ and $\{u_t\}$ are mutually independent. The table presents the rejection rate of bootstrap results for different $h$, and autocoefficient $\rho$ under the significance level of 10%, 5%, and 1% respectively with sample size$T=250,500,1000$.

Test $H2_0: E(\varepsilon_{t+h}|X_t)=0$, and $h=1,4,12,20$

| DGP | | | | | 10% | | | 5% | | | 1% | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $B.1$ | $h$ | $\beta_1$ | $\beta_2$ | $\rho$ | $T=250$ | 500 | 1000 | $T=250$ | 500 | 1000 | $T=250$ | 500 | 1000 |
| | 1 | 0.5 | 0.1 | 0.1 | 0.168 | 0.265 | 0.377 | 0.088 | 0.169 | 0.261 | 0.032 | 0.048 | 0.099 |
| | | 0.5 | 0.3 | 0.5 | 0.874 | 0.996 | 1.000 | 0.808 | 0.989 | 1.000 | 0.585 | 0.930 | 1.000 |
| | | 0.5 | 0.5 | 0.5 | 0.995 | 1.000 | 1.000 | 0.984 | 1.000 | 1.000 | 0.925 | 1.000 | 1.000 |
| | | 0.5 | 0.9 | 0.7 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.993 | 1.000 | 1.000 |
| | 4 | 0.5 | 0.1 | 0.1 | 0.178 | 0.229 | 0.394 | 0.103 | 0.155 | 0.283 | 0.024 | 0.040 | 0.101 |
| | | 0.5 | 0.3 | 0.5 | 0.877 | 0.993 | 1.000 | 0.797 | 0.988 | 1.000 | 0.551 | 0.922 | 0.999 |
| | | 0.5 | 0.5 | 0.5 | 0.997 | 1.000 | 1.000 | 0.991 | 1.000 | 1.000 | 0.921 | 0.999 | 1.000 |
| | | 0.5 | 0.9 | 0.7 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.988 | 1.000 | 1.000 |
| | 12 | 0.5 | 0.1 | 0.1 | 0.134 | 0.245 | 0.349 | 0.075 | 0.138 | 0.233 | 0.017 | 0.037 | 0.097 |
| | | 0.5 | 0.3 | 0.5 | 0.845 | 0.994 | 1.000 | 0.754 | 0.985 | 1.000 | 0.496 | 0.918 | 0.999 |
| | | 0.5 | 0.5 | 0.5 | 0.994 | 1.000 | 1.000 | 0.981 | 1.000 | 1.000 | 0.889 | 0.999 | 1.000 |
| | | 0.5 | 0.9 | 0.5 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.977 | 1.000 | 1.000 |
| | 20 | 0.5 | 0.1 | 0.1 | 0.140 | 0.235 | 0.410 | 0.079 | 0.146 | 0.286 | 0.016 | 0.035 | 0.110 |
| | | 0.5 | 0.3 | 0.5 | 0.829 | 0.994 | 1.000 | 0.739 | 0.987 | 1.000 | 0.460 | 0.920 | 1.000 |
| | | 0.5 | 0.5 | 0.5 | 0.993 | 1.000 | 1.000 | 0.975 | 1.000 | 1.000 | 0.867 | 0.997 | 1.000 |
| | | 0.5 | 0.9 | 0.5 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.973 | 0.999 | 1.000 |

## Table 3.1 Sample statistics

Panel (a) reports summary statistics of the data for S&P 500, all at a quarterly frequency. Panel (b) reports statistics for monthly frequency . Excess returns and short rates are continuously compounded. Sample means and standard deviations (SD) for excess returns, dividend, and earnings growth have been annualized by multiplying by 4(12) and $\sqrt{4}(\sqrt{12})$, respectively, for the case of quarterly (monthly) frequency data. Short rates are three-month T-bill rates. Dividend and earnings yields, and the corresponding dividend and earnings growth are computed using dividends or earnings summed up over the past year. Panel (c) reports statistics for annually frequency. The unit root test is the Phillips and Perron (1988) test for the estimated regression $x_t = \alpha + \rho x_{t-1} + u_t$ under the null $x_t = x_{t-1} + u_t$ . The critical values at the significance level 1% 2.5%, 5%, and 10% are 3:46, 3:14, 2:88, and 2:57, respectively. * and ** are at the significance leve 5% and 1%.

Panel (a) US S&P500 data, March 1936-December 2001 Quarterly

| | Excess Return | Short Rate | Dividend Yield | earnings Yield | Dividend Growth | Earnings Growth |
|---|---|---|---|---|---|---|
| Mean | 0.0803 | 0.0416 | 0.0408 | 0.0769 | 0.0566 | 0.0599 |
| S.D. | 0.3336 | 0.0320 | 0.0157 | 0.0299 | 0.1578 | 0.2303 |
| Auto | 0.092 | 0.9531 | 0.9221 | 0.9516 | 0.4306 | 0.6926 |
| Test statistics | | | | | | |
| $H_0$: unit root | $-13.97^{*}$ | $-2.9290$ | $-4.0337$ | $-1.5765$ | $-9.72^{**}$ | $-6.37^{**}$ |

Panel (b) US S&P500 data, January 1970-December 2006 monthly

| | Excess Return | Short Rate | Dividend Yield | earnings Yield |
|---|---|---|---|---|
| Mean | 0.0610 | 0.0597 | 0.0317 | 0.0694 |
| S.D. | 0.5259 | 0.0288 | 0.0129 | 0.0299 |
| Auto | 0.081 | 0.992 | 0.978 | 0.987 |
| Test statistics | | | | |
| $H_0$: unit root | $-28.34^{*}$ | $-1.28$ | $-1.51$ | $-1.17$ |

Panel (c) US S&P500 data, 1872-2005 annually

| | Excess Return | Short Rate | Dividend Yield | earnings Yield | Dividend Payout Ratio | Inflation 1919–2005 | b/m 1921–2005 | i/k 1947–2005 | ntis 1927–2005 | eqis 1927–2005 | cay 1945–2005 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean | 0.0171 | 0.0479 | 0.0451 | 0.0750 | 0.6213 | 0.0309 | 1.6924 | 0.0836 | 0.0229 | 0.1986 | $-0.0488$ |
| S.D. | 0.170 | 0.0278 | 0.0163 | 0.0273 | 0.1987 | 0.0464 | 0.5338 | 0.0343 | 0.0772 | 0.1099 | 1.4925 |
| Auto | 0.077 | 0.831 | 0.762 | 0.737 | 0.615 | 0.566 | 0.918 | 0.912 | 0.436 | 0.493 | 0.564 |
| Test statistics | | | | | | | | | | | |
| $H_0$: unit root | $-10.53^{**}$ | $-1.05$ | $-1.53$ | $-1.54$ | $-1.67$ | $-4.32^{**}$ | $-0.90$ | $-3.63^{**}$ | $-7.82^{**}$ | $-2.33^{*}$ | $-4.30^{**}$ |

# Table 4.1a. Predictability of US Excess Returns(Quarterly)

We estimate regressions of the form $Y_{t+h}=\alpha_h+\beta_h' X_t+\varepsilon_{h,t+h}$ where $Y_{t+h}=(\tau/h)[(y_{t+1}-r_t)+\cdots+(y_{t+h}-r_{t+h-1})]$ is the annualized $h$-period excess return for the aggregate stock market, $r_t$ is the risk-free rate from $t$ to $t+1$, and $y_{t+1}-r_t$ is the excess one period return from time $t$ to $t+1$, with instruments $z_t$ being log dividend yields or log earnings yields. T-statistics in parentheses are computed using Newey West(1987) and Hodrick (1992) standard errors. Horizons $h$ are quarterly. The test column reports a $p$-value for the predictictability test under two hypothesises $H2$ and $H2$. Hypothesis $H2$ is $H_0$: $E(Y_{t+h}|X_t)=E(Y_{t+h})$ and Hypothesis $H2$ is $H_0$: $E(\varepsilon_{t+h}|X_t)=0$. * and ** are at the significance leve 5% and 1% respectively. The time periods are from 1936 to 2001, from 1952 to 2001, from 1936 to 1990, and from 1952 to 1990.

| | $k$ | Univariate Regression $dy4$ (1)NW | (2)HK | Test (1)H1 | (2)H2 | $ey4$ (1)NW | (2)HK | Test (1)H1 | (2)H2 |
|---|---|---|---|---|---|---|---|---|---|
| 1936 − 2001 | 1 | 0.1441 | 0.1441 | 0.0565 | 0.0758 | 0.0630 | 0.0625 | 0.2526 | 0.0901 |
| | | (2.28)* | (1.824) | (0.000)** | (0.000)** | (0.91) | (0.70) | (0.000)** | (0.000)** |
| | 4 | 0.1618 | 0.1578 | 0.0525 | 0.1649 | 0.0918 | 0.0898 | 0.0379 | 0.0677 |
| | | (2.81)** | (2.030) | (0.000)** | (0.000)** | (1.42) | (1.30) | (0.015)* | (0.000)** |
| | 12 | 0.1522 | 0.1455 | 0.0390 | 0.0454 | 0.0648 | 0.549 | 0.0377 | 0.0537 |
| | | (2.92)** | (1.568) | (0.000)** | (0.000)** | (1.05) | (0..90) | (0.016)* | (0.000)** |
| | 20 | 0.1846 | 0.1028 | 0.0921 | 0.1552 | 0.0484 | 0.0348 | 0.0490 | 0.0549 |
| | | (3.84)** | (1.364) | (0.000)** | (0.000)** | (0.81) | (0.54) | (0.000)** | (0.000)** |
| 1952 − 2001 | 1 | 0.0417 | 0.0489 | 0.1048 | 0.1129 | 0.0467 | 0.0464 | 0.2952 | 0.0984 |
| | | (0.51) | (0.771) | (0.000)** | (0.000)** | (0.59) | (0.46) | (0.000)** | (0.000)** |
| | 4 | 0.0542 | 0.0530 | 0.1558 | 0.0910 | 0.0791 | 0.0798 | 0.1260 | 0.1178 |
| | | (0.69) | (0.773) | (0.000)** | (0.000)** | (1.41) | (1.218) | (0.000)** | (0.000)** |
| | 12 | 0.0091 | 0.0234 | 0.0244 | 0.0454 | 0.0768 | 0.0756 | 0.0683 | 0.0887 |
| | | (0.12) | (0.344) | (0.024)* | (0.000)** | (1.25) | (1.03) | (0.000)** | (0.000)** |
| | 20 | 0.0276 | 0.0287 | 0.0921 | 0.1552 | 0.0283 | 0.0223 | 0.1005 | 0.1065 |
| | | (0.35) | (0.477) | (0.000)** | (0.000)** | (1.80) | (0.320) | (0.000)** | (0.000)** |
| 1936 − 1990 | 1 | 0.2993 | 0.2203 | 0.0315 | 0.0515 | 0.0886 | 0.0896 | 0.3864 | 0.0387 |
| | | (3.19)** | (2.416)* | (0.021)* | (0.000)** | (1.57) | (1.304) | (0.000)** | (0.015)* |
| | 4 | 0.3228 | 0.2383 | 0.1560 | 0.0689 | 0.1417 | 0.1473 | 0.0333 | 0.0361 |
| | | (4.56)** | (3.097)** | (0.000)** | (0.000)** | (1.52) | (2.309)* | (0.019)* | (0.017)* |
| | 12 | 0.2744 | 0.2380 | 0.1069 | 0.1016 | 0.0957 | 0.0945 | 0.0555 | 0.0979 |
| | | (5.08)** | (4.08)** | (0.000)** | (0.000)** | (1.29) | (1.02) | (0.000)** | (0.000)** |
| | 20 | 0.2652 | 0.1787 | 0.1326 | 0.1042 | 0.1225 | 0.1078 | 0.0924 | 0.0882 |
| | | (5.97)** | (2.819)** | (0.000)** | (0.000)** | (1.38) | (1.392) | (0.000)** | (0.000)** |
| 1952 − 1990 | 1 | 0.1622 | 0.1482 | 0.0851 | 0.1006 | 0.1537 | 0.1028 | 0.2496 | 0.0223 |
| | | (1.06) | (2.783)** | (0.000)** | (0.000)** | (0.60) | (1.168) | (0.000)** | (0.022)* |
| | 4 | 0.1921 | 0.3070 | 0.0682 | 0.0639 | 0.1291 | 0.1287 | 0.0809 | 0.0685 |
| | | (1.40) | (2.50)* | (0.000)** | (0.000)** | (1.62) | (1.521) | (0.000)** | (0.000)** |
| | 12 | 0.0988 | 0.1132 | 0.1849 | 0.1318 | 0.0576 | 0.0475 | 0.0732 | 0.1356 |
| | | (0.80) | (1.453) | (0.000)** | (0.000)** | (0.75) | (0.45) | (0.000)** | (0.000)** |
| | 20 | 0.0622 | 0.0845 | 0.2669 | 0.1138 | 0.0771 | 0.0798 | 0.3169 | 0.1878 |
| | | (0.67) | (1.916) | (0.000)** | (0.000)** | (1.17) | (1.133) | (0.000)** | (0.000)** |

# Table 4.1b. Predictability of US Excess Returns(Quarterly)

We estimate regressions of the form $Y_{t+h} = \alpha_h + \beta_h' X_t + \varepsilon_{h,t+h}$ where $Y_{t+h} = (\tau/h)[(y_{t+1} - r_t) + \cdots + (\ y_{t+h} - r_{t+h-1})]$ is the annualized $h$-period excess return for the aggregate stock market, $r_t$ is the risk-free rate from $t$ to $t+1$, and $y_{t+1} - r_t$ is the excess one period return from time $t$ to $t+1$, with instruments $z_t$ being log dividend yields and risk free rate together. T-statistics in parentheses are computed using Newey West(1987) and Hodrick (1992) standard errors. Horizons $h$ are quarterly. The test column reports a $p$-value for the predictictability test under six hypothesises $H1$–$H6$. $X_1$ represents the short rate $r$ and $X_2$ represents the dividend yields. Hypothesis $H1$ ($H_0 : E(Y_{t+h}|X_{1t}) = E(Y_{t+h})$), Hypothesis $H2$ ($H_0 : E(\varepsilon_{t+h}|X_{1t}) = 0$),Hypothesis $H3$ ($H_0 : E(Y_{t+h}|X_{1t}) = E(Y_{t+h})$), Hypothesis $H4$ ($H_0 : E(\varepsilon_{t+h}|X_{2t}) = 0$), Hypothesis $H5$ ($H_0 : E(Y_{t+h}|X_{1t}, X_{2t}) = E(Y_{t+h})$), Hypothesis $H6$ ($H_0 : E(\varepsilon_{t+h}|X_{1t}, X_{2t}) = 0$). * and ** are at the significance leve 5% and 1% respectively. The time periods are from 1936 to 2001, from 1952 to 2001, from 1936 to 1990, and from 1952 to 1990.

| | | BivariateRegression | | | | Test | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $k$ | $r$ (1)NW | (2)HK | $dy4$ (1)NW | (2)HK | (1)H1 | (2)H2 | (3)H3 | (4)H4 | (5)H5 | (6)H6 |
| 1936 − 2001 | 1 | −0.9864 | −1.0888 | 0.0849 | 0.0875 | 0.0409 | 0.0495 | 0.0475 | 0.0699 | 0.0088 | 0.0137 |
| | | (−5.730)** | (−1.608) | (1.530) | (1.530) | (0.010)** | (0.008)** | (0.005)** | (0.000)** | (0.044)* | (0.032)* |
| | 4 | −0.8305 | −0.5596 | 0.1068 | 0.1032 | 0.0694 | 0.0600 | 0.0525 | 0.0761 | 0.0064 | 0.0168 |
| | | (−5.520)** | (−0.827) | (2.200)* | (1.865) | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.052) | (0.028)* |
| | 12 | −0.6993 | −0.4321 | 0.0985 | 0.0980 | 0.0533 | 0.0469 | 0.0701 | 0.0892 | 0.0059 | 0.0072 |
| | | (−8.57)** | (−0.708) | (2.530)* | (1.265) | (0.000)** | (0.006)** | (0.000)** | (0.000)** | (0.056) | (0.046)* |
| | 20 | −0.6606 | −0.6364 | 0.1246 | 0.1152 | 0.0634 | 0.1278 | 0.0913 | 0.2218 | 0.0040 | 0.0043 |
| | | (−8.83)** | (1.196) | (3.57)** | (1.255) | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.062) | (0.064) |
| 1952 − 2001 | 1 | −1.2958 | −2.1623 | 0.1355 | 0.1362 | 0.1147 | 0.1236 | (0.1593) | 0.1151 | 0.0123 | 0.0095 |
| | | (−6.74)** | (−2.912)** | (1.91)* | (2.152)* | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.034)* | (0.040)* |
| | 4 | −1.0484 | −1.4433 | 0.1291 | 0.1313 | 0.1548 | 0.1316 | 0.1558 | 0.0909 | 0.0338 | 0.0472 |
| | | (−5.85)** | (−1.930) | (1.990)* | (1.921) | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.017)* | (0.005)** |
| | 12 | −0.7141 | −0.6272 | 0.0685 | 0.0874 | 0.0712 | 0.0488 | 0.0335 | 0.0823 | 0.0108 | 0.0153 |
| | | (−6.61)** | (−0.987) | (1.17) | (0.680) | (0.000)** | (0.006)** | (0.016)* | (0.000)** | (0.035)* | (0.031)* |
| | 20 | −0.5738 | −0.4829 | 0.0596 | 0.0774 | 0.0534 | 0.0690 | 0.0759 | 0.1468 | 0.0198 | 0.0081 |
| | | (−5.99)** | (−0.745) | (1.00) | (0.600) | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.026)* | (0.048)* |
| 1936 − 1990 | 1 | −0.9709 | −1.0380 | 0.1961 | 0.1917 | 0.0325 | 0.0321 | 0.0510 | 0.0503 | 0.0091 | 0.0090 |
| | | (−5.49)** | (−1.543) | (2.44)* | (2.126)* | (0.021)* | (0.021)* | (0.000)** | (0.000)** | (0.044)* | (0.045)* |
| | 4 | −0.8131 | −0.4865 | 0.2398 | 0.2254 | 0.0386 | 0.0390 | 0.0661 | 0.0698 | 0.0142 | 0.0144 |
| | | (−5.34)** | (−0.714) | (4.21)** | (3.006)** | (0.015)* | (0.012)* | (0.000)** | (0.000)** | (0.034)* | (0.032)* |
| | 12 | −0.7121 | −0.6785 | 0.2114 | 0.1898 | 0.0827 | 0.0603 | 0.1069 | 0.0633 | 0.0126 | 0.0139 |
| | | (−8.26)** | (−0.879) | (5.65)** | (3.214)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.035)* | (0.033)* |
| | 20 | −0.6877 | −0.6458 | 0.2097 | 0.1719 | 0.0791 | 0.0222 | 0.2107 | 0.0580 | 0.0335 | 0.0045 |
| | | (−8.58)** | (−1.289) | (7.43)** | (2.832)** | (0.000)** | (0.024)* | (0.000)** | (0.000)** | (0.017)* | (0.061) |
| 1952 − 1990 | 1 | −1.4347 | −2.7329 | 0.4102 | 0.4125 | 0.0947 | 0.1134 | 0.0851 | 0.0823 | 0.0183 | 0.0224 |
| | | (−6.29)** | (3.504)** | (3.02)** | (3.672)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.029)* | (0.022)* |
| | 4 | −1.1908 | −1.9840 | 0.4030 | 0.3935 | 0.0892 | 0.1049 | 0.0683 | 0.0607 | 0.0273 | 0.0193 |
| | | (−6.26)** | (−2.508)* | (3.78)** | (3.700)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.019)* | (0.025)* |
| | 12 | −0.8431 | −0.7324 | 0.2623 | 0.2457 | 0.1210 | 0.0716 | 0.1849 | 0.1381 | 0.0628 | 0.0361 |
| | | (−9.02)** | (−1.278) | (4.28)** | (2.457)* | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.018)* |
| | 20 | −0.6679 | −0.7120 | 0.2024 | 0.2057 | 0.3003 | 0.0541 | 0.2679 | 0.1058 | 0.1149 | 0.0029 |
| | | (−7.47)** | (−1.087) | (4.09)** | (2.387)* | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.112) |

# Table 4.1c. Predictability of US Excess Returns(Quarterly)

We estimate regressions of the form $Y_{t+h} = \alpha_h + \beta_h' X_t + \varepsilon_{h,t+h}$ where $Y_{t+h} = (\tau/h)[(y_{t+1} - r_t) + \cdots + (\ y_{t+h} - r_{t+h-1})$ is the annualized $h$-period excess return for the aggregate stock market, $r_t$ is the risk-free rate from $t$ to $t+1$, and $y_{t+1} - r_t$ is the excess one period return from time $t$ to $t+1$, with instruments $z_t$ being log dividends yields and log earnings yields together. T-statistics in parentheses are computed using Newey West(1987) and Hodrick (1992) standard errors. Horizons $h$ are quarterly. The test column reports a $p$-value for the predictictability test under six hypothesises $H1$–$H6$. $X_1$ represents the short rate $r$ and $X_2$ represents the earnings yields. Hypothesis $H1$ $(H_0 : E(Y_{t+h}|X_{1t}) = E(Y_{t+h}))$, Hypothesis $H2$ $(H_0 : E(\varepsilon_{t+h}|X_{1t}) = 0)$,Hypothesis $H3$ $(H_0 : E(Y_{t+h}|X_{1t}) = E(Y_{t+h}))$, Hypothesis $H4$ $(H_0 : E(\varepsilon_{t+h}|X_{2t}) = 0)$, Hypothesis $H5$ $(H_0 : E(Y_{t+h}|X_{1t}, X_{2t}) = E(Y_{t+h}))$, Hypothesis $H6$ $(H_0 : E(\varepsilon_{t+h}|X_{1t}, X_{2t}) = 0)$. * and ** are at the significance leve 5% and 1% respectively. The time periods are from 1936 to 2001, from 1952 to 2001, from 1936 to 1990, and from 1952 to 1990.

| | | BivariateRegression | | | | Test | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $k$ | $dy4$ (1)NW | (2)HK | $ey4$ (1)NW | (2)HK | (1)H1 | (2)H2 | (3)H3 | (4)H4 | (5)H5 | (6)H6 |
| 1936 − 2001 | 1 | 0.3891 | 0.1842 | −0.3003 | −0.1019 | 0.0475 | 0.0771 | 0.0781 | 0.0887 | 0.0303 | 0.0482 |
| | | (3.08)** | (1.330) | (−2.27)* | (−0.741) | (0.008)** | (0.000)** | (0.000)** | (0.000)** | (0.018)* | (0.006)** |
| | 4 | 0.3485 | 0.1119 | −0.2289 | 0.0012 | 0.0526 | 0.0781 | 0.0379 | 0.0680 | 0.0209 | 0.0211 |
| | | (3.11)** | (0.981) | (−1.97)* | (0.011) | (0.000)** | (0.000)** | (0.015)* | (0.000)** | (0.025)* | (0.024)* |
| | 12 | 0.3503 | 0.1470 | −0.2263 | −0.1020 | 0.0701 | 0.0400 | 0.0306 | 0.0385 | 0.0148 | 0.0091 |
| | | (3.36)** | (1.40) | (−2.19)* | (−1.39) | (0.000)** | (0.011)* | (0.018)* | (0.014)* | (0.030)* | (0.044)* |
| | 20 | 0.4752 | 0.2111 | −0.3119 | −0.0912 | 0.0913 | 0.0495 | 0.1493 | 0.1495 | 0.0121 | 0.0189 |
| | | (5.05)** | (1.55) | (−3.64)** | (−1.64) | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.035)* | (0.027)* |
| 1952 − 2001 | 1 | 0.3857 | 0.1948 | −0.3953 | −0.1113 | 0.1593 | 0.0945 | 0.0941 | 0.0468 | 0.0898 | 0.0447 |
| | | (2.46)* | (1.557) | (−2.61)** | (−0.873) | (0.000)** | (0.000)** | (0.000)** | (0.009)** | (0.000)** | (0.010)** |
| | 4 | 0.3584 | 0.1615 | −0.3485 | −0.0636 | 0.1139 | 0.1161 | 0.1682 | 0.1255 | 0.0865 | 0.0810 |
| | | (2.57)* | (1.195) | (−2.83)** | (−0.450) | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** |
| | 12 | 0.2843 | 0.1821 | −0.2969 | −0.0832 | 0.0336 | 0.0202 | 0.0458 | 0.0388 | 0.0087 | 0.0071 |
| | | (1.87) | (1.21) | (−2.65)** | (−0.76) | (0.016)* | (0.026)* | (0.009)** | (0.014)* | (0.045)* | (0.046)* |
| | 20 | 0.3567 | 0.1458 | −0.3402 | −0.0765 | 0.0821 | 0.0986 | 0.0851 | 0.1001 | 0.0266 | 0.0525 |
| | | (2.23)* | (0.561) | (−3.32)** | (−0.520) | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.022)* | (0.000)** |
| 1936 − 1990 | 1 | 0.7476 | 0.3848 | −0.4719 | −0.1800 | 0.0510 | 0.0497 | 0.0351 | 0.0388 | 0.0192 | 0.0221 |
| | | (4.23)** | (1.849) | (−2.99)** | (1.062) | (0.000)** | (0.000)** | (0.019)* | (0.014)* | (0.027)* | (0.024)* |
| | 4 | 0.6859 | 0.2869 | −0.3823 | −0.0532 | 0.0662 | 0.0666 | 0.0333 | 0.0381 | 0.0121 | 0.0163 |
| | | (4.58)** | (1.872) | (−2.76)** | (−0.411) | (0.000)** | (0.000)** | (0.020)* | (0.015)* | (0.035)* | (0.031)* |
| | 12 | 0.6320 | 0.2987 | −0.3677 | −0.789 | 0.0706 | 0.0751 | 0.1741 | 0.1994 | 0.0501 | 0.0550 |
| | | (5.26)** | (1.98) | (−3.03)** | (−0.93) | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** |
| | 20 | 0.6521 | 0.1916 | −0.4005 | −0.0139 | 0.1486 | 0.1245 | 0.0901 | 0.0834 | 0.1018 | 0.0824 |
| | | (6.77)** | (1.874) | (−4.25)** | (−0.181) | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** |
| 1952 − 1990 | 1 | 1.3055 | 0.9349 | −0.94 | −0.5240 | 0.0851 | 0.0907 | 0.0271 | 0.0219 | 0.0112 | 0.0089 |
| | | (4.05)** | (3.71)** | (−3.90)** | (−2.618)** | (0.000)** | (0.000)** | (0.021)* | (0.024)* | (0.036)* | (0.045)* |
| | 4 | 1.2350 | 0.8210 | −0.8581 | −0.4212 | 0.0683 | 0.0592 | 0.0809 | 0.0564 | 0.0427 | 0.0329 |
| | | (4.35)** | (3.234)** | (−4.21)** | (−2.080)* | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.009)** | (0.016)* |
| | 12 | 1.0544 | 0.7824 | −0.7689 | −0.7320 | 0.0098 | 0.0648 | 0.0910 | 0.1074 | 0.0041 | 0.0311 |
| | | (6.89)** | (2.98)** | (−6.38)** | (−3.54)** | (0.043)* | (0.000)** | (0.000)** | (0.000)** | (0.061) | (0.017)* |
| | 20 | 0.8561 | 0.4167 | −0.6423 | −0.1999 | (0.0212) | 0.0511 | 0.0458 | 0.0875 | 0.0824 | 0.0551 |
| | | (8.14)** | (2.685)** | (−7.43)** | (−1.570) | (0.025)* | (0.000)** | (0.006)** | (0.000)** | (0.000)** | (0.000)** |

# Table 4.1d. Predictability of US Excess Returns(Quarterly)

(a)We estimate regressions of the form $Y_{t+h} = \alpha_h + \beta_h' X_t + \varepsilon_{h,t+h}$ where $Y_{t+h} = (\tau/h)[(y_{t+1} - r_t) + \cdots + (\ y_{t+h} - r_{t+h-1})$ is the annualized $h$-period excess return for the aggregate stock market, $r_t$ is the risk-free rate from $t$ to $t+1$, and $y_{t+1} - r_t$ is the excess one period return from time $t$ to $t+1$, with instruments $z_t$ being short rate, log dividends yields, and log earnings yields together. T-statistics in parentheses are computed using Newey West(1987) and Hodrick (1992) standard errors. Horizons $h$ are quarterly. The time periods are from 1936 to 2001, from 1952 to 2001, from 1936 to 1990, and from 1952 to 1990.

(b)The test column reports a $p$-value for the predictictability test under six hypothesises $H1 - H8$. $X_1$ represents the short rate $r$ , $X_2$ represents the dividends yields and $X_3$ represents the earnings yields. Hypothesis $H1$ ($H_0 : E(Y_{t+h}|X_{1t}) = E(Y_{t+h})$), Hypothesis $H2$ ($H_0 : E(\varepsilon_{t+h}|X_{1t}) = 0$),Hypothesis $H3$ ($H_0 : E(Y_{t+h}|X_{1t}) = E(Y_{t+h})$), Hypothesis $H4$ ($H_0 : E(\varepsilon_{t+h}|X_{2t}) = 0$), Hypothesis $H5$ ($H_0 : E(Y_{t+h}|X_{3t}) = E(Y_{t+h})$), Hypothesis $H6$ ($H_0 : E(\varepsilon_{t+h}|X_{3t}) = 0$),Hypothesis $H7$ ($H_0 : E(Y_{t+h}|X_{1t}, X_{2t}, X_{3t}) = E(Y_{t+h})$), Hypothesis $H8$ ($H_0 : E(\varepsilon_{t+h}|X_{1t}, X_{2t}, X_{3t}) = 0$). * and ** are at the significance leve 5% and 1% respectively.

| | $k$ | TrivariateRegression $r$ (1)NW | (2)HK | $dy4$ (1)NW | (2)HK | $ey4$ (1)NW | (2)HK |
|---|---|---|---|---|---|---|---|
| 1936 − 2001 | 1 | −1.0703 | −1.0334 | 0.1018 | 0.1000 | −0.0132 | −0.0168 |
| | | (−5.40)** | (−1.296) | (0.80) | (0.605) | (−0.090) | (−0.103) |
| | 4 | −0.9325 | −0.7759 | 0.0094 | 0.0472 | 0.1237 | 0.0654 |
| | | (−5.65)** | (−1.084) | (0.31) | (0.377) | (1.26) | (0.556) |
| | 12 | −0.7717 | −0.6836 | 0.0175 | 0.0125 | 0.0862 | 0.0632 |
| | | (−8.49)** | (−1.923) | (0.20) | (0.042) | (1.10) | (0.70) |
| | 20 | −0.6231 | −0.4190 | 0.1685 | 0.0666 | −0.0435 | 0.0291 |
| | | (−6.63)** | (−0.682) | (1.79) | (0.486) | (−0.52) | (0.361) |
| 1952 − 2001 | 1 | −1.4102 | −2.6243 | 0.0327 | 0.0337 | 0.1276 | 0.1272 |
| | | (−6.04)** | (−2.85)** | (0.21) | (0.239) | (0.80) | (0.815) |
| | 4 | −1.1074 | −1.8146 | 0.0755 | 0.0479 | 0.0662 | 0.0629 |
| | | (−5.01)** | (−2.063)* | (0.56) | (0.687) | (0.52) | (0.487) |
| | 12 | −0.6863 | −1.0213 | 0.0988 | 0.0453 | −0.0337 | −0.0134 |
| | | (−5.43)** | (−2.34)* | (0.81) | (0.43) | (−0.31) | (−0.23) |
| | 20 | −0.4535 | −0.3548 | 0.2209 | 0.1199 | −0.1591 | −0.0418 |
| | | (−3.62)** | (−0.489) | (1.54) | (0.432) | (−1.36) | (−0.249) |
| 1936 − 1990 | 1 | −0.9918 | −0.6544 | 0.3706 | 0.3051 | 0.0245 | 0.0125 |
| | | (−4.64)** | (−0.696) | (1.84) | (1.112) | (0.15) | (0.099) |
| | 4 | −0.852 | −0.4708 | 0.1932 | 0.2300 | 0.0449 | −0.0046 |
| | | (−4.98)** | (−0.031) | (1.33) | (1.336) | (0.38) | (−0.031) |
| | 12 | −0.6939 | −0.4873 | 0.2341 | 0.2032 | −0.0217 | −0.0312 |
| | | (−8.11)** | (−6.46)** | (2.33) | (1.90) | (−0.24) | (−0.32) |
| | 20 | −0.6067 | −0.4239 | 0.3085 | 0.1423 | −0.0955 | 0.0297 |
| | | (−6.43)** | (−0.714) | (3.19)** | (1.597) | (−1.10) | (0.431) |
| 1952 − 1990 | 1 | −1.2799 | −2.0892 | 0.6690 | 0.6598 | −0.2348 | −0.2418 |
| | | (−5.13)** | (−2.096)* | (1.84) | (1.78) | (−0.87) | (−0.949) |
| | 4 | −0.9802 | −1.3320 | 0.7451 | 0.6565 | −0.3121 | −0.2392 |
| | | (−4.72)** | (−1.376) | (2.65)** | (2.303)* | (−1.58) | (−0.960) |
| | 12 | −0.5595 | −0.4578 | 0.7738 | 0.6719 | −0.4558 | −0.3722 |
| | | (−6.62)** | (−2.23)* | (6.71)** | (2.13)* | (−4.14)** | (−1.80) |
| | 20 | −0.4047 | −0.3068 | 0.6553 | 0.3793 | −0.4111 | −0.1570 |
| | | (−3.27)** | (−0.374) | (5.24)** | (2.188)* | (−3.23)** | (−0.991) |

| | $k$ | Trivariate Regression Test (1)H1 | (2)H2 | (3)H3 | (4)H4 | (5)H5 | (6)H6 | (7)H7 | (8)H8 |
|---|---|---|---|---|---|---|---|---|---|
| 1936 − 2001 | 1 | 0.0409 | 0.0509 | 0.0475 | 0.0600 | 0.0781 | 0.0888 | 0.0089 | 0.0157 |
| | | (0.013)* | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.044)* | (0.026)* |
| | 4 | 0.0694 | 0.0645 | 0.0526 | 0.0781 | 0.0379 | 0.0676 | 0.0058 | 0.0166 |
| | | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.015)* | (0.000)** | (0.056) | (0.028)* |
| | 12 | 0.0233 | 0.0466 | 0.0701 | 0.0901 | 0.0306 | 0.0434 | 0.0029 | 0.0043 |
| | | (0.022)* | (0.003)** | (0.000)** | (0.000)** | (0.018)* | (0.008)** | (0.089) | (0.060) |
| | 20 | 0.0635 | 0.1266 | 0.0913 | 0.2269 | 0.1493 | 0.1108 | 0.0258 | 0.0352 |
| | | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.022)* | (0.016)* |
| 1952 − 2001 | 1 | 0.1148 | 0.1079 | 0.1593 | 0.1033 | 0.0941 | 0.0518 | 0.0111 | 0.0026 |
| | | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.035)* | (0.091) |
| | 4 | 0.1270 | 0.1056 | 0.1558 | 0.0711 | 0.1260 | 0.1016 | 0.0195 | 0.0280 |
| | | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.025)* | (0.020)* |
| | 12 | 0.0448 | 0.0504 | 0.0861 | 0.1417 | 0.0676 | 0.1111 | 0.0245 | 0.0299 |
| | | (0.006)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.023)* | (0.019)* |
| | 20 | 0.0901 | 0.0487 | 0.0751 | 0.0639 | 0.0841 | 0.0718 | 0.0047 | 0.0013 |
| | | (0.000)** | (0.002)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.058) | (0.110) |
| 1936 − 1990 | 1 | 0.0325 | 0.0268 | 0.0510 | 0.0427 | 0.0351 | 0.0287 | 0.0071 | 0.0049 |
| | | (0.019)* | (0.021)* | (0.000)** | (0.008)** | (0.018)* | (0.022)* | (0.046)* | (0.058) |
| | 4 | 0.0087 | 0.0398 | 0.0662 | 0.0703 | 0.0333 | 0.0534 | 0.0040 | 0.0145 |
| | | (0.046)* | (0.013)* | (0.000)** | (0.000)** | (0.020)* | (0.000)** | (0.060) | (0.027)* |
| | 12 | 0.0827 | 0.0599 | 0.0827 | 0.0639 | 0.0858 | 0.0535 | 0.0019 | 0.0130 |
| | | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.101) | (0.035)* |
| | 20 | 0.0817 | 0.0251 | 0.2098 | 0.0708 | 0.1070 | 0.0242 | 0.0349 | 0.0103 |
| | | (0.000)** | (0.022)* | (0.000)** | (0.000)** | (0.000)** | (0.023)* | (0.019)* | (0.037)* |
| 1952 − 1990 | 1 | 0.0947 | 0.1137 | 0.0851 | 0.0799 | 0.0271 | 0.0475 | 0.0062 | 0.0144 |
| | | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.021)* | (0.000)** | (0.050)* | (0.028)* |
| | 4 | 0.0892 | 0.0750 | 0.0683 | 0.0579 | 0.0809 | 0.0572 | 0.0233 | 0.0051 |
| | | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.024)* | (0.058) |
| | 12 | 0.1210 | 0.0679 | 0.1849 | 0.1499 | 0.1202 | 0.0942 | 0.0584 | 0.0266 |
| | | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.022)* |
| | 20 | 0.0711 | 0.1101 | 0.2362 | 0.0633 | 0.2116 | 0.0998 | 0.0957 | 0.0116 |
| | | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.000)** | (0.034)* |

# Table 4.3 Predictability of US Excess Returns(Annually)

We estimate regressions of the form $Y_{t+h} = \alpha_h + \beta_h' X_t + \varepsilon_{h,t+h}$ where $Y_{t+h} = (\tau/h)[(y_{t+1} - r_t) + \cdots + (\ y_{t+h} - r_{t+h-1})]$ is the annualized $h$-period excess return for the aggregate stock market, $r_t$ is the risk-free rate from $t$ to $t+1$, and $y_{t+1} - r_t$ is the excess one period return from time $t$ to $t+1$, with instruments $z_t$ being log dividend yields, log earnings yields, dividend payout ratio, short rate, inflation, book-to-market ratio(b/m), investment to capital ratio(i/k), corporate issuing activity (Eqis and Ntis), and consumption, wealth, and income ratio(cay). T-statistics in parentheses are computed using Newey West(1987) and Hodrick (1992) standard errors. Horizons $h$ are anuually. The test column reports a $p$-value for the predictictability test under two hypothesises $H1$ and $H2$. Hypothesis $H1$ is $H_0 : E(Y_{t+h}|X_t) = E(Y_{t+h})$ and Hypothesis $H2$ is $H_0 : E(\varepsilon_{t+h}|X_t) = 0$. * and ** are at the significance leve 5% and 1% respectively. In order to compare the prevailing predictive models with the historical mean model, we calculate a new variable $\Delta(\frac{Q_h}{\sigma^2}) = \hat{Q}_N(h)/\widehat{\sigma}_\varepsilon^2 - \hat{Q}_A(h)/\widehat{\sigma}_\varepsilon^2$, where $\hat{Q}_N(h)/\widehat{\sigma}_\varepsilon^2$ is computed by the historical mean model and $\hat{Q}_A(h)/\widehat{\sigma}_\varepsilon^2$ is computed by the prevailing predictive regression model under two hypothesises $H1$ and $H2$ . We also compute the IS $\overline{R}^2$, OOS $\overline{R}^2$, and $\Delta RMSE$ following the Goyal and Welch (2007)'s definition. The table summarizes the results for both in-sample and out-of-sample tests.

| Variable | Data | $h$ | Univariate Regression (1) NW | Univariate Regression (2) HK | Test(IS) (1) H1 | Test(IS) (2) H2 | Test(OS) (1) H1 | Test(OS) (2) H2 | (1) $\Delta(\frac{Q_h}{\sigma^2})$ | (2) $\Delta(\frac{Q_h}{\sigma^2})$ | GW IS $\overline{R}^2$ | GW OOS $\overline{R}^2$ | GW $\Delta RMSE$ (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| dy | 1872 − 2005 | 1 | 0.0795 | 0.0783 | 0.0674 | 0.0328 | 0.1613 | 0.0753 | 0.0367 | 0.0058 | 0.91 | −1.93 | −0.10 |
| | | | (2.07)* | (1.65) | (0.000)** | (0.019)* | (0.000)** | (0.000)** | | | | | |
| | | 5 | 0.0665 | 0.0634 | 0.0401 | 0.0430 | 0.0963 | 0.0921 | 0.0233 | 0.0183 | 6.04 | −4.45 | −0.68 |
| | | | (1.60) | (1.04) | (0.014)* | (0.011)* | (0.000)** | (0.000)** | | | | | |
| ey | 1872 − 2005 | 1 | 0.0527 | 0.0478 | 0.0861 | 0.0494 | 0.1266 | 0.0238 | 0.0714 | 0.0515 | 1.08 | −1.78 | −0.08 |
| | | | (1.14) | (0.78) | (0.000)** | (0.008)** | (0.000)** | (0.023)* | | | | | |
| | | 5 | 0.0344 | 0.0302 | 0.1055 | 0.0855 | 0.1030 | 0.1081 | 0.0300 | 0.0343 | 6.24 | −1.04 | −0.03 |
| | | | (0.80) | (0.46) | (0.000)** | (0.000)** | (0.000)** | (0.000)** | | | | | |
| d/e | 1872 − 2005 | 1 | 0.0759 | 0.0696 | 0.0180 | 0.0451 | 0.0760 | 0.0988 | 0.0486 | 0.0235 | −0.75 | −4.33 | −0.31 |
| | | | (1.46) | (1.12) | (0.028)* | (0.009)** | (0.000)** | (0.000)** | | | | | |
| | | 5 | 0.0698 | 0.0589 | 0.0298 | 0.0361 | 0.0938 | 0.0955 | 0.0208 | 0.0217 | 0.66 | −4.87 | −0.76 |
| | | | (1.55) | (0.89) | (0.021)* | (0.018)* | (0.000)** | (0.000)** | | | | | |
| r | 1872 − 2005 | 1 | −0.7299 | −0.6545 | 0.0232 | 0.0409 | 0.1057 | 0.0706 | 0.0189 | 0.0011 | 0.34 | −3.37 | −0.14 |
| | | | (−1.95)* | (−0.98) | (0.024)* | (0.010)** | (0.010)** | (0.000)** | | | | | |
| | | 5 | −0.5524 | −0.4982 | 0.0340 | 0.0343 | 0.1145 | 0.1004 | 0.0415 | 0.0266 | 3.83 | −17.66 | −2.78 |
| | | | (−1.02) | (−0.76) | (0.019)* | (0.019)* | (0.010)** | (0.000)** | | | | | |
| infl | 1919 − 2005 | 1 | −0.3766 | −0.3121 | 0.0409 | 0.1088 | 0.0064 | 0.0563 | 0.1182 | 0.0132 | −1.00 | −4.07 | −0.20 |
| | | | (−1.18) | (−0.87) | (0.010)** | (0.010)** | (0.052) | (0.000)** | | | | | |
| | | 5 | −0.4034 | −0.3441 | 0.1145 | 0.1001 | 0.0877 | 0.0876 | 0.0147 | 0.0138 | −1.21 | −11.25 | −1.70 |
| | | | (−1.00) | (−0.49) | (0.000)** | (0.000)** | (0.000)** | (0.000)** | | | | | |
| b/m | 1921 − 2005 | 1 | 0.0641 | 0.0546 | 0.0352 | 0.1014 | 0.0629 | 0.1436 | 0.0617 | 0.0741 | 3.20 | −1.72 | −0.01 |
| | | | (1.56) | (1.12) | (0.018)* | (0.000)** | (0.000)** | (0.000)** | | | | | |
| | | 5 | 0.0182 | 0.0124 | 0.1317 | 0.1267 | 0.1413 | 0.1324 | 0.0683 | 0.0586 | 10.78 | −13.06 | −2. − 03 |
| | | | (0.46) | (0.21) | (0.000)** | (0.000)** | (0.000)** | (0.000)** | | | | | |
| i/k | 1947 − 2005 | 1 | 0.0355 | 0.0312 | 0.1213 | 0.1035 | 0.3574 | 0.4443 | 0.2328 | 0.3748 | 6.63 | −1.77 | 0.07 |
| | | | (0.66) | (0.40) | (0.000)** | (0.000)** | (0.000)** | (0.000)** | | | | | |
| | | 5 | 0.0256 | 0.0212 | 0.0799 | 0.0821 | 0.1156 | 0.0732 | 0.0426 | 0.0006 | 33.99 | 12.99 | 3.39 |
| | | | (0.40) | (0.21) | (0.000)** | (0.000)** | (0.000)** | (0.000)** | | | | | |
| ntis | 1927 − 2005 | 1 | 0.1833 | 0.1645 | 0.0177 | 0.0265 | 0.0034 | 0.0234 | 0.1212 | 0.0461 | 8.15 | −5.07 | −0.26 |
| | | | (0.67) | (0.34) | (0.028)* | (0.021)* | (0.072) | (0.023)* | | | | | |
| | | 5 | 0.1754 | 0.1437 | 0.0729 | 0.0726 | 0.0218 | 0.0060 | 0.0512 | 0.0678 | 6.59 | −3.46 | −0.32 |
| | | | (0.65) | (0.45) | (0.000)** | (0.000)** | (0.025)* | (0.046)* | | | | | |
| eqis | 1927 − 2005 | 1 | −0.1125 | −0.0937 | 0.0464 | 0.0426 | 0.1137 | 0.0670 | 0.0109 | 0.0025 | 9.15 | 2.04 | 0.30 |
| | | | (−2.98)* | (−1.60) | (0.005)** | (0.008)** | (0.000)** | (0.000)** | | | | | |
| | | 5 | −0.0515 | −0.0332 | 0.1598 | 0.1599 | 0.1013 | 0.0813 | 0.0283 | 0.0075 | 9.50 | −2.35 | −0.11 |
| | | | (−1.26) | (−0.90) | (0.000)** | (0.000)** | (0.000)** | (0.000)** | | | | | |
| cay | 1945 − 2005 | 1 | 0.0529 | 0.0445 | 0.0128 | 0.1121 | 0.0066 | 0.1538 | 0.1194 | 0.0843 | 15.72 | 16.78 | 1.61 |
| | | | (4.08)** | (3.68)** | (0.036)* | (0.000)** | (0.043)* | (0.000)** | | | | | |
| | | 5 | 0.0133 | 0.0092 | 0.1085 | 0.1055 | 0.1451 | 0.1647 | 0.0721 | 0.0909 | 36.05 | 30.35 | 7.50 |
| | | | (1.11) | (0.86) | (0.000)** | (0.000)** | (0.000)** | (0.000)** | | | | | |

## Table 5.1a. Equity Premium Out-of-Sample Forecasting Results(Univariate, Quarterly)

Table 5.1 show the out-of-sample results of the univariate linear predictive models. Table 5.1a summarize the MSE, MAE, and RMSE of univariate linear predictive regression for dividend yield during the period 1936-2001, 1952-2001, 1936-1990, and 1952-1990. The benchmark model is historical average equity returns. The alternative models are linear predictive model and two nonparametric predictive models.

| | | | | HMM | LPM | NPM1 | NPM2 | | | | HMM | LPM | NPM1 | NPM2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $1936-2001$ | $dy4$ | 1 | $MSE$ | 0.1389 | 0.1481 | 0.1444 | 0.1262 | $1936-1990$ | 1 | $MSE$ | 0.1465 | 0.1515 | 0.1256 | 0.1253 |
| | | | $MAE$ | 0.2881 | 0.2803 | 0.2889 | 0.2728 | | | $MAE$ | 0.2719 | 0.2780 | 0.2752 | 0.2725 |
| | | | $RMSE$ | 0.3726 | 0.3848 | 0.3799 | 0.3553 | | | $RMSE$ | 0.3829 | 0.3892 | 0.3544 | 0.3539 |
| | | 4 | $MSE$ | 0.0546 | 0.0629 | 0.0576 | 0.0388 | | 4 | $MSE$ | 0.0597 | 0.0649 | 0.0396 | 0.0362 |
| | | | $MAE$ | 0.1733 | 0.1925 | 0.1897 | 0.1568 | | | $MAE$ | 0.1873 | 0.1894 | 0.1592 | 0.1502 |
| | | | $RMSE$ | 0.2336 | 0.2508 | 0.2399 | 0.1970 | | | $RMSE$ | 0.2443 | 0.2547 | 0.1990 | 0.1904 |
| | | 12 | $MSE$ | 0.0297 | 0.0384 | 0.0281 | 0.0136 | | 12 | $MSE$ | 0.0361 | 0.0474 | 0.0142 | 0.0100 |
| | | | $MAE$ | 0.1423 | 0.1642 | 0.1304 | 0.0939 | | | $MAE$ | 0.1665 | 0.1862 | 0.0938 | 0.0805 |
| | | | $RMSE$ | 0.1723 | 0.1961 | 0.1678 | 0.1168 | | | $RMSE$ | 0.1900 | 0.2176 | 0.1191 | 0.0999 |
| | | 20 | $MSE$ | 0.0146 | 0.0380 | 0.0352 | 0.0115 | | 20 | $MSE$ | 0.0346 | 0.0401 | 0.0123 | 0.0093 |
| | | | $MAE$ | 0.1042 | 0.1665 | 0.1605 | 0.0875 | | | $MAE$ | 0.1724 | 0.1687 | 0.0885 | 0.0798 |
| | | | $RMSE$ | 0.1208 | 0.1950 | 0.1876 | 0.1071 | | | $RMSE$ | 0.1861 | 0.2002 | 0.1108 | 0.0965 |
| $1952-2001$ | $dy4$ | 1 | $MSE$ | 0.1260 | 0.1652 | 0.1613 | 0.1225 | $1952-1990$ | 1 | $MSE$ | 0.1520 | 0.2018 | 0.1481 | 0.1464 |
| | | | $MAE$ | 0.2604 | 0.2996 | 0.3024 | 0.2679 | | | $MAE$ | 0.2861 | 0.3357 | 0.3077 | 0.2997 |
| | | | $RMSE$ | 0.3549 | 0.4065 | 0.4017 | 0.3500 | | | $RMSE$ | 0.3898 | 0.4492 | 0.3848 | 0.3826 |
| | | 4 | $MSE$ | 0.0435 | 0.0738 | 0.0623 | 0.0358 | | 4 | $MSE$ | 0.0552 | 0.1054 | 0.0453 | 0.0411 |
| | | | $MAE$ | 0.1589 | 0.2100 | 0.1945 | 0.1490 | | | $MAE$ | 0.1798 | 0.2631 | 0.1664 | 0.1569 |
| | | | $RMSE$ | 0.2085 | 0.2716 | 0.2496 | 0.1893 | | | $RMSE$ | 0.2349 | 0.3246 | 0.2128 | 0.2026 |
| | | 12 | $MSE$ | 0.0193 | 0.0453 | 0.0308 | 0.0127 | | 12 | $MSE$ | 0.0234 | 0.0660 | 0.0185 | 0.0098 |
| | | | $MAE$ | 0.1163 | 0.1783 | 0.1345 | 0.0898 | | | $MAE$ | 0.1248 | 0.2204 | 0.1116 | 0.0811 |
| | | | $RMSE$ | 0.1391 | 0.2129 | 0.1755 | 0.1126 | | | $RMSE$ | 0.1529 | 0.2569 | 0.1362 | 0.0989 |
| | | 20 | $MSE$ | 0.0146 | 0.0322 | 0.0167 | 0.0091 | | 20 | $MSE$ | 0.0163 | 0.0489 | 0.0096 | 0.0058 |
| | | | $MAE$ | 0.1042 | 0.1546 | 0.0989 | 0.0760 | | | $MAE$ | 0.1084 | 0.1994 | 0.0844 | 0.0661 |
| | | | $RMSE$ | 0.1208 | 0.1796 | 0.1291 | 0.0952 | | | $RMSE$ | 0.1278 | 0.2212 | 0.0978 | 0.0763 |

Table 5.2a. Equity Premium Out-of-Sample Forecasting Results(Bivariate, Quarterly)

Table 5.2 show the out-of-sample results of the bivariate linear predictive models using quarterly data. Table 5.2a summarize the MSE, MAE, and RMSE of linear predictive regression for dividend yield and short rate during the period 1936-2001, 1952-2001, 1936-1990, and 1952-1990. The benchmark model is historical average equity returns. The alternative models are linear predictive model and two nonparametric predictive models.

| | | | | HMM | LPM | NPM1 | NPM2 | | | | HMM | LPM | NPM1 | NPM2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $1936-2001$ | $dy4, r$ | 1 | $MSE$ | 0.1389 | 0.1344 | 0.1310 | 0.1226 | $1936-1990$ | 1 | $MSE$ | 0.1692 | 0.1378 | 0.1177 | 0.1210 |
| | | | $MAE$ | 0.2681 | 0.2705 | 0.2762 | 0.2701 | | | $MAE$ | 0.2999 | 0.2664 | 0.2669 | 0.2699 |
| | | | $RMSE$ | 0.3726 | 0.3667 | 0.3619 | 0.3501 | | | $RMSE$ | 0.4114 | 0.3722 | 0.3431 | 0.3479 |
| | | 4 | $MSE$ | 0.0734 | 0.0877 | 0.0699 | 0.0375 | | 4 | $MSE$ | 0.0710 | 0.1839 | 0.0698 | 0.0352 |
| | | | $MAE$ | 0.2048 | 0.2307 | 0.2098 | 0.1553 | | | $MAE$ | 0.2050 | 0.3450 | 0.2047 | 0.1483 |
| | | | $RMSE$ | 0.2710 | 0.2962 | 0.2643 | 0.1936 | | | $RMSE$ | 0.2664 | 0.4288 | 0.2642 | 0.1877 |
| | | 12 | $MSE$ | 0.0540 | 0.0289 | 0.0227 | 0.0130 | | 12 | $MSE$ | 0.0403 | 0.0885 | 0.0222 | 0.0097 |
| | | | $MAE$ | 0.1860 | 0.1397 | 0.1198 | 0.0915 | | | $MAE$ | 0.1758 | 0.2596 | 0.1120 | 0.0793 |
| | | | $RMSE$ | 0.2324 | 0.1700 | 0.1508 | 0.1141 | | | $RMSE$ | 0.2006 | 0.2976 | 0.1491 | 0.0986 |
| | | 20 | $MSE$ | 0.0538 | 0.0145 | 0.0128 | 0.0093 | | 20 | $MSE$ | 0.0384 | 0.0382 | 0.0078 | 0.0057 |
| | | | $MAE$ | 0.1893 | 0.0988 | 0.0915 | 0.0782 | | | $MAE$ | 0.1801 | 0.1747 | 0.0685 | 0.0621 |
| | | | $RMSE$ | 0.2320 | 0.1203 | 0.1130 | 0.0963 | | | $RMSE$ | 0.1959 | 0.1954 | 0.0883 | 0.0758 |
| $1952-2001$ | $dy4, r$ | 1 | $MSE$ | 0.1076 | 0.1445 | 0.1153 | 0.1050 | $1952-1990$ | 1 | $MSE$ | 0.1388 | 0.1270 | 0.1245 | 0.1358 |
| | | | $MAE$ | 0.2460 | 0.3003 | 0.2603 | 0.2508 | | | $MAE$ | 0.2572 | 0.2782 | 0.2713 | 0.2909 |
| | | | $RMSE$ | 0.3280 | 0.3802 | 0.3395 | 0.3240 | | | $RMSE$ | 0.3726 | 0.3563 | 0.3528 | 0.3685 |
| | | 4 | $MSE$ | 0.0355 | 0.0690 | 0.0568 | 0.0332 | | 4 | $MSE$ | 0.0552 | 0.0393 | 0.0328 | 0.0384 |
| | | | $MAE$ | 0.1471 | 0.2143 | 0.1897 | 0.1454 | | | $MAE$ | 0.1798 | 0.1583 | 0.1469 | 0.1519 |
| | | | $RMSE$ | 0.1885 | 0.2627 | 0.2383 | 0.1822 | | | $RMSE$ | 0.2349 | 0.1981 | 0.1811 | 0.1960 |
| | | 12 | $MSE$ | 0.0179 | 0.0300 | 0.0283 | 0.0142 | | 12 | $MSE$ | 0.0234 | 0.0218 | 0.0122 | 0.0093 |
| | | | $MAE$ | 0.1090 | 0.1404 | 0.1347 | 0.0964 | | | $MAE$ | 0.1248 | 0.1166 | 0.0908 | 0.0789 |
| | | | $RMSE$ | 0.1340 | 0.1733 | 0.1681 | 0.1190 | | | $RMSE$ | 0.1529 | 0.1477 | 0.1106 | 0.0963 |
| | | 20 | $MSE$ | 0.0126 | 0.0196 | 0.0171 | 0.0110 | | 20 | $MSE$ | 0.0163 | 0.0141 | 0.0135 | 0.0047 |
| | | | $MAE$ | 0.0930 | 0.1246 | 0.1101 | 0.0858 | | | $MAE$ | 0.1084 | 0.0995 | 0.0963 | 0.0576 |
| | | | $RMSE$ | 0.1124 | 0.1402 | 0.1309 | 0.1047 | | | $RMSE$ | 0.1278 | 0.1189 | 0.1163 | 0.0688 |

Table 5.3. Equity Premium Out-of-Sample Forecasting Results (Annually)

Table 5.3 summarize the out-of-sample results of univariate linear predictive regression for dividend yield and earning yield by using the measure MSE, MAE, and RMSE annually. The predictors are dividend yield ($D/Y$), earnings-price ratio ($E/P$), dividend-payout ratio ($D/E$), book-to-market ratio ($B/M$), net equity expansion ($NTIS$), treasure bill rate ($TBL$), Percent Equity Issuing ($eqis$), Consumption income wealth ratio ($Cay$), inflation ($INFL$), and investment-to-capital ratio ($I/K$) during the period 1972 and 2005. The benchmark model is historical average equity returns. The alternative models are linear predictive model and two nonparametric predictive models.

| | | | | HMM | LPM | NPM1 | NPM2 | | | | | HMM | LPM | NPM1 | NPM2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| dy | 1872 – 2005 | 1 | $MSE$ | 0.0261 | 0.0268 | 0.0261 | 0.0250 | b/m | 1921 – 2005 | 1 | $MSE$ | 0.0263 | 0.0385 | 0.0410 | 0.0243 |
| | | | $MAE$ | 0.1238 | 0.1284 | 0.1271 | 0.1263 | | | | $MAE$ | 0.1230 | 0.1531 | 0.1568 | 0.1227 |
| | | | $RMSE$ | 0.1615 | 0.1640 | 0.1617 | 0.1581 | | | | $RMSE$ | 0.1621 | 0.1962 | 0.2025 | 0.1559 |
| | | 5 | $MSE$ | 0.0249 | 0.0237 | 0.0239 | 0.0236 | | | 5 | $MSE$ | 0.0267 | 0.0267 | 0.0279 | 0.0250 |
| | | | $MAE$ | 0.1221 | 0.1220 | 0.1224 | 0.1148 | | | | $MAE$ | 0.1251 | 0.1291 | 0.1327 | 0.1275 |
| | | | $RMSE$ | 0.1578 | 0.1540 | 0.1547 | 0.1535 | | | | $RMSE$ | 0.1635 | 0.1633 | 0.1672 | 0.1580 |
| ey | 1872 – 2005 | 1 | $MSE$ | 0.0261 | 0.0293 | 0.0285 | 0.0259 | i/k | 1947 – 2005 | 1 | $MSE$ | 0.0257 | 0.0332 | 0.0244 | 0.0227 |
| | | | $MAE$ | 0.1238 | 0.1331 | 0.1355 | 0.1298 | | | | $MAE$ | 0.1213 | 0.1637 | 0.1252 | 0.1226 |
| | | | $RMSE$ | 0.1615 | 0.1711 | 0.1687 | 0.1608 | | | | $RMSE$ | 0.1604 | 0.1822 | 0.1561 | 0.1507 |
| | | 5 | $MSE$ | 0.0249 | 0.0233 | 0.0216 | 0.0207 | | | 5 | $MSE$ | 0.0266 | 0.0285 | 0.0264 | 0.0259 |
| | | | $MAE$ | 0.1221 | 0.1199 | 0.1174 | 0.1101 | | | | $MAE$ | 0.1250 | 0.1476 | 0.1335 | 0.1328 |
| | | | $RMSE$ | 0.1578 | 0.1528 | 0.1469 | 0.1439 | | | | $RMSE$ | 0.1630 | 0.1689 | 0.1625 | 0.1609 |
| d/e | 1872 – 2005 | 1 | $MSE$ | 0.0261 | 0.0264 | 0.0262 | 0.0243 | ntis | 1927 – 2005 | 1 | $MSE$ | 0.0268 | 0.0317 | 0.0261 | 0.0257 |
| | | | $MAE$ | 0.1238 | 0.1255 | 0.1249 | 0.1186 | | | | $MAE$ | 0.1253 | 0.1578 | 0.1338 | 0.1337 |
| | | | $RMSE$ | 0.1615 | 0.1624 | 0.1620 | 0.1558 | | | | $RMSE$ | 0.1637 | 0.1780 | 0.1615 | 0.1604 |
| | | 5 | $MSE$ | 0.0249 | 0.0233 | 0.0216 | 0.0207 | | | 5 | $MSE$ | 0.0262 | 0.0297 | 0.0277 | 0.0277 |
| | | | $MAE$ | 0.1221 | 0.1199 | 0.1174 | 0.1101 | | | | $MAE$ | 0.1239 | 0.1281 | 0.1330 | 0.1331 |
| | | | $RMSE$ | 0.1578 | 0.1528 | 0.1469 | 0.1439 | | | | $RMSE$ | 0.1618 | 0.1722 | 0.1663 | 0.1665 |
| r | 1872 – 2005 | 1 | $MSE$ | 0.0261 | 0.0268 | 0.0261 | 0.0237 | eqis | 1927 – 2005 | 1 | $MSE$ | 0.0253 | 0.0262 | 0.0248 | 0.0248 |
| | | | $MAE$ | 0.1238 | 0.1284 | 0.1267 | 0.1194 | | | | $MAE$ | 0.1205 | 0.1201 | 0.1262 | 0.1263 |
| | | | $RMSE$ | 0.1615 | 0.1636 | 0.1614 | 0.1539 | | | | $RMSE$ | 0.1591 | 0.1620 | 0.1575 | 0.1577 |
| | | 5 | $MSE$ | 0.0249 | 0.0237 | 0.0256 | 0.0236 | | | 5 | $MSE$ | 0.0262 | 0.0330 | 0.0262 | 0.0253 |
| | | | $MAE$ | 0.1221 | 0.1216 | 0.1252 | 0.1182 | | | | $MAE$ | 0.1239 | 0.1377 | 0.1269 | 0.1249 |
| | | | $RMSE$ | 0.1578 | 0.1539 | 0.1602 | 0.1535 | | | | $RMSE$ | 0.1618 | 0.1816 | 0.1618 | 0.1589 |
| infl | 1919 – 2005 | 1 | $MSE$ | 0.0249 | 0.0307 | 0.0251 | 0.0239 | cay | 1945 – 2005 | 1 | $MSE$ | 0.0249 | 0.0276 | 0.0231 | 0.0210 |
| | | | $MAE$ | 0.1181 | 0.1308 | 0.1288 | 0.1247 | | | | $MAE$ | 0.1215 | 0.1296 | 0.1158 | 0.1166 |
| | | | $RMSE$ | 0.1577 | 0.1753 | 0.1584 | 0.1547 | | | | $RMSE$ | 0.1578 | 0.1660 | 0.1521 | 0.1450 |
| | | 5 | $MSE$ | 0.0268 | 0.0368 | 0.0264 | 0.0257 | | | 5 | $MSE$ | 0.0206 | 0.0257 | 0.0247 | 0.0240 |
| | | | $MAE$ | 0.1242 | 0.1479 | 0.1337 | 0.1271 | | | | $MAE$ | 0.1191 | 0.1296 | 0.1265 | 0.1329 |
| | | | $RMSE$ | 0.1638 | 0.1917 | 0.1625 | 0.1603 | | | | $RMSE$ | 0.1436 | 0.1604 | 0.1573 | 0.1549 |

## Table 5.4. Equity Premium Out-of-Sample Forecasting Results for Combining Methods (Annually)

Table 5.4 report the equity premium out-of-sample combined forecasting results using the annual data. The combination forecasts of $Y_{t+1}$ made at time $t$ are weighted averages of the $M$ individual forecasts based on $\widehat{Y}_{c,t+h} = \sum_{i=1}^{M} \omega_{i,t} \widehat{Y}_{i,t+h}$ where $\{\omega_{i,t}\}_{i=1}^{M}$ are the ex ante combining weights formed at time $t$, and $\widehat{Y}_{i,t+h}$ is the out-of-sample forecast of the equity premium based on the individual predictive models. The first three methods use simple averaging schemes: mean, median, and trimmed mean. Their discount mean square prediction error ($DMSPE$) combining method employs the following weights: $w_{i,t} = \phi_{i,t}^{-1} / \sum_{j=1}^{M} \phi_{j,t}^{-1}, \phi_{i,t} = \sum_{s=R}^{t-1} \theta^{t-1-s} (Y_{i,t+h} - \widehat{Y}_{i,t+h})^2$ and $\theta$ is a discount factor. We consider the two values of 1.0 and 0.9 for $\theta$.

| | OS | | | HMM | LPM | NPM1 | NPM2 | | OS | | | HMM | LPM | NPM1 | NPM2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean | 1965 – 2005 | 1 | $MSE$ | 0.0261 | 0.0244 | 0.0251 | 0.0239 | DMSPE | 1965 – 2005 | 1 | $MSE$ | 0.0261 | 0.0246 | 0.0256 | 0.0240 |
| | | | $MAE$ | 0.1238 | 0.1205 | 0.1217 | 0.1206 | $\theta = 1.0$ | | | $MAE$ | 0.1238 | 0.1214 | 0.1239 | 0.1212 |
| | | | $RMSE$ | 0.1615 | 0.1563 | 0.1585 | 0.1545 | | | | $RMSE$ | 0.1615 | 0.1568 | 0.1599 | 0.1549 |
| | | 5 | $MSE$ | 0.0249 | 0.0240 | 0.0219 | 0.0215 | | | 5 | $MSE$ | 0.0249 | 0.0266 | 0.0252 | 0.0214 |
| | | | $MAE$ | 0.1221 | 0.1168 | 0.1114 | 0.1174 | | | | $MAE$ | 0.1221 | 0.1275 | 0.1247 | 0.1164 |
| | | | $RMSE$ | 0.1578 | 0.1548 | 0.1479 | 0.1466 | | | | $RMSE$ | 0.1578 | 0.1632 | 0.1589 | 0.1463 |
| Median | 1965 – 2005 | 1 | $MSE$ | 0.0261 | 0.0233 | 0.0249 | 0.0230 | DMSPE | 1965 – 2005 | 1 | $MSE$ | 0.0261 | 0.0248 | 0.0260 | 0.0240 |
| | | | $MAE$ | 0.1238 | 0.1169 | 0.1190 | 0.1170 | $\theta = 0.9$ | | | $MAE$ | 0.1238 | 0.1220 | 0.1248 | 0.1211 |
| | | | $RMSE$ | 0.1615 | 0.1528 | 0.1578 | 0.1517 | | | | $RMSE$ | 0.1615 | 0.1575 | 0.1613 | 0.1549 |
| | | 5 | $MSE$ | 0.0249 | 0.0256 | 0.0213 | 0.0213 | | | 5 | $MSE$ | 0.0249 | 0.0303 | 0.0288 | 0.0214 |
| | | | $MAE$ | 0.1221 | 0.1195 | 0.1109 | 0.1178 | | | | $MAE$ | 0.1221 | 0.1355 | 0.1332 | 0.1162 |
| | | | $RMSE$ | 0.1578 | 0.1600 | 0.1458 | 0.1460 | | | | $RMSE$ | 0.1578 | 0.1740 | 0.1698 | 0.1462 |
| Trimmed Mean | 1965 – 2005 | 1 | $MSE$ | 0.0261 | 0.0243 | 0.0248 | 0.0233 | | | | | | | | |
| | | | $MAE$ | 0.1238 | 0.1200 | 0.1200 | 0.1187 | | | | | | | | |
| | | | $RMSE$ | 0.1615 | 0.1557 | 0.1575 | 0.1526 | | | | | | | | |
| | | 5 | $MSE$ | 0.0249 | 0.0244 | 0.0217 | 0.0213 | | | | | | | | |
| | | | $MAE$ | 0.1221 | 0.1171 | 0.1117 | 0.1161 | | | | | | | | |
| | | | $RMSE$ | 0.1578 | 0.1562 | 0.1471 | 0.1461 | | | | | | | | |

# Table 5.5. Equity Premium Out-of-Sample Forecasting Results (Quarterly)

Table 5.5 report the equity premium out-of-sample forecasting results using the quarterly data from 1947:1–2007:4. The out-of-sample forecast evaluation periods are 1965:1–2007:4 consistent with Goyal and Welch (2008). Table 5.5 summarize the MSE, MAE, and RMSE of linear predictive regression for dividend-price ratio ($D/P$), dividend yield ($D/Y$), earnings-price ratio ($E/P$), dividend-payout ratio ($D/E$), stock variance ($SVAR$), book-to-market ratio ($B/M$), net equity expansion ($NTIS$), treasure bill rate ($TBL$), long-term yield ($LTY$), long-term return ($LTR$), term spread ($TMS$), default yield spread ($DFY$), default return yield ($DFR$), inflation ($INFL$), and investment-to-capital ratio ($I/K$). The benchmark model is historical average equity returns. The alternative models are linear predictive model and two nonparametric predictive models. $\Delta$ is the difference of the average utility for a mean-variance investor with relative risk aversion parameter $\gamma$ who allocates her portfolio monthly between stocks and risk-free bonds using forecasts of the equity premium based on the alternative models and the historical average.

| | | | HMM | LPM | NPM1 | NPM2 | | | | HMM | LPM | NPM1 | NPM2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1965:1 | −2007:4 | | | | | | | | | | | | |
| D/P | 1 | $MSE$ | 0.0402 | 0.0548 | 0.0466 | 0.0331 | LTY | 1 | $MSE$ | 0.0402 | 0.0475 | 0.0318 | 0.0315 |
| | | $MAE$ | 0.1469 | 0.1720 | 0.1659 | 0.1397 | | | $MAE$ | 0.1469 | 0.1791 | 0.1405 | 0.1366 |
| | | $RMSE$ | 0.2004 | 0.2340 | 0.2159 | 0.1819 | | | $RMSE$ | 0.2004 | 0.2181 | 0.1782 | 0.1776 |
| | | $\Delta$ | | 0.0188 | 0.0084 | 0.1109 | | | $\Delta$ | | −0.0865 | −0.1020 | 0.1069 |
| | 4 | $MSE$ | 0.0235 | 0.0380 | 0.0301 | 0.0157 | | 4 | $MSE$ | 0.0235 | 0.0157 | 0.0119 | 0.0115 |
| | | $MAE$ | 0.1137 | 0.1499 | 0.1376 | 0.0984 | | | $MAE$ | 0.1137 | 0.0985 | 0.0862 | 0.0851 |
| | | $RMSE$ | 0.1534 | 0.1950 | 0.1734 | 0.1255 | | | $RMSE$ | 0.1534 | 0.1254 | 0.1090 | 0.1075 |
| | | $\Delta$ | | 0.0132 | 0.0335 | 0.1241 | | | $\Delta$ | | 0.1242 | −0.0806 | 0.0774 |
| | 12 | $MSE$ | 0.0180 | 0.0318 | 0.0230 | 0.0093 | | 12 | $MSE$ | 0.0180 | 0.0094 | 0.0072 | 0.0070 |
| | | $MAE$ | 0.1099 | 0.1443 | 0.1192 | 0.0753 | | | $MAE$ | 0.1099 | 0.0756 | 0.0706 | 0.0707 |
| | | $RMSE$ | 0.1343 | 0.1783 | 0.1516 | 0.0967 | | | $RMSE$ | 0.1343 | 0.0969 | 0.0847 | 0.0837 |
| | | $\Delta$ | | 0.0142 | 0.0081 | 0.1156 | | | $\Delta$ | | −0.0919 | −0.0857 | 0.1146 |
| D/Y | 1 | $MSE$ | 0.0402 | 0.0546 | 0.0474 | 0.0331 | LTR | 1 | $MSE$ | 0.0402 | 0.0601 | 0.0320 | 0.0310 |
| | | $MAE$ | 0.1469 | 0.1708 | 0.1665 | 0.1397 | | | $MAE$ | 0.1469 | 0.1906 | 0.1375 | 0.1353 |
| | | $RMSE$ | 0.2004 | 0.2337 | 0.2176 | 0.1820 | | | $RMSE$ | 0.2004 | 0.2451 | 0.1789 | 0.1761 |
| | | $\Delta$ | | 0.0123 | 0.0030 | 0.1094 | | | $\Delta$ | | 0.1053 | 0.0936 | 0.1095 |
| | 4 | $MSE$ | 0.0235 | 0.0378 | 0.0298 | 0.0157 | | 4 | $MSE$ | 0.0235 | 0.0488 | 0.0171 | 0.0161 |
| | | $MAE$ | 0.1137 | 0.1491 | 0.1372 | 0.0985 | | | $MAE$ | 0.1137 | 0.1856 | 0.1005 | 0.0989 |
| | | $RMSE$ | 0.1534 | 0.1944 | 0.1728 | 0.1255 | | | $RMSE$ | 0.1534 | 0.2210 | 0.1308 | 0.1269 |
| | | $\Delta$ | | 0.0127 | 0.0333 | 0.1248 | | | $\Delta$ | | 0.1041 | 0.1121 | 0.1229 |
| | 12 | $MSE$ | 0.0180 | 0.0315 | 0.0221 | 0.0093 | | 12 | $MSE$ | 0.0180 | 0.0353 | 0.0097 | 0.0095 |
| | | $MAE$ | 0.1099 | 0.1423 | 0.1174 | 0.0753 | | | $MAE$ | 0.1099 | 0.1632 | 0.0764 | 0.0758 |
| | | $RMSE$ | 0.1343 | 0.1775 | 0.1487 | 0.0965 | | | $RMSE$ | 0.1343 | 0.1879 | 0.0983 | 0.0976 |
| | | $\Delta$ | | 0.0123 | 0.0052 | 0.1159 | | | $\Delta$ | | 0.0847 | 0.0831 | 0.1146 |
| E/P | 1 | $MSE$ | 0.0402 | 0.0594 | 0.0432 | 0.0337 | TMS | 1 | $MSE$ | 0.0402 | 0.0879 | 0.0390 | 0.0336 |
| | | $MAE$ | 0.1469 | 0.1789 | 0.1591 | 0.1406 | | | $MAE$ | 0.1469 | 0.2434 | 0.1544 | 0.1398 |
| | | $RMSE$ | 0.2004 | 0.2438 | 0.2079 | 0.1835 | | | $RMSE$ | 0.2004 | 0.2965 | 0.1974 | 0.1833 |
| | | $\Delta$ | | 0.0595 | 0.0482 | 0.1107 | | | $\Delta$ | | 0.1179 | 0.1316 | 0.1104 |
| | 4 | $MSE$ | 0.0235 | 0.0428 | 0.0269 | 0.0159 | | 4 | $MSE$ | 0.0235 | 0.0515 | 0.0159 | 0.0135 |
| | | $MAE$ | 0.1137 | 0.1544 | 0.1295 | 0.0994 | | | $MAE$ | 0.1137 | 0.1984 | 0.0985 | 0.0933 |
| | | $RMSE$ | 0.1534 | 0.2069 | 0.1641 | 0.1262 | | | $RMSE$ | 0.1534 | 0.2269 | 0.1263 | 0.1162 |
| | | $\Delta$ | | 0.0530 | 0.0669 | 0.1247 | | | $\Delta$ | | 0.1149 | 0.1252 | 0.1262 |
| | 12 | $MSE$ | 0.0180 | 0.0361 | 0.0206 | 0.0094 | | 12 | $MSE$ | 0.0180 | 0.0389 | 0.0095 | 0.0086 |
| | | $MAE$ | 0.1099 | 0.1482 | 0.1108 | 0.0755 | | | $MAE$ | 0.1099 | 0.1785 | 0.0754 | 0.0748 |
| | | $RMSE$ | 0.1343 | 0.1899 | 0.1435 | 0.0969 | | | $RMSE$ | 0.1343 | 0.1973 | 0.0974 | 0.0925 |
| | | $\Delta$ | | 0.0451 | 0.0254 | 0.1144 | | | $\Delta$ | | 0.0979 | 0.0793 | 0.1142 |
| D/E | 1 | $MSE$ | 0.0402 | 0.0632 | 0.0340 | 0.0335 | DFY | 1 | $MSE$ | 0.0402 | 0.0741 | 0.0364 | 0.0339 |
| | | $MAE$ | 0.1469 | 0.1947 | 0.1418 | 0.1409 | | | $MAE$ | 0.1469 | 0.2130 | 0.1464 | 0.1414 |
| | | $RMSE$ | 0.2004 | 0.2515 | 0.1844 | 0.1831 | | | $RMSE$ | 0.2004 | 0.2723 | 0.1908 | 0.1840 |
| | | $\Delta$ | | 0.1052 | 0.0946 | 0.1116 | | | $\Delta$ | | 0.1278 | 0.1150 | 0.1065 |
| | 4 | $MSE$ | 0.0235 | 0.0526 | 0.0179 | 0.0159 | | 4 | $MSE$ | 0.0235 | 0.0581 | 0.0189 | 0.0159 |
| | | $MAE$ | 0.1137 | 0.1923 | 0.1040 | 0.0987 | | | $MAE$ | 0.1137 | 0.2015 | 0.1078 | 0.0993 |
| | | $RMSE$ | | 0.1151 | 0.1173 | 0.1250 | | | $RMSE$ | 0.1534 | 0.2410 | 0.1375 | 0.1262 |
| | | $\Delta$ | 0.0944 | 0.2095 | 0.2117 | 0.2193 | | | $\Delta$ | | 0.1269 | 0.1370 | 0.1244 |
| | 12 | $MSE$ | 0.0180 | 0.0434 | 0.0114 | 0.0095 | | 12 | $MSE$ | 0.0180 | 0.0417 | 0.0106 | 0.0094 |
| | | $MAE$ | 0.1099 | 0.1819 | 0.0815 | 0.0756 | | | $MAE$ | 0.1099 | 0.1774 | 0.0814 | 0.0756 |
| | | $RMSE$ | 0.1343 | 0.2083 | 0.1069 | 0.0972 | | | $RMSE$ | 0.1343 | 0.2042 | 0.1030 | 0.0968 |
| | | $\Delta$ | | 0.1013 | 0.1040 | 0.1160 | | | $\Delta$ | | 0.0993 | 0.0968 | 0.1144 |

| | | | HMM | LPM | NPM1 | NPM2 | | | | HMM | LPM | NPM1 | NPM2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1965:1 | −2007:4 | | | | | | | | | | | | |
| SVAR | 1 | $MSE$ | 0.0402 | 0.0672 | 0.0338 | 0.0337 | DFR | 1 | $MSE$ | 0.0402 | 0.0591 | 0.0340 | 0.0338 |
| | | $MAE$ | 0.1469 | 0.1984 | 0.1438 | 0.1417 | | | $MAE$ | 0.1469 | 0.1870 | 0.1419 | 0.1419 |
| | | $RMSE$ | 0.2004 | 0.2592 | 0.1840 | 0.1837 | | | $RMSE$ | 0.2004 | 0.2431 | 0.1845 | 0.1840 |
| | | $\Delta$ | | 0.1152 | 0.1066 | 0.1090 | | | $\Delta$ | | 0.0943 | 0.0834 | 0.1098 |
| | 4 | $MSE$ | 0.0235 | 0.0482 | 0.0163 | 0.0161 | | 4 | $MSE$ | 0.0235 | 0.0425 | 0.0161 | 0.0161 |
| | | $MAE$ | 0.1137 | 0.1820 | 0.1010 | 0.0994 | | | $MAE$ | 0.1137 | 0.1699 | 0.0994 | 0.0994 |
| | | $RMSE$ | 0.1534 | 0.2194 | 0.1277 | 0.1267 | | | $RMSE$ | 0.1534 | 0.2063 | 0.1269 | 0.1269 |
| | | $\Delta$ | | 0.1080 | 0.1117 | 0.1246 | | | $\Delta$ | | 0.0936 | 0.0963 | 0.1233 |
| | 12 | $MSE$ | 0.0180 | 0.0388 | 0.0095 | 0.0095 | | 12 | $MSE$ | 0.0180 | 0.0349 | 0.0096 | 0.0095 |
| | | $MAE$ | 0.1099 | 0.1723 | 0.0767 | 0.0762 | | | $MAE$ | 0.1099 | 0.1619 | 0.0767 | 0.0769 |
| | | $RMSE$ | 0.1343 | 0.1969 | 0.0976 | 0.0976 | | | $RMSE$ | 0.1343 | 0.1869 | 0.0978 | 0.0977 |
| | | $\Delta$ | | 0.0942 | 0.0935 | 0.1135 | | | $\Delta$ | | 0.0831 | 0.0804 | 0.1143 |
| B/M | 1 | $MSE$ | 0.0402 | 0.0845 | 0.0815 | 0.0331 | INFL | 1 | $MSE$ | 0.0402 | 0.0572 | 0.0338 | 0.0329 |
| | | $MAE$ | 0.1469 | 0.2254 | 0.2258 | 0.1397 | | | $MAE$ | 0.1469 | 0.1831 | 0.1415 | 0.1355 |
| | | $RMSE$ | 0.2004 | 0.2907 | 0.2855 | 0.1820 | | | $RMSE$ | 0.2004 | 0.2392 | 0.1840 | 0.1814 |
| | | $\Delta$ | | −0.0388 | −0.0472 | 0.1111 | | | $\Delta$ | | 0.0911 | 0.0810 | 0.1098 |
| | 4 | $MSE$ | 0.0235 | 0.0718 | 0.0686 | 0.0157 | | 4 | $MSE$ | 0.0235 | 0.0411 | 0.0161 | 0.0156 |
| | | $MAE$ | 0.1137 | 0.2243 | 0.2147 | 0.0984 | | | $MAE$ | 0.1137 | 0.1667 | 0.0992 | 0.0978 |
| | | $RMSE$ | 0.1534 | 0.2679 | 0.2619 | 0.1254 | | | $RMSE$ | 0.1534 | 0.2026 | 0.1271 | 0.1251 |
| | | $\Delta$ | | −0.0516 | −0.0177 | 0.1235 | | | $\Delta$ | | 0.0905 | 0.0937 | 0.1235 |
| | 12 | $MSE$ | 0.0180 | 0.0619 | 0.0572 | 0.0093 | | 12 | $MSE$ | 0.0180 | 0.0361 | 0.0100 | 0.0095 |
| | | $MAE$ | 0.1099 | 0.2134 | 0.1984 | 0.0751 | | | $MAE$ | 0.1099 | 0.1643 | 0.0769 | 0.0754 |
| | | $RMSE$ | 0.1343 | 0.2488 | 0.2392 | 0.0964 | | | $RMSE$ | 0.1343 | 0.1901 | 0.0999 | 0.0975 |
| | | $\Delta$ | | −0.0391 | −0.0499 | 0.1162 | | | $\Delta$ | | 0.0855 | 0.0836 | 0.1147 |
| NTIS | 1 | $MSE$ | 0.0402 | 0.0580 | 0.0336 | 0.0334 | I/K | 1 | $MSE$ | 0.0402 | 0.0457 | 0.0333 | 0.0315 |
| | | $MAE$ | 0.1469 | 0.1835 | 0.1412 | 0.1408 | | | $MAE$ | 0.1469 | 0.1603 | 0.1402 | 0.1338 |
| | | $RMSE$ | 0.2004 | 0.2408 | 0.1832 | 0.1827 | | | $RMSE$ | 0.2004 | 0.2138 | 0.1825 | 0.1774 |
| | | $\Delta$ | | 0.0902 | 0.0840 | 0.1139 | | | $\Delta$ | | 0.0516 | 0.0441 | 0.1101 |
| | 4 | $MSE$ | 0.0235 | 0.0392 | 0.0164 | 0.0159 | | 4 | $MSE$ | 0.0235 | 0.0358 | 0.0158 | 0.0149 |
| | | $MAE$ | 0.1137 | 0.1584 | 0.0998 | 0.0990 | | | $MAE$ | 0.1137 | 0.1515 | 0.0983 | 0.0961 |
| | | $RMSE$ | 0.1534 | 0.1979 | 0.1280 | 0.1259 | | | $RMSE$ | 0.1534 | 0.1891 | 0.1255 | 0.1219 |
| | | $\Delta$ | | 0.0809 | 0.0853 | 0.1239 | | | $\Delta$ | | 0.0742 | 0.0802 | 0.1255 |
| | 12 | $MSE$ | 0.0180 | 0.0291 | 0.0108 | 0.0094 | | 12 | $MSE$ | 0.0180 | 0.0313 | 0.0094 | 0.0090 |
| | | $MAE$ | 0.1099 | 0.1399 | 0.0833 | 0.0766 | | | $MAE$ | 0.1099 | 0.1515 | 0.0756 | 0.0749 |
| | | $RMSE$ | 0.1343 | 0.1707 | 0.1040 | 0.0971 | | | $RMSE$ | 0.1343 | 0.1768 | 0.0969 | 0.0949 |
| | | $\Delta$ | | 0.0567 | 0.0527 | 0.1153 | | | $\Delta$ | | 0.0725 | 0.0703 | 0.1145 |
| TBL | 1 | $MSE$ | 0.0402 | 0.0341 | 0.0318 | 00268 | | | | | | | |
| | | $MAE$ | 0.1469 | 0.1457 | 0.1362 | 0.1244 | | | | | | | |
| | | $RMSE$ | 0.2004 | 0.1847 | 0.1783 | 0.1636 | | | | | | | |
| | | $\Delta$ | | −0.0198 | −0.0305 | 0.1073 | | | | | | | |
| | 4 | $MSE$ | 0.0235 | 0.0155 | 0.0159 | 0.0103 | | | | | | | |
| | | $MAE$ | 0.1137 | 0.0914 | 0.0984 | 0.0801 | | | | | | | |
| | | $RMSE$ | 0.1534 | 0.1245 | 0.1260 | 0.1015 | | | | | | | |
| | | $\Delta$ | | 0.1250 | 0.0030 | 0.0001 | | | | | | | |
| | 12 | $MSE$ | 0.0180 | 0.0127 | 0.0095 | 0.0059 | | | | | | | |
| | | $MAE$ | 0.1099 | 0.0885 | 0.0749 | 0.0608 | | | | | | | |
| | | $RMSE$ | 0.1343 | 0.1129 | 0.0973 | 0.0767 | | | | | | | |
| | | $\Delta$ | | 0.0043 | 0.0056 | 0.1149 | | | | | | | |

Table 5.6. Equity Premium Out-of-Sample Forecasting Results for Combining Methods (Quarterly)

Table 5.6 report the equity premium out-of-sample combined forecasting results using the quarterly data. The combination forecasts of $Y_{t+1}$ made at time $t$ are weighted averages of the $M$ individual forecasts based on $\widehat{Y}_{c,t+h} = \Sigma_{i=1}^{M} \omega_{i,t} \widehat{Y}_{i,t+h}$ where $\{\omega_{i,t}\}_{i=1}^{M}$ are the ex ante combining weights formed at time $t$, and $\widehat{Y}_{i,t+h}$ is the out-of-sample forecast of the equity premium based on the individual predictive models. The first three methods use simple averaging schemes: mean, median, and trimmed mean. Their discount mean square prediction error ($DMSPE$) combining method employs the following weights: $w_{i,t} = \phi_{i,t}^{-1} / \Sigma_{j=1}^{M} \phi_{j,t}^{-1}, \phi_{i,t} = \Sigma_{s=R}^{t-1} \theta^{t-1-s} (Y_{i,t+h} - \widehat{Y}_{i,t+h})^2$ and $\theta$ is a discount factor. We consider the two values of 1.0 and 0.9 for $\theta$. $\Delta$ is the difference of the average utility for a mean-variance investor with relative risk aversion parameter $\gamma$ who allocates her portfolio monthly between stocks and risk-free bonds using forecasts of the equity premium based on the alternative models and the historical average.

| | OS | | | HMM | LPM | NPM1 | NPM2 | | OS | | | HMM | LPM | NPM1 | NPM2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean | 1965:1 −2007:4 | 1 | $MSE$ | 0.0402 | 0.0290 | 0.0312 | 0.0233 | DMSPE $\theta = 1.0$ | 1965:1 −2007:4 | 1 | $MSE$ | 0.0402 | 0.0260 | 0.0303 | 0.0223 |
| | | | $MAE$ | 0.1469 | 0.1411 | 0.1462 | 0.1175 | | | | $MAE$ | 0.1469 | 0.1315 | 0.1442 | 0.1174 |
| | | | $RMSE$ | 0.2004 | 0.1703 | 0.1766 | 0.1526 | | | | $RMSE$ | 0.2004 | 0.1611 | 0.1740 | 0.1494 |
| | | | $\Delta$ | | 0.1664 | 0.1610 | 0.2298 | | | | $\Delta$ | | 0.2764 | 0.1331 | 0.2296 |
| | | 4 | $MSE$ | 0.0235 | 0.0250 | 0.0273 | 0.0228 | | | 4 | $MSE$ | 0.0235 | 0.0240 | 0.0249 | 0.0214 |
| | | | $MAE$ | 0.1137 | 0.1283 | 0.1354 | 0.1190 | | | | $MAE$ | 0.1137 | 0.1248 | 0.1280 | 0.1137 |
| | | | $RMSE$ | 0.1534 | 0.1580 | 0.1652 | 0.1511 | | | | $RMSE$ | 0.1534 | 0.1548 | 0.1578 | 0.1464 |
| | | | $\Delta$ | | 0.2022 | 0.1892 | 0.2223 | | | | $\Delta$ | | 0.1989 | 0.1720 | 0.1969 |
| | | 12 | $MSE$ | 0.0180 | 0.0126 | 0.0139 | 0.0036 | | | 12 | $MSE$ | 0.0180 | 0.0079 | 0.0086 | 0.0036 |
| | | | $MAE$ | 0.1099 | 0.1055 | 0.1115 | 0.0522 | | | | $MAE$ | 0.1099 | 0.0808 | 0.0840 | 0.0504 |
| | | | $RMSE$ | 0.1343 | 0.1122 | 0.1181 | 0.0601 | | | | $RMSE$ | 0.1343 | 0.0888 | 0.0925 | 0.0596 |
| | | | $\Delta$ | | 0.1437 | 0.1391 | 0.2066 | | | | $\Delta$ | | 0.1142 | 0.1023 | 0.2065 |
| Median | 1965:1 −2007:4 | 1 | $MSE$ | 0.0402 | 0.0254 | 0.0275 | 0.0224 | DMSPE $\theta = 0.9$ | 1965:1 −2007:4 | 1 | $MSE$ | 0.0402 | 0.0248 | 0.0260 | 0.0221 |
| | | | $MAE$ | 0.1469 | 0.1305 | 0.1356 | 0.1178 | | | | $MAE$ | 0.1469 | 0.1280 | 0.1306 | 0.1165 |
| | | | $RMSE$ | 0.2004 | 0.1594 | 0.1658 | 0.1497 | | | | $RMSE$ | 0.2004 | 0.1575 | 0.1611 | 0.1487 |
| | | | $\Delta$ | | 0.2002 | 0.1930 | 0.2294 | | | | $\Delta$ | | 0.1551 | 0.1531 | 0.2260 |
| | | 4 | $MSE$ | 0.0235 | 0.0233 | 0.0245 | 0.0208 | | | 4 | $MSE$ | 0.0235 | 0.0229 | 0.0239 | 0.0212 |
| | | | $MAE$ | 0.1137 | 0.1225 | 0.1266 | 0.1115 | | | | $MAE$ | 0.1137 | 0.1208 | 0.1246 | 0.1126 |
| | | | $RMSE$ | 0.1534 | 0.1527 | 0.1564 | 0.1444 | | | | $RMSE$ | 0.1534 | 0.1513 | 0.1544 | 0.1455 |
| | | | $\Delta$ | | 0.2135 | 0.2072 | 0.2413 | | | | $\Delta$ | | 0.1681 | 0.1262 | 0.1463 |
| | | 12 | $MSE$ | 0.0180 | 0.0080 | 0.0095 | 0.0034 | | | 12 | $MSE$ | 0.0180 | 0.0063 | 0.0074 | 0.0036 |
| | | | $MAE$ | 0.1099 | 0.0819 | 0.0910 | 0.0506 | | | | $MAE$ | 0.1099 | 0.0715 | 0.0778 | 0.0519 |
| | | | $RMSE$ | 0.1343 | 0.0893 | 0.0977 | 0.0586 | | | | $RMSE$ | 0.1343 | 0.0795 | 0.0860 | 0.0598 |
| | | | $\Delta$ | | 0.1684 | 0.1606 | 0.2080 | | | | $\Delta$ | | 0.1006 | 0.0988 | 0.2065 |
| Trimmed Mean | 1965:1 −2007:4 | 1 | $MSE$ | 0.0402 | 0.0282 | 0.0303 | 0.0224 | | | | | | | | |
| | | | $MAE$ | 0.1469 | 0.1393 | 0.1442 | 0.1177 | | | | | | | | |
| | | | $RMSE$ | 0.2004 | 0.1680 | 0.1740 | 0.1496 | | | | | | | | |
| | | | $\Delta$ | | 0.1741 | 0.1681 | 0.2293 | | | | | | | | |
| | | 4 | $MSE$ | 0.0235 | 0.0245 | 0.0259 | 0.0218 | | | | | | | | |
| | | | $MAE$ | 0.1137 | 0.1269 | 0.1316 | 0.1153 | | | | | | | | |
| | | | $RMSE$ | 0.1534 | 0.1565 | 0.1611 | 0.1477 | | | | | | | | |
| | | | $\Delta$ | | 0.2056 | 0.1966 | 0.2303 | | | | | | | | |
| | | 12 | $MSE$ | 0.0180 | 0.0111 | 0.0123 | 0.0036 | | | | | | | | |
| | | | $MAE$ | 0.1099 | 0.0980 | 0.1040 | 0.0517 | | | | | | | | |
| | | | $RMSE$ | 0.1343 | 0.1052 | 0.1108 | 0.0596 | | | | | | | | |
| | | | $\Delta$ | | 0.1512 | 0.1466 | 0.2071 | | | | | | | | |

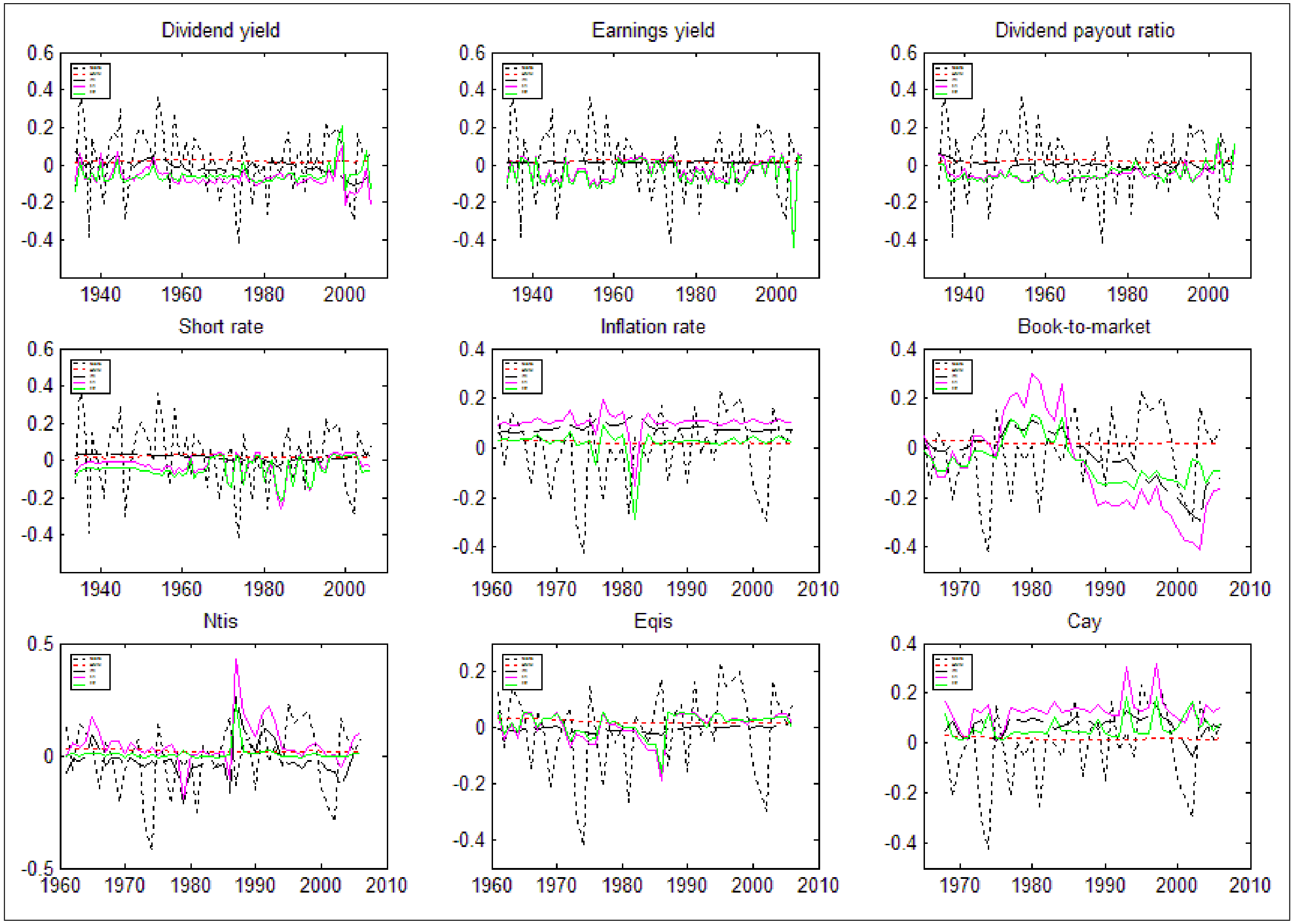


Figure 5.1 Equity Premium Out-of-Sample Forecasting Results for Individual Methods (Annually)

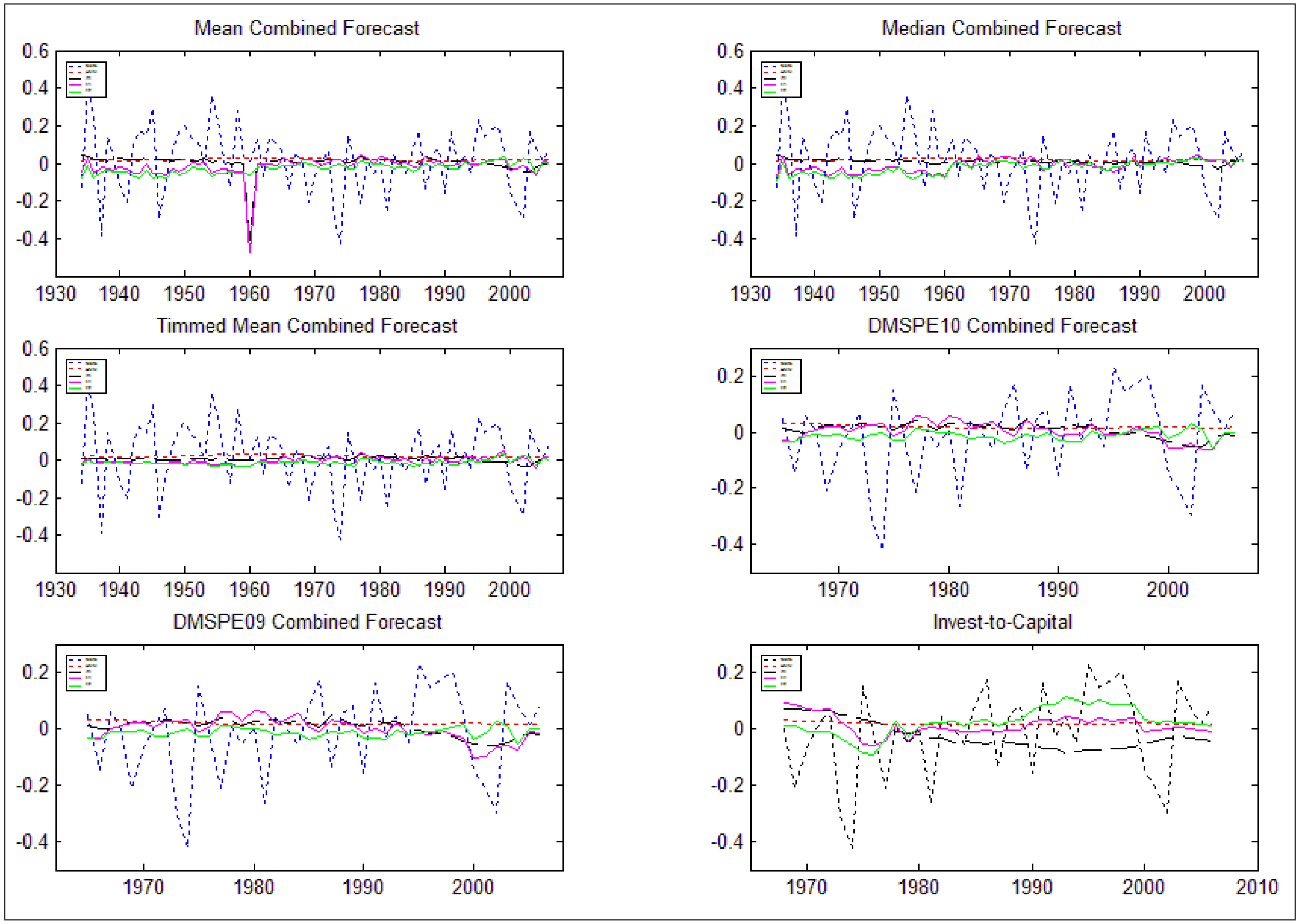


Figure 5.2 Equity Premium Out-of-Sample Forecasting Results for Combined Methods (Annually)

Figure 5.2 illustrate the out-of-sample combined performance for annual predictive regressions for combined methods using annual data.

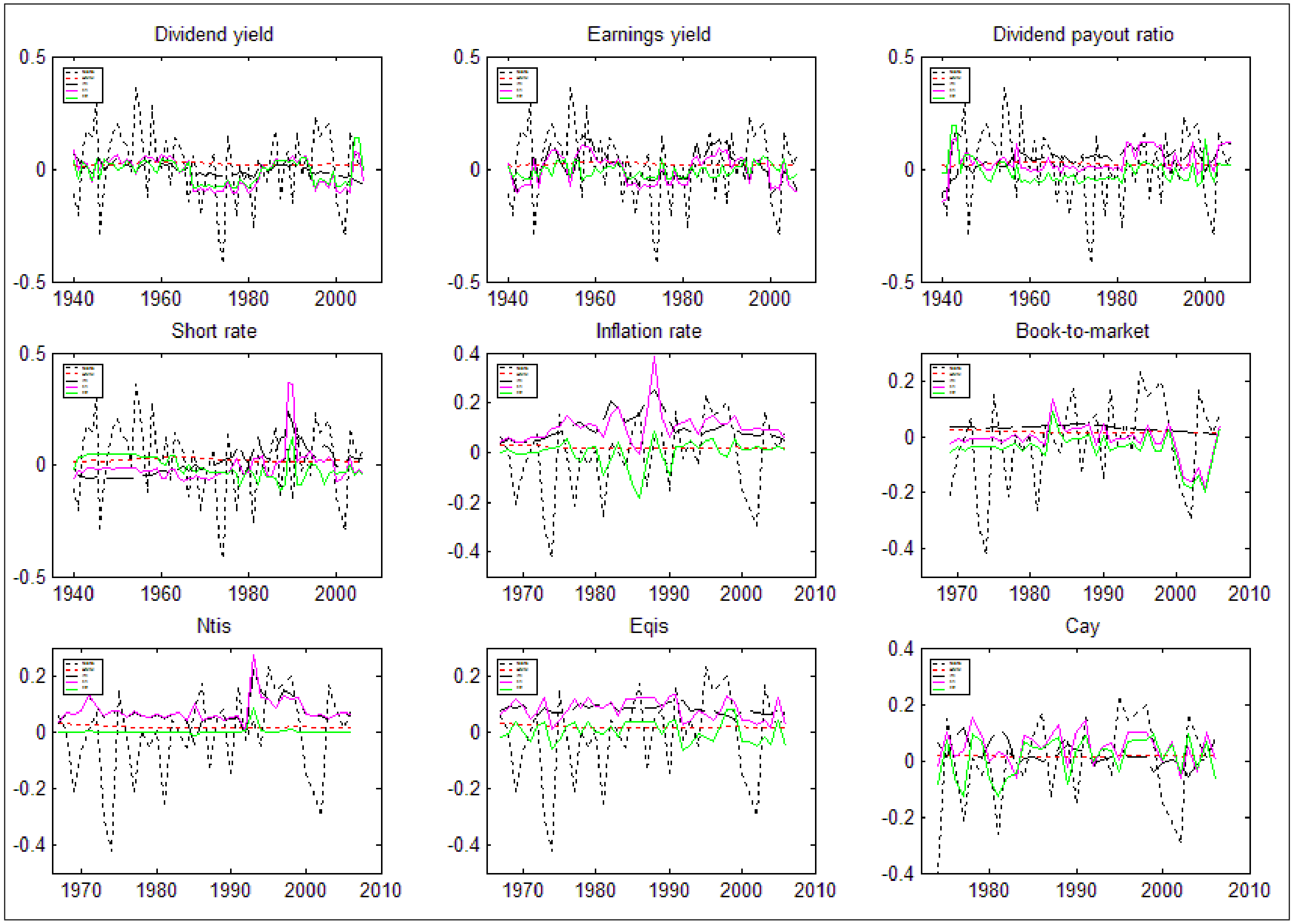


Figure 5.3 Equity Premium Out-of-Sample Forecasting Results for Individual Methods (Annually, 5-year ahead)

Figure 5.3 illustrate the out-of-sample performance for annual predictive regressions for individual methods using annual data over 5-year rolling window.

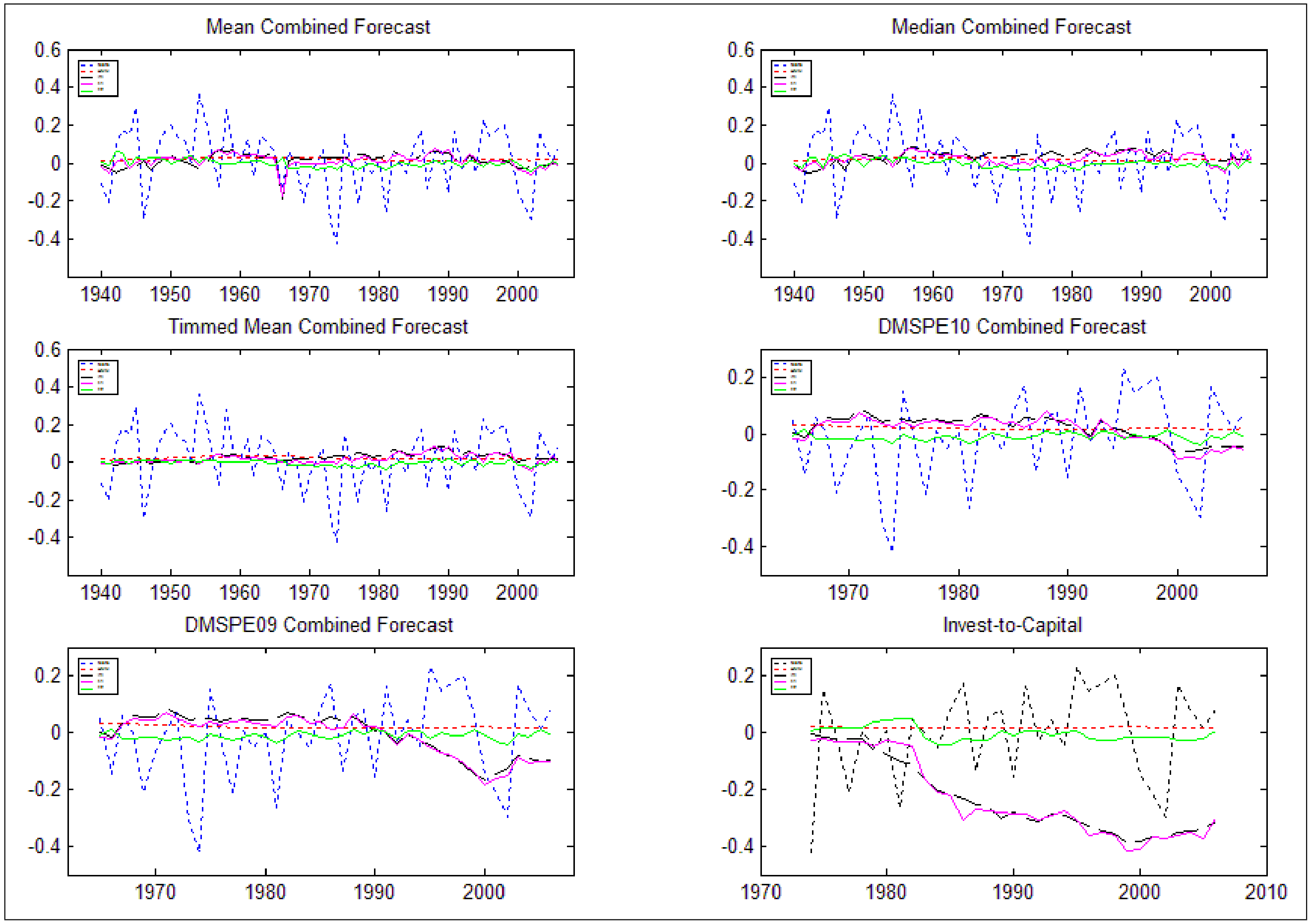


Figure 5.4 Equity Premium Out-of-Sample Forecasting Results for Combined Methods (Annually, 5-year Ahead)

Figure 5.4 illustrate the out-of-sample performance for annual predictive regressions for combined methods using annual data over 5-year rolling window.

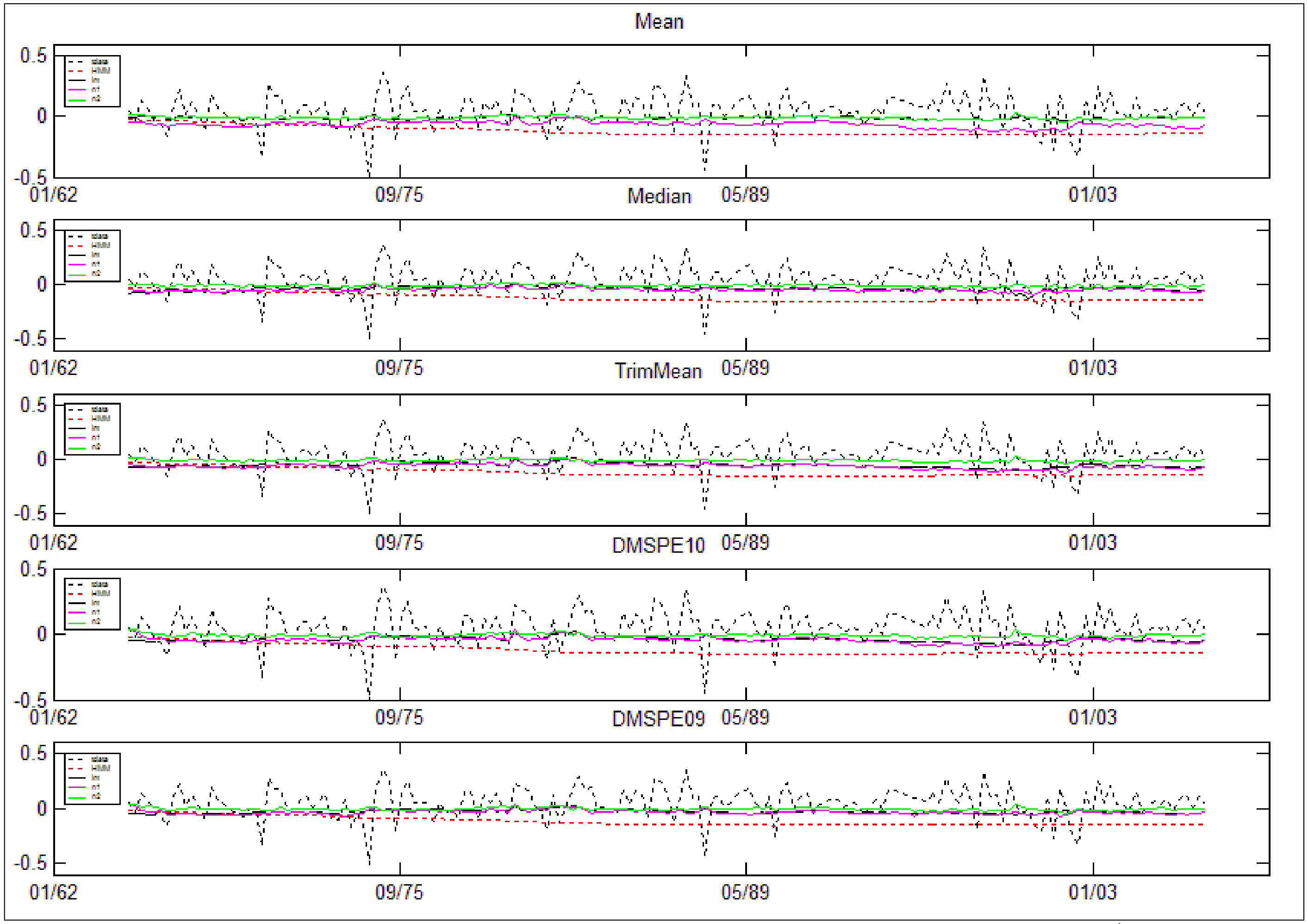


Figure 5.6 Equity Premium Out-of-Sample Forecasting Results for Combined Methods (Quarterly, 1-Period ahead)

Figure 5.6 illustrate the out-of-sample performance for quarterly predictive regressions for combined methods over 1-quarter rolling window.

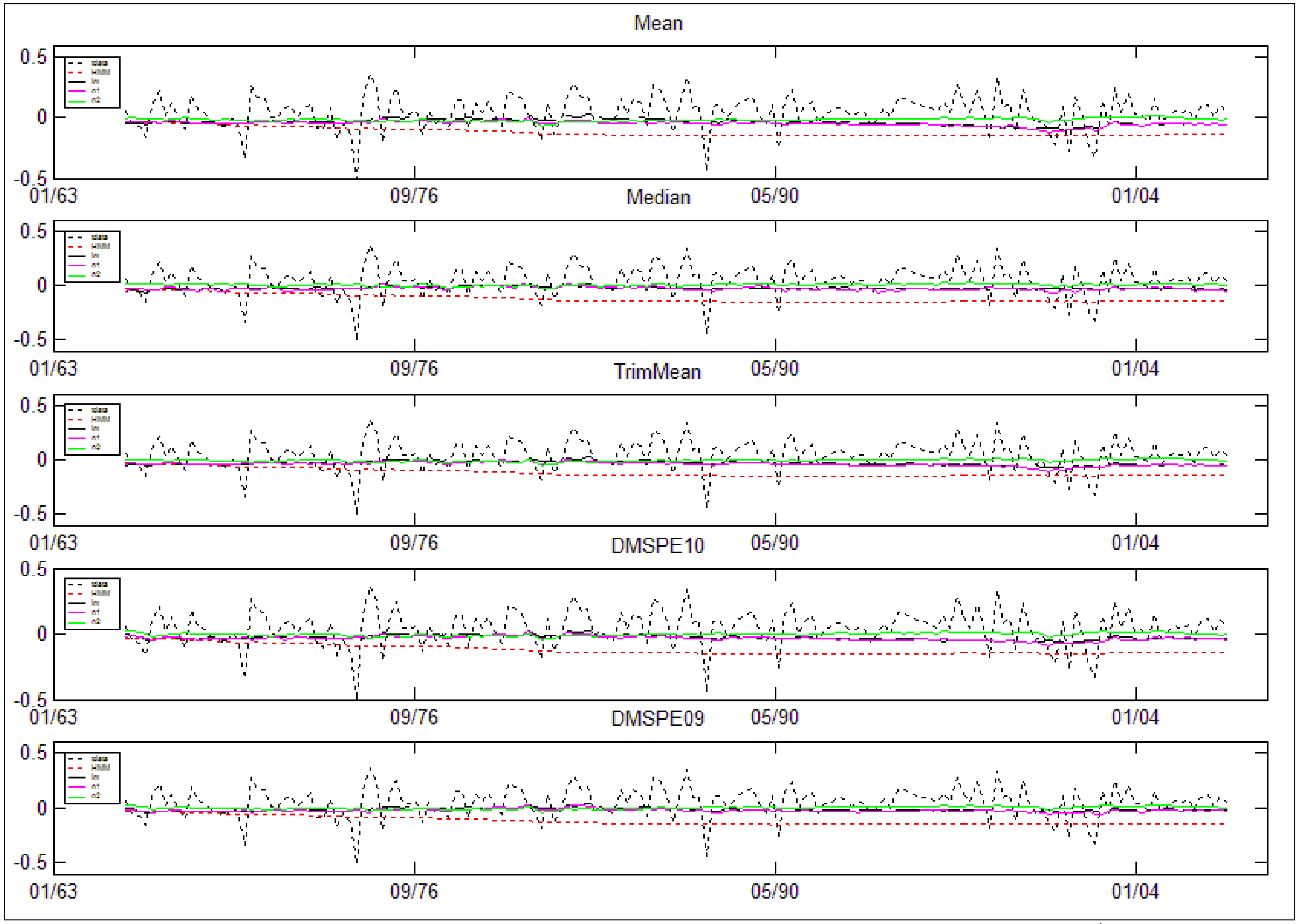


Figure 5.8 Equity Premium Out-of-Sample Forecasting Results for Combined Methods(Quarterly,4-Periodahead)
Figure 5.8 illustrate the out-of-sample performance for quarterly predictive regressions for combined methods over 4-quarter rolling window.

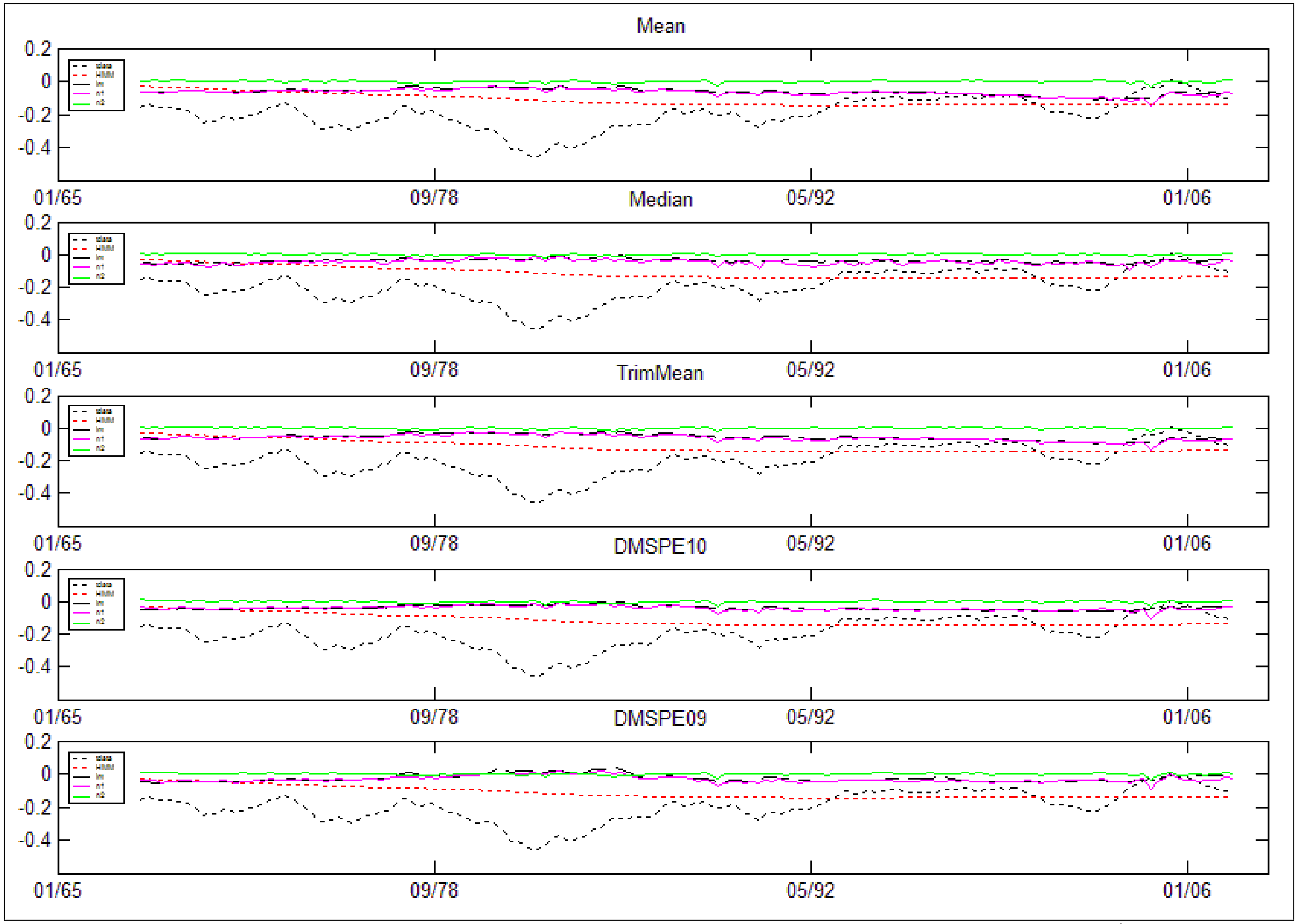


Figure 5.10 Equity Premium Out-of-Sample Forecasting Results for Combined Methods (Quarterly, 12-Period ahead)
Figure 5.10 illustrate the out-of-sample performance for quarterly predictive regressions for combined methods over 12-quarter rolling window.